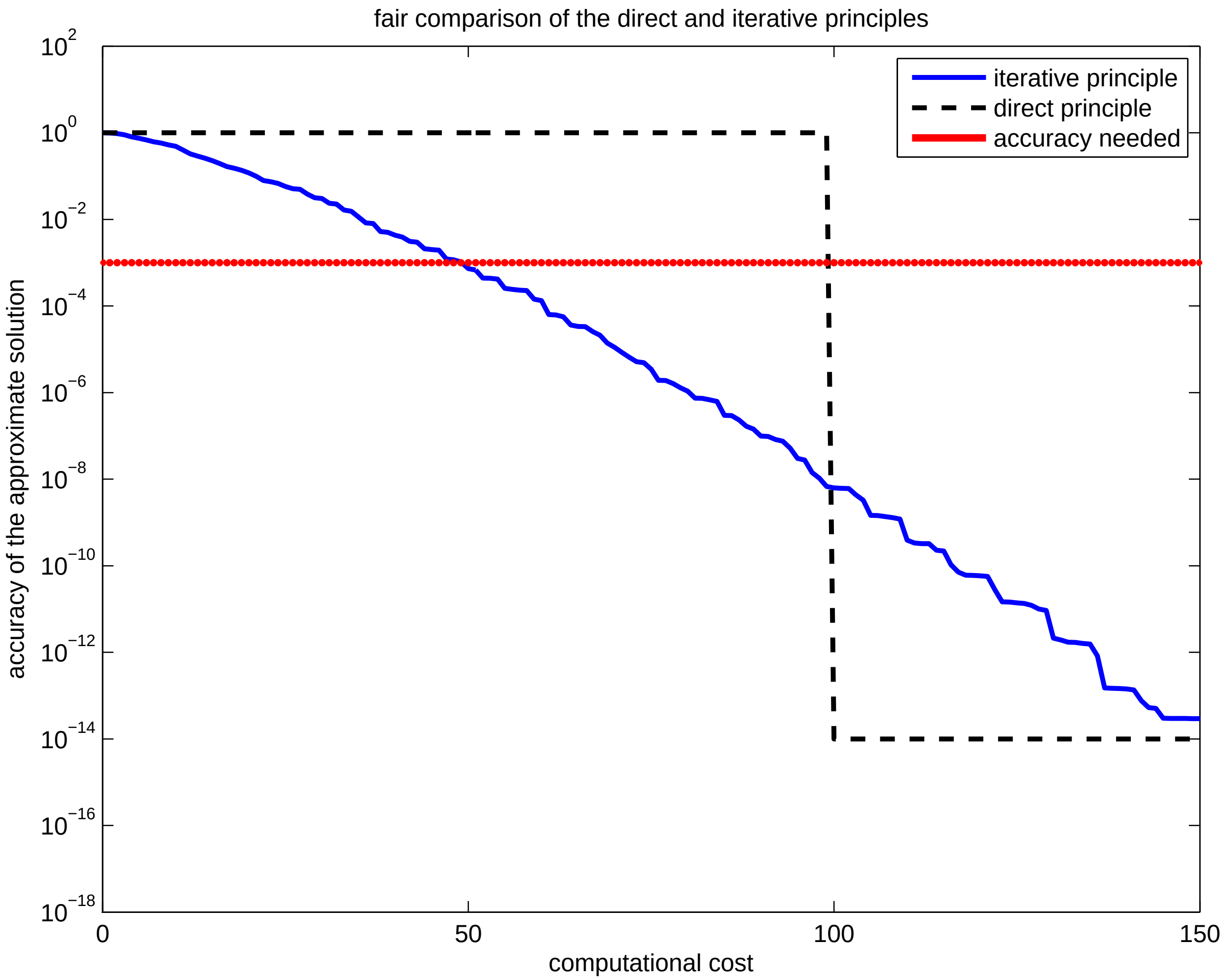
Stability and stopping criteria in iterative methods for linear algebraic systems

Zdeněk Strakoš
Institute of Computer Science AS CR

C.C. Paige, M. Rozložník, G. Meurant, P. Tichý and J. Liesen

PANM 13 in honor of Professor Ivo Babuška, Prague, May 2006.

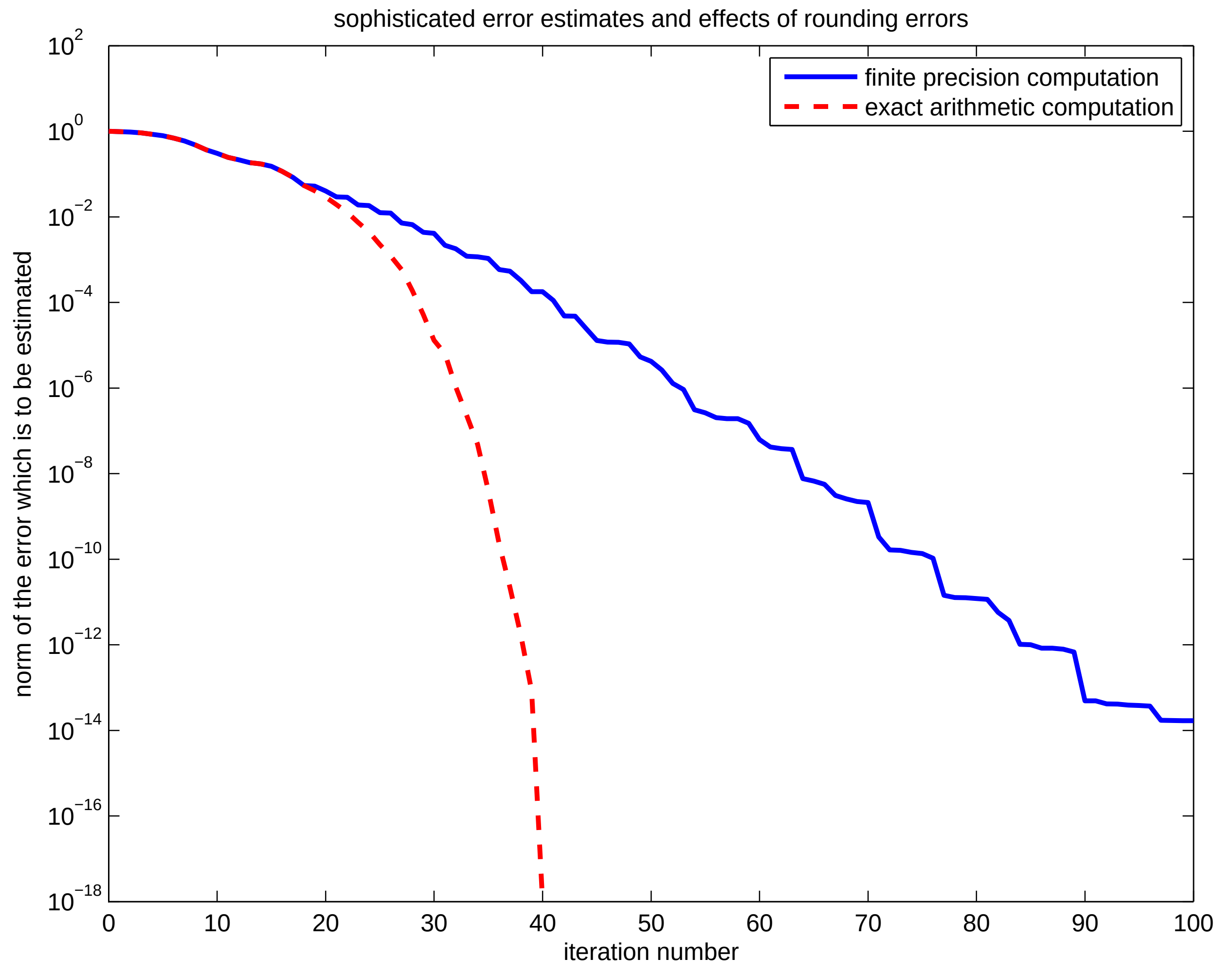
Comparison of the direct and iterative principle



Combination of the direct and iterative principle

- In order to reduce the disadvantages and profit from the advantages.
- Principal advantage of the iterative part is in stopping the computation at the desired accuracy level.
- It requires a meaningful stopping criterion. Errors due to modeling, approximation, uncertainty and computation should be under control.
- Due to difficulties with the previous point this (potential) principal advantage is often presented as a disadvantage (a need for a stopping criteria . . .).

Stopping criteria and rounding error analysis



Common misunderstandings

In modern iterative methods:

- Convergence - no limit, no asymptotics.
- Analysis in numerical linear algebra is highly nonlinear.
- Rounding error analysis of iterative methods is fundamental.

Symmetric case:

G. Meurant and Z. S., *The Lanczos and conjugate gradient method in finite precision arithmetic*, Acta Numerica, 15, Cambridge University Press, pp. 471–542, 2006

Stopping criteria based on error estimates

An example when stopping criteria on the algebraic level is related to the approximation error on the discretization level:

(Preconditioned) Conjugate gradient method
for solving discretized elliptic self-adjoint PDEs,

M. Arioli, Numerische Mathematik, 97, pp. 1-24, 2004

M. Hestenes and E. Stiefel, J. Res. Nat. Bur. St., 49, pp. 409 - 436, 1952
G. Meurant, Numerical Algorithms, 22, pp. 353–365, 1999
Z. S. and P. Tichý, Electr. Trans. on Numer. Anal., 13, pp. 56–80, 2002
Z. S. and P. Tichý, BIT Num. Math., 45, pp. 789–817, 2005
Z. S. and J. Liesen, Z. Angew. Math. Mech., 85, pp. 307–325, 2005

Two comments on numerical stability of GMRES:

1. Simpler GMRES does not work.
2. MGS GMRES is normwise backward stable.

Orthogonality and numerical stability in GMRES

$$\|r_n\| = \min_{u \in x_0 + \mathcal{K}_n(A,r_0)} \|b - Au\| = \min_{z \in A\mathcal{K}_n(A,r_0)} \|r_0 - z\|$$

$$\Leftrightarrow \quad r_n \perp A\mathcal{K}_n(A, r_0)\,,$$

$$\mathcal{K}_n \equiv \mathcal{K}_n(A, r_0) \equiv span\,\{r_0, \cdots, A^{n-1}r_0\}.$$

J. Liesen, M. Rozložník, and Z. S., SIAM J. Sci. Comput., 23, pp. 1503–1525, 2002 :

Wrong choice of the basis which is used in construction of the approximate solution can have a disastrous impact.

Simpler GMRES

Based on the orthonormal basis of $A\mathcal{K}_n(A, r_0)$. Define $w_1 \equiv Ar_0/\|Ar_0\|$, $v_1 = r_0/\|r_0\|$. Then the recursive columnwise QR factorization yields

$$[Av_1,\, AW_{n-1}] \;=\; A\,[v_1,\, W_{n-1}] \;\equiv\; W_n\, R_n\,,$$

$$W_n \;\equiv\; [w_1, \ldots, w_n], \quad W_n^T W_n \;=\; I_n\,,$$

$$\mathrm{span}\,\{w_1, \ldots, w_n\} \;=\; A\mathcal{K}_n(A, r_0)\,,$$

$$\kappa([v_1, W_{n-1}]) \;\leq\; \kappa(R_n) \;\leq\; \kappa(A)\,\kappa([v_1, W_{n-1}])\,.$$

The approximate solution

Using the orthonormal basis $w_1, \dots, w_n$ of $A\mathcal{K}_n(A, r_0)$:

$$x_n \;=\; x_0 \,+\, [v_1,\, W_{n-1}]\, t_n \;\in\; x_0 \,+\, \mathcal{K}_n(A, r_0)\,;$$

$$\Rightarrow \quad r_n \;=\; r_0 - A\,[v_1, W_{n-1}]\, t_n \;=\; r_0 \,-\, W_n R_n\, t_n\,;$$

$$\Rightarrow \quad t_n \;=\; (W_n R_n)^+\, r_0 \;=\; R_n^{-1}\, W_n^T\, r_0\,.$$

How does this affect the numerical stability?

The approximate solution is uncomputable

Theorem

$$\frac{\|r_n\|}{\|r_0\|} \;=\; \sigma_{n+1}([v_1, W_n])\,\sigma_1([v_1, W_n]) \;=\; \frac{2\kappa([v_1, W_n])}{\kappa([v_1, W_n])^2 + 1}\,,$$

$$\frac{\|r_0\|}{\|r_n\|} \;\leq\; \kappa([v_1, W_n]) \;\leq\; 2\,\frac{\|r_0\|}{\|r_n\|}\,,$$

$$\frac{\|r_0\|}{\|r_n\|} \;\leq\; \kappa(R_n) \;\leq\; 2\kappa(A)\,\frac{\|r_0\|}{\|r_n\|}\,.$$

Explanation

Consequently, $\kappa(R_n)$ must inevitably increase as $\|r_n\|$ decreases, even for small $\kappa(A)$ and with the most stable way of computing $w_1, \ldots, w_n$.

$\Rightarrow$ Computation of $t_n = R_n^{-1} W_n^T r_0$ is inherently unstable!

Finite precision computation - surprise (?):

Numerical behavior gets worse when the orthogonality of $w_1, \ldots, w_n$ is maintained better. The Householder implementation performs worse than the Modified-Gram-Schmidt implementation.

The straightforward approach is used in “Simpler GMRES” [Walker, Lu Zhou - 94], and it is related to other implementations, e.g. Orthodir.

Classical GMRES implementation

Based on the orthonormal basis of $\mathcal{K}_n(A, r_0)$. Let $v_1 \equiv r_0/\|r_0\|$. Then the Arnoldi process yields

$$AV_n = V_{n+1} H_{n+1,n} ,$$

$$V_n \equiv [v_1, \ldots, v_n], \quad V_n^T V_n = I_n ,$$

$$\mathrm{span}\,\{v_1, \ldots, v_n\} = \mathcal{K}_n(A, r_0) ,$$

$$\kappa(H_{n+1,n}) \le \kappa(A) .$$

The problem can not occur!

Using the orthonormal basis $v_1, \dots, v_n$ of $\mathcal{K}_n(A, r_0)$:

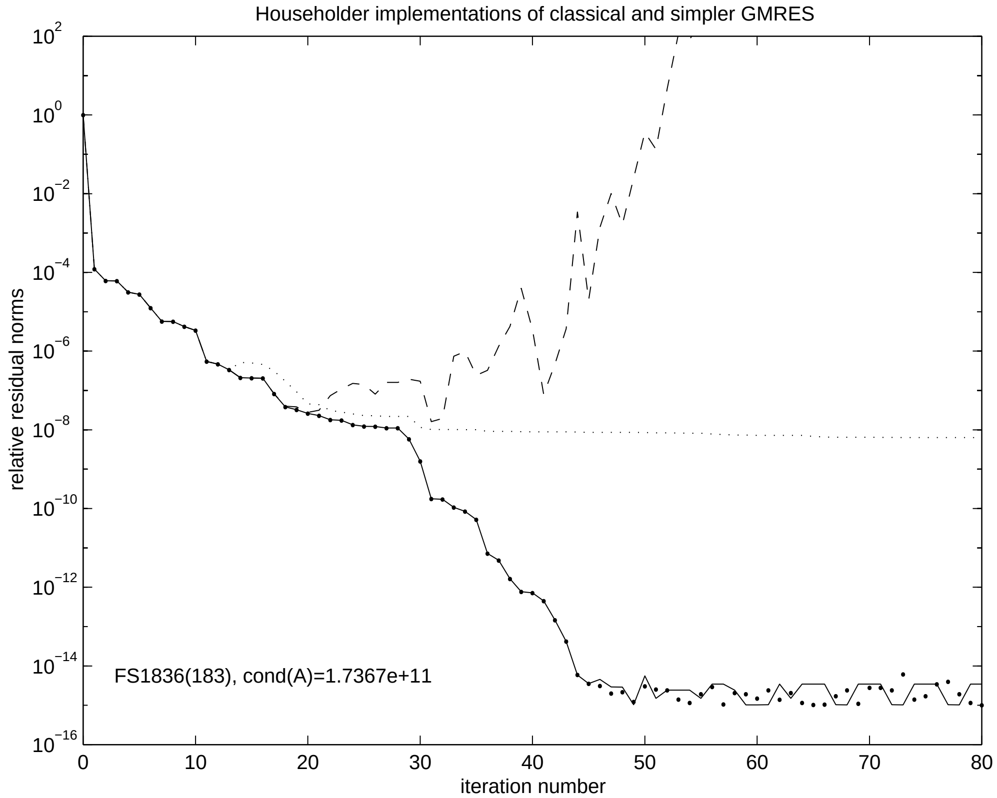
$$x_n = x_0 + V_n z_n \in \mathcal{K}_n(A, r_0),$$

$$\Rightarrow \quad r_n = r_0 - AV_n z_n = r_0 - V_{n+1} H_{n+1,n} z_n ,$$

$$\Rightarrow \quad z_n = (V_{n+1} H_{n+1,n})^+ r_0 = (H_{n+1,n})^+ V_{n+1}^T r_0 .$$

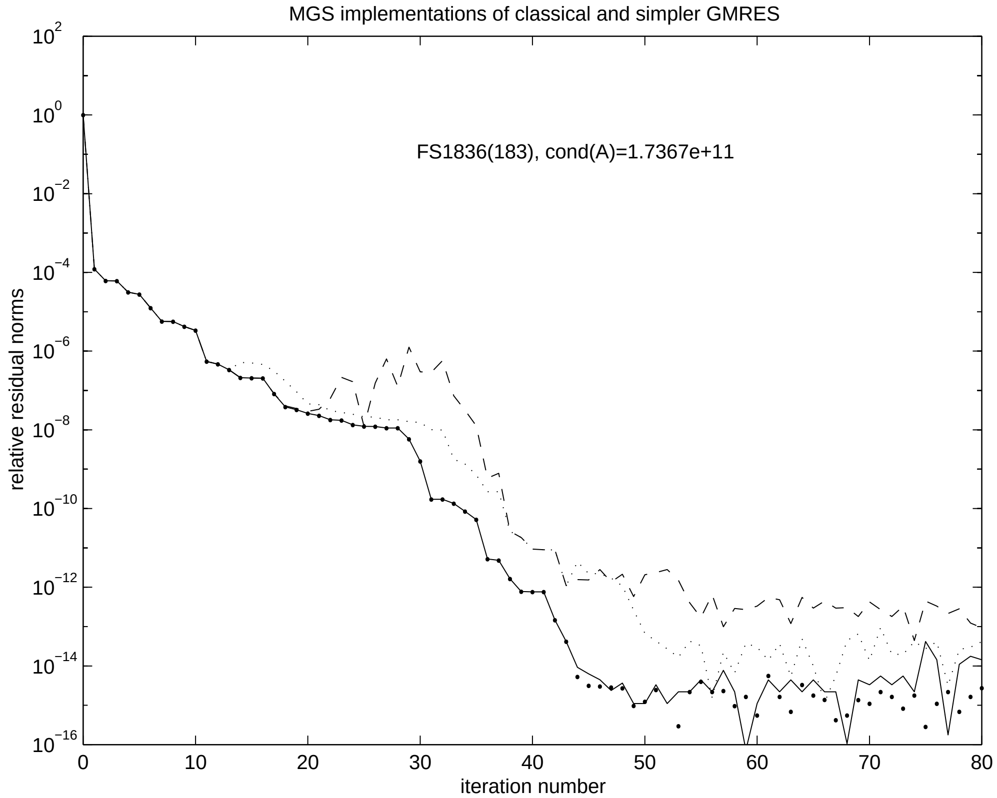
How does this affect the numerical stability?

Illustration - Householder implementation



Excessive ill-conditioning of computed R_n leads in the straightforward implementation (simpler GMRES) to divergence (dashed line).

Illustration - MGS implementation



Due to rounding errors the (exact precision) identity is violated, and the computed R_n is not so badly ill-conditioned as it ideally should be!

Moral

Choice of the right basis used for construction of the approximate solution is fundamental for numerical stability of the Krylov subspace methods.

Even the best orthogonalization technique in computation of the basis (here Householder reflections) can not compensate for instabilities artificially created due to a bad choice of the subspace. Paradoxically - preserving orthogonality of the computed basis can even make things worse!

In the rest we concentrate on the classical GMRES.

Literature

Numerical stability of the GMRES implementation based on the (ideally) orthonormal basis of $\mathcal{K}_n(A, r_0)$ – some related publications:

[Björck - 67], [Karlson - 91], [Björck, Paige - 92],

[Drkošová, Greenbaum, Rozložník, S - 95], [Arioli, Fassino - 96],

[Greenbaum, Rozložník, S - 96], [Rozložník 1997],

[Paige, S - 02a, 02b, 02c],

[Giraud, Langou, Rozložník, van der Eshof - 05], ...

Common knowledge on MGS GMRES

$$v_{n+1} = (I - v_n v_n^* - \; \ldots \; - v_1 v_1^*) \, A v_n$$

$$= (I - v_n v_n^*) \; \ldots \; (I - v_1 v_1^*) \, A v_n$$

A common belief:

MGS orthogonalization is a good compromise between propagation of errors (loss of orthogonality) and algorithm efficiency (computational cost). Price - the computation is recursive!

Comparison with classical Gram-Schmidt, Householder reflections, Givens rotations.

Another surprise

Despite the loss of orthogonality, some (good!) algorithms with MGS provide results as good as algorithms using the most stable orthogonalization (with the loss of orthogonality among the computed basis vectors kept close to the machine precision level).

Theoretical justification?

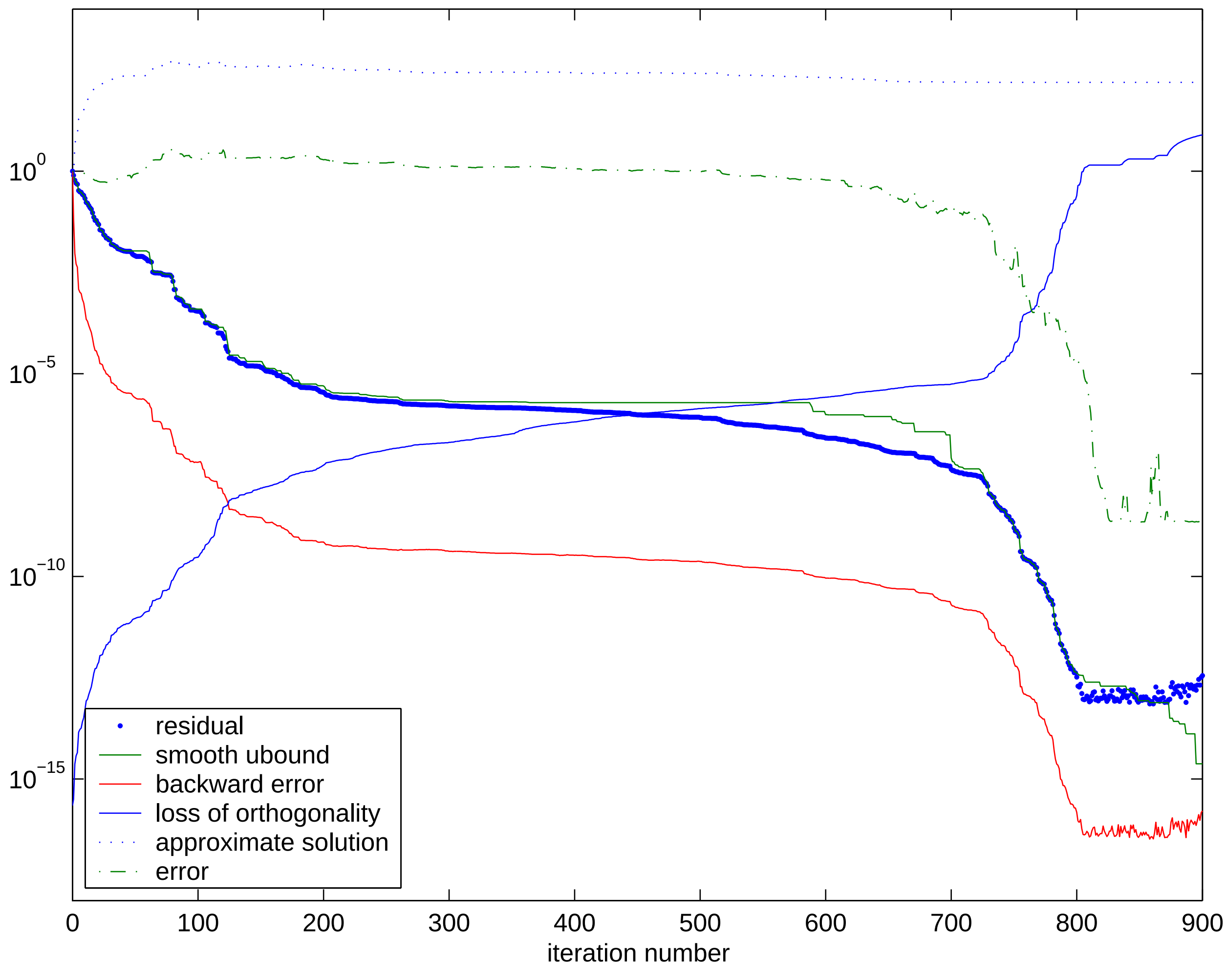
- Linear least squares – direct methods: [Björck, Paige - 92];
- Our case: MGS GMRES.

One more reference

MGS GMRES is normwise backward stable!

C. C. Paige, M. Rozložník, and Z. S., *Modified Gram-Schmidt (MGS), Least Squares, and backward stability of MGS-GMRES*, SIAM J. Matrix Anal Appl., 28, pp. 264–284, 2006

Last figure: Sherman 2, b MM, x0 = 0



Most important of all

Thank you, Professor Babuška, and many happy and fruitful years!