Questions on the concepts of numerical stability in Krylov subspace methods

Zdeněk Strakoš

Nečas Center for Mathematical Modeling
Charles University in Prague and Czech Academy of Sciences

http://www.karlin.mff.cuni.cz/~strakos

University of Strathclyde, June 2015

Thanks to very many

Hackbusch, Computing (1995)

Adaptivity has created a new paradigm in mathematical computation. In traditional numerical mathematics, the fields "discretization" (e.g., FEM), its "numerical analysis" (e.g., error estimates), and "solution algorithms" (e.g., solvers for linear systems) are well separated. Adaptive techniques, however, require a combination of all three. For example, the error estimation has become a part of the algorithm. The concrete discretization is now an outgrowth of the algorithm."

PROBLEM $\rightleftharpoons$ MODEL $\rightleftharpoons$ DISCRETIZATION $\rightleftharpoons$ COMPUTATION

Individual stages are accompanied by errors, in particular by approximation errors of the model, discretization errors, linearization errors, truncation and/or rounding errors in numerical matrix computations.

Adaptivity (take AFEM as an example)

After setting the model and its initial discrete approximation, AFEM adaptivity proceeds with iterating the step

$$\text{SOLVE} \longrightarrow \text{ESTIMATE} \longrightarrow \text{MARK} \longrightarrow \text{REFINE}$$

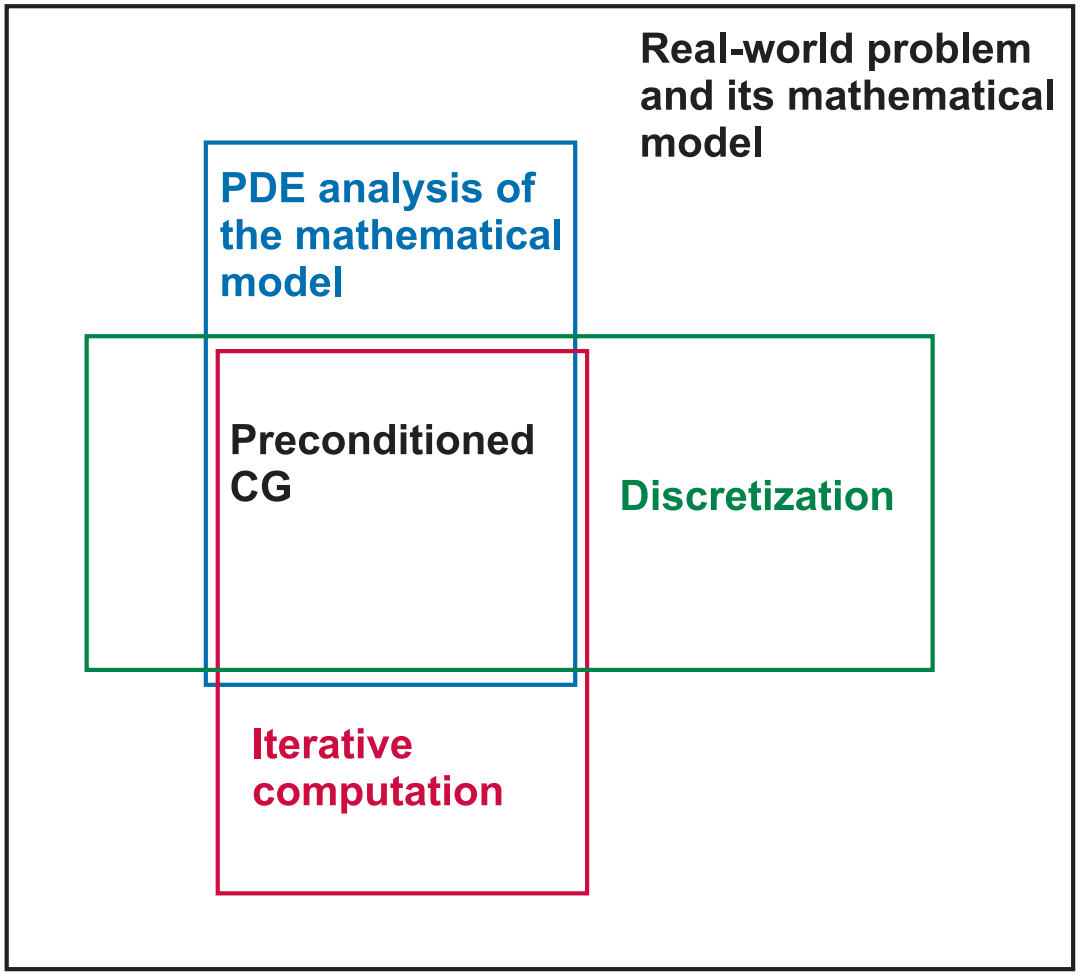
Here SOLVE can very rarely be considered in the exact sense, i.e., with a negligible computational error.

Von Neumann and Goldstine (1947)

When a problem in pure or in applied mathematics is 'solved' by numerical computation, errors, that is, deviations of the numerical 'solution' obtained from the true, rigorous one, are unavoidable. Such a 'solution' is therefore meaningless, unless there is an estimate of the total error in the above sense.

This analysis of the sources of errors should be objective and strict inasmuch as completeness is concerned, . . .

Example, Málek and S, SIAM Spotlight, 2015



Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

1 Introduction

- Is inexact computing an unpleasant reality that is going to be changed by the progress of computer architectures, or is it the way to go?
- What is meant by numerical stability analysis and do we have a theory of inexact computing?
- Are the computed results endangered by stochastic accumulation of rounding errors, or by their deterministic amplification during their propagation?
- Can we assume exact arithmetic while evaluating computational behaviour of Krylov subspace methods?

Section 8 of the paper on CG by Hestenes and Stiefel (1952) offers an instructive reading (and ideas rediscovered decades later).

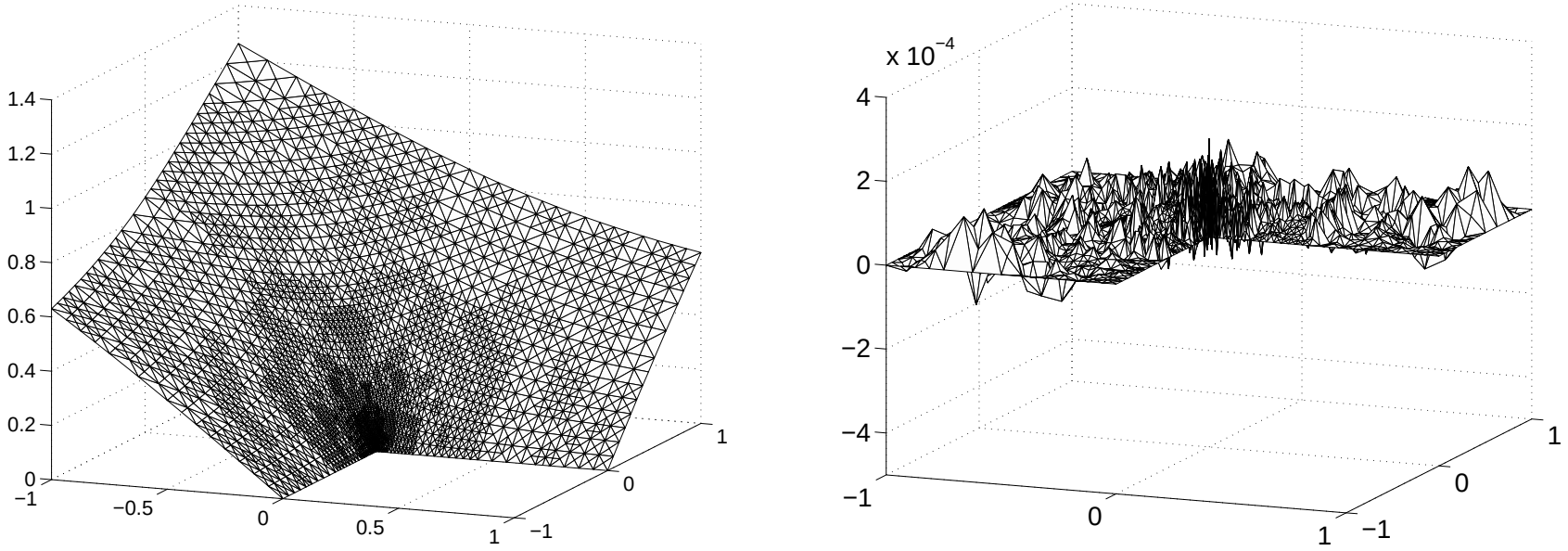
Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

2 How do we measure computational error?

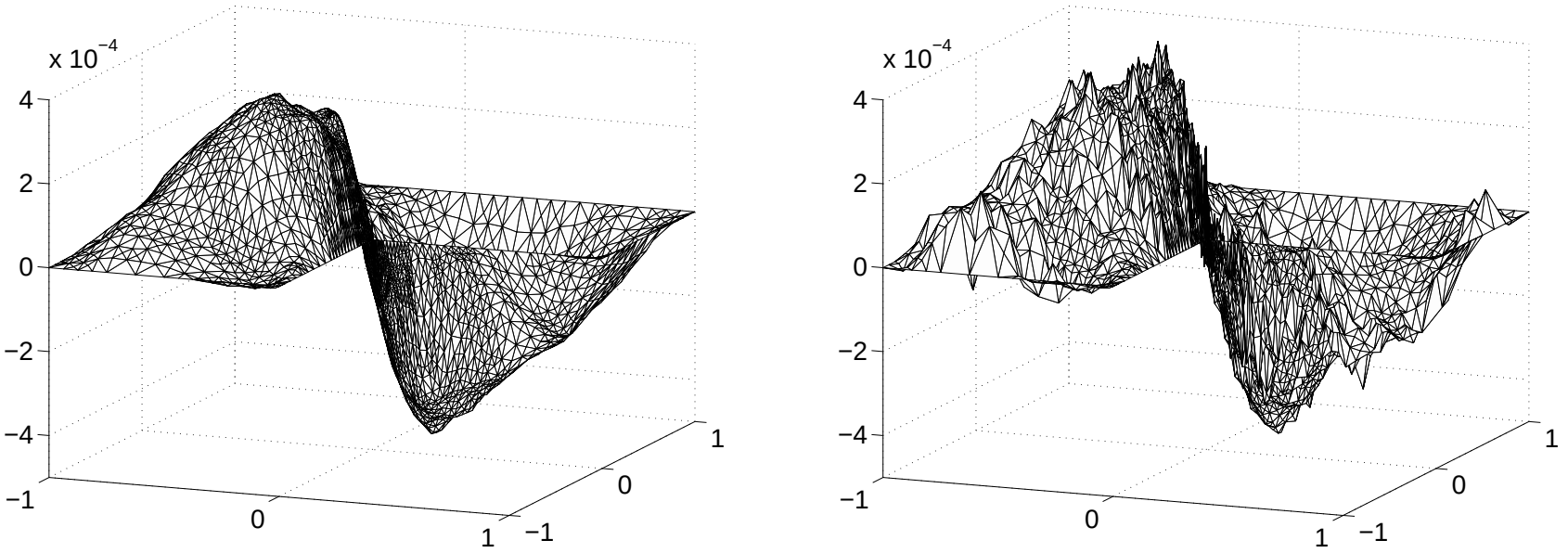
- Residual norm is a clear winner in practice. Even for CG, where it can significantly grow or oscillate and where the classical works strongly argue against using it; see Hestenes and Stiefel (1952).
- When the condition number is small, the residual norm can be trusted. But do we need Krylov subspace methods then? In particular, when considering a highly parallel (exascale) environment?
- Backward error interpretation, normwise and componentwise backward error. Backward error analysis and perturbation (sensitivity) analysis give the normwise forward error estimates.
- Does the existing methodology allow for estimation of the componentwise forward error?

2 L-shape domain, Papež, Liesen, S (2014)



Exact solution x (left) and the discretisation error $x - x_h$ (right) in the Poisson model problem, linear FEM.

2 L-shape domain, Papež, Liesen, S (2014)



Algebraic error $x_h - x_h^{(n)}$ (left) and the total error $x - x_h^{(n)}$ (right) after the number of CG iterations guaranteeing

$$\|\nabla(x - x_h)\| \;\gg\; \|\mathbf{x} - \mathbf{x}_n\|_{\mathbf{A}} \,.$$

2 Iterative and Krylov subspace methods?

- Iterative methods and fixed point theorem
- Iterative methods as a sequence of steps, with each representing a contraction
- Convergence is identified with a contraction parameter uniformly bounded from the value one.
- Iterative methods are self-correcting; see the letter of Gauss to Gerling from 1823 (?)
- Does the previous apply to Krylov subspace methods? See Lanczos (1952)!

2 Basic Krylov subspace methods

- CG and MINRES (SYMMLQ) for symmetric problems
- BiCG (CGS, BiCGStab) and QMR for nonsymmetric problems
- FOM and GMRES for nonsymmetric problems

The basic methods with powerful preconditioners will serve in most problems. In analysis, we must understand the basic methods first.

2 Computational cost and inaccurate computations

- The main distinction between direct and iterative computations
- How fast can we reach the desired accuracy? By the way, what is it?
- Can we ever get there?

The second item refers to the possible delay due to inaccurate computations.

The third item questions the maximal attainable accuracy.

Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

3 Self-adjoint bounded and coercive $\mathcal{A}$, CG

CG in Hilbert spaces : $r_0 = b - \mathcal{A}x_0 \in V^{\#}, \quad p_0 = \tau r_0 \in V$

For $n = 1, 2, \ldots, n_{\max}$

$$\alpha_{n-1} = \frac{\langle r_{n-1}, \tau r_{n-1} \rangle}{\langle \mathcal{A}p_{n-1}, p_{n-1} \rangle} = \frac{(\tau r_{n-1}, \tau r_{n-1})_V}{(\tau \mathcal{A}p_{n-1}, p_{n-1})_V}$$

$$x_n = x_{n-1} + \alpha_{n-1} p_{n-1} \,,$$

stop when the stopping criterion is satisfied

$$r_n = r_{n-1} - \alpha_{n-1} \mathcal{A}p_{n-1}$$

$$\beta_n = \frac{\langle r_n, \tau r_n \rangle}{\langle r_{n-1}, \tau r_{n-1} \rangle} = \frac{(\tau r_n, \tau r_n)_V}{(\tau r_{n-1}, \tau r_{n-1})_V}$$

$$p_n = \tau r_n + \beta_n p_{n-1}$$

End

3 Preconditioned algebraic CG $(\mathcal{A}_h x_h = b_h)$

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0, \quad$ solve $\quad \mathbf{M}\mathbf{z}_0 = \mathbf{r}_0, \ \mathbf{p}_0 = \mathbf{z}_0$

For $n = 1, \ldots, n_{\max}$

$$\alpha_{n-1} = \frac{\mathbf{z}_{n-1}^* \mathbf{r}_{n-1}}{\mathbf{p}_{n-1}^* \mathbf{A}\mathbf{p}_{n-1}}$$

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \alpha_{n-1}\mathbf{p}_{n-1}\,,$$ stop when the stopping criterion is satisfied

$$\mathbf{r}_n = \mathbf{r}_{n-1} - \alpha_{n-1}\mathbf{A}\mathbf{p}_{n-1}$$

$$\mathbf{z}_n = \mathbf{M}^{-1}\mathbf{r}_n\,,$$ solve for $\mathbf{z}_n$

$$\beta_n = \frac{\mathbf{z}_n^* \mathbf{r}_n}{\mathbf{z}_{n-1}^* \mathbf{r}_{n-1}}$$

$$\mathbf{p}_n = \mathbf{z}_n + \beta_n \mathbf{p}_{n-1}$$

End

3 CG, large outliers and condition numbers

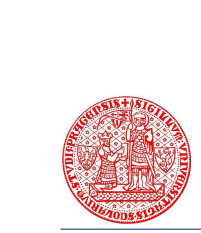
Consider the desired accuracy ϵ, $\kappa_s(\mathbf{A}) \equiv \lambda_{N-s}/\lambda_1$. Then

$$\mathbf{k} = \mathbf{s} + \left\lceil \frac{\ln(2/\epsilon)}{2} \sqrt{\kappa_s(\mathbf{A})} \right\rceil$$

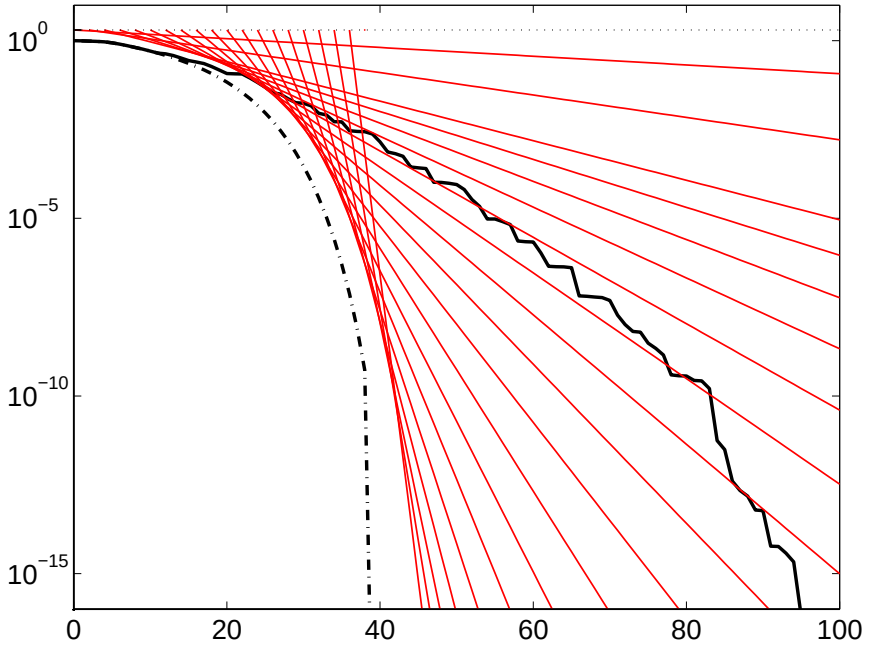
CG steps will produce the approximate solution $\mathbf{x}_n$ satisfying

$$\|\mathbf{x} - \mathbf{x}_n\|_{\mathbf{A}} \leq \epsilon \, \|\mathbf{x} - \mathbf{x}_0\|_{\mathbf{A}} \, .$$

This statement qualitatively explains superlinear convergence of CG at the presence of large outliers in the spectrum, assuming exact arithmetic.



3 Liesen, S (2013); Gergelits, S (2014)



Short recurrences typically mean in practical computations loss of (bi-)orthogonality and linear independence due to rounding errors. This means delay of convergence.

3 Clusters = fast convergence ?

- Finite precision arithmetic CG computation with a matrix having t isolated well separated eigenvalues may require for reaching a reasonable approximate solution a significantly larger number of steps than t .

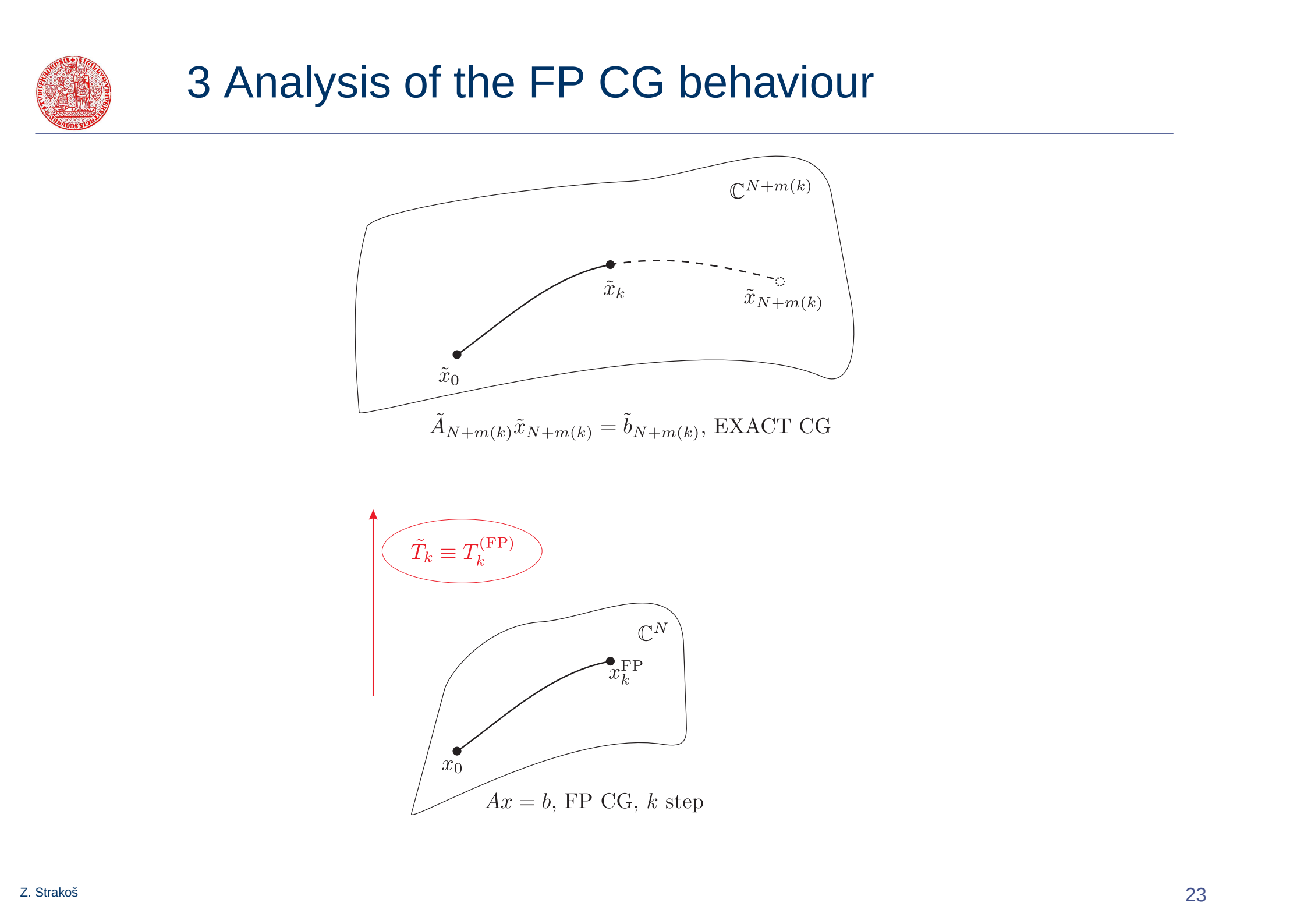
3 Self adjoint bounded operators

$$
\begin{array}{ccc}
\mathcal{B}\,u \;=\; f & \longleftrightarrow & \omega(\lambda),\ \displaystyle\int F(\lambda)\,d\omega(\lambda) \\
\uparrow & & \uparrow \\
\mathbf{T}_n\,\mathbf{y}_n \;=\; \mathbf{e}_1 & \longleftrightarrow & \omega^{(n)}(\lambda),\ \displaystyle\sum_{i=1}^{n} \omega_i^{(n)}\, F\left(\theta_i^{(n)}\right)
\end{array}
$$

Using $F(\lambda) = \lambda^{-1}$ gives

$$
\int_{\lambda_L}^{\lambda_U} \lambda^{-1}\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1} \;+\; \frac{\|u-u_n\|_a^2}{\|f\|_V^2}
$$

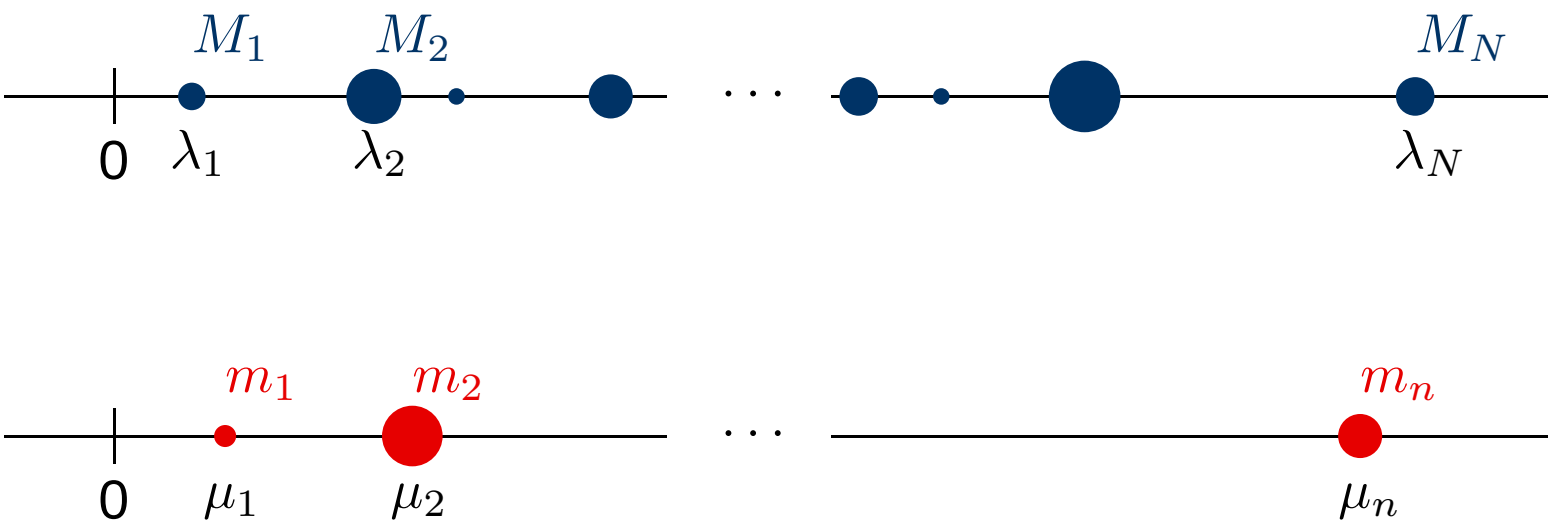
Stieltjes and Vorobyev moment problems are for self-adjoint bounded operators equivalent; infinite dimensional CG (Lanczos).



3 Analysis of the FP CG behaviour

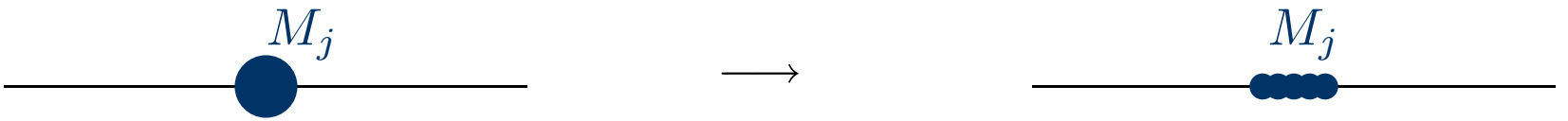
3 Moment problem illustration of exact CG

For a given n find a distribution function with n mass points in such a way that it in a best way captures the properties of the distribution function determined by the matrix A and the initial residual r_0 .



3 Finite precision CG

Paige (1971–80); Greenbaum (1989); S (1991); Greenbaum, S (1992)



single eigenvalue

λ_j

$\longrightarrow$

many close eigenvalues

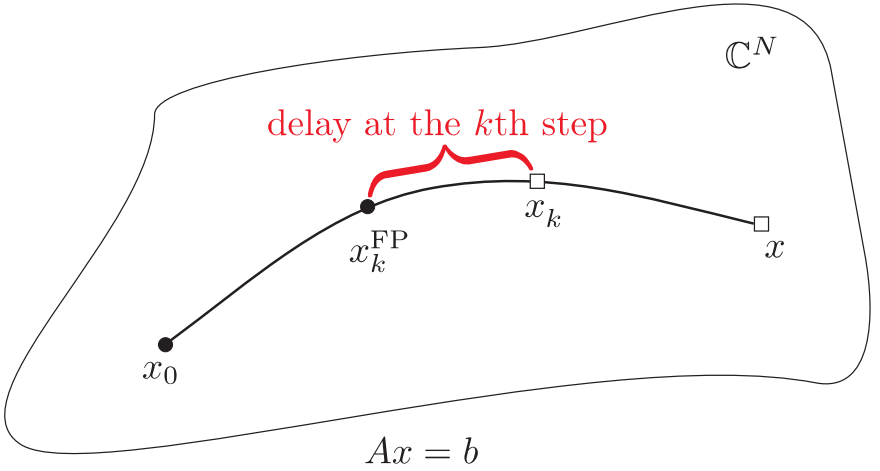
$\widehat{\lambda}_{j_1}, \widehat{\lambda}_{j_2}, \ldots, \widehat{\lambda}_{j_\ell}$

3 Omitted issue

- We skipped the issue of maximal attainable accuracy (comparison of the iteratively and directly computed residuals, Greenbaum (1997), ... , analogous methodology used in relaxing accuracy of the matrix-vector multiplication in the works on inexact Krylov subspace methods)

3 We are not done!

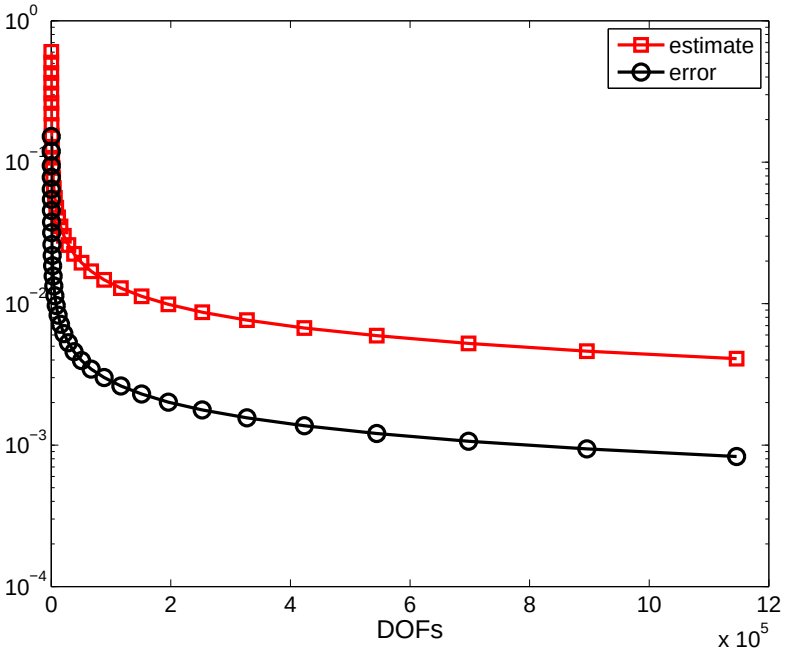
- The bounds on the cluster widths in Greenbaum (1989) are not optimal
- Other approaches based on the augmented systems due to Paige and others
- Do we remain close to the theoretical Krylov subspaces?
  Gergelits and S (?)

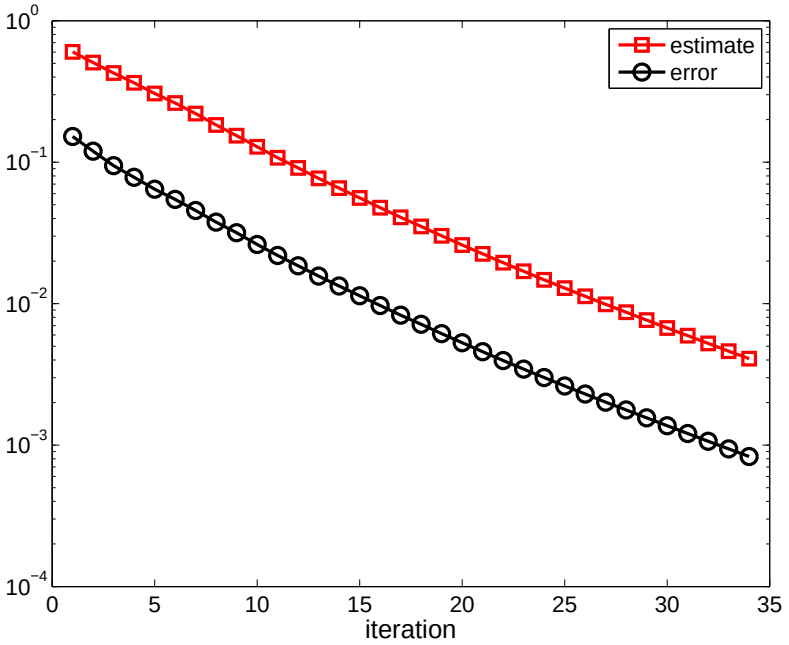


□ exact computation

● finite precision computation

3 Reaching an arbitrary accuracy in AFEM?





Inexactness and maximal attainable accuracy in matrix computations?

Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

4 Formally look similar ...

- Effects of rounding errors at individual iterations can not be bounded due to the possible (near) breakdowns in nonsymmetric Lanczos
- Look-ahead can not guarantee well conditioning of the computed Lanczos vectors
- Stabilization of the CGS (BiCG) may destabilize the computation of the recurrence coefficients
- Estimates in quadratic formulas need numerical stability justifications (that applies also to the symmetric case)

4 The BiCG approximation to $c^* A^{-1} b$

Using local biorthogonality

$$c^* A^{-1} b \;=\; \sum_{j=0}^{n-1} \alpha_j s_j^* r_j \;+\; s_n^* A^{-1} r_n \,.$$

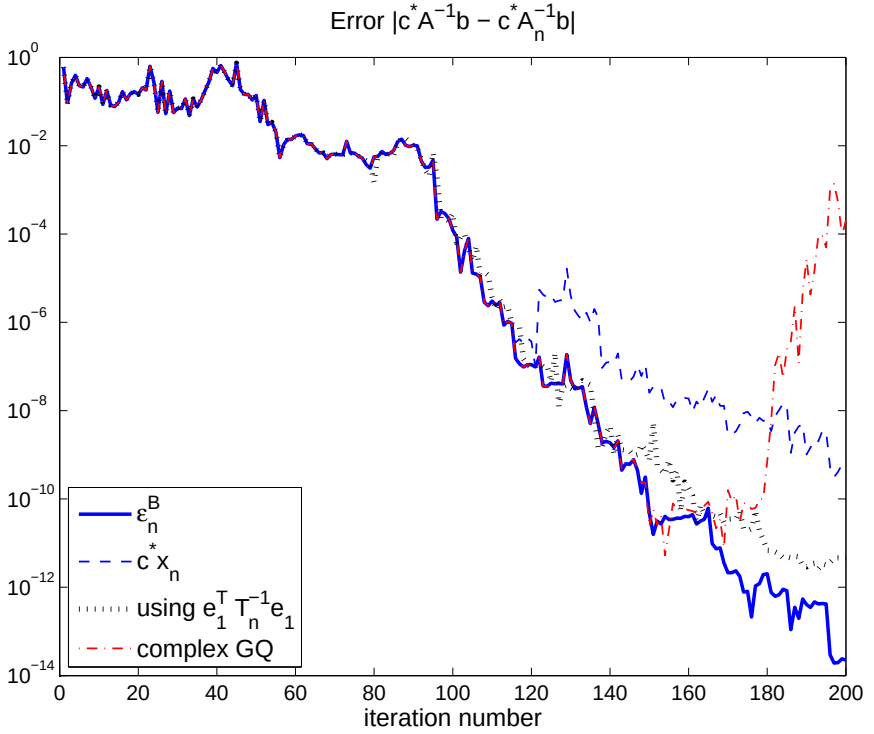
Using global biorthogonality

$$c^* A^{-1} b \;=\; c^* x_n \;+\; s_n^* A^{-1} r_n \,.$$

Finally,

$$c^* A_n^{-1} b \;=\; (c^* v_1)\, \|b\|\, (T_n^{-1})_{1,1} \;=\; c^* x_n = \sum_{j=0}^{n-1} \alpha_j s_j^* r_j \,.$$

4 RCWA problem - comparison of estimates



Comparison of mathematically equivalent estimates based on BiCG and Non-Hermitian Lanczos.

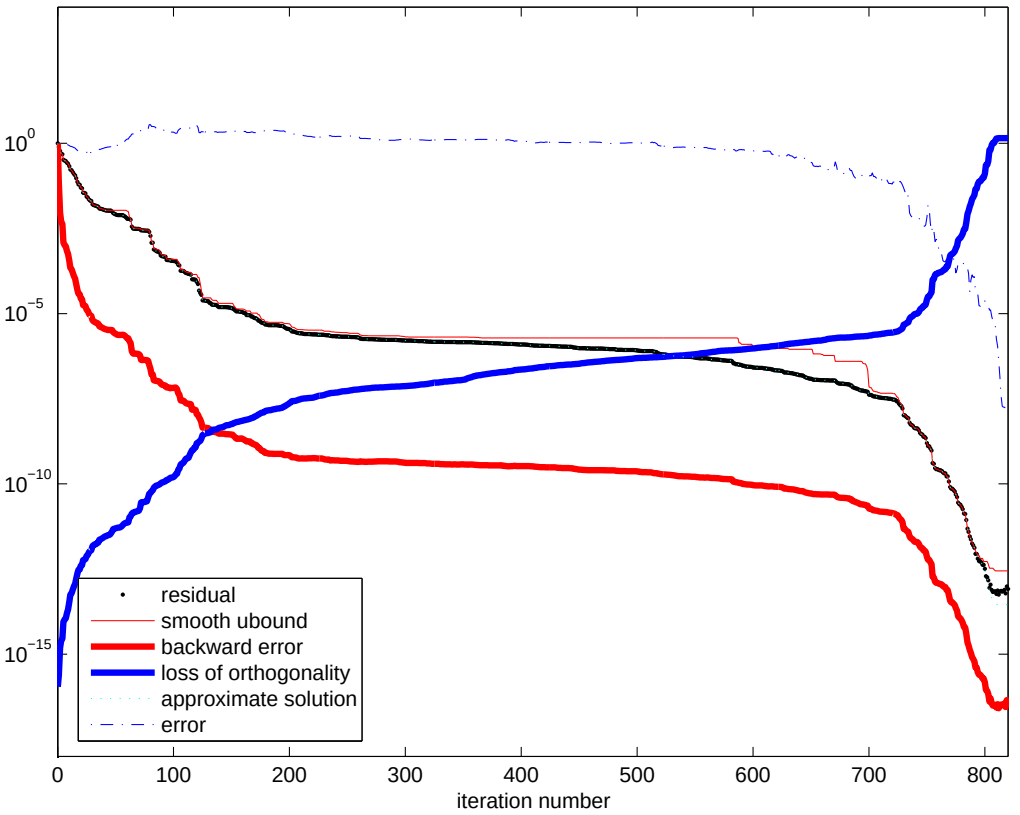
Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

5 They are!

- Using short recurrences, global loss of orthogonality leading to loss of linear independence among generating vectors means loss of information and, consequently, delay of convergence
- In modified Gram-Schmidt GMRES the orthogonality among the computed generating vectors is completely lost only after the GMRES reaches its maximal attainable accuracy
- Modified Gram-Schmidt GMRES is backward stable; see Paige, Rozložník, S (2006)
- Loss of orthogonality and the normwise backward error are in MGS GMRES inversely proportional
- Classical Gram-Schmidt GMRES? Nice results on classical Gram-Schmidt QR ...

5 MGS GMRES computation



Sherman2 from Matrix market, problem rhs.

Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

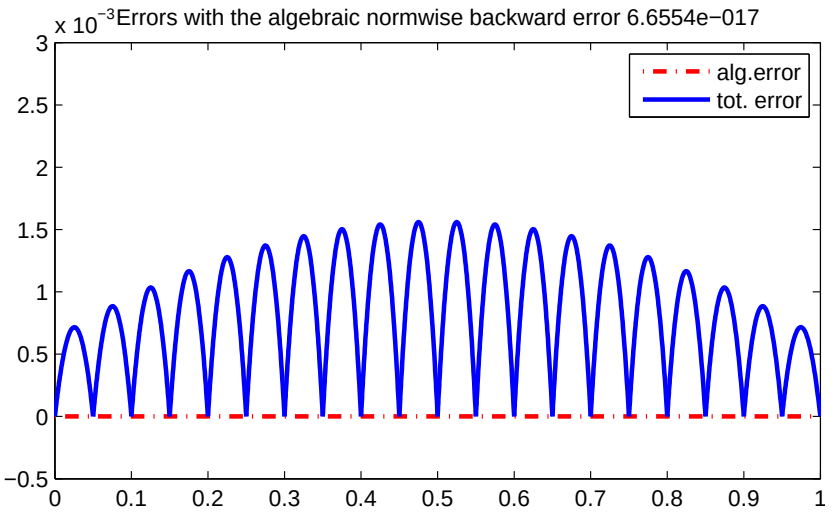
6 It depends ...

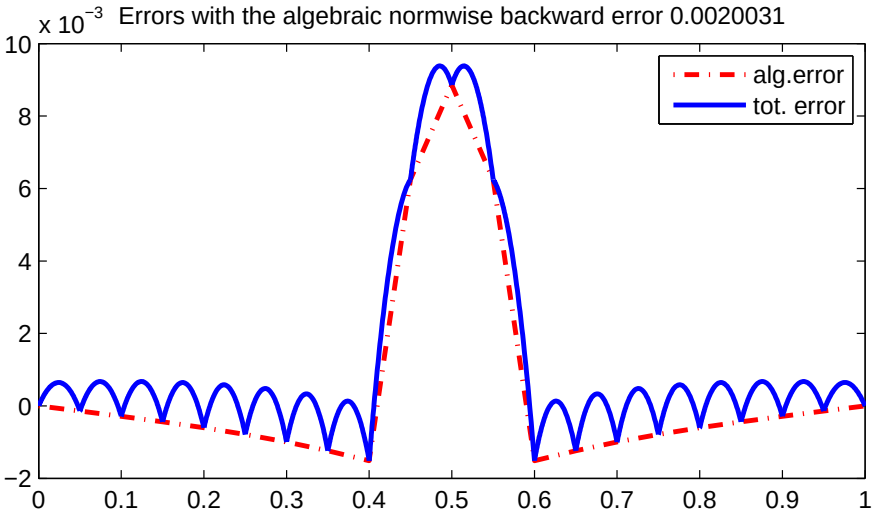
- Yes. Poisson problem is easy. CG will exhibit almost no loss of orthogonality (therefore it makes a little sense to use it as a test problem for numerical stability issues)

- No. Even with 1D Poisson problem we can observe very surprising phenomena

Here we mean using model problem for challenging the solvers. We definitely do not argue for solving the Poisson problem using the CG method.

6 1D Poisson, Papež, Liesen and S (2014)





The discretisation error (left),
the algebraic and the total error (right),

Outline

1. Introduction
2. Should common practice be questioned?
3. Symmetric problems and short recurrences
4. Nonsymmetric problems and short recurrences
5. Are long recurrences different?
6. Are model problems easy?
7. Conclusions

7 Conclusions

- Plethora of Krylov subspace methods and their implementations
- Understanding of the basic ones forms a solid ground for further algorithmic development
- While formulating results of an analysis, equal emphasize should be on the assumptions restricting their applicability. They offer a valuable guidance on what is yet to be done
- With increasing computing power, the role of rigorous analysis dramatically increases

7 Conclusions

Patrick J. Roache's book *Validation and Verification in Computational Science*, 1998, p. 387:

"With the often noted tremendous increases in computer speed and memory, and with the less often acknowledged but equally powerful increases in algorithmic accuracy and efficiency, a natural question suggest itself. What are we doing with the new computer power? with the new GUI and other set-up advances? with the new algorithms? What *should* we do? ... Get the right answer."

References

- J. Málek and Z.S., *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs.* SIAM Spolight Series, SIAM (2015)
- T. Gergelits and Z.S., *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations,* Numer. Alg. 65, 759-782 (2014)
- J. Papež, J. Liesen and Z.S., *Distribution of the discretization and algebraic error in numerical solution of partial differential equations,* Linear Alg. Appl. 449, 89-114 (2014)
- J. Liesen and Z.S., *Krylov Subspace Methods, Principles and Analysis.* Oxford University Press (2013)
- Z.S. and P. Tichý, *On efficient numerical approximation of the bilinear form* $c^* A^{-1} b$*,* SIAM J. Sci. Comput., 33 (2011), pp. 565-587

Thank you very much for kind patience!