Nečas Center for Mathematical Modeling

MOdelling REvisited + MOdel REduction
ERC-CZ project LL1202 - MORE

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

On the numerical behaviour of the CG method

Tomáš Gergelits, Zdeněk Strakoš

Department of Numerical Mathematics, Faculty of Mathematics and Physics,
Charles University in Prague

Student Krylov Day 2015, TU Delft

2 February 2015

Content of the talk

Composite polynomial bounds for CG in finite precision arithmetic

Krylov subspaces generated by CG in finite precision arithmetic

The essence of the CG method

Consider preconditioned system

$$Ax = b, \quad A \in \mathbb{C}^{N\times N} \text{ HPD matrix} \quad \text{and} \quad b \in \mathbb{C}^{N}.$$

CG is the projection method which minimizes the energy norm of the error

$$x_k \in x_0 + \mathcal{K}_k(A, r_0), \quad r_k \perp \mathcal{K}_k(A, r_0), \quad k = 1, 2, \ldots$$
$$\mathcal{K}_k(A, r_0) = \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$$

$$\|x - x_k\|_A = \min\left\{\|x - y\|_A : \; y \in x_0 + \mathcal{K}_k(A, r_0)\right\}.$$

CG is a matrix formulation of the Gauss-Christoffel quadrature

$\Rightarrow$ The CG method is nonlinear.

Linear bound for the nonlinear CG method

The error in the CG method satisfies

$$\|x - x_k\|_A = \min_{\substack{\varphi(0)=1\\ \deg(\varphi)\le k}} \left\{ \sum_{j=1}^{N} |\xi_j|^2 \lambda_j \varphi^2(\lambda_j) \right\}^{1/2} \le \min_{\substack{\varphi(0)=1\\ \deg(\varphi)\le k}} \max_{j=1,\dots,N} |\varphi(\lambda_j)| \|x - x_0\|_A .$$

The error in the Chebyshev semi-iterative (CSI) method satisfies

$$\|x - x_k^{CSI}\|_A \le |\chi_k(0)|^{-1} \|x - x_0\|_A = \min_{\substack{\varphi(0)=1\\ \deg(\varphi)\le k}} \max_{\lambda\in[\lambda_1,\lambda_N]} |\varphi(\lambda)| \|x - x_0\|_A .$$

[Flanders, Shortley (1950), Lanczos (1953), Young (1954); Markov (1884)]

Linear bound is relevant for the CSI method and trivially holds for CG

$$\|x - x_k\|_A \le \|x - x_k^{CSI}\|_A \le 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \|x - x_0\|_A .$$

[Rutishauser (1959)]

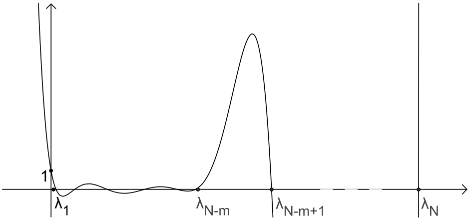
In the case of m large outlying eigenvalues the composite polynomial

$$
\begin{aligned}
&q_m(\lambda)\chi_{k-m}(\lambda)/\chi_{k-m}(0), \quad \text{where} \\
&\qquad q_m(\lambda) = (\lambda - \lambda_N)\dots(\lambda - \lambda_{N-m+1}), \\
&\qquad \chi_{k-m} \equiv (k-m)\text{th Chebyshev polynomial shifted on } [\lambda_1, \lambda_{N-m}]
\end{aligned}
$$

gives for $k \geq m$

$$\frac{\|x - x_k\|_A}{\|x - x_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_m} - 1}{\sqrt{\kappa_m} + 1}\right)^{k-m}$$

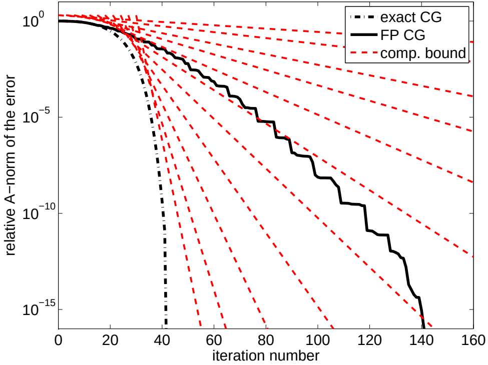
$$\kappa_m = \frac{\lambda_{N-m}}{\lambda_1}.$$



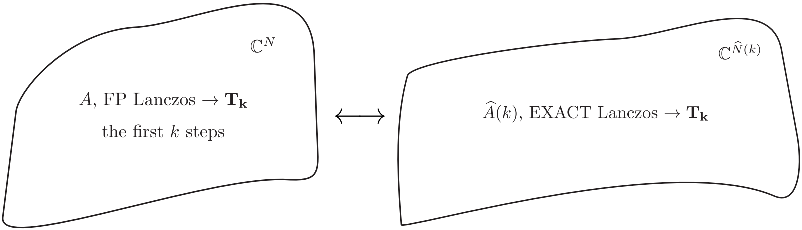
[Axelsson (1976), Jennings (1977); cf. van der Sluis, van der Vorst (1986)]

Short recurrences $\Longrightarrow$ loss of orthogonality $\Longrightarrow$ delay of convergence & rank deficiency

Failure of the composite polynomial bound



Backward-like analysis



Eigenvalues of $\widehat{A}(k)$ are tightly clustered around the eigenvalues of A.

[Paige (1980), Greenbaum (1989), Strakoš (1991)]

In numerical experiments, $\widehat{A}(k)$ can be replaced (with small inaccuracy) by an "universal" $\widehat{A}$ with sufficiently many eigenvalues in tiny clusters around the eigenvalues of A.

[Greenbaum, Strakoš (1992)]

Consequences

Minimization problem which bounds the CG convergence behaviour in finite precision arithmetic is

$$\min_{\substack{\varphi(0)=1 \\ \deg(\varphi)\le k}} \max_{\lambda\in\sigma(\widehat{A})} |\varphi(\lambda_j)|$$

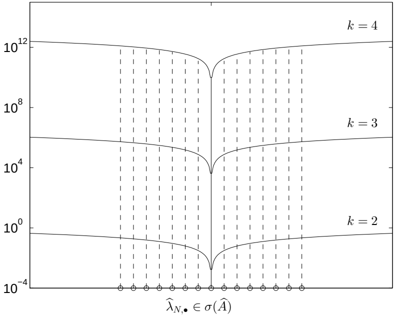
with the spectrum of the matrix $\widehat{A}$

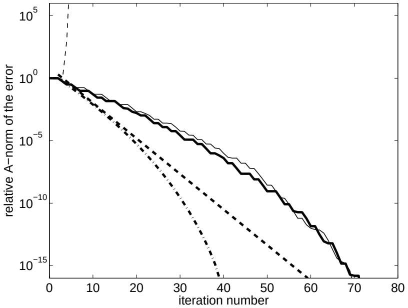
$$\sigma(\widehat{A}) \equiv \bigcup_{j=1,\ldots,N} \left\{\widehat{\lambda}_{j,1},\ldots,\widehat{\lambda}_{j,l}\right\}, \quad \widehat{\lambda}_{j,\cdot} \in [\lambda_j - \Delta, \lambda_j + \Delta] \quad \text{with tiny } \Delta.$$

Consequently, the upper bound based on the composite polynomial must be based on

$$\max_{\lambda\in\sigma(\widehat{A})} \left| q_m(\lambda) \frac{\chi_{k-m}(\lambda)}{\chi_{k-m}(0)} \right| .$$

Consequences





The composite polynomial is very steep in the neighbourhood of the outlying eigenvalues and thus the corresponding bound (dashed line) blows up.

Summary I

Analysis assuming exact arithmetic is an oxymoron.

☞ short recurrences – loss of orthogonality.

☞ long recurrences – no CG method

Uniform spectrum, small condition number – the CSI method.

Content of the talk

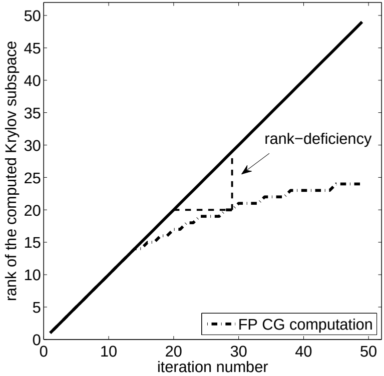
Composite polynomial bounds for CG in finite precision arithmetic

Krylov subspaces generated by CG in finite precision arithmetic

Idea of shift

We relate: k-th iteration of FP CG $\Longleftrightarrow$ l-th iteration of exact CG

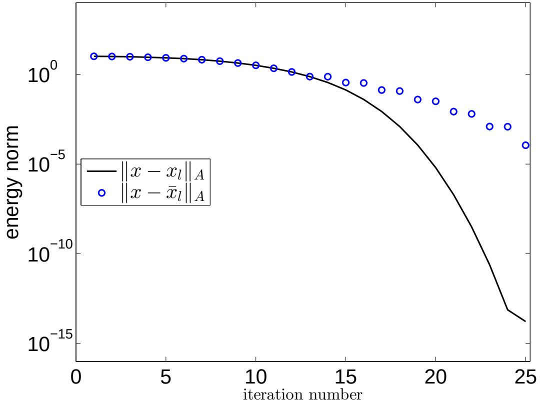
- $k - l \quad \approx \quad$ delay of convergence
- $k - l \quad \approx \quad$ rank-deficiency of computed Krylov subspace



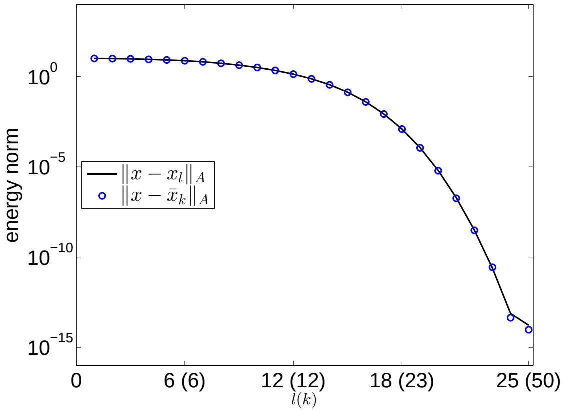
We want to study:

$$
\begin{array}{ccc}
\|x - \overline{x}_k\|_A & \times & \|x - x_l\|_A \\
\overline{x}_k & \times & x_l \\
\overline{\mathcal{K}}_k(A, r_0) & \times & \mathcal{K}_l(A, r_0)
\end{array}
$$

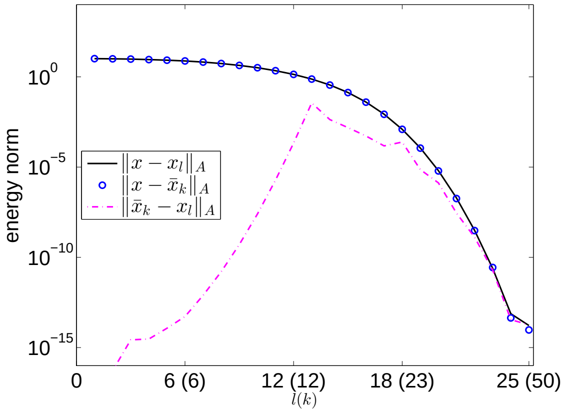
Comparison of trajectory of approximation vectors



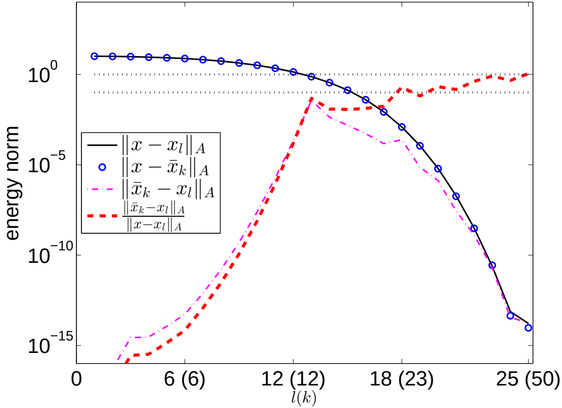
Comparison of trajectory of approximation vectors



Comparison of trajectory of approximation vectors



Comparison of trajectory of approximation vectors

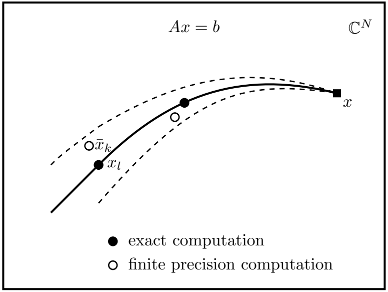


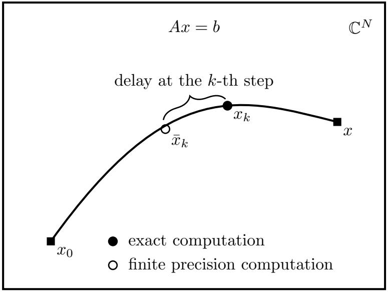
Observation

$$\frac{\|\bar{x}_k - x_l\|_A}{\|x - x_l\|_A} \ll 1$$

Trajectories of approximation vectors are very similar in space $\mathbb{C}^N$.

Comparison of trajectory of approximation vectors





Trajectory of approximations $\overline{x}_k$ generated by FP CG computations follows closely the trajectory of the exact CG approximations x_l.

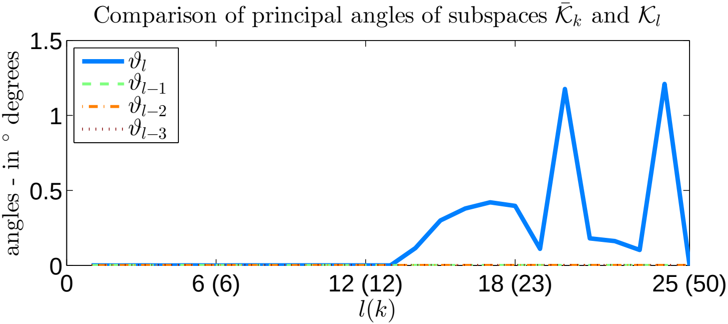
Principal angles and vectors

$$\vartheta_j = \min_{\substack{p\in\mathcal{F}_j\\ \|p\|=1}} \min_{\substack{q\in\mathcal{G}_j\\ \|q\|=1}} \arccos(\, p^* q\,) \equiv \arccos(\, p_j{}^* q_j\,)\,, \quad j = 1, 2, \ldots, l$$

where

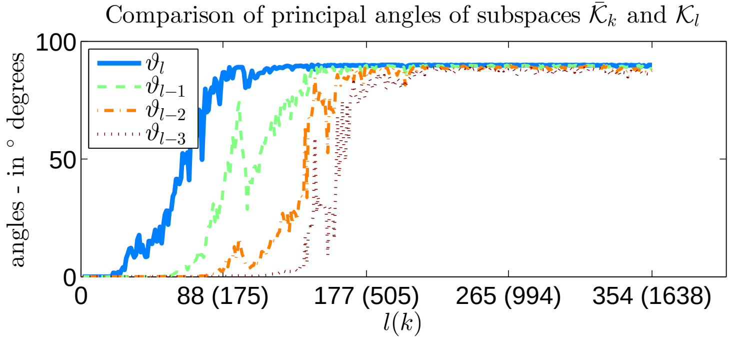
$$\mathcal{F}_j \equiv \mathcal{F} \cap \{p_1, \ldots, p_{j-1}\}^{\perp}\,, \qquad \mathcal{G}_j \equiv \mathcal{G} \cap \{q_1, \ldots, q_{j-1}\}^{\perp}\,,$$

$$\mathcal{F} = \overline{\mathcal{K}}_k(A, r_0), \qquad \mathcal{G} = \mathcal{K}_l(A, r_0).$$



Departure of subspaces

For more difficult problems, the subspaces can depart in few directions.



[data: bus494 from `MatrixMarket`]

Summary II and outlook

- ☞ The convergence rate of finite precision CG and exact CG typically significantly differs.
- ☞ The trajectories of computed approximations are enclosed in a shrinking "cone".
- ☞ Apart from the delay, the computed Krylov subspaces do not depart much from their exact arithmetic counterparts.

- ▶ Study further the properties of principal vectors, find relationship to the structure of invariant subspaces.
- ▶ Study the effect of clustered eigenvalues.

Acknowledgement

This work has been supported by the ERC-CZ project LL1202, by the GACR grant 201/13-06684S and by the GAUK grant 695612.

References

- T. Gergelits and Z. Strakoš, *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations*, Numer. Algorithms, 65 (2014), pp. 759–782.
- J. Liesen and Z. Strakoš, *Krylov Subspace Methods: Principles and Analysis*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.
- J. Málek and Z. Strakoš, *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs*, SIAM Spotlight Series, 2015.