Properties of the CG method in finite precision arithmetic

Tomáš Gergelits, Zdeněk Strakoš

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

CIME Summer School 2015, Cetraro

23 June 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

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

$$\mathcal{K}_k(A, r_0) = \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$$

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

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

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

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

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

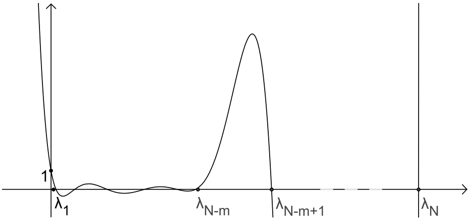
[Rutishauser (1959)]

In the case of m large outlying eigenvalues the composite polynomial

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

gives for $k \geq m$

$$\frac{\|x - x_k\|_A}{\|x - x_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_m}-1}{\sqrt{\kappa_m}+1}\right)^{k-m}$$
$$\kappa_m = \frac{\lambda_{N-m}}{\lambda_1}.$$

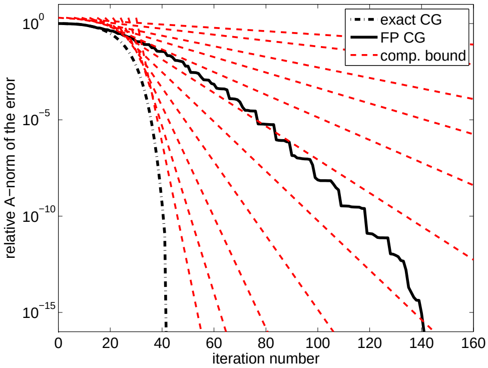


[Axelsson (1976), Jennings (1977); cf. van der Sluis, van der Vorst (1986)]

CG in finite precision arithmetic

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

Failure of the composite polynomial bound



Summary I

Points to consider:

- ☞ short recurrences – loss of orthogonality.
- ☞ long recurrences – no CG method

Linear convergence, small condition number – then why CG? 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

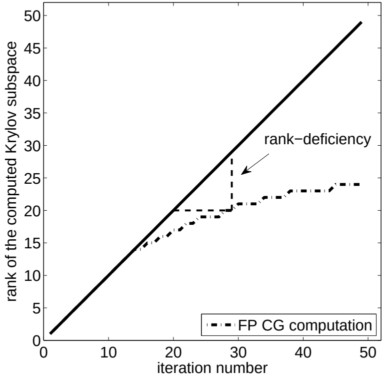
Idea of shift

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

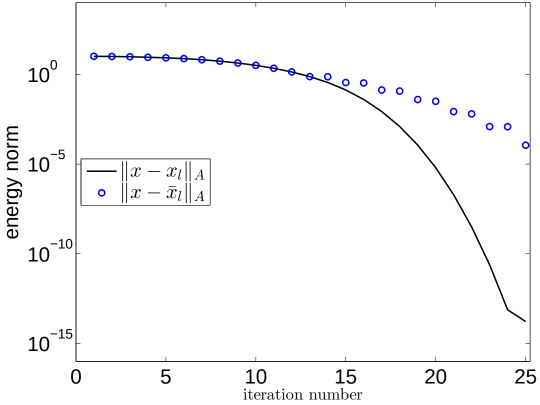
- $k - \ell \quad \approx \quad$ delay of convergence
- $k - \ell \quad \approx \quad$ rank-deficiency of computed Krylov subspace

We want to study:

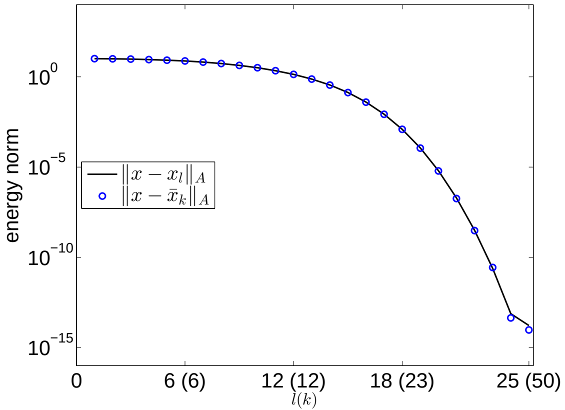
$$\|x - \overline{x}_k\|_A \quad \times \quad \|x - x_\ell\|_A$$
$$\overline{x}_k \quad \times \quad x_\ell$$
$$\overline{\mathcal{K}}_k(A, r_0) \quad \times \quad \mathcal{K}_\ell(A, r_0)$$



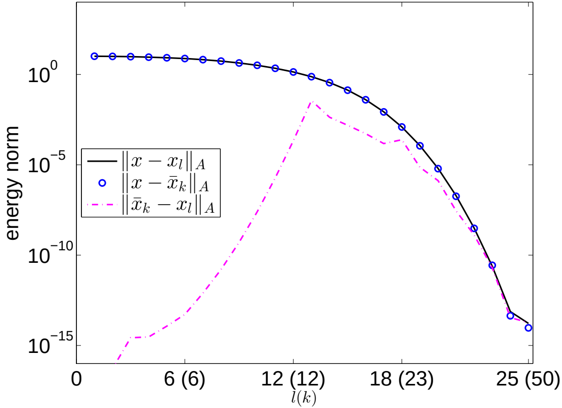
Comparison of trajectory of approximation vectors

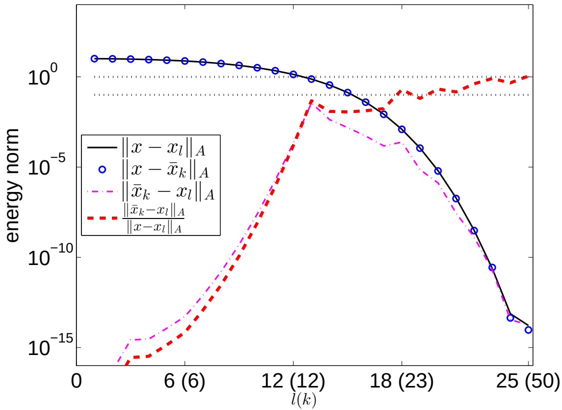


Comparison of trajectory of approximation vectors



Comparison of trajectory of approximation vectors



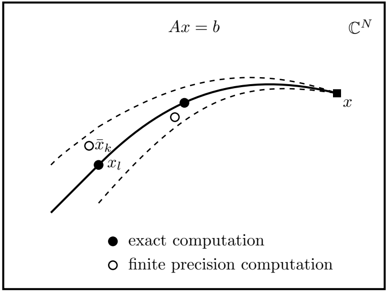


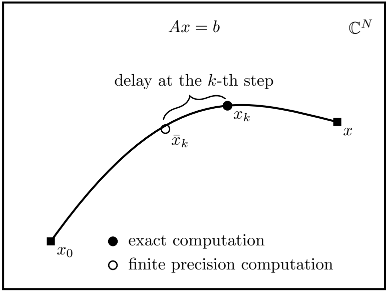
Observation

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

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

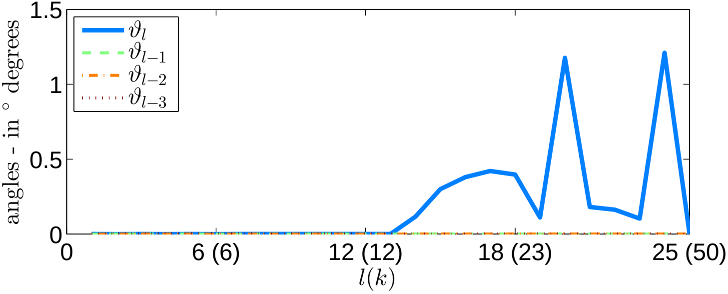
$$\vartheta_j = \min_{\substack{p \in \mathcal{F}_j \\ \|p\|=1}} \min_{\substack{q \in \mathcal{G}_j \\ \|q\|=1}} \arccos\left(p^* q \right) \equiv \arccos\left(p_j{}^* q_j \right), \quad j = 1, 2, \dots, \ell$$

where

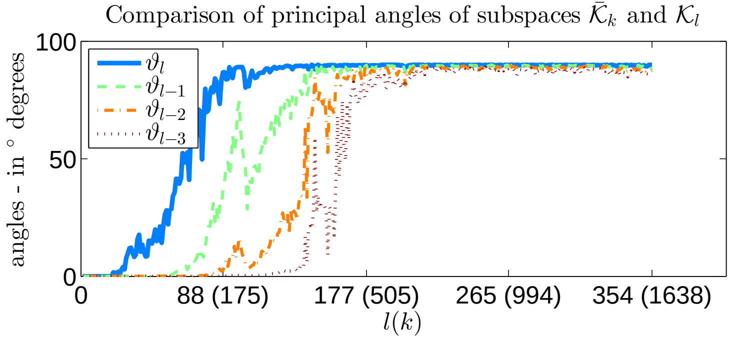
$$\mathcal{F}_j \equiv \mathcal{F} \cap \{p_1, \dots, p_{j-1}\}^\perp, \qquad \mathcal{G}_j \equiv \mathcal{G} \cap \{q_1, \dots, q_{j-1}\}^\perp,$$

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

Comparison of principal angles of subspaces $\bar{\mathcal{K}}_k$ and $\mathcal{K}_l$



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



[data: bus494 from `MatrixMarket`]

Summary II

- ☞ The convergence rate of finite precision CG and exact CG typically significantly differs. When there is no delay, then other methods can be competitive or even outperform CG computations.
- ☞ 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.

Outlook

- properties of principal vectors, relationship to the structure of invariant subspaces.
- analogous behaviour in other Krylov subspace methods based on short recurrences?

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.

Acknowledgement

This work has been supported by the ERC-CZ project LL1202 and by the GAUK grant 172915.