Nečas Center for Mathematical Modeling

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

MINISTRY OF EDUCATION, YOUTH AND SPORTS

On the properties of Krylov subspaces in finite precision CG computations

Tomáš Gergelits* <gergelits@karlin.mff.cuni.cz> **Zdeněk Strakoš*,†** <strakos@karlin.mff.cuni.cz>

*Faculty of Mathematics and Physics, Charles University in Prague

†Institute of Computer Science, Academy of Sciences of the Czech Republic

Introduction

We address the question of difference between Krylov subspaces generated by the CG method in finite precision arithmetic and their exact arithmetic counterparts. Since we study the behaviour of CG in practical computations, we concentrate on situation with significant delay of convergence. We observe that apart from the delay the computed Krylov subspaces do not depart much from their exact arithmetic counterparts.

Krylov subspaces in practical computations

The need of computing the basis of the Krylov subspace

$$\mathcal{K}_l(B,v) = \text{span}\{v, Bv, \dots, B^{l-1}v\}$$

is in the nutshell of CG and other Krylov subspace methods. However, due to rounding errors, the subspace spanned by the practically computed basis may differ and the following questions arise:

- What is the difference between the computed Krylov subspace $\overline{\mathcal{K}}_l(B,v)$ and its exact arithmetic counterpart $\mathcal{K}_l(B,v)$?
- Can we find perturbations $\Delta B, \delta v$ such that $\mathcal{K}_l(B+\Delta B, v+\delta v) = \overline{\mathcal{K}}_l(B,v)$?
- How much are the Krylov subspaces $\mathcal{K}_l(B+\Delta B, v+\delta v)$ sensitive to general small perturbations?

Results in literature, e.g., [1, 2, 3], rely on the assumption of full dimensionality of studied (computed or perturbed) Krylov subspaces.

The CG method in practical computations

The CG method is the method of choice for solving linear systems

$$Ax = b, \quad A \in \mathbb{F}^{N\times N} \text{ HPD}, \quad b \in \mathbb{F}^N, \quad \mathbb{F} \text{ is } \mathbb{R} \text{ or } \mathbb{C}.$$

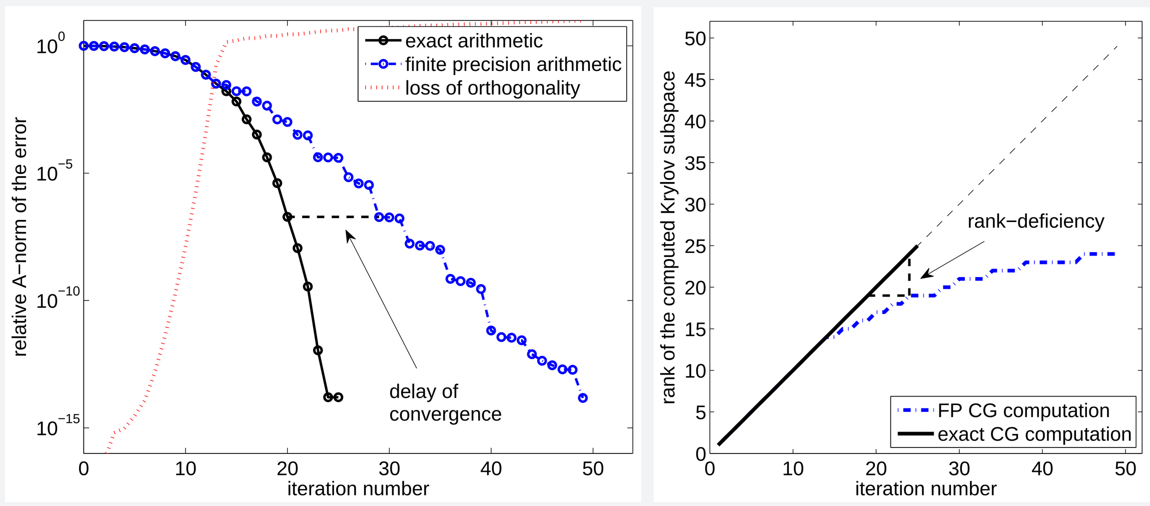
- CG is a projection method which minimizes the energy norm of the error

$$x_l \in x_0 + \mathcal{K}_l(A, r_0), \quad r_l \perp \mathcal{K}_l(A, r_0); \qquad \|x - x_l\|_A = \min\{\|x-y\|_A : \; y \in x_0 + \mathcal{K}_l(A, r_0)\}.$$

- CG is computationally based on short recurrences.

Delay of convergence & rank deficiency

Using short recurrences in practical computation leads to the loss of global orthogonality of the computed residuals and even to the loss of their linear independence. Consequently, the computed Krylov subspace spanned by these vectors can be rank-deficient which causes significant delay of convergence.



Idea of shift

Taking into the account the phenomenon of delay of convergence, we should compare different iterations when comparing CG computations in finite precision and exact arithmetic. We relate:

$$k\text{-th iteration of FP CG} \iff l\text{-th iteration of exact CG} \quad \text{where}$$

$k - l$ corresponds to delay of convergence or rank-deficiency of computed Krylov subspace.
We want to study:

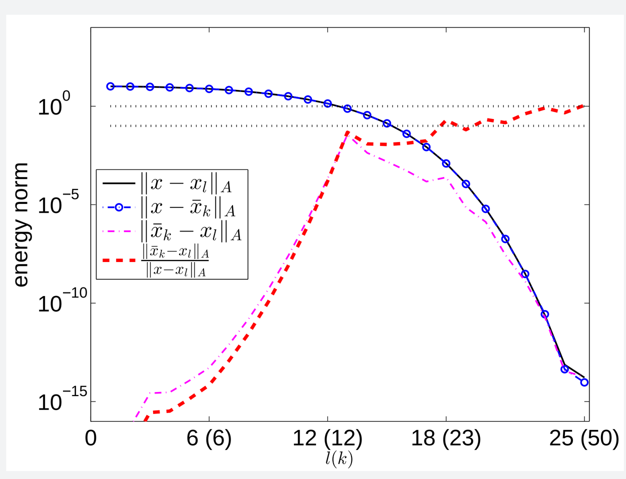
$$\|x - \bar{x}_k\|_A \quad \times \quad \|x - x_l\|_A$$
$$\bar{x}_k \quad \times \quad x_l$$
$$\overline{\mathcal{K}}_k(A, r_0) \quad \times \quad \mathcal{K}_l(A, r_0).$$

Correspondence among computed and exact approximation vectors

We see that

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

i.e., the distance between the exact and shifted FP approximations is small in comparison with the actual size of error.



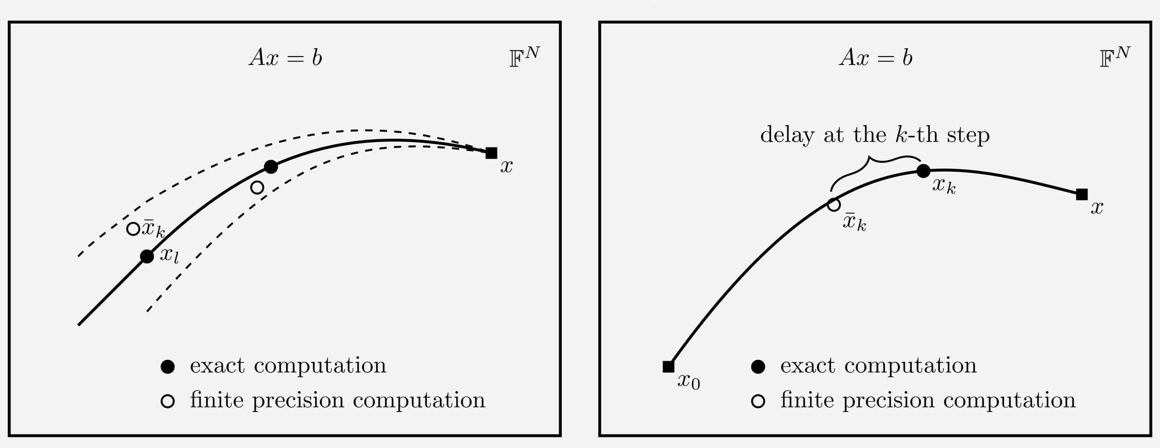


Figure : Trajectory of approximations $\bar{x}_k$ generated by FP CG computations follows closely the trajectory of the exact CG approximations x_l with a delay given by the rank-deficiency of the computed Krylov subspace.

Correspondence between computed and exact Krylov subspaces

The distance between subspaces is measured using principal angles ϑ_j and vectors p_j, q_j defined as

$$\vartheta_j = \min_{\substack{p\in\mathcal{F}_j\\ \|p\|=1}} \min_{\substack{q\in\mathcal{G}_j\\ \|q\|=1}} \arccos(p^*q) \equiv \arccos(p_j^* q_j) \quad \text{where} \quad \begin{array}{l}\mathcal{F}_j \equiv \overline{\mathcal{K}}_k(A,r_0) \cap \{p_1,\dots,p_{j-1}\}^\perp,\\ \mathcal{G}_j \equiv \mathcal{K}_l(A,r_0) \cap \{q_1,\dots,q_{j-1}\}^\perp.\end{array}$$

We compute principal angles and vectors via the computation of the SVD decomposition of the matrix $U_l^* V_l$

$$U_l^* V_l = F \Sigma G^*,$$

where U_l is the orthogonal basis of the l-dimensional restriction of $\overline{\mathcal{K}}_k(A, r_0)$ and V_l is an orthogonal basis of $\mathcal{K}_l(A, r_0)$. It holds that:

$$[p_1, \dots, p_l] \equiv P = U_l F, \quad p_j = U_l f_j,$$
$$[q_1, \dots, q_l] \equiv Q = V_l G, \quad q_j = V_l g_j,$$
$$\text{diag}(\cos(\vartheta_1), \dots, \cos(\vartheta_l)) \equiv \Sigma.$$

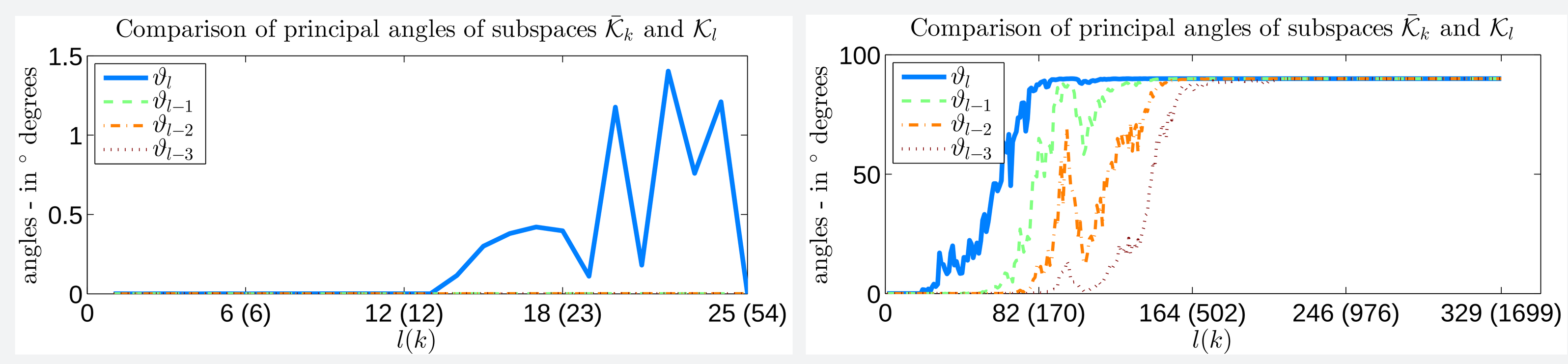


Figure : Left: In this numerical experiment, the largest canonical angle is $\approx 1^\circ$ and thus Krylov subspaces $\mathcal{K}_l(A, r_0)$ and $\overline{\mathcal{K}}_k(A, r_0)$ are nearly the same. **Right:** Things can be more complicated, as illustrated on experiment with data **Bus 494** from the MatrixMarket database. The departure of subspaces is, however, still only in few directions.

Study of departure of Krylov subspaces

Influence of clustered eigenvalues
We have observed that the quality of closeness of generated Krylov subspaces depends on the possible presence of clustered eigenvalues. The tighter cluster is, the more severe is departure of subspaces.

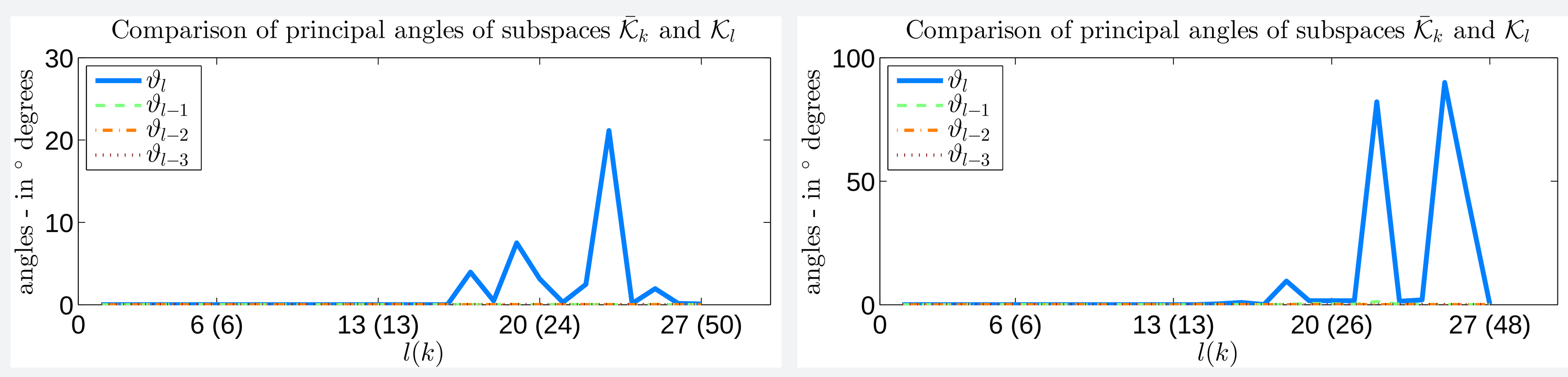


Figure : Illustration of the influence of clustered eigenvalues. We plot the largest principal angles for two different settings. Five largest eigenvalues are clustered in the interval of length $\Delta = 10^{-8}$ (left) and $\Delta = 10^{-12}$ (right).

Similar phenomenon of the loss and recapture of correlation between Arnoldi vectors computed by two implementations of the Arnoldi algorithm in finite precision arithmetic was observed in [4].

Concluding remarks

- The trajectories of computed approximations are enclosed in a shrinking "cone".
- Krylov subspaces are in general sensitive to small perturbations of the matrix A. The observed "stability" (or *inertia*) of computed Krylov subspace represents phenomenon which needs further investigation.
- There is principle difference in analysis between short and long recurrences. Using short recurrences we can not guarantee that the computed basis is well conditioned and that the computed subspaces have full dimension.

Acknowledgement

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

Bibliography

Carpraux, J.-F. and Godunov, S. K. and Kuznetsov, S. V. (1996). Condition number of the Krylov bases and subspaces. *Linear Algebra and its Applications 265*, 1–28.

Paige, C. C. and Van Dooren, P. (1999). Sensitivity analysis of the Lanczos reduction. *Numerical Linear Algebra with Applications 6*, 29–50.

Stewart, G. W. (2002). Backward error bounds for approximate Krylov subspaces. *Linear Algebra and its Applications 340*, 81–86.

Okša, G. and Rozložník, M. (2011). On numerical behavior of the Arnoldi algorithm in finite precision for matrices with close eigenvalues. In *SNA'11: Seminar on Numerical Analysis 2011*, Institute of Geonics AS CR, Ostrava, 88–91.

http://more.karlin.mff.cuni.cz