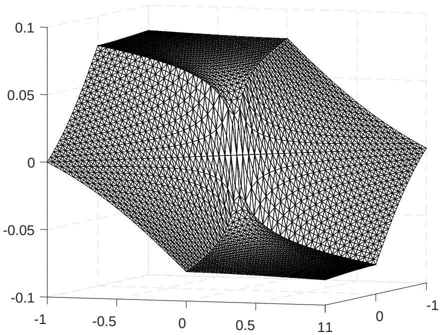
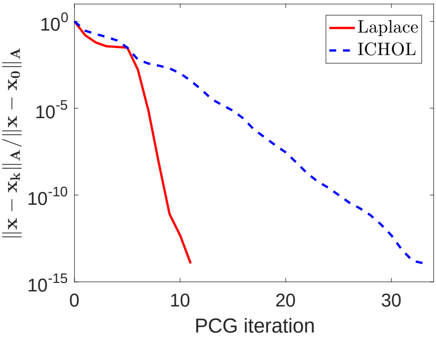
The perfidious condition number

Zdeněk Strakoš
Charles University, Prague
Jindřich Nečas Center for Mathematical Modelling

Manchester, May 2019

The class of elliptic PDEs, an example



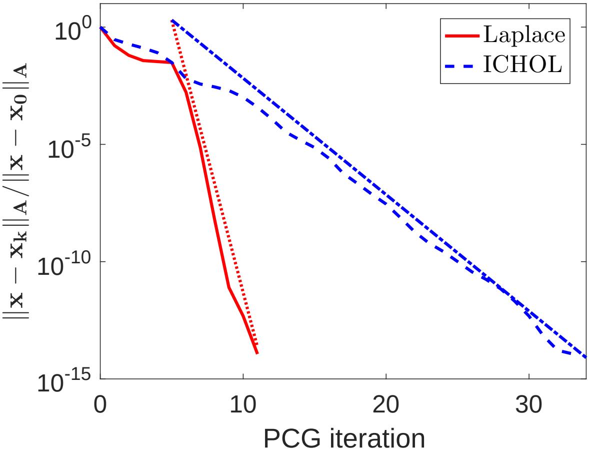


$$-\nabla \cdot (\, k(x)\, \nabla u \,) \;=\; 0\,,$$

Morin, Nocheto, Siebert, SIREV (2002),
linear FE, standard uniform triangulation, $N = 3969$ DOF.

ICHOL PCG (drop-off tolerance 1e-02), $\kappa \approx 16$;
Laplace operator PCG, $\kappa \approx 160$.

Predicting the computational cost?



Polynomial (Krylov subspace) methods

Consider a linear bounded invertible operator $\mathcal{G} : S \to S$ and the equation on Hilbert space

$$\mathcal{G}\, u \,=\, f\,, \quad f \in S\,.$$

(Infinite dimensional) Krylov subspace methods at step n implicitly construct a finite dimensional approximation $\mathcal{G}_n$ of $\mathcal{G}$ with the desired approximate solution u_n defined by ($u_0 = 0$)

$$u_n \,:=\, p_{n-1}(\mathcal{G}_n)\, f \;\approx\; u \,=\, \mathcal{G}^{-1} f\,,$$

where $p_{n-1}(\lambda)$ is the associated polynomial of degree at most $n-1$ and $\mathcal{G}_n$ is obtained by restricting and projecting $\mathcal{G}$ onto the nth Krylov subspace

$$\mathcal{K}_n(\mathcal{G}, f) \,:=\, \operatorname{span}\left\{ f, \mathcal{G}f, \ldots, \mathcal{G}^{n-1} f \right\}.$$

A.N. Krylov (1931), Gantmakher (1934), Hestenes and Stiefel (1952), Lanczos (1952-53); Karush (1952), Hayes (1954), Stesin (1954), Vorobyev (1958).

Integral representation of self-adjoint operators on Hilbert spaces

- Finite dimensional self-adjoint operators (finite Hermitian matrices)

$$
\begin{aligned}
\mathcal{G} &= \frac{1}{2\pi\iota}\int_{\Gamma}\lambda\left(\lambda I_N-\mathcal{G}\right)^{-1}d\lambda \;=\; \frac{1}{2\pi\iota}\sum_{\ell=1}^{N}\int_{\Gamma_\ell}\lambda\left(\lambda I_N-\mathcal{G}\right)^{-1}d\lambda \\
&= \sum_{\ell=1}^{N} Y\,\mathrm{diag}\left(\frac{1}{2\pi\iota}\int_{\Gamma_\ell}\frac{\lambda}{\lambda-\lambda_\ell}\,d\lambda\right)Y^* \;=\; \sum_{\ell=1}^{N}\lambda_\ell\, y_\ell y_\ell^* \\
&= \int \lambda\, dE(\lambda)\,.
\end{aligned}
$$

- Compact infinite dimensional self-adjoint operators
- Bounded infinite dimensional self-adjoint operators
- Generalization to bounded normal and non-normal operators

Using the spectral decomposition $\mathcal{G} = \sum_{\ell=1}^{N} \lambda_\ell\, y_\ell y_\ell^*$

Symbolically

$$w_1^* \mathcal{G}\, w_1 = w_1^* \left(\sum_{\ell=1}^{N} \lambda_\ell\, y_\ell y_\ell^* \right) w_1 \equiv w_1^* \left(\int \lambda\, dE(\lambda) \right) w_1$$
$$= \sum_{\ell=1}^{N} \lambda_\ell \left|(y_\ell, w_1)\right|^2 = \sum_{\ell=1}^{N} \lambda_\ell\, \omega_\ell = \int \lambda\, d\omega(\lambda)\,,$$

where the spectral function $E(\lambda)$ of $\mathcal{G}$ is understood to be a nondecreasing family of projections with increasing λ, symbolically $dE(\lambda_\ell) \equiv y_\ell y_\ell^*$ and

$$I = \sum_{\ell=1}^{N} y_\ell y_\ell^* \equiv \int dE(\lambda)\,.$$

Hilbert (1906, 1912, 1928), Von Neumann (1927, 1932), Wintner (1929) .

CG and Gauss quadrature relationship - the size of errors are equal!

At any iteration step n, CG represents the matrix formulation of the n-point Gauss quadrature of the Riemann-Stieljes integral determined by A and r_0,

$$\int_0^\infty \phi(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \phi(\theta_i^{(n)}) \;+\; R_n(\phi)\,.$$

For the function $\phi(\lambda) \equiv \lambda^{-1}$,

$$\frac{\|x - x_0\|_A^2}{\|r_0\|^2} \;=\; n\text{-th Gauss quadrature} \;+\; \frac{\|x - x_n\|_A^2}{\|r_0\|^2}\,.$$

This has become the basis for CG error estimation; see Golub, 1994; and, e.g., the surveys in S and Tichý, 2002; Meurant and S, 2006; Golub and Meurant, 2010; Liesen and S, 2013.

Condition number and contraction bound(s)??

$$\frac{\|x - x_n\|_A}{\|x - x_0\|_A} \ \leq \ 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^n$$

Condition number and contraction bound(s)??

$$\frac{\|x - x_n\|_A}{\|x - x_0\|_A} \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^n$$

- Rutishauser (1959) as well as Lanczos (1952) considered CG principally different in nature from the method based on Chebyshev polynomials.

Condition number and contraction bound(s)??

$$\frac{\|x - x_n\|_A}{\|x - x_0\|_A} \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^n$$

- Rutishauser (1959) as well as Lanczos (1952) considered CG principally different in nature from the method based on Chebyshev polynomials.
- Daniel (1967) did not identify CG convergence with the Chebyshev polynomials-based bound. He carefully writes (modifying slightly his notation)

  " . . . assuming only that the spectrum of the matrix A lies inside the interval $[\lambda_1, \lambda_N]$, we can do no better than Theorem 1.2.2." [see the bound above]

Condition number and contraction bound(s)??

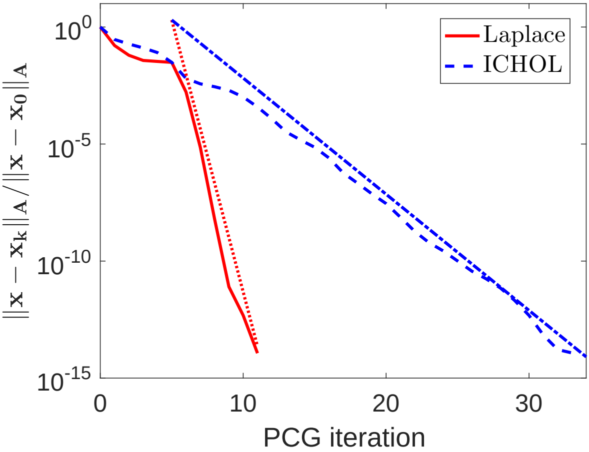
$$\frac{\|x - x_n\|_A}{\|x - x_0\|_A} \le 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^n$$

- Rutishauser (1959) as well as Lanczos (1952) considered CG principally different in nature from the method based on Chebyshev polynomials.
- Daniel (1967) did not identify CG convergence with the Chebyshev polynomials-based bound. He carefully writes (modifying slightly his notation)

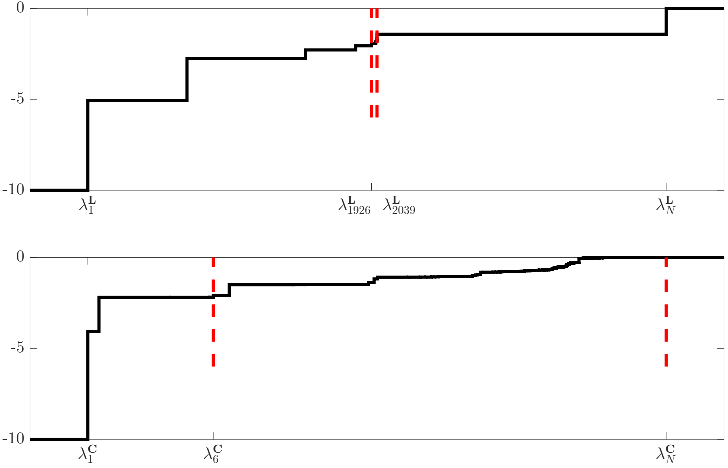
  " ... assuming only that the spectrum of the matrix A lies inside the interval $[\lambda_1, \lambda_N]$, we can do no better than Theorem 1.2.2." [see the bound above]

  That means that the Chebyshev polynomials-based bound holds for *any* distribution of eigenvalues between λ_1 and λ_N and for *any* distribution of the components of the initial residuals in the individual invariant subspaces.

Back to the elliptic PDE problem example



Various parts of the spectra and convergence behavior



Information on the whole spectrum is available a priori!

Theorem : (Pairing the eigenvalues and the FE discretization nodal values $k_j, \;\; j = 1, \ldots, N$**)**

Let $0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$ be the eigenvalues of $\mathbf{L}^{-1}\mathbf{A}$. Let $k(x)$ be bounded and piecewise continuous. Then there exists a (possibly non-unique) permutation π such that the eigenvalues above satisfy

$$\lambda_{\pi(\mathbf{j})} \;\approx\; \mathbf{k_j}\,, \quad \mathbf{j} = \mathbf{1}, \ldots, \mathbf{N}\,.$$

Gergelits, Mardal, Nielsen, S, Laplacian preconditioning of elliptic PDEs: localization of the eigenvalues of the discretized operator, SIAM J. Numer. Anal, 2019.

Here we assume for simplicity a Lagrange FE discretization. The results of the paper are more general.

Clusters of eigenvalues?

Clusters of eigenvalues?

- It is not true that CG (or other Krylov subspace methods used for solving systems of linear algebraic equations with symmetric matrices) applied to a matrix with t distinct well separated tight clusters of eigenvalues produces in general a large error reduction after t steps; see Sections 5.6.5 and 5.9.1 in Liesen, S (2013). The associated myth has been proved false more than 25 years ago; see Greenbaum (1989); S (1991); Greenbaum, S (1992). Still it is persistently repeated in the literature as an obvious fact.

Clusters of eigenvalues?

- It is not true that CG (or other Krylov subspace methods used for solving systems of linear algebraic equations with symmetric matrices) applied to a matrix with t distinct well separated tight clusters of eigenvalues produces in general a large error reduction after t steps; see Sections 5.6.5 and 5.9.1 in Liesen, S (2013). The associated myth has been proved false more than 25 years ago; see Greenbaum (1989); S (1991); Greenbaum, S (1992). Still it is persistently repeated in the literature as an obvious fact.

- As for systems with nonsymmetric matrices, without specific (strong) assumptions on the structure of invariant subspaces it can not be claimed that distribution of eigenvalues provides insight into the asymptotic behavior of Krylov subspace methods (such as GMRES); see Sections 5.7.4, 5.7.6 and 5.11 of Liesen, S (2013). As above, the relevant results Greenbaum, S (1994); Greenbaum, Pták, S (1996) and Arioli, Pták, S (1998) are more than 20 years old.

Concluding remarks and outlook

- Krylov subspace methods adapt to the problem. Exploiting this adaptation is the key to their efficient use.
- Unlike nonlinear problems and/or multilevel methods, the analysis of Krylov subspace methods can not be based, in general, on contraction arguments.
- Individual steps modeling-analysis-discretization-computation should not be considered separately within isolated disciplines. They form a single problem. Operator preconditioning follows this philosophy.
- HPC computations require handling all involved issues.
  A posteriori error analysis and stopping criteria are essential ...
- Assumptions must be honored.

Wilkinson, AEP, Chapter 2, Section 28 (1965)

"We shall describe any computing problem as 'ill conditioned' if the values to be computed are very sensitive to small changes in the data."

Wilkinson, AEP, Chapter 2, Section 28 (1965)

"We shall describe any computing problem as 'ill conditioned' if the values to be computed are very sensitive to small changes in the data."

cf. Forsythe, BAMS, 59 (1953):

"There is a great need for clarification of the group of ideas associated with 'condition.' With the concept of 'ill-conditioned' systems $Ax = b$ *goes the idea of preconditioning them. "*

Wilkinson, AEP, Chapter 2, Section 28 (1965)

"We shall describe any computing problem as 'ill conditioned' if the values to be computed are very sensitive to small changes in the data."

cf. Forsythe, BAMS, 59 (1953):

"There is a great need for clarification of the group of ideas associated with 'condition.' With the concept of 'ill-conditioned' systems $Ax = b$ goes the idea of preconditioning them. "

Málek, S, *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs,* SIAM Spotlights, SIAM (2015), Chapter 13:

"Here we do not argue against using condition numbers ... where appropriate. We argue against using them as general unquestioned tools which are considered fully descriptive ... as arguments closing the door for further investigation."

Lanczos, Why Mathematics (1966)

"We will go on pondering and meditating, the great mysteries still ahead of us, we will err and stumble on the way, and if we win a little victory, we will be jubilant and thankful, without claiming, however, that we have done something that can eliminate the contribution of all the millenia before us."

Wallis, Arithmetica Infinitorium (1656) (see Khruschev 2008)

"There remains this: we beech the skilled in these things, that we thought worth showing, they will think openly receiving, and whatever it hides, worth imparting more properly by themselves to the wider mathematical community."

Thank you very much for your kind patience!