Operator Preconditioning, Discretization and Decomposition into Subspaces

Zdeněk Strakoš

Charles University, Prague

PANM Hejnice, June 27, 2018

Many thanks

to the coauthors

Tomáš Gergelits, Jakub Hrnčíř, Josef Málek, Jan Papež, Ivana Pultarová.

Discussions with Barbara Wohlmuth and Uli Rüde have motivated the part of the work related to decompositions into subspaces.

Many thanks

to the coauthors

Tomáš Gergelits, Jakub Hrnčíř, Josef Málek, Jan Papež, Ivana Pultarová.

Discussions with Barbara Wohlmuth and Uli Rüde have motivated the part of the work related to decompositions into subspaces.

I am in a big debt to very many collaborators and friends at home and abroad. In this talk I will touch results of only very few of them. I wish to particularly mention three distinguished mathematicians and remarkable humans whom I had the privilege to meet in my life and who have influenced the way I do my work.

Many thanks

to the coauthors

Tomáš Gergelits, Jakub Hrnčíř, Josef Málek, Jan Papež, Ivana Pultarová.

Discussions with Barbara Wohlmuth and Uli Rüde have motivated the part of the work related to decompositions into subspaces.

I am in a big debt to very many collaborators and friends at home and abroad. In this talk I will touch results of only very few of them. I wish to particularly mention three distinguished mathematicians and remarkable humans whom I had the privilege to meet in my life and who have influenced the way I do my work.

I would like to devote this lecture to the memory of

Gene H. Golub, Ivo Marek and Vlastimil Pták.

Motto - interconnections, insight and understanding

Cornelius Lanczos, *Why Mathematics,* 1966

“In a recent comment on mathematical preparation an educator wanted to characterize our backwardness by the following statement: ”Is it not astonishing that a person graduating in mathematics today knows hardly more than what Euler knew already at the end of the eighteenth century?”. On its face value this sounds a convincing argument. Yet it misses the point completely. Personally I would not hesitate not only to graduate with first class honors, but to give the Ph.D. (and with summa cum laude) without asking any further questions, to anybody who knew only one quarter of what Euler knew, provided that he knew it in the way in which Euler knew it.”

Operator preconditioning ... a second thought?

Two basic questions arise when we look at the existing very extensive literature on preconditioning:

Operator preconditioning ... a second thought?

Two basic questions arise when we look at the existing very extensive literature on preconditioning:

- Why we do not have much written on an analytic theory of preconditioning? Perhaps because the problem is difficult?

Operator preconditioning ... a second thought?

Two basic questions arise when we look at the existing very extensive literature on preconditioning:

- Why we do not have much written on an analytic theory of preconditioning? Perhaps because the problem is difficult?
- Why, at the same time, from the frequently published claims on clustering eigenvalues etc. it looks like there is such a theory?

Operator preconditioning ... a second thought?

Two basic questions arise when we look at the existing very extensive literature on preconditioning:

- Why we do not have much written on an analytic theory of preconditioning? Perhaps because the problem is difficult?
- Why, at the same time, from the frequently published claims on clustering eigenvalues etc. it looks like there is such a theory?

Preconditioning is linked with iterative solution of large scale problems.
Here we will consider Krylov subspace methods, in particular the method of conjugate gradients (CG).

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

1 Hierarchy of linear problems starting at infinite dimension

Problem with bounded invertible operator $\mathcal{G}$ on the infinite dim. Hilbert space S

$$\mathcal{G}\, u \;=\; f$$

is approximated on a finite dimensional subspace $S_h \subset S$ by the problem with the finite dimensional operator

$$\mathcal{G}_h\, u_h \;=\; f_h\,,$$

represented, using an appropriate basis of S_h , by the (sparse?) matrix problem

$$\mathbf{A}\,\mathbf{x} \;=\; \mathbf{b}\,.$$

Bounded invertible operators in Hilbert spaces can be approximated by compact or finite dimensional operators only in the sense of strong convergence (pointwise limit)

$$\|\,\mathcal{G}_h\, w - \mathcal{G}\, w\,\| \;\rightarrow\; 0 \quad \text{as} \quad h \rightarrow 0 \quad \text{for all } w \in S\,;$$

The convergence $\mathcal{G}_h\, w \;\rightarrow\; \mathcal{G}\, w$ is not uniform w.r.t. w; the role of right hand sides.

1 Fundamental theorem of discretization of $\mathcal{G}\,u = f$

How closely $\mathcal{G}_h\,u_h = f_h$ approximates $\mathcal{G}\,u = f$? The residual measure

$$\mathcal{G}_h\,\pi_h u - f_h$$

gives

$$\pi_h u - u_h = \mathcal{G}_h^{-1}\,(\,\mathcal{G}_h\,\pi_h u - f_h\,).$$

If $\|\mathcal{G}_h^{-1}\|_h$ is bounded from above uniformly in h (the discretization is stable), then consistency

$$\|\,\mathcal{G}_h\,\pi_h u - f_h\,\|_h \;\rightarrow\; 0 \quad \text{as} \quad h \rightarrow 0$$

implies convergence of the discretization scheme

$$\|\,\pi_h u - u_h\,\|_h \;\rightarrow\; 0 \quad \text{as} \quad h \rightarrow 0\,.$$

Additional important point: In computations we only approximate u_h by $u_h^{(n)}$.

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

Consider, as above, a linear bounded invertible operator $\mathcal{G} : S \to S$ and the equation

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

(Infinite dimensional) Krylov subspace methods at the 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

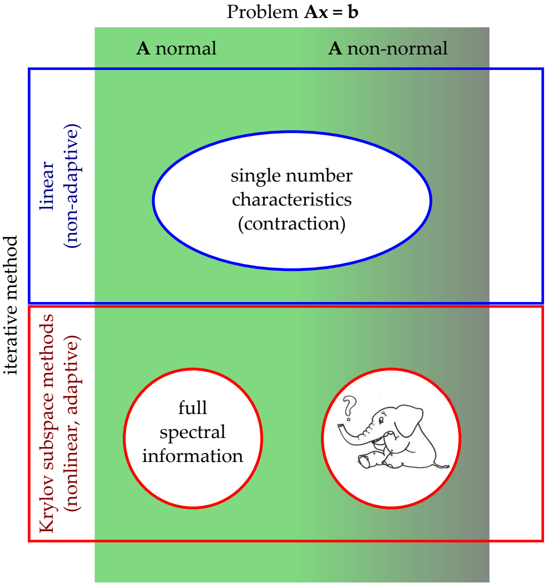
$$\mathcal{K}_n(\mathcal{G}, f) := \text{span}\left\{ f, \mathcal{G}f, \ldots, \mathcal{G}^{n-1} f \right\}.$$

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

2 Four basic questions

1. How fast the iterations u_n, $n = 1, 2, \dots$ approximate the desired solution u ?
   Nonlinear adaptation.
2. Which characteristics of $\mathcal{G}$ and f determine behaviour of the method?
   Inner nature of the problem. What is it?
3. How to handle efficiently discretization and computational issues?
   Provided that $\mathcal{K}_n(\mathcal{G}, f)$ can be computed, the projection provides discretization of the infinite dimensional problem $\mathcal{G}u = f$.
4. How to handle transformation of $\mathcal{G}u = f$ into
   an easier-to-solve problem? Preconditioning.
   This is coupled with the point 2 above.

Vlastimil Pták: Finite dimensional nonlinearity is most difficult.


Problem $\mathbf{A}\mathbf{x} = \mathbf{b}$
A normal
A non-normal
iterative method
linear
(non-adaptive)
single number
characteristics
(contraction)
Krylov subspace methods
(nonlinear, adaptive)
full
spectral
information

- Stationary Richardson (assume **A** HPD)

$$\mathbf{x} - \mathbf{x}_n \;=\; (\mathbf{I} - \omega^{-1}\mathbf{A})^n \, (\mathbf{x} - \mathbf{x}_0)$$

- Chebyshev semiiterative method

$$\mathbf{x} - \mathbf{x}_n \;=\; \frac{1}{|\chi_n(0)|}\, \chi_n(\mathbf{A})\,(\mathbf{x} - \mathbf{x}_0)\,, \qquad \frac{1}{|\chi_n(0)|} \le 2 \left(\frac{\sqrt{\kappa(\mathbf{A})} - 1}{\sqrt{\kappa(\mathbf{A})} + 1} \right)^n ;$$

$$\|\chi_n(\mathbf{A})\| \;=\; \max_{\lambda_j} |\chi_n(\lambda_j)| = \max_{\lambda \in [\lambda_1, \lambda_N]} |\chi_n(\lambda)| = 1\,.$$

Here the description of convergence is focused on asymptotic behavior and it is linear!

- Analogous reasoning is used in solving nonlinear equations (optimization).

2 The largely ignored arguments

But recall Pták, *What should be a rate of convergence?*, RAIRO Anal. Numér 11 (1977); (also Liesen (2014):)

a method of estimating the convergence of iterative processes *"should describe accurately in particular the initial stage of the process, not only its asymptotic behavior, since, after all, we are interested in keeping the number of steps necessary to obtain a good estimate as low as possible."*

But recall Pták, *What should be a rate of convergence?*, RAIRO Anal. Numér 11 (1977); (also Liesen (2014):)

a method of estimating the convergence of iterative processes *"should describe accurately in particular the initial stage of the process, not only its asymptotic behavior, since, after all, we are interested in keeping the number of steps necessary to obtain a good estimate as low as possible."*

The arguments are clear and evidence was presented decades ago. Despite this, many works still describe behavior of Krylov subspace methods asymptotically using single number characteristics and/or ignore possibly dramatic effects of rounding errors to computations based on short recurrences, while claiming generality and almost universal impact of the presented results. This includes also recent publications in top periodicals.

2 Conjugate Gradient method (CG) for $\mathbf{A}\mathbf{x} = \mathbf{b}$ with $\mathbf{A}$ HPD (1952)

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0,\ \mathbf{p}_0 = \mathbf{r}_0\,.$ For $n = 1, \ldots, n_{\max}:$

$$
\begin{aligned}
\alpha_{n-1} &= \frac{\mathbf{r}_{n-1}^*\mathbf{r}_{n-1}}{\mathbf{p}_{n-1}^*\mathbf{A}\mathbf{p}_{n-1}} \\
\mathbf{x}_n &= \mathbf{x}_{n-1} + \alpha_{n-1}\mathbf{p}_{n-1}\,, \quad \text{stop when the stopping criterion is satisfied} \\
\mathbf{r}_n &= \mathbf{r}_{n-1} - \alpha_{n-1}\mathbf{A}\mathbf{p}_{n-1} \\
\beta_n &= \frac{\mathbf{r}_n^*\mathbf{r}_n}{\mathbf{r}_{n-1}^*\mathbf{r}_{n-1}} \\
\mathbf{p}_n &= \mathbf{r}_n + \beta_n\mathbf{p}_{n-1}
\end{aligned}
$$

Here α_{n-1} ensures the minimization of $\|\mathbf{x} - \mathbf{x}_n\|_{\mathbf{A}}$ along the line

$$z(\alpha) = \mathbf{x}_{n-1} + \alpha\mathbf{p}_{n-1}\,.$$

Provided that

$$\mathbf{p}_i \perp_{\mathbf{A}} \mathbf{p}_j, \quad i \neq j,$$

the one-dimensional line minimizations at the individual steps 1 to n result in the n-dimensional minimization over the whole shifted Krylov subspace

$$\mathbf{x}_0 + \mathcal{K}_n(\mathbf{A}, \mathbf{r}_0) \;=\; \mathbf{x}_0 + \mathrm{span}\{\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{n-1}\}\,.$$

Indeed,

$$\mathbf{x} - \mathbf{x}_0 \;=\; \sum_{\ell=0}^{N-1} \alpha_\ell \mathbf{p}_\ell \;=\; \sum_{\ell=0}^{n-1} \alpha_\ell \mathbf{p}_\ell \;+\; \mathbf{x} - \mathbf{x}_n\,,$$

where

$$\mathbf{x} - \mathbf{x}_n \;\perp_{\mathbf{A}}\; K_n(\mathbf{A}, \mathbf{r}_0)\,, \quad \text{or, equivalently,} \quad \mathbf{r}_n \;\perp\; K_n(\mathbf{A}, \mathbf{r}_0)\,.$$

Mathematically, the orthogonality condition leads to short recurrences due to the relationship to the orthogonal polynomials that define the algebraic residuals and search vectors.

Numerically, rounding errors can completely destroy the orthogonality; the matrices composed of the computed search and residual vectors can be drastically rank-deficient. As a consequence of experimental observations it was believed for several decades that the beautiful mathematical structure of the exact CG (Lanczos) was in practical computations inevitably lost and the finite precision behaviour would remain a mystery.

Crucial question:
Is there any optimality of CG (Lanczos) left in the presence of rounding errors?

$$\mathbf{T}_n = \begin{pmatrix} \gamma_1 & \delta_2 & & & \\ \delta_2 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \delta_n \\ & & & \delta_n & \gamma_n \end{pmatrix}$$

is the Jacobi matrix of the Lanczos process coefficients at step n (that is implicitly present also in CG).

Assuming computations in exact arithmetic,
whenever the bottom element of a normalized eigenvector of $\mathbf{T}_n$ vanishes,
the associated eigenvalue of $\mathbf{T}_n$ closely approximates an eigenvalue of $\mathbf{A}$
and an analogous approximation must exist for $\mathbf{T}_{n+1}, \mathbf{T}_{n+2}$ etc.

The notion of *"deflation"*.

- We no longer have Krylov subspaces defined by the input data.
- Computed residuals are not orthogonal to the generated subspaces, i.e., the Galerkin orthogonality does not hold.
- The structure of Krylov subspace methods as projection processes onto nested subspaces of increasing dimensionality seems to be completely lost.

Is anything preserved?

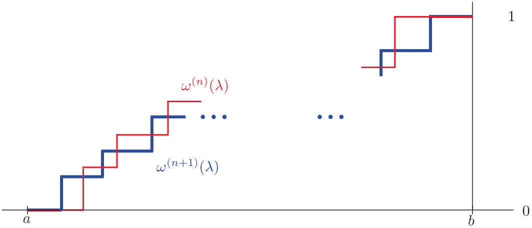
Tool to be used - full spectral information, i.e., spectral decomposition of $\mathbf{A}$ and the projections of $\mathbf{b}/\|\mathbf{b}\|$ $(\mathbf{x}_0 = 0)$ onto the invariant subspaces define the distribution function $\omega(\lambda)$.

- Practical computation generates a sequence of (nested) Jacobi matrices $\mathbf{T}_n$, $n = 1, 2, \ldots$
- Whenever the bottom element of a normalized eigenvector of $\mathbf{T}_n$ vanishes, the associated eigenvalue of $\mathbf{T}_n$ closely approximates an eigenvalue of $\mathbf{A}$ and an analogous approximation must exist for $\mathbf{T}_{n+1}, \mathbf{T}_{n+2}$ etc; see Paige (1971 -1980). This breakthrough result is highly nontrivial.

What distribution function can be associated with the amplification of local roundoff? Greenbaum (1989) gave a beautiful answer. For a given iteration step n the associated distribution function

$$\omega_{1-n}(\lambda)$$

has the points of increase close to the eigenvalues of $\mathbf{A}$, with clusters around those eigenvalues of $\mathbf{A}$ that are closely approximated by several (possibly many) eigenvalues of T_n. The clusters share the weights of the eigenvalues of $\mathbf{A}$.

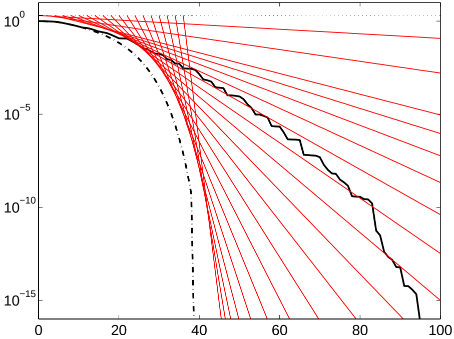


Matching moments model reduction: problem of moments and Gauss quadrature

$$\int \lambda^j \, d\omega(\lambda) \;\rightarrow\; \int \lambda^j \, d\omega_{1-n}(\lambda) \;\approx\; \int \lambda^j \, d\widehat{\omega}(\lambda)\,.$$

Liesen, S, *Krylov Subspace Methods - Principles and Analysis,* OUP (2013)

2 Presence of rounding errors is not resolved by assumptions!



The difference between the dash-dotted and the solid line?

$$\int \lambda^j \, d\omega(\lambda) \;\rightarrow\; \int \lambda^j \, d\omega_{1-n}(\lambda)$$

2 There is no help for those unwilling to listen

Referee report (2005): *"The only new items presented here have to do with analysis involving floating point operations (...). These are likely to bear very little interest to the audience of our Journal*

... the author give a misguided argument. The main advantage of iterative methods over direct methods does not primarily lie in the fact that the iteration can be stopped early (whatever this means), but that their memory (mostly) and computational requirements are moderate.

Referee report (2005): *"The only new items presented here have to do with analysis involving floating point operations (...). These are likely to bear very little interest to the audience of our Journal*

... the author give a misguided argument. The main advantage of iterative methods over direct methods does not primarily lie in the fact that the iteration can be stopped early (whatever this means), but that their memory (mostly) and computational requirements are moderate.

No trace of knowledge of the rock solid evidence and arguments present already by Hestenes and Stiefel (1952), Lanczos(1952-53), . . . , Pták (1977).

2 Opinion can not be separated from scientific evidence

- Opinion in science (and elsewhere) should be based on facts and evidence.
- Hypothesis formulates an opinion which is to be proved or disproved.
- Opinion not declared as hypothesis is without facts and evidence void.

2 Opinion can not be separated from scientific evidence

- Opinion in science (and elsewhere) should be based on facts and evidence.
- Hypothesis formulates an opinion which is to be proved or disproved.
- Opinion not declared as hypothesis is without facts and evidence void.

The following quote was presented by J. Tinsley Oden at the GAMM 2018 in Munich: E. T. Jaynes, Probability Theory, The logic of Science, 2003

"The essence of honesty and objectivity demands that we take into account all the evidence we have, not just some arbitrary subset of it."

2 Opinion can not be separated from scientific evidence

- Opinion in science (and elsewhere) should be based on facts and evidence.
- Hypothesis formulates an opinion which is to be proved or disproved.
- Opinion not declared as hypothesis is without facts and evidence void.

The following quote was presented by J. Tinsley Oden at the GAMM 2018 in Munich: E. T. Jaynes, Probability Theory, The logic of Science, 2003

"The essence of honesty and objectivity demands that we take into account all the evidence we have, not just some arbitrary subset of it."

Tibor Devényi, *Kariéra Dr. Gézy Tamhletoho*, Gondoled, Budapest (1975), Czech translation (1981), Chapter Scientific Discussion:

"We should use arguments, not sabres"

(The fairy-tale writer Lajos Pósa)

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

3 Preconditioning deals with the problem, not with the method

Preconditioning of a linear algebraic system

$$\mathbf{A}\,\mathbf{x} \;=\; \mathbf{b}$$

means its transformation to another system with more favourable properties for its numerical solution. Standard textbook introduction considers $\mathbf{A}$ SPD and takes an SPD matrix $\mathbf{B} \approx \mathbf{A}$ with decomposition $\mathbf{B} = \mathbf{L}\mathbf{L}^*$, giving

$$\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{*\,-1}\,\mathbf{L}^*\mathbf{x} \;=\; \mathbf{L}^{-1}\mathbf{b}\,.$$

In order to technically apply an iterative method (CG) to the transformed system, its algorithm is reformulated in terms of the original variables which is better resembled by

$$\mathbf{B}^{-1}\mathbf{A}\,\mathbf{x} \;=\; \mathbf{B}^{-1}\mathbf{b}\,.$$

3 Reference choice $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2}$

Given SPD matrix $\mathbf{B}$, this schema will work with any decomposition $\mathbf{B} = \mathbf{L}\mathbf{L}^*$. For later convenience, consider the special (reference) choice

$$\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{1/2}.$$

Then for any other decomposition $\mathbf{B} = \mathbf{L}\mathbf{L}^*$ we have

$$\mathbf{L}^{-1}\mathbf{B}\mathbf{L}^{*-1} = (\mathbf{L}^{-1}\mathbf{B}^{1/2})(\mathbf{B}^{1/2}\mathbf{L}^{*-1}) = \mathbf{I},$$

and taking the unitary matrix

$$\mathbf{Q} := \mathbf{L}^{-1}\mathbf{B}^{1/2}, \quad \mathbf{Q}^{-1} = \mathbf{Q}^* = \mathbf{B}^{-1/2}\mathbf{L} = \mathbf{B}^{1/2}\mathbf{L}^{*-1},$$

we have the unitary transformation from $\mathbf{L}$ to $\mathbf{B}^{1/2}$ and vice versa

$$\mathbf{L} = \mathbf{B}^{1/2}\mathbf{Q}^*, \quad \mathbf{B}^{1/2} = \mathbf{L}\mathbf{Q}.$$

3 Which goal should preconditioning target?

Preconditioning is the transformation of the original problem to the form with more favorable properties that allow fast computation. Ok, but not giving a guidance.

3 Which goal should preconditioning target?

Preconditioning is the transformation of the original problem to the form with more favorable properties that allow fast computation. Ok, but not giving a guidance.

- *If the preconditioned matrix has k distinct clusters of eigenvalues, then the backward stability of the algorithm in finite precision arithmetic together with the polynomial convergence bound based on eigenvalues ensures that, computationally, there will be a large error reduction after k steps if the algorithm is applied to the preconditioned system.*

- *Let the eigenvalues of the diagonalizable preconditioned matrix belong into a few clusters, say t of them. If the diameters of the clusters are small enough, then the preconditioned matrix behaves numerically like a matrix with t distinct eigenvalues. As a result, we would expect t iterations of a Krylov subspace method to produce a reasonably accurate approximation.*

3 Which goal should preconditioning target?

Preconditioning is the transformation of the original problem to the form with more favorable properties that allow fast computation. Ok, but not giving a guidance.

- *If the preconditioned matrix has k distinct clusters of eigenvalues, then the backward stability of the algorithm in finite precision arithmetic together with the polynomial convergence bound based on eigenvalues ensures that, computationally, there will be a large error reduction after k steps if the algorithm is applied to the preconditioned system.*
- *Let the eigenvalues of the diagonalizable preconditioned matrix belong into a few clusters, say t of them. If the diameters of the clusters are small enough, then the preconditioned matrix behaves numerically like a matrix with t distinct eigenvalues. As a result, we would expect t iterations of a Krylov subspace method to produce a reasonably accurate approximation.*

Is here anything wrong?

Preconditioning is the transformation of the original problem to the form with more favorable properties that allow fast computation. Ok, but not giving a guidance.

- *If the preconditioned matrix has k distinct clusters of eigenvalues, then the backward stability of the algorithm in finite precision arithmetic together with the polynomial convergence bound based on eigenvalues ensures that, computationally, there will be a large error reduction after k steps if the algorithm is applied to the preconditioned system.*

- *Let the eigenvalues of the diagonalizable preconditioned matrix belong into a few clusters, say t of them. If the diameters of the clusters are small enough, then the preconditioned matrix behaves numerically like a matrix with t distinct eigenvalues. As a result, we would expect t iterations of a Krylov subspace method to produce a reasonably accurate approximation.*

Is here anything wrong? Yes, the statements are conceptually and fundamentally wrong. It is easy to find theoretical as well as experimental evidence in literature.

- 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 of 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 literature as an obvious fact.

- Without an appropriate (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) applied to systems with (generally) nonsymmetric matrices; 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.

3 Condition and spectral numbers?

- It can indeed be useful to investigate condition and spectral numbers providing that this is not considered, in general, the end of the story. See Faber, Manteuffel and Parter (1990).

3 Condition and spectral numbers?

- It can indeed be useful to investigate condition and spectral numbers providing that this is not considered, in general, the end of the story. See Faber, Manteuffel and Parter (1990).

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

3 Condition and spectral numbers?

- It can indeed be useful to investigate condition and spectral numbers providing that this is not considered, in general, the end of the story. See Faber, Manteuffel and Parter (1990).

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

- Daniel (1967) did not identify the CG convergence with the Chebyshev polynomials-based bound. He carefully writes (modifyling 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."

  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.

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

4 Extensive literature related to operator preconditioning

Gunn, D'yakonov, Faber, Manteuffel, Parter, Klawonn, Arnold, Falk, Winther, Axelsson, Karátson, Hiptmair, Vassilevski, Neytcheva, Notay, Elmann, Silvester, Wathen, Zulehner, Simoncini, Oswald, Griebel, Rüde, Steinbach, Wohlmuth, Bramble, Pasciak, Xu, Kraus, Nepomnyaschikh, Dahmen, Kunoth, Yserentant, Mardal, Nordbotten, Rees, Smears, Pearson, ...

Details, proofs and (certainly far from complete) references can be found in

- J. Málek and Z.S., *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs.* SIAM Spotlight Series, SIAM (2015)
- J. Hrnčíř, I. Pultarová, Z.S., *Decomposition into subspaces preconditioning: Abstract Framework* (2018, submitted for publication)

4 Basic setting on the Hilbert space V

Inner product

$$(\cdot,\cdot)_V : V \times V \to \mathbb{R}\,, \quad \|\cdot\|_V\,,$$

dual space $V^\#$ of bounded linear functionals on V with the duality pairing and the associated Riesz map

$$\langle\cdot,\cdot\rangle : V^\# \times V \to \mathbb{R}\,, \quad \tau : V^\# \to V \quad \text{such that} \quad (\tau f, v)_V := \langle f, v\rangle \quad \text{for all } v \in V.$$

Equation in the functional space $V^\#$

$$\mathcal{A}u \;=\; b$$

with a linear, bounded, coercive, and self-adjoint operator

$$\mathcal{A} : V \to V^\#\,, \quad a(u,v) := \langle \mathcal{A}u, v\rangle\,,$$

$$C_{\mathcal{A}} \;:=\; \sup_{v\in V,\, \|v\|_V=1} \|\mathcal{A}v\|_{V^\#} < \infty\,,$$

$$c_{\mathcal{A}} \;:=\; \inf_{v\in V,\, \|v\|_V=1} \langle \mathcal{A}v, v\rangle > 0\,.$$

Linear, bounded, coercive, and self-adjoint $\mathcal{B}$ with $C_{\mathcal{B}}, c_{\mathcal{B}}$,

$$(\cdot,\cdot)_{\mathcal{B}} : V \times V \to \mathbb{R}, \qquad (w,v)_{\mathcal{B}} := \langle \mathcal{B}w, v\rangle \qquad \text{for all } w,v \in V\,,$$

$$\tau_{\mathcal{B}} : V^{\#} \to V, \qquad (\tau_{\mathcal{B}} f, v)_{\mathcal{B}} \;:=\; \langle f, v\rangle \qquad \text{for all } f \in V^{\#},\; v \in V\,.$$

Instead of the equation in the functional space $V^{\#}$

$$\mathcal{A}u \;=\; b$$

we solve the equation in the solution space V

$$\tau_{\mathcal{B}}\, \mathcal{A}\, u \;=\; \tau_{\mathcal{B}}\, b\,,$$

i.e.

$$\mathcal{B}^{-1} \mathcal{A}\, u \;=\; \mathcal{B}^{-1} b.$$

Theorem (Norm equivalence and condition number)

Assuming that the linear, bounded, coercive and self-adjoint operators $\mathcal{A}$ and $\mathcal{B}$ are $V^{\#}$-*norm equivalent* on V, i.e. there exist $0 < \alpha \leq \beta < \infty$ such that

$$\alpha \leq \frac{\|\mathcal{A}w\|_{V^{\#}}}{\|\mathcal{B}w\|_{V^{\#}}} \leq \beta, \quad \text{for all } \; w \in V, w \neq 0\,.$$

Then

$$\kappa(\mathcal{B}^{-1}\mathcal{A}) := \|\mathcal{B}^{-1}\mathcal{A}\|_{\mathcal{L}(V,V)} \|\mathcal{A}^{-1}\mathcal{B}\|_{\mathcal{L}(V,V)} \leq \frac{\beta}{\alpha}\,.$$

Theorem (Spectral equivalence and spectral number)

Assuming that the linear, bounded, coercive and self-adjoint operators $\mathcal{A}$ and $\mathcal{B}$ are *spectrally equivalent* on V, i.e. there exist $0 < \gamma \leq \delta < \infty$ such that

$$\gamma \leq \frac{\langle \mathcal{A}w, w\rangle}{\langle \mathcal{B}w, w\rangle} \leq \delta, \quad \text{for all } w \in V, w \neq 0\,.$$

Then

$$\hat{\kappa}(\mathcal{A},\mathcal{B}) := \frac{\sup_{z\in V,\, \|z\|_V=1} \left((\tau\mathcal{B})^{-1/2}\tau\mathcal{A}\,(\tau\mathcal{B})^{-1/2}z, z\right)_V}{\inf_{v\in V,\, \|v\|_V=1} \left((\tau\mathcal{B})^{-1/2}\tau\mathcal{A}\,(\tau\mathcal{B})^{-1/2}v, v\right)_V} \leq \frac{\delta}{\gamma}\,.$$

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

Consider N-dimensional subspace $V_h \subset V$ and look for $u_h \in V_h$, $u_h \approx u \in V$ such that

$$\langle \mathcal{A} u_h - b, v \rangle \; = \; 0 \qquad \text{for all } v \in V_h \, .$$

Restrictions $\mathcal{A}_h : V_h \to V_h^{\#}$, $b_h : V_h \to \mathbb{R}$ give the problem in $V_h^{\#}$

$$\mathcal{A}_h u_h \; = \; b_h , \qquad u_h \in V_h , \quad b_h \in V_h^{\#} \, .$$

With the inner product $(\cdot,\cdot)_{\mathcal{B}}$ and the associated restricted Riesz map

$$\tau_{\mathcal{B},h} : V_h^{\#} \to V_h$$

we get the abstract form of the preconditioned discretized problem in V_h

$$\tau_{\mathcal{B},h} \, \mathcal{A}_h \, u_h \; = \; \tau_{\mathcal{B},h} \, b_h \, .$$

Using the discretization basis $\Phi_h = (\phi_1, \ldots, \phi_N)$ of V_h
and the canonical dual basis $\Phi_h^{\#} = (\phi_1^{\#}, \ldots, \phi_N^{\#})$ of $V_h^{\#}$, $(\Phi_h^{\#})^* \Phi_h = \mathbf{I}_N$,

$$\mathbf{M}_h^{-1} \, \mathbf{A}_h \, \mathbf{x}_h \; = \; \mathbf{M}_h^{-1} \, \mathbf{b}_h ,$$

where

$$\begin{aligned} \mathbf{A}_h, \, \mathbf{M}_h \; &\in \; \mathbb{R}^{N \times N}, \quad \mathbf{x}_h, \mathbf{b}_h \in \mathbb{R}^N , \\ (\mathbf{x}_h)_i \; &= \; \langle \phi_i^{\#}, u_h \rangle , \quad (\mathbf{b}_h)_i \; = \; \langle b, \phi_i \rangle , \\ \mathbf{A}_h \; &= \; (a(\phi_j, \phi_i))_{i,j=1,\ldots,N} \; = \; (\langle \mathcal{A} \phi_j, \phi_i \rangle)_{i,j=1,\ldots,N} , \\ \mathbf{M}_h \; &= \; (\langle \mathcal{B} \phi_j, \phi_i \rangle)_{i,j=1,\ldots,N} , \end{aligned}$$

or

$$\mathbf{A}_h \; = \; (\mathcal{A} \Phi_h)^* \Phi_h, \qquad \mathbf{M}_h \; = \; (\mathcal{B} \Phi_h)^* \Phi_h \, .$$

Using (an arbitrary) decomposition $\mathbf{M}_h = \mathbf{L}_h \mathbf{L}_h^*$, the resulting preconditioned algebraic system can be transformed into

$$(\mathbf{L_h^{-1}} \mathbf{A}_h \mathbf{L_h^{*\,-1}})\,(\mathbf{L_h^*} \mathbf{x}_h) = \mathbf{L}_h^{-1} \mathbf{b}_h \,,$$

i.e.,

$$\mathbf{A}_{t,h}\,\mathbf{x}_h^t = \mathbf{b}_h^t \,.$$

5 Preconditioning as transformation of discretization basis

Consider

$$\Phi_h \to \tilde{\Phi}_{t,h} \quad \text{such that} \quad \mathbf{M}_{t,h} = (\mathcal{B}\tilde{\Phi}_{t,h})^* \tilde{\Phi}_{t,h} = \mathbf{I}\,,$$

i.e. orthogonalization of the basis with respect to the inner product $(\cdot,\cdot)_{\mathcal{B}}$. Then

$$\tilde{\Phi}_{t,h} = \Phi_h \mathbf{M_h}^{-1/2}, \quad \tilde{\Phi}^{\#}_{t,h} = \Phi^{\#}_h \mathbf{M_h}^{1/2}$$

gives immediately the preconditioned system $\tilde{\mathbf{A}}_{t,h}\, \tilde{\mathbf{x}}^t_h \;=\; \tilde{\mathbf{b}}^t_h$ corresponding to $\mathbf{L_h} := \mathbf{M_h}^{1/2}$. Any other choice

$$\Phi_{t,h} = \Phi_h \mathbf{L}_{\mathbf{h}}^{*\,-1}, \quad \Phi^{\#}_{t,h} = \Phi^{\#}_h \mathbf{L_h}$$

is given via orthogonal transformation

$$\Phi_{t,h} = \tilde{\Phi}_{t,h}\mathbf{Q}^*\,, \quad \mathbf{Q}^* = \mathbf{M_h}^{1/2}\mathbf{L}_{\mathbf{h}}^{*\,-1}\,, \quad \mathbf{Q}^*\mathbf{Q} = \mathbf{I}\,.$$

5 Points that are worth noticing

- Preconditioning is mathematically equivalent to orthogonalization of the discretization basis wrt the inner product $(\cdot,\cdot)_{\mathcal{B}}$. This will change the supports of the basis functions!

- Transformation of the discretization basis (preconditioning) is different from a change of the algebraic basis (similarity transformation).

- Any algebraic preconditioning can be put into the operator preconditioning framework by transformation of the discretization basis and the associated change of the inner product in the infinite dimensional Hilbert space V.

Theorem (Norm equivalence and condition number)

Let the linear, bounded, coercive and self-adjoint operators $\mathcal{A}$ and $\mathcal{B}$ from V to $V^{\#}$ be $V^{\#}$-norm equivalent with the lower and upper bounds α and β, respectively, i.e.

$$\alpha \leq \frac{\|\mathcal{A}w\|_{V^{\#}}}{\|\mathcal{B}w\|_{V^{\#}}} \leq \beta \quad \text{for all } w \in V\,,\ w \neq 0, \quad 0 < \alpha \leq \beta < \infty\,.$$

Let $\mathbf{S}_h$ be the Gram matrix of the discretization basis $\Phi_h = (\phi_1, \ldots, \phi_N)$ of $V_h \subset V$,

$$(\mathbf{S}_h)_{ij} = (\phi_i, \phi_j)_V\ .$$

Then the condition number of the matrix $\mathbf{M}_h^{-1}\mathbf{A}_h$ is bounded as

$$\kappa(\mathbf{M}_h^{-1}\mathbf{A}_h) := \|\mathbf{M}_h^{-1}\mathbf{A}_h\|\,\|\mathbf{A}_h^{-1}\mathbf{M}_h\| \leq \frac{\beta}{\alpha}\,\kappa(\mathbf{S}_h)\,.$$

Theorem (Spectral equivalence and spectral number)

Let the linear, bounded, coercive and self-adjoint operators $\mathcal{A}$ and $\mathcal{B}$ be spectrally equivalent with the lower and upper bounds γ and δ respectively, i.e.

$$\gamma \leq \frac{\langle \mathcal{A}w, w\rangle}{\langle \mathcal{B}w, w\rangle} \leq \delta \quad \text{for all } w \in V\,, \quad 0 < \gamma \leq \delta < \infty\,.$$

Then the spectral number $\hat{\kappa}(\mathbf{A}_h, \mathbf{M}_h)$, which is equal to the condition number of the matrix $\mathbf{A}_{t,h} = \mathbf{L}_h^{-1}\mathbf{A}_h(\mathbf{L}_h^*)^{-1}$ for any $\mathbf{L}_h$ such that $\mathbf{M}_h = \mathbf{L}_h\mathbf{L}_h^*$, is bounded as

$$\hat{\kappa}(\mathbf{A}_h, \mathbf{M}_h) := \frac{\sup_{\mathbf{z}\in\mathbb{R}^N,\, \|\mathbf{z}\|=1} \left(\mathbf{M}_h^{-1/2}\mathbf{A}_h\mathbf{M}_h^{-1/2}\mathbf{z}, \mathbf{z}\right)}{\inf_{\mathbf{v}\in\mathbb{R}^N,\, \|\mathbf{v}\|=1} \left(\mathbf{M}_h^{-1/2}\mathbf{A}_h\mathbf{M}_h^{-1/2}\mathbf{v}, \mathbf{v}\right)} = \kappa(\mathbf{A}_{t,h}) \leq \frac{\delta}{\gamma}\,.$$

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

Decomposition with non-unique representation of elements in V

$$V = \sum_{j \in J} V_j, \quad \text{i.e.,} \quad v = \sum_{j \in J} v_j, \ \ v_j \in V_j, \quad \text{for all } \ v \in V, \ J \ \text{ is finite;}$$

Sufficient condition for $V^{\#} \subset V_j^{\#}$:

$$c_{V_j} \, \|v\|_V^2 \ \le \ \|v\|_j^2 \quad \text{for all } \ v \in V_j, \ \ 0 < c_{V_j}, \ \ j \in J;$$

Other side inequality:

$$\|v\|_{\mathrm{S}}^2 \ := \inf_{v = \sum_{j \in J} v_j} \left\{ \sum_{j \in J} \|v_j\|_j^2 \right\} \ \le \ C_{\mathrm{S}} \, \|v\|_V^2 \, , \quad \text{for all } \ v \in V \, .$$

Consider local preconditioners

$$\mathcal{B}_j : V_j \to V_j^{\#}, \qquad \langle \mathcal{B}_j w, z \rangle \;=\; \langle \mathcal{B}_j z, w \rangle \qquad \text{for all} \quad w, z \in V_j \,,$$

with $C_{\mathcal{B}_j}$, $c_{\mathcal{B}_j}$ defined as above. Then $\mathcal{B}_j^{-1} : V_j^{\#} \to V_j$, $V^{\#} \subset V_j^{\#}$, and

$$\mathcal{M}^{-1} \;:=\; \sum_{j \in J} \mathcal{B}_j^{-1}, \qquad \mathcal{M}^{-1} : V^{\#} \to V$$

gives the global preconditioner. The preconditioned (equivalent?) problem

$$\mathcal{M}^{-1}\, \mathcal{A}\, u \;=\; \mathcal{M}^{-1}\, b\,.$$

Boundedness and coercivity of $\mathcal{M}^{-1}$

$$\|\mathcal{M}^{-1}\|_{\mathcal{L}(V^\#,V)} = \sup_{f\in V^\#,\, \|f\|_{V^\#}=1} \|\mathcal{M}^{-1}f\|_V \le C_{\mathcal{M}^{-1}} := \sum_{j\in J} \frac{1}{c_{\mathcal{B}_j} c_{V_j}} < \infty\,,$$

$$\inf_{f\in V^\#,\, \|f\|_{V^\#}=1} \langle f, \mathcal{M}^{-1} f\rangle \ge c_{\mathcal{M}^{-1}} := \frac{1}{C_{\mathrm{S}} \max_{j\in J} C_{\mathcal{B}_j}} > 0\,,$$

gives equivalence of $\mathcal{A}u = b$ and $\mathcal{M}^{-1}\mathcal{A}u = \mathcal{M}^{-1}b$.

Moreover, we can get norm equivalence and spectral equivalence of $\mathcal{A}$ and $\mathcal{M}$.

6 Bound using norms of the locally preconditioned residuals

Theorem

For any $v \in V \approx u$

$$a\left(\mathcal{M}^{-1}\mathcal{A}\,(v-u),\, v-u\right) \;=\; \sum_{j\in J} \|\bar{r}_j\|^2_{\mathcal{B}_j}\,,$$

$$\frac{\min_{j\in J} c_{\mathcal{B}_j}}{C^2_{\mathcal{A}}}\left(\sum_{k\in J}\frac{1}{c_{V_k}\,c_{\mathcal{B}_k}}\right)^{-1}\sum_{j\in J}\|\bar{r}_j\|^2_j \quad\le$$

$$\|v-u\|^2_V \quad\le\quad \frac{C_{\mathrm{S}}\,(\max_{j\in J} C_{\mathcal{B}_j})^2}{c^2_{\mathcal{A}}}\sum_{j\in J}\|\bar{r}_j\|^2_j\,,$$

where $\bar{r}_j \;:=\; \mathcal{B}_j^{-1}\mathcal{A}\,v \;-\; \mathcal{B}_j^{-1}b$ are the locally preconditioned residuals of v.

Theorem

If we consider the stable splitting

$$c_{\mathrm{S}}\|v\|_V^2 \;\leq\; \|v\|_{\mathrm{S}}^2 \;\leq\; C_{\mathrm{S}}\|v\|_V^2 \qquad \text{for all } v \in V,$$

then

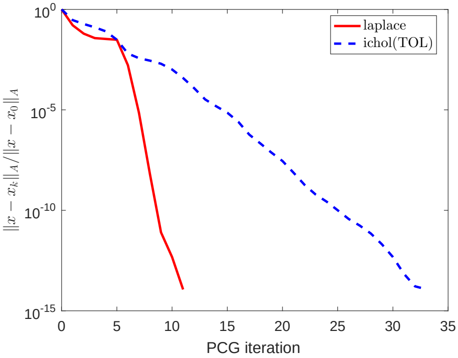
$$\frac{c_{\mathcal{A}}}{C_{\mathrm{S}} \max_{j\in J} C_{\mathcal{B}_j}} \;\leq\; \frac{\langle \mathcal{A}v, v\rangle}{\langle \mathcal{M}v, v\rangle} \;\leq\; \frac{C_{\mathcal{A}}}{c_{\mathrm{S}} \min_{j\in J} c_{\mathcal{B}_j}} \quad \text{for all } v \in V,\; v \neq 0,$$

$$\frac{c_{\mathrm{S}} \min_{j\in J} c_{\mathcal{B}_j}}{C_{\mathcal{A}}} \;\leq\; \frac{\|\mathcal{A}^{-1} f\|_V}{\|\mathcal{M}^{-1} f\|_V} \;\leq\; \frac{C_{\mathrm{S}} \max_{j\in J} C_{\mathcal{B}_j}}{c_{\mathcal{A}}} \qquad \text{for all } f \in V^{\#},\; f \neq 0.$$

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

7 Better conditioning does not necessarily mean faster convergence!

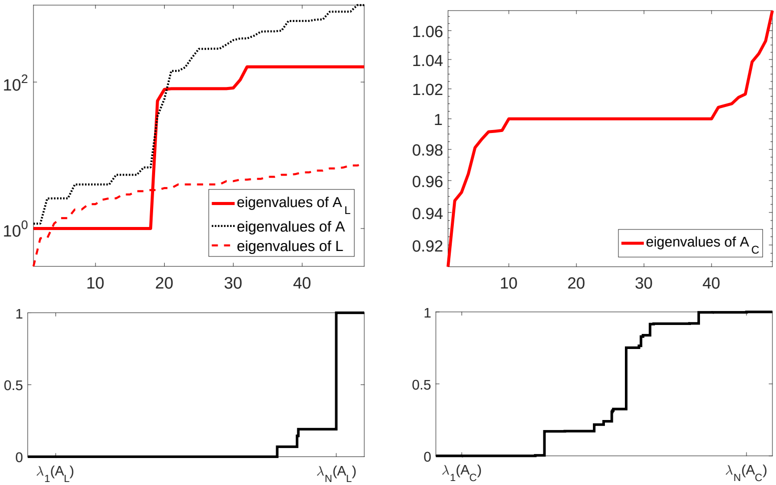


Nonhomogeneous diffusion function (Morin, Nocheto, Siebert, SIREV (2002)), uniform mesh.

ICHOLPCG (drop-off tolerance 1e-02); Laplace operator PCG.

Condition numbers of $\mathbf{A}_{t,h}$: 1.6e01, 1.61e02.

7 Gergelits, Mardal, Nielsen, S (in preparation)



Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

"Myth:
A belief given uncritical acceptance by the members of a group especially in support of existing or traditional practices and institutions."

Webster's Third New International Dictionary, Enc. Britannica Inc., Chicago (1986)

“Myth:
A belief given uncritical acceptance by the members of a group especially in support of existing or traditional practices and institutions.”

Webster’s Third New International Dictionary, Enc. Britannica Inc., Chicago (1986)

A. Einstein,
in Oxford User’s Guide to Mathematics, E. Zeidler (ed), OUP (2004), p. 3:

“Everything should be made as simple as possible, but not simpler.”

"Myth:
A belief given uncritical acceptance by the members of a group especially in support of existing or traditional practices and institutions."

Webster's Third New International Dictionary, Enc. Britannica Inc., Chicago (1986)

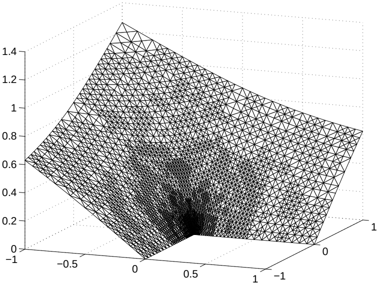
A. Einstein,
in Oxford User's Guide to Mathematics, E. Zeidler (ed), OUP (2004), p. 3:

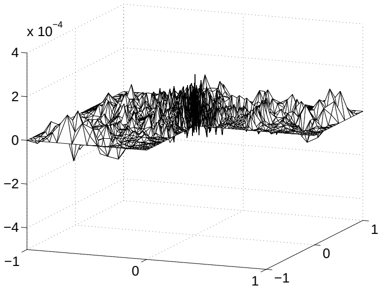
"Everything should be made as simple as possible, but not simpler."

Once a myth becomes widely promoted common "mathematical knowledge",
it is difficult to return to scientific discussion based on facts and evidence ...

8 Examples of widespread myths concern

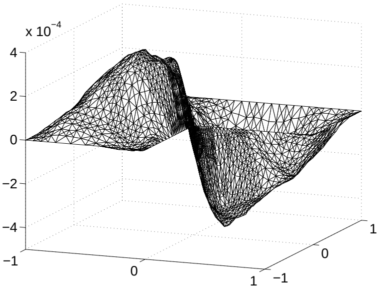
- Minimal polynomials and finite termination property
- Chebyshev bounds and CG
- Spectral information and clustering of eigenvalues
- Operator-based bounds and functional analysis arguments on convergence
- Finite precision computations seen as a minor modification of the exact considerations
- Linearization of nonlinear phenomenon
- Considering CG in matrix computations as a simplification of CG in general nonlinear optimization
- Well conditioned basis and short recurrences (look-ahead)
- Sparsity as an ultimate positive feature of the FEM discretizations
- Discretization and algebraic errors in numerical PDEs

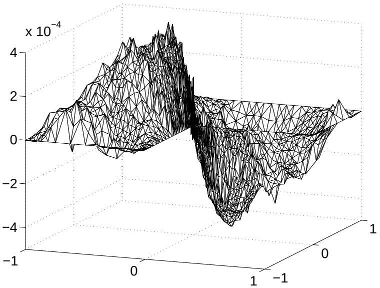




Exact solution u (left) and the discretization error $u - u_h$ (right) in the Poisson model problem, linear FEM, adaptive mesh refinement.

Quasi equilibrated discretization error over the domain.





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

$$\|\nabla(u - u_h)\| \gg \|x - x_n\|_A \,.$$

8 Insight and understanding ... Why do humans science ?

- Humans must do science in order to survive.
  Question: How to make things work?

8 Insight and understanding ... Why do humans science ?

- Humans must do science in order to survive.
  Question: How to make things work?

- Humans must do science because they are humans.
  Question: Why and how does the world work?

8 Insight and understanding ... Why do humans science ?

- Humans must do science in order to survive.
  Question: How to make things work?

- Humans must do science because they are humans.
  Question: Why and how does the world work?

Success as the only measure of the common good ??? Avalanche of performance metrics, overspecialization, fragmentation, confusion, shallowness, mythology .....

Pure against (!) applied mathematics,
basic research against (!) applied research, ...

Gen 11, 1-7

"The whole world spoke the same language, using the same words. { ... } They said to one another, "Come, let us mold bricks and harden them with fire. { ... } Then they said, "Come, let us build ourselves a city and a tower with its top in the sky, and so make a name for ourselves; otherwise we will be scattered all over the earth."

The Lord came down to see the city and the tower that men had built. Then the Lord said: { . . . } "Let us go down and confuse their language, so that one will not understand what another says." Thus the Lord scattered them from there all over the earth ... "

Outline

1. Infinite dimensional problems and finite dimensional computations
2. Krylov subspace methods: Hestenes, Stiefel, Lanczos (1950-52)
3. Preconditioning - what do we mean?
4. Operator preconditioning
5. Discretization
6. Decomposition into subspaces
7. Is there a world behind the conditioning, norm and spectral equivalence?
8. Abandoning logical principles leads to mythology (and worse)
9. Optimistic outlook

9 A way out? Being humble and admitting the failures

- J. Tinsley Oden, GAMM 2018, Munich: *"A model can never be validated as a perfect portrayal of the truth. It can only be deemed 'not invalid', contingent on its agreement with available observational data for (subjective) choices of metrics and tolerances."*

- Verification based on model problems assumes extrapolation.

- Invalidation of hypotheses ("common knowledge") using simple model problems is often considered insufficient. What kind of logic is used in such cases?

- Assumptions required in derivation of results are forgotten (by mathematicians!) in their applications.

- We should do our piece of work, irrespectively of how big or small, to the best of our abilities. This is what leads to originality. (C. S. Lewis, Fern-seed and Elephants (1975))

- We can never foresee the end of any of our action. What would the builders of the Hejnice monastery say?

9 Conclusions and outlook related to operator preconditioning

- Given operator preconditioning framework may help in comparison of existing approaches (work in progress).

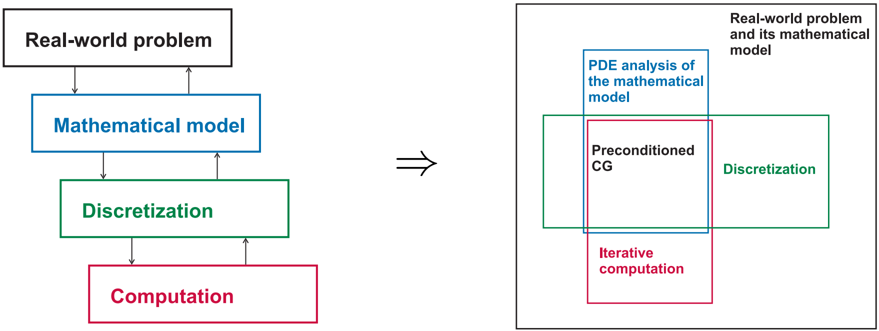
- Results guaranteeing fast convergence in practice are based on the subspace splitting and construction of preconditioning that use information on (the inner structure of) the operator $\mathcal{A}$.

- Relationship between the operators $\mathcal{A}$ and $\mathcal{B}$? What can be said about the whole spectrum of the matrix $B^{-1}A$? (Work in progress).

- Adaptation to the problem is the key to efficient solvers.
  Adaptation in many ways!

- $\mathcal{O}(n)$ reliable approximate solvers? A posteriori error analysis leading to efficient and reliable balancing the errors of various origin (including the inaccuracy of algebraic computations).

9 An optimistic view, but a very long way to go ....



Formulation of the model, discretization and algebraic computation, including the evaluation of the error, stopping criteria for the algebraic solver, adaptivity etc. are very closely related to each other.

Theorem

Let $\mathcal{A} : V \to V^{\#}$ be a linear, bounded, coercive and self-adjoint operator. Then its boundedness constant $C_{\mathcal{A}}$ and the coercivity constant $c_{\mathcal{A}}$ can be expressed as

$$C_{\mathcal{A}} = \|\mathcal{A}\|_{\mathcal{L}(V,V^{\#})} = \sup_{v\in V,\, \|v\|_V=1} \langle \mathcal{A}v, v\rangle, \tag{1}$$

$$c_{\mathcal{A}} = \inf_{v\in V,\, \|v\|_V=1} \langle \mathcal{A}v, v\rangle = \frac{1}{\sup_{f\in V^{\#},\, \|f\|_{V^{\#}}=1} \|\mathcal{A}^{-1}f\|_V} = \frac{1}{\|\mathcal{A}^{-1}\|_{\mathcal{L}(V^{\#},V)}}. \tag{2}$$

Proof of (1)

Statement (1) follows from

$$\|\mathcal{A}\|_{\mathcal{L}(V,V^{\#})} = \|\tau\mathcal{A}\|_{\mathcal{L}(V,V)} = \sup_{v\in V,\, \|v\|_V=1} (\tau\mathcal{A}v, v)_V = \sup_{v\in V,\, \|v\|_V=1} \langle \mathcal{A}v, v\rangle,$$

where we used the fact that for any self-adjoint operator S in a Hilbert space V

$$\begin{aligned} \|S\|_{\mathcal{L}(V,V)} &= \sup_{z\in V,\, \|z\|_V=1} \|Sz\|_V = \sup_{z\in V,\, \|z\|_V=1} (Sz, Sz)_V^{1/2} \\ &= \sup_{z\in V,\, \|z\|_V=1} |(Sz, z)_V|. \end{aligned}$$

Proof of (2)

$$\frac{1}{\sup_{f\in V^\#,\, \|f\|_{V^\#}=1} \|\mathcal{A}^{-1}f\|_V} = \inf_{v\in V,\, \|v\|_V=1} \|\mathcal{A}v\|_{V^\#} = \inf_{v\in V,\, \|v\|_V=1} \|\tau\mathcal{A}v\|_V$$

We have to prove

$$m_{\mathcal{A}} := \inf_{v\in V,\, \|v\|_V=1} (\tau\mathcal{A}v, v)_V = \inf_{v\in V,\, \|v\|_V=1} \|\tau\mathcal{A}v\|_V .$$

Here "$\leq$" is trivial. We will show that "$<$" leads to a contradiction. Since $m_{\mathcal{A}}$ belongs to the spectrum of $\tau\mathcal{A}$, there exists a sequence $v_1, v_2, \dots \in V$, $\|v_k\|_V = 1$, such that

$$\lim_{k\to\infty} \|\tau\mathcal{A}v_k - m_{\mathcal{A}}v_k\|_V^2 = 0. \tag{3}$$

Assuming

$$m_{\mathcal{A}} < \inf_{v\in V,\, \|v\|_V=1} \|\tau\mathcal{A}v\|_V - \triangle, \quad \triangle > 0,$$

we get

$$\begin{aligned} \|\tau\mathcal{A}v_k - m_{\mathcal{A}}v_k\|_V^2 &= \|\tau\mathcal{A}v_k\|_V^2 + m_{\mathcal{A}}^2 - 2m_{\mathcal{A}}(\tau\mathcal{A}v_k, v_k)_V \\ &\geq \|\tau\mathcal{A}v_k\|_V^2 + m_{\mathcal{A}}^2 - 2m_{\mathcal{A}}\|\tau\mathcal{A}v_k\|_V = (\|\tau\mathcal{A}v_k\|_V - m_{\mathcal{A}})^2 > \triangle^2. \end{aligned}$$

Thank you for your kind patience! (Regards also from my friend Delhi)