# Sparsity, discretization, preconditioning, and adaptivity in solving linear equations

Zdeněk Strakoš[1] Jan Papež[2] Tomáš Gergelits[1]

[1]Charles University and Czech Academy of Sciences, Prague

[2]Inria Paris

XX Householder Symposium, Virginia, June 18–23, 2017

# Motto of the talk

Mathematical modeling, related analysis and computations (here the focus is on preconditioned Krylov subspace methods) has to deal with questions that go across several fields and therefore handling them in their complexity requires extensive and thorough collaboration.

# Adaptation as the main principle

Cornelius Lanczos, March 9, 1947

*"To obtain a solution in very few steps*
*means nearly always that one has found a way*
*that does justice to the inner nature of the problem."*

Albert Einstein, March 18, 1947

*"Your remark on the importance of*
*adapted approximation methods makes very*
*good sense to me, and I am convinced*
*that this is a fruitful mathematical aspect,*
*and not just a utilitarian one."*

- Inner nature of the problem?
- Nonlinear adaptation of the iterations to linear problems?

# Thanks to very many, for current collaboration in particular to

Erin Carson,
Jakub Hrnčíř,
Jörg Liesen,
Josef Málek,
Miroslav Pranic
Stefano Pozza
Ivana Pultarová,
Miroslav Rozložník,
Petr Tichý,
Miroslav Tůma.

# Outline

1. Infinite dimensional problem and finite dimensional computations
2. Convergence behavior and spectral information
3. Mathematical structure preserved at the presence of numerical errors
4. Preconditioning and discretization
5. Decomposition into subspaces and preconditioning
6. $h$-adaptivity based on the residual-based estimator
7. Conclusions

# 1. Infinite dimensional problem and finite dimensional computations

# 1 Krylov subspace methods as polynomial methods

Consider a numerical solution of equations

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

on an infinite dimensional Hilbert space $V$ , where $\mathcal{G}$ is a linear invertible operator, $\mathcal{G} : V \to V$ .

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 a particular polynomial of degree at most $n-1$ and $\mathcal{G}_n$ is obtained by restricting and projecting $\mathcal{G}$ onto the $n$th Krylov subspace

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

# 1 Four basic questions

1. How fast $u_n,\ n = 1, 2, \ldots$ approximate the desired solution $u$ ? Nonlinear adaptation.

2. Which characteristics of $\mathcal{G}$ and $f$ can be used in investigating the previous question? Inner nature of the problem.

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.

# 1 Hierarchy of problems

Problem in infinite dimensional Hilbert space with bounded invertible operator $\mathcal{G}$

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

is approximated on the subspace $V_h \subset V$ by the problem with the finite dimensional operator

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

represented, using an appropriate basis of $V_h$, 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)

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

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

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

Consistency deals with the question 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

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

implies convergence of the discretization scheme

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

In computations we only approximate $u_h$ by $u_h^{(n)}$ .

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

$$\mathcal{A} = \frac{1}{2\pi\iota}\int_{\Gamma}\lambda\,(\lambda I_N - \mathcal{A})^{-1}\,d\lambda \;=\; \frac{1}{2\pi\iota}\sum_{j=1}^{N}\int_{\Gamma_j}\lambda\,(\lambda I_N - \mathcal{A})^{-1}\,d\lambda$$

$$= \sum_{j=1}^{N} Y\,\mathrm{diag}\left(\frac{1}{2\pi\iota}\int_{\Gamma_j}\frac{\lambda}{\lambda-\lambda_j}\,d\lambda\right)\,Y^* \;=\; \sum_{j=1}^{N}\lambda_j\,y_j y_j^*$$

$$= \int_{m(\mathcal{A})}^{M(\mathcal{A})}\lambda\,dE(\lambda)\,.$$

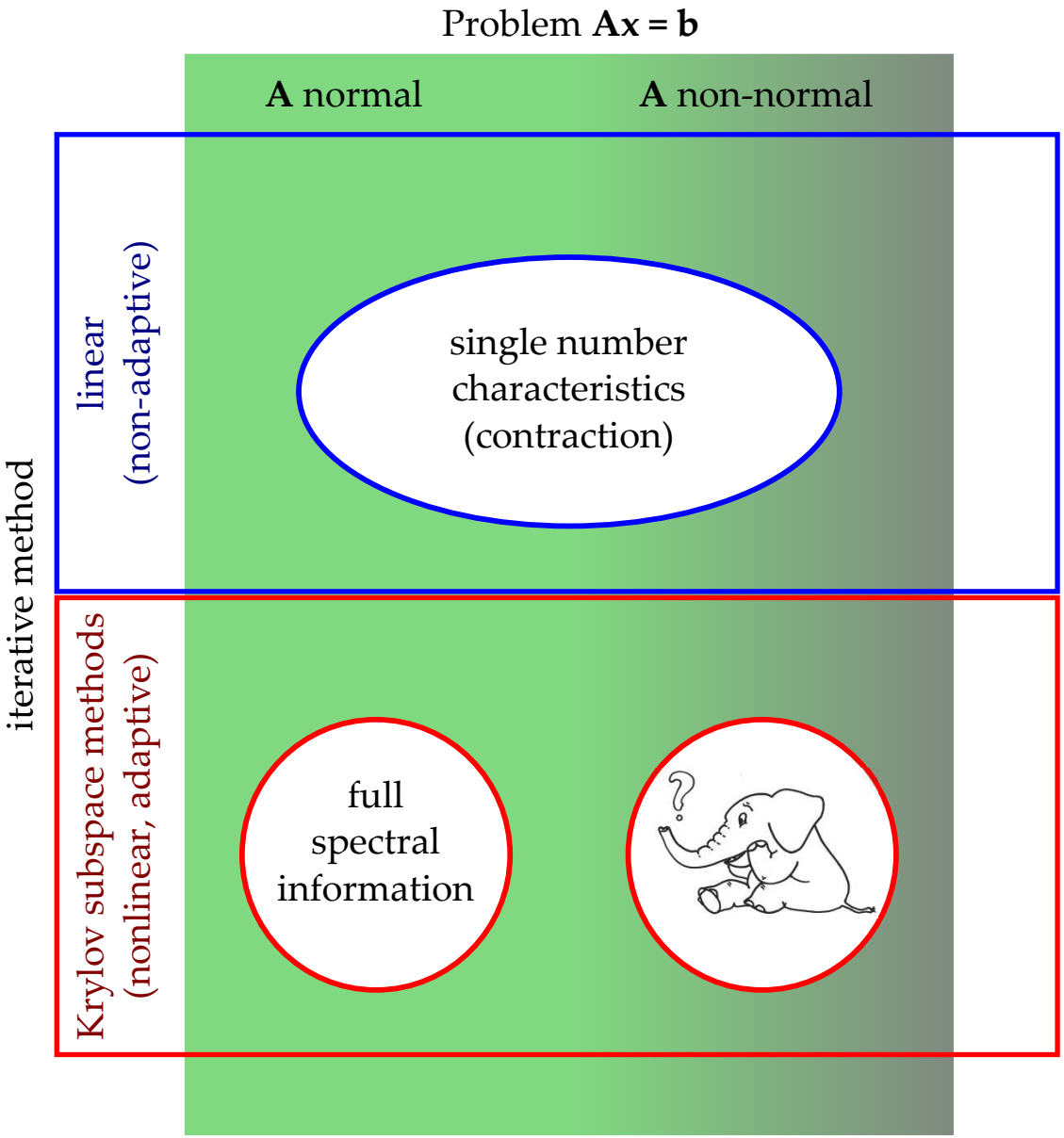
- Compact infinite dimensional self-adjoint operators
- Bounded infinite dimensional self-adjoint operators
- Generalization to bounded normal and non-normal operators

# 2. Convergence and spectral information

Some more details and references to many original works can be found in

- J. Liesen. and Z.S., *Krylov Subspace Methods, Principles and Analysis.* Oxford University Press (2013), Sections 5.1 - 5.7
- T. Gergelits and Z.S., *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations,* Numer. Alg. 65, 759-782 (2014)

# 2 Adaptation to the inner nature of the problem?



- Stationary Richardson (assume $\mathbf{A}$ HPD)

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

- Chebyshev semiiterative method

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

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

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$, $\mathbf{p}_0 = \mathbf{r}_0$. 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 $\alpha_{n-1}$ 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

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

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

**Inexact computations?**

# 2 Lanczos (CG) as the nonlinear moment problem

Let $\mathcal{G}$ be a linear bounded self-adjoint operator on a Hilbert space $V$, $f \in V$, $\|f\| = 1$. Consider the $2n$ real numbers

$$m_j = (\mathcal{G}^j f, f) = \int \lambda^j \, d\omega(\lambda), \quad j = 0, \ldots, 2n-1.$$

CG (assuming, in addition, the coercivity of $\mathcal{G}$), as well as the Lanczos method for approximating eigenvalues, solve the $2n$ equations

$$\sum_{j=1}^{n} \omega_j^{(n)} \{\theta_j^{(n)}\}^\ell = m_\ell, \qquad \ell = 0, 1, \ldots, 2n-1,$$

for the $2n$ real unknowns $\omega_j^{(n)} > 0$, $\theta_j^{(n)}$.

- Golub, Welsch (1968), Gordon (1968), ..., Vorobyev (1958, 1965)
- Generalizations to quasi-definite linear functionals and beyond, complex Gauss quadrature, relationship with the nonsymmetric Lanczos algorithm, minimal partial realization etc. are given in Draux (1983), Gragg (1974), Gragg and Lindquist (1983, ... , Pozza, Pranic, S (2017), Pozza, Pranic, S (2018?).

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

is the Jacobi matrix of the Lanczos process coefficients at step $n$.

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.

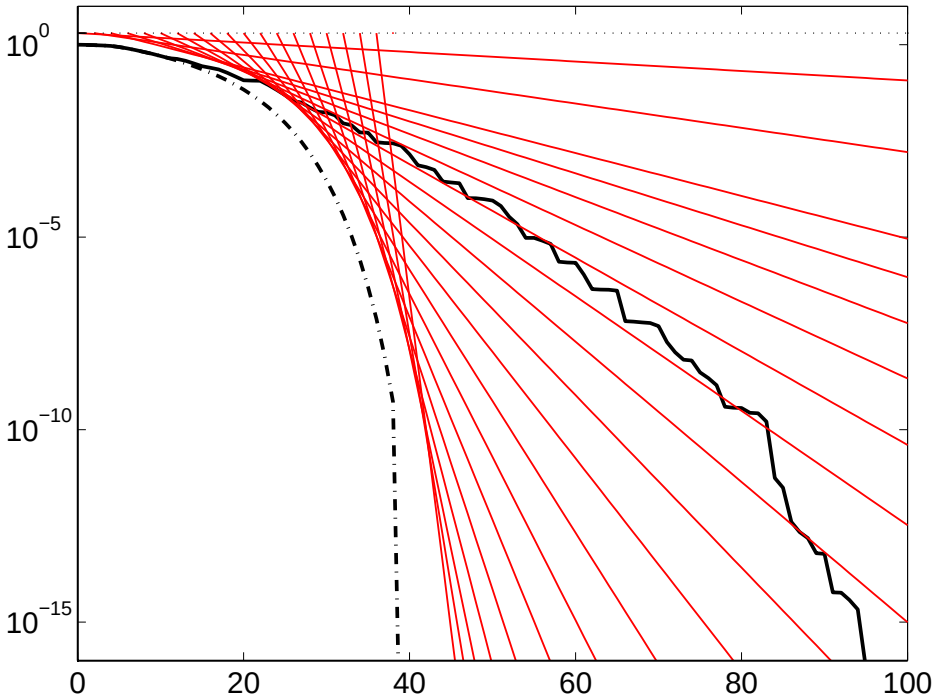
The notion of *"deflation"*.

- 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?**

# 2 Adaptive Chebyshev bound is not descriptive.




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

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

Given any spectrum and any sequence of the nonincreasing residual norms, there is a complete parametrization of the set of all GMRES associated matrices and right-hand sides.

The set of problems for which the distribution of eigenvalues alone does not correspond to convergence behavior is not of measure zero and it is not pathological.

- Widespread eigenvalues alone can not be identified with poor convergence.
- Clustered eigenvalues alone can not be identified with fast convergence.

**What does it tell us about the problem?**

**Theorem**

1° The spectrum of $\mathbf{A}$ is given by $\{\lambda_1, \ldots, \lambda_N\}$ and GMRES$(\mathbf{A}, \mathbf{b})$ yields residuals with the prescribed nonincreasing sequence $(\mathbf{x}_0 = 0)$

$$\|\mathbf{r}_0\| \geq \|\mathbf{r}_1\| \geq \cdots \geq \|\mathbf{r}_{N-1}\| > \|\mathbf{r}_N\| = 0\,.$$

2° Let $\mathbf{C}$ be the spectral companion matrix, $h = (h_1, \ldots, h_N)^T$, $h_i^2 = \|\mathbf{r}_{i-1}\|^2 - \|\mathbf{r}_i\|^2$, $i = 1, \ldots, N$. Let $\mathbf{R}$ be a nonsingular upper triangular matrix such that $\mathbf{Rs} = \mathbf{h}$ with $\mathbf{s}$ being the first column of $\mathbf{C}^{-1}$, and let $\mathbf{W}$ be a unitary matrix. Then

$$\mathbf{A} = \mathbf{WRCR}^{-1}\mathbf{W}^* \quad \text{and} \quad \mathbf{b} = \mathbf{Wh}\,.$$

Greenbaum, Pták, Arioli and S (1994 - 98); Liesen (1999); Eiermann and Ernst (2001); Meurant (2012); Meurant and Tebbens (2012, 2014); .....

# 3. Mathematical structure preserved at the presence of numerical errors

# 3 Back to the mathematical structure of Jacobi matrices

- Practical computation generates a sequence of (nested) Jacobi matrices $\mathbf{T}_n$, $n = 1, 2, \ldots$

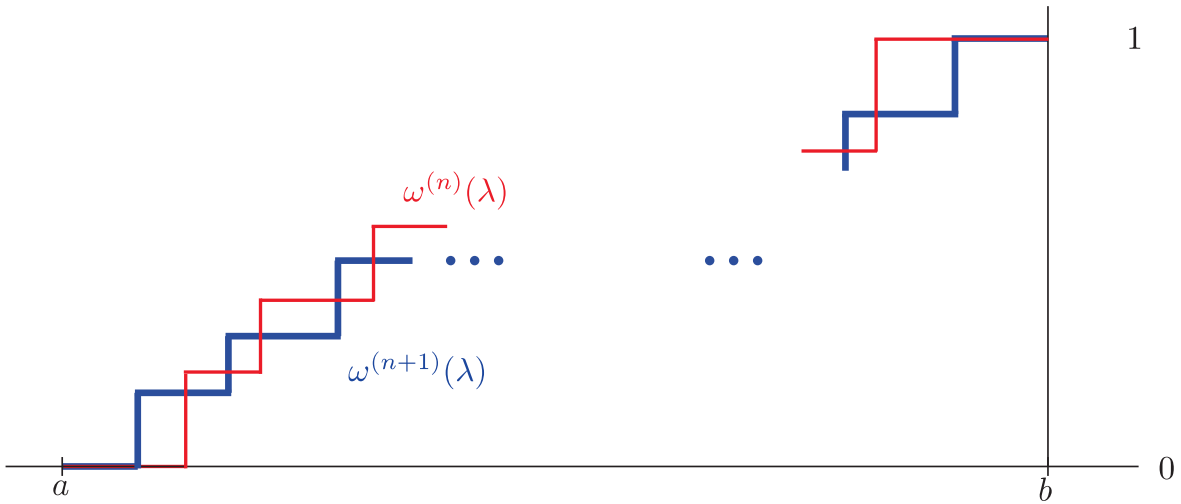
- Whenever the bottom element of a normalized eigenvector of $\mathbf{T}_n$ vanishes, the associated eigenvalue of $\mathbf{T}_n$ closely approximates an eigenvalue of $\mathbf{A}$ and an analogous approximation must exist for $\mathbf{T}_{n+1}, \mathbf{T}_{n+2}$ etc; see Paige (1971 -1980). This breakthrough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

The fundamental question: **What distribution function is behind this?** 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 the eigenvalues of $\mathbf{A}$ multiply approximated within the steps $1$ to $n$ .

# 3 Interlocking property for the modified distribution functions




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

- Mathematical structure preserved for the methods with short recurrences? Complex Jacobi matrices, Gauss quadrature in the complex plane?
- Mathematical structure preserved for Arnoldi, FOM and GMRES? Hessenberg matrices?

# 4. Preconditioning and discretization

Some more details and many references to original works 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)
- I. Pultarová, Z.S., *Decomposition into subspaces and operator preconditioning* (2017?)

Hilbert space $V$ with the 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

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

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

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

**Theorem.**

The spectrum of $\tau\mathcal{A}$ is enclosed in $[c_{\mathcal{A}}, C_{\mathcal{A}}]$ with the endpoints belonging to the spectrum,

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

$$c_{\mathcal{A}} = \inf_{v\in V,\, \|v\|_V=1} \langle \mathcal{A}v, v\rangle = \left\{ \|\mathcal{A}^{-1}\|_{\mathcal{L}(V^{\#},V)} \right\}^{-1} .$$

Linear, bounded, coercive, and self-adjoint $\mathcal{B}$ , $C_{\mathcal{B}}$ , $c_{\mathcal{B}}$ defined analogously. Define

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

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

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.

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

We are interested in the condition number (but recall *Málek, S, 2015, Chapter 11*)

$$\kappa(\mathcal{B}^{-1}\mathcal{A}) \;:=\; \|\mathcal{B}^{-1}\mathcal{A}\|_V \|\mathcal{A}^{-1}\mathcal{B}\|_V \;\leq\; \kappa(\mathcal{A})\,\kappa(\mathcal{B})$$

and in the spectral number

$$\hat{\kappa}(\mathcal{B}^{-1}\mathcal{A}) \;:=\; \frac{\sup_{u\in V,\,\|u\|_V=1}(\mathcal{B}^{-1}\mathcal{A}u,u)_V}{\inf_{v\in V,\,\|v\|_V=1}(\mathcal{B}^{-1}\mathcal{A}v,v)_V}\,.$$

Assuming the norm equivalence of $\mathcal{A}^{-1}$ and $\mathcal{B}^{-1}$, i.e.,

$$\alpha \;\leq\; \frac{\|\mathcal{A}^{-1}f\|_V}{\|\mathcal{B}^{-1}f\|_V} \;\leq\; \beta \qquad \text{for all } f\in V^{\#},\; f\neq 0\,,$$

we get

$$\kappa(\mathcal{B}^{-1}\mathcal{A}) \;\leq\; \frac{\beta}{\alpha}\,.$$

$N$-dimensional subspace $V_h \subset V$; abstract Galerkin discretization gives $u_h \in V_h$, $u_h \approx u \in V$ satisfying Galerkin orthogonality

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

The restrictions $\mathcal{A}_h : V_h \to V_h^\#$, $b_h : V_h \to \mathbb{R}$ gives 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 $V_h$

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

# 4 Preconditioning - straight consequence of the $V_h \longrightarrow V_h^{\#}$ setting

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

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

or

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

Indeed, for the restricted Riesz map $\tau_{\mathcal{B},h}$ for $\mathbf{v}$ and $\mathbf{f}$, with $f = \Phi_h^{\#}\mathbf{f}$, $v = \Phi_h \mathbf{v}$

$$(\tau_{\mathcal{B},h} f, v)_{\mathcal{B}} = (\tau_{\mathcal{B},h}\Phi_h^{\#}\mathbf{f}, \Phi_h\mathbf{v})_{\mathcal{B}} = (\Phi_h \mathbf{M}_\tau \mathbf{f}, \Phi_h \mathbf{v})_{\mathcal{B}} = \langle \mathcal{B}\Phi_h \mathbf{M}_\tau \mathbf{f}, \Phi_h \mathbf{v}\rangle = \mathbf{v}^* \mathbf{M}_h \mathbf{M}_\tau \mathbf{f},$$
$$(\tau_{\mathcal{B},h} f, v)_{\mathcal{B}} = \langle f, v\rangle = \mathbf{v}^* \mathbf{f}$$

and therefore

$$\mathbf{M}_\tau \;=\; \mathbf{M}_h^{-1}\,.$$

Using the Cholesky 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.,

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

# 4 Spectral number of the preconditioned problem

If the operators $\mathcal{A}$ and $\mathcal{B}$ are spectrally equivalent, i.e.,

$$\alpha \leq \frac{\langle \mathcal{A}w, w\rangle}{\langle \mathcal{B}w, w\rangle} \leq \beta \qquad \text{for all } w \in V,\ w \neq 0\,,$$

we get

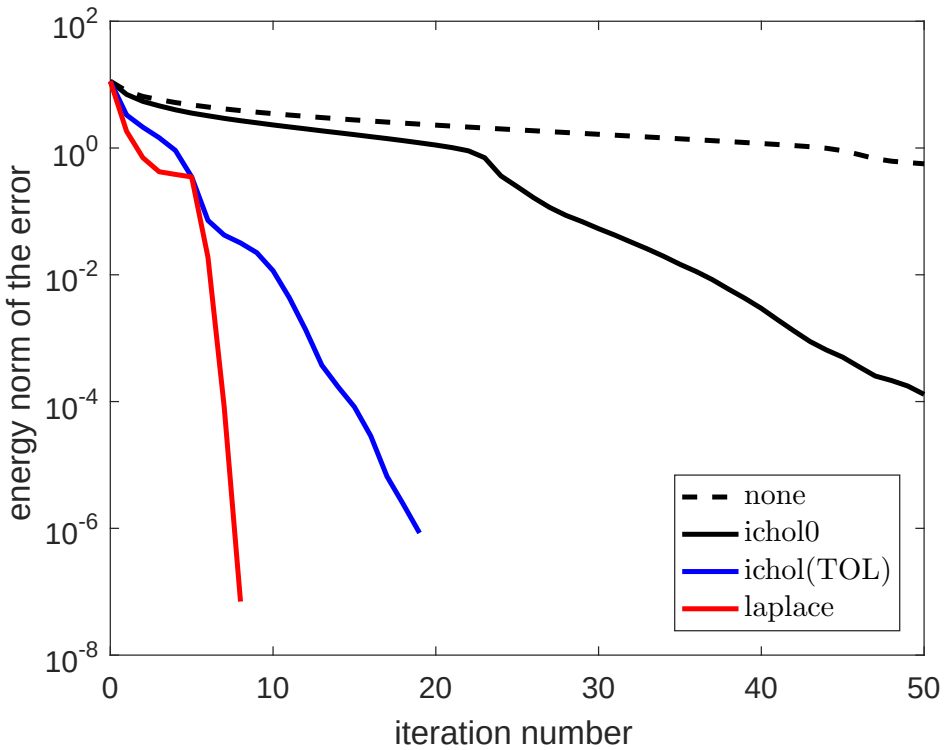
$$\hat{\kappa}(\mathbf{M}_h^{-1}\mathbf{A}_h) = \kappa(\mathbf{A}_{t,h}) \leq \frac{\beta}{\alpha}\,.$$

Recall

$$\kappa(\mathbf{M}_h^{-1}\mathbf{A}_h) = \|\mathbf{M}_h^{-1}\mathbf{A}_h\|\|\mathbf{A}_h^{-1}\mathbf{M}_h\| \neq \frac{|\lambda_{\max}(\mathbf{M}_h^{-1}\mathbf{A}_h)|}{|\lambda_{\min}(\mathbf{M}_h^{-1}\mathbf{A}_h)|} = \hat{\kappa}(\mathbf{M}_h^{-1}\mathbf{A}_h)\,.$$

Equality holds iff $\mathbf{M}_h$ and $\mathbf{A}_h$ commute (then $\mathbf{M}_h^{-1}\mathbf{A}_h = \mathbf{M}_h^{-1/2}\mathbf{A}_h\mathbf{M}_h^{-1/2}$ ).

# 4 Better conditioning does not necessarily mean faster convergence!




Nonhomogeneous diffusion tensor, uniform mesh. Unpreconditioned CG; ichol PCG (no fill-in); ichol PCG (drop-off tolerance 1e-02); Laplace operator PCG.
Condition numbers of $\mathbf{A}_{t,h}$ : 6.75e04, 4.31e02, 1.6e01, 1.61e02.

Transformation of the discretization basis

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

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

The transformed basis

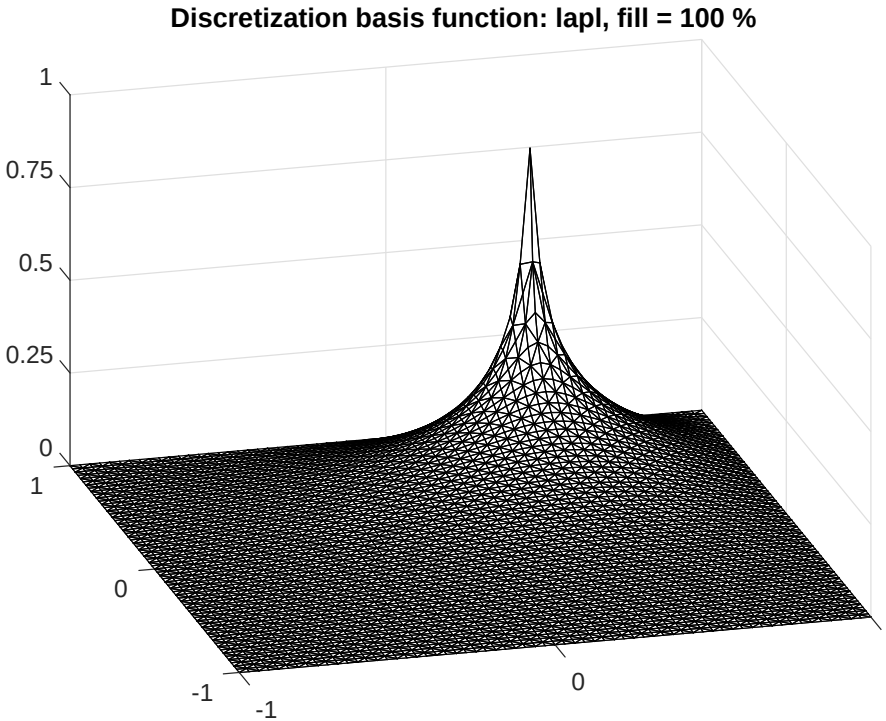
$$\Phi_{t,h} = \Phi_h (\mathbf{L}_h^*)^{-1}, \quad \Phi_{t,h}^{\#} = \Phi_h^{\#} \mathbf{L}_h$$

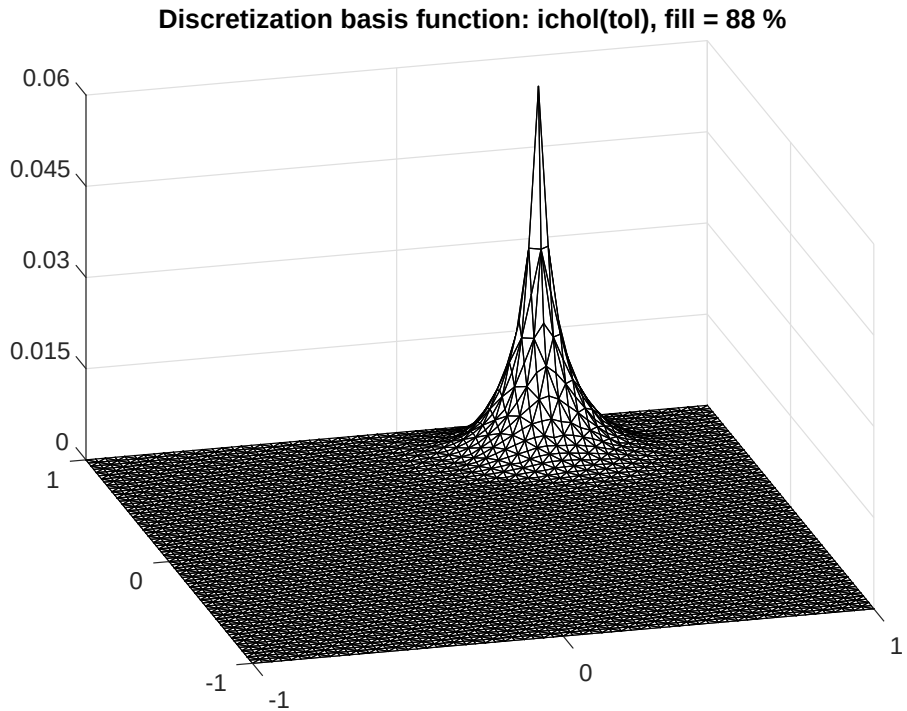
gives

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

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

# 4 Transformed FEM nodal basis elements have global support




Transformed discretization basis elements corresponding to the lapl (left) and ichol(tol) preconditioning (right).

# 5. Decomposition into subspaces and preconditioning

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

2. $$c_{V_j}\ \|u\|_V^2 \le \|u\|_j^2 \quad \text{for all } u \in V_j,\ 0 < c_{V_j},\ j \in J; \quad \text{then} \quad V^{\#} \subset V_j^{\#};$$

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

Consider

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

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

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

The preconditioned (equivalent?) problem

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

Coercivity and boundedness of $\mathcal{M}^{-1}$

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

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

gives equivalence of $\mathcal{A}\, u \;=\; b$ and $\mathcal{M}^{-1} \mathcal{A}\, u \;=\; \mathcal{M}^{-1} b$. Moreover, we have norm equivalence of $\mathcal{A}^{-1}$ and $\mathcal{M}^{-1}$ as well as spectral equivalence of $\mathcal{A}$ and $\mathcal{M}$.

**Theorem**

For any $v \in V$

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

$$\frac{\min_{j\in J} c_{\mathcal{B}_j}}{C^2_{\mathcal{A}}} \left(\sum_{k\in J} \frac{1}{c_{V_k} c_{\mathcal{B}_k}}\right)^{-1} \sum_{j\in J} \|\bar{r}_j\|^2_j \leq \|v-u\|^2_V \leq \frac{C_{\mathrm{S}} (\max_{j\in J} C_{\mathcal{B}_j})^2}{c^2_{\mathcal{A}}} \sum_{j\in J} \|\bar{r}_j\|^2_j ,$$

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

# 6. $h$-adaptivity based on the residual-based estimator

More details and references can be found in

- J. Papež, Z.S., *On a residual-based a posteriori error estimator for the total error* (IMA J. Num. Anal., accepted, 2017)

# 6 Equilibrating the discretization error over the domain?

Consider any estimator EST$(\cdot)$ that provides an upper bound for the discretization error

$$\|u - u_h\| \leq \text{EST}(u_h)\,,$$

In order to evaluate the right-hand side, we need $u_h$ that is not available.

This matter can not be resolved by simply plugging-in a computed approximation $u_h^{(n)}$ instead of $u_h$ unless we prove that this can be done;
i.e., unless it is clear that the derivation of the estimator is not based on any assumption that is violated by $u_h^{(n)}$ .

The residual-based formula derived as the lower bound for the discretization error gives with no change the lower bound for the total error. For the upper bound the situation is different.

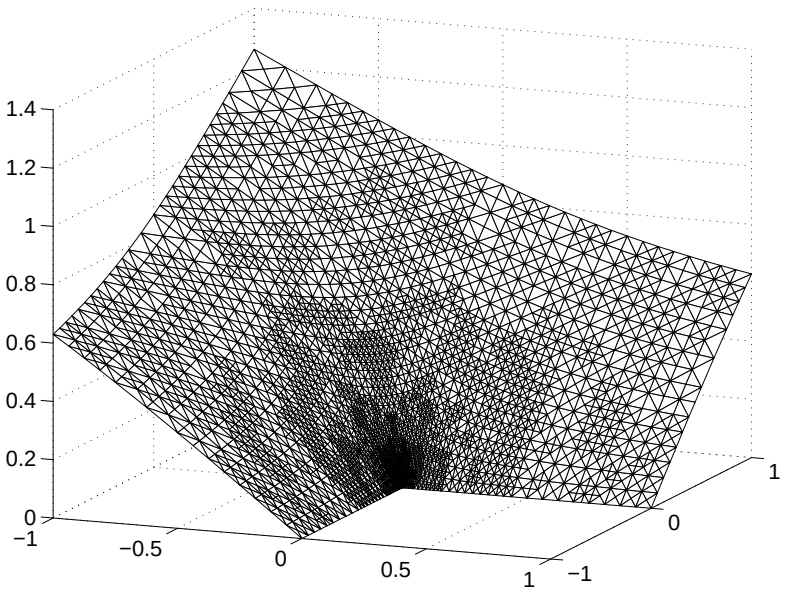
Residual-based a posteriori error bound for the total error that accounts for inexact algebraic computations, i.e., for arbitrary $v_h \in V_h$ $(v_h = u_h^{(n)})$

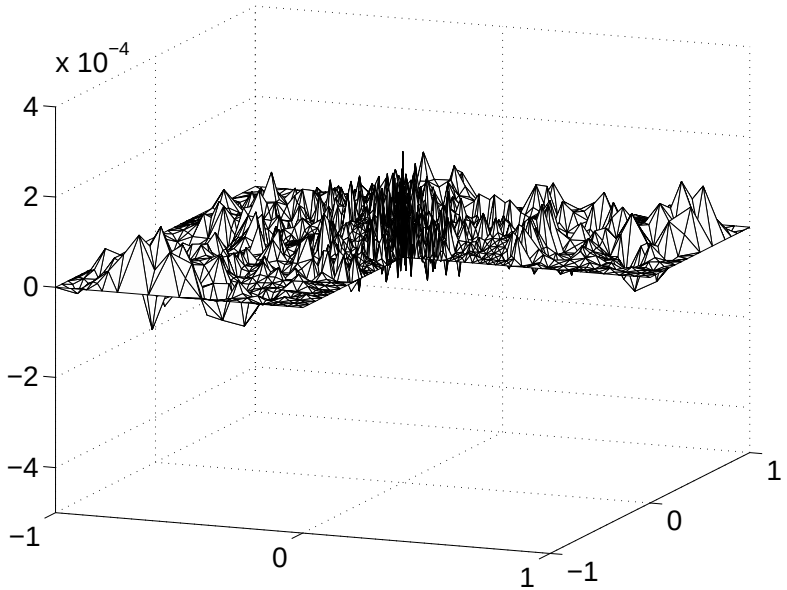
$$\|\nabla(u - v_h)\|^2 \leq 2\, C_1^2\, C_2^2 \left(J^2(v_h) + \mathrm{osc}^2\right) + 2\, \widetilde{C}_{\mathrm{intp}}^2(u, v_h)\, \|\nabla(u_h - v_h)\|^2 ,$$

where (using the linear FEM discretization basis functions)

$$J_E(v_h) \equiv |E|^{1/2} \left\| \left[ \frac{\partial v_h}{\partial n_E} \right] \right\|_E , \qquad J(v_h) \equiv \left( \sum_{E \in \mathcal{E}_{\mathrm{int}}} J_E^2(v_h) \right)^{1/2} .$$

# 6 L-shape domain

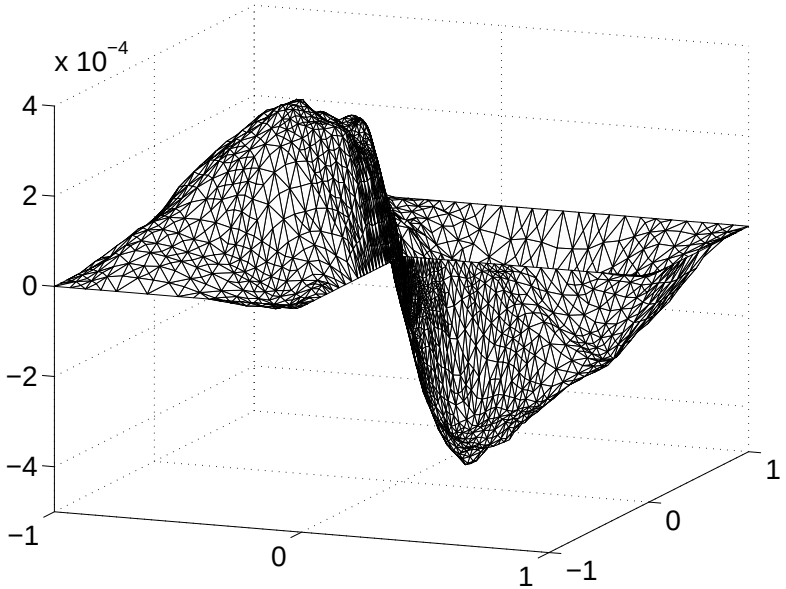


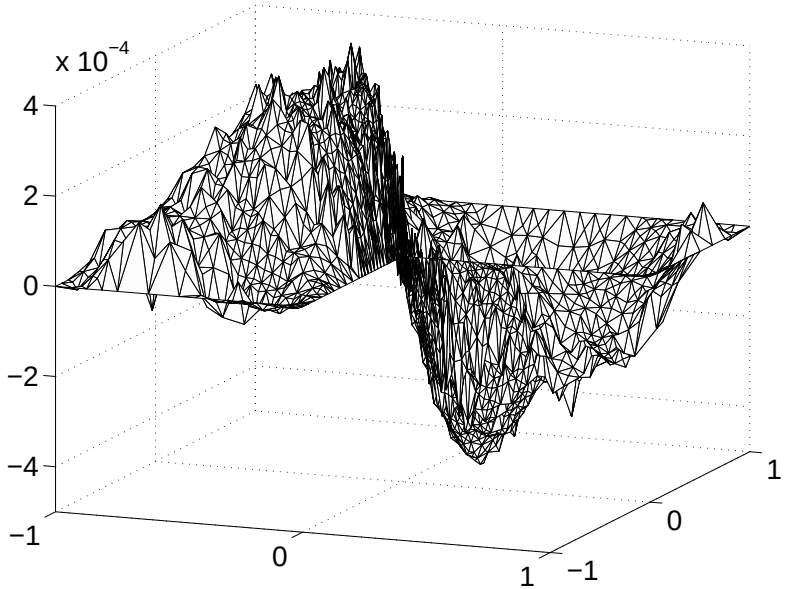




Exact solution $u$ (left) and the discretization error $u - u_h$ (right) in the Poisson model problem, linear FEM, adaptive mesh refinement.

Quasi equilibrated discretization error over the domain.

# 6 L-shape domain




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

$$\|\nabla(u - u_h)\| \gg \|\mathbf{x} - \mathbf{x}_n\|_{\mathbf{A}} .$$

# 7. Concluding remarks and outlook

# 7 Concluding remarks and outlook

- Individual steps modeling-analysis-discretization-computation should not be considered separately within isolated disciplines. They form a single problem. Operator preconditioning follows this philosophy.
- Fast HPC computations require handling all involved issues.
  A posteriori error analysis leading to efficient and reliable stopping criteria is essential ...
- Krylov subspace methods adapt to the problem. Exploiting this adaptation is the key to their efficient use.
- Daniel (1967) did not identify the CG convergence with the Chebyshev polynomials-based bound. He carefully writes (modifying slightly his notation)

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

- $\mathcal{O}(n)$ reliable approximate solvers?
  Rüde, Wohlmuth, Burstedde, Kunoth, ...

# Thank you very much for your kind patience!