Krylov subspace methods from the analytic, application and computational perspective

Zdeněk Strakoš

Charles University in Prague and Czech Academy of Sciences

http://www.karlin.mff.cuni.cz/˜strakos

Rencontres INRIA-JLJJ en calcul scientifique, Paris, April 2016

Thanks

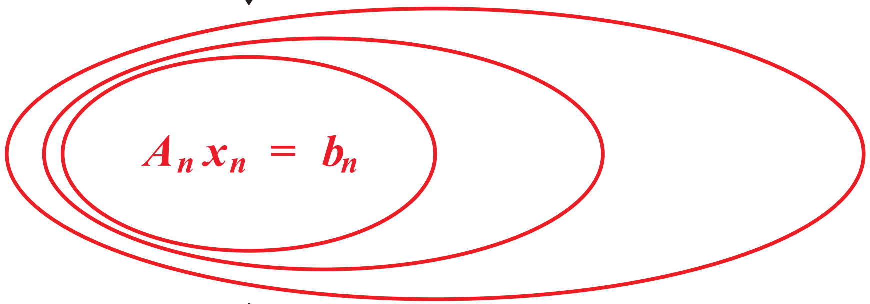
Results affected by very many authors,
and coauthored, in particular, with

Josef Málek
Tomáš Gergelits
Jan Papež
Joerg Liesen

Krylov subspace methods

$$Ax = b\,, x_0\,, \mathcal{S}_n\,, \mathcal{C}_n$$

$$\downarrow$$



$$\downarrow$$

x_n approximates the solution x in $x_0 + \mathcal{S}_n$
with $b - Ax_n$ orthogonal to $\mathcal{C}_n$

$\mathcal{S}_n\,, \mathcal{C}_n$ related to $\mathcal{K}_n(A, r_0) \equiv span\,\{r_0, Ar_0, \cdots, A^{n-1}r_0\}$

$\longrightarrow$ moments $r_0^* A^j r_0\,, j = 0, 1, 2, \cdots$

Historical development and context

1950

Krylov subspace methods
Lanczos 1950/52, Hestenes & Stiefel 1952

Modern numerical analysis
von Neumann & Goldstine 1947
Turing 1948

Krylov sequences
Gantmacher 1934

Mathematical foundations of quantum mechanics
Hilbert 1926/27,
von Neumann 1927/32,
Wintner 1929

Transformation of the characteristic equation
Krylov 1931

Representation theorem
Riesz 1909

Orthogonalisation algorithms for functions and vectors
Schmidt 1905/07, Szász 1910

Jacobi form (or matrix)
Hellinger & Toeplitz 1914

Foundations of functional analysis, including continuous spectrum, resolution of unity, self-adjoined operators, Hilbert space
Hilbert 1906-1912

Analytic theory of continued fractions, Riemann-Stieltjes integral, solution of the moment problem
Stieltjes 1894

Orthogonalisation via the Gramian
Gram 1883

Minimal polynomial
Frobenius 1878

Jordan canonical form
Weierstrass 1868,
Jordan 1870

Continued fractions and Chebyshev inequalities
Chebyshev 1855,
Markov, Stieltjes 1884

Generalisations of the Gauss quadrature, minimal partial realisation
Christoffel 1858/77

Diagonalisation of quadratic forms
Jacobi 1857

Continued fractions and three-term recurrence for orthogonal polynomials
Chebyshev 1855/59

Reduction of bilinear form to tridiagonal form
Jacobi 1848

Characteristic equation
Cauchy 1840

Gauss quadrature and orthogonal polynomials
Jacobi 1826

Real symmetric matrices have real eigenvalues, interlacing property
Cauchy 1824

Orthogonalisation idea
Laplace 1820

Gauss quadrature
Gauss 1814

Secular equation of the moon
Lagrange 1774

Infinite series expansions and continued fractions
Euler 1744/48

Mechanical quadrature
Newton, Cotes 1720s

Three-term recurrences and continued fractions
Brouncker, Wallis 1650s

1650

Lanczos, Hestenes and Stiefel

Numerical analysis

Convergence analysis   Rounding error analysis   Cost of computations   Floating point computations

Iterative methods   Polynomial preconditioning   Stopping criteria   Data uncertainty

Least squares solutions   Structure and sparsity

Gaussian elimination

Vandermonde determinant

Matrix theory

Optimisation

Convex geometry

Minimising functionals

Approximation theory

Orthogonal polynomials

Chebyshev, Jacobi and Legendre polynomials

Green's function

Gibbs oscillation

Rayleigh quotients

Fourier series

Cornelius Lanczos
An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, 1950

Solution of systems of linear equations by minimized iterations, 1952

Chebyshev polynomials in the solution of large-scale linear systems, 1952

Magnus R. Hestenes & Eduard Stiefel
Methods of conjugate gradients for solving linear systems, 1952

Linear algebra

General inner products

Cauchy-Schwarz inequality

Orthogonalisation

Projections

Functional analysis

Differential and integral operators

Liouville-Neumann expansion

Trigonometric interpolation   Continued fractions   Sturm sequences   Fredholm problem

Gauss-Christoffel quadrature   Riemann-Stieltjes integral   Dirichlet and Fejér kernel

Real analysis

Operator preconditioning

Klawonn (1995, 1996); Arnold, Falk, and Winther (1997, 1997); Steinbach and Wendland (1998); Mc Lean and Tran (1997); Christiansen and Nédélec (2000, 2000); Powell and Silvester (2003); Elman, Silvester, and Wathen (2005); Hiptmair (2006); Axelsson and Karátson (2009); Mardal and Winther (2011); Kirby (2011); Zulehner (2011); Preconditioning Conference 2013, Oxford; ...

Related ideas on spectral equivalence of operators can be found, e.g., in Faber, Manteuffel and Parter (1990) with references to D'Yakonov (1961) and Gunn(1964, 1965). .... Very nice recent work Smears (2016).

Mesh independent condition number

R. Hiptmair, CMA (2006):

“There is a continuous operator equation posed in infinite-dimensional spaces that underlines the linear system of equations [...] awareness of this connection is key to devising efficient solution strategies for the linear systems.

Operator preconditioning is a very general recipe [...]. It is simple to apply, but may not be particularly efficient, because in case of the [*condition number*] bound of Theorem 2.1 is too large, the operator preconditioning offers no hint how to improve the preconditioner. Hence, operator preconditioner may often achieve [...] the much-vaunted mesh independence of the preconditioner, but it may not perform satisfactorily on a given mesh.”

Linear asymptotic behavior?

V. Faber, T. Manteuffel and S. V. Parter, Adv. in Appl. Math. (1990):

“For a fixed h, using a preconditioning strategy based on an equivalent operator may not be superior to classical methods [...] Equivalence alone is not sufficient for a good preconditioning strategy. One must also choose an equivalent operator for which the bound is small.

There is no flaw in the analysis, only a flaw in the conclusions drawn from the analysis [...] asymptotic estimates ignore the constant multiplier. Methods with similar asymptotic work estimates may behave quite differently in practice.”

Outline

1. Numerical solution of BVPs
2. Operator preconditioning
3. Algebraic preconditioning, discretization, and problem formulation
4. Various comments
5. Conclusions

1 Notation

Let V be an infinite dimensional Hilbert space with the inner product

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

$V^\#$ be the dual space of bounded (continuous) linear functionals on V with the duality pairing

$$\langle \cdot,\cdot \rangle : V^\# \times V \to \mathbb{R}\,.$$

For each $f \in V^\#$ there exists a unique $\tau f \in V$ such that

$$\langle f, v \rangle = (\tau f, v)_V \quad \text{for all } v \in V\,.$$

In this way the inner product $(\cdot,\cdot)_V$ determines the Riesz map

$$\tau : \; V^\# \to V\,.$$

1 Weak formulation of the BVP, assumptions

Let $a(\cdot,\cdot) = V \times V \to R$ be a bounded and coercive bilinear form. For $u \in V$ we can write the bounded linear functional $a(u,\cdot)$ on V as

$$\mathcal{A}u \equiv a(u,\cdot) \in V^{\#}\,, \quad \text{i.e.}\,,$$

$$\langle \mathcal{A}u, v\rangle = a(u,v) \quad \text{for all } v \in V\,.$$

This defines the bounded and coercive operator

$$\mathcal{A} : V \to V^{\#}, \qquad \inf_{u \in V,\, \|u\|_V = 1} \langle \mathcal{A}u, u\rangle = \alpha > 0,\ \|\mathcal{A}\| = C\,.$$

The Lax-Milgram theorem ensures that for any $b \in V^{\#}$ there exists a unique solution $x \in V$ of the problem

$$a(x,v) = \langle b, v\rangle \quad \text{for all } v \in V\,.$$

1 Operator problem formulation

Equivalently,

$$\langle \mathcal{A}x - b, v \rangle = 0 \quad \text{for all } v \in V \,,$$

which can be written as the equation in $V^{\#}$,

$$\mathcal{A}x = b \,, \qquad \mathcal{A} : V \to V^{\#}, \quad x \in V, \quad b \in V^{\#} \,.$$

We will consider $\mathcal{A}$ self-adjoint with respect to the duality pairing $\langle \cdot, \cdot \rangle$.

1 Discretization using $V_h \subset V$

Let $\Phi_h = (\phi_1^{(h)}, \ldots, \phi_N^{(h)})$ be a basis of the subspace $V_h \subset V$,
let $\Phi_h^\# = (\phi_1^{(h)\#}, \ldots, \phi_N^{(h)\#})$ be the canonical basis of its dual $V_h^\#$.

The Galerkin discretization then gives

$$\mathcal{A}_h x_h = b_h\,, \quad x_h \in V_h\,, \quad b_h \in V_h^\#\,, \quad \mathcal{A}_h : V_h \to V_h^\#\,.$$

Using the coordinates $x_h = \Phi_h \mathbf{x}\,,\ b_h = \Phi_h^\# \mathbf{b}$, the discretization results in the linear algebraic system

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

1 Computation

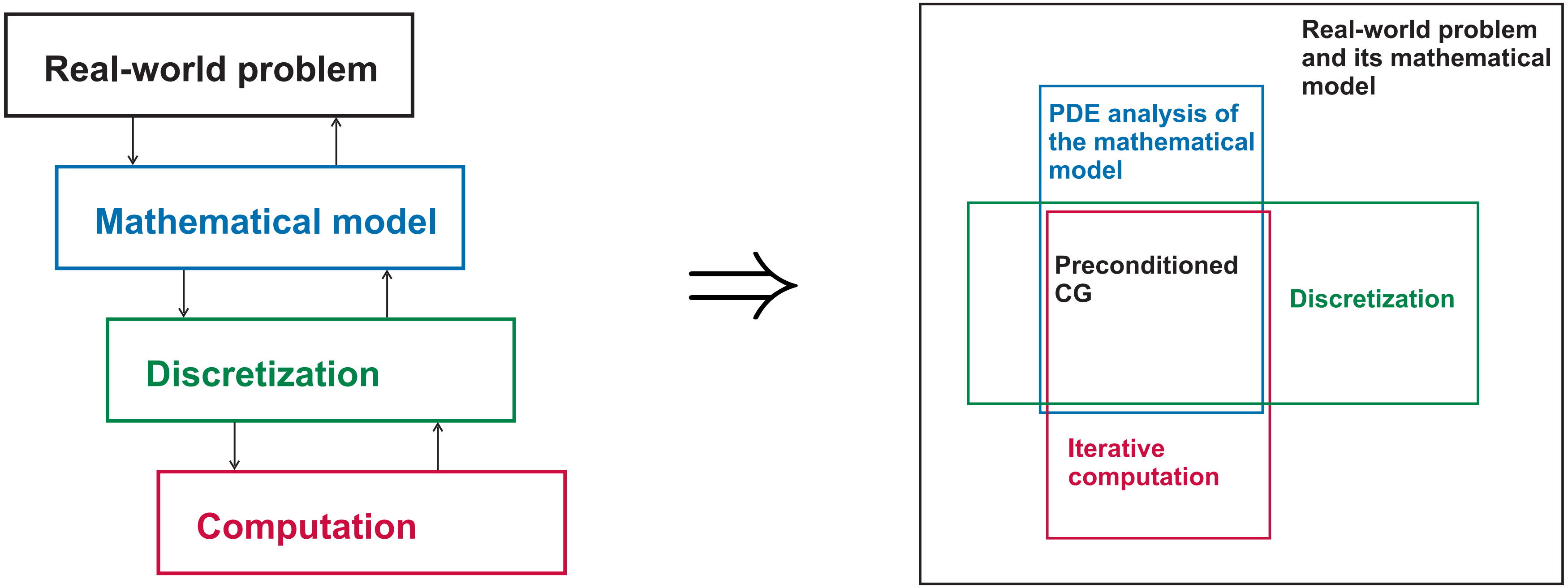
Preconditioning needed for accelerating the iterations is then often build up algebraically for the given matrix problem, giving (here illustrated as the left preconditioning)

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

Then the CG method is applied to the (symmetrized) preconditioned system, i.e., (PCG) ($\mathbf{M}$-preconditioned CG) is applied to the unpreconditioned system. The schema of the solution process:

$$\mathcal{A},\ \langle b, \cdot\rangle \ \rightarrow\ \mathbf{A}, \mathbf{b} \rightarrow \ \text{preconditioning} \ \rightarrow\ \text{PCG applied to } \mathbf{A}\mathbf{x} = \mathbf{b}\,.$$

1 This talk presents a bit different view



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.

Outline

1. Numerical solution of BVPs

2. Operator preconditioning

3. Algebraic preconditioning, discretization, and problem formulation

4. Various comments

5. Conclusions

2 Operator formulation of the problem

Recall that the inner product $(\cdot,\cdot)_V$ defines the Riesz map τ.
It can be used to transform the equation in $V^\#$

$$\mathcal{A}x = b\,, \qquad \mathcal{A} : V \rightarrow V^\#, \quad x \in V, \quad b \in V^\#\,.$$

into the equation in V

$$\tau\mathcal{A}x = \tau b, \qquad \tau\mathcal{A} : V \rightarrow V, \quad x \in V, \quad \tau b \in V\,,$$

This transformation is called operator preconditioning; see Klawonn (1995, ...), Arnold, Winther et al (1997, ...), ...

2 The mathematically best preconditioning?

With the choice of the inner product $(\cdot,\cdot)_V = a(\cdot,\cdot)$ we get

$$a(u,v) = \langle \mathcal{A}u, v\rangle = a(\tau \mathcal{A}\, u, v)$$

i.e.,

$$\tau = \mathcal{A}^{-1}\,,$$

and the preconditioned system

$$x = \mathcal{A}^{-1} b\,.$$

2 CG in infinite dimensional Hilbert spaces

$r_0 = b - \mathcal{A}x_0 \in V^{\#}, \quad p_0 = \tau r_0 \in V$. For $n = 1, 2, \ldots, n_{\max}$

$$\alpha_{n-1} = \frac{\langle r_{n-1}, \tau r_{n-1} \rangle}{\langle \mathcal{A} p_{n-1}, p_{n-1} \rangle}$$

$$x_n = x_{n-1} + \alpha_{n-1} p_{n-1} \,,$$ stop when the stopping criterion is satisfied

$$r_n = r_{n-1} - \alpha_{n-1} \mathcal{A} p_{n-1}$$

$$\beta_n = \frac{\langle r_n, \tau r_n \rangle}{\langle r_{n-1}, \tau r_{n-1} \rangle}$$

$$p_n = \tau r_n + \beta_n p_{n-1}$$

Hayes (1954); Vorobyev (1958, 1965); Karush (1952); Stesin (1954)
Superlinear convergence for (identity + compact) operators. Here the Riesz map τ indeed serves as the preconditioner.

2 Discretization of the infinite dimensional CG

Using the coordinates in the bases Φ_h and $\Phi_h^\#$ of V_h and $V_h^\#$ respectively, $(V_h^\# = \mathcal{A} V_h)$,

$$\langle f, v \rangle \to \mathbf{v}^* \mathbf{f} \,,$$

$$(u, v)_V \to \mathbf{v}^* \mathbf{M} \mathbf{u}, \qquad (\mathbf{M}_{ij}) = ((\phi_j, \phi_i)_V)_{i,j=1,\dots,N} \,,$$

$$\mathcal{A} u \to \mathbf{A} \mathbf{u} \,, \qquad \mathcal{A} u = \mathcal{A} \Phi_h \mathbf{u} = \Phi_h^\# \mathbf{A} \mathbf{u} \,; \quad (\mathbf{A}_{ij}) = (a(\phi_j, \phi_i))_{i,j=1,\dots,N} \,,$$

$$\tau f \to \mathbf{M}^{-1} \mathbf{f} \,, \qquad \tau f = \tau \Phi_h^\# \mathbf{f} = \Phi_h \mathbf{M}^{-1} \mathbf{f} \,;$$

we get with $b = \Phi_h^\# \mathbf{b}$, $x_n = \Phi_h \, \mathbf{x}_n$, $p_n = \Phi_h \, \mathbf{p}_n$, $r_n = \Phi_h^\# \, \mathbf{r}_n$
the algebraic CG formulation

2 Galerkin discretization gives matrix CG in V_h

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$, solve $\mathbf{M}\mathbf{z}_0 = \mathbf{r}_0$, $\mathbf{p}_0 = \mathbf{z}_0$. For $n = 1, \dots, n_{\max}$

$$\alpha_{n-1} = \frac{\mathbf{z}_{n-1}^* \mathbf{r}_{n-1}}{\mathbf{p}_{n-1}^* \mathbf{A} \mathbf{p}_{n-1}}$$

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \alpha_{n-1}\mathbf{p}_{n-1}\,,$$ stop when the stopping criterion is satisfied

$$\mathbf{r}_n = \mathbf{r}_{n-1} - \alpha_{n-1}\mathbf{A}\mathbf{p}_{n-1}$$

$$\mathbf{M}\mathbf{z}_n = \mathbf{r}_n\,,$$ solve for $\mathbf{z}_n$

$$\beta_n = \frac{\mathbf{z}_n^* \mathbf{r}_n}{\mathbf{z}_{n-1}^* \mathbf{r}_{n-1}}$$

$$\mathbf{p}_n = \mathbf{z}_n + \beta_n \mathbf{p}_{n-1}$$

Günnel, Herzog, Sachs (2014); Málek, S (2015)

2 Philosophy of the a-priori robust bounds

The bound

$$\kappa(\mathbf{M}^{-1}\mathbf{A}) \;\leq\; \frac{\sup_{u,v\in V,\,\|u\|_V=1,\|v\|_V=1} |\langle \mathcal{A}u, v\rangle|}{\inf_{u\in V,\,\|u\|_V=1} \langle \mathcal{A}u, u\rangle}$$

is valid independently of the discretization, see, e.g., Hiptmair (2006). If the bound is small enough, then the matter about the rate of convergence is resolved.

2 Observations

- Unpreconditioned CG, i.e. $\mathbf{M} = \mathbf{I}$, corresponds to the discretization basis Φ orthonormal wrt $(\cdot, \cdot)_V$.

- Orthogonalization of the discretization basis with respect to the given inner product in V will result in the unpreconditioned CG that is applied to the transformed (preconditioned) algebraic system. The resulting orthogonal discretization basis functions do not have local support and the transformed matrix is not sparse.

- Orthogonalization is not unique. For the same inner product we can get different bases and different discretized systems with exactly the same convergence behaviour.

Outline

1. Numerical solution of BVPs

2. Operator preconditioning

3. Algebraic preconditioning, discretization, and problem formulation

4. Various comments

5. Conclusions

3 Algebraic preconditioning?

Consider an algebraic preconditioning with the (SPD) preconditioner

$$\widehat{\mathbf{M}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\,(\mathbf{Q}\mathbf{Q}^*)\,\widehat{\mathbf{L}}^*$$

Where $\mathbf{Q}\mathbf{Q}^* = \mathbf{Q}^*\mathbf{Q} = \mathbf{I}\,.$

Question: Can any algebraic preconditioning be expressed in the operator preconditioning framework? How does it link with the discretization and the choice of the inner product in V ?

3 Change of the basis and of the inner product

Transform the discretization bases

$$\widehat{\Phi} = \Phi\,((\widehat{\mathbf{L}}\mathbf{Q})^*)^{-1}, \quad \widehat{\Phi}^{\#} = \Phi^{\#}\,\widehat{\mathbf{L}}\mathbf{Q}\,.$$

with the change of the inner product in V (recall $(u,v)_V = \mathbf{v}^*\mathbf{M}\mathbf{u}$)

$$(u,v)_{\text{new},V} = (\widehat{\Phi}\widehat{\mathbf{u}}, \widehat{\Phi}\widehat{\mathbf{v}})_{\text{new},V} := \widehat{\mathbf{v}}^*\widehat{\mathbf{u}} = \mathbf{v}^*\widehat{\mathbf{L}}\mathbf{Q}\mathbf{Q}^*\widehat{\mathbf{L}}^*\mathbf{u} = \mathbf{v}^*\widehat{\mathbf{L}}\widehat{\mathbf{L}}^*\mathbf{u} = \mathbf{v}^*\widehat{\mathbf{M}}\mathbf{u}\,.$$

The discretized Hilbert space formulation of CG gives the algebraically preconditioned matrix formulation of CG with the preconditioner $\widehat{\mathbf{M}}$

(more specifically, it gives the unpreconditioned CG applied to the algebraically preconditioned discretized system).

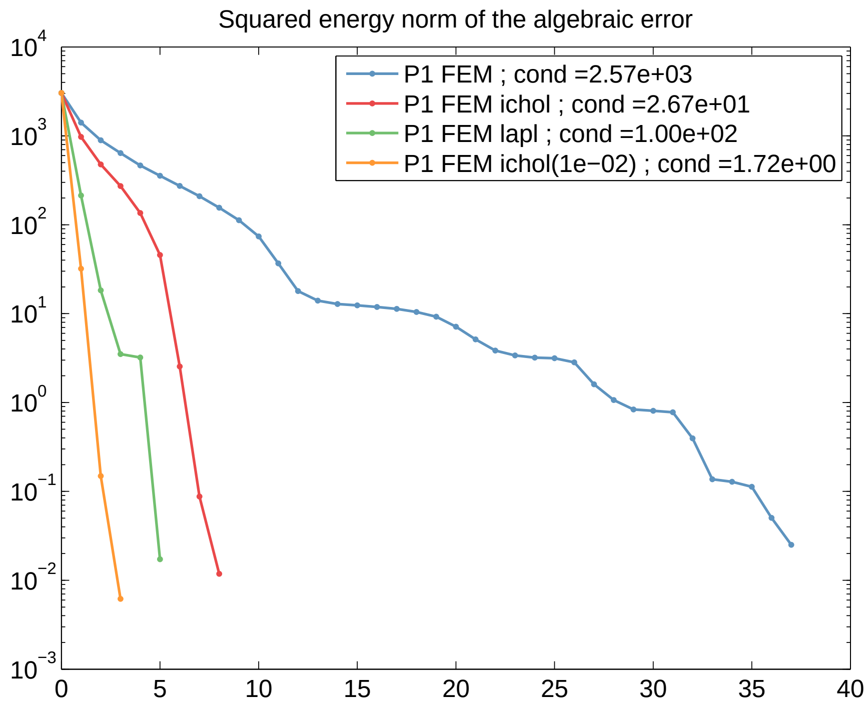
3 Sparsity, locality, global transfer of information

Sparsity of matrices of the algebraic systems is always presented as an advantage of the FEM discretizations.

Sparsity means locality of information in the individual matrix rows/columns. Getting a sufficiently accurate approximation to the solution may then require a substantial global transfer of information over the domain, i.e., a large dimension of the Krylov space.

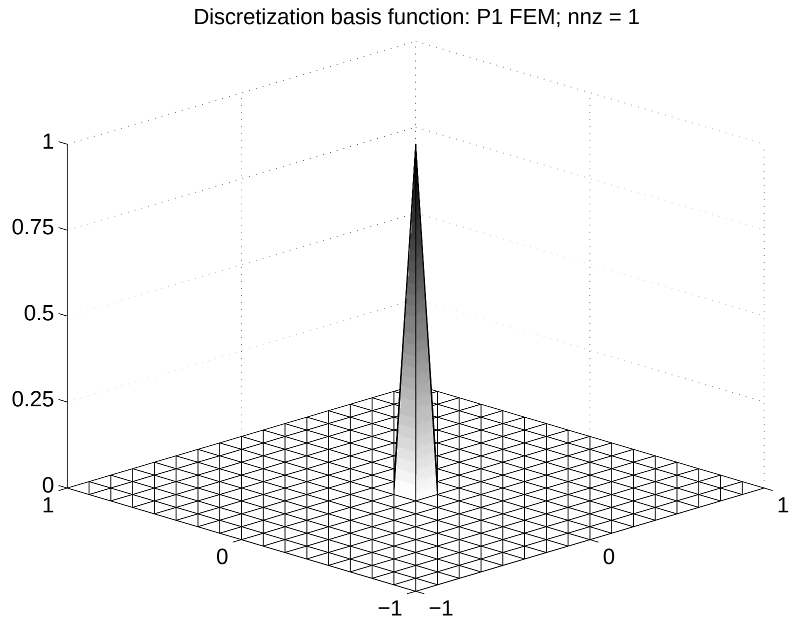
Preconditioning can be interpreted in part as addressing the unwanted consequence of sparsity (locality of the supports of the basis functions). Globally supported basis functions (hierarchical bases preconditioning, DD with coarse space components, multilevel methods, hierarchical grids etc.) can efficiently handle the transfer of global information.

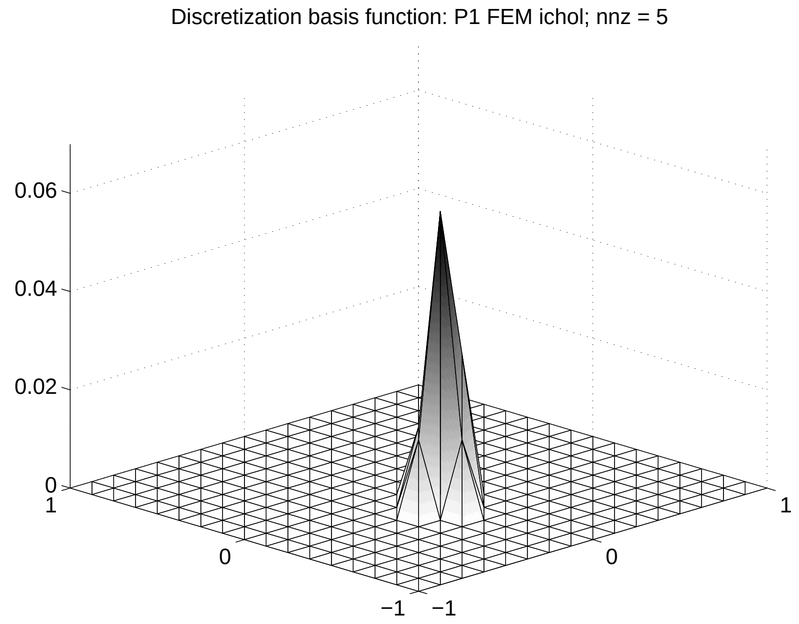
3 Example - Nonhomogeneous diffusion tensor



PCG convergence: unpreconditioned; ichol (no fill-in); Laplace operator preconditioning; ichol (drop-off tolerance 1e-02). Uniform mesh, condition numbers 2.5e03, 2.6e01, 1.0e02, 1.7e00.

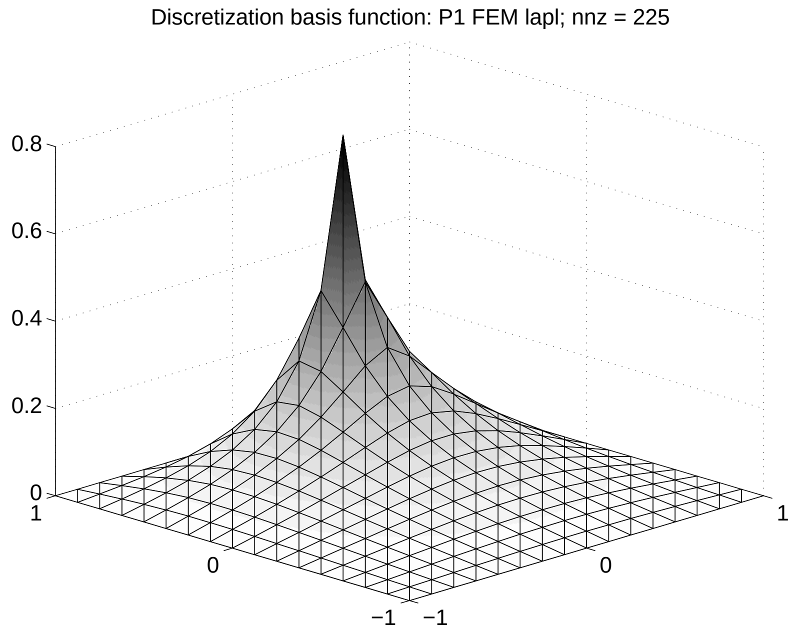
3 Transformed basis elements

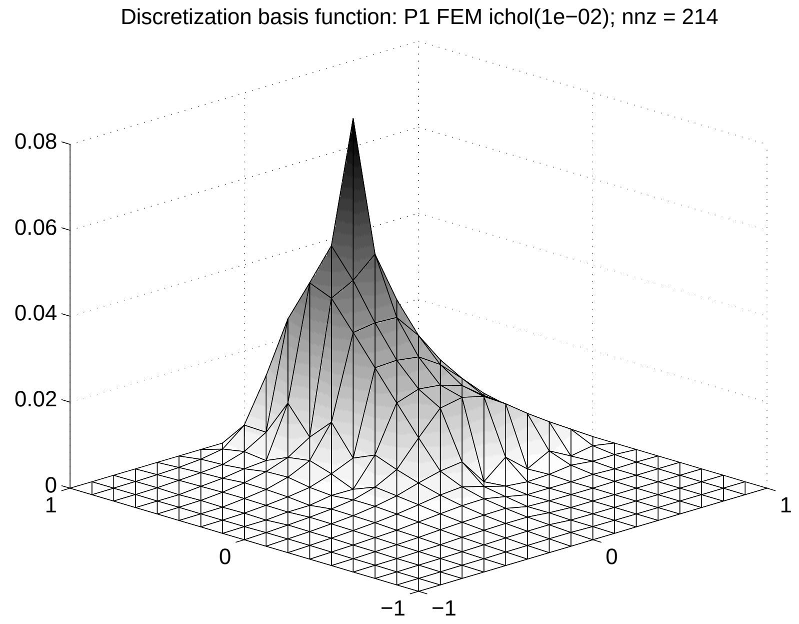




Original discretization basis element and its transformation corresponding to the ichol preconditioning.

3 Transformed basis elements





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

Outline

4 Model reduction using Krylov subspaces

Consider $\mathcal{B} = \tau\mathcal{A}\,,\ z_0 = \tau b - \tau\mathcal{A}x_0\,,$ and the Krylov sequence
$z_0,\ z_1 = \mathcal{B}z_0,\ z_2 = \mathcal{B}z_1 = \mathcal{B}^2 z_0,\ \ldots\ , z_n = \mathcal{B}z_{n-1} = \mathcal{B}^n z_0,\ \ldots$

Determine a sequence of operators $\mathcal{B}_n$ defined on the sequence of the nested subspaces $V_n =$ span $\{z_0, \ldots, z_{n-1}\}$, with the projector E_n onto V_n ,

$$\mathcal{B}_n \ = \ E_n\,\mathcal{B}\,E_n\,.$$

Convergence

$$\mathcal{B}_n \ \to \ \mathcal{B}\,?$$

4 Vorobyev moment problem

The finite dimensional operators $\mathcal{B}_n$ can be used to obtain approximate solutions to various linear problems. The choice of $z_0, z_1, \ldots$ as above gives a sequence of Krylov subspaces that are determined by the operator $\mathcal{B}$ and the initial element z_0 . In this way the Vorobyev method of moments gives the Krylov subspace methods.

Vorobyev (1958, 1965) covers bounded linear operators, bounded self-adjoint operators and some unbounded extensions. He made links to CG, Lanczos, Stieltjes moment problem, work of Markov, Gauss-Christoffel quadrature . . .

4 Conjugate gradient method - first n steps

The first n steps of the (infinite or finite dimensional) CG method are given by

$$\mathbf{T}_n \mathbf{y}_n = \|z_0\|_V \, \mathbf{e}_1 , \qquad x_n = x_0 + Q_n \mathbf{y}_n \, , \quad x_n - x_0 \in V_n \, .$$

Assume an approximation to the the n-th Krylov subspace K_n is taken as the finite dimensional discretization subspace $V_h \subset V$ in

$$\{\mathcal{A}, b, x_0, \tau\} \;\rightarrow\; \{\tau \mathcal{A}_n \,:\, K_n \rightarrow K_n\} \;\rightarrow\; \text{PCG with } \{\mathbf{A}_h, \mathbf{M}_h\} \, ?$$

Then we get a close to optimal discretization (CG minimizes the energy norm over the discretization subspaces).

4 Gauss-Christoffel quadrature

$$\mathcal{B}\,x \;=\; f \qquad \longleftrightarrow \qquad \omega(\lambda), \qquad \int F(\lambda)\,d\omega(\lambda)$$

$$\uparrow \qquad\qquad\qquad\qquad \uparrow$$

$$\mathbf{T}_n\,\mathbf{y}_n \;=\; \|z_0\|_V\,\mathbf{e}_1 \qquad \longleftrightarrow \qquad \omega^{(n)}(\lambda), \quad \sum_{i=1}^{n} \omega_i^{(n)}\,F\left(\theta_i^{(n)}\right)$$

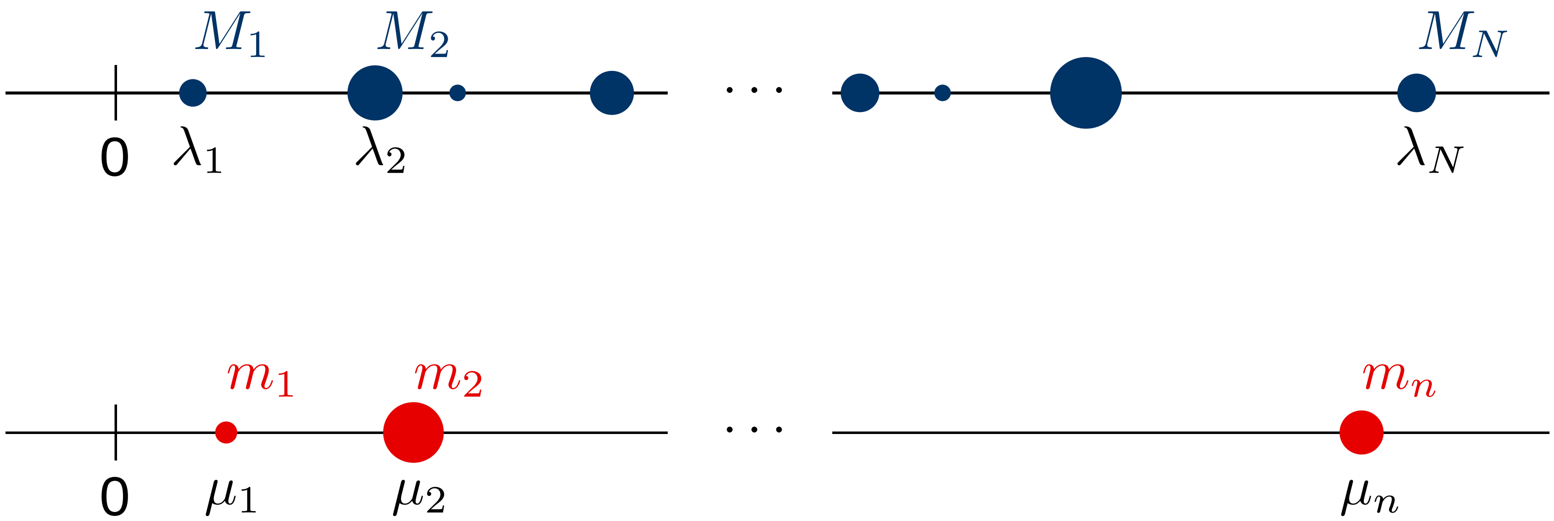
Using $F(\lambda) = \lambda^{-1}$ gives (assuming coercivity)

$$\int_{\lambda_L}^{\lambda_U} \lambda^{-1}\,d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)}\left(\theta_i^{(n)}\right)^{-1} + \frac{\|x - x_n\|_a^2}{\|f\|_V^2}$$

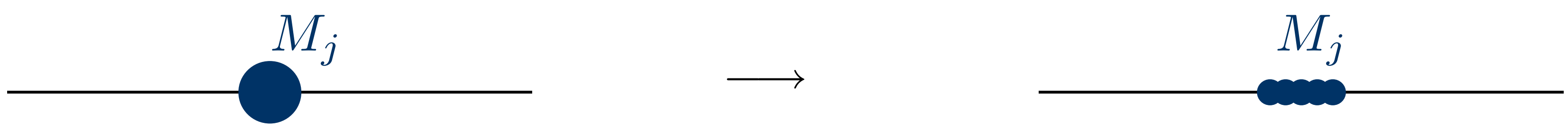
Stieltjes (1894) and Vorobyev (1958) moment problems for self-adjoint bounded operators reduce to the Gauss-Christoffel quadrature (1814).
No one would consider describing it by contraction.

4 Convergence via distribution functions

Consider the (blue) distribution function determined by the operator $\tau\mathcal{A}$ and the normalized τr_0 . For a given n , find the (red) distribution function with n mass points that matches the maximal number (2n) of the first moments $((\tau\mathcal{A})^\ell \tau r_0, \tau r_0)_V$, $\ell = 0, 1, 2, \cdots$



4 Clusters of eigenvalues mean fast convergence?



single eigenvalue λ_j $\longrightarrow$ many close eigenvalues $\widehat{\lambda}_{j_1}, \widehat{\lambda}_{j_2}, \ldots, \widehat{\lambda}_{j_\ell}$

Replacing a single eigenvalue by a tight cluster can make a substantial difference; Greenbaum (1989); Greenbaum, S (1992); Golub, S (1994).

If it does not, then it means that CG can not adapt to the problem, and it converges almost linearly. **In such cases - is it worth using?**

4 Rounding errors can be an important issue

- If preconditioning ensures getting an acceptable solution in a very few iterations, then rounding errors are not of concern.
- However, hard problems do exist. Then rounding errors can not be ignored.
- Descriptions of Krylov subspace methods that are based on contractions (condition numbers) are, in general, not descriptive.
- Analogy with a-priori and a-posteriori analysis in numerical PDEs.

The power of Krylov subspace methods is in their
self-adaptation to the problem!

4 Bounded invertible operators

Consider a bounded linear operator $\mathcal{B}$ on a Hilbert space V that has a bounded inversion, and the problem

$$\mathcal{B}\, u \;=\; f\,.$$

- Since the identity operator on an infinite dimensional Hilbert space is not compact and $\mathcal{B}\mathcal{B}^{-1} = \mathcal{I}\,,$ it follows that $\mathcal{B}$ can not be compact.

- A uniform limit (in norm) of finite dimensional (approximation) operators $\mathcal{B}_n$ is a compact operator.

- Results on strong convergence (pointwise limit); for the method of moments see Vorobyev (1958, 1965)

$$\|\mathcal{B}_n\, w - \mathcal{B}\, w\| \to 0 \quad \forall w \in V\,.$$

4 Invalid argument

Let $\mathcal{Z}_h$ be a numerical approximation of the bounded operator $\mathcal{Z}$ such that, with an appropriate extension, $\|\mathcal{Z} - \mathcal{Z}_h\| = \mathcal{O}(h)$.

Then we have $[(\lambda - \mathcal{Z})^{-1} - (\lambda - \mathcal{Z}_h)^{-1}] = \mathcal{O}(h)$ uniformly for $\lambda \in \Gamma$, where Γ surrounds the spectrum of $\mathcal{Z}$ with a distance of order $\mathcal{O}(h)$ or more. For any polynomial p

$$p(\mathcal{Z}) - p(\mathcal{Z}_h) = \frac{1}{2\pi i} \int_\Gamma p(\lambda)[(\lambda - \mathcal{Z})^{-1} - (\lambda - \mathcal{Z}_h)^{-1}]\, d\lambda\,,$$

and it seems that one can investigate $p(\mathcal{Z})$ instead of $p(\mathcal{Z}_h)$.

But the *assumption* $\|\mathcal{Z} - \mathcal{Z}_h\| = \mathcal{O}(h)\,,\ h \rightarrow 0$ does not hold for any bounded invertible infinite dimensional operator $\mathcal{Z}$.

4 Any GMRES convergence with any spectrum

1° The spectrum of $\mathbf{A}$ is given by $\{\lambda_1, \dots, \lambda_N\}$ and GMRES$(\mathbf{A}, \mathbf{b})$ yields residuals with the prescribed nonincreasing sequence $(x_0 = 0)$

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

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

$$\mathbf{A} = \mathbf{W}\mathbf{R}\mathbf{C}\mathbf{R}^{-1}\mathbf{W}^* \quad \text{and} \quad \mathbf{b} = \mathbf{W}\mathbf{h} \,.$$

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

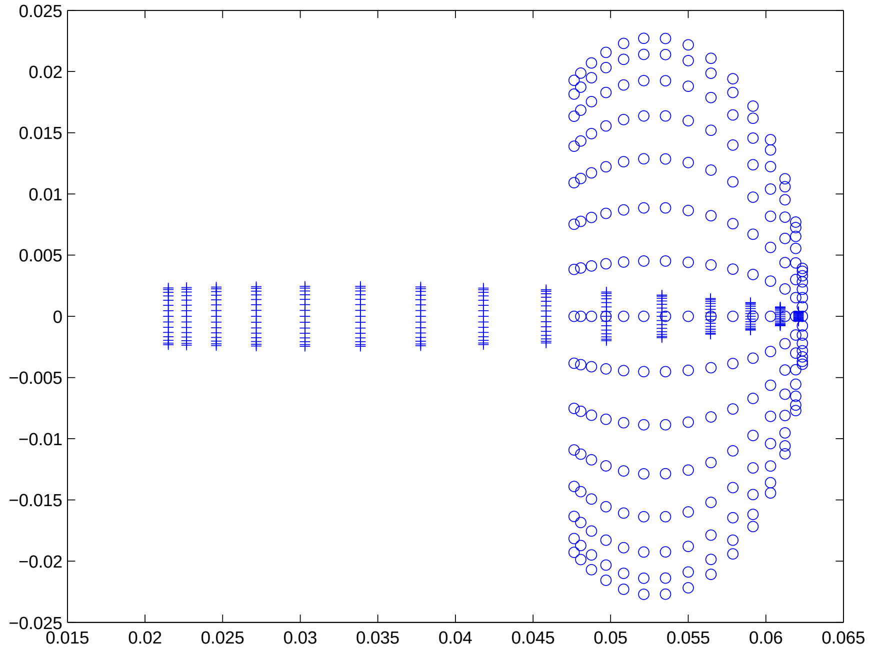
4 Interpretation?

Given any spectrum and any sequence of the nonincreasing residual norms, this gives 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 conform to convergence behaviour 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.

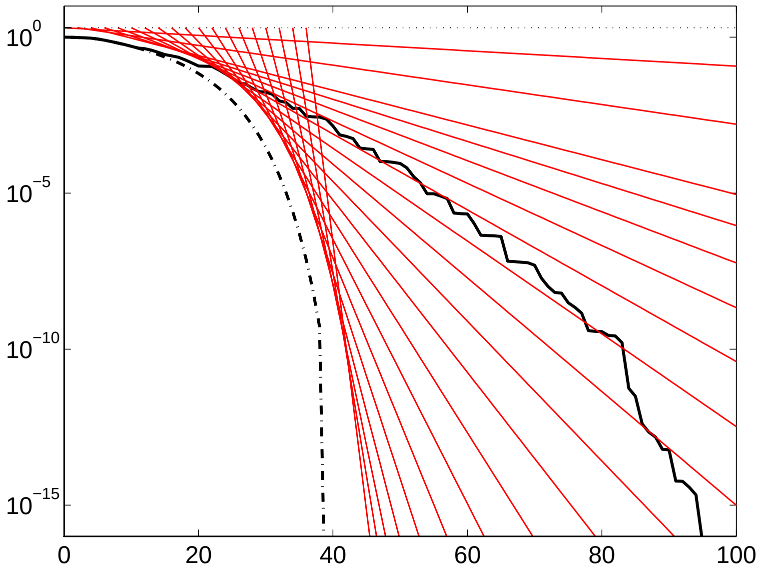
Equivalent orthogonal matrices, Greenbaum, S (1994).
Pseudospectrum indication!

4 Convection-diffusion model problem

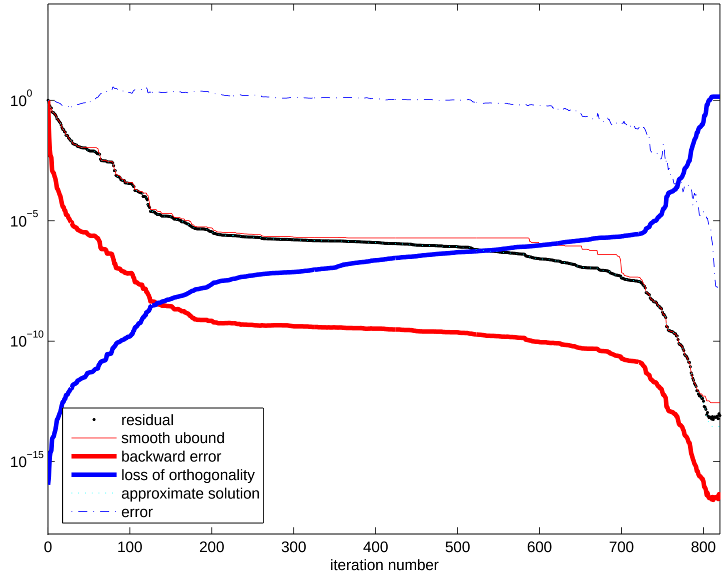


Quiz: In one case the convergence of GMRES is substantially faster than in the other; for the solution see Liesen, S (2005).

4 Delay of convergence due to inexactness

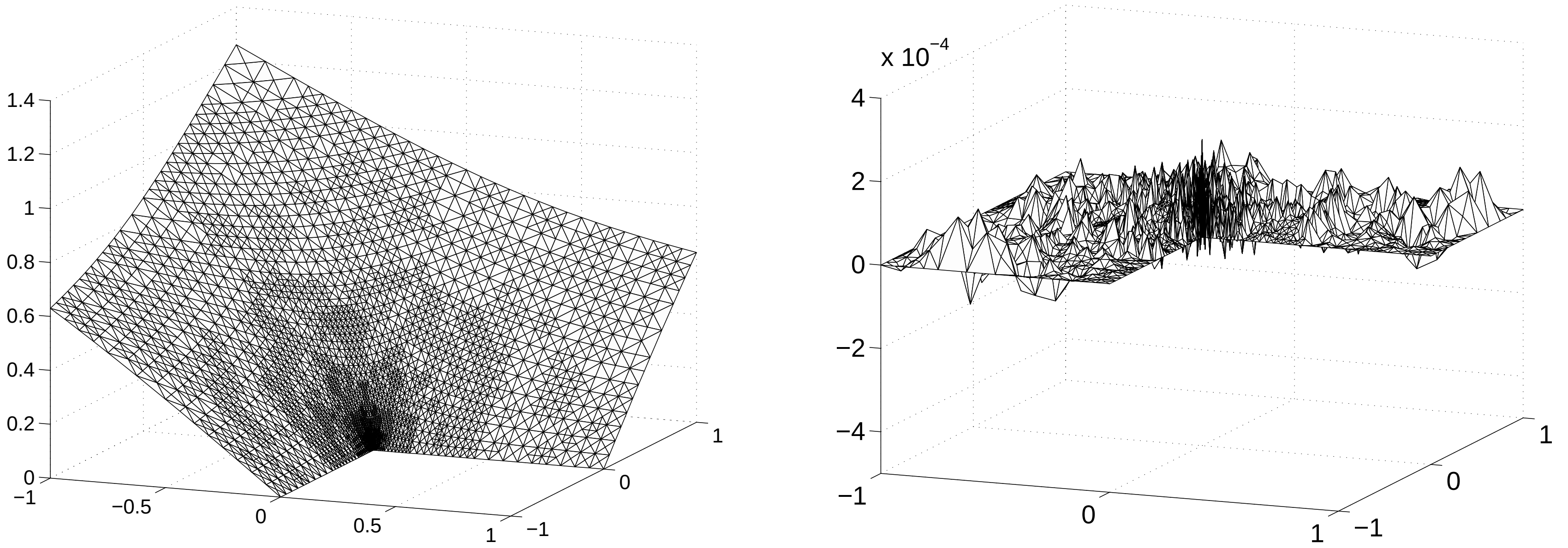


?



Here numerical inexactness due to roundoff. How much may we relax accuracy of the most costly operations without causing an unwanted delay and/or affecting the maximal attainable accuracy?

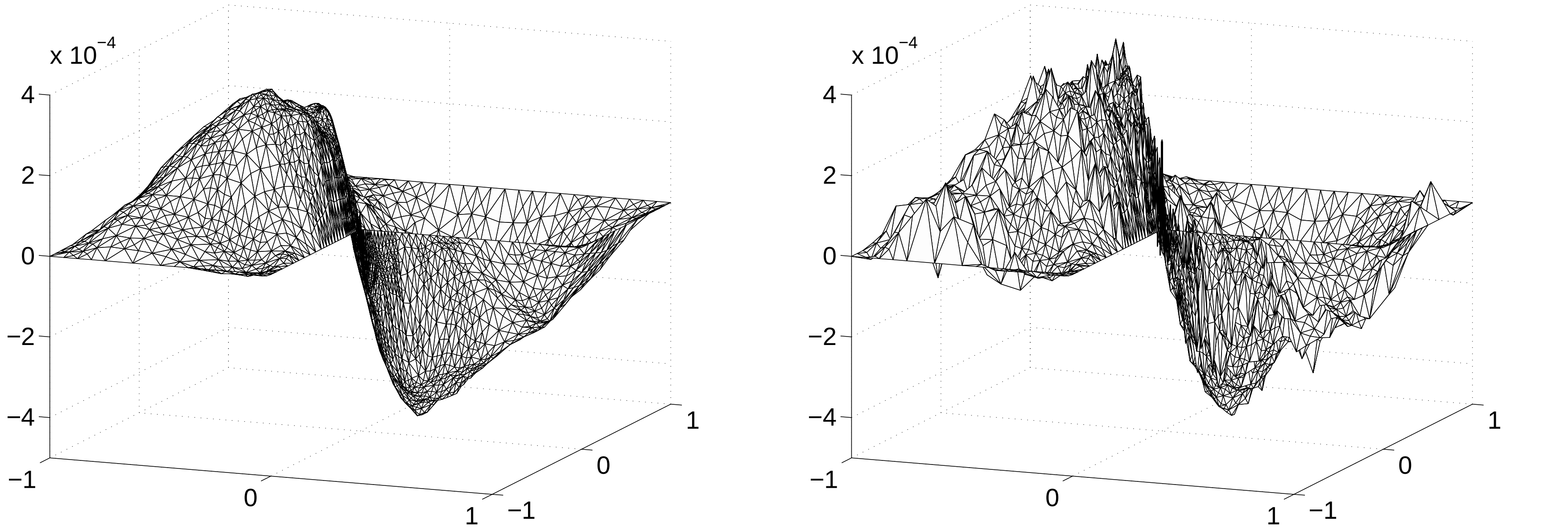
4 Stopping criteria?



Exact solution x (left) and the discretisation error $x - x_h$ (right) in the L-shape Poisson model problem, linear FEM, adaptive mesh refinement.

Quasi equilibrated discretization error over the domain.

4 L-shape domain, Papež, Liesen, S (2014)



The algebraic error $x_h - x_h^{(n)}$ (left) can dominate the total error $x - x_h^{(n)}$ (right) even while

$$\|\mathbf{x} - \mathbf{x}_n\|_{\mathbf{A}} \ll \|x - x_h\|_a = \|\nabla(x - x_h)\|\,.$$

Outline

1. Numerical solution of BVPs
2. Operator preconditioning
3. Algebraic preconditioning, discretization, and problem formulation
4. Various comments
5. Conclusions

Conclusions

- Krylov subspace methods adapt to the problem. Exploiting this adaptation is the key to their efficient use.

- 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 and stopping criteria are essential ...
  We are grateful for collaboration with Martin on these topics.

References

- J. Málek and Z.S., *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs.* SIAM Spotlight Series, SIAM (2015)
- T. Gergelits and Z.S., *Composite convergence bounds based on Chebyshev polynomials and finite precision conjugate gradient computations,* Numer. Alg. 65, 759-782 (2014)
- J. Papež, J. Liesen and Z.S., *Distribution of the discretization and algebraic error in numerical solution of partial differential equations,* Linear Alg. Appl. 449, 89-114 (2014)
- J. Liesen and Z.S., *Krylov Subspace Methods, Principles and Analysis.* Oxford University Press (2013)

Merci!