Challenges in Analysis of Algebraic Iterative Solvers

Jörg Liesen

Technical University of Berlin

and

Zdeněk Strakoš

Charles University in Prague and Czech Academy of Sciences

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

Workshop in honor of K. R. Rajagopal

Prague, March 2012

Cornelius Lanczos, March 9, 1947

"The reason why I am strongly drawn to such
approximation mathematics problems is not
the practical applicability of the solution,
but rather the fact that a very "economical"
solution is possible only when it is very "adequate".
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."

Algebraic iterative computations

- In iterative methods applied to linear algebraic problems, computational cost of finding sufficiently accurate approximation to the exact solution heavily depends on the particular data, i.e.,

  ⁕ on the underlying real world problem,
  ⁕ on the mathematical model,
  ⁕ on its discretisation.

- Any evaluation of cost in iterative computations must take into account effects of rounding errors.

- In mathematical modeling of real world phenomena, the accuracy of the computed approximation must be related to the underlying phenomena. Its evaluation can not be restricted to algebra.

Is there any algebraic error worth consideration?

Knupp and Salari, 2003:

“There may be incomplete iterative convergence (IICE) or round-off-error that is polluting the results. If the code uses an iterative solver, then one must be sure that the iterative stopping criteria is sufficiently tight so that the numerical and discrete solutions are close to one another. Usually in order-verification tests, one sets the iterative stopping criterion to just above the level of machine precision to circumvent this possibility.”

Why do we care? Is not all these algebraic stuff linear and simple?

Conjugate Gradients: A HPD, $x_0,\ r_0,\ p_0 = r_0$

$$\|x - x_n\|_A \;=\; \min_{u\,\in\, x_0 + \mathcal{K}_n(A, r_0)} \|x - u\|_A$$

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

$$\begin{aligned}
\gamma_{n-1} \;&=\; (r_{n-1}, r_{n-1})/(p_{n-1}, Ap_{n-1}) \\
x_n \;&=\; x_{n-1} + \gamma_{n-1}\, p_{n-1} \\
r_n \;&=\; r_{n-1} - \gamma_{n-1}\, Ap_{n-1} \\
\delta_n \;&=\; (r_n, r_n)/(r_{n-1}, r_{n-1}) \\
p_n \;&=\; r_n + \delta_n\, p_{n-1}.
\end{aligned}$$

Hestenes and Stiefel (1952), Lanczos (1950, 1952)

This algebraic stuff is nothing but linear!

CG is the Gauss-Christoffel Quadrature

$$Ax = b\,,\; x_0 \qquad \longleftrightarrow \qquad \omega(\lambda), \quad \int_\zeta^\xi f(\lambda)\, d\omega(\lambda)$$

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

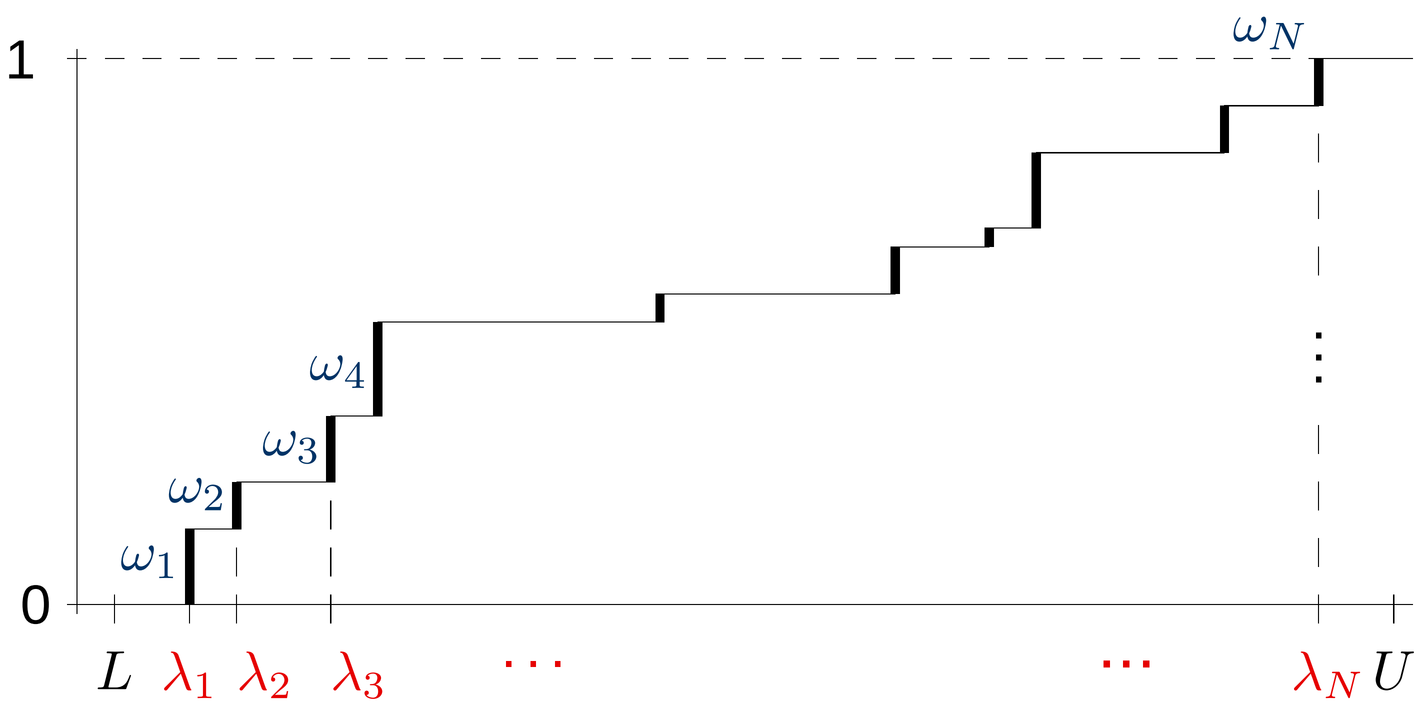
$$T_n\, y_n \;=\; \|r_0\|\, e_1 \qquad \longleftrightarrow \qquad \omega^{(n)}(\lambda), \quad \sum_{i=1}^{n} \omega_i^{(n)}\, f\left(\theta_i^{(n)}\right)$$

$$x_n \;=\; x_0 + W_n\, y_n$$

$$\boxed{\omega^{(n)}(\lambda) \longrightarrow \omega(\lambda)}$$

Distribution function $\omega(\lambda)$

λ_i, s_i are the eigenpairs of A , $\omega_i = |(s_i, w_1)|^2$, $w_1 = r_0/\|r_0\|$



Hestenes and Stiefel (1952), Lanczos (1952, almost unknown)

CG does model reduction matching $2n$ moments

$$\int_L^U \lambda^{-1}\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1} \;+\; R_n(f)$$

$$\frac{\|x - x_0\|_{\mathbf{A}}^2}{\|r_0\|^2} \;=\; n\text{-th Gauss quadrature} \;+\; \frac{\|x - x_n\|_{\mathbf{A}}^2}{\|r_0\|^2}$$

With $x_0 = 0,$ $$b^* A^{-1} b \;=\; \sum_{j=0}^{n-1} \gamma_j \|r_j\|^2 \;+\; r_n^* A^{-1} r_n \,.$$

Golub, Meurant, Reichel, Boley, Gutknecht, Saylor, Smolarski, ... ,
Meurant and S (2006), Golub and Meurant (2010), S and Tichý (2011),

Liesen, S, Krylov subspace methods, OUP (2012)

Outline

1. CG convergence bounds based on Chebyshev polynomials
2. Sensitivity of the Gauss-Christoffel quadrature
3. PDE discretizations and matrix computations

1 Linear bounds for the nonlinear method?

$$\begin{aligned}\|x - x_n\|_A &= \min_{\substack{p(0)=1\\ \deg(p)\le n}} \|A^{1/2} p(A)(x - x_0)\| \\ &= \min_{\substack{p(0)=1\\ \deg(p)\le n}} \|Y p(\Lambda) Y^* A^{1/2}(x - x_0)\| \\ &\le \left(\min_{\substack{p(0)=1\\ \deg(p)\le n}} \max_{1\le j\le N} |p(\lambda_j)| \right) \|x - x_0\|_A\end{aligned}$$

Using the shifted Chebyshev polynomials on the interval $[\lambda_1, \lambda_N]$,

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

1 Minimization property and the bound

This bound has a remarkably wiggling history:

- Markov (1890)
- Flanders and Shortley (1950)
- Lanczos (1953), Kincaid (1947), Young (1954, ...)
- Stiefel (1958), Rutishauser (1959)
- Meinardus (1963), Kaniel (1966)
- Daniel (1967a, 1967b)
- Luenberger (1969)

It is relevant to the Chebyshev method!

1 Composite bounds considering large outliers?

This bound should not be used in connection with the behaviour of CG unless $\kappa(A) = \lambda_N/\lambda_1$ is really small or unless the (very special) distribution of eigenvalues makes it relevant.

In particular, one should be very careful while using it as a part of a composite bound in the presence of the large outlying eigenvalues

$$\min_{\substack{p(0)=1 \\ \deg(p)\le n-s}} \max_{1\le j\le N} |\, q_s(\lambda_j)\, p(\lambda_j)\,| \quad \le \quad \max_{1\le j\le N} |q_s(\lambda_j)| \left| \frac{T_{n-s}(\lambda_j)|}{T_{n-s}(0)} \right|$$

$$< \quad \max_{1\le j\le N-s} \left| \frac{T_{n-s}(\lambda_j)}{T_{n-s}(0)} \right| .$$

This Chebyshev method bound on the interval $[\lambda_1, \lambda_{N-s}]$ is then valid after s initial steps.

1 Quote (2009, ...): the desired accuracy ϵ

Theorem. After

$$k \;=\; s \;+\; \left\lceil \frac{\ln(2/\epsilon)}{2} \sqrt{\frac{\lambda_{N-s}}{\lambda_1}} \right\rceil$$

iteration steps the CG will produce the approximate solution x_n satisfying

$$\|x - x_n\|_A \;\leq\; \epsilon \, \|x - x_0\|_A \, .$$

This recently republished and used statement is in finite precision arithmetic not true at all.

1 Axelsson (1976), Jennings (1977)

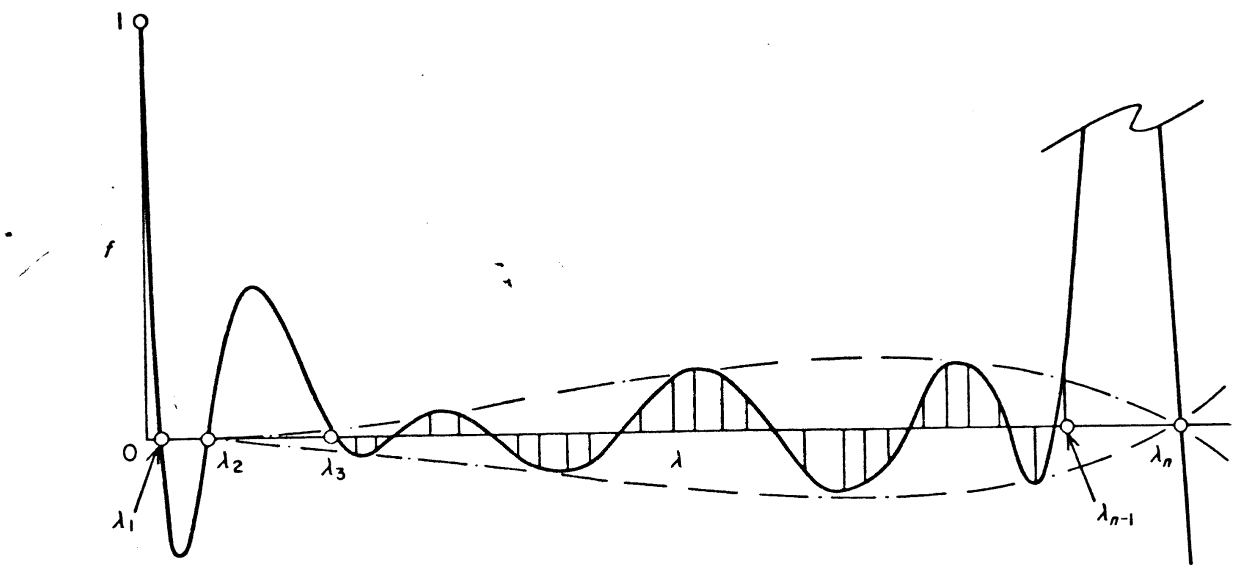
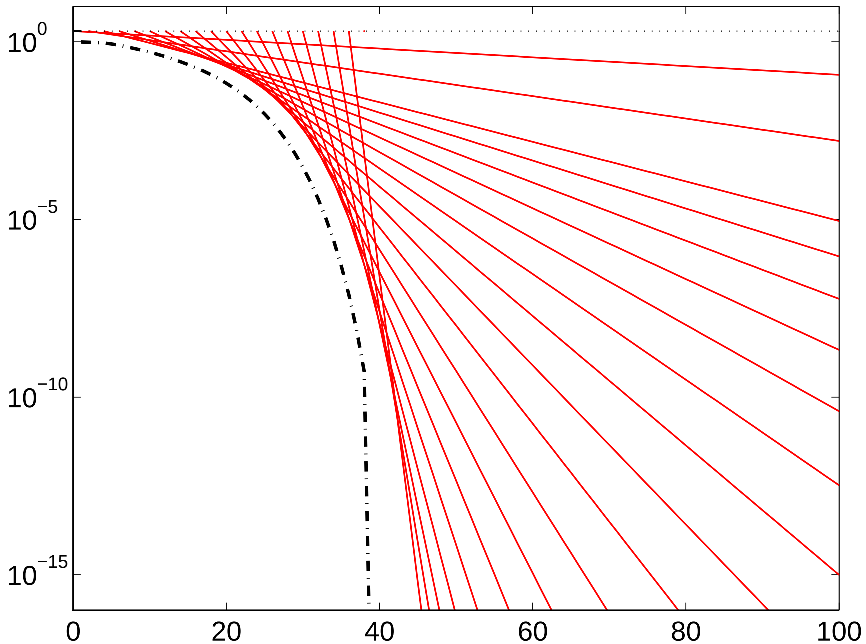
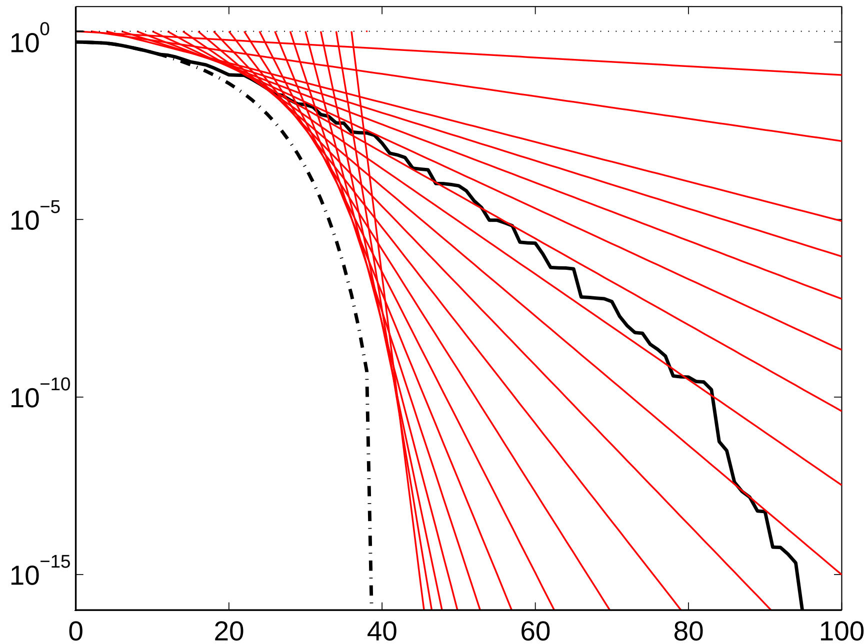


FIG. 4. A Chebyshev polynomial modified by a simple third order auxiliary polynomial having zeros at λ_1, λ_2 and λ_n.

p. 72: ... it may be inferred that rounding errors ... affects the convergence rate when large outlying eigenvalues are present.

1 The composite bounds completely fail





Composite bounds with varying number of outliers:
Exact CG (left) and FP CG (right),
Gergelits (2011).

2 CG and Gauss-Christoffel quadrature errors

$$\int_L^U \lambda^{-1}\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1} \;+\; R_n(f)$$

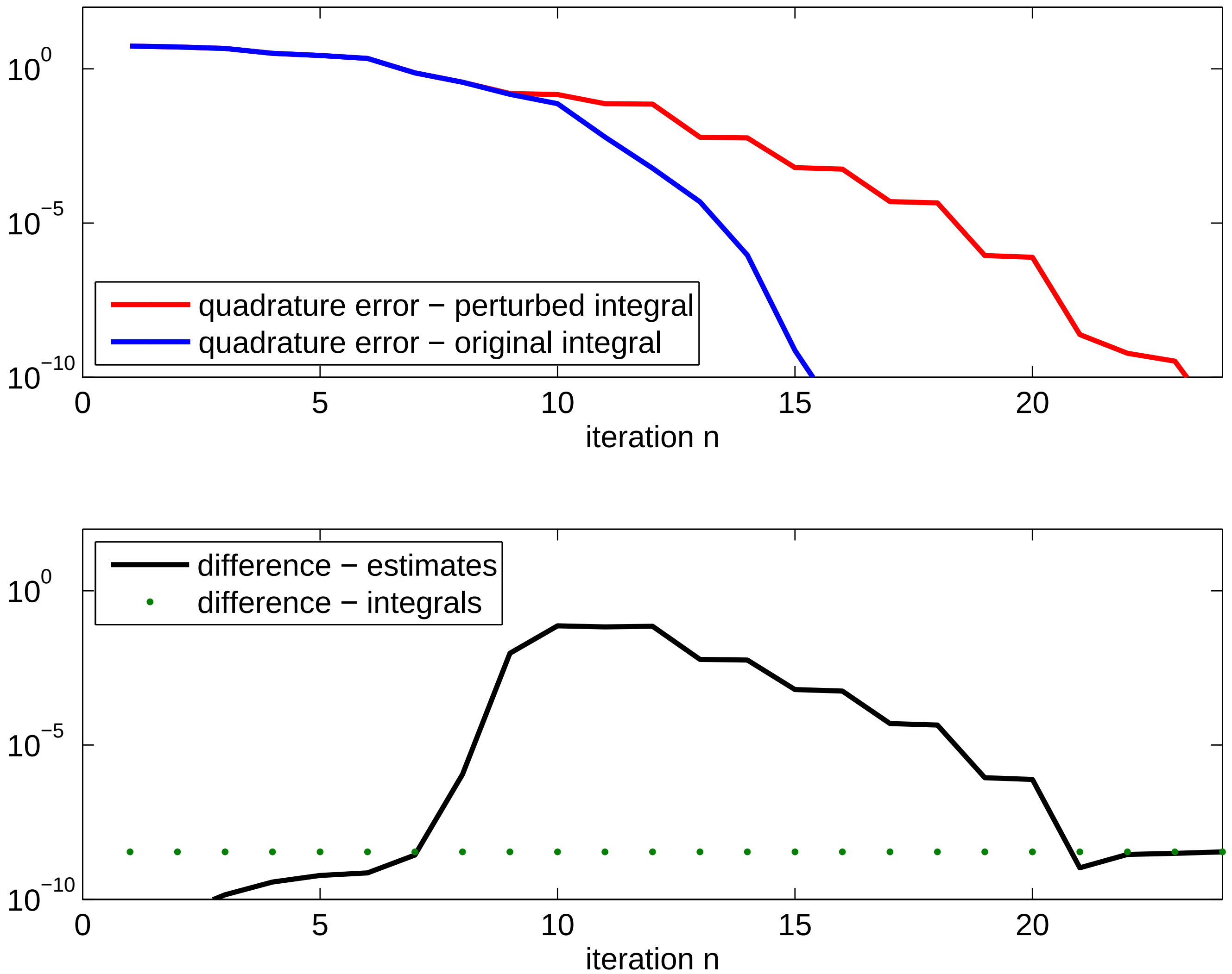
$$\frac{\|x - x_0\|_{\mathbf{A}}^2}{\|r_0\|^2} \;=\; n\text{-th Gauss quadrature} \;+\; \frac{\|x - x_n\|_{\mathbf{A}}^2}{\|r_0\|^2}$$

Consider two slightly different distribution functions with

$$I_\omega \;=\; \int_L^U \lambda^{-1}\, d\omega(\lambda) \;\approx\; I_\omega^n$$

$$I_{\tilde\omega} \;=\; \int_L^U \lambda^{-1}\, d\tilde\omega(\lambda) \;\approx\; I_{\tilde\omega}^n$$

2 Sensitivity of the Gauss-Christoffel Q.



2 The point goes back to 1814

1. Gauss-Christoffel quadrature for a small number of quadrature nodes can be highly sensitive to small changes in the distribution function that enlarge its support.

   In particular, the difference between the corresponding quadrature approximations (using the same number of quadrature nodes) can be many orders of magnitude larger than the difference between the integrals being approximated.

2. This sensitivity in Gauss-Christoffel quadrature can be observed for discontinuous, continuous, and even analytic distribution functions, and for analytic integrands uncorrelated with changes in the distribution functions, with no singularity close to the interval of integration.

2 Theorem - O'Leary, S, Tichý (2007)

Consider distribution functions $\omega(\lambda)$ and $\tilde{\omega}(\lambda)$ on $[L, U]$. Let $p_n(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$ and $\tilde{p}_n(\lambda) = (\lambda - \tilde{\lambda}_1) \dots (\lambda - \tilde{\lambda}_n)$ be the nth orthogonal polynomials corresponding to ω and $\tilde{\omega}$ respectively, with $\hat{p}_s(\lambda) = (\lambda - \xi_1) \dots (\lambda - \xi_s)$ their least common multiple.
If f'' is continuous on $[L, U]$, then the difference $\Delta^n_{\omega,\tilde{\omega}}$ between the approximation I^n_ω to I_ω and the approximation $I^n_{\tilde{\omega}}$ to $I_{\tilde{\omega}}$, obtained from the n-point Gauss-Christoffel quadrature, is bounded as

$$
\begin{aligned}
|\Delta^n_{\omega,\tilde{\omega}}| \quad &\leq \quad \left| \int_L^U \hat{p}_s(\lambda) f[\xi_1, \dots, \xi_s, \lambda] \, d\omega(\lambda) \; - \; \int_L^U \hat{p}_s(\lambda) f[\xi_1, \dots, \xi_s, \lambda] \, d\tilde{\omega}(\lambda) \right| \\
&+ \quad \left| \int_L^U f(\lambda) \, d\omega(\lambda) \; - \; \int_L^U f(\lambda) \, d\tilde{\omega}(\lambda) \right| .
\end{aligned}
$$

3 Take very simple model boundary value problem

$$-\Delta u \;=\; 16\eta_1\eta_2(1-\eta_1)(1-\eta_2)$$

on the unit square with zero Dirichlet boundary conditions. Galerkin finite element method (FEM) discretization with linear basis functions on the regular triangular grid with the mesh size $h = 1/(m+1)$, where m is the number of inner nodes in each direction. Discrete (piecewise linear) solution

$$u_h \;=\; \sum_{j=1}^{N} \zeta_j \, \phi_j(\eta_1, \eta_2)\,.$$

Computational error

$$\underbrace{u - u_h^{(n)}}_{\text{total error}} \;=\; \underbrace{u - u_h}_{\text{discretisation error}} \;+\; \underbrace{u_h - u_h^{(n)}}_{\text{algebraic error}}\,.$$

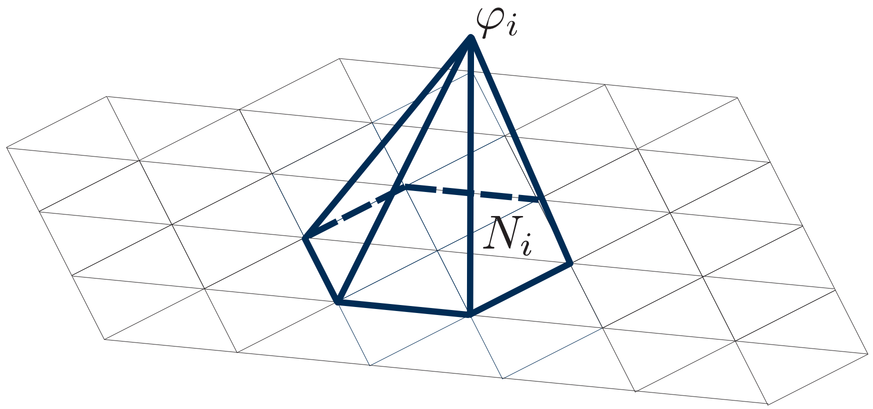
3 Local discretization and global computation

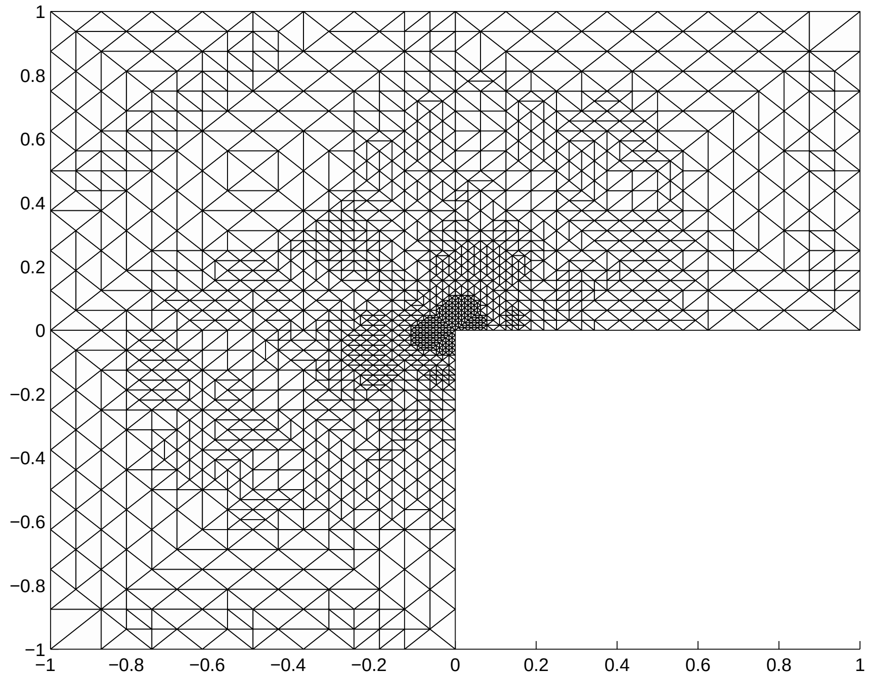
Discrete (piecewise linear) solution

$$u_h = \sum_{j=1}^{N} \zeta_j \, \phi_j(\eta_1, \eta_2) \, .$$

- If ζ_j is known exactly, then $u_h^{(n)} = u_h$, and the global information is approximated as the linear combination of the local basis functions.
- Apart from trivial cases, ζ_j , which supplies the global information, is not known exactly.

3 Local discretisation and global computation





3 Energy norm of the error

Theorem

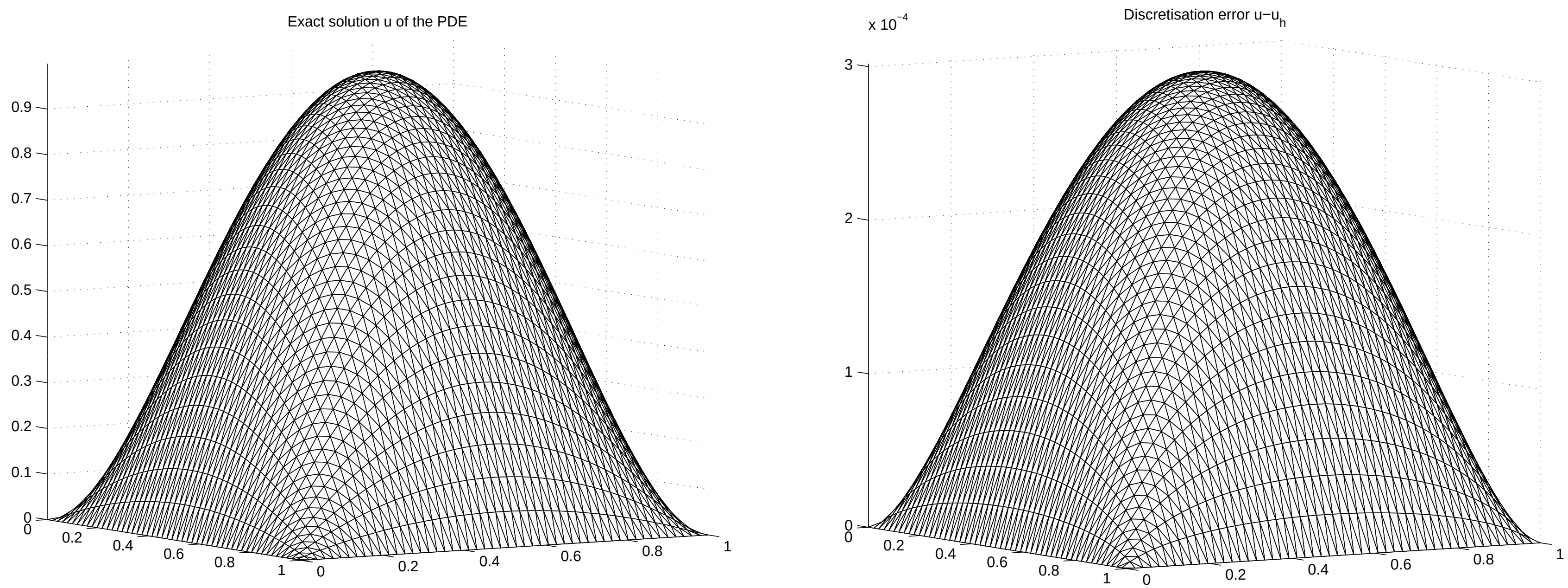
Up to a small inaccuracy proportional to machine precision,

$$\begin{aligned} \|\nabla(u - u_h^{(n)})\|^2 &= \|\nabla(u - u_h)\|^2 + \|\nabla(u_h - u_h^{(n)})\|^2 \\ &= \|\nabla(u - u_h)\|^2 + \|x - x_n\|_A^2 \, . \end{aligned}$$

Using zero Dirichlet boundary conditions,

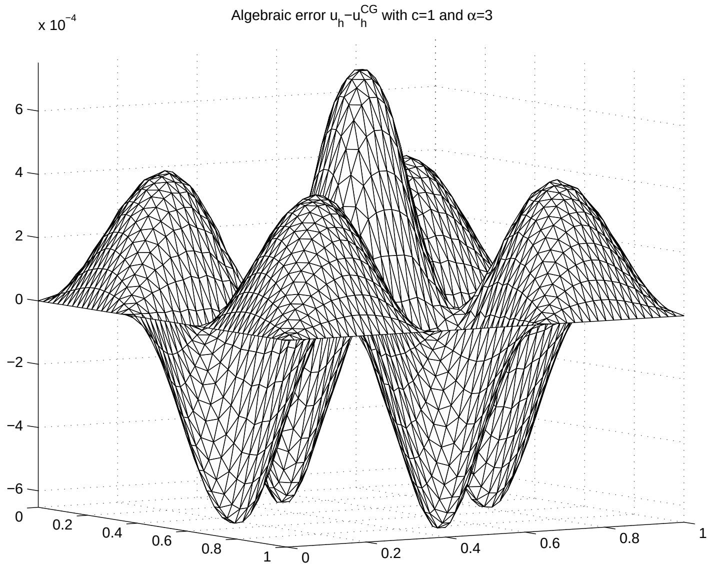
$$\|\nabla(u - u_h)\|^2 = \|\nabla u\|^2 - \|\nabla u_h\|^2 \, .$$

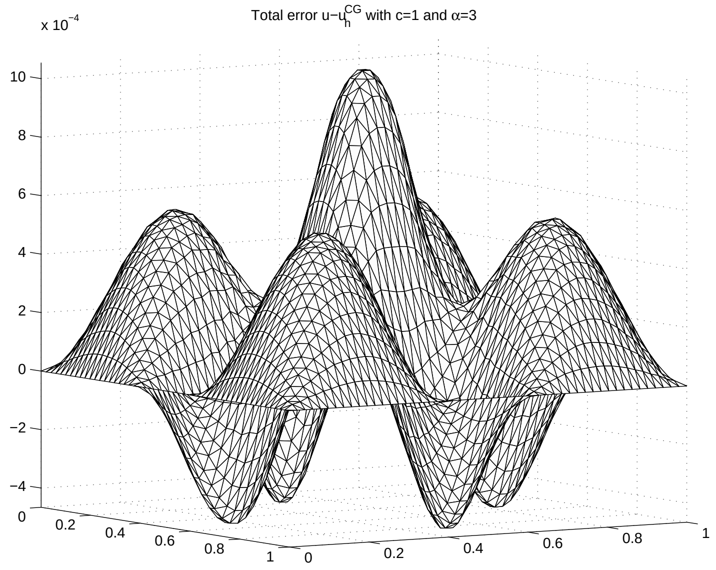
3 Solution and the discretization error



Exact solution u of the Poisson model problem (left)
and the MATLAB `trisurf` plot of the discretization error $u - u_h$ (right).

3 Algebraic and total errors

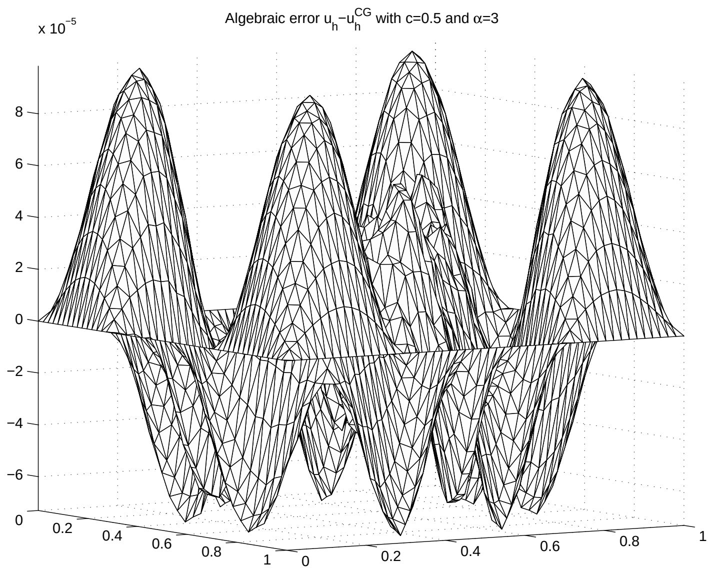


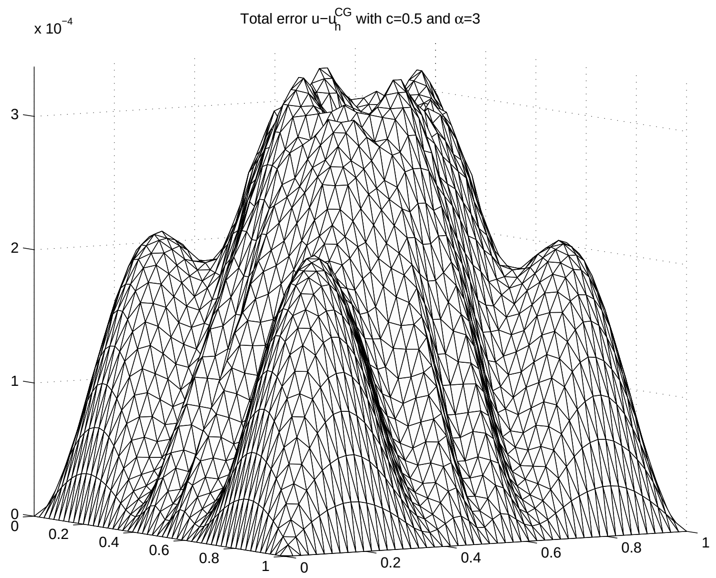


Algebraic error $u_h - u_h^{(n)}$ (left) and the MATLAB `trisurf` plot of the total error $u - u_h^{(n)}$ (right)

$$\begin{aligned} \|\nabla(u - u_h^{(n)})\|^2 &= \|\nabla(u - u_h)\|^2 + \|x - x_n\|_A^2 \\ &= 5.8444e-03 + 1.4503e-05\,. \end{aligned}$$

3 Algebraic and total errors

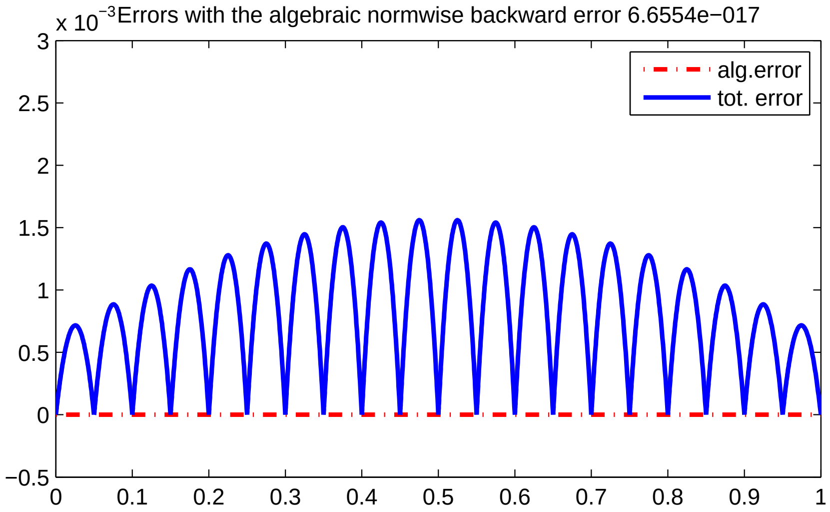


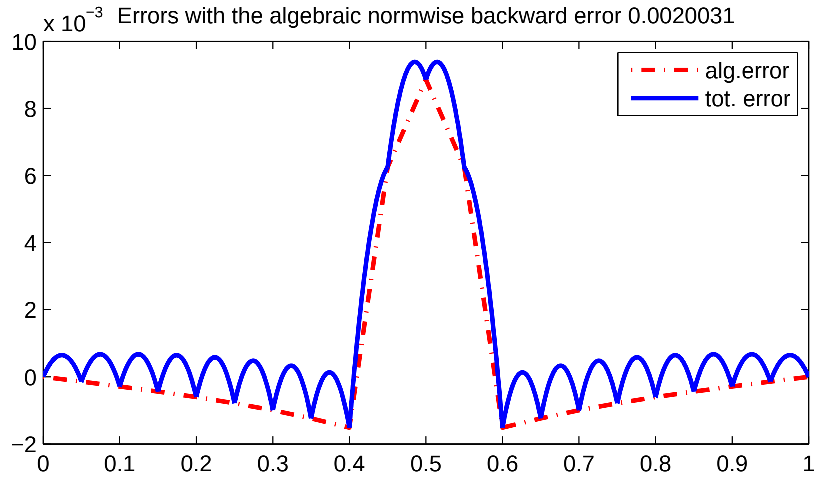


Algebraic error $u_h - u_h^{(n)}$ (left) and the MATLAB `trisurf` plot of the total error $u - u_h^{(n)}$ (right)

$$\begin{aligned} \|\nabla(u - u_h^{(n)})\|^2 &= \|\nabla(u - u_h)\|^2 + \|x - x_n\|_A^2 \\ &= 5.8444e-03 + 5.6043e-07\,. \end{aligned}$$

3 One can see 1D analogy





The discretization error (left),
the algebraic and the total error (right),
Papež (2011).

3 Why?

Krylov subspace methods represent matching moments model reduction!

3 Adaptivity?

We need *a-posteriori* error bounds which are:

- Locally efficient,
- fully computable (no hidden constants),
- and allow to compare the contribution of the discretization error and the algebraic error to the total error.

Conclusions

Patrick J. Roache's book *Validation and Verification in Computational Science*, 2006, p. 387:

"With the often noted tremendous increases in computer speed and memory, and with the less often acknowledged but equally powerful increases in algorithmic accuracy and efficiency, a natural question suggest itself. What are we doing with the new computer power? with the new GUI and other set-up advances? with the new algorithms? What *should* we do? ... *Get the right answer."*

Thank you for your work, help and friendship!