

The role of finite dimensional Hilbert spaces and their duals in the design of Krylov subspaces methods

Mario Arioli

`m.arioli@rl.ac.uk`

CCLRC-Rutherford Appleton Laboratory

- First part
 - Norms, Hyper-norms, and Dual Spaces
 - Interpolation spaces
 - Backward error (Rigal-Gaches Theory)
- Second part
 - Krylov methods and stopping criteria
 - Elliptic problems and Finite-Element method
 - Indefinite problems

- $(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \rightarrow \mathbf{R}$ **scalar product** and
 $\|u\|_{\mathcal{H}} = \sqrt{(u, u)} \quad \forall u \in \mathcal{H}$ **norm**.

■ $(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \rightarrow \mathbf{R}$ **scalar product** and

$$\|u\|_{\mathcal{H}} = \sqrt{(u, u)} \quad \forall u \in \mathcal{H} \text{ norm.}$$

■ $\exists \{\psi_i\}_{i=1, \dots, N}$ a basis for \mathcal{H}

$$\forall u \in \mathcal{H} \quad u = \sum_{i=1}^N u_i \psi_i \quad u_i \in \mathbf{R} \quad i = 1, \dots, N$$

■ $(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \rightarrow \mathbf{R}$ **scalar product** and

$$\|u\|_{\mathcal{H}} = \sqrt{(u, u)} \quad \forall u \in \mathcal{H} \text{ norm.}$$

■ $\exists \{\psi_i\}_{i=1, \dots, N}$ a basis for \mathcal{H}

$$\forall u \in \mathcal{H} \quad u = \sum_{i=1}^N u_i \psi_i \quad u_i \in \mathbf{R} \quad i = 1, \dots, N$$

■ **Representation of scalar product in \mathbf{R}^N .**

Let $u = \sum_{i=1}^N u_i \psi_i$ and $v = \sum_{i=1}^N v_i \psi_i$.

Then

$$(u, v) = \sum_{i=1}^N \sum_{j=1}^N u_i v_j (\psi_i, \psi_j) = \mathbf{v}^T \mathbf{H} \mathbf{u}$$

where $\mathbf{H}_{ij} = \mathbf{H}_{ji} = (\psi_i, \psi_j)$ and $\mathbf{u}, \mathbf{v} \in \mathbf{R}^N$.

Moreover, $\mathbf{u}^T \mathbf{H} \mathbf{u} > 0$ iff $\mathbf{u} \neq 0$ and, thus \mathbf{H} **SPD**.

- Other norms are possible on \mathbf{R}^N :

■ Other norms are possible on \mathbf{R}^N :

■ $\|\mathbf{u}\|_p = (\sum_{i=1}^N (u_i)^p)^{1/p}$ with $1 < p < \infty$

■ $\|\mathbf{u}\|_1 = (\sum_{i=1}^N |u_i|)$

■ $\|\mathbf{u}\|_\infty = \max_i |u_i|$

■ Other norms are possible on \mathbf{R}^N :

■ $\|\mathbf{u}\|_p = (\sum_{i=1}^N (u_i)^p)^{1/p}$ with $1 < p < \infty$

■ $\|\mathbf{u}\|_1 = (\sum_{i=1}^N |u_i|)$

■ $\|\mathbf{u}\|_\infty = \max_i |u_i|$

■ Hyper-norms on \mathbf{R}^N of order k .

$$\|\cdot\|_{\vec{k}} : \quad \mathbf{R}^N \rightarrow \mathbf{R}^k$$

I $\forall \lambda \in \mathbf{R} \quad \|\lambda \mathbf{u}\|_{\vec{k}} = |\lambda| \|\mathbf{u}\|_{\vec{k}}$

II $\forall \mathbf{u}, \mathbf{v} \in \mathbf{R}^N \quad \|\mathbf{u} + \mathbf{v}\|_{\vec{k}} \leq \|\mathbf{u}\|_{\vec{k}} + \|\mathbf{v}\|_{\vec{k}}$ component-wise

III $\|\mathbf{u}\|_{\vec{k}} = 0_k \Rightarrow \mathbf{u} = \mathbf{0}_N$

Dual space \mathcal{H}'

- $f \in \mathcal{H}' : \mathcal{H} \rightarrow \mathbf{R}$ (**functional**);
- $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v) \quad \forall u, v \in \mathcal{H}$
- \mathcal{H}' is the space of the linear functionals on \mathcal{H}

$$\|f\|_{\mathcal{H}'} = \sup_{u \neq 0} \frac{f(u)}{\|u\|_{\mathcal{H}}}$$

Dual space \mathcal{H}'

- $f \in \mathcal{H}' : \mathcal{H} \rightarrow \mathbf{R}$ (**functional**);
- $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v) \quad \forall u, v \in \mathcal{H}$
- \mathcal{H}' is the space of the linear functionals on \mathcal{H}

$$\|f\|_{\mathcal{H}'} = \sup_{u \neq 0} \frac{f(u)}{\|u\|_{\mathcal{H}}}$$

- If \mathcal{H} finite dimensional and $u = \sum_{i=1}^N u_i \psi_i$, then
 $f(u) = \sum_{i=1}^N u_i f(\psi_i) = \mathbf{f}^T \mathbf{u}$

- $f \in \mathcal{H}' : \mathcal{H} \rightarrow \mathbf{R}$ (**functional**);
- $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v) \quad \forall u, v \in \mathcal{H}$
- \mathcal{H}' is the space of the linear functionals on \mathcal{H}

$$\|f\|_{\mathcal{H}'} = \sup_{u \neq 0} \frac{f(u)}{\|u\|_{\mathcal{H}}}$$

- If \mathcal{H} finite dimensional and $u = \sum_{i=1}^N u_i \psi_i$, then
 $f(u) = \sum_{i=1}^N u_i f(\psi_i) = \mathbf{f}^T \mathbf{u}$

- **Dual vector**

Let $u \in \mathcal{H}$, $u \neq 0$, then $\exists f_u \in \mathcal{H}'$ such that

$$f_u(u) = \|u\|_{\mathcal{H}}$$

(Hahn-Banach).

Dual space \mathcal{H}'

- Let \mathcal{H} be a Hilbert finite dimensional space and \mathbf{H} the real $\mathbf{N} \times \mathbf{N}$ matrix identifying the scalar product.

Dual space \mathcal{H}'

- Let \mathcal{H} be a Hilbert finite dimensional space and \mathbf{H} the real $\mathbf{N} \times \mathbf{N}$ matrix identifying the scalar product.



$$f_u(u) = \mathbf{f}^T \mathbf{u} = (\mathbf{u}^T \mathbf{H} \mathbf{u})^{1/2}$$

The dual vector of u has the following representation:

Dual space \mathcal{H}'

- Let \mathcal{H} be a Hilbert finite dimensional space and \mathbf{H} the real $\mathbf{N} \times \mathbf{N}$ matrix identifying the scalar product.



$$f_u(u) = \mathbf{f}^T \mathbf{u} = (\mathbf{u}^T \mathbf{H} \mathbf{u})^{1/2}$$

The dual vector of u has the following representation:

$$\mathbf{f} = \frac{\mathbf{H} \mathbf{u}}{\|\mathbf{u}\|_{\mathbf{H}}}$$

and

$$\|f_u\|_{\mathcal{H}'}^2 = \mathbf{u}^T \mathbf{H} \mathbf{u} = \mathbf{f}^T \mathbf{H}^{-1} \mathbf{f}$$

Dual space basis

- The general definitions of a dual basis for \mathcal{H} is

$$\phi_j(\psi_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Dual space basis

- The general definitions of a dual basis for \mathcal{H} is

$$\phi_j(\psi_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- The ϕ_i are linearly independent:

$$\sum_{i=1}^N \beta_i \phi_i(\mathbf{u}) = 0 \forall \mathbf{u} \in \mathcal{H} \Rightarrow \sum_{i=1}^N \beta_i \phi_i(\psi_i) = 0 \Rightarrow \beta_i = 0.$$

Dual space basis

- The general definitions of a dual basis for \mathcal{H} is

$$\phi_j(\psi_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- The ϕ_i are linearly independent:

$$\sum_{i=1}^N \beta_i \phi_i(\mathbf{u}) = 0 \forall \mathbf{u} \in \mathcal{H} \Rightarrow \sum_{i=1}^N \beta_i \phi_i(\psi_i) = 0 \Rightarrow \beta_i = 0.$$

- $f(\psi_i) = \gamma_i$ and $f(u) = f(\sum_{i=1}^N u_i \psi_i) = \sum_{i=1}^N \gamma_i u_i$

$$\phi_i(\mathbf{u}) = \phi\left(\sum_{i=1}^N u_i \psi_i\right) = u_i \Rightarrow f = \sum_{i=1}^N \alpha_i \phi_i$$

Interpolation spaces

$$\begin{aligned}\mathcal{H} &= (\mathbf{R}^N, (\mathbf{u}, \mathbf{v})_{\mathcal{H}} = \mathbf{u}^T H \mathbf{v}) \\ \mathcal{M} &= (\mathbf{R}^N, (\mathbf{u}, \mathbf{v})_{\mathcal{M}} = \mathbf{u}^T M \mathbf{v})\end{aligned}$$

$$(\mathbf{u}, \mathbf{v})_{\mathcal{H}} = (\mathbf{u}, \mathbf{S}\mathbf{v})_{\mathcal{M}} = (\mathbf{S}\mathbf{u}, \mathbf{v})_{\mathcal{M}}$$

$$\mathbf{S} = M^{-1}H$$

S self-adjoint in the good scalar product!

$$\left\{ \mathbf{S}\mathbf{x} = \mu\mathbf{x} \quad \Leftrightarrow \quad H\mathbf{x} = \mu M\mathbf{x} \right\} \Rightarrow \mu = \delta^2 > 0$$

$$\exists W \text{ s.t. } M = W^{-T}W^{-1}, \quad H = W^{-T}\Delta^2W^{-1}, \quad \Delta \text{ diagonal}$$

Interpolation spaces

$$\Lambda = W\Delta W^{-1} \quad \Lambda^{1/2} = W\Delta^{1/2}W^{-1}$$

$$\begin{aligned} \mathbf{S} = M^{-1}H &= WW^T W^{-T} \Delta^2 W^{-1} \\ &= W\Delta W^{-1} W\Delta W^{-1} \\ &= \Lambda^2 \end{aligned}$$

$$M\Lambda = W^{-T}W^{-1}W\Delta W^T W^{-T}W^{-1} = \Lambda^T M \Rightarrow (\mathbf{u}, \Lambda\mathbf{v})_{\mathcal{M}} = (\Lambda\mathbf{u}, \mathbf{v})_{\mathcal{M}}$$

$$(\Lambda^{1/2}\mathbf{u}, \Lambda^{1/2}\mathbf{u})_{\mathcal{M}} = (\mathbf{u}, \Lambda\mathbf{u})_{\mathcal{M}}$$

$$\left[\mathcal{H}, \mathcal{M} \right]_{1/2} = \left\{ \mathbf{u} \in \mathbf{R}^N; \left((\mathbf{u}, \mathbf{u})_{\mathcal{M}} + (\mathbf{u}, \Lambda\mathbf{u})_{\mathcal{M}} \right)^{1/2} \right\}$$

Linear systems

$$A\mathbf{u} = \mathbf{b}$$

$$A \in \mathbf{R}^{N \times N} \text{ non singular}$$

Linear operator

- $A : \mathcal{H} \rightarrow \mathcal{V} \mathcal{H}$ and \mathcal{V} finite dimensional Hilbert spaces.

Linear operator

■ $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{V}$ and \mathcal{V} finite dimensional Hilbert spaces.

■

$$\|\mathbf{A}\|_{\mathcal{H},\mathcal{V}} = \sup_{\mathbf{u} \neq 0} \frac{\|\mathbf{A}\mathbf{u}\|_{\mathcal{V}}}{\|\mathbf{u}\|_{\mathcal{H}}} = \|\mathbf{V}^{1/2} \mathbf{A} \mathbf{H}^{-1/2}\|_2$$

Linear operator

■ $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{V}$ and \mathcal{H} and \mathcal{V} finite dimensional Hilbert spaces.

■

$$\|\mathbf{A}\|_{\mathcal{H},\mathcal{V}} = \sup_{\mathbf{u} \neq 0} \frac{\|\mathbf{A}\mathbf{u}\|_{\mathcal{V}}}{\|\mathbf{u}\|_{\mathcal{H}}} = \|\mathbf{V}^{1/2} \mathbf{A} \mathbf{H}^{-1/2}\|_2$$

■ The result follows from the generalized eigenvalue problem in \mathbf{R}^N

$$\mathbf{A}^T \mathbf{V} \mathbf{A} \mathbf{u} = \lambda \mathbf{H} \mathbf{u}$$

Linear systems: variational framework

■ Find $\mathbf{u} \in \mathcal{H}$ such that for all $\mathbf{v} \in \mathcal{H}$

$$\mathbf{a}(\mathbf{u}, \mathbf{v}) = L(\mathbf{v}) \quad (L(\cdot) \in \mathcal{H}' \text{ dual space of } \mathcal{H})$$

Linear systems: variational framework

- Find $\mathbf{u} \in \mathcal{H}$ such that for all $\mathbf{v} \in \mathcal{H}$

$$\mathbf{a}(\mathbf{u}, \mathbf{v}) = L(\mathbf{v}) \quad (L(\cdot) \in \mathcal{H}' \text{ dual space of } \mathcal{H})$$

- Existence and uniqueness: $\forall \mathbf{v}, \mathbf{w} \in \mathcal{H}$

$$\begin{aligned} \mathbf{a}(\mathbf{w}, \mathbf{v}) &\leq C_1 \|\mathbf{w}\|_{\mathcal{H}} \|\mathbf{v}\|_{\mathcal{H}} \\ \sup_{\mathbf{w} \in \mathcal{H} \setminus \{0\}} \frac{\mathbf{a}(\mathbf{w}, \mathbf{v})}{\|\mathbf{w}\|_{\mathcal{H}}} &\geq C_2 \|\mathbf{v}\|_{\mathcal{H}} \\ \sup_{\mathbf{v} \in \mathcal{H} \setminus \{0\}} \frac{\mathbf{a}(\mathbf{w}, \mathbf{v})}{\|\mathbf{v}\|_{\mathcal{H}}} &\geq C_2 \|\mathbf{w}\|_{\mathcal{H}} \end{aligned}$$

Backward error

We have the following equivalence in a general Hilbert (true also for a Banach):

$$\left. \begin{array}{l} \exists \mathbf{b} \in \mathcal{BL}(\mathcal{H}), \exists \delta L \in \mathcal{H}' \text{ such that:} \\ \mathbf{a}(\tilde{\mathbf{u}}, \mathbf{v}) + \mathbf{b}(\tilde{\mathbf{u}}, \mathbf{v}) = (L + \delta L)(\mathbf{v}), \\ \forall \mathbf{v} \in \mathcal{H}, \text{ and} \\ \|\mathbf{b}(\cdot, \cdot)\|_{\mathcal{BL}(\mathcal{H})} \leq \alpha, \|\delta L\|_{\mathcal{H}'} \leq \beta \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \|\rho_{\tilde{\mathbf{u}}}\|_{\mathcal{H}'} \leq \alpha \|\tilde{\mathbf{u}}\|_{\mathcal{H}} + \beta \\ \text{where } \rho_{\tilde{\mathbf{u}}} \in \mathcal{H}' \text{ is defined by} \\ \langle \rho_{\tilde{\mathbf{u}}}, \mathbf{v} \rangle_{\mathcal{H}', \mathcal{H}} = \mathbf{a}(\tilde{\mathbf{u}}, \mathbf{v}) - L(\mathbf{v}), \\ \forall \mathbf{v} \in \mathcal{H} \end{array} \right.$$

A., Noulard, and Russo (2001)

Part II

Linear systems

$$A\mathbf{u} = \mathbf{b}$$

$A \in \mathbf{R}^{N \times N}$ positive definite

A symmetric positive definite

$$\mathcal{H} = (\mathbf{R}^N, \|\cdot\|_{\mathbf{A}}) \text{ and } \mathcal{H}' = (\mathbf{R}^N, \|\cdot\|_{\mathbf{A}^{-1}})$$

At each step k the conjugate gradient method minimizes the energy norm of the error $\delta\mathbf{u}^{(k)} = \mathbf{u} - \mathbf{u}^{(k)}$ on a Krylov space $\mathbf{u}^{(0)} + \mathcal{K}_k$:

$$\min_{\mathbf{u}^{(k)} \in \mathbf{u}^{(0)} + \mathcal{K}_k} \|\delta\mathbf{u}^{(k)}\|_{\mathbf{A}}^2$$

$$\|\delta\mathbf{u}^{(k)}\|_{\mathbf{A}} = \|\rho_{\mathbf{u}^{(k)}}\|_{\mathcal{H}'} = \|\mathbf{r}^{(k)}\|_{\mathbf{A}^{-1}}$$

$$\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{u}^{(k)}$$

The symmetric case: stopping criteria

■ Classic Criterion:

$$\text{IF } \|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_2 \leq \sqrt{\varepsilon}\|\mathbf{b}\|_2 \text{ THEN STOP ,}$$

The symmetric case: stopping criteria

■ Classic Criterion:

$$\text{IF } \|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_2 \leq \sqrt{\varepsilon}\|\mathbf{b}\|_2 \text{ THEN STOP ,}$$

■ New Criterion:

$$\text{IF } \|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}} \leq \eta\|\mathbf{b}\|_{\mathbf{A}^{-1}} \text{ THEN STOP ,}$$

with $\eta < 1$ an a-priori threshold fixed by the user.

■ Classic Criterion:

$$\text{IF } \|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_2 \leq \sqrt{\varepsilon}\|\mathbf{b}\|_2 \text{ THEN STOP ,}$$

■ New Criterion:

$$\text{IF } \|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}} \leq \eta\|\mathbf{b}\|_{\mathbf{A}^{-1}} \text{ THEN STOP ,}$$

with $\eta < 1$ an a-priori threshold fixed by the user.

The choice of η will depend on the properties of the problem that we want to solve, and, in the practical cases, η can be frequently much larger than ε , the roundoff unit of the computer finite precision arithmetic.

The symmetric case: stopping criteria cont.

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}} \quad ?$$

The symmetric case: stopping criteria cont.

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}} \quad ?$$

$$\|\mathbf{b}\|_{\mathbf{A}^{-1}} \quad ?$$

The symmetric case: stopping criteria cont.

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$$

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$$

- Hestenes-Stiefel rule (1952) (see Strakoš and Tichý, 2002) **numerically stable**

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$$

- Hestenes-Stiefel rule (1952) (see Strakoš and Tichý, 2002) **numerically stable**
- Gauss quadrature rules (Golub and Meurant, 1997)

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$$

- Hestenes-Stiefel rule (1952) (see Strakoš and Tichý, 2002) **numerically stable**
- Gauss quadrature rules (Golub and Meurant, 1997)
 - Gauss equivalent to Hestenes-Stiefel rule (Strakoš and Tichý). The Gauss quadrature does not require any a-priori knowledge of the smallest and the biggest eigenvalues and computes a lower bound of $\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$.

$$\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$$

- Hestenes-Stiefel rule (1952) (see Strakoš and Tichý, 2002) **numerically stable**
- Gauss quadrature rules (Golub and Meurant, 1997)
 - Gauss equivalent to Hestenes-Stiefel rule (Strakoš and Tichý). The Gauss quadrature does not require any a-priori knowledge of the smallest and the biggest eigenvalues and computes a lower bound of $\|\mathbf{A}\mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}$.
 - Gauss-Lobatto and Gauss-Radau. They compute lower and upper bounds using the extremes eigenvalues of \mathbf{A} .

The symmetric case: Hestenes-Stiefel rule

During the conjugate gradient iterates, we compute the scalar α_k and the conjugate vectors $\mathbf{p}^{(k)}$ ($\mathbf{p}^{(j)T} \mathbf{A} \mathbf{p}^{(i)} = 0, j \neq i$) and the residuals $\mathbf{r}^{(k)}$.

Thus,

$$\mathbf{u} = \sum_{j=1}^N \alpha_j \mathbf{p}^{(j)}$$

and

$$\|\delta \mathbf{u}^{(k)}\|_{\mathbf{A}}^2 = \|\mathbf{A} \mathbf{u}^{(k)} - \mathbf{b}\|_{\mathbf{A}^{-1}}^2 = e_{\mathbf{A}}^2 = \sum_{j=k+1}^N \alpha_j \mathbf{r}^{(j)T} \mathbf{r}^{(j)}$$

The symmetric case: Hestenes-Stiefel rule

Under the assumption that $e_{\mathbf{A}}^{(k+d)} \ll e_{\mathbf{A}}^{(k)}$, where the integer d denotes a suitable delay, the Hestenes and Stiefel estimate ξ_k will be

$$\xi_k = \sum_{j=k+1}^{k+d} \alpha_j \mathbf{r}^{(j)T} \mathbf{r}^{(j)} = \sum_{j=k+1}^{k+d} \chi_j.$$

The symmetric case: Hestenes-Stiefel rule

Under the assumption that $e_{\mathbf{A}}^{(k+d)} \ll e_{\mathbf{A}}^{(k)}$, where the integer d denotes a suitable delay, the Hestenes and Stiefel estimate ξ_k will be

$$\xi_k = \sum_{j=k+1}^{k+d} \alpha_j \mathbf{r}^{(j)T} \mathbf{r}^{(j)} = \sum_{j=k+1}^{k+d} \chi_j.$$

The choice of a value for d depends on preconditioner and ill-conditioning.

$$\mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}$$

If $\|\mathbf{u} - \mathbf{u}^{(0)}\|_{\mathbf{A}}^2 \leq \|\mathbf{u}\|_{\mathbf{A}}^2$

$$\mathbf{b}^T \mathbf{A}^{-1} \mathbf{b} = \mathbf{u}^T \mathbf{A} \mathbf{u} \geq \sum_{j=0}^k \alpha_j r^{(j)T} r^{(j)} + \mathbf{b}^T \mathbf{u}^{(0)} + \mathbf{r}^{(0)T} \mathbf{u}^{(0)} = \rho_k.$$

is a stable lower bound (see Strakoš and Tichý BIT(2006))

Therefore, we use the following stopping criterion

IF $\xi_k \leq \eta^2 \rho_k$ THEN STOP .

The dual norm of the preconditioned residual is equal to the dual norm of the original residual.

A. Numer. Math. 2004, Meurant Numer Alg. 2004

Continuous problem

$$a(u, v) = \int_{\Omega} \mathcal{K}(x) \nabla u \cdot \nabla v d\mathbf{x}, \quad \forall u, v \in H_0^1(\Omega)$$

$\forall u, v \in H_0^1(\Omega)$, $\exists \gamma \in \mathbf{R}_+$ and $\exists M \in \mathbf{R}_+$ such that

$$\begin{aligned} \gamma \|u\|_{1,\Omega}^2 &\leq a(u, u) \\ a(u, v) &\leq M \|u\|_{1,\Omega} \|v\|_{1,\Omega}, \end{aligned}$$

$$L(v) = \int_{\Omega} f v d\mathbf{x}, \quad L(v) \in H^{-1}(\Omega).$$

(P) $\left\{ \begin{array}{l} \text{Find } u \in H_0^1(\Omega) \text{ such that} \\ a(u, v) = L(v), \quad \forall v \in H_0^1(\Omega), \end{array} \right.$ has a unique solution.

■ Weak formulation

$$\begin{cases} \text{Find } u_h \in \mathcal{H}_h \text{ such that} \\ a_h(u_h, v_h) = L_h(v_h), \quad \forall v_h \in \mathcal{H}_h, \end{cases}$$

Finite element methods choose \mathcal{H}_h to be a space of functions v_h defined on a subdivision Ω_h of Ω into simplices T of diameter h_T ; h denotes a piecewise constant function defined on Ω_h via $h|_T = h_T$.

■ Weak formulation

$$\left\{ \begin{array}{l} \text{Find } u_h \in \mathcal{H}_h \text{ such that} \\ a_h(u_h, v_h) = L_h(v_h), \quad \forall v_h \in \mathcal{H}_h, \end{array} \right.$$

Finite element methods choose \mathcal{H}_h to be a space of functions v_h defined on a subdivision Ω_h of Ω into simplices T of diameter h_T ; h denotes a piecewise constant function defined on Ω_h via $h|_T = h_T$.

■ Existence and uniqueness: $\mathcal{H}_h \subset \hat{\mathcal{H}} = H_0^1(\Omega)$.

■ Error Estimate: $\|u - u_h\|_{\mathcal{H}} \leq C(h)$

Finite-element framework

Solve

$$\mathbf{A} \mathbf{u}_h = \mathbf{b}$$

given

$$\sup_{\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \sup_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^t \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_{\mathbf{H}} \|\mathbf{w}\|_{\mathbf{H}}} \leq C_1 \quad (\text{sup-sup})$$

$$\inf_{\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \sup_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^t \mathbf{A} \mathbf{v}}{\|\mathbf{v}\|_{\mathbf{H}} \|\mathbf{w}\|_{\mathbf{H}}} \geq C_2 \quad (\text{inf-sup})$$

Note: $\|v_h\|_{\mathcal{H}_h} = \|\mathbf{v}\|_{\mathbf{H}}$.

Finally, assuming $h < 1$ and $t > 0$, and choosing $\eta = \mathcal{O}(h)$, we have

$$\|u - u_h^{(k)}\|_{\hat{\mathcal{H}}} \leq C^*(h^t) \|u\|_{\hat{\mathcal{H}}} + 2\|u - u_h\|_{\hat{\mathcal{H}}} \leq C(h).$$

where

- $u(\mathbf{x})$ is the exact solution of the variational problem,
- $u_h(\mathbf{x})$ is the exact solution of the approximate problem,
- $u_h^{(k)}(\mathbf{x}) = \sum_{i=1}^N \mathbf{u}_h^{(k)} \phi_i(\mathbf{x})$ is the approximate solution at step k .
($\phi_i(\mathbf{x})$ are the basis functions)

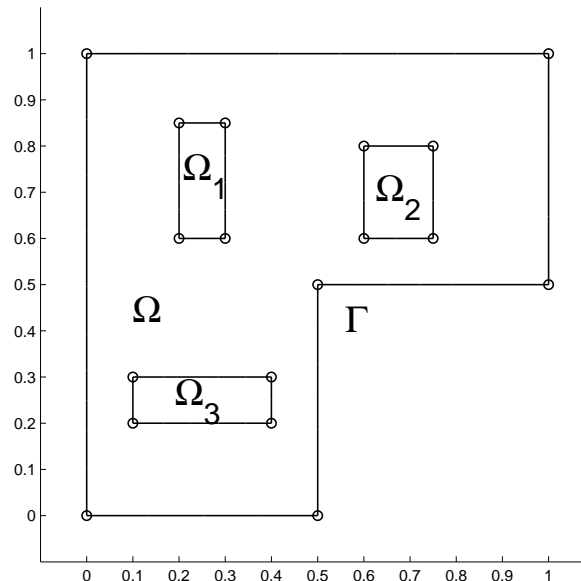
A. Numer. Math. 2004

Test problem

Problem

$$\kappa(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \Omega \setminus \{\Omega_1 \cup \Omega_2 \cup \Omega_3\}, \\ 10^6 & \mathbf{x} \in \Omega_1, \\ 10^4 & \mathbf{x} \in \Omega_2, \\ 10^2 & \mathbf{x} \in \Omega_3. \end{cases}$$

$$L(v) = \int_{\Omega} 10v d\mathbf{x}, \quad \forall v \in H_0^1(\Omega)$$



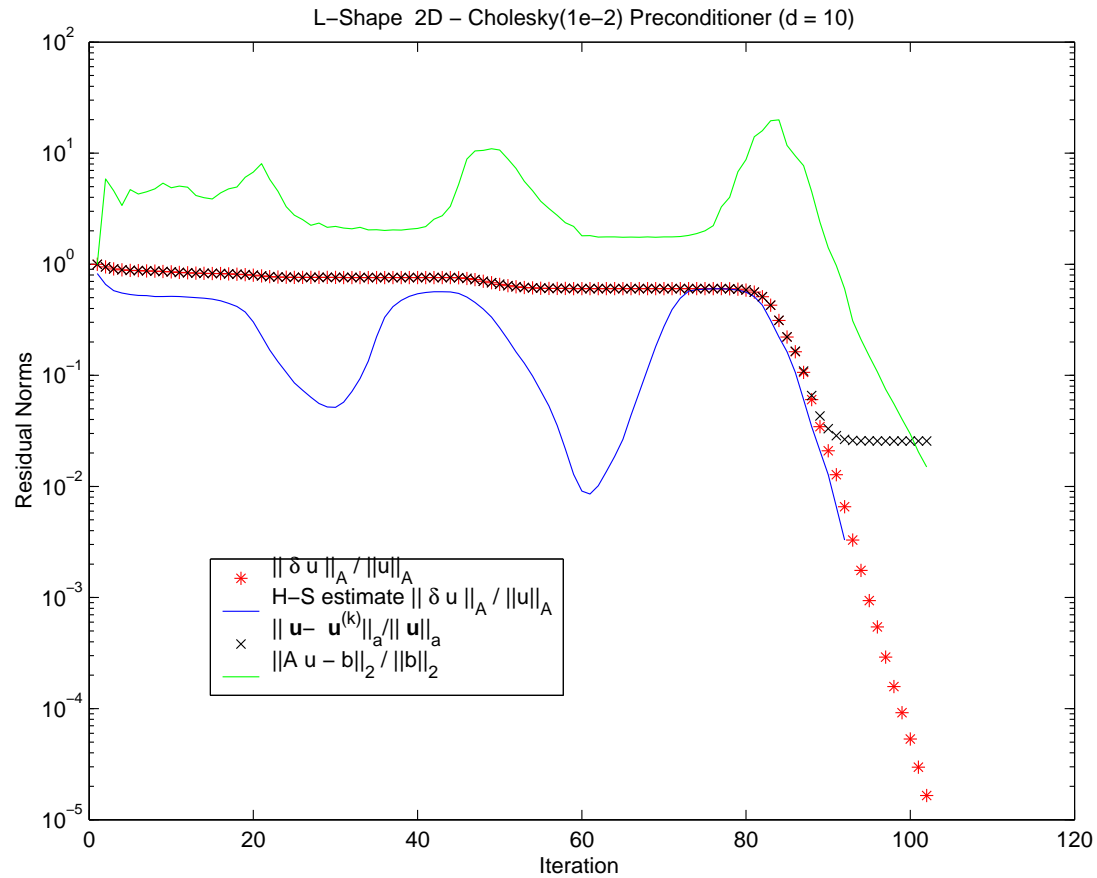
Preconditioners: estimates for $\kappa(\mathbf{M}^{-1}\mathbf{A})$

M	$\kappa(\mathbf{M}^{-1}\mathbf{A})$
I	$1.8 \cdot 10^{10}$
Jacobi	$1.5 \cdot 10^9$
Inc. Cholesky(0)	$4.3 \cdot 10^8$

$$\eta^2 \approx 10^{-5} \text{ and } \mathbf{N} = 29619.$$

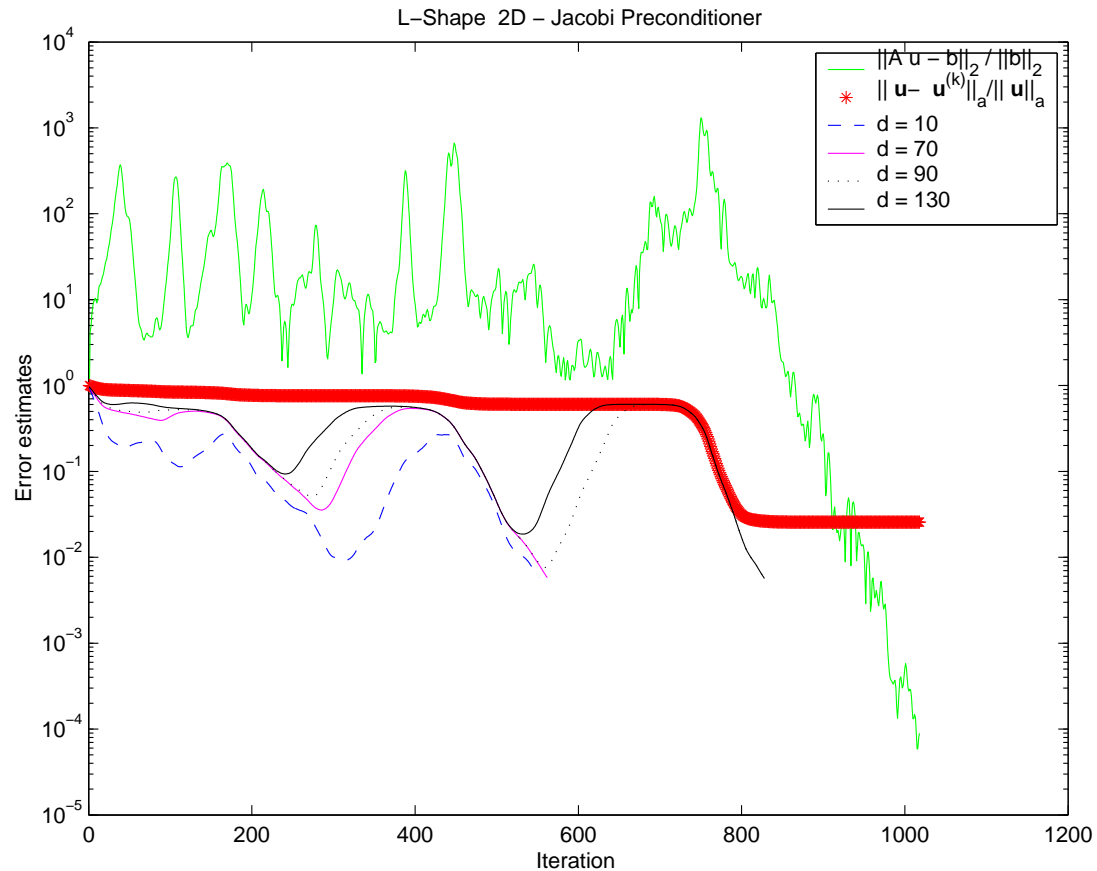
The condition numbers of the preconditioned matrices $\mathbf{M}^{-1}\mathbf{A}$ for the second problem are still very high, and only the incomplete Cholesky preconditioner with drop tolerance 10^{-2} is an effective choice.

Example: Problem 2



Behaviour of the norms of the residual for the incomplete Cholesky preconditioner with drop tolerance 10^{-2} and $d = 10$.

Example: Problem 2



Comparison of several estimates of the energy error for $d = 10, 70, 90, 130$ in Problem 2.

The positive definite problem

- $a(u, v) \neq a(v, u)$
- \mathbf{A} asymmetric but **positive definite**
- $\mathbf{H} = \frac{1}{2}(\mathbf{A}^T + \mathbf{A})$ **SPD**

See A., Login, and Wathen Numer. Math. 2005

Linear systems: variational framework

■ Find $u \in \hat{\mathcal{H}}$ such that for all $v \in \hat{\mathcal{H}}$

$$a(u, v) = L(v) \quad (L(\cdot) \in \hat{\mathcal{H}}' \text{ dual space of } \hat{\mathcal{H}})$$

Linear systems: variational framework

- Find $u \in \hat{\mathcal{H}}$ such that for all $v \in \hat{\mathcal{H}}$

$$a(u, v) = L(v) \quad (L(\cdot) \in \hat{\mathcal{H}}' \text{ dual space of } \hat{\mathcal{H}})$$

- Existence and uniqueness: $\forall v, w \in \hat{\mathcal{H}}$

$$\begin{aligned} a(w, v) &\leq C_1 \|w\|_{\hat{\mathcal{H}}} \|v\|_{\hat{\mathcal{H}}} \\ \sup_{w \in \hat{\mathcal{H}} \setminus \{0\}} \frac{a(w, v)}{\|w\|_{\hat{\mathcal{H}}}} &\geq C_2 \|v\|_{\hat{\mathcal{H}}} \\ \sup_{v \in \hat{\mathcal{H}} \setminus \{0\}} \frac{a(w, v)}{\|v\|_{\hat{\mathcal{H}}}} &\geq C_2 \|w\|_{\hat{\mathcal{H}}} \end{aligned}$$

Linear systems: variational framework

- Find $u \in \hat{\mathcal{H}}$ such that for all $v \in \hat{\mathcal{H}}$

$$a(u, v) = L(v) \quad (L(\cdot) \in \hat{\mathcal{H}}' \text{ dual space of } \hat{\mathcal{H}})$$

- Existence and uniqueness: $\forall v, w \in \hat{\mathcal{H}}$

$$a(w, v) \leq C_1 \|w\|_{\hat{\mathcal{H}}} \|v\|_{\hat{\mathcal{H}}}$$

$$\sup_{w \in \hat{\mathcal{H}} \setminus \{0\}} \frac{a(w, v)}{\|w\|_{\hat{\mathcal{H}}}} \geq C_2 \|v\|_{\hat{\mathcal{H}}}$$

$$\sup_{v \in \hat{\mathcal{H}} \setminus \{0\}} \frac{a(w, v)}{\|v\|_{\hat{\mathcal{H}}}} \geq C_2 \|w\|_{\hat{\mathcal{H}}}$$

- Finite-element approximation with $\mathcal{H}_h \subset \hat{\mathcal{H}}$

Backward error

Theorem 0.0 *Let \mathbf{u} be the solution of the weak formulation and let $\mathbf{u}, u_h = \Pi_h \mathbf{u}$ satisfy*

$$A\mathbf{u} = \mathbf{b}; \quad \frac{\|\mathbf{u} - u_h\|_{\hat{\mathcal{H}}}}{\|u_h\|_{\hat{\mathcal{H}}}} \leq C(h).$$

Then $\tilde{u}_h = \Pi_h \tilde{\mathbf{u}}$ satisfies

$$\frac{\|\mathbf{u} - \tilde{u}_h\|_{\hat{\mathcal{H}}}}{\|\tilde{u}_h\|_{\hat{\mathcal{H}}}} \leq \tilde{C}(h) = O(C(h))$$

if

$$\frac{\|\mathbf{b} - A\tilde{\mathbf{u}}\|_{H^{-1}}}{\|\tilde{\mathbf{u}}\|_H} \leq \eta C(h) C_2,$$

for some $\eta \in (0, 1)$.

New stopping criterion

$$\|\mathbf{r}^{(k)}\|_{\mathbf{H}^{-1}} \leq C_2 C(h) \|\mathbf{u}^{(k)}\|_{\mathbf{H}}.$$

How to calculate $\|\mathbf{r}^k\|_{\mathbf{H}^{-1}}$?

- Solve preconditioned system

$$\mathbf{H}^{-1/2} \mathbf{A} \mathbf{H}^{-1/2} \hat{\mathbf{u}} = \mathbf{H}^{-1/2} \mathbf{f}$$

- ▶ $\|\hat{\mathbf{r}}^k\|_{l_2} = \|\mathbf{r}^k\|_{H^{-1}}$
- ▶ 3-term recurrence.

- Approximate it from Krylov subspace information.

- Concus & Golub, Widlund: 3-term recurrences for nonsymmetric problems

- ▶ work in \mathbf{H} -inner product
- ▶ do not minimize the residual norm.

The use of \mathbf{H} as preconditioner makes the convergence of the Krylov methods independent of h .

A., Loghin, and Wathen Numer. Math. 2005, Georgoulis Loghin 2006 for discontinuous Galerkin

Elliptic problems in \mathbf{R}^2 (Ω unit square)

$$\begin{aligned} -\nabla \cdot (\mathbf{a}(\mathbf{x}) \nabla u) + \mathbf{b}(\mathbf{x}) \cdot \nabla u + c(\mathbf{x})u &= f && \text{in } \Omega \\ u &= 0 && \text{on } \Gamma. \end{aligned}$$

where

$$(\mathbf{a})_{ij}, (\mathbf{b})_i, c \in \mathbf{L}^\infty(\Omega), \quad i, j = 1, 2,$$

$$k_2(\mathbf{x}) |\boldsymbol{\xi}|^2 \leq \boldsymbol{\xi}^T \mathbf{a}(\mathbf{x}) \boldsymbol{\xi} \leq k_1(\mathbf{x}) |\boldsymbol{\xi}|^2,$$

$$c(\mathbf{x}) - \frac{1}{2} \nabla \cdot \mathbf{b}(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \Omega.$$

Examples

$$\mathcal{A}(w, v) = (\mathbf{a} \cdot \nabla w, \nabla v) + (\mathbf{b} \cdot \nabla w, v) + (cw, v),$$

is continuous and coercive with

$$C_1 = \|k_1\|_{L^\infty(\Omega)} + \|\mathbf{b}\|_{L^\infty(\Omega)} + C(\Omega)\|c\|_{L^\infty(\Omega)},$$

$$C_2 = \min_{\mathbf{x} \in \Omega} k_2(\mathbf{x}),$$

$$\text{wrt } \|\cdot\|_{\mathcal{H}} = |\cdot|_{H_0^1(\Omega)} := |\cdot|_1.$$

■ Discretization:

linear elements on uniform & adaptive meshes

Error estimate:

$$|u - u_h|_1 \leq Ch^{s-1} \|u\|_s, \quad 1 \leq s \leq 2.$$

Numerical experiments

■ Discretization:

linear elements on uniform & adaptive meshes

Error estimate:

$$|u - u_h|_1 \leq Ch^{s-1} \|u\|_s, \quad 1 \leq s \leq 2.$$

■ Estimation of parameters

$$h \sim \frac{\|\mathbf{u}^k\|_{\mathbf{M}}}{\|\mathbf{u}^k\|_{l_2}}, \quad \|u\|_s \sim \|\mathbf{A}\mathbf{u}^k\|_{l_2}$$

Stopping criteria and estimates

- Residual dual norm: $\|\mathbf{r}^k\|_{\mathbf{H}^{-1}}$

Stopping criteria and estimates

- Residual dual norm: $\|\mathbf{r}^k\|_{\mathbf{H}^{-1}}$
- Energy estimate $\|\mathbf{u}^k - \mathbf{u}^{k-1}\|_{\mathbf{H}} \leq C_2 h^2 \|\mathbf{A}\mathbf{u}^k\|_{l_2}$

Advection-diffusion problem

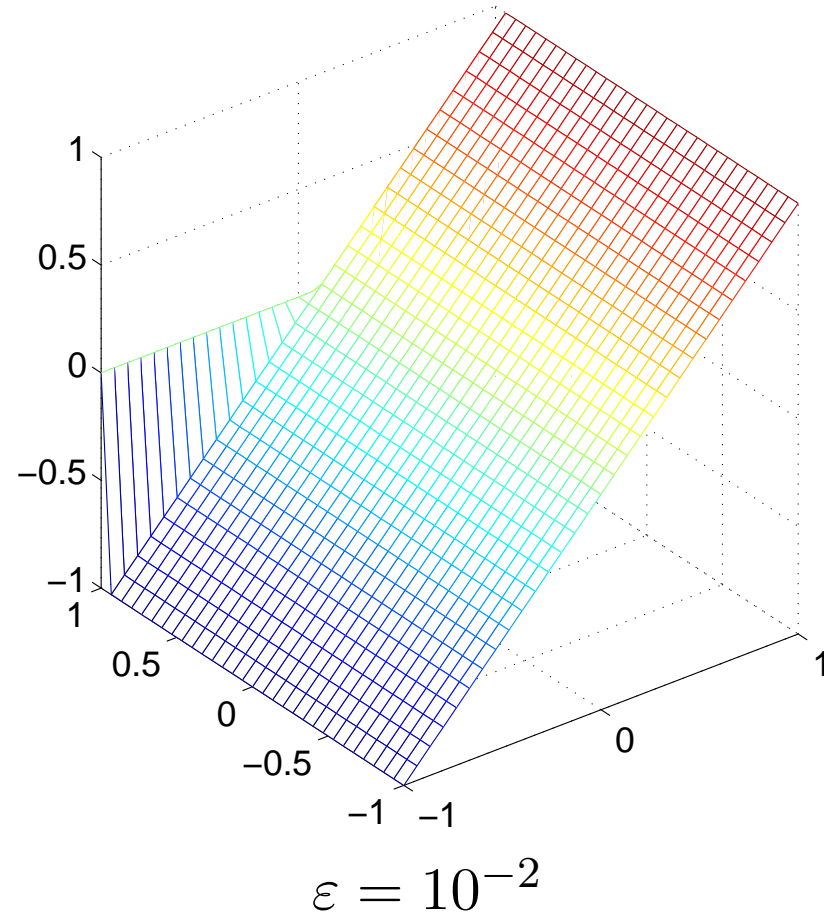
$$\begin{aligned} -\varepsilon \Delta u + \mathbf{b} \cdot \nabla u &= f && \text{in } \Omega \\ u &= g && \text{on } \Gamma. \end{aligned}$$

$$\mathbf{b} = (2y(1 - x^2), -2x(1 - y^2)),$$

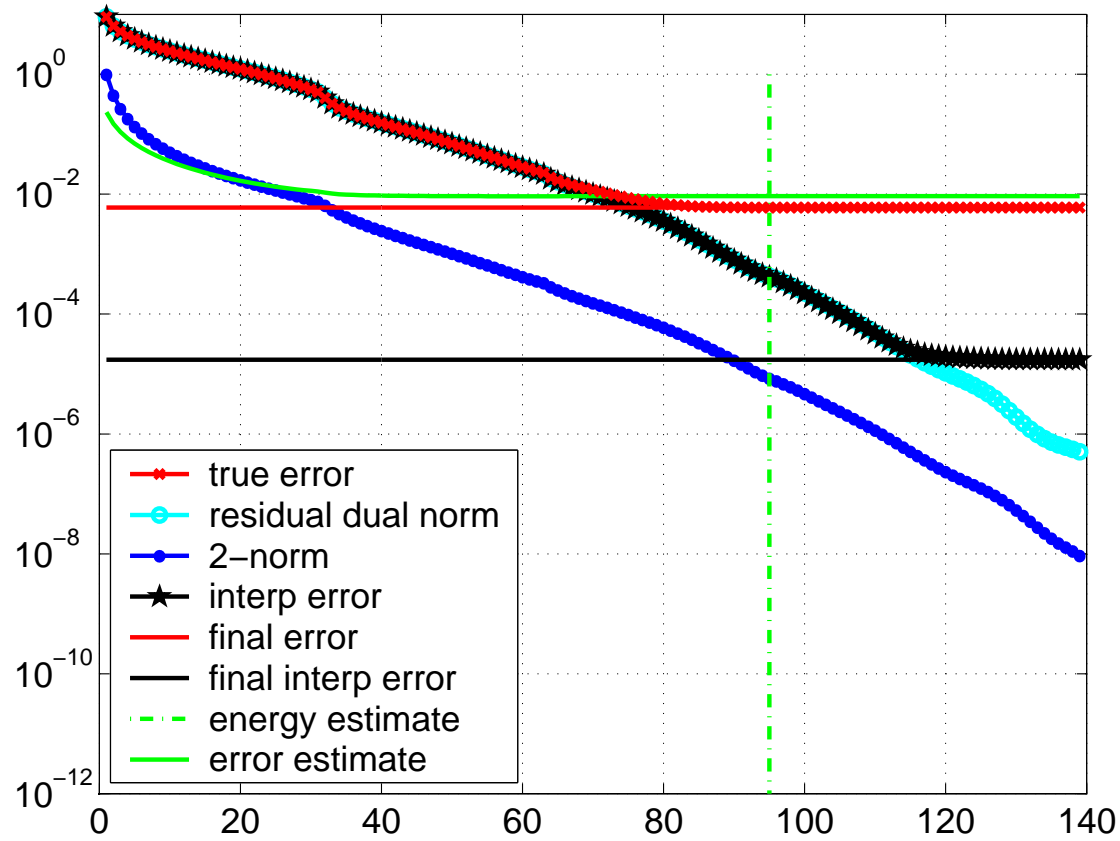
$$u(x, y) = x \left(\frac{1 - e^{\frac{y-1}{\varepsilon}}}{1 - e^{-\frac{2}{\varepsilon}}} \right),$$

$$\|v_h\|_{\mathcal{H}_h}^2 = \varepsilon |v_h|_1^2 + \sum_{T \in \mathcal{T}^h} \delta_T \|\mathbf{b} \cdot \nabla v_h\|_{0,T}^2$$

Advection-diffusion problem

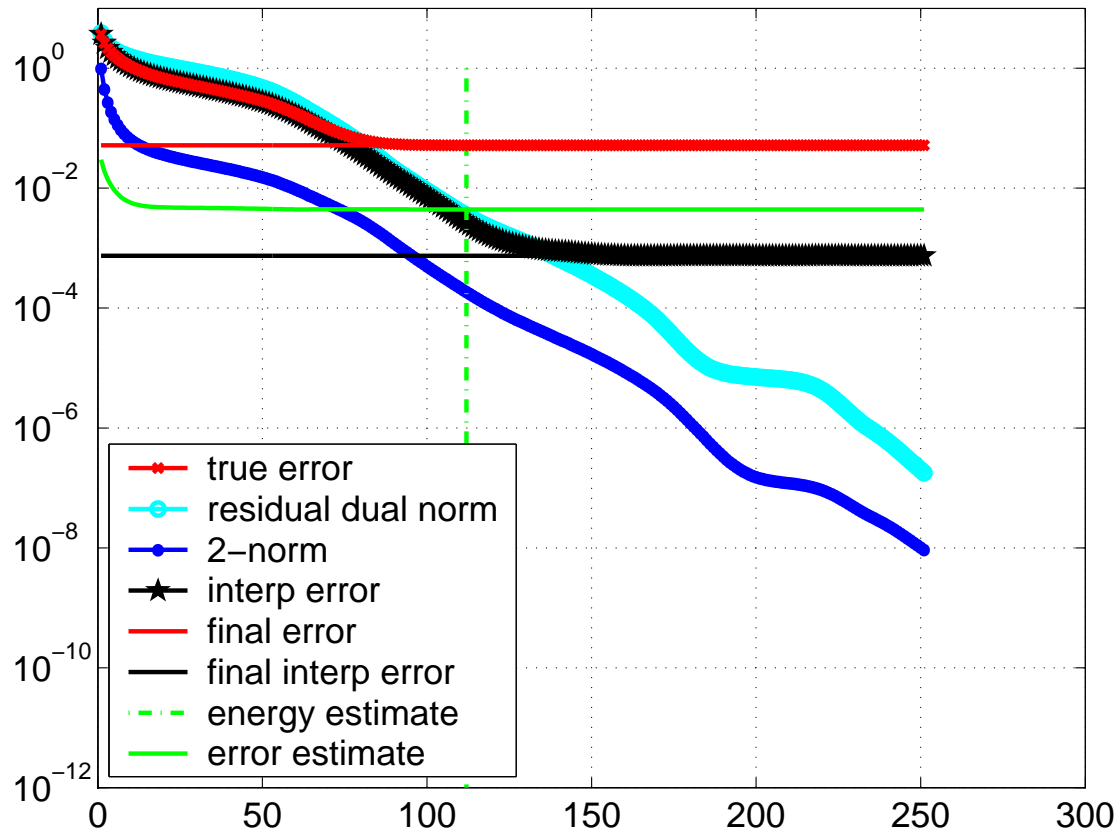


Advection-diffusion problem



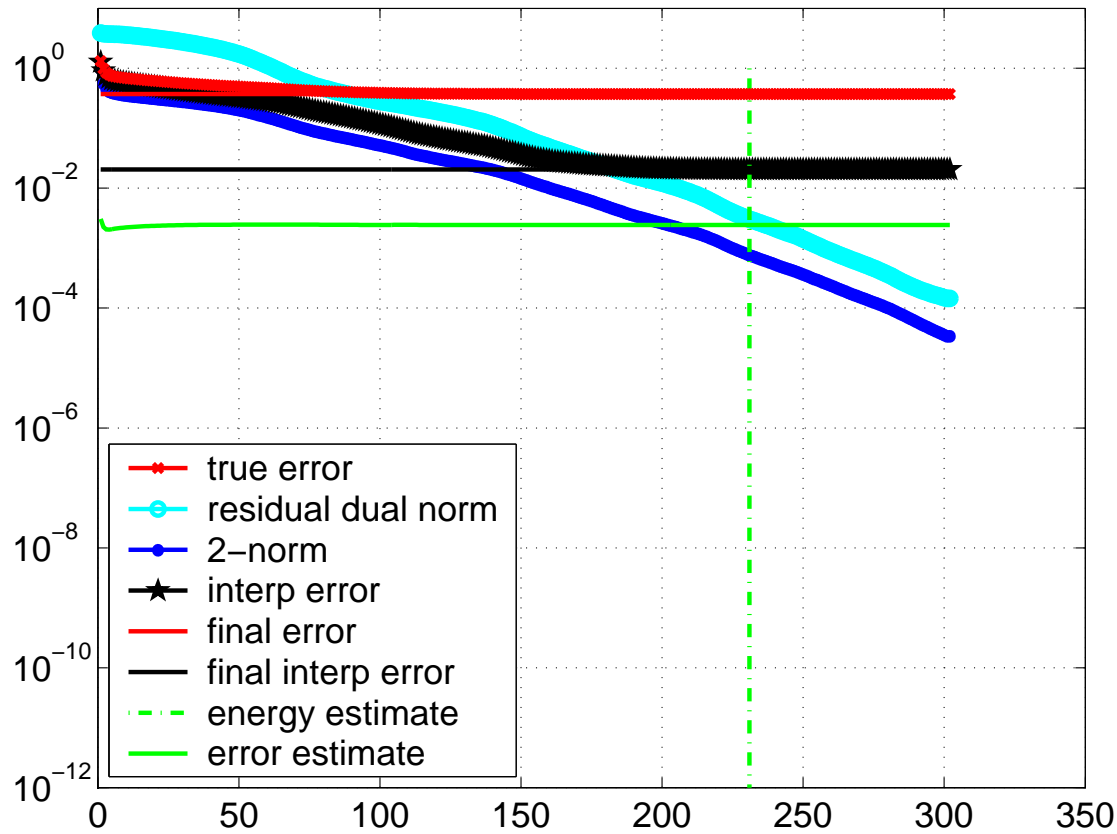
Uniform mesh; $\varepsilon = 1$

Advection-diffusion problem



Uniform mesh; $\varepsilon = 10^{-1}$

Advection-diffusion problem



Uniform mesh; $\varepsilon = 10^{-2}$

Symmetric Indefinite Case

■ Find $u \in \hat{\mathcal{H}}$ such that for all $v \in \hat{\mathcal{H}}$

$$a(u, v) = L(v) \quad (L(\cdot) \in \hat{\mathcal{H}}' \text{ dual space of } \hat{\mathcal{H}})$$

Symmetric Indefinite Case

- Find $u \in \hat{\mathcal{H}}$ such that for all $v \in \hat{\mathcal{H}}$

$$a(u, v) = L(v) \quad (L(\cdot) \in \hat{\mathcal{H}}' \text{ dual space of } \hat{\mathcal{H}})$$

- Existence and uniqueness: $\forall v, w \in \hat{\mathcal{H}}$

$$a(w, v) \leq C_1 \|w\|_{\hat{\mathcal{H}}} \|v\|_{\hat{\mathcal{H}}}$$

$$\sup_{w \in \hat{\mathcal{H}} \setminus \{0\}} \frac{a(w, v)}{\|w\|_{\hat{\mathcal{H}}}} \geq C_2 \|v\|_{\hat{\mathcal{H}}}$$

$$\sup_{v \in \hat{\mathcal{H}} \setminus \{0\}} \frac{a(w, v)}{\|v\|_{\hat{\mathcal{H}}}} \geq C_2 \|w\|_{\hat{\mathcal{H}}}$$

- Bilinear form formulation (Brooks and Hughes (1982))

Stokes

$$a(w, r; v, p) = \nu(\nabla w, \nabla v) - (r, \operatorname{div} v) - (q, \operatorname{div} w)$$

and $L(v, q) = (f, v)$

- Bilinear form formulation (Brooks and Hughes (1982))

Picard iteration, Oseen problem

$$a(w, r; v, p) = \nu(\nabla w, \nabla v) + (b \cdot \nabla w + \theta w, v) - (r, \operatorname{div} v) - (q, \operatorname{div} w)$$

$$\text{and } L(v, q) = (f, v)$$

Symmetric Indefinite Case

Using mixed finite-element discretization

- $\exists \mathbf{u} \in \mathcal{H}_h \quad \mathbf{v}^T \mathcal{A} \mathbf{u} = \mathbf{v}^T \mathbf{f}, \quad \forall \mathbf{v} \in \mathcal{H}_h$
- $\mathcal{H}_h = (\mathbf{R}^N, \|\cdot\|_H)$ and $\mathcal{H}'_h = (\mathbf{R}^N, \|\cdot\|_{H^{-1}})$ H SPD
- C_1 and C_2 independent of discretization

$$\max_{\mathbf{w} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T \mathcal{A} \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \leq C_1$$

$$\min_{\mathbf{w} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T \mathcal{A} \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \geq C_2$$

Using mixed finite-element discretization

- $\exists \mathbf{u} \in \mathcal{H}_h \quad \mathbf{v}^T \mathfrak{A} \mathbf{u} = \mathbf{v}^T \mathbf{f}, \quad \forall \mathbf{v} \in \mathcal{H}_h$
- $\mathcal{H}_h = (\mathbf{R}^N, \|\cdot\|_H)$ and $\mathcal{H}'_h = (\mathbf{R}^N, \|\cdot\|_{H^{-1}})$ H SPD
- C_1 and C_2 independent of discretization

$$\max_{\mathbf{w} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T \mathfrak{A} \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \leq C_1$$

$$\min_{\mathbf{w} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T \mathfrak{A} \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \geq C_2$$

which is equivalent to uniform conditioning of \mathfrak{A} with respect to the norm induced by H :

$$\|\mathfrak{A}\|_{H, H^{-1}} \leq C_1, \quad \|\mathfrak{A}^{-1}\|_{H^{-1}, H} \leq C_2^{-1},$$

or, $\kappa_H(\mathfrak{A}) \leq C_1/C_2$.

Symmetric Indefinite Case

Using mixed finite-element discretization

- $\exists \mathbf{u} \in \mathcal{H}_h \quad \mathbf{v}^T \mathfrak{A} \mathbf{u} = \mathbf{v}^T \mathbf{f}, \quad \forall \mathbf{v} \in \mathcal{H}_h$
- $\mathcal{H}_h = (\mathbf{R}^N, \|\cdot\|_H)$ and $\mathcal{H}'_h = (\mathbf{R}^N, \|\cdot\|_{H^{-1}})$ H SPD
- C_1 and C_2 independent of discretization

$$\max_{\mathbf{w} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T \mathfrak{A} \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \leq C_1$$

$$\min_{\mathbf{w} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbf{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T \mathfrak{A} \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \geq C_2$$

which is equivalent to uniform conditioning of \mathfrak{A} with respect to the norm induced by H :

$$\|\mathfrak{A}\|_{H, H^{-1}} \leq C_1, \quad \|\mathfrak{A}^{-1}\|_{H^{-1}, H} \leq C_2^{-1},$$

or, $\kappa_H(\mathfrak{A}) \leq C_1/C_2$. $\mathfrak{A} \mathbf{w} = \sigma \mathbf{H} \mathbf{v}$, $\mathfrak{A}^T \mathbf{v} = \sigma \mathbf{H} \mathbf{w}$ and $\sigma \in (C_2, C_1)$

Symmetric Indefinite Case

Using mixed finite-element discretization \mathfrak{A} has a block-structure

Symmetric Indefinite Case

Using mixed finite-element discretization \mathfrak{A} has a block-structure

$$\mathfrak{A} = \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}.$$

Symmetric Indefinite Case

Using mixed finite-element discretization \mathfrak{A} has a block-structure

$$\mathfrak{A} = \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}.$$

Correspondingly,

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}$$

$$\|\mathbf{v}\|_H^2 = \|\mathbf{v}_1\|_{H_1}^2 + \|\mathbf{v}_2\|_{H_2}^2.$$

Symmetric Indefinite Case

Using mixed finite-element discretization \mathfrak{A} has a block-structure

$$\mathfrak{A} = \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}.$$

Correspondingly,

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}$$

$$\|\mathbf{v}\|_H^2 = \|\mathbf{v}_1\|_{H_1}^2 + \|\mathbf{v}_2\|_{H_2}^2.$$

$$\|\mathbf{f}\|_{H_1^{-1}}^2 = \|\mathbf{f}_1\|_{H_1^{-1}}^2 + \|f_2\|_{H_2^{-1}}^2.$$

Stopping criteria and Krylov methods

The hypotheses of Theorem 0.0 are satisfied and its conclusion applies to the indefinite case.

- We should use $\|\mathbf{r}\|_{H^{-1}}$ to monitor the error in an iterative process.

Arioli Loghin RAL-TR-10-2006

Stopping criteria and Krylov methods

The hypotheses of Theorem 0.0 are satisfied and its conclusion applies to the indefinite case.

- We should use $\|\mathbf{r}\|_{H^{-1}}$ to monitor the error in an iterative process.
- MINRES preconditioned with H

Arioli Loghin RAL-TR-10-2006

The hypotheses of Theorem 0.0 are satisfied and its conclusion applies to the indefinite case.

- We should use $\|\mathbf{r}\|_{H^{-1}}$ to monitor the error in an iterative process.
- MINRES preconditioned with H
- symmetrically-preconditioned GMRES with H (this algorithm is equivalent to right-preconditioned GMRES in the H -inner product).
The GMRES residual is the required quantity, while C_2 can be estimated on a coarse mesh quite successfully .

Arioli Loghin RAL-TR-10-2006

The hypotheses of Theorem 0.0 are satisfied and its conclusion applies to the indefinite case.

- We should use $\|\mathbf{r}\|_{H^{-1}}$ to monitor the error in an iterative process.
- MINRES preconditioned with H
- symmetrically-preconditioned GMRES with H (this algorithm is equivalent to right-preconditioned GMRES in the H -inner product).
The GMRES residual is the required quantity, while C_2 can be estimated on a coarse mesh quite successfully .
- Three terms GMRES with H -norm

Arioli Loghin RAL-TR-10-2006

Conclusions and future work

- The natural norm for FEM problems is the residual dual norm;
- Estimation is possible;
- For non-symmetric problems, preconditioning with the norm is useful;
- Reliable estimates for finite element error are still required: *a posteriori* estimates?.