Convergence and Optimality by Mesh Adaptivity

Andreas Veeser

Università degli Studi di Milano

Certification of Errors in Numerical Simulations
CEA-EdF-INRIA summer school
Paris-Saclay, June 30, 2022

Goal

A posteriori estimates bound the error due to the discretization, the approximate solution, the model etc. The **influence of the mesh** is captured typically in the form

$$\left(\sum_{T \in \mathcal{T}} \mathtt{indicator}_T(\mathtt{disc\_sol}_{|T}, \mathtt{data}_{|T})^2 \right)^{1/2}$$

Our goal now:

Analyze the potential of the use of this **splitting** to construct 'optimal' meshes.

Global plan

1. Approximation based on mesh adaptivity
2. Mesh-adaptive FEMs

We **first** consider the simpler situation where the **target function** is **explicitly known** to us, not only given implicitly by a PDE problem.

This is also relevant for approximating the data in PDE problem and for coarsening in evolution problem.

Local plan

1. Approximation based on mesh adaptivity
   - Mesh adaptivity with piecewise constants
   - Equidistribution of local errors
   - A (self-)adaptive mesh construction

Let's start with the simplest situation I can think of.

Piecewise constants on uniform/arbitrary 1d-meshes

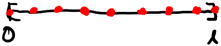
Given a 1d mesh $\mathcal{M} : 0 = x_0 < \cdots < x_N = 1$, define

$$S(\mathcal{M}) := \left\{ s : [0,1[\to \mathbb{R} \mid s_{|[x_{i-1},x_i[} \text{ is constant} \right\}$$

and write $\mathcal{M}_N$ for the uniform mesh with N intervals, ie $x_i = i/N$.

Consider approximation with elements from

$$S_N := S(\mathcal{M}_N) \quad \text{and} \quad \Sigma_N := \bigcup_{\mathcal{M}} S(\mathcal{M}).$$



An element s in

- S_N is determined by N constants,
- Σ_N is determined by N constants and $N-1$ breakpoints.

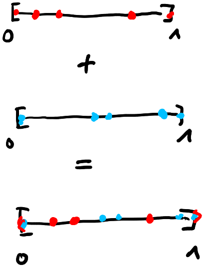
Examples of linear and nonlinear approximation

S_N is a **linear space**, while Σ_N is **not**. In fact,

$$s_1, s_2 \in S_N, \alpha \in \mathbb{R} \implies \alpha s_1 \in S_N \text{ and } s_1 + s_2 \in S_N,$$

and

$$s \in \Sigma_N,\ \alpha \in \mathbb{R} \implies \alpha s \in \Sigma_N,$$
$$s_1, s_2 \in \Sigma_N \not\Longrightarrow s_1 + s_2 \in \Sigma_N.$$



However, Σ_N is **mildly nonlinear** in the sense that

$$s_1, s_2 \in \Sigma_N \implies s_1 + s_2 \in \Sigma_{2N}.$$

Global and local max norm errors

Given a fixed **known** $v \in C^0[0,1]$ and $s \in S(\mathcal{M})$ piecewise constant, take

$$\|v - s\|_{L^\infty} := \max_{x\in[0,1[} |v(x) - s(x)|$$

as **error notion**. Given an interval $I = [a, b] \subset [0, 1[$, introduce the **cell or local error**

$$e(I) := \min_{c\in\mathbb{R}} \|v - c\|_{L^\infty(I)} = \frac{1}{2}\left(\sup_I v - \inf_I v\right)$$

which satisfies

$$\inf_{I\in\mathcal{M}} \|v - s\|_{L^\infty} = \max_{I\in\mathcal{M}} e(I),$$

$$I \subseteq I' \implies e(I) \le e(I'), \qquad \lim_{|I'\setminus I|\to 0} e(I') - e(I) = 0.$$

Equidistribution of local errors

Let $\mathrm{tol} > 0$ and denote by $\#\mathcal{M}$ the number of cells in a mesh $\mathcal{M}$.

We aim for a mesh $\mathcal{M}$ with

$$\inf_{s \in S(\mathcal{M})} \|v - s\|_{L^\infty} \leq \mathrm{tol} \quad \text{and} \quad \#\mathcal{M} \text{ minimal.}$$

Construct t_i by $t_0 := 0$ and

$$t_{i+1} := \max\{t \in [t_i, 1] \mid e[t_i, t[\leq \mathrm{tol}\} \quad \text{whenever} \quad t_i < 1,$$

which essentially **equidistributes the local errors**.

Then the mesh given by $t_0 < \cdots < t_N$ is such a optimal mesh.

Power functions and their regularity

For $\rho > 0$, consider

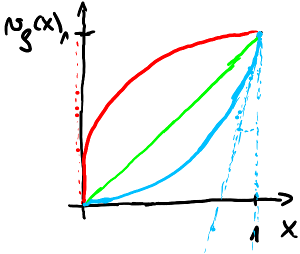
$$v_\rho(x) := x^\rho$$

Then $v_\rho \in C^0[0,1]$ and

$$\rho \geq 1 \implies \|v_\rho'\|_{L^\infty} = \rho,$$

$$\rho \in \,]0,1[\, \implies \text{merely } v_\rho \in C^{0,\rho}[0,1]$$

$$\|v_\rho'\|_{L^1} := \int_0^1 |v_\rho'| = 1.$$



Note that monotoncity of v_ρ yields

$$e_\rho(I) = \inf_{c\in\mathbb{R}} \|v - c\|_{L^\infty(I)} = \tfrac{1}{2}\big[v(\max I) - v(\min I)\big] = \tfrac{1}{2}\int_I |v_\rho'|,$$

which may be considered a special case of the **Bramble-Hilbert lemma** with a integrability shift.

About convergence rates

The **interval length** N^{-1} leads to

$$\inf_{s \in S_N} \|v_\rho - s\|_\infty \simeq \frac{1}{2} \begin{cases} \rho N^{-1}, & \rho \geq 1, \\ N^{-\rho}, & \rho \in]0,1[, \end{cases}$$

while **equilibrating** with $\left\|v'_\rho\right\|_{L^1(t_{i-1},t_i)} = N^{-1} \left\|v'_\rho\right\|_{L^1(0,1)}$ gives

$$\inf_{s \in \Sigma_N} \|v_\rho - s\|_{L^\infty} = \frac{1}{2} N^{-1}.$$

No improvement for $\rho = 1$, case without **'local features'**.

Thus, mesh adaptivity **may improve**

$$\texttt{error} \leq CN^{-r}$$

by reducing C or **by enlarging** r; the latter typically more dramatic.

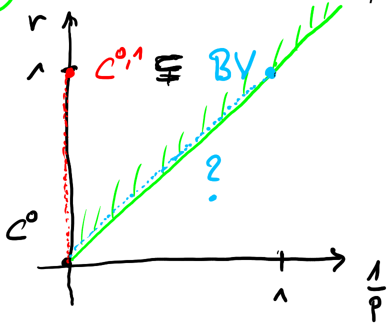
Piecewise constants in the DeVore diagram

In general, for $v \in C^0[0,1]$,

$$\inf_{s \in S_N} \|v - s\|_\infty = O(N^{-1}) \iff v \in C^{0,1}[0,1]$$

$$\inf_{s \in \Sigma_N} \|v - s\|_\infty = O(N^{-1}) \iff v \in BV[0,1] \quad \text{(Kahane '61)}$$

Recall $W^{r,p}(0,1) \subseteq C^0[0,1] \implies r \geq 0$ and $r - \frac{1}{p} \geq 0$.



Local plan

1. Approximation based on mesh adaptivity
   - Mesh adaptivity with piecewise constants
   - Equidistribution of local errors
   - A (self-)adaptive mesh construction

Let's see how far the idea of equidistributing the local errors goes ...

Abstract equidistribution

Suppose that, for any mesh $\mathcal{T}$, its global error $E(\mathcal{T})$ has the **splitting**

$$E(\mathcal{T}) \leq C_1 \left(\sum_{T\in\mathcal{T}} e(T)^p\right)^{1/p},$$

and allows for the following bound with **shifted summability**: for some fixed $q \in \,]0, p[$ and C_2, there holds

$$\left(\sum_{T\in\mathcal{T}} e(T)^q\right)^{1/q} \leq C_2.$$

If can construct a **mesh $\mathcal{T}$ with equidistributed local errors**,

$$\forall T \in \mathcal{T} \;\; e(T) = t \quad \text{for some } t > 0,$$

then $tN^{1/q} \leq C_2$ and therefore

$$E(\mathcal{T}) \leq C_1 t N^{1/p} \leq C_1 C_2 N^{1/p-1/q}.$$

Best error localizations in H^1

Given a shape-regular triangulation $\mathcal{T}$ of $\Omega \subseteq \mathbb{R}^d$, set

$$S(\mathcal{T}) := \left\{ s \in C^0(\bar{\Omega} \mid s_{|\partial\Omega} = 0,\ \forall T \in \mathcal{T}\ s_{|T} \in \mathbb{P}_\ell(T) \right\}.$$

Then, for any $v \in H_0^1(\Omega)$, we have **best error localization**

$$\inf_{s \in S(\mathcal{T})} \|\nabla(v - s)\|^2_{L^2(\Omega)} \leq C_1 \sum_{T \in \mathcal{T}} \inf_{p \in \mathbb{P}_\ell(T)} \|\nabla(v - p)\|^2_{L^2(T)};$$

cf. Veeser '16 and Aurada et al. '13.

Best error localizations - proof in 1d

Fix a cell $I = [x_{i-1}, x_i]$. A local best approximation $p_i \in \mathbb{P}_\ell[x_{i-1}, x_i]$ is determined only up to a constant, which we may choose such that $p_i(x_{i-1}) = v(x_{i-1})$.

Moreover, exploiting its Galerkin orthogonality with $q(x) = x$ yields

$$0 = \int_{x_{i-1}}^{x_i} (v - p_i)' q' = \int_{x_{i-1}}^{x_i} (v - p_i)' = (v - p_i)(x_i)$$

Thus, gluing these p_i together yields globally continuous best approximation and so even $C_1 = 1$.

Further best error localizations

There are further best error localizations for

- the L^2-norm and reaction-diffusion norm (Tantardini/Verfürth/Veeser '15),
- the H^{-1}-norm (Blechta/Málek/Vohralík '16, Tantardini/Verfürth/Veeser '17),
- other operators/elements (Ciarlet/Vohralík '17, Ern/Smears/Vohralík, ...).

Generalized Bramble-Hilbert lemmas

If $d = 2$, the Sobolev embedding $H^1(\Omega) \subseteq W^{2,1}(\Omega)$ implies the Bramble-Hilbert-like inequality

$$\inf_{p \in \mathbb{P}_\ell(T)} \|\nabla(v - p)\|_{L^2(T)} \leq C_2 \, |v|_{W^{2,1}(T)}$$

Note that, for $\Omega = B(0,1) \subset \mathbb{R}^2$, we have

$$|x|^\rho \in H^1 \iff \rho > 0 \iff |x|^\rho \in W^{2,p} \text{ for some } p > 1.$$

More generally, Veeser '16 and Gaspoz/Morin '14 show

$$\begin{aligned}\inf_{p \in \mathbb{P}_\ell(T)} \|\nabla(v - p)\|_{L^2(T)} &\leq C_2' \inf_{\tilde{p} \in \mathbb{P}_{\ell-1}(T)^d} \|\nabla v - \tilde{p}\|_{L^2(T)} \\ &\leq C_2 \, |\nabla v|_{B^r_q\left(L_q(T)\right)}\end{aligned}$$

with

$$r \leq \ell + 1, \quad q > 0 \quad \text{and} \quad r - \frac{d}{q} = 1 - \frac{d}{2}.$$

H^1-equilibration in DeVore diagram

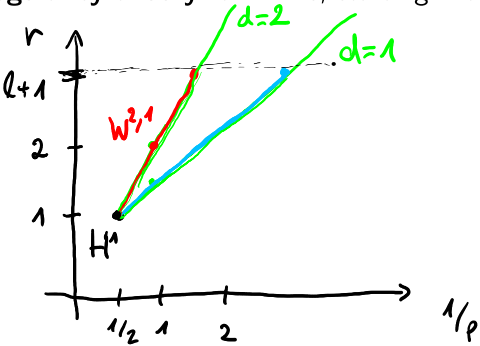
Thus, equilibration of $\inf_{p\in\mathbb{P}_\ell(T)} \|\nabla(v-p)\|_{L^2(T)}$, $T\in\mathcal{T}$, would lead to the rate

$$N^{-(r-1)/d}$$

if $|\nabla v|_{B^r_q(L_q(\Omega))}<\infty$. For 1d-splines and not only H^1, cf. Petrushev '88.

There is a **Besov regularity theory** for PDEs, starting with Dahlke/DeVore '97.

•



Local plan

1. Approximation based on mesh adaptivity
   - Mesh adaptivity with piecewise constants
   - Equidistribution of local errors
   - A (self-)adaptive mesh construction

Apart from 1d, we haven't constructed any meshes up to now.

Bisection guided by maximum strategy

For simplicity, reconsider approximation with 1-dimensional piecewise constants in maximum norm.

Given an interval I, denote by $\mathbf{bisect}(I) := \{I_1, I_2\}$ the pair of intervals generated by the subdivision of I in its midpoint.

Set $N := 0$, $\mathcal{M}_0 := \{[0,1]\}$ and iterate (cf. Birman/Solomyak '67)

1. $t_N := \mathbf{max}_{I \in \mathcal{M}_N} e(I)$
2. if $t_N = 0$, then STOP
3. pick some $I_N \in \mathcal{M}_N$ with $e(I_N) = t_N$
4. $\mathcal{M}_{N+1} := \big(\mathcal{M}_N \setminus \{I_N\}\big) \cup \mathtt{bisect}(I_N)$
5. increment N

This **iterative feedback process** can be recorded by a **binary tree**.

Regularity for best convergence rate

Let

$$A_N := \{s \mid s \text{ pw constant on mesh generated with} \leq N-1 \text{ bisections}\}$$

denote the counterpart of S_N and Σ_N.

DeVore '87 shows that

$$\inf_{s \in A_N} \|v - s\|_{L^\infty} \leq \left\|Mv'\right\|_{L^1} N^{-1}$$

where Mv' is the **maximal function** of v' satisfying, cf. Bennett/Sharpley '88,

$$\left\|Mv'\right\|_{L^1} < \infty \iff \int_0^1 |v'| \log(1 + |v'|) < \infty.$$

Some comparison and a remark

The error decay N^{-1} with pw constants is dictated by

uniform	bisection	'free'
$\lvert u \rvert_{C^{0,1}} = \lVert u' \rVert_{L^\infty}$	$\lVert Mu' \rVert_{L^1}$	$\lVert u' \rVert_{L^1} \geq \mathrm{var}(u)$

Remember: **On a computer**, we have $N \leq N_{\max}$ with $N_{\max}$ finite but growing with time ...

The above algorithm **may not fully exploit** the potential of A_N if

- the local errors 'sum' in ℓ_p, $p < \infty$ and
- a single bisection does not reduce the error at least by a fixed fraction.

Tree approximation by Binev '16 provides a remedy by applying the maximum strategy on modified, history-dependent indicators.

Getting more general

- Bisection generalizes to **conforming shape regular simplicial meshes**; cf. Binev/Dahmen/DeVore '04, Stevenson '08.
- For generalization to **piecewise polynomials** of (fixed) higher order, see Chen/Xu, Binev/Dahmen/DeVore/Petrushev '02, Gaspoz/Morin '14, ...

Bisection appears to be a good compromise between flexibility and algorithmic convenience.

Related surveys

- R. DeVore, Nonlinear approximation, Acta Numerica 7, 51-150, 1998.
- A. Veeser, Adaptive tree approximation with finite element functions: a first look. In: Daniele Di Pietro et al. (Eds.), Numerical methods for PDEs: State of the Art Techniques, SEMA SIMAI Springer Series, vol 15. Springer, Cham, 249-284, 2018.
- R. H. Nochetto, K. G. Siebert, A. Veeser, Theory of Adaptive FEM: An introduction, in: Multiscale, Nonlinear and Adaptive Approximation, DeVore/Kunoth (Eds.), Springer, 2009

Global and local plan

1. Approximation based on mesh adaptivity

2. Mesh-adaptive FEMs
   - Setting
   - Convergence
   - Rate optimality

The real game: apply mesh adaptivity to the numerical solution of PDEs – the main new issue is to deal with the **global dependence in the indicators**.

We first fix a ‘model’ setting.

Model problem and structure of algorithm

Considering the **Poisson problem**

$$-\Delta u = f \text{ in } \Omega \subset \mathbb{R}^2, \qquad u = 0 \text{ on } \partial\Omega,$$

let's **keep the iterative feedback nature** of the bisection algorithm.

Given an initial edge-to-edge triangulation $\mathcal{T}_0$, set $k := 0$ and iterate

1. $U_k :=$ **solve**$(f, \mathcal{T}_k)$
2. $\{\eta_{k;T}\}_{T\in\mathcal{T}_k} :=$ **estimate**(U_k, f)
3. $\hat{\mathcal{T}}_k :=$ **mark**$(\mathcal{T}_k, \{\eta_{k;T}\}_{T\in\mathcal{T}_k})$
4. $\mathcal{T}_{k+1} :=$ **refine**$(\mathcal{T}_k, \hat{\mathcal{T}}_k)$
5. increment k

Note each step requires (at least) $\#\mathcal{T}_k$ operations.

Step 'solve'

Let's take **linear finite elments**. Given a triangulation $\mathcal{T}$, set

$$S(\mathcal{T}) := \{v : \Omega \to \mathbb{R} \mid \forall v \in C^0(\Omega),\ v_{|\Omega} = 0, T \in \mathcal{T}\ v_{|T} \in \mathbb{P}_1\},$$

Then the **Galerkin solution**

$$U_k = \textbf{solve}(f, \mathcal{T}_k) \in S(\mathcal{T}_k)$$

verifies

$$\forall \varphi \in S(\mathcal{T}_k) \qquad \int_\Omega \nabla U_k \cdot \nabla \varphi = \langle f, \varphi \rangle\,.$$

Note that this is defined for $f \in H^{-1}(\Omega)$.

Step 'estimate' - extract information

We have that $f \leftrightarrow u$, where $f \to u$ is global, while $u \to f$ is local.

Let $\mathcal{T}$ be some refinement of $\mathcal{T}_0$. Replacing f and u by the residual $R_{\mathcal{T}} := f + \Delta U_{\mathcal{T}}$ and the error $U_{\mathcal{T}} - u$, respectively, suggests that we cannot expect better than the following.

Omitting f and $U_{\mathcal{T}}$ in the indicators, we have the **global upper bound**

$$|U_{\mathcal{T}} - u|_{\Omega} := \|\nabla(U_{\mathcal{T}} - u)\|_{L^2(\Omega)} \leq C_U \left(\sum_{T \in \mathcal{T}} \eta_{\mathcal{T}}(T)^2 \right)^{1/2}$$

and, for any $T \in \mathcal{T}$, the **local lower bound**

$$C_L \eta_{\mathcal{T}}(T) \leq |U_{\mathcal{T}} - u|_{\omega_{\mathcal{T}}(T)}$$

where $\omega_{\mathcal{T}}(T)$ is a $\mathcal{T}$-neighborhood of T; cf. Kreuzer/Veeser '21.

Step 'mark' - elaborate information

Let $\delta \in \,]0,1[$, to be chosen later. Following Dörfler '96, choose

$$\hat{\mathcal{T}}_k = \mathbf{mark}(\mathcal{T}_k, \{\eta_{k;T}\}_{T\in\mathcal{T}_k})$$

such that a **fixed fraction of the total estimator** is collected with **(near) minimal cardinality**:

$$\sum_{T\in\hat{\mathcal{T}}_k} \eta_{k;T}^2 \geq \delta^2 \sum_{T\in\mathcal{T}_k} \eta_{k;T}^2$$

Minimal cardinality is reached with the largest indicators; near minimal cardinality can be reached with **linear complexity** by 'binning' ; cf. Praetorius/Pfeiler '19.

Note that $\hat{\mathcal{T}}_k$ is invariant upon multiplying the indicators with a fixed positive factor.

Step 'refine' - generate new mesh

Generate a new triangulation

$$\mathcal{T}_{k+1} = \textbf{refine}(\mathcal{T}_k, \hat{\mathcal{T}}_k)$$

by 2d bisection such that

- each marked triangle $T \in \hat{\mathcal{T}}_k$ is at least '2.5 times' bisected,
- shape regularity is uniformily bounded,
- conformity is re-established.

Note that the last item entails additional refinements; to control their number is a nontrivial task that we do not address here; cf. Dahmen/Binev/DeVore '04, Stevenson '08.

A simplifying (and practical) assumption

Let $\mathcal{T}$ be any refinement of $\mathcal{T}_0$, z be a vertex of $\mathcal{T}$ and denote by $\omega_{\mathcal{T}}(z)$ the star around z in $\mathcal{T}$. Moreover, let $S^+_{\mathcal{T}}(z)$ the space spanned by the hat-shaped bubbles associated with each triangle and interior edge of ω_z.

Suppose, for any such vertex z,

$$\|f\|_{H^{-1}(\omega_{\mathcal{T}}(z))} \leq C_S \sup_{\varphi \in S^+_{\mathcal{T}}(z), |\varphi|_{\omega_{\mathcal{T}}(z)}=1} \langle f, \varphi \rangle .$$

This **saturation assumption** ensures that an essential part of f can be seen on the 'next' local refinement level and so yields $\|f\|_{H^{-1}(\omega_{\mathcal{T}}(z))}$ approximately computable.

It **excludes dominating data oscillation** and requires that data is essentially resolved on $\mathcal{T}_0$.

Local lower bounds for corrections

The saturation of data implies the following **variant of the local lower bound**:

For any $T \in \mathcal{T}$ such that each triangle in $\omega_{\mathcal{T}}(T)$ has vertices of $\mathcal{T}_*$ in its interior, we have the following lower bound for a **correction**:

$$\tilde{C}_L \eta_{\mathcal{T}}(T) \leq |U_{\mathcal{T}_*} - U_{\mathcal{T}}|_{\omega_{\mathcal{T}}(T)}$$

with $\tilde{C}_L \leq C_L$.

Note that this cannot hold when data, ie f, is not 'resolved' by $\mathcal{T}_*$.

Local plan

2. Mesh-adaptive FEMs
   - Setting
   - Convergence
   - Rate optimality

Following Dörfler '96 and Morin/Nochetto/Siebert '00, we first establish linear convergence w.r.t. to the iteration number.

Galerkin orthogonality and error monotonicity

Let $\mathcal{T}_*$ be a refinement of $\mathcal{T}$.

Then the **Galerkin orthogonality**

$$\int_\Omega \nabla(u - U_{\mathcal{T}_*}) \cdot \nabla(U_{\mathcal{T}} - U_{\mathcal{T}_*}) = 0$$

yields the (global) **error monotonicty**:

$$\begin{aligned} |u - U_{\mathcal{T}_*}|_\Omega^2 &= |u - U_{\mathcal{T}}|_\Omega^2 - |U_{\mathcal{T}} - U_{\mathcal{T}_*}|_\Omega^2 \\ &\geq |u - U_{\mathcal{T}}|_\Omega^2 . \end{aligned}$$

Linking error and correction

Thanks to the global upper bound, the marking strategy and the local lower bounds for the correction, we derive

$$|U_k - u|_\Omega^2 \le C_U^2 \sum_{T\in\mathcal{T}_k} \eta_{k;T}^2 \le \frac{C_U^2}{\delta^2} \sum_{T\in\hat{\mathcal{T}}_k} \eta_{k;T}^2$$
$$\le \frac{C_U^2}{\delta^2 \tilde{C}_L^2} |U_k - U_{k+1}|_\Omega^2$$

Thus, the previous orthogonality gives the **strict error reduction**

$$|U_{k+1} - u|_\Omega \le \sqrt{1 - \delta^2 \frac{\tilde{C}_L^2}{C_U^2}}\, |U_k - u|_\Omega \, .$$

Convergence - summary

Under the above assumptions and for any $\delta \in (0,1)$, we have

$$|u - U_k|_\Omega \leq C\alpha^k$$

with

$$\alpha = \sqrt{1 - \delta^2 \frac{\tilde{C}_L^2}{C_U^2}}.$$

Note

- The ratio $\frac{\tilde{C}_L}{C_U} \leq 1$ appears to be a quality measure.
- We didn't use minimal cardinality.
- The proof through strict error reduction requires some regularity on f.

Convergence - alternatives

Morin/Siebert/Veeser '08 and Siebert '11 present **alternative approaches** that

- cover inf-sup stable, conforming methods for well-posed problems and more general marking strategies,
- neither use nor conclude strict error reduction, but only plain convergence.

For discontinuous Galerkin methods in the spirit of the alternatives, see Kreuzer/Georgoulis '18.

Local plan

2. Mesh-adaptive FEMs
   - Setting
   - Convergence
   - Rate optimality

Following Stevenson '07, we quantify the convergence speed w.r.t. #DOFs.

Rate optimality

Denote by $\mathbb{T}$ the class of all meshes that can be generated from $\mathcal{T}_0$ and, for $N \in \mathbb{N}$, set

$$A_N := \bigcup_{\mathcal{T} \in \mathbb{T}_N} S(\mathcal{T}) \quad \text{with} \quad \mathbb{T}_N := \{\mathcal{T} \in \mathbb{T} \mid \#\mathcal{T} - \#\mathcal{T}_0 \leq N\}.$$

Let $r > 0$. We say that the presented algorithm is **r-rate optimal** whenever its outputs $U_k \in S_{N_k}$, $N_k = \#\mathcal{T}_k - \mathcal{T}_0$, verify the following implication:

$$\inf_{s \in A_N} |u - s|_\Omega \leq CN^{-r} \implies |u - U_k|_\Omega \leq C' CN_k^{-r}$$

Rate optimality is weaker than instance optimality.

Upper bound for correction

Consider again a refinement $\mathcal{T}^*$ of $\mathcal{T}$ and denote by

$$\mathcal{T}_R := \mathcal{T} \setminus (\mathcal{T} \cap \mathcal{T}_*)$$

the triangles of $\mathcal{T}$ that are refined in $\mathcal{T}^*$.

Then the following **upper bound for the correction** holds with $\tilde{C}_U \geq C_U$:

$$|U_{\mathcal{T}_*} - U_{\mathcal{T}}|_\Omega \leq \tilde{C}_U \left(\sum_{T \in \mathcal{T}_R} \eta_{\mathcal{T}}(T)^2 \right)^{1/2} .$$

In fact, the relevant residual norm is $\sup\{\langle R, \varphi \rangle \mid \varphi \in S(\mathcal{T}^*), |\varphi| \leq 1\}$.
and its test functions are "in $S(\mathcal{T})$ on $\cup_{T \in \mathcal{T}_* \cap \mathcal{T}} T$".

Error reduction and Dörfler strategy - the converse way

Let $\mu \in]0,1[$ be a reduction factor. If

$$|u - U_{\mathcal{T}_*}|_\Omega \leq \mu \, |u - U_{\mathcal{T}}|_\Omega ,$$

then the local lower bounds, orthogonality, and the previous upper bound imply

$$\begin{aligned}(1-\mu^2) C_L^2 \sum_{T \in \mathcal{T}} \eta_{\mathcal{T}}(T)^2 &\leq (1-\mu^2) \, |u - U_{\mathcal{T}}|_\Omega^2 \\ &\leq |u - U_{\mathcal{T}}|_\Omega^2 - |u - U_{\mathcal{T}_*}|_\Omega^2 = |U_{\mathcal{T}_*} - U_{\mathcal{T}}|_\Omega^2 \leq \tilde{C}_U^2 \sum_{T \in \mathcal{T}_R} \eta_{\mathcal{T}}(T)^2\end{aligned}$$

ie

$$\delta_\mu^2 \sum_{T \in \mathcal{T}} \eta_{\mathcal{T}}(T)^2 \leq \sum_{T \in \mathcal{T}_R} \eta_{\mathcal{T}}(T)^2 \quad \text{with} \quad \delta_\mu = \frac{C_L}{\tilde{C}_U} \sqrt{1-\mu^2} \in \Big]0, \frac{C_L}{\tilde{C}_U}\Big[.$$

Linking to approximability of u

Assume approximability with rate $r > 0$:

$$\inf_{s \in A_N} |u - s|_\Omega \leq C N^{-r}.$$

Then there exists a partition $\mathcal{T}_\mu$ such that

$$\inf_{s \in S(\mathcal{T}_\mu)} |u - s|_\Omega \leq \mu \, |u - U_\mathcal{T}|_\Omega \quad \text{and} \quad \#\mathcal{T}_\mu \leq N \leq C(\mu, r) \, |u - U_\mathcal{T}|_\Omega^{-1/r}.$$

Let $\mathcal{T}^*$ be the minimal common refinement of $\mathcal{T}_\mu$ and $\mathcal{T}$. Then

$$\#\mathcal{T}_R \leq \#\mathcal{T}^* - \#\mathcal{T} \leq \#\mathcal{T}_\mu \leq N \leq C(\mu, r) \, |u - U_\mathcal{T}|_\Omega^{-1/r},$$

which limits the number of refined elements to achieve strict error reduction in terms of actual error and approximation rate.

Combing with Dörfler marking

Assuming

$$\delta \in]0, \frac{C_L}{\tilde{C}_U}[,$$

we can choose $\mu \in]0, 1[$ such that $\delta_\mu = \delta$.

Consequently, the preceding arguments with $\mathcal{T} = \mathcal{T}_k$ and the minimal cardinality in Dörfler marking limits the marked elements by

$$\#\hat{\mathcal{T}}_k \leq \#\mathcal{T}_R \leq C \, |u - U_k|_\Omega^{-1/r} \, .$$

Iteration history

Let $K \in \mathbb{N}$ an iteration number. Combining the last result with a result on the re-establishment of conformity, $|u - U_K|_\Omega < \alpha^{K-k} |u - U_k|_\Omega$, we finally obtain

$$N_K = \mathcal{T}_K - \mathcal{T}_0 \leq C \sum_{k=0}^{K-1} \# \hat{\mathcal{T}}_k \leq C \sum_{k=0}^{K-1} |u - U_k|_\Omega^{-1/r}$$
$$\leq \sum_{k=0}^{K-1} \alpha^{(K-k)/r} |u - U_K|_\Omega^{-1/r} \leq C(\alpha, r) |u - U_K|^{-1/r}$$

ie

$$|u - U_K|_\Omega \leq C N_K^{-r}.$$

Rate optimality - summary

Under the above assumptions and

$$\delta \in \left]0, \frac{C_L}{\tilde{C}_U}\right[,$$

the presented algorithm is rate optimal, ie, for any $r > 0$, we have

$$\inf_{s \in A_N} |u - s|_\Omega \leq C N^{-r} \implies |u - U_k|_\Omega \leq C' C N_k^{-r}$$

Note

- Again, the ratio $\frac{C_L}{\tilde{C}_U} \leq 1$ appears as a quality measure.
- The choice of the parameter δ becomes delicate for a ‘bad’ estimator; cf. Diening/Kreuzer.

Rate optimality - extensions

- Feischl '19 covers stationary Stokes problem with Taylor-Hood by verifying generalized quasi-orthogonality
- Gantner/Haberl/Praetorius/Schimanko '21 proves rate optimality w.r.t. to overall computational cost
- Haberl/Praetorius/Schimanko/Vohralík '21 addresses nonlinear operators and algebraic solvers
- ...

Related surveys

- R. H. Nochetto, A. Veeser, Primer of Adaptive FEM, in: Lecture Notes in Math 2040, 2012.
- R. H. Nochetto, K. G. Siebert, A. Veeser, Theory of Adaptive FEM: An introduction, in: Multiscale, Nonlinear and Adaptive Approximation, DeVore/Kunoth (Eds.), Springer, 2009
- C. Carstensen, M. Feischl, M. Page, D. Praetorius, Axioms of adaptivity, Computers & Mathematics with Applications 67 (6), 1195-1253, 2014.

Questions?

1. Approximation based on mesh adaptivity
   - Mesh adaptivity with piecewise constants
   - Equidistribution of local errors
   - A (self-)adaptive mesh construction

2. Mesh-adaptive FEMs
   - Setting
   - Convergence
   - Rate optimality

Thank you for your attention!