On the derivation of guaranteed and p-robust a posteriori error estimates for the Helmholtz equation

Théophile Chaumont-Frelet, Alexandre Ern, and **Martin Vohralík**

Inria Paris & Ecole des Ponts

ANR APOWA kick-off, March 26, 2024

Outline

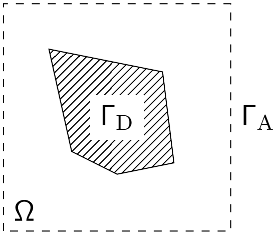
Outline

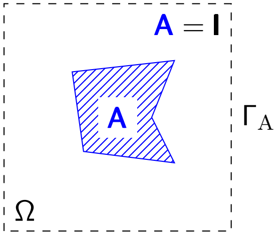
Model problem

Given $f : \Omega \to \mathbb{C}$, find $u : \Omega \to \mathbb{C}$ such that

$$\begin{cases} -\omega^2 \mu u - \nabla \cdot (\mathbf{A} \nabla u) &= \mu f & \text{in } \Omega, \\ u &= 0 & \text{on } \Gamma_{\mathrm{D}}, \\ \mathbf{A} \nabla u \cdot \boldsymbol{n} - i\omega\gamma u &= 0 & \text{on } \Gamma_{\mathrm{A}}, \end{cases}$$

where μ, $\mathbf{A}$, and γ are given strictly positive (definite) coefficients.





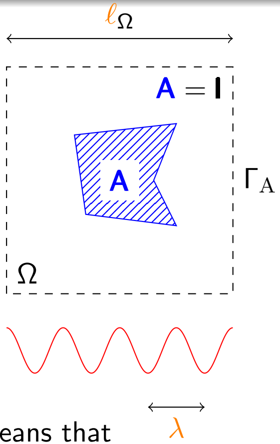
What do we mean by high frequency?

The physical meaning of μ and $\mathbf{A}$ depends on the application, but the wavespeed is always given by:

$$c_{\min} := \sqrt{\frac{\sigma_{\min}(\mathbf{A})}{\mu}}$$

The (minimal) wavelength is then given by:

$$\lambda := \frac{2\pi}{\omega} c_{\min}.$$



The "important" quantity is $N_\lambda := \ell_\Omega/\lambda$. High-frequency means that

$$\frac{\omega \ell_\Omega}{c_{\min}} \simeq N_\lambda$$

is "large" (a few tens or hundreds).

Helmholtz problem and its weak formulation

Recall the Helmholtz problem:

$$\left\{\begin{aligned} -\omega^2\mu u - \nabla\cdot(\mathbf{A}\nabla u) &= \mu f && \text{in } \Omega,\\ u &= 0 && \text{on } \Gamma_{\mathrm{D}},\\ \mathbf{A}\nabla u\cdot\boldsymbol{n} - i\omega\gamma u &= 0 && \text{on } \Gamma_{\mathrm{A}}.\end{aligned}\right.$$

Weak formulation: assuming $f \in L^2(\Omega)$, we seek $u \in H^1_{\mathrm{D}}(\Omega)$ such that

$$b(u,v) = (\mu f, v) \quad \forall v \in H^1_{\mathrm{D}}(\Omega),$$

where

$$b(u,v) := -\omega^2(\mu u, v)_\Omega - i\omega(\gamma u, v)_{\Gamma_{\mathrm{A}}} + (\mathbf{A}\nabla u, \nabla v)_\Omega.$$

Helmholtz problem and its weak formulation

Recall the Helmholtz problem:

$$\begin{cases} -\omega^2 \mu u - \nabla \cdot (\mathbf{A} \nabla u) &= \mu f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \Gamma_{\mathrm{D}}, \\ \mathbf{A} \nabla u \cdot \boldsymbol{n} - i\omega\gamma u &= 0 \quad \text{on } \Gamma_{\mathrm{A}}. \end{cases}$$

Weak formulation: assuming $f \in L^2(\Omega)$, we seek $u \in H^1_{\mathrm{D}}(\Omega)$ such that

$$b(u, v) = (\mu f, v) \quad \forall v \in H^1_{\mathrm{D}}(\Omega),$$

where

$$b(u, v) := -\omega^2 (\mu u, v)_\Omega - i\omega(\gamma u, v)_{\Gamma_{\mathrm{A}}} + (\mathbf{A} \nabla u, \nabla v)_\Omega.$$

Finite element approximation

We consider a mesh $\mathcal{T}_h$ of Ω formed by tetrahedral elements K.
Mesh parameter: $h_K := \operatorname{diam} K \leq h$.
The coefficients $\mu, \gamma, \mathbf{A}$ are constant inside each element/face.

We introduce the finite element discretization space

$$V_h := \left\{ v_h \in H^1_{\mathrm{D}}(\Omega) \mid v_h|_K \in \mathbb{P}_p(K) \ \forall K \in \mathcal{T}_h \right\}$$

with $p \geq 1$.

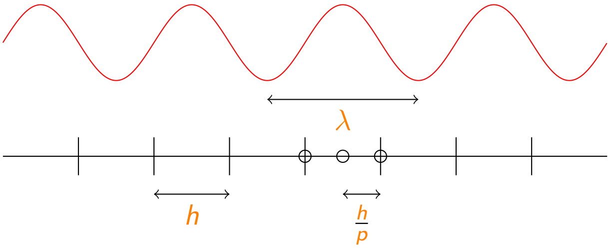
P.G. Ciarlet, 1978.

What do we mean by a fine mesh?

"The mesh is fine" means that

$$N_{\mathrm{dofs}/\lambda} = \lambda \Big/ \frac{h}{p} \simeq \left(\frac{\omega h}{c_{\min} p}\right)^{-1}$$

is large.



Finite element approximation

Recall that u is the only element of $H^1_{\mathrm{D}}(\Omega)$ such that

$$b(u, v) = (\mu f, v) \quad \forall v \in H^1_{\mathrm{D}}(\Omega).$$

Analogously, the finite element solution $u_h \in V_h$ is such that

$$b(u_h, v_h) = (\mu f, v_h) \quad \forall v_h \in V_h. \tag{1}$$

In this talk, we are interested in estimating the error

$$e_h := u - u_h$$

A priori error estimates

A priori estimate

Assume that $\omega h / c_{\min} p \leq \mathscr{C}_1$. Then

$$\|\nabla e_h\|_{\mathbf{A},\Omega} \leq \mathscr{C}_2 \left(\frac{\omega h}{c_{\min} p} \right)^p.$$

F. Ihlenburg and I. Babuška, *SIAM J. Numer. Anal.*, 1997.

J.M. Melenk and S.A. Sauter, *Math. Comp.*, 2010.

T. Chaumont-Frelet and S. Nicaise, *IMA J. Numer. Anal.*, 2019.

Some limitations:

- The above result requires important regularity assumptions.
- The error estimate is not always applicable.
- The constant $\mathscr{C}_2$ is *not computable* in general.

A priori error estimates

A priori estimates provide qualitative upper bounds.

They are important as they show that the method converges.
They also indicate how fast the convergence happens.

They are not suited to *quantitatively* estimate the error in practice.

A posteriori estimates

(Ideal) a posteriori estimate

$$\|\nabla e_h\|_{\mathbf{A},\Omega} \leq \eta.$$

Here η is a fully-computable real number called an "error estimator".
This quantity is computed as a *local* post-processing of u_h, i.e. $\eta = \eta(u_h)$.
There are *no generic constants*. We have a *guaranteed* error estimate.

Outline

The low-frequency case

We first consider the low frequency limit where $\omega = 0$.

The problem then reads: find $u : \Omega \to \mathbb{C}$ such that

$$\begin{cases} -\nabla \cdot (\mathbf{A}\nabla u) = \mu f & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_{\mathrm{D}}, \\ \mathbf{A}\nabla u \cdot \boldsymbol{n} = 0 & \text{on } \Gamma_{\mathrm{A}}. \end{cases}$$

For the sake of simplicity, let $f = f_h \in \mathbb{P}_p(\mathcal{T}_h)$.

Outline

Flux

At the continuous level, the ideal flux $\boldsymbol{\sigma} := -\mathbf{A}\nabla u$ satisfies

$$\boldsymbol{\sigma} = \arg \min_{\substack{\boldsymbol{\tau} \in \boldsymbol{H}_{\Gamma_{\mathrm{A}}}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{\tau} = \mu f_h \text{ in } \Omega}} \|\mathbf{A}^{-1}\boldsymbol{\tau} + \nabla u\|_{\mathbf{A},\Omega}.$$

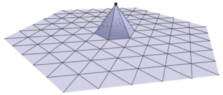
The best-possible computable flux directly mimics this definition at the discrete level

$$\boldsymbol{\sigma}_h := \arg \min_{\substack{\boldsymbol{\tau}_h \in \boldsymbol{W}_h \\ \nabla\cdot\boldsymbol{\tau}_h = \mu f_h \text{ in } \Omega}} \|\mathbf{A}^{-1}\boldsymbol{\tau}_h + \nabla u_h\|_{\mathbf{A},\Omega}.$$

Localization

Consider the set of "hat functions" $\{\psi^{\boldsymbol{a}}\}_{\boldsymbol{a}\in\mathcal{V}_h}$ of the mesh. We then have

$$\sum_{\boldsymbol{a}\in\mathcal{V}_h} \psi^{\boldsymbol{a}} = 1.$$



The ideal flux $\boldsymbol{\sigma} := -\mathbf{A}\nabla u$ can be decomposed as

$$\boldsymbol{\sigma} = \sum_{\boldsymbol{a}\in\mathcal{V}_h} \boldsymbol{\sigma}^{\boldsymbol{a}}, \qquad \boldsymbol{\sigma}^{\boldsymbol{a}} = -\psi^{\boldsymbol{a}}\mathbf{A}\nabla u.$$

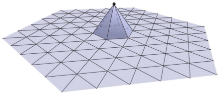
Easy computations show that

$$\boldsymbol{\sigma}^{\boldsymbol{a}} \cdot \boldsymbol{n} = 0 \text{ on } \partial\omega^{\boldsymbol{a}}, \qquad \nabla\cdot\boldsymbol{\sigma}^{\boldsymbol{a}} = \psi^{\boldsymbol{a}}\mu f_h - \mathbf{A}\nabla u\cdot\nabla\psi^{\boldsymbol{a}} \text{ in } \omega^{\boldsymbol{a}}.$$

Localization

Consider the set of "hat functions" $\{\psi^{\boldsymbol{a}}\}_{\boldsymbol{a}\in\mathcal{V}_h}$ of the mesh. We then have

$$\sum_{\boldsymbol{a}\in\mathcal{V}_h} \psi^{\boldsymbol{a}} = 1.$$



The ideal flux $\boldsymbol{\sigma} := -\mathbf{A}\boldsymbol{\nabla} u$ can be decomposed as

$$\boldsymbol{\sigma} = \sum_{\boldsymbol{a}\in\mathcal{V}_h} \boldsymbol{\sigma}^{\boldsymbol{a}}, \qquad \boldsymbol{\sigma}^{\boldsymbol{a}} = -\psi^{\boldsymbol{a}}\mathbf{A}\boldsymbol{\nabla} u.$$

Easy computations show that

$$\boldsymbol{\sigma}^{\boldsymbol{a}} \cdot \boldsymbol{n} = 0 \text{ on } \partial\omega^{\boldsymbol{a}}, \qquad \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}^{\boldsymbol{a}} = \psi^{\boldsymbol{a}}\mu f_h - \mathbf{A}\boldsymbol{\nabla} u \cdot \boldsymbol{\nabla}\psi^{\boldsymbol{a}} \text{ in } \omega^{\boldsymbol{a}}.$$

Localization

We have shown that

$$\boldsymbol{\sigma} = \sum_{\boldsymbol{a} \in \mathcal{V}_h} \boldsymbol{\sigma}^{\boldsymbol{a}}$$

and

$$\boldsymbol{\sigma}^{\boldsymbol{a}} = \arg \min_{\substack{\boldsymbol{\tau} \in \boldsymbol{H}_0(\mathrm{div}, \omega^{\boldsymbol{a}}) \\ \nabla \cdot \boldsymbol{\tau} = \psi^{\boldsymbol{a}} \mu f_h - \mathbf{A} \nabla u \cdot \nabla \psi^{\boldsymbol{a}} \text{ in } \omega^{\boldsymbol{a}}}} \|\mathbf{A}^{-1} \boldsymbol{\tau} + \psi^{\boldsymbol{a}} \nabla u\|_{\mathbf{A}, \omega^{\boldsymbol{a}}}.$$

Implicit characterization that we can mimic at the discrete level!

Localization

We have shown that

$$\boldsymbol{\sigma} = \sum_{\mathbf{a} \in \mathcal{V}_h} \boldsymbol{\sigma}^{\mathbf{a}}$$

and

$$\boldsymbol{\sigma}^{\mathbf{a}} = \arg \min_{\substack{\boldsymbol{\tau} \in \boldsymbol{H}_0(\mathrm{div}, \omega^{\mathbf{a}}) \\ \nabla\cdot\boldsymbol{\tau} = \psi^{\mathbf{a}} \mu f_h - \mathbf{A}\nabla u \cdot \nabla \psi^{\mathbf{a}} \text{ in } \omega^{\mathbf{a}}}} \|\mathbf{A}^{-1}\boldsymbol{\tau} + \psi^{\mathbf{a}} \nabla u\|_{\mathbf{A}, \omega^{\mathbf{a}}}.$$

Implicit characterization that we can mimic at the discrete level!

Localization

We thus set

$$\boldsymbol{\sigma}_h := \sum_{\boldsymbol{a}\in\mathcal{V}_h} \boldsymbol{\sigma}_h^{\boldsymbol{a}}$$

with

$$\boldsymbol{\sigma}_h^{\boldsymbol{a}} := \arg \min_{\substack{\boldsymbol{\tau}_h\in \boldsymbol{H}_0(\mathrm{div},\omega^{\boldsymbol{a}})\cap \boldsymbol{W}_h \\ \nabla\cdot\boldsymbol{\tau}_h=\psi^{\boldsymbol{a}}\mu f_h-\mathbf{A}\nabla u_h\cdot\nabla\psi^{\boldsymbol{a}} \text{ in } \omega^{\boldsymbol{a}}}} \|\mathbf{A}^{-1}\boldsymbol{\tau}_h + \psi^{\boldsymbol{a}}\nabla u_h\|_{\mathbf{A},\omega^{\boldsymbol{a}}}.$$

The compatibility condition

$$(\psi^{\boldsymbol{a}}\mu f_h - \mathbf{A}\nabla u_h\cdot\nabla\psi^{\boldsymbol{a}}, 1)_{\omega^{\boldsymbol{a}}} = (\mu f_h, \psi^{\boldsymbol{a}})_\Omega - (\mathbf{A}\nabla u_h, \nabla\psi^{\boldsymbol{a}})_\Omega = 0$$

holds true since u_h is the discrete solution and $\psi^{\boldsymbol{a}} \in V_h$.

Localization

We thus set

$$\boldsymbol{\sigma}_h := \sum_{\boldsymbol{a}\in\mathcal{V}_h} \boldsymbol{\sigma}_h^{\boldsymbol{a}}$$

with

$$\boldsymbol{\sigma}_h^{\boldsymbol{a}} := \arg \min_{\substack{\boldsymbol{\tau}_h \in \boldsymbol{H}_0(\mathrm{div},\omega^{\boldsymbol{a}})\cap \boldsymbol{W}_h \\ \nabla\cdot\boldsymbol{\tau}_h = \psi^{\boldsymbol{a}}\mu f_h - \mathbf{A}\nabla u_h\cdot\nabla\psi^{\boldsymbol{a}} \text{ in } \omega^{\boldsymbol{a}}}} \|\mathbf{A}^{-1}\boldsymbol{\tau}_h + \psi^{\boldsymbol{a}}\nabla u_h\|_{\mathbf{A},\omega^{\boldsymbol{a}}}.$$

The compatibility condition

$$(\psi^{\boldsymbol{a}}\mu f_h - \mathbf{A}\nabla u_h \cdot \nabla\psi^{\boldsymbol{a}}, 1)_{\omega^{\boldsymbol{a}}} = (\mu f_h, \psi^{\boldsymbol{a}})_\Omega - (\mathbf{A}\nabla u_h, \nabla\psi^{\boldsymbol{a}})_\Omega = 0$$

holds true since u_h is the discrete solution and $\psi^{\boldsymbol{a}} \in V_h$.

Summary of the localization process

Step 1: solve a set of small, uncoupled linear systems

$$\boldsymbol{\sigma}_h^{\boldsymbol{a}} := \arg \min_{\substack{\boldsymbol{\tau}_h \in \boldsymbol{H}_0(\mathrm{div},\omega^{\boldsymbol{a}}) \cap \boldsymbol{W}_h \\ \nabla \cdot \boldsymbol{\tau}_h = \psi^{\boldsymbol{a}} \mu f_h - \mathbf{A} \nabla u_h \cdot \nabla \psi^{\boldsymbol{a}} \text{ in } \omega^{\boldsymbol{a}}}} \|\mathbf{A}^{-1} \boldsymbol{\tau}_h + \psi^{\boldsymbol{a}} \nabla u_h\|_{\mathbf{A},\omega^{\boldsymbol{a}}}.$$

Step 2: assemble these local contributions

$$\boldsymbol{\sigma}_h := \sum_{\boldsymbol{a} \in \mathcal{V}_h} \boldsymbol{\sigma}_h^{\boldsymbol{a}}.$$

Step 3: compute the estimator

$$\eta := \|\mathbf{A}^{-1} \boldsymbol{\sigma}_h + \nabla u_h\|_{\mathbf{A},\Omega}.$$

Step 4: enjoy the guaranteed estimate by the Prager–Synge (in)equality

$$\|\nabla e_h\|_{\mathbf{A},\Omega} \leq \eta.$$

Outline

Efficiency

We can show that

$$\eta \leq C_{\mathrm{eff}} \|\nabla e_h\|_{\mathbf{A},\Omega},$$

where C_{eff} only depends on:

- the "flatness" of the tetrahedra in the mesh,
- the "contrasts" in the coefficients.

An important aspect is that C_{eff} does not depend on the polynomial degree p.
Local efficiency can be shown as well.

P. Braess, V. Pillwein, and J. Schöberl, *CMAME*, 2009.

A. Ern and M. Vohralík, *SINUM*, 2015 & *Math. Comp.*, 2020.

Outline

The high-frequency case

Back to our original problem

$$\begin{cases} -\omega^2 \mu u - \nabla \cdot (\mathbf{A} \nabla u) = \mu f & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_{\mathrm{D}}, \\ \mathbf{A} \nabla u \cdot \boldsymbol{n} - i\omega\gamma u = 0 & \text{on } \Gamma_{\mathrm{A}}. \end{cases}$$

Recall

$$b(u, v) = -\omega^2 (\mu u, v)_\Omega - i\omega(\gamma u, v)_{\Gamma_{\mathrm{A}}} + (\mathbf{A} \nabla u, \nabla v)_\Omega.$$

Energy norm and lack of coercivity

We will consider the "balanced" norm

$$|||v|||^2_{\omega,\Omega} := \omega^2 \|v\|^2_{\mu,\Omega} + \|\nabla v\|^2_{\mathbf{A},\Omega}.$$

The sesquilinear form b is *not coercive*.

Instead we have the "Gårding" inequality

$$\operatorname{Re} b(v,v) = \|\nabla v\|^2_{\mathbf{A},\Omega} - \omega^2 \|v\|^2_{\mu,\Omega} = |||v|||^2_{\omega,\Omega} - 2\omega^2 \|v\|^2_{\mu,\Omega}.$$

Outline

1. Introduction
2. Low frequencies
   - Guaranteed upper bound by local flux equilibration and the Prager–Synge equality
   - Efficiency
3. High frequencies
   - Flux equilibration
   - A coarse error estimate
   - Efficiency
   - Sharp error estimate
4. Controlling the pre-factors
   - The stability constant $\mathscr{C}_{\mathrm{st}}$
   - The approximation factor $\mathscr{C}_{\mathrm{ap}}$
5. Numerical illustrations
6. Conclusions and outlook

Definition of an equilibrated flux

Letting $\boldsymbol{\sigma} := -\mathbf{A}\nabla u$, we have

$$\boldsymbol{\sigma} \cdot \boldsymbol{n} = -i\omega\gamma u \text{ on } \Gamma_{\mathrm{A}} \qquad \nabla \cdot \boldsymbol{\sigma} = \mu f_h + \omega^2 \mu u \text{ in } \Omega.$$

Hence, natural requirements for $\boldsymbol{\sigma}_h$ are

$$\boldsymbol{\sigma}_h \cdot \boldsymbol{n} = -i\omega\gamma u_h \text{ on } \Gamma_{\mathrm{A}} \qquad \nabla \cdot \boldsymbol{\sigma}_h = \mu f_h + \omega^2 \mu u_h \text{ in } \Omega.$$

The "low frequency" reconstruction algorithm directly extends.

Prager–Synge inequality

Let $v \in H^1_{\Gamma_{\mathrm{D}}}(\Omega)$. We have

$$\begin{aligned} b(e_h, v) &= (\mu f_h, v)_\Omega - b(u_h, v) \\ &= (\mu f_h + \omega^2 \mu u_h, v)_\Omega + i\omega(\gamma u_h, v)_{\Gamma_{\mathrm{A}}} - (\mathbf{A}\nabla u_h, \nabla v)_\Omega \\ &= (\nabla \cdot \boldsymbol{\sigma}_h, v)_\Omega - (\boldsymbol{\sigma}_h \cdot \boldsymbol{n}, v)_{\Gamma_{\mathrm{A}}} - (\mathbf{A}\nabla u_h, \nabla v)_\Omega \\ &= -(\boldsymbol{\sigma}_h + \mathbf{A}\nabla u_h, \nabla v)_\Omega. \end{aligned}$$

Prager–Synge inequality

$$|b(e_h, v)| \leq \eta \|\nabla v\|_{\mathbf{A},\Omega} \quad \forall v \in H^1_{\Gamma_{\mathrm{D}}}(\Omega)$$

So far, so good!

What's the matter?

Here, we do not have

$$\|\nabla e_h\|_{\mathbf{A},\Omega}^2 \leq |b(e_h, e_h)|,$$

which is a major issue!

Instead, we only have the "Gårding" inequality

$$\operatorname{Re} b(e_h, e_h) \geq \|e_h\|_{\omega,\Omega}^2 - 2\omega^2 \|e_h\|_{\mu,\Omega}^2.$$

Outline

Stability constant

For $g \in L^2(\Omega)$, let $\mathscr{S}^\star g$ denote the unique element of $H^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$b(w, \mathscr{S}^\star g) = 2\omega^2(\mu w, g)_\Omega \quad \forall w \in H^1_{\Gamma_{\mathrm{D}}}(\Omega)$$

and let

$$\mathscr{C}_{\mathrm{st}} := \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \|\nabla(\mathscr{S}^\star g)\|_{\mathbf{A},\Omega}.$$

$\mathscr{C}_{\mathrm{st}}$ is the best constant such that

$$\|\nabla(\mathscr{S}^\star g)\|_{\mathbf{A},\Omega} \leq \mathscr{C}_{\mathrm{st}}\omega\|g\|_{\mu,\Omega} \quad \forall g \in L^2(\Omega).$$

It is closely related to resolvant estimates.

Making up for the lack of coercivity

By definition, we have

$$b(w, \mathscr{S}^\star e_h) = 2\omega^2(\mu w, e_h) \quad \forall w \in H^1_{\Gamma_{\mathrm{D}}}(\Omega).$$

Hence, in particular,

$$b(e_h, \mathscr{S}^\star e_h) = 2\omega^2 \|e_h\|^2_{\mu,\Omega},$$

which is exactly the “bad” term the Gårding inequality:

$$\operatorname{Re} b(e_h, e_h) = \|\!|e_h|\!\|^2_{\omega,\Omega} - 2\omega^2 \|e_h\|^2_{\mu,\Omega}.$$

Making up for the lack of coercivity

Using Prager–Synge inequality, we have

$$|||e_h|||_{\omega,\Omega}^2 = \operatorname{Re} b(e_h, e_h + \mathscr{S}^\star e_h) \leq \eta \|\nabla(e_h + \mathscr{S}^\star e_h)\|_{\mathbf{A},\Omega}.$$

It follows that

$$\begin{aligned} |||e_h|||_{\omega,\Omega}^2 &\leq \eta \left(\|\nabla e_h\|_{\mathbf{A},\Omega} + \|\nabla(\mathscr{S}^\star e_h)\|_{\mathbf{A},\Omega}\right) \\ &\leq \eta \left(\|\nabla e_h\|_{\mathbf{A},\Omega} + \mathscr{C}_{\mathrm{st}}\omega\|e_h\|_{\mu,\Omega}\right) \\ &\leq \eta \max(1, \mathscr{C}_{\mathrm{st}})\, |||e_h|||_{\omega,\Omega}\,, \end{aligned}$$

and

$$|||e_h|||_{\omega,\Omega} \leq \max(1, \mathscr{C}_{\mathrm{st}})\eta.$$

Coarse error estimate

We obtained the following error estimate:

Coarse error estimate

$$|||e_h||| \leq \max(1, \mathscr{C}_{\mathrm{st}})\eta$$

$\mathscr{C}_{\mathrm{st}}$ is the best constant such that:

Stability constant

$$\|\nabla(\mathscr{S}^\star g)\|_{\mathbf{A},\Omega} \leq \mathscr{C}_{\mathrm{st}}\omega\|g\|_{\mu,\Omega} \qquad \forall g \in L^2(\Omega).$$

Outline

1. Introduction
2. Low frequencies
   - Guaranteed upper bound by local flux equilibration and the Prager–Synge equality
   - Efficiency
3. High frequencies
   - Flux equilibration
   - A coarse error estimate
   - Efficiency
   - Sharp error estimate
4. Controlling the pre-factors
   - The stability constant $\mathscr{C}_{\rm st}$
   - The approximation factor $\mathscr{C}_{\rm ap}$
5. Numerical illustrations
6. Conclusions and outlook

Efficiency

We can show that

$$\eta \leq C_{\text{eff}} \left(1 + \max_{K \in \mathcal{T}_h} \frac{\omega h_K}{c_{\min K}\, p}\right) \|\!| e_h |\!\|_{\omega,\Omega} ,$$

where $c_{\min K}$ is the wavespeed in the element K.

For any reasonable discretization, we have

$$\frac{\omega h_K}{c_{\min K}\, p} \leq 1,$$

so that in practice

$$\eta \leq C_{\text{eff}} \|\!| e_h |\!\|_{\omega,\Omega} ,$$

where C_{eff} only depends on the elements "flatness" and the contrasts.

W. Dörfler, S. Sauter, *Comput. Meth. Appl. Math.*, 2013.

T. Chaumont-Frelet, A. Ern, and M. Vohralík, *Numer. Math.*, 2021.

The problem with the coarse error estimate

Recall that

$$\eta \leq C_{\text{eff}} \, |||e_h|||_{\omega,\Omega} \qquad |||e_h|||_{\omega,\Omega} \leq \max(1, \mathscr{C}_{\text{st}})\eta.$$

We have $C_{\text{eff}} \simeq 1$ and $\mathscr{C}_{\text{st}} \gtrsim \omega\ell_\Omega / c_{\min}$, so that

$$\eta \lesssim |||e_h|||_{\omega,\Omega} \lesssim \frac{\omega\ell_\Omega}{c_{\min}}\eta.$$

In practice, the coarse error estimate will largely overestimate the error in the high-frequency regime.

Outline

The approximation factor

We introduce

$$\mathscr{C}_{\rm ap} := \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \min_{v_h \in V_h} \|\nabla(\mathscr{S}^\star g - v_h)\|_{\mathbf{A},\Omega}.$$

Approximability

For all $g \in L^2(\Omega)$, there exists $v_h^\star \in V_h$ such that

$$\|\nabla(\mathscr{S}^\star g - v_h^\star)\|_{\mathbf{A},\Omega} \leq \mathscr{C}_{\rm ap}\omega\|g\|_{\mu,\Omega}.$$

Using Galerkin orthogonality

Recall that

$$\|e_h\|_{\omega,\Omega}^2 = \operatorname{Re} b(e_h, e_h + \mathscr{S}^\star e_h).$$

By Galerkin orthogonality, we have

$$\begin{aligned}
\|e_h\|_{\omega,\Omega}^2 &= \operatorname{Re} b(e_h, e_h) + \operatorname{Re} b(e_h, \mathscr{S}^\star e_h) \\
&= \operatorname{Re} b(e_h, e_h) + \operatorname{Re} b(e_h, \mathscr{S}^\star e_h - v_h^\star) \\
&\leq \eta \left(\|\boldsymbol{\nabla} e_h\|_{\mathbf{A},\Omega} + \|\boldsymbol{\nabla}(\mathscr{S}^\star e_h - v_h^\star)\|_{\mathbf{A},\Omega} \right) \\
&\leq \eta \max(1, \mathscr{C}_{\mathrm{ap}}) \|e_h\|_{\omega,\Omega}.
\end{aligned}$$

Sharp error estimate

Sharp error estimate

$$|||e_h|||_{\omega,\Omega} \leq \max(1, \mathscr{C}_{\mathrm{ap}})\eta.$$

Approximation factor

$$\mathscr{C}_{\mathrm{ap}} := \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \min_{v_h \in V_h} \|\nabla(\mathscr{S}^\star g - v_h)\|_{\mathbf{A},\Omega} \to 0.$$

W. Dörfler, S. Sauter, *Comput. Meth. Appl. Math.*, 2013.

T. Chaumont-Frelet, A. Ern, and M. Vohralík, *Numer. Math.*, 2021.

Takeaways

The "equilibration" technology is the same as for low frequencies.

Coarse error estimate

$$\|e_h\|_{\omega,\Omega} \leq \max(1, \mathscr{C}_{\rm st})\eta \quad \mathscr{C}_{\rm st} \gtrsim \omega\ell_\Omega / c_{\min}$$

Sharp error estimate

$$\|e_h\|_{\omega,\Omega} \leq \max(1, \mathscr{C}_{\rm ap})\eta \quad \mathscr{C}_{\rm ap} \to 0$$

Efficiency

$$\eta \leq C_{\rm eff}\left(1 + \frac{\omega h}{c_{\min} p}\right)\|e_h\|_{\omega,\Omega}$$

Outline

Outline

The stability constant

The stability constant is defined by

$$\mathscr{C}_{\rm st} := \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \|\nabla(\mathscr{S}^\star g)\|_{\mathbf{A},\Omega}.$$

It is only related to the PDE and independent of the numerical scheme.

Qualitative behaviour

It is known that we have "at least":

$$\mathscr{C}_{\mathrm{st}} \gtrsim \frac{\omega \ell_\Omega}{c_{\min}}.$$

For non-trapping settings (the "easier" scenario), we have

$$\mathscr{C}_{\mathrm{st}} \lesssim \frac{\omega \ell_\Omega}{c_{\min}}.$$

If strong traping happens, "extreme" behaviors can occur

$$\mathscr{C}_{\mathrm{st}} \gtrsim \exp\left(\alpha \frac{\omega \ell_\Omega}{c_{\min}}\right)$$

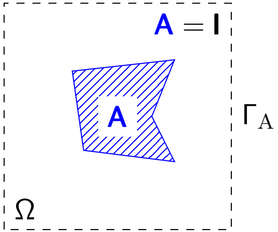
for "some" frequencies. For "most frequencies"

$$\mathscr{C}_{\mathrm{st}} \gtrsim \left(\frac{\omega \ell_\Omega}{c_{\min}}\right)^\beta.$$

D. Lafontaine, E.A. Spence, and J. Wunsch, *Comm. Pure Appl. Math.* 2021

Quantitative estimate for star-shaped non-trapping obstacles

$\Omega := (-\ell/2, \ell/2)^3$ is a cube centered at the origin.
$D \subset \Omega$ is star shaped with respect to the origin.



Assume that $\gamma = 1$ and that
$\mu = 1$ and $\mathbf{A} = \boldsymbol{I}$ in $\Omega \setminus D$.

Assume that $\mu = \mu_D \geq 1$ and $\mathbf{A} = \mathbf{A}_D \preceq \boldsymbol{I}$ in D.

This describes an obstacle made of a material with a "slow" wavespeed.

Quantitative estimate for star-shaped non-trapping obstacles

Guaranteed upper bound

$$\mathscr{C}_{\rm st} \leq 6 + \frac{3+\sqrt{3}}{\sqrt{3}} \frac{\omega \ell_\Omega}{c_{\min}}$$

The proof relies on a "Morawetz multiplier":
multiply the PDE by $\boldsymbol{x} \cdot \boldsymbol{\nabla} u$ and integrate by parts until it works!

- C.S. Morawetz, *Comm. Pure Appl. Math.*, 1962.
- J.M. Melenk, *PhD thesis*, 1995.
- H. Barucq, T. Chaumont-Frelet, and C. Gout, *Math. Comp.*, 2016.
- T. Chaumont-Frelet, A. Ern, and M. Vohralík, *Numer. Math.* 2021.

Outline

The approximation factor

The approximation factor is defined by

$$\mathscr{C}_{\mathrm{ap}} := \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \min_{v_h \in V_h} \|\nabla(\mathscr{S}^\star g - v_h)\|_{\mathbf{A},\Omega}.$$

It depends on both the PDE and the *approximation space* V_h.

Assuming that $\mathbf{A} = \boldsymbol{I}$, Ω is convex, and $\mathscr{C}_{\mathrm{st}}$ is known, we can control it.

Idea one: explicit interpolation error

R. Arcangeli and J.L. Gout, RAIRO Numer. Anal., 1976.

If $v \in H^2(\Omega)$, let $I_h^1 v \in V_h$ denotes its first-order Lagrange interpolant:

$$\|\nabla(v - I_h^1 v)\|_{\mathbf{A},\Omega} \leq \mathscr{C}_{\mathcal{T},\mathrm{i}} h \|\nabla^2 v\|_\Omega,$$

with a constant $\mathscr{C}_{\mathcal{T},\mathrm{i}}$ that is easily computable.

We then have

$$\mathscr{C}_{\rm ap} := \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \min_{v_h^\star \in V_h} \|\nabla(\mathscr{S}^\star g - v_h^\star)\|_{\mathbf{A},\Omega}$$

$$\leq \frac{1}{\omega} \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \|\nabla(\mathscr{S}^\star g - I_h^1(\mathscr{S}^\star g)\|_{\mathbf{A},\Omega}$$

$$\leq \frac{1}{\omega} \mathscr{C}_{\mathcal{T},\mathrm{i}} h \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \|\nabla^2(\mathscr{S}^\star g)\|_\Omega.$$

Idea two: estimation of the Hessian norm

P. Grisvard, 1985.

T. Chaumont-Frelet, S. Nicaise, and J. Tomezyk, *Comm. Pure Appl. Anal.*, 2020.

Because Ω is convex and $\gamma = 1$, we have

$$\|\nabla^2(\mathscr{S}^\star g)\|_\Omega \leq \|\Delta(\mathscr{S}^\star g)\|_\Omega.$$

Then, we use the facts that

$$-\Delta(\mathscr{S}^\star g) = 2\mu\omega^2 g + \mu\omega^2 \mathscr{S}^\star g$$

and

$$\omega\|\mathscr{S}^\star g\|_{\mu,\Omega} \leq 2\mathscr{C}_{\mathrm{st}}\|g\|_{\mu,\Omega}$$

to show that

$$\|\nabla^2(\mathscr{S}^\star g)\|_\Omega \leq 2\frac{\omega}{c_{\min}}\left(1 + \mathscr{C}_{\mathrm{st}}\right)\|g\|_{\mu,\Omega}.$$

Explicit control of the approximation factor

Recall that

$$\mathscr{C}_{\rm ap} \leq \frac{1}{\omega} \mathscr{C}_{\mathcal{T},\rm i} h \max_{\substack{g \in L^2(\Omega) \\ \|g\|_{\mu,\Omega}=1}} \|\boldsymbol{\nabla}^2 (\mathscr{S}^\star g)\|_\Omega$$

and

$$\|\boldsymbol{\nabla}^2 (\mathscr{S}^\star g)\|_\Omega \leq 2 \frac{\omega}{c_{\min}} \left(1 + \mathscr{C}_{\rm st}\right) \|g\|_{\mu,\Omega} \quad \forall g \in L^2(\Omega).$$

Guaranteed bound

$$\mathscr{C}_{\rm ap} \leq 2 \left(1 + \mathscr{C}_{\mathcal{T},\rm i}\right) \frac{\omega h}{c_{\min}} \mathscr{C}_{\rm st}$$

Takeaways

The estimator η needs to be "pre-factored" by $\mathscr{C}_{\mathrm{st}}$ or $\mathscr{C}_{\mathrm{ap}}$.
The "qualitative" behaviors of both quantities are relatively well known.

The behaviour of $\mathscr{C}_{\mathrm{st}}$ is only dictated by the PDE.
Explicit bounds are available for non-trapping star-shaped obstables.

The approximation factor $\mathscr{C}_{\mathrm{ap}}$ depends on the PDE and V_h.
When $\boldsymbol{A} = \boldsymbol{I}$, Ω is convex and $\mathscr{C}_{\mathrm{st}}$ is known, we can bound it nicely.

Outline

Propagation of a plane wave

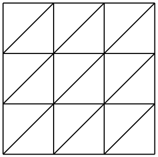
We consider the propagation of a plane wave in $\Omega = (-1,1)^2$

$$\begin{cases} -\omega^2 u - \Delta u &= 0 \quad \text{in } \Omega, \\ \nabla u \cdot \boldsymbol{n} - i\omega u &= g \quad \text{on } \Gamma_{\mathrm{A}}, \end{cases}$$

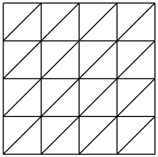
where

$$g := \nabla \xi_\theta \cdot \boldsymbol{n} - i\omega \xi_\theta, \quad \xi_\theta := e^{i\omega \boldsymbol{d}\cdot\boldsymbol{x}},$$

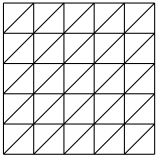
with $\boldsymbol{d} := (\cos\theta, \sin\theta)$ and $\theta = \pi/12$. The solution is $u = \xi_\theta$.



$h = \sqrt{2} \times 2/3$

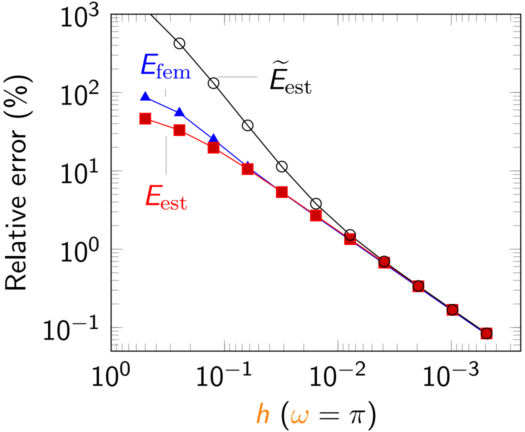


$h = \sqrt{2} \times 1/2$



$h = \sqrt{2} \times 2/5$

Plane wave experiment $p = 1$ and $\omega = \pi$

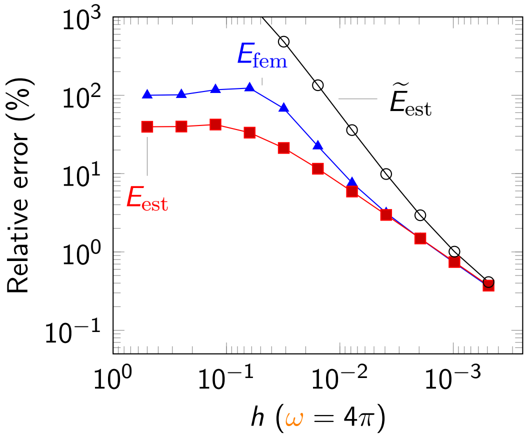


$$E_{\rm fem} := \|e_h\|_{\omega,\Omega}$$

$$E_{\rm est} := \eta$$

$$\widetilde{E}_{\rm est} := (1 + \mathscr{C}_{\rm ap})\eta$$

Plane wave experiment $p = 1$ and $\omega = 4\pi$

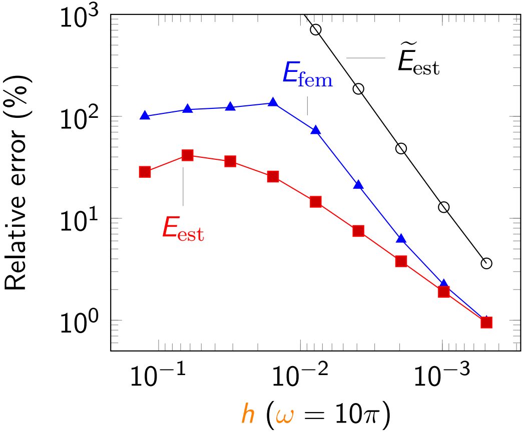


$$E_{\mathrm{fem}} := \| e_h \|_{\omega,\Omega}$$

$$E_{\mathrm{est}} := \eta$$

$$\widetilde{E}_{\mathrm{est}} := (1 + \mathscr{C}_{\mathrm{ap}})\eta$$

Plane wave experiment $p = 1$ and $\omega = 10\pi$

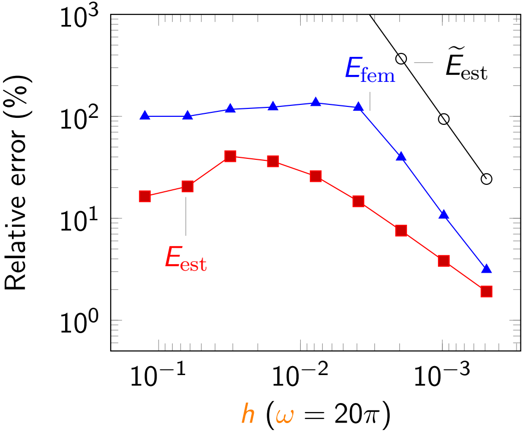


$$E_{\mathrm{fem}} := \|e_h\|_{\omega,\Omega}$$

$$E_{\mathrm{est}} := \eta$$

$$\widetilde{E}_{\mathrm{est}} := (1 + \mathscr{C}_{\mathrm{ap}})\eta$$

Plane wave experiment $p = 1$ and $\omega = 20\pi$

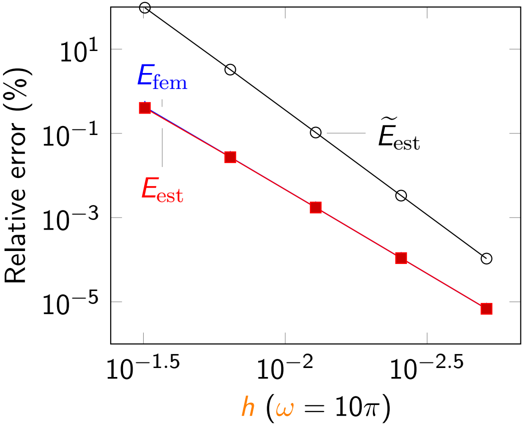


$$E_{\mathrm{fem}} := \| e_h \|_{\omega,\Omega}$$

$$E_{\mathrm{est}} := \eta$$

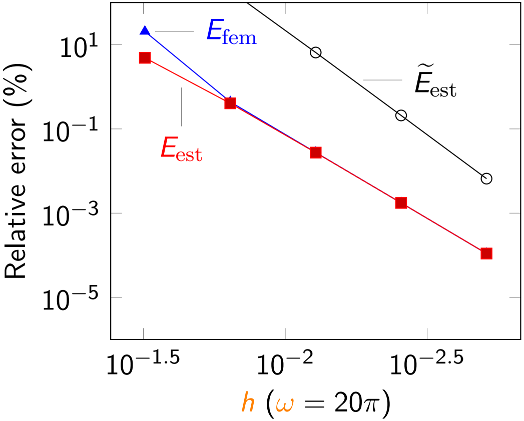
$$\widetilde{E}_{\mathrm{est}} := (1 + \mathscr{C}_{\mathrm{ap}})\eta$$

Plane wave experiment $p = 4$ and $\omega = 10\pi$



$$E_{\mathrm{fem}} := \|e_h\|_{\omega,\Omega}$$
$$E_{\mathrm{est}} := \eta$$
$$\widetilde{E}_{\mathrm{est}} := (1 + \mathscr{C}_{\mathrm{ap}})\eta$$

Plane wave experiment $p = 4$ and $\omega = 20\pi$

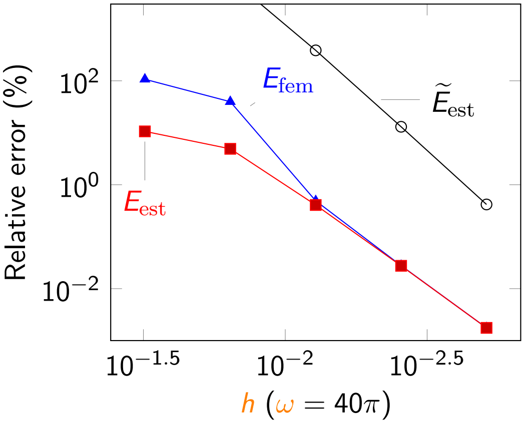


$$E_{\rm fem} := \|e_h\|_{\omega,\Omega}$$

$$E_{\rm est} := \eta$$

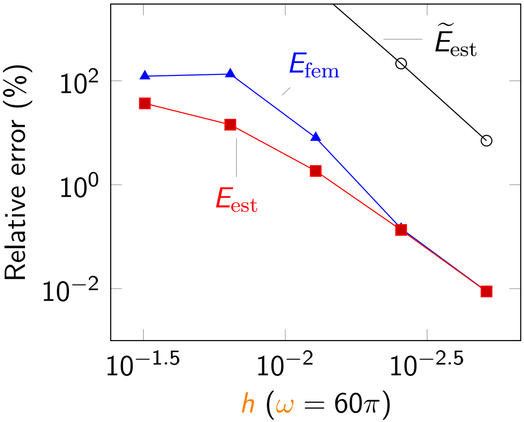
$$\widetilde{E}_{\rm est} := (1 + \mathscr{C}_{\rm ap})\eta$$

Plane wave experiment $p = 4$ and $\omega = 40\pi$



$$E_{\rm fem} := \|e_h\|_{\omega,\Omega}$$
$$E_{\rm est} := \eta$$
$$\widetilde{E}_{\rm est} := (1 + \mathscr{C}_{\rm ap})\eta$$

Plane wave experiment $p = 4$ and $\omega = 60\pi$



$$E_{\rm fem} := |||e_h|||_{\omega,\Omega}$$

$$E_{\rm est} := \eta$$

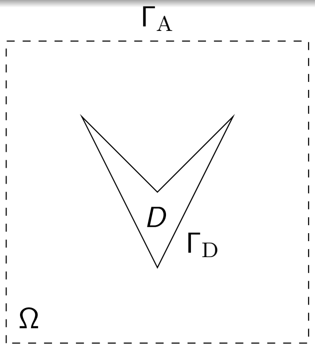
$$\widetilde{E}_{\rm est} := (1 + \mathscr{C}_{\rm ap})\eta$$

Scattering by an non-trapping obstacle

We now consider a scattering problem

$$\begin{cases} -\omega^2 u - \Delta u = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_{\mathrm{D}}, \\ \nabla u \cdot \boldsymbol{n} - i\omega u = g & \text{on } \Gamma_{\mathrm{A}}, \end{cases}$$

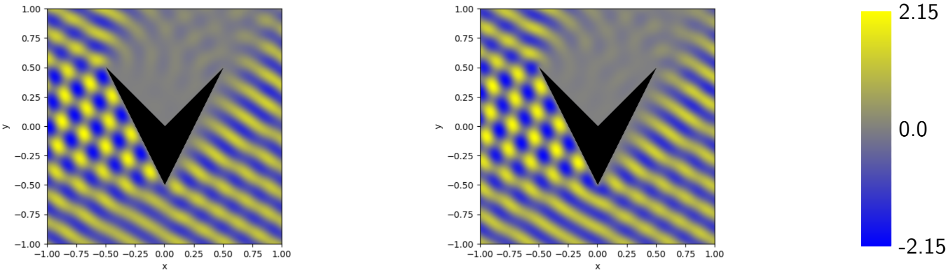
where again $g = \nabla \xi_\theta \cdot \boldsymbol{n} - i\omega\xi_\theta$.



We fix the wavenumber $\omega = 10\pi$ and employ $\mathbb{P}_3$ elements.

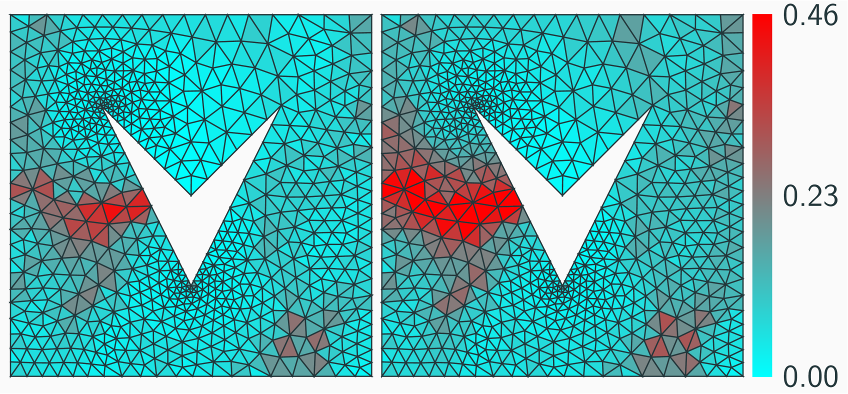
We consider a sequence of meshes that are adaptively refined using η_K.

Solution of the scattering problem



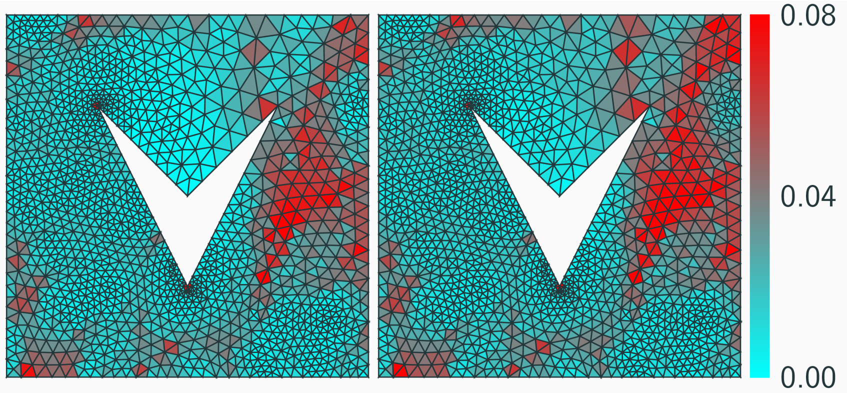
Real (left) and imaginary (right) parts of the solution

Estimated error in mesh #1



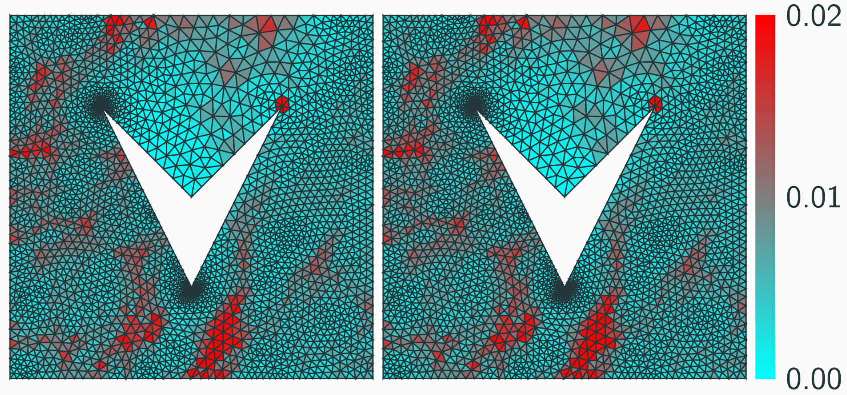
Estimator η_K (left) and elementwise error $\|e_h\|_{\omega,K}$ (right)

Estimated error in mesh #2



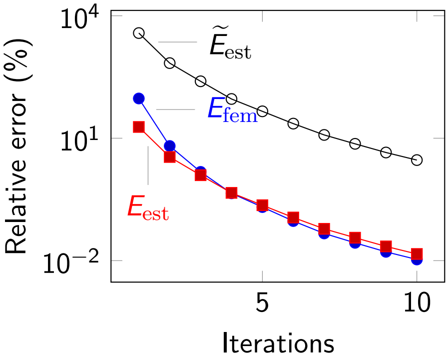
Estimator η_K (left) and elementwise error $\|e_h\|_{\omega,K}$ (right)

Estimated error in mesh #3



Estimator η_K (left) and elementwise error $\|e_h\|_{\omega,K}$ (right)

Behavior of the estimator through the adaptive procedure



$$E_{\mathrm{fem}} := \|e_h\|_{\omega,\Omega}$$

$$E_{\mathrm{est}} := \eta$$

$$\widetilde{E}_{\mathrm{est}} := (1 + \mathscr{C}_{\mathrm{ap}})\eta$$

Behaviors of the estimated and analytical errors in the adaptive procedure

Outline

Conclusions and outlook

We construct an a posteriori error estimator η *via* flux equilibration.
It directly provides guaranteed error estimates at low frequencies.

For high frequencies, η has to be pre-factored, either by $\mathscr{C}_{\rm st}$ or by $\mathscr{C}_{\rm ap}$.
The estimates are asymptotically constant-free.
In specific situations, we can provide guaranteed bounds on $\mathscr{C}_{\rm st}$ and $\mathscr{C}_{\rm ap}$.

There is still a long way toward fully reliable error estimation for high-frequency problems!

T. Chaumont-Frelet, A. Ern, and M. Vohralík, *Numer. Math.*, 2021.

Thank you for your attention!

Conclusions and outlook

We construct an a posteriori error estimator η *via* flux equilibration.
It directly provides guaranteed error estimates at low frequencies.

For high frequencies, η has to be pre-factored, either by $\mathscr{C}_{\rm st}$ or by $\mathscr{C}_{\rm ap}$.
The estimates are asymptotically constant-free.
In specific situations, we can provide guaranteed bounds on $\mathscr{C}_{\rm st}$ and $\mathscr{C}_{\rm ap}$.

There is still a long way toward fully reliable error estimation for high-frequency problems!

T. Chaumont-Frelet, A. Ern, and M. Vohralík, *Numer. Math.*, 2021.

Thank you for your attention!

Conclusions and outlook

We construct an a posteriori error estimator η *via* flux equilibration.
It directly provides guaranteed error estimates at low frequencies.

For high frequencies, η has to be pre-factored, either by $\mathscr{C}_{\mathrm{st}}$ or by $\mathscr{C}_{\mathrm{ap}}$.
The estimates are asymptotically constant-free.
In specific situations, we can provide guaranteed bounds on $\mathscr{C}_{\mathrm{st}}$ and $\mathscr{C}_{\mathrm{ap}}$.

There is still a long way toward fully reliable error estimation for high-frequency problems!

T. Chaumont-Frelet, A. Ern, and M. Vohralík, *Numer. Math.*, 2021.

Thank you for your attention!