Guaranteed error estimates for finite element discretizations of Helmholtz problems

T. Chaumont-Frelet*,†, A. Ern‡,§, M. Vohralík§,‡
June 17, 2020

*Inria Sophia-Antipolis, †Laboratoire J.A. Dieudonné, ‡CERMICS, §Inria Paris
Motivations
Motivations

Model problem
Model problem

Given \( f : \Omega \rightarrow \mathbb{C} \), find \( u : \Omega \rightarrow \mathbb{C} \) such that

\[
\begin{cases}
-\omega^2 \mu u - \nabla \cdot (A \nabla u) = \mu f & \text{in } \Omega, \\
\mu \gamma & \text{on } \Gamma_D, \\
A \nabla u \cdot n - i\omega \gamma u = 0 & \text{on } \Gamma_A
\end{cases}
\]

where \( \mu, A \) and \( \gamma \) are given coefficients strictly positive coefficients.
What do I mean by high-frequency?

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

$$c := \sqrt{\frac{\sigma_{\text{min}}(A)}{\mu}}$$

The (minimal) wavelength is then given by:

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

The “important” quantity is $N_\lambda := \frac{\ell_\Omega}{\lambda}$. High-frequency means that

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

is “large” (a few tens or hundreds).
Recall the Helmholtz problem in strong form

\[
\begin{cases}
-\omega^2 \mu u - \nabla \cdot (A \nabla u) = \mu f \quad \text{in } \Omega, \\
u = 0 \quad \text{on } \Gamma_D, \\
A \nabla u \cdot n - i\omega \gamma u = 0 \quad \text{on } \Gamma_A.
\end{cases}
\]

Assuming \( f \in L^2(\Omega) \), we seek \( u \in H^1_D(\Omega) \) such that

\[
b(u, v) = (\mu f, v) \quad \forall v \in H^1_D(\Omega),
\]

with

\[
b(u, v) := -\omega^2 (\mu u, v)_\Omega - i\omega (\gamma u, v)_{\Gamma_A} + (A \nabla u, \nabla v)_\Omega.
\]
We consider a mesh $\mathcal{T}_h$ of $\Omega$ into tetrahedral elements $K$.

The elements $K \in \mathcal{T}_h$ are “small” ($h_K := \text{diam } K \leq h$).

The coefficients $\mu, \gamma, A$ are constant inside each element/face.

We introduce a “finite element” discretization space

$$V_h := \{ v_h \in H^1_D(\Omega) \mid v_h|_K \in P_p(K) \ \forall K \in \mathcal{T}_h \}$$

with $p \geq 1$.

What do I mean by refined mesh?

When I am saying that “the mesh is fine”, I mean that

\[ \frac{N_{\text{dofs}}}{\lambda} = \frac{\lambda}{\frac{h}{p}} \approx \left( \frac{\omega h}{c_{\text{min}} p} \right)^{-1} \]

is large.
Recall that $u$ is the only element of $H^1_D(\Omega)$ such that

$$b(u, v) = (\mu f, v) \quad \forall v \in H^1_D(\Omega).$$

Analogously, we seek a discrete solution $u_h \in V_h$ such that

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

Problem (1) corresponds to a finite dimensional linear system, that we can numerically solve.

In this talk, we are especially interested in measuring the error

$$e_h := u - u_h$$
Motivations

A priori error estimates
A priori error estimates

A priori estimates
Assume that $\omega h / c_{\text{min}} p \leq C_1$, then

$$\| \nabla e_h \|_{A, \Omega} \leq C_2 \left( \frac{\omega h}{c_{\text{min}} p} \right)^p.$$ 


Some limitations:

- The above result requires important regularity assumptions.
- The error estimate is not always applicable.
- The constants $C_1$ and $C_2$ are 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.
Motivations

A posteriori error estimation
Here $\eta$ is a fully-computable real number called an “error estimator”. This quantity is computed as a post-processing of $u_h$, i.e. $\eta = \eta(u_h)$. There is no generic constants. We have a guaranteed error estimate.
1. Low frequencies
2. High frequencies
3. Controlling the pre-factors
4. Numerical illustrations
Low frequencies
We first consider the low frequency limit where $\omega = 0$.

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

$$
\begin{cases}
-\nabla \cdot (A \nabla u) = \mu f & \text{in } \Omega, \\
u = 0 & \text{on } \Gamma_D, \\
A \nabla u \cdot n = 0 & \text{on } \Gamma_A.
\end{cases}
$$

For the sake of simplicity, we will assume that $f = f_h \in \mathbb{P}_p(\mathcal{T}_h)$. 
Consider the sesquilinear form
\[
a(u, v) = (A \nabla u, \nabla v)_\Omega, \quad u, v \in H^1_{\Gamma_D}(\Omega).
\]

We can characterize \( u \) as the unique element of \( H^1_{\Gamma_D}(\Omega) \) such that
\[
a(u, v) = (\mu f_h, v)_\Omega
\]
for all \( v \in H^1_{\Gamma_D}(\Omega) \).

\( u_h \) is the unique element of \( V_h \) satisfying
\[
a(u_h, v_h) = (\mu f_h, v_h)_\Omega
\]
for all \( v_h \in V_h \).
Low frequencies

Key ideas behind a posteriori estimation
Second-order problems typically arise from two physical laws, e.g., “Faraday’s law + Gauss’ law = Poisson problem”.

The continuous solution $u$ is uniquely determined by the condition that

$$ u \in H^1(\Omega), \quad u = 0 \text{ on } \Gamma_D $$

and for the “flux” $\sigma := -A \nabla u \in H(\text{div}, \Omega)$

$$ \sigma \cdot n = 0 \text{ on } \Gamma_A, \quad \nabla \cdot \sigma = \mu f_h \text{ in } \Omega. $$

The discrete solution $u_h$ satisfies (2a) but not (2b) in general.
Consider the minimization problem

$$
\sigma := \arg \min_{\tau \in H(\text{div}, \Omega)} \| A^{-1} \tau + \nabla u_h \|_{A, \Omega}.
$$

If the above minimum is 0, then

$$
\sigma = -A \nabla u_h \in H(\text{div}, \Omega)
$$

with

$$
\sigma \cdot n = 0 \text{ on } \Gamma_A \quad \nabla \cdot \sigma = \mu f_h \text{ in } \Omega
$$

i.e. \( u_h \) satisfies (2b), which implies that \( u = u_h \).

Otherwise, it measures how “non-conforming” \( u_h \) is.
Low frequencies

The Prager-Synge theorem
Assume that $\sigma \in H(\text{div}, \Omega)$ satisfies

$$\sigma \cdot n = 0 \text{ on } \Gamma_A, \quad \nabla \cdot \sigma = \mu f_h \text{ in } \Omega.$$ 

$\sigma = -A \nabla u$ is one example.

Then

$$a(e_h, v) = (\mu f_h, v)_\Omega - (A \nabla u_h, \nabla v)_\Omega$$

$$= (\nabla \cdot \sigma, v)_\Omega - (A \nabla u_h, \nabla v)_\Omega$$

$$= - (\sigma + A \nabla u_h, \nabla v)_\Omega.$$ 

**Prager–Synge inequality**

$$|a(e_h, v)| \leq \|A^{-1} \sigma + \nabla u_h\|_{A, \Omega} \|\nabla v\|_{A, \Omega}$$
Recall that whenever $\nabla \cdot \sigma = \mu f_h$ in $\Omega$ and $\sigma \cdot n = 0$ on $\Gamma_A$

$$\left| a(e_h, v) \right| \leq \| A^{-1} \sigma + \nabla u_h \|_{A,\Omega} \| \nabla v \|_{A,\Omega} \quad \forall v \in H^1_{\Gamma_D}(\Omega).$$

Picking in particular $v = e_h$, we have

$$\| \nabla e_h \|_{A,\Omega}^2 = \left| a(e_h, e_h) \right| \leq \| A^{-1} \sigma + \nabla u_h \|_{A,\Omega} \| \nabla e_h \|_{A,\Omega}.$$

### Guaranteed upper bound

$$\| \nabla e_h \|_{A,\Omega} \leq \| A^{-1} \sigma + \nabla u_h \|_{A,\Omega}$$

---

Low frequencies

Practical flux construction
Equilibrated fluxes satisfy \( \nabla \cdot \sigma = \mu f_h \) in \( \Omega \) and \( \sigma \cdot n = 0 \) on \( \Gamma_A \). They readily provide guaranteed error bounds.

Here, because \( \mu f_h \in \mathbb{P}_p(\mathcal{T}_h) \) we can actually find discrete fluxes \( \sigma_h \).

The correct tool to do that is the Raviart–Thomas finite element space

\[
W_h := \left\{ w_h \in H_{B_h}(\text{div}, \Omega) \mid w_h|_K \in [\mathbb{P}_p(K)]^3 + x\mathbb{P}_{p-1}(K) \right\}.
\]

We are going to select a discrete flux $\sigma_h \in W_h$.

The “equilibration” constraint on the flux is

$$\nabla \cdot \sigma_h = \mu f_h \text{ in } \Omega.$$

The estimate the flux provides us is

$$\|\nabla e_h\|_{A,\Omega} \leq \|A^{-1} \sigma_h + \nabla u_h\|_{A,\Omega}.$$
Recall that $\sigma_h \in W_h$ has to satisfy

$$\nabla \cdot \sigma_h = \mu f_h \text{ in } \Omega$$

and gives us

$$\|\nabla e_h\|_{A,\Omega} \leq \|A^{-1}\sigma_h + \nabla u_h\|_{A,\Omega}.$$

The “optimal” choice is then

$$\sigma_h := \arg \min_{\tau_h \in W_h} \|A^{-1}\tau_h + \nabla u_h\|_{A,\Omega}. \quad \nabla \cdot \tau_h = \mu f_h \text{ in } \Omega.$$
After introducing a Lagrange multiplier, there exists a unique pair \((\sigma_h, q_h) \in W_h \times \mathbb{P}_p(\mathcal{T}_h)\) such that

\[
\begin{cases}
(A^{-1}\sigma_h, w_h)_\Omega + (q_h, \nabla \cdot w_h)_\Omega &= -(\nabla u_h, w_h)_\Omega \quad \forall w_h \in W_h, \\
(\nabla \cdot \sigma_h, r_h) &= (\mu f_h, r_h) \quad \forall r_h \in \mathbb{P}_p(\mathcal{T}_h).
\end{cases}
\]

This square linear system can be solved to compute the optimal flux.

Unfortunately, it is more expensive than solving the original problem, so we avoid that in practice by using a localization trick.

At the continuous level, the ideal flux is $\sigma := -A \nabla u$.

A characterization is

$$\sigma = \arg \min \limits_{\tau \in H_{A}^{\Gamma}(\text{div}, \Omega)} \| A^{-1} \tau + \nabla u \|_{A, \Omega}.$$ 

The “optimal” flux directly mimicks this definition at the discrete level

$$\sigma_h := \arg \min \limits_{\tau_h \in W_h} \| A^{-1} \tau_h + \nabla u_h \|_{A, \Omega}.$$
Consider the set of “hat functions” \(\{\psi^a\}_{a \in \mathcal{V}_h}\) of the mesh. We then have
\[
\sum_{a \in \mathcal{V}_h} \psi^a = 1.
\]

The ideal flux \(\sigma := -A \nabla u\) can be decomposed as
\[
\sigma = \sum_{a \in \mathcal{V}_h} \sigma^a, \quad \sigma^a = -\psi^a A \nabla u.
\]

Easy computations show that
\[
\sigma^a \cdot n = 0 \text{ on } \partial \omega^a \quad \nabla \cdot \sigma^a = \psi^a \mu f_h - A \nabla u \cdot \nabla \psi^a \text{ in } \omega^a
\]
Localization

We have shown that

\[ \sigma = \sum_{a \in \mathcal{V}_h} \sigma^a \]

and

\[ \sigma^a = \arg \min_{\tau \in H_0(\text{div}, \omega^a)} \| A^{-1} \tau + \nabla u \|_{A, \omega^a}. \]

\[ \nabla \cdot \tau = \psi^a \mu_f - A \nabla u \cdot \nabla \psi^a \text{ in } \omega^a \]

We can mimick that on the discrete level!
We thus set

\[ \sigma_h := \sum_{a \in \mathcal{V}_h} \sigma^a_h \]

with

\[ \sigma^a_h := \arg \min_{\tau_h \in H_0(\text{div}, \omega^a) \cap \mathcal{W}_h} \| A^{-1} \tau_h + \nabla u_h \|_{A, \omega^a} \]

\[ \nabla \cdot \tau_h = \psi^a \mu f_h - A \nabla u_h \cdot \nabla \psi^a \text{ in } \omega^a \]

The compatibility condition

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

holds true since \( u_h \) is the discrete solution and \( \psi^a \in \mathcal{V}_h \).
Summary of the localization process

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

\[ \boldsymbol{\sigma}_h^a := \arg \min_{\tau_h \in H_0(\text{div}, \omega^a) \cap W_h} \| A^{-1} \tau_h + \nabla u_h \|_{A, \omega^a}. \]

\[ \nabla \cdot \tau_h = \psi^a \mu f_h - A \nabla \psi^a \quad \text{in} \quad \omega^a \]

Step 2: assemble these local contributions

\[ \boldsymbol{\sigma}_h := \sum_{a \in V_h} \boldsymbol{\sigma}_h^a. \]

Step 3: compute the estimator

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

Step 4: enjoy the guaranteed estimate

\[ \| \nabla e_h \|_{A, \Omega} \leq \eta. \]
Low frequencies

Efficiency
For time reasons, I will not give any proofs, but we can show that

\[ \eta \leq C_{\text{eff}} \| \nabla e_h \|_{A,\Omega}, \]

where \( C_{\text{eff}} \) only depends on:

- the “flatness” of the tetrahedra in the mesh,
- the “contrasts” in the coefficients.

A nice aspect is that \( C_{\text{eff}} \) does not depend on \( p \).

Low frequencies

Takeaways
Takeaways

An equilibrated flux is an object $\sigma \in H(\text{div}, \Omega)$ such that

$$\sigma \cdot n = 0 \text{ on } \Gamma_A \quad \nabla \cdot \sigma = \mu f_h \text{ in } \Omega.$$ 

There exist efficient algorithms to build a discrete flux $\sigma_h \in W_h$.

The guaranteed error estimate

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

holds true with

$$\eta := \| A^{-1} \sigma_h + \nabla u_h \|_{A,\Omega}.$$ 

This bound cannot be too loose:

$$\eta \leq C_{\text{eff}} \| \nabla e_h \|_{A,\Omega}.$$
High frequencies
The high frequency case

Back to our original problem

\[
\begin{cases}
-\omega^2 \mu u - \nabla \cdot (A \nabla u) = \mu f \quad \text{in } \Omega, \\
u = 0 \quad \text{on } \Gamma_D, \\
A \nabla u \cdot n - i\omega \gamma u = 0 \quad \text{on } \Gamma_A.
\end{cases}
\]

We introduce

\[
b(u, v) := -\omega^2 (\mu u, v)_\Omega - i\omega (\gamma u, v)_{\Gamma_A} + (A \nabla u, \nabla v)_{\Omega}.
\]
We will consider the “balanced” norm

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

The sesquilinear form \( b \) is not coercive.

Instead we have the “Gårding” inequality

\[ \text{Re} b(v, v) = \| \nabla v \|_{A, \Omega}^2 - \omega^2 \| v \|_{\mu, \Omega}^2 = \| v \|_{\omega, \Omega}^2 - 2\omega^2 \| v \|_{\mu, \Omega}^2. \]
High frequencies

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

$$\sigma \cdot \mathbf{n} = -i\omega \gamma u \text{ on } \Gamma_A \quad \nabla \cdot \sigma = \mu f_h + \omega^2 \mu u \text{ in } \Omega.$$ 

Hence, natural requirements for $\sigma_h$ are

$$\sigma_h \cdot \mathbf{n} = -i\omega \gamma u_h \text{ on } \Gamma_A \quad \nabla \cdot \sigma_h = \mu f_h + \omega^2 \mu u_h \text{ in } \Omega.$$ 

The “standard” reconstruction algorithms directly extend.
Let \( v \in H^1_{\Gamma_D}(\Omega) \). We have

\[
\begin{align*}
\langle e_h, v \rangle & = \langle \mu f_h, v \rangle_{\Omega} - \langle u_h, v \rangle_{\Omega} \\
& = \langle \mu f_h + \omega^2 \mu u_h, v \rangle_{\Omega} + i \omega \langle \gamma u_h, v \rangle_{\Gamma_A} - \langle A \nabla u_h, \nabla v \rangle_{\Omega} \\
& = \langle \nabla \cdot \sigma_h, v \rangle_{\Omega} - \langle \sigma_h \cdot n, v \rangle_{\Gamma_A} - \langle A \nabla u_h, \nabla v \rangle_{\Omega} \\
& = -\langle \sigma_h + A \nabla u_h, \nabla v \rangle_{\Omega}.
\end{align*}
\]

**Prager-Synge inequality**

\[
|\langle e_h, v \rangle| \leq \eta \|\nabla v\|_{A,\Omega} \quad \forall v \in H^1_{\Gamma_D}(\Omega)
\]

So far, so good!
What’s the matter?

Here, we do not have

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

which is a major issue!

Instead, we only have the “Gårding” inequality

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

A coarse error estimate
For $g \in L^2(\Omega)$, let $\mathcal{I}^* g$ denote the unique element of $H^1_{\Gamma_D}(\Omega)$ such that
\[ b(w, \mathcal{I}^* g) = 2\omega^2 (\mu w, g)_\Omega \quad \forall w \in H^1_{\Gamma_D}(\Omega) \]
and let
\[ C_{st} := \frac{1}{\omega} \max_{\|g\|_{\mu, \Omega} = 1} \| \nabla (\mathcal{I}^* g) \|_{A, \Omega}. \]

$C_{st}$ is the best constant such that
\[ \| \nabla (\mathcal{I}^* g) \|_{A, \Omega} \leq C_{st} \omega \| g \|_{\mu, \Omega} \quad \forall g \in L^2(\Omega). \]

It is closely related to resolvant estimates.
Making up for the lack of coercivity

By definition, we have

\[ b(w, \mathcal{S}^* e_h) = 2\omega^2 (\mu w, e_h) \quad \forall w \in H^1_{\Gamma_D}(\Omega). \]

Hence, in particular,

\[ b(e_h, \mathcal{S}^* e_h) = 2\omega^2 \| e_h \|^2_{\mu,\Omega}, \]

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

\[ \text{Re } b(e_h, e_h) = \| e_h \|^2_{\omega,\Omega} - 2\omega^2 \| e_h \|^2_{\mu,\Omega}. \]
Making up for the lack of coercivity

Using Prager-Synge inequality, we have

\[ \| e_h \|_{\omega, \Omega}^2 = \text{Re} \ b(e_h, e_h + \mathcal{J}^* e_h) \leq \eta \| \nabla (e_h + \mathcal{J}^* e_h) \|_{A, \Omega}. \]

It follows that

\[ \| e_h \|_{\omega, \Omega}^2 \leq \eta \left( \| \nabla e_h \|_{A, \Omega} + \| \nabla (\mathcal{J}^* e_h) \|_{A, \Omega} \right) \]
\[ \leq \eta \left( \| \nabla e_h \|_{A, \Omega} + C_{\text{st}}\omega \| e_h \|_{\mu, \Omega} \right) \]
\[ \leq \eta \max(1, C_{\text{st}}) \| e_h \|_{\omega, \Omega}, \]

and

\[ \| e_h \|_{\omega, \Omega} \leq \max(1, C_{\text{st}}) \eta. \]
Coarse error estimate

We obtained the following error estimate:

\[ \| e_h \| \leq \max(1, C_{st}) \eta \]

\( C_{st} \) is the best constant such that:

\[ \| \nabla (\mathcal{I}^* g) \|_{A, \Omega} \leq C_{st} \omega \| g \|_{\mu, \Omega} \quad \forall g \in L^2(\Omega). \]
High frequencies

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 \( C_{\text{eff}} \) only depends on the elements “flatness” and the contrasts.


The problem with the coarse error estimate

Recall that

\[ \eta \leq C_{\text{eff}} \| e_h \|_{\omega, \Omega} \quad \| e_h \|_{\omega, \Omega} \leq \max(1, C_{\text{st}}) \eta. \]

We have \( C_{\text{eff}} \simeq 1 \) and \( C_{\text{st}} \gtrsim \omega l_{\Omega} / c_{\text{min}} \), so that

\[ \eta \lesssim \| e_h \|_{\omega, \Omega} \lesssim \frac{\omega l_{\Omega}}{c_{\text{min}}} \eta. \]

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

Sharp error estimate
The approximation factor

We introduce

$$C_{ap} := \frac{1}{\omega} \max_{g \in L^2(\Omega)} \min_{v_h \in V_h} \| \nabla (S^* g - v_h) \|_{A,\Omega}.$$

**Approximability**

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

$$\| \nabla (S^* g - v_h^*) \|_{A,\Omega} \leq C_{ap} \omega \| g \|_{\mu,\Omega}.$$

This was called $\eta$ in Markus’ talk.
It’s better than the stability constant!

Recall that

\[ C_{ap} := \max_{g \in L^2(\Omega)} \min_{v_h \in V_h} \omega \| \mathcal{L} g - v_h \|_{\omega,\Omega} \|g\|_{\mu,\Omega} = 1 \]

since we can take \( v_h = 0 \), we have

\[ C_{ap} \leq \max_{g \in L^2(\Omega)} \omega \| \mathcal{L} g \|_{\omega,\Omega} =: C_{st}. \]

Besides, standard approximability results for FEM show that

\[ C_{ap} \leq C \left( \frac{1}{h} \right)^s \left( \frac{1}{p \ell_{\Omega}} \right) \]

for some \( s > 0 \), so that \( C_{ap} \to 0 \).
Using Galerkin orthogonality

Recall that

$$\|e_h\|^2_{\omega, \Omega} = \text{Re} \, b(e_h, e_h + \mathcal{S}^* e_h).$$

By Galerkin orthogonality, we have

$$\|e_h\|^2_{\omega, \Omega} = \text{Re} \, b(e_h, e_h) + \text{Re} \, b(e_h, \mathcal{S}^* e_h)$$
$$= \text{Re} \, b(e_h, e_h) + \text{Re} \, b(e_h, \mathcal{S}^* e_h - \nu^*_h)$$
$$\leq \eta \left( \|\nabla e_h\|_{A, \Omega} + \|\nabla (\mathcal{S}^* e_h - \nu^*_h)\|_{A, \Omega} \right)$$
$$\leq \eta \max(1, \mathcal{C}_{ap}) \|e_h\|_{\omega, \Omega}.$$
Sharp error estimate

\[ \| e_h \|_{\omega, \Omega} \leq \max(1, C_{ap}) \eta. \]

Approximation factor

\[ C_{ap} := \max_{g \in L^2(\Omega)} \min_{v_h \in V_h} \omega \| \mathcal{I} g - v_h \|_{\omega, \Omega} \to 0. \]

\[ \| g \|_{\mu, \Omega} = 1 \]


High frequencies

Takeaways
The “equilibration” technology is the same than for low frequencies.

### Corse error estimate

\[ \| e_h \|_{\omega,\Omega} \leq \max(1, C_{st}) \eta \quad C_{st} \gtrsim \omega l \Omega / c_{\min} \]

### Sharp error estimate

\[ \| e_h \|_{\omega,\Omega} \leq \max(1, C_{ap}) \eta \quad C_{ap} \to 0 \]

### Efficiency

\[ \eta \leq C_{\text{eff}} \left( 1 + \frac{\omega h}{c_{\min} p} \right) \| e_h \|_{\omega,\Omega} \]
Controlling the pre-factors
Controlling the pre-factors

The stability constant $C_{st}$
The stability constant

The stability constant is defined by

$$C_{st} := \max_{g\in L^2(\Omega)} \|\nabla (\mathcal{L}^* g)\|_{A,\Omega}.$$  

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

It is known that we have “at least”:

$$C_{st} \gtrsim \frac{\omega \ell \Omega}{c_{\text{min}}}.$$ 

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

$$C_{st} \lesssim \frac{\omega \ell \Omega}{c_{\text{min}}}.$$ 

If strong trapping happens, “extreme” behaviors can occur

$$C_{st} \gtrsim \exp \left( \alpha \frac{\omega \ell \Omega}{c_{\text{min}}} \right)$$

for “some” frequencies. For “most frequency”

$$C_{st} \gtrsim \left( \frac{\omega \ell \Omega}{c_{\text{min}}} \right)^\beta.$$ 

---

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 \( A = I \) in \( \Omega \setminus D \).

Assume that \( \mu = \mu_D \geq 1 \) and \( A = A_D \preceq I \) in \( D \).

This describes an obstacle made of a material with a “slow” wave speed.
Guaranteed upper bound

\[ C_{st} \leq 6 + \frac{3 + \sqrt{3}}{\sqrt{3}} \frac{\omega \ell_{\Omega}}{c_{\min}} \]

The proof relies on a “Morawetz multiplier”: multiply the PDE by \( x \cdot \nabla u \) and integrate by parts until it works!

Controlling the pre-factors

The approximation factor $C_{ap}$
The approximation factor

The approximation factor is defined by

$$C_{ap} := \max_{g \in L^2(\Omega)} \min_{v^*_h \in V_h} \|\nabla (J^*g - v^*_h)\|_{A,\Omega}.$$

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

Assuming that $A = I$, $\Omega$ is convex, and $C_{st}$ is known, we can control it.
Idea one: explicit interpolation error


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

\[
\| \nabla (v - I_h^1 v) \|_{A,\Omega} \leq C_T,i h \| \nabla^2 v \|_\Omega,
\]

with a constant \( C_{T,i} \) that is easily computable.

We then have

\[
C_{ap} := \max_{g \in L^2(\Omega)} \min_{v_h^* \in V_h} \| \nabla (\mathcal{I}^* g - v_h^*) \|_{A,\Omega}
\]

\[
\leq \max_{g \in L^2(\Omega)} \| \nabla (\mathcal{I}^* g - I_h^1(\mathcal{I}^* g) \|_{A,\Omega}
\]

\[
\leq C_T,i \max_{g \in L^2(\Omega)} \| \nabla^2 (\mathcal{I}^* g) \|_\Omega.
\]
Because $\Omega$ is convex and $\gamma = 1$, we have

$$\|\nabla^2 (\mathcal{L}^* g)\|_\Omega \leq \|\Delta (\mathcal{L}^* g)\|_\Omega.$$  

Then, we use the facts that

$$-\Delta (\mathcal{L}^* g) = 2\mu \omega^2 g + \mu \omega^2 \mathcal{L}^* g$$

and

$$\omega \|\mathcal{L}^* g\|_{\mu, \Omega} \leq 2C_{st} \|g\|_{\mu, \Omega},$$

to show that

$$\|\nabla^2 (\mathcal{L}^* g)\|_\Omega \leq 2 \omega C_{\min} (1 + C_{st}) \|g\|_{\mu, \Omega}.$$
Explicit control of the approximation factor

Recall that

\[ C_{\text{ap}} \leq C_{T,i} h \max_{g \in L^2(\Omega)} \max_{\|g\|_{\mu,\Omega} = 1} \| \nabla^2 (S^* g) \|_{\Omega} \]

and

\[ \| \nabla^2 (S^* g) \|_{\Omega} \leq 2 \frac{\omega}{C_{\text{min}}} (1 + C_{\text{st}}) \| g \|_{\mu,\Omega} \quad \forall g \in L^2(\Omega). \]

Guaranteed bound

\[ C_{\text{ap}} \leq 2 \left( 1 + C_{T,i} \right) \frac{\omega h}{C_{\text{min}}} C_{\text{st}} \]
Controlling the pre-factors

Takeaways
The estimator $\eta$ needs to be “pre-factored” by $C_{st}$ or $C_{ap}$.
The “qualitative” behaviors of both quantities are relatively well known.

The behaviour of $C_{st}$ is only dictated by the PDE.
Explicit bounds are available for non-trapping star-shaped obstacles.

The approximation factor $C_{ap}$ depends on the PDE and $V_h$.
When $A = I$, $\Omega$ is convex and $C_{st}$ is known, we can bound it nicely.
Numerical illustrations
Numerical illustrations

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

\[
\begin{cases}
  -\omega^2 u - \Delta u = 0 & \text{in } \Omega, \\
  \nabla u \cdot n - i\omega u = g & \text{on } \Gamma_A,
\end{cases}
\]

where

\[
g := \nabla \xi_\theta \cdot n - i\omega \xi_\theta \quad \xi_\theta := e^{i\omega d \cdot x}
\]

with $d := (\cos \theta, \sin \theta)$ 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 $\omega = \pi$

\[
E_{\text{fem}} := \| e_h \|_{\omega, \Omega}
\]
\[
E_{\text{est}} := \eta
\]
\[
\tilde{E}_{\text{est}} := (1 + C_{\text{ap}}) \eta
\]
Plane wave experiment $p = 1$ and $\omega = 4\pi$

\[ E_{\text{fem}} := \| e_h \|_{\omega, \Omega} \]
\[ E_{\text{est}} := \eta \]
\[ \tilde{E}_{\text{est}} := (1 + C_{ap}) \eta \]
Plane wave experiment $p = 1$ and $\omega = 10\pi$

\[ E_{\text{fem}} : = \| e_h \|_{\omega, \Omega} \]
\[ E_{\text{est}} : = \eta \]
\[ \tilde{E}_{\text{est}} : = (1 + \mathcal{C}_a)\eta \]
Plane wave experiment $p = 1$ and $\omega = 20\pi$

\[ E_{\text{fem}} := \| e_h \|_{\omega, \Omega} \]
\[ E_{\text{est}} := \eta \]
\[ \tilde{E}_{\text{est}} := (1 + \mathcal{C}_{\text{ap}})\eta \]
Plane wave experiment $p = 4$ and $\omega = 10\pi$

\[ E_{\text{fem}} := \| e_h \|_{\omega, \Omega} \]
\[ E_{\text{est}} := \eta \]
\[ \tilde{E}_{\text{est}} := (1 + \mathcal{C}_{\text{ap}}) \eta \]
Plane wave experiment \( p = 4 \) and \( \omega = 20\pi \)

\[ E_{\text{fem}} := \| e_h \|_{\omega, \Omega} \]

\[ E_{\text{est}} := \eta \]

\[ \tilde{E}_{\text{est}} := (1 + C_{\text{ap}}) \eta \]
Plane wave experiment \( p = 4 \) and \( \omega = 40\pi \)

\[
E_{\text{fem}} := \| e_h \|_{\omega, \Omega}
\]

\[
\tilde{E}_{\text{est}} := (1 + \mathcal{C}_{\text{ap}}) \eta
\]

\[
E_{\text{est}} := \eta
\]
Plane wave experiment $p = 4$ and $\omega = 60\pi$

\[
E_{\text{fem}} := \| e_h \|_{\omega, \Omega}
\]

\[
E_{\text{est}} := \eta
\]

\[
\widetilde{E}_{\text{est}} := (1 + \mathcal{C}_{ap}) \eta
\]
Numerical illustrations

A more realistic example
Scattering by an non-trapping obstacle

We now consider a scattering problem

\[
\begin{cases}
-\omega^2 u - \Delta u = 0 & \text{in } \Omega, \\
u = 0 & \text{on } \Gamma_D, \\
abla u \cdot n - i\omega u = g & \text{on } \Gamma_A,
\end{cases}
\]

where again \( g = \nabla \xi_\theta \cdot n - i\omega \xi_\theta \).

We fix the wavenumber \( \omega = 10\pi \) and employ \( P_3 \) elements.

We consider a sequence of meshes that are adaptively refined using \( \eta_K \).
Solution of the scattering problem

Real (left) and imaginary (right) parts of the solution
Estimated error in mesh #1

Estimator $\eta_K$ (left) and elementwise error $\| e_h \|_{\omega, K}$ (right)
Estimated error in mesh #2

Estimator $\eta_K$ (left) and elementwise error $\|e_h\|_{\omega,K}$ (right)
Estimated error in mesh #3

Estimator $\eta_K$ (left) and elementwise error $\| e_h \|_{\omega,K}$ (right)
Behaviors of the estimated and analytical errors in the adaptive procedure

\[ E_{\text{fem}} := \| e_h \|_{\omega, \Omega} \]
\[ E_{\text{est}} := \eta \]
\[ \tilde{E}_{\text{est}} := (1 + C_{st}) \eta \]
Concluding remarks
Concluding remarks

Takeaways
We construct an a posteriori error estimator $\eta$ via flux equilibration. It directly provides guaranteed error estimates at low frequencies.

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

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

Concluding remarks

Extensions
Extensions

We can obtain guaranteed bounds on $C_{st}$ in weakly trapping geometry with “directional” Morawetz multiplier

$$(x_d e^d) \cdot \nabla u.$$ 


T. Chaumont-Frelet and E.A. Spence, *submitted*.

T. Chaumont-Frelet and Z. Kassali, *in progress*.

It is possible to obtain approximations for $C_{ap}$ in more general situations.
Extensions

Everything I presented (painfully) extends to Maxwell’s equations:

- T. Chaumont-Frelet, *will be submitted soon!*

Thanks for your attention! :-}