



Institut für Mathematik  
Universität Zürich



**computational mathematics**

Wave-Number Explicit Convergence Analysis for  
Galerkin-type discretizations  
of the Helmholtz equation

Stefan A. Sauter

Institut für Mathematik, Universität Zürich

January 2016

joint work with J.M. Melenk (TU Wien) and I.G. Graham (U Bath)

## Setting, Main Assumptions I

Let  $\Omega \subset \mathbb{R}^d$  a bounded Lipschitz domain and  $\omega \geq \omega_0 > 0$ . Let  $c \in L^\infty(\Omega)$  and  $\mathbf{A} \in \mathbf{W}^{1,\infty}(\Omega^*, \mathbb{R}_{\text{sym}}^{d \times d})$  be given for some domain  $\Omega^* \supset \bar{\Omega}$  with

$$0 < \alpha := \inf_{\mathbf{x} \in \Omega^*} \lambda_{\min}(\mathbf{A}(\mathbf{x})) \leq \sup_{\mathbf{x} \in \Omega^*} \lambda_{\max}(\mathbf{A}(\mathbf{x})) =: \beta < \infty,$$
$$0 < c_{\min} := \inf_{x \in \Omega} c(x) \leq \sup_{x \in \Omega} c(x) =: c_{\max} < \infty.$$

Let  $\mathcal{H} \subset H^1(\Omega)$  be a closed subspace with norm\*

$$\|u\|_{\mathcal{H},\Omega} := \left( \|\nabla u\|^2 + \left\| \frac{\omega}{c} u \right\|^2 \right)^{1/2}.$$

\*  $(\cdot, \cdot): L^2(\Omega)$  scalar product,  $\|\cdot\|: L^2(\Omega)$  norm.

## Variational Formulation of Helmholtz equation, Main Assumptions II

Sesquilinear form  $A : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$  can be split into  $A = a - b$  with

definition of $a$	$a(u, v) = (\mathbf{A} \nabla u, \nabla v) - \left( \frac{\omega}{c} u, \frac{\omega}{c} v \right),$
-------------------	--

continuity of $b$	$\forall u, v \in \mathcal{H} \quad  b(u, v)  \leq C_b \ u\ _{\mathcal{H}} \ v\ _{\mathcal{H}}$
-------------------	---

definiteness of $b$	$(u \in \mathcal{H} \wedge \operatorname{Im} b(u, u) = 0) \implies u _{\partial\Omega} = 0,$
---------------------	--

$$\forall u \in \mathcal{H} \quad \operatorname{Re} b(u, u) \leq 0,$$

$b$ : boundary operator	$(uv) _{\partial\Omega} = 0 \implies b(u, v) = 0,$
-------------------------	--

**Helmholtz Problem.** For  $F \in \mathcal{H}'$  seek  $u \in \mathcal{H}$  such that

$$A(u, v) = F(v) \quad \forall v \in \mathcal{H}.$$

**Theorem 1.** Let all assumptions be satisfied. The Helmholtz problem and its adjoint equation have a unique solution.

The proof (e.g., Leis) uses Fredholm theory, elliptic regularity, unique continuation principle.

## Stability results

Let  $F(v) = \int_{\Omega} f \bar{v}$  for some  $f \in L^2(\Omega)$ . Let  $u \in \mathcal{H}$  denote the solution of the Helmholtz problem. Theorem 1 implies that there is a constant  $C_{\text{stab}} > 0$  independent of  $f$  such that

$$\|u\|_{\mathcal{H}} \leq C_{\text{stab}} \|f\|.$$

**Question:** How does  $C_{\text{stab}}$  depend on data?

**Proposition (Melenk, 1995).** Let  $\Omega$  be a bounded star-shaped domain with smooth boundary or a bounded convex domain. Let  $\mathbf{A} = \mathbf{I}$  and  $c = 1$ . Let  $b(u, v) = \pm i\omega \int_{\partial\Omega} u \bar{v}$ . Then, there is  $C_{\text{stab}} > 0$  depending only on  $\Omega$  such that for any  $f \in L^2(\Omega)$  the solution  $u$  of the Helmholtz problem satisfies

$$\begin{aligned} \|u\|_{\mathcal{H}} &\leq C_{\text{stab}} \|f\|, \\ |u|_{H^2(\Omega)} &\leq C_{\text{stab}} (1 + \omega) \|f\|. \end{aligned}$$

The proof employs elliptic regularity and the “Rellich trick”, i.e., using the test function  $v = \langle \mathbf{x}, \nabla u \rangle$  and integration by parts.

**Theorem (Melenk, 2012, Melenk/Esterhazy 2012).** Let  $\Omega \subset \mathbb{R}^d$ ,  $d \in \{2, 3\}$ , be a bounded Lipschitz domain. Let  $\mathbf{A} = \mathbf{I}$  and  $c = 1$ . Let  $b(u, v) = \pm i\omega \int_{\partial\Omega} u \bar{v}$ . Then, there is  $C > 0$  depending only on  $\Omega$  such that for any  $f \in L^2(\Omega)$  the solution  $u$  of the Helmholtz problem satisfies

$$\|u\|_{\mathcal{H}} \leq C\omega^{5/2} \|f\|.$$

The proof uses a) Fourier analysis of the full space problem, b) “Helmholtz” extension operators, c) layer potentials.

**Theorem (Betcke et al. 2010).**

**Geometric assumptions:** Let  $\mathcal{E} := \left\{ \mathbf{x} \in \mathbb{R}^2 \mid \left( \frac{x_1}{a_1} \right)^2 + \left( \frac{x_2}{a_2} \right)^2 < 1 \right\}$  for some  $a_1 > a_2 > 0$  and let  $\Omega := \mathbb{R}^2 \setminus \overline{\mathcal{E}}$ .

**Functional assumption:**  $\mathbf{A} = \mathbf{I}$  and  $c = 1$ .  $\mathcal{H} := H_{w,0}^1(\Omega)$  : weighted Sobolev space for the exterior Helmholtz equation and  $b(u, v) = 0$ .

Then there exists a sequence  $0 < \omega_0 < \omega_1 < \dots \rightarrow \infty$  such that the inverse of the  $\eta$ -combined acoustic double layer operator  $A_{\omega_m, \eta}^{-1}$  satisfies

$$\|A_{\omega_m, \eta}^{-1}\| \geq C e^{\gamma \omega_m} \left( 1 + \frac{|\eta|}{\omega_m} \right)^{-1} \quad \text{for some } \gamma > 0, C > 0.$$



## Galerkin Discretization

Let  $S \subset \mathcal{H}$  a finite dimensional subspace. For given  $f \in L^2(\Omega)$  find  $u_S \in S$  such that

$$A(u_S, v) = (f, v) \quad \forall v \in \mathcal{H}.$$

## Discrete Stability

Dual problem:

For given  $f \in L^2(\Omega)$  find  $z \in \mathcal{H}$  such that  $A(v, z) = (v, f) \quad \forall v \in \mathcal{H}$ .

Dual solution operator:  $Q^* : L^2(\Omega) \rightarrow \mathcal{H}$  with  $Q^* f := z$ .

**Definition (Adjoint approximability)** (see Schatz 74, S. 06, Banjai, S. 07, Dörfler, S. 13) For a finite dimensional subspace  $S \subset \mathcal{H}$ , the *adjoint approximability* for the Helmholtz problem is

$$\sigma(S) := \sup_{g \in L^2(\Omega) \setminus \{0\}} \frac{\inf_{v \in S} \left\| Q^* \left( \left( \frac{\omega}{c} \right)^2 g \right) - v \right\|_{\mathcal{H}}}{\left\| \frac{\omega}{c} g \right\|_{L^2(\Omega)}}.$$

**Theorem (stability)** (S. 06). Let the **main assumptions** be satisfied and

$$\sigma(S) < \frac{\min\{\alpha, 1\}}{\max\{\beta, 1\} + C_b}.$$

Then, the Galerkin discretization has a unique solution.

**Theorem** (S. 06). Let the **main assumptions** be satisfied and

$$\sigma(S) < \frac{\min\{\alpha, 1\}}{4(\max\{\beta, 1\} + C_b)}.$$

Then, the Galerkin solution exists, is unique, and the error  $e = u - u_S$  satisfies

$$\|e\|_{\mathcal{H}} \leq 2 \frac{\max\{\beta, 1\} + C_b}{\min\{\alpha, 1\}} \inf_{v \in S} \|u - v\|_{\mathcal{H}},$$

$$\left\| \frac{\omega}{c} e \right\|_{L^2(\Omega)} \leq (\max\{\beta, 1\} + C_b) \sigma(S) \|e\|_{\mathcal{H}}.$$

## Estimate of adjoint approximability $\sigma(S)$

The key rôle for the estimate of the adjoint approximability  $\sigma(S)$  is played by the *splitting lemma*; see:

Melenk & S. 2010, 2011,

Melenk 2012,

Parsania, Melenk & S. 2013,

Esterhazy & Melenk 2012.

**Splitting Lemma.** Let  $\Omega \subset \mathbb{R}^d$  be a bounded Lipschitz domain; either with analytic boundary or  $\Omega$  is a bounded polygonal domain. Let  $\mathbf{A} = \mathbf{I}$ ,  $c = 1$ , and  $b(u, v) = \pm i\omega \int_{\partial\Omega} u\bar{v}$ . Then there exist constants  $C, \gamma > 0$  independent of  $\omega \geq \omega_0$  such that for every  $f \in L^2(\Omega)$  the solution of the Helmholtz equation can be written as  $u = u_{H^2} + u_{\mathcal{A}}$  with

$$\omega \|u_{H^2}\|_{\mathcal{H}} + \|u_{H^2}\|_{H^2(\Omega)} \leq C \|f\|_{L^2(\Omega)},$$

$$\|\Phi_{n,k} \nabla^{n+2} u_{\mathcal{A}}\|_{L^2(\Omega)} \leq C \frac{C_{\text{stab}}}{\omega} \max\{n, \omega\}^{n+2} \gamma^n \|f\|_{L^2(\Omega)} \quad \forall n \in \mathbb{N}_0,$$

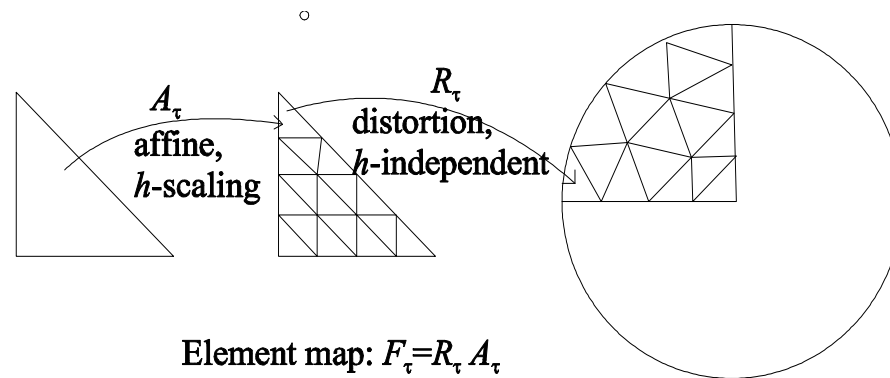
where  $\Phi_{n,k}$  denotes the weight function lifting the corner singularities. It holds  $\Phi_{n,k} = 1$  for analytic domains.

## hp-Finite Elements

$\Omega \subset \mathbb{R}^d$  is a bounded domain with analytic boundary.

$\mathcal{T} = \{\tau_i : 1 \leq i \leq q\}$  is a conforming finite element triangulation

$h_\tau := \text{diam } \tau, \quad h_{\mathcal{T}} := \max \{h_\tau, \tau \in \mathcal{T}\}.$



**Remark 1.** For polygonal domains we assume that  $\mathcal{T}$  is constructed from a quasi-uniform, shape-regular triangulation by refining those elements geometrically which touch the vertices by using  $L \approx p$  refinement layers.

### hp-Finite Element Space

$$S_{\mathcal{T}}^p := \left\{ u \in C^0(\bar{\Omega}) \mid \forall \tau \in \mathcal{T} : u|_{\tau} \circ F_{\tau} \in \mathbb{P}_p \right\}.$$



**Theorem** (Melenk, S. 2010,11). Let the assumptions of the splitting lemma be satisfied. There exist constants  $c_1, c_2$  independent of  $\omega, h$ , and  $p$  such that

$$\omega h/p \leq c_1 \quad \text{together with} \quad p \geq c_2 \log \omega$$

implies the existence and uniqueness of the finite element solution.

The minimal dimension of the corresponding  $h - \log \omega$  finite element space satisfies

$$\dim S = O(\omega^d).$$

**Theorem** (Melenk, S. 2010, 11). Let the assumptions of the previous theorem be satisfied. In the case of a polygonal domain we assume that  $\Omega$  is convex. Then,

$$\|u - u_S\|_{\mathcal{H}} \lesssim C_{f,g} \frac{h}{p}$$

## Variable Wave Speed (joint work with I.G. Graham)

Sesquilinear form  $A : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$  is given by

$$A(u, v) = (\nabla u, \nabla v) - \left( \frac{\omega}{c} u, \frac{\omega}{c} u \right) - i \left( \frac{\omega}{c} u, v \right)_{L^2(\partial\Omega)}.$$

**Question:** How does the stability constant  $C_{\text{stab}}$  behave if  $\omega$  is replaced by  $\frac{\omega}{c}$ ? Which assumptions imply  $C_{\text{stab}} \lesssim \left( \frac{\omega}{c_{\min}} \right)^\alpha$ ?

The analysis of the problem becomes particularly hard if  $c'$  is large, e.g.,

$$c(\mathbf{x}) = 2 + \prod_{i=1}^d \sin(mx_i) \text{ and } m \sim \omega.$$

Techniques of proofs:

a) Fourier Analysis

For constant wave speed we have  $\hat{u}(\xi) = \hat{G}(\xi) \hat{f}(\xi)$  with

$$\hat{G}(\xi) = \begin{cases} \frac{i}{\sqrt{2\pi\omega}} \int_0^\infty e^{i\omega r} \mu(r) \cos(\|\xi\| r) dr & d = 1, \\ \frac{i}{4} \int_0^\infty H_0^{(1)}(\omega r) \mu(r) r J_0(r \|\xi\|) dr & d = 2, \\ (2\pi)^{-3/2} \int_0^\infty e^{i\omega r} \mu(r) r \frac{\sin(r \|\xi\|)}{r \|\xi\|} dr & d = 3. \end{cases}$$

and the estimate for  $C_{\text{stab}}$  follows from estimates of  $\hat{G}$ .

This technique does **not** apply to the case of non-constant wave speed. We tried to employ modulated Fourier expansions but **failed**.

## b) The “Rellich trick”

Employ the function  $v = \langle \mathbf{x}, \nabla u \rangle$  as a test function in the variational formulation and integrate by parts. This technique does **not** work if  $\nabla c$  is “large” and “oscillating”.

## c) Analyse the Sturm-Liouville problem in one-dimension.

The system matrix for the Sturm-Liouville problem is tri-diagonal but highly indefinite. For the analysis of the inverse one needs lower estimates for its determinant.

$$\begin{aligned}
\det = & q(1)q(8) + q(2)s(1)q(8) + q(1)q(2)q(3)s(2)q(8) + q(3)s(1)s(2)q(8) + q(1)q(3)q(4)s(3)q(8) + q(2)q(3)q(4)s(1)s(3)q(8) \\
& + q(1)q(2)q(4)s(2)s(3)q(8) + q(4)s(1)s(2)s(3)q(8) + q(1)q(4)q(5)s(4)q(8) + q(2)q(4)q(5)s(1)s(4)q(8) + q(1)q(2)q(3)q(4)q(5)s(2) \\
& + q(3)q(4)q(5)s(1)s(2)s(4)q(8) + q(1)q(3)q(5)s(3)s(4)q(8) + q(2)q(3)q(5)s(1)s(3)s(4)q(8) + q(1)q(2)q(5)s(2)s(3)s(4)q(8) \\
& + q(5)s(1)s(2)s(3)s(4)q(8) + q(1)q(5)q(6)s(5)q(8) + q(2)q(5)q(6)s(1)s(5)q(8) + q(1)q(2)q(3)q(5)q(6)s(2)s(5)q(8) \\
& + q(3)q(5)q(6)s(1)s(2)s(5)q(8) + q(1)q(3)q(4)q(5)q(6)s(3)s(5)q(8) + q(2)q(3)q(4)q(5)q(6)s(1)s(3)s(5)q(8) \\
& + q(1)q(2)q(4)q(5)q(6)s(2)s(3)s(5)q(8) + q(4)q(5)q(6)s(1)s(2)s(3)s(5)q(8) + q(1)q(4)q(6)s(4)s(5)q(8) + q(2)q(4)q(6)s(1)s(4)s(5) \\
& + q(1)q(2)q(3)q(4)q(6)s(2)s(4)s(5)q(8) + q(3)q(4)q(6)s(1)s(2)s(4)s(5)q(8) + q(1)q(3)q(6)s(3)s(4)s(5)q(8) + q(2)q(3)q(6)s(1)s(3)s(5) \\
& + q(1)q(2)q(6)s(2)s(3)s(4)s(5)q(8) + q(6)s(1)s(2)s(3)s(4)s(5)q(8) + q(1)q(6)q(7)s(6)q(8) + q(2)q(6)q(7)s(1)s(6)q(8) \\
& + q(1)q(2)q(3)q(6)q(7)s(2)s(6)q(8) + q(3)q(6)q(7)s(1)s(2)s(6)q(8) + q(1)q(3)q(4)q(6)q(7)s(3)s(6)q(8) + q(2)q(3)q(4)q(6)q(7)s(1)s(3)s(6) \\
& + q(1)q(2)q(4)q(6)q(7)s(2)s(3)s(6)q(8) + q(4)q(6)q(7)s(1)s(2)s(3)s(6)q(8) + q(1)q(4)q(5)q(6)q(7)s(4)s(6)q(8) + q(2)q(4)q(5)q(6)q(7)s(1)s(2)s(4)s(6) \\
& + q(3)q(4)q(5)q(6)q(7)s(1)s(2)s(4)s(6)q(8) + q(1)q(3)q(5)q(6)q(7)s(3)s(4)s(6)q(8) + q(2)q(3)q(5)q(6)q(7)s(1)s(3)s(4)s(6)q(8)
\end{aligned}$$



$$\begin{aligned}
&+q(1)q(2)q(6)q(7)s(2)s(3)s(4)s(5)s(7)+q(6)q(7)s(1)s(2)s(3)s(4)s(5)s(7)+q(1)q(6)s(6)s(7)+q(2)q(6)s(1)s(6)s(7)+q(1)q(2)q(3)q(4)q(5)s(6)s(7) \\
&+q(1)q(3)q(4)q(6)s(3)s(6)s(7)+q(2)q(3)q(4)q(6)s(1)s(3)s(6)s(7)+q(1)q(2)q(4)q(6)s(2)s(3)s(6)s(7)+q(4)q(6)s(1)s(2)s(3)s(5)s(6)s(7) \\
&+q(2)q(4)q(5)q(6)s(1)s(4)s(6)s(7)+q(1)q(2)q(3)q(4)q(5)q(6)s(2)s(4)s(6)s(7)+q(3)q(4)q(5)q(6)s(1)s(2)s(4)s(6)s(7)+q(1)q(2)q(3)q(4)q(5)q(6)s(7) \\
&+q(2)q(3)q(5)q(6)s(1)s(3)s(4)s(6)s(7)+q(1)q(2)q(5)q(6)s(2)s(3)s(4)s(6)s(7)+q(5)q(6)s(1)s(2)s(3)s(4)s(6)s(7)+q(1)q(5)s(5)s(6)s(7) \\
&+q(1)q(2)q(3)q(5)s(2)s(5)s(6)s(7)+q(3)q(5)s(1)s(2)s(5)s(6)s(7)+q(1)q(3)q(4)q(5)s(3)s(5)s(6)s(7)+q(2)q(3)q(4)q(5)s(1)s(2)s(3)s(4)s(5)s(6)s(7) \\
&+q(4)q(5)s(1)s(2)s(3)s(5)s(6)s(7)+q(1)q(4)s(4)s(5)s(6)s(7)+q(2)q(4)s(1)s(4)s(5)s(6)s(7)+q(1)q(2)q(3)q(4)s(2)s(4)s(5)s(6)s(7) \\
&+q(1)q(3)s(3)s(4)s(5)s(6)s(7)+q(2)q(3)s(1)s(3)s(4)s(5)s(6)s(7)+q(1)q(2)s(2)s(3)s(4)s(5)s(6)s(7)+s(1)s(2)s(3)s(4)s(5)s(6)s(7)
\end{aligned}$$

Clearly, the estimate of such determinants is too complicated!



## One-dimensional case, piecewise constant wave speed

Let  $\Omega = (-1, 1)$  and introduce the points

$$-1 = x_0 < x_1 < \dots < x_n = 1$$

and define the intervals  $\tau_i = (x_{i-1}, x_i)$ ,  $1 \leq i \leq n$ . Consider piecewise constant wave speed which is given by

$$c(x) := c_i \text{ for } x \in \tau_i := (x_{i-1}, x_i), \quad 1 \leq i \leq n$$

where  $c_i$  are positive constants.

For a positive wave number  $\omega \in \mathbb{R}_{\geq \omega_0}$  we consider the homogenous Helmholtz equation in the strong form

$$-u'' - \left(\frac{\omega}{c}\right)^2 u = 0 \quad \text{in } \Omega = (-1, 1),$$

$$-u' - i \frac{\omega}{c_1} u = g_1 \quad \text{at } x = -1,$$

$$u' - i \frac{\omega}{c_n} u = g_2 \quad \text{at } x = 1.$$

**Example** (Graham, S. 2016). Let the number  $n$  of subintervals be even. Choose  $\omega = n$  and

$$x_\ell := \begin{cases} -1 + 2^{\frac{\ell-1}{n}} & \ell \text{ odd,} \\ -1 + \frac{2\ell}{n} & \ell \text{ even,} \end{cases}$$

$$c_\ell := \frac{1}{\pi} \times \begin{cases} 1 & \ell \text{ odd,} \\ 3 & \ell \text{ even.} \end{cases}$$

Then, for  $k = 0, 1, 2$ , it holds

$$\|u^{(k)}\|_{L^2(\Omega)} \leq 7n\omega^{k-1} \max\{|g_1|, |g_2|\}.$$

**Theorem** (Graham, S. 2016). For  $k = 2, 4, 6 \dots$  and  $n = 2k + 1$ , let the piecewise constant wave speed be defined as before. There exists a sequence of interval partitionings  $\left(x_i^{(k)}\right)_{i=0}^n$  and wave numbers  $\omega_k = n$  such that

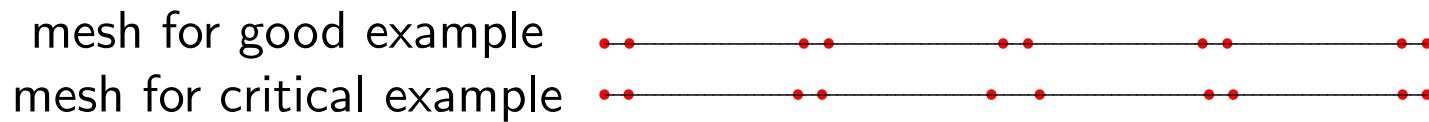
$$\|u'\|_{L^2(\Omega)} \geq c_1 e^{\gamma \omega_k} \min \{|g_1|, |g_2|\}.$$

for some  $c_1, \gamma > 0$ .

## Numerical Experiment

For some  $r \in [0, 1[$  and any  $n = 4k + 1$  let

$$\omega := \frac{\pi}{2} \times \begin{cases} (n - r) & \text{“good example”,} \\ \left(\frac{n+1}{2} - r\right) & \text{“critical example”,} \end{cases} \quad \text{and} \quad c_\ell := \begin{cases} 1 - r & \ell \text{ odd,} \\ 1 + r & \ell \text{ even,} \end{cases}$$



Robin boundary data:  $g_1 = g_2 = 1$ .



