Advances in Computational Mathematics (2024) 50:112
https://doi.org/10.1007/s10444-024-10193-w

Helmholtz FEM solutions are locally quasi-optimal modulo low frequencies

M. Averseng[1] · **J. Galkowski**[2] · **E.A. Spence**[1]

Received: 1 June 2023 / Accepted: 9 August 2024 / Published online: 18 November 2024

Abstract

For h-FEM discretisations of the Helmholtz equation with wavenumber k, we obtain k-explicit analogues of the classic local FEM error bounds of Nitsche and Schatz (Math. Comput. **28**(128), 937–958 1974), Wahlbin (1991, §9), Demlow et al.(Math. Comput. **80**(273), 1–9 2011), showing that these bounds hold with constants independent of k, provided one works in Sobolev norms weighted with k in the natural way. We prove two main results: (i) a bound on the local H^1 error by the best approximation error plus the L^2 error, both on a slightly larger set, and (ii) the bound in (i) but now with the L^2 error replaced by the error in a negative Sobolev norm. The result (i) is valid for shape-regular triangulations, and is the k-explicit analogue of the main result of Demlow et al. (Math. Comput. **80**(273), 1–9 2011). The result (ii) is valid when the mesh is locally quasi-uniform on the scale of the wavelength (i.e., on the scale of k^{-1}) and is the k-explicit analogue of the results of Nitsche and Schatz (Math. Comput. **28**(128), 937–958 1974), Wahlbin (1991, §9). Since our Sobolev spaces are weighted with k in the natural way, the result (ii) indicates that the Helmholtz FEM solution is locally quasi-optimal modulo low frequencies (i.e., frequencies $\lesssim k$). Numerical experiments confirm this property, and also highlight interesting propagation phenomena in the Helmholtz FEM error.

Keywords Finite element method · Helmholtz equation

Mathematics Subject Classification (2010) 35J05 · 65N15 · 65N30 · 78A45

Communicated by: Ilaria Perugia

✉ M. Averseng
M.Averseng@bath.ac.uk

J. Galkowski
J.Galkowski@ucl.ac.uk

E.A. Spence
E.A.Spence@bath.ac.uk

[1] Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

[2] Department of Mathematics, University College London, 25 Gordon Street, London WC1H 0AY, UK

1 Introduction: the main result in a simple setting

1.1 The PML approximation to the Helmholtz exterior Dirichlet problem and its FEM discretisation

Let $\Omega_- \subset \mathbb{R}^d$ be a bounded Lipschitz open set with its open complement $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$ connected. Let $\widetilde{u}$ be the solution of the variable-coefficient exterior Dirichlet problem for the Helmholtz equation

$$-k^{-2}\nabla \cdot \big(A_{\mathrm{scat}}(x)\nabla \widetilde{u}(x)\big) - (c_{\mathrm{scat}}(x))^{-2}\widetilde{u}(x) = g(x) \quad \text{in } \Omega_+, \qquad \widetilde{u} = 0 \quad \text{on } \partial\Omega_+$$

satisfying the Sommerfeld radiation condition, and with the supports of g, $I - A_{\mathrm{scat}}$, and $1 - c_{\mathrm{scat}}$ compact. Let $u \in H^1_0(\Omega)$ be the radial perfectly-matched-layer (PML) approximation to $\widetilde{u}$, where Ω is the truncated domain; i.e., u is the solution to the variational problem

$$\text{find } u \in H^1_0(\Omega) \text{ such that } a(u, v) = G(v) \text{ for all } v \in H^1_0(\Omega), \tag{1.1}$$

where a is the sesquilinear form given by

$$a(u, v) = \int_\Omega k^{-2} A \nabla u \cdot \overline{\nabla v} - c^{-2} u \overline{v}\,,$$

$G(v) = \int_\Omega g\overline{v}$, and the coefficients A and c are defined in §7.1.5 in terms of the PML scaling function and (respectively) A_{scat} and c_{scat}.

We consider the Galerkin discretisation of (1.1) using the standard conforming Lagrange finite-element spaces $\{V_h\}_{h>0}$ of continuous piecewise polynomials of degree p on a family of shape-regular triangulations $(\mathcal{T}_h)_{h>0}$ of Ω; i.e.,

$$\text{find } u_h \in V_h \text{ such that } a(u_h, v_h) = G(v_h) \quad \text{for all } v_h \in V_h. \tag{1.2}$$

Subtracting (1.2) from (1.1) we find that Galerkin orthogonality holds:

$$a(u - u_h, v_h) = 0 \quad \text{for all } v_h \in V_h. \tag{1.3}$$

1.2 First result: bound on the local FEM error in H^1_k with an L^2 error term

Given two subsets $\Omega_0 \subset \Omega_1 \subset \Omega$, let

$$\partial_<(\Omega_0, \Omega_1) := \mathrm{dist}\big(\partial\Omega_0 \setminus \partial\Omega, \partial\Omega_1 \setminus \partial\Omega\big), \tag{1.4}$$

with the convention $\partial_<(\Omega_0, \Omega_1) = +\infty$ when $\Omega_1 = \Omega$; see Fig. 1. Working with this notion of distance allows us to consider subdomains that go up to the boundary.

The Sobolev norms $\|\cdot\|_{H^s_k(D)}$ for D a bounded Lipschitz domain are defined as for $\|\cdot\|_{H^s(D)}$ (via restriction of $\|\cdot\|_{H^s(\mathbb{R}^d)}$ to D, with this second norm defined by the Fourier transform), except that now we weight the jth derivative with k^{-j}; see §8.

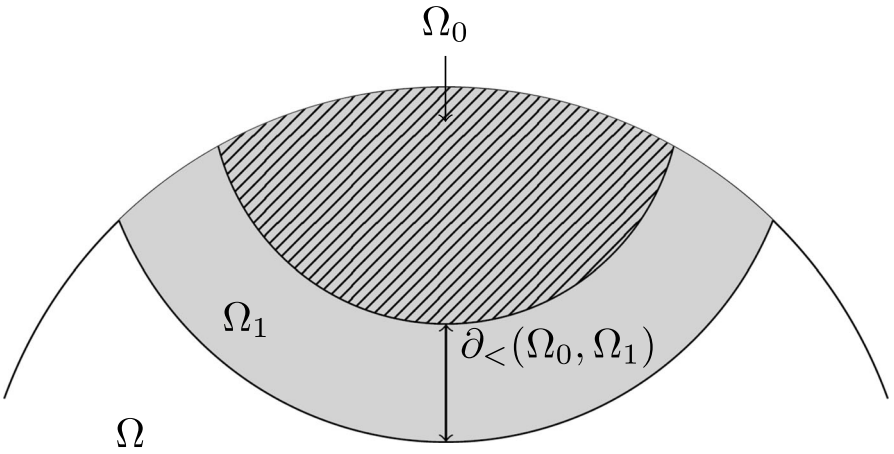


Fig. 1 Illustration of the distance $\partial_<(\Omega_0, \Omega_1)$ defined by (1.4), with Ω_0 hatched and Ω_1 shaded

Theorem 1.1 (Local quasioptimality in H^1_k up to an L^2 error term) *Suppose that A and c are $L^\infty(\Omega)$ and the PML scaling function f_θ (defined in §7.1.5) is $W^{1,\infty}(\Omega)$. Given $C_0 > 0$ there exists $C_1, C_* > 0$ such that the following is true. Let $\Omega_0 \subset \Omega_1 \subset \Omega$ be such that $\Omega_0 \neq \Omega_1$,*

$$d := \partial_<(\Omega_0, \Omega_1) \geq \frac{C_0}{k}, \quad \text{and} \quad \max_{K \cap \Omega_1 \neq \emptyset} h_K \leq \frac{C_1}{k}. \tag{1.5}$$

Given $k > 0$, let $u \in H^1_0(\Omega)$ and $u_h \in V_h$ satisfy the Galerkin orthogonality (1.3). Then,

$$\|u - u_h\|_{H^1_k(\Omega_0)} \leq C_* \Big(\min_{w_h \in V_h} \|u - w_h\|_{H^1_k(\Omega_1)} + \|u - u_h\|_{L^2(\Omega_1)} \Big). \tag{1.6}$$

We make two remarks (which also hold for Theorem 1.2 below).

1. Recall that, for standard finite-element spaces with p fixed, $h_K k$ must be chosen as a decreasing function of k to maintain accuracy (see [14] and the references therein); thus the second condition in (1.5) is not restrictive.
2. The assumption that u and u_h satisfy the Galerkin orthogonality (1.3) can be weakened to u and u_h satisfying a local version of Galerkin orthogonality on Ω_1; see (4.2) below.

1.3 Second result: bound on the local FEM error in H^1_k with an error term in a negative norm

Theorem 1.1 holds for all shape-regular meshes $\mathcal{T}_h$ (i.e., the mesh elements are uniformly "well-shaped"), and thus the mesh can in principle be highly non-uniform. However, if the mesh $\mathcal{T}_h$ is *locally quasi-uniform on a scale of* k^{-1} (i.e., in every ball of radius proportional to k^{-1}, the mesh elements diameters are comparable) and the

PDE coefficients and boundary of the domain have sufficient regularity, then the L^2 norm of the error on the right-hand side of (1.6) can be replaced by a negative Sobolev norm.

Theorem 1.2 (Local quasioptimality in H^1_k up to an error term in a negative norm) *Suppose that, for some $\ell \in \mathbb{Z}^+$, A and c are $C^{\ell,1}(\overline{\Omega})$ (i.e., their ℓth derivatives are Lipschitz), the PML scaling function f_θ (defined in §7.1.5) is $C^{\ell+1,1}(\overline{\Omega})$, and $\partial\Omega$ is $C^{\ell+1,1}$. Given $C_0, C_{\mathrm{qu}} > 0$ there exists $C_1, C_2, C_* > 0$ such that the following is true. Let $\Omega_0 \subset \Omega_1 \subset \Omega$ be such that $\Omega_0 \neq \Omega_1$ and (1.5) holds. Assume further that $\mathcal{T}$ is* quasi-uniform on scale k^{-1}*, in the sense that, for every ball B of radius at most $C_2 k^{-1}$,*

$$\frac{\max_{K\cap B\neq\emptyset} h_K}{\min_{K\cap B\neq\emptyset} h_K} \leq C_{\mathrm{qu}}. \tag{1.7}$$

Given $k > 0$, let $u \in H^1_0(\Omega)$ and $u_h \in V_h$ satisfy (1.3). Then,

$$\|u-u_h\|_{H^1_k(\Omega_0)} \leq C_*\Big(\min_{w_h\in V_h} \|u-w_h\|_{H^1_k(\Omega_1)} + \|u-u_h\|_{(H_k^{\min\{\ell+1,p\},<}(\Omega_1))^*}\Big). \tag{1.8}$$

Note that meshes satisfying (1.7) can be highly non-uniform on Ω; indeed, the ratio between the largest and smallest mesh elements on an $\mathcal{O}(1)$ scale can be proportional to $\exp(\alpha k)$ for some $\alpha \in \mathbb{R}$ (depending on C_{qu}).

We now define the norm $\|\cdot\|_{(H_k^{s,<}(\Omega_1))^*}$ appearing in the last term on the right-hand side of (1.8), but highlight that when Ω_1 is an interior subset of Ω, $\|\cdot\|_{(H_k^{s,<}(\Omega_1))^*}$ is equivalent to $\|\cdot\|_{H_k^{-s}(\Omega_1)}$. For $s \geq 0$ and $D \subset \Omega$ Lipschitz, let

$$H^{s,<}(D) := \overline{\Big\{ v\in H^1_0(\Omega) \,:\, v|_D \in H^s(D),\ \operatorname{supp} v \subset \overline{D},\ \partial_<(\operatorname{supp} v, D) > 0\Big\}}$$

(where the closure is taken with respect to the H^s norm) and

$$\|v\|_{(H_k^{s,<}(D))^*} := \sup_{w\in H^{s,<}(D),\ \|w\|_{H^s_k(D)}=1} \big|v(w)\big|.$$

When $\partial D \cap \partial\Omega = \emptyset$ (i.e., D is an interior subset of Ω), $\|\cdot\|_{(H_k^{s,<}(D))^*}$ is equivalent to $\|\cdot\|_{H_k^{-s}(D)}$ by (8.4) (see also [29, Equation 9.18]). When $\partial D \cap \partial\Omega \neq \emptyset$,

$$H^s_0(D) \subset H^{s,<}(D) \subset H^s(D) \cap H^1_0(D)$$

(where $H^s_0(D)$ is the closure of $C^\infty_{\mathrm{comp}}(D)$ in $H^s(D)$), and so

$$\|v\|_{H_k^{-s}(D)} \leq \|v\|_{(H_k^{s,<}(D))^*} \leq C \max\left\{ \|v\|_{H_k^{-1}(D)}, \|v\|_{\widetilde{H}_k^{-s}(D)} \right\}.$$

To informally understand the differences between these norms, we note that $\|\cdot\|_{H_k^{-s}(D)}$ doesn't "see" the boundary of D, $\|v\|_{\widetilde{H}_k^{-s}(D)}$ sees the boundary of D, and $\|v\|_{(H_k^{s,<}(D))^*}$ sees only the parts of ∂D that coincide with $\partial\Omega$.

1.4 The relationship of Theorems 1.1 and 1.2 to other results in the literature

Estimates on the local FEM error for second-order linear elliptic PDEs were pioneered by Nitsche and Schatz in [26]; see also [10, 27, 28], [29, Chapter 9], and [9]. These arguments use that the sesquilinear forms of second-order linear elliptic PDEs are coercive (i.e., sign definite) on sufficiently small balls. This property is used to prove, again on small balls, a discrete analogue of the classic Caccioppoli estimate (bounding the H^1 norm of the PDE solution in terms of the L^2 norm and the data on a slightly larger set). The Caccioppoli estimate is the main ingredient required to prove a bound of the form (1.6) on small balls. A covering argument is then used to obtain the bound on an arbitrary domain from the bound on small balls. These arguments combined with a duality argument and elliptic regularity then produce a bound of the form (1.8).

Although these classic results apply to the Helmholtz equation, they don't use norms weighted with k and the constants in the bounds are not explicit in k. The main motivation for the present paper was to obtain the analogues of the results in [26], [29, Chapter 9], and [9] applied to the Helmholtz equation in k-weighted norms, and with constants explicit in k. Roughly speaking, we show that the results of [26], [29, Chapter 9], and [9] hold for the Helmholtz equation with constants independent of k, provided that one works in k-weighted norms. In more detail,

- Theorem 1.1 is a k-explicit version of [9, Theorem 3.4], with both proved under the assumption that the mesh is shape-regular. This result (in non k-explicit form) for quasi-uniform meshes appears as [10, Theorem 4], [28, Theorem 4.1], and [29, Theorem 9.1].
- Theorem 1.2 is, roughly speaking, a k-explicit version of [26, Theorem 5.1(i)] and [29, Theorem 9.2]. The differences are
  - in [26, Theorem 5.1(i)] the best approximation error on Ω_1 (i.e., the first term on the right-hand sides of (1.6) and (1.8)) is estimated by the standard polynomial approximation result ((3.24) below) and the subdomains are assumed not to touch the boundary, and
  - the results in [26] and [29, Chapter 9] are geared toward quasi-uniform meshes (see [26, Assumption A3], [29, Equation 9.6]) whereas Theorem 1.2 requires quasi-uniformity only on the scale of the wavelength (i.e., on a scale of k^{-1}).

An additional difference between the results of the present paper and existing results is that we cover Helmholtz transmission problems, i.e., those with discontinuous A_{scat} and c_{scat}, and the analogue of Theorem 1.2 in this context appears to be new (independent of the k-explicitness); indeed, [10, 26] consider second-order linear elliptic PDEs with smooth coefficients and [28, 29] cover Poisson's equation.

1.5 Interpreting the results of Theorems 1.1 and 1.2

The standard interpretation of the non-k-explicit versions of the bounds (1.6) and (1.8) is that the FEM solution is locally quasi-optimal up to a lower-order term that allows error to propagate into Ω_0 from the rest of the domain. (This second term is sometimes called the "pollution" or "slush" term in the literature; later in the paper, we refer to

it as the "slush" term to avoid confusion with the pollution effect for the Helmholtz h-FEM.)

The fact that (1.6) and (1.8) are proved in k-weighted norms with k-independent constants allows this interpretation to be refined in the Helmholtz context to *Helmholtz FEM solutions are locally quasi-optimal modulo low frequencies*, where "low frequencies" here means "frequencies $\leq Ck$ for some $C > 1$".

We now show (albeit heuristically) how this property can be inferred from the bounds (1.6) (1.8), with this property illustrated by numerical experiments in §2.

The key point is that a bound on a high k-weighted Sobolev norm of a function in terms of a low k-weighted Sobolev norm, with the constant independent of k, implies that the function is controlled by its frequencies $\lesssim k$. This is illustrated by the following simple lemma. This lemma uses the k-weighted Fourier transform $\mathcal{F}_k$, defined by (8.1) below, with Fourier variable ξ. The weighting by k implies that "low frequencies" are now $\{\xi : |\xi| \leq C\}$ for some $C > 1$.

Lemma 1.3 (Bound on high Sobolev norm by low Sobolev norm implies function controlled by its low frequencies) *Suppose a family of functions $(f(k))_{k>0}$ is such that $f(k) \in H^1_k(\mathbb{R}^d)$ and there exists $C_2 > 0$ such that given $s > 0$ there exists $C_1 > 0$ such that*

$$\|f\|^2_{H^1_k(\mathbb{R}^d)} \leq C_1\Big(C_2 + \|f\|^2_{H^{-s}_k(\mathbb{R}^d)}\Big) \text{ for all } k \geq k_0. \tag{1.9}$$

If

$$\frac{C_1}{\langle R\rangle^{2(s+1)}} \leq \frac{1}{2}, \tag{1.10}$$

then, for all $k \geq k_0$,

$$\left(\frac{k}{2\pi}\right)^d \int_{\mathbb{R}^d} \langle\xi\rangle^2 |\mathcal{F}_k f(\xi)|^2 \,\mathrm{d}\xi \leq 2C_1\left(C_2 + \left(\frac{k}{2\pi}\right)^d \int_{|\xi|\leq R} \langle\xi\rangle^{-2s} |\mathcal{F}_k f(\xi)|^2 \,\mathrm{d}\xi\right).$$

In the "ideal" situation that $C_1 = \langle C\rangle^{2(s+1)}$, (1.10) can be satisfied by taking $R > C$ and s large; we call this the "ideal" situation, since (1.9) holds with this value of C_1 (and $C_2 = 0$) when $|\mathcal{F}_k f(\xi)| = 0$ for $|\xi| \geq C$.

Proof *of Lemma 1.3* By the definition of $\|\cdot\|_{H^s_k(\mathbb{R}^d)}$ (8.2), (1.9) implies that

$$\begin{aligned}&\int_{\mathbb{R}^d} \langle\xi\rangle^2 |\mathcal{F}_k f(\xi)|^2 \,\mathrm{d}\xi \\ &\quad\leq C_1\left(\left(\frac{2\pi}{k}\right)^d C_2 + \int_{|\xi|\leq R} \langle\xi\rangle^{-2s} |\mathcal{F}_k f(\xi)|^2 \,\mathrm{d}\xi + \frac{1}{\langle R\rangle^{2(s+1)}} \int_{|\xi|\geq R} \langle\xi\rangle^2 |\mathcal{F}_k f(\xi)|^2 \,\mathrm{d}\xi\right),\end{aligned}$$

and the result follows. □

The bound (1.8) is conceptually similar to the setting of Lemma 1.3 of a high Sobolev norm of $u - u_h$ being bounded by one of its low Sobolev norms, except that (i) the norm on the right-hand side is over a slightly larger domain (Ω_1 vs Ω_0) (ii) the order of the negative norm (i.e., s in (1.9)) cannot be arbitrarily large.

Nevertheless, we expect from (1.8) that $u - u_h$ is controlled locally by the local best approximation error and the low frequencies of $u - u_h$; i.e., the Galerkin solution is locally quasi-optimal, modulo low frequencies.

1.6 Outline of the paper

Section 2 contains numerical experiments illustrating local quasi-optimality modulo low frequencies of Helmholtz FEM solutions.

Section 3 describes a general framework, with Sect. 7 then showing how a variety of Helmholtz problems fit in this framework (including the exterior Dirichlet problem and transmission problem discussed above).

Section 4 states the main results applied to the general framework, with Theorem 4.1 the generalisation of Theorems 1.1 and 4.2 the generalisation of Theorem 1.2

Section 6 proves Theorems 1.1 and 1.2, with Sect. 5 proving auxiliary results (Caccioppoli estimates) used in the proofs.

The appendix (§1) recaps the definitions of Sobolev spaces weighted by k.

2 Numerical experiments illustrating local quasi-optimality modulo low frequencies of Helmholtz FEM solutions

This section presents numerical experiments where the error in the Helmholtz FEM solution behaves differently in different subsets of the domain due to either

1. a non-uniform mesh, see §2.1, or
2. the solution being zero in some part of the domain (so that the local best approximation error is immediately zero in this part of the domain), see §2.2.

These situations are manufactured so that each of the two terms on the right-hand side of (1.8) is dominant in a different subset of the domain.

Our ultimate aim is to study the local behaviour of the FEM error for scattering problems. However, a proper understanding of this situation requires understanding the k-dependence of the two terms on the right-hand side of (1.8) as a function of the position of Ω_1 relative to the scatterer; this is a work in progress and will be reported elsewhere.

In all the experiments, the Galerkin equations (1.2) are formulated with the software FreeFem++ [18] using continuous Lagrange elements of degree $p = 1, \ldots, 4$. The resulting linear systems are then solved using the parallel domain decomposition toolbox HPDDM. The code used to produce these numerical results is available at https://github.com/MartinAverseng/local_qo_experiments.

In the experiments, we compute "low" and "high" frequency components of the FEM error; how we do this is described in §2.3 below, with the "low" and "high" frequencies corresponding to the components of the solution with the absolute value of the (unweighted) Fourier variable $\leq 2k$ and $\geq 2k$, respectively.

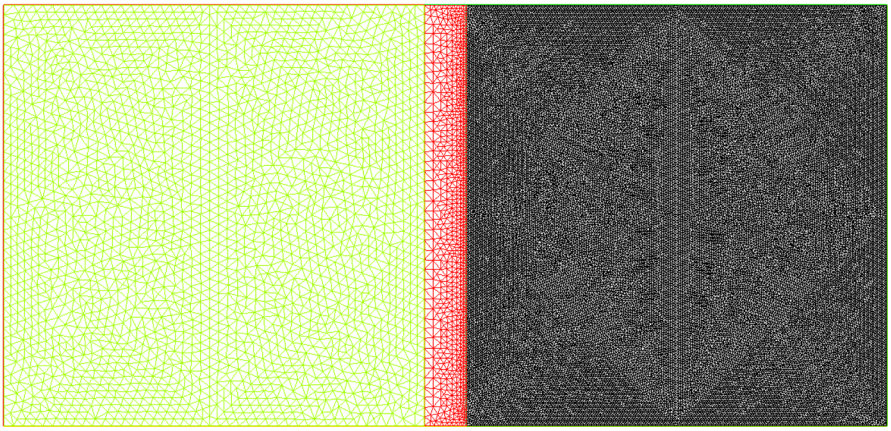


Fig. 2 A mesh with two square regions S_1 and S_2 each with a uniform mesh size (different to each other)

2.1 Experiments with a non-uniform mesh

We solve the interior impedance problem in a rectangular domain $\Omega = [0, 2.1] \times [0, 1]$, i.e.,

$$(k^{-2}\Delta + 1)u = 0 \quad \text{in } \Omega \quad \text{and} \quad k^{-1}\partial_n u - \mathrm{i}u = g \quad \text{on } \partial\Omega,$$

with data g is chosen so that the exact solution is the plane wave $u = \exp(\mathrm{i}k(\cos(\theta)x + \sin(\theta)y))$.

This problem falls into the class of Helmholtz problems described in §7.1, with the sesquilinear form given by (7.8) with $A_{\text{scat}} \equiv I$, $c_{\text{scat}} \equiv 1$, $\Omega_- = \emptyset$ (no impenetrable obstacle), $\Omega_{\text{p}} = \emptyset$ (no penetrable obstacle), and $\Omega_{\text{tr}} = \Omega$.

Experiment 1. We consider two different mesh sizes $h_1 > h_2$ and consider the following three different meshes on Ω:

- a globally uniform mesh of size h_1,
- a globally uniform mesh of size h_2,
- a mesh with sizes h_1 in the left-hand square $[0, 1] \times [0, 1]$, h_2 in the right-hand square $[1.1, 2.1] \times [0, 1]$, and some non-constant mesh size in the transition region $[1, 1.1] \times [0, 1]$.[1]

Figure 2 shows an example of the third type of mesh.

Figures 3 and 4 plot the FEM error $u - u_h$ for all three meshes with $k = 50$ (i.e., the wavelength ≈ 0.13), $p = 4$, $h_1 = 1.78k^{-1}$, $h_2 = 0.38k^{-1}$, $\theta = 0$ in Fig. 3, and $\theta = \pi/2$ in Fig. 4. The errors are plotted on a logarithmic scale, with the scale kept the same for all the plots.

[1] More specifically, the mesh is designed by first meshing the boundaries of the two squares and transition region, which is partitioned into a total of 10 segments. On the 4 segments bounding the left-hand (respectively right-hand) square, the mesh size is uniform equal to h_1 (respectively h_2), and on the two segments on top and bottom of the transition region, the mesh size is constant equal to $\sqrt{h_1 h_2}$. The command `buildmesh` from FreeFem++ then generates a mesh for the global domain respecting the boundary mesh. As a result, the mesh size is uniform equal to h_1 in the left-hand square, h_2 in the right-hand square, and non-uniform in the transition region.

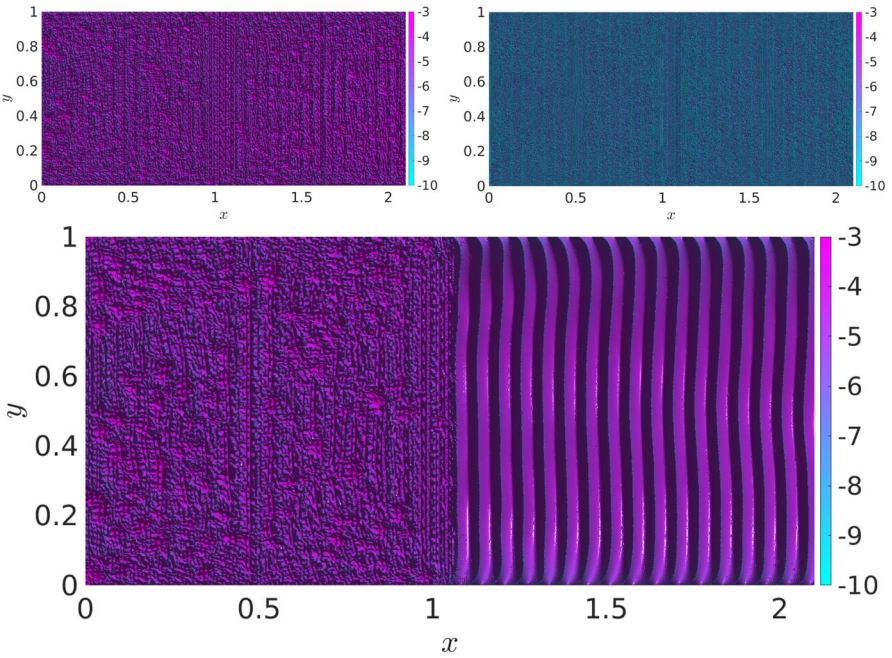


Fig. 3 For Experiment 1 in §2.1, plot of the quantity $\log(10^{-12} + |\Re(u - u_h)|)/\log(10)$, for $k = 50$, $p = 4$ and $\theta = 0$. Top left: globally uniform mesh with $h = 1.78k^{-1}$. Top right: globally uniform mesh with $h = 0.38k^{-1}$. Bottom: non-uniform mesh, with a coarse region $h_1 = 1.78k^{-1}$ and a fine region with $h_2 = 0.38k^{-1}$

In both Figs. 3 and 4, for the third mesh the error in the right-hand square (with mesh width h_2) is between 10 to 100 times larger than the error in the right-hand square for the second (globally h_2) mesh; i.e., in the right-hand square for the third mesh, the error is dominated by the "slush" term on the right-hand side of (1.8) and not the local best-approximation error. Furthermore, by eye, this "slush" error is low frequency.

The difference between Figs. 3 and 4 is that in the first, the exact solution propagates from left to right, whereas in the second the exact solution propagates upwards. These figures show that the error is affected by the direction of propagation of the solution (or more precisely, its localisation in Fourier space), but does not "inherit" these properties; indeed, in Fig. 4, the error still propagates from left to right, even though the exact solution propagates upwards.

Experiment 2 This experiments considers the non-uniform mesh from Experiment 1 with h_1 and h_2 now chosen to decrease with k at different rates, with the FEM error computed at a sequence of values of k. We do this so that the terms on the right-hand side of (1.8) have different k-dependence in the left and right squares.

We take $p = 2$, $h_1 = \sqrt{2}/k$, $h_2 = 1/k^{3/2}$, and $k \in [5, 60]$, and Figs. 5 and 6 plot the H^1_k norm of the error on the left and right, respectively, as well as their high- and low-frequency components (see §2.3 below for how these components are computed).

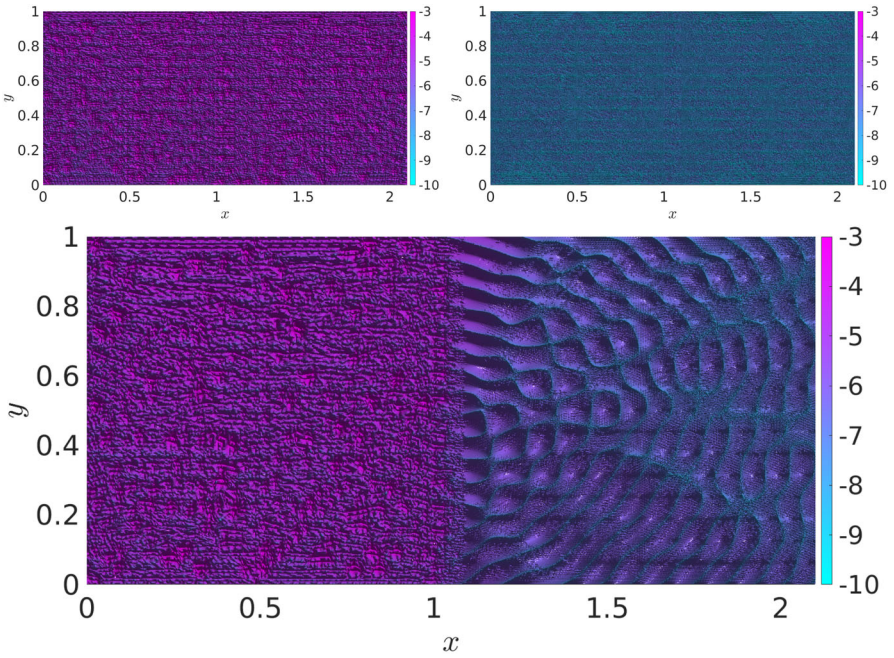


Fig. 4 Same as Fig. 3 (including all parameter values) but with $\theta = \pi/2$

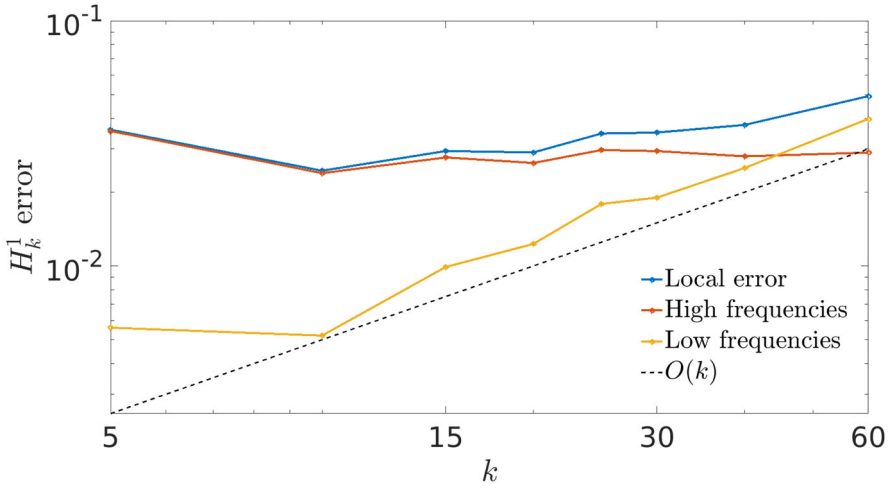


Fig. 5 For Experiment 2 in §2.1, plots of the H^1_k error and its low- and high-frequency components in the left square (i.e., the square with the coarser mesh)

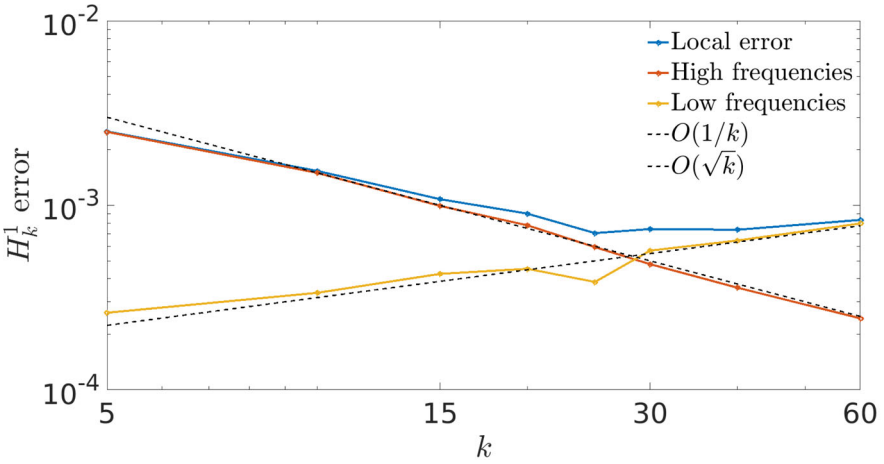


Fig. 6 For Experiment 2 in §2.1, plots of the H^1_k error and its low- and high-frequency components in the right square (i.e., the square with the finer mesh)

The key point is that, *on both the left and right, the FEM solution is locally quasi-optimal, modulo low frequencies, with the low frequencies on the left caused by the pollution effect, and the low frequencies on the right coming from the "slush" term propagating from the left.*

In more detail: in Fig. 5, we see that

(i) the error in the left square grows with k,

(ii) the high frequencies of the error in the left square are roughly constant in k, and

(iii) the growth in the error is caused by the low-frequency components, which are proportional to k.

Points (i) and (iii) are expected from [20, Corollary 3.2 and Point 1 in the following discussion] which proves that

$$\|u - u_h\|_{H^1_k} \sim k(hk)^{2p}\|u\|_{H^1_k} \sim k,$$

for the 1-d impedance problem with hk sufficiently small. Point (ii) is expected because the local best approximation error on the left $\sim (hk)^p\|u\|_{H^{p+1}} \sim (hk)^2 \sim 1$; i.e., modulo low frequencies the FEM solution is locally quasi-optimal.

In Fig. 6, we see that

(iv) The high frequencies of the error in the right square decrease like k^{-1}.

(v) The low frequencies of the error in the right square grow like $k^{1/2}$.

Point (iv) is expected since the local best approximation error on the right $\sim (hk)^p\|u\|_{H^{p+1}} \sim (hk)^2 \sim k^{-1}$.

A heuristic argument for Point (v) is the following. Let χ be a compactly supported cutoff function supported on the left square, and let $e := \chi(u - u_h)$. For sufficiently large k, Points (i) to (iii) above show that the low frequencies dominate, and thus we assume that, first, $\mathcal{F}_k e(\xi)$ is negligible except for $|\xi| \le C$ for some constant $C > 1$ and, second, $\mathcal{F}_k e(\xi)$ is approximately evenly distributed in $\{\xi : |\xi| \le C\}$. Since the

local H^1_k error on the left $\sim k$,

$$k^d \int_{|\xi|\leq C} \langle \xi \rangle^2 |\mathcal{F}_k e(\xi)|^2 \, \mathrm{d}\xi \sim k^2$$

which, under the assumptions, implies that $|\mathcal{F}_k e(\xi)| \sim k^{1-d/2}$ uniformly for $|\xi| \leq C$. Within $\{\xi : |\xi| \leq C\}$, only the components in $\{\xi : 1-\eta/k \leq |\xi| \leq 1+\eta/k\}$ propagate, where $\eta = O(1)$; this is because the Helmholtz operator is (semiclassically) elliptic away from frequency k (see, e.g., the discussion in [13, §1.8]). Therefore, the squared H^1_k norm of the error propagated to the right square is approximately

$$k^d \int_{1-\eta/k\leq|\xi|\leq 1+\eta/k} \langle \xi \rangle^2 |\mathcal{F}_k e(\xi)|^2 \, \mathrm{d}\xi \sim k^2 \int_{1-\eta/k\leq|\xi|\leq 1+\eta/k} \langle \xi \rangle^2 \, \mathrm{d}\xi \sim k^2 \frac{\eta}{k} \sim k;$$

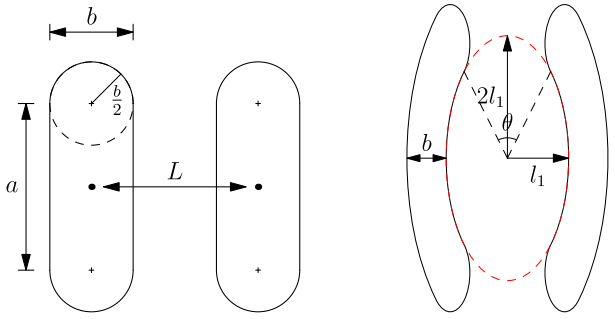
i.e., Point (v).

2.2 Experiments with an artificial source term

Geometry of the obstacles considered in the simulations We work with the Helmholtz problem in §1.1, i.e., the exterior Dirichlet problem, with $A_{\text{scat}} \equiv I$ and $c_{\text{scat}} \equiv 1$. We consider the following four obstacles Ω_-.

- No obstacle, i.e., $\Omega_- = \emptyset$.
- Two flat mirrors, as shown in Fig. 7a.
- One flat mirror, i.e., the obstacle in Fig. 7a with the left mirror removed.
- Two curved mirrors with an inscribed ellipse, as shown in Fig. 7b.

In all experiments, the computational domain is a disk of radius 1.5, and the PML region is the annulus $\{1 \leq r \leq 1.5\}$. For the one flat mirror, $a = 0.4$, $b = 0.2$, and $L = 0.6$. For the two flat mirrors, $a = 0.6$, $b = 0.2$, and $L = 0.8$. For the two curved mirrors, $l_1 = 0.265$, $l_2 = 2l_1$, $b = 0.17$, and $\theta = \arctan(2\tan(\pi/3))$.



(a) Flat mirrors (b) Elliptic mirrors

Fig. 7 Flat mirror and elliptic mirror geometries

Quasi-resonant wavenumbers for the trapping obstacles Whereas there is no trapping for the case of no obstacle or one flat mirror, the two flat mirrors and two curved mirrors trap rays. For these latter two geometries, the norm of the solution operator grows faster through an increasing sequence of ks (often called "quasi resonances") than the nontrapping solution operator.

The experiments below for the two flat mirrors and two curved mirrors are conducted at quasi resonances for these obstacles.

For the flat mirrors, the quasi resonances are

$$k_n := n\frac{\pi}{L-b}, \quad n = 1, 2, \ldots.$$

Indeed, for each of those frequencies, one can manufacture a "quasi-mode" of the Laplace operator, in the form $u_n(x, y) = \chi(y)\psi(x)\sin(k_n x)$ where χ is a cutoff function supported in $(-a/2, a/2)$, and ψ is a cutoff function with $\psi \equiv 1$ on $[-L/2 + b/2, L/2 - b/2]$ and supported in $[-L - b/2, L + b/2]$.

For the curved mirrors, the Helmholtz solution operator grows exponentially through the square roots of eigenvalues of the Laplace operator with Dirichlet conditions in the domain U equal to the ellipse inscribed between the mirrors. By separation of variables, these quasi resonances can be expressed as zeros of some special Mathieu functions (see [23, Appendix E]) and computed accordingly (see, e.g.,[30] and associated Matlab toolbox).

The artificial source term. Let $g := k^{-2}\Delta u + u$, where

$$u(x, y) := \chi(x)\chi(y)\exp(\mathrm{i}kx) \quad \text{and} \quad \chi(x) = \begin{cases} 0 & \text{for } |x| \geq 0.1 \\ \exp\left(\frac{5x^2}{x^2-0.01}\right) & \text{otherwise.} \end{cases}$$

Observe that u is supported in $\Omega_\square := [-0.1, 0.1]^2$.

The FEM error Figure 8 plots the FEM error (on a logarithmic scale) with $hk = 1$ and $p = 4$. For the two nontrapping obstacles (i.e., no obstacle and the one flat mirror) $k = 100$. For the two trapping obstacles k is taken to be the closest quasi resonance to 100, namely $k = 104.72$ for the two flat mirrors and $k = 95.838$ for the two curved mirrors.

All four of the figures show a large high-frequency error on the support of u (i.e., on $\Omega_\square$) and a small low-frequency error away from the support of u. Since the best approximation error is zero away from the support of u, this is consistent with the claim that Helmholtz FEM solutions are quasi-optimal modulo low frequencies.

The four figures also show that, whereas the high-frequency error in $\Omega_\square$ is roughly the same in all four cases, the low-frequency error is affected by the shape of the obstacles. Indeed, for the two trapping obstacles the "slush" is larger inside the trapping regions than elsewhere, and for the three non-empty obstacles the "slush" is smaller in the shadow regions.

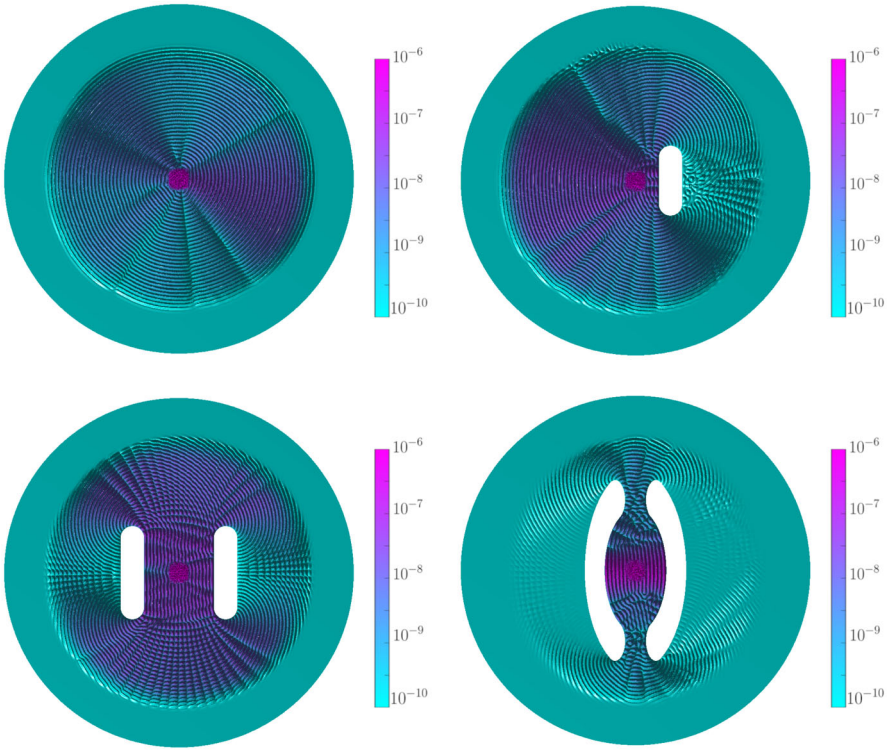


Fig. 8 For the experiment in §2.2 with $k \approx 100$, $hk = 1$ and $p = 4$, the plots show $\log(|\Re(u - u_h)|)/\log(10)$ (the rationale for plotting the real part and not the absolute value is to give a sense of the wave-length). Plots obtained with FFMATLIB

To quantify the high- and low-frequency behaviour more precisely, for each obstacle, we consider two boxes: $\Omega_\square$ and a location away from the source (described for each obstacle in the caption of Table 1). In each box, we let $v_h = \chi(u - u_h)$, where χ is a cut-off function compactly supported in each box, and we compute the high- and low-frequency components of v_h as described in §2.3 below.

For the "one mirror" obstacle, v_h in the two different boxes is plotted in Fig. 9 for $k = 50$. Figure 10 plots the Discrete Fourier Transform (DFT) of v_h in base 10 log scale; these plots confirm that the error away from $\Omega_\square$ (in the left plot) is dominated by low frequencies, compared to the error in $\Omega_\square$ (in the right plot).

Tables 1 and 2 plot the quantity ρ defined by (2.1) below, which measures the proportion of the high-frequencies of v_h. As expected, the values of ρ are much smaller in the "away" region than in $\Omega_\square$.

2.3 Computing the high- and low-frequency components

To compute the high- and low-frequency components of a finite-element function v_h restricted to a square, we compute the 2-dimensional discrete Fourier transform of the

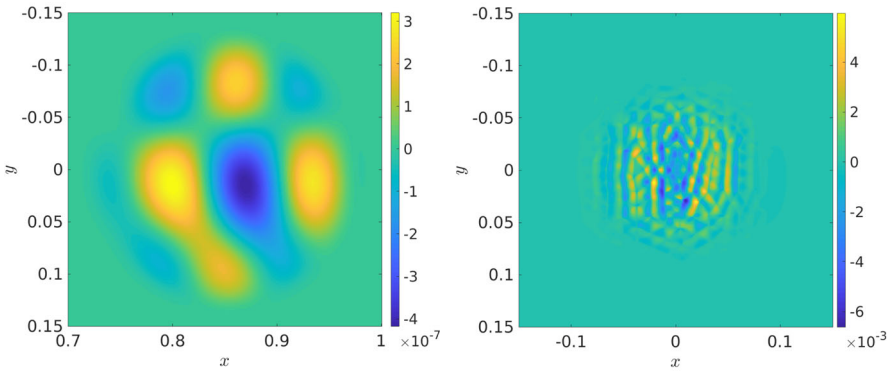


Fig. 9 The signal V corresponding to the scattering by "one wall" (see Fig. 8, top right) with $k = 50$ and $\Delta = 10^{-3}$. Left: $x_0 = 0.7$, $y_0 = -0.15$ (thus, the box is on the right of the wall, and the error is "pure slush." Right: $x_0 = y_0 = -0.15$ (thus, the box contains $\Omega_\square$)

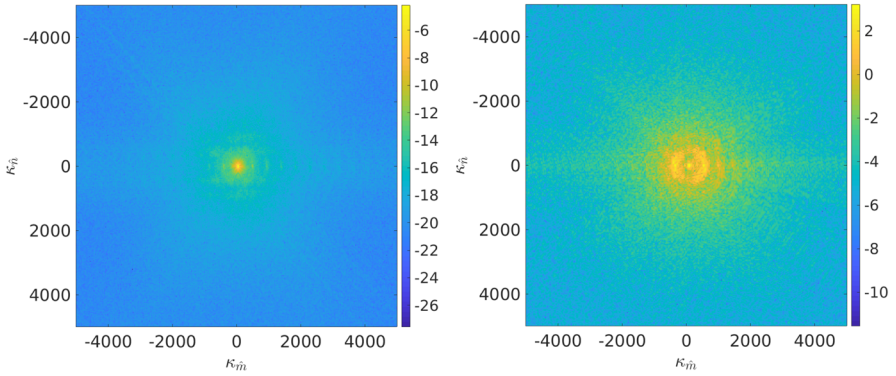


Fig. 10 Graph of $\log(|\widehat{V}|)$, where in each figure, $\widehat{V}$ is the Discrete Fourier Transform of the corresponding V in Fig. 9. This confirms that error away from $\Omega_\square$ (left) is dominated by low frequencies, compared to the error in $\Omega_\square$ (right)

Table 1 Values of ρ (defined by (2.1)) in experiments a)-d) for a box away from $\Omega_\square$

Experiment	$k = 50$	$k = 75$	$k = 100$
a) no obstacle	0.0593	0.0165	0.0091
b) one flat mirror	0.0625	0.0183	0.0103
c) two flat mirrors	0.0471	0.0251	0.0122
d) two curved mirrors	0.0460	0.0236	0.0186

For the experiments a)-b), the box is defined by $x_0 = 0.7$, $y_0 = -0.15$. In c) and d), we use $x_0 = -0.15$ and $y_0 = 0.2$ (thus the boxes are inside the cavities). We take $\Delta = (20k)^{-1}$. In all cases, $N\Delta = 0.3$ (i.e., the box sides are 0.3). The polynomial degree of the FEM is set to $p = 2$ in these experiments

Table 2 Values of ρ (defined by (2.1)) in experiments a)-d) for the box $[-0.15, 0.15]^2 \supset \Omega_\square$

Experiment	$k = 50$	$k = 75$	$k = 100$
a) no obstacle	0.9379	0.8421	0.8142
b) one flat mirror	0.9375	0.8700	0.8001
c) two flat mirrors	0.9342	0.8726	0.8043
d) two curved mirrors	0.9532	0.9177	0.5817

matrix V defined by

$$V_{m,n} = v_h(x_m, y_n)\,, \quad 0 \le m, n \le N-1$$

where $x_m = x_0 + m\Delta$, $y_n = y_0 + n\Delta$ and the sampling rate Δ^{-1} is chosen sufficiently large (in our tests, we use $\Delta = 2\pi/(200k)$ and compare with $\Delta = 2\pi/(400k)$ to confirm that the results are not very sensitive to this value). This interpolation from a triangular mesh to a Cartesian grid is performed with the help of the FFMATLIB toolbox.[2]

Let $\widehat{V}$ be the discrete Fourier Transform of V, that is

$$\widehat{V}_{\widehat{m},\widehat{n}} = \sum_{m=0}^{N}\sum_{n=0}^{N} V_{m,n} \mathrm{e}^{-\frac{2\mathrm{i}\pi}{N}(\widehat{m}m+\widehat{n}n)}\,,$$

so that

$$v_h(x_m, y_n) = \frac{1}{N^2}\sum_{\widehat{m}=1}^{N}\sum_{\widehat{n}=1}^{N} \widehat{W}_{\widehat{m},\widehat{n}} \mathrm{e}^{\mathrm{i}\kappa_{\widehat{m}}x_m} \mathrm{e}^{\mathrm{i}\kappa_{\widehat{n}}y_n}\,, \quad 0 \le m, n \le N-1\,,$$

with

$$\widehat{W}_{\widehat{m},\widehat{n}} = \widehat{V}_{\widehat{m},\widehat{n}} \mathrm{e}^{-\mathrm{i}(\kappa_{\widehat{m}}x_0+\kappa_{\widehat{n}}y_0)}\,, \quad \kappa_{\widehat{m}} = \frac{2\pi}{\Delta}\frac{\widehat{m}}{N}\,.$$

That v_h is low frequency means that one can represent it accurately as a linear combination of waves of the form $\mathrm{e}^{\mathrm{i}\kappa_1 x}\mathrm{e}^{\mathrm{i}\kappa_2 y}$ with $\kappa_1, \kappa_2 \lesssim k$. Here, we check this by computing the discrete signal

$$\widetilde{v}_h(x_n, y_n) := \frac{1}{N^2}\sum_{\widehat{m}=1}^{N}\sum_{\widehat{n}=1}^{N} \widehat{H}_{\widehat{m}}\widehat{H}_{\widehat{n}}\widehat{W}_{\widehat{m},\widehat{n}} \mathrm{e}^{\mathrm{i}\kappa_{\widehat{n}}x_n} \mathrm{e}^{\mathrm{i}\kappa_{\widehat{m}}y_n}\,,$$

where H is a "low-pass filter," i.e.,

$$\widehat{H}_{\widehat{m}} = \begin{cases} 1 & \text{if } \kappa_{\widehat{m}} \le \alpha k \text{ or } \frac{2\pi}{\Delta} - \kappa_{\widehat{m}} \le \alpha k\,, \\ 0 & \text{otherwise.} \end{cases}$$

[2] https://github.com/samplemaker/freefem_matlab_octave_plot/blob/public/README.md

Note that, by periodicity, for all $0 \le m, \widehat{m} \le N-1$, one has

$$\mathrm{e}^{\mathrm{i}\kappa_{\widehat{m}} x_m} = \mathrm{e}^{-\mathrm{i}(\frac{2\pi}{\Delta} - \kappa_{\widehat{m}})x_m},$$

hence $\widehat{H}$ effectively removes all frequencies outside the interval $[-\alpha k, \alpha k]$. Here, α is a parameter, set to 2 in our tests.

The relative l^2 norm of the high-frequency components of v_h is

$$\rho := \frac{\|v_h - \widetilde{v}_h\|_{l^2}}{\|v_h\|_{l^2}}, \tag{2.1}$$

where $\|\cdot\|_{l^2}$ is the discrete l^2 norm. By Parseval's theorem for discrete Fourier transforms, this ratio is equal to

$$\rho = \frac{\sum_{\widehat{m},\widehat{n}=1}^{N} (1 - \widehat{H}_{\widehat{m}}\widehat{H}_{\widehat{n}})^2 \left|\widehat{V}_{\widehat{m},\widehat{n}}\right|^2}{\sum_{\widehat{m},\widehat{n}=1}^{N} \left|\widehat{V}_{\widehat{m},\widehat{n}}\right|^2}.$$

3 Abstract framework

3.1 Function spaces

Given a Hilbert space $\mathcal{Y}$, let $\mathcal{Y}^*$ be the Hilbert space of bounded *anti-linear* functionals $G : \mathcal{Y} \to \mathbb{C}$ (i.e., $G(\lambda y) = \overline{\lambda} G(y)$) equipped with the norm

$$\|G\|_{\mathcal{Y}^*} := \sup_{y \in \mathcal{Y}\setminus\{0\}} \frac{|G(y)|}{\|y\|_{\mathcal{Y}}}. \tag{3.1}$$

Let $\Omega \subset \mathbb{R}^d$ be a bounded open set and let $\mathcal{H} := L^2(\Omega)$. As usual, $\mathcal{H}$ is identified with its dual $\mathcal{H}^*$. Furthermore, let $k_0 > 0$ and let $(\mathcal{Z}_k)_{k \ge k_0}$ be a family of Hilbert spaces such that for each $k \ge k_0$, the inclusion $\mathcal{Z}_k \subset \mathcal{H}$ is dense. In all the examples below, $\mathcal{Z}_k$ is either $H^1_k(\Omega)$ or this space with a zero Dirichlet boundary condition prescribed on a subset of its boundary. To quantify abstractly the "regularity" of elements of $\mathcal{H}$, given $\ell \in \mathbb{N}$, we introduce a scale of Hilbert spaces $(\mathcal{Z}_k^j)_{0 \le j \le \ell+2}$ with $\mathcal{Z}_k^0 = \mathcal{H}$, $\mathcal{Z}_k^1 = \mathcal{Z}_k$, and with dense inclusions $\mathcal{Z}_k^j \subset \mathcal{Z}_k^{j-1}$ for $j = 1, \ldots, \ell+2$.

For $j \ge 0$, each anti-linear functional g on $\mathcal{Z}_k^j$ also defines an anti-linear functional g' on $\mathcal{Z}_k^{j'}$ for $j' \ge j$, via restriction since $\mathcal{Z}_k^{j'} \subset \mathcal{Z}_k^j$. This restriction map is moreover continuous and injective, by density of the previous inclusions and hence we identify g and g'. This identification is compatible with the identification of $\mathcal{H}$ to its dual, and gives the chain of continuous and dense inclusions

$$(\mathcal{Z}_k^{\ell+1})^* \supset (\mathcal{Z}_k^{\ell})^* \supset \ldots \supset (\mathcal{Z}_k^1)^* \supset \underbrace{\mathcal{Z}_k^0}_{=\mathcal{H}} \supset \underbrace{\mathcal{Z}_k^1}_{=\mathcal{Z}_k} \supset \ldots \supset \mathcal{Z}_k^{\ell} \supset \mathcal{Z}_k^{\ell+1}. \tag{3.2}$$

We assume that there exists $C_{\text{emb}} > 0$ such that, for all $0 \le j \le j' \le \ell + 2$,

$$\|u\|_{\mathcal{Z}_k^j} \le C_{\text{emb}} \|u\|_{\mathcal{Z}_k^{j'}} \quad \text{for all } u \in \mathcal{Z}_k^{j'} \text{ and } k \ge k_0. \tag{3.3}$$

which also implies that $\|u\|_{(\mathcal{Z}_k^{j'})^*} \le C_{\text{emb}} \|u\|_{(\mathcal{Z}_k^j)^*}$ for $0 \le j \le j' \le \ell + 2$.

For technical reasons (to be able to treat transmission problems), we introduce another scale $(\mathcal{W}_k^j)_{0 \le j \le \ell+1}$ of Hilbert spaces with the property that $\mathcal{Z}_k^j \subset \mathcal{W}_k^j \subset \mathcal{H}$ with continuous inclusions (in particular $\mathcal{W}_k^0 = \mathcal{H}$), and, for all $0 \le j \le \ell + 2$,

$$\|u\|_{\mathcal{W}_k^j} \le C_{\text{emb}} \|u\|_{\mathcal{Z}_k^j} \quad \text{for all } u \in \mathcal{Z}_k^j \text{ and } k \ge k_0. \tag{3.4}$$

Example 3.1 *For the Helmholtz transmission problem with the outgoing condition approximated by a perfectly-matched layer,*

$$\mathcal{Z}_k = H_0^1(\Omega), \quad \mathcal{W}_k^j = L^2(\Omega) \cap \big(H^j(\Omega_{\text{in}}) \oplus H^j(\Omega_{\text{out}} \cap \Omega)\big), \quad \textit{and} \quad \mathcal{Z}_k^j = \mathcal{W}_k^j \cap H_0^1(\Omega),$$

where Ω_{in} is the penetrable obstacle, Ω_{out} its exterior, and Ω the (truncated) computational domain containing Ω_{in}; see §7.2 below.

3.2 Local properties

Lemma 3.2 *Let B_j, $j = 1, \dots, N$ be open sets, and let*

$$C_{\text{over}} := \max \Big\{ |J| : B_{j_1} \cap \dots \cap B_{j_J} \ne \emptyset \textit{ with } j_1, \dots, j_J \textit{ distinct} \Big\}. \tag{3.5}$$

Then,

$$(C_{\text{over}})^{-1} \sum_{j=1}^N \|v\|^2_{\mathcal{H}(B_j)} \le \|v\|^2_{\mathcal{H}(\cup_{j=1}^N B_j)} \le \sum_{j=1}^N \|v\|^2_{\mathcal{H}(B_j)} \tag{3.6}$$

for all $v \in \mathcal{H}(\cup_{j=1}^N B_j)$.

The notation B_j is used because we use Lemma 3.2 below with the B_j either balls or balls intersected with some larger (fixed) open set.

Proof of Lemma 3.2 The second inequality in (3.6) follows immediately from the fact that the $\mathcal{H}$ norm is the L^2 norm. Given $x \in \cup_{j=1}^N B_j$ let $m(x)$ be the number of distinct $B_1, \dots, B_N$ that contain x. Then,

$$\sum_{j=1}^N \|v\|^2_{\mathcal{H}(B_j)} = \int_{\cup_{j=1}^N B_j} m(x) |v(x)|^2 \, \mathrm{d}x$$

and the first inequality in (3.6) follows since $|m(x)| \le C_{\text{over}}$. □

Assumption 3.3 *The following holds with $\mathcal{Y}$ equal to either $\mathcal{W}$ or $\mathcal{Z}$. Given open sets $B_1, \ldots, B_N$, with C_{over} as in (3.5),*

$$(C_{\text{over}})^{-1} \sum_{j=1}^{N} \|v\|^2_{\mathcal{Y}_k^j(B_j)} \leq \|v\|^2_{\mathcal{Y}_k^j(\cup_{j=1}^N B_j)} \leq \sum_{j=1}^{N} \|v\|^2_{\mathcal{Y}_k^j(B_j)} \tag{3.7}$$

for all $v \in \mathcal{Y}_k^j(\cup_{j=1}^N B_j)$. Furthermore, if $\operatorname{supp} u \cap \operatorname{supp} v = \emptyset$, then

$$\|u+v\|^2_{\mathcal{Y}_k^j} = \|u\|^2_{\mathcal{Y}_k^j} + \|v\|^2_{\mathcal{Y}_k^j}. \tag{3.8}$$

With $\partial_<$ defined by (1.4), for $U \subset \Omega$ open, let

$$\begin{aligned} \mathcal{Z}_k^{j,<}(U) &:= \overline{\left\{v \in \mathcal{Z}_k^j \text{ s.t. } \operatorname{supp} v \subset \overline{U},\ \partial_<(\operatorname{supp} v, \overline{U}) > 0\right\}} \\ &= \overline{\left\{v \in \mathcal{Z}_k^j \text{ s.t. } \operatorname{supp} v \subset U \cup (\partial U \cap \partial\Omega)\right\}} \end{aligned} \tag{3.9}$$

(where the closures are taken with respect to the $\mathcal{Z}_k^j$ norm). Observe that the convention that $\partial_<(A, B) = +\infty$ when $B = \Omega$ implies that $\mathcal{Z}_k^{j,<}(\Omega) = \mathcal{Z}_k^j$.

For any $U \subset \Omega$, let

$$\|u\|_{\mathcal{Z}_k^j(U)} := \inf\left\{\|v\|_{\mathcal{Z}_k^j} \,:\, v|_U = u|_U,\ v \in \mathcal{Z}_k^j\right\}; \tag{3.10}$$

observe that this definition implies that $\|u\|_{\mathcal{Z}_k^j(U)} \leq \|u\|_{\mathcal{Z}_k^j(V)}$ for $U \subset V$.

For an open set $U \subset \Omega$ and $j \geq 0$, observe that (3.1) implies that

$$\|u\|_{(\mathcal{Z}_k^{j,<}(U))^*} = \sup_{v \in \mathcal{Z}_k^{j,<}(U)\setminus\{0\}} \frac{|u(v)|}{\|v\|_{\mathcal{Z}_k^j}}. \tag{3.11}$$

Since $\mathcal{Z}_k^{j,<}(\Omega) = \mathcal{Z}_k^j$, $\|\cdot\|_{(\mathcal{Z}_k^{j,<}(\Omega))^*} = \|\cdot\|_{(\mathcal{Z}_k^j)^*}$.

We define $\mathcal{W}_k^{j,<}(U)$ and $\|\cdot\|_{(\mathcal{W}_k^{j,<}(U))^*}$ analogously to (3.9) and (3.11).

Lemma 3.4 *For any open set $U \subset \Omega$ and any function $u \in \mathcal{Z}_k$,*

$$\|u\|_{(\mathcal{Z}_k^{j,<}(U))^*} \leq C_{\text{emb}} \|u\|_{(\mathcal{W}_k^{j,<}(U))^*}. \tag{3.12}$$

Proof Let $v \in \mathcal{Z}_k^j$ be non-zero, supported on $\overline{U}$, and such that $\partial_<(\operatorname{supp} v, U) > 0$. Then, by (3.4) and the definition of $\mathcal{W}_k^{j,<}(U)$, $v \in \mathcal{W}_k^{j,<}(U)$ and furthermore

$$\frac{|\langle u, v\rangle_{\mathcal{H}}|}{\|v\|_{\mathcal{Z}_k^j}} \leq C_{\text{emb}} \frac{|\langle u, v\rangle_{\mathcal{H}}|}{\|u\|_{\mathcal{W}_k^j}}.$$

By definition of the dual $(\mathcal{W}_k^{j,<}(U))^*$ norm (defined analogously to (3.11)),

$$\frac{|\langle u, v\rangle_{\mathcal{H}}|}{\|v\|_{\mathcal{Z}_k^j}} \leq C_{\text{emb}} \|u\|_{(\mathcal{W}_k^{j,<}(U))^*},$$

and the result follows since the set $\left\{v \in \mathcal{Z}_k^j \text{ s.t. } \operatorname{supp} v \subset \overline{U},\ \partial_<(\operatorname{supp} v, \overline{U}) > 0\right\}$ is dense in $\mathcal{Z}_k^{j,<}(U)$ by its definition (3.9). □

3.3 Sesquilinear forms

We consider a family $(a_k)_{k\geq k_0}$ of *sesquilinear* forms $a_k : \mathcal{Z}_k \times \mathcal{Z}_k \to \mathbb{C}$ (i.e., $a_k(\lambda u, \mu v) = \lambda\overline{\mu} a_k(u, v)$), satisfying the following assumptions.

Assumption 3.5 (Continuity and local coercivity) *There exist positive constants C_{cont}, c_{coer}, and C_{coer} such that*

$$|a_k(u, v)| \leq C_{\text{cont}} \|u\|_{\mathcal{Z}_k} \|v\|_{\mathcal{Z}_k} \quad \text{for all } u, v \in \mathcal{Z}_k \text{ and } k \geq k_0 \tag{3.13}$$

and if $x_0 \in \Omega$ and $r \leq c_{\text{coer}} k^{-1}$ then

$$\Re\left\{a_k(v, v)\right\} \geq C_{\text{coer}} \|v\|^2_{\mathcal{Z}_k} \quad \text{for all } v \in \mathcal{Z}_k^<(B(x_0, r) \cap \Omega) \text{ and } k \geq k_0. \tag{3.14}$$

Assumption 3.6 (Elliptic regularity up to $\partial\Omega$ for the adjoint problem on $\mathcal{O}(k^{-1})$ balls) *Given $x_0 \in \Omega$, $r > 0$, and $d > 0$, let*

$$U_0 := B(x_0, r) \cap \Omega \quad \text{and} \quad U_1 := B(x_0, r + d) \cap \Omega \tag{3.15}$$

(so that $\partial_<(U_0, U_1) = d$).

Given $c > 0$, and $\ell \in \mathbb{Z}^+$, there exists $C_{\text{ell}} > 0$ such that if $r + d \leq c_{\text{coer}} k^{-1}$ and $r, d \geq ck^{-1}$, then, for all $u \in \mathcal{Z}_k^<(U_1)$,

$$\|u\|_{\mathcal{Z}_k^{j+2}(U_0)} \leq C_{\text{ell}}\Big(\|u\|_{\mathcal{H}} + \sup_{v \in \mathcal{Z}_k^<(U_1),\ \|v\|_{(\mathcal{W}_k^j)^*}=1} |a_k(v, u)|\Big), \quad j = 0, \ldots, \ell. \tag{3.16}$$

The following assumption involves "localising" operators that commute with a_k in a weak sense. In all the specific examples in §7, these operators are cut-off functions, but for transmission problems these functions must be defined piecewise and satisfy certain properties across the interface; see Lemma 7.7.

Assumption 3.7 (Compatible localisers) *Given $x_0 \in \Omega$, $r > 0$, and $d > 0$, let U_0 and U_1 be as in (3.15). There exist constants $C_\dagger > 0$ and $C_{\text{com}} > 0$ and a family $\{\psi(U_0, U_1)\}_{U_0 \subset U_1}$ of localisers indexed by $U_0 \subset U_1 \subset \Omega$, and hence indexed by x_0, r, d, where each $\psi = \psi(U_0, U_1) : \mathcal{H} \to \mathcal{H}$ is a self-adjoint operator with the following properties*

(i) With $U_0' := B(x_0, r + d/4) \cap \Omega$ *and* $U_1' := B(x_0, r + 3d/4) \cap \Omega$, *for all* $u \in \mathcal{H}$,

$$u \equiv \psi u \text{ on } U_0', \quad \psi u \equiv 0 \text{ on } (U_1')^c .$$

(ii) ψ *maps* $\mathcal{Z}_k^j$ *to itself continuously, with*

$$\|\psi u\|_{\mathcal{Z}_k^j} \le C_\dagger \sum_{m=0}^{j} (kd)^{-(j-m)} \|u\|_{\mathcal{Z}_k^m(U_1)} \quad \text{for all } j \in \{0, 1, \dots, \ell + 2\}, \tag{3.17}$$

(iii) for all $j = 1, \dots, \ell + 1$, *and* $u, v \in \mathcal{Z}_k$,

$$\begin{aligned} &|a_k(\psi u, v) - a_k(u, \psi v)| \\ &\le \frac{C_{\mathrm{com}}}{kd} \left(\sum_{m=0}^{j-1} (kd)^{-m} \right) \min \left(\|u\|_{\mathcal{Z}_k^j(U_1 \setminus \overline{U_0})} \|v\|_{(\mathcal{W}_k^{(j-1),<}(U_1 \setminus \overline{U_0}))^*}, \|u\|_{(\mathcal{W}_k^{(j-1),<}(U_1 \setminus \overline{U_0}))^*} \|v\|_{\mathcal{Z}_k^j(U_1 \setminus \overline{U_0})} \right). \end{aligned} \tag{3.18}$$

Corollary 3.8 (Mapping properties of the adjoint solution operator on $\mathcal{O}(k^{-1})$ balls)
Let U_0 *and* U_1 *be given by* (3.15). *Suppose that* $r + d \le c_{\mathrm{coer}} k^{-1}$. *Let* $\mathcal{R}^* : (\mathcal{Z}_k^<(U_1))^* \to \mathcal{Z}_k^<(U_1)$ *be the operator defined by the variational problem*

$$a_k(v, \mathcal{R}^* g) = \overline{\langle g, v \rangle} \quad \text{for all } v \in \mathcal{Z}_k^<(U_1) .$$

Then, there exists $C_{\mathrm{res}} > 0$ *such that*

$$\left\| \mathcal{R}^* g \right\|_{\mathcal{Z}_k^{j+2}(U_0)} \le C_{\mathrm{res}} \|g\|_{\mathcal{W}_k^j} \quad \text{for all } g \in \mathcal{W}_k^{j,<}(U_1) , \ j = 0, \dots, \ell .$$

Proof Since $r + d \le c_{\mathrm{coer}} k^{-1}$, $\mathcal{R}^*$ is well-defined by Assumption 3.5 and the Lax–Milgram lemma, and

$$\left\| \mathcal{R}^* g \right\|_{\mathcal{Z}_k} \le \frac{1}{C_{\mathrm{coer}}} \|g\|_{(\mathcal{Z}_k^<(U_1))^*} . \tag{3.19}$$

Let $g \in \mathcal{W}_k^{j,<}(U_1)$. By (3.16), the definition of $\mathcal{R}^*$, and the triangle inequality,

$$\left\| \mathcal{R}^* g \right\|_{\mathcal{Z}_k^{j+2}(U_0)} \le C_{\mathrm{ell}} \Big(\left\| \mathcal{R}^* g \right\|_{\mathcal{H}} + \sup_{v \in \mathcal{Z}_k^<(U_1),\, \|v\|_{(\mathcal{W}_k^j)^*} = 1} |\langle g, v \rangle| \Big) \le C_{\mathrm{ell}} \big(\left\| \mathcal{R}^* g \right\|_{\mathcal{H}} + \|g\|_{\mathcal{W}_k^j} \big),$$

and the result follows from (3.19). □

3.4 Triangulation and finite-dimensional subspaces

Let $\mathcal{T}$ be a regular triangulation (in the sense of, e.g., [6, Page 61]) of Ω. For each element $K \in \mathcal{T}$, let $h_K := \operatorname{diam}(K)$. For simplicity, we assume that

$$h_K k \leq C_{\text{ppw}} \tag{3.20}$$

for some $C_{\text{ppw}} > 0$ (where "ppw" standard for "points per wavelength"); as discussed after Theorem 1.1 for standard finite-element spaces with p fixed, $h_K k$ must be chosen as a decreasing function of k to maintain accuracy and thus this assumption is not restrictive.

Assumption 3.9 (Broken norms) *For each $u \in \mathcal{Z}_k$ and each element K of $\mathcal{T}$, the restriction of u to K belongs to $H^1_k(K)$, and*

$$\|u\|_{\mathcal{Z}_k(K)} = \|u\|_{H^1_k(K)} \quad \text{for all } K \in \mathcal{T}\,. \tag{3.21}$$

For each $K \in \mathcal{T}$, we assume that $C_0^\infty(K) \subset \mathcal{Z}_k^j$ and that

$$\|u\|_{\mathcal{Z}_k^j} = \|u\|_{H^j_k(K)} \quad \text{for all } u \in C_0^\infty(K) \tag{3.22}$$

(where the first norm is defined by (3.10)). Furthermore, there exists a constant C_{loc} such that

$$|a_k(u,v)| \leq C_{\text{loc}} \sum_{K\in\mathcal{T}} \|u\|_{H^1_k(K)} \|v\|_{H^1_k(K)} \quad \text{for all } u, v \in \mathcal{Z}_k\,. \tag{3.23}$$

We fix a finite-dimensional space $V_h \subset \mathcal{Z}_k$ consisting of functions whose restrictions to each $K \in \mathcal{T}$ is in $C^\infty(\overline{K})$. For any open subset $U \subset \Omega$, let

$$V_h^<(U) := \mathcal{Z}_k^<(U) \cap V_h\,.$$

We introduce the following standard assumptions on V_h:

Assumption 3.10 (Approximation property) *There exist constants $\kappa > 0$, $p \in \mathbb{Z}^+$ and $C_{\text{approx}} > 0$ such that the following holds. For each $j \in \{1, \dots, p+1\}$, given $u \in \mathcal{Z}_k^j$, there exists $u_h \in V_h$ such that*

$$\sum_{K\in\mathcal{T}} (h_K k)^{2(m-j)} \|u - u_h\|^2_{H^m_k(K)} \leq C_{\text{approx}} \|u\|^2_{\mathcal{Z}_k^j} \quad \text{for all } u_h \in V_h\,,\ 0 \leq m \leq j \leq p+1. \tag{3.24}$$

Furthermore, if $U_0 \subset U_1$ are such that

$$\partial_<(U_0, U_1) > \kappa \max_{K\cap U_1 \neq \emptyset} h_K$$

and $\operatorname{supp} u \subset U_0$*, then u_h can be chosen in $V_h^<(U_1)$.*

Assumption 3.11 (Super-approximation property) *For each $C_\dagger > 0$, there exists constant $C_{\text{super}} > 0$ such that, with p and κ as in* Assumption 3.10*, the following property holds for sets $U_0 \subset U_1 \subset \Omega$ given by* (3.15) *and satisfying*

$$d = \partial_<(U_0, U_1) > 4\kappa \max_{K \cap U_1 \neq \emptyset} h_K \tag{3.25}$$

Let $\chi = \psi(U_0, U_1)$ be the localiser associated to U_0 and U_1 (given by Assumption 3.7*). Then, for each $u_h \in V_h$, there exists $v_h \in V_h^<(U_1)$ such that, for all $K \in \mathcal{T}$,*

$$\left\| \chi^2 u_h - v_h \right\|_{H_k^1(K)} \leq C_{\text{super}} \frac{h_K}{d} \left[\left(1 + \frac{1}{kd} \right) \|u_h\|_{L^2(K)} + \|\chi u_h\|_{H_k^1(K)} \right]. \tag{3.26}$$

The constant κ in Assumptions 3.10 and 3.11 is related to the "stencil" of the chosen finite element, i.e., how large the support the finite-element basis functions is. For Lagrange finite-elements $\kappa = 1$; see §7.4 below.

For any open subset $U \subset \mathbb{R}^d$ and any $s \in \mathbb{R}$, we recall that $H^s(U)$ is defined as the set of restrictions to U of elements of $H^s(\mathbb{R}^d)$, with a Hilbert structured inherited via

$$\|v\|_{H_k^s(U)} := \inf_{V \in H_k^s(\mathbb{R}^d) \,:\, V|_U = v} \|V\|_{H_k^s(\mathbb{R}^d)}. \tag{3.27}$$

When U is a Lipschitz domain,

$$\|u\|_{H_k^{-s}(U)} \sim \sup_{\substack{\|v\|_{H_k^s(\mathbb{R}^d)}=1, \\ \text{supp} v \subset U}} |(u, v)_{L^2}|, \tag{3.28}$$

where $\sim$ denotes norm equivalence; i.e., $H_k^{-s}(U)$ is dual to $\widetilde{H}_k^s(U)$ defined as the closure of $C_{\text{comp}}^\infty(U)$ in $H^s(\mathbb{R}^d)$; see [24, Page 77 and Theorem 3.30(i), Page 92].

Assumption 3.12 (Inverse inequalities) *Given $p \in \mathbb{Z}^+$ as in* Assumption 3.10*, there exists $C_{\text{inv}} > 0$ such that, for all $K \in \mathcal{T}$ and $u_h \in V_h$,*

$$\|u_h\|_{H_k^1(K)} \leq \frac{C_{\text{inv}}}{h_K k} \|u_h\|_{L^2(K)} \tag{3.29}$$

and, for $0 \leq s \leq p$,

$$\|u_h\|_{L^2(K)} \leq \frac{C_{\text{inv}}}{(h_K k)^s} \|u_h\|_{H_k^{-s}(K)}. \tag{3.30}$$

4 Statement of the main results

Theorem 4.1 (General version of Theorem 1.1) *Given positive constants C_{cont}, C_{coer}, c_{coer}, κ, p, C_{inv}, C_{pw}, C_{super}, $C_{\text{com}} > 0$, and some $C_0 > 0$, there exists a*

constant $C_\star > 0$ such that the following holds. Let $(a_k)_{k\geq k_0}$ and $V_h \subset \mathcal{Z}_k$ satisfy Assumptions 3.5, 3.7, 3.10, 3.11, 3.12, and 3.23*, with the constants above. Let $\Omega_0 \subset \Omega_1 \subset \Omega$ be arbitrary subsets such that*

$$d := \partial_<(\Omega_0, \Omega_1) \geq \frac{C_0}{k} \quad \text{and} \quad \max_{K\cap\Omega_1\neq\emptyset} h_K \leq \frac{C_1}{k} \tag{4.1}$$

where

$$C_1 := \frac{1}{4\max\{1, 8\kappa\}} \min\left\{\frac{C_0}{2}, \frac{4c_{\text{coer}}}{3}\right\}.$$

If $k \geq k_0$ and $u \in \mathcal{Z}_k$ and $u_h \in V_h$ are such that

$$a_k(u - u_h, v_h) = 0 \quad \text{for all } v_h \in V_h^<(\Omega_1), \tag{4.2}$$

then

$$\|u - u_h\|_{\mathcal{Z}_k(\Omega_0)} \leq C_*\Big(\min_{w_h\in V_h} \|u - w_h\|_{\mathcal{Z}_k(\Omega_1)} + \|u - u_h\|_{\mathcal{H}(\Omega_1)}\Big). \tag{4.3}$$

If the triangulation $\mathcal{T}$ is furthermore locally quasi-uniform and Assumption 3.6 (local elliptic regularity) holds, then the result can be improved by weakening the $\mathcal{H}$ norm of the error on the right-hand side.

Theorem 4.2 (General version of Theorem 1.2) *Given positive constants C_{cont}, C_{coer}, c_{coer}, C_{ell}, κ, p, C_{inv}, C_{pw}, C_{super}, C_{com}, $C_{\text{approx}} > 0$, and some $C_0 > 0$, there exists a constant $C_\star > 0$ such that the following holds. Let $(a_k)_{k\geq k_0}$ and $V_h \subset \mathcal{Z}_k$ satisfy* Assumptions 3.5, 3.6, 3.7, 3.10, 3.11, 3.12, and 3.23*, with the constants above. Let $\Omega_0 \subset \Omega_1 \subset \Omega$ be arbitrary subsets such that*

$$d := \partial_<(\Omega_0, \Omega_1) \geq \frac{C_0}{k} \quad \text{and} \quad \max_{K\cap\Omega_1\neq\emptyset} h_K \leq \frac{C_1}{k}$$

where

$$C_1 := \frac{1}{48(\ell+2)\max\{1, 8\kappa\}} \min\left\{\frac{C_0}{2}, 2c_{\text{coer}}\right\}.$$

Assume further that $\mathcal{T}$ is quasi-uniform on scale k^{-1}*, in the sense that, for every ball B of radius at most $3c_{\text{coer}}/(2k)$, the bound (1.7) holds. If $k \geq k_0$ and $u \in \mathcal{Z}_k$ and $u_h \in V_h$ are such that (4.2) holds, then*

$$\|u - u_h\|_{\mathcal{Z}_k(\Omega_0)} \leq C_*\left(\min_{w_h\in V_h} \|u - w_h\|_{\mathcal{Z}_k(\Omega_1)} + \|u - u_h\|_{(\mathcal{W}_k^{s+1,<}(\Omega_1))^*}\right) \tag{4.4}$$

where $s := \min\{\ell, p-1\}$.

Remark 4.3 (Galerkin orthogonality) *Given $G \in (\mathcal{Z}_k)^*$, if $u \in \mathcal{Z}_k$ satisfies*

$$a_k(u, v) = G(v) \quad \text{for all } v \in \mathcal{Z}_k$$

and if u_h is a solution to the Galerkin equations

$$\text{find } u_h \in V_h \text{ such that } a(u_h, v_h) = G(v_h) \quad \text{for all } v_h \in V_h, \tag{4.5}$$

then

$$a(u - u_h, v_h) = 0 \quad \text{for all } v_h \in V_h,$$

and thus $u - u_h$ satisfies (4.2) *since $V_h^<(U_1) \subset V_h$.*

Remark 4.4 (The dependence of the constant C^* on the subspace V_h) *We have stated* Theorem 4.1 *for a fixed finite-dimensional subspace $V_h \subset \mathcal{Z}_k$. The key point is that the constant C^* depends quantitatively on V_h only through the constants in the assumptions of* Sect. 3. *In general, one is interested in applying* Theorem 4.1 *to a* family *of subspaces indexed by some parameter (e.g., the mesh parameter h of a sequence of quasi-uniform triangulations $(\mathcal{T}_h)_{h>0}$ of Ω). The idea is that the bound* (4.3) *will hold with a common constant C^* for all subspaces of the sequence, provided that the assumptions of* Sect. 3 *hold uniformly for all of those subspaces. We show in* §7.4 *that these assumptions do indeed hold uniformly for standard choices of discretisation.*

In the rest of the paper, we use the letter C in estimates of the form $a \leq Cb$ to represent a generic positive constant, whose numerical value is allowed to change from one place to another, but which can be expressed as a function of the constants in the assumptions of Sect. 3.

5 Caccioppoli estimates

The central idea in the proofs of Theorems 4.1-4.2 is to apply a discrete version of the classical Caccioppoli inequality to a solution of the Helmholtz equation at the discrete level.

In this section, we always assume (without stating explicitly) that $(a_k)_{k\geq k_0}$ and V_h satisfy the assumptions of Sect. 3.

Lemma 5.1 (Caccioppoli estimate in the L^2 norm on $\mathcal{O}(k^{-1})$ balls) *There exists a constant $C_{\mathrm{ca}} > 0$ (whose value depends only on the constants appearing in the assumptions of* Sect. 3 *apart from C_{ell}) such that, given $r, d > 0$, if U_0 and U_1 are as in* (3.15) *(so that $d = \partial_<(U_0, U_1)$),*

$$d \geq \max\{1, 8\kappa\} \max_{K \cap U_1 \neq \emptyset} h_K, \tag{5.1}$$

and

$$r + \frac{d}{2} \leq \frac{c_{\mathrm{coer}}}{k}, \tag{5.2}$$

then the following holds. If $z_h \in V_h$ satisfies

$$a_k(z_h, v_h) = 0 \quad \text{for all } v_h \in V_h^<(U_1) \tag{5.3}$$

and $k \geq k_0$, then

$$\|z_h\|_{\mathcal{Z}_k(U_0)} \leq \frac{C_{\mathrm{ca}}}{kd} \|z_h\|_{\mathcal{H}(U_1)}. \tag{5.4}$$

Note that the combination of (5.1) and (5.2) imply that $\max_{K \in U_1} h_K$ is bounded by a constant multiple of k^{-1}.

Proof of Lemma 5.1. Let

$$U_{1/2} := B(x_0, r + d/2) \cap \Omega$$

so that

$$\partial_<(U_0, U_{1/2}) = d/2, \quad \partial_<(U_{1/2}, U_1) = d/2. \tag{5.5}$$

Let $\chi = \psi(U_0, U_{1/2})$ be the localiser defined in Assumption 3.7. If we can show that

$$\|\chi z_h\|^2_{\mathcal{Z}_k} \leq \frac{C}{(kd)^2} \|z_h\|^2_{\mathcal{H}(U_1)}, \tag{5.6}$$

then the result follows since, by (3.10),

$$\|z_h\|^2_{\mathcal{Z}_k(U_0)} = \|\chi z_h\|^2_{\mathcal{Z}_k(U_0)} \leq \|\chi z_h\|^2_{\mathcal{Z}_k}.$$

The inequality (5.2) and the assumption (3.14) imply that a_k is coercive on $\mathcal{Z}_k^<(U_{1/2})$. Since $\chi z_h \in \mathcal{Z}_k^<(U_{1/2})$,

$$\|\chi z_h\|^2_{\mathcal{Z}_k} \leq (C_{\mathrm{coer}})^{-1} \left| a_k(\chi z_h, \chi z_h) \right|.$$

Then, by (3.18) with $j = 1$,

$$\|\chi z_h\|^2_{\mathcal{Z}_k} \leq (C_{\mathrm{coer}})^{-1} \left| a_k(z_h, \chi^2 z_h) \right| + C_{\mathrm{com}} (kd)^{-1} \|z_h\|_{\mathcal{H}(U_{1/2})} \|\chi z_h\|_{\mathcal{Z}_k}, \tag{5.7}$$

so that, using the inequality

$$2ab \leq \varepsilon a^2 + b^2/\varepsilon \quad \text{for all } a, b, \varepsilon > 0, \tag{5.8}$$

we have, for all $\varepsilon_1 > 0$,

$$\|\chi z_h\|^2_{\mathcal{Z}_k} \leq C \left(\left| a_k(z_h, \chi^2 z_h) \right| + \varepsilon_1^{-1} (kd)^{-2} \|z_h\|^2_{\mathcal{H}(U_{1/2})} \right) + \varepsilon_1 \|\chi z_h\|^2_{\mathcal{Z}_k}. \tag{5.9}$$

Let $w_h \in V_h^<(U_{1/2})$ be the finite-element super-approximation of $\chi^2 z_h$ provided by Assumption 3.11 applied to the pair of sets $U_0, U_{1/2}$; note that the condition (3.25) needed to apply Assumption 3.11 becomes $d > 8\kappa \max_{K \cap U_1 \neq \emptyset} h_K$ by (5.5), which holds by (5.1).

By the Galerkin orthogonality (5.3), the property (3.23), and the fact that both $\chi^2 z_h$ and w_h are supported on $U_{1/2}$,

$$\left|a_k(z_h, \chi^2 z_h)\right| = \left|a_k(z_h, \chi^2 z_h - w_h)\right| \leq C_{\mathrm{loc}} \sum_{K \cap U_{1/2} \neq \emptyset} \|z_h\|_{H_k^1(K)} \left\|\chi^2 z_h - w_h\right\|_{H_k^1(K)}. \tag{5.10}$$

Now, by (3.26), the fact that $kd \leq 2c_{\mathrm{coer}}$ by (5.2), (3.29), and (5.8),

$$\begin{aligned}
\|z_h\|_{H_k^1(K)} \left\|\chi^2 z_h - w_h\right\|_{H_k^1(K)} &\leq C_{\mathrm{super}} \|z_h\|_{H_k^1(K)} \frac{h_K}{d} \left[\frac{(1+2c_{\mathrm{coer}})}{kd} \|z_h\|_{L^2(K)} + \|\chi z_h\|_{H_k^1(K)}\right] \\
&\leq \frac{C_{\mathrm{super}} C_{\mathrm{inv}}}{kd} \left[\frac{(1+2c_{\mathrm{coer}})}{kd} \|z_h\|^2_{L^2(K)} + \|\chi z_h\|_{H_k^1(K)} \|z_h\|_{L^2(K)}\right] \\
&\leq \frac{C}{(kd)^2} (1 + \varepsilon_2^{-1}) \|z_h\|^2_{L^2(K)} + \varepsilon_2 \|\chi z_h\|^2_{H_k^1(K)}
\end{aligned}$$

for all $\varepsilon_2 > 0$. Combining this last inequality with (5.10) and (5.9) and then using (3.21), we obtain

$$\|\chi z_h\|^2_{\mathcal{Z}_k} \leq \frac{C}{(kd)^2} (1 + \varepsilon_1^{-1} + \varepsilon_2^{-1}) \|z_h\|^2_{\mathcal{H}(U_1)} + (\varepsilon_1 + \varepsilon_2) \|\chi z_h\|^2_{\mathcal{Z}_k}.$$

Choosing $\varepsilon_1 = \varepsilon_2 = 1/4$, the last term on the right-hand side can be absorbed in the left-hand side, leading to (5.6), and hence the result (5.4) follows. □

To prove the Caccioppoli estimate with a negative norm on the right-hand side (Lemma 5.5), we need the following lemma.

Lemma 5.2 *If $\Omega_0 \subset \Omega_1 \subset \Omega$ are arbitrary sets such that*

$$\bigcup \{K \in \mathcal{T} \text{ s.t. } K \cap \Omega_0 \neq \emptyset\} \subset \Omega_1,$$

then

$$\sum_{K \cap U_0 \neq \emptyset} \|u\|^2_{H_k^{-j}(K)} \leq \|u\|^2_{(\mathcal{Z}_k^{j,<}(\Omega_1))^*}. \tag{5.11}$$

Proof The proof is very similar to that of [27, Lemma 1.1]. Let $\varphi_K \in C^\infty_{\mathrm{comp}}(K)$ be such that $\|\varphi_K\|_{H_k^j(K)} = 1$, let $\theta_K := \|u\|_{H_k^{-j}(K)}$, and let

$$\varphi := \sum_{K \cap \Omega_0 \neq \emptyset} \theta_K \varphi_K .$$

By linearity, $\varphi \in \mathcal{Z}_k^j$, $\operatorname{supp} \varphi \subset \Omega_1$, so that $\varphi \in \mathcal{Z}_k^{j,<}(\Omega_1)$. By (3.8) and (3.22),

$$\|\varphi\|^2_{\mathcal{Z}_k^j} \leq \sum_{K \cap U_0 \neq \emptyset} \theta_K^2 .$$

Hence,

$$\|u\|^2_{(\mathcal{Z}_k^{j,<}(\Omega_1))^*} \geq \frac{|(u,\varphi)_{\mathcal{H}}|^2}{\|\varphi\|^2_{\mathcal{Z}_k^j}} \geq \frac{\left|\sum_{K\cap\Omega_0\neq\emptyset}\theta_K(u,\varphi_K)_{\mathcal{H}}\right|^2}{\sum_{K\cap\Omega_0\neq\emptyset}\theta_K^2} = \frac{\left|\sum_{K\cap\Omega_0\neq\emptyset}\|u\|_{H_k^{-j}(K)}(u,\varphi_K)_{\mathcal{H}}\right|^2}{\sum_{K\cap\Omega_0\neq\emptyset}\|u\|^2_{H_k^{-j}(K)}}.$$

Taking the supremum over φ_K in the right-hand side, we obtain the result. □

Lemma 5.3 *Let B_i, $i = 1,\dots,N$ be open sets and let C_{over} be as in (3.5). Then,*

$$\sum_{i=1}^N \|u\|^2_{(\mathcal{W}_k^{j,<}(B_i))^*} \leq C_{\text{over}}\|u\|^2_{(\mathcal{W}_k^{j,<}(\cup_{i=1}^N B_i))^*} \tag{5.12}$$

Proof The proof is very similar to that of Lemma 5.2, with the following modifications. The function φ_i is now an arbitrary element of $\mathcal{W}_k^{j,<}(B_i)$ with unit $\mathcal{W}_k^j$ norm, and $\theta_i := \|u\|_{\mathcal{W}_k^j}$. We now let

$$\varphi := \sum_i \theta_i\varphi_i, \quad \text{so that} \quad \|\varphi\|^2_{\mathcal{W}_k^j} \leq C_{\text{over}}\sum_{i=1}^N \theta_i^2$$

by Assumption 3.3, and the rest of the proof is unchanged.

Remark 5.4 *It is clear from the proof of* Lemma 5.2 *that the bound also holds when the $\|\cdot\|_{(\mathcal{Z}_k^{j,<}(U))^*}$ norm is defined with a supremum ranging over the smaller subset of functions supported in U, instead of all functions in $\mathcal{Z}_k^{j,<}(U)$. As a consequence,* Theorem 4.2 *also holds with this changed definition of $\|\cdot\|_{(\mathcal{Z}_k^{j,<}(U))^*}$.*

Lemma 5.5 (Caccioppoli estimate in negative norms) *Given $C_{\text{qu}}, c > 0$, there exists a constant $C'_{\text{ca}} > 0$ (whose value depends on the constants appearing in the assumptions of* Sect. 3*) such that the following holds. Let V_h satisfy the assumptions of* Sect. 3*, and let the sets $U_0 \subset U_1 \subset \Omega$ be as in (3.15) (so that $d = \partial_<(U_0,U_1)$) with*

$$d \geq 12(\ell+2)\max\{1,8\kappa\}\max_{K\cap U_1\neq\emptyset} h_K, \qquad r,d \geq ck^{-1} \quad \text{and} \quad r+d \leq \frac{c_{\text{coer}}}{k}, \tag{5.13}$$

and

$$\frac{\max_{K\cap U_1\neq\emptyset} h_K}{\min_{K\cap U_1\neq\emptyset} h_K} \leq C_{\text{qu}}. \tag{5.14}$$

If $z_h \in V_h$ satisfies (5.3), then

$$\|z_h\|_{\mathcal{Z}_k(U_0)} \leq C'_{\text{ca}}\left(\frac{1}{kd}\right)^{\alpha_*}\|z_h\|_{(\mathcal{W}_k^{s+1,<}(U_1))^*},$$

where $s := \min\{\ell, p-1\}$ and $\alpha_* = (s+2)(s+3)/2$.

Proof Let $\widetilde{U}_0 := B(x_0, r + d/2) \cap \Omega$. Let $\widetilde{d} := \partial_<(\widetilde{U}_0, U_1)$ and note that $\widetilde{d} = d/2$. Later in the proof, we apply Corollary 3.8 with $U_0 \to \widetilde{U}_0$, $U_1 \to U_1$, $d \to \widetilde{d} = d/2$ and $r \to r + d/2$. Note that $r + d \to r + d$, so that the condition $r + d \le c_{\mathrm{coer}}/k$ remains the same.

By Lemma 5.1, it suffices to show that for $j = 0, \dots, s := \min\{\ell, p-1\}$,

$$\|z_h\|_{(\mathcal{W}_k^{j,<}(U_j'))^*} \le C \left(\frac{1}{k\widetilde{d}}\right)^{j+2} \|z_h\|_{(\mathcal{W}_k^{j+1,<}(U_{j+1}'))^*} \tag{5.15}$$

where the sets

$$U_j' := B\left(x_0, r + \frac{j+1}{\ell+2}\widetilde{d}\right) \cap \Omega$$

so that

$$U_0 \subset U_0' \subset U_1' \subset \ldots \subset U_{\ell+1}' = \widetilde{U}_0.$$

In proving (5.15), we use five nested sets between U_j' and U_{j+1}'; we therefore let

$$U_{j+\nu/6}' := B\left(x_0, r + \frac{j+1+\nu/6}{\ell+2}\widetilde{d}\right) \cap \Omega, \quad \text{for } \nu = 0, \dots, 5.$$

Since $\partial_<(U_{j+\nu/6}', U_{j+(\nu+1)/6}') = (6(\ell+2))^{-1}\widetilde{d} = (6(\ell+2))^{-1}d/2$, the first condition in (5.13) implies that

$$\partial_<(U_{j+\nu/6}', U_{j+(\nu+1)/6}') > \max\left\{8\kappa, 1\right\} \max_{K \cap U_1 \neq \emptyset} h_K \quad \text{for } \nu = 0, \dots, 5. \tag{5.16}$$

We introduce the localiser $\chi = \psi(U_{j+1/6}', U_{j+2/6}')$. Let $v \in \mathcal{W}_k^{j,<}(U_j')$ and note that $\chi v = v$. Let $\mathcal{R}^*$ be the solution operator on U_1 defined in Corollary 3.8. Then,

$$|(z_h, v)_{\mathcal{H}}| = |(\chi z_h, v)_{\mathcal{H}}| = \left|a_k(\chi z_h, \mathcal{R}^* v)\right|.$$

By the orthogonality (5.3), for all $w_h \in V_h$,

$$a_k(\chi z_h, \mathcal{R}^* v) = a_k(z_h, \chi\mathcal{R}^* v - w_h) + \Big(a_k(\chi z_h, \mathcal{R}^* v) - a_k(z_h, \chi\mathcal{R}^* v)\Big). \tag{5.17}$$

By Assumption 3.10, (5.16), and the fact that $\mathrm{supp}\chi \subset U_{j+2/6}'$, we can choose $w_h \in V_h^<(U_{j+3/6}')$ as an approximation of $\chi\mathcal{R}^* v$ satisfying (3.24). Using (in this order) the locality of a_k ((3.23) in Assumption 3.9), the Cauchy–Schwarz inequality, the approximation property (3.24) (noting that, by the definition of s, $s + 2 \le p + 1$), (3.17) (with j replaced by $j + 2$), and the elliptic regularity for $\mathcal{R}^*$ (Corollary 3.8), we find

$$\begin{aligned}
&\left|a_k(z_h, \chi\mathcal{R}^* v - w_h)\right| \le C \sum_{K \in \mathcal{T}} \|z_h\|_{H_k^1(K)} \left\|\chi\mathcal{R}^* v - w_h\right\|_{H_k^1(K)} \\
&\le C \Bigg(\sum_{K \cap U_{j+3/6}' \neq \emptyset} (h_K k)^{2(j+1)} \|z_h\|_{H_k^1(K)}^2 \Bigg)^{1/2} \Bigg(\sum_{K \cap U_{j+3/6}' \neq \emptyset} (h_K k)^{-2(j+1)} \left\|\chi\mathcal{R}^* v - w_h\right\|_{H_k^1(K)}^2 \Bigg)^{1/2}
\end{aligned}$$

$$\leq C(hk)^{j+1}\Big(\sum_{K\cap U'_{j+3/6}\neq\emptyset}\|z_h\|^2_{H^1_k(K)}\Big)^{1/2}\big\|\chi\mathcal{R}^*v\big\|_{\mathcal{Z}_k^{j+2}}$$
$$\leq C\left(\frac{1}{k\widetilde{d}}\right)^{j+2}(hk)^{j+1}\|z_h\|_{\mathcal{Z}_k(U'_{j+4/6})}\|v\|_{\mathcal{W}_k^j}. \tag{5.18}$$

where $h := \max_{K\cap U'_{j+3/6}\neq\emptyset} h_K$ and we have used (5.16) in the last step.

We then use Lemma 5.1 to bound $\|z_h\|_{\mathcal{Z}_k(U'_{j+4/6})}$ by $\|z_h\|_{\mathcal{H}(U'_{j+5/6})}$. The condition (5.1) now becomes

$$\widetilde{d} \geq 6(\ell+2)\max\{1, 8\kappa\}\max_{K\cap U'_{5/6}\neq\emptyset} h_K,$$

which is satisfied by the first condition in (5.13). By the second condition in (5.13), the condition (5.2) is satisfied; i.e., the assumptions of Lemma 5.1 are satisfied.

We next use the local quasi-uniformity assumption (5.14), the inverse estimate (3.30) (noting that $s+1\leq p$) and (5.11) to obtain

$$(hk)^{2(j+1)}\|z_h\|^2_{\mathcal{H}(U'_{j+5/6})} \leq C\sum_{K\cap U'_{j+5/6}\neq\emptyset}(h_Kk)^{2(j+1)}\|z_h\|^2_{L^2(K)} \leq C\sum_{K\cap U'_{j+5/6}\neq\emptyset}\|z_h\|^2_{H_k^{-(j+1)}(K)}$$
$$\leq C\|z_h\|^2_{(\mathcal{Z}_k^{j+1,<}(U'_{j+1}))^*}, \tag{5.19}$$

where, in using (5.11), we have used that $\partial_<(U'_{j+5/6}, U'_{j+1}) > \max_{K\cap U_1\neq\emptyset} h_K$ by (5.16). Combining (5.18) and (5.19), we obtain the following bound on the first term on the right-hand side of (5.17):

$$\big|a_k(z_h, \chi\mathcal{R}^*v - w_h)\big| \leq C\left(\frac{1}{k\widetilde{d}}\right)^{j+2}\|z_h\|_{(\mathcal{Z}_k^{j+1,<}(U'_{j+1}))^*}\|v\|_{\mathcal{W}_k^j},$$
$$\leq C\left(\frac{1}{k\widetilde{d}}\right)^{j+2}\|z_h\|_{(\mathcal{W}_k^{j+1,<}(U'_{j+1}))^*}\|v\|_{\mathcal{W}_k^j}, \tag{5.20}$$

where we have used (3.12) in the last step.

To bound the second term on the right-hand side of in (5.17), we use (3.18) (with j replaced by $j+2$), and the mapping properties of $\mathcal{R}^*$ from Corollary 3.8 to find that

$$\Big|a_k(\chi z_h, \mathcal{R}^*v) - a_k(z_h, \chi\mathcal{R}^*v)\Big|$$
$$\leq \frac{C_{\mathrm{com}}}{k\widetilde{d}}\left(\sum_{r=0}^{j+1}(k\widetilde{d})^{-r}\right)\|z_h\|_{(\mathcal{W}_k^{j+1,<}(U'_{j+2/6}\setminus\overline{U'_{j+1/6}}))^*}\big\|\mathcal{R}^*v\big\|_{\mathcal{Z}_k^{j+2}(U'_{j+2/6}\setminus\overline{U'_{j+1/6}})}$$
$$\leq \frac{C}{k\widetilde{d}}\left(\frac{1}{k\widetilde{d}}\right)^{j+1}\|z_h\|_{(\mathcal{W}_k^{j+1,<}(U'_{j+1}))^*}\|v\|_{\mathcal{W}_k^j}. \tag{5.21}$$

Combining (5.17), (5.20), and (5.21), we obtain that

$$|(z_h, v)_{\mathcal{H}}| \le C \left(\frac{1}{k\widetilde{d}}\right)^{j+2} \|z_h\|_{(\mathcal{W}_k^{j+1,<}(U'_{j+1}))^*} \|v\|_{\mathcal{W}_k^j}$$

for all $v \in \mathcal{W}_k^{j,<}(U'_j)$, which implies the result (5.15). □

6 Proofs of Theorems 4.1 and 4.2

Lemma 6.1 (Analogue of Theorem 4.1 for small balls close together) *Given positive constants C_{cont}, C_{coer}, c_{coer}, κ, p, C_{inv}, C_{pw}, C_{super}, $C_{\mathrm{com}} > 0$, there exists a constant $C_\star > 0$ such that the following holds. Let $(a_k)_{k\ge k_0}$ and $V_h \subset \mathcal{Z}_k$ satisfy* Assumptions 3.5, 3.7, 3.10, 3.11, 3.12, and 3.23, *with the constants above. For $x_0 \in \Omega$, let*

$$\Omega_0 = B(x_0, d/4) \cap \Omega \quad \text{and} \quad \Omega_1 = B(x_0, 3d/4) \cap \Omega \tag{6.1}$$

with

$$d \le \frac{4}{3}\frac{c_{\mathrm{coer}}}{k} \quad \text{and} \quad d \ge 4\max\{1, 8\kappa\} \max_{K\cap\Omega_1\neq\emptyset} h_K. \tag{6.2}$$

If $k \ge k_0$, $u \in \mathcal{Z}_k$, and $u_h \in V_h$ are such that

$$a_k(u - u_h, v_h) = 0 \quad \text{for all } v_h \in V_h^<(\Omega_1), \tag{6.3}$$

then

$$\|u - u_h\|_{\mathcal{Z}_k(\Omega_0)} \le \frac{C}{kd}\left(\|u\|_{\mathcal{Z}_k(\Omega_1)} + \frac{1}{kd}\|u\|_{\mathcal{H}(\Omega_1)} + \|u - u_h\|_{\mathcal{H}(\Omega_1)}\right). \tag{6.4}$$

Proof *of Theorem 4.1 using Lemma 6.1.* We first show using a covering argument that Lemma 6.1 implies that the bound (6.4) holds for general sets $\Omega_0 \subset \Omega_1 \subset \Omega$ satisfying the assumptions of Theorem 4.1, i.e., (4.1). First, we find a subset $\Omega_1' \subset \Omega_1$ such that

$$d' = \partial_<(\Omega_0, \Omega_1') = \min\left\{\frac{C_0}{2}, \frac{4c_{\mathrm{coer}}}{3}\right\}\frac{1}{k}. \tag{6.5}$$

This definition implies that $d' \le 4c_{\mathrm{coer}}/(3k)$, and also that $d' \le d/2$ (by the first condition in (4.1)), so that Ω_1' is indeed a subset of Ω_1. Observe that (6.5) and the second condition in (4.1) imply that

$$d' \ge 4\max\{1, 8\kappa\} \max_{K\cap\Omega_1\neq\emptyset} h_K;$$

i.e., the inequalities in (6.2) are satisfied with d replaced by d'.

Next, we introduce $x_1, \dots, x_N \in \Omega_0$ such that

$$\Omega_0 \subset \bigcup_{j=1}^{N} \Big(B(x_j, d'/4) \cap \Omega \Big) \subset \bigcup_{j=1}^{N} \Big(B(x_j, 3d'/4) \cap \Omega \Big) \subset \Omega_1', \tag{6.6}$$

and such that the intersection between m distinct balls is empty when $m \geq C$, for some constant C depending only on the space dimension n. Note that the intersections with Ω are needed when Ω_0 is near the boundary of Ω.

We now apply Lemma 6.1 with d replaced by d', and thus

$$\Omega_0 = B(x_j, d'/4) \cap \Omega, \quad \text{and} \quad \Omega_1 = B(x_j, d'/4) \cap \Omega.$$

Note that the orthogonality assumption (4.2) on the large domain Ω_1 implies the analogous orthogonality (6.3) on each ball. Therefore,

$$\begin{aligned} &\|u - u_h\|_{\mathcal{Z}_k(B(x_j,d'/4))} \\ &\quad \leq \frac{C}{kd'} \Big(\|u\|_{\mathcal{Z}_k(B(x_j,3d'/4))} + \frac{1}{kd'} \|u\|_{\mathcal{H}(B(x_j,3d'/4))} + \|u - u_h\|_{\mathcal{H}(B(x_j,3d'/4))} \Big). \end{aligned}$$

Summing with respect to j and using (3.6)-(3.7),

$$\|u - u_h\|_{\mathcal{Z}_k(\Omega_0)} \leq \frac{C}{kd'} \Big(\|u\|_{\mathcal{Z}_k(\Omega_1')} + \frac{1}{kd'} \|u\|_{\mathcal{H}(\Omega_1')} + \|u - u_h\|_{\mathcal{H}(\Omega_1')} \Big) \tag{6.7}$$

By (6.5), the instances of $(kd')^{-1}$ on the right-hand side of (6.7) are bounded by a constant; then, by (3.10),

$$\|u - u_h\|_{\mathcal{Z}_k(\Omega_0)} \leq C \Big(\|u\|_{\mathcal{Z}_k(\Omega_1)} + \|u - u_h\|_{\mathcal{H}(\Omega_1)} \Big). \tag{6.8}$$

To obtain (4.3) from (6.8), we observe that if $u \in \mathcal{Z}_k$ and $u_h \in V_h$ satisfy the assumptions of the theorem, then so do $\widetilde{u} := u - w_h \in \mathcal{Z}_k$ and $\widetilde{u}_h := u_h - w_h \in V_h$, where $w_h \in V$ is arbitrary. Indeed, the key point is that u and u_h enter the assumptions of the theorem only via $u - u_h$, and $u - u_h = \widetilde{u} - \widetilde{u}_h$. Therefore, in (6.8), the norms of u on the right-hand side can be replaced by the norms of $u - w_h$ for arbitrary $w_h \in V_h$, and this gives (4.3). □

Proof *of Lemma 6.1.* Let

$$\Omega_{1/2} := B(x_0, d/2) \cap \Omega$$

and let $\chi = \psi(\Omega_{1/2}, \Omega_1)$ be the localiser associated to $\Omega_{1/2}, \Omega_1$ via Assumption 3.7 (with $(r, d) = (d/2, d/4)$).

Let the operator $\Pi_h : \mathcal{Z}_k \to V_h^<(\Omega_1)$ be defined as the solution of the variational problem

$$a_k(\Pi_h \zeta, v_h) = a_k(\zeta, v_h) \quad \text{for all } v_h \in V_h^<(\Omega_1). \tag{6.9}$$

Since $3d/4 < c_{\mathrm{coer}}k^{-1}$ (by (6.2)), a_k is continuous and coercive on $V_h^<(\Omega_1)$ by (3.13) and (3.14), and thus Π_h is well-defined by the Lax-Milgram lemma.

By the definition of χ and the triangle inequality,

$$\begin{aligned}\|u-u_h\|_{\mathcal{Z}_k(\Omega_0)} = \|\chi u - u_h\|_{\mathcal{Z}_k(\Omega_0)} &\le \|\chi u - \Pi_h(\chi u)\|_{\mathcal{Z}_k(\Omega_0)} + \|\Pi_h(\chi u) - u_h\|_{\mathcal{Z}_k(\Omega_0)} \\ &\le \|(\mathrm{Id}-\Pi_h)(\chi u)\|_{\mathcal{Z}_k} + \|z_h\|_{\mathcal{Z}_k(\Omega_0)}, \end{aligned} \tag{6.10}$$

where $z_h := \Pi_h(\chi u) - u_h$. To bound the first term on the right-hand side of (6.10), we use Céa's lemma, which follows from the continuity and coercivity of a_k on $V_h^<(\Omega_1)$, to obtain

$$\|(\mathrm{Id}-\Pi_h)(\chi u)\|_{\mathcal{Z}_k} \le C \inf_{w_h \in V_h^<(\Omega_1)} \|\chi u - w_h\|_{\mathcal{Z}_k} \le C\|\chi u\|_{\mathcal{Z}_k} \le C\Big(\|u\|_{\mathcal{Z}_k(\Omega_1)} + \frac{1}{kd}\|u\|_{\mathcal{H}(\Omega_1)}\Big), \tag{6.11}$$

where we have used (3.17) in the last inequality.

We now bound $\|z_h\|_{\mathcal{Z}_k(\Omega_0)}$ in (6.10) using the Caccioppoli inequality (5.4) applied with $U_0 = \Omega_0$ and $U_1 = \Omega_{1/2}$. The distance between these two sets is $d/4$, and so the condition (5.1) is ensured by the second condition in (6.2) since $K \cap \Omega_{1/2} \subset K \cap \Omega_1$. The condition (5.2) becomes that $d/4 + d/8 \le c_{\mathrm{coer}}/k$, and is thus ensured by the first condition in (6.2). Then, the definition $z_h := \Pi_h(\chi u) - u_h$, the definition of Π_h (6.9), the fact that $\chi \equiv 1$ on $\Omega_{1/2}$, the locality of a_k ((3.23) in Assumption 3.9), and the orthogonality (6.3) imply that

$$a(z_h, v_h) = a\big(\Pi_h(\chi u) - u_h, v_h\big) = a(\chi u - u_h, v_h) = 0 \quad \text{for all } v_h \in V_h^<(\Omega_{1/2}),$$

i.e., (5.3) holds (with $U_1 = \Omega_{1/2}$). Therefore, by (5.4),

$$\|z_h\|_{\mathcal{Z}_k(\Omega_0)} \le \frac{C}{kd}\|z_h\|_{\mathcal{H}(\Omega_{1/2})}. \tag{6.12}$$

Using (in this order) the definition of z_h, the triangle inequality, the fact that the $\mathcal{Z}_k$ norm is stronger than the $\mathcal{H}$ norm, and (6.11), we find that

$$\begin{aligned}\|z_h\|_{\mathcal{H}(\Omega_{1/2})} \le \|u-u_h\|_{\mathcal{H}(\Omega_{1/2})} + \|u-\Pi_h(\chi u)\|_{\mathcal{H}(\Omega_{1/2})} &\le C\Big(\|u-u_h\|_{\mathcal{H}(\Omega_1)} + \|(\mathrm{Id}-\Pi_h)(\chi u)\|_{\mathcal{Z}_k}\Big) \\ &\le C\Big(\|u-u_h\|_{\mathcal{H}(\Omega_1)} + \|u\|_{\mathcal{Z}_k(\Omega_1)} + \frac{1}{kd}\|u\|_{\mathcal{H}(\Omega_1)}\Big). \end{aligned} \tag{6.13}$$

The bound (6.4) then follows from combining (6.10), (6.11), (6.12), and (6.13). □

Lemma 6.2 (Analogue of Theorem 4.2 for small balls close together) *Given positive constants C_{cont}, C_{coer}, c_{coer}, κ, p, C'_{inv}, C_{pw}, C'_{super}, C_{com}, $C_{\mathrm{qu}} > 0$, there exists a constant $C_\star > 0$ such that the following holds. Let $(a_k)_{k\ge k_0}$ and $V_h \subset \mathcal{Z}_k$ satisfy* Assumptions 3.5, 3.6, 3.7, 3.10, 3.11, 3.12, and 3.23, *with the constants above. For*

$x_0 \in \Omega$, let Ω_0 and Ω_1 be as in (6.1) with

$$d \leq \frac{2c_{\mathrm{coer}}}{k} \quad \text{and} \quad d \geq 48(\ell+2)\max\{1, 8\kappa\} \max_{K \cap \Omega_1 \neq \emptyset} h_K. \tag{6.14}$$

Assume further that

$$\frac{\max_{K\cap\Omega_1\neq\emptyset} h_K}{\min_{K\cap\Omega_1\neq\emptyset} h_K} \leq C_{\mathrm{qu}}.$$

If $k \geq k_0$, $u \in \mathcal{Z}_k$, and $u_h \in V_h$ are such that (6.3) holds, then

$$\|u - u_h\|_{\mathcal{Z}_k(\Omega_0)} \leq \frac{C}{kd}\left(\|u\|_{\mathcal{Z}_k(\Omega_1)} + \frac{1}{kd}\|u\|_{\mathcal{H}(\Omega_1)} + \|u - u_h\|_{(\mathcal{W}_k^{s+1,<}(\Omega_1))^*} \right), \tag{6.15}$$

where $s := \min\{\ell, p-1\}$

Proof *of Theorem 4.2 using* Lemma 6.2. This exactly parallels the proof of Theorem 4.1 using Lemma 6.1, with the first step choosing $\Omega_1' \subset \Omega_1$ such that

$$d' = \partial_<(\Omega_0, \Omega_1') = \min\left\{ \frac{C_0}{2}, 2c_{\mathrm{coer}} \right\} \frac{1}{k}.$$

□

Proof *of Lemma 6.2.* This exactly parallels the proof of Lemma 6.1, except that now (i) we use Lemma 5.5 instead of Lemma 5.1, and (ii) when obtaining the analogue of (6.7) we use additionally (5.12). □

7 Examples of Helmholtz problems fitting in the abstract framework

Summary. In §7.1-7.3, we show that the general framework in which Theorems 4.1 and 4.2 hold includes

- truncation of the unbounded exterior domain by *either* a PML, *or* an impedance boundary condition, *or* the exact Dirichlet-to-Neumann map for the exterior of a ball,
- scattering by Dirichlet or Neumann impenetrable obstacles, and
- scattering by penetrable obstacles.

In §7.4, we show that the assumptions on the finite-dimension subspace V_h are satisfied for shape-regular Lagrange finite elements.

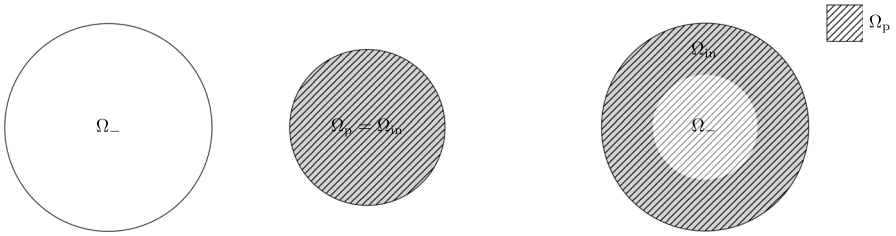


Fig. 11 The two configurations of Ω_- and $\Omega_{\rm p}$ considered. In both cases, $\Omega_{\rm p}$ is hatched, and $\Omega_{\rm in}$ is uniformly shaded

7.1 Definitions of the sesquilinear forms $a(\cdot,\cdot)$ and spaces $\mathcal{Z}_k$

7.1.1 The geometry and coefficients for scattering by a combination of an impenetrable Dirichlet or Neumann obstacle and a penetrable obstacle

Let $\Omega_{\rm p}, \Omega_- \subset B_{R_0} := \{x : |x| < R_0\} \subset \mathbb{R}^d$, $d = 2, 3$, be bounded open sets with Lipschitz boundaries, $\Gamma_{\rm p}$ and Γ_-, respectively, such that $\Gamma_{\rm p} \cap \Gamma_- = \emptyset$ and $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$ is connected. Let $\Omega_{\rm out} := \Omega_+ \setminus \overline{\Omega_{\rm p}}$ and $\Omega_{\rm in} := \Omega_+ \cap \Omega_{\rm p}$.

The obstacle $\Omega_{\rm p}$ is the penetrable obstacle, across whose boundary we impose transmission conditions, and Ω_- is the impenetrable obstacle, on which we impose either a zero Dirichlet or a zero Neumann condition. The condition $\Gamma_{\rm p} \cap \Gamma_- = \emptyset$ allows two configurations: the first, illustrated in Fig. 11a, is when the penetrable and impenetrable obstacles are disjoint. The second, illustrated in Fig. 11b, is when the impenetrable obstacle is inside the penetrable obstacle.

For simplicity, we do not cover the case when Ω_- is disconnected, with Dirichlet boundary conditions on some connected components and Neumann boundary conditions on others, but the main results hold for this problem too (at the cost of introducing more notation).

Let $A_{\rm out} \in C^{0,1}(\Omega_{\rm out}, \mathbb{R}^{d\times d})$ and $A_{\rm in} \in C^{0,1}(\Omega_{\rm in}, \mathbb{R}^{d\times d})$ be symmetric positive definite, let $c_{\rm out} \in L^\infty(\Omega_{\rm out}; \mathbb{R})$, $c_{\rm in} \in L^\infty(\Omega_{\rm in}; \mathbb{R})$ be strictly positive, and let $A_{\rm out}$ and $c_{\rm out}$ be such that there exists $R_{\rm scat} > R_0 > 0$ such that

$$\overline{\Omega_-} \cup \operatorname{supp}(I - A_{\rm out}) \cup \operatorname{supp}(1 - c_{\rm out}) \Subset B_{R_{\rm scat}}.$$

Let

$$A_{\rm scat} := \begin{cases} A_{\rm in} & \text{in } \Omega_{\rm in}, \\ A_{\rm out} & \text{in } \Omega_{\rm out}, \end{cases} \quad \text{and} \quad \frac{1}{c_{\rm scat}^2} := \begin{cases} c_{\rm in}^{-2} & \text{in } \Omega_{\rm in}, \\ c_{\rm out}^{-2} & \text{in } \Omega_{\rm out} \end{cases}.$$

7.1.2 The scattering problem

Given $g \in L^2(\Omega_+)$ (where recall that $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$) with $\operatorname{supp} g \subset B_{R_{\rm scat}}$ and $k, \gamma > 0$, let $u = (u_{\rm in}, u_{\rm out})$ be the solution of

$$k^{-2} c_{\rm out}^2 \nabla \cdot (A_{\rm out} \nabla u_{\rm out}) + u_{\rm out} = -g \quad \text{in } \Omega_{\rm out}, \tag{7.1a}$$

$$k^{-2}c_{\text{in}}^2\nabla\cdot(A_{\text{in}}\nabla u_{\text{in}})+u_{\text{in}}=-g \quad \text{in } \Omega_{\text{in}}, \tag{7.1b}$$

$$u_{\text{in}}=u_{\text{out}} \quad \text{and} \quad \partial_{n,A_{\text{in}}}u_{\text{in}}=\gamma\,\partial_{n,A_{\text{out}}}u_{\text{out}} \quad \text{on } \Gamma_{\text{p}}, \tag{7.1c}$$

$$\text{either} \quad u_{\text{in}}=0 \quad \text{or} \quad \partial_{n,A_{\text{in}}}u_{\text{in}}=0 \quad \text{on } \Gamma_- \quad \text{if } \Omega_-\subset\Omega_{\text{in}}, \text{ or}, \tag{7.1d}$$

$$\text{either} \quad u_{\text{out}}=0 \quad \text{or} \quad \partial_{n,A_{\text{out}}}u_{\text{out}}=0 \quad \text{on } \Gamma_- \quad \text{if } \Omega_-\subset\Omega_{\text{out}}, \tag{7.1e}$$

where $u_{\text{in}}\in H^1(\Omega_{\text{in}})$, $u_{\text{out}}\in H^1(\Omega_{\text{out}}\cap B_R)$ for every $R>0$, and u_{out} satisfies the Sommerfeld radiation condition

$$k^{-1}\frac{\partial u_{\text{out}}}{\partial r}(x)-\mathrm{i}u_{\text{out}}(x)=o\Big(\frac{1}{r^{(d-1)/2}}\Big) \tag{7.2}$$

as $r:=|x|\to\infty$ (uniformly in $\widehat{x}:=x/r$). The solution of this problem exists and is unique; see, e.g., [16] and the references therein.

7.1.3 The variational formulation

Given $R>R_{\text{scat}}$, let

$$\Omega:=\Omega_{\text{in}}\cup(\Omega_{\text{out}}\cap B_R)$$

and let

$$\mathcal{Z}_k:=\Big\{v\in H^1(\Omega)\,:\,v=0 \text{ on } \Gamma_-\Big\} \quad \text{or} \quad H^1(\Omega), \tag{7.3}$$

in both cases equipped with the H^1_k norm defined by (8.3), with the former space corresponding to zero Dirichlet boundary conditions on Γ_- and the latter corresponding to zero Neumann boundary conditions on Γ_-.

Let $\mathrm{DtN}_k : H^{1/2}(\partial B_R)\to H^{-1/2}(\partial B_R)$ be the Dirichlet-to-Neumann map, $u\mapsto k^{-1}\partial_r u$, for the Helmholtz equation $(k^{-2}\Delta+1)u=0$ posed in the exterior of B_R and satisfying the Sommerfeld radiation condition (7.2); i.e., when $d=2$, given $g\in H^{1/2}(\partial B_R)$,

$$\mathrm{DtN}_k g(\varphi):=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\frac{H_n^{(1)'}(kR)}{H_n^{(1)}(kR)}\exp(\mathrm{i}n\varphi)\int_0^{2\pi}\exp(-\mathrm{i}n\theta)g(R,\theta)\,\mathrm{d}\theta; \tag{7.4}$$

for the analogous expression when $d=3$, see, e.g., e.g., [5, Equation 3.6], [25, Equation 3.7].

The variational formulation of the scattering problem (7.1)-(7.2) is

$$\text{find } u\in\mathcal{Z}_k \text{ such that } a(u,v)=G(v) \text{ for all } v\in\mathcal{Z}_k, \tag{7.5}$$

where

$$a(u,v):=\left(\int_{\Omega_{\text{out}}\cap B_R}+\frac{1}{\gamma}\int_{\Omega_{\text{in}}}\right)\Big(k^{-2}(A_{\text{scat}}\nabla u)\cdot\overline{\nabla v}-c_{\text{scat}}^{-2}u\overline{v}\Big)-k^{-1}\big\langle\mathrm{DtN}_k u,\,v\big\rangle_{\partial B_R} \tag{7.6}$$

and

$$G(v) := \left(\int_{\Omega_{\text{out}} \cap B_R} + \frac{1}{\gamma} \int_{\Omega_{\text{in}}} \right) c^{-2} g \overline{v}. \tag{7.7}$$

7.1.4 Approximation of DtN_k by an impedance boundary condition

A commonly-used approximation of DtN_k is to impose that $k^{-1}\partial_r u = \mathrm{i}u$ on ∂B_R, i.e., impose an impedance boundary condition. In this case, one also often removes the requirement that the outer boundary is a ball. Let $R_{\text{tr}} > R_{\text{scat}}$, let $\Omega_{\text{tr}} \subset \mathbb{R}^d$ be a bounded Lipschitz open set with $B_{R_{\text{tr}}} \subset \Omega_{\text{tr}} \subset B_{CR_{\text{tr}}}$ for some $C > 0$ (i.e., Ω_{tr} has characteristic length scale R_{tr}), and let $\Gamma_{\text{tr}} := \partial\Omega_{\text{tr}}$. The impedance problem is (7.5) with now (7.6) replaced by

$$a(u, v) := \left(\int_{\Omega_{\text{out}} \cap \Omega_{\text{tr}}} + \frac{1}{\gamma} \int_{\Omega_{\text{in}}} \right) \left(k^{-2} (A_{\text{scat}} \nabla u) \cdot \overline{\nabla v} - c_{\text{scat}}^{-2} u \overline{v} \right) - \mathrm{i}k^{-1} \langle u, v \rangle_{\Gamma_{\text{tr}}}, \tag{7.8}$$

(7.7) replaced by the analogous expression with integration over $\Omega_{\text{out}} \cap B_R$ replaced by integration over $\Omega_{\text{out}} \cap \Omega_{\text{tr}}$, and $\mathcal{Z}_k$ still defined by (7.3), but now with

$$\Omega = \Omega_{\text{in}} \cup (\Omega_{\text{out}} \cap \Omega_{\text{tr}}). \tag{7.9}$$

See [11] for k-explicit bounds on the error incurred by this approximation (showing, in particular, how the error depends on $\partial\Omega_{\text{tr}}$).

7.1.5 Approximation of DtN_k by a radial PML

Let $R_{\text{tr}} > R_{\text{PML},-} > R_{\text{scat}}$, let Ω_{tr} and $\Gamma_{\text{tr}} := \partial\Omega_{\text{tr}}$ be as above, and let Ω be defined by (7.9). For $0 \le \theta < \pi/2$, let the PML scaling function $f_\theta \in C^1([0, \infty); \mathbb{R})$ be defined by $f_\theta(r) := f(r) \tan\theta$ for some f satisfying

$$\{f(r) = 0\} = \{f'(r) = 0\} = \{r \le R_{\text{PML},-}\}, \quad f'(r) \ge 0, \quad f(r) \equiv r \text{ on } r \ge R_{\text{PML},+}; \tag{7.10}$$

i.e., the scaling "turns on" at $r = R_{\text{PML},-}$, and is linear when $r \ge R_{\text{PML},+}$. We emphasise that R_{tr} can be $< R_{\text{PML},+}$, i.e., we allow truncation before linear scaling is reached. Indeed, $R_{\text{PML},+} > R_{\text{PML},-}$ can be arbitrarily large and therefore, given any bounded interval $[0, R]$ and any function $\widetilde{f} \in C^1([0, R])$ satisfying

$$\{\widetilde{f}(r) = 0\} = \{\widetilde{f}'(r) = 0\} = \{r \le R_{\text{PML},-}\}, \qquad \widetilde{f}'(r) \ge 0,$$

we can choose an f with $f|_{[0,R]} = \widetilde{f}$. Given $f_\theta(r)$, let

$$\alpha(r) := 1 + \mathrm{i} f_\theta'(r) \quad \text{and} \quad \beta(r) := 1 + \mathrm{i} f_\theta(r)/r, \tag{7.11}$$

and let

$$A := \begin{cases} A_{\text{in}} & \text{in } \Omega_{\text{in}}, \\ A_{\text{out}} & \text{in } \Omega_{\text{out}} \cap B_{R_{\text{PML},-}}, \\ HDH^T & \text{in } (B_{R_{\text{PML},-}})^c \end{cases} \text{ and } \frac{1}{c^2} := \begin{cases} c_{\text{in}}^{-2} & \text{in } \Omega_{\text{in}}, \\ c_{\text{out}}^{-2} & \text{in } \Omega_{\text{out}} \cap B_{R_{\text{PML},-}}, \\ \alpha(r)\beta(r)^{d-1} & \text{in } (B_{R_{\text{PML},-}})^c, \end{cases} \tag{7.12}$$

where, in polar coordinates (r, φ),

$$D = \begin{pmatrix} \beta(r)\alpha(r)^{-1} & 0 \\ 0 & \alpha(r)\beta(r)^{-1} \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \text{ for } d = 2, \tag{7.13}$$

and, in spherical polar coordinates (r, ϕ, φ),

$$D = \begin{pmatrix} \beta(r)^2\alpha(r)^{-1} & 0 & 0 \\ 0 & \alpha(r) & 0 \\ 0 & 0 & \alpha(r) \end{pmatrix} \text{ and } H = \begin{pmatrix} \sin\phi\cos\varphi & \cos\phi\cos\varphi & -\sin\varphi \\ \sin\phi\sin\varphi & \cos\phi\sin\varphi & \cos\varphi \\ \cos\phi & -\sin\phi & 0 \end{pmatrix} \tag{7.14}$$

for $d = 3$. Observe that $A_{\text{out}} = I$ and $c_{\text{out}}^{-2} = 1$ when $r = R_{\text{PML},-}$ and thus A and c^{-2} are continuous at $r = R_{\text{PML},-}$.

We highlight that, in other papers on PMLs, the scaled variable, which in our case is $r + \mathrm{i} f_\theta(r)$, is often written as $r(1 + \mathrm{i}\widetilde{\sigma}(r))$ with $\widetilde{\sigma}(r) = \sigma_0$ for r sufficiently large; see, e.g., [19, §4], [2, §2]. Therefore, to convert from our notation, set $\widetilde{\sigma}(r) = f_\theta(r)/r$ and $\sigma_0 = \tan\theta$.

Let $\mathcal{Z}_k$ be still defined by (7.3), but now with Ω given by (7.9). Given $g \in L^2(\Omega)$ with $\operatorname{supp} g \subset B_{R_{\text{scat}}}$, a variational formulation of the PML problem is then (7.5) with

$$a(u, v) := \left(\int_{\Omega \cap \Omega_{\text{out}}} + \frac{1}{\gamma} \int_{\Omega \cap \Omega_{\text{in}}} \right) \left(k^{-2} (A \nabla u) \cdot \overline{\nabla v} - c^{-2} u \overline{v} \right) \tag{7.15}$$

and $G(v)$ given by (7.7); this variational formulation is obtained by multiplying the PDEs in (7.1) by $c_{\text{in/out}}^{-2} \alpha \beta^{d-1}$ and integrating by parts.

Assumption 7.1 *When $d = 3$, $f_\theta(r)/r$ is nondecreasing.*

Assumption 7.1 is standard in the literature; e.g., in the alternative notation described above it is that $\widetilde{\sigma}$ is non-decreasing – see [2, §2]. We record for later the following sign property of A under Assumption 7.1.

Lemma 7.2 *Suppose that f_θ satisfies* Assumption 7.1. *With A defined by* (7.12), *given $\epsilon > 0$ there exists $A_- > 0$ such that, for all $\epsilon \le \theta \le \pi/2 - \epsilon$,*

$$\Re\big(A(x)\xi, \xi\big)_2 \ge A_- \|\xi\|_2^2 \quad \textit{for all } \xi \in \mathbb{C}^d \textit{ and } x \in \Omega.$$

Reference for the proof. See, e.g., [13, Lemma 2.3]. □

Remark 7.3 (Existence and uniqueness of the solution of the PML problem) *Using the fact that the solution of the true scattering problem exists and is unique with*

A_{out}, A_{in}, c_{out}, c_{in}, Ω_-, and Ω_{in} described above, the solution of the PML variational formulation above exists and is unique (i) for fixed k and sufficiently large $R_{\text{tr}} - R_1$ by [21, Theorem 2.1], [22, Theorem A], [19, Theorem 5.8] *and (ii) for fixed $R_{\text{tr}} > R_1$ and sufficiently large k by* [12, Theorem 1.5] *under the additional assumption that $f_\theta \in C^3$.*

Remark 7.4 (Accuracy of the PML approximation) *For the particular data G (7.7) (i.e., coming from a function supported in $B_{R_{\text{scat}}}$), it is well-known that, for fixed k, the error $\|u - v\|_{H^1_k(B_{R_{\text{PML},-}}\setminus\Omega)}$ decays exponentially in $R_{\text{tr}} - R_{\text{PML},-}$ and $\tan\theta$; see* [21, Theorem 2.1], [22, Theorem A], [19, Theorem 5.8]. *It was recently proved in* [12, Theorems 1.2 and 1.5] *that the error $\|u - v\|_{H^1_k(B_{R_{\text{PML},-}}\setminus\Omega)}$ also decreases exponentially in k (again under the assumption that $f_\theta \in C^3$).*

7.1.6 Summary of the sesquilinear forms $a(\cdot,\cdot)$ and spaces $\mathcal{Z}_k$

For truncation by the exact Dirichlet-to-Neumann map and the truncation boundary equal to the boundary of a ball, the sesquilinear form $a(\cdot,\cdot)$ is defined by (7.6) and the space $\mathcal{Z}_k$ is defined by (7.3) with $\Omega := \Omega_{\text{in}} \cup (\Omega_{\text{out}} \cap B_R)$.

For truncation by an impedance boundary condition, the sesquilinear form $a(\cdot,\cdot)$ is defined by (7.8) and the space $\mathcal{Z}_k$ is defined by (7.3) with $\Omega := \Omega_{\text{in}} \cup (\Omega_{\text{out}} \cap \Omega_{\text{tr}})$.

For truncation by a perfectly-matched layer, the sesquilinear form $a(\cdot,\cdot)$ is defined by (7.15) and the space $\mathcal{Z}_k$ is defined by (7.3) with $\Omega := \Omega_{\text{in}} \cup (\Omega_{\text{out}} \cap \Omega_{\text{tr}})$.

7.2 The spaces $\mathcal{Z}_k^j$ and $\mathcal{W}_k^j$ and Assumption 3.3

With *either* $\Omega := \Omega_{\text{in}} \cup (\Omega_{\text{out}} \cap B_R)$ (for DtN truncation) *or* $\Omega := \Omega_{\text{in}} \cup (\Omega_{\text{out}} \cap \Omega_{\text{tr}})$ (for impedance or PML truncation), let

$$\mathcal{W}_k^j := L^2(\Omega) \cap \left(H_k^j(\Omega_{\text{in}}) \oplus H_k^j(\Omega_{\text{out}} \cap \Omega)\right) \quad \text{and} \quad \mathcal{Z}_k^j := \mathcal{W}_k^j \cap \mathcal{Z}_k. \tag{7.16}$$

The embedding inequality (3.4) immediately holds, and Assumption 3.3 holds via standard properties of the H_k^j norm.

7.3 The assumptions on the sesquilinear form (Assumptions 3.5, 3.6, and 3.7)

Lemma 7.5 (Satisfying Assumption 3.5) Assumption 3.5 *is satisfied for the sesquilinear forms defined by* (7.6), (7.8), *and* (7.15).

Proof We first establish the continuity property (3.13). For the sesquilinear form (7.15) from PML truncation, continuity follows by the Cauchy–Schwarz inequality. For the sesquilinear form (7.8) from impedance truncation, continuity follows by the Cauchy–Schwarz inequality, and the weighted trace inequality

$$\|v\|_{L^2(\partial D)} \le Ck^{1/2}\|v\|_{H^1_k(D)};$$

see, e.g., [17, Theorem 1.5.1.10, last formula on page 41]. For the sesquilinear form (7.6) from truncation by DtN_k, continuity follows by the Cauchy–Schwarz inequality, and the inequality

$$k^{-1}\left|\langle \mathrm{DtN}_k u, v\rangle_{\partial B_R}\right| \leq C\,\|u\|_{\mathcal{Z}_k}\,\|v\|_{\mathcal{Z}_k} \quad \text{for all } u, v \in \mathcal{Z}_k,$$

which holds by, e.g., [25, Equation 3.4a] (taking into account that [25] use a different k-weighting in the H^1 norm to (8.7) – see the comments after (8.7)).

For the local coercivity (3.14), we claim that, for all three sesquilinear forms, given A_{scat}, c_{scat}, and $\gamma > 0$, there exist $C_1, C_2 > 0$ such that

$$\Re\big\{a_k(v, v)\big\} \geq C_1\,\|v\|^2_{\mathcal{Z}_k} - C_2\,\|v\|^2_{\mathcal{H}} \quad \text{for all } v \in \mathcal{Z}_k \text{ and for all } k \geq k_0. \tag{7.17}$$

Once (7.17) is established, the local coercivity follows from the Poincaré–Friedrichs inequality. Indeed,

$$\|v\|^2_{\mathcal{H}} \leq C_{\mathrm{PF}}(kr)^2 \left\|k^{-1}\nabla v\right\|^2_{\mathcal{H}} \quad \text{for all } v \in \mathcal{Z}_k^{<}(B(x_0, r) \cap \Omega),$$

so that if $C_1/2 \geq C_2 C_{\mathrm{PF}}(kr)^2$ then

$$\Re\big\{a_k(v, v)\big\} \geq \frac{C_1}{2}\,\|v\|^2_{\mathcal{Z}_k} \quad \text{for all } v \in \mathcal{Z}_k \text{ and for all } k \geq k_0,$$

where we have used that $\|\cdot\|_{\mathcal{Z}_k} = \|\cdot\|_{H^1(\Omega)}$ in all three cases. Thus, (3.14) holds with

$$c_{\mathrm{coer}} := \left(\frac{C_1}{2C_2 C_{\mathrm{PF}}}\right)^{1/2} \quad \text{and} \quad C_{\mathrm{coer}} := \frac{C_1}{2}.$$

For the sesquilinear form (7.8) from impedance truncation, the proof of (7.17) is immediate. For the sesquilinear form (7.8) from PML truncation, the proof follows from Lemma 7.2. For the sesquilinear form (7.6) from truncation by DtN_k, the proof follows from the inequality $\Re\langle \mathrm{DtN}_k\phi, \phi\rangle_{\partial B_R} \leq 0$ for all $\phi \in H^{1/2}(\partial B_R)$; see [5, Second inequality in Equation 2.8], [25, Equation 3.4b]. □

Lemma 7.6 (Satisfying Assumption 3.6) *Suppose that* $A_{\mathrm{out}}, c_{\mathrm{out}} \in C^{\ell,1}(\overline{\Omega_{\mathrm{out}}})$, $A_{\mathrm{in}}, c_{\mathrm{in}} \in C^{\ell,1}(\overline{\Omega_{\mathrm{in}}})$, *the PML scaling function* f_θ *is* $C^{\ell+1,1}(\overline{\Omega})$, *and both* $\partial\Omega$ *and* Γ_{p} *are* $C^{\ell+1,1}$. *Then,* Assumption 3.6 *is satisfied for the sesquilinear forms defined by* (7.6), (7.8), *and* (7.15).

Proof By the definition of $\mathcal{W}_k^j$ and $\mathcal{Z}_k^j$, the required bound (3.16) is

$$\begin{aligned}&\|u\|_{H_k^{j+2}(U_0\cap\Omega_{\mathrm{in}})} + \|u\|_{H_k^{j+2}(U_0\cap\Omega_{\mathrm{out}})}\\ &\leq C\Big(\|u\|_{L^2(U_1)} + \left\|k^{-2}\nabla\cdot(A\nabla u) + c^{-2}u\right\|_{H_k^j(U_1\cap\Omega_{\mathrm{in}})} + \left\|k^{-2}\nabla\cdot(A\nabla u) + c^{-2}u\right\|_{H_k^j(U_1\cap\Omega_{\mathrm{out}})}\Big)\end{aligned} \tag{7.18}$$

for all $u \in \mathcal{Z}_k^<(U_1)$ and $j = 0, \dots, \ell$, where u satisfies the transmission conditions (7.1c) across Γ_p, either a Dirichlet or Neumann boundary condition on Γ_-, and either a Dirichlet or impedance boundary condition on Γ_tr.

By assumption, the characteristic length scales of both U_0 and U_1 are proportional to k^{-1}; denote this length scale (temporarily) by L.

We now claim that there exists $C > 0$ such that, for all $u \in \mathcal{Z}_k^<(U_1)$ and $j = 0, \dots, \ell$,

$$\begin{aligned}&\|u\|_{H^{j+2}(U_0\cap\Omega_\mathrm{in})} + \|u\|_{H^{j+2}(U_0\cap\Omega_\mathrm{out})}\\ &\quad\le C\Big(L^{-j-1}\,\|u\|_{H^1(U_1)} + \|\nabla\cdot(A\nabla u)\|_{H^j(U_1\cap\Omega_\mathrm{in})} + \|\nabla\cdot(A\nabla u)\|_{H^j(U_1\cap\Omega_\mathrm{out})}\Big).\end{aligned} \tag{7.19}$$

Indeed, this result without the L dependence is proved

(i) away from the boundary in, e.g., [24, Theorems 4.7, 4.16],

(ii) locally next to a Dirichlet or Neumann boundary in [24, Theorem 4.18],

(iii) locally next to a transmission boundary with $\gamma = 1$ in [24, Theorem 4.20] and for general γ in [8, Theorem 5.2.1(i)],

(iv) locally next to an impedance boundary in [14, Theorem 3.4], and

(v) locally next to ∂B_R on which $k^{-1}\partial_n u = \mathrm{DtN}_k u$ in [14, Theorem 3.5].

In all cases, the L-dependence can be inserted, either by keeping track of the constants in the proofs, or by a scaling argument (using the fact that the constant in the bound on the $\mathcal{O}(1)$ domain depends only on the $C^{\ell,1}$ norms of A_out, A_in, c_out, and c_in, the $C^{\ell+1,1}$ norm of f_θ, and the $C^{\ell+1,1}$ norms of $\partial\Omega$ and Γ_p).

Multiplying (7.19) by $k^{-(j+2)}$, we obtain that

$$\begin{aligned}&k^{-j-2}|u|_{H^{j+2}(U_0\cap\Omega_\mathrm{in})} + k^{-j-2}|u|_{H^{j+2}(U_0\cap\Omega_\mathrm{out})}\\ &\quad\le C\Big((kL)^{-j-1}\,\|u\|_{H_k^1(U_1)} + \left\|k^{-2}\nabla\cdot(A\nabla u)\right\|_{H_k^j(U_1\cap\Omega_\mathrm{in})} + \left\|k^{-2}\nabla\cdot(A\nabla u)\right\|_{H_k^j(U_1\cap\Omega_\mathrm{out})}\Big)\end{aligned}$$

for $j = 0, \dots, \ell$, and thus, using that $L = Ck^{-1}$,

$$\begin{aligned}&\|u\|_{H_k^{j+2}(U_0\cap\Omega_\mathrm{in})} + \|u\|_{H_k^{j+2}(U_0\cap\Omega_\mathrm{out})}\\ &\quad\le C\Big(\,\|u\|_{H_k^1(U_1)} + \left\|k^{-2}\nabla\cdot(A\nabla u)\right\|_{H_k^j(U_1\cap\Omega_\mathrm{in})} + \left\|k^{-2}\nabla\cdot(A\nabla u)\right\|_{H_k^j(U_1\cap\Omega_\mathrm{out})}\Big)\end{aligned} \tag{7.20}$$

for $j = 0, \dots, \ell$. Since $r + d \le c_\mathrm{coer}k^{-1}$ the coercivity (3.14) holds. We then obtain (7.18) from (7.20) by using the Lax–Milgram lemma, the triangle inequality, and the fact that multiplication by a $C^{\ell,1}$ function is continuous from H^ℓ to H^ℓ (see, e.g., [1, Theorem 7.4] with $s_1 = \ell - 1$, $s_2 = \ell$, $s = \ell$, $p_1 = \infty$, $p_2 = 2$, and $p = 2$). □

Lemma 7.7 (Satisfying Assumption 3.7) *Suppose that* $A_\mathrm{out}, c_\mathrm{out} \in C^{\ell,1}(\overline{\Omega_\mathrm{out}}) \cap C^{\ell+1,1}(\Gamma_\mathrm{p})$, $A_\mathrm{in}, c_\mathrm{in} \in C^{\ell,1}(\overline{\Omega_\mathrm{in}}) \cap C^{\ell+1,1}(\Gamma_\mathrm{p})$, *the PML scaling function* f_θ *is* $C^{\ell,1}(\overline{\Omega})$, *and both* $\partial\Omega$ *and* Γ_p *are* $C^{\ell+1,1}$. *Then, Parts (i), (ii), and (iii) of* Assumption 3.7 *are satisfied for the sesquilinear forms defined by* (7.6), (7.8), *and* (7.15).

To prove Lemma 7.7, we need the following lemma.

Lemma 7.8 *Given an open set $U \subset \mathbb{R}^d$ with outward-pointing unit-normal vector ν and $C^{m+1,1}$ compact boundary, symmetric positive-definite functions $A_{\text{in}} \in C^{m,1}(\partial U, \mathbb{R}^{d\times d})$, $A_{\text{out}} \in C^{m,1}(\partial U, \mathbb{R}^{d\times d})$, $x_0 \in \mathbb{R}^d$, and $0 < r < R$, the following is true. There exists $\psi : \mathbb{R}^d \to \mathbb{R}$ that is continuous on $\mathbb{R}^d$, $C^{m,1}$ on both $\overline{U}$ and $\mathbb{R}^d \setminus U$, and such that*

$$\psi \equiv 0 \text{ on } B(x_0, R)^c, \qquad \psi \equiv 1 \text{ on } B(x_0, r),$$

and

$$\nabla(\psi|_U) \cdot (A_{\text{in}}\nu) = \nabla(\psi|_{\mathbb{R}^d\setminus\overline{U}}) \cdot (A_{\text{out}}\nu) = 0 \quad \text{on} \quad \partial U. \tag{7.21}$$

We postpone the proof of Lemma 7.8 until after the proof of Lemma 7.7. However, we highlight here that the construction of such a ψ is possible since, by positive definiteness of A_{in} and A_{out}, $(A_{\text{in}}\nu)\cdot\nu$ and $(A_{\text{out}}\nu)\cdot\nu$ are not zero for any $x \in \partial U$; the vectors $A_{\text{in}}\nu$ and $A_{\text{out}}\nu$ are therefore never tangent to ∂U and so prescribing that ψ is constant in these directions at ∂U is possible.

Proof *of Lemma 7.7 using Lemma 7.8.* Let $\psi \in C_{\text{comp}}(\mathbb{R}^d; [0, 1])$ satisfy the following four conditions:
(1)

$$\operatorname{supp}\psi \subset B(x_0, r + 3d/4) \quad \text{and} \quad \psi \equiv 1 \quad \text{on } B(x_0, r + d/4), \tag{7.22}$$

(2)

$$\psi \in C^{\ell+1,1}\big(B(x_0, r + 3d/4) \cap \Omega_{\text{in}}\big) \cap C^{\ell+1,1}\big(B(x_0, r + 3d/4) \cap \Omega_{\text{out}} \cap \Omega_{\text{tr}}\big)$$

with $\|\partial^j\psi\|_{L^\infty} \le Cd^{-j}$ in each of the two regions for $j = 1, \dots, \ell + 2$, and
(3) with ν the outward-pointing unit-normal vector to Ω_{in},

$$(A_{\text{in}}\nabla\psi)\cdot\nu = 0 \quad \text{and} \quad (A_{\text{out}}\nabla\psi)\cdot\nu = 0 \quad \text{on } \Gamma_{\text{p}} \tag{7.23}$$

when the limits are taken from Ω_{in} and Ω_{out}, respectively, and
(4) with ν the outward-pointing unit-normal vector to Ω,

$$(A_{\text{out}}\nabla\psi)\cdot\nu = 0 \quad \text{on } \Gamma_{\text{tr}} \tag{7.24}$$

and

$$\text{either} \quad (A_{\text{in}}\nabla\psi)\cdot\nu = 0 \quad \text{on } \Gamma_- \quad \text{or} \quad (A_{\text{out}}\nabla\psi)\cdot\nu = 0 \quad \text{on } \Gamma_- \tag{7.25}$$

depending on whether Ω_- is inside Ω_{in} or Ω_{out}.

Conditions (3) and (4) imply there are at most three interfaces across which $\nabla\psi$ jumps — see Fig. 12. For each interface, we construct such a ψ by Lemma 7.8, and then a ψ satisfying the three jumps in Conditions (3) and (4), along with Conditions

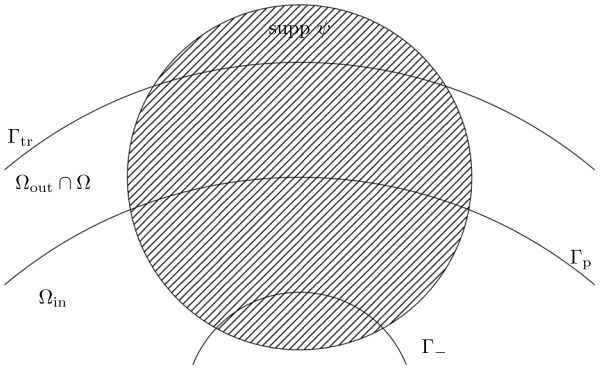


Fig. 12 The relative locations of $\Gamma_{\rm tr}$, $\Gamma_{\rm p}$, Γ_-, and supp ψ (for particular choices of x_0 and large r and d), where Ω_- and $\Omega_{\rm p}$ are as in Fig. 11b (i.e., Ω_- is inside $\Omega_{\rm in}$). The condition (7.24) is imposed on $\Gamma_{\rm tr}$, the condition (7.23) is imposed on $\Gamma_{\rm p}$, and the first condition in (7.25) is imposed on Γ_-

(1) and (2), is constructed using a partition of unity (where each of the three partition of unity functions is one near one of the three interfaces).

Part (i) of Assumption 3.7 holds from (7.22). Part (ii) of Assumption 3.7 holds by the Leibnitz rule applied piecewise in $B(x_0, r + 3d/4) \cap \Omega_{\rm in}$ and $B(x_0, r + 3d/4) \cap (\Omega_{\rm out} \cap \Omega)$.

We now check Part (iii). For all three of the sesquilinear forms,

$$\begin{aligned} &k^2\big(a_k(\psi u, v) - a_k(u, \psi v)\big) \\ &\quad = \int_{\Omega\cap\Omega_{\rm out}} A\nabla(\psi u)\cdot\overline{\nabla v} - A\nabla u\cdot\nabla(\psi\overline{v}) + \gamma^{-1}\int_{\Omega_{\rm in}} A_{\rm in}\nabla(\psi u)\cdot\overline{\nabla v} - A_{\rm in}\nabla u\cdot\nabla(\psi\overline{v}) \\ &\quad = \int_{(\Omega\cap\Omega_{\rm out})\cap(U_1\setminus U_0)} \nabla\psi\cdot\big(u(A\overline{\nabla v}) - \overline{v}(A\nabla u)\big) + \gamma^{-1}\int_{\Omega_{\rm in}\cap(U_1\setminus U_0)} \nabla\psi\cdot\big(u(A_{\rm in}\overline{\nabla v}) - \overline{v}(A_{\rm in}\nabla u)\big). \end{aligned} \tag{7.26}$$

We now prove (3.18) with the first argument of the minimum on the right-hand side; the proof for the second argument follows by swapping the roles of u and v.

We first bound the fourth term on the right-hand side of (7.26); the analogous term over $(\Omega \cap \Omega) \cap (U_1 \setminus U_0)$ (i.e., the second term on the right-hand side of (7.26)) is bounded in an identical way. We use (in the following order) the definition of $\|\cdot\|_{(\mathcal{W}_k^{j-1,<}(U_1\setminus U_0))^*}$, the analogue of (3.17) in the $\mathcal{W}_k^j$ norms (and with ψ replaced by $k^{-1}\nabla\psi$), the fact that $A_{\rm in}$ is $C^{\ell,1}$, and the fact that $k^{-1}\partial$ maps $\mathcal{Z}_k^j$ to $\mathcal{W}_k^{j-1}$ with norm bounded independent of k (with $\mathcal{W}_k^j$ and $\mathcal{Z}_k^j$ given by (7.16)) to obtain that, for $j = 1, \dots, \ell + 1$,

$$\begin{aligned} k^{-2}\left|\int_{\Omega_{\rm in}\cap(U_1\setminus U_0)} \overline{v}(\nabla\psi)\cdot(A_{\rm in}\nabla u)\right| &\le \|v\|_{(\mathcal{W}_k^{j-1,<}(U_1\setminus U_0))^*}\big\|(k^{-1}\nabla\psi)\cdot(k^{-1}A_{\rm in}\nabla u)\big\|_{\mathcal{W}_k^{j-1}(\Omega_{\rm in}\cap(U_1\setminus U_0))} \\ &\le \|v\|_{(\mathcal{W}_k^{j-1,<}(U_1\setminus U_0))^*}\frac{CC_\dagger}{kd}\Big(\sum_{m=0}^{j-1}(kd)^{-(j-1-m)}\Big)\|u\|_{\mathcal{Z}_k^j(U_1\setminus U_0)}. \end{aligned} \tag{7.27}$$

It therefore remains to bound the first and third terms on the right-hand side of (7.26). By the symmetry of A_{in} and the divergence theorem,

$$\begin{aligned}\int_{\Omega_{\text{in}}\cap(U_1\setminus U_0)} u(\nabla\psi)\cdot(A_{\text{in}}\overline{\nabla v}) &= \int_{\Omega_{\text{in}}\cap(U_1\setminus U_0)} (\overline{\nabla v})\cdot(uA_{\text{in}}\nabla\psi)\\ &= \int_{\partial(\Omega_{\text{in}}\cap(U_1\setminus U_0))} u\overline{v}(A_{\text{in}}\nabla\psi)\cdot\nu - \int_{\Omega_{\text{in}}\cap(U_1\setminus U_0)} \overline{v}\nabla\cdot(uA_{\text{in}}\nabla\psi).\end{aligned}$$

We now claim that the boundary integral over $\partial(\Omega_{\text{in}}\cap(U_1\setminus U_0))$ is equal to zero. This boundary integral can be split into integrals over $\partial(U_1\setminus U_0)$, Γ_{p}, Γ_-, and Γ_{tr}. (Note that, since U_1 has characteristic length scale k^{-1} and Γ_-, Γ_{p}, and Γ_{tr} are all a k-independent distance apart, there will never be boundary integrals over any two of Γ_-, Γ_{p}, and Γ_{tr} at the same time for k sufficiently large.) The integrals over $U_1\setminus U_0$, vanish because $\nabla\psi$ is zero here, and the integrals over Γ_{p} and $\partial\Omega$ vanish because of the first conditions in (7.23) and (7.24).

Therefore

$$\begin{aligned}k^{-2}\int_{\Omega_{\text{in}}\cap(U_1\setminus U_0)} u(\nabla\psi)\cdot(A_{\text{in}}\overline{\nabla v}) &= k^{-2}\int_{\Omega_{\text{in}}\cap(U_1\setminus U_0)} \overline{v}\nabla\cdot(uA_{\text{in}}\nabla\psi)\\ &= k^{-2}\int_{\Omega_{\text{in}}\cap(U_1\setminus U_0)} \Big(\overline{v}(\nabla u)\cdot(A_{\text{in}}\nabla\psi) + \overline{v}u\nabla\cdot(A_{\text{in}}\nabla\psi)\Big).\end{aligned}$$

The first term on the right-hand side is bounded exactly as in (7.27); the second term is bounded similarly, and the proof is complete. □

It therefore remains to prove Lemma 7.8.

Proof *of Lemma 7.8.* Let $\chi\in C^\infty_{\text{comp}}(\mathbb{R}^d)$ be supported in $B(x_0,R')$ and identically 1 in $B(x_0,r')$ with $r<r'<R'<R$.

Since $A_{\text{in}}\nu$ and $A_{\text{out}}\nu$ are uniformly transverse to ∂U, we now claim that there exists a neighborhood U_ε of ∂U in $\mathbb{R}^d$ such that

$$X_{\text{in}}:\partial U\times[-\varepsilon,\varepsilon]\to U_\varepsilon,\quad X_{\text{in}}(x,t):=x+tA_{\text{in}}(x)\nu(x)$$

and

$$X_{\text{out}}:\partial U\times[-\varepsilon,\varepsilon]\to U_\varepsilon,\quad X_{\text{out}}(x,t):=x+tA_{\text{out}}(x)\nu(x)$$

are bijections. Indeed, this follows from (a) the inverse function theorem (valid for Lipschitz maps — and hence for $A_{\text{in/out}}\in C^{0,1}$ and $\nu\in C^{1,1}$ – by [7, Theorem 1]) and (b) the fact that, for ε small, $X_{\text{in/out}}$ are injective.

Furthermore, since X_{in} maps into U for $t<0$ and X_{out} maps into $\mathbb{R}^d\setminus\overline{U}$ for $t>0$, we define

$$X(x,\varepsilon):=\begin{cases}X_{\text{in}}(x,t), & t<0,\\ X_{\text{out}}(x,t), & t\geq 0.\end{cases}$$

Define $\widetilde{\psi}$ on U_ε by $\widetilde{\psi}(X(x,t))=\widetilde{\psi}(X(x,0))=\chi(x)$ and observe that (7.21) holds.

Now, since $\chi\equiv 1$ on $\partial U\cap B(x_0,r')$, by reducing ε if necessary, we can ensure that $\widetilde{\psi}\equiv 1$ on $B(x_0,r)\cap U_\varepsilon$. Similarly, since $\chi\equiv 0$ on $\partial U\cap B(x_0,R')$, by reducing ε if necessary, we can ensure that $\widetilde{\psi}\equiv 0$ on $U_\varepsilon\cap B(x_0,R)$.

We extend $\widetilde{\psi}$ by 0 outside U_ε, and define $\psi := \eta\widetilde{\psi} + (1-\eta)\chi$ where η is smooth, supported on $U_{\varepsilon/2}$ (so that ψ is smooth away from ∂U and $\equiv 1$ on $B(x_0, r)$), and identically equal to one in $U_{\varepsilon/4}$ (to preserve the condition on ∂U). □

7.4 The assumptions on V_h (Assumptions 3.10, 3.11, and 3.12)

Lemma 7.9 Assumptions 3.9, 3.10, 3.11, and 3.12 *hold when V_h is a space of Lagrange finite elements that resolves $\partial\Omega$ and Γ_{p}, with p the polynomial degree, $\kappa = 1$, the constants C_{approx}, C_{super}, and C'_{inv} independent of h_K and k, and (for* Assumption 3.10*) the space $\mathcal{Z}_k^j$ given as in* §7.2.

Proof In Assumption 3.9, (3.21) and (3.23) are satisfied by the definition of $\mathcal{Z}_k$ (7.3), and (3.22) is satisfied by the assumption that V_h resolves $\partial\Omega$ and Γ_{p}.

The approximation property of Assumption 3.10 with $\kappa = 1$ and u_h equal to the Lagrange interpolant of u holds by, e.g., [4, Theorem 4.4.20] and the definition (8.3) of the k-weighted norms.

When χ is a smooth function on K and v_h is the Lagrange interpolant of $\chi^2 u_h$,

$$\left\| \chi^2 u_h - v_h \right\|_{L^2(K)} \le C'_{\mathrm{super}} \frac{h_K}{d} \|u_h\|_{L^2(K)} \quad \text{and} \tag{7.28}$$

$$\left\| \nabla(\chi^2 u_h - v_h) \right\|_{L^2(K)} \le C'_{\mathrm{super}} \left(\frac{h_K}{d} \|\nabla(\chi u_h)\|_{L^2(K)} + \frac{h_K}{d^2} \|u_h\|_{L^2(K)} \right) \tag{7.29}$$

by [9, Theorem 2.1], [3, Theorem 1]. The bound (3.26) then follows from (8.3), (7.28), (7.29), and the inequality

$$\sqrt{a^2 + b^2} \le a + b \quad \text{for all } a, b > 0. \tag{7.30}$$

That (3.25) holds with $\kappa = 1$ follows the fact that $\operatorname{supp}\chi$ and $B(0, d)$ are $d/4$ apart, and thus $v_h \in V_h^<(U_1)$ is ensured by $\max_{K \cap U_1 \neq \emptyset} h_K < d/4$.

By, e.g., [4, Lemma 4.5.3], given $p \in \mathbb{Z}^+$, there exists a constant $C'_{\mathrm{inv}} > 0$ such that

$$\|u_h\|_{H^s(K)} \le C'_{\mathrm{inv}} h_K^{-s} \|u_h\|_{L^2(K)} \tag{7.31}$$

for all $K \in \mathcal{T}$, $u_h \in V_h$, and $0 \le s \le p$. Then, by (7.31), (8.3), and (3.20), there exists $C_{\mathrm{inv}} > 0$ such that, for $0 \le s \le p$,

$$\|u_h\|_{H_k^s(K)} \le \frac{C_{\mathrm{inv}}}{(h_K k)^s} \|u_h\|_{L^2(K)}; \tag{7.32}$$

the inequality (3.29) is then this with $s = 1$

The inequality (3.30) follows by repeating the proof of [15, Theorem 3.6] with the inverse inequality (7.31) replaced by (7.32) in [15, Equation 3.23]. Note that in [15], the $H^{-s}(K)$ norm is defined with a supremum ranging over all elements of $H^s(K)$, but, as stated in [15, Remark 3.8], the proof works without modification for the $H^{-s}(K)$ norm defined by (3.27) with $k = 1$ and equivalent to a norm defined with a supremum over elements of $H^s(K)$ supported inside K as in (3.28). □

Appendix 1. Recap of Sobolev spaces weighted by k

Given $k > 0$, let

$$\mathcal{F}_k\phi(\xi) := \int_{\mathbb{R}^d} \exp\big(-\mathrm{i}kx\cdot\xi\big)\phi(x)\,\mathrm{d}x \tag{8.1}$$

and, for $s \in \mathbb{R}$, let

$$H^s_k(\mathbb{R}^d) := \Big\{u \in \mathcal{S}^*(\mathbb{R}^d),\ \langle\xi\rangle^s \mathcal{F}_k u \in L^2(\mathbb{R}^d)\Big\} \quad \text{with} \quad \|u\|^2_{H^s_k(\mathbb{R}^d)} := \left(\frac{k}{2\pi}\right)^d \int_{\mathbb{R}^d} \langle\xi\rangle^{2s} |\mathcal{F}_k u(\xi)|^2\,\mathrm{d}\xi, \tag{8.2}$$

where $\langle\xi\rangle := (1+|\xi|^2)^{1/2}$. For an open $D \subset \mathbb{R}^d$, let

$$\|v\|_{H^s_k(D)} := \inf_{V \in H^s_k(\mathbb{R}^d) : V|_D = v} \|V\|_{H^s_k(\mathbb{R}^d)}. \tag{8.3}$$

Recall that, for $s \geq 0$ and D a bounded Lipschitz domain, by, e.g., [24, Page 77 and Theorem 3.30(i), Page 92],

$$\|v\|_{H^{-s}_k(D)} \sim \sup_{\|w\|_{H^s_k(D)}=1\,,\,\mathrm{supp} w \subset D} |v(w)|, \tag{8.4}$$

where $\langle\cdot,\cdot\rangle_D$ is the duality pairing in D and $\sim$ denotes norm equivalence, and

$$\|v\|_{\widetilde{H}^{-s}_k(D)} \sim \sup_{\|w\|_{H^s_k(D)}=1} |v(w)|. \tag{8.5}$$

We highlight that, for $s = m \in \mathbb{Z}^+$, (8.3) is equivalent to the norm

$$|||v|||^2_{H^m_k(D)} := \sum_{0\leq|\alpha|\leq m} k^{-2|\alpha|} \big\|\partial^\alpha v\big\|^2_{L^2(D)}, \tag{8.6}$$

and thus, in particular,

$$\|v\|^2_{H^1_k(D)} \sim |||v|||^2_{H^1_k(D)} = k^{-2}\|\nabla v\|^2_{L^2(D)} + \|v\|^2_{L^2(D)}, \tag{8.7}$$

where $\sim$ again denotes norm equivalence. Many papers on numerical analysis of the Helmholtz equation use the weighted H^1 norm $\|v\|^2_{H^1_k(D)} := \|\nabla v\|^2_{L^2(D)} + k^2\|v\|^2_{L^2(D)}$; we use (8.6)/(8.3) instead since weighting the jth derivative by k^{-j} is easier to keep track of than weighting it by k^{-j+1} (especially for high derivatives).

Acknowledgements The authors thank the referees for their careful reading of the paper and numerous suggestions for improvement. The idea of looking at the local FEM error for the Helmholtz equation came out of discussions EAS had with Ralf Hiptmair (ETH Zürich). MA and EAS were supported by EPSRC grant EP/R005591/1 and JG was supported by EPSRC grants EP/V001760/1 and EP/V051636/1.

Funding Open access funding provided by Université d'Angers.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Behzadan, A., Holst, M.: Multiplication in Sobolev spaces, revisited. Arkiv för Matematik **59**(2), 275–306 (2021)
2. Bramble, J., Pasciak, J.: Analysis of a finite PML approximation for the three dimensional time-harmonic Maxwell and acoustic scattering problems. Math. Comput. **76**(258), 597–614 (2007)
3. Brenner, S.C.: A general superapproximation result. Comput. Methods Appl. Math. **20**(4), 763–767 (2020)
4. Brenner, S.C., Scott, L.R.: *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, 3rd edition, (2008)
5. Chandler-Wilde, S.N., Monk, P.: Wave-number-explicit bounds in time-harmonic scattering. SIAM J. Math. Anal. **39**(5), 1428–1455 (2008)
6. Ciarlet, P.G.: Basic error estimates for elliptic problems. In: Handbook of Numerical Analysis, Vol. II, pages 17–351. North-Holland, Amsterdam, (1991)
7. Clarke, F.: On the inverse function theorem. Pac. J. Math. **64**(1), 97–102 (1976)
8. Costabel, M., Dauge, M., Nicaise, S.: Corner singularities and analytic regularity for linear elliptic systems. Part I: Smooth domains. (2010) https://hal.archives-ouvertes.fr/file/index/docid/453934/filename/CoDaNi_Analytic_Part_I.pdf
9. Demlow, A., Guzmán, J., Schatz, A.H.: Local energy estimates for the finite element method on sharply varying grids. Math. Comput. **80**(273), 1–9 (2011)
10. Descloux, J.: Interior regularity and local convergence of Galerkin finite element approximations for elliptic equations. In: J.J.H. Miller, editor, *Topics in Numerical Analysis II*, pages 27–41. Academic Press, (1975)
11. Galkowski, J., Lafontaine, D., Spence, E.A.: Local absorbing boundary conditions on fixed domains give order-one errors for high-frequency waves. IMA. J. Num. Anal. (2023) https://doi.org/10.1093/imanum/drad058
12. Galkowski, J., Lafontaine, D., Spence, E.A.: Perfectly-matched-layer truncation is exponentially accurate at high frequency. SIAM J. Math. Anal. **55**(4), 3344–3394 (2023)
13. Galkowski, J., Lafontaine, D., Spence, E.A., Wunsch, J.: The hp-FEM applied to the Helmholtz equation with PML truncation does not suffer from the pollution effect. Comm. Math. Sci., to appear, (2024)
14. Galkowski, J., Spence, E.A.: Sharp preasymptotic error bounds for the Helmholtz h-FEM. arXiv 2301.03574, (2023)
15. Graham, I.G., Hackbusch, W., Sauter, S.A.: Finite elements on degenerate meshes: inverse-type inequalities and applications. IMA J. Numer. Anal. **25**(2), 379–407 (2005)
16. Graham, I.G., Pembery, O.R., Spence, E.A.: The Helmholtz equation in heterogeneous media: a priori bounds, well-posedness, and resonances. J. Differential Equations **266**(6), 2869–2923 (2019)
17. Grisvard, P.: Elliptic problems in nonsmooth domains. Pitman, Boston (1985)
18. Hecht, F.: New development in freefem++. J. Numer. Math. **20**(3–4), 251–265 (2012)
19. Hohage, T., Schmidt, F., Zschiedrich, L.: Solving time-harmonic scattering problems based on the pole condition II: convergence of the PML method. SIAM J. Math. Anal. **35**(3), 547–560 (2003)

20. Ihlenburg, F., Babuska, I.: Finite element solution of the Helmholtz equation with high wave number part II: the *hp* version of the FEM. SIAM J. Numer. Anal. **34**(1), 315–358 (1997)
21. Lassas, M., Somersalo, E.: On the existence and convergence of the solution of PML equations. Computing **60**(3), 229–241 (1998)
22. Lassas, M., Somersalo, E.: Analysis of the PML equations in general convex geometry. Proceedings of the Royal Society of Edinburgh Section A: Mathematics **131**(5), 1183–1207 (2001)
23. Marchand, P., Galkowski, J., Spence, E.A., Spence, A.: Applying GMRES to the Helmholtz equation with strong trapping: how does the number of iterations depend on the frequency? Adv. Comput. Math. **48**(4), 1–63 (2022)
24. McLean, W.:Strongly elliptic systems and boundary integral equations. Cambridge University Press, (2000)
25. Melenk, J.M., Sauter, S.: Convergence analysis for finite element discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary conditions. Math. Comp. **79**(272), 1871–1914 (2010)
26. Nitsche, J.A., Schatz, A.H.: Interior estimates for Ritz-Galerkin methods. Math. Comput. **28**(128), 937–958 (1974)
27. Schatz, A.H., Wahlbin, L.B.: Interior maximum norm estimates for finite element methods. Math. Comput. **31**(138), 414–442 (1977)
28. Schatz, A.H., Wahlbin, L.B.: On the quasi-optimality in L_∞ of the $\overset{\circ}{H}_1$-projection into finite element spaces. Math. Comput. **38**(157), 1–22 (1982)
29. Wahlbin, L.B.: Local behavior in finite element methods. In: Handbook of numerical analysis, Vol. II, pages 353–522. North-Holland, Amsterdam, (1991)
30. Wilson, H.B., Scharstein, R.W.: Computing elliptic membrane high frequencies by Mathieu and Galerkin methods. J. Eng. Math. **57**(1), 41–55 (2007)