A SHARP RELATIVE-ERROR BOUND FOR THE HELMHOLTZ h-FEM AT HIGH FREQUENCY

D. LAFONTAINE*, E. A. SPENCE†, J. WUNSCH‡

Abstract. For the h-finite-element method (h-FEM) applied to the Helmholtz equation, the question of how quickly the meshwidth h must decrease with the frequency k to maintain accuracy as k increases has been studied since the mid 80's. Nevertheless, there still do not exist in the literature any k-explicit bounds on the *relative error* of the FEM solution (the measure of the FEM error most often used in practical applications), apart from in one dimension. The main result of this paper is the sharp result that, for the lowest fixed-order conforming FEM (with polynomial degree, p, equal one), the condition "h^2k^3 sufficiently small" is sufficient for the relative error of the FEM solution in 2 or 3 dimensions to be controllably small (independent of k) for scattering of a plane wave by a nontrapping obstacle and/or a nontrapping inhomogeneous medium. We also prove relative-error bounds on the FEM solution for arbitrary fixed-order methods applied to scattering by a nontrapping obstacle, but these bounds are not sharp for $p \geq 2$. A key ingredient in our proofs is a result describing the oscillatory behaviour of the solution of the plane-wave scattering problem, which we prove using semiclassical defect measures.

Key words. Helmholtz equation, high frequency, pollution effect, finite element method, error estimate, semiclassical analysis.

AMS subject classifications. 35J05, 65N15, 65N30, 78A45

1. Introduction and informal statement of the main results.

1.1. Introduction. When solving the Helmholtz equation $\Delta u + k^2 u = 0$ with the h version of the finite-element method (where accuracy is increased by decreasing the meshwidth h while keeping the polynomial degree p constant), h must decrease faster than k^{-1} to maintain accuracy as k increases; this is the so-called "pollution effect" [2].

A thorough investigation of how quickly h must decrease with the frequency k to maintain accuracy as k increases was performed by Ihlenburg and Babuška in the mid 90's [31, 32] on the 1-d model problem.

$$u'' + k^2 u = -f \quad \text{in } (0,1), \quad u(0) = 0 \quad \text{and} \quad u'(1) - \mathrm{i}ku(1) = 0. \tag{1.1}$$

An explicit expression for the discrete Green's function for this problem is available, and Ihlenburg and Babuška used this to prove the following two sets of results:

1. The h-FEM is quasioptimal in the H^1 semi-norm, with quasioptimality constant independent of k, if (hk^2/p) is sufficiently small; i.e. there exists $c, C > 0$, independent of h, k, and p such that, if $hk^2/p \leq c$, then

$$\|\nabla(u - u_h)\|_{L^2(0,1)} \leq C \min_{v_h \in \mathcal{H}_h} \|\nabla(u - v_h)\|_{L^2(0,1)},$$

where $\mathcal{H}_h$ is the appropriate conforming subspace of $H^1(0,1)$ of piecewise polynomials of degree p on meshes of width h, and u_h is the Galerkin solution; see [31, Theorem 3], [30, Theorem 4.13], [32, Theorem 3.5] (when $p = 1$ this

*Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK, D.Lafontaine@bath.ac.uk

†Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK, E.A.Spence@bath.ac.uk

‡Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston IL 60208-2730, US, jwunsch@math.northwestern.edu

result was proved earlier in [1, Theorem 3.2]). The numerical experiments in [31, Figures 8 and 9] then indicated that, when $p = 1$, the condition "hk^2 sufficiently small" for quasi-optimality is necessary.

2. Under an assumption on the data f (discussed below), the relative error in the h-FEM can be made arbitrarily small by, when $p = 1$, making $hk^{3/2}$ sufficiently small and, when $p \geq 2$ (and assuming that the data is sufficiently smooth, see [30, Remark 4.28]), making $h^{2p}k^{2p+1}$ sufficiently small. More precisely, [31, Equation 3.25], [32, Theorem 3.7], [30, Equation 4.5.15, §4.6.4, and Theorem 4.27] prove that there exists $C > 0$, independent of h and k (but dependent on p) such that, if hk is sufficiently small, then the Galerkin solution u_h exists and

$$\frac{\|u - u_h\|_{H^1_k(0,1)}}{\|u\|_{H^1_k(0,1)}} \leq C \left(\left(\frac{hk}{p} \right)^p + k \left(\frac{hk}{p} \right)^{2p} \right), \tag{1.2}$$

where the weighted H^1 norm $\|\cdot\|_{H^1_k(0,1)}$ is defined by (3.2) below. The numerical experiments in [31, Figure 11], and [30, Figure 4.13] then indicated that, when $p = 1$, the condition "h^2k^3 sufficiently small" is necessary for the relative error to be bounded (in agreement with the earlier numerical experiments in [5] for small k).

The quasi-optimality results in Point 1 above have since been generalised to Helmholtz problems in 2 and 3 dimensions (and improved in the case $p \geq 2$). Indeed, the fact that the h-FEM with $p = 1$ is quasioptimal (with constant independent of k) in the full H^1_k norm when hk^2 is sufficiently small was proved for the homogeneous Helmholtz equation on a bounded domain with impedance boundary conditions in [36, Proposition 8.2.7] (in the case of constant coefficients) and [27, Theorem 4.5 and Remark 4.6(ii)] (in the case of variable coefficients), and for scattering problems with variable coefficients in [23, Theorem 3]. The fact that the h-FEM for $p \geq 2$ is quasioptimal when h^pk^{p+1} is sufficiently small was proved in for a variety of constant coefficient Helmholtz problems in [37, Corollary 5.6], [38, Proof of Theorem 5.8], and [24, Theorem 5.1], and for a variety of problems including variable-coefficient Helmholtz problems in [14, Theorem 2.15]; the condition "h^pk^{p+1} sufficiently small" is indicated to be sharp for quasi-optimality by, e.g., the numerical experiments in [14, §4.4].

In contrast, the relative-error bound (1.2) in Point 2 above has *not* been obtained for any Helmholtz problem in 2 or 3 dimensions, even though numerical experiments indicate that the condition "$h^{2p}k^{2p+1}$ sufficiently small" is necessary and sufficient for the relative error to be controllably small; see, e.g., [19, Left-hand side of Figure 3]. The closest-available result is that, if $h^{2p}k^{2p+1}$ is sufficiently small, then

$$\|u - u_h\|_{H^1_k(D)} \leq C \left((hk)^p + k(hk)^{2p} \right) \|f\|_{L^2(D)}, \tag{1.3}$$

for the Helmholtz problem $\Delta u + k^2 u = -f$ posed in a domain D with either impedance boundary conditions on ∂D or a perfectly matched layer (PML). Indeed, for the PML problem, (1.3) is proved for $p = 1$ in [34, Theorem 4.4 and Remark 4.5(iv)] and [24, Theorem 5.4]. For the impedance problem, (1.3) is proved for $p = 1$ in [46, Theorem 6.1], for $p \geq 1$ in [19, Corollary 5.2] (following earlier work by [48]), and for $p \geq 1$ for the variable-coefficient Helmholtz equation $\nabla \cdot (\mathsf{A}\nabla u) + k^2 n u = -f$ in [40, §2.3] (under a nontrapping condition on A and n). We highlight that, while [19] and [34] prove results of the form (1.3), all the numerical results in [19] and [34] concern the

relative H^1 error, illustrating that relative error is indeed the quantity of interest in practice.

1.2. The main results of this paper. The two main results are the following:

(a) Theorem 4.1 proves the relative-error bound (1.2) when $p = 1$ for scattering of a plane wave by a nontrapping obstacle and/or a nontrapping inhomogeneous medium (modelled by the PDE $\nabla \cdot (\mathsf{A}\nabla u) + k^2 n u = 0$ with variable A and n) in 2 or 3 dimensions (see Definition 2.2 below for the precise definition of the boundary-value problems considered). As highlighted above, the numerical experiments in [5, 31, 30] show that "h^2k^3 sufficiently small" is necessary for the relative error of the h-FEM with $p = 1$ to be controllably small (independent of k), and so the result of Theorem 4.1 is the sharp bound to which the title of the paper refers.

(b) Theorem 4.2 proves for $p \geq 2$ a slightly-weaker bound than (1.2), namely that

$$\frac{\|u - u_h\|_{H^1_k(\Omega_R)}}{\|u\|_{H^1_k(\Omega_R)}} \leq C\big(hk + k(hk)^{p+1}\big), \tag{1.4}$$

for scattering of a plane wave by a nontrapping obstacle in 2 or 3 dimensions. As highlighted above, these are the first-ever frequency-explicit relative-error bounds on the Helmholtz h-FEM in 2 or 3 dimensions.

An additionally novelty of our results is that all the constants in the relative-error bound of Theorem 4.1 are explicit, not only in k and h, but also in the coefficients A and n. The only other coefficient-explicit finite-element analyses of the Helmholtz equation with variable A and n are in [27], [23], and [40]. Indeed, [27, Theorems 4.2 and 4.5] prove quasioptimality for the interior impedance problem under the condition "hk^2 sufficient small" when $p = 1$, [23, Theorem 3] proves the analogous result for scattering by a nontrapping Dirichlet obstacle, and [40, Theorem 2.39] proves the bound (1.3) for the interior impedance problem when $h^{2p}k^{2p+1}$ is sufficiently small. The constants in the bounds of [27, Theorems 4.2 and 4.5], [23, Theorem 3], and [40, Theorem 2.39] are expressed in terms of analogous quantities to those appearing in Theorem 4.1 (with these quantities defined in §3).

The two main results, Theorems 4.1 and 4.2, are proved for a particular class of Helmholtz problems, namely those corresponding to scattering by a plane wave, and not for the equation $\Delta u + k^2 u = -f$ with general $f \in L^2$. We highlight that, for this latter class of problems, it is unreasonable to expect a relative-error bound such as (1.2) to hold, and thus the best one can do is prove bounds for a particular class of realistic data (as we do here). For example, consider the 1-d problem (1.1) with

$$f(x) := -\big[\exp(\mathrm{i}k^n x)\chi(x)\big]'' - k^2\big[\exp(\mathrm{i}k^n x)\chi(x)\big], \tag{1.5}$$

where χ has compact support in $(0,1)$. The solution to (1.1) is then $u(x) = \exp(\mathrm{i}k^n x)\chi(x)$, which oscillates on a scale of k^{-n}, i.e., a smaller scale than k^{-1} when $n > 1$. The finite-element method with, say, $p = 1$ and $hk^{3/2}$ small (and independent of k) will therefore not resolve this solution, and hence a bound such as (1.2) does not hold. This example is nevertheless consistent with the previous results recalled in §1.1 since (i) the assumptions on the solution u in [31, First equation in §3.4] and [32, Definition 3.2] exclude such data f, and (ii) with f given by (1.5), $\|f\|_{L^2(0,1)} \sim k^{2n}$ and $\|u\|_{H^1_k(0,1)} \sim k^n$, so that $\|f\|_{L^2(0,1)} \gg \|u\|_{H^1_k(0,1)}$, and the error estimate (1.3) holds in this case because, although the absolute error on left-hand side of (1.3) is large, the right-hand side of (1.3) is larger.

2. Formulation of the problem.

ASSUMPTION 2.1 (Assumptions on the domain and coefficients).

(i) $\Omega_- \subset \mathbb{R}^d, d = 2, 3$, *is a bounded open Lipschitz set such that its open complement* $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$ *is connected.*

(ii) $\mathsf{A} \in C^{0,1}(\Omega_+, \mathsf{SPD})$ *(where* SPD *is the set of* $d \times d$ *real, symmetric, positive-definite matrices) is such that* $\operatorname{supp}(\mathsf{I}-\mathsf{A})$ *is compact in* $\mathbb{R}^d$ *and there exist* $0 < A_{\min} \leq A_{\max} < \infty$ *such that, in the sense of quadratic forms,*

$$A_{\min} \leq \mathsf{A}(\mathbf{x}) \leq A_{\max} \quad \text{for almost every } \mathbf{x} \in \Omega_+. \tag{2.1}$$

(iii) $n \in L^\infty(\Omega_+, \mathbb{R})$ *is such that* $\operatorname{supp}(1-n)$ *is compact in* $\mathbb{R}^d$ *and there exist* $0 < n_{\min} \leq n_{\max} < \infty$ *such that*

$$n_{\min} \leq n(\mathbf{x}) \leq n_{\max} \quad \text{for almost every } \mathbf{x} \in \Omega_+. \tag{2.2}$$

Let the scatterer Ω_{sc} be defined by $\Omega_{\mathrm{sc}} := \Omega_- \cup \operatorname{supp}(\mathsf{I}-\mathsf{A}) \cup \operatorname{supp}(1-n)$. Given $R > 0$ such that $\Omega_{\mathrm{sc}} \subset B_R$, where B_R denotes the ball of radius R about the origin, let $\Omega_R := \Omega_+ \cap B_R$. Let $\Gamma_R := \partial B_R$ and let $\Gamma := \partial\Omega_-$. Let $\mathbf{n}$ denote the outward-pointing unit normal vector field on both Γ and Γ_R. We denote by $\partial_{\mathbf{n}}$ the corresponding Neumann trace on Γ or Γ_R and $\partial_{\mathbf{n},\mathsf{A}}$ the corresponding conormal-derivative trace. We denote by γu the Dirichlet trace on Γ or Γ_R.

DEFINITION 2.2 (Helmholtz plane-wave scattering problem). *Given* $k > 0$ *and* $\mathbf{a} \in \mathbb{R}^d$ *with* $|\mathbf{a}| = 1$, *let* $u^I(\mathbf{x}) := \mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{a}}$. *Given* Ω_-, A, *and* n, *as in Assumption 2.1, we say* $u \in H^1_{\mathrm{loc}}(\Omega_+)$ *satisfies the* Helmholtz plane-wave scattering problem *if*

$$\nabla \cdot (\mathsf{A}\nabla u) + k^2 n u = 0 \quad \text{in } \Omega_+, \quad \text{either} \quad \gamma u = 0 \quad \text{or} \quad \partial_{\mathbf{n},\mathsf{A}} u = 0 \quad \text{on } \Gamma, \tag{2.3}$$

and $u^S := u - u^I$ *satisfies the Sommerfeld radiation condition*

$$\frac{\partial u^S}{\partial r}(\mathbf{x}) - \mathrm{i}k u^S(\mathbf{x}) = o\left(\frac{1}{r^{(d-1)/2}}\right) \tag{2.4}$$

as $r := |\mathbf{x}| \to \infty$, *uniformly in* $\widehat{\mathbf{x}} := \mathbf{x}/r$.

We call a solution of the Helmholtz equation satisfying the Sommerfeld radiation condition (2.4) an *outgoing* solution (so, in Definition 2.2, u^S is outgoing).

Define $\mathrm{DtN}_k : H^{1/2}(\Gamma_R) \to H^{-1/2}(\Gamma_R)$ to be the Dirichlet-to-Neumann map for the equation $\Delta u + k^2 u = 0$ posed in the exterior of B_R with the Sommerfeld radiation condition (2.4). When $\Gamma_R = \partial B_R$, for some $R > 0$, the definition of DtN_k in terms of Hankel functions and polar coordinates (when $d = 2$)/spherical polar coordinates (when $d = 3$) is given in, e.g., [37, Equations 3.7 and 3.10]. Let

$$H^1_{0,D}(\Omega_R) := \big\{ v \in H^1(\Omega_R) : \gamma v = 0 \text{ on } \Gamma \big\}.$$

When Dirichlet boundary conditions are prescribed in (2.3), let $\mathcal{H} := H^1_{0,D}(\Omega_R)$; when Neumann boundary conditions are prescribed, let $\mathcal{H} := H^1(\Omega_R)$.

LEMMA 2.3 (Variational formulation of the Helmholtz plane-wave scattering problem). *With* u^I, Ω_-, A, n, Ω_R, *and* $\mathcal{H}$ *as above, define* $\widetilde{u} \in \mathcal{H}$ *as the solution of the variational problem*

$$\text{find } \widetilde{u} \in \mathcal{H} \text{ such that} \quad a(\widetilde{u}, v) = F(v) \quad \text{for all } v \in \mathcal{H}, \tag{2.5}$$

where

$$a(\widetilde{u}, v) := \int_{\Omega_R} \Big((\mathsf{A}\nabla\widetilde{u}) \cdot \overline{\nabla v} - k^2 n \widetilde{u}\overline{v} \Big) - \big\langle \mathrm{DtN}_k(\gamma\widetilde{u}), \gamma v \big\rangle_{\Gamma_R}, \quad \text{and} \tag{2.6}$$

$$F(v) := \int_{\Gamma_R} \left(\partial_{\mathbf{n}} u^I - \mathrm{DtN}_k(\gamma u^I)\right) \overline{\gamma v}.$$

where $\langle \cdot, \cdot \rangle_{\Gamma_R}$ denotes the duality pairing on Γ_R that is linear in the first argument and antilinear in the second. Then $\widetilde{u} = u|_{\Omega_R}$, where u is the solution of the Helmholtz plane-wave scattering problem of Definition 2.2.

For a proof of Lemma 2.3, see, e.g., [26, Lemma 3.3]. From here on we denote the solution of the variational problem (2.5) by u, so that u satisfies

$$a(u, v) = F(v) \quad \text{for all } v \in \mathcal{H}. \tag{2.7}$$

LEMMA 2.4. *The solution of the Helmholtz plane-wave scattering problem of Definition 2.2 exists and is unique.*

Proof. Uniqueness follows from the unique continuation principle; see [26, §1], [27, §2] and the references therein. Since $a(\cdot, \cdot)$ satisfies a Gårding inequality (see (10.6) below), Fredholm theory then gives existence. ∎

The h finite-element method. Let $\mathcal{T}_h$ be a family of triangulations of Ω_R (in the sense of, e.g., [16, Page 67]) that is shape regular (see, e.g., [8, Definition 4.4.13], [16, Page 128]). When Neumann boundary conditions are prescribed in (2.3), let

$$\mathcal{H}_h := \{v \in C(\overline{\Omega_R}) : v|_K \text{ is a polynomial of degree } p \text{ for each } K \in \mathcal{T}_h\}; \tag{2.8}$$

when Dirichlet boundary conditions are prescribed we impose the additional condition that elements of $\mathcal{H}_h$ are zero on Γ. In both cases we then have $\mathcal{H}_h \subset \mathcal{H}$, with the dimension of $\mathcal{H}_h$ proportional to h^{-d}. Our main results, Theorems 4.1 and 4.2 below require Γ to be at least $C^{1,1}$. For such Ω_R it is not possible to fit $\partial\Omega_R$ exactly with simplicial elements (i.e. when each element of $\mathcal{T}_h$ is a simplex), and fitting $\partial\Omega_R$ with isoparametric elements (see, e.g, [16, Chapter VI]) or curved elements (see, e.g., [6]) is impractical. Some analysis of non-conforming error is therefore necessary, but since this is very standard (see, e.g., [8, Chapter 10]), we ignore this issue here.

The finite-element method for the variational problem (2.5) is the Galerkin method applied to the variational problem (2.5), i.e.

$$\text{find } u_h \in \mathcal{H}_h \text{ such that } \; a(u_h, v_h) = F(v_h) \;\text{ for all } v_h \in \mathcal{H}_h. \tag{2.9}$$

Observe that setting $v = v_h$ in (2.7) and combining this with (2.9) we obtain the *Galerkin orthogonality* that

$$a(u - u_h, v_h) = 0 \quad \text{for all } v_h \in \mathcal{H}_h. \tag{2.10}$$

3. Definitions of quantities involved in the statement of the main results. Throughout the paper we assume that $R \geq R_0 > 0$ for some fixed $R_0 > 0$ and $k \geq k_0$ for some fixed $k_0 > 0$. For simplicity we assume throughout that

$$k_0 R_0 \geq 1 \quad \text{and} \quad hk \leq 1. \tag{3.1}$$

Given a bounded open set D, we let the weighted H^1 norm, $\|\cdot\|_{H^1_k}$ be defined by

$$\|u\|^2_{H^1_k(D)} := \|\nabla u\|^2_{L^2(D)} + k^2 \|u\|^2_{L^2(D)}. \tag{3.2}$$

We now define quantities $C_{\mathrm{DtN}\,j}, j = 1, 2, C_{\mathrm{sol}}, C_{\mathrm{osc}}, C_{\mathrm{PF}}, C_{H^2}, C_{\mathrm{int}}$, and C_{MS} that appear in the main results (Theorems 4.1 and 4.2). All of these are dimensionless quantities, independent of k, h, and p, but dependent on one or more of A, n, Ω_- (indicated below).

$C_{\mathrm{DtN}j}$, $j = 1, 2$. By [37, Lemma 3.3], there exist $C_{\mathrm{DtN}j} = C_{\mathrm{DtN}j}(k_0 R_0)$, $j = 1, 2$, such that

$$\left|\left\langle \mathrm{DtN}_k(\gamma u), \gamma v\right\rangle_{\Gamma_R}\right| \leq C_{\mathrm{DtN}1} \left\|u\right\|_{H^1_k(\Omega_R)} \left\|v\right\|_{H^1_k(\Omega_R)} \tag{3.3}$$

for all $u, v \in H^1(\Omega_R)$ and for all $k \geq k_0$, and

$$-\Re\left\langle \mathrm{DtN}_k\phi, \phi\right\rangle_{\Gamma_R} \geq C_{\mathrm{DtN}2} R^{-1} \left\|\phi\right\|^2_{L^2(\Gamma_R)} \qquad \text{for all } \phi \in H^{1/2}(\Gamma_R) \text{ and for all } k \geq k_0. \tag{3.4}$$

C_{sol}. We assume that A, n, and Ω_- are *nontrapping* in the sense that there exists $C_{\mathrm{sol}} = C_{\mathrm{sol}}(\mathsf{A}, n, \Omega_-, R, k_0)$ such that, given $f \in L^2(\Omega_R)$, the solution of the boundary value problem (BVP)

$$\nabla \cdot (\mathsf{A}\nabla v) + k^2 n v = -f \ \text{ in } \Omega_+, \qquad \text{either } \gamma v = 0 \ \text{ or } \ \partial_{\mathbf{n},\mathsf{A}} v = 0 \ \text{ on } \Gamma,$$

and v satisfies the Sommerfeld radiation condition (2.4) (with u^S replaced by v), satisfies the bound

$$\left\|v\right\|_{H^1_k(\Omega_R)} \leq C_{\mathrm{sol}} R \left\|f\right\|_{L^2(\Omega_+)} \qquad \text{for all } k \geq k_0; \tag{3.5}$$

observe that the factor R on the right-hand side makes C_{sol} dimensionless. (Remark 4.4 discusses the situation where this nontrapping assumption is removed and C_{sol} depends on k.) This assumption holds if the obstacle Ω_- and the coefficients A and n are nontrapping in the sense that all billiard trajectories (or, more precisely, Melrose–Sjöstrand generalized bicharacteristics [29, Section 24.3]) starting in an exterior neighbourhood of Ω_- and evolving according to the Hamiltonian flow defined by the symbol of (2.3) escape from that neighbourhood after some uniform time. For this flow to be well-defined, Γ must be C^∞, and A and n must be globally $C^{1,1}$ and C^∞ in a neighbourhood of Γ; note that the flow may in general be set-valued rather than unique in cases where the boundary is permitted to be infinite-order flat. Assuming the uniqueness of the flow, an explicit expression for C_{sol} in terms of A, n, Ω_-, and R is then given in [23, Theorems 1 and 2, and Equation 6.32]. However, the bound (3.5) can be established in situations with much less smoothness; indeed, [26, Theorems 2.5, 2.7, and 2.19] establishes (3.5) for a Dirichlet C^0 star-shaped obstacle and L^∞ A and n satisfying certain monotonicity assumptions. Furthermore, our arguments in the rest of the paper do not need the flow to be well-defined on $\Omega_{\mathrm{sc}} := \Omega_- \cup \operatorname{supp}(\mathsf{I} - \mathsf{A}) \cup \operatorname{supp}(1 - n)$, they only require that the bound (3.5) holds. We can therefore define nontrapping in this weaker sense, and work with scatterers of much lower smoothness than in standard microlocal-analysis settings.

C_{osc}. By Theorem 9.1 below, if A, n, and Ω_- are nontrapping then there exists $C_{\mathrm{osc}} = C_{\mathrm{osc}}(\mathsf{A}, n, \Omega_-)$ ('osc' standing for 'oscillation') such that for u a solution to the Helmholtz plane-wave scattering problem of Definition 2.2,

$$\left|u\right|_{H^2(\Omega_R)} \leq C_{\mathrm{osc}} k \left\|u\right\|_{H^1_k(\Omega_R)}, \tag{3.6}$$

where $|\cdot|_{H^2(\Omega_R)}$ denotes the H^2 semi-norm; i.e. $|u|_{H^2(\Omega_R)} := \sum_{|\alpha|=2} \int_{\Omega_R} |\partial^\alpha u|^2$.

C_{PF}. By [8, §5.3], [45, Corollary A.15], there exists $C_{\mathrm{PF}} = C_{\mathrm{PF}}(\Omega_-)$ ('PF' standing for 'Poincaré–Friedrichs') such that

$$R^{-2} \left\|v\right\|^2_{L^2(\Omega_R)} \leq C_{\mathrm{PF}} \left(R^{-1} \left\|\gamma v\right\|^2_{L^2(\Gamma_R)} + \left\|\nabla v\right\|^2_{L^2(\Omega_R)}\right) \tag{3.7}$$

for all $v \in H^1(\Omega_R)$.

C_{H^2}. By Theorem 6.1 below, there exists $C_{H^2} = C_{H^2}(\mathsf{A}, \Omega_-)$ such that, if $f \in L^2(\Omega_R)$ and $v \in H^1(\Omega_R)$ satisfy

$$\nabla \cdot (\mathsf{A}\nabla v) = -f \quad \text{in } \Omega_R, \qquad \partial_{\mathbf{n}} v = \mathrm{DtN}_k(\gamma v) \text{ on } \Gamma_R, \text{ and} \tag{3.8a}$$

$$\text{either } \gamma v = 0 \text{ or } \partial_{\mathbf{n}} v = 0 \quad \text{on } \Gamma, \tag{3.8b}$$

then

$$|v|_{H^2(\Omega_R)} \le C_{H^2} \left(\|f\|_{L^2(\Omega_R)} + R^{-1} \|\nabla v\|_{L^2(\Omega_R)} + R^{-2} \|v\|_{L^2(\Omega_R)} \right). \tag{3.9}$$

The key point in (3.9) is that, although v in (3.8) depends on k via the boundary condition on Γ_R, C_{H^2} is independent of k.

C_{int}. By, e.g., [8, Equation 4.4.28], [43, Theorem 4.1] the *nodal interpolant* $I_h : C(\overline{\Omega_R}) \to \mathcal{H}_h$ is well-defined for functions in $H^2(\Omega_R)$ (for $d = 2, 3$) and satisfies

$$\|v - I_h v\|_{L^2(\Omega_R)} + h \|\nabla(v - I_h v)\|_{L^2(\Omega_R)} \le C_{\mathrm{int}} h^2 |v|_{H^2(\Omega_R)}, \tag{3.10}$$

for all $v \in H^2(\Omega_R)$, for some C_{int} that depends only on the shape-regularity constant of the mesh. As a consequence of (3.10), the definition of $\|\cdot\|_{H^1_k(\Omega_R)}$ (3.2), and the assumption that $hk \le 1$ (3.1), we have

$$\|v - I_h v\|_{H^1_k(\Omega_R)} \le \sqrt{2} C_{\mathrm{int}} h |v|_{H^2(\Omega_R)}. \tag{3.11}$$

C_{MS}. By [38, Lemma 3.4 and Proposition 5.3] there exists $C_{\mathrm{MS}} = C_{\mathrm{MS}}(\Omega_-)$ ('MS' standing for 'Melenk–Sauter') so that, if Γ is analytic, $\mathsf{A} = \mathsf{I}$, $n = 1$, and Ω_+ is nontrapping, then the bound (8.6) below holds.

4. Statement of the main results. The first theorem holds for any $p \ge 1$, but is most relevant in the case $p = 1$.

THEOREM 4.1. *Let u be the solution of the Helmholtz plane-wave scattering problem (Definition 2.2). Assume that both Assumption 2.1 and (3.1) hold, Ω_- is $C^{1,1}$, and A, n, and Ω_- are nontrapping. If $p \ge 1$ and*

$$h^2 k^3 \le C_1, \tag{4.1}$$

then the Galerkin solution u_h to the variational problem (2.9) exists, is unique, and satisfies the bound

$$\|u - u_h\|_{H^1_k(\Omega_R)} \le \left[C_2 hk + C_3 h^2 k^3 \right] \|u\|_{H^1_k(\Omega_R)}, \tag{4.2}$$

where

$$C_1 := \frac{1}{4(A_{\max} + C_{\mathrm{DtN}1})(C_{H^2})^2 (C_{\mathrm{int}})^2 C_{\mathrm{sol}} R} \left(1 + \frac{2\sqrt{2}}{\min\left\{ A_{\min}, C_{\mathrm{DtN}2}, (C_{\mathrm{PF}})^{-1} \right\}} \right)^{-1} \cdot \left(n_{\max} + \frac{1}{k_0 R_0 C_{\mathrm{sol}}} + 2 \right)^{-1},$$

$$C_2 := \frac{\sqrt{2} C_{\mathrm{int}} C_{\mathrm{osc}}}{A_{\min}} \left(\max\left\{ A_{\max}, n_{\max} \right\} + C_{\mathrm{DtN}1} \right)$$

and

$$C_3 := \frac{4\sqrt{2}}{\sqrt{A_{\min}}}\big(A_{\max} + C_{\mathrm{DtN}1}\big)(C_{\mathrm{int}})^2 C_{H^2} C_{\mathrm{sol}} R C_{\mathrm{osc}} \sqrt{n_{\max} + A_{\min}} \cdot \left(n_{\max} + \frac{1}{k_0 R_0 C_{\mathrm{sol}}} + 2\right).$$

THEOREM 4.2. *Let u be the solution of the Helmholtz plane-wave scattering problem (Definition 2.2). Assume that both Assumption 2.1 and (3.1) hold, $\mathsf{A} = \mathsf{I}$, $n = 1$, Ω_- is a nontrapping Dirichlet obstacle, Γ is analytic, and the triangulation $\mathcal{T}_h$ in the definition of $\mathcal{H}_h$ (2.8) satisfies the quasi-uniformity assumption [38, Assumption 5.1]. If*

$$\frac{(hk)^2}{p} + C_{\mathrm{sol}} R \frac{k(hk)^{p+1}}{p^p} \le \widetilde{C}_1 \tag{4.3}$$

then the Galerkin solution u_h to the variational problem (2.9) exists, is unique, and satisfies the bound

$$\|u - u_h\|_{H^1_k(\Omega_R)} \le \left[\left(\widetilde{C}_2 + \frac{\widetilde{C}_3 C_{\mathrm{MS}}}{p}\right) hk + \widetilde{C}_3 C_{\mathrm{MS}} C_{\mathrm{sol}} R \, \frac{k(hk)^{p+1}}{p^p}\right] \|u\|_{H^1_k(\Omega_R)}, \tag{4.4}$$

where

$$\widetilde{C}_1 := \frac{1}{2\sqrt{2}(1 + C_{\mathrm{DtN}1}) C_{H^2} C_{\mathrm{MS}}} \left(1 + \frac{2\sqrt{2}}{\min\left\{A_{\min}, C_{\mathrm{DtN}2}, (C_{\mathrm{PF}})^{-1}\right\}}\right)^{-1},$$

$$\widetilde{C}_2 := \sqrt{2} C_{\mathrm{cont}} C_{\mathrm{int}} C_{\mathrm{osc}} \quad \text{and} \quad \widetilde{C}_3 := 4\big(1 + C_{\mathrm{DtN}1}\big) C_{\mathrm{int}} C_{\mathrm{osc}}.$$

Observe that (i) the condition (4.3) is satisfied if $h^{p+1}k^{p+2}$ is sufficiently small, and (ii) the bound (4.4) is of the form (1.4).

The result of Theorem 4.2 might appear not to be a high-order result, since the lowest-order terms in (4.3) and (4.4) are h^2 and h, respectively. Nevertheless, if $k(hk)^{p+1}$ is sufficiently small, so that (4.3) is satisfied, then

$$h \sim k^{-1-1/(p+1)} \quad \text{so} \quad hk \sim k^{-1/(p+1)} \ll 1 \quad \text{as } k \to \infty,$$

and the dominant term on the right-hand side of (4.4) is that involving $k(hk)^{p+1}$.

4.1. The ideas behind the proofs. Theorems 4.1 and 4.2 are proved by adapting the so-called elliptic-projection argument, used to prove the bound (1.3) on the solution in terms of the data, to instead prove relative-error bounds. The elliptic-projection argument was introduced in the Helmholtz context in [20, 21] for interior-penalty discontinuous Galerkin methods, used for the standard FEM and continuous interior-penalty methods in [46, 48], subsequently used by [4, 47, 13, 24, 34], and then augmented with an error splitting argument in [19] (see, e.g., the literature review in [40, §2.3]). The elliptic-projection argument itself is a modification of the classic duality argument, coming out of ideas introduced in [42], which was used to prove quasi-optimality of the Helmholtz FEM in [1, 31, 36, 41, 37, 38, 13, 14, 24, 27, 23].

Our modifications of the elliptic-projection argument are outlined in §5. Our three new ingredients are (i) keeping track of how all the constants in this argument depend

on A, n, Ω_-, and R, (ii) a rigorous proof, using microlocal/semiclassical analysis, of the bound (3.6) describing the oscillatory behaviour of the solution of the plane-wave scattering problem (see Theorem 9.1 below), and (iii) the proof of H^2 regularity, with constant independent of k, of the solution of Poisson's equation with the boundary condition $\partial_{\mathbf{n}} v = \mathrm{DtN}_k(\gamma v)$; see (3.9) and Theorem 6.1.

Regarding (i): while the standard duality argument applied to the Helmholtz equation discussed above has recently been made explicit in A, n, and Ω_- in [27, 23] (as mentioned in §1.2), the only places in the literature where the elliptic-projection argument is made explicit in A, n, and Ω_-, are the present paper and [40, §2.3].

Regarding (ii): oscillatory behaviour similar to (3.6) of Helmholtz solutions has been an assumption in many analyses of finite- and boundary-element methods; see, e.g., [31, First equation in §3.4], [32, Definition 3.2], [9, Definition 4.6], [3, Definition 3.5], [18, Assumption 3.4]. However, to our knowledge, the only existing rigorous proofs of such behaviour are [25, Theorems 1.1 and 1.2] and [22, Theorem 1.11(c)], with both results concerning the Neumann trace of the solution of the Helmholtz plane-wave scattering problem with $\mathsf{A} = \mathsf{I}$ and $n = 1$ (and therefore applicable to boundary-element methods applied to this problem).

Regarding (iii): the analogous result (H^2 regularity with constant independent of k) for Poisson's equation with the *impedance boundary condition* $\partial_{\mathbf{n}} v = \mathrm{i}k\gamma v$ is central to the elliptic-projection argument for the Helmholtz equation with impedance boundary conditions. This result was explicitly assumed in [21, Lemma 4.3], implicitly assumed in [46, 48, 4, 13], and recently proved in [15].

4.2. Why does Theorem 4.2 not cover scattering by an inhomogeneous medium?. In both the elliptic-projection argument and the standard duality argument, a key role is played by the quantity $\eta(\mathcal{H}_h)$ defined by (8.3) below, which describes how well solutions of the (adjoint of the) Helmholtz equation can be approximated in $\mathcal{H}_h$.

In the case $p = 1$ we estimate $\eta(\mathcal{H}_h)$ using H^2 regularity of the solution (which holds when A and Ω_- satisfy the assumptions of Theorem 4.1), leading to the bound (8.5) below. When $p \geq 1$, $\mathsf{A} = \mathsf{I}$, $n = 1$, Ω_- is a Dirichlet obstacle, and Γ is analytic, [38] proved the bound (8.6) on $\eta(\mathcal{H}_h)$, and we use this result to prove Theorem 4.2. The bound (8.6) was proved via a judicious splitting of the solution [38, Theorem 4.20] into an analytic but oscillating part, and an H^2 part that behaves "well" for large frequencies, and this splitting is only available for the exterior Dirichlet problem with $\mathsf{A} = \mathsf{I}$ and $n = 1$.

We highlight that an alternative splitting procedure valid for Helmholtz problems with variable coefficients was recently developed in [14], leading to an alternative proof of the bound on $\eta(\mathcal{H}_h)$ (8.6) [14, Lemma 2.13]. However, this alternative procedure requires that DtN_k be approximated by $\mathrm{i}k$ on Γ_R. Indeed, in [14, Proof of Lemma 2.13] the solution is expanded in powers of k, i.e. $u = \sum_{j=0}^{\infty} k^j u_j$, and then on Γ_R one has $\partial_{\mathbf{n}} u_{j+1} = \mathrm{i}\gamma u_j$; this relationship between u_{j+1} and u_j on Γ_R no longer holds if DtN_k is not approximated by $\mathrm{i}k$.

4.3. Approximating DtN_k. Implementing the operator DtN_k is computationally expensive, and so in practice one seeks to approximate this operator by *either* imposing an absorbing boundary condition on Γ_R, *or* using a PML. In this paper we follow the precedent established in [37, 38] of, when proving new results about the FEM for exterior Helmholtz problems, first assuming that DtN_k is realised exactly. We remark, however, that if the two key ingredients in Remark 4.1 (a proof of the

oscillatory behaviour (3.6) and H^2-regularity, independent of k, of a Poisson problem) can be established when DtN_k is replaced by an absorbing boundary condition on Γ_R, then the result of Theorem 4.1 carry over to this case. When an impedance boundary condition (i.e. the simplest absorbing boundary condition) is imposed on Γ_R, the necessary Poisson H^2-regularity result is proved in [15], but we discuss below in Remark 9.9 the difficulties in proving (3.6) in this case.

4.4. Removing the nontrapping assumption. The only place in the proofs of Theorems 4.1 and 4.2 where the nontrapping assumption (i.e. the fact that C_{sol} in (3.5) is independent of k) is used is in the proof of the bound (3.6) (in Theorem 9.1 below). We sketch in Remark 9.10 below how (3.6) can be proved in the trapping case (i.e. when C_{sol} is not independent of k); the rest of the proofs of Theorems 4.1 and 4.2 then go through as before. In the case of Theorem 4.1, the requirement for the relative error to be bounded independently of k would then be that $h^2k^3C_{\mathrm{sol}}$ be sufficiently small. Under the strongest form of trapping, C_{sol} can grow exponentially through a sequence of ks [7, §2.5], but is bounded polynomially in k if a set of frequencies of arbitrarily-small measure is excluded [33, Theorem 1.1]. However, it is not clear how sharp the requirement "$h^2k^3C_{\mathrm{sol}}$ sufficiently small" for the relative error to be bounded is in these cases.

5. Outline of the proof and connection to existing arguments. As in the standard duality argument coming out of ideas introduced in [42] and then formalised in [41], our starting point is the fact that, since $a(\cdot,\cdot)$ satisfies the Gårding inequality (10.6), Galerkin orthogonality (2.10) and continuity of $a(\cdot,\cdot)$ (10.4) imply that, for any $v_h \in \mathcal{H}_h$,

$$\begin{aligned} A_{\min}\,\|u-u_h\|^2_{H^1_k(\Omega_R)} &\le \Re a(u-u_h,u-v_h) + k^2\big(n_{\max}+A_{\min}\big)\,\|u-u_h\|^2_{L^2(\Omega_R)}\,,\\ &\le C_{\mathrm{cont}}\,\|u-u_h\|_{H^1_k(\Omega_R)}\,\|u-v_h\|_{H^1_k(\Omega_R)} + k^2\big(n_{\max}+A_{\min}\big)\,\|u-u_h\|^2_{L^2(\Omega_R)}\,. \end{aligned} \tag{5.1}$$

Recall (from, e.g., [41, Theorem 2.5], [37, Theorem 4.3], [44, Theorem 6.32]) that the standard duality argument shows that

$$\|u-u_h\|_{L^2(\Omega_R)} \le C_{\mathrm{cont}}\eta(\mathcal{H}_h)\,\|u-u_h\|_{H^1_k(\Omega_R)}\,, \tag{5.2}$$

where $\eta(\mathcal{H}_h)$, defined by (8.3) below, describes how well solutions of the adjoint problem are approximated in the space $\mathcal{H}_h$. Inputting (5.2) into (5.1) one obtains quasioptimality, with constant independent of k, if $k\eta(\mathcal{H}_h)$ is sufficiently small; the bounds on $\eta(\mathcal{H}_h)$ described in Lemma 8.2 below then imply that this condition is satisfied if h^pk^{p+1} is sufficiently small.

In contrast, the elliptic-projection argument, which we follow, shows that

$$\|u-u_h\|_{L^2(\Omega_R)} \lesssim \eta(\mathcal{H}_h)\,\|u-w_h\|_{H^1_k(\Omega_R)} \qquad \text{for all } w_h \in \mathcal{H}_h, \tag{5.3}$$

provided that $hk^2\eta(\mathcal{H}_h)$ is sufficiently small (see Lemma 10.1 below), where in this overview discussion we use the notation $a \lesssim b$ when $a \le Cb$ with C independent of k, h, and p, but dependent on A, n, Ω_-, and R. Observe that (5.3) is a stronger bound than (5.2), since w_h on the right-hand side of (5.3) is arbitrary. The proof of (5.3) in our setting of the plane-wave scattering problem requires the new Poisson H^2-regularity bound (3.9), which we prove in Theorem 6.1 below.

Inputting (5.3) into (5.1), choosing $w_h = v_h$, and using the inequality

$$2\alpha\beta \le \varepsilon\alpha^2 + \varepsilon^{-1}\beta^2, \qquad \text{for all } \alpha,\beta,\varepsilon > 0, \tag{5.4}$$

on the first term on the right-hand side of (5.1), we obtain that, if $hk^2\eta(\mathcal{H}_h)$ is sufficiently small, then, for any $v_h \in \mathcal{H}_h$,

$$\|u-u_h\|^2_{H^1_k(\Omega_R)} \lesssim \left(1+k^2(\eta(\mathcal{H}_h))^2\right)\|u-v_h\|^2_{H^1_k(\Omega_R)}\,.$$

Assuming H^2 regularity of the solution, and using (3.11), we obtain that, if $hk^2\eta(\mathcal{H}_h)$ is sufficiently small, then

$$\|u-u_h\|^2_{H^1_k(\Omega_R)} \lesssim \left(1+k^2(\eta(\mathcal{H}_h))^2\right)h^2|u|^2_{H^2(\Omega_R)}. \tag{5.5}$$

In the standard elliptic-projection argument (see, e.g., [13, §5.5]) applied to the PDE $\Delta u + k^2 u = -f$, an H^2-regularity bound similar to (3.5) and the nontrapping bound (3.5) are combined to give $|u|_{H^2(\Omega_R)} \lesssim k\|f\|_{L^2(\Omega_R)}$, and combining this with both (5.5) and the bound $\eta(\mathcal{H}_h) \lesssim hk$ (see (8.5) below) proves the bound (1.3) with $p=1$ on the Galerkin error in terms of the data when h^2k^3 is sufficiently small.

In contrast, in this paper we prove, using microlocal/semiclassical analysis, that the solution the plane-wave scattering problem satisfies $|u|_{H^2(\Omega_R)} \lesssim k\,\|u\|_{H^1_k(\Omega_R)}$ (see Theorem 9.1 below), and using this in (5.5), along with the bounds on $\eta(\mathcal{H}_h)$ in Lemma 8.2, we obtain the relative-error bounds (4.2) and (4.4).

6. Proof of the Poisson H^2-regularity result (3.9).

THEOREM 6.1. *With* A, Ω_-, Γ, *and* Ω_R *as in Section 2, let* $v \in H^1(\Omega_R)$ *be the solution of the Poisson boundary value problem* (3.8). *If* Γ *is* $C^{1,1}$, *then* $v \in H^2(\Omega_R)$ *and the bound* (3.9) *holds.*

We follow the recent proof of the related regularity result (with DtN_k replaced by $\mathrm{i}k$ and $\mathsf{A}=\mathsf{I}$) [15, Theorem 3.1] and start by recalling results due to Grisvard [28].

LEMMA 6.2. *Let* D *be a bounded, convex, open set of* $\mathbb{R}^n$ *with* C^2 *boundary. Then, for all* $\mathbf{v} \in H^1(D;\mathbb{C}^d)$,

$$\int_D \left(|\nabla\cdot\mathbf{v}|^2 - \sum_{i,j=1}^n \int_D \frac{\partial v_i}{\partial x_j}\overline{\frac{\partial v_j}{\partial x_i}} \right) \geq -2\Re\big\langle(\gamma\mathbf{v})_T, \nabla_T(\gamma\mathbf{v}\cdot\mathbf{n})\big\rangle_{\partial D}, \tag{6.1}$$

where ∇_T *is the surface gradient on* ∂D *and* $(\gamma\mathbf{v})_T := \gamma\mathbf{v} - \mathbf{n}(\gamma\mathbf{v}\cdot\mathbf{n})$ *is the tangential component of* $\gamma\mathbf{v}$.

Proof. The result with $\mathbf{v}$ real follows from [28, Theorem 3.1.1.1] and the fact that the second fundamental form of ∂D (defined in, e.g., [28, §3.1.1]), is non-positive (see [28, Proof of Theorem 3.1.2.3]). The result with $\mathbf{v}$ complex follows in a straightforward way by repeating the argument in [28, Theorem 3.1.1.1] for complex $\mathbf{v}$. ∎

LEMMA 6.3. ([28, Lemma 3.1.3.4].) *If* $\mathsf{A} \in C^{0,1}(D,\mathsf{SPD})$ *satisfies* (2.1) *(with* Ω_+ *replaced by* D*), then, for all* $v \in H^2(D)$,

$$(A_{\min})^2 \sum_{i,j=1}^d \left|\frac{\partial^2 v}{\partial x_i \partial x_j}\right|^2 \leq \sum_{i,j,\ell,m=1}^d A_{i\ell}A_{jm}\frac{\partial^2 v}{\partial x_j\partial x_\ell}\frac{\partial^2 \overline{v}}{\partial x_i \partial x_m}. \tag{6.2}$$

As a first step to proving Theorem 6.1, we prove it in the case when $\Omega_- = \emptyset$.

LEMMA 6.4. *Let* $\mathsf{A} \in C^{0,1}(B_R,\mathsf{SPD})$ *satisfy* (2.1) *(with* Ω_+ *replaced by* B_R*) and be such that* $\mathrm{supp}(\mathsf{I}-\mathsf{A}) \subset\subset B_R$. *Given* $f \in L^2(B_R)$, *let* $v \in H^1(B_R)$ *be the solution of*

$$\nabla\cdot(\mathsf{A}\nabla v) = -f \quad \text{in } B_R, \qquad \partial_{\mathbf{n}} v = \mathrm{DtN}_k(\gamma v) \text{ on } \Gamma_R. \tag{6.3}$$

Then $v \in H^2(B_R)$ and

$$|v|_{H^2}^2 \leq \frac{2}{(A_{\min})^2}\left[\|f\|_{L^2}^2 + \left(d^4\|DA\|_{L^\infty}^2 + \frac{2}{(A_{\min})^2}d^8\|A\|_{L^\infty}^2\|DA\|_{L^\infty}^2\right)\|\nabla v\|_{L^2}^2\right].$$

Proof. Let $w \in H^1(\mathbb{R}^d)$ be the outgoing solution of the following transmission problem

$$\begin{aligned} \nabla\cdot(\mathsf{A}\nabla w) = -f \quad \text{in } B_R, \qquad \Delta w + k^2 w = 0 \quad \text{in } \mathbb{R}^d \setminus \overline{B_R}, \\ \gamma w_+ = \gamma w_- \ \text{ and } \ \partial_{\mathbf{n}} w_+ = \partial_{\mathbf{n}} w_- \quad \text{on } \Gamma_R, \end{aligned}$$

where $w_- := w|_{B_R}$ and $w_+ := w|_{\mathbb{R}^d\setminus B_R}$. (Note that it is important here that $\mathsf{A} = \mathsf{I}$ in a neighbourhood of Γ_R, so that $\partial_{\mathbf{n},\mathsf{A}} w_- = \partial_{\mathbf{n}} w_-$.) By the definition of the operator DtN_k, $w_- = v$. Since Γ_R is C^2, the regularity result [17, Theorem 5.2.1 and §5.4b] implies that $w_- \in H^2(B_R)$ and $w_+ \in H^2_{\mathrm{loc}}(\mathbb{R}^d \setminus \overline{B_R})$; therefore $v \in H^2(B_R)$.

Since $v \in H^2(B_R)$ and A is Lipschitz, $\mathsf{A}\nabla v \in H^1(B_R)$ and we can apply Lemma 6.2 with $\mathbf{v} := \mathsf{A}\nabla v$. Since $\mathsf{A} = \mathsf{I}$ near Γ_R, $\mathbf{v} = \nabla v$ near Γ_R and so the right-hand side of (6.1) becomes

$$-2\Re\big\langle \nabla_T(\gamma v), \nabla_T(\partial_{\mathbf{n}} v)\big\rangle_\Gamma = -2\Re\big\langle \nabla_T(\gamma v), \nabla_T(\mathrm{DtN}_k(\gamma v))\big\rangle_\Gamma,$$

where we have used the boundary condition in (6.3).

Now, DtN_k and ∇_T commute on Γ_R; this can be seen either by rotation invariance, or by using the definition of DtN_k and ∇_T in terms of Fourier series on Γ_R. Therefore, the inequality (3.4) implies that the right-hand side of (6.1) is non-negative, hence

$$\sum_{i,j,\ell,m=1}^{d} \int_{B_R} \frac{\partial}{\partial x_j}\left(A_{i\ell}\frac{\partial v}{\partial x_\ell}\right)\frac{\partial}{\partial x_i}\left(A_{jm}\frac{\partial \overline{v}}{\partial x_m}\right) \leq \|f\|^2_{L^2(B_R)}. \tag{6.4}$$

The left-hand side of (6.4) equals

$$\sum_{i,j,\ell,m=1}^{d} \int_\Omega A_{i\ell}A_{jm}\frac{\partial^2 v}{\partial x_j \partial x_\ell}\frac{\partial^2 \overline{v}}{\partial x_i \partial x_m} + \sum_{i,j,\ell,m=1}^{d}\int_\Omega R_{i,j,\ell,m}, \tag{6.5}$$

where

$$\begin{aligned} R_{i,j,\ell,m} &= \frac{\partial A_{i\ell}}{\partial x_j}\frac{\partial v}{\partial x_\ell}A_{jm}\frac{\partial^2 \overline{v}}{\partial x_i \partial x_m} + A_{i\ell}\frac{\partial^2 v}{\partial x_j \partial x_\ell}\frac{\partial A_{jm}}{\partial x_i}\frac{\partial \overline{v}}{\partial x_m} + \frac{\partial A_{i\ell}}{\partial x_j}\frac{\partial v}{\partial x_\ell}\frac{\partial A_{jm}}{\partial x_i}\frac{\partial \overline{v}}{\partial x_m} \\ &=: R^1_{i,j,\ell,m} + R^2_{i,j,\ell,m} + R^3_{i,j,\ell,m}. \end{aligned}$$

By the Cauchy-Schwarz inequality

$$\left|\int_{B_R} R^1_{i,j,\ell,m}\right| + \left|\int_{B_R} R^2_{i,j,\ell,m}\right| \leq 2\|A\|_{L^\infty}\|DA\|_{L^\infty}\|\nabla v\|_{L^2}|v|_{H^2}$$

and

$$\left|\int_{B_R} R^3_{i,j,\ell,m}\right| \leq \|DA\|^2_{L^\infty}\|\nabla v\|^2_{L^2}.$$

We therefore obtain

$$\left| \sum_{i,j,\ell,m=1}^{d} \int_{B_R} R_{i,j,\ell,m} \right| \leq 2d^4 \|A\|_{L^\infty} \|DA\|_{L^\infty} \|\nabla v\|_{L^2} |v|_{H^2} + d^4 \|DA\|_{L^\infty}^2 \|\nabla v\|_{L^2}^2.$$

Combining this with (6.2), (6.4), and (6.5), we obtain

$$(A_{\min})^2 |v|_{H^2}^2 \leq \|f\|_{L^2}^2 + 2d^4 \|A\|_{L^\infty} \|DA\|_{L^\infty} \|\nabla v\|_{L^2} |v|_{H^2} + d^4 \|DA\|_{L^\infty}^2 \|\nabla v\|_{L^2}^2.$$

Using (5.4) on the second term on the right-hand side, we obtain the result. ∎

We now use Lemma 6.4 to prove Theorem 6.1.

Proof. [Proof of Theorem 6.1] Let $0 < R_0 < R_1 < R$ be such that $\overline{\Omega_-} \subset B_{R_0}$, and let $\chi \in C^\infty(\mathbb{R}^d)$ be such that $0 \leq \chi \leq 1$ and

$$\chi = 0 \ \text{ in } B_{R_0} \quad \text{ and } \quad \chi = 1 \ \text{ in } \mathbb{R}^d \setminus \overline{B_{R_1}}.$$

We decompose v as

$$v = \chi v + (1-\chi)v =: v_1 + v_2. \tag{6.6}$$

Then $v_1 \in H^1(B_R)$ and satisfies

$$\nabla \cdot (\mathsf{A} \nabla v_1) = -\chi f + \nabla \chi \cdot (\mathsf{A} \nabla v) + \nabla v \cdot (\mathsf{A} \nabla \chi) + v \nabla \cdot (\mathsf{A} \nabla \chi) \quad \text{in } B_R,$$

and $\partial_{\mathbf{n}} v_1 = \mathrm{DtN}_k(\gamma v_1)$ on Γ_R. Lemma 6.4 implies that $v_1 \in H^2(B_R)$ and that there exists $C_4 = C_4(\mathsf{A}, d, \chi) > 0$ such that

$$|v_1|_{H^2(\Omega_R)} \leq C_4 \left(\|f\|_{L^2(\Omega_R)} + R^{-1} \|\nabla v\|_{L^2(\Omega_R)} + R^{-2} \|v\|_{L^2(\Omega_R)} \right), \tag{6.7}$$

where (i) we have used the fact that $\nabla \chi = 0$ in a neighbourhood of Ω_- to write all the norms as norms over Ω_R, and (ii) we have inserted the inverse powers of R on the right-hand side to keep C_4 a dimensionless quantity. On the other hand, v_2 satisfies

$$\nabla \cdot (\mathsf{A} \nabla v_2) = -(1-\chi) f - \nabla \chi \cdot (\mathsf{A} \nabla v) - \nabla v \cdot (\mathsf{A} \nabla \chi) - v \nabla \cdot (\mathsf{A} \nabla \chi) \quad \text{in } B_R,$$

$v_2 = 0$ in $B_R \setminus B_{R_1}$, and *either* $\gamma v_2 = 0$ *or* $\partial_{\mathbf{n}} v_2 = 0$ on Γ.

Since A is Lipschitz, $A_{\min} > 0$, and both Γ and Γ_R are $C^{1,1}$, [28, Theorems 2.3.3.2, 2.4.2.5, and 2.4.2.7] imply that, if $w \in H^1(\Omega_R)$, $\nabla \cdot (\mathsf{A} \nabla w) \in L^2(\Omega_R)$, and *either* $\gamma w = 0$ *or* $\partial_{\mathbf{n}} w = 0$ on $\partial \Omega_R$, then $w \in H^2(\Omega_-)$ and there exists $C_5 = C_5(\mathsf{A}, \Omega_-, d, R) > 0$ such that

$$|w|_{H^2(\Omega_R)} \leq C_5 \left(\|\nabla \cdot (\mathsf{A} \nabla w) - w\|_{L^2(\Omega_R)} + R^{-1} \|\nabla w\|_{L^2(\Omega_R)} + R^{-2} \|w\|_{L^2(\Omega_R)} \right).$$

Applying this with $w = v_2$, we obtain that

$$|v_2|_{H^2(\Omega_R)} \leq C_6 \left(\|f\|_{L^2(\Omega_R)} + R^{-1} \|\nabla v\|_{L^2(\Omega_R)} + R^{-2} \|v\|_{L^2(\Omega_R)} \right), \tag{6.8}$$

and the bound (3.9) follows from combining (6.7) and (6.8) using (6.6). ∎

7. The elliptic projection and associated results. Define the sesquilinear form $a_\star(\cdot,\cdot)$ by

$$a_\star(u,v) := \int_{\Omega_R} \mathsf{A}\nabla u \cdot \overline{\nabla v} - \left\langle \mathrm{DtN}_k \gamma u, \gamma v \right\rangle_{\Gamma_R}. \tag{7.1}$$

Recall from Lemma 2.3 the notation that $\mathcal{H}$ equals *either* $H^1_{0,D}(\Omega_R)$ (with Dirichlet conditions in (2.3)) *or* $H^1(\Omega_R)$ (with Neumann conditions).

LEMMA 7.1 (Continuity and coercivity of $a_\star(\cdot,\cdot)$). *For all $u, v \in \mathcal{H}$,*

$$\left|a_\star(u,v)\right| \le C_{\mathrm{cont}\star} \left\|u\right\|_{H^1_k(\Omega_R)} \left\|v\right\|_{H^1_k(\Omega_R)} \quad \textit{and} \quad \Re a_\star(v,v) \ge C_{\mathrm{coer}\star} \left\|v\right\|^2_{H^1_R(\Omega_R)}, \tag{7.2}$$

where

$$C_{\mathrm{cont}\star} := A_{\max} + C_{\mathrm{DtN}1}, \qquad C_{\mathrm{coer}\star} := \frac{1}{2}\min\left\{A_{\min}, C_{\mathrm{DtN}2}, (C_{\mathrm{PF}})^{-1}\right\},$$

and

$$\left\|v\right\|^2_{H^1_R(\Omega_R)} := \left\|\nabla u\right\|^2_{L^2(\Omega_R)} + \frac{1}{R^2}\left\|v\right\|^2_{L^2(\Omega_R)}. \tag{7.3}$$

Proof. The first inequality in (7.2) follows from the inequality (3.3) and the Cauchy–Schwarz inequality. The second inequality in (7.2) follows from (3.4) and (3.7). ∎

As a corollary of Lemma 7.1 we have

$$C_{\mathrm{coer}\star}\left\|v\right\|^2_{H^1_R(\Omega_R)} \le \left|a_\star(v,v)\right| \le C_{\mathrm{cont}\star}\left\|v\right\|^2_{H^1_k(\Omega_R)} \qquad \text{for all } v \in \mathcal{H}, \tag{7.4}$$

and we then define the new norm on $\mathcal{H}$,

$$\left\|v\right\|_\star := \sqrt{a_\star(v,v)}.$$

LEMMA 7.2 (Bounds on the solution of the variational problem associated with $a_\star(\cdot,\cdot)$). *The solution of the variational problem*

$$\textit{find } u \in \mathcal{H} \textit{ such that } a_\star(u,v) = (f,v)_{L^2(\Omega_R)} \quad \textit{for all } v \in \mathcal{H}$$

satisfies

$$\left\|u\right\|_{H^1_R(\Omega_R)} \le \frac{R}{C_{\mathrm{coer}\star}}\left\|f\right\|_{L^2(\Omega_R)} \qquad \textit{and} \qquad \left|u\right|_{H^2(\Omega_R)} \le C_{H^2\star}\left\|f\right\|_{L^2(\Omega_R)}, \tag{7.5}$$

where

$$C_{H^2\star} := C_{H^2}\left(1 + \sqrt{2}(C_{\mathrm{coer}\star})^{-1}\right).$$

Proof. Since $a_\star(\cdot,\cdot)$ is continuous and coercive in $\mathcal{H}$, the first bound in (7.5) follows from the Lax–Milgram theorem and the fact that

$$\sup_{v\in\mathcal{H}} \frac{\left|(f,v)_{L^2(\Omega_R)}\right|}{\left\|v\right\|_{H^1_R(\Omega_R)}} \le R\left\|f\right\|_{L^2(\Omega_R)},$$

by the definition of $\|\cdot\|_{H^1_k(\Omega_R)}$ (7.3). The second bound in (7.5) follows from combining the first bound in (7.5) and the bound (3.9). ∎

DEFINITION 7.3 (Elliptic projection $\mathcal{P}_h$). *Given $u \in \mathcal{H}$, define $\mathcal{P}_h u \in \mathcal{H}_h$ by*

$$a_\star(v_h, \mathcal{P}_h u) = a_\star(v_h, u) \quad \text{for all } v_h \in \mathcal{H}_h.$$

Since $a_\star(\cdot,\cdot)$ is continuous and coercive in $H^1(\Omega_R)$ by Lemma 7.1, the Lax–Milgram theorem implies that $\mathcal{P}_h$ is well defined. The definition of $\mathcal{P}_h$ then immediately implies the Galerkin-orthogonality property that

$$a_\star(v_h, u - \mathcal{P}_h u) = 0 \quad \text{for all } v_h \in \mathcal{H}_h. \tag{7.6}$$

LEMMA 7.4 (Approximation properties of $\mathcal{P}_h$). *The elliptic projection $\mathcal{P}_h$ satisfies*

$$\|u - \mathcal{P}_h u\|_\star \le \sqrt{C_{\mathrm{cont}\star}} \min_{v_h \in \mathcal{H}_h} \|u - v_h\|_{H^1_k(\Omega_R)} \qquad \text{and} \tag{7.7}$$

$$\|u - \mathcal{P}_h u\|_{L^2(\Omega_R)} \le h\sqrt{2} C_{\mathrm{int}} C_{H^2\star} \sqrt{C_{\mathrm{cont}\star}} \|u - \mathcal{P}_h u\|_\star \tag{7.8}$$

for all $u \in \mathcal{H}$.

Proof. By the Cauchy–Schwarz inequality $a_\star(\cdot,\cdot)$ is continuous in the $\|\cdot\|_\star$ norm, and by definition, $a_\star(\cdot,\cdot)$ is coercive in this norm. Therefore Céa's lemma implies that

$$\|u - \mathcal{P}_h u\|_\star \le \min_{v_h \in \mathcal{H}_h} \|u - v_h\|_\star ,$$

and (7.7) follows from the norm equivalence (7.4).

To prove (7.8) we use a standard duality argument. Given $u \in \mathcal{H}$, let ξ be the solution of the variational problem

$$\text{find } \xi \in \mathcal{H} \text{ such that } a_\star(\xi, v) = (u - \mathcal{P}_h u, v)_{L^2(\Omega_R)} \quad \text{for all } v \in \mathcal{H}. \tag{7.9}$$

Then, by Galerkin orthogonality (7.6) and continuity of $a_\star(\cdot,\cdot)$, for all $v_h \in \mathcal{H}_h$,

$$\|u - \mathcal{P}_h u\|^2_{L^2(\Omega_R)} = a_\star(\xi, u - \mathcal{P}_h u) = a_\star(\xi - v_h, u - \mathcal{P}_h u) \le \|\xi - v_h\|_\star \|u - \mathcal{P}_h u\|_\star \tag{7.10}$$

By the norm equivalence (7.4), the consequence (3.11) of the definition of C_{int}, the definition of ξ (7.9), and the second bound in (7.5),

$$\begin{aligned} \|\xi - I_h\xi\|_\star \le \sqrt{C_{\mathrm{cont}\star}} \|\xi - I_h\xi\|_{H^1_k(\Omega_R)} &\le \sqrt{C_{\mathrm{cont}\star}}\sqrt{2} C_{\mathrm{int}} h |\xi|_{H^2(\Omega_R)}, \\ &\le \sqrt{C_{\mathrm{cont}\star}}\sqrt{2} C_{\mathrm{int}} h C_{H^2\star} \|u - \mathcal{P}_h u\|_{L^2(\Omega_R)} , \end{aligned}$$

and the result (7.8) follows from combining this last inequality with (7.10). ∎

8. Adjoint approximability.

DEFINITION 8.1 (Adjoint solution operator $\mathcal{S}^*$). *Given $f \in L^2(\Omega_R)$, let $\mathcal{S}^* f$ be defined as the solution of the variational problem*

$$\textit{find } \mathcal{S}^* f \in \mathcal{H} \quad \textit{such that} \quad a(v, \mathcal{S}^* f) = (v, f)_{L^2(\Omega_R)} \quad \textit{for all } v \in \mathcal{H}. \tag{8.1}$$

$\mathcal{S}^*$ can be therefore understood as the solution operator of the adjoint problem to the variational problem (2.5) with data in $L^2(\Omega_R)$.

Green's second identity applied to outgoing solutions of the Helmholtz equation implies that $\langle \mathrm{DtN}_k\psi, \overline{\phi}\rangle_{\Gamma_R} = \langle \mathrm{DtN}_k\phi, \overline{\psi}\rangle_{\Gamma_R}$ (see, e.g., [44, Lemma 6.13]); thus $a(\overline{v}, u) = a(\overline{u}, v)$ and so the definition (8.1) implies that

$$a(\overline{\mathcal{S}^* f}, v) = (\overline{f}, v)_{L^2(\Omega_R)} \quad \text{for all } v \in \mathcal{H}; \tag{8.2}$$

i.e. $\mathcal{S}^* f$ is the complex-conjugate of an outgoing Helmholtz solution.

Following [41], we define the quantity $\eta(\mathcal{H}_h)$ by

$$\eta(\mathcal{H}_h) := \sup_{f\in L^2(\Omega_R)} \min_{v_h\in\mathcal{H}_h} \frac{\|\mathcal{S}^* f - v_h\|_{H^1_k(\Omega_R)}}{\|f\|_{L^2(\Omega_R)}}; \tag{8.3}$$

observe that this definition implies that, given $f \in L^2(\Omega_R)$,

$$\text{there exists } w_h \in \mathcal{H}_H \text{ such that } \|\mathcal{S}^* f - w_h\|_{H^1_k(\Omega_R)} \le \eta(\mathcal{H}_h)\, \|f\|_{L^2(\Omega_R)}\,. \tag{8.4}$$

Lemma 8.2. *Assume that* A, n, *and* Ω_- *are nontrapping (and so* (3.5) *holds with* C_{sol} *independent of* k*).*

(i) If $\Gamma \in C^{1,1}$, $\mathsf{A} \in C^{1,1}$, *and* $p = 1$, *then*

$$\eta(\mathcal{H}_h) \le hk \left[\sqrt{2} C_{\mathrm{int}} C_{H^2} C_{\mathrm{sol}} R \left(n_{\max} + \frac{1}{k_0 R_0 C_{\mathrm{sol}}} + 2 \right)\right]. \tag{8.5}$$

(ii) If Ω_- *is a Dirichlet obstacle (so that* $\mathcal{H} = H^1_{0,D}(\Omega_R)$*),* Γ *is analytic,* $\mathsf{A} = \mathsf{I}$, $n = 1$, $p \ge 1$, *and the triangulation* $\mathcal{T}_h$ *in the definition of* $\mathcal{H}_h$ (2.8) *satisfies the quasi-uniformity assumption [38, Assumption 5.1], then there exists* $C_{\mathrm{MS}} = C_{\mathrm{MS}}(\Omega_-)$ *such that*

$$\eta(\mathcal{H}_h) \le C_{\mathrm{MS}} \left[\frac{h}{p} + C_{\mathrm{sol}} R \left(\frac{hk}{p} \right)^p \right]. \tag{8.6}$$

Proof. Part (ii) is proved in [38, Lemma 3.4 and Proposition 5.3]: see [38, Proof of Theorem 5.8], and observe that the nontrapping assumption implies that α in [38] equals zero. We now prove Part (i).

By the consequence (3.11) of the definition of C_{int} (3.10), there exists $v_h \in \mathcal{H}_h$ such that

$$\|\mathcal{S}^* f - v_h\|_{H^1_k(\Omega_R)} \le \sqrt{2} C_{\mathrm{int}} h |\mathcal{S}^* f|_{H^2(\Omega_R)}$$

(indeed, we can take $v_h = I_h(\mathcal{S}^* f)$). By (8.2), the BVP (3.8) is satisfied with $v := \mathcal{S}^* f$ and $\widetilde{f} := f + k^2 n \mathcal{S}^* f$. Applying the bounds (3.9) and (3.5), we obtain

$$\begin{aligned} |\mathcal{S}^* f|_{H^2(\Omega_R)} &\le C_{H^2} \Big(k^2 n_{\max} \|\mathcal{S}^* f\|_{L^2(\Omega_R)} + \|f\|_{L^2(\Omega_R)} \\ &\qquad + \frac{1}{R} \|\nabla(\mathcal{S}^* f)\|_{L^2(\Omega_R)} + \frac{1}{R^2} \|\mathcal{S}^* f\|_{L^2(\Omega_R)} \Big), \\ &\le C_{H^2} C_{\mathrm{sol}} kR \left(n_{\max} + \frac{1}{kR\, C_{\mathrm{sol}}} + \frac{1}{kR} + \frac{1}{(kR)^2} \right) \|f\|_{L^2(\Omega_R)}\,, \end{aligned}$$

and the result (8.5) follows from the assumption that $kR \ge k_0 R_0 \ge 1$ (see (3.1)). ∎

9. Proof of the oscillatory-behaviour bound (3.6).

THEOREM 9.1. *If* A, n, *and* Ω_- *are nontrapping (in the sense that the bound* (3.5) *holds), and additionally* A *and* n *are both* $C^{1,1}$, *then the bound* (3.6) *holds.*

LEMMA 9.2. *To prove Theorem 9.1, it is sufficient to prove that there exists* $k_0 > 0$ *and* $C_{\rm mass} = C_{\rm mass}(\mathsf{A}, n, \Omega_-, R) > 0$ *such that*

$$\|u\|_{L^2(\Omega_{R+1})} \leq C_{\rm mass} \|u\|_{L^2(\Omega_R)} \qquad \text{for all } k \geq k_0. \tag{9.1}$$

Proof. By the well-posedness of the plane-wave scattering problem, H^2 regularity, and linearity, the map $k \mapsto u$ is continuous from $(1,\infty)$ to $H^2(\Omega_R)$. Therefore, the function $k \mapsto \|u\|_{H^2(\Omega_R)} \left(k \|u\|_{H^1_k(\Omega_R)}\right)^{-1}$ is continuous on $[1,\infty)$, and it is sufficient to prove that the bound (3.6) holds for k sufficiently large.

Let $\chi \in C^\infty(\mathbb{R}^d)$ be such that $0 \leq \chi \leq 1$, $\chi = 1$ on Ω_R and $\chi = 0$ on $\mathbb{R}^d \setminus B_{R+1/2}$. Applying the H^2-regularity results [28, Theorems 2.3.3.2, 2.4.2.5, and 2.4.2.7] to χu (with these results valid since A is Lipschitz, $A_{\min} > 0$, both Γ and Γ_R are $C^{1,1}$, and either $\gamma u = 0$ or $\partial_{\mathbf{n}} u = 0$ on Γ), we obtain, in a similar way to the proof of Theorem 6.1, that there exists $C_1 = C_1(\mathsf{A}, n, \Omega_-, R) > 0$, such that

$$|u|_{H^2(\Omega_R)} \leq C_1 k \|u\|_{H^1_k(\Omega_{R+1})}.$$

Therefore to prove (3.9) it is sufficient to prove that there exists $C_2 = C_2(\mathsf{A}, n, \Omega_-, R) > 0$, such that

$$\|u\|_{H^1_k(\Omega_{R+1})} \leq C_2 \|u\|_{H^1_k(\Omega_R)}. \tag{9.2}$$

We now need to show that we can prove (9.2) from (9.1). We claim that

$$\|\nabla u\|_{L^2(\Omega_{R+1})} \leq \sqrt{\frac{n_{\max}}{A_{\min}}} k \|u\|_{L^2(\Omega_{R+1})} \qquad \text{for all } k > 0. \tag{9.3}$$

Indeed, applying Green's identity in Ω_R (which is justified by [35, Theorem 4.4] since $u \in H^1(\Omega_R)$) and recalling that either $\gamma u = 0$ or $\partial_{\mathbf{n}} u = 0$ on Γ, we have that

$$\int_{\Omega_{R+1}} (\mathsf{A}\nabla u) \cdot \overline{\nabla u} - k^2 n |u|^2 = \Re \int_{\Gamma_{R+1}} \overline{u} \frac{\partial u}{\partial r}.$$

By (3.4), the right-hand side is ≤ 0, and (9.3) follows using the inequalities (2.1) and (2.2). Therefore, using (9.3) and (9.1),

$$\|u\|_{H^1_k(\Omega_{R+1})} \leq \sqrt{\frac{n_{\max}}{A_{\min}} + 1}\; k \|u\|_{L^2(\Omega_{R+1})} \leq C_{\rm mass} \sqrt{\frac{n_{\max}}{A_{\min}} + 1}\; k \|u\|_{L^2(\Omega_R)}$$

which implies the bound (9.2), and the result follows. ∎

9.1. Overview of the ideas used in the rest of this section to prove (9.1).

We have therefore reducing proving the oscillatory-behaviour bound (3.6) to proving the bound (9.1), which we prove using *defect measures*. The precise definition of a defect measure is given in Theorem 9.3 below, but the idea is that the defect measure of a Helmholtz solution describes where the mass in phase space $(\mathbf{x}, \boldsymbol{\xi})$ of the solution is concentrated in the high-frequency limit. Two examples of this feature are (i) the defect measure of the plane wave $u^I(\mathbf{x}) := \exp(\mathrm{i}k\mathbf{x}\cdot\mathbf{a})$ is the product of a delta function in phase space, at $\boldsymbol{\xi} = \mathbf{a}$, and Lebesgue measure in $\mathbf{x}$ (see (9.7) below), reflecting the

fact that, at high frequency (and in fact at any frequency), all the "mass" of the plane wave is travelling in the direction $\mathbf{a}$, and (ii) the defect measure of an outgoing solution of the Helmholtz equation is zero on the so-called "directly incoming set" [11, Proposition 3.5], [23, Lemma 3.4], where this set is defined in (9.13) below as points in phase space whose rays under backward propagation don't hit the scatterer.

A key feature of the defect measure of a Helmholtz solution is that it is invariant under the Hamiltonian flow defined by the symbol of the PDE, as long as the flow doesn't encounter the boundary (see Theorem 9.6 below). This is analogous to results about propagation of singularities of the wave equation, where singularities travel along the trajectories of the flow (the *bicharacteristics*), and the projection of these trajectories in space are the *rays*.

For extensive discussion of defect measures in $\mathbb{R}^d$ see [49, Chapter 5], and for material on defect measures on manifolds with boundary see [11], [39], [23]. For discussion on the history of defect measures, see [10].

9.2. Recap of results about defect measures. Before defining defect measures, we need to define the functions on phase space (i.e. the set of positions $\mathbf{x}$ and momenta $\boldsymbol{\xi}$) that may be dually paired with the defect measure. These elements are called *symbols*, defined as functions on the cotangent bundle $T^*\Omega_+$. On $T^*\mathbb{R}^d = \{(\mathbf{x},\boldsymbol{\xi}) : \mathbf{x}\in\mathbb{R}^d, \boldsymbol{\xi}\in\mathbb{R}^d\}$ (and, more generally, locally away from the boundary of $\overline{\Omega}_+$) the *quantisation* of a symbol $b(\mathbf{x},\boldsymbol{\xi}) \in C^\infty_{\rm comp}(T^*\mathbb{R}^d)$ is defined by

$$b\big(\mathbf{x}, (\mathrm{i}k)^{-1}\partial_{\mathbf{x}}\big)u(\mathbf{x}) := \frac{k^d}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}k(\mathbf{x}-\mathbf{y})\cdot\boldsymbol{\xi}}\, b(\mathbf{x},\boldsymbol{\xi})u(\mathbf{y})\,\mathrm{d}\mathbf{y}\mathrm{d}\boldsymbol{\xi}; \tag{9.4}$$

see, e.g., [49, §4]. The analogous definition near the boundary is more involved; see [11, §4.2] (where it involves the so-called *compressed cotangent bundle* of Ω_+, $T^*_{\rm b}\overline{\Omega_+}$) and [39, §1.2]. We will not, in any event, require any specifics of the measure at the boundary in proving Theorem 9.1.

THEOREM 9.3. (Existence of defect measures [49, Theorem 5.2], [11, §4.2].) *Suppose $\{v(k)\}_{k_0\le k<\infty}$ is a collection of functions that is uniformly locally bounded in $L^2(\Omega_+)$, i.e. given $\chi\in C^\infty_{\rm comp}(\mathbb{R}^d)$ there exists $C>0$, depending on χ and k_0 but independent of k, such that*

$$\|\chi v(k)\|_{L^2(\Omega_+)} \le C \quad \text{for all } k\ge k_0. \tag{9.5}$$

Then there exists a sequence $k_\ell\to\infty$ and a non-negative Radon measure μ on $T^*_{\rm b}\overline{\Omega_+}$ (depending on k_ℓ) such that, for any symbol $b(\mathbf{x},\boldsymbol{\xi})\in C^\infty_{\rm comp}(T^*_{\rm b}\overline{\Omega_+})$

$$\big\langle b\big(\mathbf{x},(\mathrm{i}k_\ell)^{-1}\partial_{\mathbf{x}}\big)v(k_\ell), v(k_\ell)\big)\big\rangle_{\Omega_+} \longrightarrow \int b\,\mathrm{d}\mu \quad \text{as } \ell\to\infty. \tag{9.6}$$

In the case of a plane wave $u^I(\mathbf{x}) := \exp(\mathrm{i}k\mathbf{x}\cdot\mathbf{a})$, a direct calculation using (9.4) and the definition of the Fourier transform shows that, for all k,

$$\begin{aligned}\big\langle b\,u^I, u^I\big\rangle_{\mathbb{R}^d} &:= \frac{k^d}{(2\pi)^d}\int_{\mathbb{R}^d}\mathrm{d}\mathbf{x}\int_{\mathbb{R}^d}\mathrm{d}\mathbf{y}\int_{\mathbb{R}^d}\mathrm{d}\boldsymbol{\xi}\,\mathrm{e}^{\mathrm{i}k(\mathbf{x}-\mathbf{y})\cdot\boldsymbol{\xi}}\,\mathrm{e}^{\mathrm{i}k\mathbf{y}\cdot\mathbf{a}}\,\mathrm{e}^{-\mathrm{i}k\mathbf{x}\cdot\mathbf{a}}b(\mathbf{x},\boldsymbol{\xi})\\ &= \int_{\mathbb{R}^d} b(\mathbf{x},\mathbf{a})\mathrm{d}\mathbf{x};\end{aligned} \tag{9.7}$$

i.e. for any sequence $k_\ell\to\infty$, the corresponding defect measure of u^I is the product of the Lebesgue measure in $\mathbf{x}$ by a delta measure at $\boldsymbol{\xi}=\mathbf{a}$; we therefore talk about *the* (as opposed to *a*) defect measure of u^I.

The next lemma proves that, if u is the solution of the plane-wave scattering problem and χ is an arbitrary cut-off function, then χu is uniformly bounded in k (on compact subsets of Ω_+); existence of a defect measure of u then follows from Theorem 9.3. In the rest of this section, to emphasise the k-dependence of u, we write $u = u(k)$.

LEMMA 9.4. *Let $u(k)$ be the solution of the plane-wave scattering problem of Definition 2.2. Assume that A, n, and Ω_- are nontrapping. Then there exists $C(\mathsf{A}, n, \Omega_-, R, k_0) > 0$ such that*

$$\|u(k)\|_{L^2(\Omega_R)} \leq C \quad \text{for all } k \geq k_0. \tag{9.8}$$

Proof. Let $\chi \in C^\infty_{\mathrm{comp}}(\mathbb{R}^d)$ be such that $\chi = 1$ in a neighbourhood of the scatterer Ω_{sc}. Let $v := u^S + \chi u^I$, so that $u = (1-\chi)u^I + v$. Since $\|u^I(k)\|_{L^2(\Omega_R)} \leq C_1(R)$ for all $k > 0$, the result (9.8) will follow if we prove a uniform bound on $\|v(k)\|_{L^2(\Omega_R)}$. The definition of v implies that v satisfies the Sommerfeld radiation condition, either $\gamma v = 0$ or $\partial_{\mathbf{n}} v = 0$ on Γ, and, with $\mathcal{L}_{\mathsf{A},n} w := \nabla\cdot(\mathsf{A}\nabla w) + k^2 n w$ and $[A, B] := AB - BA$,

$$\mathcal{L}_{\mathsf{A},n} v = -\mathcal{L}_{\mathsf{A},n}\big((1-\chi)u^I\big) = \big[\mathcal{L}_{\mathsf{A},n}, \chi\big]u^I - (1-\chi)\mathcal{L}_{\mathsf{A},n}u^I = \big[\mathcal{L}_{\mathsf{A},n}, \chi\big]u^I,$$

since $\mathcal{L}_{\mathsf{A},n}u^I = 0$ when $1 - \chi \neq 0$. By explicit calculation, using the fact that $u^I(\mathbf{x}) = \exp(\mathrm{i}k\mathbf{x}\cdot\mathbf{a})$,

$$\big\|\big[\mathcal{L}_{\mathsf{A},n}, \chi\big]u^I\big\|_{L^2(\Omega_R)} \leq C_1(\|\mathsf{A}\|_{L^\infty(\Omega_R)}, \|D\mathsf{A}\|_{L^\infty(\Omega_R)}, \chi)\, k;$$

the nontrapping bound (3.5) then implies that $\|v(k)\|_{L^2(\Omega_R)} \leq C_2$ with C_2 independent of k, and the result follows. ∎

Away from Γ, the flow $\varphi_t = (\mathbf{x}(t), \boldsymbol{\xi}(t))$ is defined as the solution of the Hamiltonian system

$$\dot{x_i}(t) = \partial_{\xi_i} p\big(\mathbf{x}(t), \boldsymbol{\xi}(t)\big), \qquad \dot{\xi}_i(t) = -\partial_{x_i} p\big(\mathbf{x}(t), \boldsymbol{\xi}(t)\big),$$

where the Hamiltonian is given by the semi-classical principal symbol of the Helmholtz equation (2.3), namely

$$p(\mathbf{x}, \boldsymbol{\xi}) := \sum_{i=1}^{d}\sum_{j=1}^{d} A_{ij}(\mathbf{x})\xi_i\xi_j - n(\mathbf{x}).$$

Near Γ, φ_t must be understood in terms of Melrose–Sjöstrand generalized bicharacteristics; see [29, Section 24.3]. Since the flow over the interior is, by definition, generated by the Hamilton vector field H_p, we have $\partial_t(b\circ\varphi_t) = H_p b$ for any symbol b supported away from the boundary.

Observe that, away from Ω_{sc}, $p(\mathbf{x}, \boldsymbol{\xi}) = |\boldsymbol{\xi}|^2 - 1$ and so $\dot{x_i} = 2\xi_i$ and $\dot{\xi}_i = 0$. In our arguments below we consider the flow when $p = 0$, i.e. $|\boldsymbol{\xi}| = 1$. This is because of the following result.

THEOREM 9.5. (Support of defect measure [49, Theorem 5.4], [11, Equation 3.17].) *Suppose $u(k)$ satisfies (9.8), and let μ be any defect measure of $u(k)$. Then* $\operatorname{supp}\mu \subset \{(\mathbf{x}, \boldsymbol{\xi}) : p(\mathbf{x}, \boldsymbol{\xi}) = 0\}$.

Therefore, away from Ω_{sc}, μ is only non-zero when the flow has $|\dot{\mathbf{x}}| = 2|\boldsymbol{\xi}| = 2$, i.e. the flow has speed 2.

THEOREM 9.6. (Invariance of defect measure under the flow [49, Theorem 5.4], [11, Proposition 4.4].) *Suppose $u(k)$ satisfies (9.8), and let μ be any defect measure of*

$u(k)$. *For any* $b(\mathbf{x},\boldsymbol{\xi}) \in C^\infty_{\mathrm{comp}}(T^*\mathbb{R}^d)$ *supported away from* $T^*\Omega_{\mathrm{sc}}$, *we have* $\mu(H_p b) = 0$.

In the proof of (9.1), we use the consequence of Theorem 9.6 that given $A \subset T^*\mathbb{R}^d$ such that $\pi_{\mathbf{x}}(\varphi_s(A)) \cap \Omega_{\mathrm{sc}} = \emptyset$ for s between 0 and t, we have the invariance

$$\mu(\varphi_t(A)) = \mu(A); \tag{9.9}$$

here $\pi_{\mathbf{x}}$ denotes projection in the $\mathbf{x}$ variables (i.e. $\pi_{\mathbf{x}}((\mathbf{x},\boldsymbol{\xi})) = \mathbf{x}$).

9.3. Proof of (9.1) using defect measures. The following lemma reduces proving the bound (9.1) to proving a statement about defect measures.

LEMMA 9.7. *Let* $0 < R_0 < R$ *be such that* $\Omega_{\mathrm{sc}} \subset\subset B_{R_0}$. *If every defect measure of* u *is non-zero and there exists* $C_{R,R_0} > 0$ *such that, for every defect measure* μ *of* u,

$$\mu(T^*\Omega_{R+2}) \leq C_{R,R_0}\mu(T^*\Omega_{R_0}), \tag{9.10}$$

then the bound (9.1) *holds.*

Proof. We prove the contrapositive. Suppose (9.1) fails; we aim to exhibit a defect measure associated to u for which (9.10) fails. Then, for any $C_1 > 0$, there exists a sequence $(k_n)_{n=1}^\infty$, with $k_n \to \infty$, such that

$$\|u(k_n)\|_{L^2(\Omega_{R+1})} \geq C_1\|u(k_n)\|_{L^2(\Omega_R)}; \tag{9.11}$$

we choose $C_1 := 2C_{R,R_0}$. By Lemma 9.4, the sequence $\{u(k_n)\}_{n=1}^\infty$ is locally uniformly bounded and Theorem 9.3 implies that, by passing to a subsequence, there exists a defect measure μ of u associated to the subsequence, which we again denote k_n. Let $\chi_0, \chi_1 \in C^\infty(\mathbb{R}^d)$ be such that $0 \leq \chi_0, \chi_1 \leq 1$, and

$$\operatorname{supp}\chi_1 \subset B_{R+2}, \quad \chi_1 = 1 \text{ in } B_{R+1}, \quad \operatorname{supp}\chi_0 \subset B_R, \quad \chi_0 = 1 \text{ in } B_{R_0}.$$

The bound (9.11) then implies that

$$\|\chi_1 u(k_n)\|_{L^2(\Omega_+)} \geq 2C_{R,R_0}\|\chi_0 u(k_n)\|_{L^2(\Omega_+)}. \tag{9.12}$$

Passing to the limit $n \to \infty$ and using the property of defect measure (9.6), we obtain that

$$\int \chi_1^2 \,\mathrm{d}\mu \geq 2C_{R,R_0}\int \chi_0^2 \,\mathrm{d}\mu.$$

The definitions of χ_0 and χ_1 imply that

$$\int \chi_0^2 \,\mathrm{d}\mu \geq \int 1_{T^*\Omega_{R_0}} \,\mathrm{d}\mu = \mu(T^*\Omega_{R_0})$$

(where 1_A denotes the indicator function of a set A) and

$$\int \chi_1^2 \,\mathrm{d}\mu \leq \int 1_{T^*\Omega_{R+2}} \,\mathrm{d}\mu = \mu(T^*\Omega_{R+2});$$

hence

$$\mu(T^*\Omega_{R+2}) \geq 2C_{R,R_0}\mu(T^*\Omega_{R_0}),$$

contradicting (9.10). ∎

Let $\mathcal{I}$ denote the *directly incoming set* defined by

$$\mathcal{I} := \left\{ \rho \in T^*(\Omega_+ \backslash \Omega_{\rm sc}),\ \text{s.t. } \pi_{\mathbf{x}}\Big(\bigcup_{t\geq 0} \varphi_{-t}(\rho) \Big) \cap \Omega_{\rm sc} = \emptyset \right\}; \tag{9.13}$$

where $\pi_{\mathbf{x}}$ denotes projection in the $\mathbf{x}$ variables (i.e. $\pi_{\mathbf{x}}((\mathbf{x},\boldsymbol{\xi})) = \mathbf{x}$). that is, $\mathcal{I}$ is everything that never hits the scatterer under backward flow. Let $\Gamma_+ := (T^*\Omega_+)\backslash\mathcal{I}$. These definitions do not require the generalized bicharacteristic flow φ_t to be defined in $T^*\Omega_{\rm sc}$, but when the flow is defined everywhere, Γ_+ is the forward generalized bicharacteristic flowout of $\Omega_{\rm sc}$, that is

$$\Gamma_+ = \left\{ \bigcup_{t\geq 0} \varphi_t(\rho) \ :\ \rho \in T^*\Omega_{\rm sc} \right\} \text{ when } \varphi_t \text{ is defined everywhere.}$$

The following lemma uses outgoingness of u^S to show that, given a set E in phase space, the mass of u lying over E is *either* in the forward flowout Γ_+ *or* associated to the incident wave u^I.

LEMMA 9.8. *For any Borel set $E \subset T^*\Omega$, $\mu(E \setminus \Gamma_+) = \mu^I(E \setminus \Gamma_+)$, where μ is any defect measure of u, and μ^I is the defect measure of u^I.*

Proof. Let k_ℓ be the sequence associated to the particular defect measure of u. By Lemma 9.4, $u^S(k_\ell)$ is uniformly locally bounded, and so there exists a subsequence k_{ℓ_m} and a defect measure associated to u^S, denoted by μ^S. Then, by linearity and (9.6), $\mu = \mu^S + \mu^I$. It is therefore sufficient to prove that $\mu^S(E \setminus \Gamma_+) = 0$. But, by the definition of Γ_+, $E \setminus \Gamma_+ \subset \mathcal{I}$, and $\mu^S(\mathcal{I}) = 0$ by [11, Proposition 3.5], [23, Lemma 3.4], since u^S is outgoing. ∎

Proof. [Proof of Theorem 9.1] By Lemmas 9.2 and 9.7 it is sufficient to prove the bound (9.10) (observe that the hypothesis in Lemma 9.7 that every defect measure of u is non-zero holds by Lemma 9.8 since $\mu^I(\mathcal{I}) \neq 0$). Let $R_{\rm sc} := \max_{\mathbf{x}\in\Omega_{\rm sc}} |\mathbf{x}|$. We claim that it is sufficient to show that, for any $\rho > R_{\rm sc}$ there exists $\varepsilon = \varepsilon(R_{\rm sc}, \rho)$, with $\varepsilon(R_{\rm sc}, \rho)$ is an increasing function of ρ, and $C = C(\rho, \varepsilon) > 0$ such that

$$\mu(T^*(B_{\rho+\varepsilon} \setminus B_\rho)) \leq C(\rho,\varepsilon)\mu(T^*\Omega_\rho). \tag{9.14}$$

Indeed, we now show that the bound (9.10) then follows by using (9.14) repeatedly. Since $\varepsilon(R_{\rm sc}, \rho)$ is an increasing function of ρ, if $\varepsilon^* := \varepsilon(R_{\rm sc}, R_0)$, then (9.14) implies, with $C(\rho) := C(\rho, \varepsilon(R_{\rm sc}, \rho))$,

$$\mu(T^*(B_{\rho+\varepsilon^*} \setminus B_\rho)) \leq C(\rho)\,\mu(T^*\Omega_\rho) \quad \text{ for all } \rho \geq R_0. \tag{9.15}$$

The bound (9.10) then follows by applying (9.15) with $\rho = R_0$, $\rho = R_0 + \varepsilon^*$, ..., $\rho = R_0 + m\varepsilon^*$, where $m = \lceil (R + 2 - R_0)/\varepsilon^* \rceil$.

It is therefore sufficient to prove the bound (9.14); we introduce the notation that $A := B_{\rho+\varepsilon} \setminus B_\rho$, and observe that (9.14) then reads $\mu(T^*A) \leq C_{\rho,\varepsilon}\mu(T^*\Omega_\rho)$. We prove this bound by combining the following three inequalities:

$$\mu(T^*A) \leq \mu(T^*A \cap \Gamma_+) + \mu_I(T^*A) = \mu(T^*A \cap \Gamma_+) + |A| \tag{9.16}$$

(where $|\cdot|$ denotes Lebesgue measure in $\mathbb{R}^d$),

$$\mu(T^*A \cap \Gamma_+) \leq \mu(T^*(B_\rho \setminus B_{\rho_0})) \leq \mu(T^*\Omega_\rho), \tag{9.17}$$

where $\rho_0 := (\rho + R_{\rm sc})/2$, and

$$\mu(T^*\Omega_\rho) \geq \delta|\Omega_\rho| \tag{9.18}$$

for some $\delta > 0$. Indeed, using (9.16), (9.17), and (9.18), we have

$$\mu(T^*A) \leq \Big(1 + |A|(\delta|\Omega_\rho|)^{-1}\Big)\mu(T^*\Omega_\rho),$$

which is (9.14). We prove (9.16) and (9.18) using Lemma 9.8 and the structure of μ^I, and (9.17) using invariance of defect measures under the flow outside of $T^*\Omega_{\rm sc}$ (i.e. Theorem 9.6).

Proof of (9.16). Lemma 9.8 implies that

$$\mu(T^*A) = \mu(T^*A \cap \Gamma_+) + \mu(T^*A \setminus \Gamma_+) \leq \mu(T^*A \cap \Gamma_+) + \mu_I(T^*A).$$

By (9.7), μ_I is a δ-measure on $\boldsymbol{\xi} = \mathbf{a}$ times Lebesgue measure in $\mathbf{x}$, so $\mu_I(T^*A) = |A|$, (where $|\cdot|$ denotes Lebesgue measure in $\mathbb{R}^d$) and (9.16) follows.

Proof of (9.17). Recall that, for $X \subset\subset \mathbb{R}^d \setminus \overline{\Omega_{\rm sc}}$, $S^*X := \{(\mathbf{x},\boldsymbol{\xi}) : x \in X, \boldsymbol{\xi} \in \mathbb{R}^d \text{ with } |\boldsymbol{\xi}| = 1\}$, and observe that, by Theorem 9.5, $\mu(T^*A \cap \Gamma_+) = \mu(S^*A \cap \Gamma_+)$ and $\mu(T^*(B_\rho \setminus B_{\rho_0})) = \mu(S^*(B_\rho \setminus B_{\rho_0}))$; we therefore only need to prove that

$$\mu(S^*A \cap \Gamma_+) \leq \mu(S^*(B_\rho \setminus B_{\rho_0})). \tag{9.19}$$

We first introduce some notation that allows us to bound $\mu(S^*A \cap \Gamma_+)$ using only the invariance of defect measure (9.9) in the exterior of $\Omega_{\rm sc}$. Given $\mathbf{b} \in \mathbb{R}^d$ with $|\mathbf{b}| = 1$ and $\widetilde{\rho} > R_{\rm sc}$, let $\Omega_{{\rm sc},\widetilde{\rho},\mathbf{b}} \subset \mathbb{R}^d$ and $\Lambda_{{\rm sc},\widetilde{\rho},\mathbf{b}} \subset S^*\Omega_+$ be defined by

$$\Omega_{{\rm sc},\widetilde{\rho},\mathbf{b}} := \Big(\bigcup_{t\geq 0}\big(\Omega_{\rm sc} + t\mathbf{b}\big)\Big) \cap \Omega_{\widetilde{\rho}} \quad \text{and} \quad \Lambda_{{\rm sc},\rho+\varepsilon,\mathbf{b}} := \Omega_{{\rm sc},\widetilde{\rho},\mathbf{b}} \times \{\mathbf{b}\};$$

i.e. $\Omega_{{\rm sc},\widetilde{\rho},\mathbf{b}}$ equals the union of all possible translations of $\Omega_{\rm sc}$ in the direction $\mathbf{b}$, intersected with $\Omega_{\widetilde{\rho}}$, and $\Lambda_{{\rm sc},\widetilde{\rho},\mathbf{b}}$ equals these points paired with the direction $\mathbf{b}$. Since the spatial projections of the flow outside $\Omega_{\rm sc}$ are straight lines,

$$\Gamma_+ \cap S^*\Omega_{\widetilde{\rho}} \cap \{\boldsymbol{\xi} = \mathbf{b}\} = \Big\{(\mathbf{x},\mathbf{b}) \in S^*\Omega_{\widetilde{\rho}} : \exists s \geq 0 \text{ s.t. } \mathbf{x} - s\mathbf{b} \in \Omega_{\rm sc}\Big\}.$$

Therefore

$$\Gamma_+ \cap S^*\Omega_{\widetilde{\rho}} \cap \{\boldsymbol{\xi} = \mathbf{b}\} \subset \Lambda_{{\rm sc},\widetilde{\rho},\mathbf{b}}, \qquad \Gamma_+ \cap S^*\Omega_{\widetilde{\rho}} \subset \bigcup_{\mathbf{b}\in\mathbb{R}^d,|\mathbf{b}|=1} \Lambda_{{\rm sc},\widetilde{\rho},\mathbf{b}}, \tag{9.20}$$

and thus, for any $\varepsilon > 0$,

$$S^*A \cap \Gamma_+ = S^*A \cap S^*\Omega_{\rho+\varepsilon} \cap \Gamma_+ \subset S^*A \cap \Big(\bigcup_{\mathbf{b}\in\mathbb{R}^d,|\mathbf{b}|=1} \Lambda_{{\rm sc},\rho+\varepsilon,\mathbf{b}}\Big), \tag{9.21}$$

Recall that $\rho_0 := (\rho + R_{\rm sc})/2$. Let

$$t_0 := \frac{\rho_0 - R_{\rm sc}}{4} = \frac{\rho - R_{\rm sc}}{8} \tag{9.22}$$

and

$$\varepsilon := -\rho + \sqrt{R_{\rm sc}^2 + \left(\frac{\rho - R_{\rm sc}}{4} + \sqrt{\rho^2 - R_{\rm sc}^2}\right)^2}; \tag{9.23}$$

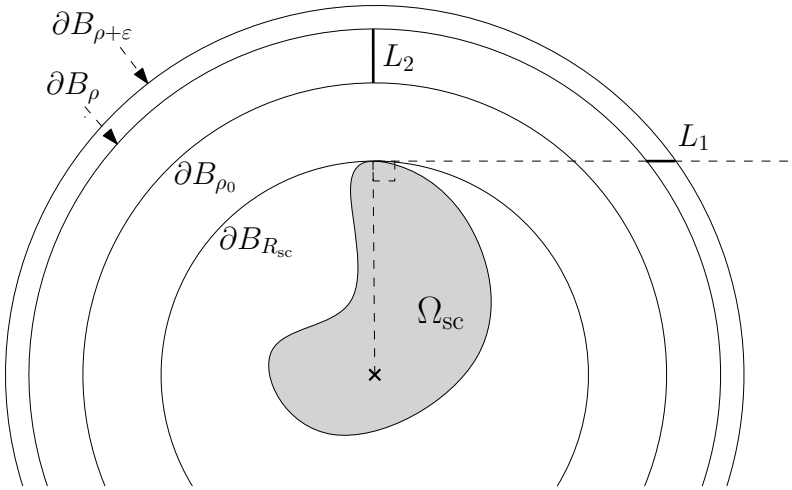

FIG. 9.1. *Figure showing the lengths L_1 and L_2 defined by (9.27).*

observe that $\varepsilon > 0$ and ε is an increasing function of ρ, as claimed underneath (9.14). We now claim that, with these definitions of t_0 and ε,

$$\bigcup_{0\le t\le t_0} \varphi_t\big(S^*(B_\rho \setminus B_{\rho_0})\big) \cap \Omega_{\rm sc} = \emptyset \tag{9.24}$$

(i.e., the forward flowout of the annulus $B_\rho \setminus B_{\rho_0}$ does not hit the scatterer for $0 \le t \le t_0$) and

$$S^*A \cap \Bigg(\bigcup_{\mathbf{b}\in\mathbb{R}^d, |\mathbf{b}|=1} \Lambda_{{\rm sc},\rho+\varepsilon,\mathbf{b}} \Bigg) \subset \varphi_{t_0}\big(S^*(B_\rho \setminus B_{\rho_0})\big). \tag{9.25}$$

(Since $S^*A \cap \Gamma_+$ is contained in the left-hand side of (9.25) by (9.21), (9.25) says that the forward flowout of $B_\rho \setminus B_{\rho_0}$ in time t_0 covers all points in S^*A that are ever reached by flowout from $T^*\Omega_{sc}$.) Outside $\Omega_{\rm sc}$ the flow has speed 2, and its spatial projections are straight lines. Therefore (9.24) is ensured if $t_0 < (\rho_0 - R_{\rm sc})/2$, which is ensured by (9.22).

We now show that (9.25) holds. Since

$$(\mathbf{x},\mathbf{b}) = (\mathbf{x} - 2t_0\mathbf{b} + 2t_0\mathbf{b}, \mathbf{b}) = \varphi_{t_0}(\mathbf{x} - 2t_0\mathbf{b}, \mathbf{b}),$$

(9.25) follows from showing that $(\mathbf{x}-2t_0\mathbf{b},\mathbf{b}) \in S^*(B_\rho \setminus B_{\rho_0})$, i.e. $\mathbf{x}-2t_0\mathbf{b} \in B_\rho \setminus B_{\rho_0}$, for all $(\mathbf{x},\mathbf{b})$ belonging to the left-hand side of (9.25). For such $(\mathbf{x},\mathbf{b})$, by definition,

$$\rho \le |\mathbf{x}| \le \rho + \epsilon, \text{ and } \mathbf{x} - s\mathbf{b} \in \Omega_{\rm sc} \tag{9.26}$$

for some $s \ge 0$. We now claim that for such $(\mathbf{x},\mathbf{b})$,

$$\mathbf{x} - \ell\mathbf{b} \in B_\rho \setminus B_{\rho_0} \quad \text{for all} \quad L_1 < \ell \le L_2,$$

where

$$L_1 := \sqrt{(\rho+\varepsilon)^2 - R_{\rm sc}^2} - \sqrt{\rho^2 - R_{\rm sc}^2}, \qquad L_2 := \rho - \rho_0. \tag{9.27}$$

This is because, on the one hand, a ray of length $> L_1$ starting from a point $\mathbf{x}$ in a direction $-\mathbf{b}$, with $(\mathbf{x}, \mathbf{b})$ satisfying (9.26), will automatically enter B_ρ. Indeed, the longest such ray that does not intersect B_ρ has length L_1, as shown in Figure 9.1. On the other hand, a ray of length $\leq L_2$ starting from a point $\mathbf{x}$ in a direction $-\mathbf{b}$, with $(\mathbf{x}, \mathbf{b})$ satisfying (9.26), will not intersect B_{ρ_0}. Indeed, the shortest such ray that enters $\overline{B_{\rho_0}}$ has length L_2, as shown in Figure 9.1. It is then straightforward to check that $L_1 < 2t_0 \leq L_2$ when t_0 is given by (9.22) and ε is given by (9.23), so that (9.25) holds.

We now prove the bound (9.19) on $\mu(S^*A \cap \Gamma_+)$ using (9.24) and (9.25). Because of (9.24), we can use (9.9) to find that

$$\mu\big(\varphi_{t_0}(S^*(B_\rho \setminus B_{\rho_0}))\big) = \mu(S^*(B_\rho \setminus B_{\rho_0}));$$

using this with (9.21) and (9.25), we obtain (9.19), and thus (9.17).

Proof of (9.18). Using Lemma 9.8 and the structure of μ_I, we have

$$\begin{aligned}\mu(T^*\Omega_\rho) &\geq \mu(T^*\Omega_\rho \setminus \Gamma_+) = \mu_I(T^*\Omega_\rho \setminus \Gamma_+)\\ &= \mu_I\big((T^*\Omega_\rho \setminus \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\big) + \mu_I\big((T^*\Omega_\rho \setminus \Gamma_+) \cap \{\boldsymbol{\xi} \neq \mathbf{a}\}\big)\\ &= \Big|\pi_{\mathbf{x}}\Big((T^*\Omega_\rho \setminus \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\Big)\Big|. \end{aligned} \tag{9.28}$$

Since

$$\pi_{\mathbf{x}}\Big((T^*\Omega_\rho \setminus \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\Big) \cup \pi_{\mathbf{x}}\Big((T^*\Omega_\rho \cap \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\Big) \supset \Omega_\rho.$$

we obtain

$$\Big|\pi_{\mathbf{x}}\Big((T^*\Omega_\rho \setminus \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\Big)\Big| \geq |\Omega_\rho| - \Big|\pi_{\mathbf{x}}\Big((T^*\Omega_\rho \cap \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\Big)\Big|. \tag{9.29}$$

By the first inclusion in (9.20),

$$\big|\pi_{\mathbf{x}}\big((T^*\Omega_\rho \cap \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\big)\big| \leq \big|\Omega_{\mathrm{sc},R,\mathbf{a}}\big|, \tag{9.30}$$

with this inequality expressing the fact that any parts of the scattered wave travelling in direction $\mathbf{a}$ must lie in $\Omega_{\mathrm{sc},R,\mathbf{a}}$. Combining (9.29) with (9.30) yields

$$\Big|\pi_{\mathbf{x}}\big((T^*\Omega_\rho \setminus \Gamma_+) \cap \{\boldsymbol{\xi} = \mathbf{a}\}\big)\Big| \geq |\Omega_\rho| - |\Omega_{\mathrm{sc},R,\mathbf{a}}|. \tag{9.31}$$

Since $\Omega_{\mathrm{sc},R,\mathbf{a}} \subsetneq \Omega_\rho$, there exists $\delta > 0$ such that $|\Omega_\rho| - |\Omega_{\mathrm{sc},R,\mathbf{a}}| \geq \delta|\Omega_\rho|$, and thus (9.28) and (9.31) imply that (9.18) holds; the proof is complete. □

Remark 9.9 (What if impedance boundary conditions are imposed on Γ_R?). *If the impedance boundary condition $\partial_{\mathbf{n}} u^S - iku^S = 0$ is imposed on Γ_R (as an approximation of DtN_k), then there are additional reflections on Γ_R [39], μ^S has support on the incoming set, and Lemma 9.8 no longer holds.*

Remark 9.10 (Proving Theorem 9.1 in the trapping case). *In the trapping case, $\|u(k)\|_{L^2(\Omega_R)}$ may no longer be uniformly bounded, as it is in Lemma 9.4, since (3.5) no longer holds with C_{sol} bounded independently of k. If a subsequence of k's exists along which $\|u(k)\|_{L^2(\Omega_R)}$ is uniformly bounded, we may obtain a contradiction by the same argument as above by considering this subsequence. Thus, we can assume, without loss of generality, that $\|u(k)\|_{L^2(\Omega_R)} \to \infty$. Now instead of defining defect measures of $u(k)$, one can instead define defect measures of $u(k)/\|u(k)\|_{L^2(\Omega_R)}$. If*

R is sufficiently large, then the bound in [12, Theorem 1.1] (i.e. the fact that the nontrapping cut-off resolvent estimate holds, even under trapping, if the supports of the cut-offs on both sides are sufficiently far away from the scatterer) implies that $v := u(k)/\|u(k)\|_{L^2(\Omega_R)}$ satisfies (9.5). Any defect measure of v is then immediately non-zero, since $\mu(\chi^2) \geq 1$ for any χ with $\operatorname{supp}\chi \supset B_R$. Lemma 9.7 goes through as before after multiplying both sides of (9.12) by $\|u(k)\|^{-2}_{L^2(\Omega_R)}$. The main change needed to the rest of the proof is to take into account the fact that a defect measure of $u^I(k)/\|u(k)\|_{L^2(\Omega_R)}$ is zero when $\|u(k)\|_{L^2(\Omega_R)}$ grows through the sequence k_ℓ associated with that measure. In this situation, however, the bound (9.16) becomes $\mu(T^*A) \leq \mu(T^*A \cap \Gamma_+)$; combining this with (9.17) we obtain $\mu(T^*A) \leq 2\mu(T^*\Omega_R)$, from which the key bound (9.14) (and hence the result of the theorem) follows.

10. Proof of Theorems 4.1 and 4.2.

LEMMA 10.1 (Aubin-Nitsche analogue via elliptic projection). *Assuming that the Galerkin solution u_h to the variational problem (2.9) exists, if*

$$hk^2\eta(\mathcal{H}_h) \leq \mathcal{C}_1, \qquad \text{where} \qquad \mathcal{C}_1 := \frac{1}{2\sqrt{2}C_{\mathrm{cont}\star}C_{H^2\star}C_{\mathrm{int}}}, \tag{10.1}$$

then

$$\|u-u_h\|_{L^2(\Omega_R)} \leq 2C_{\mathrm{cont}\star}\eta(\mathcal{H}_h)\,\|u-w_h\|_{H^1_k(\Omega_R)} \qquad \text{for all } w_h \in \mathcal{H}_h.$$

Proof. Let $\xi = \mathcal{S}^*(u-u_h)$; i.e. ξ is the solution of variational problem

$$\text{find } \xi \in \mathcal{H} \text{ such that } a(v,\xi) = (v, u-u_h)_{L^2(\Omega_R)} \qquad \text{for all } v \in \mathcal{H}.$$

Then, by Galerkin orthogonality (7.6) and the definition of $a_\star(\cdot,\cdot)$ (7.1),

$$\begin{aligned}\|u-u_h\|^2_{L^2(\Omega_R)} = a(u-u_h,\xi) &= a(u-u_h,\xi-v_h) \qquad \text{for all } v_h \in \mathcal{H}_h, \\ &= a_\star(u-u_h,\xi-v_h)_{L^2(\Omega_R)} - k^2(u-u_h,\xi-v_h)_{L^2(\Omega_R)}.\end{aligned} \tag{10.2}$$

We choose $v_h = \mathcal{P}_h\xi$, and then use (in the following order) (i) the Galerkin orthogonality (7.6), (ii) continuity of $a_\star(\cdot,\cdot)$, (iii) the bound (7.8), (iv) the upper bound in the norm equivalence (7.4) and the bound (7.7), and (v) the consequence (8.4) of the definition of η to obtain that, for all $w_h \in \mathcal{H}_H$,

$$\begin{aligned}&\|u-u_h\|^2_{L^2(\Omega_R)} = a_\star(u-w_h,\xi-\mathcal{P}_h\xi)_{L^2(\Omega_R)} - k^2(u-u_h,\xi-v_h)_{L^2(\Omega_R)}, \\ &\leq \|u-w_h\|_\star\,\|\xi-\mathcal{P}_h\xi\|_\star + k^2\,\|u-u_h\|_{L^2(\Omega_R)}\,\|\xi-\mathcal{P}_h\xi\|_{L^2(\Omega_R)}, \\ &\leq \left(\|u-w_h\|_\star + hk^2\sqrt{2}C_{\mathrm{int}}C_{H^2\star}\sqrt{C_{\mathrm{cont}\star}}\,\|u-u_h\|_{L^2(\Omega_R)}\right)\|\xi-\mathcal{P}_h\xi\|_\star, \\ &\leq \left(\sqrt{C_{\mathrm{cont}\star}}\,\|u-w_h\|_{H^1_k(\Omega_R)} + hk^2\sqrt{2}C_{\mathrm{int}}C_{H^2\star}\sqrt{C_{\mathrm{cont}\star}}\,\|u-u_h\|_{L^2(\Omega_R)}\right) \\ &\qquad\qquad \cdot \sqrt{C_{\mathrm{cont}\star}}\min_{v_h\in\mathcal{H}_h}\|\xi-v_h\|_{H^1_k(\Omega_R)}, \\ &\leq \left(\sqrt{C_{\mathrm{cont}\star}}\,\|u-w_h\|_{H^1_k(\Omega_R)} + hk^2\sqrt{2}C_{\mathrm{int}}C_{H^2\star}\sqrt{C_{\mathrm{cont}\star}}\,\|u-u_h\|_{L^2(\Omega_R)}\right) \\ &\qquad\qquad \cdot \sqrt{C_{\mathrm{cont}\star}}\eta(\mathcal{H}_h)\,\|u-u_h\|_{L^2(\Omega_R)};\end{aligned} \tag{10.3}$$

the result then follows. ∎

REMARK 10.2 (Advantage of elliptic-projection over standard duality argument). *Comparing (10.2) and (10.3) we see the advantage of the elliptic-projection argument*

over the standard duality argument: in (10.3), Galerkin orthogonality for $a_\star(\cdot,\cdot)$ has allowed us to obtain $u - w_h$ (with w_h arbitrary) as opposed to $u - u_h$ in the first argument of the sesquilinear form on the right-hand side, leading to the bound (5.3) instead of (5.2). The price for this is that we have an additional L^2 inner product on the right-hand side of (10.3), and controlling this leads to the condition (10.1).

Recall that, by the Cauchy–Schwarz inequality and the inequality (3.3), $a(\cdot,\cdot)$ is continuous, i.e., for all $u, v \in \mathcal{H}$,

$$\left|a(u,v)\right| \leq C_{\mathrm{cont}} \left\|u\right\|_{H^1_k(\Omega_R)} \left\|v\right\|_{H^1_k(\Omega_R)} \quad \text{where} \quad C_{\mathrm{cont}} := \max\left\{A_{\max}, n_{\max}\right\} + C_{\mathrm{DtN1}}. \tag{10.4}$$

LEMMA 10.3. *Assuming that the Galerkin solution u_h to the variational problem (2.9) exists, if (10.1) holds, then*

$$\left\|u - u_h\right\|_{H^1_k(\Omega_R)} \leq \Big(\mathcal{C}_2 hk + \mathcal{C}_3 hk^2 \eta(\mathcal{H}_h)\Big) \left\|u\right\|_{H^1_k(\Omega_R)}, \tag{10.5}$$

where

$$\mathcal{C}_2 := \frac{\sqrt{2} C_{\mathrm{cont}} C_{\mathrm{int}} C_{\mathrm{osc}}}{A_{\min}} \quad \text{and} \quad \mathcal{C}_3 := \frac{4 C_{\mathrm{cont}\star} C_{\mathrm{int}} C_{\mathrm{osc}} \sqrt{n_{\max} + A_{\min}}}{\sqrt{A_{\min}}}.$$

Proof. Since DtN_k satisfies the inequality (3.4), and A and n satisfy the inequalities (2.1) and (2.2), $a(\cdot,\cdot)$ (2.6) satisfies the Gårding inequality

$$\Re a(v,v) \geq A_{\min} \left\|v\right\|^2_{H^1_k(\Omega_R)} - k^2 (n_{\max} + A_{\min}) \left\|v\right\|^2_{L^2(\Omega_R)}. \tag{10.6}$$

Using Galerkin orthogonality (2.10) and continuity of $a(\cdot,\cdot)$ (10.4), we find that that (5.1) holds for any $v_h \in \mathcal{H}_h$. Using first the inequality (5.4) with $\alpha = \|u - u_h\|_{H^1_k(\Omega_R)}$, $\beta = C_{\mathrm{cont}} \|u - v_h\|_{H^1_k(\Omega_R)}$, $\varepsilon = A_{\min}$, and then Lemma 10.1, we find that if (10.1) holds, then, for any $v_h \in \mathcal{H}_h$,

$$\begin{aligned} \frac{A_{\min}}{2} \left\|u - u_h\right\|^2_{H^1_k(\Omega_R)} &\leq \frac{(C_{\mathrm{cont}})^2}{2A_{\min}} \left\|u - v_h\right\|^2_{L^2(\Omega_R)} + k^2 \big(n_{\max} + A_{\min}\big) \left\|u - u_h\right\|^2_{L^2(\Omega_R)}, \\ &\leq \left[\frac{(C_{\mathrm{cont}})^2}{2A_{\min}} + 4k^2 \big(n_{\max} + A_{\min}\big) (C_{\mathrm{cont}\star})^2 \big(\eta(\mathcal{H}_h)\big)^2\right] \left\|u - v_h\right\|^2_{H^1_k(\Omega_R)}, \end{aligned} \tag{10.7}$$

By the consequence (3.11) of the definition of C_{int} and the bound (3.6),

$$\left\|u - I_h u\right\|_{H^1_k(\Omega_R)} \leq \sqrt{2} h C_{\mathrm{int}} |u|_{H^2(\Omega_R)} \leq \sqrt{2} hk C_{\mathrm{int}} C_{\mathrm{osc}} \left\|u\right\|_{H^1_k(\Omega_R)}. \tag{10.8}$$

Choosing $v_h = I_h u$ in (10.7), using (10.8), taking the square root and using the inequality $\sqrt{a^2 + b^2} \leq a + b$ for all $a, b > 0$, we find the result (10.5). ∎

Proof. [Proof of Theorem 4.1] Under the assumption that the Galerkin solution u_h exists, the fact that the bound (4.2) holds under the condition (4.1) follows from combining Lemma 10.3 with the bound (8.5) on η. To prove that u_h exists under the condition (4.1), recall that, since the variational problem (2.9) is equivalent to a linear system of equations in a finite-dimensional space, existence of a solution follows from uniqueness. Suppose that there exists a $\widetilde{u}_h \in \mathcal{H}_h$ such that $a(\widetilde{u}_h, v_h) = 0$ for all $v_h \in \mathcal{H}_h$; to prove uniqueness, we need to show that $\widetilde{u}_h = 0$. Let $\widetilde{u}$ be such that $a(\widetilde{u}, v) = 0$ for all $v \in \mathcal{H}$, so that $\widetilde{u}_h$ is the Galerkin approximation to $\widetilde{u}$. Repeating the argument in the first part of the proof we see that the condition (4.1) holds then the bound (4.2) holds (with u replaced by $\widetilde{u}$ and u_h replaced by $\widetilde{u}_h$). By Lemma 2.4, $\widetilde{u} = 0$, so (4.2) implies that $\widetilde{u}_h = 0$ and the proof is complete. ∎

Proof. [Proof of Theorem 4.2] This is very similar to the proof of Theorem 4.1, except that we use the bound (8.6) on $\eta(\mathcal{H}_h)$ instead of (8.5). ∎

Acknowledgements. The authors thank Théophile Chaumont-Frelet (INRIA, Nice), Ivan Graham (University of Bath), and particularly Owen Pembery (University of Bath) for useful discussions. DL and EAS acknowledge support from EPSRC grant EP/1025995/1. JW was partly supported by Simons Foundation grant 631302.

REFERENCES

[1] A. K. Aziz, R. B. Kellogg, and A. B. Stephens, *A two point boundary value problem with a rapidly oscillating solution*, Numer. Math., 53 (1988), pp. 107–121.
[2] I. M. Babuška and S. A. Sauter, *Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?*, SIAM Review, (2000), pp. 451–484.
[3] L. Banjai and S. Sauter, *A refined Galerkin error and stability analysis for highly indefinite variational problems*, SIAM J. Numer. Anal., 45 (2007), pp. 37–53.
[4] H. Barucq, T. Chaumont-Frelet, and C. Gout, *Stability analysis of heterogeneous Helmholtz problems and finite element solution based on propagation media approximation*, Mathematics of Computation, 86 (2017), pp. 2129–2157.
[5] A. Bayliss, C. I. Goldstein, and E. Turkel, *On accuracy conditions for the numerical computation of waves*, Journal of Computational Physics, 59 (1985), pp. 396–404.
[6] C. Bernardi, *Optimal finite-element interpolation on curved domains*, SIAM Journal on Numerical Analysis, 26 (1989), pp. 1212–1240.
[7] T. Betcke, S. N. Chandler-Wilde, I. G. Graham, S. Langdon, and M. Lindner, *Condition number estimates for combined potential boundary integral operators in acoustics and their boundary element discretisation*, Numer. Methods Partial Differential Eq., 27 (2011), pp. 31–69.
[8] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 15 of Texts in Applied Mathematics, Springer, 3rd ed., 2008.
[9] A. Buffa and S. Sauter, *On the acoustic single layer potential: stabilization and Fourier analysis*, SIAM Journal on Scientific Computing, 28 (2006), pp. 1974–1999.
[10] N. Burq, *Mesures semi-classiques et mesures de défaut*, Astérisque, 245 (1997), pp. 167–195.
[11] ———, *Semi-classical estimates for the resolvent in nontrapping geometries*, International Mathematics Research Notices, 2002 (2002), pp. 221–241.
[12] F. Cardoso and G. Vodev, *Uniform estimates of the resolvent of the Laplace-Beltrami operator on infinite volume Riemannian manifolds. II*, Annales Henri Poincaré, 3 (2002), pp. 673–691.
[13] T. Chaumont-Frelet and S. Nicaise, *High-frequency behaviour of corner singularities in Helmholtz problems*, ESAIM: Math. Model. Numer. Anal., 52 (2018), pp. 1803–1845.
[14] ———, *Wavenumber explicit convergence analysis for finite element discretizations of general wave propagation problem*, IMA J. Numer. Anal., 40 (2020), p. 15031543.
[15] T. Chaumont-Frelet, S. Nicaise, and J. Tomezyk, *Uniform a priori estimates for elliptic problems with impedance boundary conditions*, Communications on Pure & Applied Analysis, 19 (2020), p. 2445.
[16] P. G. Ciarlet, *Basic error estimates for elliptic problems*, in Handbook of numerical analysis, Vol. II, North-Holland, Amsterdam, 1991, pp. 17–351.
[17] M. Costabel, M. Dauge, and S. Nicaise, *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.
[18] G. C. Diwan, A. Moiola, and E. A. Spence, *Can coercive formulations lead to fast and accurate solution of the Helmholtz equation?*, J. Comp. Appl. Math., 352 (2019), pp. 110–131.
[19] Y. Du and H. Wu, *Preasymptotic error analysis of higher order FEM and CIP-FEM for Helmholtz equation with high wave number*, SIAM J. Numer. Anal., 53 (2015), pp. 782–804.
[20] X. Feng and H. Wu, *Discontinuous Galerkin methods for the Helmholtz equation with large wave number*, SIAM J. Numer. Anal., 47 (2009), pp. 2872–2896.
[21] ———, *hp-Discontinuous Galerkin methods for the Helmholtz equation with large wave number*, Math. Comp., 80 (2011), pp. 1997–2024.
[22] J. Galkowski, E. H. Müller, and E. A. Spence, *Wavenumber-explicit analysis for the Helmholtz h-BEM: error estimates and iteration counts for the Dirichlet problem*, Numer. Math., 142 (2019), pp. 329–357.

[23] J. Galkowski, E. A. Spence, and J. Wunsch, *Optimal constants in nontrapping resolvent estimates*, Pure and Applied Analysis, 2 (2020), pp. 157–202.
[24] D. Gallistl, T. Chaumont-Frelet, S. Nicaise, and J. Tomezyk, *Wavenumber explicit convergence analysis for finite element discretizations of time-harmonic wave propagation problems with perfectly matched layers*, hal preprint 01887267, (2018).
[25] I. G. Graham, M. Löhndorf, J. M. Melenk, and E. A. Spence, *When is the error in the h-BEM for solving the Helmholtz equation bounded independently of k?*, BIT Numer. Math., 55 (2015), pp. 171–214.
[26] I. G. Graham, O. R. Pembery, and E. A. Spence, *The Helmholtz equation in heterogeneous media: a priori bounds, well-posedness, and resonances*, Journal of Differential Equations, 266 (2019), pp. 2869–2923.
[27] I. G. Graham and S. A. Sauter, *Stability and finite element error analysis for the Helmholtz equation with variable coefficients*, Mathematics of Computation, 89 (2020), pp. 105–138.
[28] P. Grisvard, *Elliptic problems in nonsmooth domains*, Pitman, Boston, 1985.
[29] L. Hörmander, *The analysis of linear partial differential operators. III*, Classics in Mathematics, Springer, Berlin, 2007. Pseudo-differential operators, Reprint of the 1994 edition.
[30] F. Ihlenburg, *Finite element analysis of acoustic scattering*, Springer Verlag, 1998.
[31] F. Ihlenburg and I. Babuška, *Finite element solution of the Helmholtz equation with high wave number Part I: The h-version of the FEM*, Comput. Math. Appl., 30 (1995), pp. 9–37.
[32] F. Ihlenburg and I. Babuska, *Finite element solution of the Helmholtz equation with high wave number part II: the hp version of the FEM*, SIAM J. Numer. Anal., 34 (1997), pp. 315–358.
[33] D. Lafontaine, E. A. Spence, and J. Wunsch, *For most frequencies, strong trapping has a weak effect in frequency-domain scattering*, Communications on Pure and Applied Mathematics, to appear (2020).
[34] Y. Li and H. Wu, *FEM and CIP-FEM for Helmholtz Equation with High Wave Number and Perfectly Matched Layer Truncation*, SIAM J. Numer. Anal., 57 (2019), pp. 96–126.
[35] W. McLean, *Strongly elliptic systems and boundary integral equations*, Cambridge University Press, 2000.
[36] J. M. Melenk, *On generalized finite element methods*, PhD thesis, The University of Maryland, 1995.
[37] J. M. Melenk and S. Sauter, *Convergence analysis for finite element discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary conditions*, Math. Comp, 79 (2010), pp. 1871–1914.
[38] ———, *Wavenumber explicit convergence analysis for Galerkin discretizations of the Helmholtz equation*, SIAM J. Numer. Anal., 49 (2011), pp. 1210–1243.
[39] L. Miller, *Refraction of high-frequency waves density by sharp interfaces and semiclassical measures at the boundary*, J. Math. Pures Appl. (9), 79 (2000), pp. 227–269.
[40] O. R. Pembery, *The Helmholtz Equation in Heterogeneous and Random Media: Analysis and Numerics*, PhD thesis, University of Bath, 2020.
[41] S. A. Sauter, *A refined finite element convergence theory for highly indefinite Helmholtz problems*, Computing, 78 (2006), pp. 101–115.
[42] A. H. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Mathematics of Computation, 28 (1974), pp. 959–962.
[43] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Mathematics of Computation, 54 (1990), pp. 483–493.
[44] E. A. Spence, *Overview of Variational Formulations for Linear Elliptic PDEs*, in Unified transform method for boundary value problems: applications and advances, A. S. Fokas and B. Pelloni, eds., SIAM, 2015, pp. 93–159.
[45] A. Toselli and O. Widlund, *Domain Decomposition Methods: Algorithms and Theory*, Springer, 2005.
[46] H. Wu, *Pre-asymptotic error analysis of CIP-FEM and FEM for the Helmholtz equation with high wave number. Part I: linear version*, IMA J. Numer. Anal., 34 (2014), pp. 1266–1288.
[47] H. Wu and J. Zou, *Finite element method and its analysis for a nonlinear Helmholtz equation with high wave numbers*, SIAM Journal on Numerical Analysis, 56 (2018), pp. 1338–1359.
[48] L. Zhu and H. Wu, *Preasymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part II: hp version*, SIAM J. Numer. Anal., 51 (2013), pp. 1828–1852.
[49] M. Zworski, *Semiclassical analysis*, American Mathematical Society, Providence, RI, 2012.