"When all else fails, integrate by parts" – an overview of new and old variational formulations for linear elliptic PDEs

E. A. Spence*

December 31, 2014

Abstract

We give an overview of variational formulations of second-order linear elliptic PDEs that are based on integration by parts (or, equivalently, on Green's identities).

Keywords: variational formulation, Green's identities, Helmholtz equation, finite element method, Discontinuous Galerkin methods, Trefftz-Discontinuous Galerkin methods, Ultra-Weak Variational Formulation, Fokas transform method, null-field method, boundary integral equation.

1 Introduction

Although the motto "when all else fails, integrate by parts" is applicable in a wide range of situations, the author first heard it in the context of the finite element method (FEM). The motto is relevant here since this method is based on the weak form of the PDE that one is trying to solve, and the weak form is obtained by multiplying the PDE by a test function and integrating by parts.

The weak form of a PDE is an example of a *variational problem*: given a Hilbert space $\mathcal{H}$, a sesquilinear form $a(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ and a continuous anti-linear functional $F : \mathcal{H} \to \mathbb{C}$,

$$\text{find } u \in \mathcal{H} \text{ such that } a(u,v) = F(v) \text{ for all } v \in \mathcal{H}. \tag{1.1}$$

We highlight immediately that the term "variational problem" is also used to describe the problem of minimising a functional (as in the Calculus of Variations). We see the link between these two notions of variational problem in §5 below (in Lemma 5.15), but we emphasise that this article is only concerned with problems of the form (1.1).

There are several different ways of converting a boundary value problem (BVP) for a linear PDE into a variational problem of the form (1.1), and many of them are based on integration by parts; this brings us to the first goal of the paper.

Goal 1: To give an overview of variational formulations for second-order linear elliptic PDEs based on multiplying by a test function and integrating by parts (or, equivalently, based on Green's identities).

We restrict attention to the particular second-order linear elliptic PDE

$$\Delta u + \lambda u = -f, \tag{1.2}$$

where $\lambda \in \mathbb{R}$, although many of the ideas that we describe also apply to general second-order linear elliptic PDEs, higher-order elliptic PDEs, and other linear PDEs. When $\lambda = 0$, (1.2) is Poisson's equation (and when $f = 0$ it is Laplace's equation), when $\lambda > 0$ it is the Helmholtz equation, and when $\lambda < 0$ it is often called the modified Helmholtz equation. We concentrate on the Helmholtz equation, since out of these three equations it currently commands the most attention from the research community.

The variational formulations that we discuss are

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

- the standard variational formulation, i.e. the weak form of the PDE (which is the basis of the FEM),
- *Discontinuous Galerkin (DG) methods* and *Trefftz-Discontinuous Galerkin (TDG) methods*, including the *Ultra-Weak Variational Formulation (UWVF)*,
- a variational formulation based on a quadratic functional introduced by Després in [Des97],
- *boundary integral equations* (which are the basis of the boundary element method (BEM)), and
- the *null-field method*.

This paper appears in a collection of articles about the so-called "unified transform method" or "Fokas transform method" introduced by Fokas in 1997 [Fok97] and developed further by Fokas and collaborators since then (see the monograph [Fok08] and the review papers [FS12] and [DTV14]). This method was introduced in the context of certain nonlinear PDES called *integrable* PDEs (with the defining property that they possess a so-called *Lax pair* formulation), but the method is also applicable to linear PDEs; this brings us to the second goal of the paper.

Goal 2: To show how the Fokas transform method applied to second-order linear elliptic PDEs can be placed into the framework established in Goal 1.

The heart of this paper is the "map" shown in Figure 1. To understand the map, note that $\mathcal{L} := \Delta + \lambda$ and recall that Green's first identity (G1) is

$$\int_D \overline{v}\mathcal{L}u = \int_{\partial D} \overline{v}\,\frac{\partial u}{\partial n} - \int_D \left(\nabla u \cdot \overline{\nabla v} - \lambda u \overline{v}\right), \tag{1.3}$$

and Green's second identity (G2) is

$$\int_D \left(\overline{v}\mathcal{L}u - u\overline{\mathcal{L}v}\right) = \int_{\partial D} \left[\overline{v}\,\frac{\partial u}{\partial n} - u\overline{\frac{\partial v}{\partial n}}\right]. \tag{1.4}$$

G1 therefore corresponds to multiplying the differential operator by (the complex conjugate of) a test function v and integrating by parts once (i.e. moving one derivative from u onto v), and G2 corresponds to multiplying by a test function and integrating by parts twice (i.e. moving two derivatives onto v).

To make the integration by parts completely explicit, recall that the divergence theorem

$$\int_D \nabla \cdot \mathbf{F} = \int_{\partial D} \mathbf{F} \cdot \mathbf{n} \tag{1.5}$$

applied with $\mathbf{F} = \phi\mathbf{G}$ gives the integration by parts formula

$$\int_D \phi\, \nabla \cdot \mathbf{G} = \int_{\partial D} \phi\, \mathbf{G} \cdot \mathbf{n} - \int_D \mathbf{G} \cdot \nabla\phi, \tag{1.6}$$

where $\mathbf{n}$ is the outward-pointing unit normal vector to D. Letting $\phi = \overline{v}$ and $\mathbf{G} = \nabla u$ in (1.6), we obtain G1 (1.3) (with the $\lambda u\overline{v}$ term on each side removed).

The methods in the dashed box in Figure 1 require the notion of a triangulation $\mathcal{T}$ of Ω (where Ω is the domain in which the PDE is posed), i.e. $\overline{\Omega}$ is divided into a finite number of subsets (not necessarily triangles) satisfying certain conditions; see Definition 2.1 below.

Finally, observe that Figure 1 concerns the homogeneous PDE $\mathcal{L}u = 0$. This is because, whereas methods based on G1 can be used to solve BVPs involving the inhomogeneous PDE $\mathcal{L}u = -f$, methods based on G2 are limited (either by definition or in practice) to BVPs involving the homogeneous PDE $\mathcal{L}u = 0$.

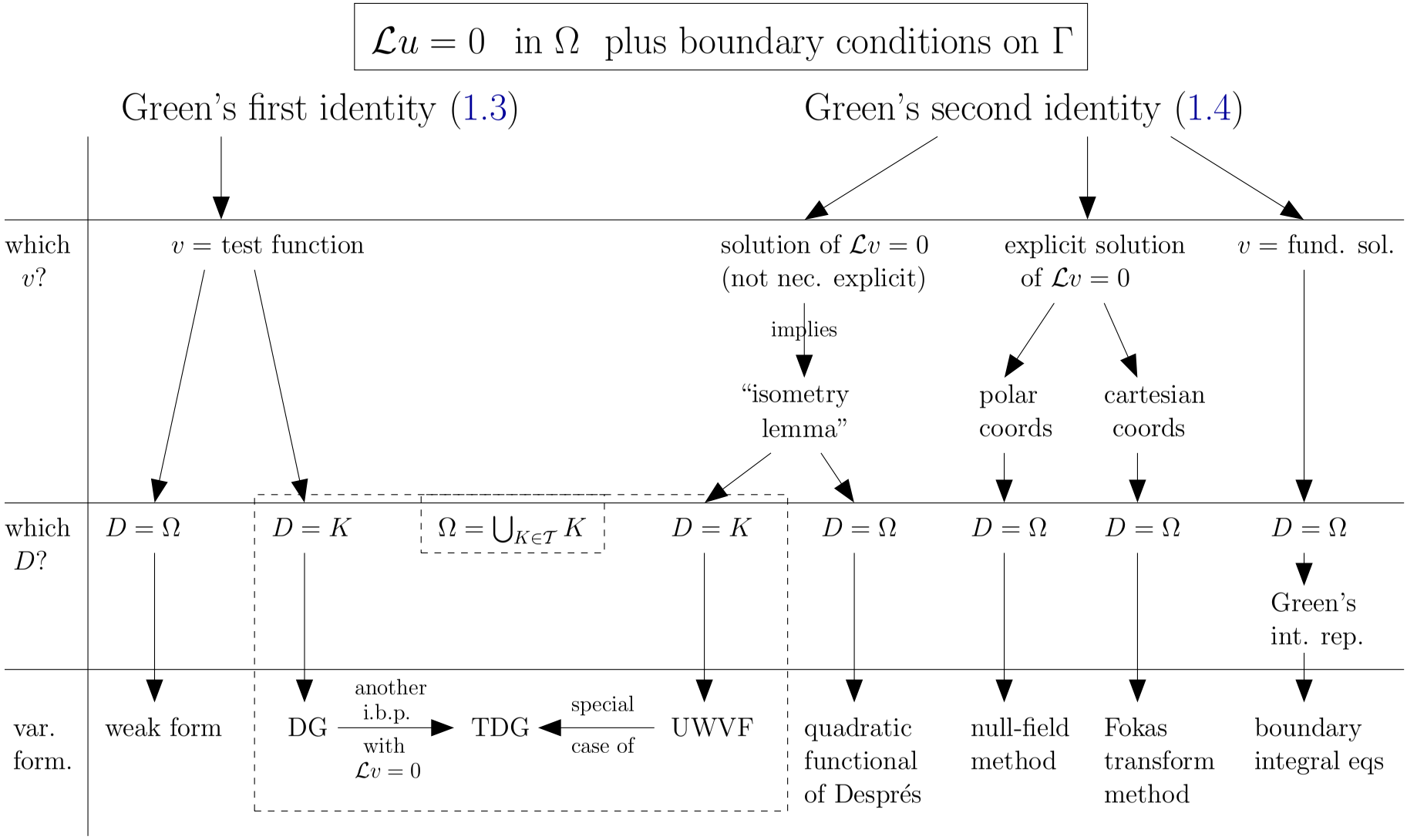

Figure 1: "Map" of the variational formulations discussed in this paper

Outline of the paper. In §2 we define some notation and recap some standard results (mainly concerning function spaces). In §3 we define some BVPs for the operator $\mathcal{L} := \Delta + \lambda$.

In §4 we derive Green's first and second identities and then discuss whether the BVPs introduced in §3 are self-adjoint; we do this because BVPs for the Helmholtz equation in exterior domains are not self-adjoint (because of the Sommerfeld radiation condition) and this has important implications for the variational formulations discussed in §10.

In §5 we recap standard results about variational problems. When one encounters a variational problem of the form (1.1) the following two questions naturally arise:

Q1. Does the variational problem (1.1) have a solution and, if so, is the solution unique?

Q2. Can we obtain a bound on the solution u in terms of F?

Furthermore, the variational problem (1.1) is the basis of the Galerkin method. Indeed, given a finite-dimensional subspace of $\mathcal{H}$, $\mathcal{H}_N$, the Galerkin method is

$$\text{find } u_N \in \mathcal{H}_N \text{ such that } a(u_N, v_N) = F(v_N) \text{ for all } v_N \in \mathcal{H}_N. \tag{1.7}$$

This leads us to a third question,

Q3. Do the Galerkin equations (1.7) have a unique solution (after imposing that $N \geq N_0$, for some N_0, if necessary)? If so, is the solution *quasi-optimal*, i.e. does there exists a $C_{qo} > 0$ such that

$$\|u - u_N\|_{\mathcal{H}} \leq C_{qo} \min_{v_N \in \mathcal{H}_N} \|u - v_N\|_{\mathcal{H}} \tag{1.8}$$

(again, after imposing that $N \geq N_0$ if necessary)? [1]

The results outlined in §5 focus on the key properties of *continuity*, *coercivity*, and *coercivity up to a compact perturbation*, since using these concepts we can give conditions under which the answers to the questions Q1, Q2, and Q3 are all "yes".

In §6–§10 we describe the variational formulations shown in Figure 1. Our goals in these sections are to

1. derive the variational formulations, and
2. discuss, and sometimes prove, whether the formulations are continuous, coercive, or coercive up to a compact perturbation (and thus establish which of the results in §5 are applicable).

We go through the variational formulations more or less in the order that they appear in Figure 1 (reading left to right). The only exception is that we discuss Green's integral representation and boundary integral equations before the null-field method and the Fokas transform method (this is because we use some integral-equation results in the section on the null-field method). Note that we do *not* discuss the relative merits of the different variational formulations (since this would make the paper substantially longer).

In §11 we conclude by discussing the two goals of the paper in a wider context.

Contents

1 Introduction **1**

2 Notation and background results. **5**

3 Boundary value problems for the operator $\mathcal{L} := \Delta + \lambda$ **7**

[1] Strictly speaking, quasi-optimality is a property of a *sequence* of Galerkin solutions, $(u_N)_{N\in\mathbb{Z}}$, corresponding to a sequence of finite-dimensional subspaces $(\mathcal{H}_N)_{N\in\mathbb{Z}}$ (since for a single u_N one can always find a C_{qo} such that (1.8) holds, and thus we are really interested in obtaining a C_{qo} such that (1.8) holds for all sufficiently large N). Nevertheless, when (1.8) holds we will be slightly cavalier and describe both the Galerkin solution u_N and the Galerkin method itself as quasi-optimal.

4 Green's identities and self-adjointness **9**
4.1 Green's identities . . . 9
4.2 Self-adjointness . . . 11

5 Recap of variational problems **12**
5.1 The inf-sup condition . . . 14
5.2 Coercivity . . . 16
5.3 Coercivity up to a compact perturbation . . . 17
5.4 Advantages of coercivity over the inf-sup condition and coercivity up to a compact perturbation . . . 20

6 Standard variational formulation (i.e. the weak form) **21**
6.1 The interior Dirichlet problem . . . 21
6.2 The interior impedance problem for the Helmholtz equation . . . 23

7 Discontinuous Galerkin (DG) formulation and Trefftz-Discontinuous Galerkin (TDG) formulation **26**
7.1 The DG formulation . . . 26
7.2 The Trefftz-DG formulation . . . 28

8 The UWVF and the quadratic functional of Després **29**
8.1 The isometry lemma . . . 29
8.2 The Ultra-Weak Variational Formulation (UWVF) . . . 29
8.3 The quadratic functional of Després . . . 32

9 Green's integral representation and boundary integral equations (BIEs) **34**
9.1 Green's integral representation . . . 34
9.2 Boundary integral equations (BIEs) . . . 36

10 The null-field method and the Fokas transform method **40**
10.1 Interior Dirichlet problem for the modified Helmholtz equation . . . 40
10.2 The Fokas transform method for the IDP for the modified Helmholtz equation . . . 42
10.3 Interior Dirichlet problem for the Helmholtz equation . . . 45
10.4 Exterior Dirichlet problem for the Helmholtz equation . . . 46
10.5 The null-field method for the Helmholtz exterior Dirichlet problem . . . 47
10.6 The method of Aziz, Dorr, and Kellogg for the Helmholtz exterior Dirichlet problem 50

11 Concluding remarks **51**
11.1 Variational formulations based identities other than Green's identities. . . . 52
11.2 Variational formulations not based on any identities. . . . 53
11.3 From Green to Lax. . . . 53

A note on style and rigor. When writing this paper, I tried to strike a balance between keeping everything mathematically honest and not getting bogged down in technicalities. The outcome is that the results are all stated rigorously, but I hope that even a reader who is not fully comfortable with Sobolev spaces (such as, perhaps, a graduate student) will still be able to understand the ideas behind the variational formulations.

For the reader trying to see the "forest from the trees" regarding the function spaces needed in the variational formulations: since all the variational formulations discussed in the paper arise from Green's identities, the definitions of the function spaces are all consequences of the conditions on u and v needed for G1 (1.3) to hold, and these are given in Lemma 4.1 (see also Remark 4.4).

2 Notation and background results.

Differential operators. We defined above $\mathcal{L} := \Delta + \lambda$ for $\lambda \in \mathbb{R}$. It will be useful later to have specific notation for this operator when we restrict λ to be either > 0 or < 0. Therefore, for

$k, \mu > 0$, we define

$$\mathcal{L}_k := \Delta + k^2 \quad \text{and} \quad \mathcal{L}_\mu := \Delta - \mu^2$$

(so $\mathcal{L}_k$ is the operator in the Helmholtz equation and $\mathcal{L}_\mu$ is the operator in the modified Helmholtz equation).

Notation for domains. We use the word "domain" to mean a connected open set, and we let D denote a generic bounded Lipschitz domain (see, e.g., [McL00, Definition 3.28] for the definition of Lipschitz) [2].

In this paper we consider BVPs in both bounded and unbounded Lipschitz domains (and we call these *interior* and *exterior* BVPs respectively). We use Ω to denote the domain in which the PDE is posed in interior BVPs, and Ω_+ to denote the domain in exterior BVPs.

Notation for interior BVPs. Let $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) be a bounded Lipschitz domain. Let $\Gamma := \partial\Omega$ and let $\mathbf{n}$ be the outward-pointing unit normal vector to Ω. As usual,

$$L^2(\Omega) := \left\{ v : \Omega \to \mathbb{C} \,:\, v \text{ is Lebesgue-measurable and } \int_\Omega |v|^2 < \infty \right\}.$$

For $v : \Omega \to \mathbb{C}$, let $\partial^\alpha v$ denote the weak derivative of v with multi-index α. Let

$$H^m(\Omega) := \big\{ v \in L^2(\Omega) \,:\, \partial^\alpha v \text{ exists and is in } L^2(\Omega) \text{ for all } |\alpha| \le m \big\},$$

and

$$H^1(\Omega, \Delta) := \big\{ v \in H^1(\Omega) \,:\, \Delta v \text{ exists in a weak sense and is in } L^2(\Omega) \big\}.$$

For the definition of the Sobolev spaces $H^s(\Gamma)$ for $|s| \le 1$ see, e.g., [McL00, Page 96], [CWGLS12, §A.3].

For $u \in C^\infty(\overline{\Omega}) := \{v|_{\overline{\Omega}} : v \in C^\infty(\mathbb{R}^d)\}$ we define the trace of u, γu, by $\gamma u = u|_\Gamma$. Recall that this operator extends to a bounded linear operator from $H^1(\Omega)$ to $H^{1/2}(\Gamma)$, i.e. there exists a $C > 0$ such that

$$\|\gamma v\|_{H^{1/2}(\Gamma)} \le C \, \|v\|_{H^1(\Omega)} \qquad \text{for all } v \in H^1(\Omega); \tag{2.1}$$

see, e.g., [McL00, Theorem 3.37]. We also have the multiplicative trace inequality, i.e. there exists a $C_1 > 0$ such that

$$\|\gamma v\|^2_{L^2(\Gamma)} \le C_1 \, \|v\|_{L^2(\Omega)} \, \|v\|_{H^1(\Omega)} \qquad \text{for all } v \in H^1(\Omega); \tag{2.2}$$

see [Gri85, Theorem 1.5.1.10, last formula on Page 41], [BS00, Theorem 1.6.6].

Let $\partial_n u$ denote the normal-derivative trace on Γ. Recall that if $u \in H^2(\Omega)$ then $\partial_n u := \mathbf{n} \cdot \gamma(\nabla u)$ and for $u \in H^1(\Omega, \Delta)$, $\partial_n u$ is defined as an element of $H^{-1/2}(\Gamma)$ so that Green's first identity holds; see Lemma 4.1 below. We sometimes call γu the *Dirichlet trace* of u, and $\partial_n u$ the *Neumann trace* of u.

Notation for exterior BVPs. Let Ω_- be a bounded Lipschitz open set such that the open complement $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$ is connected (this condition rules out Ω_- being, e.g., an annulus). Let $\Gamma := \partial\Omega_-$, and let $\mathbf{n}$ denote the outward-pointing unit normal vector to Ω_- (thus $\mathbf{n}$ points *into* Ω_+). Given $R > \sup_{\mathbf{x} \in \Omega_-} |\mathbf{x}|$, let $B_R := \{\mathbf{x} : |\mathbf{x}| < R\}$, $\Omega_R := \Omega_+ \cap B_R$, and $\Gamma_R := \partial B_R = \{\mathbf{x} : |\mathbf{x}| = R\}$.

Let

$$H^1_{\text{loc}}(\Omega_+) := \big\{ v : \chi v \in H^1(\Omega_+) \text{ for every } \chi \in C^\infty_{\text{comp}}(\overline{\Omega_+}) \big\},$$

where $C^\infty_{\text{comp}}(\overline{\Omega_+}) := \{v|_{\overline{\Omega_+}} : v \in C^\infty_{\text{comp}}(\mathbb{R}^d)\}$. Let

$$H^1_{\text{loc}}(\Omega_+, \Delta) := \big\{ v : v \in H^1_{\text{loc}}(\Omega_+) \text{ and } \Delta v|_G \in L^2(G) \text{ for every bounded measurable set } G \subset \Omega_+ \big\}.$$

Let $\gamma_\pm$ denote the exterior and interior traces from $\Omega_\pm$ to Γ, which satisfy $\gamma_+ : H^1_{\text{loc}}(\Omega_+) \to H^{1/2}(\Gamma)$ and $\gamma_- : H^1(\Omega_-) \to H^{1/2}(\Gamma)$. Let $\partial_n^\pm$ denote the exterior and interior normal-derivative traces, which satisfy $\partial_n^+ : H^1_{\text{loc}}(\Omega_+, \Delta) \to H^{-1/2}(\Gamma)$ and $\partial_n^- : H^1(\Omega_-, \Delta) \to H^{-1/2}(\Gamma)$.

[2] Note that some authors, including [McL00], allow both connected *and* disconnected sets in the definition of "domain".

Inequalities. We have the inequality

$$2ab \leq \frac{a^2}{\varepsilon} + \varepsilon b^2 \quad \text{for } a, b, \varepsilon > 0, \tag{2.3}$$

and its consequence

$$(a+b) \leq \sqrt{2}\sqrt{a^2+b^2}. \tag{2.4}$$

We recall the Cauchy-Schwarz inequality

$$\left| \int_D u\, \overline{v} \right| \leq \|u\|_{L^2(D)}\, \|v\|_{L^2(D)} \quad \text{for all } u, v \in L^2(D) \tag{2.5}$$

(we use this inequality with D equal either Ω or Γ).

Miscellaneous. We write $a \lesssim b$ if $a \leq Cb$ for some $C > 0$ that is independent of all parameters of interest (these will usually be k and μ, but also sometimes h). We write $a \gtrsim b$ if $b \lesssim a$, and $a \sim b$ if both $a \gtrsim b$ and $a \lesssim b$.

We only display the surface and volume measures in integrals when it might be ambiguous as to which variable the integration is respect to (this turns out only to be in §9 and §10.5).

As one can see from Figure 1, some of the variational formulations require the notion of a *triangulation* of Ω (also called a *mesh* or *partition*).

Definition 2.1 (Triangulation) *Following [Cia91, Page 61], we say that a finite collection of sets $\mathcal{T}$ is a triangulation of Ω if the following properties hold.*

1. $\overline{\Omega} = \bigcup_{K \in \mathcal{T}} K$
2. *Each $K \in \mathcal{T}$ is closed and its interior, $\mathring{K}$, is non-empty and connected.*
3. *If $K_1, K_2 \in \mathcal{T}$ and $K_1 \neq K_2$ then $\mathring{K}_1 \cap \mathring{K}_2 = \emptyset$.*
4. *Each $K \in \mathcal{T}$ is Lipschitz.*

One then defines $h_K := \operatorname{diam}(K) = \max_{\mathbf{x},\mathbf{y} \in \overline{K}} |\mathbf{x}-\mathbf{y}|$ and $h := \max_{K \in \mathcal{T}} h_K$. One usually thinks of a family of triangulations $\mathcal{T}_h$, with $0 < h \leq h_0$ for some h_0. Let ρ_K be the diameter of the largest ball contained in K (so $\rho_K \leq h_K$). The family $\mathcal{T}_h$ is *regular* (or *non-degenerate*) if $h_K \lesssim \rho_K$ for all $K \in \mathcal{T}_h$ and for all $0 < h \leq h_0$.

3 Boundary value problems for the operator $\mathcal{L} := \Delta + \lambda$

Definition 3.1 (Interior Dirichlet problem (IDP)) *Let Ω be a bounded Lipschitz domain. Given $g_D \in H^{1/2}(\Gamma)$ and $f \in L^2(\Omega)$, we say that $u \in H^1(\Omega)$ satisfies the* interior Dirichlet problem *(IDP) if*

$$\mathcal{L}u = -f \quad \text{in } \Omega \quad \text{and} \quad \gamma u = g_D \quad \text{on } \Gamma. \tag{3.1}$$

Some remarks on the IDP:

- The most general situation is where $f \in (H^1(\Omega))^*$ (where $(H^1(\Omega))^*$ is the anti-dual space to $H^1(\Omega)$) instead of $f \in L^2(\Omega)$.
- The PDE in (3.1) is understood as holding in a distributional sense, i.e.

$$\int_\Omega u\, \mathcal{L}\phi = -\int_\Omega f\, \phi \quad \text{for all } \phi \in C^\infty_{\text{comp}}(\Omega).$$

- The uniqueness of the IDP depends on λ. Indeed, if $\lambda \leq 0$ then the solution is unique (this can be proved using Green's first identity). On the other hand, if $\lambda = \lambda_j$, where λ_j is the jth Dirichlet eigenvalue of the negative Laplacian in Ω, i.e. there exists a $u_j \in H^1(\Omega) \setminus \{0\}$ such that

$$-\Delta u_j = \lambda_j u_j \quad \text{in } \Omega \quad \text{and} \quad \gamma u_j = 0 \quad \text{on } \Gamma, \tag{3.2}$$

  then the solution of the IDP is not unique.

The only exterior problem that we consider is the exterior Dirichlet problem for the Helmholtz equation.

Definition 3.2 (Exterior Dirichlet problem (EDP) for Helmholtz) *Let Ω_- be a bounded Lipschitz open set such that the open complement $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$ is connected. Given $g_D \in H^{1/2}(\Gamma)$ and $f \in L^2(\Omega_+)$ with compact support, we say that $u \in H^1_{\mathrm{loc}}(\Omega_+)$ satisfies the* exterior Dirichlet problem *(EDP) for the Helmholtz equation if*

$$\mathcal{L}_k u := \Delta u + k^2 u = -f \quad \text{in } \Omega_+, \tag{3.3}$$

$\gamma_+ u = g_D$ on Γ, and u satisfies the Sommerfeld radiation condition

$$\frac{\partial u}{\partial r}(\mathbf{x}) - \mathrm{i}ku(\mathbf{x}) = o\left(\frac{1}{r^{(d-1)/2}}\right) \tag{3.4}$$

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

Remark 3.3 (How should (3.3) and (3.4) be understood?) *As in the case of the IDP, the PDE (3.3) is understood in a distributional sense. To impose the pointwise condition (3.4), we need u to be in $C^1(\mathbb{R}^d \setminus \Omega_R)$ for some $R > 0$. This is ensured, however, by interior regularity of the Helmholtz operator, which implies that u is C^∞ outside the support of f and away from Γ (see Remark 3.6 below).*

An alternative way of defining the EDP is to fix R such that $\operatorname{supp} f \subset B_R$ and impose the PDE (3.3) in Ω_R in its weak form. The sesquilinear form of that variational formulation involves an operator on Γ_R that (roughly speaking) encodes the information that u satisfies the Sommerfeld radiation condition outside Ω_R. One can then show that the solution of this variational problem has an extension from Ω_R to Ω_+ that satisfies the EDP defined by Definition 3.2; for more details see, e.g., the discussion in [Spe14, §4.3] (this discussion is for the exterior impedance problem, however the case of the EDP is almost identical).

Further remarks on the EDP:

- The solution of the EDP for the Helmholtz equation is unique for all k; see [CK83, Theorem 3.13] [CWGLS12, Corollary 2.9].

The majority of practical applications of the Helmholtz equation involve posing the PDE in exterior domains. Although there are many standard ways of dealing numerically with both unbounded domains and the radiation condition, most investigations of numerical methods for solving the Helmholtz equation begin by considering the *interior impedance problem*. This problem has the advantage that it is posed on a bounded domain, but (unlike the IDP) the solution is unique for all k (as it is for the EDP); see Theorem 3.5 below.

Definition 3.4 (Interior impedance problem (IIP) for Helmholtz) *Let Ω be a bounded Lipschitz domain. Given $g \in L^2(\Gamma)$, $f \in L^2(\Omega)$, and $\eta \in \mathbb{R} \setminus \{0\}$, we say that $u \in H^1(\Omega)$ satisfies the* interior impedance problem *(IIP) if*

$$\mathcal{L}_k u := \Delta u + k^2 u = -f \quad \text{in } \Omega \tag{3.5}$$

and

$$\partial_n u - \mathrm{i}\eta\gamma u = g \quad \text{on } \Gamma. \tag{3.6}$$

Some remarks on the IIP:

- We could consider the more general situation where $f \in (H^1(\Omega))^*$ and $g \in H^{-1/2}(\Gamma)$.
- The PDE (3.5) is understood in a distributional sense and the boundary condition (3.6) is understood as saying that $\partial_n u$ (as an element of $H^{-1/2}(\Gamma)$) equals the L^2-function $g + \mathrm{i}\eta\gamma u$.

- One often chooses $\eta = k$; this is because the impedance boundary condition is often used as an approximation to the radiation condition. For example, an approximation of the EDP is: given $g_D \in H^{1/2}(\Gamma)$ and $f \in L^2(\Omega_+)$ with compact support, choose R such that $\mathrm{supp} f \subset B_R$ and find $u \in H^1(\Omega_R)$ such that $\mathcal{L}_k u = -f$ in Ω_R, $\gamma_+ u = g_D$ on Γ and

$$\partial_n u - \mathrm{i}ku = 0 \quad \text{on } \Gamma_R.$$

Theorem 3.5 (Uniqueness of the Helmholtz IIP) *The solution of the Helmholtz IIP is unique.*

Proof. This proof uses some results from later sections, namely Green's first identity (Lemma 4.1) and Green's integral representation (Theorem 9.1).

We need to show that if $\mathcal{L}_k u = 0$ in Ω and $\partial_n u - \mathrm{i}\eta\gamma u = 0$ on Γ, then $u = 0$ in Ω. Since $u \in H^1(\Omega, \Delta)$, Green's first identity ((4.4) below) with $v = u$ and $D = \Omega$ implies that

$$\int_\Gamma \overline{\gamma u}\, \partial_n u - \int_\Omega \left(|\nabla u|^2 - k^2 |u|^2\right) = 0.$$

Using the boundary condition, we obtain

$$\mathrm{i}\eta \int_\Gamma |\gamma u|^2 - \int_\Omega \left(|\nabla u|^2 - k^2 |u|^2\right) = 0. \tag{3.7}$$

If $\eta \in \mathbb{R} \setminus \{0\}$, the imaginary part of (3.7) implies that $\gamma u = 0$, and then the boundary condition implies that $\partial_n u = 0$. Green's integral representation (Theorem 9.1) implies that if $\gamma u = 0$ and $\partial_n u = 0$ then $u = 0$ in Ω. ∎

This theorem can also be understood as saying that, under the impedance boundary condition $\partial_n u - \mathrm{i}\eta\gamma u = 0$ with $\eta \in \mathbb{R} \setminus \{0\}$, the eigenvalues of the negative Laplacian are complex.

Remark 3.6 (Interior regularity for solutions of $\mathcal{L}u = -f$) *The PDE $\mathcal{L}u = -f$ states that the Laplacian of u equals a linear combination of u and f, and so if $f \in C^m$, say, we might expect u to be in C^{m+2}. The rigorous version of this observation is the following: if $f \in H^m(\Omega)$ and $\mathcal{L}u = -f$ then $u \in H^{m+2}(\widetilde{\Omega})$ for any $\widetilde{\Omega}$ that is compactly contained in Ω (i.e. there exists a compact set Ω' such that $\widetilde{\Omega} \subset \Omega' \subset \Omega$); see, e.g., [Eva98, §6.3.1, Theorem 2]. The Sobolev imbedding theorem can then be used to prove that (i) if $f \in L^2(\Omega)$ then $u \in C(\Omega)$ [Eva98, §5.6.3, Theorem 6], and (ii) if $f \in C^\infty(\Omega)$ then $u \in C^\infty(\Omega)$ [Eva98, §6.3.1, Theorem 3].*

4 Green's identities and self-adjointness

4.1 Green's identities

Informal discussion. We multiply $\mathcal{L}u$ by the complex conjugate of a test function v (where by "test function" we mean a function with, as yet, unspecified properties) and move either one or two derivatives onto v to obtain

$$\overline{v}\mathcal{L}u = \nabla \cdot \left[\overline{v}\nabla u\right] - \nabla u \cdot \overline{\nabla v} + \lambda u \overline{v} \tag{4.1}$$

and

$$\overline{v}\mathcal{L}u - u\overline{\mathcal{L}v} = \nabla \cdot \left[\overline{v}\nabla u - u\overline{\nabla v}\right] \tag{4.2}$$

respectively. (Note that one can obtain (4.2) from two copies of (4.1), with the roles of u and v swapped in the second copy.) We then integrate (4.1) and (4.2) over a domain D (assuming u is sufficiently smooth on D) and use the divergence theorem (1.5).

In the rest of the paper, we will be a bit cavalier and refer to both (4.1) and its integrated form as "G1", and similarly we refer to both (4.2) and integrated form as "G2".

Statement of Green's first and second identities (G1 and G2 respectively).

Lemma 4.1 (Green's first identity (G1)) *Let D be a bounded Lipschitz domain. If $u \in H^1(D,\Delta)$ and $v \in H^1(D)$ then*

$$\int_D v\mathcal{L}u = \int_{\partial D} \gamma v\, \partial_n u - \int_D \big(\nabla u \cdot \nabla v - \lambda uv\big) \tag{4.3}$$

and

$$\int_D \overline{v}\mathcal{L}u = \int_{\partial D} \overline{\gamma v}\, \partial_n u - \int_D \big(\nabla u \cdot \overline{\nabla v} - \lambda u\overline{v}\big). \tag{4.4}$$

Lemma 4.2 (Green's second identity (G2)) *Let D be a bounded Lipschitz domain. If u and v are both in $H^1(D,\Delta)$ then*

$$\int_D \big(v\mathcal{L}u - u\mathcal{L}v\big) = \int_{\partial D} \big[v\partial_n u - u\partial_n v\big] \tag{4.5}$$

and

$$\int_D \big(\overline{v}\mathcal{L}u - u\overline{\mathcal{L}v}\big) = \int_{\partial D} \big[\overline{\gamma v}\, \partial_n u - u\overline{\partial_n v}\big] \tag{4.6}$$

Remark 4.3 (How should one understand the integrals over ∂D?) *By the results recapped in §2, when $u \in H^1(D,\Delta)$ and $v \in H^1(D)$, $\partial_n u \in H^{-1/2}(\partial D)$ and $\gamma v \in H^{1/2}(\partial D)$. The integral $\int_{\partial D} \overline{\gamma v}\, \partial_n u$ in (4.4) therefore does not make sense as a usual (Lebesgue) integral.*

Recall that there exists a continuous sesquilinear form (see Definition 5.1 below for the definition of a sesquilinear form) $\langle\cdot,\cdot\rangle_{\partial D} : H^{-s}(\partial D) \times H^s(\partial D) \to \mathbb{C}$ for $|s| \le 1$, such that

$$\langle \phi, \psi\rangle_{\partial D} = \int_{\partial D} \phi\, \overline{\psi} \tag{4.7}$$

when $\phi, \psi \in L^2(\partial D)$ (i.e. (4.7) defines $\langle\cdot,\cdot\rangle_{\partial D}$ for $s = 0$). The integral $\int_{\partial D} \overline{\gamma v}\, \partial_n u$ in (4.4) should be understood as $\langle \partial_n u, \gamma v\rangle_{\partial D}$. (Therefore, if $\partial_n u \in L^2(\partial D)$ then $\langle \partial_n u, \gamma v\rangle_{\partial D} = \int_{\partial D} \overline{\gamma v}\, \partial_n u$ in the usual sense.) The notation $\langle\cdot,\cdot\rangle_{\partial D}$ is used for this sesquilinear form since $H^{-s}(\partial D)$ can be understood as a realisation of the dual space of $H^s(\partial D)$, with $\langle\cdot,\cdot\rangle_{\partial D}$ then equivalent to the duality pairing $\langle\cdot,\cdot\rangle_{(H^s(\partial D))^* \times H^s(\partial D)}$; see [McL00, Pages 76 and 98], [CWGLS12, Page 279] for more details.

Proofs of Lemmas 4.1 and 4.2. It is sufficient to prove (4.3), since (4.4) follows from (4.3) by replacing v by $\overline{v}$, (4.5) follows from (4.3), and (4.6) follows from (4.5).

The divergence theorem (1.5) holds when $\mathbf{F} \in C^1(\overline{D})$ [McL00, Theorem 3.34], and then (4.3) holds for $u \in H^2(D)$ and $v \in H^1(D)$ by (i) the density of $C^k(\overline{D})$ in $H^k(D)$ for $k \in \mathbb{N}$ [McL00, Page 77], (ii) the boundedness of the trace operator from $H^1(D)$ to $H^{1/2}(\partial D)$, (iii) the fact that $\partial_n u = \mathbf{n}\cdot\gamma(\nabla u)$ for $u \in H^2(D)$, and (iv) the Cauchy-Schwarz inequality (2.5). The proof of how to lessen the condition $u \in H^2(D)$ to $u \in H^1(D,\Delta)$ is given in, e.g., [CWGLS12, Pages 280 and 281], [McL00, Lemma 4.3]; the key point is that the normal derivative $\partial_n u$ is *defined* for $u \in H^1(D,\Delta)$ so that (4.3) holds (with $\int_{\partial D} \partial_n u\, \overline{\gamma v}$ understood as $\langle \partial_n u, \gamma v\rangle_{\partial D}$ as discussed in Remark 4.3). ∎

Remark 4.4 (If you only remember one thing from this section...) *The key point to take from this section regarding function spaces is that if $u \in H^1(D)$ satisfies $\mathcal{L}u = -f$ (in a distributional sense) with $f \in L^2(D)$, then*

$$-\int_D f\,\overline{v} = \int_{\partial D} \overline{\gamma v}\, \partial_n u - \int_D \big(\nabla u \cdot \overline{\nabla v} - \lambda u\overline{v}\big) \tag{4.8}$$

for all $v \in H^1(D)$ (with the integral over ∂D understood as a duality pairing as discussed in Remark 4.3).

4.2 Self-adjointness

Definition 4.5 (Formal adjoint of a differential operator) *If $\mathcal{L}$ is a general linear differential operator of order p, then its* formal adjoint*, denoted by $\mathcal{L}^*$, is defined so that the identity*

$$\overline{v}\mathcal{L}u - u\overline{\mathcal{L}^*v} = \nabla \cdot \mathbf{J}(u,v) \tag{4.9}$$

holds with $\mathbf{J}(u,v)$ a sesquilinear form involving derivatives of u and v of order $p-1$ or less [Sta68, §5.7], [Kee95, §4.3.2], [Nai67, §5].

The identity (4.9) (which can be seen as a generalisation of G2 (4.2)) is often called *Lagrange's identity* or *Lagrange's formula*.

If $\mathcal{L}^* = \mathcal{L}$ then $\mathcal{L}$ is *formally self-adjoint*. The identity G2 (4.2) therefore shows that the operator $\mathcal{L} := \Delta + \lambda$ is formally self-adjoint when $\lambda \in \mathbb{R}$. Note that formal self-adjointness is a condition on the differential operator itself (i.e. *without* boundary conditions), whereas self-adjointness is a condition on the BVP, i.e. on the differential operator *with* boundary conditions.

We now introduce the notion of adjoint boundary conditions. To keep things simple, we give the definition only for the second-order operator $\mathcal{L} := \Delta + \lambda$. This assumption ensures that the adjoint boundary conditions consist of one condition on each part of the boundary, but the same idea generalises to higher-order equations (for which the adjoint boundary conditions can involve several conditions on each part of the boundary); see, e.g., [Nai67, §1.6], [Sta79, Chapter 3, §3]. Note that, up to now, D has always been a bounded Lipschitz domain, but in this subsection (and this subsection only) we allow D to be unbounded.

Definition 4.6 (Adjoint boundary conditions for BVPs involving $\mathcal{L} := \Delta + \lambda$) *Let D be a Lipschitz domain (either bounded or unbounded), with outward-pointing normal vector $\boldsymbol{\nu}$. Let $\mathcal{L} := \Delta + \lambda$ and let u satisfy the BVP*

$$\mathcal{L}u = f \ \text{ in } D \quad \text{ and } \quad \mathcal{B}u = 0 \ \text{ on } \partial D, \tag{4.10}$$

for some linear operator $\mathcal{B}$ involving the trace operators γ and ∂_ν. The adjoint boundary-condition *operator $\mathcal{B}^*$ is such that*

$$\int_D \overline{v}\mathcal{L}u - u\overline{\mathcal{L}^*v} = 0 \tag{4.11}$$

when $\mathcal{B}u = 0$ and $\mathcal{B}^*v = 0$.

When D is bounded the Lagrange identity (4.9) and the divergence theorem (1.5) imply that (4.11) is equivalent to

$$\int_{\partial D} \mathbf{J}(u,v) \cdot \boldsymbol{\nu} = 0,$$

but when D is unbounded things are more complicated (see Example 4.9 below).

Definition 4.7 (Self-adjoint) *A BVP is* self-adjoint *if the differential operator is formally self-adjoint (i.e. $\mathcal{L} = \mathcal{L}^*$) and the boundary-condition operator $\mathcal{B}$ equals the adjoint boundary-condition operator $\mathcal{B}^*$.*

Example 4.8 (Self-adjointness of the Helmholtz IDP) *In this case $D = \Omega$, $\boldsymbol{\nu} = \mathbf{n}$, $\mathcal{L}u = (\Delta + k^2)u$ and $\mathcal{B}u = \gamma u$. G2 (4.2) implies that $\mathcal{L}^* = \mathcal{L}$ and $\mathbf{J}(u,v) = \left[\overline{v}\nabla u - u\overline{\nabla v}\right]$, and thus*

$$\int_\Omega \overline{v}\mathcal{L}u - u\overline{\mathcal{L}^*v} = \int_{\partial\Omega} \mathbf{J}(u,v)\cdot\mathbf{n} = \int_{\partial\Omega} \left[\overline{\gamma v}\,\partial_n u - \gamma u\,\overline{\partial_n v}\right].$$

When $\mathcal{B}u = 0$, $\int_{\partial\Omega} \mathbf{J}(u,v)\cdot\mathbf{n} = \int_\Gamma \overline{\gamma v}\,\partial_n u$. The condition on v that causes this last integral to vanish is $\gamma v = 0$, and thus $\mathcal{B}^*v = \gamma v$. Since $\mathcal{L} = \mathcal{L}^*$ and $\mathcal{B} = \mathcal{B}^*$, the Helmholtz IDP is self-adjoint.

Example 4.9 (Non-self-adjointness of the Helmholtz EDP) *In this case $D = \Omega_+$, $\mathcal{L}u = (\Delta + k^2)u$ and the boundary condition operator $\mathcal{B}$ is understood as the trace operator on Γ and the Sommerfeld radiation condition (3.4) at infinity. G2 (4.2) again implies that $\mathcal{L}^* = \mathcal{L}$ and $\mathbf{J}(u,v) = \left[\overline{v}\nabla u - u\overline{\nabla v}\right]$.*

To find $\mathcal{B}^*$ from the condition (4.11), we observe that

$$\int_{\Omega_+} \overline{v}\mathcal{L}u - u\overline{\mathcal{L}^*v} = \lim_{R\to\infty} \int_{\Omega_R} \overline{v}\mathcal{L}u - u\overline{\mathcal{L}^*v} = \lim_{R\to\infty} \int_{\partial\Omega_R} \mathbf{J}(u,v)\cdot\boldsymbol{\nu},$$

where $\boldsymbol{\nu}$ is the outward-pointing unit normal vector to Ω_R (so $\boldsymbol{\nu} = -\mathbf{n}$ on Γ and $\boldsymbol{\nu} = \widehat{\mathbf{x}}$ on Γ_R). Therefore, from (4.9) and the divergence theorem (1.5),

$$\int_{\Omega_+} \overline{v}\mathcal{L}u - u\overline{\mathcal{L}^*v} = -\int_\Gamma \left[\overline{\gamma_+ v}\,\partial_n^+ u - \gamma_+ u\,\overline{\partial_n^+ v}\right] + \lim_{R\to\infty}\int_{\Gamma_R}\left[\overline{v}\,\frac{\partial u}{\partial r} - u\,\overline{\frac{\partial v}{\partial r}}\right]. \tag{4.12}$$

Just like in the case of the IDP, $\mathcal{B}^* = \gamma$ on Γ.

To determine $\mathcal{B}^*$ at infinity, we need to consider the integral over Γ_R in (4.12). More specifically, $\mathcal{B}^*v$ will consist of the conditions on v that ensure that

$$\lim_{R\to\infty}\int_{\Gamma_R}\left[\overline{v}\,\frac{\partial u}{\partial r} - u\,\overline{\frac{\partial v}{\partial r}}\right] = 0 \tag{4.13}$$

when u satisfies the Sommerfeld radiation condition (3.4). The radiation condition (3.4) implies that, when $\mathbf{x}\in\Gamma_R$,

$$\frac{\partial u}{\partial r}(\mathbf{x}) - \mathrm{i}ku(\mathbf{x}) = o\left(\frac{1}{R^{(d-1)/2}}\right), \quad u(\mathbf{x}) = \mathcal{O}\left(\frac{1}{R^{(d-1)/2}}\right), \quad \text{and} \quad \frac{\partial u}{\partial r}(\mathbf{x}) = \mathcal{O}\left(\frac{1}{R^{(d-1)/2}}\right) \tag{4.14}$$

as $R\to\infty$ (see [CK83, Theorem 3.6] for how the second and third conditions in (4.14) follow from the first). Since $\partial u/\partial r - \mathrm{i}ku$ is smaller than $\partial u/\partial r$ as $r\to\infty$, it makes sense to introduce $\partial u/\partial r - \mathrm{i}ku$ in the integral in (4.13) by adding and subtracting $-\mathrm{i}ku\overline{v}$,

$$\int_{\Gamma_R}\left[\overline{v}\frac{\partial u}{\partial r} - u\overline{\frac{\partial v}{\partial r}}\right] = \int_{\Gamma_R}\left[\overline{v}\left(\frac{\partial u}{\partial r} - \mathrm{i}ku\right) - u\overline{\left(\frac{\partial v}{\partial r} + \mathrm{i}kv\right)}\right]. \tag{4.15}$$

Recalling that $\int_{\Gamma_R} = \mathcal{O}(R^{d-1})$, we see that the right-hand side of (4.15) tends to zero as $R\to\infty$ if v satisfies the adjoint radiation condition

$$\frac{\partial v}{\partial r}(\mathbf{x}) + \mathrm{i}kv(\mathbf{x}) = o\left(\frac{1}{R^{(d-1)/2}}\right) \quad \text{and} \quad v(\mathbf{x}) = \mathcal{O}\left(\frac{1}{R^{(d-1)/2}}\right) \tag{4.16}$$

as $R\to\infty$. (In a similar way to how $u = \mathcal{O}(R^{-(d-1)/2})$ follows from $\partial u/\partial r - \mathrm{i}ku = o(R^{-(d-1)/2})$ when $\mathcal{L}u = 0$ [CK83, Theorem 3.6], the second condition in (4.16) follows from the first when $\mathcal{L}^*v = 0$.)

Therefore, $\mathcal{B}^*v = \gamma v$ on Γ and $\mathcal{B}^*$ equals the adjoint radiation condition (4.16) at infinity.

Note that we sometimes call (3.4) the *outgoing* radiation condition and (4.16) the *incoming* radiation condition (with this terminology assuming that we create a solution of the wave equation from a solution of the Helmholtz equation by multiplying by $\mathrm{e}^{-\mathrm{i}\omega t}$).

In §10 we need the following lemma, which can be proved using the calculations at the end of Example 4.9.

Lemma 4.10 *If u and v both satisfy the Sommerfeld radiation condition (3.4) then*

$$\lim_{R\to\infty}\int_{\Gamma_R}\left[v\frac{\partial u}{\partial r} - u\frac{\partial v}{\partial r}\right] = 0 \quad \text{and} \quad \lim_{R\to\infty}\int_{\Gamma_R}\left[\overline{v}\frac{\partial u}{\partial r} - u\overline{\frac{\partial v}{\partial r}}\right] = 2\mathrm{i}k\lim_{R\to\infty}\int_{\Gamma_R} u\,\overline{v} \neq 0.$$

5 Recap of variational problems

In this section we recap some of the standard theory of variational problems, focusing on how to obtain the following three key things (which correspond to the three questions Q1–Q3 in §1):

K1. Existence and uniqueness of a solution to the variational problem (1.1).

K2. A bound on the solution u in terms of F.

K3. Existence and uniqueness of a solution to the Galerkin equations (1.7) (when $N \geq N_0$, for some N_0, if necessary) and quasi-optimality of the Galerkin solution, i.e. the bound

$$\|u - u_N\|_{\mathcal{H}} \leq C_{qo} \min_{v_N \in \mathcal{H}_N} \|u - v_N\|_{\mathcal{H}} \tag{5.1}$$

for some $C_{qo} > 0$ (again, when $N \geq N_0$ if necessary).

This theory is contained in many texts; we follow [SS11] and use mostly the same notation.

The variational problem (1.1) concerns a sesquilinear form mapping $\mathcal{H} \times \mathcal{H} \to \mathbb{C}$. We now consider the more general situation of a sesquilinear form mapping $\mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are two (not necessarily equal) Hilbert spaces.

Definition 5.1 (Sesquilinear form) *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces over $\mathbb{C}$. The mapping $a(\cdot,\cdot) : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ is called a* sesquilinear form *if it is linear in its first argument and anti-linear in its second argument. That is, if $u_1, u_2 \in \mathcal{H}_1$, $v_1, v_2 \in \mathcal{H}_2$, and $\lambda_1, \lambda_2 \in \mathbb{C}$, then*

$$a(\lambda_1 u_1 + \lambda_2 u_2, v_1) = \lambda_1 a(u_1, v_1) + \lambda_2 a(u_2, v_1)$$

and

$$a(u_1, \lambda_1 v_1 + \lambda_2 v_2) = \overline{\lambda_1} a(u_1, v_1) + \overline{\lambda_2} a(u_1, v_2).$$

The appropriate variational problem for a sesquilinear form $a : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ is then, given a continuous anti-linear functional $F : \mathcal{H}_2 \to \mathbb{C}$,

$$\boxed{\text{find } u \in \mathcal{H}_1 \text{ such that } a(u,v) = F(v) \quad \text{for all } v \in \mathcal{H}_2.} \tag{5.2}$$

The Hilbert space $\mathcal{H}_1$ is often called the *trial space* and $\mathcal{H}_2$ is then called the *test space*.

Definition 5.2 (Bilinear form) *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces over $\mathbb{R}$. The mapping $a(\cdot,\cdot) : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{R}$ is called a* bilinear form *if it is linear in both arguments. That is, if $u_1, u_2 \in \mathcal{H}_1$, $v_1, v_2 \in \mathcal{H}_2$, and $\lambda_1, \lambda_2 \in \mathbb{R}$, then*

$$a(\lambda_1 u_1 + \lambda_2 u_2, v_1) = \lambda_1 a(u_1, v_1) + \lambda_2 a(u_2, v_1)$$

and

$$a(u_1, \lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 a(u_1, v_1) + \lambda_2 a(u_1, v_2).$$

In the rest of this section we focus only on sesquilinear forms (since the majority of variational problems in the rest of the paper involve these), but all the results below have analogues involving bilinear forms.

Definition 5.3 (Continuity of sesquilinear forms) *A sesquilinear form $a(\cdot,\cdot) : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ is* continuous *(or* bounded*) if there exists a $C_c < \infty$ such that*

$$\left|a(u,v)\right| \leq C_c \|u\|_{\mathcal{H}_1} \|v\|_{\mathcal{H}_2} \tag{5.3}$$

for all $u \in \mathcal{H}_1$ and $v \in \mathcal{H}_2$. We define the smallest C_c for which (5.3) holds to be the norm of $a(\cdot,\cdot)$ denoted by $\|a\|$.

The next lemma shows that we can identify a sesquilinear form mapping $\mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ with a linear operator from $\mathcal{H}_1 \to \mathcal{H}_2^*$, where $\mathcal{H}_2^*$ is the anti-dual space of $\mathcal{H}_2$ (i.e. the set of all continuous, anti-linear functionals on $\mathcal{H}_2$, see [SS11, §2.1.2 and Lemma 2.1.38]). Given $F \in \mathcal{H}_2^*$ we introduce the notation that

$$F(v) = \langle F, v \rangle_{\mathcal{H}_2^* \times \mathcal{H}_2}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}_2^* \times \mathcal{H}_2}$ is called a *duality pairing*. (Note that $\langle \cdot, \cdot \rangle_{\mathcal{H}_2^* \times \mathcal{H}_2}$ is then a sesquilinear form on $\mathcal{H}_2^* \times \mathcal{H}_2$.) We use $L(\mathcal{H}_1, \mathcal{H}_2^*)$ to denote the set of all continuous (i.e. bounded) linear operators from $\mathcal{H}_1$ to $\mathcal{H}_2^*$.

Lemma 5.4 (Sesquilinear forms ↔ linear operators) *For every continuous sesquilinear form $a : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ there exists a unique $A \in L(\mathcal{H}_1, \mathcal{H}_2^*)$ such that*

$$a(u, v) = \langle Au, v \rangle_{\mathcal{H}_2^* \times \mathcal{H}_2}$$

for all $u \in \mathcal{H}_1$, $v \in \mathcal{H}_2$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}_2^* \times \mathcal{H}_2}$ denotes the duality pairing between $\mathcal{H}_2^*$ and $\mathcal{H}_2$. Furthermore,

$$\|A\|_{\mathcal{H}_1 \to \mathcal{H}_2^*} \leq \|a\| .$$

Proof. See [SS11, Lemma 2.1.38]. ∎

In what follows we make no distinction between a sesquilinear form $a : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ and the associated operator $A : \mathcal{H}_1 \to \mathcal{H}_2^*$ (i.e. we state results in terms of whichever object is most natural).

Given $A \in L(\mathcal{H}_1, \mathcal{H}_2^*)$ we say that A is invertible if it is both injective and surjective (so there is an algebraic inverse A^{-1}). Observe that if $A^{-1} \in L(\mathcal{H}_2^*, \mathcal{H}_1)$ then K1 and K2 both hold, and vice versa.

Our goal in this section is to find conditions on $a(\cdot, \cdot)$ that ensure the properties K1-K3 above hold (to ensure K3 holds we might also need conditions on the finite-dimensional subspaces of the Galerkin method). We discuss three properties $a(\cdot, \cdot)$ can have that ensure K1-K3 hold (in various degrees). These three properties are

(a) $a : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ satisfies the *inf-sup condition*,

(b) $a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ is *coercive*,

(a) $a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ is *coercive up to a compact perturbation*.

Property (a) is the weakest, (b) is the strongest, and (c) is in the middle.

5.1 The inf-sup condition

Definition 5.5 (The inf-sup condition) *The sesquilinear form $a : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ satisfies the* inf-sup condition *if there exists a $\gamma > 0$ such that*

$$\inf_{u \in \mathcal{H}_1 \setminus \{0\}} \sup_{v \in \mathcal{H}_2 \setminus \{0\}} \frac{\left|a(u, v)\right|}{\|u\|_{\mathcal{H}_1} \|v\|_{\mathcal{H}_2}} \geq \gamma \tag{5.4a}$$

and

$$\sup_{u \in \mathcal{H}_1 \setminus \{0\}} \left|a(u, v)\right| > 0 \quad \text{for all } v \in \mathcal{H}_2 \setminus \{0\}. \tag{5.4b}$$

The following theorem is fundamental.

Theorem 5.6 (K1 and K2 under inf-sup condition) *The following are equivalent:*
(a) The sesquilinear form $a(\cdot, \cdot)$ satisfies the inf-sup condition (5.4).
(b) A is invertible and $A^{-1} \in L(\mathcal{H}_2^*, \mathcal{H}_1)$ with $\|A^{-1}\|_{\mathcal{H}_2^* \to \mathcal{H}_1} \leq 1/\gamma$; i.e. for each $F \in \mathcal{H}_2^*$, the variational problem (5.2) has a unique solution which satisfies

$$\|u\|_{\mathcal{H}_1} \leq \frac{1}{\gamma} \|F\|_{\mathcal{H}_2^*} . \tag{5.5}$$

Proof. See [SS11, Theorem 2.1.44 and Remark 2.1.46]. ∎

Remark 5.7 *In functional analysis texts one often encounters the result that $B \in L(H_1, H_2)$ is invertible (with bounded inverse) if (i) B is bounded below, i.e.*

$$\|Bu\|_{H_2} \geq \gamma \|u\|_{H_1} \quad \text{for all } u \in H_1, \tag{5.6}$$

and (ii) the range of B is dense in H_2; see, e.g., [AA02, Lemma 2.8]. The first part of the inf-sup condition (5.4a) is equivalent to $A \in L(\mathcal{H}_1, \mathcal{H}_2^*)$ being bounded below (i.e. (5.6) holds with B replaced by A, H_1 replaced by $\mathcal{H}_1$, and H_2 replaced by $\mathcal{H}_2^*$). The second part (5.4b) ensures that the range of A is dense in $\mathcal{H}_2^*$ (see [SS11, Part (iii) of the proof of Theorem 2.1.44]).

Before we state the next result (which gives conditions under which K3 holds), we need to define the discrete inf-sup condition. We first introduce some notation. For $i = 1, 2$ let $(\mathcal{H}_N^i)_{N\in\mathbb{Z}^+}$ be a sequence of finite-dimensional, nested subspaces of $\mathcal{H}_i$, whose union is dense in $\mathcal{H}_i$; i.e. for all $N \geq 1$,

$$\mathcal{H}_N^i \subset \mathcal{H}_{N+1}^i, \quad \dim \mathcal{H}_N^i < \infty, \quad \overline{\bigcup_{N\in\mathbb{Z}^+} \mathcal{H}_N^i} = \mathcal{H}_i. \tag{5.7}$$

We assume that $\dim \mathcal{H}_N^i = N$ for $i = 1, 2$.

The Galerkin equations for the variational problem (5.2) are then

$$\boxed{\text{find } u_N \in \mathcal{H}_N^1 \text{ such that } a(u_N, v_N) = F(v_N) \quad \text{for all } v_N \in \mathcal{H}_N^2.} \tag{5.8}$$

Note that

- Galerkin methods with different trial and test spaces, i.e. those of the form (5.8), are often called *Petrov-Galerkin* methods (with methods that use the same trial and test spaces, i.e. those of the form (1.7), then called *Bubnov-Galerkin* methods), and
- from the variational problem (5.2) and the Galerkin equations (5.8), one obtains the *Galerkin orthogonality* condition that

$$a(u - u_N, v_N) = 0 \quad \text{for all } v_N \in \mathcal{H}_N^2. \tag{5.9}$$

Definition 5.8 (The discrete inf-sup condition) *The sesquilinear form $a(\cdot,\cdot)$ and finite-dimensional subspaces $(\mathcal{H}_N^i)_{N\in\mathbb{Z}^+}$, $i = 1, 2$, satisfy the* discrete inf-sup condition *if, for each N, there exists a $\gamma_N > 0$ such that*

$$\inf_{u\in\mathcal{H}_N^1\setminus\{0\}} \sup_{v\in\mathcal{H}_N^2\setminus\{0\}} \frac{\left|a(u,v)\right|}{\|u\|_{\mathcal{H}_1}\|v\|_{\mathcal{H}_2}} \geq \gamma_N \tag{5.10a}$$

and

$$\sup_{u\in\mathcal{H}_N^1\setminus\{0\}} \left|a(u,v)\right| > 0 \quad \text{for all } v \in \mathcal{H}_N^2 \setminus \{0\}. \tag{5.10b}$$

Note that

- since $\mathcal{H}_N^1$ and $\mathcal{H}_N^2$ are finite dimensional, the infima and suprema in Definition 5.8 can be replaced by minima and maxima respectively, and
- (although it is not important for the rest of this paper) the second condition (5.10b) can be obtained from the first (5.10a) (recalling that $N = \dim \mathcal{H}_N^1 = \dim \mathcal{H}_N^2$); see, e.g., [Gra09, Page 16].

Theorem 5.9 (K3 under discrete inf-sup condition) *Assume that $a(\cdot,\cdot)$ is continuous and satisfies the inf-sup condition (5.4). Assume that there exists a N_0 such that $(\mathcal{H}_N^i)_{N\in\mathbb{Z}^+}$, $i = 1, 2$, satisfy the discrete inf-sup condition (5.10) when $N \geq N_0$. Then, for each $F \in \mathcal{H}_2^*$, the variational problem (5.2) has a unique solution, the Galerkin equations (5.8) have a unique solution $u_N \in \mathcal{H}_N^1$ when $N \geq N_0$, and*

$$\|u - u_N\|_{\mathcal{H}_1} \leq \left(\frac{\|a\|}{\gamma_N}\right) \min_{v_N\in\mathcal{H}_N^1} \|u - v_N\|_{\mathcal{H}_1} \tag{5.11}$$

when $N \geq N_0$.

Proof. This result is usually proved with the factor $(1 + \|a\|/\gamma_N)$ in front of the best approximation error in the quasi-optimality bound; see, e.g., [SS11, Theorem 4.2.1]. It was observed in [XZ03], however, that a result of Kato on projection operators [Kat60, Lemma 4] means that $(1 + \|a\|/\gamma_N)$ can be replaced by $(\|a\|/\gamma_N)$ (see [Dem06] for a good introduction to these results). Note that (5.11) implies that, if γ_N is independent of N, then $u_N \to u$ as $N \to \infty$. ∎

Having established that K1, K2, and K3 hold if the sesquilinear form satisfies the inf-sup condition (and the finite-dimensional spaces satisfy the discrete inf-sup condition), we now consider sesquilinear forms that correspond to *compact perturbations* of invertible operators and use Fredholm theory (see, e.g., [SS11, §2.1.3] for the definition of a compact operator).

Note for the next theorem that $Au = 0$ (as an element of $\mathcal{H}_2^*$) if and only if $a(u, v) = 0$ for all $v \in \mathcal{H}_2$.

Theorem 5.10 (K1 and K2 for invertible + compact) *Let $A \in L(\mathcal{H}_1, \mathcal{H}_2^*)$ be such that $A = B+T$ where $B \in L(\mathcal{H}_1, \mathcal{H}_2^*)$ is such that $B^{-1} \in L(\mathcal{H}_2^*, \mathcal{H}_1)$ (i.e. $b(\cdot, \cdot)$ satisfies the inf-sup condition) and $T \in L(\mathcal{H}_1, \mathcal{H}_2^*)$ is compact. If A is injective, i.e.*

$$Au = 0 \quad \text{implies that} \quad u = 0,$$

then A is invertible and $A^{-1} \in L(\mathcal{H}_2^*, \mathcal{H}_1)$ (i.e. $a(\cdot, \cdot)$ satisfies the inf-sup condition (5.4)).

Proof. This is proved using Fredholm theory in [SS11, Theorem 4.2.7] when $\mathcal{H}_1 = \mathcal{H}_2$, but the proof when $\mathcal{H}_1 \neq \mathcal{H}_2$ follows in the same way. (Note that this argument does not give us any information about the norm of A^{-1}, other than that it is finite.) ∎

Combining Theorems 5.10 and 5.9, we obtain the following corollary.

Corollary 5.11 (K3 for invertible + compact under the discrete inf-sup condition) *Let A satisfy the assumptions of Theorem 5.10 and assume that there exists an N_0 such that $a(\cdot, \cdot)$ and $(\mathcal{H}_N^i)_{N \in \mathbb{Z}^+}$, $i = 1, 2$, satisfy the discrete inf-sup condition (5.10) when $N \geq N_0$. Then, when $N \geq N_0$, the Galerkin equations (5.8) have a unique solution $u_N \in \mathcal{H}_N^1$ and the error estimate (5.11) holds.*

5.2 Coercivity

We now assume that $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$.

Definition 5.12 *The sesquilinear form $a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ is coercive if there exists an $\alpha > 0$ such that*

$$\left|a(v, v)\right| \geq \alpha \left\|v\right\|_{\mathcal{H}}^2 \quad \text{for all } v \in \mathcal{H}. \tag{5.12}$$

The property (5.12) is sometimes called "$\mathcal{H}$-ellipticity" (as in, e.g., [SS11, Page 39], [Ste08, §3.2], and [HW08, Definition 5.2.2]), with "coercivity" then used to mean *either* that $a(\cdot, \cdot)$ is a compact perturbation of a coercive operator (as in, e.g., [Ste08, §3.6] and [HW08, §5.2]) *or* that $a(\cdot, \cdot)$ satisfies a Gårding inequality (as in [SS11, Definition 2.1.54]).

Remark 5.13 (The numerical range and alternative definitions of coercivity) *Define the* numerical range *of A by*

$$W(A) := \left\{ \frac{a(v, v)}{\left\|v\right\|_{\mathcal{H}}^2} : v \in \mathcal{H} \setminus \{0\} \right\}. \tag{5.13}$$

This set is convex (see, e.g., [GR97, Theorem 1.1-2]) and so if $a(\cdot, \cdot)$ is coercive then there exists a $\theta \in [0, 2\pi)$ such that

$$\Re\big(\mathrm{e}^{\mathrm{i}\theta} a(v, v)\big) \geq \alpha \left\|v\right\|_{\mathcal{H}}^2 \quad \text{for all } v \in \mathcal{H}. \tag{5.14}$$

Although the convexity of the numerical range is well known, the implication (5.14) does not usually appear in the literature on variational problems. (Note that (5.14) is sometimes used as a definition of coercivity; see [SS11, Equation (2.43)].)

Theorem 5.14 (The Lax-Milgram theorem) *If $a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ is a continuous and coercive sesquilinear form, then $A^{-1} \in L(\mathcal{H}^*, \mathcal{H})$ with $\|A^{-1}\|_{\mathcal{H}^* \to \mathcal{H}} \leq 1/\alpha$; i.e., for each $F \in \mathcal{H}^*$, the variational problem (1.1) has a unique solution which satisfies*

$$\left\|u\right\|_{\mathcal{H}} \leq \frac{1}{\alpha} \left\|F\right\|_{\mathcal{H}^*}.$$

Proof. One can *either* show that $a(\cdot,\cdot)$ satisfies the inf-sup condition (5.4) with $\gamma = \alpha$ and then obtain the result from Theorem 5.6 (see, e.g., [SS11, Proof of Lemma 2.1.51]) *or* prove the result directly (see, e.g., [Eva98, §6.2.1], [McL00, Lemma 2.32], [HW08, Theorem 5.2.3], [Ste08, Theorem 3.4]). ∎

Observe that the Lax-Milgram theorem therefore gives K1 and K2 for continuous and coercive sesquilinear forms.

The following lemma shows the link with the other definition of a "variational problem" mentioned in §1 (the problem of minimising a functional).

Lemma 5.15 (The link between symmetric coercive bilinear forms and minimising functionals) *Assume that $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ (note the change from a sesquilinear form to a bilinear form) is continuous, coercive, and symmetric (i.e. $a(u,v) = a(v,u)$ for all $u, v \in \mathcal{H}$). Then, given $F \in \mathcal{H}'$, the unique solution of the variational problem (1.1) is also a solution of the problem of minimising the quadratic functional $J : \mathcal{H} \to \mathbb{R}$ defined by*

$$J(v) := \frac{1}{2}a(v,v) - F(v). \tag{5.15}$$

Conversely, if $u \in \mathcal{H}$ minimises (5.15) then u solves the variational problem (1.1).

Proof. See [SS11, Proposition 2.1.53]. ∎

The following result gives K3 when $a(\cdot,\cdot)$ is continuous and coercive.

Theorem 5.16 (Céa's lemma) *Let $a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ be continuous and coercive. If $\mathcal{H}_N$ is any finite-dimensional subspace of $\mathcal{H}$ then the Galerkin equations (1.7) have a unique solution and the quasi-optimality bound (1.8) holds with $C_{qo} = C_c/\alpha$.*

Proof. This is proved in [SS11, Proposition 4.1.25] for particular $a(\cdot,\cdot)$ and $\mathcal{H}$, although the proof also applies to the general case; see also [Ste08, Theorem 8.1], [Cia91, Theorem 13.1]. ∎

5.3 Coercivity up to a compact perturbation

We saw in §5.1 that Fredholm theory gives us K1 and K2 for compact perturbations of invertible operators (assuming that the resulting operator is injective), and that we can obtain K3 if the finite-dimensional subspaces satisfy the discrete inf-sup condition (usually under the assumption that N is sufficiently large).

We see below that for compact perturbations of *coercive* (as opposed to just invertible) operators we can obtain K3 (if N is large enough) *without* the finite-dimensional subspaces satisfying the discrete inf-sup condition.

Definition 5.17 (Coercivity up to a compact perturbation) *$A \in L(\mathcal{H}, \mathcal{H}^*)$ is a compact perturbation of a coercive operator if there exists a compact operator $T \in L(\mathcal{H}, \mathcal{H}^*)$ and $\beta > 0$ such that*

$$\left|\left\langle (A-T)v, v\right\rangle_{\mathcal{H}^*\times\mathcal{H}}\right| \geq \beta \|v\|_{\mathcal{H}}^2 \qquad \text{for all } v \in \mathcal{H}$$

(i.e. $A = B + T$ where B is coercive).

Theorem 5.18 (K1, K2, and K3 for coercive + compact) *Let $\mathcal{H}$ be a Hilbert space over $\mathbb{C}$ and let $(\mathcal{H}_N)_{N\in\mathbb{Z}^+}$ be a dense sequence of finite-dimensional nested subspaces (i.e. (5.7) holds). Let $A \in L(\mathcal{H}, \mathcal{H}^*)$ be a compact perturbation of a coercive operator and assume that A is injective. Then*

(i) A is invertible and $A^{-1} \in L(\mathcal{H}^*, \mathcal{H})$.

(ii) There exists an $N_0 > 0$ such that, for all $N \geq N_0$ and all $F \in \mathcal{H}^*$, the Galerkin equations (1.7) have a unique solution and there exists a C_{qo} (independent of N) such that the quasi-optimality bound (1.8) holds.

Proof. Part (i) follows from Theorem 5.10, since every coercive operator satisfies the inf-sup condition. See [SS11, Theorem 4.2.9] for the proof of (ii). This proof verifies that the discrete inf-sup condition holds (with γ_N independent of N for N sufficiently large) by a contradiction argument, and then uses Corollary 5.11. Note that the proof does not give an explicit value for γ_N, and so we do not get an explicit value for C_{qo}. ∎

A common situation in which A is a compact perturbation of a coercive operator is when $a(\cdot,\cdot)$ satisfies a *Gårding inequality*.

Definition 5.19 (Gårding inequality) *Let $\mathcal{H}$ and $\mathcal{V}$ be Hilbert spaces such that $\mathcal{H} \subset \mathcal{V}$ and the inclusion map is continuous. We say that $a(\cdot,\cdot)$ satisfies a Gårding inequality if there exist $\alpha > 0$ and $C_{\mathcal{V}} \in \mathbb{R}$ such that*

$$\Re a(v,v) \geq \alpha \|v\|_{\mathcal{H}}^2 - C_{\mathcal{V}} \|v\|_{\mathcal{V}}^2 \qquad \text{for all } v \in \mathcal{H}. \tag{5.16}$$

(Note that if $C_{\mathcal{V}} \leq 0$ then $a(\cdot,\cdot)$ is coercive, and so the definition is only really useful when $C_{\mathcal{V}} > 0$.)

The standard example is $\mathcal{H} = H^1(\Omega)$ (or a closed subspace of it such as $H_0^1(\Omega)$) and $\mathcal{V} = L^2(\Omega)$, and we encounter this situation in §6.

Theorem 5.20 (Gårding inequality $\implies$ coercive + compact) *Let $\mathcal{H}$ and $\mathcal{V}$ be as in Definition 5.19, and assume further that the inclusion map is compact. If $a(\cdot,\cdot)$ satisfies a Gårding inequality then $A : \mathcal{H} \to \mathcal{H}^*$ is a compact perturbation of a coercive operator.*

In the literature, the Gårding inequality (5.16) is usually only stated for the case where $\mathcal{H} = H^1(\Omega)$ or $H_0^1(\Omega)$ and $\mathcal{V} = L^2(\Omega)$ (exceptions are [McL00, Page 44] and [Néd01, Theorem 5.4.5] [3]). We have been unable to find an explicit proof of Theorem 5.20 in the general case, and so we give one here.

Proof of Theorem 5.20. Let $\imath$ denote the inclusion map from $\mathcal{H}$ to $\mathcal{V}$. Define $T_1 : \mathcal{V} \to \mathcal{H}^*$ by

$$\langle T_1 u, v\rangle_{\mathcal{H}^*\times\mathcal{H}} = (u, \imath v)_{\mathcal{V}},$$

where $(\cdot,\cdot)_{\mathcal{V}}$ denotes the inner product on $\mathcal{V}$. Then, by the Cauchy-Schwarz inequality and the boundedness of $\imath$,

$$\left|\langle T_1 u, v\rangle_{\mathcal{H}^*\times\mathcal{H}}\right| \leq \|u\|_{\mathcal{V}} \|\imath\|_{\mathcal{H}\to\mathcal{V}} \|v\|_{\mathcal{H}};$$

this implies that $\|T_1 u\|_{\mathcal{H}^*} \leq \|u\|_{\mathcal{V}} \|\imath\|_{\mathcal{H}\to\mathcal{V}}$ and thus $T_1 : \mathcal{V} \to \mathcal{H}^*$ is continuous.

Next define $T_2 = T_1 \circ \imath : \mathcal{H} \to \mathcal{H}^*$. Observe that T_2 is compact since it is the composition of a continuous operator with a compact operator (see [SS11, Lemma 2.1.29]).

Define $B : \mathcal{H} \to \mathcal{H}^*$ by

$$\langle Bu, v\rangle_{\mathcal{H}^*\times\mathcal{H}} := a(u,v) + C_{\mathcal{V}}(\imath u, \imath v)_{\mathcal{V}}.$$

Then, by (5.16), $\Re\langle Bu, v\rangle_{\mathcal{H}^*\times\mathcal{H}} \geq \alpha \|v\|_{\mathcal{H}}^2$, so B is coercive. Furthermore, $B = A + C_{\mathcal{V}} T_2$, since

$$\langle T_2 u, v\rangle_{\mathcal{H}^*\times\mathcal{H}} = \langle T_1 \imath u, v\rangle_{\mathcal{H}^*\times\mathcal{H}} = (\imath u, \imath v)_{\mathcal{V}},$$

and thus A is a compact perturbation of a coercive operator since T_2 is compact. ∎

If $a(\cdot,\cdot)$ is injective and satisfies a Gårding inequality (with $\mathcal{H}$ compactly contained in $\mathcal{V}$) then Theorems 5.20 and 5.18 give quasi-optimality of the Galerkin method (i.e. K3), but *without* explicit expressions for C_{qo} or N_0. The advantage of $a(\cdot,\cdot)$ satisfying a Gårding inequality (over just knowing that A is a compact perturbation of a coercive operator) is that we can obtain information about C_{qo} and N_0. We can do this by *either* verifying that the discrete inf-sup condition holds *or* proving quasi-optimality directly. It turns out that the second method gives stronger results, and so we only present this method here (but see Remark 5.24 for a comparison of the results of the two methods).

[3] In both these references $\mathcal{V}$ is used to denote the smaller space, and $\mathcal{H}$ is used to denote the larger one. Here we use the opposite notation, so that A remains an operator from $\mathcal{H}$ to $\mathcal{H}^*$ (as it has been up to now).

Theorem 5.21 (K3 for Gårding inequality) *Let $\mathcal{H}$ and $\mathcal{V}$ be as in Definition 5.19 and assume further that the inclusion map is compact. Let $a : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ be continuous, injective, and satisfy a Gårding inequality (so that by Theorems 5.18 and 5.20 the variational problem (1.1) has a unique solution).*

Given $f \in \mathcal{V}$, define $S^* f \in \mathcal{H}$ as the solution of the variational problem

$$a(v, S^* f) = (v, f)_{\mathcal{V}} \quad \text{for all } v \in \mathcal{H} \tag{5.17}$$

(observe that $(v, f)_{\mathcal{V}}$ is a linear functional on $\mathcal{H}$ since $\mathcal{H}$ is continuously embedded in $\mathcal{V}$). Let $(\mathcal{H}_N)_{N \in \mathbb{Z}^+}$ be a dense sequence of finite-dimensional, nested subspaces, and let

$$\eta(\mathcal{H}_N) := \sup_{f \in \mathcal{V}} \min_{v_N \in \mathcal{H}_N} \frac{\|S^* f - v_N\|_{\mathcal{H}}}{\|f\|_{\mathcal{V}}}. \tag{5.18}$$

If

$$\eta(\mathcal{H}_N) \leq \frac{1}{C_c} \sqrt{\frac{\alpha}{2C_{\mathcal{V}}}}, \tag{5.19}$$

then the Galerkin equations (1.7) have a unique solution which satisfies

$$\|u - u_N\|_{\mathcal{H}} \leq \frac{2C_c}{\alpha} \min_{v_N \in \mathcal{H}_N} \|u - v_N\|_{\mathcal{H}} \tag{5.20}$$

(i.e. quasi-optimality with $C_{qo} = 2C_c/\alpha$).

Note that the operator $S^* : \mathcal{V} \to \mathcal{H}$ in Theorem 5.21 is related to the adjoint of A (hence the reason for the * notation). In what follows we refer to the variational problem (5.17) as the *adjoint variational problem*, although it is really the adjoint problem with a certain class of right-hand sides (see, e.g., [SS11, Exercise 2.1.41] for more on adjoint operators).

Proof of Theorem 5.21. Since this argument usually appears in the literature only for the case when $\mathcal{H} = H^1(\Omega)$ and $\mathcal{V} = L^2(\Omega)$ (see, e.g., [BS00, §5.7], [EM12, Lemma 4.1]), we give the details for the general situation here.

We first assume that a solution u_N exists. The Gårding inequality (5.16) applied to $u - u_N$ implies that

$$\alpha \|u - u_N\|_{\mathcal{H}}^2 - C_{\mathcal{V}} \|u - u_N\|_{\mathcal{V}}^2 \leq \Re a(u - u_N, u - u_N).$$

By Galerkin orthogonality (5.9), the right-hand side of this inequality can be replaced by $\Re a(u - u_N, u - v_N)$ for any $v_N \in \mathcal{H}_N$. Using this fact along with continuity of $a(\cdot, \cdot)$, we find that

$$\alpha \|u - u_N\|_{\mathcal{H}}^2 - C_{\mathcal{V}} \|u - u_N\|_{\mathcal{V}}^2 \leq C_c \|u - u_N\|_{\mathcal{H}} \|u - v_N\|_{\mathcal{H}} \quad \text{for all } v_N \in \mathcal{H}_N.$$

Therefore, the quasi-optimality (5.20) follows if we can show that

$$\sqrt{C_{\mathcal{V}}} \|u - u_N\|_{\mathcal{V}} \leq \sqrt{\frac{\alpha}{2}} \|u - u_N\|_{\mathcal{H}}. \tag{5.21}$$

Now, by the definition of S^* (5.17), Galerkin orthogonality (5.9), and continuity,

$$\begin{aligned} \|u - u_N\|_{\mathcal{V}}^2 = a\big(u - u_N, S^*(u - u_N)\big) &= a\big(u - u_N, S^*(u - u_N) - v_N\big) \\ &\leq C_c \|u - u_N\|_{\mathcal{H}} \|S^*(u - u_N) - v_N\|_{\mathcal{H}} \end{aligned} \tag{5.22}$$

for any $v_N \in \mathcal{H}_N$. The definition of $\eta(\mathcal{H}_N)$ (5.18) implies that there exists a $w_N \in \mathcal{H}_N$ such that

$$\|S^*(u - u_N) - w_N\|_{\mathcal{H}} \leq \eta(\mathcal{H}_N) \|u - u_N\|_{\mathcal{V}}$$

and using this fact in (5.22) we obtain that

$$\|u - u_N\|_{\mathcal{V}} \leq C_c \eta(\mathcal{H}_N) \|u - u_N\|_{\mathcal{H}}. \tag{5.23}$$

Therefore, the condition (5.19) implies that (5.21), and thus also (5.20), holds.

We have so far assumed that u_N exists. Recall that an $N \times N$ matrix A is invertible if and only if A has full rank, which is the case if and only if the only solution of $A\mathbf{x} = 0$ is $\mathbf{x} = 0$. Therefore, to show that u_N exists, we only need to show that u_N is unique. Seeking a contradiction, suppose that there exists a $\widetilde{u}_N \in \mathcal{H}_N$ such that

$$a(\widetilde{u}_N, v_N) = 0 \quad \text{for all } v_N \in \mathcal{H}_N.$$

Let $\widetilde{u}$ be such that

$$a(\widetilde{u}, v) = 0 \quad \text{for all } v \in \mathcal{H}; \tag{5.24}$$

thus $\widetilde{u}_N$ is the Galerkin approximation to $\widetilde{u}$. Repeating the argument in the first part of the proof we see that if (5.19) holds then the quasi-optimality (5.20) holds (with u replaced by $\widetilde{u}$ and u_N replaced by $\widetilde{u}_N$). By assumption, the only solution to the variational problem (5.24) is $\widetilde{u} = 0$, and then (5.20) implies that $\widetilde{u}_N = 0$. We have therefore shown that the solution u_N exists under the condition (5.19) and the proof is complete. ∎

Remark 5.22 (The history of the argument in Theorem 5.21) *The argument that obtains (5.23) from (5.22) was first introduced in the coercive case by Nitsche in [Nit68] and Aubin in [Aub67, Theorem 3.1], and is thus often referred to as the "Aubin-Nitsche lemma" or the "Aubin-Nitsche duality argument" (see, e.g., [Cia91, Theorem 19.1]). Schatz [Sch74] was then the first to use this argument in conjunction with a Gårding inequality to prove quasi-optimality of the corresponding Galerkin method.*

Remark 5.23 (Can one obtain information about N_0 via the condition on $\eta(\mathcal{H}_N)$?) *The usefulness of Theorem 5.21 depends on how easy it is to estimate $\eta(\mathcal{H}_N)$ and thus determine a bound on the threshold N_0 for quasi-optimality to hold. The quantity $\eta(\mathcal{H}_N)$ measures how well the finite-dimensional subspaces $\mathcal{H}_N$ approximate S^*f; therefore, to estimate $\eta(\mathcal{H}_N)$ we need information about (i) the subspaces, and (ii) S^*f.*

Regarding (ii), for the standard variational formulations of BVPs involving the PDE $\mathcal{L}u = -f$ (discussed in §6), $\mathcal{H} = H^1(\Omega)$ or $H^1_0(\Omega)$ and $\mathcal{V} = L^2(\Omega)$. The solution of the adjoint variational problem (5.17) can then be shown to be the solution of the adjoint BVP with $f \in L^2(\Omega)$ and zero boundary conditions, and thus bounds on S^*f in terms of f can be obtained using PDE techniques. (We give the resulting bound on $\eta(\mathcal{H}_N)$ in the case of the Helmholtz IIP in Remark 6.7 below.) This argument relies on the fact that the anti-linear functionals in the standard variational formulations are already of the form $(f, v)_{\mathcal{V}}$ for some $f \in \mathcal{V}$. This is not always the case; see, e.g., the variational formulation of the Helmholtz IDP in §10.3. For this variational formulation it is not clear what BVP (if any) the solution of the adjoint variational problem (5.17) corresponds to, and thus it is not clear how to obtain a bound on $\eta(\mathcal{H}_N)$ in this case.

Remark 5.24 *As mentioned before Theorem 5.21, one can also use the Gårding inequality to verify that the discrete inf-sup condition (5.10) holds and obtain bounds on C_{qo} and N_0 this way. Using an argument that also involves the adjoint variational problem (5.17), one finds that $a(\cdot, \cdot)$ satisfies the first discrete inf-sup condition (5.10a) if*

$$\eta(\mathcal{H}_N) \leq \frac{\alpha}{2C_c C_{\mathcal{V}}} \tag{5.25}$$

(as with the argument in Theorem 5.21, the 2 in (5.25) can be replaced by any number > 1). Recall that (as noted under Definition 5.8) one can obtain (5.10b) from (5.10a), and thus under (5.25) both parts of the inf-sup condition hold.

The condition (5.25) is more restrictive than (5.19). Indeed, when $a(\cdot, \cdot)$ satisfies a Gårding inequality, the worst-case scenario is when $C_{\mathcal{V}}$ is large and α is small, and then $1/C_{\mathcal{V}}$ is smaller than $1/\sqrt{C_{\mathcal{V}}}$ and $\alpha/2$ is smaller than $\sqrt{\alpha/2}$.

5.4 Advantages of coercivity over the inf-sup condition and coercivity up to a compact perturbation

We noted above that, out of the inf-sup condition, coercivity, and coercivity up to a compact perturbation, coercivity is the strongest property. Comparing the results obtained in §5.1-5.3, we see that coercivity has the following two advantages over the other two conditions:

1. The constant in the bound on u in terms of F is known explicitly (if C_c and α are given explicitly).

2. We obtain quasi-optimality for *any* finite dimensional subspace (i.e. without any constraint on the subspace dimension) and with an explicit expression for C_{qo}.

There is a third advantage of coercivity over the other two conditions, and this concerns the solution of the Galerkin equations.

3. The matrix of the Galerkin method inherits analogous continuity and coercivity properties from $a(\cdot,\cdot)$. These allow one to prove results about the number of iterations required to solve the Galerkin equations using iterative methods. For symmetric, coercive sesquilinear forms one can use the conjugate gradient method, and then the well-known bounds on the number of iterations can be found in, e.g., [Gre97, Chapter 3]. For non-symmetric, coercive sesquilinear forms one must use more general iterative methods such as the generalised minimum residual method (GMRES), and in this case a result of Elman [Elm82], [EES83, Theorem 3.3] gives bounds on the number of iterations (see, e.g, [GGS15, §1.3] and [SKS15, §1.3] for alternative statements of this result and applications of it to Helmholtz problems).

6 Standard variational formulation (i.e. the weak form)

Summary (linking to Figure 1): G1 with v a test function and $D = \Omega$.

We consider the interior Dirichlet problem (IDP) (Definition 3.1) in each of the three cases $\lambda = 0$, $\lambda < 0$, and $\lambda > 0$, and the interior impedance problem (IIP) for the Helmholtz equation (Definition 3.4).

6.1 The interior Dirichlet problem

Since λ is real, we can take f and g_D to be real-valued. Indeed, if f and g_D are complex then the solution $u = u_1 + \mathrm{i}u_2$, where the real valued functions u_1 and u_2 are the solutions of the BVPs

$$\mathcal{L}u_1 = -\Re f \ \text{ on } \Omega \quad \text{and} \quad \gamma u_1 = \Re g \ \text{ on } \Gamma$$

and

$$\mathcal{L}u_2 = -\Im f \ \text{ on } \Omega \quad \text{and} \quad \gamma u_2 = \Im g \ \text{ on } \Gamma.$$

Having real-valued f and g_D means that we can restrict attention to spaces of real-valued functions.

Multiplying the PDE $\mathcal{L}u = -f$ by a (real) test function v and integrating by parts (i.e. using G1 (4.3) with $D = \Omega$), we obtain that

$$-\int_\Omega f\, v = \int_{\partial\Omega} \gamma v\, \partial_n u - \int_\Omega \big(\nabla u \cdot \nabla v - \lambda u v\big) \tag{6.1}$$

for all $v \in H^1(\Omega)$ (recall Remark 4.4).

We first consider the case $g_D = 0$; see Remark 6.3 below for the case $g_D \neq 0$. Since $\gamma u = 0$, it is natural to work in the Hilbert space $H^1_0(\Omega) := \{v \in H^1(\Omega) : \gamma v = 0\}$. If v in (6.1) is in $H^1_0(\Omega)$ then the integral over Γ vanishes, and (6.1) becomes the assertion that $a_D(u,v) = F(v)$, where

$$a_D(u,v) := \int_\Omega \nabla u \cdot \nabla v - \lambda u v \quad \text{and} \quad F(v) := \int_\Omega f v \tag{6.2}$$

(with the subscript D standing for "Dirichlet"). Note that $a_D(\cdot,\cdot)$ is a bilinear form on $H^1_0(\Omega) \times H^1_0(\Omega)$, and $F(\cdot)$ is a linear functional on $H^1_0(\Omega)$.

The standard variational formulation of the IDP is therefore, given $f \in L^2(\Omega)$,

$$\boxed{\text{find } \ u \in H^1_0(\Omega) \ \text{ such that } \ a_D(u,v) = F(v) \ \text{ for all } \ v \in H^1_0(\Omega).} \tag{6.3}$$

We now turn our attention to the continuity and coercivity properties of $a_D(\cdot,\cdot)$, and we consider the three cases $\lambda = 0$, $\lambda < 0$, and $\lambda > 0$ separately.

Poisson's equation ($\lambda = 0$) We impose the usual H^1–norm on the space $H^1_0(\Omega)$:

$$\|v\|^2_{H^1(\Omega)} := \|\nabla v\|^2_{L^2(\Omega)} + \|v\|^2_{L^2(\Omega)}\,. \tag{6.4}$$

By the Cauchy-Schwarz inequality (2.5), $a_D(\cdot,\cdot)$ is continuous with $C_c = 1$ in this norm.

Moving to the question of whether or not $a_D(\cdot,\cdot)$ is coercive, we recall the Poincaré inequality, namely that there exists a $C > 0$ such that

$$\|v\|_{L^2(\Omega)} \le C\,\|\nabla v\|_{L^2(\Omega)} \qquad \text{for all } v \in H^1_0(\Omega) \tag{6.5}$$

[Eva98, §5.6.1, Theorem 3], [BS00, §5.3]. This inequality implies that $a_D(\cdot,\cdot)$ is coercive with $\alpha = 1/(1+C^2)$, and existence and uniqueness of a solution to the variational problem (6.3) and continuous dependence of the solution on the data (i.e. properties K1 and K2 of §5) then follow from the Lax-Milgram theorem (Theorem 5.14). Existence and uniqueness of a quasi-optimal solution of the Galerkin equations (i.e. property K3) for any finite dimensional subspace then follows from Céa's lemma (Theorem 5.16).

Since $a_D(\cdot,\cdot)$ is also symmetric, Lemma 5.15 implies that the unique solution of the variational problem (6.3) minimises the functional

$$J(v) = \frac{1}{2}a_D(v,v) - F(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - fv;$$

this fact is called Dirichlet's principle [Eva98, §2.2.5, §8.1.2].

The modified Helmholtz equation ($\lambda < 0$) We now let $\lambda = -\mu^2$ for some $\mu > 0$ (so that $\mathcal{L} = \mathcal{L}_\mu$). It is then convenient to introduce the weighted norm

$$\|v\|^2_{1,\mu,\Omega} := \|\nabla v\|^2_{L^2(\Omega)} + \mu^2\,\|v\|^2_{L^2(\Omega)}\,, \tag{6.6}$$

which is equivalent to the usual H^1-norm (6.4) when $\mu > 0$.

We weight the norm in this way for two reasons. The first is that if v is a solution of $\mathcal{L}_\mu v = 0$ then we expect that $|\nabla v| \sim \mu|v|$ (this can be verified when v is a separable solution of $\mathcal{L}_\mu v = 0$ in cartesian or polar coordinates) and then the two terms on the right-hand side of (6.6) are comparable. The second reason is that the constant C_c for $a_D(\cdot,\cdot)$ is independent of μ in the norm $\|\cdot\|_{1,\mu,\Omega}$ (indeed, $C_c = 2$ in this case), but it is dependent on μ in the usual H^1-norm ($C_c = 2\mu^2$ in this case). To prove these facts about C_c, we use the Cauchy-Schwarz inequality (2.5) and the inequality (2.4).

Turning to coercivity we see that $a_D(v,v) = \|v\|^2_{1,\mu,\Omega}$, and thus $a_D(\cdot,\cdot)$ is coercive with $\alpha = 1$ in the norm $\|\cdot\|_{1,\mu,\Omega}$; properties K1 and K2 then follow from the Lax-Milgram theorem and property K3 follows from Céa's lemma.

The Helmholtz equation ($\lambda > 0$) We now let $\lambda = k^2$ for some $k > 0$ (so that $\mathcal{L} = \mathcal{L}_k$). For exactly the same reasons as in the case of the modified Helmholtz equation, we introduce the weighted norm

$$\|v\|^2_{1,k,\Omega} := \|\nabla v\|^2_{L^2(\Omega)} + k^2\,\|v\|^2_{L^2(\Omega)}\,, \tag{6.7}$$

and in this norm $a_D(\cdot,\cdot)$ is continuous with $C_c = 2$.

The question of whether or not $a_D(\cdot,\cdot)$ is coercive is covered by the following two lemmas.

Lemma 6.1 ($a_D(\cdot,\cdot)$ is coercive when k is sufficiently small) *Let C be the constant in the Poincaré inequality (6.5). If*

$$k \le \frac{1}{2C}$$

then $a_D(\cdot,\cdot)$ is coercive (in the norm $\|\cdot\|_{1,k,\Omega}$) with $\alpha = 1/2$.

Proof. The definition of $a_D(\cdot,\cdot)$ and the Poincaré inequality (6.5) imply that, for any $v \in H^1_0(\Omega)$,

$$\begin{aligned} a_D(v,v) &= \|\nabla v\|^2_{L^2(\Omega)} + k^2\,\|v\|^2_{L^2(\Omega)} - 2k^2\,\|v\|^2_{L^2(\Omega)} \\ &\ge \|\nabla v\|^2_{L^2(\Omega)} + k^2\,\|v\|^2_{L^2(\Omega)} - 2k^2C^2\,\|\nabla v\|^2_{L^2(\Omega)}\,, \end{aligned} \tag{6.8}$$

and the result follows. ∎

Lemma 6.2 ($a_D(\cdot,\cdot)$ is not coercive when k is sufficiently large) *Let λ_1 be the first Dirichlet eigenvalue of the negative Laplacian in Ω. If $k^2 \geq \lambda_1$ then there exists a $v \in H^1_0(\Omega)$ with $a_D(v,v) = 0$.*

Proof. If λ_j is an eigenvalue of the negative Laplacian with eigenfunction $u_j \in H^1_0(\Omega)$, i.e. (3.2) holds, then

$$\|\nabla u_j\|^2_{L^2(\Omega)} - \lambda_j \|u_j\|^2_{L^2(\Omega)} = 0 \tag{6.9}$$

from G1 (6.1). The definition of $a_D(\cdot,\cdot)$ then implies that

$$a_D(u_j,u_j) = \|\nabla u_j\|^2_{L^2(\Omega)} - \lambda_j \|u_j\|^2_{L^2(\Omega)} + (\lambda_j - k^2)\|u_j\|^2_{L^2(\Omega)} = (\lambda_j - k^2)\|u_j\|^2_{L^2(\Omega)}\,. \tag{6.10}$$

Therefore if $k^2 = \lambda_1$, then $a_D(u_1,u_1) = 0$.

We now need to show that if $k^2 > \lambda_1$ then there exists a $v \in H^1_0(\Omega)$ with $a_D(v,v) = 0$. The expression (6.10) implies that if $\lambda_1 < k^2 < \lambda_j$ then

$$a_D(u_1,u_1) < 0 < a_D(u_j,u_j). \tag{6.11}$$

Since the numerical range of an operator is convex (see Remark 5.13), (6.11) implies there exists a $v \in H^1_0(\Omega)$ such that $a_D(v,v) = 0$. ∎

The equation (6.8) shows that $a_D(\cdot,\cdot)$ satisfies a Gårding inequality (see Definition 5.19) with $\mathcal{H} = H^1_0(\Omega)$, $\mathcal{V} = L^2(\Omega)$, $\alpha = 1$, and $C_{\mathcal{V}} = 2k^2$. Since $H^1_0(\Omega)$ is compactly contained in $L^2(\Omega)$ (i.e. the inclusion map from $H^1_0(\Omega)$ to $L^2(\Omega)$ is continuous and compact, see, e.g., [Eva98, §5.7], [McL00, Theorem 3.27]), Theorem 5.20 shows that the operator associated with $a_D(\cdot,\cdot)$ is a compact perturbation of a coercive operator. Theorem 5.18 then implies that, if $k^2 \neq \lambda_j$ for every $j = 1,2,\ldots$, the variational problem (6.3) has a unique solution which depends continuously on f (i.e. properties K1 and K2 hold).

Remark 6.3 (The case $g_D \neq 0$) *We now show how to reduce the case $g_D \neq 0$ to the case $g_D = 0$. Given $g_D \in H^{1/2}(\Gamma)$, let $u_1 \in H^1(\Omega)$ be such that $\gamma u_1 = g_D$; the existence of such a u_1 is guaranteed by [McL00, Theorem 3.37]. With u the solution of the IDP, we define $u_0 \in H^1_0(\Omega)$ by $u_0 := u - u_1$. Then u_0 satisfies the BVP*

$$\mathcal{L}u_0 = -f - \mathcal{L}u_1 \quad \text{in } \Omega \qquad \text{and} \qquad \gamma u_0 = 0 \quad \text{on } \Gamma.$$

Multiplying the PDE by a test function $v \in H^1_0(\Omega)$ and integrating by parts, we find that the variational problem for u_0 is

$$\boxed{\text{find } u_0 \in H^1_0(\Omega) \text{ such that } a_D(u_0,v) = F(v) - a_D(u_1,v) \text{ for all } v \in H^1_0(\Omega),}$$

which is just the variational problem (6.3) with a different linear functional.

6.2 The interior impedance problem for the Helmholtz equation

Since the impedance boundary condition (3.6) contains the complex number i, the solution of the IIP is (in general) complex-valued, and we therefore need to consider complex-valued function spaces.

Multiplying the PDE (3.5) by a test function and integrating by parts (i.e. using G1 (4.4) with $D = \Omega$) we obtain

$$-\int_\Omega f\,\overline{v} = \int_{\partial\Omega} \overline{\gamma v}\,\partial_n u - \int_\Omega \left(\nabla u \cdot \overline{\nabla v} - k^2 u\overline{v}\right) \tag{6.12}$$

for all $v \in H^1(\Omega)$. Using the impedance boundary condition (3.6) in the integral over Γ we find that (6.12) is equivalent to $a_I(u,v) = F(v)$, where

$$a_I(u,v) := \int_\Omega \left(\nabla u \cdot \overline{\nabla v} - k^2 u\overline{v}\right) - \mathrm{i}\eta \int_\Gamma \gamma u\,\overline{\gamma v} \quad \text{and} \quad F(v) := \int_\Omega f\overline{v} + \int_\Gamma g\overline{v} \tag{6.13}$$

(with the subscript I standing for "impedance"). Note that $a_I(\cdot,\cdot)$ is a sesquilinear form on $H^1(\Omega)\times H^1(\Omega)$, and $F(\cdot)$ is an anti-linear functional on $H^1(\Omega)$.

The variational formulation of the IIP is therefore, given $f\in L^2(\Omega)$, $g\in L^2(\Gamma)$, and $\eta\in\mathbb{R}\setminus\{0\}$,

$$\boxed{\text{find } u\in H^1(\Omega) \text{ such that } a_I(u,v)=F(v) \text{ for all } v\in H^1(\Omega).} \tag{6.14}$$

As in the case of the Dirichlet problem, we use the weighted norm $\|\cdot\|_{1,k,\Omega}$ defined by (6.7). The Cauchy-Schwarz inequality (2.5) and the multiplicative trace inequality (2.2) imply that $a_I(\cdot,\cdot)$ is continuous with

$$C_c\sim\left(1+\frac{|\eta|}{k}\right). \tag{6.15}$$

We now turn to coercivity, and prove analogues of Lemmas 6.1 and 6.2 For the first, we need an appropriate analogue of the Poincaré inequality (6.5) for functions in $H^1(\Omega)$ and not just $H^1_0(\Omega)$; this is

$$\|v\|^2_{L^2(\Omega)}\le C\big(\,\|\nabla v\|^2_{L^2(\Omega)}+\|\gamma v\|^2_{L^2(\Gamma)}\,\big)\qquad\text{for all } v\in H^1(\Omega). \tag{6.16}$$

The inequality (6.16) follows from the inequality

$$\|v\|_{L^2(\Omega)}\lesssim\big(\,\|\nabla v\|_{L^2(\Omega)}+\|\gamma v\|_{L^1(\Gamma)}\,\big)\qquad\text{for all } v\in H^1(\Omega) \tag{6.17}$$

by using the Cauchy-Schwarz inequality (to show that $\|\gamma v\|_{L^1(\Gamma)}\lesssim\|\gamma v\|_{L^2(\Gamma)}$) and the inequality (2.4). The inequality (6.17) can be proved using the inequality

$$\|v-(v)_\Omega\|_{L^2(\Omega)}\lesssim\|\nabla v\|_{L^2(\Omega)}\qquad\text{for all } v\in H^1(\Omega), \tag{6.18}$$

where $(v)_\Omega$ denotes the average of v over Ω (see, e.g., [Eva98, §5.8.1, Theorem 1] for the proof of (6.18), and [BS00, §5.3] for the proof of (6.17) from (6.18)).

Lemma 6.4 ($a_I(\cdot,\cdot)$ is coercive when k is sufficiently small) *Let C be the constant in the inequality (6.16). If*

$$2Ck^2+C^2\frac{k^4}{|\eta|^2}\le\frac{1}{2} \tag{6.19}$$

then $a_I(\cdot,\cdot)$ is coercive (in the norm $\|\cdot\|_{1,k,\Omega}$) with $\alpha=1/2$. (Note that if $|\eta|=k$ then (6.19) reduces to the requirement that k be sufficiently small.)

Sketch proof. We have

$$\Re a_I(v,v)=\|\nabla v\|^2_{L^2(\Omega)}-k^2\|v\|^2_{L^2(\Omega)}\qquad\text{and}\qquad\Im a_I(v,v)=-\eta\|\gamma v\|^2_{L^2(\Gamma)}. \tag{6.20}$$

The result follows after forming $|a_I(v,v)|^2$ using the expressions (6.20), and then using the inequalities (6.16), (2.3), and (2.4) (in that order). ∎

For the IDP we knew in advance that $a_D(\cdot,\cdot)$ could not be coercive for all $k>0$ since the variational problem (6.3) does not have a unique solution when $k^2=\lambda_j$. In contrast, the solution of the IIP is unique for every $k>0$ by Theorem 3.5 (assuming $\eta\in\mathbb{R}\setminus\{0\}$) and thus is it not immediately obvious whether $a_I(\cdot,\cdot)$ is coercive for every $k>0$ or not.

Lemma 6.5 ($a_I(\cdot,\cdot)$ is not coercive when k is sufficiently large) *Let λ_1 be the first Dirichlet eigenvalue of the negative Laplacian in Ω. If $k^2\ge\lambda_1$ then there exists a $v\in H^1(\Omega)$ with $a_I(v,v)=0$.*

Proof. If $v\in H^1_0(\Omega)$ then

$$a_D(v,v)=a_I(v,v)$$

(the integral over Γ in $a_I(v,v)$ vanishes since $\gamma v=0$). The result then follows from Lemma 6.2, since $H^1_0(\Omega)\subset H^1(\Omega)$. ∎

The definition of $a_I(\cdot,\cdot)$ implies that

$$\Re a_I(v,v) = \|v\|^2_{1,k,\Omega} - 2k^2\|v\|^2_{L^2(\Omega)},$$

and thus $a_I(\cdot,\cdot)$ satisfies a Gårding inequality with $\mathcal{H} = H^1(\Omega)$, $\mathcal{V} = L^2(\Omega)$, $\alpha = 1$, and $C_{\mathcal{V}} = 2k^2$. Since $H^1(\Omega)$ is compactly contained in $L^2(\Omega)$ [Eva98, §5.7], [McL00, Theorem 3.27] and the solution of the IIP is unique for every $k > 0$ (by Theorem 3.5), Theorems 5.20 and 5.18 give properties K1, K2, and K3 (with K3 holding once the subspace dimension is large enough).

Remark 6.6 (Finite dimensional subspaces) *The standard choices of finite-dimensional subspaces of $H^1(\Omega)$ (or $H^1_0(\Omega)$) are subspaces consisting of piecewise-polynomials, i.e. polynomials supported on each element of a triangulation of Ω. We note, however, that the* partition of unity finite element method *(PUFEM) for the Helmholtz equation introduced in [Mel95], [MB96], [BM97] uses piecewise polynomials multiplied by local solutions of the homogeneous PDE (i.e. solutions of $\mathcal{L}_k v = 0$ on each element). This is an example of a "wave-based method"; see Remark 7.4.*

Remark 6.7 (Illustration of Theorem 5.21 – Quasi-optimality of the Galerin method for the IIP) *We now show how Theorem 5.21 can be used to find bounds on both the constant of quasi-optimality C_{qo} and the threshold N_0 after which quasi-optimality holds.*

Consider the case $\eta = k$. Then, with the explicit values of α and $C_{\mathcal{V}}$ given above, the condition (5.19) becomes

$$\eta(\mathcal{H}_N) \leq \frac{1}{2kC_c} \tag{6.21}$$

(and recall from (6.15) that C_c is independent of k).

Let $\mathcal{T}_h$ be a family of regular triangulations of Ω (see Definition 2.1) and assume that each element of $\mathcal{T}_h$ is a simplex, i.e. a triangle in 2-d and a tetrahedron in 3-d. Fix $p \in \mathbb{N}$ and let $\mathcal{H}_N := \{v \in C(\overline{\Omega}) : v|_K$ is a polynomial of degree $\leq p$ for each $K \in \mathcal{T}_h\}$ (note that the subspace dimension, N, is then proportional to h^{-d}). Then, by [SZ90, Theorem 4.1],

$$\min_{v_N \in \mathcal{H}_N} \|w - v_N\|_{1,k,\Omega} \lesssim h\|w\|_{H^2(\Omega)} + hk\|w\|_{H^1(\Omega)}, \tag{6.22}$$

where the omitted constant is independent of h and k.

Given $f \in L^2(\Omega)$, let $w = S^* f$ (where $S^* f$ satisfies the variational problem (5.17)). Using the definition of $a_I(\cdot,\cdot)$, one can then show that w satisfies the BVP

$$\Delta w + k^2 w = -f \quad \text{in } \Omega \qquad \text{and} \qquad \partial_n w + \mathrm{i}kw = 0 \quad \text{on } \Gamma$$

(recalling that we're taking $\eta = k$). The question of bounding $\eta(\mathcal{H}_N)$ is then reduced to bounding the H^2- and H^1-norms of w in terms of the L^2-norm of f, with the k-dependence of the constants known explicitly. It can be shown that, under some geometric assumptions, given $k_0 > 0$,

$$\|w\|_{H^2(\Omega)} \lesssim k\|f\|_{L^2(\Omega)} \qquad \text{and} \qquad \|w\|_{H^1(\Omega)} \lesssim \|f\|_{L^2(\Omega)} \tag{6.23}$$

for all $k \geq k_0$ (where the omitted constants are independent of k) [Mel95, Proposition 8.1.4], [CF06, Theorem 1] (see also [GGS15, Remark 2.5, Theorem 2.9, and Lemma 2.12]). Using the bounds (6.23) along with (6.22) in the definition of $\eta(\mathcal{H}_N)$ (5.18), we find that

$$\eta(\mathcal{H}_N) \lesssim hk.$$

This last bound, along with (6.21), implies that there exists a $C > 0$ (independent of h and k) such that if

$$hk^2 \leq C$$

then the quasi-optimality bound (5.20) holds.

7 Discontinuous Galerkin (DG) formulation and Trefftz-Discontinuous Galerkin (TDG) formulation

Summary: (effectively) G1 with v a test function and $D = K$, where K is an element of a triangulation of Ω.

The term "discontinuous Galerkin method" encompasses many different methods, and so we first give a brief (and broad) overview of these. The overall idea is to use finite-dimensional spaces consisting of functions that are not continuous (and thus might not be in $H^1(\Omega)$). Since the solutions of the BVPs introduced in §3 are continuous (by Remark 3.6), their Galerkin approximations should ideally be continuous (or "close-to-continuous"). There are then three standard ways of imposing this continuity:

1. using *numerical fluxes*,
2. using *interior penalty terms*, or
3. using *Lagrange multipliers*.

(Note that there is some overlap between 1 and 2, since many interior penalty methods can be obtained via special choices of the numerical fluxes through the framework of DG methods established in [ABCM02].)

The plan of this section is the following. In §7.1 we sketch the DG approach based on 1. above (i.e. numerical fluxes) for the Helmholtz IIP. We do not state explicitly the variational problem because defining the relevant $\mathcal{H}$, $a(\cdot,\cdot)$, and $F(\cdot)$ would involve introducing a fairly substantial amount of new notation, with relatively little insight in return. (For the reader interested in the precise details, the discussion below contains appropriate references.) We then briefly discuss approaches based on 2. and 3. in Remarks 7.1 and 7.2 respectively. We also briefly mention a new discontinuous-Galerkin-type method, the so-called discontinuous Petrov-Galerkin (DPG) method, in Remark 7.3.

In §7.2 we discuss Trefftz-Discontinuous Galerkin (TDG) methods for the Helmholtz IIP. The idea behind these methods is that solutions of $\mathcal{L}_k u = 0$ in Ω can be approximated well by piecewise-solutions of the homogeneous Helmholtz equation, i.e. functions v such that $\mathcal{L}_k v = 0$ on each K of a triangulation of Ω. (Note that requiring such a v to be continuous on Ω is very restrictive, and thus the discontinuous setting is essential.) TDG methods arise from considering DG methods with these particular subspaces.

Finally, we note that in this section (and this section only) we do not use the notation γu for the trace of u, and write only u instead. Similarly we write the normal derivative of u as $\nabla u \cdot \mathbf{n}$ instead of $\partial_n u$. We do this to keep the notation consistent with that in the references we give.

7.1 The DG formulation

The starting point of the DG formulation is the IIP written as a first-order system. With $\boldsymbol{\sigma}$ defined by

$$\mathrm{i}k\boldsymbol{\sigma} := \nabla u \quad \text{in } \Omega, \tag{7.1}$$

the IIP becomes

$$\mathrm{i}ku - \nabla \cdot \boldsymbol{\sigma} = \frac{1}{\mathrm{i}k} f \quad \text{in } \Omega \tag{7.2}$$

$$\mathrm{i}\boldsymbol{\sigma} \cdot \mathbf{n} - \mathrm{i}\eta\gamma u = g \quad \text{on } \Gamma. \tag{7.3}$$

Multiplying (7.1) by the complex conjugate of the vector test function $\boldsymbol{\tau}$, multiplying (7.2) by the complex conjugate of the scalar test function v, integrating both equations over $K \in \mathcal{T}$ (where $\mathcal{T}$ is a triangulation of Ω, see Definition 2.1), and integrating by parts (i.e using (1.6)), we obtain

$$\mathrm{i}k \int_K \boldsymbol{\sigma} \cdot \overline{\boldsymbol{\tau}} + \int_K u \overline{\nabla \cdot \boldsymbol{\tau}} - \int_{\partial K} u \, \overline{\boldsymbol{\tau}} \cdot \mathbf{n} = 0, \tag{7.4a}$$

and

$$\mathrm{i}k\int_K u\overline{v} + \int_K \boldsymbol{\sigma}\cdot\overline{\nabla v} - \int_{\partial K}\boldsymbol{\sigma}\cdot\mathbf{n}\,\overline{v} - \frac{1}{\mathrm{i}k}\int_K f\overline{v} = 0. \tag{7.4b}$$

If $u \in H^1(K)$ and $\boldsymbol{\sigma} \in L^2(K)$, then the equation (7.4a) holds for all $\boldsymbol{\tau} \in H(\mathrm{div};K) := \{\boldsymbol{\tau} : \boldsymbol{\tau} \in L^2(K), \nabla\cdot\boldsymbol{\tau} \in L^2(K)\}$ and the equation (7.4b) holds for all $v \in H^1(K)$.

We now introduce finite-dimensional spaces $V_p(K)$ and $\mathbf{V}_p(K)$ such that $V_p(K) \subset H^1(K)$ and $\mathbf{V}_p(K) \subset H(\mathrm{div};K)$, and replace u, v by $u_p, v_p \in V_p(K)$ and replace $\boldsymbol{\sigma}, \boldsymbol{\tau}$ by $\boldsymbol{\sigma}_p, \boldsymbol{\tau}_p \in \mathbf{V}_p(K)$. Finally, we approximate the traces of u and $\boldsymbol{\sigma}$ on K by the *numerical fluxes* $\widehat{u}_p$ and $\widehat{\boldsymbol{\sigma}}_p$ (which will be specified later in terms of u_p and ∇u_p). Doing all this, we find that (7.4) becomes

$$\mathrm{i}k\int_K \boldsymbol{\sigma}_p\cdot\overline{\boldsymbol{\tau}_p} + \int_K u_p\,\overline{\nabla\cdot\boldsymbol{\tau}_p} - \int_{\partial K}\widehat{u}_p\overline{\boldsymbol{\tau}_p}\cdot\mathbf{n} = 0 \tag{7.5a}$$

and

$$\mathrm{i}k\int_K u_p\overline{v}_p + \int_K \boldsymbol{\sigma}_p\cdot\overline{\nabla v}_p - \int_{\partial K}\widehat{\boldsymbol{\sigma}}_p\cdot\mathbf{n}\,\overline{v}_p - \frac{1}{\mathrm{i}k}\int_K f\overline{v} = 0. \tag{7.5b}$$

We now integrate by parts back in (7.5a) to obtain

$$\mathrm{i}k\int_K \boldsymbol{\sigma}_p\cdot\overline{\boldsymbol{\tau}_p} - \int_K \nabla u_p\cdot\overline{\boldsymbol{\tau}_p} - \int_{\partial K}(\widehat{u}_p - u_p)\overline{\boldsymbol{\tau}_p}\cdot\mathbf{n} = 0, \tag{7.6}$$

Observe that if $\widehat{u}_p = u_p$ and $\mathrm{i}k\boldsymbol{\sigma}_p = \nabla u_p$ then (7.6) states that $0 = 0$.

We now recombine (7.5b) and (7.6), and seek to eliminate the variable $\boldsymbol{\sigma}$ and the test function $\boldsymbol{\tau}$. We do this by choosing $\boldsymbol{\tau}_p = \nabla v_p$ on each element, and to do this we need to assume that $\nabla_h V_p(K) \subset \mathbf{V}_p(K)$, where ∇_h denotes the element-wise gradient. Choosing $\boldsymbol{\tau}_p = \nabla v_p$ in (7.6) and substituting the resulting expression for $\int_K \boldsymbol{\sigma}_p\cdot\nabla v_p$ into (7.5b), we find that

$$\int_K \nabla u_p\cdot\overline{\nabla v_p} - k^2 u_p\overline{v_p} - \int_{\partial K}(u_p - \widehat{u}_p)\overline{\nabla v_p}\cdot\mathbf{n} - \mathrm{i}k\int_{\partial K}\widehat{\boldsymbol{\sigma}}_p\cdot\mathbf{n}\,\overline{v_p} = \int_K f\overline{v}. \tag{7.7}$$

Note that (a) we have succeeded in eliminating $\boldsymbol{\sigma}_p$ and $\boldsymbol{\tau}_p$, but $\widehat{\boldsymbol{\sigma}}_p$ remains, and (b) if $\widehat{u}_p = u_p$ and $\mathrm{i}k\widehat{\boldsymbol{\sigma}}_p = \nabla u_p$ then (7.7) is just G1 (4.4) with $D = K$. (One can therefore understand (7.7) as arising from applying G1 on K to v_p and u_p but, firstly, making ∇u_p on ∂K a new variable $\widehat{\boldsymbol{\sigma}}_p/ik$ and, secondly, introducing a term on ∂K with a new variable $\widehat{u}_p$ such that the term is zero when $\widehat{u}_p = u_p$.)

The equation (7.7) is the basis of the DG variational formulation. To obtain this variational formulation from (7.7) one must

(i) sum (7.7) over $K \in \mathcal{T}$,

(ii) decide on the definitions of $\widehat{\boldsymbol{\sigma}}_p$ and $\widehat{u}_p$ in terms of u_p and $\nabla_h u_p$ on adjacent elements (these definitions will be different when ∂K contains part of Γ than when $\partial K \subset \Omega$), and

(iii) use the boundary condition (7.3) on Γ.

We omit the details; these can be found (for the Helmholtz IIP) in, e.g., [EM12, §6.3].

Continuity and coercivity properties. For certain choices of the numerical fluxes, the sesquilinear form of the DG formulation is continuous (in an appropriate norm) and satisfies a Gårding inequality; see, e.g., [ABCM02, §4.1-4.2], [PS02, Proposition 3.1], and [BBD13, Theorem 3.3].

Remark 7.1 (Interior penalty methods) *Interior penalty methods impose continuity of the solution via terms that penalise the jump of the solution (and possibly its derivatives) across element edges. There are several different interior penalty methods for the Helmholtz equation, but we mention here the methods introduced in [FW09], [FW11], and [FX13]. The methods introduced in the first two papers do not fit in the framework of [ABCM02] (i.e. they do not arise from an appropriate choice of numerical fluxes), but those introduced in [FX13] do. The novelty of these methods is that, although the sesquilinear forms are not coercive, some of the consequences*

of coercivity hold; namely, the Galerkin equations have a unique solution without any constraint on the dimension of the (piecewise polynomial) approximation space, and error estimates can be obtained that are explicit in k, h, and p (where p is the polynomial degree) [FW09, Remarks 4.3 and 5.1], [FW11, Remark 3.2].

Note that one can also add similar penalty terms to the sesquilinear form of the standard variational formulation of the Helmholtz IIP ($a_I(\cdot,\cdot)$ defined by (6.13)), and some of the resulting methods share the features just described; see [Wu14, Corollary 3.5 and Theorem 4.4].

Remark 7.2 (Methods using Lagrange multipliers) *Probably the most well-known method for solving the Helmholtz equation that imposes continuity via Lagrange multipliers is the* discontinuous enrichment method *(DEM) (and related discontinuous Galerkin method) of Farhat and collaborators. The DEM, introduced in [FHF01], consists of imposing the standard variational formulation on each element K and then imposing inter-element continuity weakly via Lagrange multipliers. The main idea in [FHF01] is then to "enrich" the piecewise-polynomial approximation space with plane waves* $\exp(\mathrm{i}k\mathbf{x}\cdot\mathbf{d})$*, where* $\mathbf{d} = (\cos\theta, \sin\theta)^T$*. In [FHH03] (and subsequent papers) the polynomial part of the space was dropped, and then since there was no longer any "enrichment", the method was called a discontinuous Galerkin method (and abbreviated to DGM); see [FHH03, §1.2] or [Moi11, §1.2.2] for more details. (Note that with the choice of plane-wave subspaces, the DG method of [FHH03] falls under the category of TDG methods described in §7.2.)*

Remark 7.3 (The discontinuous Petrov-Galerkin (DPG) method) *The DPG method applied to the Helmholtz IIP uses equations (7.4) to formulate a variational problem with unknowns u, $\boldsymbol{\sigma}$, $\widehat{u}$, and $\widehat{\boldsymbol{\sigma}}$ (where $\widehat{u}$ and $\widehat{\boldsymbol{\sigma}}$ are numerical fluxes); see [DG11, §3.1], [DGMZ12, §2.4]. This variational problem has different trial and test spaces (i.e. it is of the form (5.2) with $\mathcal{H}_1 \neq \mathcal{H}_2$, hence the "Petrov" in the name "DPG"). The key point of the DPG method is that, given finite-dimensional subspaces $\mathcal{H}_N^1$ of $\mathcal{H}_1$, the method defines so-called "optimal test spaces" $\mathcal{H}_N^2$ of $\mathcal{H}_2$; these optimal test spaces are such that, when the sesquilinear form satisfies the inf-sup condition (5.4), the discrete inf-sup condition (5.10) is* automatically *satisfied; see, e.g., [Gop13, Proposition 5]. These optimal test spaces admit functions with no continuity constraints across element interfaces, hence the "discontinuous" in "DPG" [Gop13, Definition 16], [DG14, §3]. In practice, the optimal test spaces are not computed exactly (since this would require solving BVPs on each $K \in \mathcal{T}$), but instead are approximated [Gop13, §3]. The DPG method can also be thought of as a least-squares method in a non-standard inner product; see, e.g., [Gop13, Theorem 13], [DG14, §2].*

7.2 The Trefftz-DG formulation

The idea behind the Trefftz-DG formulation is to assume that $V_p(K)$ satisfies the *Trefftz property*

$$\mathcal{L}_k v_p = 0 \ \text{ in } K \quad \text{for all } v_p \in V_p(K). \tag{7.8}$$

This property allows us to perform another integration by parts in (7.7), i.e. we use

$$\int_K \nabla u_p \cdot \overline{\nabla v_p} = -\int_K u_p \overline{\Delta v_p} + \int_{\partial K} u_p \overline{\nabla v_p} \cdot \mathbf{n}$$

to obtain

$$-\int_K \overline{(\Delta v_p + k^2 v_p)} u_p + \int_{\partial K} \widehat{u}_p \overline{\nabla v_p} \cdot \mathbf{n} - \mathrm{i}k \int_{\partial K} \widehat{\boldsymbol{\sigma}}_p \cdot \mathbf{n}\, \overline{v_p} = \int_K f \overline{v}. \tag{7.9}$$

The fact that $V_p(K)$ satisfies the Trefftz property means that (7.9) becomes

$$\int_{\partial K} \widehat{u}_p \overline{\nabla v_p} \cdot \mathbf{n} - \mathrm{i}k \int_{\partial K} \widehat{\boldsymbol{\sigma}}_p \cdot \mathbf{n}\, \overline{v_p} = \int_K f \overline{v}. \tag{7.10}$$

Just as (7.7) is the basis of the DG variational formulation for the IIP, (7.10) is the basis of the Trefftz-DG variational formulation of the IIP. To obtain this variational formulation from (7.10) we carry out the steps (i)–(iii) on the previous page; see [EM12, §6.3] or [Moi11, §4.2] for the details.

Although we have formulated the variational problem with $f \neq 0$, the TDG formulation is usually only considered for the case $f = 0$. This is because, whereas solutions of $\mathcal{L}_k u = 0$ are well approximated by functions in $V_p(K)$ [Moi11, Chapter 3] (which was the original motivation for considering these spaces), solutions of the inhomogeneous Helmholtz equation $\mathcal{L}_k u = -f$ are (in general) poorly approximated by functions in $V_p(K)$ (see [Moi11, Remark 3.5.11]).

Continuity and coercivity properties. The sesquilinear form of the TDG formulation is continuous and coercive in a norm consisting of jumps across element boundaries; see [HMP11, §3.1], [Moi11, §4.3], [BM08, Lemma 3.4]. The error estimates given by Céa's lemma (Theorem 5.16) can be then be used to obtain estimates in the L^2-norm through duality arguments (i.e. arguments involving solutions of certain adjoint problems). The key result that allows one to do this in a Trefftz framework is [MW99, Theorem 3.1], with this result then used in [BM08, §4], [Moi11, §4.3.1].

Remark 7.4 ("Wave-based methods") *TDG methods for the Helmholtz equation come under the broad heading of "wave-based methods". Roughly speaking, there are two main classes of such methods. The first class consists of methods that use basis functions that are locally solutions of the PDEs (with these bases used either in a DG setting, as discussed above, or in a least-squares setting); these methods are then called Trefftz methods. The second class consists of methods that use "modulated bases", i.e. the basis functions are local solutions of the PDEs multiplied by non-oscillatory functions (such as low-degree polynomials); the standard example of a method in this class is the partition of unity finite element method (PUFEM) mentioned in Remark 6.6. We refer the reader to the theses [Moi11] and [Luo13] for good overviews of wave-based methods.*

8 The UWVF and the quadratic functional of Després

Summary: G2 with v a solution of $\mathcal{L}v = 0$ can be used to prove the so-called "isometry lemma". This lemma is then applied with $D = K$ (where K is an element of a triangulation of Ω) to obtain the UWVF, or with $D = \Omega$ to obtain the quadratic functional.

8.1 The isometry lemma

Both the Ultra-Weak Variational Formulation (UWVF) and the quadratic functional introduced by Després are based on the following lemma, which is a consequence of G2 (4.6) when $\mathcal{L}v = 0$.

Lemma 8.1 ("Isometry lemma") *Let D be a bounded Lipschitz domain. If $u, v \in H^1(D, \Delta)$ with $\mathcal{L}u = 0$, $\mathcal{L}v = 0$, $\partial_n u \in L^2(\partial D)$, and $\partial_n v \in L^2(\partial D)$, then, for $\eta \in \mathbb{R}$,*

$$\int_{\partial D} \big(\partial_n u + \mathrm{i}\eta\gamma u\big)\overline{\big(\partial_n v + \mathrm{i}\eta\gamma v\big)} = \int_{\partial D} \big(\partial_n u - \mathrm{i}\eta\gamma u\big)\overline{\big(\partial_n v - \mathrm{i}\eta\gamma v\big)}. \tag{8.1}$$

Proof. Expanding the brackets on both sides, we see that (8.1) reduces to two copies of G2 (4.6). ∎

This result is called the isometry lemma because when $u = v$ (8.1) becomes

$$\|\partial_n u + \mathrm{i}\eta\gamma u\|^2_{L^2(\partial D)} = \|\partial_n u - \mathrm{i}\eta\gamma u\|^2_{L^2(\partial D)}\,, \tag{8.2}$$

i.e. the map from one impedance trace to the other is an isometry.

Although the isometry lemma is valid when $\mathcal{L} = \Delta + \lambda$ with $\lambda \in \mathbb{R}$, in what follows we only consider the Helmholtz equation (i.e. $\mathcal{L} = \mathcal{L}_k$). This is because, although the two formulations discussed in this section can be used to solve BVPs for the Laplace and modified Helmholtz equations, the formulations are most naturally applied to certain Helmholtz BVPs.

8.2 The Ultra-Weak Variational Formulation (UWVF)

We restrict attention to the Helmholtz IIP. Furthermore, since the isometry lemma concerns solutions of the homogeneous Helmholtz equation, we take $f = 0$.

First some notation: let $\mathcal{T}$ be a triangulation of Ω, with elements $\{K_j\}_{j=1}^N$. Let $\boldsymbol{\nu}^{K_j}$ be the outward-pointing unit normal vector to K_j, let γ_j be the trace on ∂K_j, and let ∂^j_ν be the normal

derivative trace on ∂K_j. Let $\Sigma_{jl} := \overline{K_j} \cap \overline{K_l}$ with normal $\boldsymbol{\nu}^{K_j}$ (so $\Sigma_{jl} = \Sigma_{lj}$ but the normal on Σ_{lj} equals $\boldsymbol{\nu}^{K_l} = -\boldsymbol{\nu}^{K_j}$). Let $\Gamma_j := \partial K_j \cap \Gamma$. These definitions imply that

$$\int_{\partial K_j} = \sum_{l=1,\, l\neq j}^{N} \int_{\Sigma_{jl}} + \int_{\Gamma_j}. \tag{8.3}$$

The unknowns in the variational problem will be

$$\mathcal{X}_j := \partial_\nu^j u + \mathrm{i}\eta\gamma_j u, \quad j = 1, \ldots, N.$$

Observe that if we know all the $\mathcal{X}_j$s then we know $\partial_n u + \mathrm{i}\eta\gamma u$ on Γ, and combining this with the impedance boundary condition (3.6) we then know both $\partial_n u$ and γu on Γ. We can then find u in Ω using Green's integral representation (Theorem 9.1 below).

We now assume that

(i) $\partial_\nu^j u \in L^2(\partial K_j)$ for all j, and

(ii) when $\Sigma_{jl} \neq \emptyset$, $\partial_\nu^j u = -\partial_\nu^l u$.

Regarding (i), $\partial_\nu^j u$ is always in $L^2(\partial K_j)$ for elements K_j away from Γ by interior regularity (see Remark 3.6). The reason we assume (i) is that we then have $\mathcal{X}_j \in L^2(\partial K_j)$ for all j. Regarding (ii), this is certainly the case if $u \in H^2(\Omega)$, since then $\partial_\nu^j u = \boldsymbol{\nu}^{K_j} \cdot \gamma(\nabla u)$ and when $\Sigma_{jl} \neq \emptyset$, $\boldsymbol{\nu}^{K_j} = -\boldsymbol{\nu}^{K_l}$ (and the traces of ∇u from either side of Σ_{jl} are equal).

Given $v^{K_j} : K_j \to \mathbb{C}$ with $\mathcal{L}_k v^{K_j} = 0$ and $\gamma_j v^{K_j}, \partial_\nu^j v^{K_j} \in L^2(\partial K_j)$ for all j, we define

$$\mathcal{Y}_j := \big(\partial_\nu^j v^{K_j} + \mathrm{i}\eta\gamma_j v^{K_j}\big) \in L^2(\partial K_j). \tag{8.4}$$

We then define $F_j(\mathcal{Y}_j)$ to be the other impedance trace, i.e.

$$F_j(\mathcal{Y}_j) := \big(-\partial_\nu^j v^{K_j} + \mathrm{i}\eta\gamma_j v^{K_j}\big) \in L^2(\partial K_j).$$

Therefore $F_j : L^2(\partial K_j) \to L^2(\partial K_j)$ and is an isometry by Lemma 8.1. Applying the result of the isometry lemma (8.1) with $D = K_j$, $u = u$, and $v = v^{K_j}$, and rewriting the resulting equation using the notation introduced above and the expression (8.3), we obtain

$$\int_{\partial K_j} \mathcal{X}_j \overline{\mathcal{Y}_j} = \sum_{l=1,\, l\neq j}^{N} \int_{\Sigma_{jl}} \big(-\partial_\nu^j u + \mathrm{i}\eta\gamma_j u\big)\overline{F_j(\mathcal{Y}_j)} + \int_{\Gamma_j} \big(-\partial_\nu^j u + \mathrm{i}\eta\gamma_j u\big)\overline{F_j(\mathcal{Y}_j)}. \tag{8.5}$$

When $\Sigma_{jl} \neq \emptyset$, $\gamma_j u = \gamma_l u$, and then this fact, along with the assumption (ii) above, implies that

$$\big(-\partial_\nu^j u + \mathrm{i}\eta\gamma_j u\big)\Big|_{\Sigma_{jl}} = \big(\partial_\nu^l u + \mathrm{i}\eta\gamma_l u\big)\Big|_{\Sigma_{lj}} = \mathcal{X}_l.$$

The boundary condition (3.6) implies that

$$\big(-\partial_\nu^j u + \mathrm{i}\eta\gamma_j u\big)\Big|_{\Gamma_j} = -g,$$

and using these last two equations in (8.5) we obtain

$$\int_{\partial K_j} \mathcal{X}_j \overline{\mathcal{Y}_j} = \sum_{l=1,\, l\neq j}^{N} \int_{\Sigma_{lj}} \mathcal{X}_l\, \overline{F_j(\mathcal{Y}_j)} - \int_{\Gamma_j} g\, \overline{F_j(\mathcal{Y}_j)}. \tag{8.6}$$

Define the Hilbert space $\mathcal{H}$ by

$$\mathcal{H} := \prod_{j=1}^{N} L^2(\partial K_j)$$

and let $\mathcal{X} := (\mathcal{X}_1, \ldots, \mathcal{X}_N)$; assumption (i) above then implies that $\mathcal{X} \in \mathcal{H}$. Define an inner product on $\mathcal{H}$ by

$$\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{j=1}^{N} \int_{\partial K_j} \mathcal{X}_j \overline{\mathcal{Y}_j}.$$

Then, summing up (8.6) over j, we find that

$$a(\mathcal{X},\mathcal{Y}) = G(\mathcal{Y}) \quad \text{for all } \mathcal{Y} \text{ of the form (8.4)} \tag{8.7}$$

(i.e. impedance traces of piecewise solutions of $\mathcal{L}_k v = 0$), where

$$a(\mathcal{X},\mathcal{Y}) := \langle \mathcal{X},\mathcal{Y}\rangle - \sum_{j=1}^{N}\sum_{l=1,\, l\neq j}^{N} \int_{\Sigma_{lj}} \mathcal{X}_l \overline{F_j(\mathcal{Y}_j)}$$

and

$$G(\mathcal{Y}) := -\sum_{j=1}^{N} \int_{\Gamma_j} g\, \overline{F_j(\mathcal{Y}_j)}.$$

We then consider the variational problem

$$\boxed{\text{find } \mathcal{X}\in\mathcal{H} \text{ such that} \quad a(\mathcal{X},\mathcal{Y}) = G(\mathcal{Y}) \quad \text{for all } \mathcal{Y}\in\mathcal{H}.} \tag{8.8}$$

Given $\mathcal{Y}\in\mathcal{H}$, there exist $\{v^{K_j}\}_{j=1}^N$ such that $\mathcal{L}_k v^{K_j} = 0$ in K_j and $(\partial_\nu^j v^{K_j} + \mathrm{i}\eta\gamma_j v^{K_j}) = \mathcal{Y}_j$ for $j = 1,\dots,N$ (using existence of a solution to the IIP on each K_j). Therefore, by (8.7), the $\mathcal{X}$ corresponding to the solution of the IIP in Ω satisfies the variational problem (8.8). The continuity and coercivity properties of $a(\cdot,\cdot)$ discussed below imply that the variational problem (8.8) has a unique solution, and thus the only solution of (8.8) is $\mathcal{X}$ given by $\mathcal{X}_j := \partial_\nu^j u + \mathrm{i}\eta\gamma_j u$, $j = 1,\dots,N$, where u is the solution of the IIP.

Discretising the variational problem (8.8). Seeking to discretise (8.8) by choosing a finite-dimensional subspace $\mathcal{H}_N \subset \mathcal{H}$, we see that we need to be able to find easily $\{F_j(\mathcal{Y}_j)\}_{j=1}^N$ for $\mathcal{Y}\in\mathcal{H}_N$. Given a $\mathcal{Y}\in\mathcal{H}_N$, finding $\{F_j(\mathcal{Y}_j)\}_{j=1}^N$ means solving an impedance BVP on each K_j. These BVPs are less oscillatory than the original Helmholtz BVP (since the K_j are smaller than Ω), and are therefore easier to solve. However, we can make the BVPs on K_j even easier to solve by letting $\mathcal{H}_N$ consist of impedance traces of explicit solutions of $\mathcal{L}_k v = 0$ on each K_j, since finding $\{F_j(\mathcal{Y}_j)\}_{j=1}^N$ is then straightforward. More precisely, for $K\in\mathcal{T}$ let M_P^K be a finite-dimensional subspace of P explicit solutions of $\mathcal{L}_k v = 0$ on K (and we allow the number P to vary from element to element). Let $\mathcal{H}_N := \prod_{j=1}^N \mathcal{H}_{N,j}$, where

$$\mathcal{H}_{N,j} := \left\{ \mathcal{Y}_j : \mathcal{Y}_j = \partial_\nu^j v + \mathrm{i}\eta\gamma_j v \text{ for } v\in M_{P_j}^{K_j} \right\}.$$

Given $\mathcal{Y}_j\in\mathcal{H}_{N,j}$, we obtain $F_j(\mathcal{Y}_j)$ by finding the $v\in M_{P_j}^{K_j}$ that gave rise to that particular $\mathcal{Y}_j$ and then setting $F_j(\mathcal{Y}_j) = -\partial_\nu^j v + \mathrm{i}\eta\gamma_j v$.

Common choices of the functions in M_P^K are separable solutions of the Helmholtz equation, e.g. plane waves $\exp(\mathrm{i}k\mathbf{x}\cdot\mathbf{d})$, where $\mathbf{d} = (\cos\theta,\sin\theta)^T$ (the separable solutions in cartesian coordinates), spherical waves (the separable solutions in polar coordinates), or a mixture of the two; see, e.g., the review [LHM09, §3] for more details.

Continuity and coercivity properites. With $\mathcal{H}_N$ constructed as above, the UWVF can be recast as a particular Trefftz-DG method; this fact was noticed in [Gab07, §2.6], [BM08, §2], and [GHP09, §3]. The sesquilinear form is then continuous and coercive in a norm consisting of jumps across element boundaries; see the discussion in §7.2. (Note that a slightly weaker result was proved in the original analysis of the UWVF; see [CD98, Lemma 3.3, Equation 3.30].)

Remark 8.2 (Connection between UWVF and least squares) *When written in operator form, the UWVF is equivalent to a factorisation of the normal equations for the least squares method consisting of minimising the jumps in impedance traces across element boundaries; see [LHM09, Page 135]. (This helps explain why the UWVF is better conditioned than the least squares method.)*

8.3 The quadratic functional of Després

This variational formulation arises most naturally for problems with an impedance boundary condition (although it can be modified to work for other boundary conditions). It was introduced for the exterior impedance problem for the Helmholtz equation in [Des97] (see also [Des98] and [BC00]) and for the analogous problem for the time-harmonic Maxwell equations in [CD03].

Here we give the formulation for the Helmholtz IIP, and then we talk briefly about the exterior impedance problem at the end. Since the formulation is based on the isometry lemma (Lemma 8.1), it is only applicable when $f = 0$ and η is a real constant.

The formulation is based on the isometry lemma with $u = v$, i.e. (8.2), and this result boils down to the fact that

$$\Im \int_{\partial D} \gamma u \, \overline{\partial_n u} = 0. \tag{8.9}$$

The equation (8.9) can be proved only using G1 (4.4) (i.e. one doesn't need G2). However, we keep this formulation under G2 in the map in Figure 1 because of the conceptual link with the UWVF.

Given $g \in L^2(\Gamma)$ and $\eta \in \mathbb{R} \setminus \{0\}$, define the quadratic functional $I(\cdot)$ by

$$I(v) := \frac{1}{2} \left\| \partial_n v - \mathrm{i}\eta\gamma v \right\|^2_{L^2(\Gamma)} + \frac{1}{2} \left\| \partial_n v + \mathrm{i}\eta\gamma v \right\|^2_{L^2(\Gamma)} - 2\Re\big(\partial_n v - \mathrm{i}\eta\gamma v, g\big)_{L^2(\Gamma)} \tag{8.10}$$

and the space $\mathcal{H}$ by

$$\mathcal{H} := \big\{ v \in H^1(\Omega, \Delta) : \, \mathcal{L}_k v = 0, \, \partial_n v \in L^2(\Gamma) \big\} \tag{8.11}$$

(note that this is the space needed for the isometry lemma to hold).

Theorem 8.3 *Given $g \in L^2(\Gamma)$ and $\eta \in \mathbb{R} \setminus \{0\}$, let u be the solution of Helmholtz IIP with $f = 0$. Then $u \in \mathcal{H}$ and $I(u) \le I(v)$ for all $v \in \mathcal{H}$.*

Proof. By the definition of the IIP (Definition 3.4), $u \in H^1(\Omega)$. The PDE (3.5) implies that $\mathcal{L}_k u = 0$ and thus $\Delta u \in L^2(\Omega)$. Since $g \in L^2(\Gamma)$, the boundary condition (3.6) implies that $\partial_n u \in L^2(\Gamma)$; therefore $u \in \mathcal{H}$.

If $v \in \mathcal{H}$ then, by (8.2) with $D = \Omega$,

$$I(v) = \left\| \partial_n v - \mathrm{i}\eta\gamma v \right\|^2_{L^2(\Gamma)} - 2\Re\big(\partial_n v - \mathrm{i}\eta\gamma v, g\big)_{L^2(\Gamma)}.$$

Now, since $|a - b|^2 = |a|^2 + |b|^2 - 2\Re(a\bar{b})$,

$$\left\| \partial_n v - \mathrm{i}\eta\gamma v - g \right\|^2_{L^2(\Gamma)} = \left\| \partial_n v - \mathrm{i}\eta\gamma v \right\|^2_{L^2(\Gamma)} + \|g\|^2_{L^2(\Gamma)} - 2\Re\big(\partial_n v - \mathrm{i}\eta\gamma v, g\big)_{L^2(\Gamma)},$$

so

$$I(v) = \left\| \partial_n v - \mathrm{i}\eta\gamma v - g \right\|^2_{L^2(\Gamma)} - \|g\|^2_{L^2(\Gamma)}.$$

Therefore, the minimum of $I(v)$ is $-\|g\|^2_{L^2(\Gamma)}$ and this is reached when $\partial_n v - \mathrm{i}\eta\gamma v = g$; by uniqueness of the solution of the IIP, this is when $v = u$. ∎

Define the *impedance trace operators* $h^{\pm}(\cdot)$ by

$$h^{\pm}(v) := \partial_n v \pm \mathrm{i}\eta\gamma v,$$

so that

$$I(v) = \frac{1}{2} \left\| h^-(v) \right\|^2_{L^2(\Gamma)} + \frac{1}{2} \left\| h^+(v) \right\|^2_{L^2(\Gamma)} - 2\Re\big(h^-(v), g\big)_{L^2(\Gamma)}.$$

Define the sesquilinear form $a(\cdot, \cdot)$ and anti-linear functional $F(\cdot)$ by

$$a(v, w) := \frac{1}{2}\big(h^+(v), h^+(w)\big)_{L^2(\Gamma)} + \frac{1}{2}\big(h^-(v), h^-(w)\big)_{L^2(\Gamma)} \quad \text{and} \quad F(w) = \big(g, h^-(w)\big)_{L^2(\Gamma)}. \tag{8.12}$$

Lemma 8.4 *Given $g \in L^2(\Gamma)$ and $\eta \in \mathbb{R} \setminus \{0\}$, define $h^{\pm}(\cdot)$, $a(\cdot, \cdot)$, $F(\cdot)$, and $\mathcal{H}$ as above. If u is the solution of the IIP (with $f = 0$) then*

$$a(u, v) = F(v) \quad \text{for all } v \in \mathcal{H}. \tag{8.13}$$

Proof. Let $\varepsilon \in \mathbb{R}$. Theorem 8.3 implies that $I(u) \leq I(v)$ for all $v \in \mathcal{H}$, and therefore

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} I(u + \varepsilon v)\bigg|_{\varepsilon=0} = 0 \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}\varepsilon} I(u + \mathrm{i}\varepsilon v)\bigg|_{\varepsilon=0} = 0 \tag{8.14}$$

for all $v \in \mathcal{H}$. The first condition in (8.14) simplifies to the real part of (8.13), and the second condition in (8.14) simplifies to the imaginary part of (8.13). ∎

We now need to specify a norm on $\mathcal{H}$. We would like to let

$$\|w\|_{\mathcal{H}}^2 := \left\|h^+(w)\right\|_{L^2(\Gamma)}^2 + \left\|h^-(w)\right\|_{L^2(\Gamma)}^2, \tag{8.15}$$

since, if this is a norm, then $a(\cdot,\cdot)$ defined by (8.12) is continuous with $C_c = 1$ (via the inequality (2.4)) and coercive with $\alpha = 1/2$.

Lemma 8.5 *$\mathcal{H}$ is a Hilbert space with norm $\|\cdot\|_{\mathcal{H}}$ defined by (8.15).*

Proof. Define

$$\mathcal{V} := \left\{v \in H^1(\Omega, \Delta) : \partial_n v \in L^2(\Gamma),\, \gamma v \in H^1(\Gamma)\right\}. \tag{8.16}$$

A regularity result of Nečas [McL00, Lemma 4.24] implies that if $v \in H^1(\Omega, \Delta)$ then the conditions $\partial_n v \in L^2(\Gamma)$ and $\gamma v \in H^1(\Gamma)$ are equivalent (i.e. if one holds then so does the other, and vice versa). Therefore, we can append the condition $\gamma v \in H^1(\Gamma)$ to the definition of $\mathcal{H}$ (8.11), and thus see that $\mathcal{H} \subset \mathcal{V}$.

Now $\mathcal{V}$ is a Hilbert space with norm

$$\|v\|_{\mathcal{V}}^2 := k^2 \|v\|_{L^2(\Omega)}^2 + \|\nabla v\|_{L^2(\Omega)}^2 + k^{-2} \|\Delta v\|_{L^2(\Omega)}^2 + k^2 \|\gamma v\|_{L^2(\Gamma)}^2 + \|\nabla_\Gamma \gamma v\|_{L^2(\Gamma)}^2 + \|\partial_n v\|_{L^2(\Gamma)}^2,$$

where we have weighted the terms with k in a similar way to that in the norm (6.7).

It is straightforward to show that $\mathcal{H}$ is closed in $\mathcal{V}$ with the norm $\|\cdot\|_{\mathcal{V}}$; therefore if we can show that $\|\cdot\|_{\mathcal{V}}$ is equivalent to $\|\cdot\|_{\mathcal{H}}$ then we are done. From the definitions of the norms and the operators $h^{\pm}(\cdot)$, there exists a $C_1(k,\eta)$ such that $\|v\|_{\mathcal{H}} \leq C_1(k,\eta)\|v\|_{\mathcal{V}}$ for all $v \in \mathcal{V}$ (if $\eta = k$ then $C_1(k,\eta) \sim 1$). Furthermore, by the well-posedness of the IIP, there exists a $C_2(k,\eta)$ such that

$$\|v\|_{\mathcal{V}} \leq C_2(k,\eta) \left\|h^+(v)\right\|_{L^2(\Gamma)}$$

for all $v \in \mathcal{V}$. Since $\|h^+(v)\|_{L^2(\Gamma)} \leq \|v\|_{\mathcal{H}}$, the other half of the norm equivalence follows. ∎

In summary, with $\mathcal{H}$ defined by (8.11), $a(\cdot,\cdot)$ and $F(\cdot)$ defined by (8.12), and $\|\cdot\|_{\mathcal{H}}$ defined by (8.15), the variational problem (1.1) is a continuous and coercive variational formulation of the Helmholtz IIP. The disadvantage of this formulation is that it is difficult to obtain piecewise-polynomial finite-dimensional subspaces of $\mathcal{H}$ (the difficulty arises from requiring that $\mathcal{L}_k v = 0$). For the analogous formulation of the exterior impedance problem (discussed briefly below), bypassing this difficulty is investigated in [Des97], [Des98], and [BC00]. In particular, from the variational formulation of the exterior impedance problem one can derive a system of integral equations (which can also be derived starting from Green's integral representation); see [BC00, §3.2–3.4], [Des98, §3-4], [Néd01, §3.4.4] for more details.

Remark 8.6 (The analogous formulation for the exterior impedance problem) *We consider the 2-d exterior impedance problem for the homogeneous Helmholtz equation with $\eta = k$, i.e. given $g \in L^2(\Gamma)$, find u satisfying*

$$\mathcal{L}_k u = 0 \quad \text{in } \Omega_+, \qquad \partial_n^+ u + \mathrm{i}k\gamma_+ u = g \quad \text{on } \Gamma,$$

and the Sommerfeld radiation condition (3.4) (the solution to this problem is unique by, e.g., [CK83, Theorem 3.12] [CWGLS12, Corollary 2.9]). The space $\mathcal{H}$ now consists of solutions of $\mathcal{L}_k v = 0$ in Ω_+ with $\partial_n^+ v \pm \mathrm{i}k\gamma_+ v \in L^2(\Gamma)$ and

$$v(r,\theta) = \frac{\mathrm{e}^{\mathrm{i}kr} a(\theta)}{\sqrt{r}} + \frac{\mathrm{e}^{-\mathrm{i}kr} b(\theta)}{\sqrt{r}} + o\left(\frac{1}{\sqrt{r}}\right) \qquad \text{as } r \to \infty,$$

for $a, b \in L^2[0, 2\pi]$, i.e. the space contains both outgoing and incoming solutions (and linear combinations of the two). (Note that we have chosen to use the notation of [BC00] over that in [Des97] and [Des98], since [BC00] share our conventions for "outgoing" and "incoming".) The analogue of the isometry property (8.2) is now

$$4k^2 \|a\|^2_{L^2[0,2\pi]} + \left\|\partial_n^+ v - \mathrm{i}k\gamma_+ v\right\|^2_{L^2(\Gamma)} = 4k^2 \|b\|^2_{L^2[0,2\pi]} + \left\|\partial_n^+ v + \mathrm{i}k\gamma_+ v\right\|^2_{L^2(\Gamma)}; \tag{8.17}$$

see [BC00, Lemma 3.3], [Des97, Lemma 3.1], and [Des98, Theorem 1]. Just as (8.2) gives rise to the quadratic functional (8.10) (which is minimised at the solution of the IIP), (8.17) gives rise to a quadratic functional that is minimised at the solution of the exterior impedance problem.

9 Green's integral representation and boundary integral equations (BIEs)

Summary: G2 with v equal to the fundamental solution of $\mathcal{L}$ and $D = \Omega_\pm$ gives Green's integral representation (from which boundary integral equations can be obtained).

9.1 Green's integral representation

In this section we let $\mathcal{L}$ denote a general linear differential operator. Given such an $\mathcal{L}$, E is a *fundamental solution* for $\mathcal{L}$ if

$$\mathcal{L}_{\mathbf{x}} E(\mathbf{x}, \mathbf{y}) = -\delta(\mathbf{x} - \mathbf{y}), \tag{9.1}$$

where the subscript $\mathbf{x}$ on $\mathcal{L}_{\mathbf{x}}$ indicates that the differentiation is in the $\mathbf{x}$-variable. The δ on the right-hand side of (9.1) is the Dirac delta function, and thus equation (9.1) is understood in a distributional sense. Note that E is *a* fundamental solution, and not *the* fundamental solution, since there can be several different fundamental solutions for a given $\mathcal{L}$ (and we see examples of this below).

We now discuss fundamental solutions for the operator $\Delta + \lambda$ in two and three dimensions (the situation in higher dimensions is similar). We begin with the case $\lambda = -\mu^2$ for $\mu > 0$ (i.e. the modified Helmholtz equation). In 3-d, two solutions of (9.1) with $\mathcal{L} = \Delta - \mu^2$ are

$$\frac{\mathrm{e}^{-\mu|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} \quad \text{and} \quad \frac{\mathrm{e}^{\mu|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} \tag{9.2}$$

(see, e.g., [Sta68, §5.8, Pages 53–55]), illustrating the fact that fundamental solutions are not unique. It is usually convenient to have the fundamental solution tending to zero as $|\mathbf{x} - \mathbf{y}| \to \infty$ (if this is possible), and therefore we choose the first fundamental solution in (9.2).

In 2-d, solutions of (9.1) with $\mathcal{L} = \Delta - \mu^2$ are given by modified Bessel functions (see, e.g., [Sta68, §5.8, Pages 53-55]), and we encounter a similar situation to that in 3-d; one solution tends to infinity as $|\mathbf{x} - \mathbf{y}| \to \infty$, the other tends to zero. Choosing the one that tends to zero, we therefore define $\Phi_\mu(\mathbf{x}, \mathbf{y})$ by

$$\Phi_\mu(\mathbf{x}, \mathbf{y}) := \begin{cases} \frac{1}{2\pi} K_0(\mu|\mathbf{x}-\mathbf{y}|), & d = 2, \\ \dfrac{\mathrm{e}^{-\mu|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}, & d = 3, \end{cases}$$

where K_0 is defined by, e.g., [NIS14, Equations 10.25.3 and 10.27.4].

When $\mathcal{L} = \Delta$ (i.e. the PDE is Laplace's/Poisson's equation) we define $\Phi_0(\mathbf{x}, \mathbf{y})$ by

$$\Phi_0(\mathbf{x}, \mathbf{y}) := \begin{cases} \dfrac{1}{2\pi} \log\left(\dfrac{a}{|\mathbf{x}-\mathbf{y}|}\right), & d = 2, \\ \dfrac{1}{4\pi|\mathbf{x}-\mathbf{y}|}, & d = 3, \end{cases}$$

where $a \in \mathbb{R}$. We immediately see that $\Phi_0(\mathbf{x}, \mathbf{y})$ does not tend to zero as $|\mathbf{x} - \mathbf{y}| \to \infty$ in 2-d; this means that the theory for boundary integral equations (BIEs) for the Laplace equation in 2-d is

more awkward than in 3-d. Furthermore, $\Phi_0(\mathbf{x},\mathbf{y})$ contains the arbitrary parameter a when $d=2$. Usually one lets $a=1$, but considering other values of a can sometimes be useful (for an example in the theory of BIEs see [McL00, Theorem 8.16])

Finally, moving on to the Helmholtz equation, we find that the two solutions of (9.1) in 3-d when $\mathcal{L}=\Delta+k^2$ with $k>0$ are

$$\frac{\mathrm{e}^{\mathrm{i}k|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} \quad \text{and} \quad \frac{\mathrm{e}^{-\mathrm{i}k|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}.$$

Both of these functions decay at the same rate as $|\mathbf{x}-\mathbf{y}|\to\infty$. The first, however, satisfies the outgoing radiation condition (3.4), whereas the second satisfies the incoming radiation condition (4.16), and we therefore choose the first (but see Remark 9.3). We find a similar situation in 2-d; here there are two solutions of (9.1) given in terms of the Hankel functions $H_0^{(1)}$ and $H_0^{(2)}$ (defined by, e.g., [NIS14, Equations 10.4.3, 10.2.5, and 10.2.6]). $H_0^{(1)}$ satisfies the outgoing radiation condition, $H_0^{(2)}$ satisfies the incoming radiation condition, and so we chose the first. We therefore define $\Phi_k(\mathbf{x},\mathbf{y})$ by

$$\Phi_k(\mathbf{x},\mathbf{y}) := \begin{cases} \frac{\mathrm{i}}{4}H_0^{(1)}(k|\mathbf{x}-\mathbf{y}|), & d=2, \\ \dfrac{\mathrm{e}^{\mathrm{i}k|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}, & d=3. \end{cases} \tag{9.3}$$

Green's integral representation arises from applying G2 with $v=\Phi$ and $D=\Omega_\pm$. From now on we restrict our attention to the homogeneous Helmholtz equation, but we note that analogous results hold for Laplace's equation and the modified Helmholtz equation, as well as for the inhomogeneous counterparts of all three equations (see, e.g, [McL00, Theorem 7.5], [Eva98, §2.2.4]).

Recall the notation introduced in §2, namely that Ω_- is a bounded Lipschitz open set such that $\Omega_+ := \mathbb{R}^d \setminus \overline{\Omega_-}$ is connected.

Theorem 9.1 (Green's integral representation for Ω_-) *If $u\in H^1(\Omega_-)\cap C^2(\Omega_-)$ satisfies $\mathcal{L}_k u=0$ then*

$$\int_\Gamma \left(\Phi_k(\mathbf{x},\mathbf{y})\partial_n^- u(\mathbf{y}) - \frac{\partial\Phi_k(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{y})}\gamma_- u(\mathbf{y})\right)\mathrm{d}s(\mathbf{y}) = \begin{cases} u(\mathbf{x}), & \mathbf{x}\in\Omega_-, \\ 0, & \mathbf{x}\in\Omega_+. \end{cases} \tag{9.4}$$

Proof. For $\mathbf{x}\in\Omega_+$, this is an immediate consequence of G2 (without the complex conjugate) (4.5) applied with $v(\cdot)=\Phi_k(\mathbf{x},\cdot)$ and $D=\Omega_-$. For $\mathbf{x}\in\Omega_-$ we apply G2 (4.5) with $D=\Omega_-\setminus B_\varepsilon(\mathbf{x})$ and then let $\varepsilon\to 0$; see [CK83, Theorem 3.1] for the details. ∎

Note that if $u\in H^1(\Omega_-)$ satisfies $\mathcal{L}_k u=0$ then $u\in C^\infty(\Omega_-)$ by Remark 3.6, so the assumption in Theorem 9.1 that $u\in C^2(\Omega_-)$ is not restrictive.

Theorem 9.2 (Green's integral representation for Ω_+) *If $u\in H^1_{\mathrm{loc}}(\Omega_+)\cap C^2(\Omega_+)$ satisfies $\mathcal{L}_k u=0$ and the Sommerfeld radiation condition (3.4) then*

$$-\int_\Gamma \left(\Phi_k(\mathbf{x},\mathbf{y})\partial_n^+ u(\mathbf{y}) - \frac{\partial\Phi_k(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{y})}\gamma_+ u(\mathbf{y})\right)\mathrm{d}s(\mathbf{y}) = \begin{cases} 0, & \mathbf{x}\in\Omega_-, \\ u(\mathbf{x}), & \mathbf{x}\in\Omega_+. \end{cases} \tag{9.5}$$

Proof. This follows in a similar way to Theorem 9.1 (recalling that the normal vector $\mathbf{n}$ points into Ω_+), with the integral at infinity vanishing by Lemma 4.10 (note that if we had used G2 *with* the complex conjugate then this would not happen). ∎

Remark 9.3 (Chase the complex conjugate) *We chose the fundamental solution $\Phi_k(\mathbf{x},\mathbf{y})$ (9.3) on the basis that it should satisfy the outgoing radiation condition (3.4), but this meant that to obtain Green's integral representation in Ω_+ (9.5) we had to use G2 without the complex conjugate (because of Lemma 4.10). If we had chosen the fundamental solution satisfying the* incoming *radiation condition (4.16) and used this in Ω_+ with the version of G2* with *the complex conjugate, then we would have also obtained the representation formulae (9.5).*

In Ω_- it doesn't matter which fundamental solution we use. This is because the difference between the two is a solution of the homogeneous Helmholtz equation, and then G2 implies that the two different representation formulae (one with each fundamental solution) are actually equivalent.

The key point about Green's integral representation is that once you know $\gamma_\pm u$ and $\partial_n^\pm u$ on Γ then you know u in $\Omega_\pm$. Therefore, to find the solution of the IDP, one only needs to find $\partial_n^- u$ on Γ, thereby reducing the problem from one posed in a d-dimensional domain to one posed in a $(d-1)$-dimensional domain. The case of the EDP is similar, except that now there is the additional advantage that, whereas Ω_+ is unbounded, Γ is bounded.

It is convenient to introduce the following notation. The *single-layer potential* $\mathcal{S}_k$ is defined by

$$\mathcal{S}_k\phi(\mathbf{x}) := \int_\Gamma \Phi_k(\mathbf{x},\mathbf{y})\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}), \quad \mathbf{x}\in\mathbb{R}^d\setminus\Gamma, \tag{9.6}$$

and the *double-layer potential* $\mathcal{D}_k$ is defined by

$$\mathcal{D}_k\phi(\mathbf{x}) := \int_\Gamma \frac{\partial\Phi_k(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{y})}\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}), \quad \mathbf{x}\in\mathbb{R}^d\setminus\Gamma.$$

(Observe that both these operators take functions defined on Γ to functions defined on $\mathbb{R}^d\setminus\Gamma$.) Green's integral representation for Ω_+ (9.5) then reads

$$-\mathcal{S}_k\partial_n^+u(\mathbf{x}) + \mathcal{D}_k\gamma_+u(\mathbf{x}) = \begin{cases} 0, & \mathbf{x}\in\Omega_-,\\ u(\mathbf{x}), & \mathbf{x}\in\Omega_+.\end{cases} \tag{9.7}$$

Remark 9.4 (Kupradze's method) *The following formulation of the EDP is based on viewing the first equation in (9.7) as an equation to be solved for ∂_n^+u. That is, given $g_D\in H^{1/2}(\Gamma)$, we find a $\psi\in H^{-1/2}(\Gamma)$ such that*

$$\mathcal{S}_k\psi(\mathbf{x}) = \mathcal{D}_kg_D(\mathbf{x}) \quad \text{for all } \mathbf{x}\in\Omega_- \tag{9.8}$$

and then set $\partial_n^+u=\psi$. Note that this is not a variational problem in the sense of §5, since the trial space is a space of functions on Γ, and the test space consists of all points in Ω_-.

We now show that the equation (9.8) has a unique solution. Assume that $\mathcal{S}_k\psi(\mathbf{x})=0$ for all $\mathbf{x}\in\Omega_-$. Let $u:=\mathcal{S}\psi$. Then $u=0$ in Ω_- and thus $\gamma_-u=0$ on Γ. The jump relations for the single-layer potential ((9.9) below) then imply that $\gamma_+u=0$ on Γ. u is then a solution of the EDP for the Helmholtz equation with zero right-hand side and zero Dirichlet trace. By uniqueness of the solution to this BVP, $u=0$ in Ω_+. Then, by the jump relations (9.9), $\psi=\partial_n^-u-\partial_n^+u=0$.

Although (9.8) is uniquely solvable, non-uniqueness arises when one tries to discretise the equation; see [Mar06, §7.3] and the references therein.

9.2 Boundary integral equations (BIEs)

We now focus on the Helmholtz EDP. We saw above that Green's integral representation means that we only need to find ∂_n^+u, and we now show how to do this using BIEs.

We first need to understand how $\mathcal{S}_k\phi(\mathbf{x})$, $\mathcal{D}_k\phi(\mathbf{x})$, and their normal derivatives behave as $\mathbf{x}\to\Gamma$ from either Ω_- or Ω_+ (in other words, we need to know what the interior and exterior Dirichlet and Neumann traces of $\mathcal{S}_k\phi$ and $\mathcal{D}_k\phi$ are). Define the operators S_k, D_k, D_k', and H_k by

$$S_k\phi(\mathbf{x}) = \int_\Gamma \Phi_k(\mathbf{x},\mathbf{y})\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}), \quad D_k\phi(\mathbf{x}) = \int_\Gamma \frac{\partial\Phi_k(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{y})}\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}), \quad \mathbf{x}\in\Gamma,$$

and

$$D_k'\phi(\mathbf{x}) = \int_\Gamma \frac{\partial\Phi_k(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{x})}\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}), \quad H_k\phi(\mathbf{x}) = \frac{\partial}{\partial n(\mathbf{x})}\int_\Gamma \frac{\partial\Phi_k(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{y})}\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}), \quad \mathbf{x}\in\Gamma.$$

Observe that all these operators take functions defined on Γ to functions defined on Γ. The notation D_k and D_k' expresses the fact that these two operators are adjoint with respect to the real-valued $L^2(\Gamma)$-inner product.

Lots of technical difficulties immediately arise. Indeed, if Γ is not smooth then the integrals defining D_k and D_k' must be understood as Cauchy Principal Values, and even when Γ is smooth H_k has to be understood as a limit (see, e.g., [CWGLS12, Equation (2.36)], [McL00, Theorem 7.4 (iii)], [SS11, §3.3.4]); we ignore all these difficulties here.

With the definitions of S_k, D_k, D'_k, and H_k above, the *jump relations* for $\mathcal{S}_k$ are

$$\gamma_\pm \mathcal{S}_k = S_k, \quad \partial_n^\pm \mathcal{S}_k = \mp\frac{1}{2}I + D'_k, \tag{9.9}$$

and those for $\mathcal{D}_k$ are

$$\gamma_\pm \mathcal{D}_k = \pm\frac{1}{2}I + D_k, \quad \partial_n^\pm \mathcal{D}_k = H_k. \tag{9.10}$$

Taking the exterior Dirichlet trace of the second equation in (9.7) (or, equivalently, taking the interior Dirichlet trace of the first equation in (9.7)) and using the jump relations (9.9) and (9.10), we obtain the integral equation

$$S_k \partial_n^+ u = \left(-\frac{1}{2}I + D_k\right) g_D \tag{9.11}$$

(where we have used the fact that $\gamma_+ u = g_D$).

Similarly, taking the exterior Neumann trace of the second equation in (9.7) (or, equivalently, taking the interior Neumann trace of the first equation in (9.7)), we obtain the integral equation

$$\left(\frac{1}{2}I + D'_k\right) \partial_n^+ u = H_k g_D. \tag{9.12}$$

Unfortunately, neither of the equations (9.11) and (9.12) is uniquely solvable for all $k > 0$. Indeed, equation (9.11) does not have a unique solution for all $k > 0$ because the Neumann trace of the solution of the Helmholtz IDP also satisfies an integral equation of the form $S_k \partial_n^- u = \ldots$ (this can be seen by taking the interior Dirichlet trace of the first equation in (9.4)); therefore, since the Helmholtz IDP does not have a unique solution when k^2 is a Dirichlet eigenvalue of the Laplacian, neither do integral equations of the form $S_k \phi = \ldots$. Similarly, the Dirichlet trace of the solution of the Helmholtz interior Neumann problem satisfies an integral equation of the form $(\frac{1}{2}I + D_k)\gamma_- u = \ldots$ and since $(\frac{1}{2}I + D_k)$ is, roughly speaking, adjoint to $(\frac{1}{2}I + D'_k)$, this implies that integral equations of the form $(\frac{1}{2}I + D'_k)\phi = \ldots$ do not have a unique solution when k^2 is a Neumann eigenvalue of the Laplacian. (In the literature, the particular values of k for which a boundary integral equation is not uniquely solvable are often called "spurious frequences" or "spurious resonances" of that integral equation.)

There are at least four different ways around this non-uniqueness, with [Mar06, §6.8] providing a good overview.

1. Supplement one of the integral equations with additional equations, such as the null-field equations (discussed in §10.5) or (9.8) evaluated at certain points in Ω_-; see, e.g., [Mar06, §6.11]
2. Modify the fundamental solution; see, e.g., [Mar06, §6.9], [CK83, §3.6].
3. Modify the integral representation (this can be understood as using a particular *indirect* BIE method; see Remark 9.5).
4. Combine the two integral equations (9.11) and (9.12).

The third and fourth options have won out, at least from the point of view of mathematicians, and the relationship between *direct* and *indirect* methods (defined and discussed in Remark 9.5) means that these two options are, in some sense, equivalent. We now describe the fourth option.

Subtracting $\mathrm{i}\eta$ times (9.11) from (9.12), we obtain

$$A'_{k,\eta} \partial_n^+ u = \left[H_k - \mathrm{i}\eta\left(-\frac{1}{2}I + D_k\right)\right] g_D, \tag{9.13}$$

where

$$A'_{k,\eta} := \frac{1}{2}I + D'_k - \mathrm{i}\eta S_k$$

is the so-called *combined field* or *combined potential* integral equation.

We saw above that the integral operators in each of the equations (9.11) and (9.12) could also be used to solve an interior BVP (for (9.11) the BVP was the IDP, and for (9.12) the BVP was the interior Neumann problem [4]). The interior BVP for $A'_{k,\eta}$ is the Helmholtz IIP; see [CWGLS12, Theorem 2.30]. By Theorem 3.5, the solution of the IIP is unique for all $k > 0$ (when $\eta \in \mathbb{R} \setminus \{0\}$) and this fact can be used to prove that the integral operator $A'_{k,\eta}$ is invertible for all $k > 0$ when $\eta \in \mathbb{R} \setminus \{0\}$ [CWGLS12, Theorem 2.27], thus overcoming the problem of non-uniqueness.

9.2.1 Variational formulations of the EDP using boundary integral equations

We consider two different variational formulations of the EDP, the first based on the integral equation (9.11) and the second based on the equation (9.13). In this section we slightly abuse notation, and use $\langle \cdot, \cdot \rangle_\Gamma$ to denote *both* the sesquilinear form on $H^{-1/2}(\Gamma) \times H^{1/2}(\Gamma)$ discussed in Remark 4.3 *and* its complex conjugate, which can be thought of as a sesquilinear form on $H^{1/2}(\Gamma) \times H^{-1/2}(\Gamma)$ after interchanging the arguments. Which one we're using will be clear from the arguments of the form, but the important point is that we always assume $\langle \cdot, \cdot \rangle_\Gamma$ is *sesquilinear*.

Formulation based on (9.11) (involving S_k). Define

$$a(\phi, \psi) := \langle S_k \phi, \psi \rangle_\Gamma \quad \text{and} \quad F(\psi) := \left\langle \left(-\frac{1}{2} I + D_k \right) g_D, \psi \right\rangle_\Gamma . \tag{9.14}$$

By the mapping properties $S_k : H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$ and $D_k : H^{1/2}(\Gamma) \to H^{1/2}(\Gamma)$ [CWGLS12, Theorem 2.17], [McL00, Theorem 7.1], $a(\cdot, \cdot)$ is a sesquilinear form on $H^{-1/2}(\Gamma) \times H^{-1/2}(\Gamma)$ and $F(\cdot)$ is an anti-linear functional on $H^{-1/2}(\Gamma)$. The continuity of the sesquilinear form $\langle \cdot, \cdot \rangle_\Gamma : H^{1/2}(\Gamma) \times H^{-1/2}(\Gamma) \to \mathbb{C}$ and the continuity of $S_k : H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$ imply that $a(\cdot, \cdot)$ is continuous with $\|a\| \leq \|S_k\|_{H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)}$.

When $k = 0$, $a(\cdot, \cdot)$ is coercive; indeed

$$\langle S_0 \phi, \phi \rangle_\Gamma \gtrsim \|\phi\|^2_{H^{-1/2}(\Gamma)} \qquad \text{for all } \phi \in H^{-1/2}(\Gamma) \tag{9.15}$$

by, e.g., [SS11, Theorem 3.5.3], [McL00, Corollary 8.13 and Theorem 8.16], [Ste08, Theorems 6.22 and 6.23], [SKS15, §1.4]. We saw above that, when k^2 is a Dirichlet eigenvalue of the Laplacian, S_k is not invertible, and thus $a(\cdot, \cdot)$ cannot be coercive for these particular k. However, the difference $S_k - S_0$ is compact (this follows from the bounds in, e.g., [CWGLS12, Equation (2.25)]), and thus S_k is a compact perturbation of a coercive operator. Therefore, Theorem 5.18 implies that, when k^2 is not a Dirichlet eigenvalue of the Laplacian, the variational problem (1.1) with $a(\cdot, \cdot)$ and $F(\cdot)$ defined by (9.14) and $\mathcal{H} = H^{1/2}(\Gamma)$ has a unique solution that depends continuously on the data (i.e. the properties K1 and K2 hold).

Formulation based on (9.13) (involving $A'_{k,\eta}$). The operator $A'_{k,\eta} : H^s(\Gamma) \to H^s(\Gamma)$ for $|s| \leq 1/2$ [CWGLS12, Theorem 2.27]. Since the unknown Neumann trace $\partial_n^+ u$ is in $H^{-1/2}(\Gamma)$, it is natural to pose equation (9.13) in the space $H^{-1/2}(\Gamma)$. That is, we define

$$a(\phi, \psi) := \left(A'_{k,\eta} \phi, \psi \right)_{H^{-1/2}(\Gamma)} \quad \text{and} \quad F(\psi) = \left(\left[H_k - \mathrm{i}\eta \left(-\frac{1}{2} I + D_k \right) \right] g_D, \psi \right)_{H^{-1/2}(\Gamma)}, \tag{9.16}$$

where $(\cdot, \cdot)_{H^{-1/2}(\Gamma)}$ denotes the $H^{-1/2}(\Gamma)$ inner product.

Since $A'_{k,\eta} : H^{-1/2}(\Gamma) \to H^{-1/2}(\Gamma)$, $a(\cdot, \cdot)$ is a continuous sesquilinear form on $H^{-1/2}(\Gamma) \times H^{-1/2}(\Gamma)$ (and similar mapping properties imply that $F(\cdot)$ is an anti-linear functional on $H^{-1/2}(\Gamma)$). Turning to coercivity, we have that $A'_{0,0}$ is a compact perturbation of a coercive operator (see Remark 9.6 below), and since the difference $A'_{k,\eta} - A'_{0,0}$ is compact, $A'_{k,\eta}$ itself is then a compact perturbation of a coercive operator. Since $A'_{k,\eta}$ is injective for all $k > 0$ (as long as $\eta \in \mathbb{R} \setminus \{0\}$), Theorem 5.18 implies that the variational problem (1.1) with $a(\cdot, \cdot)$ and $F(\cdot)$ defined

[4] More precisely, we sketched above how the "adjoint" of $(\frac{1}{2} I + D'_k)$, namely $(\frac{1}{2} I + D_k)$, could be used to solve the interior Neumann problem, but the same is true for $(\frac{1}{2} I + D'_k)$ itself; see [CWGLS12, Table 2.1].

by (9.16) and $\mathcal{H} = H^{-1/2}(\Gamma)$ has a unique solution for all $k > 0$ (i.e. the properties K1 and K2 hold).

The drawback to posing (9.13) as an equation in $H^{-1/2}(\Gamma)$ is that the $H^{-1/2}(\Gamma)$-inner product is difficult to implement practically. If the Dirichlet trace g_D is in $H^1(\Gamma)$ (as it is in the case of plane-wave or point-source scattering) the right-hand side of (9.13) is in $L^2(\Gamma)$ [CWGLS12, Theorem 2.12]. We can therefore consider the integral equation (9.13) as an equation in $L^2(\Gamma)$, i.e. we define

$$a(\phi,\psi) := \left(A'_{k,\eta}\phi,\psi\right)_{L^2(\Gamma)} \quad \text{and} \quad F(\psi) = \left(\left[H_k - \mathrm{i}\eta\left(-\frac{1}{2}I + D_k\right)\right] g_D, \psi\right)_{L^2(\Gamma)}$$

and let $\mathcal{H} = L^2(\Gamma)$. The corresponding Galerkin method is then much easier to implement than the one in $H^{-1/2}(\Gamma)$.

Since $A'_{k,\eta} : L^2(\Gamma) \to L^2(\Gamma)$, $a(\cdot,\cdot)$ is continuous. Turning to coercivity, we have that, when Γ is C^1, the operators S_k and D'_k are compact as mappings from $L^2(\Gamma)$ to itself (for S_k this follows from the mapping properties given in, e.g., [CWGLS12, Theorem 2.17], and for D_k this is proved in [FJR78, Theorem 1.2]). Therefore, when Γ is C^1, $A'_{k,\eta}$ is a compact perturbation of a coercive operator, namely the identity. The question of whether $A'_{k,\eta}$ is a compact perturbation of a coercive operator in $L^2(\Gamma)$ when Γ is Lipschitz is still open. However under certain geometric restrictions $A'_{k,\eta}$ itself is coercive on $L^2(\Gamma)$; see [SKS15, Theorem 1.2].

Remark 9.5 (Indirect boundary integral equations) *We focused here on using Green's integral representation to obtain BIEs, and this is often called the* direct *method. Alternatively one can seek the solution of a BVP as either a single- or double-layer potential (or a linear combination of the two) and take appropriate traces of these potentials to reformulate the BVP as an integral equation on Γ; this is often called the* indirect *method. The BIEs obtained using the indirect method involve, roughly speaking, the adjoints of the operators in the BIEs obtained using the direct method; see, e.g., [CWGLS12, §2.5–2.6].*

Remark 9.6 (Coercivity up to a compact perturbation of the operator $A'_{k,\eta}$ on $H^{-1/2}(\Gamma)$) *There are a few subtleties with considering $A'_{k,\eta}$ as a operator on $H^{-1/2}(\Gamma)$ that we glossed over above. The standard way to realise the $H^{-1/2}(\Gamma)$-inner product and norm is via the Laplace single-layer potential. Indeed, by the coercivity of S_0 (9.15), $\|\phi\|^2_{S_0} := \langle\phi, S_0\phi\rangle_\Gamma$ is an equivalent norm on $H^{-1/2}(\Gamma)$. We then let*

$$\widetilde{a}(\phi,\psi) := \langle A'_{k,\eta}\phi, S_0\psi\rangle_\Gamma \quad \text{and} \quad \widetilde{F}(\psi) = \left\langle\left[H_k - \mathrm{i}\eta\left(-\frac{1}{2}I + D_k\right)\right] g_D, S_0\psi\right\rangle_\Gamma,$$

and these give rise to a variational problem on $H^{-1/2}(\Gamma)$. Using G1 (4.4), the single-layer jump relations (9.9), and the Poincaré-type inequality (6.16), Elschner showed that

$$\left\langle\left(\frac{1}{2}I + D'_0\right)\phi, S_0\phi\right\rangle_\Gamma \gtrsim \|\phi\|^2_{H^{-1/2}(\Gamma)} - \|S_0\phi\|^2_{L^2(\Gamma)} \quad \text{for all } \phi \in H^{-1/2}(\Gamma) \tag{9.17}$$

[Els92, Proposition A1]. [5] *Therefore, there exists a $c > 0$ such that*

$$\left\langle\left(\frac{1}{2}I + D'_0 + cS_0\right)\phi, S_0\phi\right\rangle_\Gamma \gtrsim \|\phi\|^2_{H^{-1/2}(\Gamma)} \quad \text{for all } \phi \in H^{-1/2}(\Gamma)$$

and then, since S_0, $D'_k - D'_0$, and S_k are compact operators on $H^{-1/2}(\Gamma)$, the operator associated with $\widetilde{a}(\cdot,\cdot)$ is a compact perturbation of a coercive operator.

[5] More precisely, Elschner showed that (9.17) holds with $\|S_0\phi\|^2_{L^2(\Gamma)}$ replaced by $|\int_\Gamma S_0\phi|^2$, but (9.17) follows from this by using the Cauchy-Schwarz inequality (2.5).

10 The null-field method and the Fokas transform method

Summary: G2 with v a separable solution of $\mathcal{L}v = 0$.

The null-field method is this idea applied to exterior Helmholtz problems with v a separable solution in polar coordinates. The Fokas transform method can be understood as this idea applied with v a separable solution in cartesian coordinates; it turns out that the vs in cartesian coordinates are most suited to interior BVPs, and thus the Fokas transform method has been applied to interior BVPs for the Laplace, modified Helmholtz, and Helmholtz equations. An extension of this method has recently been successfully applied to exterior BVPs for the modified Helmholtz equation; see Remark 10.3 below.

We begin by introducing these ideas in their simplest possible setting, namely the IDP for the modified Helmholtz equation.

10.1 Interior Dirichlet problem for the modified Helmholtz equation

We consider the homogeneous interior Dirichlet problem, i.e. Definition 3.1 with $\mathcal{L} = \mathcal{L}_\mu$ and $f = 0$. Let

$$\mathcal{R} := \Big\{ v \in H^1(\Omega, \Delta) : \mathcal{L}_\mu v = 0 \Big\}. \tag{10.1}$$

If u and $v \in \mathcal{R}$, then G2 (4.6) with $D = \Omega$ implies that

$$\int_\Gamma \partial_n u \, \overline{\gamma v} = \int_\Gamma \gamma u \, \overline{\partial_n v}. \tag{10.2}$$

(Note that the integrals in (10.2), as well as all the other integrals over Γ in this section, should be understood as duality pairings between $H^{-1/2}(\Gamma)$ and $H^{1/2}(\Gamma)$ unless we add the condition that $\partial_n v \in L^2(\Gamma)$ to the space $\mathcal{R}$.)

If we define

$$a(u, v) := \int_\Gamma \partial_n u \, \overline{\gamma v} \quad \text{and} \quad F(v) := \int_\Gamma g_D \overline{\partial_n v}, \tag{10.3}$$

then the equation (10.2) gives rise to the following variational formulation of the IDP: given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } u \in \mathcal{R} \text{ such that} \quad a(u, v) = F(v) \quad \text{for all } v \in \mathcal{R}.} \tag{10.4}$$

We now seek a variational problem where the Hilbert space consists of functions on Γ (as opposed to functions in Ω_-). Define $P_{\mathrm{DtN}} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ to be the Dirichlet-to-Neumann map. Of course, given $\phi \in H^{1/2}(\Gamma)$, finding $P_{\mathrm{DtN}}\phi$ is equivalent to solving the IDP, and so if the variational problem involves the operator P_{DtN} then we need to check at the end that it is practical to implement. We now define

$$b(\phi, \psi) := \int_\Gamma P_{\mathrm{DtN}}\phi \, \overline{\psi}, \tag{10.5}$$

so that $a(u, v) = b(\gamma u, \gamma v)$. The variational problem (10.4) therefore becomes

$$\boxed{\text{find } u \in \mathcal{R} \text{ such that} \quad b(\gamma u, \gamma v) = F(v) \quad \text{for all } v \in \mathcal{R}.} \tag{10.6}$$

The next lemma concerns the properties of $b(\cdot, \cdot)$ as a sesquilinear form on $H^{1/2}(\Gamma) \times H^{1/2}(\Gamma)$. Note that everything so far in this section has only relied on G2, which holds for each of the Laplace, modified Helmholtz, and Helmholtz equations. The results of the lemma, however, are specific to the modified Helmholtz equation.

Theorem 10.1 (Continuity and coercivity of $b(\cdot, \cdot)$ for modified Helmholtz)
(i) For any $\phi, \psi \in H^{1/2}(\Gamma)$,

$$\big|b(\phi, \psi)\big| \le \|P_{\mathrm{DtN}}\|_{H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)} \, \|\phi\|_{H^{1/2}(\Gamma)} \, \|\psi\|_{H^{1/2}(\Gamma)}.$$

(ii) Given $\mu_0 > 0$, there exists a C, independent of μ (but dependent on μ_0), such that

$$\big|b(\psi, \psi)\big| \ge C \, \|\psi\|^2_{H^{1/2}(\Gamma)}$$

for all $\mu \ge \mu_0$.

Proof. Continuity is straightforward (note that, given $\mu_0 > 0$, $\|P_{\mathrm{DtN}}\|_{H^{1/2}(\Gamma)\to H^{-1/2}(\Gamma)} \lesssim \mu$ for $\mu \geq \mu_0$; see [Say13, Proposition 2.5.2]). For coercivity, given $\psi \in H^{1/2}(\Gamma)$ there exists a unique $u \in \mathcal{R}$ such that $\gamma u = \psi$. Then, using G1 (4.4),

$$b(\psi,\psi) = \int_\Gamma \partial_n u\, \overline{\gamma u} = \int_\Omega |\nabla u|^2 + \mu^2 |u|^2,$$

and the result follows from the bound on the trace operator (2.1). ∎

We now consider the alternative variational problem, given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } \phi \in H^{1/2}(\Gamma) \text{ such that } \; b(\phi,\psi) = G(\psi) \quad \text{for all } \psi \in H^{1/2}(\Gamma),} \tag{10.7}$$

where

$$G(\psi) := \int_\Gamma g_D \overline{P_{\mathrm{DtN}}\psi}. \tag{10.8}$$

Since $G(\gamma v) = F(v)$ when $v \in \mathcal{R}$, it is straightforward to show that if u is a solution of the variational problem (10.6) then γu is a solution of the variational problem (10.7). Theorem 10.1 and the Lax-Milgram theorem (Theorem 5.14) imply that the solution of (10.7) exists and is unique, and therefore the *only* solution of (10.7) is $\phi = \gamma u$. We can therefore abandon the variational problem (10.6) and focus on (10.7).

Stepping back a moment, it may seem like we have achieved nothing. Indeed, the solution to the variational problem (10.7) is $\phi = \gamma u$, but we're given γu in the boundary condition and instead want to find $\partial_n u = P_{\mathrm{DtN}}(\gamma u)$. This shows us that when we discretise (10.7) we need to chose finite-dimensional subspaces where it is straightforward to apply P_{DtN}, since after we have computed an approximation to ϕ, ϕ_N, we want to approximate $\partial_n u$ by $P_{\mathrm{DtN}}\phi_N$.

Discretising the variational problem (10.7). The discussion in the previous paragraph indicated that, although we could in principle choose *any* finite-dimensional subspace of $H^{1/2}(\Gamma)$ when applying the Galerkin method to (10.7), we need to choose basis functions for which it is easy to apply P_{DtN}. We also come to this conclusion when we observe that to compute $b(\phi,\psi)$ and $F(\psi)$ for given $\phi,\psi \in H^{1/2}(\Gamma)$ we need to be able to compute $P_{\mathrm{DtN}}\phi$ and $P_{\mathrm{DtN}}\psi$.

This situation is completely analogous to the situation encountered with the UWVF in §8.2, where we needed to choose subspaces in which it is easy to apply the operator F_j (which involved calculating the "impedance-to-impedance" map).

Let $\mathcal{R}_N$ be a finite-dimensional subspace of $\mathcal{R}$ containing explicit solutions of $\mathcal{L}_\mu v = 0$ obtained by separation of variables. For example, the function

$$v(x,y) = \mathrm{e}^{m_1 x + m_2 y} \tag{10.9}$$

satisfies $\mathcal{L}_\mu v = 0$ in 2-d if $m_1^2 + m_2^2 - \mu^2 = 0$. A natural parametrisation of this last equation is $m_1 = \mu\cos\theta$, $m_2 = \mu\sin\theta$. Letting $\nu = \mathrm{e}^{\mathrm{i}\theta}$, we find that

$$m_1 = \frac{\mu}{2}\left(\nu + \frac{1}{\nu}\right) \quad \text{and} \quad m_2 = \frac{\mu}{2\mathrm{i}}\left(\nu - \frac{1}{\nu}\right),$$

and thus (10.9) becomes

$$v(x,y) = \exp\left[\frac{\mu}{2}\left(\nu + \frac{1}{\nu}\right)x + \frac{\mu}{2\mathrm{i}}\left(\nu - \frac{1}{\nu}\right)y\right] \tag{10.10}$$

where ν can be any complex number; we therefore write $v(x,y) = v(x,y;\nu)$.

With $\mathcal{R}_N$ such a subspace, let

$$\mathcal{Q}_N := \{\gamma v_N : v_N \in \mathcal{R}_N\}, \tag{10.11}$$

and note that $\mathcal{Q}_N \subset H^{1/2}(\Gamma)$ since $\mathcal{R}_N \subset \mathcal{R} \subset H^1(\Omega)$. Observe also that $\mathcal{Q}_N$ is a *global basis* of $H^{1/2}(\Gamma)$, i.e. each basis function has support on all of Γ. If $\phi_N \in \mathcal{Q}_N$ then there exists a $v_N \in \mathcal{R}_N$

such that $\phi_N = \gamma v_N$. Then $P_{\mathrm{DtN}}\phi_N = \partial_n v_N$, and this can be found easily from the explicit expression for v_N.

The Galerkin method for the variational problem (10.7) therefore reads: given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } \phi_N \in \mathcal{Q}_N \text{ such that } b(\phi_N, \psi_N) = G(\psi_N) \quad \text{for all } \psi_N \in \mathcal{Q}_N.} \tag{10.12}$$

Existence, uniqueness, and quasi-optimality of ϕ_N (i.e. property K3) follow from using Theorem 10.1 and the Lax-Milgram theorem (Theorem 5.14). Once we have found ϕ_N, an approximation of $\partial_n u$ is given by $P_{\mathrm{DtN}}\phi_N$ (which, from above, is straightforward to find).

The key point to take away from this discussion on discretisations is that (similar to the case of the UWVF) the Galerkin method applied to the variational formulation (10.7) is only practical when the finite-dimensional subspaces consist of (traces of) explicit solutions of $\mathcal{L}_\mu v = 0$.

10.2 The Fokas transform method for the IDP for the modified Helmholtz equation

We now discuss the Fokas transform method. We focus on the IDP for the modified Helmholtz equation, but we also discuss how these ideas have been applied to interior BVPs for the Laplace and Helmholtz equations, and to exterior BVPs for the modified Helmholtz equation.

The global relation. The Fokas transform method for the IDP for the modified Helmholtz equation is based on the variational problem (10.6), i.e. the fact that if u is the solution to the IDP then

$$\int_\Gamma \partial_n u\, \overline{\gamma v} = \int_\Gamma g_D\, \overline{\partial_n v} \quad \text{for all } v \in \mathcal{R}.$$

Furthermore, one restricts attention to functions in $\mathcal{R}$ of the form (10.10), i.e.

$$v(\mathbf{x};\nu) := \exp\left[\frac{\mu}{2}\left(\nu + \frac{1}{\nu}\right)x + \frac{\mu}{2\mathrm{i}}\left(\nu - \frac{1}{\nu}\right)y\right] \tag{10.13}$$

for $\mathbf{x} \in \Omega_-$ and $\nu \in \mathbb{C} \setminus \{0\}$. The crux of the Fokas method is therefore the relation

$$\int_\Gamma \partial_n u\, \overline{\gamma v(\cdot\,;\nu)} = \int_\Gamma g_D\, \overline{\partial_n v(\cdot\,;\nu)} \quad \text{for all } \nu \in \mathbb{C} \setminus \{0\}, \tag{10.14}$$

with this one-parameter family of equations then called the *global relation*. (The word "global" reflects the fact that (10.14) contains information about certain integrals of the unknown Neumann trace, as opposed to information about the Neumann trace itself.)

With the notation introduced in the previous section (i.e. $b(\cdot,\cdot)$ and $G(\cdot)$ are defined by (10.5) and (10.8) respectively), solving the global relation (10.14) can be written as:

$$\boxed{\text{find } \phi \in H^{1/2}(\Gamma) \text{ such that } b(\phi, \gamma v(\cdot\,;\nu)) = G(\gamma v(\cdot\,;\nu)) \quad \text{for all } \nu \in \mathbb{C}} \tag{10.15}$$

and then set $\partial_n u = P_{\mathrm{DtN}}\phi$.

A natural question is then, is the set of equations (10.14) sufficient to determine $\partial_n u$? In [Ash12] it is proved that if there exists a χ such that

$$\int_\Gamma \chi\, \overline{\gamma v(\cdot\,;\nu)} = \int_\Gamma g_D\, \overline{\partial_n v(\cdot\,;\nu)} \quad \text{for all } \nu \in \mathbb{C} \setminus \{0\}, \tag{10.16}$$

then there exists a solution of the IDP for the modified Helmholtz equation with $\partial_n u = \chi$; the proof considers smooth convex domains and continuous functions, but can, in principle, be extended to non-smooth (but still convex) domains and functions in appropriate Sobolev spaces. Since the solution of the IDP for the modified Helmholtz equation is unique, there is then only one χ satisfying (10.16) (and this is equal to the Neumann trace of the solution).

For certain domains one can solve global relation explicitly; i.e. one can *either* find the unknown Neumann trace $\partial_n u$ *or* find the solution u in Ω_- (for this second option one also needs to use the analogue of Green's integral representation in the Fourier, or spectral, space); for more information see the review articles [FS12], [DTV14] and the book [Fok08]. For general domains, however, the global relation must be solved numerically.

Solving the global relation numerically (i.e. discretising the variational problem (10.15)). The global relation (10.14) holds for all $\nu \in \mathbb{C}$, and the first step in solving this equation numerically is to enforce that (10.14) at a discrete set of points $(\nu_j)_{j=1}^N$ (we discuss the question of how to choose the points below).

Roughly speaking we then have two options.

1. The first option is to *define* $\mathcal{Q}_N$ *so that the Galerkin equations* (10.12) *are a discretisation of* (10.15). Indeed, recalling that $\mathcal{R}_N$ denotes a finite-dimensional subspace of $\mathcal{R}$, and $\mathcal{Q}_N$ consists of Dirichlet traces of functions in $\mathcal{R}_N$, we let

$$\mathcal{R}_N := \big\{v(\cdot\,;\nu_j),\, j = 1,\dots,N\big\} \quad \text{and} \quad \mathcal{Q}_N := \big\{\gamma v(\cdot\,;\nu_j),\, j = 1,\dots,N\big\}, \tag{10.17}$$

   where $(\nu_j)_{j=1}^N$ are the points at which the global relation holds. With this particular $\mathcal{Q}_N$, the Galerkin equations (10.12) are then a discretisation of the variational problem (10.15), and existence and uniqueness of a quasi-optimal Galerkin solution (property K3 of §5) can then be obtained using Theorem 10.1 and the Lax-Milgram theorem (Theorem 5.14).

2. The second option is to *use different trial and test spaces*; in particular we can use the "operator-adapted" space $\mathcal{Q}_N$ in (10.17) for the test space, but use a non-operator-adapted space for the trial space. (This allows us to use *local* trial spaces, i.e. trial spaces whose basis functions are supported only on parts of Γ.) The disadvantage of this option compared to the first is that to obtain K3 we would need to verify the discrete inf-sup condition (5.10), and this is difficult.

All the numerical implementations of the Fokas method in the literature have chosen the second option, and so we now discuss this option further.

Since the trial space no longer consists of traces of solutions of $\mathcal{L}_\mu v = 0$, it no longer makes sense to have γu as the unknown of the variational problem (since computing $P_{\mathrm{DtN}}\phi_N$, where ϕ_N is the approximation to ϕ, will no longer be straightforward). We therefore consider the equivalent variational problem, given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } \chi \in H^{-1/2}(\Gamma) \text{ such that } \langle \chi, \gamma v(\cdot\,;\nu)\rangle_\Gamma = G(\gamma v(\cdot\,;\nu)) \quad \text{for all } \nu \in \mathbb{C},} \tag{10.18}$$

where $\langle\cdot,\cdot\rangle_\Gamma$ is the sesquilinear form on $H^{-1/2}(\Gamma) \times H^{1/2}(\Gamma)$ discussed in Remark 4.3 (i.e. when both arguments are in $L^2(\Gamma)$, $\langle\cdot,\cdot\rangle_\Gamma$ is the L^2-inner product). The result in [Ash12] discussed above shows that if χ satisfies (10.18) then $\chi = \partial_n u$, where u is the solution of the IDP for the modified Helmholtz equation.

To discretise (10.18) one needs to

1. choose a finite dimensional subspace $\mathcal{V}_M \subset H^{-1/2}(\Gamma)$ (with dimension M),
2. choose points $(\nu_j)_{j=1}^N \in \mathbb{C}$,
3. solve the problem

$$\boxed{\text{find } \chi_M \in \mathcal{V}_M \text{ such that } \langle \chi_M, \gamma v(\cdot\,;\nu_j)\rangle_\Gamma = G(\gamma v(\cdot\,;\nu_j)) \quad \text{for } j = 1,\dots,N.}$$

Observe that, since the trial and test spaces are now independent, one can take the dimension of the test space, N, larger than the dimension of the trial space, M, i.e. create an overdetermined system that can be solved by, e.g., least squares.

Table 1 gives an overview of the different implementations in the literature of the preceding three steps. To understand this table we note the following:

- All the investigations consider BVPs in 2-d polygons.
- L stands for "Laplace", MH stands for "modified Helmholtz", and H stands for "Helmholtz". L* means that the paper considers the $\overline{\partial}$-equation discussed in Remark 10.2 below (which can be understood as a reduction of Laplace's equation).

- Regarding the location of the $\{\nu_j\}_{j=1}^N$, the rays on which the $\{\nu_j\}_{j=1}^N$ are chosen in [FFX04], [SFFS08], [SSF10], and [FIS15] are specified by the exponentials appearing in the global relation (which in turn are specified by the geometry of the polygon); see, e.g., [SSF10, Equation 2.9 and Remark 2.3].
- For the definition of the *Halton notes* see [FF11, Appendix A], [DF14, Page 5] and the references therein. These are quasi-random points in the complex plane, with the particular feature that different points never get very close to each other.
- Regarding the basis functions in the trial space, "Fourier" means that a Fourier basis was used on each side of the polygon, and "polynomial" means that a polynomial basis (consisting of either Chebychev or Legendre polynomials) was used on each side.
- The only works on the numerical implementation of the Fokas method not included in Table 1 are the five papers [SPS07], [SFPS09], [SSP12], [FL15], and [Ash13]. The first three of these are, in some sense, continuations of the method in [SFFS08] (they analyse the properties of the system of linear equations for specific geometries). The fourth paper, [FL15], concerns the exterior Dirichlet problem for the modified Helmholtz equation and is discussed in Remark 10.3. The fifth paper, [Ash13], concerns interior problems, but uses a completely different variational formulation than those described above; see Remark 10.4.

	PDEs considered	Dimension of test and trial spaces	Location of $\{\nu_j\}_{j=1}^N$	Basis functions in trial space
[FFX04]	L*	$N = M$	on rays in $\mathbb{C}$	Fourier
[SFFS08]	L*	$N = M$	on rays in $\mathbb{C}$	Fourier, polynomial
[SSF10]	L, MH	$N = M$	on rays in $\mathbb{C}$	Fourier, polynomial
[FF11]	L	$N > M$	at Halton nodes	polynomial
[DF14]	MH, H	$N > M$	at Halton nodes	polynomial
[FIS15]	MH	$N > M$	on rays in $\mathbb{C}$	polynomial

Table 1: Overview of the different numerical implementations of the Fokas method (see the explanatory notes in the text)

Remark 10.2 (The "reduction" of Laplace's equation to the $\overline{\partial}$-equation) *The papers [FFX04], [SFFS08], [SPS07], [SFPS09], [SSP12], and [Ash13] all effectively consider solving BVPs involving the operator $\partial/\partial\overline{z}$. The simplest such BVP is the following: let Ω be a bounded domain in $\mathbb{C}$ with boundary Γ. Given $g : \Gamma \to \mathbb{C}$, find $U : \mathbb{C} \to \mathbb{C}$ such that $\Re U = g$ on Γ, and*

$$\frac{\partial U}{\partial \overline{z}} = 0 \quad \text{in } \Omega \tag{10.19}$$

(i.e. U is analytic; see [AF03, §2.6.3]). Note that it is sufficient to find $\Im U$ on Γ, since once one knows U on Γ one can then find U in Ω using Cauchy's integral formula. (Note the similarity with finding u in Ω using Green's integral representation once one knows both γu and $\partial_n u$.)

How is this relevant to Laplace's equation? Identifying $\mathbb{R}^2$ with $\mathbb{C}$ (and thus writing $z = x+iy$), we have that

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y}\right) \quad \text{and} \quad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)$$

[AF03, §2.6.3], and thus

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = 4\frac{\partial^2}{\partial z \partial \overline{z}}.$$

Therefore, if u satisfies Laplace's equation then

$$\frac{\partial}{\partial \overline{z}}\left(\frac{\partial u}{\partial z}\right) = 0.$$

In some physical applications of Laplace's equation one is interested in finding the gradient of u, i.e. finding both $\partial u/\partial x$ and $\partial u/\partial y$ (for example, in fluid dynamics, the velocity potential for two-dimensional, irrotational, incompressible flow satisfies Laplace's equation, and then the velocity is given by the gradient of the potential; see, e.g., [AF03, §2.1.2]). Since $\partial u/\partial z = (\partial u/\partial x - \mathrm{i}\partial u/\partial y)/2$, if u is real, one can find ∇u from $\partial u/\partial z$.

In summary, for some applications it is sufficient to find the gradient of the solution of Laplace's equation, and this can be found by solving the $\overline{\partial}$-equation (10.19).

Remark 10.3 (Numerical implementation of the Fokas method for exterior problems) *The paper [FL15] contains a preliminary numerical implementation of the Fokas transform method for the Dirichlet problem for the modified Helmholtz equation in the exterior of a square. The appropriate global relation is formed by applying G2 in the exterior domain with v given by (10.13). However, since these vs do not tend to zero as $r := |\mathbf{x}| \to \infty$, there is a non-zero contribution from the integral at infinity (containing the coefficient of the first term in the asymptotic expansion of the solution as $r \to \infty$). The presence of this extra unknown means that one must supplement the global relation with an additional equation, and [FL15] uses the analogue of the boundary integral equation (9.11) in the transform space (i.e. the equation involves the transform of the unknown Neumann trace as opposed to the Neumann trace itself). Indeed, Equation (4.1) in [FL15] is the transform-space analogue of the first equation in (9.5) (but now for modified Helmholtz instead of Helmholtz), and taking the limit as $\mathbf{x} \to \Gamma$ in Equation (4.1) in [FL15] then yields the analogue of (9.11) in the transform space.*

Remark 10.4 (The variation formulation in [Ash13]) *The paper [Ash13] introduces a new variational formulation of the Dirichlet problem for Laplace's equation in a convex polygon (after reducing Laplace's equation to the $\overline{\partial}$-equation as discussed in Remark 10.2), with this variational formulation based on the analogue of the global relation (10.14) for this particular problem. The difference between the variational formulation in [Ash13] and the variational formulations based on (10.14) discussed above is that, in [Ash13], the unknowns are a set of transforms of the Neumann trace $\partial_n u$ (as opposed to $\partial_n u$ itself), and therefore the Hilbert space is a space of analytic functions (as opposed to a space of functions on Γ).*

10.3 Interior Dirichlet problem for the Helmholtz equation

Having discussed the Fokas transform method, our next goal is to discuss the null-field method. Although this second method is designed to solve exterior Helmholtz BVPs, it is helpful to first consider interior Helmholtz BVPs, and thus in this subsection we consider the Helmholtz IDP.

Repeating the steps we performed for the modified Helmholtz IDP in §10.1, we see that everything up to Theorem 10.1 is the same for the Helmholtz equation, i.e. the space $\mathcal{R}$ and sesquilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ are defined by (10.1), (10.3), and (10.5) respectively. However, the analogue of Theorem 10.1 is now the following.

Theorem 10.5 (Continuity, Gårding inequality, and injectivity for $b(\cdot,\cdot)$ for Helmholtz) *(i) For any $\phi, \psi \in H^{1/2}(\Gamma)$,*

$$\left|b(\phi,\psi)\right| \leq \|P_{\mathrm{DtN}}\|_{H^{1/2}(\Gamma)\to H^{-1/2}(\Gamma)} \|\phi\|_{H^{1/2}(\Gamma)} \|\psi\|_{H^{1/2}(\Gamma)}.$$

(ii) Assume that k^2 is not a Dirichlet eigenvalue of the Laplacian in Ω and also that, given $\theta \in (1/2,1)$, there exists a $C_1 > 0$ (depending on k and θ) such that if $u \in H^1(\Omega)$ satisfies $\mathcal{L}_k u = 0$ then

$$\|u\|_{H^\theta(\Omega)} \leq C_1 \|\gamma u\|_{H^{\theta-1/2}(\Gamma)}. \tag{10.20}$$

Then, there exists a $C_2 > 0$ (independent of k) such that

$$b(\psi,\psi) \geq \frac{1}{C_2^2}\|\psi\|^2_{H^{1/2}(\Gamma)} - (1+k^2)C_1^2\|\psi\|^2_{H^{\theta-1/2}(\Gamma)}$$

for all $\psi \in H^{1/2}(\Gamma)$.
(iii) If k^2 is not a Dirichlet eigenvalue of the Laplacian in Ω and $b(\phi,\psi) = 0$ for all $\psi \in H^{1/2}(\Gamma)$, then $\phi = 0$ (and thus the operator associated with $b(\cdot,\cdot)$ is injective).

Proof. The proof of (i) is identical to the modified Helmholtz case. For (ii), given $\psi \in H^{1/2}(\Gamma)$ there exists a unique $u \in \mathcal{R}$ such that $\gamma u = \psi$ (u is unique because k^2 is not a Dirichlet eigenvalue). Then, by G1 (4.4),

$$b(\psi,\psi) = \int_\Gamma \partial_n u \overline{\gamma u} = \int_\Omega |\nabla u|^2 - k^2|u|^2 = \|u\|^2_{H^1(\Omega)} - (1+k^2)\,\|u\|^2_{L^2(\Omega)}\,. \tag{10.21}$$

Let C_2 be the constant in the inequality (2.1) (concerning the trace theorem). Then, using the inequalities (2.1) and (10.20) in (10.21), we obtain the result.

For (iii), the hypothesis and the definition of $b(\cdot,\cdot)$ (10.5) imply that

$$\langle P_{\mathrm{DtN}}\phi, \psi\rangle_\Gamma = 0 \quad \text{for all } \psi \in H^{1/2}(\Gamma),$$

which implies that $P_{\mathrm{DtN}}\phi = 0$ as an element of $H^{-1/2}(\Gamma)$ (if $F(\cdot)$ is an anti-linear functional on $H^{1/2}(\Gamma)$ and $F(\psi) = 0$ for all $\psi \in H^{1/2}(\Gamma)$ then $F = 0$ as an element of $(H^{1/2}(\Gamma))^* \cong H^{-1/2}(\Gamma)$). When k^2 is not a Dirichlet eigenvalue of the Laplacian in Ω_-, the operator $P_{\mathrm{DtN}} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is invertible (this is perhaps most easily proved via boundary integral equations, see [SS11, Remark 3.7.5]), and thus $\phi = 0$. ∎

Remark 10.6 (The bound (10.20)) *In [ADK82] it is stated that the bound (10.20) follows "by a standard inequality for elliptic problems in bounded, smooth domains", however we have been unable to find a reference. Note that, when the solution of the IDP for $\mathcal{L}$ is unique, the bound with $\theta = 1$ holds by Fredholm theory (see, e.g., [SS11, Theorem 2.10.4]), and the bound with $\theta = 1/2$ holds for the Laplace equation by [JK95, Corollary 5.5] (which is also stated as [CWGLS12, Theorem A.6]).*

The results of Theorem 10.5 combined with Theorem 5.20 and Part (i) of Theorem 5.18 then imply that the variational problem (10.7) has a unique solution (i.e. the properties K1 and K2 hold) when k^2 is not a Dirichlet eigenvalue of the Laplacian in Ω. Similar to the case of the modified Helmholtz equation we can then abandon the variational problem (10.6) and focus on (10.7).

As before, we need to use the finite-dimensional subspaces $\mathcal{Q}_N$ defined by (10.11) for the Galerkin method to be practical (with these subspaces now consisting of traces of explicit solutions of $\mathcal{L}_k v = 0$). Part (ii) of Theorem 5.18 can then be used to prove that the Galerkin equations (10.12) have a unique, quasi-optimal solution (i.e. property K3 holds), provided that N is large enough and that $(\mathcal{Q}_N)_{N\in\mathbb{Z}^+}$ is a sequence of finite dimensional nested subspaces of $H^{1/2}(\Gamma)$ whose union is dense in $H^{1/2}(\Gamma)$ (i.e. (5.7) holds with $\mathcal{H}_i = H^{1/2}(\Gamma)$ and $\mathcal{H}_N^i = \mathcal{Q}_N$). Since $\mathcal{Q}_N$ consists of traces of particular solutions of $\mathcal{L}_k v = 0$, this density property is not immediate and needs to be checked; we discuss this more for the case of the EDP in §10.6 below, and we refer the reader to [CWL15, §4.1] for discussion of this density property for the IDP. Indeed, [CWL15] (which appears in the same collection of articles as the present paper) also discusses the Fokas transform method for the Helmholtz IDP, with [CWL15, Lemma 4.1] proving an appropriate density result and then [CWL15, Theorem 4.2] applying this to the Galerkin method.

10.4 Exterior Dirichlet problem for the Helmholtz equation

We now let

$$\mathcal{R} := \Big\{ u \in H^1_{\mathrm{loc}}(\Omega_+,\Delta) : \mathcal{L}_k u = 0 \text{ in } \Omega_+ \text{ and } u \text{ satisfies the radiation condition (3.4)} \Big\}. \tag{10.22}$$

For $u, v \in \mathcal{R}$, applying G2 in Ω_R, letting $R \to \infty$, and using Lemma 4.10, we find that

$$\int_\Gamma \partial_n^+ u\, \gamma_+ v = \int_\Gamma \gamma_+ u\, \partial_n^+ v, \tag{10.23}$$

but

$$\int_\Gamma \partial_n^+ u\, \overline{\gamma_+ v} \neq \int_\Gamma \gamma_+ u\, \overline{\partial_n^+ v}$$

(since the integral over Γ_R does not tend to zero in the second case).

One can use (10.23) as the basis of a variational formulation of the Helmholtz EDP. Indeed, if

$$a(u,v) := \int_\Gamma \partial_n u\, \gamma v \quad \text{and} \quad F(v) = \int_\Gamma g_D \partial_n v,$$

then the equation (10.23) gives rise to the following variational formulation of the EDP: given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } u \in \mathcal{R} \text{ such that} \quad a(u,v) = F(v) \quad \text{for all } v \in \mathcal{R}.} \tag{10.24}$$

Defining $b(\cdot,\cdot)$ by

$$b(\phi,\psi) := \int_\Gamma P_{\mathrm{DtN}}\phi\, \psi, \tag{10.25}$$

where $P_{\mathrm{DtN}} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is now the exterior Dirichlet-to-Neumann map for the Helmholtz equation, we see that the variational problem (10.24) becomes

$$\boxed{\text{find } u \in \mathcal{R} \text{ such that} \quad b(\gamma u, \gamma v) = F(v) \quad \text{for all } v \in \mathcal{R}.} \tag{10.26}$$

The difficulty is that $b(\cdot,\cdot)$ is now a bilinear form on a complex-valued space (and $F(\cdot)$ is a linear functional). This is not a problem in and of itself, because the Lax-Milgram theorem, for example, still holds for such $b(\cdot,\cdot)$ and $F(\cdot)$. However, the lack of complex-conjugate in $b(\cdot,\cdot)$ makes it difficult to prove coercivity or a Gårding inequality. Indeed, with $R > \sup_{\mathbf{x}\in\Omega_-} |\mathbf{x}|$, applying G1 (4.3) in Ω_R implies that

$$b(\psi,\psi) = \int_{\Gamma_R} u \frac{\partial u}{\partial r} - \int_{\Omega_R} \big(\nabla u \cdot \nabla u - k^2 u^2\big), \tag{10.27}$$

where $u \in \mathcal{R}$ is such that $\gamma_+ u = \psi$ (u is unique by the uniqueness of the solution to the Helmholtz EDP). The lack of complex conjugate in the bilinear form means that one cannot get a norm of u from the terms in Ω_R on the right-hand side of (10.27). (Note that this difficulty can be understood as a consequence of the fact that the Helmholtz EDP is not self-adjoint; see Example 4.9.)

We now have two options: (i) continue with the variational problem (10.26) and accept that it will be difficult to prove that the properties K1, K2, and K3 hold, or (ii) modify $b(\cdot,\cdot)$. The first option is the null-field method, and we discuss this in §10.5. The second option was introduced by Aziz, Dorr, and Kellogg [ADK82], and we discuss this in §10.6.

10.5 The null-field method for the Helmholtz exterior Dirichlet problem

We start from (10.23) and choose a family of solutions of $\mathcal{L}_k v = 0$ that we hope span (in some sense) $\mathcal{R}$. In 2-d we choose

$$v_n(\mathbf{x}) = H_n^{(1)}(kr)\mathrm{e}^{\mathrm{i}n\theta} \quad \text{for } n \in \mathbb{Z}, \text{ where } \mathbf{x} = (r,\theta), \tag{10.28}$$

and in 3-d

$$v_{n,m}(\mathbf{x}) = h_n^{(1)}(kr) Y_{n,m}(\theta,\phi) \quad \text{for } n \in \mathbb{Z} \text{ and } m = -n,\dots,n, \text{ where } \mathbf{x} = (r,\theta,\phi), \tag{10.29}$$

where $H_n^{(1)}$ are the Hankel functions defined by, e.g., [NIS14, Equations 10.4.3 and 10.2.5], $h_n^{(1)}$ are the spherical Bessel functions defined by [NIS14, Equation 10.47.5], and $Y_{n,m}$ are the spherical harmonics defined by [NIS14, Equation 14.30.1].

In what follows we consider the 2-d case, but everything follows in the same way for 3-d. Choosing v in (10.23) to be v_n defined by (10.28), we obtain the *null-field equations*

$$\int_\Gamma \partial_n^+ u\, \gamma_+ v_n = \int_\Gamma \gamma_+ u\, \partial_n^+ v_n \quad \text{for all } n \in \mathbb{Z}. \tag{10.30}$$

The null-field method for the EDP is then, given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } \phi \in H^{-1/2}(\Gamma) \text{ such that} \quad \int_\Gamma \phi\, \gamma_+ v_n = \int_\Gamma g_D\, \partial_n^+ v_n \quad \text{for all } n \in \mathbb{Z},} \tag{10.31}$$

and then set $\partial_n^+ u = \phi$. (We can think of (10.31) as a moment problem: we know the moments of ϕ with respect to v_n and the task is to find ϕ itself.) We prove below that the solution to (10.31) is unique for every $k > 0$. The unique solvability for every $k > 0$ of this boundary-based formulation of the EDP (in contrast to the "spurious frequencies" of the integral equations (9.11) and (9.12)) was the main novelty of the null-field equations when they were first introduced.

To practically implement (10.31), one chooses a family of functions $\{\psi_j\}_{j\in\mathbb{Z}} : H^{-1/2}(\Gamma) \to \mathbb{C}$, approximates ϕ by

$$\phi_N := \sum_{j=1}^{N} a_j \psi_j,$$

and then solves the linear system

$$\sum_{j=1}^{N} \left(\int_\Gamma \psi_j \, \gamma_+ v_n \right) a_j = \int_\Gamma g_D \, \partial_n^+ v_n \quad \text{for } n = 1, \ldots, N. \tag{10.32}$$

Common choices of the functions ψ_j are then $\gamma_+ v_j$ and $\partial_n^+ v_j$; see [Mar06, §7.7.2] for more discussion.

Remark 10.7 (Connection with least squares [Mar06, §7.8.2]) *If we seek to approximate the solution of the exterior Dirichlet problem by*

$$u_N(\mathbf{x}) := \sum_{j=1}^{N} a_j \, v_j(\mathbf{x}),$$

with v_j defined by (10.28) and $a_j \in \mathbb{C}$, and impose the Dirichlet boundary condition $\gamma_+ u = g_D$ in a least-squares sense, then we are led to minimising the functional

$$J(a_1, \ldots, a_N) = \int_\Gamma \left| \sum_{j=1}^{N} a_j \gamma_+ v_j - g_D \right|^2 .$$

Taking the derivatives of $J(a_1, \ldots, a_N)$ with respect to the real and imaginary parts of a_j for $j = 1, \ldots, N$ we arrive at the linear system

$$\sum_{j=1}^{N} \left(\int_\Gamma \overline{\gamma_+ v_n} \, \gamma_+ v_j \, \mathrm{d}s \right) a_j = \int_\Gamma g_D \, \overline{\gamma_+ v_n}, \quad \text{for } n = 1, \ldots, N. \tag{10.33}$$

If we choose $\psi_j = \overline{\gamma_+ v_j}$ in the discretised null-field equations (10.32), then the corresponding (Hermitian) matrix is the transpose (or complex-conjugate) of the matrix in (10.33).

Alternative derivation of the null-field equations and a proof of uniqueness. We now give an alternative derivation of the null-field equations that can be "reversed" to give a proof of uniqueness.

We start from the first equation in (9.5), i.e.

$$-\int_\Gamma \left(\Phi_k(\mathbf{x}, \mathbf{y}) \partial_n^+ u(\mathbf{y}) - \frac{\partial \Phi_k(\mathbf{x}, \mathbf{y})}{\partial n(\mathbf{y})} \gamma_+ u(\mathbf{y}) \right) \mathrm{d}s(\mathbf{y}) = 0, \quad \mathbf{x} \in \Omega_-. \tag{10.34}$$

We introduce polar coordinates $\mathbf{x} = (r, \theta)$ and $\mathbf{y} = (\rho, \phi)$, and recall the Fourier series expansion of the fundamental solution

$$\Phi_k(\mathbf{x}, \mathbf{y}) := \frac{\mathrm{i}}{4} H_0^{(1)}\big(k|\mathbf{x} - \mathbf{y}|\big) = \frac{\mathrm{i}}{4} \sum_{n=-\infty}^{\infty} H_n^{(1)}(kr_>) J_n(kr_<) \mathrm{e}^{\mathrm{i}n(\phi-\theta)}, \tag{10.35}$$

where $r_> = \max(r, \rho)$ and $r_< = \min(r, \rho)$; see, e.g., [Mar06, Equation (2.29)][6]

[6] To convert [Mar06, Equation (2.29)] into an expression equivalent to (10.35) (after relabelling variables), let $\widetilde{\mathbf{b}} = -\mathbf{b}$ and observe that (i) $\mathbf{r}_2 = \mathbf{r}_1 - \widetilde{\mathbf{b}}$, and (ii) the angular co-ordinate of $\widetilde{\mathbf{b}}$ is $\pi + \beta$.

Assume that Ω_- contains the ball of radius a centred at the origin, B_a, and let $\mathbf{x}$ be in this ball. Substituting (10.35) into (10.34) and noting that $r_< = r$ and $r_> = \rho$, we obtain that

$$\sum_{n=-\infty}^{\infty} J_n(kr)\mathrm{e}^{-\mathrm{i}n\theta} \int_\Gamma \left(-H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\partial_n^+ u(\mathbf{y}) + \frac{\partial}{\partial n(\mathbf{y})}\big(H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\big)\gamma_+ u(\mathbf{y}) \right) \mathrm{d}s(\mathbf{y}) = 0 \quad \text{for } r < a. \tag{10.36}$$

(Interchanging the sum and integral is justified since the series (10.35) converges absolutely due to the asymptotics

$$H_n^{(1)}(kr_>)J_n(kr_<) \sim \frac{1}{\mathrm{i}\pi n}\left(\frac{r_<}{r_>}\right)^n \quad \text{as } n \to \infty$$

[AS64, Equation 9.31], [NIS14, §10.19].)

We now claim that

$$\int_\Gamma \left(-H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\partial_n^+ u(\mathbf{y}) + \frac{\partial}{\partial n(\mathbf{y})}\big(H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\big)\gamma_+ u(\mathbf{y}) \right) \mathrm{d}s(\mathbf{y}) = 0 \quad \text{for all } n \in \mathbb{Z}. \tag{10.37}$$

Indeed, this is a consequence of applying the following lemma to (10.36).

Lemma 10.8 *If $(\alpha_n)_{n\in\mathbb{Z}}$ are such that (i) the series $\sum_{n=-\infty}^{\infty} |\alpha_n||J_n(kr)|$ converges, and (ii) there exists an $a > 0$ such that*

$$\sum_{n=-\infty}^{\infty} \alpha_n J_n(kr)\,\mathrm{e}^{-\mathrm{i}n\theta} = 0 \quad \textit{for } r < a, \tag{10.38}$$

then $\alpha_n = 0$ for all $n \in \mathbb{Z}$.

Proof. Multiplying the sum (10.38) by $\mathrm{e}^{\mathrm{i}m\theta}$ and integrating over $\theta \in (0, 2\pi)$, we find that

$$J_m(kr)\alpha_m = 0 \quad \text{for all } m \in \mathbb{Z} \text{ and for all } r < a.$$

If $J_m(kr) \neq 0$ then $\alpha_m = 0$. However, if $k = \beta_{m,l}/r$, where $\beta_{m,l}$ is the lth zero of J_m, then $J_m(kr) = 0$. (Recall that the Dirichlet eigenvalues of the negative Laplacian in B_r are $(\beta_{m,l}/r)^2$). We therefore need to choose an $r < a$ such that $J_m(kr) \neq 0$ for all $m \in \mathbb{Z}$.

Recall that $0 < \beta_{0,1} \leq \beta_{m,l}$ for all $m = 0, 1, \ldots$ and $l = 1, \ldots$ [NIS14, Equation 10.21.2]. Therefore, given $k > 0$, choose $r^* < \min(\beta_{0,1}/k, a)$. Then $J_m(kr^*) \neq 0$ for all $m \in \mathbb{Z}$ and we are done. ∎

The definition of v_n (10.28) implies that (10.37) can be rewritten as

$$\int_\Gamma \big(-\gamma_+ v_n(\mathbf{y})\partial_n^+ u(\mathbf{y}) + \partial_n^+ v_n(\mathbf{y})\gamma_+ u(\mathbf{y})\big)\,\mathrm{d}s(\mathbf{y}) = 0 \quad \text{for all } n \in \mathbb{Z},$$

which is (10.30). (This is how the null-field equations were originally derived by Waterman in [Wat65] for the time-harmonic Maxwell equations, and in [Wat69] for the Helmholtz equation.)

For uniqueness, we need to show that if $\phi \in H^{-1/2}(\Gamma)$ is such that

$$\int_\Gamma \phi\,\gamma_+ v_n = 0 \quad \text{for all } n \in \mathbb{Z} \tag{10.39}$$

then $\phi = 0$. To begin, we assume that Ω_- contains B_a and rewrite (10.39) as

$$\int_\Gamma \phi(\mathbf{y}) H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\mathrm{d}s(\mathbf{y}) = 0 \quad \text{for all } n \in \mathbb{Z}$$

(where, as above, $\mathbf{y} = \rho\mathrm{e}^{\mathrm{i}\phi}$). Multiplying each of these equations by $J_n(kr)\mathrm{e}^{-\mathrm{i}n\theta}$, with $r < a$, and summing them up, we obtain that

$$0 = \sum_{n=-\infty}^{\infty} J_n(kr)\mathrm{e}^{-\mathrm{i}n\theta} \int_\Gamma \phi(\mathbf{y}) H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\mathrm{d}s(\mathbf{y}) \quad \text{for } r < a.$$

However, the expansion of the fundamental solution (10.35) then implies that

$$\sum_{n=-\infty}^{\infty} J_n(kr)\mathrm{e}^{-\mathrm{i}n\theta}\int_\Gamma \phi(\mathbf{y})H_n^{(1)}(k\rho)\mathrm{e}^{\mathrm{i}n\phi}\mathrm{d}s(\mathbf{y}) = \int_\Gamma \Phi_k(\mathbf{x},\mathbf{y})\phi(\mathbf{y})\,\mathrm{d}s(\mathbf{y}) \quad \text{for } \mathbf{x}\in B_a \tag{10.40}$$

(i.e. we have "reversed" the original derivation of the null-field equations).

Recalling the definition of the single-layer potential $\mathcal{S}_k$ (9.6), we see that if ϕ is such that (10.40) holds, then $u := \mathcal{S}_k\phi$ is zero in B_a. Now u is in $C^2(\Omega_-)$ and satisfies $\mathcal{L}_k u = 0$ and, by Green's integral representation, C^2 solutions of the Helmholtz equation are analytic (see, e.g., [CK83, Theorem 3.5]). Therefore, if $u = 0$ in B_a then $u = 0$ in Ω_-. By the jump relations (9.9), u is continuous across Γ, and so $\gamma_+u = \gamma_-u = 0$. By uniqueness of the Helmholtz EDP, $u = 0$ in Ω_+. The jump relations (9.9) imply that $\phi = \partial_n^- u - \partial_n^+ u$, and then (since $u = 0$ in both Ω_- and Ω_+) $\phi = 0$ and we have uniqueness.

10.6 The method of Aziz, Dorr, and Kellogg for the Helmholtz exterior Dirichlet problem

The method introduced in [ADK82] reformulates the Helmholtz EDP as the following variational problem: given $g_D \in H^{1/2}(\Gamma)$,

$$\boxed{\text{find } u\in\mathcal{R} \text{ such that } \int_\Gamma \gamma_+u\,\overline{\partial_n^+ v} = \int_\Gamma g_D\,\overline{\partial_n^+ v} \quad \text{for all } v\in\mathcal{R},} \tag{10.41}$$

where $\mathcal{R}$ is defined by (10.22). This variational problem can be understood as imposing the Dirichlet boundary condition on Γ in a weak sense, i.e. there is no use of Green's identities.

Define the sesquilinear form $\widetilde{b}(\cdot,\cdot)$ by

$$\widetilde{b}(\phi,\psi) := \int_\Gamma \phi\,\overline{P_{\mathrm{DtN}}\psi},$$

where $P_{\mathrm{DtN}} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is the exterior Dirichlet-to-Neumann map for the Helmholtz equation. With this definition, the variational problem (10.41) can be written as

$$\boxed{\text{find } u\in\mathcal{R} \text{ such that } \widetilde{b}(\gamma_+u,\gamma_+v) = \int_\Gamma g_D\,\overline{\partial_n^+ v} \quad \text{for all } v\in\mathcal{R}.} \tag{10.42}$$

The next theorem allows us to reduce the variational problem (10.42) to the following one:

$$\boxed{\text{find } \phi\in H^{1/2}(\Gamma) \text{ such that } \widetilde{b}(\phi,\psi) = G(\psi) \quad \text{for all } \psi\in H^{1/2}(\Gamma),} \tag{10.43}$$

where $G(\cdot)$ is defined by (10.8).

Theorem 10.9 (Continuity, Gårding inequality, and injectivity for $\widetilde{b}(\cdot,\cdot)$)
(i) For any $\phi,\psi\in H^{1/2}(\Gamma)$,

$$\left|\widetilde{b}(\phi,\psi)\right| \le \|P_{\mathrm{DtN}}\|_{H^{1/2}(\Gamma)\to H^{-1/2}(\Gamma)}\,\|\phi\|_{H^{1/2}(\Gamma)}\,\|\psi\|_{H^{1/2}(\Gamma)}\,.$$

(ii) Assume that, given $\theta\in(1/2,1)$ and $R > \sup_{\mathbf{x}\in\Omega_-}|\mathbf{x}|$, there exists a $C_1 > 0$ (depending on k and θ) such that if $u\in H^1_{\mathrm{loc}}(\Omega_+)$ satisfies $\mathcal{L}_k u = 0$ then

$$\|u\|_{H^\theta(\Omega_R)} \le C_1\,\|\gamma_+u\|_{H^{\theta-1/2}(\Gamma)}\,. \tag{10.44}$$

Then, there exists a $C_2 > 0$ (independent of k) such that

$$\left|\widetilde{b}(\psi,\psi)\right| \ge \frac{1}{C_2^2}\,\|\psi\|^2_{H^{1/2}(\Gamma)} - (1+k^2)C_1^2\,\|\psi\|^2_{H^{\theta-1/2}(\Gamma)} \tag{10.45}$$

for all $\psi\in H^{1/2}(\Gamma)$.
(iii) If $\widetilde{b}(\phi,\psi) = 0$ for all $\psi\in H^{1/2}(\Gamma)$ then $\phi = 0$ (and thus the operator associated with $\widetilde{b}(\cdot,\cdot)$ is injective).

Proof. (Note that Part (ii) of this theorem was first proved in [ADK82, Theorem 3.1].)

The proof of (i) is straightforward (note that if Ω_+ is nontrapping then, given $k_0 > 0$, $\|P_{\mathrm{DtN}}\|_{H^{1/2}(\Gamma)\to H^{-1/2}(\Gamma)} \lesssim k$ for all $k \geq k_0$ by [BSW15]). The proof of (ii) is very similar to the proof of Part (ii) of Theorem 10.5. Indeed, given $\psi \in H^{1/2}(\Gamma)$ there exists a unique $u \in \mathcal{R}$ such that $\gamma_+ u = \psi$. Then, by G1 (4.4),

$$\left|\widetilde{b}(\psi,\psi)\right| \geq -\Re\widetilde{b}(\psi,\psi) = \int_{\Omega_R} \left(|\nabla u|^2 - k^2|u|^2\right) - \Re\int_{\Gamma_R} \overline{u}\frac{\partial u}{\partial r}. \tag{10.46}$$

Now, if $u \in \mathcal{R}$ then $\Re\int_{\Gamma_R} \overline{u}\,\partial u/\partial r\,\mathrm{d}s \leq 0$ [Néd01, Theorem 2.6.4]. After using this inequality in (10.46), we find that the proof of (10.45) is identical to the proof of Part (ii) of Theorem 10.5.

The proof of (iii) is also very similar to the proof of Part (iii) of Theorem 10.5. The exterior Dirichlet-to-Neumann map $P_{\mathrm{DtN}} : H^{1/2}(\Gamma) \to H^{1/2}(\Gamma)$ is invertible for all $k > 0$ by uniqueness of the solution to the Helmholtz EDP (although the invertibility is perhaps most easily proved via integral equations; see [CWGLS12, Theorem 2.31]). ∎

Similar to the case of the Helmholtz IDP in §10.3, Theorems 10.9 and 5.18 show that the variational problem (10.43) has a unique solution. It is then straightforward to show that this solution is $\gamma_+ u$, where u is the solution to (10.42). We can therefore forget about the variational problem (10.42) and instead concentrate on (10.43).

Similar to the interior problems in §10.1 and §10.2, given $\phi, \psi \in H^{1/2}(\Gamma)$, to find $\widetilde{b}(\phi,\psi)$ and $G(\psi)$ one needs to find $P_{\mathrm{DtN}}\psi$. Moreover (and exactly as before), the unknown ϕ in the variational problem (10.43) equals $\gamma_+ u$, which is already given by the boundary condition. Therefore, once we have found the Galerkin approximation to ϕ, ϕ_N, we need to be able to work out $P_{\mathrm{DtN}}\phi_N$ easily.

Therefore, for the Galerkin method applied to (10.43) to be practical, we need to choose finite-dimensional subspaces where it is easy to apply the Dirichlet-to-Neumann map (for example, subspaces consisting of traces of explicit solutions to the Helmholtz equation in Ω_+). A natural choice in 2-d is

$$\mathcal{Q}_N := \operatorname{span}\left\{\gamma_+ v_n : n = -N, \ldots, N\right\}, \tag{10.47}$$

where v_n are defined by (10.28). (Note that $\mathcal{Q}_N \subset \mathcal{Q}_{N+1}$ by definition.)

Theorem 10.10 (Asymptotic density of $\mathcal{Q}_N$ in $H^{1/2}(\Gamma)$) *With $\mathcal{Q}_N$ defined by (10.47), their union over $N \in \mathbb{Z}^+$ is dense in $H^{1/2}(\Gamma)$ (i.e. (5.7) holds with $\mathcal{H}_i = H^{1/2}(\Gamma)$ and $\mathcal{H}_N^i = \mathcal{Q}_N$).*

Proof. The analogue of this result in 3-d (with v_n replaced by $v_{n,m}$ defined by (10.29)) is proved in [ADK82, Theorem 5.1]; the proof for the 2-d case is completely analogous. ∎

This density result and Part (ii) of Theorem 5.18 then imply that, when N is sufficiently large, the Galerkin method applied to the variational problem (10.43) has a unique, quasi-optimal solution (i.e. the property K3 holds).

Remark 10.11 (The method of DeSanto for scattering by diffraction gratings) *In this paper, the only BVP for the Helmholtz equation posed on an* unbounded *domain that we considered was the EDP. Another such BVP is the Helmholtz equation posed above an infinite rough surface, and in the case when the surface is periodic the surface is known as a* diffraction grating. *A method for solving scattering by diffraction gratings was introduced by DeSanto in [DeS81], and further developed by DeSanto and co-workers in [DEHM98], [DEHM01], [DEH+01], and [ACWD06]. In [CWL15, §4.2] it is shown that this method can be viewed as an implementation of the Fokas transform method to diffraction grating problems.*

11 Concluding remarks

This paper had the following two goals:

Goal 1: To give an overview of variational formulations for second-order linear elliptic PDEs based on multiplying by a test function and integrating by parts (or, equivalently, based on Green's identities).

Goal 2: To show how the Fokas transform method applied to second-order linear elliptic PDEs can be placed into the framework established in Goal 1.

Stepping back to look at the bigger picture around Goal 1, we might ask the following two questions:

1. Goal 1 has assumed that all variational formulations based on multiplying by a test function and integrating by parts are equivalent to using Green's identities, but are there any formulations based on other identities?

2. What about variational formulations that are not based on any identites?

Similarly, concerning the wider context of Goal 2 we might ask the question

3. Since the Fokas transform method arose from investigations of certain non-linear PDEs (the so-called *integrable* PDEs), what have integrable non-linear PDEs got to do with Green's identities?

In this final section we answer these three questions.

11.1 Variational formulations based identities other than Green's identities.

New variational formulations of the interior impedance and exterior Dirichlet problems for the Helmholtz equation were introduced in [MS14] and these variational formulations are based on identities other than Green's identities. The idea behind these variational formulations is the following.

We saw in §6 that the standard variational formulations of the interior Dirichlet and impedance problems for the Helmholtz equation, (6.3) and (6.14) respectively, arose from integrating over Ω the identity

$$\overline{v}\,\mathcal{L}_k u = \nabla\cdot\left[\overline{v}\nabla u\right] - \nabla u\cdot\overline{\nabla v} + k^2 u\overline{v}, \tag{11.1}$$

i.e. G1 (4.1). Recall that the sesquilinear forms in these variational formulations, $a_D(\cdot,\cdot)$ and $a_I(\cdot,\cdot)$, are not coercive when k is sufficiently large (see Lemmas 6.2 and 6.5). On one level, the reason for this is that, when $u=v$, the non-divergence terms on the right-hand side of (11.1) equal $-|\nabla v|^2+k^2|v|^2$, and this expression is *not* single-signed (i.e. for some v it will be positive, and for some v it will be negative).

This observation motivates the following question: if the identity (11.1) is replaced by a different identity with the property that when $u=v$ the non-divergence terms on the right-hand side *are* single-signed, will this lead to coercive variational formulations? The paper [MS14] shows that the answer is yes for the interior impedance and exterior Dirichlet problems. (Note that there cannot exist a formulation of the interior Dirichlet problem that is coercive for all $k>0$ because the solution of this problem is not unique for all $k>0$.)

The formulations in [MS14] are based on a class of identities for the Helmholtz equation introduced by Morawetz in [ML68] and [Mor75] (building on the earlier work [Mor61] concerning the wave equation). The simplest such identity involving two functions u and v is

$$\overline{\mathcal{M}v}\mathcal{L}u + \mathcal{M}u\overline{\mathcal{L}v} = \nabla\cdot\left[\overline{\mathcal{M}v}\,\nabla u + \mathcal{M}u\,\overline{\nabla v} + \mathbf{x}(k^2u\overline{v} - \nabla u\cdot\overline{\nabla v})\right] - \nabla u\cdot\overline{\nabla v} - k^2u\overline{v}, \tag{11.2}$$

where the multiplier $\mathcal{M}$ is defined by

$$\mathcal{M}v := \mathbf{x}\cdot\nabla v - \mathrm{i}k\beta v + \frac{d-1}{2}v, \tag{11.3}$$

and β is an arbitrary real number. We see that, when $u=v$, the non-divergence terms of (11.2) equal $-|\nabla v|^2-k^2|v|^2$ and this expression is single-signed. Integrating the identity (11.2) over Ω and using the PDE (3.5) and boundary conditions (3.6) gives rise to a variational formulation of the Helmholtz interior impedance problem that is coercive for all $k>0$ when Ω is star-shaped with respect to a ball [MS14, Theorem 1.1 and §3].

The only disadvantage of the new formulation compared to the standard variational formulation is that the Hilbert space of the new formulation is $\mathcal{V}$ defined by (8.16) (we need to work in this space because of the $\mathcal{M}u\overline{\mathcal{L}v}$ term on the left-hand side of (11.2)). Because the space $\mathcal{V}$ is smaller than $H^1(\Omega)$, it is harder to create piecewise-polynomial finite-dimensional subspaces of $\mathcal{V}$ than it is for $H^1(\Omega)$ (see [MS14, §5] for more details).

11.2 Variational formulations not based on any identities.

The main class of variational formulations that are not based on integrating identities over the domain (or over elements of a triangulation of the domain) are *least-squares methods*.

Given a linear PDE, $\mathcal{L}u = f$, on a domain Ω, with boundary conditions $\mathcal{B}u = g$ on $\Gamma := \partial\Omega$, the simplest least-squares method consists of minimising the functional

$$J(v) = \|\mathcal{L}v - f\|_\Omega^2 + \|\mathcal{B}v - g\|_\Gamma^2$$

(where $\|\cdot\|_\Omega$ is an appropriate norm on Ω arising from an inner product $(\cdot,\cdot)_\Omega$, and $\|\cdot\|_\Gamma$ is an appropriate norm on Γ arising from an inner product $(\cdot,\cdot)_\Gamma$).

The problem of minimising $J(\cdot)$ can be shown to be equivalent to the variational problem (1.1) with

$$a(u,v) = (\mathcal{L}u, \mathcal{L}v)_\Omega + (\mathcal{R}u, \mathcal{R}v)_\Gamma \quad \text{and} \quad F(v) = (f, \mathcal{L}v)_\Omega + (g, \mathcal{R}v)_\Gamma,$$

and this formulation does not (in general) arise from integrating an identity over Ω.

We refer the reader to [BG09] for a good introduction to least-squares methods in general, and to [EM12, §6] for a review of least-squares methods for the Helmholtz equation that use subspaces satisfying the Trefftz property (7.8).

11.3 From Green to Lax.

The Fokas transform method arose from attempts to solve BVPs for integrable non-linear PDEs, where we say that a PDE is *integrable* if it possesses a Lax pair formulation. The reason that the Fokas transform method is also applicable to linear PDEs is that linear PDEs possess Lax pair formulations; this fact was first noted by Fokas and Gelfand in [FG94].

The second goal of this paper was to place the Fokas transform method into the framework of variational formulations arising from Green's identities. The reason that we have been able to do this is that *Lax pairs for linear PDEs arise naturally from the differential (as opposed to integrated) form of Green's second identity.* Indeed, a Lax pair formulation for the second-order linear elliptic PDE (1.2) can be obtained from (4.2), and a Lax pair formulation for a general linear PDE can be obtained from (4.9); see [FS12, §7] for more details.

Acknowledgments. For useful discussions and comments, the author thanks Anthony Ashton (University of Cambridge), Simon Chandler-Wilde (University of Reading), Thanasis Fokas (University of Cambridge), Ivan Graham (University of Bath), Ralf Hiptmair (ETH Zürich), Paul Martin (Colorado School of Mines)[7], Peter Monk (University of Delaware), and especially Andrea Moiola (University of Reading).

The author was supported by EPSRC grant EP/1025995/1.

References

[AA02] Y. A. Abramovich and C. D. Aliprantis. *An invitation to operator theory.* American Mathematical Society Providence, 2002.

[ABCM02] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM Journal on Numerical Analysis*, 39(5):1749–1779, 2002.

[ACWD06] T. Arens, S. N. Chandler-Wilde, and J. A. DeSanto. On integral equation and least squares methods for scattering by diffraction gratings. *Communications in Computational Physics*, 1(6):1010–1042, 2006.

[ADK82] A. K. Aziz, M. R. Dorr, and R. B. Kellogg. A new approximation method for the Helmholtz equation in an exterior domain. *SIAM Journal on Numerical Analysis*, 19(5):899–908, 1982.

[7] Regarding the title of this paper, Paul Martin told the author that Fritz Ursell was famous for asking the question of graduate students and seminar speakers "Have you tried using Green's theorem?"

[AF03] M. J. Ablowitz and A. S. Fokas. *Complex Variables: Introduction and Applications*. CUP, 2nd edition, 2003.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.

[Ash12] A. C. L. Ashton. On the rigorous foundations of the Fokas method for linear elliptic partial differential equations. *Proc. R. Soc. Lond. A*, 468(2141):1325–1331, 2012.

[Ash13] A. C. L. Ashton. The spectral Dirichlet-Neumann map for Laplace's equation in a convex polygon. *SIAM J. Math. Anal.*, 45(6):3575–3591, 2013.

[Aub67] J. P. Aubin. Behavior of the error of the approximate solutions of boundary value problems for linear elliptic operators by Galerkin's and finite difference methods. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 21(4):599–637, 1967.

[BBD13] T. P. Barrios, R. Bustinza, and V. Domínguez. On the discontinuous Galerkin method for solving boundary value problems for the Helmholtz equation: A priori and a posteriori error analyses. *arXiv preprint arXiv:1310.2847*, 2013.

[BC00] N. Bartoli and F. Collino. Integral equations via saddle point problem for 2d electromagnetic problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 34(05):1023–1049, 2000.

[BG09] P. B. Bochev and M. D. Gunzburger. *Least-squares finite element methods*. Springer Verlag, 2009.

[BM97] I. Babuška and J. M. Melenk. The partition of unity method. *International Journal for Numerical Methods in Engineering*, 40(4):727–758, 1997.

[BM08] A. Buffa and P. Monk. Error estimates for the ultra weak variational formulation of the Helmholtz equation. *M2AN Math. Model. Numer. Anal.*, 42(6):925–940, 2008.

[BS00] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, 2000.

[BSW15] D. Baskin, E. A. Spence, and J. Wunsch. Sharp high-frequency estimates for the Helmholtz equation and applications to boundary integral equations. *In preparation*, 2015.

[CD98] O. Cessenat and B. Després. Application of an Ultra Weak Variational Formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM Journal on Numerical Analysis*, 35(1):255–299, 1998.

[CD03] F. Collino and B. Després. Integral equations via saddle point problems for time-harmonic Maxwell's equations. *Journal of Computational and Applied Mathematics*, 150(1):157–192, 2003.

[CF06] P. Cummings and X. Feng. Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations. *Mathematical Models and Methods in Applied Sciences*, 16(1):139–160, 2006.

[Cia91] P. G. Ciarlet. Basic error estimates for elliptic problems. In *Handbook of numerical analysis, Vol. II*, pages 17–351. North-Holland, Amsterdam, 1991.

[CK83] D. Colton and R. Kress. *Integral Equation Methods in Scattering Theory*. John Wiley & Sons Inc., New York, 1983.

[CWGLS12] S. N. Chandler-Wilde, I. G. Graham, S. Langdon, and E. A. Spence. Numerical-asymptotic boundary integral methods in high-frequency acoustic scattering. *Acta Numerica*, 21(1):89–305, 2012.

[CWL15] S. N. Chandler-Wilde and S. Langdon. Acoustic scattering: high frequencey boundary element methods and unified transform methods. In A. S. Fokas and B. Pelloni, editors, *Unified transform method for boundary value problems: applications and advances*. SIAM, 2015. to appear.

[DEH+01] J. DeSanto, G. Erdmann, W. Hereman, B. Krause, M. Misra, and E. Swim. Theoretical and computational aspects of scattering from periodic surfaces: two-dimensional perfectly reflecting surfaces using the spectral-coordinate method. *Waves in Random Media*, 11(4):455–488, 2001.

[DEHM98] J. DeSanto, G. Erdmann, W. Hereman, and M. Misra. Theoretical and computational aspects of scattering from rough surfaces: one-dimensional perfectly reflecting surfaces. *Waves in Random Media*, 8(4):385–414, 1998.

[DEHM01] J. DeSanto, G. Erdmann, W. Hereman, and M. Misra. Theoretical and computational aspects of scattering from periodic surfaces: one-dimensional transmission interface. *Waves in Random Media*, 11(4):425–454, 2001.

[Dem06] L. Demkowicz. Babuška ↔ Brezzi?? Technical Report 0608, University of Texas at Austin, Institute for Computational Engineering and Sciences, 2006. Available at http://www.ices.utexas.edu/media/reports/2006/0608.pdf.

[DeS81] J. A. DeSanto. Scattering from a perfectly reflecting arbitrary periodic surface: An exact theory. *Radio Science*, 16(6):1315–1326, 1981.

[Des97] B. Després. Fonctionnelle quadratique et equations integrales pour les problemes d'onde harmonique en domaine exterieur. *Modélisation mathématique et analyse numérique*, 31(6):679–732, 1997.

[Des98] B. Despres. Quadratic functional and integral equations for harmonic wave equations. In *Mathematical and Numerical Aspects of Wave Propagation (Golden, CO)*, pages 56–64. SIAM, Philadelphia, 1998.

[DF14] C-I. R. Davis and B. Fornberg. A spectrally accurate numerical implementation of the Fokas transform method for Helmholtz-type PDEs. *Complex Variables and Elliptic Equations*, 59(4):564–577, 2014.

[DG11] L. Demkowicz and J. Gopalakrishnan. Analysis of the DPG method for the Poisson equation. *SIAM Journal on Numerical Analysis*, 49(5):1788–1809, 2011.

[DG14] L. F. Demkowicz and J. Gopalakrishnan. An Overview of the Discontinuous Petrov Galerkin Method. In X. Feng, O. Karakashian, and Y. Xing, editors, *Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations*, pages 149–180. Springer International Publishing, 2014.

[DGMZ12] L. Demkowicz, J. Gopalakrishnan, I. Muga, and J. Zitelli. Wavenumber explicit analysis for a DPG method for the multidimensional Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, pages 126–138, 2012.

[DTV14] B. Deconinck, T. Trogdon, and V. Vasan. The method of Fokas for solving linear partial differential equations. *SIAM Review*, 56(1):159–186, 2014.

[EES83] S. C. Eisenstat, H. C. Elman, and M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis*, pages 345–357, 1983.

[Elm82] H. C. Elman. *Iterative Methods for Sparse Nonsymmetric Systems of Linear Equations*. PhD thesis, Yale University, 1982.

[Els92] J. Elschner. The double layer potential operator over polyhedral domains I: Solvability in weighted Sobolev spaces. *Applicable Analysis*, 45(1):117–134, 1992.

[EM12] S. Esterhazy and J. M. Melenk. On stability of discretizations of the Helmholtz equation. In I. G. Graham, T. Y. Hou, O. Lakkis, and R. Scheichl, editors, *Numerical Analysis of Multiscale Problems*, volume 83 of *Lecture Notes in Computational Science and Engineering*, pages 285–324. Springer, 2012.

[Eva98] L. C. Evans. *Partial differential equations*. American Mathematical Society Providence, RI, 1998.

[FF11] B. Fornberg and N. Flyer. A numerical implementation of Fokas boundary integral approach: Laplace's equation on a polygonal domain. *Proc. Roy. Soc. A*, 467(2134):2983–3003, 2011.

[FFX04] S. R. Fulton, A. S. Fokas, and C. A. Xenophontos. An analytical method for linear elliptic PDEs and its numerical implementation. *Journal of Computational and Applied Mathematics*, 167(2):465–483, 2004.

[FG94] A. S. Fokas and I. M. Gelfand. Integrability of linear and nonlinear evolution equations and the associated nonlinear Fourier transforms. *Lett. Math. Phys.*, 32:189–210, 1994.

[FHF01] C. Farhat, I. Harari, and L. P. Franca. The discontinuous enrichment method. *Computer Methods in Applied Mechanics and Engineering*, 190(48):6455–6479, 2001.

[FHH03] C. Farhat, I. Harari, and U. Hetmaniuk. A discontinuous Galerkin method with Lagrange multipliers for the solution of Helmholtz problems in the mid-frequency regime. *Computer Methods in Applied Mechanics and Engineering*, 192(11):1389–1419, 2003.

[FIS15] A. S. Fokas, A. Iserles, and S. A. Smitheman. The Unified Method in Polygonal Domains via the Explicit Fourier Transform of Legendre Polynomials. In A. S. Fokas and B. Pelloni, editors, *Unified transform method for boundary value problems: applications and advances*. SIAM, 2015. to appear.

[FJR78] E. B. Fabes, M. Jodeit, and N. M. Riviere. Potential techniques for boundary value problems on C^1 domains. *Acta Mathematica*, 141(1):165–186, 1978.

[FL15] A. S. Fokas and J. Lenells. The unified transform for the modified Helmholtz equation in the exterior of a square. In A. S. Fokas and B. Pelloni, editors, *Unified transform method for boundary value problems: applications and advances*. SIAM, 2015. to appear.

[Fok97] A. S. Fokas. A unified transform method for solving linear and certain nonlinear PDEs. *Proc. R. Soc. Lond. A*, 453:1411–1443, 1997.

[Fok08] A. S. Fokas. *A Unified Approach to Boundary Value Problems*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, 2008.

[FS12] A. S. Fokas and E. A. Spence. Synthesis, as opposed to separation, of variables. *SIAM Review*, 54(2):291–324, 2012.

[FW09] X. Feng and H. Wu. Discontinuous Galerkin methods for the Helmholtz equation with large wave number. *SIAM Journal on Numerical Analysis*, 47(4):2872–2896, 2009.

[FW11] X. Feng and H. Wu. *hp*-Discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Mathematics of Computation*, 80(276):1997–2024, 2011.

[FX13] X. Feng and Y. Xing. Absolutely stable local discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Mathematics of Computation*, 82(283):1269–1296, 2013.

[Gab07] G. Gabard. Discontinuous Galerkin methods with plane waves for time-harmonic problems. *Journal of Computational Physics*, 225(2):1961–1984, 2007.

[GGS15] M. J. Gander, I. G. Graham, and E. A. Spence. Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: What is the largest shift for which wavenumber-independent convergence is guaranteed? *Numerische Mathematik*, to appear, 2015.

[GHP09] C. J. Gittelson, R. Hiptmair, and I. Perugia. Plane wave discontinuous Galerkin methods: analysis of the h-version. *M2AN Math. Model. Numer. Anal.*, 2:297–331, 2009.

[Gop13] J. Gopalakrishnan. Five lectures on DPG methods. *arXiv preprint arXiv:1306.0557*, 2013.

[GR97] K. E. Gustafson and D. K. M. Rao. *Numerical range; The field of values of linear operators and matrices*. Universitext. Springer-Verlag, New York, 1997.

[Gra09] I. G. Graham. Advanced finite element methods. Available at `http://www.maths.bath.ac.uk/~masigg/ma60202/lectures.pdf`, 2009.

[Gre97] A. Greenbaum. *Iterative methods for solving linear systems*. SIAM, 1997.

[Gri85] P. Grisvard. *Elliptic problems in nonsmooth domains*. Pitman, Boston, 1985.

[HMP11] R. Hiptmair, A. Moiola, and I. Perugia. Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the p-version. *SIAM J. Numer. Anal.*, 49:264–284, 2011.

[HW08] G. C. Hsiao and W. L. Wendland. *Boundary integral equations*, volume 164 of *Applied Mathematical Sciences*. Springer, 2008.

[JK95] D. Jerison and C. E. Kenig. The inhomogeneous Dirichlet problem in Lipschitz domains. *J. Funct. Anal.*, 130:161–219, 1995.

[Kat60] T. Kato. Estimation of iterated matrices, with application to the von Neumann condition. *Numerische Mathematik*, 2(1):22–29, 1960.

[Kee95] J. P. Keener. *Principles of Applied Mathematics*. Perseus Books, 1995.

[LHM09] T. Luostari, T. Huttunen, and P. Monk. Plane wave methods for approximating the time harmonic wave equation. In B. Engquist, A. Fokas, E. Hairer, and A. Iserles, editors, *Highly Oscillatory Problems: Computation, Theory and Applications*. Cambridge University Press, 2009.

[Luo13] T. Luostari. *Non-polynomial approximation methods in acoustics and elasticity*. PhD thesis, University of Eastern Finland, 2013.

[Mar06] P. A. Martin. *Multiple scattering: interaction of time-harmonic waves with N obstacles*. Cambridge University Press, 2006.

[MB96] J. M. Melenk and I. Babuška. The partition of unity finite element method: Basic theory and applications. *Comput. Method Appl. M.*, 139:289–314, 1996.

[McL00] W. McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, 2000.

[Mel95] J. M. Melenk. *On generalized finite element methods*. PhD thesis, The University of Maryland, 1995.

[ML68] C. S. Morawetz and D. Ludwig. An inequality for the reduced wave operator and the justification of geometrical optics. *Communications on Pure and Applied Mathematics*, 21:187–203, 1968.

[Moi11] A. Moiola. *Trefftz-discontinuous Galerkin methods for time-harmonic wave problems*. PhD thesis, Seminar for applied mathematics, ETH Zürich, 2011. Available at http://e-collection.library.ethz.ch/view/eth:4515.

[Mor61] C. S. Morawetz. The decay of solutions of the exterior initial-boundary value problem for the wave equation. *Communications on Pure and Applied Mathematics*, 14(3):561–568, 1961.

[Mor75] C. S. Morawetz. Decay for solutions of the exterior problem for the wave equation. *Communications on Pure and Applied Mathematics*, 28(2):229–264, 1975.

[MS14] A. Moiola and E. A. Spence. Is the Helmholtz equation really sign-indefinite? *SIAM Review*, 56(2):274–312, 2014.

[MW99] P. Monk and D. Q. Wang. A least-squares method for the Helmholtz equation. *Computer Methods in Applied Mechanics and Engineering*, 175(1):121–136, 1999.

[Nai67] M. A. Naimark. *Linear differential operators. Part I: Elementary theory of linear differential operators*. Frederick Ungar Publishing Co, New York, 1967.

[Néd01] J. C. Nédélec. *Acoustic and electromagnetic equations: integral representations for harmonic problems*. Springer Verlag, 2001.

[NIS14] NIST. Digital Library of Mathematical Functions. Digital Library of Mathematical Functions, http://dlmf.nist.gov/, 2014.

[Nit68] J. Nitsche. Ein kriterium für die quasi-optimalität des ritzschen verfahrens. *Numerische Mathematik*, 11(4):346–348, 1968.

[PS02] I. Perugia and D. Schötzau. An *hp*-analysis of the local discontinuous Galerkin method for diffusion problems. *Journal of Scientific Computing*, 17(1-4):561–571, 2002.

[Say13] F. J. Sayas. Retarded potentials and time domain boundary integral equations: a road-map. *preprint*, 2013. Available at http://www.math.udel.edu/~fjsayas/TDBIEclassnotes2012.pdf.

[Sch74] A. H. Schatz. An observation concerning Ritz-Galerkin methods with indefinite bilinear forms. *Mathematics of Computation*, 28(128):959–962, 1974.

[SFFS08] A. G. Sifalakis, A. S. Fokas, S. R. Fulton, and Y. G. Saridakis. The generalized Dirichlet-Neumann map for linear elliptic PDEs and its numerical implementation. *J. Comp. Appl. Math.*, 219(1):9–34, 2008.

[SFPS09] A. G. Sifalakis, S. R. Fulton, E. P. Papadopoulou, and Y. G. Saridakis. Direct and iterative solution of the generalized Dirichlet–Neumann map for elliptic PDEs on square domains. *Journal of Computational and Applied Mathematics*, 227(1):171–184, 2009.

[SKS15] E. A. Spence, I. V. Kamotski, and V. P. Smyshlyaev. Coercivity of combined boundary integral equations in high frequency scattering. *Comm. Pure Appl. Math.*, to appear, 2015.

[Spe14] E. A. Spence. Wavenumber-explicit bounds in time-harmonic acoustic scattering. *SIAM Journal on Mathematical Analysis*, 46(4):2987–3024, 2014.

[SPS07] A. G. Sifalakis, E. P. Papadopoulou, and Y. G. Saridakis. Numerical study of iterative methods for the solution of the Dirichlet-Neumann map for linear elliptic PDEs on regular polygon domains. *Int. J. Appl. Math. Comput. Sci*, 4:173–178, 2007.

[SS11] S. A. Sauter and C. Schwab. *Boundary Element Methods*. Springer-Verlag, Berlin, 2011.

[SSF10] S. A. Smitheman, E. A. Spence, and A. S. Fokas. A spectral collocation method for the Laplace and modified Helmholtz equations in a convex polygon. *IMA J. Num. Anal.*, 30(4):1184–1205, 2010.

[SSP12] Y. G. Saridakis, A. G. Sifalakis, and E. P. Papadopoulou. Efficient numerical solution of the generalized Dirichlet–Neumann map for linear elliptic PDEs in regular polygon domains. *Journal of Computational and Applied Mathematics*, 236(9):2515–2528, 2012.

[Sta68] I. Stakgold. *Boundary value problems of mathematical physics, Volume II.* New York: The Macmillan Company; London: Collier-Macmillan Ltd, 1968.

[Sta79] I. Stakgold. *Green's functions and boundary value problems.* Wiley, New York, 1979.

[Ste08] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements.* Springer, New York, 2008.

[SZ90] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Mathematics of Computation*, 54(190):483–493, 1990.

[Wat65] P. C. Waterman. Matrix formulation of electromagnetic scattering. *Proceedings of the IEEE*, 53(8):805–812, 1965.

[Wat69] P. C. Waterman. New formulation of acoustic scattering. *The Journal of the Acoustical Society of America*, 45:1417, 1969.

[Wu14] H. Wu. Pre-asymptotic error analysis of CIP-FEM and FEM for the Helmholtz equation with high wave number. Part I: linear version. *IMA Journal of Numerical Analysis*, 34(3):1266–1288, 2014.

[XZ03] J. Xu and L. Zikatanov. Some observations on Babuška and Brezzi theories. *Numerische Mathematik*, 94(1):195–202, 2003.