ANALYSIS OF A SPECTRAL-GALERKIN APPROXIMATION TO THE HELMHOLTZ EQUATION IN EXTERIOR DOMAINS

JIE SHEN[1]   LI-LIAN WANG[2]

ABSTRACT. An error analysis is presented for the spectral-Galerkin method to the Helmholtz equation in 2-D and 3-D exterior domains. The problem in unbounded domains is first reduced to a problem on a bounded domain via the Dirichlet-to-Neumann operator, then a spectral-Galerkin method is employed to approximate the reduced problem. The error analysis is based on exploring delicate asymptotic behaviors of the Hankel functions and on deriving *a priori* estimates with explicit dependence on the wave number for both the continuous and discrete problems. Explicit error bounds with respect to the wave number are derived and some illustrative numerical examples are also presented.

1. INTRODUCTION

We consider in this paper the acoustic wave scattering from a bounded obstacle $D \subset \mathbb{R}^d$, $d = 2, 3$. In this case, the scattered wave satisfies the Helmholtz equation

$$(1.1) \qquad -\Delta U - k^2 U = F, \text{ in } \mathbb{R}^d \backslash \bar{D},$$

along with the Sommerfeld radiation condition at infinity

$$(1.2) \qquad \partial_r U - \mathrm{i}kU = o\big(r^{\frac{1-d}{2}}\big) \quad \text{as } r \to \infty, \ d = 2, 3,$$

which ensures that waves do not reflect from far field. On the surface of the scatterer D, a Dirichlet (sound soft) or Neumann (sound hard) condition is assumed.

Although the Helmholtz equation with (1.2) is linear, its numerical approximation and associated analysis is notoriously difficult due to (i) the domain is unbounded; (ii) the system is not positive definite; and (iii) when the wave number $k \gg 1$, the solution is highly oscillatory. In particular, it remains a challenge to design numerical algorithms which are robust and efficient for moderate to high wave numbers.

There have been extensive research work devoted to overcoming these difficulties (see, for instance, [13, 18, 17] and the references therein). In particular, it has been shown, at least for some simple cases, that errors of p-th order numerical methods for the Helmholtz equation behave like $O(k^{p+1}h^p)$ (see, for instance, [15, 2, 25]). Hence, high-order methods are particularly preferable for this type of problems over low-order methods. Moreover, the linear system from a discretization of the Helmholtz equation with moderate to high wave numbers is usually highly indefinite and difficult to solve. It is with these considerations in mind that we choose to use the transformed field expansion (TFE) method (cf. [21]),

which improves over the classical field expansion method [22, 3, 4], coupled with a fast spectral-Galerkin solver (cf. [23, 24, 19]).

We now briefly describe the TFE method for a 2-D obstacle enclosed by $\{r = a + g(\theta) : 0 \le \theta < 2\pi\}$. The TFE algorithm consists of the following steps:

- Assuming F is compactly supported and choosing b such that $b > a + \max_{0 \le \theta < 2\pi} |g(\theta)|$ and supp$F \subset \Omega_g := \{(r, \theta) : a + g(\theta) < r < b\}$, we then use the Dirichlet-to-Neumann operator T (see [12, 10] and the next section), to reduce the problem in the unbounded domain to

$$(1.3) \qquad \begin{aligned} & -\Delta U - k^2 U = F, \quad \text{in } \Omega_g, \\ & U|_{r=a+g(\theta)} = \xi, \qquad (\partial_r U + T(U))|_{r=b} = 0. \end{aligned}$$

- Make a change of variables,

$$(1.4) \qquad r' = \frac{(b-a)r - bg(\theta)}{(b-a) - g(\theta)}, \qquad \theta' = \theta,$$

which maps Ω_g to an annulus Ω_0. To simplify the notation, we still use (r, θ) to denote (r', θ'), and U, F, ξ to denote the functions U, F, ξ after the change of variables. Then, the problem (1.3) becomes

$$(1.5) \qquad \begin{aligned} & -\Delta U - k^2 U = F + J(g, U) \quad \text{in } \Omega_0, \\ & U(a, \theta) = \xi(\theta), \quad (\partial_r U + TU)|_{r=b} = \eta(g, U), \end{aligned}$$

where $J(g, U)$ and $\eta(g, U)$ contain differential operators with non-constant coefficients for which fast direct/iterative solver is not available.

- Consider the obstacle $\{(r, \theta) : r < a + g(\theta)\}$ as a perturbation of the disk $\{r < a\}$, i.e., write $g = \varepsilon h$ and expand u as

$$U(r, \theta; \varepsilon) = \sum_{n=0}^{\infty} U_n(r, \theta) \, \varepsilon^n.$$

Plug the above expansion into (1.5) and collect terms with ε^n, we find that [19]

$$(1.6) \qquad \begin{aligned} & -\Delta U_n - k^2 U_n = \delta_{n,0} F + \tilde{J}(g, U_{n-4}, \cdots, U_{n-1}) \quad \text{in } \Omega_0, \\ & U_n(a, \theta) = \delta_{n,0} \xi(\theta), \quad (\partial_r U_n + TU_n)|_{r=b} = \tilde{\eta}(g, U_{n-1}). \end{aligned}$$

- Solve (1.6) for $n = 0, 1, 2, \cdots$, and sum up the series by using a Padé approximation.

It is shown in [20, 21] that this TFE method is stable and robust at high-order, and it is demonstrated in [19] that this method, coupled with a spectral-Galerkin solver for (1.6), is very efficient and capable of providing very accurate results for bounded obstacle scattering with moderate to high wave numbers.

Notice that the whole algorithm boils down to solving a sequence of the following non-homogeneous Helmholtz equation in an annulus (2-D) or a spherical shell (3-D):

$$(1.7) \qquad \begin{aligned} & -\Delta U - k^2 U = F \quad \text{in } \Omega_0, \\ & U(a, \theta) = \xi(\theta), \quad (\partial_r U + TU)|_{r=b} = \eta(\theta). \end{aligned}$$

The purpose of this paper is to present a detailed error analysis of the spectral-Galerkin method for (1.7). The main difficulty here is to obtain error estimates with explicit dependence on the wave number. Among the very few results available in this regards are those in [15, 25] where the Helmholtz equation in bounded domains with a first-order approximation to the radiation boundary condition was considered and error estimates with explicit dependence on the wave numbers were derived. To the authors' best knowledge,

there seems to be no rigorous error estimate available with explicit dependence on the wave number for a numerical scheme to bounded obstacle scattering.

We now introduce some notations to be used throughout this paper. Let ϖ be a given positive weight function in $I := (a, b)$. We denote by $L_\varpi^2(I)$ a Hilbert space of real or complex functions with inner product and norm

$$(u, v)_\varpi = \int_I u(r)\overline{v(r)}\varpi(r)dr, \qquad \|u\|_\varpi = \sqrt{(u, u)_\varpi},$$

where $\bar{v}$ is the complex conjugate of v. Then, the weighted Sobolev spaces $H_\varpi^s(I)$ ($s = 0, 1, 2, \cdots$) can be defined as usual with inner products, norms and semi-norms denoted by $(\cdot, \cdot)_{s,\varpi}$, $\|\cdot\|_{s,\varpi}$ and $|\cdot|_{s,\varpi}$, respectively. For real $s > 0$, $H_\varpi^s(I)$ is defined by space interpolation (cf. [16]). The subscript ϖ will be omitted from the notations in case of $\varpi \equiv 1$. For simplicity, we denote $\partial_r^l v = \frac{d^l v}{dr^l}$, $l \geq 1$. We shall also use $(\cdot, \cdot)_\omega$ and $\|\cdot\|_\omega$ to denote the weighted inner product and the weighted L^2-norm in two and three dimensions.

Let S be the unit circle in 2-D and unit sphere in 3-D, we also use the non-isotropic periodic-type Sobolev space on $\Omega = S \times I$: $H_p^{s'}(S; H_\varpi^s(I))$, $s' \geq 0$ (subscript p stands for periodicity in the θ-direction) with the norm

$$(1.8) \qquad \|U\|_{H_p^{s'}(S;H_\varpi^s(I))} = \begin{cases} \left(\sum_{|m|=0}^\infty (1+m^2)^{s'} \|\hat{u}_m\|_{s,\varpi}^2\right)^{1/2}, & \text{if } d = 2, \\ \left(\sum_{m=0}^\infty \sum_{l=-m}^m (1+m)^{2s'} \|\hat{u}_{lm}\|_{s,\varpi}^2\right)^{1/2}, & \text{if } d = 3, \end{cases}$$

where $\{\hat{u}_m\}$ (resp. $\{\hat{u}_{lm}\}$) are the expansion coefficients of U in terms of Fourier (resp. spherical harmonic) basis, i.e.,

$$(1.9) \qquad U = \sum_{|m|=0}^\infty \hat{u}_m e^{im\theta} \quad \text{or} \quad U = \sum_{m=0}^\infty \sum_{l=-m}^m \hat{u}_{lm} Y_m^l(\theta, \phi).$$

The norm of the Sobolev space $H_p^{s'}(S)$ on S can be defined in the same fashion by replacing the norm $\|\cdot\|_{s,\varpi}$ by the absolute value $|\cdot|$. In particular, we have

$$L_p^2(\Omega) = H_p^0(S, H_\varpi^0(I)) \text{ with } \varpi = r^{d-1}, \ d = 2, 3.$$

We assume that $a > 0$ is a fixed parameter representing the radius of the scatterer. Throughout this paper, we denote by c a generic positive constant which only depends on a and possibly on a fixed $k_0 > 0$. We use the expression $A \lesssim B$ to mean that $A \leq cB$.

2. Dirichlet-to-Neumann (DtN) map

The error analysis relies heavily on the properties of DtN map which we investigate below.

2.1. **Formulation of DtN operator.** We start with the 3-D case and consider an "auxiliary" exterior problem

$$(2.1) \qquad \begin{cases} -\Delta U - k^2 U = 0, & \text{in } \Omega_{\text{ext}} := \mathbb{R}^3 \setminus \bar{B}, \\ U = \Psi, & \text{on } \partial B, \end{cases}$$

where B is a ball of radius b. This problem can be solved analytically via separation of variables, namely, we can express its solution as

$$(2.2) \qquad U(r, \theta, \phi) = \sum_{m=0}^\infty h_m^{(1)}(kr) \sum_{l=-m}^m \hat{u}_{lm} Y_m^l(\theta, \phi),$$

where $(r, \theta, \phi) \in [b, \infty) \times [0, 2\pi) \times [0, \pi)$, $h_m^{(1)}(z)$ is the spherical Hankel functions of the first kind of order m, and $\{Y_m^l\}$ are the spherical harmonic functions (see Appendix for their properties). To determine the coefficients $\{\hat{u}_{lm}\}$, we expand the Dirichlet boundary value Ψ on the sphere ∂B as

$$(2.3) \qquad U(b, \theta, \phi) = \Psi(\theta, \phi) = \sum_{m=0}^{\infty} \sum_{l=-m}^{m} \hat{\psi}_{lm} Y_m^l(\theta, \phi).$$

Letting $r = b$ in (2.2) and comparing the coefficients of the two expansions yield that

$$(2.4) \qquad \hat{u}_{lm} = \frac{\hat{\psi}_{lm}}{h_m^{(1)}(kb)}, \quad \text{for} \quad m \geq |l| \geq 0.$$

Plugging it into (2.2) leads to the exact solution of (2.1):

$$(2.5) \qquad U(r, \theta, \phi) = \sum_{m=0}^{\infty} \frac{h_m^{(1)}(kr)}{h_m^{(1)}(kb)} \sum_{l=-m}^{m} \hat{\psi}_{lm} Y_m^l(\theta, \phi).$$

Differentiating (2.5) with respect to r and setting $r = b$, we find

$$(2.6) \qquad \partial_r U(b, \theta, \phi) = \sum_{m=0}^{\infty} k \frac{h_m^{(1)'}(kb)}{h_m^{(1)}(kb)} \sum_{l=-m}^{m} \hat{\psi}_{lm} Y_m^l(\theta, \phi).$$

Hence, the DtN map is defined explicitly as

$$(2.7) \qquad T(U) = \frac{\partial U}{\partial \mathbf{n}}\Big|_{\partial B} = -\frac{\partial U}{\partial r}\Big|_{r=b} = -\sum_{m=0}^{\infty} k \frac{h_m^{(1)'}(kb)}{h_m^{(1)}(kb)} \sum_{l=-m}^{m} \hat{\psi}_{lm} Y_m^l(\theta, \phi),$$

where $\mathbf{n}$ is the outward normal of Ω_{ext}.

The counterpart of (2.1) in 2-D is

$$(2.8) \qquad \begin{cases} -\Delta U - k^2 U = 0, & \text{in } \Omega_{\text{ext}} := \mathbb{R}^2 \setminus \bar{B}, \\ U = \Phi, & \text{on } \partial B, \end{cases}$$

where B is a circle of radius b which can be solved analytically with the exact solution

$$(2.9) \qquad U(r, \theta) = \sum_{|m|=0}^{\infty} \hat{u}_m H_m^{(1)}(kr) e^{im\theta}, \quad \forall (r, \theta) \in [b, \infty) \times [0, 2\pi).$$

Here, $H_m^{(1)}(z)$ is the Hankel function of the first kind of order m. The coefficients $\{\hat{u}_m\}$ are determined by the boundary value $\Phi(\theta)$ with the expansion

$$(2.10) \qquad U(b, \theta) = \Phi(\theta) = \sum_{|m|=0}^{\infty} \hat{\phi}_m e^{im\theta}.$$

Hence, letting $r = b$ in (2.9), and comparing the coefficients of the above two expansions lead to $\hat{u}_m = \hat{\phi}_m / H_m^{(1)}(kb)$. As a consequence, the exact solution of (2.8) is

$$(2.11) \qquad U(r, \theta) = \sum_{|m|=0}^{\infty} \frac{H_m^{(1)}(kr)}{H_m^{(1)}(kb)} \hat{\phi}_m e^{im\theta}, \quad \forall (r, \theta) \in [b, \infty) \times [0, 2\pi).$$

The 2-D DtN map is given by

$$(2.12) \qquad T(U) = \frac{\partial U}{\partial \mathbf{n}}\Big|_{\partial B} = -\frac{\partial U}{\partial r}\Big|_{r=b} = -\sum_{|m|=0}^{\infty} k \frac{\partial_z H_m^{(1)}(kb)}{H_m^{(1)}(kb)} \hat{\phi}_m e^{im\theta}.$$

By using the DtN map T and choosing b sufficiently large so that B contains both D and supp$F$, the original problem (1.1)–(1.2) with Dirichlet boundary condition is reduced to:

(2.13)
$$\begin{cases} -\Delta U - k^2 U = F, & \text{in } \Omega := B \cap \mathbb{R}^d \setminus \bar{D}, \ d = 2, 3, \\ U = \xi, & \text{on } \partial D, \\ \partial_r U + TU = 0, & \text{on } \partial B. \end{cases}$$

To fix the idea, we prescribed a Dirichlet boundary condition on the scatterer D, other type of boundary conditions can be used as well.

2.2. Properties of the DtN kernel.

In order to carry out a rigorous mathematical analysis for the problem (2.13), we need to study carefully the properties of the DtN kernel associated with (2.7) and (2.12), i.e., the properties of the coefficients:

(2.14)
$$\mathcal{T}_{m,\kappa} = \begin{cases} \dfrac{\partial_z H_m^{(1)}(\kappa)}{H_m^{(1)}(\kappa)}, & \text{if } d = 2, \\[3mm] \dfrac{\partial_z h_m^{(1)}(\kappa)}{h_m^{(1)}(\kappa)}, & \text{if } d = 3. \end{cases}$$

2.2.1. *Behavior of the 3-D kernel.* In this case, we have $\kappa > 0$ and $m \geq 0$. We recall that

(2.15)
$$h_m^{(1)}(\kappa) = j_m(\kappa) + \mathrm{i} y_m(\kappa) = \sqrt{\frac{\pi}{2\kappa}} J_{m+1/2}(\kappa) + \mathrm{i}\sqrt{\frac{\pi}{2\kappa}} Y_{m+1/2}(\kappa),$$

where J_ν & Y_ν (resp. j_ν & y_ν) are the Bessel (resp. spherical Bessel) functions of the first and second kinds of order ν. Using the relevant properties of the Bessel functions (cf. [26]), one verifies that

(2.16a)
$$\begin{aligned} \mathrm{Re}(\mathcal{T}_{m,\kappa}) &= \frac{m}{\kappa} - \frac{j_m(\kappa)j_{m+1}(\kappa) + y_m(\kappa)y_{m+1}(\kappa)}{\left|h_m^{(1)}(\kappa)\right|^2} \\ &= \frac{m}{\kappa} - \frac{J_{m+1/2}(\kappa)J_{m+3/2}(\kappa) + Y_{m+1/2}(\kappa)Y_{m+3/2}(\kappa)}{J_{m+1/2}^2(\kappa) + Y_{m+1/2}^2(\kappa)}; \end{aligned}$$

(2.16b)
$$\mathrm{Im}(\mathcal{T}_{m,\kappa}) = \frac{1}{\kappa^2 \left|h_m^{(1)}(\kappa)\right|^2} = \frac{2}{\pi\kappa} \frac{1}{J_{m+1/2}^2(\kappa) + Y_{m+1/2}^2(\kappa)}.$$

An explicit expression of $\mathcal{T}_{m,\kappa}$ is given by Theorem 2.6.1 of [18]:

(2.17)
$$\mathcal{T}_{m,\kappa} = \mathrm{Re}(\mathcal{T}_{m,\kappa}) + \mathrm{i}\,\mathrm{Im}(\mathcal{T}_{m,\kappa}) = -\frac{P_m(\kappa)}{\kappa Q_m(\kappa)} + \frac{\mathrm{i}}{Q_m(\kappa)},$$

where

(2.18)
$$\begin{aligned} P_m(\kappa) &= 1 + 2a_1^m \frac{1}{\kappa^2} + \cdots + (m+1)a_m^m \frac{1}{\kappa^{2m}}, \\ Q_m(\kappa) &= 1 + a_1^m \frac{1}{\kappa^2} + \cdots + a_m^m \frac{1}{\kappa^{2m}}, \end{aligned}$$

with

(2.19)
$$a_j^m = \frac{(m+j)!(2j)!}{4^j(j!)^2(m-j)!}.$$

We now study the monotonic property of $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ with respect to m and κ. We observe from (2.17)–(2.19) that for a fixed $m \geq 0$, $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ is an *increasing* function of κ, as

illustrated by Figure 2.1 (b). However, for a given $\kappa > 0$, $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ is a *decreasing* function of m, which follows from the Nicholson's formula (see P. 444 of [26])

$$(2.20) \qquad J_{m+1/2}^2(\kappa) + Y_{m+1/2}^2(\kappa) = \frac{8}{\pi^2} \int_0^{+\infty} \mathcal{K}_0(2\kappa\sinh t)\cosh\big((2m+1)t\big)dt,$$

where $\mathcal{K}_0(\xi) > 0$ is the Kelvin's function defined by (A.2) in the appendix.

We next consider the bounds and asymptotic behavior of $\mathcal{T}_{m,\kappa}$. An immediate consequence of (2.17)–(2.19) is that

$$(2.21) \qquad \mathrm{Re}(\mathcal{T}_{m,\kappa}) < 0, \qquad \mathrm{Im}(\mathcal{T}_{m,\kappa}) > 0,$$

which ensures the well-posedness of the problem (2.13) (cf. [8]). Moreover, we have the following bounds (see, e.g., P.87 of [18]):

$$(2.22) \qquad -\frac{m+1}{\kappa} \le \mathrm{Re}(\mathcal{T}_{m,\kappa}) \le -\frac{1}{\kappa}, \qquad 0 < \mathrm{Im}(\mathcal{T}_{m,\kappa}) \le 1,$$

in particular, by (2.17)–(2.19),

$$(2.23) \qquad \mathrm{Re}(\mathcal{T}_{0,\kappa}) = -\frac{1}{\kappa}, \qquad \mathrm{Im}(\mathcal{T}_{0,\kappa}) = 1.$$
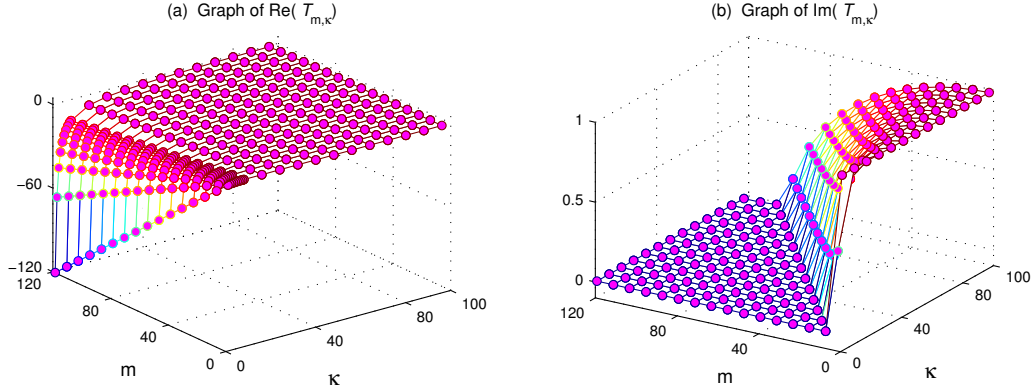

FIGURE 2.1. Graphs of $\mathrm{Re}(\mathcal{T}_{m,\kappa})$ and $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ with $(\kappa, m) \in [1, 100] \times [0, 120]$ in the 3-D case.

We now seek more precise estimates of $\mathrm{Im}(\mathcal{T}_{m,\kappa})$, and proceed separately with three cases:

(i) $\underline{\kappa > m + 1/2}$. We first recall the estimate (see P. 447 of [26]):

$$(2.24) \qquad \frac{2}{\pi\kappa} < J_\nu^2(\kappa) + Y_\nu^2(\kappa) < \frac{2}{\pi\sqrt{\kappa^2 - \nu^2}}, \quad \text{if} \quad \frac{1}{2} \le \nu < \kappa,$$

which, together with (2.16b), implies that

$$(2.25) \qquad E_{m,\kappa} := \frac{\sqrt{\kappa^2 - (m+1/2)^2}}{\kappa} < \mathrm{Im}(\mathcal{T}_{m,\kappa}) < 1, \quad \text{if} \quad \kappa > m + \frac{1}{2}.$$

We observe from Figure 2.2 that the lower bound $E_{m,\kappa}$ provides an acceptable approximation to $\mathrm{Im}(\mathcal{T}_{m,\kappa})$.

(ii) $\underline{\kappa = m + 1/2}$. Using the formulas (see P. 232 of [26]):

$$(2.26) \qquad J_\nu(\nu) = C_1\nu^{-1/3} + O(\nu^{-5/3}), \qquad Y_\nu(\nu) = -C_2\nu^{-1/3} + O(\nu^{-5/3}),$$

with

$$C_1 = \frac{\Gamma(1/3)}{2^{2/3}3^{1/6}\pi} \approx 0.4473, \quad C_2 = \frac{3^{1/3}\Gamma(1/3)}{2^{2/3}\pi} \approx 0.7748,$$

we obtain from (2.16b) that

$$(2.27) \qquad \mathrm{Im}\big(\mathcal{T}_{m,m+1/2}\big) \sim C_\circ\big(m+1/2\big)^{-1/3} := E_{m,m+1/2},$$

with $C_\circ = 2/\big(\pi(C_1^2 + C_2^2)\big) \approx 0.7954$.

In Figure 2.2 (a), we plot $\mathrm{Im}(\mathcal{T}_{m,m+1/2})$ against $E_{m,m+1/2}$ for $1 \le m \le 200$, which shows that even for small m, the asymptotic estimate $C_\circ\big(m+1/2\big)^{-1/3}$ provides a very good approximation to $\mathrm{Im}\big(\mathcal{T}_{m,m+1/2}\big)$.

(iii) $\underline{\kappa < m + 1/2}$. By the asymptotic formulas (see P. 243 of [26]),

$$(2.28) \qquad J_\nu(\nu\mathrm{sech}\alpha) \sim \frac{e^{\nu(\tanh\alpha - \alpha)}}{\sqrt{2\pi\nu\tanh\alpha}}, \qquad Y_\nu(\nu\mathrm{sech}\alpha) \sim -\frac{e^{\nu(\alpha - \tanh\alpha)}}{\sqrt{\frac{1}{2}\pi\nu\tanh\alpha}},$$

one verifies that for $m + 1/2 = \kappa\cosh\alpha$ with $\alpha > 0$,

$$(2.29) \qquad \mathrm{Im}(\mathcal{T}_{m,\kappa}) \sim \frac{2(2m+1)\tanh\alpha}{\kappa\big[e^{(2m+1)(\tanh\alpha - \alpha)} + 4e^{(2m+1)(\alpha - \tanh\alpha)}\big]} := E_{m,\kappa}.$$

Hence, $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ becomes exponentially small for large m. The exponential decay of $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ is shown more clearly from the asymptotic estimate

$$(2.30) \qquad \mathrm{Im}(\mathcal{T}_{m,\kappa}) \sim \Big(\frac{e\kappa}{2m+1}\Big)^{2m}, \quad m \gg \kappa,$$

which follows from formula 9.3.1 of [1],

$$(2.31) \qquad J_\nu(\kappa) \sim \frac{1}{\sqrt{2\pi\nu}}\Big(\frac{e\kappa}{2\nu}\Big)^\nu, \quad Y_\nu(\kappa) \sim -\frac{2}{\sqrt{\pi\nu}}\Big(\frac{e\kappa}{2\nu}\Big)^{-\nu}, \quad \nu \gg \kappa.$$

We plot in Figure (2.2) (b) the estimate $E_{m,\kappa}$ (defined in (2.25), (2.27) and (2.29)) versus $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ with $\kappa = 20, 30, 40, 50$ and various m, which indicates that $E_{m,\kappa}$ provides an accurate picture of $\mathrm{Im}(\mathcal{T}_{m,\kappa})$.

2.2.2. *Behavior of the 2-D kernel.* The identity $H_{-\nu}^{(1)}(z) = e^{\nu\pi i}H_\nu^{(1)}(z)$ and the definition (2.14) imply that

$$(2.32) \qquad \mathcal{T}_{-m,\kappa} = \frac{\partial_z H_{-m}^{(1)}(\kappa)}{H_{-m}^{(1)}(\kappa)} = \frac{(-1)^m\partial_z H_m^{(1)}(\kappa)}{(-1)^m H_m^{(1)}(\kappa)} = \mathcal{T}_{m,\kappa}.$$

Hence, it suffices to consider $\mathcal{T}_{m,\kappa}$ with $m \ge 0$. Using the recursion formulas of the Bessel functions, one verifies that

$$(2.33a) \qquad \mathrm{Re}(\mathcal{T}_{m,\kappa}) = \frac{m}{\kappa} - \frac{J_m(\kappa)J_{m+1}(\kappa) + Y_m(\kappa)Y_{m+1}(\kappa)}{J_m^2(\kappa) + Y_m^2(\kappa)};$$

$$(2.33b) \qquad \mathrm{Im}(\mathcal{T}_{m,\kappa}) = \frac{2}{\pi\kappa}\frac{1}{|H_m^{(1)}(\kappa)|^2} = \frac{2}{\pi\kappa}\frac{1}{J_m^2(\kappa) + Y_m^2(\kappa)}.$$

We observe that the 2-D kernel has an expression similar to that of the 3-D kernel (cf. (2.16)). In fact, they share similar properties and asymptotic behaviors except for $m = 0$ (comparison: Figure 2.1 (a) versus Figure 2.3 (a), and Figure 2.2 (b) versus Figure 2.3 (b)).
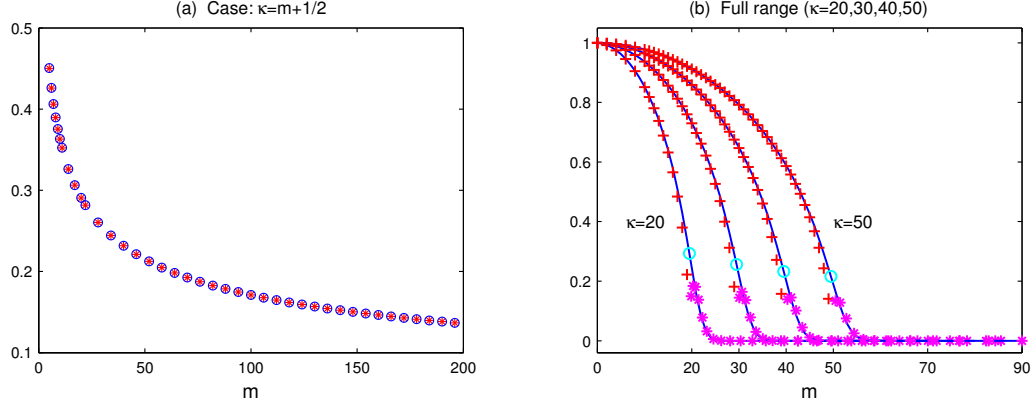
FIGURE 2.2. (a) $\mathrm{Im}(\mathcal{T}_{m,m+1/2})$ ("$\star$") against $E_{m,m+1/2}$ ("$\circ$") with $m \in [1, 200]$; (b) $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ ("solid line") against $E_{m,\kappa}$ ("+" for $\kappa > m + 1/2$; "$\circ$" for $k = m + 1/2$ and "$\star$" for $\kappa < m + 1/2$) with $\kappa = 20, 30, 40, 50$.

Indeed, we notice that the same monotonic property holds for the 2-D $\mathrm{Im}(\mathcal{T}_{m,\kappa})$:

(**i**) for a given $m \geq 1$, $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ is a strictly *increasing* function of κ, which follows from (2.33b) and the fact that $\kappa|H_m^{(1)}(\kappa)|^2$ is a strictly decreasing function of κ (cf. P. 446 of [26]);

(**ii**) for a fixed $\kappa > 0$, $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ is a strictly *decreasing* function of m, which is a direct consequence of the Nicholson's formula (A.3a).

As in (2.22), we have the following bound for the 2-D kernel (see Appendix for the proof):

$$(2.34a) \qquad\qquad\qquad 0 < \mathrm{Im}(\mathcal{T}_{m,\kappa}) < 1, \ m \geq 1;$$

$$(2.34b) \qquad -\frac{1}{2\kappa} \leq \mathrm{Re}(\mathcal{T}_{0,\kappa}) < 0, \quad -\frac{m}{\kappa} \leq \mathrm{Re}(\mathcal{T}_{m,\kappa}) \leq -\frac{1}{2\kappa}, \quad m \geq 1;$$

$$(2.34c) \qquad\qquad\qquad \mathrm{Im}(\mathcal{T}_{0,\kappa}) > 1, \ \ \forall \kappa > 0.$$

As in the 3-D case, applying the general formulas (2.24), (2.26) and (2.28) to the 2-D $\mathrm{Im}(\mathcal{T}_{m,\kappa})$, we find that an accurate approximation for $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ is

$$(2.35) \qquad E_{m,\kappa} := \begin{cases} \sqrt{1 - m^2/\kappa^2}, & \text{if } \kappa > m \geq 1, \\ C_\circ m^{-1/3}, & \text{if } \kappa = m, \\ \dfrac{4m\tanh\alpha}{\kappa\left[e^{2m(\tanh\alpha - \alpha)} + 4e^{2m(\alpha - \tanh\alpha)}\right]}, & \text{if } \kappa = m\,\mathrm{sech}\alpha, \ \alpha > 0, \end{cases}$$

where the constant $C_\circ$ is defined by (2.27).

In Figure 2.3 (b), we plot $E_{m,\kappa}$ against $\mathrm{Im}(\mathcal{T}_{m,\kappa})$, which indicates that the estimate $E_{m,\kappa}$ gives an accurate picture of the behavior of $\mathrm{Im}(\mathcal{T}_{m,\kappa})$.

3. *a priori* ESTIMATES

In order to carry out error analysis for the spectral-Galerkin approximation to (1.7), we need to establish some *a priori* estimates for the solution of (1.7). Without loss of generality, we shall set $\xi = 0$ since the non-homogeneous boundary condition at $r = a$ can be simply converted to a homogeneous one by subtracting a suitable function from the solution.
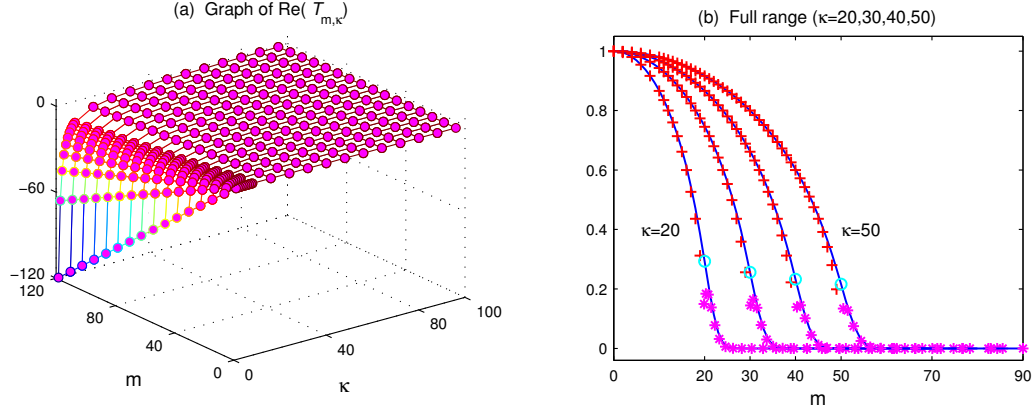
FIGURE 2.3. (a) Graph of 2-D $\mathrm{Re}\big(\mathcal{T}_{m,\kappa}\big)$ with $(\kappa, m) \in [1, 100] \times [1, 120]$; (b) 2-D $\mathrm{Im}(\mathcal{T}_{m,\kappa})$ ("solid line") against $E_{m,\kappa}$ defined in (2.35) ("+" for $\kappa > m + 1/2$; "○" for $\kappa = m + 1/2$ and "⋆" for $\kappa < m + 1/2$) with $\kappa = 20, 30, 40, 50$.

3.1. Dimension reduction.

We now rewrite the equation (1.7) with $\xi = 0$ in polar coordinates (r, θ) or spherical coordinates (r, θ, ϕ):

$$(3.1) \qquad \begin{cases} -\left(\dfrac{1}{r^{d-1}}\partial_r\big(r^{d-1}\partial_r U\big) + \dfrac{1}{r^2}\Delta_S U\right) - k^2 U = F, & \text{in } \Omega = (a, b) \times S, \\[2mm] U\big|_{r=a} = 0, \quad \big[\partial_r U + T(U)\big]\big|_{r=b} = \eta, \end{cases}$$

where

$$(3.2) \qquad \Delta_S U = \begin{cases} \partial_\theta^2 U, & \text{if } d = 2, \\[2mm] \dfrac{1}{\sin^2\phi}\partial_\theta^2 U + \dfrac{1}{\sin\phi}\partial_\phi\big(\sin\phi\,\partial_\phi U\big), & \text{if } d = 3, \end{cases}$$

and its eigenfunctions are the Fourier basis $\{e^{im\theta}\}$ (in 2-D) or the spherical harmonic functions $\{Y_m^l(\theta, \phi)\}$ (in 3-D), i.e.,

$$(3.3) \qquad -\Delta_S e^{im\theta} = m^2 e^{im\theta}\ (d = 2); \quad -\Delta_S Y_m^l(\theta, \phi) = m(m+1)Y_m^l(\theta, \phi)\ (d = 3).$$

We shall denote

$$(3.4) \qquad \beta_m = \begin{cases} m^2, & m = 0, \pm 1, \pm 2, \cdots, & \text{if } d = 2, \\[2mm] m(m+1), & m = 0, 1, 2, \cdots, & \text{if } d = 3. \end{cases}$$

Expanding the solution and given data in terms of the eigenfunctions of Δ_S :

$$(3.5) \qquad \big(U,\ F,\ eta\big) = \begin{cases} \sum_{|m|=0}^{\infty} \big(\hat{u}_m(r),\ \hat{f}_m(r),\ \hat{h}_m\big)e^{im\theta}, & \text{if } d = 2, \\[2mm] \sum_{m=0}^{\infty}\sum_{l=-m}^{m} \big(\hat{u}_{lm}(r),\ \hat{f}_{lm}(r),\ \hat{h}_{lm}\big)Y_m^l(\theta, \phi), & \text{if } d = 3, \end{cases}$$

we find from (3.3) that the problem (3.1)–(3.2) is reduced to the following sequence of one-dimensional equations (for brevity, we use u to denote $\hat{u}_m$ or $\hat{u}_{lm}$, and likewise for f and h

below):

(3.6)
$$\begin{cases} -\dfrac{1}{r^{d-1}}\dfrac{d}{dr}\left[r^{d-1}\dfrac{du}{dr}\right] + \beta_m\dfrac{u}{r^2} - k^2 u = f, \quad r \in (a,b),\ d = 2,3, \\ u(a) = 0, \qquad u'(b) - kT_{m,k}u(b) = h, \end{cases}$$

where $T_{m,k}$ is derived from (2.7) and (2.12):

(3.7)
$$T_{m,k} = \begin{cases} \dfrac{\partial_z H_m^{(1)}(kb)}{H_m^{(1)}(kb)}, & \text{if } d = 2, \\ \dfrac{\partial_z h_m^{(1)}(kb)}{h_m^{(1)}(kb)}, & \text{if } d = 3. \end{cases}$$

Notice that $T_{m,k} = \mathcal{T}_{m,kb}$ (defined by (2.14)).

3.2. **Variational formulation and** *a priori* **estimates.** We denote the weight functions $\omega^\alpha(r) = r^\alpha$ and $\omega(r) = r$. Define the 1-D weighted space

(3.8)
$$X := X(d) = \left\{ u \in H^1_{\omega^{d-1}}(I) \cap L^2_{\omega^{d-3}}(I) : \ u(a) = 0 \right\}.$$

We define a bilinear form on $H^1_p(S;X) \times H^1_p(S;X)$:

(3.9)
$$\begin{aligned} \mathcal{B}(U,V) =& \left(\partial_r U, \partial_r V\right)_{\omega^{d-1}} + \left(\nabla_S U, \nabla_S V\right)_{\omega^{d-3}} - k^2\left(U,V\right)_{\omega^{d-1}} \\ &+ b^{d-1}\left\langle T(U)(\cdot), V(b,\cdot)\right\rangle_S, \end{aligned}$$

where $\langle \cdot, \cdot \rangle_S$ is the $L^2(S)$-inner product (cf. Appendix), and the gradient operator ∇_S is defined by

(3.10)
$$\nabla_S U = \begin{cases} \partial_\theta U, & \text{if } d = 2, \\ \left(\dfrac{1}{\sin\theta}\partial_\theta U\right)\vec{e}_\theta + \left(\partial_\phi U\right)\vec{e}_\phi, & \text{if } d = 3. \end{cases}$$

The variational formulation of (3.1) is: Given $F \in L^2_{\omega^{d-1}}(\Omega)$ and $\eta \in L^2(S)$, find $U \in H^1_p(S;X)$ such that

(3.11)
$$\mathcal{B}(U,V) = (F,V)_{\omega^{d-1}} + b^{d-1}\left\langle \eta, V(b,\cdot)\right\rangle_S, \quad \forall V \in H^1_p(S;X),\ d = 2,3,$$

which admits a unique solution (see, e.g., [18]). The first main result of this paper is the following *a priori* estimate.

Theorem 3.1. *Let U be the solution of (3.11). If $F \in L^2(\Omega)$ and $H \in L^2(S)$, then we have*

(3.12)
$$\|\nabla U\| + k\|U\|_\Omega \lesssim \left(\sqrt{b^d} + \sqrt{b|I|}(kb)^{1/3}\right)\|\eta\|_{L^2(S)} + (kb)^{1/3}|I|\|F\|,$$

where $|I| = b - a$.

The rest of this section is devoted to the proof of this estimate. Observe that for each mode m or (l,m), the expansion coefficient $u = \hat{u}_m$ or $\hat{u}_{lm}$ (cf. (3.5)) satisfies the following reduced problem (i.e., the variational formulation of (3.6)–(3.7)):

(3.13)
$$\begin{aligned} &\text{Given } f \in L^2_{\omega^{d-1}}(I) \text{ and } h \in , \text{ find } u \in X \text{ such that} \\ &\mathcal{B}_m(u,v) = (f,v)_{\omega^{d-1}} + b^{d-1}h\overline{v(b)}, \quad \forall v \in X,\ d = 2,3, \end{aligned}$$

where $f = \hat{f}_m$ or $\hat{f}_{lm}$, $h = \hat{h}_m$ or $\hat{h}_{lm}$, and the sesquilinear form

(3.14)
$$\mathcal{B}_m(u,v) := (\partial_r u, \partial_r v)_{\omega^{d-1}} + \beta_m(u,v)_{\omega^{d-3}} - k^2(u,v)_{\omega^{d-1}} - kb^{d-1}T_{m,k}u(b)\overline{v(b)},$$

where β_m is defined in (3.4).

An essential step is to derive *a priori* estimates for each $u = \hat{u}_m$ or $\hat{u}_{lm}$, and then combine these estimates to get the desired result for the original problem (3.11).

We have the following *a priori* estimate for the solution of (3.13)–(3.14).

Lemma 3.1. *Let* $|I| = b - a$ *be the length of the interval* $I = (a, b)$. *If* $f \in L^2_{\omega^{d-1}}(I)$, *then given* $k_0 > 0$, *we have that for* $k \geq k_0$ *and* $d = 2, 3$,

$$
(3.15) \qquad
\begin{aligned}
\|\partial_r u\|_{\omega^{d-1}} &+ \sqrt{\beta_m}\|u\|_{\omega^{d-3}} + k\|u\|_{\omega^{d-1}} \\
&\lesssim \left(\sqrt{b^d} + \sqrt{b|I|}C_{m,k}\right)|h| + C_{m,k}|I|\|f\|_{\omega^{d-1}},
\end{aligned}
$$

where

$$
(3.16) \qquad
C_{m,k} = \begin{cases} (kb)^{\frac{1}{3}}, & \text{if} \quad |m| \leq kb, \\ 1, & \text{if} \quad |m| > kb. \end{cases}
$$

Proof. Some early work (cf. [9, 14, 15]) in this direction relies on the explicit form of Green's function which is very difficult, if not possible, to extend to more general cases. Our proof is based on an argument in [7] (see also [25, 6]). More precisely, we take two test functions $v = u$, $(r - a)\partial_r u \in X$ in (3.13) successively to obtain *a priori* estimates without using Green's functions. In the following, $\varepsilon_j > 0$ $(j = 1, \cdots, 5)$ are some suitable real numbers.

We first take $v = u$ in (3.13). The imaginary and real parts are

$$
(3.17a) \qquad -kb^{d-1}\mathrm{Im}(T_{m,k})|u(b)|^2 = b^{d-1}\mathrm{Im}(h\overline{u(b)}) + \mathrm{Im}(f, u)_{\omega^{d-1}},
$$

$$
(3.17b) \qquad
\begin{aligned}
\|\partial_r u\|^2_{\omega^{d-1}} &+ \beta_m\|u\|^2_{\omega^{d-3}} - k^2\|u\|^2_{\omega^{d-1}} - kb^{d-1}\mathrm{Re}(T_{m,k})|u(b)|^2 \\
&= b^{d-1}\mathrm{Re}(h\overline{u(b)}) + \mathrm{Re}(f, u)_{\omega^{d-1}}.
\end{aligned}
$$

In order to derive a upper bound for $\|\partial_r u\|^2_{\omega^{d-1}} + \beta_m\|u\|^2_{\omega^{d-3}}$, we proceed separately with two cases: (i) $d = 2$, $|m| > 0$ or $d = 3$, $m \geq 0$; and (ii) $d = 2$, $m = 0$. In the first case, we have from (2.22) and (2.34b) with $|m| \geq 1$ (note that $\kappa = kb$) that

$$
(3.18) \qquad \frac{1}{k|\mathrm{Re}(T_{m,k})|} \leq b, \quad \text{for } d = 2, \, |m| > 0 \text{ or } d = 3, \, m \geq 0.
$$

In what follows, we shall repeatedly use the inequality $2AB \leq \varepsilon A^2 + \frac{B^2}{\varepsilon}$ for all $A, B, \varepsilon > 0$. Applying the Cauchy-Schwarz inequality to (3.17b) leads to

$$
(3.19) \qquad
\begin{aligned}
\|\partial_r u\|^2_{\omega^{d-1}} &+ \beta_m\|u\|^2_{\omega^{d-3}} - kb^{d-1}\mathrm{Re}(T_{m,k})|u(b)|^2 \\
&\leq k^2\|u\|^2_{\omega^{d-1}} + \frac{kb^{d-1}|\mathrm{Re}(T_{m,k})|}{2}|u(b)|^2 + \frac{b^{d-1}}{2k|\mathrm{Re}(T_{m,k})|}|h|^2 \\
&\quad + \varepsilon_1 k^2\|u\|^2_{\omega^{d-1}} + \frac{1}{4\varepsilon_1 k^2}\|f\|^2_{\omega^{d-1}}.
\end{aligned}
$$

Thus, by (3.18), the estimate (3.19) becomes (for $d = 2$, $|m| > 0$ or $d = 3$, $m \geq 0$):

$$
(3.20) \qquad
\begin{aligned}
\|\partial_r u\|^2_{\omega^{d-1}} &+ \beta_m\|u\|^2_{\omega^{d-3}} - \frac{kb^{d-1}\mathrm{Re}(T_{m,k})}{2}|u(b)|^2 \\
&\leq (1 + \varepsilon_1)k^2\|u\|^2_{\omega^{d-1}} + \frac{b^d}{2}|h|^2 + \frac{1}{4\varepsilon_1 k^2}\|f\|^2_{\omega^{d-1}}.
\end{aligned}
$$

To treat the only remaining case: (ii) $d = 2$ and $m = 0$, we apply the Cauchy-Schwarz inequality to (3.17a), and get that

(3.21)
$$kb\text{Im}(T_{0,k})|u(b)|^2 \le \frac{kb\text{Im}(T_{0,k})}{2}|u(b)|^2 + \frac{b}{2k\text{Im}(T_{0,k})}|h|^2$$
$$+ \frac{\varepsilon_2 k\text{Im}(T_{0,k})}{2}\|u\|_\omega^2 + \frac{1}{2\varepsilon_2 k\text{Im}(T_{0,k})}\|f\|_\omega^2,$$

which implies that

(3.22)
$$k^2 b|u(b)|^2 \le \varepsilon_2 k^2\|u\|_\omega^2 + \frac{b}{|\text{Im}(T_{0,k})|^2}|h|^2 + \frac{1}{\varepsilon_2|\text{Im}(T_{0,k})|^2}\|f\|_\omega^2.$$

Thanks to (2.34c), we can rewrite the inequality (3.22) as

(3.23)
$$k^2 b|u(b)|^2 \le \varepsilon_2 k^2\|u\|_\omega^2 + b|h|^2 + \frac{1}{\varepsilon_2}\|f\|_\omega^2.$$

We now apply Cauchy-Schwarz inequality to (3.17b) (with $d = 2$ and $m = 0$), and use (3.23) to bound the term involving $|u(b)|^2$, to get

(3.24)
$$\|\partial_r u\|_\omega^2 + \beta_0\|u\|_{\omega^{-1}}^2 - kb\text{Re}(T_{0,k})|u(b)|^2$$
$$\le k^2\|u\|_\omega^2 + k^2 b|u(b)|^2 + \frac{b}{4k^2}|h|^2 + \frac{\varepsilon_1 k^2}{2}\|u\|_\omega^2 + \frac{1}{2\varepsilon_1 k^2}\|f\|_\omega^2$$
$$\le (1 + \varepsilon_1)k^2\|u\|_\omega^2 + cb|h|^2 + \frac{1}{\varepsilon_1}\|f\|_\omega^2,$$

where we took $\varepsilon_2 = \varepsilon_1/2$ in (3.23). In view of (3.20) and (3.24), we have the following estimate which is valid for all cases:

(3.25)
$$\|\partial_r u\|_{\omega^{d-1}}^2 + \beta_m\|u\|_{\omega^{d-3}}^2 \le (1 + \varepsilon_1)k^2\|u\|_{\omega^{d-1}}^2 + cb^d|h|^2 + c\|f\|_{\omega^{d-1}}^2.$$

Now, the main difficulty is how to bound the term $k^2\|u\|_{\omega^{d-1}}^2$. To do this, we need to derive further estimates by testing (3.13) with another function. Using a standard regularity argument, one can easily verify that for $f \in L_{\omega^{d-1}}^2(I)$, the weak solution of (3.13) satisfies $(r - a)\partial_r u \in X$. Hence, we can take $v = 2(r - a)\partial_r u$ in (3.13), and after integration by parts and thanks to the identity

(3.26)
$$(u, v)_\omega + (v, u)_\omega = 2\text{Re}(u, v)_\omega,$$

we find that the first three terms of the real part of (3.13) with $v = 2(r - a)\partial_r u$ are

(3.27a)
$$2\text{Re}\big(\partial_r u, \partial_r((r - a)\partial_r u)\big)_{\omega^{d-1}} = b^{d-1}|I||\partial_r u(b)|^2$$
$$+ \int_a^b \left[(2 - d) + (d - 1)\frac{a}{r}\right]|\partial_r u|^2 r^{d-1}dr;$$

(3.27b)
$$2\beta_m\text{Re}\big(u, (r - a)\partial_r u\big)_{\omega^{d-3}} = \beta_m b^{d-3}|I||u(b)|^2$$
$$- \beta_m\int_a^b \left[(d - 2) - (d - 3)\frac{a}{r}\right]|u|^2 r^{d-3}dr;$$

(3.27c)
$$-2k^2\text{Re}\big(u, (r - a)\partial_r u\big)_{\omega^{d-1}} = -k^2 b^{d-1}|I||u(b)|^2$$
$$+ k^2\int_a^b \left[d - (d - 1)\frac{a}{r}\right]|u|^2 r^{d-1}dr.$$

Accordingly, we find that the real part of (3.13) with $v = 2(r-a)\partial_r u$ becomes

$$b^{d-1}|I|\left(|\partial_r u(b)|^2 + \beta_m b^{-2}|u(b)|^2\right) + a(d-1)\|\partial_r u\|_{\omega^{d-2}}^2$$

$$+ k^2 \int_a^b \left[d - (d-1)\frac{a}{r}\right]|u|^2 r^{d-1} dr$$

(3.28)

$$\leq k^2 b^{d-1}|I||u(b)|^2 + (d-2)\|\partial_r u\|_{\omega^{d-1}}^2 + \beta_m \int_a^b \left[(d-2) + (3-d)\frac{a}{r}\right]|u|^2 r^{d-3} dr$$

$$+ 2b^{d-1}|I||\mathrm{Re}(h\overline{\partial_r u(b)})| + 2|\mathrm{Re}(f,(r-a)\partial_r u)_{\omega^{d-1}}|.$$

Note that in the third term, the factor $d - (d-1)\frac{a}{r} > 1$ for all $r \in (a,b)$, so we can use this term to bound $k^2\|u\|_{\omega^{d-1}}^2$ in (3.25).

By the Cauchy-Schwarz inequality, we can treat the last two terms at the right-hand side of (3.28) as

(3.29) $$2b^{d-1}|I||\mathrm{Re}(h\overline{\partial_r u(b)})| \leq \frac{b^{d-1}|I|}{2}|\partial_r u(b)|^2 + 2b^{d-1}|I||h|^2,$$

and

(3.30) $$2|\mathrm{Re}(f,(r-a)\partial_r u)_{\omega^{d-1}}| \leq \varepsilon_3\|\partial_r u\|_{\omega^{d-1}}^2 + \frac{|I|^2}{\varepsilon_3}\|f\|_{\omega^{d-1}}^2.$$

We now proceed separately for $d = 2$ and $d = 3$.

Case I: $d = 2$. In this case, a combination of (3.28)–(3.30) leads to

$$b|I|\left(|\partial_r u(b)|^2 + \beta_m b^{-2}|u(b)|^2\right) + a\|\partial_r u\|^2 + k^2 \int_a^b \left[2 - \frac{a}{r}\right]|u|^2 r dr$$

(3.31)

$$\leq k^2 b|I||u(b)|^2 + \left(\varepsilon_3\|\partial_r u\|_\omega^2 + a\beta_m\|u\|_{\omega^{-2}}^2\right) + \frac{b|I|}{2}|\partial_r u(b)|^2$$

$$+ 2b|I||h|^2 + \frac{|I|^2}{\varepsilon_3}\|f\|_\omega^2.$$

Use (3.25) with $d = 2$, we have that for $\varepsilon_3 < 1$ and for certain $\xi_1 \in (a,b)$,

(3.32)
$$\varepsilon_3\|\partial_r u\|_\omega^2 + a\beta_m\|u\|_{\omega^{-2}}^2 \leq \max\left\{\varepsilon_3, \frac{a}{\xi_1}\right\}\left(\|\partial_r u\|_\omega^2 + \beta_m\|u\|_{\omega^{-1}}^2\right)$$

$$\leq (1+\varepsilon_1)k^2\|u\|_\omega^2 + cb^2|h|^2 + c\|f\|_\omega^2.$$

Hence, it remains to bound the term $k^2 b|I||u(b)|^2$ in (3.31).

(i) $|m| > kb$. In this case, the term $b|I|k^2|u(b)|^2$ can be absorbed by $b^{-1}|I|\beta_m|u(b)|^2$ at the left-hand side of (3.31). Hence, a combination of (3.31)–(3.32) leads to the desired result:

(3.33)
$$b|I|\left(\frac{1}{2}|\partial_r u(b)|^2 + (\beta_m b^{-2} - k^2)|u(b)|^2\right) + a\|\partial_r u\|^2 + Ck^2\|u\|_\omega^2$$

$$\lesssim b^2|h|^2 + (1 + |I|^2)\|f\|_\omega^2,$$

where, with a suitable choice of ε_1, the constant

(3.34) $$C = 1 - \frac{a}{\xi_2} - \varepsilon_1 > 0, \quad \text{for certain } \xi_2 \in (a,b).$$

(ii) $|m| \leq kb$. Similar to the derivation of (3.22), we apply the Cauchy-Schwarz inequality to (3.17a),

(3.35) $$k^2 b|I||u(b)|^2 \leq \varepsilon_3|I|k^2\|u\|_\omega^2 + \frac{b|I|}{|\mathrm{Im}(T_{m,k})|^2}|h|^2 + \frac{|I|}{\varepsilon_3|\mathrm{Im}(T_{m,k})|^2}\|f\|_\omega^2.$$

Then a combination of the estimates (3.31), (3.32) and (3.35) leads to

(3.36)
$$b|I|\Big(\frac{1}{2}|\partial_r u(b)|^2 + \beta_m b^{-2}|u(b)|^2\Big) + a\|\partial_r u\|^2 + \widetilde{C}k^2\|u\|_\omega^2$$
$$\lesssim C_{m,k}^{(1)}|h|^2 + C_{m,k}^{(2)}\|f\|_\omega^2,$$

where, with a suitable choice of ε_1 & ε_3 and using the fact that $\mathrm{Im}(T_{m,k}) < 1$, the constants are

$$\widetilde{C} = 1 - \frac{a}{\xi_3} - \varepsilon_1 - \varepsilon_3|I| > 0, \quad \text{for certain } \xi_3 \in (a,b),$$

(3.37)
$$C_{m,k}^{(1)} = b^2 + \frac{b|I|}{|\mathrm{Im}(T_{m,k})|^2},$$

$$C_{m,k}^{(2)} = |I|^2 + \frac{|I|}{\varepsilon_3|\mathrm{Im}(T_{m,k})|^2} \lesssim |I|^2\Big(1 + \frac{1}{|\mathrm{Im}(T_{m,k})|^2}\Big).$$

Notice that we have

(3.38)
$$\mathrm{Im}(T_{m,k}) \geq c(kb)^{-\frac{1}{3}}, \quad \text{for } |m| \leq kb,$$

since $\mathrm{Im}(T_{m,k})$ is a decreasing function of m and the estimate (2.35).

Therefore, the desired result (3.15) with $d = 2$ follows from (3.25), (3.33) and (3.36).

Case II: $d = 3$. In this case, a combination of (3.28)–(3.30) leads to

(3.39)
$$b^2|I|\big(|\partial_r u(b)|^2 + \beta_m b^{-2}|u(b)|^2\big) + 2a\|\partial_r u\|_\omega^2 + k^2\int_a^b\Big[3 - \frac{2a}{r}\Big]|u|^2 r\, dr$$
$$\leq k^2 b^2|I||u(b)|^2 + \frac{b^2|I|}{2}|\partial_r u(b)|^2 + \Big(\|\partial_r u\|_{\omega^2}^2 + \beta_m\|u\|^2\Big)$$
$$+ 2b^2|I||h|^2 + \frac{|I|^2}{\varepsilon_3}\|f\|_{\omega^2}^2.$$

By (3.25),

(3.40)
$$\|\partial_r u\|_{\omega^2}^2 + \beta_m\|u\|^2 \leq (1+\varepsilon_1)k^2\|u\|_{\omega^2}^2 + cb^3|h|^2 + c\|f\|_{\omega^2}^2.$$

The rest of the proof is essentially the same as that in the case 2-D case. More precisely, we can derive the 3-D version of inequalities (3.33)—(3.38) with slightly different constants

(3.41)
$$C = 2 - \frac{2a}{\xi_3} - \varepsilon_1 > 0, \quad \xi_3 \in (a,b), \quad \text{if } m \geq kb,$$
$$\tilde{C} = 2 - \frac{2a}{\xi_3} - \varepsilon_1 - 2\varepsilon_3|I|, \quad \xi_3 \in (a,b), \quad \text{if } m < kb.$$

Finally, since $\mathrm{Im}(T_{m,k})$ is a decreasing function of m and $\mathrm{Im}(T_{m,k}) = \mathrm{Im}(\mathcal{T}_{m,kb})$ (cf. (2.14) and (3.7)), the desired bound follows from (2.27) and (2.35). $\square$

The proof of Theorem 3.1. Since the proof of the 2-D and 3-D cases is essentially the same, we only prove (3.12) with $d = 3$. Thanks to the orthogonality of the spherical

harmonic functions, we deduce from Lemma 3.1 that

$$\|\nabla U\|^2 + k^2\|U\|^2 \lesssim \|\partial_r U\|^2_{\omega^2} + \|\nabla_S U\|^2 + k^2\|U\|^2_{\omega^2}$$

$$\lesssim \sum_{m=0}^{\infty}\sum_{l=-m}^{m}\left(\|\partial_r\hat{u}_{lm}\|^2_{\omega^2} + \beta_m\|\hat{u}_{lm}\|^2 + k^2\|\hat{u}_{lm}\|^2_{\omega^2}\right)$$

$$\lesssim \sum_{m=0}^{\infty}\sum_{l=-m}^{m}\left(\left(\sqrt{b^3}+\sqrt{b|I|}C_{m,k}\right)^2|\hat{h}_{lm}|^2 + C^2_{m,k}|I|^2\|\hat{f}_{lm}\|^2_{\omega^2}\right)$$

$$\lesssim \sum_{m=0}^{\infty}\sum_{l=-m}^{m}\left(\left(\sqrt{b^3}+\sqrt{b|I|}(kb)^{1/3}\right)^2\hat{h}^2_{lm} + (kb)^{2/3}|I|^2\|\hat{f}_{lm}\|^2_{\omega^2}\right)$$

$$\lesssim \left(\sqrt{b^3}+\sqrt{b|I|}(kb)^{1/3}\right)^2\|\eta\|^2 + (kb)^{2/3}|I|^2\|F\|^2.$$

This ends the proof.

4. Spectral-Galerkin approximation

4.1. **The spectral-Galerkin method and its well-posedness.** Let P_N be the space of all complex polynomials of degree at most N on $\bar{I}$. Define $X_N := \{u \in P_N : u(a) = 0\}$, and

$$(4.1) \qquad Y_M := \begin{cases} \text{span}\{e^{\mathrm{i}m\theta} : -M \le m \le M\}, & \text{if } d = 2, \\ \text{span}\{Y_m^l(\theta,\phi) : 0 \le |l| \le m \le M\}, & \text{if } d = 3, \end{cases}$$

where $\mathcal{B}(\cdot,\cdot)$ is defined in (3.9).

The spectral-Galerkin approximation to (3.11) is:

$$(4.2) \qquad \begin{aligned} &\text{Find } U_{MN} \in \mathcal{V}_{MN} := X_N \times Y_M \text{ such that} \\ &\mathcal{B}(U_{MN}, V_{MN}) = (F, V_{MN})_{\omega^{d-1}} + b^{d-1}\langle\eta, V_{MN}(b,\cdot)\rangle_S, \quad \forall V_{MN} \in \mathcal{V}_{MN}. \end{aligned}$$

Since the sesquilinear form $\mathcal{B}(\cdot,\cdot)$ is not coercive in $\mathcal{V}_{MN}\times\mathcal{V}_{MN}$ even for small wave number k, an important issue is to prove the well-posedness of the discrete scheme (4.4).

Expanding the numerical solution and test function as

$$(4.3) \qquad \left(U_{MN},\ V_{MN}\right) = \begin{cases} \sum_{|m|=0}^{M}\left(\hat{u}_m^N(r),\ \hat{v}_m^N(r)\right)e^{\mathrm{i}m\theta}, & \text{if } d = 2, \\ \sum_{m=0}^{M}\sum_{l=-m}^{m}\left(\hat{u}_{lm}^N(r),\ \hat{v}_{lm}^N(r)\right)Y_m^l(\theta,\phi), & \text{if } d = 3, \end{cases}$$

one verifies that $u_N := \hat{u}_m^N$ or $\hat{u}_{lm}^N$ satisfies the reduced problem:

$$(4.4) \qquad \begin{cases} \text{Find } u_N \in X_N \text{ such that} \\ \mathcal{B}_m(u_N, v_N) = (f, v_N)_{\omega^{d-1}} + b^{d-1}h\overline{v_N(b)}, \quad \forall v_N \in X_N,\ d = 2,3, \end{cases}$$

where $\mathcal{B}_m(\cdot,\cdot)$ is defined in (3.14), and for brevity, we denote $v_N := \hat{v}_m^N$ or v_{lm}^N, and f and h are the same as those in (3.13).

It is important to note that, unlike in the Galerkin finite-element method, the spectral-Galerkin approximation space X_N has the following property: *For $u_N \in X_N$, we have $(r-a)\partial_r u_N \in X_N$. Hence, the proof of Lemma 3.1 is also valid for the discrete system* (4.4). *In particular, Theorem 3.1 holds with u_N in the place of u.* As a consequence, the problem (4.4) has at most one solution. Since (4.4) is finite dimensional, we then derive from s simple fact in linear algebra that the problem (4.4) admits a unique solution.

Therefore, following the same procedure as in the proof of Theorem 3.1 leads to the following result:

Theorem 4.2. *If $F \in L^2(\Omega)$ and $H \in L^2(S)$. The problem (4.2) admits a unique solution satisfying*

$$(4.5) \qquad \|\nabla U_{MN}\| + k\|U_{MN}\| \lesssim \left(\sqrt{b^d} + \sqrt{b|I|}(kb)^{1/3}\right)\|\eta\| + (kb)^{1/3}|I|\|F\|.$$

4.2. **Error estimates.** In this part, we shall estimate the error between U (solution of (3.11)) and U_{MN} (solution of (4.2)). Our starting point is to analyze the error of one-dimensional approximation (4.4).

4.2.1. *Analysis of 1-D scheme.* In order to carry out the error analysis, we define the orthogonal projection $_0\pi_N^1 : X \to X_N$ by

$$(4.6) \qquad \left(\partial_r(u - {}_0\pi_N^1 u),\ \partial_r v_N\right) = 0, \quad \forall v_N \in X_N.$$

For $s \geq 1$ and $s \in \mathbb{N}$, we introduce the weighted Sobolev space

$$B^s(I) := \left\{u \in L^2(I)\ :\ [(r-a)(b-r)]^{\frac{l-1}{2}}\partial_r^l u \in L^2(I),\ 1 \leq l \leq s\right\},$$

with the norm and semi-norm

$$\|u\|_{B^s} = \left(\|u\|^2 + \sum_{l=1}^{s} \left\|[(r-a)(b-r)]^{\frac{l-1}{2}}\partial_r^l u\right\|^2\right)^{\frac{1}{2}},$$

$$|u|_{B^s} = \left\|[(r-a)(b-r)]^{\frac{s-1}{2}}\partial_r^s u\right\|.$$

Lemma 4.2. *For any $u \in X \cap B^s(I)$ with $s \geq 1$ and $s \in \mathbb{N}$,*

$$(4.7) \qquad \|\partial_r({}_0\pi_N^1 u - u)\| + N|I|^{-1}\|{}_0\pi_N^1 u - u\| \lesssim N^{1-s}|u|_{B^s}.$$

Proof. This result is a direct consequence of the Legendre polynomial approximation (with a scaling and a direct extension to complex functions), which can be found, for instance, in [5], with an improvement of the weighted semi-norm in the upper bound given by [11]. □

With the aid of Lemmas 3.1 and 4.2, we are able to obtain the following error estimates.

Theorem 4.3. *Let u and u_N be respectively the solutions of (3.13) and (4.4). If $u \in X \cap B^s(I)$ with integer $s \geq 1$, then for $d = 2, 3$,*

$$(4.8) \qquad \begin{aligned} \|\partial_r(u - u_N)\|_{\omega^{d-1}} &+ \sqrt{\beta_m}\|u - u_N\|_{\omega^{d-3}} + k\|u - u_N\|_{\omega^{d-1}} \\ &\lesssim C_\star(m, N, k; a, b, d)N^{1-s}|u|_{B^s}. \end{aligned}$$

where

$$(4.9) \qquad \begin{aligned} C_\star(m, N, k; a, b, d) := &(1 + \sqrt{\beta_m})b^{(d-1)/2} + \sqrt{\beta_m}a^{\frac{d-3}{2}}|I|N^{-1} \\ &+ k^{1/3}(\sqrt{\beta_m}b^{3d/2-2}\sqrt{|I|}N^{-1/2} + |I|^2 b^{d/2}k^2 N^{-1}). \end{aligned}$$

Proof. Let $e_N = u_N - {}_0\pi_N^1 u$ and $\tilde{e}_N = u - {}_0\pi_N^1 u$. By (3.13) and (4.4),

$$(4.10) \qquad \mathcal{B}_m(u - u_N, v_N) = 0, \quad \forall v_N \in X_N.$$

Then, we derive from (3.14), (4.6) and (4.10) that for any $v_N \in X_N$,

$$(4.11) \qquad \begin{aligned} \mathcal{B}_m(e_N, v_N) =&\mathcal{B}_m(\tilde{e}_N, v_N) = \beta_m(\tilde{e}_N, v_N)_{\omega^{d-3}} \\ &- k^2(\tilde{e}_N, v_N)_{\omega^{d-1}} - kb^{d-1}T_{m,k}\tilde{e}_N(b)\overline{v_N(b)}. \end{aligned}$$

Hence, we can view (4.11) in the form of (3.13) with $u = e_N$, $h = -kb^{d-1}T_{m,k}\tilde{e}_N(b)$, $f = -k^2\tilde{e}_N$, and with an additional term $\beta_m(\tilde{e}_N, v_N)_{\omega^{d-3}}$. As with the proof of Theorem 3.1,

we take two different test functions $v_N = e_N, 2(r-a)\partial_r e_N \in X_N$ and treat the extra term as

$$(4.12) \qquad \beta_m |(\tilde{e}_N, e_N)_{\omega^{d-3}}| \le \varepsilon_6 \beta_m \|e_N\|^2_{\omega^{d-3}} + \frac{\beta_m}{4\varepsilon_6} \|\tilde{e}_N\|^2_{\omega^{d-3}},$$

and

$$
\begin{aligned}
(4.13) \quad & 2\beta_m |(\tilde{e}_N, (r-a)\partial_r e_N)_{\omega^{d-3}}| \le 2\beta_m \Big\{ b^{d-3}|I| |\tilde{e}_N(b)\overline{e_N(b)}| \\
& \quad + \big|(\partial_r \tilde{e}_N, (1-ar^{-1})e_N)_{\omega^{d-2}}\big| + \big|(\tilde{e}_N, ((d-2)-a(d-3)r^{-1})e_N)_{\omega^{d-3}}\big| \Big\} \\
& \le \varepsilon_7 \beta_m b^{d-3}|I| |e_N(b)|^2 + \frac{\beta_m b^{d-3}|I|}{\varepsilon_7} |\tilde{e}_N(b)|^2 + \varepsilon_8 \beta_m \|e_N\|^2_{\omega^{d-3}} \\
& \quad + \frac{c\beta_m}{\varepsilon_8} \Big(\|\partial_r \tilde{e}_N\|^2_{\omega^{d-1}} + \|\tilde{e}_N\|^2_{\omega^{d-3}} \Big).
\end{aligned}
$$

Thus, choosing suitable constants $\{\varepsilon_j\}_{j=6}^8$, and following the same lines as for the proof of Theorem 3.1 (with $u = e_N$, $h = -kb^{d-1}T_{m,k}\tilde{e}_N(b)$ and $f = -k^2\tilde{e}_N$), we can derive that

$$
\begin{aligned}
(4.14) \quad & \|\partial_r e_N\|^2_{\omega^{d-1}} + \beta_m \|e_N\|^2_{\omega^{d-3}} + k^2 \|e_N\|^2_{\omega^{d-1}} \\
& \lesssim \beta_m \big(\|\partial_r \tilde{e}_N\|^2_{\omega^{d-1}} + \|\tilde{e}_N\|^2_{\omega^{d-3}} \big) + \beta_m b^{d-3}|I| |\tilde{e}_N(b)|^2 \\
& \quad + k^2 b^{2(d-1)}|T_{m,k}|^2 \big(\sqrt{b^d} + \sqrt{b|I|}C_{m,k} \big)^2 |\tilde{e}_N(b)|^2 \\
& \quad + k^4 |I|^2 C^2_{m,k} \|\tilde{e}_N\|^2_{\omega^{d-1}}.
\end{aligned}
$$

To estimate the term $|\tilde{e}_N(b)|$, we use the Sobolev inequality and Lemma 4.2 to obtain that

$$(4.15) \qquad |\tilde{e}_N(b)|^2 \lesssim \big(2 + |I|^{-1}\big) \|\tilde{e}_N\| \|\tilde{e}_N\|_1 \lesssim N^{1-2s}|I| |u|^2_{B^s}.$$

Next, using the inequality $\|v\|^2_{\omega^\alpha} \le \max\{b^\alpha, a^\alpha\}\|v\|^2$, and Lemma 4.2 leads to

$$
\begin{aligned}
(4.16) \quad & \|\partial_r^\mu \tilde{e}_N\|^2_{\omega^{d-1}} \le b^{d-1} \|\partial_r^\mu \tilde{e}_N\|^2 \lesssim b^{d-1}|I|^{2-2\mu}N^{2\mu-2s}|u|^2_{B^s}, \quad \mu = 0, 1, \\
& \|\tilde{e}_N\|^2_{\omega^{d-3}} \le a^{d-3}\|\tilde{e}_N\|^2 \lesssim a^{d-3}|I|^2 N^{-2s}|u|^2_{B^s}.
\end{aligned}
$$

Hence, by the triangle inequality, (4.14)–(4.16) and Lemma 4.2, we have that

$$
\begin{aligned}
(4.17) \quad & \|\partial_r(u-u_N)\|^2_{\omega^{d-1}} + \beta_m \|u-u_N\|^2_{\omega^{d-3}} + k^2 \|u-u_N\|^2_{\omega^{d-1}} \\
& \le \Big(\|\partial_r e_N\|^2_{\omega^{d-1}} + \beta_m \|e_N\|^2_{\omega^{d-3}} + k^2 \|e_N\|^2_{\omega^{d-1}} \Big) \\
& \quad + \Big(\|\partial_r \tilde{e}_N\|^2_{\omega^{d-1}} + \beta_m \|\tilde{e}_N\|^2_{\omega^{d-3}} + k^2 \|\tilde{e}_N\|^2_{\omega^{d-1}} \Big) \\
& \lesssim (1+\beta_m)\|\partial_r \tilde{e}_N\|^2_{\omega^{d-1}} + \beta_m \|\tilde{e}_N\|^2_{\omega^{d-3}} + \beta_m b^{d-3}|I| |\tilde{e}_N(b)|^2 \\
& \quad + k^2 b^{2(d-1)}|T_{m,k}|^2 \big(\sqrt{b^d} + \sqrt{b|I|}C_{m,k} \big)^2 |\tilde{e}_N(b)|^2 \\
& \quad + k^4 |I|^2 C^2_{m,k} \|\tilde{e}_N\|^2_{\omega^{d-1}} \\
& \lesssim C^*(m,N,k;a,b,d)N^{2-2s}|u|^2_{B^s},
\end{aligned}
$$

where

$$
\begin{aligned}
(4.18) \quad C^*(m,N,k;a,b,d) := & (1+\beta_m)b^{d-1} + \beta_m a^{d-3}|I|^2 N^{-2} \\
& + \beta_m b^{d-3}|I|^2 N^{-1} + k^2 b^{2(d-1)}|T_{m,k}|^2 \big(\sqrt{b^d} + \sqrt{b|I|}C_{m,k} \big)^2 |I|N^{-1} \\
& + k^4 |I|^4 b^{d-1} C^2_{m,k} N^{-2}.
\end{aligned}
$$

We now derive an upper bound for $C^*(m, N, k; a, b, d)$. Since by (2.22) and (2.34),

$$|T_{m,k}|^2 \leq 1 + \frac{(m+1)^2}{(kb)^2},$$

and by (3.16), $C_{m,k} \lesssim (kb)^{1/3}$, we deduce that for $N \gg 1$,

$$C^*(m, N, k; a, b, d) \lesssim (1 + \beta_m) b^{d-1} + \beta_m a^{d-3} |I|^2 N^{-2}$$
$$+ \beta_m b^{3d-4} |I| k^{2/3} N^{-1} + |I|^4 b^d k^{4+2/3} N^{-2}.$$

This implies the desired result.   □

Remark 4.1. Note that in the error estimate (4.8), $N^{1-s} |u|_{B^s}$ is the best approximation error, $k^2 N^{-1}$ in C_* is the so called "pollution error" which is typical for the numerical approximations to the Helmholtz equation (cf. [2]). The extra term $k^{1/3}$ in C_* is due to the asymptotic behavior of the DtN kernel (see Section 2) and is unlikely that this extra term can be removed.

Remark 4.2. To illustrate how the error behaves with respect to N, k and b with $a > 0$ being fixed, we consider a typical oscillatory function $u(r) = e^{ikr} - e^{ika}$. Then, for any $s > 0$, we have

$$\text{(4.19)} \quad \begin{aligned} |u|_{B^s}^2 &= \int_a^b |\partial_r^s u|^2 \big((r-a)(b-r)\big)^{s-1} dr \\ &\leq k^{2s} \int_a^b \big((r-a)(b-r)\big)^{s-1} dr \lesssim k \Big(k \frac{b-a}{2}\Big)^{2s-1}. \end{aligned}$$

Plug this into (4.8), we find that for this particular, but typical solution, we have that for any $s \geq 1$,

$$\text{(4.20)} \quad \begin{aligned} \|\partial_r(u - u_N)\|_{\omega^{d-1}} &+ \sqrt{\beta_m} \|u - u_N\|_{\omega^{d-3}} + k \|u - u_N\|_{\omega^{d-1}} \\ &\lesssim C_\star(m, N, k; a, b, d) k \sqrt{\frac{b-a}{2}} \Big(\frac{k(b-a)}{2N}\Big)^{1-s}. \end{aligned}$$

Hence, the error will decay exponentially as soon as $\frac{k(b-a)}{2N} < 1$, as opposed to the usual condition $\frac{kb}{2N} < 1$. Hence, we can significantly reduce the computational cost by choosing b as close to a as we wish (note however that for scattering from a general obstacle $D = r > a + g(\theta)$ in 2-D or $D = r > a + g(\theta, \phi)$ in 3-D, we have to make sure that $b > a + \|g\|_{L^\infty}$).

With the above preparations, we are ready to perform the error analysis of the full scheme (4.2).

4.2.2. *Multidimensional cases.* To describe the error, we introduce the following non-isotropic Sobolev space

$$\text{(4.21)} \quad \mathcal{H}_{p,\omega^{d-1}}^{s,s'}(\Omega) = L_p^2\big(S; B^s(I)\big) \cap H_p^{s'-1}\big(S; H_{\omega^{d-1}}^1(I)\big) \cap H_p^{s'}\big(S; L_{\omega^{d-3}}^2(I) \cap L_{\omega^{d-1}}^2(I)\big),$$

with $d = 2, 3$; $s, s' \geq 1$ and the norm

$$\|U\|_{\mathcal{H}^{s,s'}_{p,\omega}(\Omega)} = \Big(\sum_{|m|=0}^{\infty} \Big[|\hat{u}_m|^2_{B^s} + (1+m^2)^{s'-1} \|\partial_r \hat{u}_m\|^2_\omega $$

$$+ (1+m^2)^{s'} \big(\|\hat{u}_m\|^2_{\omega^{-1}} + \|\hat{u}_m\|^2_\omega \big) \Big] \Big)^{\frac{1}{2}};$$

(4.22)

$$\|U\|_{\mathcal{H}^{s,s'}_{p,\omega^2}(\Omega)} = \Big(\sum_{m=0}^{\infty} \sum_{l=-m}^{m} \Big[|\hat{u}_{lm}|^2_{B^s} + (1+m)^{2s'-s} \|\partial_r \hat{u}_{lm}\|^2_{\omega^2} $$

$$+ (1+m)^{2s'} \big(\|\hat{u}_{lm}\|^2 + \|\hat{u}_{lm}\|^2_{\omega^2} \big) \Big] \Big)^{\frac{1}{2}}.$$

Theorem 4.4. *Let U and U_{MN} be the solutions of (3.11) and (4.2), respectively. If $U \in L^2_p(S; X) \cap \mathcal{H}^{s,s'}_{p,\omega^{d-1}}(\Omega)$ with $d = 2, 3$ and $s, s' \geq 1$, then we have*

$$\|\nabla(U - U_{MN})\| + k\|U - U_{MN}\|$$

(4.23)

$$\lesssim \Big(C_*(M, N, k; a, b, d) N^{1-s} + (1 + kM^{-1}) M^{1-s'} \Big) \|U\|_{\mathcal{H}^{s,s'}_{p,\omega^{d-1}}(\Omega)},$$

where

(4.24)

$$C_\star(M, N, k; a, b, d) := (1 + M) b^{(d-1)/2} + M a^{\frac{d-3}{2}} |I| N^{-1}$$

$$+ k^{1/3} (M b^{3d/2-2} \sqrt{|I|} N^{-1/2} + |I|^2 b^{d/2} k^2 N^{-1}).$$

Proof. Since the proof of $d = 2, 3$ is quite similar, we shall only prove the case $d = 2$. For notational convenience, let $E_{MN} = U - U_{MN}$ and $\hat{e}_m = \hat{u}_m - \hat{u}^N_m$. Thanks to the orthogonality of the Fourier series, we have that

(4.25)

$$\|\nabla E_{MN}\|^2 + k^2 \|E_{MN}\|^2 \lesssim \sum_{|m|=0}^{M} \Big(\|\partial_r \hat{e}_m\|^2_\omega + m^2 \|\hat{e}_m\|^2_{\omega^{-1}}$$

$$+ k^2 \|\hat{e}_m\|^2_\omega \Big) + \sum_{|m|>M} \Big(\|\partial_r \hat{u}_m\|^2_\omega + m^2 \|\hat{u}_m\|^2_{\omega^{-1}} + k^2 \|\hat{u}_m\|^2_\omega \Big) := S_1^2 + S_2^2.$$

Using Theorem 4.3 leads to

(4.26)

$$S_1 \lesssim \Big(\max_{0 \leq |m| \leq M} \{ C_\star(m, \cdots) \} \Big) N^{1-s} \Big(\sum_{|m|=0}^{M} |\hat{u}_m|^2_{B^s} \Big)^{\frac{1}{2}}$$

$$\lesssim C_\star(M, N, k; a, b, d) N^{1-s} \|U\|_{\mathcal{H}^{s,s'}_{p,\omega}(\Omega)}.$$

We treat S_2 as

(4.27)

$$S_2 \lesssim M^{1-s'} \Big(\sum_{|m|>M} m^{2s'-2} \big(\|\partial_r \hat{u}_m\|^2_\omega + m^2 \|\hat{u}_m\|^2_{\omega^{-1}} \big) \Big)^{\frac{1}{2}}$$

$$+ k M^{-s'} \Big(\sum_{|m|>M} m^{2s'} \|\hat{u}_m\|^2_\omega \Big)^{\frac{1}{2}}$$

$$\lesssim \big(1 + kM^{-1} \big) M^{1-s'} \|U\|_{\mathcal{H}^{s,s'}_{p,\omega}(\Omega)}.$$

Hence, a combination of (4.25)–(4.27) yields the desired result. $\square$

5. NUMERICAL RESULTS AND DISCUSSIONS

We now present some numerical results to complement our error estimates for the spectral-Galerkin scheme (4.2). We consider the problem (3.1) in 2-D and take

(5.1) $F(r, \theta) = 0,\ \eta(\theta) = 0,\ \xi(\theta) = H_m^{(1)}(ka)e^{im\theta}.$

In this case the exact solution is: $U(r, \theta) = H_m^{(1)}(kr)e^{im\theta}$. Since for a given m, $e^{im\theta}$ can be exactly determined with number of mode $M = N_\theta \geq 2m$, we will concentrate on the approximation behavior of our scheme with respect to the frequency k and the thickness of the annulus $b - a$.

In the first set of tests, we take $a = 1$ and $b = 2$. In Figure 5.4, we present the relative L^2 error versus the number of mode $N = N_r$ for a wide range of wave numbers. We note that as soon as $N_r > k(b-a)/2$, the errors start to decay, and for moderate to large wave numbers, the errors decay slowly until about $N_r \sim k(b-a)$, and finally for $N_r > k(b-a)$, all errors converge to zero at an exponentially rate.

In the second set of tests, we take $a = 1$ and $b = 1.25$. The results are plotted in Figure 5.5. We observe similar behaviors as in the first set except that now we have $b - a = \frac{1}{4}$ and only about 1/4 of the modes are needed to achieve a similar accuracy. These behaviors are consistent with our error estimates (cf. Remark 4.2).
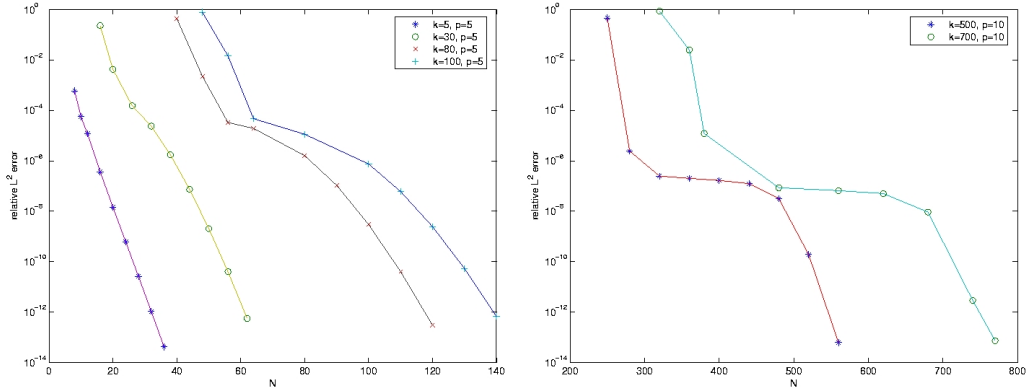

FIGURE 5.4. Relative L^2−error versus N_r as compared to an exact solution: $a = 1, b = 2$.

These results indicate that (i) the approximate solution U_{N_r, N_θ} will converge to the exact solution $U(r, \theta)$ exponentially fast as $N_r, N_\theta \to +\infty$ provided that all $F(r, \theta)$, $\xi(\theta)$ and $\eta(\theta)$ are analytic in Ω; and (ii) our numerical scheme is stable for large N_r, and capable of providing accurate results for moderate to large wave numbers .

To summarize, we have presented a complete analysis for the spectral-Galerkin method to the Helmholtz equation in exterior domains. We first studied asymptotic behaviors of the Hankel functions which play essential roles for our error analysis. Using these asymptotic estimates, we then derived *a priori* estimates with explicit dependence on the wave number for both the continuous and discrete problems. Finally, we performed an error analysis and derived error bounds with explicit dependence on the wave number. To the authors' best knowledge, our error estimates seem to be the first of its kind, i.e., with explicit dependence on the wave number for a numerical method on bounded obstacle scattering via the DtN map. A particular advantage of this approach, verified by our error estimates and numerical
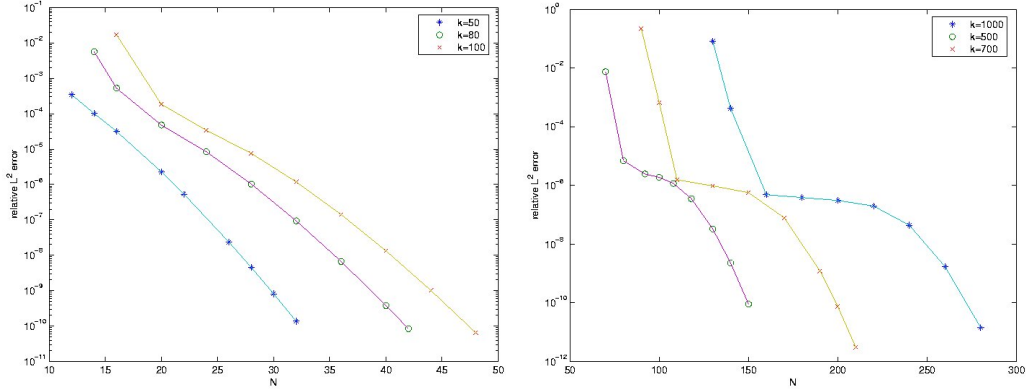
FIGURE 5.5. Relative L^2−error versus N_r as compared to an exact solution: $a = 1, b = 1.25$.

results, is that we can choose the artificial boundary very close to the scatterer while still maintaining the spectral accuracy.

APPENDIX A. THE PROOF OF (2.34)

We first prove (2.34a). It is clear that by (2.33b), $\text{Im}(\mathcal{T}_{m,\kappa}) > 0$. On the other hand, since $\text{Im}(\mathcal{T}_{m,\kappa})$ is a strictly increasing (resp. decreasing) function of κ (resp. m), we have

$$\text{Im}(\mathcal{T}_{m,\kappa}) \leq \text{Im}(\mathcal{T}_{1,\kappa}) < \text{Im}(\mathcal{T}_{1,\infty}) = 1,$$

due to the asymptotic formula $|H_1^{(1)}(\kappa)|^2 \sim \frac{2}{\pi\kappa}$ for $\kappa \gg 1$ (see Formula 9.2.3 of [1]).

We now turn to the proof of (2.34b). Recall that the modified Bessel function of the second kind of order ν is defined by

$$\text{(A.1)} \qquad \mathcal{K}_\nu(z) = \int_0^\infty e^{-z\cosh t} \cosh(\nu t) dt.$$

In particular, we have

$$\text{(A.2)} \qquad \mathcal{K}_0(z) = \int_0^\infty e^{-z\cosh t} dt, \qquad \mathcal{K}_1(z) = -\mathcal{K}_0'(z).$$

By Formula (4) on P. 445 of [26],

$$\text{(A.3a)} \qquad J_m^2(\kappa) + Y_m^2(\kappa) = \frac{8}{\pi^2} \int_0^\infty \mathcal{K}_0(2\kappa \sinh t) \cosh(2mt) dt,$$

$$\text{(A.3b)} \qquad \big[J_m J_{m+1} + Y_m Y_{m+1}\big](\kappa) = \frac{8}{\pi^2} \int_0^\infty \mathcal{K}_1(2\kappa \sinh t) \sinh\big((2m+1)t\big) dt.$$

Using the identity: $\mathcal{K}_1(z) = -\mathcal{K}_0'(z)$ and integration by parts leads to

$$\big[J_m J_{m+1} + Y_m Y_{m+1}\big](\kappa) = -\frac{8}{\pi^2} \int_0^\infty \frac{\sinh\big((2m+1)t\big)}{2\kappa(\sinh t)'} d\big(\mathcal{K}_0(2\kappa\sinh t)\big)$$

$$= -\frac{4}{\kappa\pi^2} \frac{\sinh\big((2m+1)t\big)}{\cosh t}\mathcal{K}_0(2\kappa\sinh t)\Big|_0^\infty$$

$$+ \frac{4}{\kappa\pi^2}\int_0^\infty \mathcal{K}_0(2\kappa\sinh t)\Big(\frac{\sinh\big((2m+1)t\big)}{\cosh t}\Big)' dt$$

$$= \frac{4}{\kappa\pi^2}\int_0^\infty \mathcal{K}_0(2\kappa\sinh t)\cosh(2mt)W_m(t)dt,$$

where

$$W_m(t) = \frac{1}{\cosh(2mt)}\Big(\frac{\sinh\big((2m+1)t\big)}{\cosh t}\Big)'.$$

Note that in the last step, we used the asymptotic formula (see Formula 9.7.2 of [1])

(A.4) $$\mathcal{K}_0(2\kappa\sinh t) \sim \sqrt{\frac{\pi}{2\kappa\sinh t}}e^{-2\kappa\sinh t} \sim e^{-\kappa e^t - t/2}, \qquad t \gg 1,$$

to claim that

$$\frac{\sinh\big((2m+1)t\big)}{\cosh t}\mathcal{K}_0(2\kappa\sinh t) \to 0, \quad \text{as } t \to \infty.$$

Using the identities of the Hyperbolic functions, $W_m(t)$ can be written as

(A.5) $$W_m(t) = 2m\big(1 + (\tanh t)\tanh(2mt)\big) + \text{sech}^2 t, \quad 0 \le t < \infty.$$

We now seek the maximum and minimum values of $W_m(t)$. Taking the derivative of $W_m(t)$ yields

$$W_m'(t) = 2m\big((\text{sech}^2 t)\tanh(2mt) + 2m(\tanh t)\text{sech}^2(2mt)\big) - 2(\tanh t)\text{sech}^2 t$$

$$= 2\big(m\tanh(2mt) - \tanh t\big)\text{sech}^2 t + 4m^2(\tanh t)\text{sech}^2(2mt).$$

It is obvious that

$$m\tanh(2mt) - \tanh t > 0, \quad \forall t > 0, \ \forall m \ge 1.$$

Hence, $W_m(t)$ is an increasing function of t, and consequently,

(A.6) $$2m + 1 = W_m(0) \le W_m(t) \le W_m(\infty) = 4m, \quad \forall t \ge 0, \ \forall m \ge 1.$$

Therefore, for $m \ge 1$,

(A.7) $$\frac{2m+1}{2\kappa} \le \frac{J_m(\kappa)J_{m+1}(\kappa) + Y_m(\kappa)Y_{m+1}(\kappa)}{J_m^2(\kappa) + Y_m^2(\kappa)} \le \frac{2m}{\kappa},$$

which, together with (2.33a), yields the bounds

$$-\frac{m}{\kappa} \le \text{Re}(\mathcal{T}_{m,\kappa}) \le -\frac{1}{2\kappa}, \quad \text{for } m \ge 1.$$

Finally, for $m = 0$, (A.5) implies that $0 < W_0(t) \le 1$. Accordingly, we find that

$$-\frac{1}{2\kappa} \le \text{Re}(\mathcal{T}_{0,\kappa}) < 0.$$

It remains to prove (2.34c).

Let us first show that $\text{lm}(\mathcal{T}_{0,\kappa})$ is a strictly decreasing function of κ. By (2.33b), it suffices to show that

$$f(x) := x|H_0^{(1)}(x)|^2 = x\big(J_0^2(x) + Y_0^2(x)\big), \quad x > 0,$$

is a strictly increasing function of x. Indeed, by (A.3a),

$$f(x) = \frac{8}{\pi^2} \int_0^\infty x \mathcal{K}_0(2x \sinh t) dt.$$

Differentiating it gives

$$f'(x) = \frac{8}{\pi^2} \int_0^\infty \left\{ \mathcal{K}_0(2x \sinh t) + 2x \sinh t \, \mathcal{K}_0'(2x \sinh t) \right\} dt.$$

Integrating the second term by parts leads to

$$f'(x) = \frac{8}{\pi^2} \mathcal{K}_0(2x \sinh t) \tanh t \Big|_0^\infty + \frac{8}{\pi^2} \int_0^\infty \mathcal{K}_0(2x \sinh t) \tanh^2 t dt.$$

Notice that the first term is zero due to the decay property of $\mathcal{K}_0$ (cf. (A.4)). Therefore, we have

$$f'(x) > 0, \quad \forall x > 0.$$

Finally, (2.34c) follows immediately from (2.33b), and the facts that $\mathrm{lm}(\mathcal{T}_{0,\kappa})$ is a strictly decreasing function of κ, and $\kappa |H_0^{(1)}(\kappa)|^2 \to \frac{2}{\pi}$ as $\kappa \to \infty$.

This ends the proof of (2.34).

References

[1] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. A Wiley-Interscience Publication. John Wiley & Sons Inc., New York, 1984. Reprint of the 1972 edition, Selected Government Publications.

[2] I. Babuška and S. A. Sauter. Is the pollution effect of the FEM avoidable for the Helmholtze equation considering high wave number? *SIAM J. Numer. Anal.*, 34:2392–2432, 1997.

[3] Oscar P. Bruno and Fernando Reitich. Numerical solution of diffraction problems: A method of variation of boundaries. *J. Opt. Soc. Am. A*, 10(6):1168–1175, 1993.

[4] Oscar P. Bruno and Fernando Reitich. Numerical solution of diffraction problems: A method of variation of boundaries. II. Finitely conducting gratings, Padé approximants, and singularities. *J. Opt. Soc. Am. A*, 10(11):2307–2316, 1993.

[5] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral Methods in Fluid Dynamics*. Springer-Verlag, 1987.

[6] Simon N. Chandler-Wilde and Peter Monk. Existence, uniqueness, and variational methods for scattering by unbounded rough surfaces. *SIAM J. Math. Anal.*, 37(2):598–618 (electronic), 2005.

[7] Peter Cummings and Xiaobing Feng. Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations. *Math. Models Methods Appl. Sci.*, 16(1):139–160, 2006.

[8] Leszek Demkowicz and Frank Ihlenburg. Analysis of a coupled finite-infinite element method for exterior Helmholtz problems. *Numer. Math.*, 88(1):43–73, 2001.

[9] J. Douglas, J. E. Santos, D. Sheen, and L. S. Bennethum. Frequency domain treatment of one-dimensional scalar waves. *Mathematical Models ad Methods in Applied Sciences*, 3:171–194, 1993.

[10] Marcus J. Grote and Joseph B. Keller. On nonreflecting boundary conditions. *J. Comput. Phys.*, 122(2):231–243, 1995.

[11] Ben-yu Guo and Li-lian Wang. Jacobi approximations in non-uniformly Jacobi-weighted Sobolev spaces. *J. Approx. Theory*, 128(1):1–41, 2004.

[12] Isaac Harari and Thomas J. R. Hughes. Analysis of continuous formulations underlying the computation of time-harmonic acoustics in exterior domains. *Comput. Methods Appl. Mech. Engrg.*, 97(1):103–124, 1992.

[13] Frank Ihlenburg. *Finite element analysis of acoustic scattering*, volume 132 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1998.

[14] Frank Ihlenburg and Ivo Babuška. Finite element solution of the Helmholtz equation with high wave number, part I: the h-version of FEM. *Computers Math. Applic.*, 30:9–37, 1995.

[15] Frank Ihlenburg and Ivo Babuška. Finite element solution of the Helmholtz equation with high wave number. II. The *h-p* version of the FEM. *SIAM J. Numer. Anal.*, 34(1):315–358, 1997.

[16] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I.* Springer-Verlag, New York, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.

[17] Peter Monk. *Finite element methods for Maxwell's equations.* Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 2003.

[18] Jean-Claude Nédélec. *Acoustic and electromagnetic equations*, volume 144 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2001. Integral representations for harmonic problems.

[19] D. Nicholls and Jie Shen. A stable, high–order method for two–dimensional bounded–obstacle scattering. *SIAM J. Sci. Comput.*, 28:1398–1419, 2006.

[20] David P. Nicholls and Fernando Reitich. Stability of high-order perturbative methods for the computation of Dirichlet-Neumann operators. *J. Comput. Phys.*, 170(1):276–298, 2001.

[21] David P. Nicholls and Fernando Reitich. Shape deformations in rough surface scattering: Improved algorithms. *J. Opt. Soc. Am. A*, 21(4):606–621, 2004.

[22] Lord Rayleigh. On the dynamical theory of gratings. *Proc. Roy. Soc. London*, A79:399–416, 1907.

[23] Jie Shen. Efficient spectral-Galerkin method I. direct solvers for second- and fourth-order equations by using Legendre polynomials. *SIAM J. Sci. Comput.*, 15:1489–1505, 1994.

[24] Jie Shen. Efficient spectral-Galerkin methods III. polar and cylindrical geometries. *SIAM J. Sci. Comput.*, 18:1583–1604, 1997.

[25] Jie Shen and Li-Lian Wang. Spectral approximation of the Helmholtz equation with high wave numbers. *SIAM J. Numer. Anal.*, 43(2):623–644 (electronic), 2005.

[26] G. N. Watson. *A Treatise on the Theory of Bessel Functions (second edition).* Cambridge University Press, Cambridge, UK, 1966.