Joint Analyticity of the Transformed Field and Dirichlet–Neumann Operator in Periodic Media

by

Matthew Kehoe
B.A. (Oakland University) 2010
M.S. (University of Michigan at Dearborn) 2015

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics (Applied Mathematics)
in the Graduate College of the
University of Illinois at Chicago, 2023

Chicago, Illinois

Defense Committee:
David Nicholls, Chair and Advisor
Gerard Awanou
Mimi Dai
Jan Verschelde
Thomas Royston, Bioengineering

ACKNOWLEDGMENTS

I would like to thank my advisor, David Nicholls. David has been my mentor since January 2018 and has done an outstanding job teaching me about the fields of numerical analysis, electromagnetics, and spectral element methods. His skills in computational mathematics (and in programming languages such as Matlab) have been invaluable in completing this thesis. I was lucky to have an extremely supportive advisor and couldn't have asked for a kinder person. I would also like to thank the other members of my thesis committee, including Gerard Awanou, Mimi Dai, Jan Verschelde, and Thomas Royston. I thank them for agreeing to review my thesis and have benefited from their mathematical knowledge and the face-to-face conversations we had at UIC.

I've also been fortunate to work with a lot of instructors and professors at UIC. I would like to thank Rafail Abramov, Jerry Bona, Alexey Cheskidov, Christof Sparber, Ian Tobasco, and Gyorgy Turan for speaking with me inside and outside of office hours. I also want to thank Gerard Awanou for teaching me about the theory and applications of the finite element and finite difference methods, as well as Mimi Dai for helping me improve my intuition for elliptic partial differential equations. Additionally, I am grateful to Jan Verschelde for mentoring me during my first semester as a graduate assistant for numerical analysis. I enjoyed working as a graduate teaching assistant and would like to thank Martina Bode, Matthew Lee, Jennifer Ross, Andrew Schulman, and John Steenbergen. Because of their constructive feedback, I have become a more experienced math instructor and am significantly more confident in my teaching abilities than when I started four years ago.

As a computational mathematician, I was graciously offered two summer internships through the NSF MSGI program. My first internship was at Argonne National Laboratory where I learned about the fields of tomography, the Radon transform, and inverse problems. I am thankful for the guidance from my mentor, Wendy Dai, and also for the programming help from Matthew Otten. I enjoyed working on parallel programming techniques and am fascinated by the work done on the Beebop supercomputer. My second internship was at the Cold Regions Research and Engineering Laboratory (CRREL) where I worked on two thermodynamic models (FROST and FROSTb) which predict frost-depth penetration. I would like to thank my mentor Scott Michael Slone for his assistance with the Elmer FEM software and his positive personality which made programming in Fortran more enjoyable. I would also like to thank the organizers of the MSRI Summer School programs for the interesting material presented in the summer schools on water waves, microlocal analysis, and the mathematics of big data.

In the engineering department at UIC, I have developed friendships with many graduate students who enjoy application–oriented mathematics. I fondly remember speaking

with Daniel Wen about the relationship between mathematicians and engineers and our weekly badminton sessions. I enjoyed arguing with Dan about everything and wish him the best of luck with his family in California. I also enjoyed speaking with Sam Englender throughout my brief visits to Michigan and with Neelima Borade, Joondo Chang, Jacob Mayle, and Margaret Hoeller in SEO and throughout the different math classes we had together.

My family has always been supportive of me throughout my education. I would like to thank my mother and father for the continued emotional support and for putting up with old conversations about "calculating zeros of the Riemann zeta function." My twin brother Jeff is also an engineer and is interested in mathematics. I wish him well and hope that his wedding in August goes smoothly. Both of my sisters and their husbands have also been very kind to me and I know that they will do a good job raising a family together. I'd also like to thank Frank Massey, Paul Howard, and all the other math instructors who taught me that math was interesting when I was initially focused on becoming a computer programmer.

MSK

CONTRIBUTIONS OF AUTHORS

[1] M. Kehoe and D. P. Nicholls. A Stable High–Order Perturbation of Surfaces/Asymptotic Waveform Evaluation Method for the Numerical Solution of Grating Scattering Problems. Submitted

M. Kehoe participated in the analysis of the partial differential equation, expansion around singularities, developed software in Matlab, generated figure data, and wrote part of the manuscript.

[2] M. Kehoe and D. P. Nicholls. Joint Geometry/Frequency Analyticity of Fields Scattered by Periodic Layered Media. Submitted

M. Kehoe participated in the analysis of both boundary and frequency perturbations, simplifications through the TFE method, analysis leading to the proof of joint analyticity, and wrote part of the manuscript.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF FIGURES

FIGURE **PAGE**

FIGURE **PAGE**

LIST OF TABLES

LIST OF ABBREVIATIONS

AWE	Asymptotic Waveform Evaluation
D	Energy Defect
DDM	Domain Decomposition Method
DFT	Discrete Fourier Transform
DNO	Dirichlet–Neuman Operator
DPC	Downward Propagating Condition
FE	Field Expansion
FFT	Fast Fourier Transform
HOPS	High–Order Perturbation of Surfaces
IFFT	Inverse Fast Fourier Transform
MMS	Method of Manufactured Solutions
OE	Operator Expansion
OWC	Outgoing Wave Condition
R	Reflectivity Map
SPR	Surface Plasmon Resonance
TE	Transverse Electric
TEM	Transverse Electric and Magnetic
TFE	Transformed Field Expansion
TM	Transverse Magnetic
UPC	Upward Propagating Condition

SUMMARY

This thesis presents rigorous analytical and numerical results necessary for the numerical analysis of a class of High–Order Perturbation of Surfaces/Asymptotic Waveform Evaluation (HOPS/AWE) methods in a laterally periodic two–layer structure. Numerical simulations of scattering returns from periodic diffraction gratings are crucial to a large number of applications in physics and engineering, and the work presented here examines methods for numerically modeling scattering returns from such structures. The strategies presented in this thesis represent the results of our efforts towards the dual goals of 1) Proving a theorem on the existence and uniqueness of solutions to a system of partial differential equations which model the interaction of linear waves in periodic layered media and 2) Developing a numerical algorithm to record scattered energy through a novel interfacial method that is perturbative in nature.

The first of our goals is established through classical methods based on the theory of Sobolev spaces and regular perturbation theory. The proof involves several rigorous analyses, and we formulate the scattering problem in terms of Dirichlet–Neumann Operators which are computed using the Transformed Field Expansion (TFE) methodology. A novelty of our approach is the joint analyticity of solutions with respect to both geometry and frequency perturbations. The theory itself is then validated through our second goal which is the development of a joint HOPS/AWE algorithm. For this, we develop a special class of interfacial numerical algorithms that are well–suited to periodic diffraction problems. Our algorithm calculates the Reflectivity Map, R, which measures the response (reflected energy) of a periodically corrugated grating structure as a function of its illumination frequency. Moreover, we present a series of challenging and physically relevant numerical experiments to validate the scattering results expected by our algorithm.

Forthcoming research will focus on extending the proof of analyticity to additional parameters relevant to the geometry of the structure, increasing the complexity of the structure through generalizing our results to any finite number of layered interfaces, implementing parallel programming techniques to handle multilayered surfaces, and reducing the computational cost of our HOPS/AWE algorithm. The analysis of multilayered periodic structures with numerous perturbation parameters will be an area of substantial interest for practitioners in the electromagnetic and engineering communities.

CHAPTER 1

INTRODUCTION

The theory of waves and wave propagation in periodic media has influenced many fields in the physical sciences, dating back to the influential work of Lord Rayleigh. The advancement of information and communication technologies was made possible by the use of various manifestations of waves, most notably electromagnetic waves to transmit information around the world and electrons to process information in computers. The role of wave phenomena will continue to grow in the future, especially in emerging fields of science and technology such as cryptography, medical imaging, and quantum computing. Because of the increased availability of parallel processing and high-performance computing, computer simulation has become an essential component of wave simulation, supplementing both theory and experiment. As a result, this thesis aims to extend a class of fast and robust numerical methods (known as HOPS) to simulate certain wave phenomena in a regime that is characterized by a periodic structure.

The remainder of this introductory chapter will give a brief history of the field of wave scattering and discuss early achievements of scientists and practitioners. The mathematical notation used in later sections will be introduced alongside the geometry of a two–layer periodic structure. We will also introduce the Rayleigh expansions, electromagnetic waves, TE polarization, TM polarization, and discuss the motivation behind our High–Order Perturbation of Surfaces (HOPS) schemes.

1.1 History

The scattering of acoustic and electromagnetic waves by rough interfaces has been the subject of considerable study for more than a century (1). Lord Rayleigh first investigated this problem in 1881 (2) and provided the foundation on which almost all subsequent work is based. It is possible to gain a good understanding of the mechanics of this field of scientific study and its application in light scattering by reading the works of van de Hülst (1957) (3), Twersky (1964) (4), Kerker (1969) (5), Petit (1980) (6), and Wilcox (1984) (7). For the interested reader, we recommend the Habilitationsschrift of T. Arens (2009) (8) as a definitive reference for periodic layered media problems and for the the state-of-the-art analysis of solutions to the Helmholtz and Maxwell equations in two and three dimensions.

Scattering is a process that alters the direction of light and is commonly associated with light's interaction with small particles (9). Light scatters and travels in many directions other than the propagating direction as a result of this. Light is scattered by reflection and refraction in relatively large particles, such as pigments with dimensions greater than 2.0 μm. Diffraction occurs when light is scattered by relatively small

particles with dimensions less than about 0.3 μm. When the sun is high in the sky during the day, the sky appears blue because blue light is scattered more effectively by very small particles in the atmosphere than light of longer wavelengths. When the sun is low on the horizon at sunrise and sunset, we see more of the non-scattered light, and the sky appears red.

The majority of the objects we see are visible due to light scattering from their surfaces. This is, after all, our fundamental physical observation method (5). Light scattering is determined by the wavelength or frequency of the light being scattered. Because visible light has a wavelength on the order of a micron, objects much smaller than this cannot be seen, even with a microscope. Lord Rayleigh was among the first to explain light scattering by very small particles. Rayleigh's observations show that the intensity of light scattered varies (9):

- Directly based on the intensity of incident light.
- Directly based on the average volume of scattering particles.

Lord Rayleigh also discovered that light can scatter without the use of scattering particles. This is due to the fact that changes in refractive index at different parts of a material can be sufficient to cause scattering. If a material is homogeneous, then the composition of all infinitesimal volume elements is the same and optical properties which define the material response to the incident radiation, such as transmissivity, reflectivity, and absorptivity are also the same. The aforementioned properties vary in different directions in a heterogeneous material, resulting in light scattering. On a macroscopic scale, optical properties vary over distances less than the wavelength of the incident light, resulting in the scattering of energy away from the direction of propagation.

The result of Rayleigh's observations that scattering depends on the wavelength and, thus, the color of the light is now known as the Rayleigh scattering law (10; 11). To answer the question: "Why is the sky blue in the afternoon and red at sunset or sunrise?" one may observe that blue light has a wavelength of around 400 nanometers, while red light has a wavelength of about 700 nanometers. The scattering law states that the percentage of light that will be scattered is inversely proportional to the fourth power of the wavelength. Therefore, blue light, which is at the short wavelength end of the visible spectrum, will be scattered much more strongly than red light, which is at the long wavelength end of the visible spectrum. The white light from the sun scatters and splits into different components due to particles in our environment that are roughly the same size as the wavelength of visible light. Because of their small size, oxygen and nitrogen (the major components of our atmosphere) scatter violet and blue light. This results in the blue color of the afternoon sky, since, in directions other than towards the Sun, the observer sees predominantly scattered light. In contrast, the distance that light must travel from the Sun to an observer is highest at sunrise and dusk. This signifies

that a substantial proportion of blue and violet light has been scattered, resulting in light that is predominantly of a longer wavelength and appears red to an observer.

(a) Blue Sky (Afternoon) (b) Red Sky (Dawn and Dusk)

Figure 1: Rayleigh scattering is responsible for the sky's blue tint during the day and the Sun's reddening at sunset and sunrise.

Research in the twentieth century focused on the subject of scattering from particles. In this, numerous authors contributed to the general theory of scattering by acoustic and electromagnetic waves. L. Foldy was among the first to present a complete framework (12) for the multiple scattering of a random distribution of particles. He considered the multiple scattering of scalar waves by a random distribution of isotropic scatterers through averaging a medium of uncorrelated, isotropic, point scatterers. M. Lax then expanded Foldy's work by including anisotropic scattering and pairwise correlation between particles (13). V. Twersky later extended this work through investigating the scattering of waves by multiple spheres and cylinders in a fluid, which would later lead to research in the scattering of multiple dense objects (14; 15), grating scattering (16), and the propagation of plane–compressible waves in fiber-reinforced composites (17; 18; 19).

In 1952, Twersky published a sequence of manuscripts (20; 21; 22) describing a solution to the problem of multiple scattering of radiation by an arbitrary configuration of parallel cylinders. He developed a formal model in terms of cylindrical wave functions for the scattering of an acoustic or electromagnetic wave by an array of parallel cylindrical structures which takes into account all contributions to the excitation of one cylinder by radiation scattered by the others. He then extended his solution to the case where all axes of the cylinders lie in the same plane (23). In addition, Twersky introduced methods based on Green's function (24) to describe the relationship between the scattered amplitude of an infinite grating in terms of the scattered amplitude of a single isolated cylinder. In 1961, Twersky found a method of representing the scattering coefficients in terms of elementary functions based on Schlömilch series (25). Since then, numerous studies have been conducted to confirm Twersky's findings and to expand his analysis

on cylindrical gratings (including the research on wave propagation by G. Brown in the 1980s (26; 27)).

Many other authors have contributed to the study of multiple scattering effects. J. Keller investigated wave propagation in continuous media through use of stochastic linear differential equations (28) and by including terms up to the third order in a perturbation expansion (29). U. Frisch then extended this work by developing a theory of multiple scattering of waves by a continuous random medium through perturbation expansions and approximation methods (30). Frisch applied the Feyman diagram method to identify the scattering interaction between the random surface and the random medium and demonstrated how to obtain the exact solution of a scalar wave equation by the means of functional space integration. P. Waterman and R. Truell created a rule (15) to relate the scattered wave and exciting field by defining a linear scattering operator T in a homogeneous isotropic medium governed by the Helmholtz equation. The application of Waterman's rule to scattering characteristics of particles is now known as T–matrix formalism in the engineering literature. V. Varadan, V. Bringi, and Y. Ma (31; 32) then considered vector electromagnetic waves in three–dimensions and investigated various shapes and configurations of particles. In terms of quantum mechanical scattering, L. Tsang, J. Kong, and T. Habashy applied the method of coherent potential (33; 34) to the study of multiple scattering of electromagnetic waves by a random distribution of discrete scatterers. They found that the approach of quasicrystalline approximation was particularly effective in treating electromagnetic scattering by discrete scatterers and can accurately calculate the effective propagation constants of the coherent wave. Further research is being performed by numerous authors in both the applied mathematics and engineering communities.

1.2 Motivation

The scattering of linear electromagnetic waves by a layered structure is a central model in many problems of scientific and engineering interest. Examples arise in areas such as geophysics (35; 36), imaging (37), materials science (38), nanoplasmonics (39; 40; 41), and oceanography (42). In the case of nanoplasmonics, there are many topics of interest such as extraordinary optical transmission (43), surface enhanced spectroscopy (44), and surface plasmon resonance biosensing (45; 46) and (47; 48; 49; 50). In all of the physical problems it is necessary to approximate scattering returns in a fast, robust, and highly accurate fashion. This thesis will expand upon a novel HOPS algorithm (51; 52; 53) designed for the numerical simulation of the layered periodic media (diffraction or scattering) problem.

A variety of classical algorithms have been used for simulation of this problem. However, recent studies have demonstrated (54; 51; 55; 53) that volumetric approaches (such as finite difference and finite/spectral element methods) are greatly disadvantaged

when dealing with layered media problems because of the large number of unknowns. Another natural candidate is an interfacial method based upon integral equations (IEs) (56). There are, however, also difficulties associated with these, as discussed in (54; 51; 55; 53). In the past few years, a number of these have been addressed through various techniques such as (i) the use of sophisticated quadrature rules to deliver high order spectral accuracy, (ii) the design of preconditioned iterative solvers with appropriate acceleration (57), and (iii) new tactics to avoid periodizing the Green function (58; 59; 60). Despite these alternatives (see, e.g., (61)), there are two properties that make these strategies noncompetitive in our parametrized setting. These are:

[1] For configurations parameterized by a real value ε (in our scheme the height/slope of the interface), an IE solver will return the scattering returns for only one particular value of ε. If this is changed, the solver must be run again.

[2] IE solvers require inverting a dense, nonsymmetric positive definite system of linear equations for every simulation.

In contrast, the HOPS approach (51; 55; 53) can effectively address these concerns. More specifically, in (55; 53) an alternative known as the Field Expansion (FE) method is proposed which is based on the low-order calculations of Rayleigh (62) and Rice (63). An expansion to high order was first introduced by Bruno and Reitich (64; 65; 66) and then was later enhanced and stabilized by Nicholls, Reitich, and Malcolm (67; 68; 69; 70). This latter method is known as the TFE method. The TFE method maintains all of the classical advantages of IE formulations (such as surface formulation and exact enforcement of far–field and quasi–periodic boundary conditions) while effectively addressing the two shortcomings listed above:

[1] The method is built upon expanding in the boundary parameter ε. Once the Taylor coefficients are known for the scattering quantities, the TFE method can recover all of the returns by summing the Taylor coefficients. It is unnecessary to begin a new summation for every value of ε.

[2] The scheme is based on a perturbation of the interface which, at every perturbation order, requires the inversion of a single, sparse operator corresponding to the flat-interface solution.

For a single incident wavelength, the TFE method is among the most efficient available in our layered media setting. A generalization of the HOPS approach developed by Bruno and Reitich is known as an Asymptotic Waveform Evaluation (AWE). The AWE methods (71; 72; 73; 74; 75) are built upon an additional expansion in wavelength (frequency) about a base value and will be a major source of analysis in the second half of our thesis. Our aim is to develop a novel interfacial method using a combined HOPS/AWE algorithm that provides a stable numerical scheme and a rigorous convergence

result. We will carefully show that our new algorithm is highly accurate, rapid, robust, and is jointly analytic with respect to two smallness assumptions: (i) an interfacial deformation and (ii) a frequency deformation.

1.3 Preliminaries and Notation

We consider a y–invariant, doubly layered structure with a periodic interface separating two materials; see Figure 2.

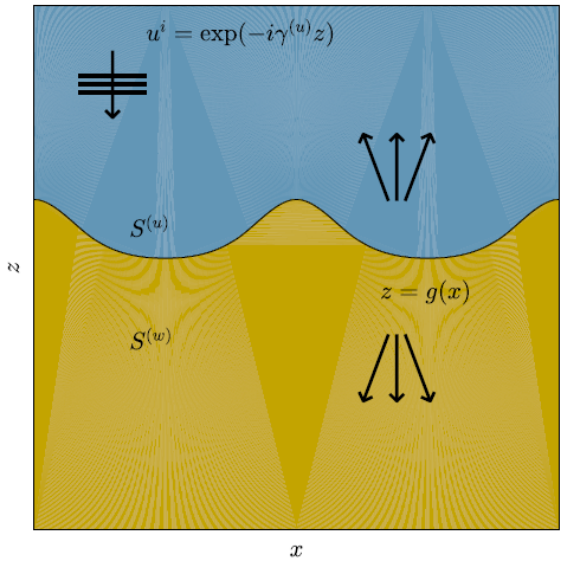


Figure 2: A two-layer structure with a periodic interface, $z = g(x)$, separating two material layers, $S^{(u)}$ and $S^{(w)}$, illuminated by plane–wave incidence.

The d–periodic interface shape is specified by the graph of the function $z = g(x)$, $g(x + d) = g(x)$. A dielectric (with refractive index n^u) occupies the domain above the interface

$$S^{(u)} := \{z > g(x)\},$$

while a material of refractive index n^w is in the lower layer

$$S^{(w)} := \{z < g(x)\}.$$

The subscripts are chosen to conform to the notation of (76; 77). The structure is illuminated from above by monochromatic plane–wave incident radiation of frequency ω and wavenumber $k^u = n^u\omega/c_0 = \omega/c^u$ (c_0 is the speed of light) aligned with the grooves

$$\underline{\mathbf{E}}^{\text{inc}}(x, z, t) = \mathbf{A}e^{-i\omega t+i\alpha x-i\gamma z}, \quad \underline{\mathbf{H}}^{\text{inc}}(x, z, t) = \mathbf{B}e^{-i\omega t+i\alpha x-i\gamma z}.$$

We consider the reduced incident fields

$$\mathbf{E}^{\text{inc}}(x,z) = e^{i\omega t}\underline{\mathbf{E}}^{\text{inc}}(x,z,t), \quad \mathbf{H}^{\text{inc}}(x,z) = e^{i\omega t}\underline{\mathbf{H}}^{\text{inc}}(x,z,t),$$
$$\alpha := k^u \sin(\theta), \quad \gamma^u := k^u \cos(\theta),$$

where the time dependence $\exp(-i\omega t)$ has been factored out. As shown in (6), the reduced electric and magnetic fields $\{\mathbf{E}, \mathbf{H}\}$ are α–quasiperiodic like the incident radiation. To close the problem we specify that the scattered radiation is "outgoing," upward propagating in $S^{(u)}$ and downward propagating in $S^{(w)}$.

It is well known (see, e.g., §1.5 – §1.7 and Petit (6)) that in this two–dimensional setting, the time–harmonic Maxwell equations decouple into two scalar Helmholtz problems which govern the Transverse Electric and Transverse Magnetic polarizations. We define the invariant (y) direction of the scattered (electric or magnetic) fields by $\tilde{u} = \tilde{u}(x,z)$ and $\tilde{w} = \tilde{w}(x,z)$ in $S^{(u)}$ and $S^{(w)}$, respectively. The incident radiation in the upper field is defined as $\tilde{u}^i(x,z)$ (which we will also denote by $\tilde{u}^{inc}(x,z)$). In Chapters 2 and 3 we will factor out the phase $\exp(i\alpha x)$ from the fields $\tilde{u}$ and $\tilde{w}$

$$u(x,z) = e^{-i\alpha x}\tilde{u}(x,z), \quad w(x,z) = e^{-i\alpha x}\tilde{w}(x,z),$$

which, we note, are d–periodic. This will simplify notation and, as discussed in Chapters 2 and 3, also remove the phase from the relevant quantities in our governing equations.

1.4 Electromagnetic Waves, Polarization, and Parameters

A wave can be described as a disturbance that travels through a medium from one location to another location (78). Waves can transfer energy from one point in space to another point in space. Therefore, there are two mechanisms which specify wave properties: The disturbance which defines the wave, and the propagation of the wave. With these, we may classify waves by the following two categories:

[1] Longitudinal Waves: When the disturbances in a wave are parallel to the wave's propagation direction, the wave is said to be a longitudinal wave. Sound waves, for example, are longitudinal waves because the pressure change occurs parallel to the wave's propagation direction.

[2] Transverse Waves: When the disturbances in a wave are perpendicular (at right angles) to the wave's propagation direction, the wave is called a transverse wave. Light is an example of a transverse wave, in which energy vibrates in a direction perpendicular to the wave's direction of motion.

Electromagnetic waves are transverse waves where both the electric and magnetic fields are perpendicular to each other and the direction of wave propagation.

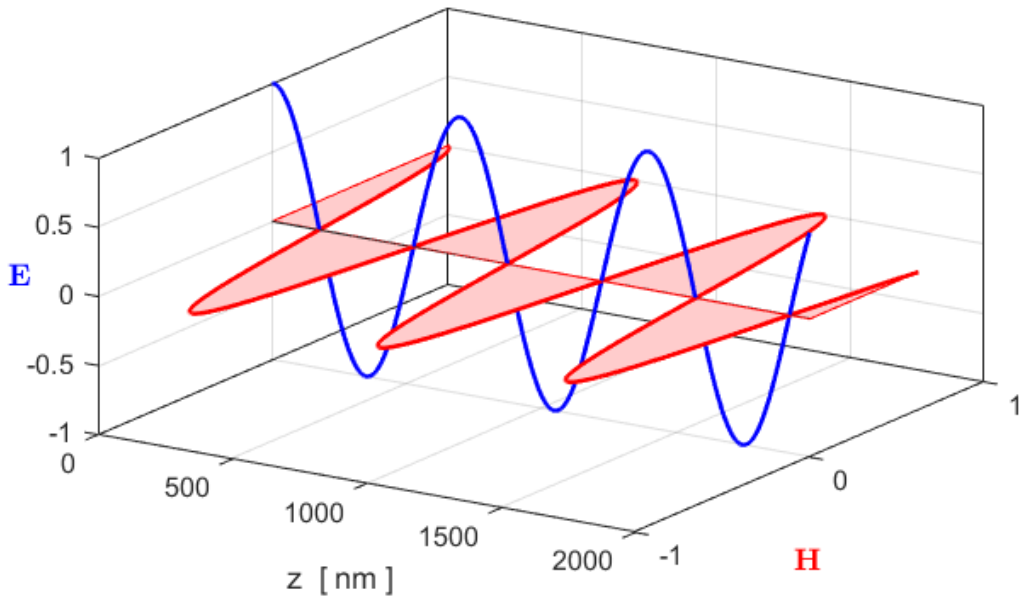


Figure 3: A light wave is an electromagnetic wave with an electric and a magnetic component. In our scenario, the electric field **E** (in blue) oscillates in the vertical direction. The magnetic field **H** (in red) is at a right angle to the electric field and oscillates in the horizontal direction. Both are perpendicular to the direction of wave propagation (**z**).

Electromagnetic energy is transmitted in waves and an electromagnetic field can propagate along various modes (79; 80; 81). The three most common modes are Transverse Electric and Magnetic (TEM), Transverse Electric (TE), and Transverse Magnetic (TM) where

- TEM Mode: In the Transverse Electric and Magnetic mode, both the electric field and the magnetic field (which, in free space, are always perpendicular to one another) are transverse (at right angles) to the direction of wave propagation (see Figure 3).
- TE Mode: In the Transverse Electric mode, the electric field is transverse to the direction of propagation while the magnetic field is parallel to the direction of propagation.
- TM Mode: In the Transverse Magnetic mode, the magnetic field is transverse to the direction of propagation while the electric field is parallel to the direction of propagation.

For various reasons the TM mode is of extraordinary importance (e.g., by the classical study of Surface Plasmon Resonance (SPR) in Raether (39)) and thus we concentrate our attention on the TM case in Chapters 5 and 6.

Throughout this thesis, we are interested in solving electromagnetic problems involving linear, homogeneous, nonmagnetic media. Our strategy, which will be discussed in detail in §1.5, is to work in the frequency domain by simplifying Maxwell's equations in matter through considering solutions where both the electric and magnetic fields are composed of time–harmonic solutions. These are solutions that have a $e^{-i\omega t}$ time–dependence for a single angular frequency ω. For frequency domain problems, two key material parameters are the permittivity ϵ and permeability μ. In vacuum, these two quantities are represented by ϵ_0 and μ_0. In addition, we are interested in representing c, the speed of light in vacuum, in terms of c_0. The index of refraction characterizes the speed of propagation of light in a medium by $n = c_0/c \geq 1$ and allows us to specify the relations

$$
\begin{aligned}
c &= \frac{c_0}{n}, && \text{(Speed in Dielectric Material)} \\
c_0 &= \frac{1}{\sqrt{\epsilon_0 \mu_0}}, && \text{(Speed of Light in Vacuum)} \\
n &= \sqrt{\frac{\epsilon \mu}{\epsilon_0 \mu_0}}, && \text{(Refractive Index)} \\
k_0 &= \frac{\omega}{c_0}, && \text{(Wavenumber in Vacuum)} \\
k &= n k_0, && \text{(Wavenumber and Refractive Index)} \\
\lambda &= \frac{2\pi c_0}{\omega}. && \text{(Wavelength)}
\end{aligned}
$$

In many cases, it is enough to specify the quantities ω and n so that the remaining dielectric parameters can be found through the permittivity and permeability of the corresponding medium. We will measure the wavelength in microns (where 1 μm = 10^{-6} m), as is common in many applications in engineering and photonics. An alternative would be to use nanometers where 1 nm = 10^{-9} m or 1 μm = 10^3 nm. In vacuum, we have $\epsilon_0 = 8.854187817 \times 10^{-12}$ F/m (farards per meter), $\mu_0 = 1.256637061 \times 10^{-6}$ H/m (henry per meter), and the speed of light becomes $c_0 = 299,792,458$ m/s (82). Additionally, we will assume that every material layer is piecewise homogeneous and isotropic, so that ϵ and μ are uniform throughout all directions of the medium.

1.5 Maxwell Equations

Following (81; 83; 84; 6; 85), we consider a region S and take as a starting point Maxwell's equations of macroscopic electromagnetism in the following form:

$$
\begin{aligned}
\nabla \times \underline{\mathbf{E}} &= -\frac{\partial \underline{\mathbf{B}}}{\partial t}, && \text{(Faraday's Law of Induction)} && \text{(1.1a)} \\
\nabla \times \underline{\mathbf{H}} &= \mathbf{J} + \frac{\partial \underline{\mathbf{D}}}{\partial t}, && \text{(Ampère's Law)} && \text{(1.1b)} \\
\nabla \cdot \underline{\mathbf{D}} &= \rho, && \text{(Gauss's Law)} && \text{(1.1c)} \\
\nabla \cdot \underline{\mathbf{B}} &= 0, && \text{(Gauss's Law for Magnetism)} && \text{(1.1d)}
\end{aligned}
$$

where $\mathbf{J}$ is the current density and ρ is the charge density. These equations link the four (time dependent) macroscopic fields

- $\underline{\mathbf{E}} = \underline{\mathbf{E}}(x, y, z, t)$: The Electric field.
- $\underline{\mathbf{H}} = \underline{\mathbf{H}}(x, y, z, t)$: The Magnetic field.
- $\underline{\mathbf{D}} = \underline{\mathbf{D}}(x, y, z, t)$: The Electric Displacement field.
- $\underline{\mathbf{B}} = \underline{\mathbf{B}}(x, y, z, t)$: The Magnetic Induction field.

The four fields are further linked via the polarization $\mathbf{P}$ and magnetization $\mathbf{M}$ by

$$\underline{\mathbf{D}} = \epsilon_0 \underline{\mathbf{E}} + \mathbf{P},$$
$$\underline{\mathbf{H}} = \frac{1}{\mu_0} \underline{\mathbf{B}} - \mathbf{M},$$

where ϵ_0 and μ_0 are the electric permittivity and magnetic permeability of vacuum. The connections between the fields depend on material properties that are defined by the quantities

- Polarization $\mathbf{P}$: The electric dipole moment per unit volume.
- Magnetization $\mathbf{M}$: The magnetic dipole moment per unit volume.

Limiting ourselves to linear, isotropic, homogenuous, nonmagnetic media, we define the constitutive relations

$$\underline{\mathbf{D}} = \epsilon_0 \epsilon_r \underline{\mathbf{E}}, \tag{1.2a}$$
$$\underline{\mathbf{B}} = \mu_0 \mu_r \underline{\mathbf{H}}, \tag{1.2b}$$

where ϵ_r is a dielectric constant representing the relative permittivity and $\mu_r = 1$ is relative permeability of the nonmagnetic medium. The linear relationship between $\underline{\mathbf{D}}$ and $\underline{\mathbf{E}}$ is often implicitly defined using the dielectric susceptibility χ, which describes the linear relationship between $\mathbf{P}$ and $\underline{\mathbf{E}}$ via

$$\mathbf{P} = \epsilon_0 \chi \underline{\mathbf{E}}.$$

From here, one finds

$$\underline{\mathbf{D}} = \epsilon_0 \underline{\mathbf{E}} + \mathbf{P} = \epsilon_0 (1 + \chi) \underline{\mathbf{E}},$$

which from (1.2a) yields $\epsilon_r = 1 + \chi$. Substituting (1.2) into (1.1) produces

$$\nabla \times \underline{\mathbf{E}} = -\mu_0 \frac{\partial \underline{\mathbf{H}}}{\partial t}, \tag{1.3a}$$

$$\nabla \times \underline{\mathbf{H}} = \mathbf{J} + \epsilon_0 \epsilon_r \frac{\partial \underline{\mathbf{E}}}{\partial t}, \tag{1.3b}$$

$$\nabla \cdot \underline{\mathbf{E}} = \rho/(\epsilon_0 \epsilon_r), \tag{1.3c}$$

$$\nabla \cdot \underline{\mathbf{H}} = 0. \tag{1.3d}$$

In consideration of our particular scenario, we assume there are no free charges (requiring $\rho \equiv 0$). We model the current density with the linear relationship

$$\mathbf{J} = \sigma \underline{\mathbf{E}},$$

which is known as Ohm's law. The scalar σ represents the conductivity of an isotropic material. To work in the frequency domain and obtain time-harmonic solutions of the form

$$\underline{\mathbf{E}}(x,y,z,t) = \mathbf{E}(x,y,z)e^{-i\omega t}, \quad \underline{\mathbf{H}}(x,y,z,t) = \mathbf{H}(x,y,z)e^{-i\omega t}, \tag{1.4}$$

we insert (1.4) into (1.3) to obtain

$$\nabla \times \mathbf{E} = i\omega\mu_0 \mathbf{H}, \tag{1.5a}$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon_0 \epsilon \mathbf{E}, \tag{1.5b}$$

$$\nabla \cdot \mathbf{E} = 0, \tag{1.5c}$$

$$\nabla \cdot \mathbf{H} = 0, \tag{1.5d}$$

where

$$\epsilon := \epsilon' + i\epsilon'', \quad \epsilon' = \epsilon_r, \quad \epsilon'' = \sigma/(\omega\epsilon_0),$$

is the complex permittivity. A dielectric (or insulator) is the name given to a material for which

$$\sigma/(\omega\epsilon_0) \ll \epsilon' \implies \mathrm{Im}(\epsilon) \approx 0,$$

and a perfect insulator is a material where $\sigma = 0$ which implies $\mathrm{Im}(\epsilon) = 0$. An example is vacuum where $\epsilon = 1$. A metal (or conductor) is the name given to a material which satisfies

$$\epsilon'' = \sigma/(\omega\epsilon_0) \approx \epsilon_r.$$

Examples of good conductors are copper and silver. We call a material a perfect conductor if $\sigma \to \infty$. To arrive at the governing equations for scattered grating, we also demand that solutions are quasiperiodic

$$\mathbf{E}(x+d_1, y+d_2, z) = e^{i\alpha d_1 + i\beta d_2}\mathbf{E}(x,y,z), \tag{1.6a}$$

$$\mathbf{H}(x+d_1, y+d_2, z) = e^{i\alpha d_1 + i\beta d_2}\mathbf{H}(x,y,z), \tag{1.6b}$$

and outgoing. Then at the material interface (83) we have continuity of both the tangential components of the electric field and the normal components of the magnetic field. Finally, we recognize that jumps in the normal electric field and tangential magnetic field are specified by

$$\mathbf{N}\times\mathbf{E} = 0, \quad \mathbf{N}\times\mathbf{H} = \mathbf{j}_s, \quad \mathbf{N}\cdot(\epsilon\mathbf{E}) = \rho_s, \quad \mathbf{N}\cdot\mathbf{H} = 0, \tag{1.7}$$

where $\mathbf{N}$ is normal to the interface, $\mathbf{j}_s$ represents the surface current density, and ρ_s is the surface charge density. In the case that all of the permittivities and permeabilities are finite, the surface current density is zero. This allows us to enforce tangential continuity of the fields $\mathbf{E}$ and $\mathbf{H}$ as

$$\mathbf{N}\times\mathbf{E} = 0, \quad \mathbf{N}\times\mathbf{H} = 0. \tag{1.8}$$

In the setting of grating structures, we choose an interface shaped by $z = g(x,y)$ and define the normal of the interface as $\mathbf{N} := (-\partial_x g, -\partial_y g, 1)^T$. Therefore in a doubly layered medium our governing equations become

$$\nabla\times\mathbf{E}^{(u)} = i\omega\mu_0\mathbf{H}^{(u)}, \qquad z > g(x,y), \tag{1.9a}$$

$$\nabla\times\mathbf{H}^{(u)} = -i\omega\epsilon_0\epsilon^{(u)}\mathbf{E}^{(u)}, \qquad z > g(x,y), \tag{1.9b}$$

$$\nabla\times\mathbf{E}^{(w)} = i\omega\mu_0\mathbf{H}^{(w)}, \qquad z < g(x,y), \tag{1.9c}$$

$$\nabla\times\mathbf{H}^{(w)} = -i\omega\epsilon_0\epsilon^{(w)}\mathbf{E}^{(w)}, \qquad z < g(x,y), \tag{1.9d}$$

$$\mathbf{N}\times\left[\mathbf{E}^{(u)} - \mathbf{E}^{(w)}\right] = -\mathbf{N}\times\mathbf{E}^{\text{inc}}, \qquad z = g(x,y), \tag{1.9e}$$

$$\mathbf{N}\times\left[\mathbf{H}^{(u)} - \mathbf{H}^{(w)}\right] = -\mathbf{N}\times\mathbf{H}^{\text{inc}}, \qquad z = g(x,y). \tag{1.9f}$$

Here, $\{\mathbf{E}^{(u)}, \mathbf{H}^{(u)}\}$ and $\{\mathbf{E}^{(w)}, \mathbf{H}^{(w)}\}$ represent outgoing, quasiperiodic, divergence free electric and magnetic fields defined in the upper $\left(S^{(u)} = \{z > g\}\right)$ and lower $\left(S^{(w)} = \{z < g\}\right)$ media. The constants $\epsilon^{(u)}$ and $\epsilon^{(w)}$ represent the permittivities which fill the two material layers.

To simplify future developments, we make two assumptions which allow us to focus on scalar solutions in two dimensions. Our first assumption is that the grating structure is invariant in the y–direction so that the interface shape becomes

$$z = g(x).$$

This implies that $-\partial_y g = 0$ and the interface normal becomes

$$\mathbf{N} = \begin{pmatrix} -\partial_x g \\ 0 \\ 1 \end{pmatrix}. \tag{1.10}$$

The second assumption is that the incident radiation is aligned with the invariant grooves of the grating structure. In this case, in TE polarization the electric field takes the form

$$\mathbf{E}^{\text{inc}} = \mathbf{E}^{\text{inc}}(x,z) = \mathbf{A}e^{i\alpha x - i\gamma z}, \quad \mathbf{A} = \begin{pmatrix} 0 \\ A \\ 0 \end{pmatrix}, \tag{1.11}$$

while in TM polarization the magnetic field can be written as

$$\mathbf{H}^{\text{inc}} = \mathbf{H}^{\text{inc}}(x,z) = \mathbf{B}e^{i\alpha x - i\gamma z}, \quad \mathbf{B} = \begin{pmatrix} 0 \\ B \\ 0 \end{pmatrix}. \tag{1.12}$$

1.6 Tranverse Electric (TE) Polarization

Suppose that the electric field is transverse to the direction of propagation while the magnetic field is parallel to the direction of propagation. Then the electric field has only a transverse component and we seek solutions satisfying

$$\mathbf{E} = \mathbf{E}(x,z) = \begin{pmatrix} 0 \\ \tilde{v}(x,z) \\ 0 \end{pmatrix}, \quad \mathbf{H} = \mathbf{H}(x,z) = \begin{pmatrix} H^x(x,z) \\ 0 \\ H^z(x,z) \end{pmatrix}. \tag{1.13}$$

In order to satisfy the time–harmonic Maxwell equations we calculate

$$\nabla \times \mathbf{E} = \begin{pmatrix} -\partial_z \tilde{v} \\ 0 \\ \partial_x \tilde{v} \end{pmatrix},$$

which implies

$$\mathbf{H} = \frac{1}{i\omega\mu_0} \nabla \times \mathbf{E} = \begin{pmatrix} -\partial_z \tilde{v}/(i\omega\mu_0) \\ 0 \\ \partial_x \tilde{v}/(i\omega\mu_0) \end{pmatrix}.$$

Similarly,

$$\nabla \times \mathbf{H} = \begin{pmatrix} 0 \\ \partial_z H^x - \partial_x H^z \\ 0 \end{pmatrix},$$

so that we can reduce (1.9b) and (1.9d) from

$$-i\omega\epsilon_0\epsilon\mathbf{E} = \nabla \times \mathbf{H},$$

to one equation in the y–component

$$\operatorname{div}\left[-\frac{1}{i\omega\mu_0}\nabla\tilde{v}\right] = -i\omega\epsilon_0\epsilon\tilde{v}.$$

As the divergence of the gradient is the Laplacian and μ_0 is constant, we obtain

$$0 = \Delta\tilde{v} + \omega^2\mu_0\epsilon_0\epsilon\tilde{v} = \Delta\tilde{v} + \frac{\omega^2}{c_0^2}\epsilon\tilde{v} = \Delta\tilde{v} + k_0^2\epsilon\tilde{v} = \Delta\tilde{v} + k^2\tilde{v}. \tag{1.14}$$

The boundary conditions become

$$0 = \mathbf{N} \times \mathbf{E} = \begin{pmatrix} -\tilde{v} \\ 0 \\ -(\partial_x g)\tilde{v} \end{pmatrix}, \tag{1.15}$$

and

$$0 = \mathbf{N} \times \mathbf{H} = \begin{pmatrix} 0 \\ (\partial_x g)H^z + H^x \\ 0 \end{pmatrix} = \frac{1}{i\omega\mu_0}\begin{pmatrix} 0 \\ (\partial_x g)\partial_x\tilde{v} - \partial_z\tilde{v} \\ 0 \end{pmatrix}. \tag{1.16}$$

The first boundary condition (1.15) shows that $\tilde{v}$ is continuous across interfaces while the second boundary condition (1.16) warrants that $\partial_N\tilde{v}$ is also continuous across interfaces. This follows from the fact that μ_0 is a constant equal to the permeability of vacuum in all media and is therefore constant across boundaries. As a consequence, in a doubly layered medium the TE governing equations are

$$\Delta\tilde{u} + (k^u)^2\tilde{u} = 0, \qquad z > g(x), \tag{1.17a}$$

$$\Delta\tilde{w} + (k^w)^2\tilde{w} = 0, \qquad z < g(x), \tag{1.17b}$$

$$\tilde{u} - \tilde{w} = -\tilde{u}^{inc}, \qquad z = g(x), \tag{1.17c}$$

$$\partial_N\tilde{u} - \tau^2\partial_N\tilde{w} = -\partial_N\tilde{u}^{inc}, \qquad z = g(x), \tag{1.17d}$$

where

$$\tau^2 = \frac{\varepsilon^u}{\varepsilon^w} = \frac{(k^u)^2}{(k^w)^2} = \frac{(n^u)^2}{(n^w)^2},$$

and $\tilde{u}$ and $\tilde{w}$ are defined as outgoing, quasiperiodic solutions (in the y–component) of the electric field in the upper and lower layers. To clarify what is meant by solutions that are bounded, outgoing, and quasiperiodic, we will introduce an Outgoing Wave Condition (OWC) in §1.8.

1.7 Tranverse Magnetic (TM) Polarization

If we instead assume that the magnetic field is transverse to the direction of propagation while the electric field is parallel to the direction of propagation, then the magnetic field is composed entirely of a transverse component and we seek solutions satisfying

$$\mathbf{H} = \mathbf{H}(x,z) = \begin{pmatrix} 0 \\ \tilde{v}(x,z) \\ 0 \end{pmatrix}, \quad \mathbf{E} = \mathbf{E}(x,z) = \begin{pmatrix} E^x(x,z) \\ 0 \\ E^z(x,z) \end{pmatrix}. \tag{1.18}$$

We may once again satisfy the time–harmonic Maxwell equations by calculating

$$\nabla \times \mathbf{H} = \begin{pmatrix} -\partial_z \tilde{v} \\ 0 \\ \partial_x \tilde{v} \end{pmatrix},$$

which implies

$$\mathbf{E} = \frac{-1}{i\omega\epsilon_0\epsilon} \nabla \times \mathbf{H} = \frac{1}{\epsilon} \begin{pmatrix} \partial_z \tilde{v}/(i\omega\epsilon_0) \\ 0 \\ -\partial_x \tilde{v}/(i\omega\epsilon_0) \end{pmatrix}.$$

Similarly to TE polarization, we find

$$\nabla \times \mathbf{E} = \begin{pmatrix} 0 \\ \partial_z E^x - \partial_x E^z \\ 0 \end{pmatrix},$$

and we can reduce (1.9a) and (1.9c)

$$i\omega\mu_0 \mathbf{H} = \nabla \times \mathbf{E},$$

to one equation in the y–component

$$\operatorname{div}\left[\frac{1}{i\omega\epsilon_0\epsilon}\nabla\tilde{v}\right] = i\omega\mu_0\tilde{v}.$$

As ϵ changes value between layers, we find

$$0 = \operatorname{div}\left[\frac{1}{\epsilon}\nabla\tilde{v}\right] + \omega^2\mu_0\epsilon_0\tilde{v} = \operatorname{div}\left[\frac{1}{\epsilon}\nabla\tilde{v}\right] + \frac{\omega^2}{c^2}\tilde{v} = \operatorname{div}\left[\frac{1}{\epsilon}\nabla\tilde{v}\right] + k_0^2\tilde{v}. \tag{1.19}$$

If the layers are homogeneous then we may reduce (1.19) in each layer to

$$0 = \Delta\tilde{v} + k^2\tilde{v}. \tag{1.20}$$

The boundary conditions become

$$0 = \mathbf{N} \times \mathbf{H} = \begin{pmatrix} -\tilde{v} \\ 0 \\ -(\partial_x g)\tilde{v} \end{pmatrix}, \tag{1.21}$$

and

$$0 = \mathbf{N} \times \mathbf{E} = \begin{pmatrix} 0 \\ (\partial_x g)E^z + E^x \\ 0 \end{pmatrix} = \frac{-1}{i\omega\epsilon_0} \begin{pmatrix} 0 \\ \left[(\partial_x g)\partial_x \tilde{v} - \partial_z \tilde{v}\right]/\epsilon \\ 0 \end{pmatrix}. \tag{1.22}$$

Similarly to TE polarization, the first boundary condition (1.21) shows that $\tilde{v}$ is continuous across interfaces. However, the second boundary condition (1.22) mandates that $(1/\epsilon)\partial_N \tilde{v}$ is continuous across interfaces. This follows from the fact that ϵ_0 is constant everywhere and ϵ is allowed to jump across layer interfaces. Therefore in a doubly layered medium the TM governing equations are

$$\Delta\tilde{u} + (k^u)^2\tilde{u} = 0, \qquad z > g(x), \tag{1.23a}$$

$$\Delta\tilde{w} + (k^w)^2\tilde{w} = 0, \qquad z < g(x), \tag{1.23b}$$

$$\tilde{u} - \tilde{w} = -\tilde{u}^{inc}, \qquad z = g(x), \tag{1.23c}$$

$$\partial_N\tilde{u} - \tau^2\partial_N\tilde{w} = -\partial_N\tilde{u}^{inc}, \qquad z = g(x), \tag{1.23d}$$

where, as in TE polarization, $\tilde{u}$ and $\tilde{w}$ are defined as outgoing, quasiperiodic solutions (in the y–component) of the magnetic field in the upper and lower layers.

1.8 Rayleigh Expansions

In order to make precise the far–field boundary conditions we desire, we study solutions of the following boundary value problem

$$\Delta\tilde{u} + (k^u)^2\tilde{u} = 0, \qquad \text{in } S^{(u)}, \tag{1.24a}$$

$$\tilde{u}(x, g(x)) = \tilde{\zeta}^u(x), \qquad \text{at } z = g(x), \tag{1.24b}$$

$$\tilde{u}(x + d, z) = e^{i\alpha d}\tilde{u}(x, z), \tag{1.24c}$$

$$\mathrm{OWC}[\tilde{u}] = 0, \qquad z \to \infty, \tag{1.24d}$$

which are outgoing, bounded, and quasiperiodic. The fourth condition (1.24d) mandates that solutions are both outgoing and bounded and is known as the Outgoing Wave Condition. To make these boundary conditions more precise, we first observe that for $z > a > |g|_\infty$ the solution to (1.24) in $S^{(u)}$ is given by

$$\tilde{u}(x, z) = \sum_{p=-\infty}^{\infty} a_p e^{i\alpha_p x + i\gamma_p^u z} + \sum_{p=-\infty}^{\infty} b_p e^{i\alpha_p x - i\gamma_p^w z}. \tag{1.25}$$

In this setting (and in many other places in this thesis), we let $p \in \mathbb{Z},\ q \in \{u, w\}$, and define

$$\alpha_p := \alpha + \left(\frac{2\pi}{d}\right) p, \quad \gamma_p^q := \begin{cases} \sqrt{(k^q)^2 - \alpha_p^2}, & p \in \mathcal{U}^q, \\ i\sqrt{\alpha_p^2 - (k^q)^2}, & p \notin \mathcal{U}^q, \end{cases} \tag{1.26}$$

and

$$\mathcal{U}^q := \{p \in \mathbb{Z} \mid \alpha_p^2 < (k^q)^2\}. \tag{1.27}$$

To enforce the requirement that our solution (1.25) is outgoing and bounded, we require $b_p \equiv 0$. If $b_p \not\equiv 0$ then solutions will be inward propagating for $p \in \mathcal{U}^u$, and unbounded for $p \notin \mathcal{U}^u$. To clarify what is meant by the Outgoing Wave Condition, we observe that for $p \in \mathcal{U}^u$, solutions are outgoing and the modes are "propagating." In contrast, if $p \notin \mathcal{U}^u$, then solutions decay exponentially and the modes are known as "evanescent." Therefore the solutions of (1.24a) which satisfy the Outgoing Wave Condition, (1.24d), are

$$\tilde{u}(x, z) = \sum_{p=-\infty}^{\infty} a_p e^{i\alpha_p x + i\gamma_p^u z}. \tag{1.28}$$

A similar argument for $z < -b < -|g|_\infty$ will show that solutions in the lower field which satisfy the Outgoing Wave Condition are given by

$$\tilde{w}(x, z) = \sum_{p=-\infty}^{\infty} b_p e^{i\alpha_p x - i\gamma_p^w z}. \tag{1.29}$$

This leads to a domain decomposition of $S^{(u)}$ where we introduce an "Artificial Boundary" at $\{z = a\}$ and define the truncated domain

$$S_{g,a} := \{g(x) < z < a\}.$$

We similarly define an "Artificial Boundary" in the lower field, $S^{(w)}$, at $\{z = -b\}$ and define the truncated domain

$$S_{g,-b} := \{-b < z < g(x)\}.$$

We can now state a new boundary value problem that is equivalent to (1.24) as

$$\Delta\tilde{u} + (k^u)^2\tilde{u} = 0, \qquad \text{in } S_{g,a}, \tag{1.30a}$$
$$\tilde{u}(x, g(x)) = \tilde{\zeta}^u(x), \qquad \text{at } z = g(x), \tag{1.30b}$$
$$\tilde{u}(x + d, z) = e^{i\alpha d}\tilde{u}(x, z), \tag{1.30c}$$
$$\Delta\tilde{v} + (k^u)^2\tilde{v} = 0, \qquad z > a, \tag{1.30d}$$
$$\tilde{u} = \tilde{v}, \qquad \text{at } z = a, \tag{1.30e}$$
$$\partial_z\tilde{u} = \partial_z\tilde{v}, \qquad \text{at } z = a, \tag{1.30f}$$
$$\tilde{v}(x + d, z) = e^{i\alpha d}\tilde{v}(x, z), \tag{1.30g}$$
$$\mathrm{OWC}[\tilde{v}] = 0, \qquad z \to \infty, \tag{1.30h}$$

By the same analysis leading to (1.28), solutions to (1.30d) are of the form

$$\tilde{v}(x, z) = \sum_{p=-\infty}^{\infty} c_p e^{i\alpha_p x + i\gamma_p^u z}. \tag{1.31}$$

From (1.30e) it is clear that if we define $\psi(x) := \tilde{u}(x, a)$ and use $\tilde{v}(x, a) = \tilde{u}(x, a)$ then

$$\tilde{v}(x, z) = \sum_{p=-\infty}^{\infty} \left(c_p e^{i\gamma_p^u a}\right) e^{i\alpha_p x + i\gamma_p^u (z-a)} = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\alpha_p x + i\gamma_p^u (z-a)},$$

where $\hat{\psi}_p$ are the Fourier coefficients of ψ. To enforce (1.30f) we compute

$$\partial_z\tilde{v}(x, a) = \sum_{p=-\infty}^{\infty} \left(i\gamma_p^u\right) \hat{\psi}_p e^{i\alpha_p x},$$

and define the Dirichlet–Neuman Operator (DNO)

$$\tilde{T}^u : \tilde{v}(x, a) \to \left(\partial_z\tilde{v}\right)(x, a), \tag{1.32}$$

Equation (1.30f) now implies, at $z = a$,

$$0 = \partial_z\tilde{u} - \partial_z\tilde{v} = \partial_z\tilde{u} - \tilde{T}^u[\tilde{v}] = \partial_z\tilde{u} - \tilde{T}^u[\tilde{u}],$$

where

$$\tilde{T}^u[\psi(x)] := \sum_{p=-\infty}^{\infty} \left(i\gamma_p^u\right) \hat{\psi}_p e^{i\alpha_p x}.$$

A similar calculation can be performed in the lower field. At $z = -b$ and $\psi(x) := \tilde{v}(x, -b)$ we find

$$\tilde{v}(x, z) = \sum_{p=-\infty}^{\infty} \left(d_p e^{i\gamma_p^w b}\right) e^{i\alpha_p x - i\gamma_p^w (z+b)} = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\alpha_p x - i\gamma_p^w (z+b)},$$

where $\hat{\psi}_p$ are the Fourier coefficients of ψ. For $z = -b$ and an equivalent representation of (1.30) in the lower field, we deduce

$$\partial_z \tilde{v}(x, -b) = \sum_{p=-\infty}^{\infty} \left(-i\gamma_p^w\right) \hat{\psi}_p e^{i\alpha_p x}. \tag{1.33}$$

Hence, we may state the boundary value problem in (1.24) and (1.30) equivalently as

$$\begin{aligned}
\Delta\tilde{u} + (k^u)^2\tilde{u} &= 0, && \text{in } S_{g,a}, && \text{(1.34a)}\\
\tilde{u}(x, g(x)) &= \tilde{\zeta}^u(x), && \text{at } z = g(x), && \text{(1.34b)}\\
\tilde{u}(x + d, z) &= e^{i\alpha d}\tilde{u}(x, z), && && \text{(1.34c)}\\
\partial_z\tilde{u} - \tilde{T}^u[\tilde{u}] &= 0, && \text{at } z = a. && \text{(1.34d)}
\end{aligned}$$

The final condition (1.34d) is known as a Transparent Boundary Condition at the Artificial Boundary $\{z = a\}$. A similar analysis in the lower field shows that downward propagating solutions which satisfy the Outgoing Wave Condition satisfy

$$\partial_z\tilde{w} - \tilde{T}^w[\tilde{w}] = 0, \quad \text{at } z = -b,$$

where

$$\tilde{T}^w[\psi(x)] := \sum_{p=-\infty}^{\infty} \left(-i\gamma_p^w\right) \hat{\psi}_p e^{i\alpha_p x},$$

and the corresponding boundary value problem becomes

$$\begin{aligned}
\Delta\tilde{w} + (k^w)^2\tilde{w} &= 0, && \text{in } S_{g,-b}, && \text{(1.35a)}\\
\tilde{w}(x, g(x)) &= \tilde{\zeta}^w(x), && \text{at } z = g(x), && \text{(1.35b)}\\
\tilde{w}(x + d, z) &= e^{i\alpha d}\tilde{w}(x, z), && && \text{(1.35c)}\\
\partial_z\tilde{w} - \tilde{T}^w[\tilde{w}] &= 0, && \text{at } z = -b. && \text{(1.35d)}
\end{aligned}$$

1.9 Thesis Outline

This thesis is divided into four parts, specifically: (1) Background and Introduction, (2) Joint Analyticity of the Upper and Lower Fields, (3) Numerical Results and Scattering, and (4) Concluding Remarks and Future Research. The "Joint Analyticity of the Upper and Lower Fields" part is contained in Chapters 2–4. In Chapters 2 and 3, we discuss our general strategy for establishing analyticity results which are based on a special change of variables and the elliptic theory of Sobolev spaces. The concluding section of these chapters details the mechanics of our HOPS algorithm and provides example profiles where our algorithm is highly accurate and robust. Chapter 4 combines these results to establish our main theorem (Theorem 4.6.1) which proves the existence and

uniqueness of solutions to a system of partial differential equations with respect to both interfacial and frequency deformations.

The "Numerical Results and Scattering" part is composed of Chapters 5 and 6. In Chapter 5, we describe the Method of Manufactured Solutions as a tool to demonstrate the accuracy of our HOPS algorithm. Then, in Chapter 6, we define the Reflectivity Map as a way of computing the reflected energy stored by a periodic structure. We simulate a large variety of scattering problems using a range of wavelengths and dielectric constants in both the TE and TM propagation modes. The "Concluding Remarks" part is given in Chapter 7 where we discuss several different possibilities for future research. Among these, the most interesting is Section 7.4 where we analyze the necessary steps to extend our analyticality theorem from two parameters to any finite integer $M > 0$.

CHAPTER 2

ANALYTICITY OF THE UPPER FIELD

2.1 Introduction

We now present all of the information necessary for establishing the analyticity of the upper field and the upper layer DNO. Our strategy is to first remove the phase from our governing equations in §2.2, introduce a domain–flattening change of variables in §2.4, and then seek solutions as a joint Taylor series in two small perturbation variables: an interfacial deformation (§2.4) and a frequency deformation (§2.5). These lead to the TFE recursions (§2.6) from which we can use Sobolev space theory to establish analyticity results. The analyticity of the upper field with respect to a single interfacial deformation is established in §2.8 while the joint analyticity in two small perturbations is established in §2.9. The third case for a single frequency deformation follows directly as a special case of Theorem 2.9.2 and the analyticity of the upper layer DNO is proven in Theorem 2.10.2. The concluding section of this chapter demonstrates a Fourier–Chebyshev approach for simulating the TFE recursions giving a HOPS/AWE algorithm whose advantageous numerical properties we explore.

2.2 Governing Equations Without Phase

In the upper field, we defined the geometry $S_{g,a} := \{g(x) < z < a\}$ where z is bounded between a constant imposed by the Artificial Boundary $\{z = a\}$ and the lower surface $g(x)$. The boundary value problem (1.34) defined in §1.8 gives an equivalent representation of the governing equations of linear wave propagation in a single homogeneous material layer

$$\Delta\tilde{u} + (k^u)^2\tilde{u} = 0, \qquad g(x) < z < a, \tag{2.1a}$$

$$\tilde{u}(x, g(x)) = \tilde{\zeta}^u(x), \qquad \text{at } z = g(x), \tag{2.1b}$$

$$\tilde{u}(x + d, z) = e^{i\alpha d}\tilde{u}(x, z), \tag{2.1c}$$

$$\partial_z\tilde{u} - \tilde{T}^u[\tilde{u}] = 0, \qquad \text{at } z = a, \tag{2.1d}$$

In our subsequent developments it will be convenient to consider periodic unknowns rather than quasiperiodic ones. This can be readily achieved by the simple phase extraction

$$u(x, z) := e^{-i\alpha x}\tilde{u}(x, z), \tag{2.2}$$

where by (2.1c)

$$\begin{aligned} u(x+d,z) = e^{-i\alpha(x+d)}\tilde{u}(x+d,z) &= e^{-i\alpha(x+d)}e^{i\alpha d}\tilde{u}(x,z) \\ &= e^{-i\alpha x}\tilde{u}(x,z) \\ &= u(x,z), \end{aligned}$$

so the α-quasiperiodicity of $\tilde{u}(x,z)$ implies that $u(x,z)$ is d–periodic. We also compute

$$\partial_x\tilde{u}(x,z) = (i\alpha)e^{i\alpha x}u(x,z) + e^{i\alpha x}\partial_x u(x,z),$$

$$\partial_x^2\tilde{u}(x,z) = (i\alpha)^2e^{i\alpha x}u(x,z) + 2(i\alpha)e^{i\alpha x}\partial_x u(x,z) + e^{i\alpha x}\partial_x^2 u(x,z),$$

$$\partial_z\tilde{u}(x,z) = e^{i\alpha x}\partial_z u(x,z),$$

$$\partial_z^2\tilde{u}(x,z) = e^{i\alpha x}\partial_z^2 u(x,z).$$

These in turn imply

$$\begin{aligned} 0 = \Delta\tilde{u} + (k^u)^2\tilde{u} &= -\alpha^2e^{i\alpha x}u + 2(i\alpha)e^{i\alpha x}\partial_x u + e^{i\alpha x}\partial_x^2 u + e^{i\alpha x}\partial_z^2 u + (k^u)^2e^{i\alpha x}u \\ &= e^{i\alpha x}\big(\Delta u + 2i\alpha\partial_x u + \big((k^u)^2 - \alpha^2\big)\,u\big), \end{aligned}$$

and, setting $(\gamma^u)^2 = (k^u)^2 - \alpha^2$, (2.1a) becomes

$$\Delta u + 2i\alpha\partial_x u + \gamma^2 u = 0, \quad g(x) < z < a. \tag{2.3}$$

Similarly, the boundary condition (2.1b) becomes

$$u(x,g(x)) = e^{-i\alpha x}\tilde{u}(x,g(x)) = e^{-i\alpha x}\tilde{\zeta}^u(x) =: \zeta^u(x), \quad \text{at } z = g(x).$$

As we will show in §2.3, the Transparent Boundary Condition for (2.1d) becomes

$$\partial_z\left[u(x,a)\right] - T^u[u(x,a)] = 0, \quad \text{at } z = a. \tag{2.4}$$

Our governing equations are now

$$\Delta u + 2i\alpha\partial_x u + (\gamma^u)^2 u = 0, \qquad g(x) < z < a, \tag{2.5a}$$

$$u(x,g(x)) = \zeta^u(x), \qquad \text{at } z = g(x), \tag{2.5b}$$

$$u(x+d,z) = u(x,z), \tag{2.5c}$$

$$\partial_z\left[u(x,a)\right] - T^u[u(x,a)] = 0, \qquad \text{at } z = a. \tag{2.5d}$$

2.3 Fourier Multipliers and the Dirichlet–Neumann Operator

In this section we examine the concept of a Fourier Multiplier and its relation to the DNO $\tilde{T}^u$ defined in §1.8 by (1.32). Our goal is to give an explicit representation of the DNO $\tilde{T}^u$ and Transparent Boundary Condition, (2.1d), when we remove the phase.

We define a Fourier Multiplier, $m(D)$, as the operator with the property that

$$m(D)\left[\psi(x)\right] := \sum_{p=-\infty}^{\infty} m(p)\hat{\psi}_p e^{i\alpha_p x}.$$

A classical derivative can be expressed as

$$\partial_x \psi = \sum_{p=-\infty}^{\infty} (i\alpha_p)\hat{\psi}_p e^{i\alpha_p x} = (i\alpha_D)\psi,$$

and similarly for the operator $\tilde{T}^u$

$$\tilde{T}^u\left[\psi\right] = \sum_{p=-\infty}^{\infty} (i\gamma_p^u)\hat{\psi}_p e^{i\alpha_p x} = (i\gamma_D)\psi.$$

Due to the linear growth of α_p and γ_p^u, it is easy to show that each maps the Sobolev Space H^{s+1} to H^s. We recall our earlier definition of the DNO in §1.8 as

$$\tilde{T}^u : \tilde{u}(x,a) \to \left(\partial_z \tilde{u}\right)(x,a), \tag{2.6}$$

where, above $z = a$,

$$\tilde{u}(x,z) = \sum_{p=-\infty}^{\infty} \left(a_p e^{i\gamma_p^u a}\right) e^{i\alpha_p x + i\gamma_p^u (z-a)} = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\alpha_p x + i\gamma_p^u (z-a)}, \tag{2.7}$$

and

$$\tilde{u}(x,a) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\alpha_p x} = \psi(x), \quad \partial_z \tilde{u}(x,a) = \sum_{p=-\infty}^{\infty} (i\gamma_p^u)\hat{\psi}_p e^{i\alpha_p x}. \tag{2.8}$$

We now define

$$\alpha_p = \alpha + \left(\frac{2\pi}{d}\right) p := \alpha + \tilde{p}, \quad \tilde{p} := \left(\frac{2\pi}{d}\right) p, \tag{2.9}$$

so that

$$\psi(x) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i(\alpha+\tilde{p})x} = e^{i\alpha x} \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x} = e^{i\alpha x}\zeta^u(x),$$

where

$$\zeta^u(x) := \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x}.$$

Writing

$$\zeta^u(x+d) = e^{-i\alpha(x+d)}\psi(x+d) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}(x+d)} = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x}$$

shows that $\zeta^u(x+d) = \zeta^u(x)$ and therefore ζ^u is d–periodic. As in §2.2, we perform the phase extraction

$$u(x,z) := e^{-i\alpha x}\tilde{u}(x,z),$$

where, above $z = a$, equation (2.7) delivers

$$u(x,z) = e^{-i\alpha x}\tilde{u}(x,z) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x + i\gamma_p^u(z-a)}.$$

and, by equation (2.8),

$$u(x,a) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x}, \quad \partial_z u(x,a) = \sum_{p=-\infty}^{\infty} (i\gamma_p^u)\hat{\psi}_p e^{i\tilde{p}x}. \tag{2.10}$$

We then define the upper layer DNO without phase as

$$T^u : u(x,a) \to (\partial_z u)(x,a), \tag{2.11}$$

so that by equation (2.10)

$$T^u\left[u(x,a)\right] = T^u\left[\sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x}\right] = \sum_{p=-\infty}^{\infty} (i\gamma_p^u)\hat{\psi}_p e^{i\tilde{p}x}. \tag{2.12}$$

With this, we see that equations (2.10) and (2.12) satisfy the Transparent Boundary Condition

$$\partial_z u(x,a) - T^u[u(x,a)] = 0. \tag{2.13}$$

2.4 Boundary Perturbation

We now apply the change of variables from Appendix C to (2.5) and start by focusing on

$$\Delta u + 2i\alpha\partial_x u + (\gamma^u)^2 u = 0. \tag{2.14}$$

The transformation rules produce the following transformation in the upper field

$$x' = x, \quad z' = a\left(\frac{z-g(x)}{a-g(x)}\right).$$

This transformation maps the perturbed geometry $S_{g,a}$ to the separable geometry $S_{0,a}$. We will later show that the transformation enables not only a rigorous proof of analyticity and convergence, but also provides a stable and highly accurate numerical scheme.

We then invert the change of variables to find

$$x = x', \quad z = \left(\frac{a - g(x')}{a}\right) z' + g(x'),$$

which we use to define the transformed field as

$$u(x', z') := u'\left(x', \left(\frac{a - g(x')}{a}\right) z' + g(x')\right).$$

In Appendix C we discuss the effects of this change of variables on the Helmholtz equation, its derivatives, and the associated boundary conditions. In the upper layer we have a domain $S_{L,U}$, (C.1), where

$$\overline{\ell} = 0, \quad \ell(x) = g(x), \quad \overline{u} = a, \quad u(x) \equiv 0, \quad \overline{h} = \overline{u} - \overline{\ell} = a.$$

Therefore

$$C(x) = 1 + \frac{0 - g(x)}{a} = 1 - \frac{g(x)}{a}, \quad D(x) = \frac{ag(x) - 0^2}{a} = g(x),$$

and

$$E = (\partial_x g)\left(\frac{a - z'}{a}\right), \quad Z_U = \frac{a - z'}{a}.$$

(We omit Z_L since $u \equiv 0$). In Appendix C we show that the change of variables changes the derivatives to

$$C\partial_x = C\partial_{x'} - E\partial_{z'}, \quad C\partial_z = \partial_{z'},$$

and the upper layer Helmholtz equation becomes

$$0 = \mathrm{div}'[A\nabla' u'] + B \cdot \nabla' u' + 2C^2 i\alpha \partial_{x'} u' + C^2(\gamma^{u'})^2 u',$$

where, for $S = C^2$,

$$A = \begin{pmatrix} S & -EC \\ -EC & 1 + E^2 \end{pmatrix}, \qquad B = (\partial_{x'} C)\begin{pmatrix} -C \\ E \end{pmatrix}.$$

For simplicity we drop the primed variables to realize

$$0 = \mathrm{div}[A\nabla u] + B \cdot \nabla u + 2Si\alpha \partial_x u + S(\gamma^u)^2 u,$$

and take a boundary perturbation approach by setting

$$g(x) = \varepsilon f(x),\ \varepsilon \in \mathbb{R},\ \varepsilon \ll 1, \tag{2.15}$$

where, by following Appendix C, discover

$$\begin{aligned} A &= A(\varepsilon) = A_0 + A_1\varepsilon + A_2\varepsilon^2, \\ B &= B(\varepsilon) = B_1\varepsilon + B_2\varepsilon^2, \\ S &= S(\varepsilon) = S_0 + S_1\varepsilon + S_2\varepsilon^2. \end{aligned}$$

Since $a = \overline{h}$ and $\ell(x) = \varepsilon f(x)$, we find

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.16a}$$

$$A_1 = \begin{pmatrix} A_1^{xx} & A_1^{xz} \\ A_1^{zx} & A_1^{zz} \end{pmatrix} = \frac{1}{a}\begin{pmatrix} -2f & -(a-z)(\partial_x f) \\ -(a-z)(\partial_x f) & 0 \end{pmatrix}, \tag{2.16b}$$

$$A_2 = \begin{pmatrix} A_2^{xx} & A_2^{xz} \\ A_2^{zx} & A_2^{zz} \end{pmatrix} = \frac{1}{a^2}\begin{pmatrix} f^2 & (a-z)f(\partial_x f) \\ (a-z)f(\partial_x f) & (a-z)^2(\partial_x f)^2 \end{pmatrix}, \tag{2.16c}$$

and

$$B_1 = \begin{pmatrix} B_1^x \\ B_1^z \end{pmatrix} = \frac{1}{a}\begin{pmatrix} \partial_x f \\ 0 \end{pmatrix}, \tag{2.16d}$$

$$B_2 = \begin{pmatrix} B_2^x \\ B_2^z \end{pmatrix} = \frac{1}{a^2}\begin{pmatrix} -f(\partial_x f) \\ -(a-z)(\partial_x f)^2 \end{pmatrix}, \tag{2.16e}$$

and

$$S_0 = 1, \quad S_1 = -\frac{2}{a}f, \quad S_2 = \frac{1}{a^2}f^2. \tag{2.16f}$$

So (2.14) becomes

$$\Delta u + 2i\alpha\partial_x u + \gamma^2 u = F(x, z; f, u, \alpha, \gamma), \quad 0 < z < a, \tag{2.17}$$

where

$$\begin{aligned} F(x, z; f, u, \alpha, \gamma) = &-\mathrm{div}[A_1\nabla u] - \mathrm{div}[A_2\nabla u] - B_1\nabla u - B_2\nabla u \\ &- 2S_1 i\alpha\partial_x u - S_1\gamma^2 u - 2S_2 i\alpha\partial_x u - S_2\gamma^2 u. \end{aligned} \tag{2.18}$$

By (2.5d) the Transparent Boundary Condition for our governing equations without phase is

$$\partial_z\left[u(x, a)\right] - T^u[u(x, a)] = 0, \quad \text{at } z = a. \tag{2.19}$$

For this boundary condition we begin with the top boundary and recall that such boundaries are flat for simplicity, i.e., $u \equiv 0$. Therefore, we can multiply (2.19) by $C = C(x)$ to realize

$$C\partial_z \left[u(x,a)\right] - CT^u[u(x,a)] = 0.$$

So by the transformation rules in Appendix C for ∂_z and ∂_x (which induces the rule $T^u \to T^{u'}$ and $u \to u'$) with $u \equiv 0$ we find

$$\partial_{z'} \left[u'(x',a)\right] - (1 - \ell(x')/\overline{h})T^{u'}[u'(x',a)] = 0.$$

We rearrange to form

$$\partial_{z'} \left[u'(x',a)\right] - T^{u'}[u'(x',a)] = P(x';g,u'),$$

where

$$P(x';g,u') = -\frac{1}{a}g(x')T^{u'}\left[u'(x',a)\right].$$

We then drop the primed variables and write the boundary condition as

$$\partial_z \left[u(x,a)\right] - T^u[u(x,a)] = P(x;g,u).$$

These changes transform the governing equations without phase in (2.5) to

$$\Delta u + 2i\alpha\partial_x u + (\gamma^u)^2 u = F\left(x,z;f,u,\alpha,\gamma^u\right), \qquad 0 < z < a, \tag{2.20a}$$

$$u(x,0) = \zeta^u(x), \qquad \text{at } z = 0, \tag{2.20b}$$

$$u(x+d,z) = u(x,z), \tag{2.20c}$$

$$\partial_z \left[u(x,a)\right] - T^u[u(x,a)] = P(x;g,u), \qquad \text{at } z = a. \tag{2.20d}$$

2.5 Frequency Perturbation

We now perform an Asymptotic Waveform Evaluation by writing the illumination frequency as

$$\omega = (1+\delta)\underline{\omega} = \underline{\omega} + \delta\underline{\omega},\ \delta \in \mathbb{R},\ \delta \ll 1. \tag{2.21}$$

With this we see that

$$k^u = \omega/c^u = (1+\delta)\underline{\omega}/c^u =: (1+\delta)\underline{k}^u = \underline{k}^u + \delta\underline{k}^u, \tag{2.22a}$$

$$\alpha = k^u \sin(\theta) = (1+\delta)\underline{k}^u \sin(\theta) =: (1+\delta)\underline{\alpha} = \underline{\alpha} + \delta\underline{\alpha}, \tag{2.22b}$$

$$\gamma^u = k^u \cos(\theta) = (1+\delta)\underline{k}^u \cos(\theta) =: (1+\delta)\underline{\gamma}^u = \underline{\gamma}^u + \delta\underline{\gamma}^u. \tag{2.22c}$$

We can relate the constants in the underscore variables by the relationship

$$\underline{\alpha}^2 + (\underline{\gamma}^u)^2 = (\underline{k}^u)^2. \tag{2.23}$$

Then, since $(\gamma^u)^2 = (\delta+1)^2\left((\underline{k}^u)^2 - \underline{\alpha}^2\right) = (\delta+1)^2(\underline{\gamma}^u)^2$, (2.18) becomes

$$\begin{aligned} F\left(x,z;f,u,\underline{\alpha},\underline{\gamma}^u\right) = &-\mathrm{div}[A_1\nabla u] - \mathrm{div}[A_2\nabla u] - B_1\nabla u - B_2\nabla u \\ &- 2S_1 i\underline{\alpha}\partial_x u - 2S_1 i\underline{\alpha}\delta\partial_x u - S_1\delta^2(\underline{\gamma}^u)^2 u - 2S_1\delta(\underline{\gamma}^u)^2 u - S_1(\underline{\gamma}^u)^2 u \\ &- 2S_2 i\underline{\alpha}\partial_x u - 2S_2 i\underline{\alpha}\delta\partial_x u - S_2\delta^2(\underline{\gamma}^u)^2 u - 2S_2\delta(\underline{\gamma}^u)^2 u - S_2(\underline{\gamma}^u)^2 u. \end{aligned}$$

Also, the left-hand side of (2.20a) becomes

$$\Delta u + 2i\underline{\alpha}\partial_x u + 2i\underline{\alpha}\delta\partial_x u + \delta^2(\underline{\gamma}^u)^2 u + 2\delta(\underline{\gamma}^u)^2 u + (\underline{\gamma}^u)^2 u,$$

and the boundary condition for (2.20a) becomes

$$\Delta u + 2i\underline{\alpha}\partial_x u + (\underline{\gamma}^u)^2 u = \tilde{F}\left(x,z;f,u,\underline{\alpha},\underline{\gamma}^u\right), \quad 0 < z < a. \tag{2.24}$$

We move all terms with δ to the right–hand side to form

$$\begin{aligned} \tilde{F}\left(x,z;f,u,\underline{\alpha},\underline{\gamma}^u\right) = &-\mathrm{div}[A_1\nabla u] - \mathrm{div}[A_2\nabla u] - B_1\nabla u - B_2\nabla u \\ &- 2i\underline{\alpha}\delta\partial_x u - \delta^2(\underline{\gamma}^u)^2 u - 2\delta(\underline{\gamma}^u)^2 u \\ &- 2S_1 i\underline{\alpha}\partial_x u - 2S_1 i\underline{\alpha}\delta\partial_x u - S_1\delta^2(\underline{\gamma}^u)^2 u - 2S_1\delta(\underline{\gamma}^u)^2 u - S_1(\underline{\gamma}^u)^2 u \\ &- 2S_2 i\underline{\alpha}\partial_x u - 2S_2 i\underline{\alpha}\delta\partial_x u - S_2\delta^2(\underline{\gamma}^u)^2 u - 2S_2\delta(\underline{\gamma}^u)^2 u - S_2(\underline{\gamma}^u)^2 u. \end{aligned}$$

The boundary condition (2.20d) becomes

$$\partial_z\left[u(x,a)\right] - T_0^u[u(x,a)] = \tilde{P}(x;f,u),$$

where $T_0^u = i\underline{\gamma}_D^u$ corresponds to the case where $\delta = 0$ and

$$\tilde{P}(x;f,u) = -\frac{1}{a}(\varepsilon f(x))T^u\left[u(x,a)\right] + \left(T^u - T_0^u\right)\left[u(x,a)\right].$$

Our governing equations are now

$$\Delta u + 2i\underline{\alpha}\partial_x u + (\underline{\gamma}^u)^2 u = \tilde{F}\left(x,z;g,u,\underline{\alpha},\underline{\gamma}^u\right), \qquad 0 < z < a, \tag{2.25a}$$

$$u(x,0) = \zeta^u(x), \qquad \text{at } z = 0, \tag{2.25b}$$

$$u(x+d,z) = u(x,z), \tag{2.25c}$$

$$\partial_z\left[u(x,a)\right] - T_0^u[u(x,a)] = \tilde{P}(x;f,u), \qquad \text{at } z = a. \tag{2.25d}$$

2.6 Transformed Field Expansions

In the previous two sections we made two smallness assumptions:

[1] Boundary Perturbation: $g(x) = \varepsilon f(x)$, $\varepsilon \in \mathbb{R}$, $\varepsilon \ll 1$,

[2] Frequency Perturbation: $\omega = (1+\delta)\underline{\omega} = \underline{\omega} + \delta\underline{\omega}$, $\delta \in \mathbb{R}$, $\delta \ll 1$.

We now apply both of these assumptions and seek solutions of the form

$$u = u(x, z; \varepsilon, \delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} u_{n,m}(x, z)\varepsilon^n\delta^m. \tag{2.26}$$

We will later show that these solutions are strongly convergent in Theorems 2.8.4 and 2.9.2. Inserting these into (2.25) produces the Transformed Field Expansions (TFE) recursions

$$\Delta u_{n,m} + 2i\underline{\alpha}\partial_x u_{n,m} + (\underline{\gamma}^u)^2 u_{n,m} = \tilde{F}_{n,m}\left(x, z; f, u, \underline{\alpha}, \underline{\gamma}^u\right), \qquad 0 < z < a, \tag{2.27a}$$

$$u_{n,m}(x, 0) = \zeta^u_{n,m}(x), \qquad \text{at } z = 0, \tag{2.27b}$$

$$u_{n,m}(x + d, z) = u_{n,m}(x, z), \tag{2.27c}$$

$$\partial_z\left[u_{n,m}(x, a)\right] - T^u_0[u_{n,m}(x, a)] = \tilde{P}_{n,m}(x; f, u), \qquad \text{at } z = a, \tag{2.27d}$$

where

$$\begin{aligned}\tilde{F}_{n,m}\left(x, z; f, u, \underline{\alpha}, \underline{\gamma}^u\right) = &-\mathrm{div}[A_1\nabla u_{n-1,m}] - \mathrm{div}[A_2\nabla u_{n-2,m}] - B_1\nabla u_{n-1,m}\\ &- B_2\nabla u_{n-2,m} - 2i\underline{\alpha}\partial_x u_{n,m-1} - (\underline{\gamma}^u)^2 u_{n,m-2}\\ &- 2(\underline{\gamma}^u)^2 u_{n,m-1} - 2S_1 i\underline{\alpha}\partial_x u_{n-1,m} - 2S_1 i\underline{\alpha}\partial_x u_{n-1,m-1}\\ &- S_1(\underline{\gamma}^u)^2 u_{n-1,m-2} - 2S_1(\underline{\gamma}^u)^2 u_{n-1,m-1} - S_1(\underline{\gamma}^u)^2 u_{n-1,m}\\ &- 2S_2 i\underline{\alpha}\partial_x u_{n-2,m} - 2S_2 i\underline{\alpha}\partial_x u_{n-2,m-1} - S_2(\underline{\gamma}^u)^2 u_{n-2,m-2}\\ &- 2S_2(\underline{\gamma}^u)^2 u_{n-2,m-1} - S_2(\underline{\gamma}^u)^2 u_{n-2,m},\end{aligned} \tag{2.28}$$

and

$$\tilde{P}_{n,m}(x; f, u) = -\frac{f}{a}\sum_{r=0}^{m} T^u_{m-r}[u_{n-1,r}(x, a)] + \sum_{r=0}^{m-1} T^u_{m-r}[u_{n,r}(x, a)]. \tag{2.29}$$

This is a method for computing the transformed corrections to the scattered field, $u_{n,m}$, with respect to both interfacial and frequency deformations. A major advantage of the TFE recursions is that (2.27) never takes derivatives of $u_{n,m}$ higher than second order. This allows us to take advantage of the classical theory of elliptic boundary value problems where we will carefully show that our solutions, u, are jointly analytic in the appropriate Sobolev space.

2.7 Sobolev Spaces and Elliptic Theory

We summarize the characterization of the Sobolev spaces $H^s = W^{s,2}$ applied in laterally d–periodic functions relevant to scattering problems of interest to us. We know that any d–periodic L^2 function

$$\mu(x+d) = \mu(x),$$

can be expressed as

$$\mu(x) = \sum_{p=-\infty}^{\infty} \hat{\mu}_p e^{i\tilde{p}x},$$

where

$$\tilde{p} := \left(\frac{2\pi}{d}\right) p, \quad \hat{\mu}_p = \frac{1}{d}\int_0^d \mu(x) e^{-i\tilde{p}x}\,\mathrm{d}x.$$

We then define our x–periodic norms. For any L^2 function $\mu = \mu(x)$, we recall the classical Sobolev norm for any real $s \geq 0$:

$$\|\mu\|^2_{H^s_x} := \sum_{p=-\infty}^{\infty} \langle \tilde{p}\rangle^{2s} |\hat{\mu}_p|^2, \quad \langle \tilde{p}\rangle^2 := 1 + |\tilde{p}|^2.$$

For the L^2 function $u = u(x,z)$ we require the classical Sobolev norm for any integer $s \geq 0$ and $a > 0$

$$\|u\|^2_{H^s_{x,z}} := \sum_{\ell=0}^{s} \sum_{p=-\infty}^{\infty} \langle \tilde{p}\rangle^{2(s-\ell)} \int_0^a |\hat{u}_p(z)|^2\,\mathrm{d}z = \sum_{\ell=0}^{s} \sum_{p=-\infty}^{\infty} \langle \tilde{p}\rangle^{2(s-\ell)} \|\hat{u}_p\|^2_{L^2([0,a])}.$$

With these norms, we define the following function spaces. First, for real $s \geq 0$,

$$H^s\big([0,d]\big) := \left\{\mu(x) \in L^2\big([0,d]\big) \;\middle|\; \|\mu\|_{H^s_x} < \infty\right\}.$$

Also, for any integer $s \geq 0$,

$$H^s\big([0,d]\times[0,a]\big) := \left\{u(x) \in L^2\big([0,d]\times[0,a]\big) \;\middle|\; \|u\|_{H^s_{x,z}} < \infty\right\}.$$

With these we can now establish the following three properties based on classical elliptic theory. The first property is the "Algebra Property" of Sobolev spaces which allows us to estimate products of functions in our function classes. The second property is a theorem which gives a rigorous statement of the "Elliptic Estimate." The final property provides a method of bounding translated elements in our function spaces.

Lemma 2.7.1. *Given an integer $s \geq 0$ and any $\sigma > 0$, there exists a constant $\mathcal{M} = \mathcal{M}(s)$ such that if $f \in C^s([0,d]), u \in H^s([0,d]\times[0,a])$ then*

$$\|fu\|_{H^s} \leq \mathcal{M}|f|_{C^s}\|u\|_{H^s}, \tag{2.30}$$

and if $\tilde{f} \in C^{s+1/2+\sigma}([0,d]), \tilde{u} \in H^{s+1/2}([0,d])$ *then there exists a constant* $\tilde{\mathcal{M}} = \tilde{\mathcal{M}}(s)$ *such that*

$$\|\tilde{f}\tilde{u}\|_{H^{s+1/2}} \leq \tilde{\mathcal{M}}|\tilde{f}|_{C^{s+1/2+\sigma}}\|\tilde{u}\|_{H^{s+1/2}}. \tag{2.31}$$

Theorem 2.7.2. *Given an integer* $s \geq 0$*, if* $F \in H^s([0,d]) \times [0,a])$*,* $\zeta^u \in H^{s+3/2}([0,d])$*,* $P \in H^{s+1/2}([0,d])$*, then there exists a unique solution* $u \in H^{s+2}([0,d]) \times [0,a])$ *of*

$$\Delta u(x,z) + 2i\underline{\alpha}\partial_x u(x,z) + (\underline{\gamma}^u)^2 u(x,z) = F(x,z), \qquad 0 < z < a, \tag{2.32a}$$

$$u(x,0) = \zeta^u(x,0), \qquad \text{at } z = 0, \tag{2.32b}$$

$$u(x+d,z) = u(x,z), \tag{2.32c}$$

$$\partial_z u(x,a) - T_0^u[u(x,a)] = P(x), \qquad \text{at } z = a, \tag{2.32d}$$

satisfying

$$\|u\|_{H^{s+2}} \leq C_e\{\|F\|_{H^s} + \|\zeta^u\|_{H^{s+3/2}} + \|P\|_{H^{s+1/2}}\}, \tag{2.33}$$

for some constant $C_e = C_e(s) > 0$*.*

Lemma 2.7.3. *Given an integer* $s \geq 0$*, if* $F \in H^s([0,d]) \times [0,a])$*, then* $(a-z)F \in H^s([0,d]) \times [0,a])$ *and there exists a positive constant* $Z_a = Z_a(s)$ *such that*

$$\|(a-z)F\|_{H^s} \leq Z_a\|F\|_{H^s}.$$

The proof of these three properties is established in Appendix B.

2.8 Analyticity of the Boundary Perturbation

Before proceeding to the analyticity of the upper field, u, we present an analyticity estimate for the Dirichlet data

$$\zeta^u(x;\varepsilon) = \sum_{n=0}^{\infty} \zeta_{n,0}^u(x)\varepsilon^n.$$

The following three lemmas will be invaluable in our subsequent analysis.

Lemma 2.8.1. *Given any integer* $s \geq 0$*, if* $u \in H^s([0,d])$ *then*

$$\|\partial_x u\|_{H^s} \leq \|u\|_{H^{s+1}}.$$

Proof. [Lemma 2.8.1] By the definition of our Sobolev norms,

$$\|\partial_x u\|_{H^s}^2 = \sum_{p=-\infty}^{\infty} \langle\tilde{p}\rangle^{2s}|\widehat{\partial_x u}_p|^2 = \sum_{p=-\infty}^{\infty} \langle\tilde{p}\rangle^{2s}|(i\tilde{p})\hat{u}_p|^2 \leq \sum_{p=-\infty}^{\infty} \langle\tilde{p}\rangle^{2s+2}|\hat{u}_p|^2 = \|u\|_{H^{s+1}}^2.$$

□

Lemma 2.8.2. *Let T_0^q, $q \in \{u, w\}$, be the DNO defined by $(i\underline{\gamma}_D^q)$ and $s \geq 0$ a positive integer. Then for $\psi \in H^{s+1}([0,d])$, we have*

$$\|T_0^q \psi\|_{H^s} \leq C_{([0,d])} \|\psi\|_{H^{s+1}},$$

for some $C_{([0,d])} > 0$.

Proof. [Lemma 2.8.2] Let $T_0^q = (i\underline{\gamma}_D^q)$ where $\psi \in H^{s+1}([0,d])$. By (B.15), (2.9), and the definition of our Sobolev norms,

$$\begin{aligned}
\|T_0^q \psi\|_{H^s}^2 &= \sum_{p=-\infty}^{\infty} \left|(i\underline{\gamma}_p^q)\hat{\psi}_p\right|^2 \langle \tilde{p} \rangle^{2s} \\
&= \sum_{p \in \underline{\mathcal{U}}^q} \left|\sqrt{(\underline{k}^q)^2 - \underline{\alpha}_p^2}\hat{\psi}_p\right|^2 \langle \tilde{p} \rangle^{2s} + \sum_{p \notin \underline{\mathcal{U}}^q} \left|\sqrt{\underline{\alpha}_p^2 - (\underline{k}^q)^2}\hat{\psi}_p\right|^2 \langle \tilde{p} \rangle^{2s} \\
&\leq \sum_{p \in \underline{\mathcal{U}}^q} C|\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2s} + \sum_{p \notin \underline{\mathcal{U}}^q} \left||\underline{\alpha}_p|\sqrt{1 - (\underline{k}^q)^2/\underline{\alpha}_p^2}\hat{\psi}_p\right|^2 \langle \tilde{p} \rangle^{2s}, \ \mathrm{C} = \max_{p \in \underline{\mathcal{U}}^q} \left[(\underline{k}^q)^2 - \underline{\alpha}_p^2\right] \\
&\leq \sum_{p \in \underline{\mathcal{U}}^q} C|\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2s} + \sum_{p \notin \underline{\mathcal{U}}^q} \tilde{C}|\underline{\alpha}_p|^2 |\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2s}, \quad \tilde{C} = \max_{p \notin \underline{\mathcal{U}}^q} \left[1 - (\underline{k}^q)^2/\underline{\alpha}_p^2\right] \\
&\leq \sum_{p=-\infty}^{\infty} \max\left\{C, 2\alpha^2 \tilde{C}\right\} |\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2s} + \sum_{p \notin \underline{\mathcal{U}}^q} 2\tilde{p}^2 \tilde{C}|\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2s} \\
&\leq \sum_{p=-\infty}^{\infty} \tilde{\tilde{C}} \langle \tilde{p} \rangle^2 |\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2s}, \quad \tilde{\tilde{C}} = \max\left\{C, 2\alpha^2 \tilde{C}, 2\tilde{C}\right\} \\
&= \sum_{p=-\infty}^{\infty} \tilde{\tilde{C}} |\hat{\psi}_p|^2 \langle \tilde{p} \rangle^{2(s+1)} \\
&= \tilde{\tilde{C}} \|\psi\|_{H^{s+1}}^2.
\end{aligned}$$

□

Lemma 2.8.3. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ then*

$$\|\zeta_{n,0}^u\|_{H^{s+3/2}} \leq K_\zeta B_\zeta^n \tag{2.34}$$

for constants $K_\zeta, B_\zeta > 0$.

Proof. [Lemma 2.8.3] We work by induction and begin with $n = 0$ where we choose

$$K_\zeta := \|\zeta_{0,0}^u\|_{H^{s+3/2}}.$$

We now assume the estimate (2.34) for all $n < \overline{n}$ and note that

$$\zeta_{\overline{n},0}^u = (-i\gamma^u)\left(\frac{f}{\overline{n}}\right)\zeta_{\overline{n}-1,0}^u.$$

From this and $\overline{n} \geq 1$ we find the bound

$$\begin{aligned}\|\zeta^u_{\overline{n},0}\|_{H^{s+3/2}} &\leq |\gamma^u|\ \mathcal{M}\ |f|_{C^{s+3/2+\sigma}} \|\zeta^u_{\overline{n}-1,0}\|_{H^{s+3/2}} \\ &\leq |\gamma^u|\ \mathcal{M}\ |f|_{C^{s+2}} K_\zeta B_\zeta^{\overline{n}-1},\end{aligned}$$

and we are done provided

$$B_\zeta > |\gamma^u|\ \mathcal{M}\ |f|_{C^{s+2}}. \qquad \square$$

We can now state our desired result for the analyticity of the transformed field $u = u(x, z; \varepsilon)$ with respect to the single perturbation parameter ε.

Theorem 2.8.4. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0, d])$ and $\zeta^u_{n,0} \in H^{s+3/2}([0, d])$ such that*

$$\|\zeta^u_{n,0}\|_{H^{s+3/2}} \leq K_\zeta B^n_\zeta, \tag{2.35}$$

for constants $K_\zeta, B_\zeta > 0$, then $u_{n,0} \in H^{s+2}([0, d] \times [0, a])$ and

$$\|u_{n,0}\|_{H^{s+2}} \leq K B^n, \tag{2.36}$$

for constants $K, B > 0$.

To establish this result we work by induction. They key estimate is encapsulated in the following lemma.

Lemma 2.8.5. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0, d])$ and*

$$\|u_{n,0}\|_{H^{s+2}} \leq K B^n, \quad \forall n < \overline{n}, \tag{2.37}$$

for constants $K, B > 0$, then there exists a constant $\overline{C} > 0$ such that

$$\max\left\{\|\tilde{F}_{\overline{n},0}\|_{H^s}, \|\tilde{P}_{\overline{n},0}\|_{H^{s+1/2}}\right\} \leq K\overline{C}\Big\{|f|_{C^{s+2}} B^{\overline{n}-1} + |f|^2_{C^{s+2}} B^{\overline{n}-2}\Big\}. \tag{2.38}$$

Proof. [Lemma 2.8.5] We begin with $\tilde{F}_{\overline{n},0}$ and recall from (2.28) that

$$\begin{aligned}\tilde{F}_{\overline{n},0}\left(x, z; f, u, \underline{\alpha}, \underline{\gamma}^u\right) = &-\mathrm{div}[A_1 \nabla u_{\overline{n}-1,0}] - \mathrm{div}[A_2 \nabla u_{\overline{n}-2,0}] - B_1 \nabla u_{\overline{n}-1,0} \\ &- B_2 \nabla u_{\overline{n}-2,0} - 2S_1 i\underline{\alpha}\partial_x u_{\overline{n}-1,0} - S_1(\underline{\gamma}^u)^2 u_{\overline{n}-1,0} \\ &- 2S_2 i\underline{\alpha}\partial_x u_{\overline{n}-2,0} - S_2(\underline{\gamma}^u)^2 u_{\overline{n}-2,0}.\end{aligned} \tag{2.39}$$

Then from (2.16) we have

$$\begin{aligned}\|\tilde{F}_{\overline{n},0}\|_{H^s}^2 &\le \|A_1^{xx}\partial_x u_{\overline{n}-1,0}\|_{H^{s+1}}^2 + \|A_1^{xz}\partial_z u_{\overline{n}-1,0}\|_{H^{s+1}}^2 + \|A_1^{zx}\partial_x u_{\overline{n}-1,0}\|_{H^{s+1}}^2 \\ &+ \|A_1^{zz}\partial_z u_{\overline{n}-1,0}\|_{H^{s+1}}^2 + \|A_2^{xx}\partial_x u_{\overline{n}-2,0}\|_{H^{s+1}}^2 + \|A_2^{xz}\partial_z u_{\overline{n}-2,0}\|_{H^{s+1}}^2 \\ &+ \|A_2^{zx}\partial_x u_{\overline{n}-2,0}\|_{H^{s+1}}^2 + \|A_2^{zz}\partial_z u_{\overline{n}-2,0}\|_{H^{s+1}}^2 + \|B_1^{x}\partial_x u_{\overline{n}-1,0}\|_{H^{s}}^2 \\ &+ \|B_1^{z}\partial_z u_{\overline{n}-1,0}\|_{H^{s}}^2 + \|B_2^{x}\partial_x u_{\overline{n}-2,0}\|_{H^{s}}^2 + \|B_2^{z}\partial_z u_{\overline{n}-2,0}\|_{H^{s}}^2 \\ &+ \|2S_1 i\underline{\alpha}\partial_x u_{\overline{n}-1,0}\|_{H^s}^2 + \|S_1(\underline{\gamma}^u)^2 u_{\overline{n}-1,0}\|_{H^s}^2 + \|2S_2 i\underline{\alpha}\partial_x u_{\overline{n}-2,0}\|_{H^s}^2 \\ &+ \|S_2(\underline{\gamma}^u)^2 u_{\overline{n}-2,0}\|_{H^s}^2.\end{aligned}$$

We now estimate each of these and apply Lemmas 2.7.1, 2.7.3, and 2.8.1. We begin with

$$\begin{aligned}\|A_1^{xx}\partial_x u_{\overline{n}-1,0}\|_{H^{s+1}} &= \|-(2/a)f\partial_x u_{\overline{n}-1,0}\|_{H^{s+1}} \\ &\le (2/a)\mathcal{M}|f|_{C^{s+1}}\|u_{\overline{n}-1,0}\|_{H^{s+2}} \\ &\le (2/a)\mathcal{M}|f|_{C^{s+1}}KB^{\overline{n}-1},\end{aligned}$$

and in a similar fashion

$$\begin{aligned}\|A_1^{xz}\partial_z u_{\overline{n}-1,0}\|_{H^{s+1}} &= \|-((a-z)/a)(\partial_x f)\partial_z u_{\overline{n}-1,0}\|_{H^{s+1}} \\ &\le (Z_a/a)\mathcal{M}|\partial_x f|_{C^{s+1}}\|u_{\overline{n}-1,0}\|_{H^{s+2}} \\ &\le (Z_a/a)\mathcal{M}|f|_{C^{s+2}}KB^{\overline{n}-1}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_1^{zx}\partial_x u_{\overline{n}-1,0}\|_{H^{s+1}} &= \|-((a-z)/a)(\partial_x f)\partial_x u_{\overline{n}-1,0}\|_{H^{s+1}} \\ &\le (Z_a/a)\mathcal{M}|\partial_x f|_{C^{s+1}}\|u_{\overline{n}-1,0}\|_{H^{s+2}} \\ &\le (Z_a/a)\mathcal{M}|f|_{C^{s+2}}KB^{\overline{n}-1},\end{aligned}$$

and we recall that $A_1^{zz} \equiv 0$. Moving to the second order

$$\begin{aligned}\|A_2^{xx}\partial_x u_{\overline{n}-2,0}\|_{H^{s+1}} &= \|(1/a^2)f^2\partial_x u_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\le (1/a^2)\mathcal{M}^2|f|_{C^{s+1}}^2\|u_{\overline{n}-2,0}\|_{H^{s+2}} \\ &\le (1/a^2)\mathcal{M}^2|f|_{C^{s+1}}^2KB^{\overline{n}-2}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_2^{xz}\partial_z u_{\overline{n}-2,0}\|_{H^{s+1}} &= \|((a-z)/a^2)f(\partial_x f)\partial_x u_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\le (Z_a/a^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|u_{\overline{n}-2,0}\|_{H^{s+2}} \\ &\le (Z_a/a^2)\mathcal{M}^2|f|_{C^{s+2}}^2KB^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zx}\partial_x u_{\overline{n}-2,0}\|_{H^{s+1}} &= \|((a-z)/a^2)f(\partial_x f)\partial_z u_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|u_{\overline{n}-2,0}\|_{H^{s+2}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zz}\partial_z u_{\overline{n}-2,0}\|_{H^{s+1}} &= \|((a-z)^2/a^2)(\partial_x f)^2\partial_z u_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (Z_a^2/a^2)\mathcal{M}^2|\partial_x f|^2_{C^{s+1}}\|u_{\overline{n}-2,0}\|_{H^{s+2}}\\ &\leq (Z_a^2/a^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{\overline{n}-2}.\end{aligned}$$

Next for the B_1 terms

$$\begin{aligned}\|B_1^x\partial_x u_{\overline{n}-1,0}\|_{H^s} &= \|(1/a)(\partial_x f)\partial_x u_{\overline{n}-1,0}\|_{H^s}\\ &\leq (1/a)\mathcal{M}|\partial_x f|_{C^s}\|u_{\overline{n}-1,0}\|_{H^{s+1}}\\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1}}KB^{\overline{n}-1},\end{aligned}$$

and $B_1^z \equiv 0$. Moving to the second order

$$\begin{aligned}\|B_2^x\partial_x u_{\overline{n}-2,0}\|_{H^s} &= \|(-1/a^2)f(\partial_x f)\partial_x u_{\overline{n}-2,0}\|_{H^s}\\ &\leq (1/a^2)\mathcal{M}^2|f|_{C^s}|\partial_x f|_{C^s}\|u_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (1/a^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|B_2^z\partial_z u_{\overline{n}-2,0}\|_{H^s} &= \|(-1/a^2)(a-z)(\partial_x f)^2\partial_z u_{\overline{n}-2,0}\|_{H^s}\\ &\leq (Z_a/a^2)\mathcal{M}^2|\partial_x f|^2_{C^s}\|u_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{\overline{n}-2}.\end{aligned}$$

To address the S_0, S_1, S_2 terms we have

$$\begin{aligned}\|2S_1 i\underline{\alpha}\partial_x u_{\overline{n}-1,0}\|_{H^s} &= \|(-4/a)i\underline{\alpha}f\partial_x u_{\overline{n}-1,0}\|_{H^s}\\ &\leq (4/a)\underline{\alpha}\mathcal{M}|f|_{C^s}\|u_{\overline{n}-1,0}\|_{H^{s+1}}\\ &\leq (4/a)\underline{\alpha}\mathcal{M}|f|_{C^s}KB^{\overline{n}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_1(\underline{\gamma}^u)^2 u_{\overline{n}-1,0}\|_{H^s} &= \|(-2/a)(\underline{\gamma}^u)^2 f u_{\overline{n}-1,0}\|_{H^s} \\ &\leq (2/a)(\underline{\gamma}^u)^2 \mathcal{M}|f|_{C^s}\|u_{\overline{n}-1,0}\|_{H^s} \\ &\leq (2/a)(\underline{\gamma}^u)^2 \mathcal{M}|f|_{C^s} K B^{\overline{n}-1},\end{aligned}$$

and

$$\begin{aligned}\|2S_2 i\underline{\alpha}\partial_x u_{\overline{n}-2,0}\|_{H^s} &= \|(2/a^2) i\underline{\alpha} f^2 \partial_x u_{\overline{n}-2,0}\|_{H^s} \\ &\leq (2/a^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}\|u_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\leq (2/a^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s} K B^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|S_2(\underline{\gamma}^u)^2 u_{\overline{n}-2,0}\|_{H^s} &= \|(1/a^2)(\underline{\gamma}^u)^2 f^2 u_{\overline{n}-2,0}\|_{H^s} \\ &\leq (1/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2|f|^2_{C^s}\|u_{\overline{n}-2,0}\|_{H^s} \\ &\leq (1/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2|f|^2_{C^s} K B^{\overline{n}-2}.\end{aligned}$$

We satisfy the estimate for $\|\tilde{F}_{\overline{n},0}\|_{H^s}$ provided that we choose

$$\overline{C} > \max\left\{\left(\frac{3+2Z_a+4\underline{\alpha}+2(\underline{\gamma}^u)^2}{a}\right)\mathcal{M}, \left(\frac{2+3Z_a+Z_a^2+2\underline{\alpha}+(\underline{\gamma}^u)^2}{a^2}\right)\mathcal{M}^2\right\}.$$

The estimate for $\tilde{P}_{\overline{n},0}$ follows from Lemma 2.8.2

$$\begin{aligned}\|\tilde{P}_{\overline{n},0}\|_{H^{s+1/2}} &= \| -(1/a) f T_0^u\left[u_{\overline{n}-1,0}\right]\|_{H^{s+1/2}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\sigma}}\|T_0^u\left[u_{\overline{n}-1,0}\right]\|_{H^{s+1/2}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\sigma}} C_{T_0^u}\|u_{\overline{n}-1,0}\|_{H^{s+3/2}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\sigma}} C_{T_0^u} K B^{\overline{n}-1},\end{aligned}$$

and provided that

$$\overline{C} > (1/a)\mathcal{M} C_{T_0^u},$$

we are done. □

With this information, we can now prove Theorem 2.8.4.

Proof. [Theorem 2.8.4] We proceed by induction in n. At order $n = m = 0$ (2.27) becomes

$$\Delta u_{0,0} + 2i\underline{\alpha}\partial_x u_{0,0} + (\underline{\gamma}^u)^2 u_{0,0} = 0, \qquad 0 < z < a, \tag{2.40a}$$

$$u_{0,0}(x, g) = \zeta^u_{0,0}(x), \qquad \text{at } z = 0, \tag{2.40b}$$

$$u_{0,0}(x + d, z) = u_{0,0}(x, z), \tag{2.40c}$$

$$\partial_z \left[u_{0,0}(x, a)\right] - T^u_0[u_{0,0}(x, a)] = 0, \qquad \text{at } z = a, \tag{2.40d}$$

and Theorem 2.7.2 guarantees a unique solution such that

$$\|u_{0,0}\|_{H^{s+2}} \leq C_e \|\zeta^u_{0,0}\|_{H^{s+3/2}}.$$

So we choose $K \geq C_e\|\zeta^u_{0,0}\|_{H^{s+3/2}}$. We now assume the estimate (2.36) for all $n < \overline{n}$ and study $u_{\overline{n},0}$. From Theorem 2.7.2 we have a unique solution satisfying

$$\|u_{\overline{n},0}\|_{H^{s+2}} \leq C_e\{\|\tilde{F}_{\overline{n},0}\|_{H^s} + \|\zeta^u_{\overline{n},0}\|_{H^{s+3/2}} + \|\tilde{P}_{\overline{n},0}\|_{H^{s+1/2}}\},$$

and appealing to Lemmas 2.8.3 (with the hypothesis (2.35)) and 2.8.5 we find

$$\|u_{\overline{n},0}\|_{H^{s+2}} \leq C_e \left\{ K_\zeta B^{\overline{n}}_\zeta + 2K\overline{C}\Big[|f|_{C^{s+2}} B^{\overline{n}-1} + |f|^2_{C^{s+2}} B^{\overline{n}-2}\Big]\right\}.$$

We are done provided we choose $K \geq 3C_e K_\zeta$ and

$$B > \max\left\{B_\zeta, 6C_e\overline{C}|f|_{C^{s+2}}, \sqrt{6C_e\overline{C}}|f|_{C^{s+2}}\right\}.$$

□

2.9 Joint Analyticity of the Upper Field

We now turn to the joint analyticity estimate for the Dirichlet data

$$\zeta^u(x; \varepsilon, \delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \zeta^u_{n,m}(x)\varepsilon^n\delta^m.$$

The following lemma expands on Lemma 2.8.3.

Lemma 2.9.1. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0, d])$ then*

$$\|\zeta^u_{n,m}\|_{H^{s+3/2}} \leq K_\zeta B^n_\zeta D^m_\zeta, \quad \forall n \geq 0,\ m \geq 0, \tag{2.41}$$

for constants $K_\zeta, B_\zeta, D_\zeta > 0$.

Proof. [Lemma 2.9.1] We begin with an induction on m where for $m = 0$ we need to show that

$$\|\zeta^u_{n,0}\|_{H^{s+3/2}} \leq K_\zeta B^n_\zeta, \quad \forall n \geq 0,\ m = 0.$$

This result has previously been established in Lemma 2.8.3. Next, we assume the estimate (2.41) for all $n, m < \overline{m}$ and note that

$$\zeta^u_{n,\overline{m}} = (-i\underline{\gamma}^u)\left(\frac{f}{\overline{m}}\right)\zeta^u_{n-1,\overline{m}-1},$$

where we have used

$$\binom{n}{\overline{m}} = \frac{n}{\overline{m}}\binom{n-1}{\overline{m}-1}.$$

From this and $\overline{m} \geq 1, D_\zeta > 1$ we find the bound

$$\begin{aligned}\|\zeta^u_{n,\overline{m}}\|_{H^{s+3/2}} &\leq |\underline{\gamma}^u|\ \mathcal{M}\ |f|_{C^{s+3/2+\sigma}}\|\zeta_{n-1,\overline{m}-1}\|_{H^{s+3/2}} \\ &\leq |\underline{\gamma}^u|\ \mathcal{M}\ |f|_{C^{s+2}}K_\zeta B^{n-1}_\zeta D^{\overline{m}-1}_\zeta \\ &\leq |\underline{\gamma}^u|\ \mathcal{M}\ |f|_{C^{s+2}}K_\zeta B^{n-1}_\zeta D^{\overline{m}}_\zeta,\end{aligned}$$

and we are done provided

$$B_\zeta > |\underline{\gamma}^u|\ \mathcal{M}\ |f|_{C^{s+2}}.$$

□

We can now establish our desired result for the joint analyticity of the transformed field $u = u(x, z; \varepsilon, \delta)$ with respect to the perturbation parameters ε and δ.

Theorem 2.9.2. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0, d])$ and $\zeta^u_{n,m} \in H^{s+3/2}([0, d])$ such that*

$$\|\zeta^u_{n,m}\|_{H^{s+3/2}} \leq K_\zeta B^n_\zeta D^m_\zeta, \tag{2.42}$$

for constants $K_\zeta, B_\zeta, D_\zeta > 0$, then $u_{n,m} \in H^{s+2}([0, d] \times [0, a])$ and

$$\|u_{n,m}\|_{H^{s+2}} \leq KB^n D^m, \tag{2.43}$$

for constants $K, B, D > 0$.

To establish this result we work by induction. The key estimate is encapsulated in the following lemma.

Lemma 2.9.3. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0, d])$ and*

$$\|u_{n,m}\|_{H^{s+2}} \leq KB^n D^m, \quad \forall\ n \geq 0,\ m < \overline{m}, \tag{2.44}$$

for constants $K, B, D > 0$ then there exists a constant $\overline{C} > 0$ such that

$$\begin{aligned}\max\{\|\tilde{F}_{n,\overline{m}}\|_{H^s}, \|\tilde{P}_{n,\overline{m}}\|_{H^{s+1/2}}\} \leq K\overline{C}\Big\{ & B^n D^{\overline{m}-1} + B^n D^{\overline{m}-2} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}} + \\ & |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-1} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-2} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}} + \\ & |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-1} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-2} \Big\}.\end{aligned}$$

Proof. [Lemma 2.9.3] We begin with $\tilde{F}_{n,\overline{m}}$ and recall from (2.28) that

$$\begin{aligned}\tilde{F}_{n,\overline{m}}\left(x, z; f, u, \underline{\alpha}, \underline{\gamma}^u\right) = & -\mathrm{div}[A_1 \nabla u_{n-1,\overline{m}}] - \mathrm{div}[A_2 \nabla u_{n-2,\overline{m}}] - B_1 \nabla u_{n-1,\overline{m}} \\ & - B_2 \nabla u_{n-2,\overline{m}} - 2i\underline{\alpha}\partial_x u_{n,\overline{m}-1} - (\underline{\gamma}^u)^2 u_{n,\overline{m}-2} \\ & - 2(\underline{\gamma}^u)^2 u_{n,\overline{m}-1} - 2S_1 i\underline{\alpha}\partial_x u_{n-1,\overline{m}} - 2S_1 i\underline{\alpha}\partial_x u_{n-1,\overline{m}-1} \\ & - S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}-2} - 2S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}-1} - S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}} \\ & - 2S_2 i\underline{\alpha}\partial_x u_{n-2,\overline{m}} - 2S_2 i\underline{\alpha}\partial_x u_{n-2,\overline{m}-1} - S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}-2} \\ & - 2S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}-1} - S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}}.\end{aligned} \tag{2.45}$$

Then from (2.16) we have

$$\begin{aligned}\|\tilde{F}_{n,\overline{m}}\|^2_{H^s} \leq & \|A_1^{xx}\partial_x u_{n-1,\overline{m}}\|^2_{H^{s+1}} + \|A_1^{xz}\partial_z u_{n-1,\overline{m}}\|^2_{H^{s+1}} + \|A_1^{zx}\partial_x u_{n-1,\overline{m}}\|^2_{H^{s+1}} \\ & + \|A_1^{zz}\partial_z u_{n-1,\overline{m}}\|^2_{H^{s+1}} + \|A_2^{xx}\partial_x u_{n-2,\overline{m}}\|^2_{H^{s+1}} + \|A_2^{xz}\partial_z u_{n-2,\overline{m}}\|^2_{H^{s+1}} \\ & + \|A_2^{zx}\partial_x u_{n-2,\overline{m}}\|^2_{H^{s+1}} + \|A_2^{zz}\partial_z u_{n-2,\overline{m}}\|^2_{H^{s+1}} + \|B_1^x\partial_x u_{n-1,\overline{m}}\|^2_{H^s} \\ & + \|B_1^z\partial_z u_{n-1,\overline{m}}\|^2_{H^s} + \|B_2^x\partial_x u_{n-2,\overline{m}}\|^2_{H^s} + \|B_2^z\partial_z u_{n-2,\overline{m}}\|^2_{H^s} \\ & + \|2i\underline{\alpha}\partial_x u_{n,\overline{m}-1}\|^2_{H^s} + \|(\underline{\gamma}^u)^2 u_{n,\overline{m}-2}\|^2_{H^s} + \|2(\underline{\gamma}^u)^2 u_{n,\overline{m}-1}\|^2_{H^s} \\ & + \|2S_1 i\underline{\alpha}\partial_x u_{n-1,\overline{m}}\|^2_{H^s} + \|2S_1 i\underline{\alpha}\partial_x u_{n-1,\overline{m}-1}\|^2_{H^s} + \|S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}-2}\|^2_{H^s} \\ & + \|2S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}-1}\|^2_{H^s} + \|S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}}\|^2_{H^s} + \|2S_2 i\underline{\alpha}\partial_x u_{n-2,\overline{m}}\|^2_{H^s} \\ & + \|2S_2 i\underline{\alpha}\partial_x u_{n-2,\overline{m}-1}\|^2_{H^s} + \|S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}-2}\|^2_{H^s} + \|2S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}-1}\|^2_{H^s} \\ & + \|S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}}\|^2_{H^s}.\end{aligned}$$

We now estimate each of these and apply Lemmas 2.7.1, 2.7.3, and 2.8.1, beginning with

$$\begin{aligned}\|A_1^{xx}\partial_x u_{n-1,\overline{m}}\|_{H^{s+1}} &= \|-(2/a) f \partial_x u_{n-1,\overline{m}}\|_{H^{s+1}} \\ &\leq (2/a)\mathcal{M}|f|_{C^{s+1}} \|u_{n-1,\overline{m}}\|_{H^{s+2}} \\ &\leq (2/a)\mathcal{M}|f|_{C^{s+1}} K B^{n-1} D^{\overline{m}},\end{aligned}$$

and in a similar fashion

$$\begin{aligned}\|A_1^{xz}\partial_z u_{n-1,\overline{m}}\|_{H^{s+1}} &= \|-((a-z)/a)(\partial_x f)\partial_z u_{n-1,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_a/a)\mathcal{M}|\partial_x f|_{C^{s+1}}\|u_{n-1,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_a/a)\mathcal{M}|f|_{C^{s+2}}KB^{n-1}D^{\overline{m}}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_1^{zx}\partial_x u_{n-1,\overline{m}}\|_{H^{s+1}} &= \|-((a-z)/a)(\partial_x f)\partial_x u_{n-1,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_a/a)\mathcal{M}|\partial_x f|_{C^{s+1}}\|u_{n-1,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_a/a)\mathcal{M}|f|_{C^{s+2}}KB^{n-1}D^{\overline{m}},\end{aligned}$$

and we recall that $A_1^{zz}\equiv 0$. Moving to the second order

$$\begin{aligned}\|A_2^{xx}\partial_x u_{n-2,\overline{m}}\|_{H^{s+1}} &= \|(1/a^2)f^2\partial_x u_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (1/a^2)\mathcal{M}^2|f|^2_{C^{s+1}}\|u_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (1/a^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{n-2}D^{\overline{m}}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_2^{xz}\partial_z u_{n-2,\overline{m}}\|_{H^{s+1}} &= \|((a-z)/a^2)f(\partial_x f)\partial_x u_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|u_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{n-2}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zx}\partial_x u_{n-2,\overline{m}}\|_{H^{s+1}} &= \|((a-z)/a^2)f(\partial_x f)\partial_z u_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|u_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_a/a^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{n-2}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zz}\partial_z u_{n-2,\overline{m}}\|_{H^{s+1}} &= \|((a-z)^2/a^2)(\partial_x f)^2\partial_z u_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_a^2/a^2)\mathcal{M}^2|\partial_x f|^2_{C^{s+1}}\|u_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_a^2/a^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{n-2}D^{\overline{m}}.\end{aligned}$$

Next for the B_1 terms

$$\begin{aligned}\|B_1^x \partial_x u_{n-1,\overline{m}}\|_{H^s} &= \|(1/a)(\partial_x f)\partial_x u_{n-1,\overline{m}}\|_{H^s} \\ &\leq (1/a)\mathcal{M}|\partial_x f|_{C^s}\|u_{n-1,\overline{m}}\|_{H^{s+1}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1}} K B^{n-1} D^{\overline{m}},\end{aligned}$$

and $B_1^z \equiv 0$. Moving to the second order

$$\begin{aligned}\|B_2^x \partial_x u_{n-2,\overline{m}}\|_{H^s} &= \|(-1/a^2) f(\partial_x f)\partial_x u_{n-2,\overline{m}}\|_{H^s} \\ &\leq (1/a^2)\mathcal{M}^2|f|_{C^s}|\partial_x f|_{C^s}\|u_{n-2,\overline{m}}\|_{H^{s+1}} \\ &\leq (1/a^2)\mathcal{M}^2|f|^2_{C^{s+1}} K B^{n-2} D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|B_2^z \partial_z u_{n-2,\overline{m}}\|_{H^s} &= \|(-1/a^2)(a-z)(\partial_x f)^2\partial_z u_{n-2,\overline{m}}\|_{H^s} \\ &\leq (Z_a/a^2)\mathcal{M}^2|\partial_x f|^2_{C^s}\|u_{n-2,\overline{m}}\|_{H^{s+1}} \\ &\leq (Z_a/a^2)\mathcal{M}^2|f|^2_{C^{s+1}} K B^{n-2} D^{\overline{m}}.\end{aligned}$$

To address the S_0, S_1, S_2 terms we have

$$\begin{aligned}\|2i\underline{\alpha}\partial_x u_{n,\overline{m}-1}\|_{H^s} &\leq 2\underline{\alpha}\|u_{n,\overline{m}-1}\|_{H^{s+1}} \\ &\leq 2\underline{\alpha} K B^n D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|(\underline{\gamma}^u)^2 u_{n,\overline{m}-2}\|_{H^s} &\leq (\underline{\gamma}^u)^2\|u_{n,\overline{m}-2}\|_{H^s} \\ &\leq (\underline{\gamma}^u)^2 K B^n D^{\overline{m}-2},\end{aligned}$$

and

$$\begin{aligned}\|2(\underline{\gamma}^u)^2 u_{n,\overline{m}-1}\|_{H^s} &\leq 2(\underline{\gamma}^u)^2\|u_{n,\overline{m}-1}\|_{H^s} \\ &\leq 2(\underline{\gamma}^u)^2 K B^n D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|2S_1 i\underline{\alpha}\partial_x u_{n-1,\overline{m}}\|_{H^s} &= \|(-4/a)i\underline{\alpha} f\partial_x u_{n-1,\overline{m}}\|_{H^s} \\ &\leq (4/a)\underline{\alpha}\mathcal{M}|f|_{C^s}\|u_{n-1,\overline{m}}\|_{H^{s+1}} \\ &\leq (4/a)\underline{\alpha}\mathcal{M}|f|_{C^s} K B^{n-1} D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|2S_1 i\underline{\alpha}\partial_x u_{n-1,\overline{m}-1}\|_{H^s} &= \|(-4/a)i\underline{\alpha}f\partial_x u_{n-1,\overline{m}-1}\|_{H^s}\\ &\leq (4/a)\underline{\alpha}\mathcal{M}|f|_{C^s}\|u_{n-1,\overline{m}-1}\|_{H^{s+1}}\\ &\leq (4/a)\underline{\alpha}\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}-2}\|_{H^s} &= \|(-2/a)(\underline{\gamma}^u)^2 f u_{n-1,\overline{m}-2}\|_{H^s}\\ &\leq (2/a)(\underline{\gamma}^u)^2\mathcal{M}|f|_{C^s}\|u_{n-1,\overline{m}-2}\|_{H^s}\\ &\leq (2/a)(\underline{\gamma}^u)^2\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}-2},\end{aligned}$$

and

$$\begin{aligned}\|2S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}-1}\|_{H^s} &= \|(-4/a)(\underline{\gamma}^u)^2 f u_{n-1,\overline{m}-1}\|_{H^s}\\ &\leq (4/a)(\underline{\gamma}^u)^2\mathcal{M}|f|_{C^s}\|u_{n-1,\overline{m}-1}\|_{H^s}\\ &\leq (4/a)(\underline{\gamma}^u)^2\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_1(\underline{\gamma}^u)^2 u_{n-1,\overline{m}}\|_{H^s} &= \|(-2/a)(\underline{\gamma}^u)^2 f u_{n-1,\overline{m}}\|_{H^s}\\ &\leq (2/a)(\underline{\gamma}^u)^2\mathcal{M}|f|_{C^s}\|u_{n-1,\overline{m}}\|_{H^s}\\ &\leq (2/a)(\underline{\gamma}^u)^2\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|2S_2 i\underline{\alpha}\partial_x u_{n-2,\overline{m}}\|_{H^s} &= \|(2/a^2)i\underline{\alpha}f^2\partial_x u_{n-2,\overline{m}}\|_{H^s}\\ &\leq (2/a^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}\|u_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (2/a^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}KB^{n-2}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|2S_2 i\underline{\alpha}\partial_x u_{n-2,\overline{m}-1}\|_{H^s} &= \|(2/a^2)i\underline{\alpha}f^2\partial_x u_{n-2,\overline{m}-1}\|_{H^s}\\ &\leq (2/a^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}\|u_{n-2,\overline{m}-1}\|_{H^{s+1}}\\ &\leq (2/a^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}KB^{n-2}D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}-2}\|_{H^s} &= \|(1/a^2)(\underline{\gamma}^u)^2 f^2 u_{n-2,\overline{m}-2}\|_{H^s} \\ &\leq (1/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2 |f|^2_{C^s} \|u_{n-2,\overline{m}-2}\|_{H^s} \\ &\leq (1/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{n-2} D^{\overline{m}-2},\end{aligned}$$

and

$$\begin{aligned}\|2S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}-1}\|_{H^s} &= \|(2/a^2)(\underline{\gamma}^u)^2 f^2 u_{n-2,\overline{m}-1}\|_{H^s} \\ &\leq (2/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2 |f|^2_{C^s} \|u_{n-2,\overline{m}-1}\|_{H^s} \\ &\leq (2/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{n-2} D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_2(\underline{\gamma}^u)^2 u_{n-2,\overline{m}}\|_{H^s} &= \|(1/a^2)(\underline{\gamma}^u)^2 f^2 u_{n-2,\overline{m}}\|_{H^s} \\ &\leq (1/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2 |f|^2_{C^s} \|u_{n-2,\overline{m}}\|_{H^s} \\ &\leq (1/a^2)(\underline{\gamma}^u)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{n-2} D^{\overline{m}}.\end{aligned}$$

We satisfy the estimate for $\|\tilde{F}_{n,\overline{m}}\|_{H^s}$ provided that we choose

$$\begin{aligned}\overline{C} > \max\Bigg\{ &\left(2\underline{\alpha} + 3(\underline{\gamma}^u)^2\right), \left(\frac{3 + 2Z_a + 8\underline{\alpha} + 8(\underline{\gamma}^u)^2}{a}\right)\mathcal{M}, \\ &\left(\frac{2 + 3Z_a + Z_a^2 + 4\underline{\alpha} + 4(\underline{\gamma}^u)^2}{a^2}\right)\mathcal{M}^2\Bigg\}.\end{aligned}$$

The estimate for $\tilde{P}_{n,\overline{m}}$ follows from Lemma 2.8.2

$$\begin{aligned}\left\|\tilde{P}_{n,\overline{m}}\right\|_{H^{s+1/2}} &= \left\| -\frac{1}{a} f(x) \sum_{r=0}^{\overline{m}} T^u_{\overline{m}-r}\left[u_{n-1,r}\right] + \sum_{r=0}^{\overline{m}-1} T^u_{\overline{m}-r}\left[u_{n,r}\right] \right\|_{H^{s+1/2}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\eta}} \sum_{r=0}^{\overline{m}} \left\|T^u_{\overline{m}-r}\left[u_{n-1,r}\right]\right\|_{H^{s+1/2}} + \sum_{r=0}^{\overline{m}-1} \left\|T^u_{\overline{m}-r}\left[u_{n,r}\right]\right\|_{H^{s+1/2}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\eta}} C_{T^u} \sum_{r=0}^{\overline{m}} \|u_{n-1,r}\|_{H^{s+3/2}} + C_{T^u} \sum_{r=0}^{\overline{m}-1} \|u_{n,r}\|_{H^{s+3/2}} \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\eta}} C_{T^u} K B^{n-1} \left(\frac{D^{\overline{m}+1} - 1}{D-1}\right) + C_{T^u} K B^n \left(\frac{D^{\overline{m}} - 1}{D-1}\right) \\ &\leq (1/a)\mathcal{M}|f|_{C^{s+1/2+\eta}} C_{T^u} K B^{n-1} D^{\overline{m}} \left(\frac{D}{D-1}\right) + C_{T^u} K B^n D^{\overline{m}-1} \left(\frac{D}{D-1}\right),\end{aligned}$$

and we are done provided that $D > 2$ and

$$\overline{C} > \max\left\{(1/a)\mathcal{M} C_{T^u}, C_{T^u}\right\}. \qquad \square$$

With this information, we can now prove Theorem 2.9.2.

Proof. [Theorem 2.9.2] We proceed by induction in m. At order $m = 0$ (2.27) becomes

$$\Delta u_{n,0} + 2i\underline{\alpha}\partial_x u_{n,0} + (\underline{\gamma}^u)^2 u_{n,0} = \tilde{F}_{n,0}\left(x, z; f, u, \underline{\alpha}, \underline{\gamma}^u\right), \qquad 0 < z < a, \tag{2.46a}$$

$$u_{n,0}(x, g) = \zeta^u_{n,0}(x), \qquad \text{at } z = 0, \tag{2.46b}$$

$$u_{n,0}(x + d, z) = u_{n,0}(x, z), \tag{2.46c}$$

$$\partial_z\left[u_{n,0}(x, a)\right] - T^u_0[u_{n,0}(x, a)] = \tilde{P}_{n,0}(x), \qquad \text{at } z = a, \tag{2.46d}$$

and Theorem 2.8.4 guarantees a unique solution such that

$$\|u_{n,0}\|_{H^{s+2}} \leq KB^n, \quad \forall n \geq 0.$$

We now assume the estimate (2.43) for all $n, m < \overline{m}$ and study $u_{n,\overline{m}}$. From Theorem 2.7.2 we have a unique solution satisfying

$$\|u_{n,\overline{m}}\|_{H^{s+2}} \leq C_e\{\|\tilde{F}_{n,\overline{m}}\|_{H^s} + \|\zeta^u_{n,\overline{m}}\|_{H^{s+3/2}} + \|\tilde{P}_{n,\overline{m}}\|_{H^{s+1/2}}\},$$

and appealing to Lemmas 2.9.1 (with the hypothesis (2.42)) and 2.9.3 we find

$$\begin{aligned}\|u_{n,\overline{m}}\|_{H^{s+2}} \leq C_e\Bigg\{ & K_\zeta B^n_\zeta D^{\overline{m}}_\zeta + 2K\overline{C}\Big(B^n D^{\overline{m}-1} + B^n D^{\overline{m}-2} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}} + \\ & |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-1} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-2} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}} + \\ & |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-1} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-2}\Big)\Bigg\}.\end{aligned}$$

We are done provided we choose $K \geq 9C_e K_\zeta$ and

$$\begin{aligned} B &> \max\left\{B_\zeta, 18C_e\overline{C}|f|_{C^{s+2}}, \sqrt{18C_e\overline{C}}|f|_{C^{s+2}}\right\}, \\ D &> \max\left\{1, D_\zeta, 18C_e\overline{C}, \sqrt{18C_e\overline{C}}\right\}. \end{aligned}$$

□

2.10 Analyticity of the Upper Layer DNO

Now that we have established the analyticity of the transformed field, $u = u(x, z; \varepsilon, \delta)$, we move on to establishing the analyticity of the DNO, $G(g) = G(\varepsilon f)$. The Dirichlet trace is defined by

$$\zeta^u(x) := u(x, g(x)),$$

and its exterior Neumann counterpart is defined as

$$\nu^u(x) := [-N \cdot \nabla u](x, g(x)) = [-\partial_z u + (\partial_x g)\partial_x u](x, g(x)), \tag{2.47}$$

where $N = (-\partial_x g, 1)^T$. With this we define the DNO

$$G(g) := \zeta^u \to \nu^u, \tag{2.48}$$

which maps the Dirichlet data, ζ^u, to the Neumann trace, ν^u,

$$G(g)[\zeta^u] = [-\partial_z u + (\partial_x g)\partial_x u](x, g(x)). \tag{2.49}$$

We now analyze how the operator $G(g)$ behaves under the change of variables in Appendix C. To do this, we multiply (2.49) by $C(x)$ to realize

$$CG = -C\partial_z u + (\partial_x g)C\partial_x u.$$

The differentiation rules for the change of variables, (C.5),

$$C\partial_x = C\partial_{x'} - E\partial_{z'}, \quad C\partial_z = \partial_{z'},$$

produces

$$CG = -\partial_{z'} u' + (\partial_{x'} g)\{C\partial_{x'} u' - E\partial_{z'} u'\}.$$

These are evaluated at the lower boundary, $z' = 0$, where we observe that

$$C(x') = 1 - \frac{g}{a}, \quad E(x', 0) = \partial_{x'} g,$$

to find

$$\left(1 - \frac{g}{a}\right) G = -\partial_{z'} u' + (\partial_{x'} g)\left\{\left(1 - \frac{g}{a}\right)\partial_{x'} u' - (\partial_{x'} g)\partial_{z'} u'\right\}.$$

We solve for G and drop the primes to find

$$G(g)[\zeta^u] = -\partial_z u(x, 0) + H(x; g, u), \tag{2.50}$$

with

$$\begin{aligned} H(x; g, u) := {} & (\partial_x g)\partial_x u(x, 0) + \frac{1}{a} g G(g)[\zeta^u] \\ & - \frac{1}{a} g(\partial_x g)\partial_x u(x, 0) - (\partial_x g)^2 \partial_z u(x, 0). \end{aligned} \tag{2.51}$$

Upon setting $g(x) = \varepsilon f(x)$ and seeking an expansion of the form

$$G = G(\varepsilon f, \delta) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} G_{n,m}(f) \varepsilon^n \delta^m,$$

the equations (2.50) and (2.51) deliver

$$G_{n,m}(f)[\zeta^u] = -\partial_z u_{n,m}(x,0) + H_{n,m}(x; f, u), \tag{2.52}$$

by which

$$\begin{aligned} H_{n,m}(x; f, u) := (\partial_x f)\partial_x u_{n-1,m}(x,0) + \frac{1}{a} f G_{n-1,m}(f)[\zeta^u] \\ - \frac{1}{a} f(\partial_x f)\partial_x u_{n-2,m}(x,0) - (\partial_x f)^2 \partial_z u_{n-2,m}(x,0). \end{aligned} \tag{2.53}$$

To prove the analyticity of the DNO we will need the following recursive estimate for $H_{n,m}$.

Lemma 2.10.1. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and*

$$\|u_{n,m}\|_{H^{s+2}} \leq K B^n D^m, \quad \|G_{n,m}\|_{H^{s+1/2}} \leq \tilde{K} \tilde{B}^n \tilde{D}^m, \quad \forall\ n < \overline{n},\ m \geq 0, \tag{2.54}$$

for constants $K, B, D, \tilde{K}, \tilde{B}, \tilde{D} > 0$ where $\tilde{K} \geq K, \tilde{B} \geq B, \tilde{D} \geq D$, then there exists a constant $\tilde{C} > 0$ such that

$$\|H_{\overline{n},m}\|_{H^{s+1/2}} \leq \tilde{K}\tilde{C} \left\{ |f|_{C^{s+2}} \tilde{B}^{n-1} \tilde{D}^m + |f|^2_{C^{s+2}} \tilde{B}^{n-2} \tilde{D}^m \right\}. \tag{2.55}$$

Proof. [Lemma 2.10.1] From (2.53) and Lemma 2.7.1 we estimate

$$\begin{aligned} \|H_{\overline{n},m}\|_{H^{s+1/2}} &\leq \mathcal{M}|\partial_x f|_{C^{s+1/2+\sigma}} \|\partial_x u_{\overline{n}-1,m}(x,0)\|_{H^{s+1/2}} \\ &+ \frac{1}{a}\mathcal{M}|f|_{C^{s+1/2+\sigma}} \|G_{\overline{n}-1,m}(f)[\zeta^u]\|_{H^{s+1/2}} \\ &+ \frac{1}{a}\mathcal{M}^2|f|_{C^{s+1/2+\sigma}} |\partial_x f|_{C^{s+1/2+\sigma}} \|\partial_x u_{\overline{n}-2,m}(x,0)\|_{H^{s+1/2}} \\ &+ \mathcal{M}^2|\partial_x f|^2_{C^{s+1/2+\sigma}} \|\partial_z u_{\overline{n}-2,m}(x,0)\|_{H^{s+1/2}}. \end{aligned}$$

This gives

$$\begin{aligned} \|H_{\overline{n},m}\|_{H^{s+1/2}} \leq \tilde{K}\Big\{ &\mathcal{M}|f|_{C^{s+2}} \tilde{B}^{\overline{n}-1} \tilde{D}^m + \frac{1}{a}\mathcal{M}|f|_{C^{s+2}} \tilde{B}^{\overline{n}-1} \tilde{D}^m \\ &+ \frac{1}{a}\mathcal{M}^2|f|^2_{C^{s+2}} \tilde{B}^{\overline{n}-2} \tilde{D}^m + \mathcal{M}^2|f|^2_{C^{s+2}} \tilde{B}^{\overline{n}-2} \tilde{D}^m \Big\}, \end{aligned}$$

and we are done provided

$$\tilde{C} \geq \left(1 + \frac{1}{a}\right) \max\{\mathcal{M}, \mathcal{M}^2\}.$$

□

We now have everything we need to prove the analyticity of the upper layer DNO.

Theorem 2.10.2. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and $\zeta^u_{n,m} \in H^{s+3/2}([0,d])$ such that*

$$\|\zeta^u_{n,m}\|_{H^{s+3/2}} \leq K_\zeta B_\zeta^n D_\zeta^m,$$

for constants $K_\zeta, B_\zeta, D_\zeta > 0$, then $G_{n,m} \in H^{s+1/2}([0,d])$ and

$$\|G_{n,m}\|_{H^{s+1/2}} \leq \tilde{K}\tilde{B}^n\tilde{D}^m, \tag{2.56}$$

for constants $\tilde{K}, \tilde{B}, \tilde{D} > 0$.

Proof. [Theorem 2.10.2] As before, we work by induction in n. At $n = 0$ we have from (2.52) that

$$G_{0,m} = -\partial_z u_{0,m}(x, 0),$$

and from Theorem 2.9.2 we have

$$\|G_{0,m}\|_{H^{s+1/2}} = \|\partial_z u_{0,m}(x,0)\|_{H^{s+1/2}} \leq \|u_{0,m}\|_{H^{s+2}} \leq KD^m.$$

So we choose $\tilde{K} \geq K$ and $\tilde{D} \geq D$. We now assume $\tilde{B} \geq B$ and the estimate (2.56) for all $n < \overline{n}$ and estimate (2.52)

$$\|G_{\overline{n},m}(f)[\zeta^u]\|_{H^{s+1/2}} \leq \|\partial_z u_{\overline{n},m}(x,0)\|_{H^{s+1/2}} + \|H_{\overline{n},m}(x)\|_{H^{s+1/2}}.$$

Using the inductive hypothesis, Lemma 2.10.1, and Theorem 2.9.2 we have

$$\|G_{\overline{n},m}(f)[\zeta^u]\|_{H^{s+1/2}} \leq KB^{\overline{n}}D^m + \tilde{K}\tilde{C}\left\{|f|_{C^{s+2}}\tilde{B}^{\overline{n}-1}\tilde{D}^m + |f|^2_{C^{s+2}}\tilde{B}^{\overline{n}-2}\tilde{D}^m\right\}.$$

We are done provided $\tilde{K} \geq 2K$ and

$$\tilde{B} \geq \max\left\{B, 4\tilde{C}|f|_{C^{s+2}}, 2\sqrt{\tilde{C}}|f|_{C^{s+2}}\right\}.$$ □

2.11 Numerical Method

We explain in detail in Chapter 5 how our numerical scheme is validated using the Method of Manufactured Solutions. In this section, we present the process of simulating

a manufactured solution in order to evaluate the accuracy of our numerical scheme in the upper field. We start by considering the basis function

$$v_p^u(x,z) := e^{i\tilde{p}x+i\gamma_p^u z}, \quad \tilde{p} = \frac{2\pi p}{d},$$

where the phase $\exp(i\alpha x)$ is removed. In order to test our algorithm we will utilize the exact Dirichlet/Neumann pairs defined below by $\{\zeta_r^u, \nu_r^u\}$. Our strategy will be to select a particular wavenumber, say $p = r$, and a profile $g(x) = \varepsilon f(x)$ where $\varepsilon > 0$ is small and our manufactured solutions are

$$\zeta_r^u(x) := A_r e^{i\tilde{r}x+i\gamma_r^u g(x)}, \tag{2.57a}$$

$$\begin{aligned}\nu_r^u(x) &:= [-\partial_z u_r + (\partial_x g)\partial_x u_r](x, g(x)) \\ &= [-(i\gamma_r^u) + \varepsilon(\partial_x f)(i\tilde{r})]A_r e^{i\tilde{r}x+i\gamma_r^u \varepsilon f(x)}.\end{aligned} \tag{2.57b}$$

To perform our tests, we will first send ζ_r^u to our algorithm and then compare our approximation to ν_r^u. Our algorithm is a Fourier spectral method (86; 87; 88) where we sample ζ_r^u at equally spaced gridpoints on $[0, d]$, and use the TFE recursions to generate ν_r^u at the same, equally spaced gridpoints. To make the specification precise we solve, at every desired perturbation order n and m, the elliptic boundary value problem, (2.27),

$$\Delta u_{n,m} + 2i\underline{\alpha}\partial_x u_{n,m} + (\underline{\gamma}^u)^2 u_{n,m} = \tilde{F}_{n,m}\left(x, z; f, u, \underline{\alpha}, \underline{\gamma}^u\right), \qquad 0 < z < a, \tag{2.58a}$$

$$u_{n,m}(x,0) = \zeta_{n,m}^u(x), \qquad \text{at } z = 0, \tag{2.58b}$$

$$u_{n,m}(x+d,z) = u_{n,m}(x,z), \tag{2.58c}$$

$$\partial_z\left[u_{n,m}(x,a)\right] - T_0^u[u_{n,m}(x,a)] = \tilde{P}_{n,m}(x), \qquad \text{at } z = a, \tag{2.58d}$$

followed by the simulation of the n–th and m–th correction of the DNO, (2.52),

$$G_{n,m}(f)[\zeta^u] = -\partial_z u_{n,m}(x,0) + H_{n,m}(x; f, u).$$

We begin by choosing the maximum perturbation orders, N and M, and then approximate

$$u(x,z;\varepsilon,\delta) \approx u^{N,M}(x,z;\varepsilon,\delta) := \sum_{n=0}^{N}\sum_{m=0}^{M} u_{n,m}(x,z)\varepsilon^n\delta^m, \tag{2.59}$$

$$G(x;\varepsilon,\delta) \approx G^{N,M}(x;\varepsilon,\delta) := \sum_{n=0}^{N}\sum_{m=0}^{M} G_{n,m}(x)\varepsilon^n\delta^m, \tag{2.60}$$

where, by the periodicity of solutions, we write

$$u_{n,m}(x,z) = \sum_{p=-\infty}^{\infty} \hat{u}_{n,m,p}(z)e^{i\tilde{p}x}, \quad G_{n,m}(x) = \sum_{p=-\infty}^{\infty} \hat{G}_{n,m,p}e^{i\tilde{p}x}. \tag{2.61}$$

Each of these $u_{n,m}(x,z)$ are then simulated by a Fourier–Chebyshev approach which posits the form

$$u_{n,m}(x,z) \approx u_{n,m}^{N_x,N_z}(x,z) := \sum_{p=-N_x/2}^{N_x/2-1} \sum_{\ell=0}^{N_z} \hat{u}_{n,m,p,\ell} e^{i\tilde{p}x} T_\ell\left(\frac{2z-a}{a}\right),$$

where T_ℓ is the ℓ–th Cheybshev polynomial. The unknowns, $\hat{u}_{n,m,p,\ell}$ are recovered from (2.58) by the collocation approach (86; 87; 89; 90; 91). More specifically, our HOPS/AWE algorithm requires $N_x \times N_z$ unknowns at every perturbation order (n,m). As our problem is x–periodic, the Fourier spectral method in the lateral direction requires N_x equally–spaced gridpoints. However, our problem is not z–periodic, so the Chebyshev spectral method in the vertical direction requires N_z+1 collocation points where

$$z_\ell = \frac{2\tilde{z}_\ell - a}{a}, \quad \tilde{z}_\ell = \cos\left(\ell\pi/N_z\right), \quad \ell = 0,\ldots,N_z.$$

We then simulate the upper layer DNO from (2.52), where the coefficients $G_{n,m}$ from (2.61) are approximated by

$$G_{n,m}(x) \approx G_{n,m}^{N_x}(x) := \sum_{p=-N_x/2}^{N_x/2-1} \hat{G}_{n,m,p} e^{i\tilde{p}x}, \tag{2.62}$$

and the $\hat{G}_{n,m,p}$ are recovered from the $\hat{u}_{n,m,p,\ell}$. Inserting the expansions (2.61) into (2.58) gives

$$\partial_z^2 \hat{u}_{n,m,p}(z) + \left((\underline{\gamma}_p^u)^2 - \tilde{p}^2 - 2\underline{\alpha}\tilde{p}\right)\hat{u}_{n,m,p}(z) = \hat{\tilde{F}}_{n,m,p}(z), \qquad 0<z<a, \tag{2.63a}$$

$$\hat{u}_{n,m,p}(0) = \hat{\zeta}_{n,m,p}^u, \qquad \text{at } z=0, \tag{2.63b}$$

$$\partial_z\left[\hat{u}_{n,m,p}(a)\right] - \hat{T}_0^u[\hat{u}_{n,m,p}(a)] = \hat{\tilde{P}}_{n,m,p}, \qquad \text{at } z=a. \tag{2.63c}$$

We can solve this two–point boundary value problem by the Chebyshev collocation method and we now turn to a numerical implementation of our HOPS/AWE algorithm in Matlab. To begin, we approximate the upper layer Dirichlet and Neumann data through the expansions

$$\zeta_{\text{TFE}}^{N_x,N_z,N,M} = \sum_{n=0}^{N}\sum_{m=0}^{M} \zeta_{n,m}^{N_x,N_z}(x)\varepsilon^n\delta^m, \quad \nu_{\text{TFE}}^{N_x,N_z,N,M} = \sum_{n=0}^{N}\sum_{m=0}^{M} \nu_{n,m}^{N_x,N_z}(x)\varepsilon^n\delta^m,$$

from which we can compute the relative errors

$$\text{Error } \zeta_r^u = \text{Error}_{\text{TFE}}(N_x,N_z,N,M) := \frac{\left|\zeta_r^u - \zeta_{TFE}^{Nx,Nz,N,M}\right|_{L^\infty}}{|\zeta_r^u|_{L^\infty}},$$

$$\text{Error } \nu_r^u = \text{Error}_{\text{TFE}}(N_x, N_z, N, M) := \frac{\left|\nu_r^u - \nu_{TFE}^{Nx,Nz,N,M}\right|_{L^\infty}}{|\nu_r^u|_{L^\infty}}.$$

2.12 Padé Approximation

We conclude our discussion of numerics by considering how the Taylor series in (ε, δ) are summed. For example, regarding the DNO, G, the approximation of $\hat{G}_p(\varepsilon, \delta)$ by

$$\hat{G}_p^{N,M}(\varepsilon, \delta) := \sum_{n=0}^{N} \sum_{m=0}^{M} \hat{G}_{n,m,p} \varepsilon^n \delta^m,$$

cf. (2.62). The technique of Padé approximation (92) has been used with HOPS methods to great advantage in the past (65; 93) and we advocate its use here. Classically, this approach seeks to estimate the truncated Taylor series of a single variable

$$Q^N(\rho) := \sum_{n=0}^{N} Q_n \rho^n \approx Q(\rho),$$

by the rational function

$$[L/M](\rho) := \frac{a^L(\rho)}{b^M(\rho)} = \frac{\sum_{\ell=0}^{L} a_\ell \rho^\ell}{1 + \sum_{m=1}^{M} b_m \rho^m}, \quad L + M = N,$$

and

$$[L/M](\rho) = Q^N(\rho) + \mathcal{O}\left(\rho^{L+M+1}\right),$$

where well–known formulas for the coefficients $\{a_\ell, b_m\}$ can be found in (92). Padé approximation enjoys greatly enhanced convergence properties and we refer the interested reader to §2.2 of Baker & Graves–Morris (92) and the insightful calculations of §8.3 of Bender & Orszag (94) for a thorough discussion of the capabilities and limitations of Padé approximants.

In the current context of functions analytic with respect to two perturbation variables we utilize the polar coordinates

$$\varepsilon = \rho \cos(\theta), \quad \delta = \rho \sin(\theta),$$

and write the function

$$\begin{aligned} \hat{G}_p(\varepsilon, \delta) &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \hat{G}_{n,m,p} \varepsilon^n \delta^m \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \left(\hat{G}_{n,m,p} \cos^n(\theta) \sin^m(\theta)\right) \rho^{n+m}. \end{aligned}$$

Letting $\ell = n + m$ and $s = m$ we can write this as

$$\hat{G}_p(\varepsilon, \delta) = \sum_{\ell=0}^{\infty} \left\{ \sum_{s=0}^{\ell} \hat{G}_{\ell-s,s,p} \cos^{\ell-s}(\theta) \sin^s(\theta) \right\} \rho^\ell =: \sum_{\ell=0}^{\infty} \tilde{G}_{\ell,p}(\theta) \rho^\ell.$$

We then select particular values of $\theta = \theta_j$ between 0 and 2π and apply classical Padé approximation on the resulting $\{\tilde{G}_{\ell,p}(\theta_j)\}$ as a function of ρ alone.

2.13 Numerical Results

For our first simulation we considered an analytic profile with the following parameters

$$f(x) = e^{\cos(x)}, \quad \alpha = 0, \quad \varepsilon = 10^{-6}, \quad \delta = 10^{-8}, \quad d = 2\pi, \quad r = 2,$$

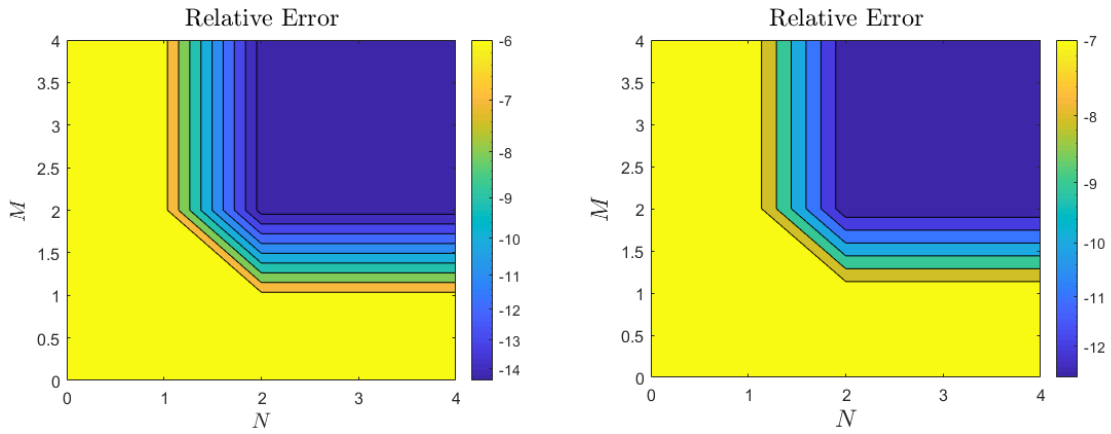
$$A_r = -3, \quad \gamma^u = 1.21, \quad N_x = 32, \quad N_z = 32, \quad N = M = 4.$$

In Table I we report the results of our tests using both Padé and Taylor summation.

N	M	Error ζ_r^w (Taylor)	Error ζ_r^w (Padé)	Error ν_r^w (Taylor)	Error ν_r^w (Padé)
0	2	2.05326e-06	2.05326e-06	7.82969e-07	7.82969e-07
0	4	2.05326e-06	2.05326e-06	7.82969e-07	7.82969e-07
1	2	2.05326e-06	2.05326e-06	7.82969e-07	7.82969e-07
1	4	2.05326e-06	2.05326e-06	7.82969e-07	7.82969e-07
2	2	2.98167e-12	3.84881e-15	9.68942e-13	2.11309e-13
2	4	2.98167e-12	3.84881e-15	9.68942e-13	2.11309e-13
3	2	2.98167e-12	3.84881e-15	9.68942e-13	2.11309e-13
3	4	2.98167e-12	3.84881e-15	9.68942e-13	2.11309e-13
4	2	2.98167e-12	3.84881e-15	9.68942e-13	2.11309e-13
4	4	3.84181e-15	3.85621e-15	2.11309e-13	2.11301e-13

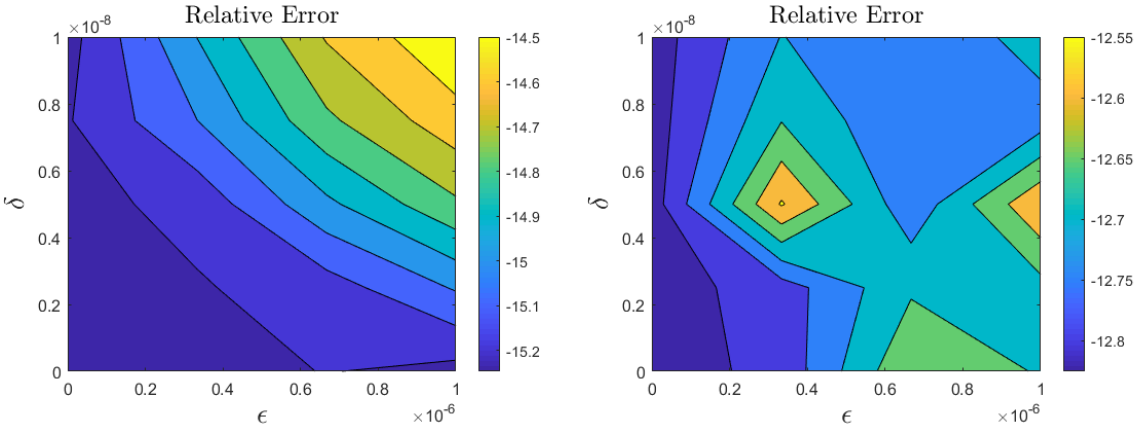
TABLE I: Relative Error, Error ζ_r^u and Error ν_r^u, versus perturbation orders N and M, for the TFE approximations to the Dirichlet data, ζ_r^u (2.57a), and the Neumann data, ν_r^u (2.57b), where we used both Taylor Series and Padé approximants. Parameter choices are specified above and both ε and δ are small.

As we can see by expanding through orders $0 \leq N, M \leq 4$, our HOPS/AWE algorithm quickly obtains spectral accuracy provided we have small values of ε and δ. Expanding on this, we generate two figures with the existing parameters in our analytic profile. In Figure 4, we keep $\varepsilon = 10^{-6}$ and $\delta = 10^{-8}$ fixed while we plot the Relative Error for ζ_r^u and ν_r^u as we expand through $0 \leq N, M \leq 4$ Padé orders. In Figure 5, we keep $N = M = 4$ fixed and plot the Relative Error for ζ_r^u and ν_r^u as we expand through $0 \leq \varepsilon \leq 10^{-6}$ and $0 \leq \delta \leq 10^{-8}$ with Padé summation.



(a) Error ζ_r^u (Padé) (b) Error ν_r^u (Padé)

Figure 4: Plot of Relative Error for ζ_r^u and ν_r^u with fixed $\varepsilon = 10^{-6}$ and $\delta = 10^{-8}$. Our HOPS/AWE algorithm used Padé summation and expanded through $0 \leq N, M \leq 4$ Padé orders. Physical parameters are reported in the analytic profile above.



(a) Error ζ_r^u (Padé) (b) Error ν_r^u (Padé)

Figure 5: Plot of Relative Error for ζ_r^u and ν_r^u with $N = M = 4$ fixed. Our HOPS/AWE algorithm used Padé summation to expand through $0 \leq \varepsilon \leq 10^{-6}$ and $0 \leq \delta \leq 10^{-8}$. Physical parameters are reported in the analytic profile above.

In light of this, we ask the natural question - what is the maximum size of the grating deformation, ε, and frequency perturbation, δ, necessary to achieve spectral accuracy? To investigate we considered a significantly larger perturbation in both ε and δ and simulated the same profile with the following parameters

$$f(x) = e^{\cos(x)}, \quad \alpha = 0, \quad \varepsilon = 0.02, \quad \delta = 0.001, \quad d = 2\pi, \quad r = 2,$$

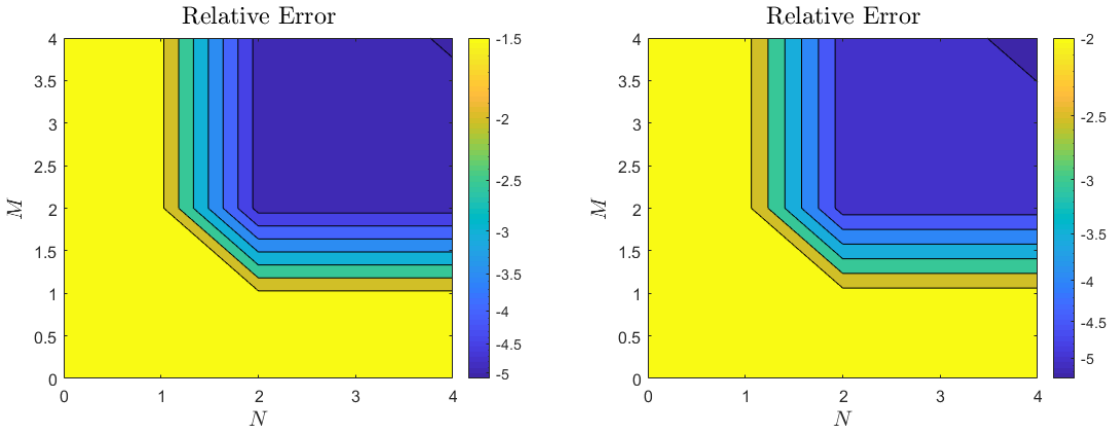
$$A_r = -3, \quad \gamma^u = 1.21, \quad N_x = 32, \quad N_z = 32, \quad N = M = 4.$$

In Table II we report the results of our tests using both Padé and Taylor summation.

N	M	Error ζ_r^w (Taylor)	Error ζ_r^w (Padé)	Error ν_r^w (Taylor)	Error ν_r^w (Padé)
0	2	0.0393875	0.0393875	0.0151348	0.0151348
0	4	0.0393875	0.0393875	0.0151348	0.0151348
1	2	0.0393875	0.0393875	0.0151348	0.0151348
1	4	0.0393875	0.0393875	0.0151348	0.0151348
2	2	0.00110548	2.06154e-05	0.000398635	1.88162e-05
2	4	0.00110548	2.06154e-05	0.000398635	1.88162e-05
3	2	0.00110548	2.06154e-05	0.000398635	1.88162e-05
3	4	0.00110548	2.06154e-05	0.000398635	1.88162e-05
4	2	0.00110548	2.06154e-05	0.000398635	1.88162e-05
4	4	3.23201e-05	8.02125e-06	1.26113e-05	5.0552e-06

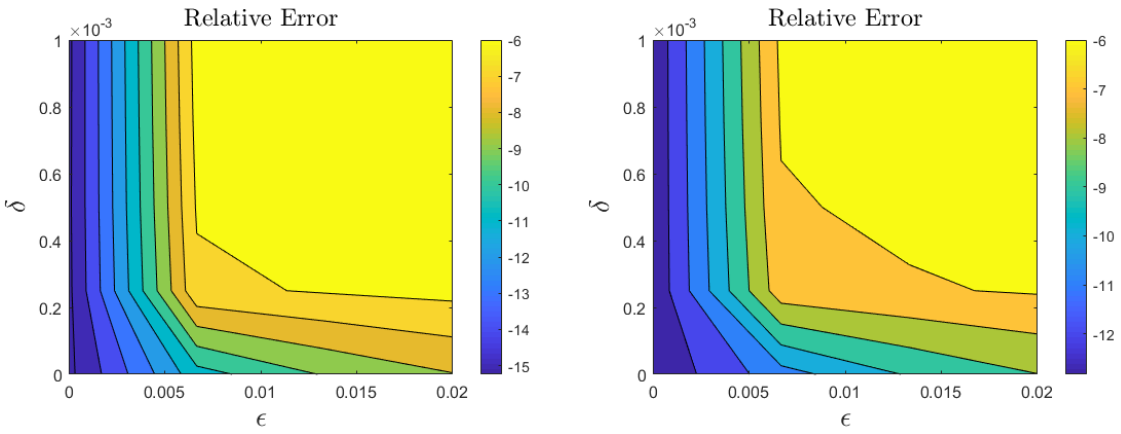
TABLE II: Relative Error, Error ζ_r^u and Error ν_r^u, versus perturbation orders N and M, for the TFE approximations to the Dirichlet data, ζ_r^u (2.57a), and the Neumann data, ν_r^u (2.57b), where we used both Taylor Series and Padé approximants. Parameter choices are specified above and both ε and δ are large.

At first, these results are slightly alarming. Continuing in the same manner as the analytic profile, we provide two more figures for the same test parameters. In Figure 6, we keep $\varepsilon = 0.02$ and $\delta = 0.001$ fixed while we plot the Relative Error for ζ_r^u and ν_r^u as we expand through $0 \leq N, M \leq 4$ Padé orders. In Figure 7, we keep $N = M = 4$ fixed and plot the Relative Error for ζ_r^u and ν_r^u as we expand through $0 \leq \varepsilon \leq 0.02$ and $0 \leq \delta \leq 0.001$.



(a) Error ζ_r^u (Padé) (b) Error ν_r^u (Padé)

Figure 6: Plot of Relative Error for ζ_r^u and ν_r^u with fixed $\varepsilon = 0.02$ and $\delta = 0.001$. Our HOPS/AWE algorithm used Padé summation and expanded through $0 \leq N, M \leq 4$ Padé orders. Physical parameters are reported above where both ε and δ are large.



(a) Error ζ_r^u (Padé) (b) Error ν_r^u (Padé)

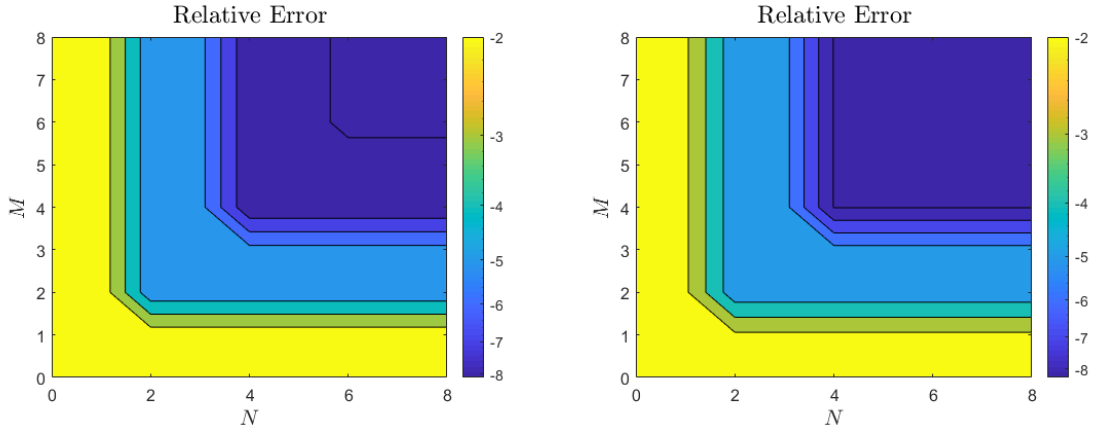
Figure 7: Plot of Relative Error for ζ_r^u and ν_r^u with $N = M = 4$ fixed. Our HOPS/AWE algorithm used Padé summation to expand through $0 \leq \varepsilon \leq 0.02$ and $0 \leq \delta \leq 0.001$. Physical parameters are reported above where both ε and δ are large.

We then simulated the same profile with

$$f(x) = e^{\cos(x)}, \quad \alpha = 0, \quad \varepsilon = 0.02, \quad \delta = 10^{-6}, \quad d = 2\pi, \quad r = 2,$$

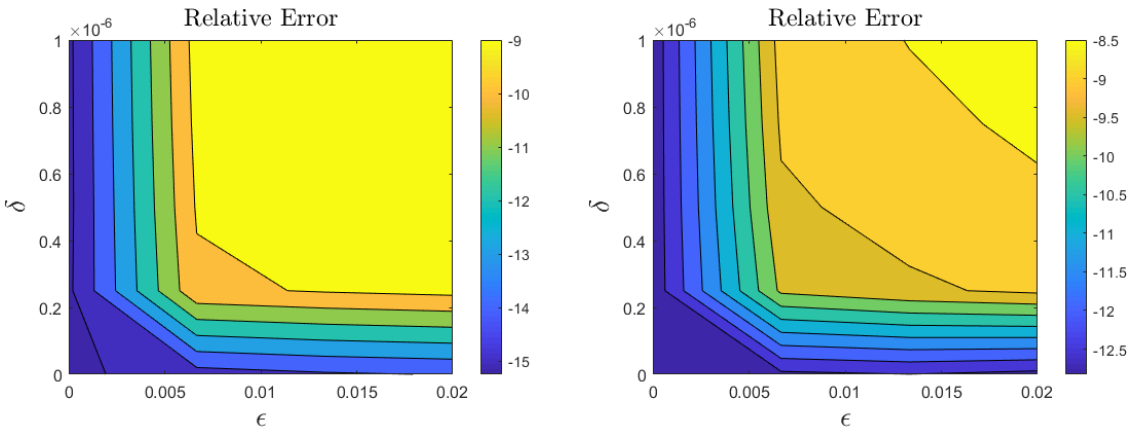
$$A_r = -3, \quad \gamma^u = 1.21, \quad N_x = 32, \quad N_z = 32, \quad N = M = 8,$$

where we now expand through $0 \leq N, M \leq 8$ Padé orders and have a smaller frequency perturbation. In Figure 8, we keep $\varepsilon = 0.02$ and $\delta = 10^{-6}$ fixed while we plot the Relative Error for ζ_r^u and ν_r^u as we expand through $0 \leq N, M \leq 8$ Padé orders. In Figure 9, we keep $N = M = 8$ fixed and plot the Relative Error for ζ_r^u and ν_r^u as we expand through $0 \leq \varepsilon \leq 0.02$ and $0 \leq \delta \leq 10^{-6}$.



(a) Error ζ_r^u (Padé) (b) Error ν_r^u (Padé)

Figure 8: Plot of Relative Error for ζ_r^u and ν_r^u with fixed $\varepsilon = 0.02$ and $\delta = 10^{-6}$. Our HOPS/AWE algorithm used Padé summation to expand through $0 \leq N, M \leq 8$ Padé orders. Physical parameters are reported above where ε is large and δ is small.



(a) Error ζ_r^u (Padé)

(b) Error ν_r^u (Padé)

Figure 9: Plot of Relative Error for ζ_r^u and ν_r^u with $N = M = 8$ fixed. Our HOPS/AWE algorithm used Padé summation to expand through $0 \leq \varepsilon \leq 0.02$ and $0 \leq \delta \leq 10^{-6}$. Physical parameters are reported above where ε is large and δ is small.

Further testing shows that our HOPS/AWE algorithm is better suited towards larger perturbations of ε, which deforms the surface $z = g(x)$, in comparison to large deformations of δ, the frequency. Several factors may be contributing to this effect, including the method by which the DNO, G, recovers surface data from the transformed field. As a result, a more detailed analysis will be performed in Chapter 7. In Appendix A, we provide several code samples which highlight some of our Matlab work. The first subroutine discusses the strategy used to calculate the transformed field, $u = u(x, y; \varepsilon, \delta)$, while the second subroutine explains how we calculate interfacial data by the upper layer DNO, and the third subroutine clarifies how we invert the operator $\mathbf{A}_{0,0}$. We are now ready to analyze the lower field.

CHAPTER 3

ANALYTICITY OF THE LOWER FIELD

3.1 Introduction

In the same spirit as the upper field, we present all of the necessary information to establish the analyticity of the lower field and the lower layer DNO. Our strategy is to first remove the phase from our governing equations in §3.2, introduce a domain–flattening change of variables in §3.3, and then seek solutions as a joint Taylor series in two small perturbation variables: an interfacial deformation (§3.3) and a frequency deformation (§3.4). These lead to the TFE recursions (§3.5) from which we can use Sobolev space theory to establish analyticity results. The analyticity of the lower field with respect to a single interfacial deformation is established in §3.7 while the joint analyticity in two small perturbations is established in §3.8. The case for a single frequency deformation follows directly as a special case of Theorem 3.8.1 and the analyticity of the lower layer DNO is proven in Theorem 3.9.2. The chapter's conclusion employs a Fourier–Chebyshev approach to highlight interesting numerical features of our HOPS/AWE algorithm.

3.2 Governing Equations Without Phase

In the lower field, we defined the geometry $S_{-b,g} := \{-b < z < g(x)\}$ where z is bounded between a constant imposed by the Artificial Boundary $\{z = -b\}$ and the upper surface $g(x)$. By the boundary value problem (1.35) defined in §1.8, we arrive at the governing equations of linear wave propagation in a single homogeneous material layer

$$\Delta\tilde{w} + (k^w)^2\tilde{w} = 0, \qquad -b < z < g(x), \tag{3.1a}$$

$$\tilde{w}(x, g(x)) = \tilde{\zeta}^w(x), \qquad \text{at } z = g(x), \tag{3.1b}$$

$$\tilde{w}(x + d, z) = e^{i\alpha d}\tilde{w}(x, z), \tag{3.1c}$$

$$\partial_z\tilde{w} - \tilde{T}^w[\tilde{w}] = 0, \qquad \text{at } z = b. \tag{3.1d}$$

Following the analysis performed in §2.2 we define phase extraction

$$w(x, z) := e^{-i\alpha x}\tilde{w}(x, z),$$

and consider periodic unknowns. Following the same procedure as the upper field, our governing equations become

$$\Delta w + 2i\alpha\partial_x w + (\gamma^w)^2 w = 0, \qquad -b < z < g(x), \tag{3.2a}$$

$$w(x, g(x)) = \zeta^w(x), \qquad \text{at } z = g(x), \tag{3.2b}$$

$$w(x + d, z) = w(x, z), \tag{3.2c}$$

$$\partial_z [w(x, -b)] - T^w[w(x, -b)] = 0, \qquad \text{at } z = -b. \tag{3.2d}$$

3.3 Boundary Perturbation

As in the upper field, we apply the change of variables from Appendix C to (3.2) and start by focusing on

$$\Delta w + 2i\alpha\partial_x w + (\gamma^w)^2 w = 0. \tag{3.3}$$

The transformation rules produce the following transformation in the lower field

$$x' = x, \quad z' = b\left(\frac{z - g(x)}{b + g(x)}\right).$$

This transformation maps the perturbed geometry $S_{-b,g}$ to the separable one $S_{-b,0}$. The change of variables can be inverted

$$x = x', \quad z = \left(\frac{b + g(x')}{b}\right) z' + g(x'),$$

which we use to define the transformed lower field

$$w(x', z') := w'\left(x', \left(\frac{b + g(x')}{b}\right) z' + g(x')\right).$$

In Appendix C we discuss the effects of this change of variables on the Helmholtz equation, its derivatives, and the associated boundary conditions. In the lower layer we have a domain $S_{L,U}$, (C.1), where

$$\overline{\ell} = -b, \quad \ell(x) \equiv 0, \quad \overline{u} = 0, \quad u(x) = g(x), \quad \overline{h} = \overline{u} - \overline{\ell} = b.$$

Therefore

$$C(x) = 1 + \frac{g(x) - 0}{b} = 1 + \frac{g(x)}{b}, \quad D(x) = \frac{0^2 + bg(x)}{b} = g(x),$$

and

$$E = (\partial_x g)\left(\frac{z' + b}{b}\right), \quad Z_L = \frac{z' + b}{b}.$$

(We omit Z_U since $\ell \equiv 0$). In Appendix C we show that the change of variables changes the derivatives to

$$C\partial_x = C\partial_{x'} - E\partial_{z'}, \quad C\partial_z = \partial_{z'},$$

and the lower layer Helmholtz equation becomes

$$0 = \operatorname{div}'[A\nabla' w'] + B \cdot \nabla' w' + 2C^2 i\alpha \partial_{x'} w' + C^2 (\gamma^{w'})^2 w', \tag{3.4}$$

where, for $S = C^2$,

$$A = \begin{pmatrix} S & -EC \\ -EC & 1 + E^2 \end{pmatrix}, \qquad B = (\partial_{x'} C) \begin{pmatrix} -C \\ E \end{pmatrix}.$$

We drop the primed variables so that (3.4) becomes

$$0 = \operatorname{div}[A\nabla w] + B \cdot \nabla w + 2Si\alpha \partial_x w + S(\gamma^w)^2 w,$$

We then take a boundary perturbation approach by setting

$$g(x) = \varepsilon f(x),\ \varepsilon \in \mathbb{R},\ \varepsilon \ll 1, \tag{3.5}$$

where, by following Appendix C, discover

$$\begin{aligned} A &= A(\varepsilon) = A_0 + A_1\varepsilon + A_2\varepsilon^2, \\ B &= B(\varepsilon) = B_1\varepsilon + B_2\varepsilon^2, \\ S &= S(\varepsilon) = S_0 + S_1\varepsilon + S_2\varepsilon^2. \end{aligned}$$

Since $b = \overline{h}$ and $u(x) = \varepsilon f(x)$, we find

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.6a}$$

$$A_1 = \begin{pmatrix} A_1^{xx} & A_1^{xz} \\ A_1^{zx} & A_1^{zz} \end{pmatrix} = \frac{1}{b} \begin{pmatrix} 2f & -(b+z)(\partial_x f) \\ -(b+z)(\partial_x f) & 0 \end{pmatrix}, \tag{3.6b}$$

$$A_2 = \begin{pmatrix} A_2^{xx} & A_2^{xz} \\ A_2^{zx} & A_2^{zz} \end{pmatrix} = \frac{1}{b^2} \begin{pmatrix} f^2 & -(b+z)f(\partial_x f) \\ -(b+z)f(\partial_x f) & (b+z)^2(\partial_x f)^2 \end{pmatrix}. \tag{3.6c}$$

Also

$$B_1 = \begin{pmatrix} B_1^x \\ B_1^z \end{pmatrix} = \frac{1}{b} \begin{pmatrix} -(\partial_x f) \\ 0 \end{pmatrix}, \tag{3.6d}$$

$$B_2 = \begin{pmatrix} B_2^x \\ B_2^z \end{pmatrix} = \frac{1}{b^2} \begin{pmatrix} -f(\partial_x f) \\ (b+z)(\partial_x f)^2 \end{pmatrix}, \tag{3.6e}$$

and

$$S_0 = 1, \quad S_1 = \frac{2}{b} f, \quad S_2 = \frac{1}{b^2} f^2. \tag{3.6f}$$

As a result, (3.3) becomes

$$\Delta w + 2i\alpha\partial_x w + \gamma^2 u = Y(x, z; g, w, \alpha, \gamma), \quad -b < z < 0, \tag{3.7}$$

where

$$\begin{aligned} Y(x, z; g, w, \alpha, \gamma) = &-\mathrm{div}[A_1\nabla w] - \mathrm{div}[A_2\nabla w] - B_1\nabla w - B_2\nabla w \\ &- 2S_1 i\alpha\partial_x w - S_1\gamma^2 w - 2S_2 i\alpha\partial_x w - S_2\gamma^2 w. \end{aligned} \tag{3.8}$$

By (3.2d) the Transparent Boundary Condition is

$$\partial_z\left[w(x, -b)\right] - T^w[w(x, -b)] = 0, \quad \text{at } z = -b. \tag{3.9}$$

For this boundary condition we begin with the lower boundary and recall that such boundaries are flat in the lower field, i.e., $\ell \equiv 0$. Therefore, we can multiply (3.9) by $C = C(x)$ to realize

$$C\partial_z\left[w(x, -b)\right] - CT^w[w(x, -b)] = 0.$$

So by the transformation rules for ∂_z and ∂_x (which induces the rule $T^w \to T^{w'}$ and $w \to w'$) with $\ell \equiv 0$ we find

$$\partial_{z'}\left[w'(x', -b)\right] - (1 + g(x')/\overline{h})T^{w'}[w'(x', -b)] = 0.$$

We rearrange to form

$$\partial_{z'}\left[w'(x', -b)\right] - T^{w'}[w'(x', -b)] = Q(x'; g, w'),$$

where

$$Q(x'; g, w') = \frac{1}{b}g(x')T^{w'}\left[w'(x', -b)\right].$$

We then drop the primed variables and write the boundary condition as

$$\partial_z\left[w(x, -b)\right] - T^w[w(x, -b)] = Q(x; g, w).$$

These changes transform the governing equations without phase in (3.2) to

$$\Delta w + 2i\alpha\partial_x w + (\gamma^w)^2 w = Y\left(x, z; g, w, \alpha, \gamma^w\right), \qquad -b < z < 0, \tag{3.10a}$$

$$w(x, 0) = \zeta^w(x), \qquad \text{at } z = 0, \tag{3.10b}$$

$$w(x + d, z) = w(x, z), \tag{3.10c}$$

$$\partial_z\left[w(x, -b)\right] - T^w[w(x, -b)] = Q(x; g, w), \qquad \text{at } z = -b. \tag{3.10d}$$

3.4 Frequency Perturbation

We now write the illumination frequency as

$$\omega = (1+\delta)\underline{\omega} = \underline{\omega} + \delta\underline{\omega},\ \delta \in \mathbb{R},\ \delta \ll 1, \tag{3.11}$$

where

$$k^w = \omega/c^w = (1+\delta)\underline{\omega}/c^w =: (1+\delta)\underline{k}^w = \underline{k}^w + \delta\underline{k}^w, \tag{3.12a}$$
$$\alpha = k^u \sin(\theta) = (1+\delta)\underline{k}^u \sin(\theta) =: (1+\delta)\underline{\alpha} = \underline{\alpha} + \delta\underline{\alpha}, \tag{3.12b}$$
$$\gamma^w = k^w \cos(\theta) = (1+\delta)\underline{k}^w \cos(\theta) =: (1+\delta)\underline{\gamma}^w = \underline{\gamma}^w + \delta\underline{\gamma}^w. \tag{3.12c}$$

These form the following relationship between the underscore variables

$$\underline{\alpha}^2 + (\underline{\gamma}^w)^2 = (\underline{k}^w)^2. \tag{3.13}$$

As the transformation rules between the upper and lower fields do not change (C.5), we may follow the analysis done in §2.5. It is not hard to show that (3.10a) becomes

$$\Delta w + 2i\underline{\alpha}\partial_x w + (\underline{\gamma}^w)^2 w = \tilde{Y}\left(x, z; g, w, \underline{\alpha}, \underline{\gamma}^w\right), \quad -b < z < 0, \tag{3.14}$$

where

$$\begin{aligned}\tilde{Y}\left(x, z; g, w, \underline{\alpha}, \underline{\gamma}^w\right) = &-\mathrm{div}[A_1\nabla w] - \mathrm{div}[A_2\nabla w] - B_1\nabla w - B_2\nabla w \\ &- 2i\underline{\alpha}\delta\partial_x w - \delta^2(\underline{\gamma}^w)^2 w - 2\delta(\underline{\gamma}^w)^2 w \\ &- 2S_1 i\underline{\alpha}\partial_x w - 2S_1 i\underline{\alpha}\delta\partial_x w - S_1\delta^2(\underline{\gamma}^w)^2 w - 2S_1\delta(\underline{\gamma}^w)^2 w - S_1(\underline{\gamma}^w)^2 w \\ &- 2S_2 i\underline{\alpha}\partial_x w - 2S_2 i\underline{\alpha}\delta\partial_x w - S_2\delta^2(\underline{\gamma}^w)^2 w - 2S_2\delta(\underline{\gamma}^w)^2 w - S_2(\underline{\gamma}^w)^2 w.\end{aligned}$$

The boundary condition (3.10d) becomes

$$\partial_z\left[w(x,-b)\right] - T_0^w[w(x,-b)] = \tilde{Q}(x; g, w),$$

where $T_0^w = i\underline{\gamma}_D^w$ corresponds to the case where $\delta = 0$ and

$$\tilde{Q}(x; g, w) = \frac{1}{b}(\varepsilon f(x))T^w\left[w(x,-b)\right] + \left(T^w - T_0^w\right)\left[w(x,-b)\right].$$

Proceeding in the same manner as in §2.5, our governing equations without phase and two small perturbations become

$$\Delta w + 2i\underline{\alpha}\partial_x w + (\underline{\gamma}^w)^2 w = \tilde{Y}\left(x, z; g, w, \underline{\alpha}, \underline{\gamma}^w\right), \qquad -b < z < 0, \tag{3.15a}$$

$$w(x, 0) = \zeta^w(x), \qquad \text{at } z = 0, \tag{3.15b}$$

$$w(x + d, z) = w(x, z), \tag{3.15c}$$

$$\partial_z\left[w(x, -b)\right] - T_0^w[w(x, -b)] = \tilde{Q}(x; g, w), \qquad \text{at } z = -b. \tag{3.15d}$$

3.5 Transformed Field Expansions

As in the upper field, we have made two smallness assumptions:

[1] Boundary Perturbation: $g(x) = \varepsilon f(x)$, $\varepsilon \in \mathbb{R}$, $\varepsilon \ll 1$,

[2] Frequency Perturbation: $\omega = (1 + \delta)\underline{\omega} = \underline{\omega} + \delta\underline{\omega}$, $\delta \in \mathbb{R}$, $\delta \ll 1$.

We now apply both of these assumptions and seek solutions of the form

$$w = w(x, z; \varepsilon, \delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} w_{n,m}(x, z)\varepsilon^n\delta^m, \tag{3.16}$$

which we will show are strongly convergent in Theorems 3.7.1 and 3.8.1. Inserting these into (3.15) produces the TFE recursions

$$\Delta w_{n,m} + 2i\underline{\alpha}\partial_x w_{n,m} + (\underline{\gamma}^w)^2 w_{n,m} = \tilde{Y}_{n,m}\left(x, z; f, w, \underline{\alpha}, \underline{\gamma}^w\right), \qquad -b < z < 0, \tag{3.17a}$$

$$w_{n,m}(x, 0) = \zeta^w_{n,m}(x), \qquad \text{at } z = 0, \tag{3.17b}$$

$$w_{n,m}(x + d, z) = w_{n,m}(x, z), \tag{3.17c}$$

$$\partial_z\left[w_{n,m}(x, -b)\right] - T_0^w[w_{n,m}(x, -b)] = \tilde{Q}_{n,m}(x), \qquad \text{at } z = -b, \tag{3.17d}$$

where

$$\begin{aligned}
\tilde{Y}_{n,m}\left(x, z; f, w, \underline{\alpha}, \underline{\gamma}^w\right) = &-\text{div}[A_1\nabla w_{n-1,m}] - \text{div}[A_2\nabla w_{n-2,m}] - B_1\nabla w_{n-1,m} \\
&- B_2\nabla w_{n-2,m} - 2i\underline{\alpha}\partial_x w_{n,m-1} - (\underline{\gamma}^w)^2 w_{n,m-2} \\
&- 2(\underline{\gamma}^w)^2 w_{n,m-1} - 2S_1 i\underline{\alpha}\partial_x w_{n-1,m} - 2S_1 i\underline{\alpha}\partial_x w_{n-1,m-1} \\
&- S_1(\underline{\gamma}^w)^2 w_{n-1,m-2} - 2S_1(\underline{\gamma}^w)^2 w_{n-1,m-1} - S_1(\underline{\gamma}^w)^2 w_{n-1,m} \\
&- 2S_2 i\underline{\alpha}\partial_x w_{n-2,m} - 2S_2 i\underline{\alpha}\partial_x w_{n-2,m-1} - S_2(\underline{\gamma}^w)^2 w_{n-2,m-2} \\
&- 2S_2(\underline{\gamma}^w)^2 w_{n-2,m-1} - S_2(\underline{\gamma}^w)^2 w_{n-2,m},
\end{aligned} \tag{3.18}$$

and

$$\tilde{Q}_{n,m}(x) = \frac{f}{b}\sum_{r=0}^{m} T^w_{m-r}[w_{n-1,r}(x, -b)] + \sum_{r=0}^{m-1} T^w_{m-r}[w_{n,r}(x, -b)]. \tag{3.19}$$

This is a method for computing the transformed corrections to the scattered field, $w_{n,m}$, with respect to both interfacial and frequency deformations. As stated in §2.6, a major advantage of the TFE recursions is that (3.17) never takes derivatives of $w_{n,m}$ higher than second order. To make use of this advantage, we will once again turn to classical elliptic theory.

3.6 Elliptic Theory

To prove the joint analyticity of the lower field, w, we require minor modifications to the Elliptic Estimate and Algebra Property of Sobolev spaces presented in §2.7.

Lemma 3.6.1. *Given an integer $s \geq 0$ and any $\sigma > 0$, there exists a constant $\mathcal{M} = \mathcal{M}(s)$ such that if $f \in C^s([0,d]), w \in H^s([0,d] \times [-b,0])$ then*

$$\|fw\|_{H^s} \leq \mathcal{M}|f|_{C^s}\|w\|_{H^s}, \tag{3.20}$$

and if $\tilde{f} \in C^{s+1/2+\sigma}([0,d]), \tilde{w} \in H^{s+1/2}([0,d])$ then there exists a constant $\tilde{\mathcal{M}} = \tilde{\mathcal{M}}(s)$ such that

$$\|\tilde{f}\tilde{w}\|_{H^{s+1/2}} \leq \tilde{\mathcal{M}}|\tilde{f}|_{C^{s+1/2+\sigma}}\|\tilde{w}\|_{H^{s+1/2}}. \tag{3.21}$$

Theorem 3.6.2. *Given an integer $s \geq 0$, if $Y \in H^s([0,d]) \times [-b,0])$, $\zeta^w \in H^{s+3/2}([0,d])$, $Q \in H^{s+1/2}([0,d])$, then there exists a unique solution of $w \in H^{s+2}([0,d] \times [-b,0])$ of*

$$\Delta w(x,z) + 2i\underline{\alpha}\partial_x w(x,z) + (\underline{\gamma}^w)^2 w(x,z) = Y(x,z), \qquad -b < z < 0, \tag{3.22a}$$

$$w(x,0) = \zeta^w(x,0), \qquad \text{at } z = 0, \tag{3.22b}$$

$$w(x+d,z) = w(x,z), \tag{3.22c}$$

$$\partial_z w(x,-b) - T_0^w[w(x,-b)] = Q(x), \qquad \text{at } z = -b, \tag{3.22d}$$

satisfying

$$\|w\|_{H^{s+2}} \leq C_e\{\|Y\|_{H^s} + \|\zeta^w\|_{H^{s+3/2}} + \|Q\|_{H^{s+1/2}}\}, \tag{3.23}$$

for some constant $C_e = C_e(s) > 0$.

Lemma 3.6.3. *Given an integer $s \geq 0$, if $Y \in H^s([0,d]) \times [-b,0])$, then $(b+z)Y \in H^s([0,d]) \times [-b,0])$ and there exists a positive constant $Z_b = Z_b(s)$ such that*

$$\|(b+z)Y\|_{H^s} \leq Z_b\|Y\|_{H^s}.$$

With these, we now have everything we need to prove our desired result on the analyticity of the lower transformed field $w = w(x,z;\varepsilon)$ with respect to the single perturbation parameter ε.

3.7 Analyticity of the Boundary Perturbation

Theorem 3.7.1. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and $\zeta_{n,0}^w \in H^{s+3/2}([0,d])$ such that*

$$\|\zeta_{n,0}^w\|_{H^{s+3/2}} \leq K_\zeta B_\zeta^n, \tag{3.24}$$

for constants $K_\zeta, B_\zeta > 0$, then $w_{n,0} \in H^{s+2}([0,d] \times [-b,0])$ and

$$\|w_{n,0}\|_{H^{s+2}} \leq K B^n, \tag{3.25}$$

for constants $K, B > 0$.

To establish this result we work by induction. The key estimate is encapsulated in the following lemma.

Lemma 3.7.2. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and*

$$\|w_{n,0}\|_{H^{s+2}} \leq K B^n, \quad \forall n < \overline{n}, \tag{3.26}$$

for constants $K, B > 0$, then there exists a constant $\overline{C} > 0$ such that

$$\max\left\{\|\tilde{Y}_{\overline{n},0}\|_{H^s}, \|\tilde{Q}_{\overline{n},0}\|_{H^{s+1/2}}\right\} \leq K\overline{C}\Big\{|f|_{C^{s+2}} B^{\overline{n}-1} + |f|_{C^{s+2}}^2 B^{\overline{n}-2}\Big\}. \tag{3.27}$$

Proof. [Lemma 3.7.2] We begin with $\tilde{Y}_{\overline{n},0}$ and recall from (3.18) that

$$\begin{aligned}\tilde{Y}_{\overline{n},0}\left(x,z;f,w,\underline{\alpha},\underline{\gamma}^w\right) = &-\mathrm{div}[A_1 \nabla w_{\overline{n}-1,0}] - \mathrm{div}[A_2 \nabla w_{\overline{n}-2,0}] - B_1 \nabla w_{\overline{n}-1,0} \\ &- B_2 \nabla w_{\overline{n}-2,0} - 2S_1 i\underline{\alpha}\partial_x w_{\overline{n}-1,0} - S_1(\underline{\gamma}^w)^2 w_{\overline{n}-1,0} \\ &- 2S_2 i\underline{\alpha}\partial_x w_{\overline{n}-2,0} - S_2(\underline{\gamma}^w)^2 w_{\overline{n}-2,0}.\end{aligned} \tag{3.28}$$

Then from (3.6) we have

$$\begin{aligned}\|\tilde{Y}_{\overline{n},0}\|_{H^s}^2 \leq &\|A_1^{xx}\partial_x w_{\overline{n}-1,0}\|_{H^{s+1}}^2 + \|A_1^{xz}\partial_z w_{\overline{n}-1,0}\|_{H^{s+1}}^2 + \|A_1^{zx}\partial_x w_{\overline{n}-1,0}\|_{H^{s+1}}^2 \\ &+ \|A_1^{zz}\partial_z w_{\overline{n}-1,0}\|_{H^{s+1}}^2 + \|A_2^{xx}\partial_x w_{\overline{n}-2,0}\|_{H^{s+1}}^2 + \|A_2^{xz}\partial_z w_{\overline{n}-2,0}\|_{H^{s+1}}^2 \\ &+ \|A_2^{zx}\partial_x w_{\overline{n}-2,0}\|_{H^{s+1}}^2 + \|A_2^{zz}\partial_z w_{\overline{n}-2,0}\|_{H^{s+1}}^2 + \|B_1^x\partial_x w_{\overline{n}-1,0}\|_{H^s}^2 \\ &+ \|B_1^z\partial_z w_{\overline{n}-1,0}\|_{H^s}^2 + \|B_2^x\partial_x w_{\overline{n}-2,0}\|_{H^s}^2 + \|B_2^z\partial_z w_{\overline{n}-2,0}\|_{H^s}^2 \\ &+ \|2S_1 i\underline{\alpha}\partial_x w_{\overline{n}-1,0}\|_{H^s}^2 + \|S_1(\underline{\gamma}^w)^2 w_{\overline{n}-1,0}\|_{H^s}^2 + \|2S_2 i\underline{\alpha}\partial_x w_{\overline{n}-2,0}\|_{H^s}^2 \\ &+ \|S_2(\underline{\gamma}^w)^2 w_{\overline{n}-2,0}\|_{H^s}^2.\end{aligned}$$

We now estimate each of these and apply Lemmas 2.8.1 (with $u = w$), 3.6.1, and 3.6.3. We begin with

$$\begin{aligned}\|A_1^{xx}\partial_x w_{\overline{n}-1,0}\|_{H^{s+1}} &= \|(2/b)f\partial_x w_{\overline{n}-1,0}\|_{H^{s+1}} \\ &\leq (2/b)\mathcal{M}|f|_{C^{s+1}}\|w_{\overline{n}-1,0}\|_{H^{s+2}} \\ &\leq (2/b)\mathcal{M}|f|_{C^{s+1}}KB^{\overline{n}-1},\end{aligned}$$

and in a similar fashion

$$\begin{aligned}\|A_1^{xz}\partial_z w_{\overline{n}-1,0}\|_{H^{s+1}} &= \|-((b+z)/b)(\partial_x f)\partial_z w_{\overline{n}-1,0}\|_{H^{s+1}} \\ &\leq (Z_b/b)\mathcal{M}|\partial_x f|_{C^{s+1}}\|w_{\overline{n}-1,0}\|_{H^{s+2}} \\ &\leq (Z_b/b)\mathcal{M}|f|_{C^{s+2}}KB^{\overline{n}-1}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_1^{zx}\partial_x w_{\overline{n}-1,0}\|_{H^{s+1}} &= \|-((b+z)/b)(\partial_x f)\partial_x w_{\overline{n}-1,0}\|_{H^{s+1}} \\ &\leq (Z_b/b)\mathcal{M}|\partial_x f|_{C^{s+1}}\|w_{\overline{n}-1,0}\|_{H^{s+2}} \\ &\leq (Z_b/b)\mathcal{M}|f|_{C^{s+2}}KB^{\overline{n}-1},\end{aligned}$$

and we recall that $A_1^{zz} \equiv 0$. Moving to the second order

$$\begin{aligned}\|A_2^{xx}\partial_x w_{\overline{n}-2,0}\|_{H^{s+1}} &= \|(1/b^2)f^2\partial_x w_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\leq (1/b^2)\mathcal{M}^2|f|^2_{C^{s+1}}\|w_{\overline{n}-2,0}\|_{H^{s+2}} \\ &\leq (1/b^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{\overline{n}-2}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_2^{xz}\partial_z w_{\overline{n}-2,0}\|_{H^{s+1}} &= \|(-(b+z)/b^2)f(\partial_x f)\partial_x w_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\leq (Z_b/b^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|w_{\overline{n}-2,0}\|_{H^{s+2}} \\ &\leq (Z_b/b^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zx}\partial_x w_{\overline{n}-2,0}\|_{H^{s+1}} &= \|(-(b+z)/b^2)f(\partial_x f)\partial_z w_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\leq (Z_b/b^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|w_{\overline{n}-2,0}\|_{H^{s+2}} \\ &\leq (Z_b/b^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zz}\partial_z w_{\overline{n}-2,0}\|_{H^{s+1}} &= \|((b+z)^2/b^2)(\partial_x f)^2\partial_z w_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (Z_b^2/b^2)\mathcal{M}^2|\partial_x f|^2_{C^{s+1}}\|w_{\overline{n}-2,0}\|_{H^{s+2}}\\ &\leq (Z_b^2/b^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{\overline{n}-2}.\end{aligned}$$

Next for the B_1 terms

$$\begin{aligned}\|B_1^x\partial_x w_{\overline{n}-1,0}\|_{H^s} &= \|(-1/b)(\partial_x f)\partial_x w_{\overline{n}-1,0}\|_{H^s}\\ &\leq (1/b)\mathcal{M}|\partial_x f|_{C^s}\|w_{\overline{n}-1,0}\|_{H^{s+1}}\\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1}}KB^{\overline{n}-1},\end{aligned}$$

and $B_1^z \equiv 0$. Moving to the second order

$$\begin{aligned}\|B_2^x\partial_x w_{\overline{n}-2,0}\|_{H^s} &= \|(-1/b^2)f(\partial_x f)\partial_x w_{\overline{n}-2,0}\|_{H^s}\\ &\leq (1/b^2)\mathcal{M}^2|f|_{C^s}|\partial_x f|_{C^s}\|w_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (1/b^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|B_2^z\partial_z w_{\overline{n}-2,0}\|_{H^s} &= \|(1/b^2)(b+z)(\partial_x f)^2\partial_z w_{\overline{n}-2,0}\|_{H^s}\\ &\leq (Z_b/b^2)\mathcal{M}^2|\partial_x f|^2_{C^s}\|w_{\overline{n}-2,0}\|_{H^{s+1}}\\ &\leq (Z_b/b^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{\overline{n}-2}.\end{aligned}$$

To address the S_0, S_1, S_2 terms we have

$$\begin{aligned}\|2S_1 i\underline{\alpha}\partial_x w_{\overline{n}-1,0}\|_{H^s} &= \|(4/b)i\underline{\alpha}f\partial_x w_{\overline{n}-1,0}\|_{H^s}\\ &\leq (4/b)\underline{\alpha}\mathcal{M}|f|_{C^s}\|w_{\overline{n}-1,0}\|_{H^{s+1}}\\ &\leq (4/b)\underline{\alpha}\mathcal{M}|f|_{C^s}KB^{\overline{n}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_1(\underline{\gamma}^w)^2 w_{\overline{n}-1,0}\|_{H^s} &= \|(2/b)(\underline{\gamma}^w)^2 f w_{\overline{n}-1,0}\|_{H^s}\\ &\leq (2/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}\|w_{\overline{n}-1,0}\|_{H^s}\\ &\leq (2/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}KB^{\overline{n}-1},\end{aligned}$$

and

$$\begin{aligned}\|2S_2 i\underline{\alpha}\partial_x w_{\overline{n}-2,0}\|_{H^s} &= \|(2/b^2) i\underline{\alpha} f^2 \partial_x w_{\overline{n}-2,0}\|_{H^s} \\ &\leq (2/b^2)\underline{\alpha}\mathcal{M}^2 |f|^2_{C^s} \|w_{\overline{n}-2,0}\|_{H^{s+1}} \\ &\leq (2/b^2)\underline{\alpha}\mathcal{M}^2 |f|^2_{C^s} K B^{\overline{n}-2},\end{aligned}$$

and

$$\begin{aligned}\|S_2(\underline{\gamma}^w)^2 w_{\overline{n}-2,0}\|_{H^s} &= \|(1/b^2)(\underline{\gamma}^w)^2 f^2 w_{\overline{n}-2,0}\|_{H^s} \\ &\leq (1/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} \|w_{\overline{n}-2,0}\|_{H^s} \\ &\leq (1/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{\overline{n}-2}.\end{aligned}$$

We satisfy the estimate for $\|\tilde{Y}_{\overline{n},0}\|_{H^s}$ provided that we choose

$$\overline{C} > \max\left\{ \left(\frac{3 + 2Z_b + 4\underline{\alpha} + 2(\underline{\gamma}^w)^2}{b} \right) \mathcal{M}, \left(\frac{2 + 3Z_b + Z_b^2 + 2\underline{\alpha} + (\underline{\gamma}^w)^2}{b^2} \right) \mathcal{M}^2 \right\}.$$

The estimate for $\tilde{Q}_{\overline{n},0}$ follows from Lemma 2.8.2

$$\begin{aligned}\|\tilde{Q}_{\overline{n},0}\|_{H^{s+1/2}} &= \|(1/b) f T_0^w [w_{\overline{n}-1,0}]\|_{H^{s+1/2}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\sigma}} \|T_0^w [w_{\overline{n}-1,0}]\|_{H^{s+1/2}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\sigma}} C_{T_0^w} \|w_{\overline{n}-1,0}\|_{H^{s+3/2}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\sigma}} C_{T_0^w} K B^{\overline{n}-1},\end{aligned}$$

and provided that

$$\overline{C} > (1/b)\mathcal{M} C_{T_0^w},$$

we are done. □

With this information, we can now prove Theorem 3.7.1.

Proof. [Theorem 3.7.1] We proceed by induction in n. At order $n = m = 0$ (3.17) becomes

$$\Delta w_{0,0} + 2i\underline{\alpha}\partial_x w_{0,0} + (\underline{\gamma}^w)^2 w_{0,0} = 0, \qquad -b < z < 0, \tag{3.29a}$$

$$w_{0,0}(x, g) = \zeta^w_{0,0}(x), \qquad \text{at } z = 0, \tag{3.29b}$$

$$w_{0,0}(x + d, z) = w_{0,0}(x, z), \tag{3.29c}$$

$$\partial_z [w_{0,0}(x, -b)] - T_0^w [w_{0,0}(x, -b)] = 0, \qquad \text{at } z = -b, \tag{3.29d}$$

and Theorem 3.6.2 guarantees a unique solution such that

$$\|w_{0,0}\|_{H^{s+2}} \leq C_e \|\zeta^w_{0,0}\|_{H^{s+3/2}}.$$

So we choose $K \geq C_e \|\zeta^w_{0,0}\|_{H^{s+3/2}}$. We now assume the estimate (3.25) for all $n < \overline{n}$ and study $w_{\overline{n},0}$. From Theorem 3.6.2 we have a unique solution satisfying

$$\|w_{\overline{n},0}\|_{H^{s+2}} \leq C_e \{ \|\tilde{Y}_{\overline{n},0}\|_{H^s} + \|\zeta^w_{\overline{n},0}\|_{H^{s+3/2}} + \|\tilde{Q}_{\overline{n},0}\|_{H^{s+1/2}} \},$$

and appealing to Lemmas 2.8.3 (with $\zeta^u = \zeta^w$ and the hypothesis (3.24)) and 3.7.2 we find

$$\|w_{\overline{n},0}\|_{H^{s+2}} \leq C_e \left\{ K_\zeta B_\zeta^{\overline{n}} + 2K\overline{C} \Big[|f|_{C^{s+2}} B^{\overline{n}-1} + |f|^2_{C^{s+2}} B^{\overline{n}-2} \Big] \right\}.$$

We are done provided we choose $K \geq 3C_e K_\zeta$ and

$$B > \max \Big\{ B_\zeta, 6C_e\overline{C}|f|_{C^{s+2}}, \sqrt{6C_e\overline{C}}|f|_{C^{s+2}} \Big\}.$$

□

We can now establish the joint analyticity of the transformed field $w = w(x, z; \varepsilon, \delta)$ with respect to the perturbation parameters ε and δ.

3.8 Joint Analyticity of the Lower Field

Theorem 3.8.1. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and $\zeta^w_{n,m} \in H^{s+3/2}([0,d])$ such that*

$$\|\zeta^w_{n,m}\|_{H^{s+3/2}} \leq K_\zeta B_\zeta^n D_\zeta^m, \tag{3.30}$$

for constants $K_\zeta, B_\zeta, D_\zeta > 0$, then $w_{n,m} \in H^{s+2}([0,d] \times [-b,0])$ and

$$\|w_{n,m}\|_{H^{s+2}} \leq K B^n D^m, \tag{3.31}$$

for constants $K, B, D > 0$.

To establish this result we work by induction. The key estimate is encapsulated in the following lemma.

Lemma 3.8.2. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and*

$$\|w_{n,m}\|_{H^{s+2}} \leq K B^n D^m, \quad \forall\ n \geq 0,\ m < \overline{m}, \tag{3.32}$$

for constants $K, B, D > 0$ then there exists a constant $\overline{C} > 0$ such that

$$\begin{aligned}\max\left\{\|\tilde{Y}_{n,\overline{m}}\|_{H^s}, \|\tilde{Q}_{n,\overline{m}}\|_{H^{s+1/2}}\right\} \leq K\overline{C}\Big\{ & B^n D^{\overline{m}-1} + B^n D^{\overline{m}-2} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}} + \\ & |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-1} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-2} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}} + \\ & |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-1} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-2}\Big\}.\end{aligned}$$

Proof. [Lemma 3.8.2] We begin with $\tilde{Y}_{n,\overline{m}}$ and recall from (3.18) that

$$\begin{aligned}\tilde{Y}_{n,\overline{m}}\left(x, z; g, w, \underline{\alpha}, \underline{\gamma}^w\right) = & -\mathrm{div}[A_1 \nabla w_{n-1,\overline{m}}] - \mathrm{div}[A_2 \nabla w_{n-2,\overline{m}}] - B_1 \nabla w_{n-1,\overline{m}} \\ & - B_2 \nabla w_{n-2,\overline{m}} - 2i\underline{\alpha}\partial_x w_{n,\overline{m}-1} - (\underline{\gamma}^w)^2 w_{n,\overline{m}-2} \\ & - 2(\underline{\gamma}^w)^2 w_{n,\overline{m}-1} - 2S_1 i\underline{\alpha}\partial_x w_{n-1,\overline{m}} - 2S_1 i\underline{\alpha}\partial_x w_{n-1,\overline{m}-1} \\ & - S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}-2} - 2S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}-1} - S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}} \\ & - 2S_2 i\underline{\alpha}\partial_x w_{n-2,\overline{m}} - 2S_2 i\underline{\alpha}\partial_x w_{n-2,\overline{m}-1} - S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}-2} \\ & - 2S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}-1} - S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}}.\end{aligned} \tag{3.33}$$

Then from (3.6) we have

$$\begin{aligned}\|\tilde{Y}_{n,\overline{m}}\|^2_{H^s} \leq & \|A_1^{xx}\partial_x w_{n-1,\overline{m}}\|^2_{H^{s+1}} + \|A_1^{xz}\partial_z w_{n-1,\overline{m}}\|^2_{H^{s+1}} + \|A_1^{zx}\partial_x w_{n-1,\overline{m}}\|^2_{H^{s+1}} \\ & + \|A_1^{zz}\partial_z w_{n-1,\overline{m}}\|^2_{H^{s+1}} + \|A_2^{xx}\partial_x w_{n-2,\overline{m}}\|^2_{H^{s+1}} + \|A_2^{xz}\partial_z w_{n-2,\overline{m}}\|^2_{H^{s+1}} \\ & + \|A_2^{zx}\partial_x w_{n-2,\overline{m}}\|^2_{H^{s+1}} + \|A_2^{zz}\partial_z w_{n-2,\overline{m}}\|^2_{H^{s+1}} + \|B_1^x\partial_x w_{n-1,\overline{m}}\|^2_{H^s} \\ & + \|B_1^z\partial_z w_{n-1,\overline{m}}\|^2_{H^s} + \|B_2^x\partial_x w_{n-2,\overline{m}}\|^2_{H^s} + \|B_2^z\partial_z w_{n-2,\overline{m}}\|^2_{H^s} \\ & + \|2i\underline{\alpha}\partial_x w_{n,\overline{m}-1}\|^2_{H^s} + \|(\underline{\gamma}^w)^2 w_{n,\overline{m}-2}\|^2_{H^s} + \|2(\underline{\gamma}^w)^2 w_{n,\overline{m}-1}\|^2_{H^s} \\ & + \|2S_1 i\underline{\alpha}\partial_x w_{n-1,\overline{m}}\|^2_{H^s} + \|2S_1 i\underline{\alpha}\partial_x w_{n-1,\overline{m}-1}\|^2_{H^s} + \|S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}-2}\|^2_{H^s} \\ & + \|2S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}-1}\|^2_{H^s} + \|S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}}\|^2_{H^s} + \|2S_2 i\underline{\alpha}\partial_x w_{n-2,\overline{m}}\|^2_{H^s} \\ & + \|2S_2 i\underline{\alpha}\partial_x w_{n-2,\overline{m}-1}\|^2_{H^s} + \|S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}-2}\|^2_{H^s} + \|2S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}-1}\|^2_{H^s} \\ & + \|S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}}\|^2_{H^s}.\end{aligned}$$

We now estimate each of these and apply Lemmas 2.8.1 (with $u = w$), 3.6.1, and 3.6.3, beginning with

$$\begin{aligned}\|A_1^{xx}\partial_x w_{n-1,\overline{m}}\|_{H^{s+1}} &= \|(2/b) f \partial_x w_{n-1,\overline{m}}\|_{H^{s+1}} \\ &\leq (2/b)\mathcal{M}|f|_{C^{s+1}}\|w_{n-1,\overline{m}}\|_{H^{s+2}} \\ &\leq (2/b)\mathcal{M}|f|_{C^{s+1}} K B^{n-1} D^{\overline{m}},\end{aligned}$$

and in a similar fashion

$$\begin{aligned}\|A_1^{xz}\partial_z w_{n-1,\overline{m}}\|_{H^{s+1}} &= \|-((b+z)/b)(\partial_x f)\partial_z w_{n-1,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_b/b)\mathcal{M}|\partial_x f|_{C^{s+1}}\|w_{n-1,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_b/b)\mathcal{M}|f|_{C^{s+2}}KB^{n-1}D^{\overline{m}}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_1^{zx}\partial_x w_{n-1,\overline{m}}\|_{H^{s+1}} &= \|-((b+z)/b)(\partial_x f)\partial_x w_{n-1,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_b/b)\mathcal{M}|\partial_x f|_{C^{s+1}}\|w_{n-1,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_b/b)\mathcal{M}|f|_{C^{s+2}}KB^{n-1}D^{\overline{m}},\end{aligned}$$

and we recall that $A_1^{zz}\equiv 0$. Moving to the second order

$$\begin{aligned}\|A_2^{xx}\partial_x w_{n-2,\overline{m}}\|_{H^{s+1}} &= \|(1/b^2)f^2\partial_x w_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (1/b^2)\mathcal{M}^2|f|^2_{C^{s+1}}\|w_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (1/b^2)\mathcal{M}^2|f|^2_{C^{s+1}}KB^{n-2}D^{\overline{m}}.\end{aligned}$$

Also,

$$\begin{aligned}\|A_2^{xz}\partial_z w_{n-2,\overline{m}}\|_{H^{s+1}} &= \|(-(b+z)/b^2)f(\partial_x f)\partial_x w_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_b/b^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|w_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_b/b^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{n-2}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zx}\partial_x w_{n-2,\overline{m}}\|_{H^{s+1}} &= \|(-(b+z)/b^2)f(\partial_x f)\partial_z w_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_b/b^2)\mathcal{M}^2|f|_{C^{s+1}}|\partial_x f|_{C^{s+1}}\|w_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_b/b^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{n-2}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|A_2^{zz}\partial_z w_{n-2,\overline{m}}\|_{H^{s+1}} &= \|((b+z)^2/b^2)(\partial_x f)^2\partial_z w_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\leq (Z_b^2/b^2)\mathcal{M}^2|\partial_x f|^2_{C^{s+1}}\|w_{n-2,\overline{m}}\|_{H^{s+2}}\\ &\leq (Z_b^2/b^2)\mathcal{M}^2|f|^2_{C^{s+2}}KB^{n-2}D^{\overline{m}}.\end{aligned}$$

Next for the B_1 terms

$$\begin{aligned}\|B_1^x \partial_x w_{n-1,\overline{m}}\|_{H^s} &= \|(-1/b)(\partial_x f)\partial_x w_{n-1,\overline{m}}\|_{H^s} \\ &\leq (1/b)\mathcal{M}|\partial_x f|_{C^s}\|w_{n-1,\overline{m}}\|_{H^{s+1}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1}} K B^{n-1} D^{\overline{m}},\end{aligned}$$

and $B_1^z \equiv 0$. Moving to the second order

$$\begin{aligned}\|B_2^x \partial_x w_{n-2,\overline{m}}\|_{H^s} &= \|(-1/b^2) f(\partial_x f)\partial_x w_{n-2,\overline{m}}\|_{H^s} \\ &\leq (1/b^2)\mathcal{M}^2|f|_{C^s}|\partial_x f|_{C^s}\|w_{n-2,\overline{m}}\|_{H^{s+1}} \\ &\leq (1/b^2)\mathcal{M}^2|f|^2_{C^{s+1}} K B^{n-2} D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|B_2^z \partial_z w_{n-2,\overline{m}}\|_{H^s} &= \|(1/b^2)(b+z)(\partial_x f)^2\partial_z w_{n-2,\overline{m}}\|_{H^s} \\ &\leq (Z_b/b^2)\mathcal{M}^2|\partial_x f|^2_{C^s}\|w_{n-2,\overline{m}}\|_{H^{s+1}} \\ &\leq (Z_b/b^2)\mathcal{M}^2|f|^2_{C^{s+1}} K B^{n-2} D^{\overline{m}}.\end{aligned}$$

To address the S_0, S_1, S_2 terms we have

$$\begin{aligned}\|2i\underline{\alpha}\partial_x w_{n,\overline{m}-1}\|_{H^s} &\leq 2\underline{\alpha}\|w_{n,\overline{m}-1}\|_{H^{s+1}} \\ &\leq 2\underline{\alpha} K B^n D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|(\underline{\gamma}^w)^2 w_{n,\overline{m}-2}\|_{H^s} &\leq (\underline{\gamma}^w)^2\|w_{n,\overline{m}-2}\|_{H^s} \\ &\leq (\underline{\gamma}^w)^2 K B^n D^{\overline{m}-2},\end{aligned}$$

and

$$\begin{aligned}\|2(\underline{\gamma}^w)^2 w_{n,\overline{m}-1}\|_{H^s} &\leq 2(\underline{\gamma}^w)^2\|w_{n,\overline{m}-1}\|_{H^s} \\ &\leq 2(\underline{\gamma}^w)^2 K B^n D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|2S_1 i\underline{\alpha}\partial_x w_{n-1,\overline{m}}\|_{H^s} &= \|(4/b)i\underline{\alpha} f\partial_x w_{n-1,\overline{m}}\|_{H^s} \\ &\leq (4/b)\underline{\alpha}\mathcal{M}|f|_{C^s}\|w_{n-1,\overline{m}}\|_{H^{s+1}} \\ &\leq (4/b)\underline{\alpha}\mathcal{M}|f|_{C^s} K B^{n-1} D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|2S_1 i\underline{\alpha}\partial_x w_{n-1,\overline{m}-1}\|_{H^s} &= \|(4/b)i\underline{\alpha}f\partial_x w_{n-1,\overline{m}-1}\|_{H^s}\\ &\le (4/b)\underline{\alpha}\mathcal{M}|f|_{C^s}\|w_{n-1,\overline{m}-1}\|_{H^{s+1}}\\ &\le (4/b)\underline{\alpha}\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}-2}\|_{H^s} &= \|(2/b)(\underline{\gamma}^w)^2 f w_{n-1,\overline{m}-2}\|_{H^s}\\ &\le (2/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}\|w_{n-1,\overline{m}-2}\|_{H^s}\\ &\le (2/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}-2},\end{aligned}$$

and

$$\begin{aligned}\|2S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}-1}\|_{H^s} &= \|(4/b)(\underline{\gamma}^w)^2 f w_{n-1,\overline{m}-1}\|_{H^s}\\ &\le (4/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}\|w_{n-1,\overline{m}-1}\|_{H^s}\\ &\le (4/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_1(\underline{\gamma}^w)^2 w_{n-1,\overline{m}}\|_{H^s} &= \|(2/b)(\underline{\gamma}^w)^2 f w_{n-1,\overline{m}}\|_{H^s}\\ &\le (2/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}\|w_{n-1,\overline{m}}\|_{H^s}\\ &\le (2/b)(\underline{\gamma}^w)^2\mathcal{M}|f|_{C^s}KB^{n-1}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|2S_2 i\underline{\alpha}\partial_x w_{n-2,\overline{m}}\|_{H^s} &= \|(2/b^2)i\underline{\alpha}f^2\partial_x w_{n-2,\overline{m}}\|_{H^s}\\ &\le (2/b^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}\|w_{n-2,\overline{m}}\|_{H^{s+1}}\\ &\le (2/b^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}KB^{n-2}D^{\overline{m}},\end{aligned}$$

and

$$\begin{aligned}\|2S_2 i\underline{\alpha}\partial_x w_{n-2,\overline{m}-1}\|_{H^s} &= \|(2/b^2)i\underline{\alpha}f^2\partial_x w_{n-2,\overline{m}-1}\|_{H^s}\\ &\le (2/b^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}\|w_{n-2,\overline{m}-1}\|_{H^{s+1}}\\ &\le (2/b^2)\underline{\alpha}\mathcal{M}^2|f|^2_{C^s}KB^{n-2}D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}-2}\|_{H^s} &= \|(1/b^2)(\underline{\gamma}^w)^2 f^2 w_{n-2,\overline{m}-2}\|_{H^s} \\ &\leq (1/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} \|w_{n-2,\overline{m}-2}\|_{H^s} \\ &\leq (1/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{n-2} D^{\overline{m}-2},\end{aligned}$$

and

$$\begin{aligned}\|2S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}-1}\|_{H^s} &= \|(2/b^2)(\underline{\gamma}^w)^2 f^2 w_{n-2,\overline{m}-1}\|_{H^s} \\ &\leq (2/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} \|w_{n-2,\overline{m}-1}\|_{H^s} \\ &\leq (2/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{n-2} D^{\overline{m}-1},\end{aligned}$$

and

$$\begin{aligned}\|S_2(\underline{\gamma}^w)^2 w_{n-2,\overline{m}}\|_{H^s} &= \|(1/b^2)(\underline{\gamma}^w)^2 f^2 w_{n-2,\overline{m}}\|_{H^s} \\ &\leq (1/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} \|w_{n-2,\overline{m}}\|_{H^s} \\ &\leq (1/b^2)(\underline{\gamma}^w)^2 \mathcal{M}^2 |f|^2_{C^s} K B^{n-2} D^{\overline{m}}.\end{aligned}$$

We satisfy the estimate for $\|\tilde{Y}_{n,\overline{m}}\|_{H^s}$ provided that we choose

$$\begin{aligned}\overline{C} > \max\Bigg\{ &\left(2\underline{\alpha} + 3(\underline{\gamma}^w)^2\right), \left(\frac{3 + 2Z_b + 8\underline{\alpha} + 8(\underline{\gamma}^w)^2}{b}\right)\mathcal{M}, \\ &\left(\frac{2 + 3Z_b + Z_b^2 + 4\underline{\alpha} + 4(\underline{\gamma}^w)^2}{b^2}\right)\mathcal{M}^2\Bigg\}.\end{aligned}$$

The estimate for $\tilde{Q}_{n,\overline{m}}$ follows from Lemma 2.8.2

$$\begin{aligned}\left\|\tilde{Q}_{n,\overline{m}}\right\|_{H^{s+1/2}} &= \left\|\frac{1}{b} f(x) \sum_{r=0}^{\overline{m}} T^w_{\overline{m}-r}\left[w_{n-1,r}\right] + \sum_{r=0}^{\overline{m}-1} T^w_{\overline{m}-r}\left[w_{n,r}\right]\right\|_{H^{s+1/2}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\eta}} \sum_{r=0}^{\overline{m}} \left\|T^w_{\overline{m}-r}\left[w_{n-1,r}\right]\right\|_{H^{s+1/2}} + \sum_{r=0}^{\overline{m}-1} \left\|T^w_{\overline{m}-r}\left[w_{n,r}\right]\right\|_{H^{s+1/2}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\eta}} C_{T^w} \sum_{r=0}^{\overline{m}} \|w_{n-1,r}\|_{H^{s+3/2}} + C_{T^w} \sum_{r=0}^{\overline{m}-1} \|w_{n,r}\|_{H^{s+3/2}} \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\eta}} C_{T^w} K B^{n-1} \left(\frac{D^{\overline{m}+1} - 1}{D-1}\right) + C_{T^w} K B^n \left(\frac{D^{\overline{m}} - 1}{D-1}\right) \\ &\leq (1/b)\mathcal{M}|f|_{C^{s+1/2+\eta}} C_{T^w} K B^{n-1} D^{\overline{m}} \left(\frac{D}{D-1}\right) + C_{T^w} K B^n D^{\overline{m}-1} \left(\frac{D}{D-1}\right),\end{aligned}$$

and we are done provided that $D > 2$ and

$$\overline{C} > \max\left\{(1/b)\mathcal{M} C_{T^w}, C_{T^w}\right\}. \qquad \square$$

With this information, we can now prove Theorem 3.8.1.

Proof. [Theorem 3.8.1] We proceed by induction in m. At order $m = 0$ (3.17) becomes

$$\Delta w_{n,0} + 2i\underline{\alpha}\partial_x w_{n,0} + (\underline{\gamma}^w)^2 w_{n,0} = \tilde{Y}_{n,0}\left(x, z; f, w, \underline{\alpha}, \underline{\gamma}^w\right), \qquad -b < z < 0, \tag{3.34a}$$

$$w_{n,0}(x, g) = \zeta^w_{n,0}(x), \qquad \text{at } z = 0 \tag{3.34b}$$

$$w_{n,0}(x + d, z) = w_{n,0}(x, z), \tag{3.34c}$$

$$\partial_z\left[w_{n,0}(x, -b)\right] - T_0^w[w_{n,0}(x, -b)] = \tilde{Q}_{n,0}(x), \qquad \text{at } z = -b, \tag{3.34d}$$

and Theorem 3.7.1 guarantees a unique solution such that

$$\|w_{n,0}\|_{H^{s+2}} \leq KB^n, \quad \forall n \geq 0.$$

We now assume the estimate (3.31) for all $n, m < \overline{m}$ and study $w_{n,\overline{m}}$. From Theorem 3.6.2 we have a unique solution satisfying

$$\|w_{n,\overline{m}}\|_{H^{s+2}} \leq C_e\{\|\tilde{Y}_{n,\overline{m}}\|_{H^s} + \|\zeta^w_{n,\overline{m}}\|_{H^{s+3/2}} + \|\tilde{Q}_{n,\overline{m}}\|_{H^{s+1/2}}\},$$

and appealing to Lemmas 2.9.1 (with $\zeta^u = \zeta^w$ and the hypothesis (3.30)) and 3.8.2 we find

$$\begin{aligned}\|w_{n,\overline{m}}\|_{H^{s+2}} \leq C_e\Bigg\{ & K_\zeta B_\zeta^n D_\zeta^{\overline{m}} + 2K\overline{C}\Bigg(B^n D^{\overline{m}-1} + B^n D^{\overline{m}-2} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}} + \\ & |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-1} + |f|_{C^{s+2}} B^{n-1} D^{\overline{m}-2} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}} + \\ & |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-1} + |f|^2_{C^{s+2}} B^{n-2} D^{\overline{m}-2}\Bigg)\Bigg\}.\end{aligned}$$

We are done provided we choose $K \geq 9C_e K_\zeta$ and

$$\begin{aligned} B &> \max\left\{B_\zeta, 18C_e\overline{C}|f|_{C^{s+2}}, \sqrt{18C_e\overline{C}}|f|_{C^{s+2}}\right\}, \\ D &> \max\left\{1, D_\zeta, 18C_e\overline{C}, \sqrt{18C_e\overline{C}}\right\}. \end{aligned}$$

□

3.9 Analyticity of the Lower Layer DNO

Now that we have established the analyticity of the field, $w = w(x, z; \varepsilon, \delta)$, we move on to establishing the analyticity of the DNO, $J(g) = J(\varepsilon f)$. As in the upper field, we apply an unnormalized normal, $N = (-\partial_x g, 1)^T$, to define the DNO

$$J(g) := \zeta^w \to \nu^w, \tag{3.35}$$

which maps the Dirichlet data, ζ^w, to the exterior Neumann trace, ν^w,

$$J(g)[\zeta^w] := [N \cdot \nabla w](x, g(x)) = [\partial_z w - (\partial_x g)\partial_x w](x, g(x)). \tag{3.36}$$

To understand how the operator $J(g)$ behaves under the change of variables in Appendix C, we multiply (3.36) by $C(x)$ to realize

$$CJ = C\partial_z w - (\partial_x g)C\partial_x w.$$

The differentiation rules for the change of variables, (C.5), produces

$$CJ = \partial_{z'} w' - (\partial_{x'} g)\{C\partial_{x'} w' - E\partial_{z'} w'\}.$$

These are evaluated at the upper boundary, $z' = 0$, where we observe that

$$C(x') = 1 + \frac{g}{b}, \quad E(x', 0) = \partial_{x'} g,$$

to find

$$\left(1 + \frac{g}{b}\right) J = \partial_{z'} w' - (\partial_{x'} g)\left\{\left(1 + \frac{g}{b}\right)\partial_{x'} w' - (\partial_{x'} g)\partial_{z'} w'\right\}.$$

We solve for J and drop the primes to find

$$J(g)[\zeta^w] = \partial_z w(x, 0) - L(x; g, w), \tag{3.37}$$

with

$$\begin{aligned} L(x; g, w) := (\partial_x g)\partial_x w(x, 0) - \frac{1}{b} g J(g)[\zeta^w] \\ + \frac{1}{b} g(\partial_x g)\partial_x w(x, 0) - (\partial_x g)^2 \partial_z w(x, 0). \end{aligned} \tag{3.38}$$

Upon setting $g(x) = \varepsilon f(x)$ and seeking an expansion of the form

$$J = J(\varepsilon f, \delta) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} J_{n,m}(f)\varepsilon^n \delta^m,$$

the equations (3.37) and (3.38) deliver

$$J_{n,m}(f)[\zeta^w] = \partial_z w_{n,m}(x, 0) - L_{n,m}(x; f, w), \tag{3.39}$$

where

$$\begin{aligned} L_{n,m}(x;f,w) := (\partial_x f)\partial_x w_{n-1,m}(x,0) - \frac{1}{b} f J_{n-1,m}(f)[\zeta^w] \\ + \frac{1}{b} f(\partial_x f)\partial_x w_{n-2,m}(x,0) - (\partial_x f)^2 \partial_z w_{n-2,m}(x,0). \end{aligned} \tag{3.40}$$

To prove the analyticity of the DNO we will need the following recursive estimate for $L_{n,m}$.

Lemma 3.9.1. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and*

$$\|w_{n,m}\|_{H^{s+2}} \leq K B^n D^m, \quad \|J_{n,m}\|_{H^{s+1/2}} \leq \tilde{K}\tilde{B}^n\tilde{D}^m, \quad \forall\ n < \overline{n},\ m \geq 0, \tag{3.41}$$

for constants $K, B, D, \tilde{K}, \tilde{B}, \tilde{D} > 0$ where $\tilde{K} \geq K, \tilde{B} \geq B, \tilde{D} \geq D$, then there exists a constant $\tilde{C} > 0$ such that

$$\|L_{\overline{n},m}\|_{H^{s+1/2}} \leq \tilde{K}\tilde{C}\left\{|f|_{C^{s+2}}\tilde{B}^{n-1}\tilde{D}^m + |f|^2_{C^{s+2}}\tilde{B}^{n-2}\tilde{D}^m\right\}. \tag{3.42}$$

Proof. [Lemma 3.9.1] From (3.40) and Lemma 3.6.1 we estimate

$$\begin{aligned} \|L_{\overline{n},m}\|_{H^{s+1/2}} &\leq \mathcal{M}|\partial_x f|_{C^{s+1/2+\sigma}}\|\partial_x w_{\overline{n}-1,m}(x,0)\|_{H^{s+1/2}} \\ &+ \frac{1}{b}\mathcal{M}|f|_{C^{s+1/2+\sigma}}\|G_{\overline{n}-1,m}(f)[\zeta^w]\|_{H^{s+1/2}} \\ &+ \frac{1}{b}\mathcal{M}^2|f|_{C^{s+1/2+\sigma}}|\partial_x f|_{C^{s+1/2+\sigma}}\|\partial_x w_{\overline{n}-2,m}(x,0)\|_{H^{s+1/2}} \\ &+ \mathcal{M}^2|\partial_x f|^2_{C^{s+1/2+\sigma}}\|\partial_z w_{\overline{n}-2,m}(x,0)\|_{H^{s+1/2}}. \end{aligned}$$

This gives

$$\begin{aligned} \|L_{\overline{n},m}\|_{H^{s+1/2}} \leq \tilde{K}\Big\{&\mathcal{M}|f|_{C^{s+2}}\tilde{B}^{\overline{n}-1}\tilde{D}^m + \frac{1}{b}\mathcal{M}|f|_{C^{s+2}}\tilde{B}^{\overline{n}-1}\tilde{D}^m \\ &+ \frac{1}{b}\mathcal{M}^2|f|^2_{C^{s+2}}\tilde{B}^{\overline{n}-2}\tilde{D}^m + \mathcal{M}^2|f|^2_{C^{s+2}}\tilde{B}^{\overline{n}-2}\tilde{D}^m\Big\}, \end{aligned}$$

and we are done provided

$$\tilde{C} \geq \left(1 + \frac{1}{b}\right)\max\{\mathcal{M}, \mathcal{M}^2\}.$$

□

We now have everything we need to prove the analyticity of the lower layer DNO.

Theorem 3.9.2. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and $\zeta^w_{n,m} \in H^{s+3/2}([0,d])$ such that*

$$\|\zeta^w_{n,m}\|_{H^{s+3/2}} \leq K_\zeta B^n_\zeta D^m_\zeta,$$

for constants $K_\zeta, B_\zeta, D_\zeta > 0$*, then* $J_{n,m} \in H^{s+1/2}([0,d])$ *and*

$$\|J_{n,m}\|_{H^{s+1/2}} \leq \tilde{K}\tilde{B}^n\tilde{D}^m, \tag{3.43}$$

for constants $\tilde{K}, \tilde{B}, \tilde{D} > 0$.

Proof. [Theorem 3.9.2] We work by induction in n. At $n = 0$ we have from (3.39) that

$$J_{0,m} = \partial_z w_{0,m}(x,0),$$

and from Theorem 3.8.1 we have

$$\|J_{0,m}\|_{H^{s+1/2}} = \|\partial_z w_{0,m}(x,0)\|_{H^{s+1/2}} \leq \|w_{0,m}\|_{H^{s+2}} \leq KD^m.$$

So we choose $\tilde{K} \geq K$ and $\tilde{D} \geq D$. We now assume $\tilde{B} \geq B$ and the estimate (3.43) for all $n < \overline{n}$ and estimate (3.39)

$$\|J_{\overline{n},m}(f)[\zeta^w]\|_{H^{s+1/2}} \leq \|\partial_z w_{\overline{n},m}(x,0)\|_{H^{s+1/2}} + \|L_{\overline{n},m}(x)\|_{H^{s+1/2}}.$$

Using the inductive hypothesis, Lemma 3.9.1, and Theorem 3.8.1 we have

$$\|J_{\overline{n},m}(f)[\zeta^w]\|_{H^{s+1/2}} \leq KB^{\overline{n}}D^m + \tilde{K}\tilde{C}\left\{|f|_{C^{s+2}}\tilde{B}^{\overline{n}-1}\tilde{D}^m + |f|^2_{C^{s+2}}\tilde{B}^{\overline{n}-2}\tilde{D}^m\right\}.$$

We are done provided $\tilde{K} \geq 2K$ and

$$\tilde{B} \geq \max\left\{B, 4\tilde{C}|f|_{C^{s+2}}, 2\sqrt{\tilde{C}}|f|_{C^{s+2}}\right\}.$$

□

3.10 Numerical Method

As in the upper field, we will simulate a manufactured solution in order to verify the accuracy of our numerical scheme. We start by considering the basis function

$$v_p^w(x,z) := e^{i\tilde{p}x - i\gamma_p^w z}, \quad \tilde{p} = \frac{2\pi p}{d},$$

where the phase $\exp(i\alpha x)$ is removed. We then utilize the exact Dirichlet/Neumann pairs $\{\zeta_r^w, \nu_r^w\}$ defined at the wavenumber $p = r$ and a profile $g(x) = \varepsilon f(x)$ where $\varepsilon > 0$ and our manufactured solutions are

$$\zeta_r^w(x) := B_r e^{i\tilde{r}x - i\gamma_r^w g(x)}, \tag{3.44a}$$

$$\begin{aligned}\nu_r^w(x) &:= [-\partial_z w_r + (\partial_x g)\partial_x w_r](x, g(x)) \\ &= [(i\gamma_r^w) + \varepsilon(\partial_x f)(i\tilde{r})]B_r e^{i\tilde{r}x - i\gamma_r^w \varepsilon f(x)}.\end{aligned} \tag{3.44b}$$

To make the specification precise we solve, at every desired perturbation order n and m, the elliptic boundary value problem, (3.17),

$$\Delta w_{n,m} + 2i\underline{\alpha}\partial_x w_{n,m} + (\underline{\gamma}^w)^2 w_{n,m} = \tilde{Y}_{n,m}\left(x, z; f, w, \underline{\alpha}, \underline{\gamma}^w\right), \quad -b < z < 0, \tag{3.45a}$$

$$w_{n,m}(x, 0) = \zeta_{n,m}^w(x), \qquad \text{at } z = 0, \tag{3.45b}$$

$$w_{n,m}(x + d, z) = w_{n,m}(x, z), \tag{3.45c}$$

$$\partial_z\left[w_{n,m}(x, -b), \right] - T_0^w[w_{n,m}(x, -b)] = \tilde{Q}_{n,m}(x), \qquad \text{at } z = -b, \tag{3.45d}$$

followed by the simulation of the n–th and m–th correction of the DNO, (3.39),

$$J_{n,m}(f)[\zeta^w] = \partial_z w_{n,m}(x, 0) - L_{n,m}(x; f, w).$$

We begin by choosing the maximum perturbation orders, N and M, and then approximate

$$w(x, z; \varepsilon, \delta) \approx w^{N,M}(x, z; \varepsilon, \delta) := \sum_{n=0}^{N}\sum_{m=0}^{M} w_{n,m}(x, z)\varepsilon^n\delta^m, \tag{3.46}$$

$$J(x; \varepsilon, \delta) \approx J^{N,M}(x; \varepsilon, \delta) := \sum_{n=0}^{N}\sum_{m=0}^{M} J_{n,m}(x)\varepsilon^n\delta^m, \tag{3.47}$$

where, by the periodicity of solutions,

$$w_{n,m}(x, z) = \sum_{p=-\infty}^{\infty} \hat{w}_{n,m,p}(z)e^{i\tilde{p}x}, \quad J_{n,m}(x) = \sum_{p=-\infty}^{\infty} \hat{J}_{n,m,p}e^{i\tilde{p}x}. \tag{3.48}$$

Each of these $w_{n,m}(x, z)$ are then simulated by a Fourier–Chebyshev approach which posits the form

$$w_{n,m}(x, z) \approx w_{n,m}^{N_x,N_z}(x, z) := \sum_{p=-N_x/2}^{N_x/2-1}\sum_{\ell=0}^{N_z} \hat{w}_{n,m,p,\ell}e^{i\tilde{p}x}T_\ell\left(\frac{2z + b}{b}\right),$$

where T_ℓ is the ℓ–th Cheybshev polynomial. The unknowns, $\hat{w}_{n,m,p,\ell}$ are recovered from (3.45) by the collocation approach. As in §2.11, our HOPS/AWE algorithm requires $N_x \times N_z$ unknowns at every perturbation order (n, m). We apply a Fourier spectral method in the lateral direction where we require N_x equally–spaced gridpoints. In the vertical direction we use a Chebyshev spectral method where we choose N_z+1 collocation

points. We then simulate the lower layer DNO from (3.39), where the coefficients $J_{n,m}$ from (3.48) are approximated by

$$J_{n,m}(x) \approx J_{n,m}^{N_x}(x) := \sum_{p=-N_x/2}^{N_x/2-1} \hat{J}_{n,m,p} e^{i\tilde{p}x},$$

and the $\hat{J}_{n,m,p}$ are recovered from the $\hat{w}_{n,m,p,\ell}$. Inserting the expansions (3.48) into (3.45) gives

$$\partial_z^2 \hat{w}_{n,m,p}(z) + \left((\underline{\gamma}_p^w)^2 - \tilde{p}^2 - 2\underline{\alpha}\tilde{p} \right) \hat{w}_{n,m,p}(z) = \hat{\hat{Y}}_{n,m,p}(z), \qquad -b < z < 0, \tag{3.49a}$$

$$\hat{w}_{n,m,p}(0) = \hat{\zeta}_{n,m,p}^w, \qquad \text{at } z = 0, \tag{3.49b}$$

$$\partial_z \left[\hat{w}_{n,m,p}(-b) \right] - \hat{T^w}[\hat{w}_{n,m,p}(-b)] = \hat{\hat{Q}}_{n,m,p}, \qquad \text{at } z = -b, \tag{3.49c}$$

Through this, we can solve our two–point boundary value problem through our Chebyshev collocation method and we now turn to a numerical implementation of our HOPS/AWE algorithm in Matlab. To begin, we approximate the lower layer Dirichlet and Neumann data through the expansions

$$\zeta_{\text{TFE}}^{N_x,N_z,N,M} = \sum_{n=0}^{N} \sum_{m=0}^{M} \zeta_{n,m}^{N_x,N_z}(x) \varepsilon^n \delta^m, \quad \nu_{\text{TFE}}^{N_x,N_z,N,M} = \sum_{n=0}^{N} \sum_{m=0}^{M} \nu_{n,m}^{N_x,N_z}(x) \varepsilon^n \delta^m,$$

from which we can compute the relative errors

$$\text{Error } \zeta_r^w = \text{Error}_{\text{TFE}}(N_x, N_z, N, M) := \frac{\left| \zeta_r^w - \zeta_{TFE}^{Nx,Nz,N,M} \right|_{L^\infty}}{|\zeta_r^w|_{L^\infty}},$$

$$\text{Error } \nu_r^w = \text{Error}_{\text{TFE}}(N_x, N_z, N, M) := \frac{\left| \nu_r^w - \nu_{TFE}^{Nx,Nz,N,M} \right|_{L^\infty}}{|\nu_r^w|_{L^\infty}}.$$

3.11 Numerical Results

For our first simulation we considered a profile with moderate interfacial and frequency perturbations and the following paramaters

$$f(x) = e^{\cos(x)}, \quad \alpha = 0, \quad \varepsilon = 10^{-4}, \quad \delta = 10^{-4}, \quad d = 2\pi, \quad r = 2,$$

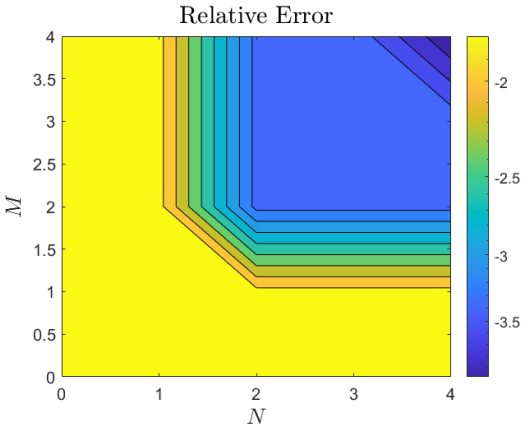
$$B_r = -4.8, \quad \gamma^w = 1.15, \quad N_x = 32, \quad N_z = 32, \quad N = M = 4.$$

In Table III we report the results of our tests using both Padé and Taylor summation.

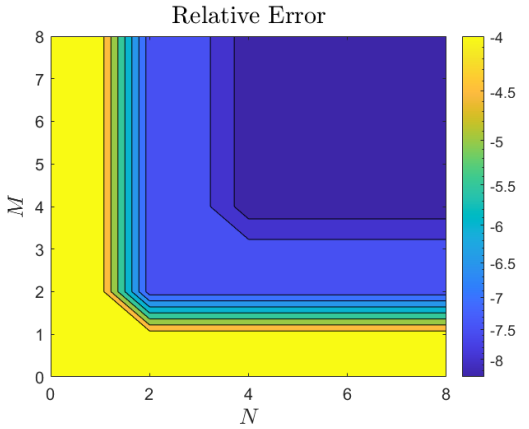
N	M	Error ζ_r^w (Taylor)	Error ζ_r^w (Padé)	Error ν_r^w (Taylor)	Error ν_r^w (Padé)
0	2	0.000187563	0.000187563	8.47895e-05	8.47895e-05
0	4	0.000187563	0.000187563	8.47895e-05	8.47895e-05
1	2	0.000187563	0.000187563	8.47895e-05	8.47895e-05
1	4	0.000187563	0.000187563	8.47895e-05	8.47895e-05
2	2	5.4509e-08	5.02104e-09	2.65222e-08	4.42536e-09
2	4	5.4509e-08	5.02104e-09	2.65222e-08	4.42536e-09
3	2	5.4509e-08	5.02104e-09	2.65222e-08	4.42536e-09
3	4	5.4509e-08	5.02104e-09	2.65222e-08	4.42536e-09
4	2	5.4509e-08	5.02104e-09	2.65222e-08	4.42536e-09
4	4	5.00595e-09	4.98285e-09	4.41455e-09	4.39911e-09

TABLE III: Relative Error, Error ζ_r^w and Error ν_r^w, versus perturbation orders N and M, for the TFE approximations to the Dirichlet data, ζ_r^w (3.44a), and the Neumann data, ν_r^w (3.44b), where we used both Taylor Series and Padé approximants. Parameter choices are specified above where we investigated moderate boundary and frequency perturbations.

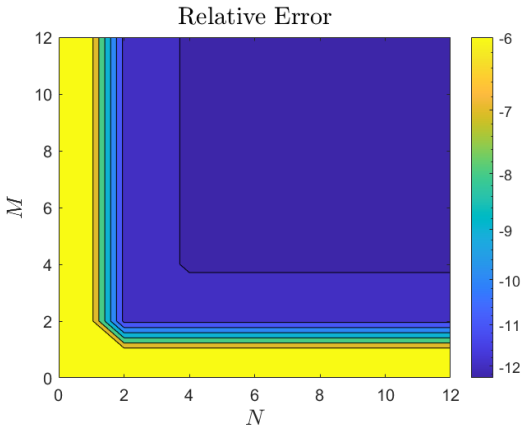
As expected from spectral methods, our HOPS/AWE algorithm reaches reasonable accuracy at $N = M = 4$ Taylor or Padé orders. We then simulated new results by increasing the number of Padé orders and decreasing the size of the interfacial and frequency perturbations. In Figure 10, we considered $N = M = 4, 8, 12, 16$ Padé orders and plotted the Relative Error for ζ_r^w as we expanded up to $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ and $\delta = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ simultaneously with Padé summation. In Figure 11, we kept $N = M = 4$ Padé orders fixed and plotted the Relative Error for ζ_r^w as we expanded up to $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ and $\delta = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ simultaneously with Padé summation.



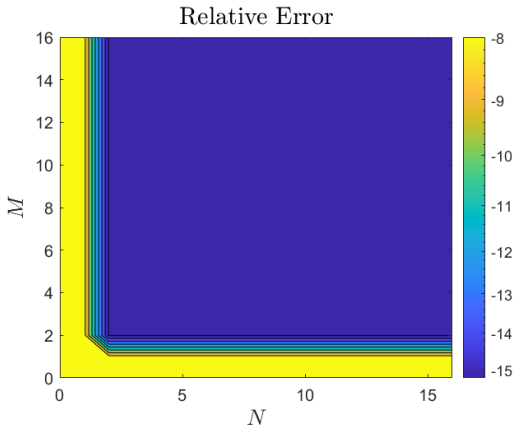
(a) Error ζ_r^w (Padé), $N = M = 4$, $\varepsilon = \delta = 10^{-2}$



(b) Error ζ_r^w (Padé), $N = M = 8$, $\varepsilon = \delta = 10^{-4}$

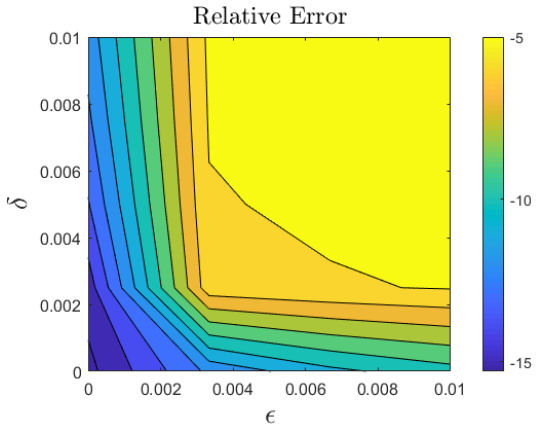


(c) Error ζ_r^w (Padé), $N = M = 12$, $\varepsilon = \delta = 10^{-6}$

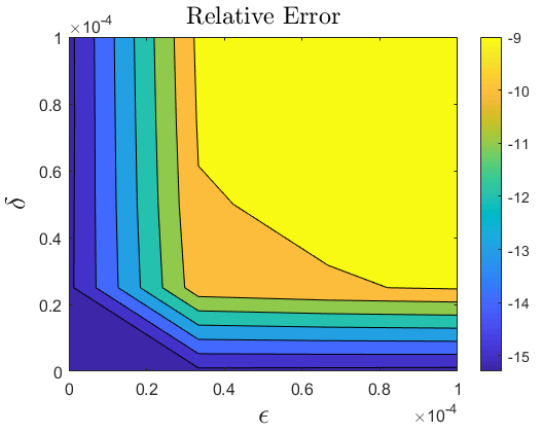


(d) Error ζ_r^w (Padé), $N = M = 16$, $\varepsilon = \delta = 10^{-8}$

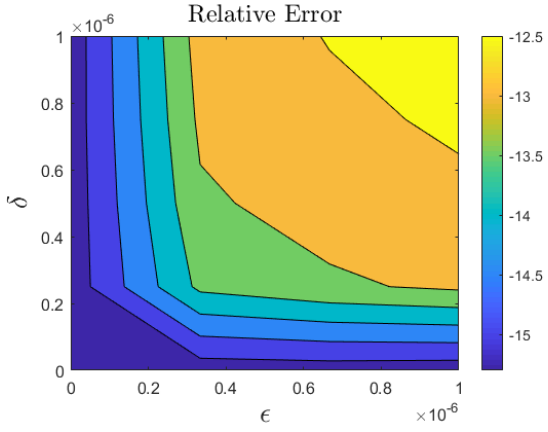
Figure 10: Plot of Relative Error for ζ_r^w. Our HOPS/AWE algorithm used Padé summation with $N = M = 4, 8, 12, 16$ Padé orders to expand up to $\varepsilon = \delta = 10^{-2}, 10^{-4}, 10^{-6}$, 10^{-8} simultaneously. Physical parameters are reported in the profile above.



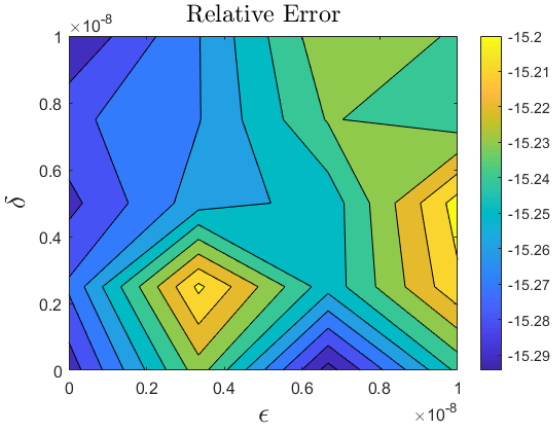
(a) Error ζ_r^w (Padé), $\varepsilon = \delta = 10^{-2}$



(b) Error ζ_r^w (Padé), $\varepsilon = \delta = 10^{-4}$



(c) Error ζ_r^w (Padé), $\varepsilon = \delta = 10^{-6}$



(d) Error ζ_r^w (Padé), $\varepsilon = \delta = 10^{-8}$

Figure 11: Plot of Relative Error for ζ_r^w with $N = M = 4$ Padé orders fixed. Our HOPS/AWE algorithm used Padé summation to expand up to $\varepsilon = \delta = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ simultaneously. Physical parameters are reported in the profile above.

We then simulated an analytic profile with a smaller frequency perturbation and the following parameters

$$f(x) = \frac{1}{4}\sin(4x), \quad \alpha = 0, \quad \varepsilon = 10^{-4}, \quad \delta = 10^{-8}, \quad d = 2\pi, \quad r = 2,$$

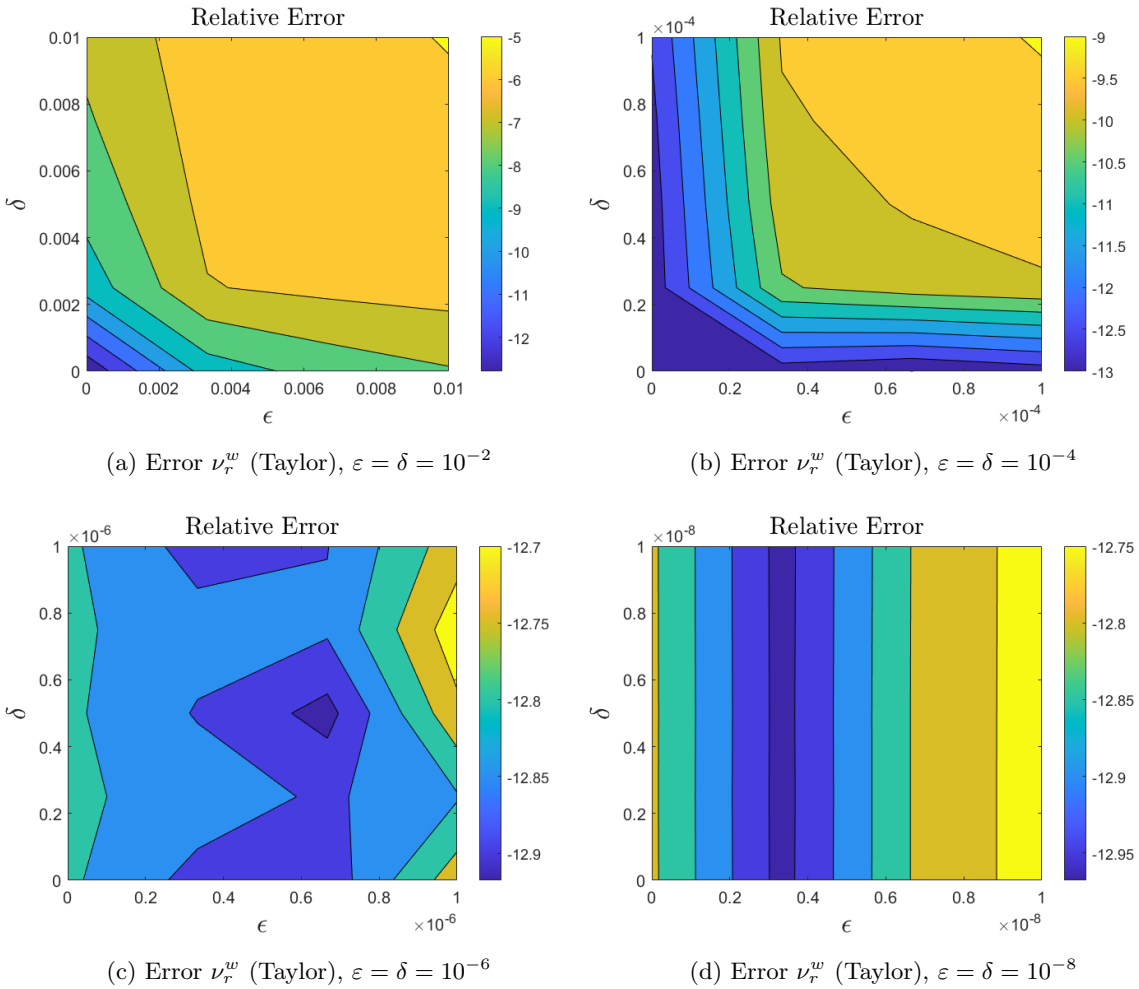
$$B_r = -4.8, \quad \gamma^w = 1.15, \quad N_x = 32, \quad N_z = 32, \quad N = M = 4.$$

In Table IV we report the results of our tests using both Padé and Taylor summation.

N	M	Error ζ_r^w (Taylor)	Error ζ_r^w (Padé)	Error ν_r^w (Taylor)	Error ν_r^w (Padé)
0	2	1.04998e-05	1.04998e-05	6.9068e-05	6.9068e-05
0	4	1.04998e-05	1.04998e-05	6.9068e-05	6.9068e-05
1	2	1.04998e-05	1.04998e-05	6.9068e-05	6.9068e-05
1	4	1.04998e-05	1.04998e-05	6.9068e-05	6.9068e-05
2	2	9.21284e-10	2.7964e-13	2.88241e-09	2.5317e-13
2	4	9.21284e-10	2.7964e-13	2.88241e-09	2.5317e-13
3	2	9.21284e-10	2.7964e-13	2.88241e-09	2.5317e-13
3	4	9.21284e-10	2.7964e-13	2.88241e-09	2.5317e-13
4	2	9.21284e-10	2.7964e-13	2.88241e-09	2.5317e-13
4	4	8.25908e-14	9.3863e-14	3.1523e-13	2.50276e-13

TABLE IV: Relative Error, Error ζ_r^w and Error ν_r^w, versus perturbation orders N and M, for the TFE approximations to the Dirichlet data, ζ_r^w (3.44a), and the Neumann data, ν_r^w (3.44b), where we used both Taylor Series and Padé approximants. Parameter choices are specified by the analytic profile above.

In Figure 12, we kept $N = M = 4$ Taylor orders fixed and computed the Relative Error for ν_r^w as we expanded up to $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ and $\delta = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ simultaneously. Jointly decreasing both perturbation variables simultaneously increased the accuracy of our HOPS/AWE algorithm and returned favorable convergence results for both $\varepsilon = \delta = 10^{-6}$ and $\varepsilon = \delta = 10^{-8}$.



(a) Error ν_r^w (Taylor), $\varepsilon = \delta = 10^{-2}$

(b) Error ν_r^w (Taylor), $\varepsilon = \delta = 10^{-4}$

(c) Error ν_r^w (Taylor), $\varepsilon = \delta = 10^{-6}$

(d) Error ν_r^w (Taylor), $\varepsilon = \delta = 10^{-8}$

Figure 12: Plot of Relative Error for ν_r^w with $N = M = 4$ Taylor orders fixed. Our HOPS/AWE algorithm used Taylor summation to expand up to $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ and $\delta = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ simultaneously with the analytic profile above.

CHAPTER 4

EXISTENCE, UNIQUENESS, AND JOINT ANALYTICITY OF SOLUTIONS TO THE TWO–LAYER PROBLEM

4.1 Introduction

This chapter combines the analysis performed in Chapters 2 and 3 to fully establish the existence, uniqueness, and analyticity of solutions to our scattering problem. In §4.2 we summarize the equations which govern the propagation of linear waves in a two-dimensional periodic structure, and discuss how the far–field boundary conditions can be enforced through the use of Transparent Boundary Conditions. Then in §4.3 we restate our governing equations in terms of interfacial quantities via a Non–Overlapping Domain Decomposition phrased in terms of the DNOs from Chapters 2 and 3. In §4.4 we present a rather general and rigorously justifiable perturbative scheme for solving systems of linear systems of equations in Banach spaces. The appropriate analyticity theorems for a single perturbation parameter and two perturbations parameters are presented and proven in §4.5, respectively. The application of these results to the governing equations presented in §4.3 is made in §4.6 where our novel result (Theorem 4.6.1) is established. The proof requires several rigorous analyses, and the first of these is given in §4.7 with the analyticity of the surface data. Then, §4.8 presents the invertibility of the linearized operator representing the flat–interface solution. The bulk of the analysis has previously been performed in Chapters 2 and 3 where we established the analyticity of the transformed fields with a single, geometric, boundary perturbation, ε, alone in Theorems 2.8.4 and 3.7.1. The joint analyticity of the transformed fields was proven in Theorems 2.9.2 and 3.8.1 and the analyticity of the DNOs in Theorems 2.10.2 and 3.9.2.

4.2 Governing Equations and Propagating Conditions

The two–layer scattering problem is composed of outgoing, quasiperiodic solutions to

$$\Delta\tilde{u} + (k^u)^2\tilde{u} = 0, \qquad z > g(x), \tag{4.1a}$$

$$\Delta\tilde{w} + (k^w)^2\tilde{w} = 0, \qquad z < g(x), \tag{4.1b}$$

$$\tilde{u} - \tilde{w} = \tilde{\zeta}, \qquad \text{at } z = g(x), \tag{4.1c}$$

$$\partial_N\tilde{u} - \tau^2\partial_N\tilde{w} = \tilde{\psi}, \qquad \text{at } z = g(x). \tag{4.1d}$$

The Dirichlet and Neumann data are

$$\tilde{\zeta}(x) := -e^{i\alpha x - i\gamma^u g(x)}, \tag{4.1e}$$

$$\tilde{\psi}(x) := (i\gamma^u + i\alpha(\partial_x g))e^{i\alpha x - i\gamma^u g(x)}, \tag{4.1f}$$

and

$$\tau^2 = \begin{cases} 1, & \text{TE}, \\ (k^u/k^w)^2 = (n^u/n^w)^2, & \text{TM}. \end{cases}$$

Following our analysis in Chapters 2 and 3, we start by removing the phase in (4.1) through the relationship $v(x,z) = e^{-i\alpha x}\tilde{v}(x,z)$, $v \in \{u, w\}$, and $\zeta(x) = e^{-i\alpha x}\tilde{\zeta}(x)$, $\psi(x) = e^{-i\alpha x}\tilde{\psi}(x)$. This yields outgoing, d–periodic solutions of

$$\Delta u + 2i\alpha\partial_x u + (\gamma^u)^2 u = 0, \qquad z > g(x), \tag{4.2a}$$

$$\Delta w + 2i\alpha\partial_x w + (\gamma^w)^2 w = 0, \qquad z < g(x), \tag{4.2b}$$

$$u - w = \zeta, \qquad \text{at } z = g(x), \tag{4.2c}$$

$$\partial_N u - i\alpha(\partial_x g)u - \tau^2\left[\partial_N w - i\alpha(\partial_x g)w\right] = \psi, \qquad \text{at } z = g(x), \tag{4.2d}$$

where

$$\zeta(x) := -e^{-i\gamma^u g(x)}, \tag{4.2e}$$

$$\psi(x) := (i\gamma^u + i\alpha(\partial_x g))e^{-i\gamma^u g(x)}, \tag{4.2f}$$

and the left–hand side of (4.2d) follows from

$$\begin{aligned} \partial_N\tilde{u} - \tau^2\partial_N\tilde{w} &= \partial_N\left(e^{i\alpha x}u\right) - \tau^2\partial_N\left(e^{i\alpha x}w\right) \\ &= e^{i\alpha x}\Big(\partial_z u + (-\partial_x g)\partial_x u - (i\alpha)(\partial_x g)u - \\ &\qquad \tau^2\Big[\partial_z w + (-\partial_x g)\partial_x w - (i\alpha)(\partial_x g)w\Big]\Big) \\ &= e^{i\alpha x}\Big(\partial_N u - i\alpha(\partial_x g)u - \tau^2\left[\partial_N w - i\alpha(\partial_x g)w\right]\Big). \end{aligned}$$

The Upward Propagating Condition (UPC) and Downward Propagating Condition (DPC) (8) rigorously enforce the Outgoing Wave Conditions which we mentioned in §1.8. We now demonstrate how these can be stated in terms of Transparent Boundary Conditions which also truncate the bi–infinite problem domain to one of finite size. As discussed in §1.8, we choose values a and b such that

$$a > |g|_\infty, \quad -b < -|g|_\infty,$$

and define the artificial boundaries $\{z = a\}$ and $\{z = -b\}$. In $\{z > a\}$ the Rayleigh expansions (6) tell us that upward propagating solutions of (4.2a) are

$$u(x,z) = \sum_{p=-\infty}^{\infty} a_p e^{i\tilde{p}x + i\gamma_p^u z}, \tag{4.3}$$

where, for $p \in \mathbb{Z}$ and $q \in \{u, w\}$,

$$\tilde{p} := \frac{2\pi p}{d}, \quad \alpha_p := \alpha + \tilde{p}, \quad \gamma_p^q := \sqrt{(k^q)^2 - \alpha_p^2}, \quad \operatorname{Im}\{\gamma_p^q\} \geq 0. \tag{4.4}$$

In a similar fashion, downward propagating solutions of (4.2b) in $\{z < -b\}$ can be expressed as

$$w(x,z) = \sum_{p=-\infty}^{\infty} d_p e^{i\tilde{p}x - i\gamma_p^w z}. \tag{4.5}$$

With these we can define the Transparent Boundary Conditions in the following way: Focusing on the UPC we rewrite (4.3) as

$$u(x,z) = \sum_{p=-\infty}^{\infty} \left(a_p e^{i\gamma_p^u a}\right) e^{i\tilde{p}x + i\gamma_p^u (z-a)} = \sum_{p=-\infty}^{\infty} \hat{\xi}_p e^{i\tilde{p}x + i\gamma_p^u (z-a)},$$

and note that,

$$u(x,a) = \sum_{p=-\infty}^{\infty} \hat{\xi}_p e^{i\tilde{p}x} =: \xi(x),$$

and

$$\partial_z u(x,a) = \sum_{p=-\infty}^{\infty} (i\gamma_p^u)\hat{\xi}_p e^{i\tilde{p}x} =: T^u[\xi(x)],$$

which defines the order–one Fourier multiplier T^u. For the DPC we rewrite (4.5) as

$$w(x,z) = \sum_{p=-\infty}^{\infty} \left(d_p e^{i\gamma_p^w b}\right) e^{i\tilde{p}x - i\gamma_p^w (z+b)} = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x - i\gamma_p^w (z+b)},$$

and keep in mind,

$$w(x,-b) = \sum_{p=-\infty}^{\infty} \hat{\psi}_p e^{i\tilde{p}x} =: \psi(x),$$

and

$$\partial_z w(x,-b) = \sum_{p=-\infty}^{\infty} (-i\gamma_p^w)\hat{\psi}_p e^{i\tilde{p}x} =: T^w[\psi(x)],$$

which defines the order–one Fourier multiplier T^w. From this we state that upward–propagating solutions of (4.2a) satisfy the Transparent Boundary Condition at $z = a$

$$\partial_z u(x,a) - T^u[u(x,a)] = 0, \quad z = a. \tag{4.6}$$

Similarly, downward–propagating solutions of (4.2b) satisfy the Transparent Boundary Condition at $z = -b$

$$\partial_z w(x,-b) - T^w[w(x,-b)] = 0, \quad z = -b. \tag{4.7}$$

We also point out that solutions which satisfy (4.6) and (4.7) equivalently satisfy the UPC and DPC, respectively (8). With these we now state the full set of governing equations as

$$\begin{aligned}
&\Delta u + 2i\alpha\partial_x u + (\gamma^u)^2 u = 0, && z > g(x), && \text{(4.8a)}\\
&\Delta w + 2i\alpha\partial_x w + (\gamma^w)^2 w = 0, && z < g(x), && \text{(4.8b)}\\
&u - w = \zeta, && z = g(x), && \text{(4.8c)}\\
&\partial_N u - i\alpha(\partial_x g)u - \tau^2\left[\partial_N w - i\alpha(\partial_x g)w\right] = \psi, && z = g(x), && \text{(4.8d)}\\
&\partial_z u(x,a) - T^u[u(x,a)] = 0, && z = a, && \text{(4.8e)}\\
&\partial_z w(x,-b) - T^w[w(x,-b)] = 0, && z = -b, && \text{(4.8f)}\\
&u(x+d,z) = u(x,z), && && \text{(4.8g)}\\
&w(x+d,z) = w(x,z). && && \text{(4.8h)}
\end{aligned}$$

4.3 A Non–Overlapping Domain Decomposition Method

We now restate our governing equations (4.8) in terms of surface quantities via a Non–Overlapping Domain Decomposition Method (DDM) (95; 96; 97). For this we define

$$\begin{aligned}
U(x) &:= u(x,g(x)), \quad \tilde{U}(x) := -\partial_N u(x,g(x)),\\
W(x) &:= w(x,g(x)), \quad \tilde{W}(x) := \partial_N w(x,g(x)),
\end{aligned}$$

where u is a d–periodic solution of (4.8a) and (4.8e), and w is a d–periodic solution of (4.8b) and (4.8f). In terms of these our full governing equations (4.8) are equivalent to the pair of boundary conditions, (4.8c) & (4.8d),

$$U - W = \zeta, \quad -\tilde{U} - (i\alpha)(\partial_x g)U - \tau^2\left[\tilde{W} - (i\alpha)(\partial_x g)W\right] = \psi. \tag{4.9}$$

This set of two equations for four unknowns can be closed by noting that the pairs $\{U, \tilde{U}\}$ and $\{W, \tilde{W}\}$ are connected, e.g., by the DNOs

$$G : U \to \tilde{U}, \quad J : W \to \tilde{W}.$$

Definition 4.3.1. We recall the precise definition of the upper layer DNO (98): Given an integer $s \geq 0$, if $g \in C^{s+2}$ the unique solution of

$$\Delta u + 2i\alpha\partial_x u + (\gamma^u)^2 u = 0, \qquad g(x) < z < a, \tag{4.10a}$$
$$u(x, g(x)) = U(x), \qquad z = g(x), \tag{4.10b}$$
$$\partial_z u(x, a) - T^u[u(x, a)] = 0, \qquad z = a, \tag{4.10c}$$
$$u(x + d, z) = u(x, z), \tag{4.10d}$$

defines the Upper Layer DNO

$$G(g) : U \to \tilde{U} := -(\partial_N u)(x, g(x)). \tag{4.11}$$

Definition 4.3.2. Similarly, we recall the definition of the lower layer DNO: Given an integer $s \geq 0$, if $g \in C^{s+2}$ the unique solution of

$$\Delta w + 2i\alpha\partial_x w + (\gamma^w)^2 w = 0, \qquad -b < z < g(x), \tag{4.12a}$$
$$w(x, g(x)) = W(x), \qquad z = g(x), \tag{4.12b}$$
$$\partial_z w(x, -b) - T^w[w(x, -b)] = 0, \qquad z = -b, \tag{4.12c}$$
$$w(x + d, z) = w(x, z), \tag{4.12d}$$

defines the Lower Layer DNO

$$J(g) : W \to \tilde{W} := (\partial_N w)(x, g(x)). \tag{4.13}$$

We now write (4.9) as

$$\mathbf{AV} = \mathbf{R}, \tag{4.14}$$

where

$$\mathbf{A} = \begin{pmatrix} I & -I \\ G + (\partial_x g)(i\alpha) & \tau^2 J - \tau^2(\partial_x g)(i\alpha) \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} U \\ W \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \zeta \\ -\psi \end{pmatrix}. \tag{4.15}$$

For later use, the trivial flat–interface version of (4.15) is $\mathbf{A}_{0,0}\mathbf{V}_{0,0} = \mathbf{R}_{0,0}$ where

$$\mathbf{A}_{0,0} = \begin{pmatrix} I & -I \\ -G_{0,0} & -\tau^2 J_{0,0} \end{pmatrix}, \quad \mathbf{V}_{0,0} = \begin{pmatrix} U_{0,0} \\ W_{0,0} \end{pmatrix}, \quad \mathbf{R}_{0,0} = \begin{pmatrix} \zeta_{0,0} \\ -\psi_{0,0} \end{pmatrix}. \tag{4.16}$$

4.4 Analyticity of Solutions to Linear Systems

Following our analysis in Chapters 2 and 3, we pursue a jointly perturbative approach to solving (4.14) based on the assumptions

$$g(x) = \varepsilon f(x), \quad \omega = \underline{\omega} + \delta\underline{\omega} = (1 + \delta)\underline{\omega},$$

where upon performing a join Taylor expansion the DNO G takes the form (2.52) (cf. §2.10) and the DNO J takes the form (3.39) (cf. §3.9).

With this we establish the existence, uniqueness, and analyticity of solutions to (4.14). To accomplish this we consider systems of linear equations of the form

$$\mathbf{A}(\varepsilon,\delta)\mathbf{V}(\varepsilon,\delta)=\mathbf{R}(\varepsilon,\delta), \tag{4.17}$$

and show how such equations can be solved by regular perturbation theory.

4.5 Rigorous Regular Perturbation Theory

To begin, we assume

$$\mathbf{A}(\varepsilon,\delta)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbf{A}_{n,m}\varepsilon^n\delta^m,\quad \mathbf{R}(\varepsilon,\delta)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbf{R}_{n,m}\varepsilon^n\delta^m,$$

in (4.17) and seek a solution of the form

$$\mathbf{V}(\varepsilon,\delta)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathbf{V}_{n,m}\varepsilon^n\delta^m. \tag{4.18}$$

From (4.17) we find at order $\mathcal{O}(\varepsilon^n,\delta^m)$

$$\begin{aligned}\mathbf{A}_{0,0}\mathbf{V}_{n,m}=\mathbf{R}_{n,m}&-\sum_{\ell=0}^{n-1}\mathbf{A}_{n-\ell,0}\mathbf{V}_{\ell,m}-\sum_{r=0}^{m-1}\mathbf{A}_{0,m-r}\mathbf{V}_{n,r}\\&-\sum_{\ell=0}^{n-1}\sum_{r=0}^{m-1}\mathbf{A}_{n-\ell,m-r}\mathbf{V}_{\ell,r},\end{aligned}$$

or

$$\begin{aligned}\mathbf{V}_{n,m}=\mathbf{A}_{0,0}^{-1}\Bigg(\mathbf{R}_{n,m}&-\sum_{\ell=0}^{n-1}\mathbf{A}_{n-\ell,0}\mathbf{V}_{\ell,m}-\sum_{r=0}^{m-1}\mathbf{A}_{0,m-r}\mathbf{V}_{n,r}\\&-\sum_{\ell=0}^{n-1}\sum_{r=0}^{m-1}\mathbf{A}_{n-\ell,m-r}\mathbf{V}_{\ell,r}\Bigg).\end{aligned} \tag{4.19}$$

With these we can establish an existence theorem (98) for this problem depending on two parameters.

Theorem 4.5.1. *Given two Banach spaces X and Y, suppose that*

[1] $\mathbf{R}_{n,m}\in Y$ *for all* $n,m\geq 0$, *and there exists constants* $B_R>0, C_{R,N}>0, C_{R,M}>0, D_R>0$ *such that*

$$\|\mathbf{R}_{n,m}\|_Y\leq C_{R,N}C_{R,M}B_R^nD_R^m,$$

[2] $\mathbf{A}_{n,m} : X \to Y$ *for all* $n, m \geq 0$, *and there exists constants* $B_A > 0, C_{A,N} > 0, C_{A,M} > 0, D_A > 0$ *such that*

$$\|\mathbf{A}_{n,m}\|_{X\to Y} \leq C_{A,N} C_{A,M} B_A^n D_A^m,$$

[3] $\mathbf{A}_{0,0}^{-1} : Y \to X$ *for all* $n, m \geq 0$, *and there exists a constant* $C_e > 0$ *such that*

$$\left\|\mathbf{A}_{0,0}^{-1}\right\|_{Y\to X} \leq C_e.$$

Then the equation (4.17) *has a unique solution,* (4.18)*, and there exists constants* $B_V > 0, C_{V,N} > 0, C_{V,M} > 0,$ *and* $D_V > 0$ *such that*

$$\|\mathbf{V}_{n,m}\|_X \leq C_{V,N} C_{V,M} B_V^n D_V^m, \tag{4.20}$$

for all $n, m \geq 0$ *and any*

$$C_{V,N} \geq 2C_e C_{R,N}, \quad C_{V,M} \geq 2C_e C_{R,M},$$
$$B_V \geq \max\{B_R, 2B_A, 8C_e C_{A,N} B_A\}, \quad D_V \geq \max\{D_R, 2D_A, 8C_e C_{A,M} D_A\}.$$

This implies that, for any $0 \leq \rho, \sigma < 1$, (4.18) *converges for all* ε *such that* $B\varepsilon < \rho$, *i.e.,* $\varepsilon < \rho/B$ *and all* δ *such that* $D\delta < \sigma$, *i.e.,* $\delta < \sigma/D$.

Proof. [Theorem 4.5.1] We work by induction, where we want to establish

$$\|\mathbf{V}_{n,m}\|_X \leq C_{V,N} C_{V,M} B_V^n D_V^m, \quad \forall n, m \geq 0.$$

We start by an induction on m. The base case $m = 0$:

$$\|\mathbf{V}_{n,0}\|_X \leq C_{V,N} B_V^n, \quad \forall n \geq 0, \tag{4.21}$$

is established through an induction on n. We start with $n = 0$ where (4.19) becomes

$$\mathbf{V}_{0,0} = \mathbf{A}_{0,0}^{-1} \mathbf{R}_{0,0},$$

and, from the properties of $\mathbf{A}_{0,0}^{-1}$, we have

$$\|\mathbf{V}_{0,0}\|_X = \left\|\mathbf{A}_{0,0}^{-1}\mathbf{R}_{0,0}\right\|_X \leq C_e \|\mathbf{R}_{0,0}\|_Y =: C_V.$$

Now, assuming estimate (4.20) for all $n < \bar{n}$ we use (4.19) and the mapping properties of $\mathbf{A}_{0,0}^{-1}$ to find

$$\|\mathbf{V}_{\bar{n},0}\|_X \leq C_e \left\{ \|\mathbf{R}_{\bar{n},0}\|_Y + \sum_{\ell=0}^{\bar{n}-1} \|\mathbf{A}_{\bar{n}-\ell,0}\mathbf{V}_{\ell,0}\|_Y \right\}.$$

Now, using the estimates on $\mathbf{R}_{n,0}$ and $\mathbf{A}_{n,0}$ (for all n) and $\mathbf{V}_{n,0}$ ($n < \bar{n}$) we have

$$\begin{aligned}
\|\mathbf{V}_{\bar{n},0}\|_X &\leq C_e \left\{ C_R B_R^{\bar{n}} + \sum_{\ell=0}^{\bar{n}-1} C_A B_A^{\bar{n}-\ell} C_V B_V^{\ell} \right\} \\
&= C_e C_R B_R^{\bar{n}} + C_e C_A C_V \left(\frac{B_A}{B_V}\right) B_V^{\bar{n}} \sum_{\ell=0}^{\bar{n}-1} \left(\frac{B_A}{B_V}\right)^{\bar{n}-\ell-1} \\
&\leq C_e C_R B_R^{\bar{n}} + C_e C_A C_V \left(\frac{B_A}{B_V}\right) B_V^{\bar{n}} \left(\frac{1}{1-1/2}\right),
\end{aligned}$$

if $B_A/B_V \leq 1/2$ (implying $B_V \geq 2B_A$). We are done if we demand that

$$B_V \geq B_R, \quad C_e C_R \leq C_V/2, \quad 2C_e C_A C_V (B_A/B_V) \leq C_V/2.$$

All of this can be achieved provided

$$C_V \geq 2C_e C_R, \quad B_V \geq \max\{B_R, 2B_A, 4C_e C_A B_A\},$$

which establishes (4.21). We now assume

$$\|\mathbf{V}_{n,m}\|_X \leq C_{V,N} C_{V,M} B_V^n D_V^m, \quad \forall n \geq 0, \quad \forall m < \bar{m},$$

and seek

$$\|\mathbf{V}_{n,\bar{m}}\|_X \leq C_{V,N} C_{V,M} B_V^n D_V^{\bar{m}}, \quad \forall n \geq 0.$$

This can be obtained through a second induction on n. The base case $n = 0$:

$$\|\mathbf{V}_{0,\bar{m}}\|_X \leq C_{V,M} D_V^{\bar{m}}, \quad \forall \bar{m} \geq 0,$$

is established through an induction on $\bar{m}$ analogous to (4.21). Finally, we assume

$$\|\mathbf{V}_{n,\bar{m}}\|_X \leq C_{V,N} C_{V,M} B_V^n D_V^{\bar{m}}, \quad \forall n \leq \bar{n}, \quad \forall \bar{m} \geq 0,$$

and seek

$$\|\mathbf{V}_{\bar{n},\bar{m}}\|_X \leq C_{V,N} C_{V,M} B_V^{\bar{n}} D_V^{\bar{m}}.$$

We now use (4.19) and the mapping properties of $\mathbf{A}_{0,0}^{-1}$ to find

$$\|\mathbf{V}_{\bar{n},\bar{m}}\|_X \le C_e\Bigg\{ \|\mathbf{R}_{\bar{n},\bar{m}}\|_Y + \sum_{\ell=0}^{\bar{n}-1} \|\mathbf{A}_{\bar{n}-\ell,0}\mathbf{V}_{\ell,\bar{m}}\|_Y + \sum_{r=0}^{\bar{m}-1} \|\mathbf{A}_{0,\bar{m}-r}\mathbf{V}_{\bar{n},r}\|_Y + \sum_{\ell=0}^{\bar{n}-1}\sum_{r=0}^{\bar{m}-1} \|\mathbf{A}_{\bar{n}-\ell,\bar{m}-r}\mathbf{V}_{\ell,r}\|_Y \Bigg\}.$$

Using the estimates on $\mathbf{R}_{n,m}$ and $\mathbf{A}_{n,m}$ (for all n,m) and $\mathbf{V}_{n,m}$ ($n<\bar{n}, m<\bar{m}$) we define

$$\tilde{C}_A := C_{A,N}C_{A,M}, \quad \tilde{C}_R := C_{R,N}C_{R,M}, \quad \tilde{C}_V := C_{V,N}C_{V,M},$$

to form

$$\begin{aligned}
\|\mathbf{V}_{\bar{n},\bar{m}}\|_X &\le C_e\Bigg\{\tilde{C}_R B_R^{\bar{n}} D_R^{\bar{m}} + \sum_{\ell=0}^{\bar{n}-1} C_{A,N} B_A^{\bar{n}-\ell} C_{V,N} B_V^{\ell} + \sum_{r=0}^{\bar{m}-1} C_{A,M} D_A^{\bar{m}-\ell} C_{V,M} D_V^{\ell} \\
&\quad + \sum_{\ell=0}^{\bar{n}-1}\sum_{r=0}^{\bar{m}-1} \tilde{C}_A B_A^{\bar{n}-\ell} D_A^{\bar{m}-\ell} \tilde{C}_V B_V^{\ell} D_V^{r} \Bigg\} \\
&= C_e\tilde{C}_R B_R^{\bar{n}} D_R^{\bar{m}} + C_e C_{A,N} C_{V,N} \left(\frac{B_A}{B_V}\right) B_V^{\bar{n}} \sum_{\ell=0}^{\bar{n}-1} \left(\frac{B_A}{B_V}\right)^{\bar{n}-\ell-1} \\
&\quad + C_e C_{A,M} C_{V,M} \left(\frac{D_A}{D_V}\right) D_V^{\bar{m}} \sum_{r=0}^{\bar{m}-1} \left(\frac{D_A}{D_V}\right)^{\bar{m}-r-1} \\
&\quad + C_e \tilde{C}_A \tilde{C}_V \left(\frac{B_A}{B_V}\right) B_V^{\bar{n}} \left(\frac{D_A}{D_V}\right) D_V^{\bar{m}} \sum_{\ell=0}^{\bar{n}-1} \left(\frac{B_A}{B_V}\right)^{\bar{n}-\ell-1} \sum_{r=0}^{\bar{m}-1} \left(\frac{D_A}{D_V}\right)^{\bar{m}-r-1} \\
&\le C_e\tilde{C}_R B_V^{\bar{n}} D_V^{\bar{m}} + C_e C_{A,N} C_{V,N} \left(\frac{B_A}{B_V}\right) B_V^{\bar{n}} \left(\frac{1}{1-1/2}\right) \\
&\quad + C_e C_{A,M} C_{V,M} \left(\frac{D_A}{D_V}\right) D_V^{\bar{m}} \left(\frac{1}{1-1/2}\right) \\
&\quad + C_e \tilde{C}_A \tilde{C}_V \left(\frac{B_A}{B_V}\right) B_V^{\bar{n}} \left(\frac{D_A}{D_V}\right) D_V^{\bar{m}} \left(\frac{1}{1-1/2}\right)^2,
\end{aligned}$$

if $B_A/B_V \le 1/2$ and $D_A/D_V \le 1/2$ (implying $B_V \ge 2B_A$ and $D_V \ge 2D_A$). We are done if we demand that

$$\begin{aligned}
&B_V \ge B_R, \quad D_V \ge D_R, \quad C_e C_{R,N} \le C_{V,N}/2, \quad C_e C_{R,M} \le C_{V,M}/2, \\
&4C_e C_{A,N} C_{V,N} (B_A/B_V) \le C_{V,N}/2, \quad 4C_e C_{A,M} C_{V,M} (D_A/D_V) \le C_{V,M}/2.
\end{aligned}$$

This can be realized if

$$\begin{aligned}
&C_{V,N} \ge 2C_e C_{R,N}, \quad B_V \ge \max\{B_R, 2B_A, 8C_e C_{A,N} B_A\}, \\
&C_{V,M} \ge 2C_e C_{R,M}, \quad D_V \ge \max\{D_R, 2D_A, 8C_e C_{A,M} D_A\}.
\end{aligned}$$

□

4.6 Joint Analyticity of Solutions of the Two–Layer Problem

We recall the surface formulation of our scattering problem,

$$\mathbf{AV} = \mathbf{R},$$

cf. (4.14), where the operator $\mathbf{A}$ and vector $\mathbf{R}$ are given in (4.15). As discussed in §4.3, $\mathbf{V}$ is a vector of unknowns which contains solutions U and W to the scattering problem. As mentioned in the Introduction, our solution procedure is perturbative in nature and we can directly invoke Theorem 4.5.1 from §4.5 to obtain our desired result. For this we may formally expand

$$\mathbf{A}(\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbf{A}_{n,m}\varepsilon^n\delta^m, \quad \mathbf{R}(\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbf{R}_{n,m}\varepsilon^n\delta^m,$$

which we will justify rigorously, and seek a solution to (4.14) in the form

$$\mathbf{V}(\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbf{V}_{n,m}\varepsilon^n\delta^m, \tag{4.22}$$

where $\varepsilon,\delta \in \mathbb{R}$. Recalling our definitions from §2.7, we define the vector–valued spaces for $s \geq 0$

$$X^s := \left\{ \mathbf{V} = \begin{pmatrix} U \\ W \end{pmatrix} \middle| \, U, W \in H^{s+3/2}([0,d]) \right\},$$

and

$$Y^s := \left\{ \mathbf{R} = \begin{pmatrix} \zeta \\ -\psi \end{pmatrix} \middle| \, \zeta \in H^{s+3/2}([0,d]), \psi \in H^{s+1/2}([0,d]) \right\}.$$

These have the norms

$$\|\mathbf{V}\|_{X^s}^2 = \left\| \begin{pmatrix} U \\ W \end{pmatrix} \right\|_{X^s}^2 := \|U\|_{H^{s+3/2}}^2 + \|W\|_{H^{s+3/2}}^2 ,$$
$$\|\mathbf{R}\|_{Y^s}^2 = \left\| \begin{pmatrix} \zeta \\ -\psi \end{pmatrix} \right\|_{Y^s}^2 := \|\zeta\|_{H^{s+3/2}}^2 + \|\psi\|_{H^{s+1/2}}^2 .$$

We now state our main result.

Theorem 4.6.1. *Given an integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ then the equation* (4.14) *has a unique solution,* (4.22)*, and there exist constants $B, C, D > 0$ such that*

$$\|\mathbf{V}_{n,m}\|_{X^s} \leq CB^nD^m,$$

for all $n, m \geq 0$. This implies that for any $0 \leq \rho, \sigma < 1$, (4.22) *converges for all ε such that $B\varepsilon < \rho$, i.e., $\varepsilon < \rho/B$ and all δ such that $D\delta < \sigma$, i.e., $\delta < \sigma/D$.*

Proof. [Theorem 4.6.1] As mentioned above, our strategy is to invoke Theorem 4.5.1, thus we must verify the relevant hypotheses. To begin, we consider the spaces

$$X = X^s, \quad Y = Y^s.$$

In §4.7 we will show that the vector $\mathbf{R}_{n,m}$, consisting of $\zeta_{n,m}$ and $\psi_{n,m}$, is bounded in Y^s for any $s \geq 0$ provided that $f \in C^{s+2}([0,d])$. This implies that the $\mathbf{R}_{n,m}$ satisfies the estimates of Item 1 in Theorem 4.5.1.

In §2.10 (Theorem 2.10.2) and §3.9 (Theorem 3.9.2), we have previously shown that the operators $G_{n,m}$ and $J_{n,m}$ in the Taylor series expansions of the DNOs satisfy appropriate bounds provided that $f \in C^{s+2}([0,d])$. With these, it is clear that the $\mathbf{A}_{n,m}$ satisfy the estimates of Item 2 in Theorem 4.5.1.

Finally, in §4.8 we show that the estimates and mapping properties of $\mathbf{A}_{0,0}^{-1}$ for Item 3 in Theorem 4.5.1 hold where $\mathbf{A}_{0,0}$ is defined in (4.16) as the flat–interface version of our governing equations. □

4.7 Analyticity of the Surface Data

To establish the analyticity of the Dirichlet and Neumann data obeying suitable estimates, we begin by defining

$$\mathcal{E}(x;\varepsilon,\delta) := e^{-i(1+\delta)\underline{\gamma}^u \varepsilon f(x)},$$

and note that we can write (4.2e) and (4.2f) as

$$\begin{aligned}
\zeta(x) &= \zeta(x;\varepsilon,\delta) = -\mathcal{E}(x;\varepsilon,\delta),\\
\psi(x) &= \psi(x;\varepsilon,\delta) = \left\{ i(1+\delta)\underline{\gamma}^u + i(1+\delta)\underline{\alpha}(\varepsilon\partial_x f)\right\}\mathcal{E}(x;\varepsilon,\delta).
\end{aligned}$$

We will now demonstrate that the function $\mathcal{E}$ is jointly analytic in ε and δ, and subject to appropriate estimates, which clearly demonstrates the joint analytic dependence of the data, $\zeta(x;\varepsilon,\delta)$ and $\psi(x;\varepsilon,\delta)$.

Lemma 4.7.1. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ then the function $\mathcal{E}(x;\varepsilon,\delta)$ is jointly analytic in ε and δ. Therefore*

$$\mathcal{E}(x;\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathcal{E}_{n,m}(x)\varepsilon^n\delta^m, \tag{4.23}$$

and, for constants $C_{\mathcal{E}}, B_{\mathcal{E}}, D_{\mathcal{E}} > 0$,

$$\|\mathcal{E}_{n,m}\|_{H^{s+3/2}} \leq C_{\mathcal{E}} B_{\mathcal{E}}^n D_{\mathcal{E}}^m, \tag{4.24}$$

for all $n, m \geq 0$.

Proof. [Lemma 4.7.1] We begin by observing the classical fact that the composition of jointly (real) analytic functions is also jointly (real) analytic (99) so that (4.23) holds, and move to expressions and estimates for the $\mathcal{E}_{n,m}$. By evaluating at $\varepsilon = 0$ we find that

$$\mathcal{E}(x; 0, \delta) = 1,$$

so that

$$\mathcal{E}_{0,m}(x) = \begin{cases} 1, & m = 0, \\ 0, & m > 0. \end{cases}$$

For $\varepsilon > 0$ we use the straightforward computation

$$\partial_\varepsilon \mathcal{E} = \left\{ -i(1+\delta)\underline{\gamma}^u f \right\} \mathcal{E},$$

and the expansion (4.23) to learn that, for $m = 0$,

$$\mathcal{E}_{n+1,0} = \left(\frac{-i\underline{\gamma}^u f}{n+1} \right) \mathcal{E}_{n,0}, \tag{4.25}$$

and, for $m > 0$,

$$\mathcal{E}_{n+1,m} = \left(\frac{-i\underline{\gamma}^u f}{n+1} \right) \left\{ \mathcal{E}_{n,m} + \mathcal{E}_{n,m-1} \right\}. \tag{4.26}$$

We work by induction in n and begin by establishing (4.24) at $n = 0$ for all $m \geq 0$. This is immediate as

$$\|\mathcal{E}_{0,0}\|_{H^{s+3/2}} = 1, \quad \|\mathcal{E}_{0,m}\|_{H^{s+3/2}} = 0.$$

We now assume (4.24) for all $n < \bar{n}$ and all $m \geq 0$, and seek this estimate in the case $n = \bar{n}$ and all $m \geq 0$. For this we conduct another induction on m, and for $m = 0$ we use (4.25) (together with Lemma 2.7.1 with $\tilde{s} = s + 1$) to discover

$$\begin{aligned} \|\mathcal{E}_{\bar{n},0}\|_{H^{s+3/2}} &\leq \mathcal{M} \left(\frac{\left|\underline{\gamma}^u\right| |f|_{C^{s+3/2+\eta}}}{\bar{n}} \right) \|\mathcal{E}_{\bar{n}-1,0}\|_{H^{s+3/2}} \\ &\leq \mathcal{M} \left(\frac{\left|\underline{\gamma}^u\right| |f|_{C^{s+2}}}{\bar{n}} \right) C_{\mathcal{E}} B_{\mathcal{E}}^{\bar{n}-1} \leq C_{\mathcal{E}} B_{\mathcal{E}}^{\bar{n}}, \end{aligned}$$

provided that

$$B_{\mathcal{E}} \geq \mathcal{M} \left|\underline{\gamma}^u\right| |f|_{C^{s+2}} \geq \mathcal{M} \left(\frac{\left|\underline{\gamma}^u\right| |f|_{C^{s+2}}}{\bar{n}} \right).$$

Finally, we assume the estimate (4.24) for $n = \bar{n}$ and $m < \bar{m}$, and use (4.26) to learn that

$$\begin{aligned}\|\mathcal{E}_{\bar{n},\bar{m}}\|_{H^{s+3/2}} &\leq \mathcal{M}\left(\frac{\left|\underline{\gamma}^u\right||f|_{C^{s+3/2+\eta}}}{\bar{n}}\right)\left\{\|\mathcal{E}_{\bar{n}-1,\bar{m}}\|_{H^{s+3/2}} + \|\mathcal{E}_{\bar{n}-1,\bar{m}-1}\|_{H^{s+3/2}}\right\}\\ &\leq \mathcal{M}\left(\frac{\left|\underline{\gamma}^u\right||f|_{C^{s+2}}}{\bar{n}}\right) C_{\mathcal{E}}\left\{B_{\mathcal{E}}^{\bar{n}-1}D_{\mathcal{E}}^{\bar{m}} + B_{\mathcal{E}}^{\bar{n}-1}D_{\mathcal{E}}^{\bar{m}-1}\right\}\\ &\leq C_{\mathcal{E}}B_{\mathcal{E}}^{\bar{n}}D_{\mathcal{E}}^{\bar{m}},\end{aligned}$$

provided that

$$\mathcal{M}\left(\frac{\left|\underline{\gamma}^u\right||f|_{C^{s+2}}}{\bar{n}}\right) \leq \frac{B_{\mathcal{E}}}{2}, \quad \mathcal{M}\left(\frac{\left|\underline{\gamma}^u\right||f|_{C^{s+2}}}{\bar{n}}\right) \leq \frac{B_{\mathcal{E}}D_{\mathcal{E}}}{2},$$

which can be accomplished, e.g., with

$$B_{\mathcal{E}} \geq 2\mathcal{M}\left|\underline{\gamma}^u\right||f|_{C^{s+2}} \geq 2\mathcal{M}\left(\frac{\left|\underline{\gamma}^u\right||f|_{C^{s+2}}}{\bar{n}}\right), \quad D_{\mathcal{E}} \geq 1.$$

□

With Lemma 4.7.1 it is straightforward to prove the following analyticity result for the Dirichlet and Neumann data.

Lemma 4.7.2. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ then the functions $\zeta(x;\varepsilon,\delta)$ and $\psi(x;\varepsilon,\delta)$ are jointly analytic in ε and δ. Therefore*

$$\{\zeta,\psi\}(x;\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\{\zeta_{n,m},\psi_{n,m}\}(x)\varepsilon^n\delta^m \tag{4.27}$$

and, for constants $C_\zeta, B_\zeta, D_\zeta > 0$, and $C_\psi, B_\psi, D_\psi > 0$,

$$\|\zeta_{n,m}\|_{H^{s+3/2}} \leq C_\zeta B_\zeta^n D_\zeta^m, \quad \|\psi_{n,m}\|_{H^{s+1/2}} \leq C_\psi B_\psi^n D_\psi^m, \tag{4.28}$$

for all $n, m \geq 0$.

4.8 The Flat–Interface Problem

As we outlined in Theorem 4.6.1, the key to our developments (as with all regular perturbation arguments) is the flat–interface version of (4.14)

$$\mathbf{A}_{0,0}\mathbf{V}_{0,0} = \mathbf{R}_{0,0},$$

in particular the invertibility of $\mathbf{A}_{0,0}$ and the mapping properties of $\mathbf{A}_{0,0}^{-1}$. From (4.15), it is not hard to see that the formulas for $\mathbf{A}$ and $\mathbf{R}$ are

$$\mathbf{A}_{0,0} = \begin{pmatrix} I & -I \\ G_{0,0} & \tau^2 J_{0,0} \end{pmatrix}, \tag{4.29a}$$

$$\mathbf{A}_{n,m} = \begin{pmatrix} 0 & 0 \\ G_{n,m} & \tau^2 J_{n,m} \end{pmatrix} + \delta_{n,1}\left\{1+\delta_{m,1}\right\}(\partial_x f)(i\underline{\alpha}) \begin{pmatrix} 0 & 0 \\ 1 & -\tau^2 \end{pmatrix}, \qquad n \neq 0 \text{ or } m \neq 0, \tag{4.29b}$$

$$\mathbf{R}_{n,m} = \begin{pmatrix} \zeta_{n,m} \\ -\psi_{n,m} \end{pmatrix}, \tag{4.29c}$$

where $\delta_{n,m}$ is the Kronecker delta function. We note that $\mathbf{A}_{0,0}$ is diagonalized by the Fourier transform so that $\mathbf{A}_{0,0}\mathbf{V}_{n,m} = \mathbf{R}_{n,m}$ can be expressed as

$$\sum_{p=-\infty}^{\infty} \widehat{\mathbf{A}}_{0,0}(p)\widehat{\mathbf{V}}_{n,m}(p)e^{i\tilde{p}x} = \sum_{p=-\infty}^{\infty} \widehat{\mathbf{R}}_{n,m}(p)e^{i\tilde{p}x},$$

which implies

$$\widehat{\mathbf{V}}_{n,m}(p) = \left[\widehat{\mathbf{A}}_{0,0}(p)\right]^{-1} \widehat{\mathbf{R}}_{n,m}(p).$$

It is not difficult to see

$$\widehat{\mathbf{A}}_{0,0}(p) = \begin{pmatrix} 1 & -1 \\ (-i\gamma_p^u) & \tau^2(-i\gamma_p^w) \end{pmatrix},$$

cf. (4.29), implying

$$\left[\widehat{\mathbf{A}}_{0,0}(p)\right]^{-1} = \frac{1}{\hat{\Delta}_p} \begin{pmatrix} \tau^2(-i\gamma_p^w) & 1 \\ (i\gamma_p^u) & 1 \end{pmatrix}, \quad \hat{\Delta}_p := -(i\gamma_p^u + \tau^2(i\gamma_p^w)).$$

Remark 4.8.1. From these formulas it becomes obvious that the operator $\mathbf{A}_{0,0}$ is always invertible and our algorithm is well–defined. Recalling that we assume a dielectric in the upper layer (so that the incident radiation propagates) we have that γ_p^u is either real and positive or purely imaginary (with positive imaginary part). If a dielectric fills the lower layer then we have the same state of affairs for γ_p^w so that, given that τ^2 will be positive and real, $\Delta_p \neq 0$. Alternatively, if a metal fills the lower layer then γ_p^w will be complex with positive imaginary part. While it is less obvious, this ensures that, once again, $\Delta_p \neq 0$.

We now verify Item 3 in Theorem 4.5.1. By the analysis above, we know that

$$\mathbf{A}_{0,0} = \begin{pmatrix} I & -I \\ G_{0,0} & \tau^2 J_{0,0} \end{pmatrix}, \tag{4.30}$$

where

$$G_{0,0} = -i\gamma_D^u, \quad J_{0,0} = -i\gamma_D^w, \tag{4.31}$$

are order–one Fourier multipliers defined by

$$G_{0,0}[U] = \sum_{p=-\infty}^{\infty} (-i\gamma_p^u)\hat{U}_p e^{i\tilde{p}x}, \quad J_{0,0}[W] = \sum_{p=-\infty}^{\infty} (-i\gamma_p^w)\hat{W}_p e^{i\tilde{p}x}. \tag{4.32}$$

Lemma 4.8.1. *The linear operator* $\mathbf{A}_{0,0}$ *maps* X^s *to* Y^s *boundedly, is invertible, and its inverse maps* Y^s *to* X^s *boundedly.*

Proof. [Lemma 4.8.1] We begin by defining the operator

$$\Delta := G_{0,0} + \tau^2 J_{0,0} = (-i\gamma_D^u) + \tau^2(-i\gamma_D^w),$$

which has Fourier symbol

$$\hat{\Delta}_p = (-i\gamma_p^u) + \tau^2(-i\gamma_p^w),$$

and noting that there exist positive constants C_G, C_J, and C_Δ such that

$$\left|-i\gamma_p^u\right| \leq C_G \left\langle \tilde{p} \right\rangle, \quad \left|-i\gamma_p^w\right| \leq C_J \left\langle \tilde{p} \right\rangle, \quad \left|\hat{\Delta}_p\right| \leq C_\Delta \left\langle \tilde{p} \right\rangle.$$

Importantly, provided that $n^u \neq n^w$, it is not difficult to establish the crucial fact that $\hat{\Delta}_p \neq 0$. Finally, one can also find a positive constant $C_{\Delta^{-1}}$ such that

$$\left|\frac{1}{\hat{\Delta}_p}\right| \leq C_{\Delta^{-1}} \left\langle \tilde{p} \right\rangle^{-1}.$$

With this it is a simple matter to realize that Δ^{-1} exists and that

$$\Delta : H^{s+3/2} \to H^{s+1/2}, \quad \Delta^{-1} : H^{s+1/2} \to H^{s+3/2}.$$

Next, we write generic elements of X^s and Y^s as

$$\mathbf{V} = \begin{pmatrix} U \\ W \end{pmatrix} \in X^s, \quad \mathbf{R} = \begin{pmatrix} \zeta \\ -\psi \end{pmatrix} \in Y^s.$$

Using the definitions of the norms of X^s and Y^s, and the facts

$$2ab \leq a^2 + b^2, \quad \|A + B\|^2 \leq (\|A\| + \|B\|)^2,$$

we find that

$$\begin{aligned}
\|\mathbf{A}_{0,0}\mathbf{V}\|_{Y^s}^2 &= \|U-W\|_{H^{s+3/2}}^2 + \left\|G_{0,0}U+\tau^2 J_{0,0}W\right\|_{H^{s+1/2}}^2 \\
&\leq 2\|U\|_{H^{s+3/2}}^2 + 2\|W\|_{H^{s+3/2}}^2 + C_G^2\|U\|_{H^{s+3/2}}^2 \\
&\quad + \tau^2 C_G C_J\left(\|U\|_{H^{s+3/2}}^2 + \|W\|_{H^{s+3/2}}^2\right) + C_J^2\tau^4\|W\|_{H^{s+3/2}}^2 \\
&\leq \max\{2, C_G^2, \tau^2 C_G C_J, \tau^4 C_J^2\}\left(\|U\|_{H^{s+3/2}}^2 + \|W\|_{H^{s+3/2}}^2\right) \\
&= \max\{2, C_G^2, \tau^2 C_G C_J, \tau^4 C_J^2\}\|\mathbf{V}\|_{X^s}^2 ,
\end{aligned}$$

so that $\mathbf{A}_{0,0}$ does indeed map X^s to Y^s boundedly. We define the operator

$$\mathbf{B} := \Delta^{-1}\begin{pmatrix}\tau^2 J_{0,0} & I\\ -G_{0,0} & I\end{pmatrix},$$

and note that

$$\mathbf{B}\mathbf{A}_{0,0} = \mathbf{A}_{0,0}\mathbf{B} = \begin{pmatrix}I & 0\\ 0 & I\end{pmatrix},$$

so that the inverse of $\mathbf{A}_{0,0}$ exists and $\mathbf{A}_{0,0}^{-1} = \mathbf{B}$. Furthermore, as above,

$$\begin{aligned}
\left\|\mathbf{A}_{0,0}^{-1}\mathbf{R}\right\|_{X^s}^2 &= \left\|\Delta^{-1}(\tau^2 J_{0,0}\zeta - \psi)\right\|_{H^{s+3/2}}^2 + \left\|\Delta^{-1}(-G_{0,0}\zeta - \psi)\right\|_{H^{s+3/2}}^2 \\
&\leq C_{\Delta^{-1}}^2\tau^4 C_J^2\|\zeta\|_{H^{s+3/2}}^2 + C_{\Delta^{-1}}^2\tau^2 C_J\left(\|\zeta\|_{H^{s+3/2}}^2 + \|\psi\|_{H^{s+1/2}}^2\right) \\
&\quad + C_{\Delta^{-1}}^2 C_G^2\|\zeta\|_{H^{s+3/2}}^2 + C_{\Delta^{-1}}^2 C_G\left(\|\zeta\|_{H^{s+3/2}}^2 + \|\psi\|_{H^{s+1/2}}^2\right) \\
&\quad + 2C_{\Delta^{-1}}^2\|\psi\|_{H^{s+1/2}}^2 \\
&\leq C_{\Delta^{-1}}^2\max\{2, C_G, C_G^2, \tau^2 C_J, \tau^4 C_J^2\}\left(\|\zeta\|_{H^{s+3/2}}^2 + \|\psi\|_{H^{s+1/2}}^2\right) \\
&= C_{\Delta^{-1}}^2\max\{2, C_G, C_G^2, \tau^2 C_J, \tau^4 C_J^2\}\|\mathbf{R}\|_{Y^s}^2 ,
\end{aligned}$$

and $\mathbf{A}_{0,0}^{-1}$ maps Y^s to X^s boundedly. □

CHAPTER 5

VALIDATION OF THE NUMERICAL SCHEME

5.1 Introduction

Verification of codes that numerically approximate solutions of partial differential equations entails establishing that the code is free of coding mistakes and capable of reaching exact mathematical solutions given appropriate discretization (100; 101). This necessitates the assessment of discretization errors using well–known benchmark solutions. Exact analytical solutions with a sufficiently complicated solution structure are the ideal benchmarks; they don't have to be physically realistic because verification is a purely mathematical endeavor. The Method of Manufactured Solutions (MMS) describes a simple and general procedure for producing such solutions and we now focus on applying the MMS to our HOPS/AWE algorithm.

5.2 The Method of Manufactured Solutions

To validate our numerical scheme we utilized the MMS (102; 103; 104). To summarize, we considered a general system of partial differential equations subject to generic boundary conditions

$$\begin{aligned} \mathcal{P}v &= 0, && \text{in } \Omega, \\ \mathcal{B}v &= 0, && \text{at } \partial\Omega. \end{aligned}$$

It is typically easy to implement a numerical algorithm to solve the nonhomogeneous version of this set of equations

$$\begin{aligned} \mathcal{P}v &= \mathcal{F}, && \text{in } \Omega, \\ \mathcal{B}v &= \mathcal{J}, && \text{at } \partial\Omega. \end{aligned}$$

To test an implementation we began with the "manufactured solution," $\tilde{v}$, and set

$$\mathcal{F}_v := \mathcal{P}\tilde{v}, \quad \mathcal{J}_v := \mathcal{J}\tilde{v}.$$

Thus, given the pair $\{\mathcal{F}_v, \mathcal{J}_v\}$ we had an exact solution of the nonhomogeneous problem, namely $\tilde{v}$. While this does not prove an implementation to be correct, if the function $\tilde{v}$ is chosen to imitate the behavior of anticipated solutions (e.g., satisfying the boundary conditions exactly) then this gives us confidence in our algorithm.

5.3 Manufactured Solutions

We considered periodic, outgoing solutions of the Helmholtz equation (4.8a)

$$u_p(x,z) := A_p e^{i\tilde{p}x} e^{i\gamma_p^u z}, \quad p \in \mathbb{Z}, \quad A_p \in \mathbb{C}, \tag{5.1}$$

and their counterparts for (4.8b)

$$w_p(x,z) := B_p e^{i\tilde{p}x} e^{-i\gamma_p^w z}, \quad p \in \mathbb{Z}, \quad B_p \in \mathbb{C}. \tag{5.2}$$

We then defined, for a particular choice of p,

$$\xi^u := u_p(x,g(x)), \quad \nu^u := -\partial_N u_p(x,g(x)), \tag{5.3a}$$

$$\xi^w := w_p(x,g(x)), \quad \nu^w := \partial_N w_p(x,g(x)). \tag{5.3b}$$

To validate the two–layer solver we set

$$\zeta = \xi^u - \xi^w, \quad \psi = -\nu^u - \tau^2 \nu^w. \tag{5.4}$$

In order to test our implementation of the recursions, (5.4), we required the joint expansion of ξ^u, ξ^w, ν^u, and ν^w in ε and δ. In analogy to our developments in §4.7 we defined

$$\mathcal{E}^{q,p}(x;\varepsilon,\delta) := \exp\left\{\pm i\gamma_p^q(\delta)\varepsilon f(x)\right\}, \quad q \in \{u,w\},$$

and then derived the terms $\mathcal{E}^{q,p}_{n,m}$ in the expansion

$$\mathcal{E}^{q,p}(x;\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathcal{E}^{q,p}_{n,m}(x)\varepsilon^n\delta^m.$$

From these it was clear that

$$\xi^u_{n,m}(x) = A_p e^{i\tilde{p}x}\mathcal{E}^{u,p}_{n,m}(x), \quad \xi^w_{n,m}(x) = B_p e^{i\tilde{p}x}\mathcal{E}^{w,p}_{n,m}(x), \tag{5.5a}$$

$$\nu^u_{n,m}(x) = \left(-i\gamma_p^u + i\tilde{p}\varepsilon f_x(x)\right)\xi^u_{n,m}(x), \quad \nu^w_{n,m}(x) = \left(i\gamma_p^w + i\tilde{p}\varepsilon f_x(x)\right)\xi^w_{n,m}(x). \tag{5.5b}$$

We then considered the surface data for our two–layer scattering problem (cf. §4.2)

$$\zeta(x) = -e^{-i\gamma^u g(x)}, \quad \psi(x) = (i\gamma^u + i\alpha(\partial_x g))e^{-i\gamma^u g(x)}, \tag{5.6}$$

and performed a joint expansion of ζ and ψ in ε and δ

$$\zeta_{n,m}(x) = -\mathcal{E}^{u,p}_{n,m}(x), \tag{5.7a}$$

$$\psi_{n,m}(x) = \left(i\gamma_p^u + i\alpha\varepsilon f_x(x)\right)\mathcal{E}^{u,p}_{n,m}(x), \tag{5.7b}$$

where we outline our procedure to find the $\mathcal{E}_{n,m}^{q,p}$ in §5.5.

5.4 Taylor Series for $\gamma_p^q(\delta)$

A key step in the development of our algorithm is to derive the Taylor series expansion for γ_p^q, where

$$\gamma_p^q = \gamma_p^q(\delta) = \sum_{m=0}^{\infty} \gamma_{p,m}^q \delta^m. \tag{5.8}$$

We started with the relationship

$$\alpha_p^2 + (\gamma_p^q)^2 = (k^q)^2,$$

which implies

$$\left(\sum_{m=0}^{\infty} \gamma_{p,m}^q \delta^m\right)\left(\sum_{r=0}^{\infty} \gamma_{p,r}^q \delta^r\right) = (1+\delta^2)(\underline{k}^q)^2 - (\underline{\alpha}_p + \delta\underline{\alpha})^2.$$

This gives

$$\begin{aligned}\sum_{m=0}^{\infty} \delta^m \sum_{r=0}^{m} \gamma_{p,m-r}^q \gamma_{p,r}^q &= \{(\underline{k}^q)^2 - (\underline{\alpha}_p)^2\} + 2\delta\ \{(\underline{k}^q)^2 - \underline{\alpha}\ \underline{\alpha}_p\} + \delta^2\ \{(\underline{k}^q)^2 - (\underline{\alpha})^2\} \\ &= (\underline{\gamma}_p^q)^2 + 2\delta\ \{(\underline{k}^q)^2 - \underline{\alpha}\ \underline{\alpha}_p\} + \delta^2(\underline{\gamma}^q)^2.\end{aligned}$$

Therefore at order $\mathcal{O}(\delta^0)$ we required

$$\gamma_{p,0}^q = \pm\underline{\gamma}_p^q, \tag{5.9}$$

and at order $\mathcal{O}(\delta^1)$ we required

$$\gamma_{p,1}^q = \frac{2((\underline{k}^q)^2 - \underline{\alpha}\ \underline{\alpha}_p)}{2\gamma_{p,0}^q}, \quad \gamma_{p,0}^q \neq 0. \tag{5.10}$$

This implies that it is crucial that $\underline{\gamma}_p^q \neq 0$ for all p in order to have a valid expansion of (5.8). The $\underline{\gamma}_p^q$ satisying $\underline{\gamma}_p^q = 0$ are known as a Rayleigh singularity (or Wood's anomaly). So we made this assumption, $\underline{\gamma}_p^q \neq 0$, and continued our development to $\mathcal{O}(\delta^2)$ where

$$\gamma_{p,2}^q = \frac{(\underline{\gamma}^q)^2 - (\gamma_{p,1}^q)^2}{2\gamma_{p,0}^q}, \quad \gamma_{p,0}^q \neq 0, \tag{5.11}$$

and for $\mathcal{O}(\delta^m)$, $m > 2$, we required

$$\gamma_{p,2}^q = \frac{-\sum_{r=1}^{m-1} \gamma_{p,m-r}^q \gamma_{p,r}^q}{2\gamma_{p,0}^q}, \quad \gamma_{p,0}^q \neq 0. \tag{5.12}$$

Remark 5.4.1. As discussed in (75) we must be away from a Rayleigh singularity, $\underline{\gamma}_p^q = 0$, for all p in order for our expansion to be valid. See the final section of (75) for a discussion of the behavior of the function $\gamma_p^q(\delta)$ in the neighborhood of a Rayleigh singularity.

5.5 Taylor Series for $\mathcal{E}^{q,p}(x;\varepsilon,\delta)$

Returning to our joint expansion

$$\mathcal{E}^{q,p}(x;\varepsilon,\delta) = \sum_{n=0}^{\infty}\sum_{n=0}^{\infty} \mathcal{E}_{n,m}^{q,p}(x)\varepsilon^n\delta^m,$$

we first calculated the Dirichlet data, (5.5a), when $n = 0$. We have

$$\mathcal{E}^{q,p}(x;0,\delta) = \exp\{\pm 0\} = 1,$$

therefore

$$\mathcal{E}_{0,m}^{q,p}(x) = \begin{cases} 1, & m = 0, \\ 0, & m > 0, \end{cases}$$

and

$$\xi_{0,m}^{u} = \begin{cases} A_p e^{i\tilde{p}x}, & m = 0, \\ 0, & m > 0, \end{cases}, \quad \xi_{0,m}^{w} = \begin{cases} B_p e^{i\tilde{p}x}, & m = 0, \\ 0, & m > 0. \end{cases}$$

We then evaluated (5.5a) when $n > 0$. Following the technique of Pourahmadi (105) (and of Marchant and Roberts (106; 107)), we observed that

$$\partial_\varepsilon \mathcal{E}^{q,p}(x;\varepsilon,\delta) = \left(\pm i\gamma_p^u(\delta)f(x)\right)\mathcal{E}^{q,p}(x;\varepsilon,\delta). \tag{5.13}$$

Inserting the Taylor series expansions for $\mathcal{E}^{q,p}$ and γ_p^q gives

$$\sum_{n=1}^{\infty}\sum_{m=0}^{\infty} \mathcal{E}_{n,m}^{q,p} n\varepsilon^{n-1}\delta^m = (\pm if)\left(\sum_{r=0}^{\infty}\gamma_{p,r}^q\delta^r\right)\left(\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathcal{E}_{n,m}^{q,p}\varepsilon^n\delta^m\right).$$

Re–indexing the left–hand side and rearranging the order of terms on the right–hand side forms

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathcal{E}_{n+1,m}^{q,p}(n+1)\varepsilon^n\delta^m = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left((\pm if)\sum_{r=0}^{m}\gamma_{p,m-r}^q\mathcal{E}_{n,r}^{q,p}\right)\varepsilon^n\delta^m.$$

Upon equating like orders we found

$$\mathcal{E}_{n+1,m}^{q,p} = \pm\frac{if}{n+1}\sum_{r=0}^{m}\gamma_{p,m-r}^q\mathcal{E}_{n,r}^{q,p}.$$

Therefore we have

$$\begin{aligned}\xi^u_{n+1,m} &= A_p e^{i\tilde{p}x}\frac{if}{n+1}\sum_{r=0}^{m}\gamma^u_{p,m-r}\mathcal{E}^{u,p}_{n,r},\\ \xi^w_{n+1,m} &= -B_p e^{i\tilde{p}x}\frac{if}{n+1}\sum_{r=0}^{m}\gamma^w_{p,m-r}\mathcal{E}^{w,p}_{n,r},\end{aligned} \tag{5.14}$$

where the initial data is

$$\xi^u_{0,m} = \begin{cases} A_p e^{i\tilde{p}x}, & m=0,\\ 0, & m>0,\end{cases}, \quad \xi^w_{0,m} = \begin{cases} B_p e^{i\tilde{p}x}, & m=0,\\ 0, & m>0.\end{cases} \tag{5.15}$$

As (5.14) and (5.15) are valid for all values of m, we see that to find the coefficient at order $(n+1,m)$, one only needs the values of $(n,0),\ldots,(n,m)$. As an example, we have $\xi^q_{0,m}$ from (5.15) which can be used to obtain $\xi^q_{1,m}$ by (5.14). We can then recover all of the $\xi^q_{n,m}$.

We then calculated the Neumann data, (5.5b), when $n=0$. We have

$$\nu^u_{0,m} = \begin{cases} -i\gamma^u_{p,0}\xi^u_{0,0}, & m=0,\\ -i\gamma^u_{p,m}\xi^u_{0,m}, & m>0,\end{cases}, \quad \nu^w_{0,m} = \begin{cases} i\gamma^w_{p,0}\xi^w_{0,0}, & m=0,\\ i\gamma^w_{p,m}\xi^w_{0,m}, & m>0,\end{cases}$$

therefore

$$\nu^u_{0,m} = \begin{cases} -i\gamma^u_{p,0}A_p e^{i\tilde{p}x}, & m=0,\\ 0, & m>0,\end{cases} \quad \nu^w_{0,m} = \begin{cases} i\gamma^w_{p,0}B_p e^{i\tilde{p}x}, & m=0,\\ 0, & m>0.\end{cases}$$

For (5.5b) and $n>0$ we inserted the Taylor series expansions for ξ^q and γ^q_p and used (5.13) to deduce

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}\nu^q_{n,m}n\varepsilon^{n-1}\delta^m &= f\left(\sum_{r=0}^{\infty}\gamma^q_{p,r}\delta^r\right)\left(\sum_{k=0}^{\infty}\gamma^q_{p,k}\delta^k\right)\left(\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\xi^q_{n,m}\varepsilon^n\delta^m\right)\\ &\quad \mp(\tilde{p}ff_x)\left(\sum_{r=0}^{\infty}\gamma^q_{p,r}\delta^r\right)\left(\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}\xi^q_{n-1,m}\varepsilon^n\delta^m\right).\end{aligned}$$

Re–indexing the left–hand side and rearranging the order of terms on the right–hand side forms

$$\begin{aligned}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\nu^q_{n+1,m}(n+1)\varepsilon^n\delta^m &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left(f\sum_{r=0}^{m}\sum_{k=0}^{r}\gamma^q_{p,m-r}\gamma^q_{p,r-k}\xi^q_{n,k}\right)\varepsilon^n\delta^m\\ &\quad \mp\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}\left((\tilde{p}ff_x)\sum_{r=0}^{m}\gamma^q_{p,m-r}\xi^q_{n-1,r}\right)\varepsilon^n\delta^m.\end{aligned}$$

Upon equating like orders we found

$$\begin{aligned}\nu^u_{n+1,m} &= \frac{f}{n+1}\sum_{r=0}^{m}\sum_{k=0}^{r}\gamma^u_{p,m-r}\gamma^u_{p,r-k}\xi^u_{n,k} - \frac{\tilde{p}ff_x}{n+1}\sum_{r=0}^{m}\gamma^u_{p,m-r}\xi^u_{n-1,r},\\ \nu^w_{n+1,m} &= \frac{f}{n+1}\sum_{r=0}^{m}\sum_{k=0}^{r}\gamma^w_{p,m-r}\gamma^w_{p,r-k}\xi^w_{n,k} + \frac{\tilde{p}ff_x}{n+1}\sum_{r=0}^{m}\gamma^w_{p,m-r}\xi^w_{n-1,r},\end{aligned} \tag{5.16}$$

where our initial data is

$$\nu^u_{0,m} = \begin{cases} -i\gamma^u_{p,0}A_p e^{i\tilde{p}x}, & m=0,\\ 0, & m>0,\end{cases} \qquad \nu^w_{0,m} = \begin{cases} i\gamma^w_{p,0}B_p e^{i\tilde{p}x}, & m=0,\\ 0, & m>0.\end{cases} \tag{5.17}$$

Analogously to the Dirichlet data, we see that (5.16) and (5.17) are valid for all values of m. Therefore we can recover the coefficient at order $(n+1,m)$ by the values of the coefficients at order $(n,0),\ldots,(n,m)$.

Finally, we calculated the surface data, (5.7a), when $n=0$. We have

$$\mathcal{E}^{u,p}(x;0,\delta) = \exp\{-0\} = 1,$$

therefore

$$\mathcal{E}^{u,p}_{0,m}(x) = \begin{cases} 1, & m=0,\\ 0, & m>0,\end{cases}$$

and

$$\zeta_{0,m} = \begin{cases} -1, & m=0,\\ 0, & m>0.\end{cases}$$

We then evaluated (5.7a) when $n>0$. Inserting the Taylor series expansions for $\mathcal{E}^{u,p}$ and γ^u_p and applying (5.13) gives

$$\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}\mathcal{E}^{u,p}_{n,m}n\varepsilon^{n-1}\delta^m = (-if)\left(\sum_{r=0}^{\infty}\gamma^u_{p,r}\delta^r\right)\left(\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathcal{E}^{u,p}_{n,m}\varepsilon^n\delta^m\right).$$

Re–indexing the left–hand side and rearranging the order of terms on the right–hand side forms

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mathcal{E}^{u,p}_{n+1,m}(n+1)\varepsilon^n\delta^m = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left((-if)\sum_{r=0}^{m}\gamma^u_{p,m-r}\mathcal{E}^{u,p}_{n,r}\right)\varepsilon^n\delta^m.$$

Upon equating like orders we found

$$\mathcal{E}^{u,p}_{n+1,m} = -\frac{if}{n+1}\sum_{r=0}^{m}\gamma^u_{p,m-r}\mathcal{E}^{u,p}_{n,r}.$$

Therefore we have

$$\zeta_{n+1,m} = \frac{if}{n+1} \sum_{r=0}^{m} \gamma_{p,m-r}^{u} \mathcal{E}_{n,r}^{u,p}, \tag{5.18}$$

where the initial data is

$$\zeta_{0,m} = \begin{cases} -1, & m = 0, \\ 0, & m > 0. \end{cases} \tag{5.19}$$

We then evaluated (5.7b) when $n = 0$. We have

$$\psi_{0,m} = \begin{cases} i\gamma_{p,0}^{u} \mathcal{E}_{0,m}^{u,p}, & m = 0, \\ i\gamma_{p,m}^{u} \mathcal{E}_{0,m}^{u,p}, & m > 0, \end{cases}$$

therefore

$$\psi_{0,m} = \begin{cases} i\gamma_{p,0}^{u}, & m = 0, \\ 0, & m > 0. \end{cases}$$

For (5.7b) and $n > 0$ we expanded

$$\alpha = \alpha(\delta) = \sum_{k=0}^{\infty} \alpha_k \delta^k,$$

and inserted the Taylor series expansions for α, ξ^q, and γ_p^q and used (5.13) to deduce

$$\begin{aligned} \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \psi_{n,m} n \varepsilon^{n-1} \delta^m = f \left(\sum_{r=0}^{\infty} \gamma_{p,r}^{u} \delta^r \right) \left(\sum_{k=0}^{\infty} \gamma_{p,k}^{u} \delta^k \right) \left(\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \mathcal{E}_{n,m}^{u} \varepsilon^n \delta^m \right) \\ + f f_x \left(\sum_{r=0}^{\infty} \gamma_{p,r}^{u} \delta^r \right) \left(\sum_{k=0}^{\infty} \alpha_k \delta^k \right) \left(\sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \mathcal{E}_{n-1,m}^{u} \varepsilon^n \delta^m \right). \end{aligned}$$

Re–indexing the left–hand side and rearranging the order of terms on the right–hand side forms

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \psi_{n+1,m} (n+1) \varepsilon^n \delta^m = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \left((f) \sum_{r=0}^{m} \sum_{k=0}^{r} \gamma_{p,m-r}^{u} \gamma_{p,r-k}^{u} \mathcal{E}_{n,k}^{u} \right) \varepsilon^n \delta^m \\ + \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \left((f f_x) \sum_{r=0}^{m} \sum_{k=0}^{r} \gamma_{p,m-r}^{u} \alpha_{r-k} \mathcal{E}_{n-1,k}^{u} \right) \varepsilon^n \delta^m. \end{aligned}$$

Upon equating like orders we found

$$\begin{aligned} \psi_{n+1,m} = \frac{f}{n+1} \sum_{r=0}^{m} \sum_{k=0}^{r} \gamma_{p,m-r}^{u} \gamma_{p,r-k}^{u} \mathcal{E}_{n,k}^{u} \\ + \frac{f f_x}{n+1} \sum_{r=0}^{m} \sum_{k=0}^{r} \gamma_{p,m-r}^{u} \alpha_{r-k} \mathcal{E}_{n-1,k}^{u}, \end{aligned} \tag{5.20}$$

where the initial data is

$$\psi_{0,m} = \begin{cases} i\gamma^u_{p,0}, & m = 0, \\ 0, & m > 0. \end{cases} \tag{5.21}$$

As before, we can find the coefficient at order $(n+1, m)$ by the values of the coefficients at $(n, 0), \ldots, (n, m)$.

5.6 The Domain of Analyticity

While the precise domain of analyticity of our solutions in (ε, δ) cannot be specified, it is clear that the expansion of $\gamma^q_p(\delta)$ only converges for δ away from the Rayleigh singularities. Therefore, our expansions are only valid on subsets of the (ε, δ)–plane away from Rayleigh singularities. For instance, in the upper layer, Rayleigh singularities occur when $\underline{\alpha}^2_p = (\underline{k}^u)^2$ which implies

$$\underline{\omega} = \pm \frac{c_0}{n^u} \left\{ \underline{\alpha} + \frac{2\pi p}{d} \right\}, \quad \text{for any } p \in \mathbb{Z}. \tag{5.22}$$

In the interest of maximizing our choice of δ we selected a "mid–point" value of $\underline{\omega}$ which is as far away as possible from consecutive Rayleigh singularities

$$\underline{\omega}_q := \frac{c_0}{n^u} \left\{ \underline{\alpha} + \frac{2\pi(q + 1/2)}{d} \right\}. \tag{5.23}$$

About this value the nearest singularities are

$$\begin{aligned} \underline{\omega}^-_q &:= \frac{c_0}{n^u} \left\{ \underline{\alpha} + \frac{2\pi q}{d} \right\} = \underline{\omega}_q - \frac{\pi c_0}{n^u d}, \\ \underline{\omega}^+_q &:= \frac{c_0}{n^u} \left\{ \underline{\alpha} + \frac{2\pi(q+1)}{d} \right\} = \underline{\omega}_q + \frac{\pi c_0}{n^u d}, \end{aligned}$$

so to maximize our range of ω we choose, for some filling fraction $0 < \sigma < 1$,

$$\underline{\omega}_q - \sigma \left(\frac{\pi c_0}{n^u d} \right) < \omega < \underline{\omega}_q + \sigma \left(\frac{\pi c_0}{n^u d} \right).$$

To express this in terms of δ we recall that $\omega = (1 + \delta)\underline{\omega}_q$ which gives

$$-\sigma \left(\frac{\pi c_0}{\underline{\omega}_q n^u d} \right) < \delta < \sigma \left(\frac{\pi c_0}{\underline{\omega}_q n^u d} \right).$$

Simplifying gives

$$-\left(\frac{\sigma}{(\underline{\alpha} d/\pi) + 2q + 1} \right) < \delta < \left(\frac{\sigma}{(\underline{\alpha} d/\pi) + 2q + 1} \right). \tag{5.24}$$

5.7 Numerical Results

Following our analysis in §5.3, we considered a wavenumber $p = r$ and defined the Dirichlet and Neumann traces

$$\xi_r^u(x) := u_r(x, g(x)), \quad \nu_r^u(x) := -\partial_N u_r(x, g(x)), \tag{5.25a}$$

$$\xi_r^w(x) := w_r(x, g(x)), \quad \nu_r^w(x) := \partial_N w_r(x, g(x)). \tag{5.25b}$$

From these we defined the two–layer data to be provided to our algorithm

$$\zeta_r := \xi_r^u - \xi_r^w, \quad \psi_r := -\nu_r^u - \tau^2 \xi_r^w, \tag{5.25c}$$

cf. (5.4). We selected the profile

$$g(x) = \varepsilon f(x) = \varepsilon \left(\frac{\cos(4x)}{4} \right), \tag{5.26}$$

with the following physical parameters

$$d = 2\pi, \quad \alpha = 0, \quad \epsilon^u = 1, \quad \epsilon^w = 1.1, \quad r = 4, \quad A_r = 5, \quad B_r = 3, \tag{5.27}$$

in TM polarization, and the numerical parameters

$$N_x = 32, \quad N_z = 32, \quad a = 1, \quad b = -1. \tag{5.28}$$

With a rescaling of the frequency (e.g., via a change of the time variable, $t' = t/c_0$) we arrange for $c_0 = 1$ and considered the base frequency

$$\underline{\omega}_1 = 3/2,$$

and filling fraction $\sigma = 0.99$. To illuminate the behavior of our scheme we studied four choices of the numerical parameter

$$N = M = 4, 8, 12, 16,$$

and the physical quantities

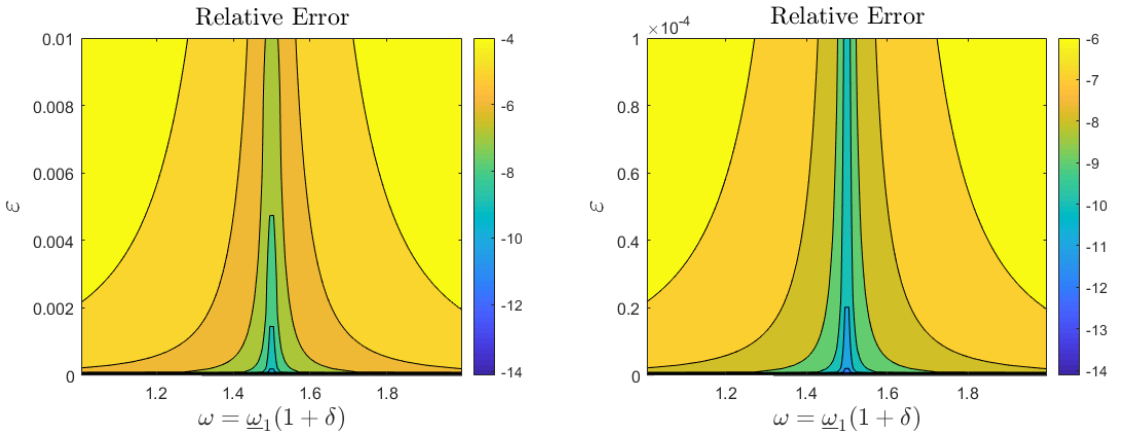
$$\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8},$$

in (5.26). For this we supplied the "exact" input data, $\{\zeta_r, \psi_r\}$, from (5.25) to our HOPS/AWE algorithm to simulate solutions of the two–layer problem giving $\{\xi_r^{u,\text{approx}}, \xi_r^{w,\text{approx}}\}$.

We compared this with the "exact" solutions $\{\xi_r^{u,\text{exact}}, \xi_r^{w,\text{exact}}\}$ and computed the relative error

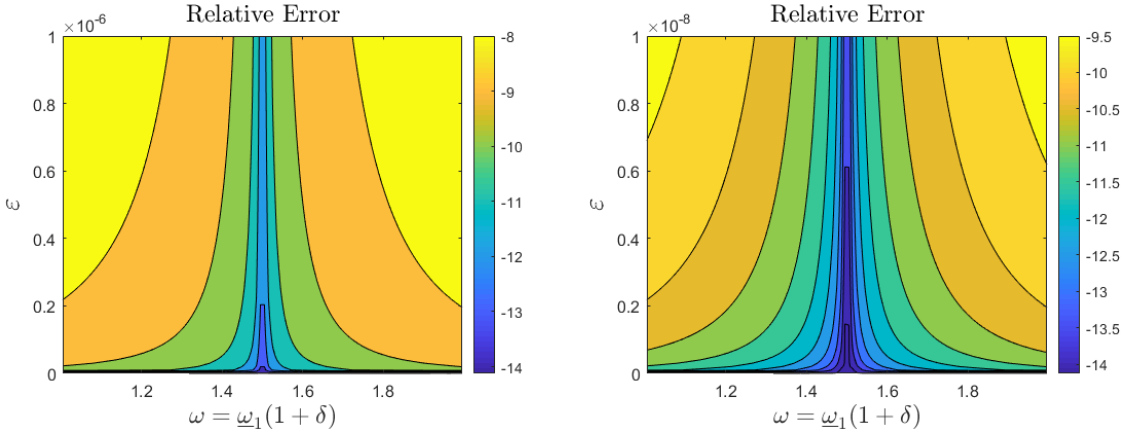
$$\text{Error}_{\text{rel}} := \frac{\left|\xi_r^{u,\text{exact}} - \xi_r^{u,\text{approx}}\right|_\infty}{\left|\xi_r^{u,\text{exact}}\right|_\infty}.$$

The results of our simulations are shown in Figures 13 and 14. More specifically, Figure 13 displays both the rapid and stable decay of the relative error for fixed N and M, and how this rate of decay improves as (ε, δ) decrease. Figure 14 shows both how the error shrinks as (ε, δ) become smaller, and that this rate is enhanced as both N and M are increased.



(a) $N = M = 4, \varepsilon = 10^{-2}$

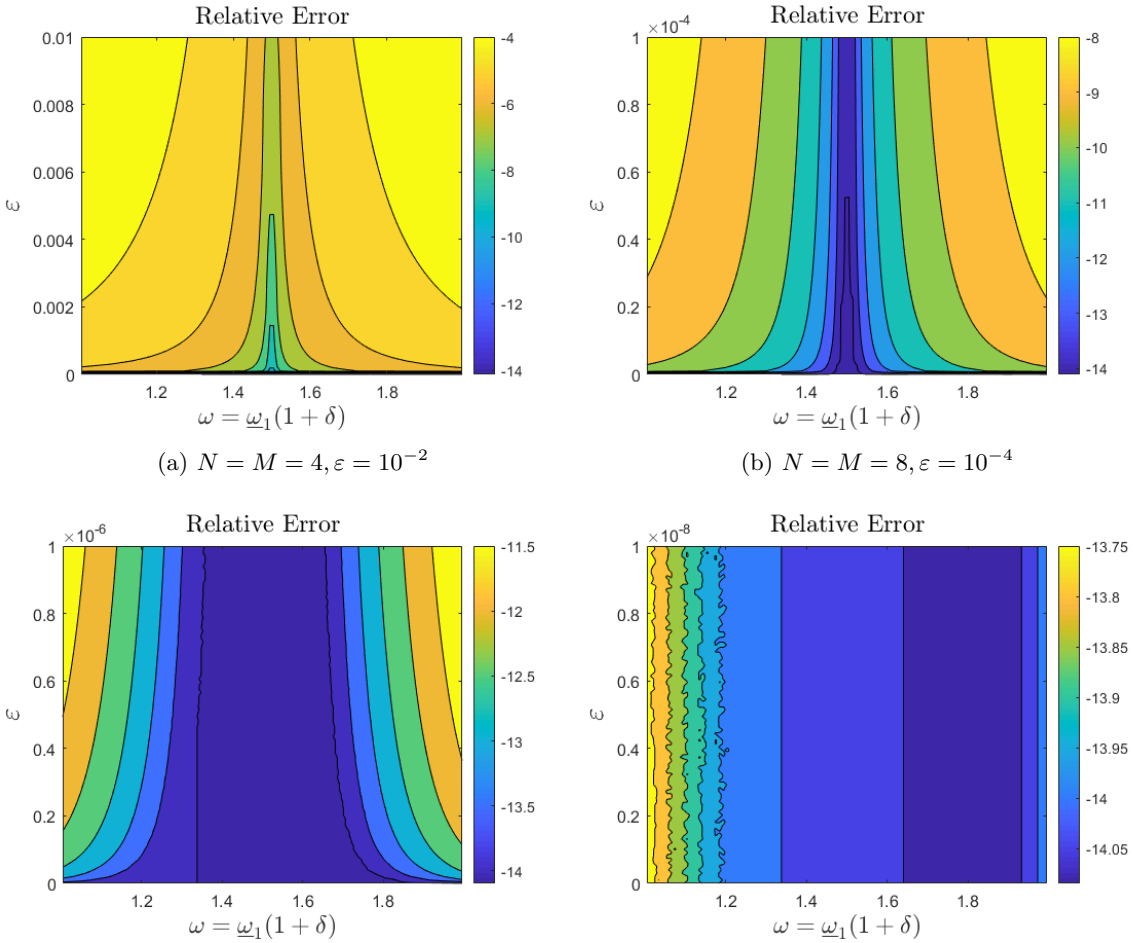
(b) $N = M = 4, \varepsilon = 10^{-4}$



(c) $N = M = 4, \varepsilon = 10^{-6}$

(d) $N = M = 4, \varepsilon = 10^{-8}$

Figure 13: Plot of relative error in the upper layer with fixed $N = M = 4$ and four choices of $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ with Taylor summation. Physical parameters were (5.27) and numerical discretization was (5.28).



(a) $N = M = 4, \varepsilon = 10^{-2}$

(b) $N = M = 8, \varepsilon = 10^{-4}$

(c) $N = M = 12, \varepsilon = 10^{-6}$

(d) $N = M = 16, \varepsilon = 10^{-8}$

Figure 14: Plot of relative error in the upper layer with four choices of $N = M = 4, 8, 12, 16$ and four choices of $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ with Taylor summation. Physical parameters were (5.27) and numerical discretization was (5.28).

We then analyzed the lower lower–layer Dirichlet data with the sinusoidal profile

$$g(x) = \varepsilon f(x) = \varepsilon \left(\frac{\sin(3x)}{3} \right), \tag{5.29}$$

We used the following physical parameters

$$d = 2\pi, \quad \alpha = 0, \quad \epsilon^u = 1, \quad \epsilon^w = 1.1, \quad r = 8, \quad A_r = 4, \quad B_r = 5, \tag{5.30}$$

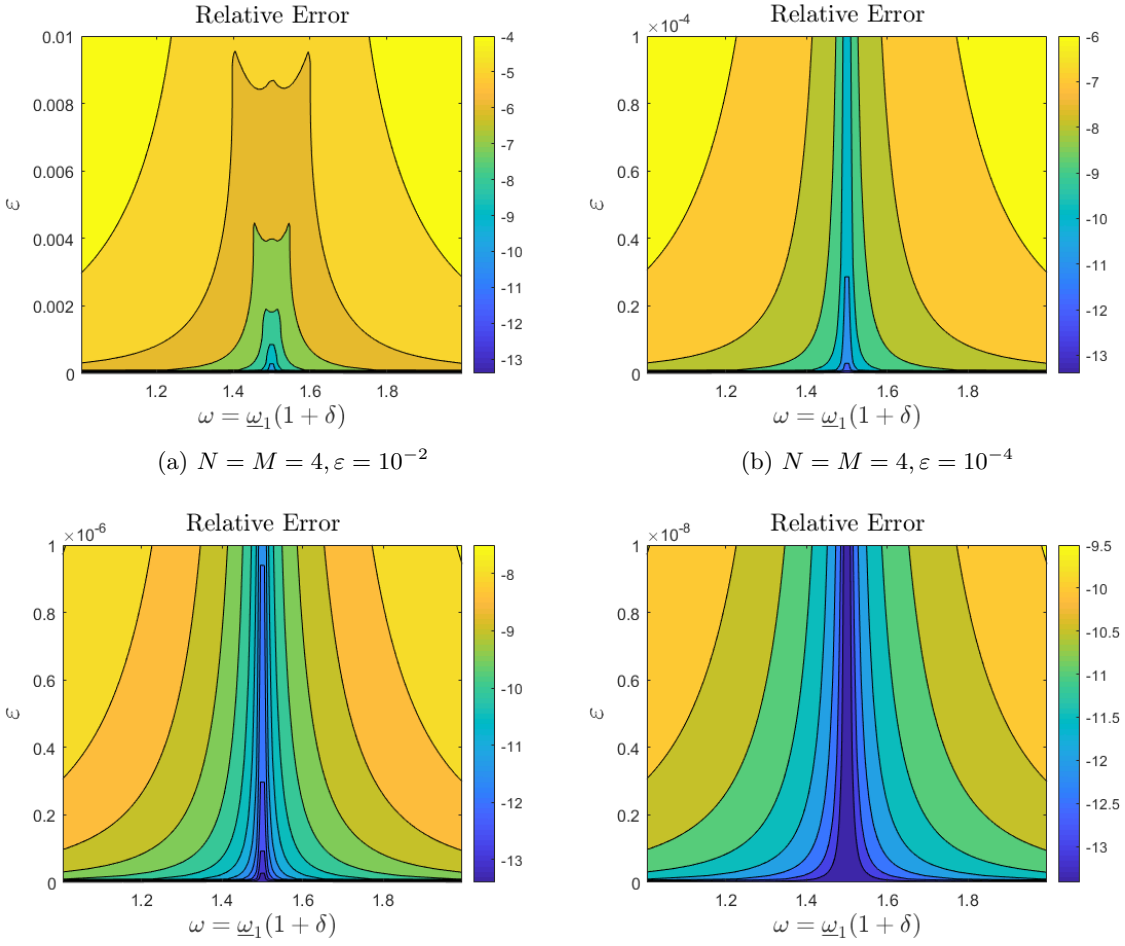
in TM polarization, and the numerical parameters

$$N_x = 32, \quad N_z = 32, \quad a = 1, \quad b = -1. \tag{5.31}$$

With these, we computed the relative error

$$\text{Error}_{\text{rel}} := \frac{\left|\xi_r^{w,\text{exact}} - \xi_r^{w,\text{approx}}\right|_\infty}{\left|\xi_r^{w,\text{exact}}\right|_\infty}.$$

The results of our simulations are shown in Figures 15 and 16. More specifically, Figure 15 displays both the rapid and stable decay of the relative error for fixed N and M, and how this rate of decay improves as (ε, δ) decrease. Figure 16 shows both how the error shrinks as (ε, δ) become smaller, and that this rate is enhanced as both N and M are increased.



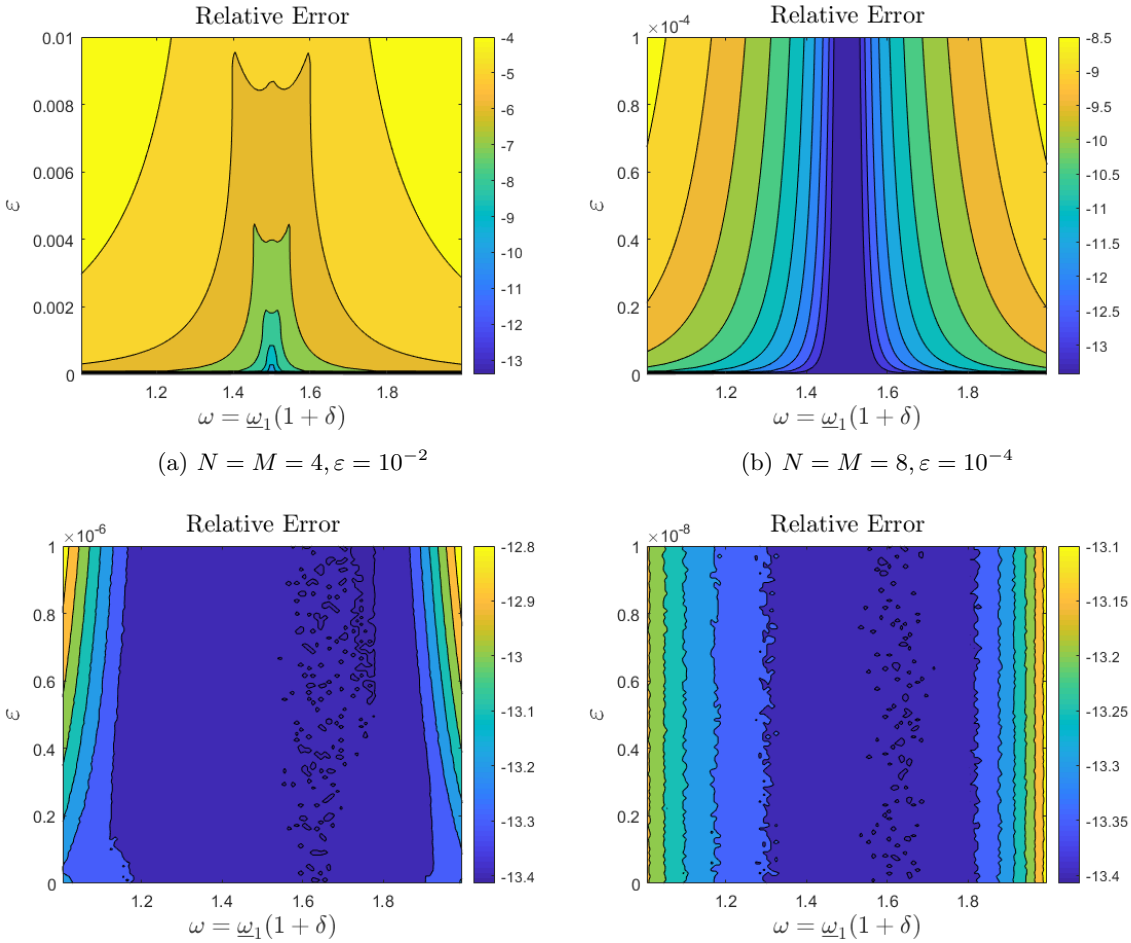
(a) $N = M = 4, \varepsilon = 10^{-2}$

(b) $N = M = 4, \varepsilon = 10^{-4}$

(c) $N = M = 4, \varepsilon = 10^{-6}$

(d) $N = M = 4, \varepsilon = 10^{-8}$

Figure 15: Plot of relative error in the lower layer with fixed $N = M = 4$ and four choices of $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ with Taylor summation. Physical parameters were (5.30) and numerical discretization was (5.31).



(a) $N = M = 4, \varepsilon = 10^{-2}$

(b) $N = M = 8, \varepsilon = 10^{-4}$

(c) $N = M = 12, \varepsilon = 10^{-6}$

(d) $N = M = 16, \varepsilon = 10^{-8}$

Figure 16: Plot of relative error in the lower layer with four choices of $N = M = 4, 8, 12, 16$ and four choices of $\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$ with Taylor summation. Physical parameters were (5.30) and numerical discretization was (5.31).

CHAPTER 6

SCATTERING AND REFLECTIVITY

6.1 Introduction

We can now define one of our primary objects of study, the Reflectivity Map. The Reflectivity Map (R) measures the response (reflected energy) of a periodically corrugated grating structure as a function of illumination frequency, ω, and corrugation amplitude, h. A HOPS method takes a perturbative view towards the geometric dependence of R on $h = \varepsilon$, $\varepsilon \ll 1$, by seeking the terms in the expansion about $\varepsilon = 0$,

$$R = R(\varepsilon) = \sum_{n=0}^{\infty} R_n \varepsilon^n.$$

With this, we realize an enormous savings in computational effort by conducting a new computation only for each choice of ω and then summing the formula above for any desired value of ε. Taking this philosophy to its natural conclusion, we consider $\omega = (1+\delta)\underline{\omega} = \underline{\omega} + \delta\underline{\omega}$ and perform a joint expansion of this map about $(\varepsilon = 0, \omega = \underline{\omega})$

$$R = R(\varepsilon, \delta) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} R_{n,m} \varepsilon^n \delta^m.$$

One would assume that a single computation, recovering all of the $R_{n,m}$, should be sufficient to compute the entire Reflectivity Map. However, the situation is not so simple as these expansions are not valid for all values of (ε, δ) and we found in §5.4 that the Rayleigh singularities (often called Wood's anomalies) enforced finite–size domains of convergence in δ. Nonetheless, we now undertake a more in–depth investigation and will focus on applying our HOPS/AWE algorithm based on the TFE methodology to the TE and TM polarizations. In §6.2 we state the mathematical meaning of the Reflectivity Map. Then in §6.3, §6.4, and §6.5 we perform an extensive series of numerical simulations to test the fidelity of the Reflectivity Map in both the TE and TM polarizations.

6.2 The Reflectivity Map

Recalling the solution (4.3) to the Helmholtz equation in the upper layer

$$u(x,z) = \sum_{p=-\infty}^{\infty} a_p e^{i\tilde{p}x + i\gamma_p^u z},$$

we note the very different character of the solution for wavenumbers p in the set

$$\mathcal{U}^u := \left\{ p \in \mathbb{Z} \mid \alpha_p^2 < (k^u)^2 \right\},$$

and those that are not. From our choice of the branch of the square root, components of $u(x,z)$ corresponding to $p \in \mathcal{U}^u$ propagate away from the layer interface, while those not in this set decay exponentially from $z = g(x)$. The latter are called evanescent waves while the former are propagating (defining the set of propagating modes $\mathcal{U}^u$) and carry energy away from the grating. With this in mind we define the efficiencies (6)

$$e_p^u := (\gamma_p^u/\gamma^u)\,|a_p|^2\,, \quad p \in \mathcal{U}^u,$$

and the Reflectivity Map

$$R := \sum_{p\in\mathcal{U}^u} e_p^u. \tag{6.1}$$

Similar quantities can be defined in the lower layer (6), and with these the principle of conservation of energy can be stated for structures composed entirely of dielectrics

$$\sum_{p\in\mathcal{U}^u} e_p^u + \tau^2 \sum_{p\in\mathcal{U}^w} e_p^w = 1.$$

In this situation a useful diagnostic of convergence for a numerical scheme (which we will utilize in our simulations) is the Energy Defect

$$D := 1 - \sum_{p\in\mathcal{U}^u} e_p^u - \tau^2 \sum_{p\in\mathcal{U}^w} e_p^w, \tag{6.2}$$

which should be zero for a purely dielectric structure.

6.3 Simulations of the Reflectivity Map: TM Mode

Using our novel HOPS/AWE approach in TM polarization (cf. §1.7) we computed

$$R_{\text{HOPS/AWE}}^{N,M,N_x,N_z,TM} \approx R,$$

for a range of ε and δ. As in our previous work (75), we show the kind of simulations this HOPS/AWE method can produce with modest computational effort. For this we selected $\underline{\omega}_q$, cf. (5.23), for $1 \le q \le 6$ and simulated R in the following frequency/wavelength ranges

$$\begin{aligned}
q = 1: &\quad \omega \in [1.005, 1.995] &\implies &\quad \lambda \in [3.14947, 6.25193],\\
q = 2: &\quad \omega \in [2.005, 2.995] &\implies &\quad \lambda \in [2.09789, 3.13376],\\
q = 3: &\quad \omega \in [3.005, 3.995] &\implies &\quad \lambda \in [1.57276, 2.09091],\\
q = 4: &\quad \omega \in [4.005, 4.995] &\implies &\quad \lambda \in [1.25789, 1.56884],\\
q = 5: &\quad \omega \in [5.005, 5.995] &\implies &\quad \lambda \in [1.04807, 1.25538],\\
q = 6: &\quad \omega \in [6.005, 6.995] &\implies &\quad \lambda \in [0.89824, 1.04633],
\end{aligned} \tag{6.3}$$

cf. (5.24). In addition, we selected

$$g(x) = \varepsilon f(x), \quad f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \tag{6.4}$$

with the parameters

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 1.1, \quad N_x = N_z = 32, \quad N = M = 16. \tag{6.5}$$

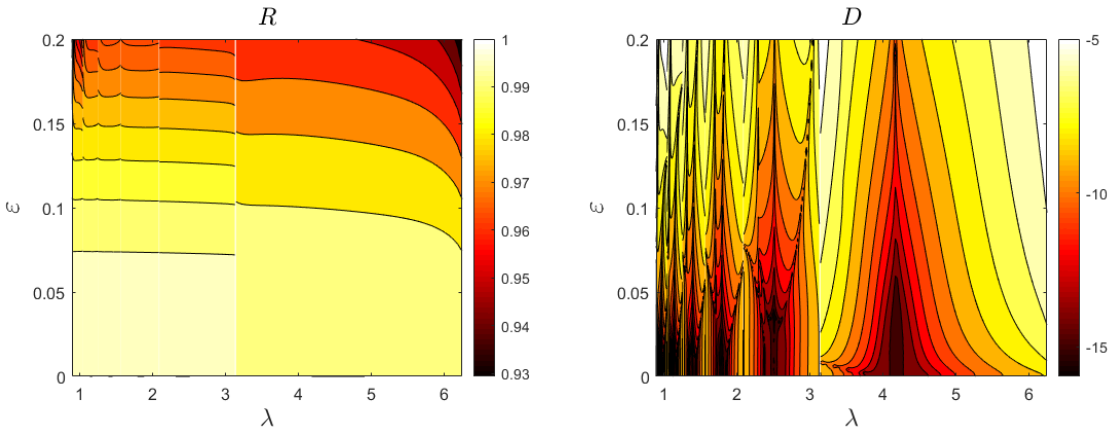
For all of our simulations in the TE and TM modes in §6.3, §6.4, and §6.5 we selected

$$c_0 = 1, \quad d = 2\pi,$$

where c_0 is the speed of light and d is the periodicity of the grating. For all of the simulations in §6.3 we enforced the artificial boundaries in the computational domain

$$a = 1, \quad b = -1.$$

In Figure 17(a) we plot all six of these subsets of the Reflectivity Map on one set of coordinate axes, and in Figure 17(b) we plot the Energy Defect, (6.2), to verify the accuracy of our expansions.



(a) Reflectivity Map (b) Energy Defect

Figure 17: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Taylor summation. We set $N = M = 16$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.4) and physical parameters were (6.5).

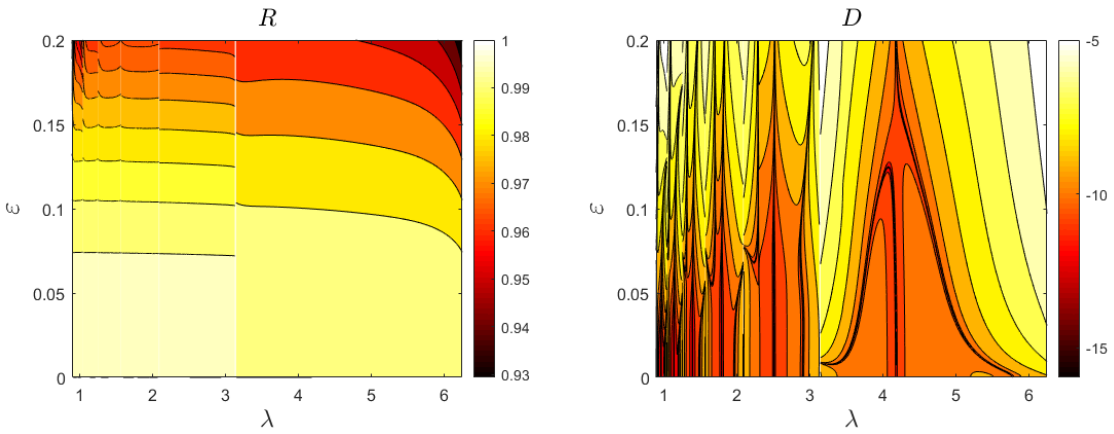
We then changed to non–normal incidence ($\alpha \neq 0$) and increased the granularity to $N_\varepsilon = N_\delta = 1000$ per invocation. In Chapter 7 we will discuss the advantageous computational complexity our HOPS/AWE algorithm enjoys in this situation of large N_ε and N_δ. We selected

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \tag{6.6}$$

with the parameters

$$\alpha = 10^{-4}, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 1.1, \quad N_x = N_z = 32, \quad N = M = 16. \tag{6.7}$$

In Figure 18(a) we plot six different subsets of the Reflectivity Map on a single coordinate axis, and in Figure 18(b) we plot the Energy Defect to demonstrate the accuracy of our scheme with a nonzero value of α.



(a) Reflectivity Map (b) Energy Defect

Figure 18: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Taylor summation. We set $N = M = 16$ with a granularity of $N_\varepsilon = N_\delta = 1000$ per invocation. The grating surface was (6.6) and physical parameters were (6.7).

Next, we considered normal incidence ($\alpha = 0$) and changed the lower index of refraction n^w to match representative values of silver (Ag) and gold (Au) as reported by Johnson & Christy (108), in particular

$$n_{\text{Ag}} = 0.05 + 2.275i, \quad n_{\text{Au}} = 1.48 + 1.883i.$$

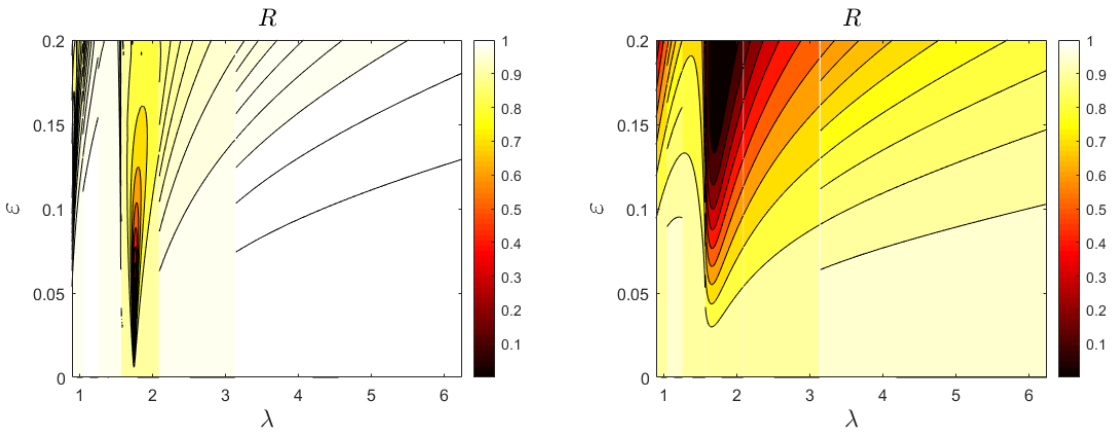
Using the same frequency and wavelength ranges, we studied

$$f(x) = \cos(4x), \quad \varepsilon_{\max} = 0.2, \tag{6.8}$$

with the parameters

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.9}$$

In Figure 19(a) we plot six different subsets of the Reflectivity Map where the lower index of refraction is selected to model the optical constant of silver. In Figure 19(b) we plot six different subsets of the Reflectivity Map where the lower index of refraction is changed to the optical constant for gold.



(a) Reflectivity Map for Silver

(b) Reflectivity Map for Gold

Figure 19: The Reflectivity Map, $R(\varepsilon, \delta)$, for silver (left) and gold (right) with Padé summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.8) and physical parameters were (6.9) with $n^w = n_{\text{Ag}}$ (left) and $n^w = n_{\text{Au}}$ (right).

We then changed the lower index of refraction n^w to match representative values of tungsten (W) and iron (Fe) as reported by Ordal et al. (109), where

$$n_{\text{W}} = 3.8313 + 2.9043i, \quad n_{\text{Fe}} = 4.274 + 9.579i.$$

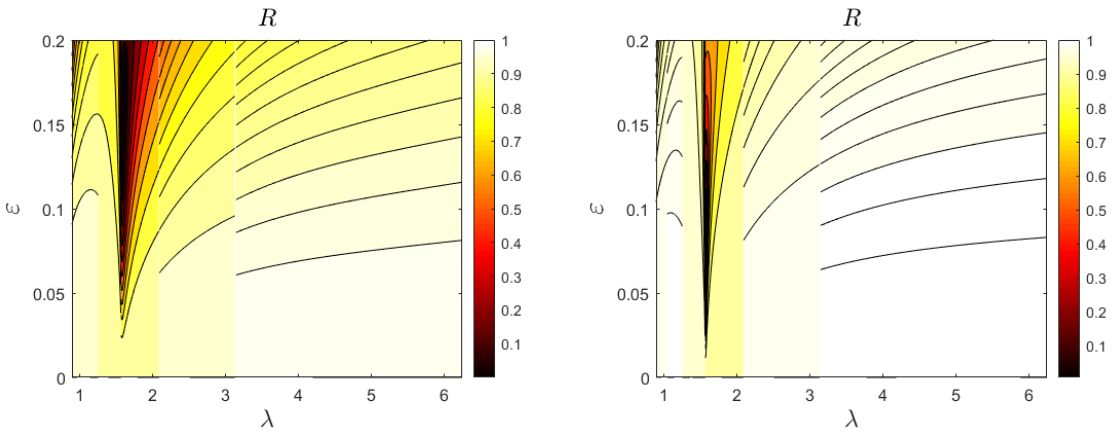
From these, we studied

$$f(x) = \sin(4x), \quad \varepsilon_{\max} = 0.2, \tag{6.10}$$

with the parameters

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.11}$$

In Figure 20(a) we plot six different subsets of the Reflectivity Map where the lower index of refraction is selected to model the optical constant of tungsten. In Figure 20(b) we plot six different subsets of the Reflectivity Map where the lower index of refraction is changed to the optical constant for iron.



(a) Reflectivity Map for Tungsten

(b) Reflectivity Map for Iron

Figure 20: The Reflectivity Map, $R(\varepsilon, \delta)$, for tungsten (left) and iron (right) with Padé summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.10) and physical parameters were (6.11) with $n^w = n_W$ (left) and $n^w = n_{\text{Fe}}$ (right).

We then changed back to non–normal incidence ($\alpha \neq 0$) and reduced our total computation time by simulating R in the following frequency/wavelength ranges

$$\begin{aligned} q = 1: &\quad \omega \in [1.005, 1.995] &\implies &\quad \lambda \in [3.14947, 6.25193], \\ q = 2: &\quad \omega \in [2.005, 2.995] &\implies &\quad \lambda \in [2.09789, 3.13376], \\ q = 3: &\quad \omega \in [3.005, 3.995] &\implies &\quad \lambda \in [1.57276, 2.09091]. \end{aligned}$$

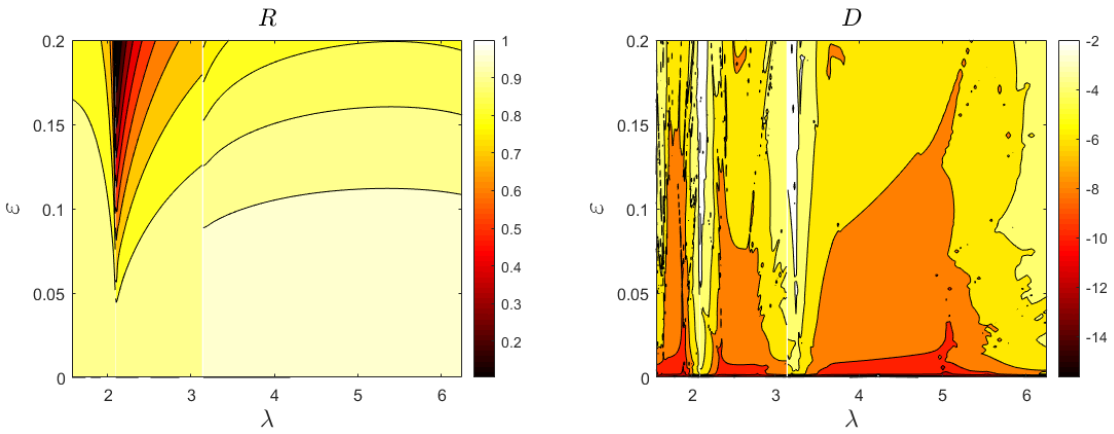
We selected

$$f(x) = \cos(3x), \quad \varepsilon_{\max} = 0.2, \tag{6.12}$$

with the parameters

$$\alpha = 0.01, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 3.1874, \quad N_x = N_z = 64, \quad N = M = 13, \tag{6.13}$$

and the value of n^w is meant to model Zinc germanium phosphide (110). In Figure 21(a) we plot three different subsets of the Reflectivity Map on one set of coordinate axes. In Figure 21(b) we plot the Energy Defect to show the accuracy of our scheme in the case $\alpha \neq 0$.



(a) Reflectivity Map (b) Energy Defect

Figure 21: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 13$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.12) and physical parameters were (6.13) with $n^w = 3.1874$ (Zinc germanium phosphide).

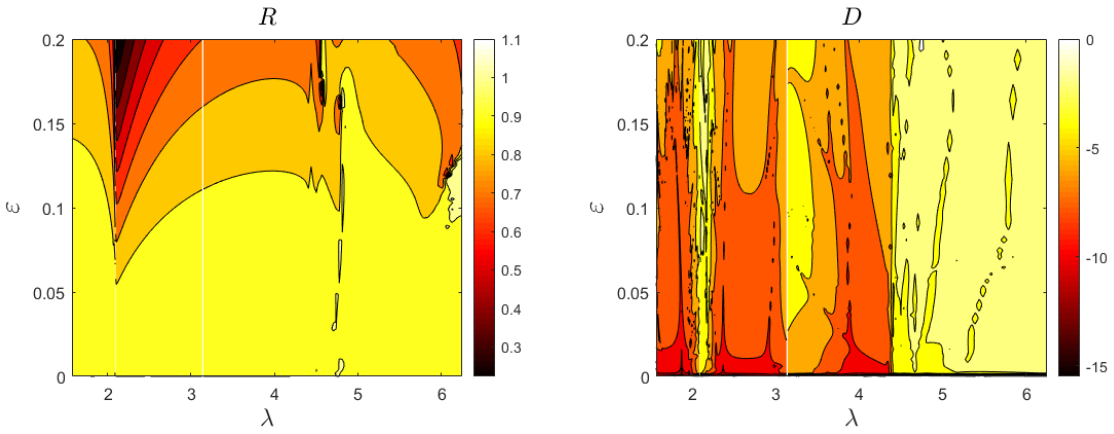
Next, we studied

$$f(x) = \sin(3x), \quad \varepsilon_{\max} = 0.2, \tag{6.14}$$

with the parameters

$$\alpha = 0.01, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 2.1054, \quad N_x = N_z = 64, \quad N = M = 13, \tag{6.15}$$

and the value of n^w is meant to model Zinc monoxide (111). In Figure 22(a) we plot three different subsets of the Reflectivity Map on one set of coordinate axes. In Figure 22(b) we plot the Energy Defect to show the accuracy of our scheme in the case $\alpha \neq 0$.



(a) Reflectivity Map (b) Energy Defect

Figure 22: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 13$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.14) and physical parameters were (6.15) with $n^w = 2.1054$ (Zinc monoxide).

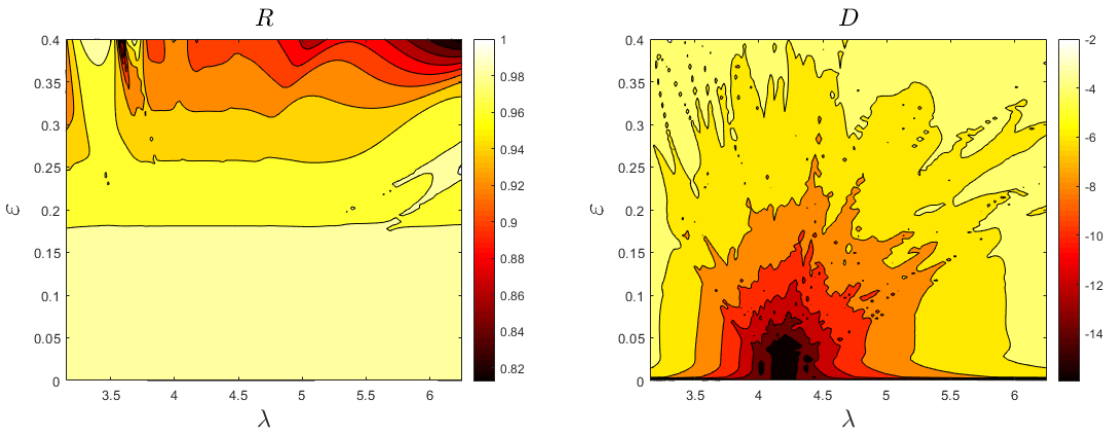
Seeking to understand what occurs for non–physical values of the dielectric constants, we simulated R with the first frequency/wavelength range in (6.3) and selected

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.4, \tag{6.16}$$

with the high refractive indices

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 5, \quad n^w = 8.1, \quad N_x = N_z = 32, \quad N = M = 12, \tag{6.17}$$

In Figure 23(a) we plot a single subset of the Reflectivity Map on a single coordinate axis, and in Figure 23(b) we plot the Energy Defect. Our choice of dielectric constants produces an interesting pattern in the computation of R.



(a) Reflectivity Map (b) Energy Defect

Figure 23: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 12$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.16) and physical parameters were (6.17).

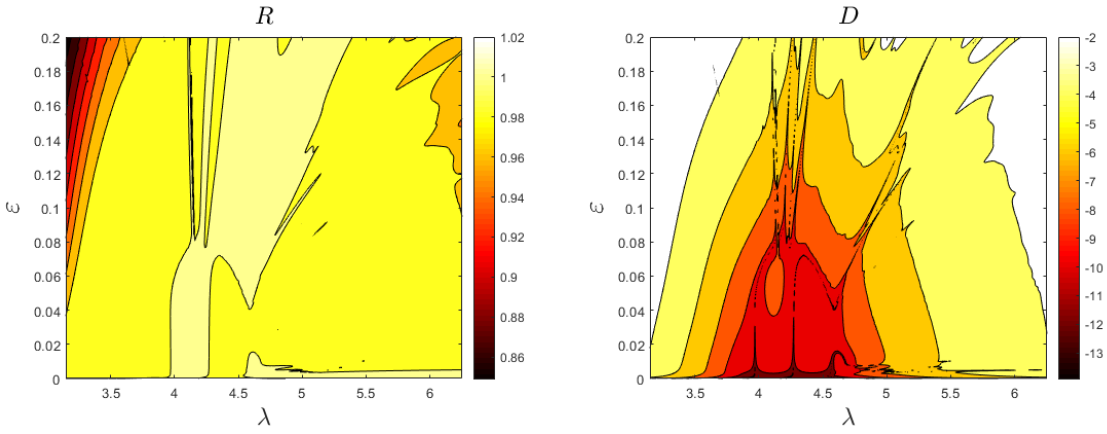
We then studied

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \tag{6.18}$$

with a purely imaginary index of refraction in the lower layer

$$\alpha = 0.1, \quad \sigma = 0.99, \quad n^u = 15, \quad n^w = 20i, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.19}$$

In Figure 24(a) we plot a single subset of the Reflectivity Map on a single coordinate axis, and in Figure 24(b) we plot the Energy Defect, where we once again observe an interesting pattern generated through our choice of dielectric constants. As the lower refractive index, n^w, is purely imaginary, we do not expect that (6.2) holds and $D \neq 0$. Nonetheless, we still found scattered energy in the far–field and small values of D when the value of n^w is purely imaginary.



(a) Reflectivity Map (b) Energy Defect

Figure 24: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 1000$ per invocation. The grating surface was (6.18) and physical parameters were (6.19).

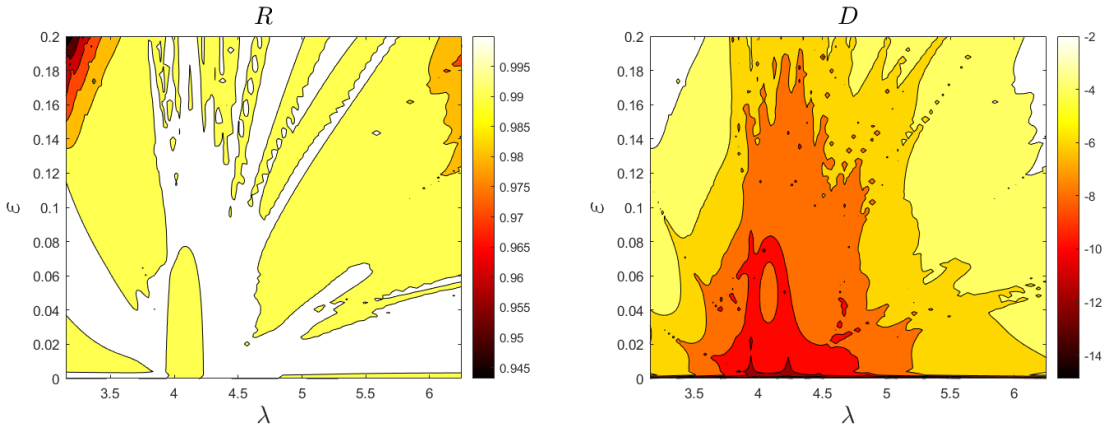
Finally, we selected

$$f(x) = \sin(x), \quad \varepsilon_{\max} = 0.2, \tag{6.20}$$

with the parameters

$$\alpha = 0.1, \quad \sigma = 0.99, \quad n^u = 10, \quad n^w = 40i, \quad N_x = N_z = 32, \quad N = M = 20. \tag{6.21}$$

In Figure 25(a) we plot a single subset of the Reflectivity Map and in Figure 25(b) we plot the Energy Defect.



(a) Reflectivity Map (b) Energy Defect

Figure 25: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 20$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.20) and physical parameters were (6.21).

6.4 Simulations of the Reflectivity Map: Smooth, Rough, and Lipschitz Profiles

We then simulated R with the first frequency/wavelength range in (6.3) with TM polarization and selected two–dimensional domains whose upper boundaries are shaped by the profiles

$$f_{s_1}(x) = \frac{\cos(4x)}{4}, \tag{6.22a}$$

$$f_{s_2}(x) = \frac{\exp\left(\cos(3x)\right)}{3} - c_0, \tag{6.22b}$$

$$f_r(x) = \left(2 \times 10^{-4}\right) x^4 \left(2\pi - x^4\right) - c_1, \tag{6.22c}$$

$$f_L(x) = \begin{cases} -2x/\pi + 1, & 0 \le x \le \pi, \\ 2x/\pi - 3, & \pi \le x \le 2\pi, \end{cases} \tag{6.22d}$$

where f_{s_1}, f_{s_2} represent a smooth (C^∞) boundary and f_r, f_L depict moderately smooth (C^4) and Lipschitz boundaries. Following (112), the constant c_0 in (6.22b) is chosen so that f_{s_2} has zero mean (as does f_r with the appropriate choice of c_1). The Fourier series representation of f_r and f_L are

$$f_r(x) = \sum_{k=1}^{\infty} \frac{96\left(2k^2\pi^2 - 21\right)}{125k^8} \cos(kx), \tag{6.23a}$$

$$f_L(x) = \sum_{k=1}^{\infty} \frac{8}{\pi^2(2k-1)^2} \cos\left((2k-1)x\right), \tag{6.23b}$$

and to minimize the effect of aliasing errors we approximated f_r and f_L by the truncated Fourier series

$$f_{r,P}(x) = \sum_{k=1}^{P} \frac{96\left(2k^2\pi^2 - 21\right)}{125k^8} \cos(kx), \tag{6.24a}$$

$$f_{L,P}(x) = \sum_{k=1}^{P/2} \frac{8}{\pi^2(2k-1)^2} \cos\left((2k-1)x\right). \tag{6.24b}$$

If $P \ll N_x/2$ then the effects of aliasing are minimal and we chose $P = 120$ for all of our simulations. For the smooth profiles, we selected

$$f(x) = f_{s_1}(x), \quad \varepsilon_{\max} = 4.0, \quad a = 10, \quad b = -10, \tag{6.25}$$

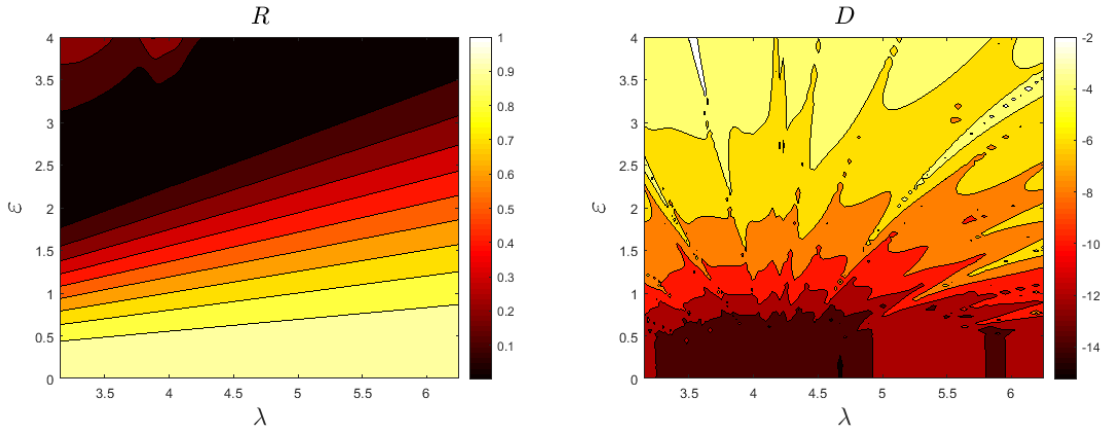
and

$$f(x) = f_{s_2}(x), \quad \varepsilon_{\max} = 2.0, \quad a = 4, \quad b = -4, \tag{6.26}$$

with the parameters

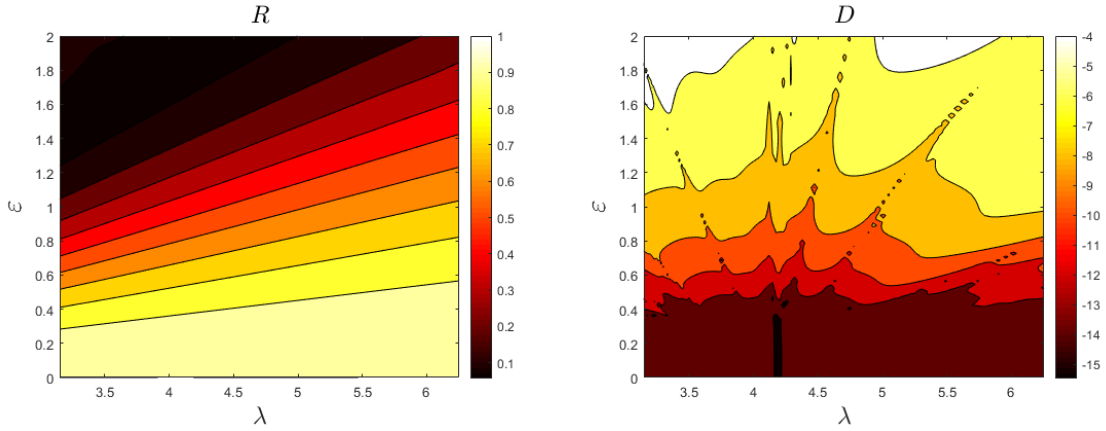
$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 1.1, \quad N_x = 256, \quad N_z = 128, \quad N = M = 20. \quad (6.27)$$

In Figures 26(a) and 27(a) we plot a single subset of the Reflectivity Map on a coordinate axis and in Figures 26(b) and 27(b) we plot the Energy Defect.



(a) Reflectivity Map (b) Energy Defect

Figure 26: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 20$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.25) and physical parameters were (6.27).



(a) Reflectivity Map (b) Energy Defect

Figure 27: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 20$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.26) and physical parameters were (6.27).

Next, for the rough profile, we selected

$$f(x) = f_{r,P}(x), \quad \varepsilon_{\max} = 2.0, \quad a = 4, \quad b = -4, \quad (6.28)$$

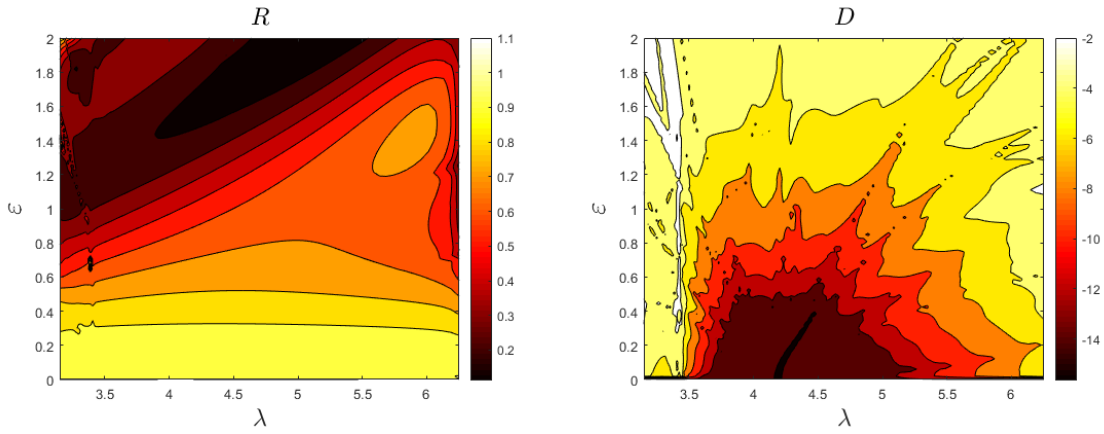
and for the Lipschitz profile, we selected

$$f(x) = f_{L,P}(x), \quad \varepsilon_{\max} = 2.0, \quad a = 4, \quad b = -4, \tag{6.29}$$

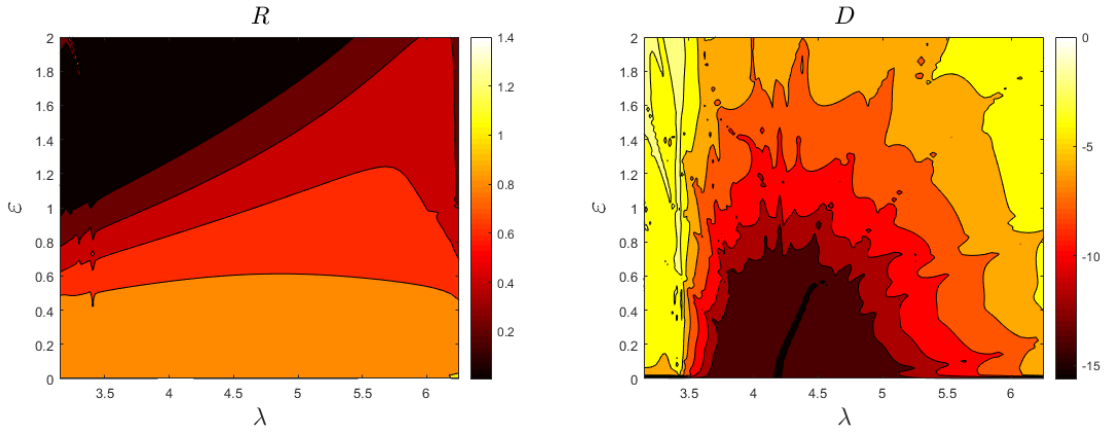
with the parameters

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 1.1, \quad N_x = 1024, \quad N_z = 128, \quad N = M = 20. \tag{6.30}$$

In Figures 28(a) and 28(b) we plot the Reflectivity Map and Energy Defect for the rough profile on a single coordinate axis and compare this to an equivalent simulation for the Lipschitz profile in Figures 28(c) and 28(d).



(a) Reflectivity Map (b) Energy Defect



(c) Reflectivity Map (d) Energy Defect

Figure 28: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 20$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. (Top) The rough profile with grating surface, (6.28), and physical parameters, (6.30). (Bottom) The Lipschitz profile with grating surface, (6.29), and physical parameters, (6.30).

6.5 Simulations of the Reflectivity Map: TE Mode

We then changed to TE polarization (cf. §1.6) and turned back to computing

$$R_{\text{HOPS/AWE}}^{N,M,N_x,N_z,TE} \approx R,$$

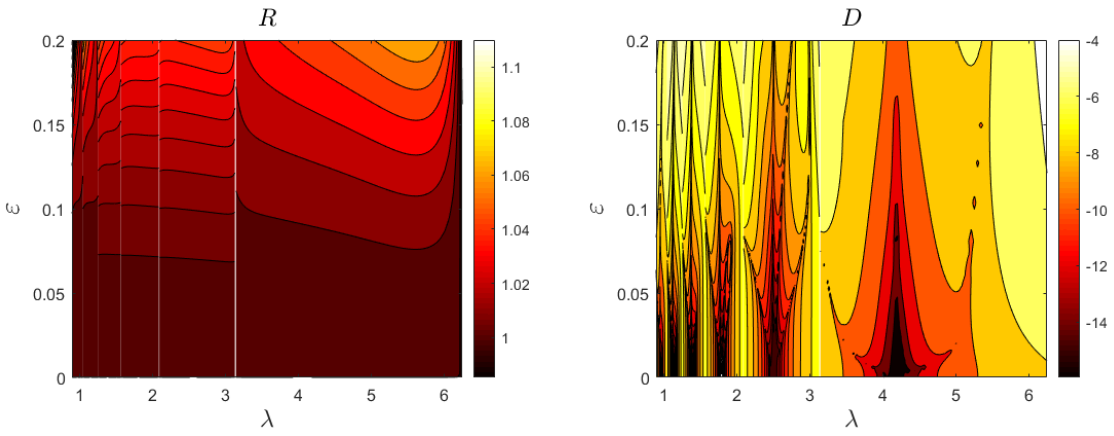
for a range of ε and δ. As in TM polarization, we simulated R with the frequency/wavelength ranges in (6.3). For our first simulation, we studied

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \quad a = 1, \quad b = -1, \tag{6.31}$$

with the parameters

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 1.1, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.32}$$

In Figure 29(a) we plot all six of these subsets of the Reflectivity Map on one set of coordinate axes, and in Figure 29(b) we plot the Energy Defect.



(a) Reflectivity Map (b) Energy Defect

Figure 29: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Taylor summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.31) and physical parameters were (6.32).

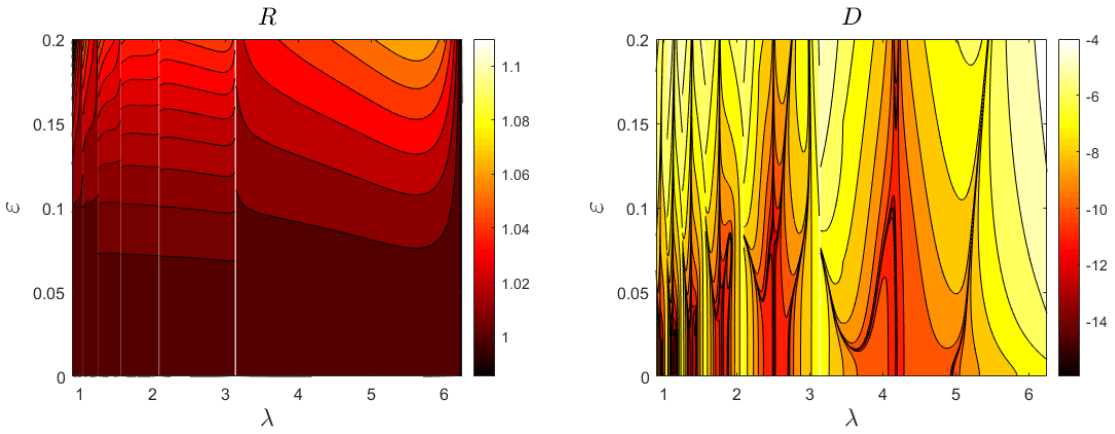
We then changed to non–normal incidence ($\alpha \neq 0$) and increased the granularity to $N_\varepsilon = N_\delta = 1000$ per invocation. We once again studied

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \quad a = 1, \quad b = -1, \tag{6.33}$$

with the parameters

$$\alpha = 10^{-4}, \quad \sigma = 0.99, \quad n^u = 1, \quad n^w = 1.1, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.34}$$

In Figure 30(a) we plot all six of these subsets of the Reflectivity Map on one set of coordinate axes, and in Figure 30(b) we plot the Energy Defect, to verify the accuracy of our expansions.



(a) Reflectivity Map (b) Energy Defect

Figure 30: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Taylor summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 1000$ per invocation. The grating surface was (6.33) and physical parameters were (6.34).

Next, we considered normal incidence ($\alpha = 0$) and changed the lower index of refraction n^w to match representative values of copper (Cu) and cobalt (Co) as reported by Johnson & Christy (108; 113), in particular

$$n_{\text{Cu}} = 0.94 + 1.337i, \quad n_{\text{Co}} = 2.1396 + 3.9840i.$$

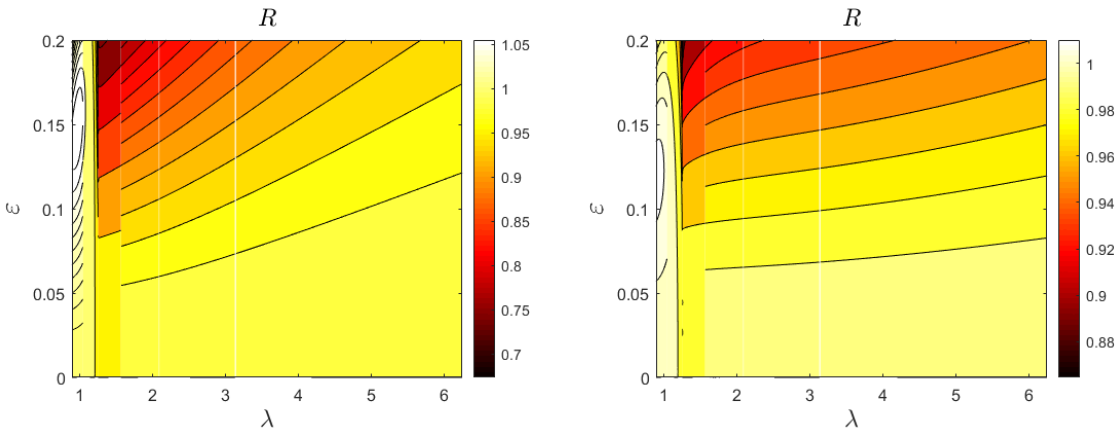
Using the same frequency and wavelength ranges, we studied

$$f(x) = \sin(5x), \quad \varepsilon_{\max} = 0.2, \quad a = 2/\pi, \quad b = -2/\pi, \tag{6.35}$$

with the parameters

$$\alpha = 0, \quad \sigma = 0.99, \quad n^u = 1, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.36}$$

In Figure 31(a) we plot six different subsets of the Reflectivity Map where the lower index of refraction is selected to model the optical constant of copper. In Figure 31(b) we plot six different subsets of the Reflectivity Map where the lower index of refraction is changed to the optical constant for cobalt.



(a) Reflectivity Map for Copper (b) Reflectivity Map for Cobalt

Figure 31: The Reflectivity Map, $R(\varepsilon, \delta)$, for copper (left) and cobalt (right) with Padé summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.35) and physical parameters were (6.36) with $n^w = n_{\text{Cu}}$ (left) and $n^w = n_{\text{Co}}$ (right).

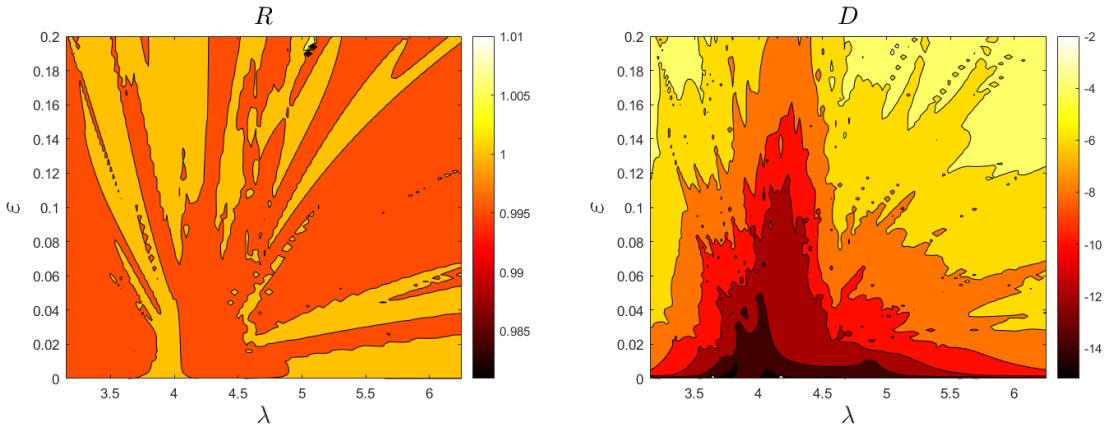
We then analyzed non–physical values of the dielectric constants. We simulated R with the first frequency/wavelength range in (6.3) and selected

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \quad a = \pi/2, \quad b = -\pi/2, \tag{6.37}$$

with a purely imaginary index of refraction in the lower layer

$$\alpha = 0.001, \quad \sigma = 0.99, \quad n^u = 5, \quad n^w = 20i, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.38}$$

In Figure 32 we plot the Reflectivity Map and Energy Defect on a single coordinate axis to demonstrate the accuracy of our scheme with a non–physical dielectric constant.



(a) Reflectivity Map (b) Energy Defect

Figure 32: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. The grating surface was (6.37) and physical parameters were (6.38).

Lastly, we selected

$$f(x) = \sin(x), \quad \varepsilon_{\max} = 0.2, \quad a = 2/\pi, \quad b = -2/\pi, \tag{6.39}$$

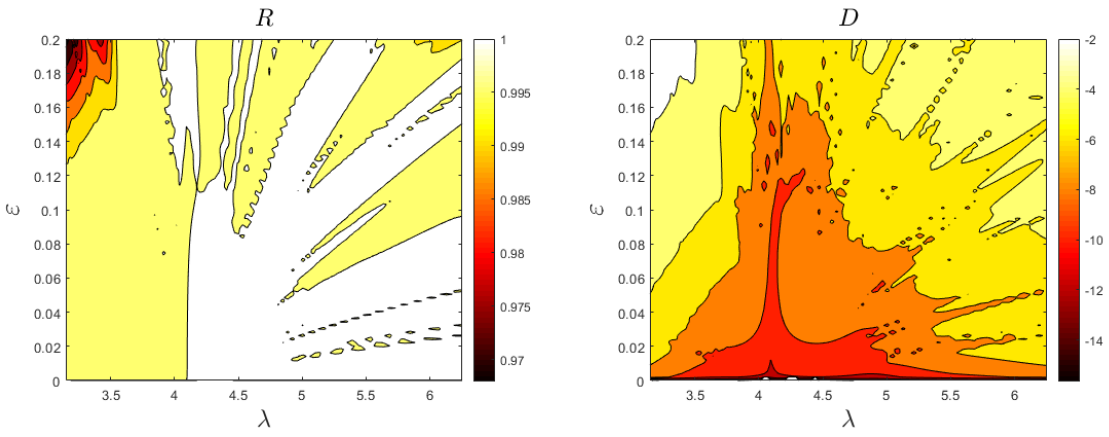
and

$$f(x) = \cos(x), \quad \varepsilon_{\max} = 0.2, \quad a = 2/\pi, \quad b = -2/\pi, \tag{6.40}$$

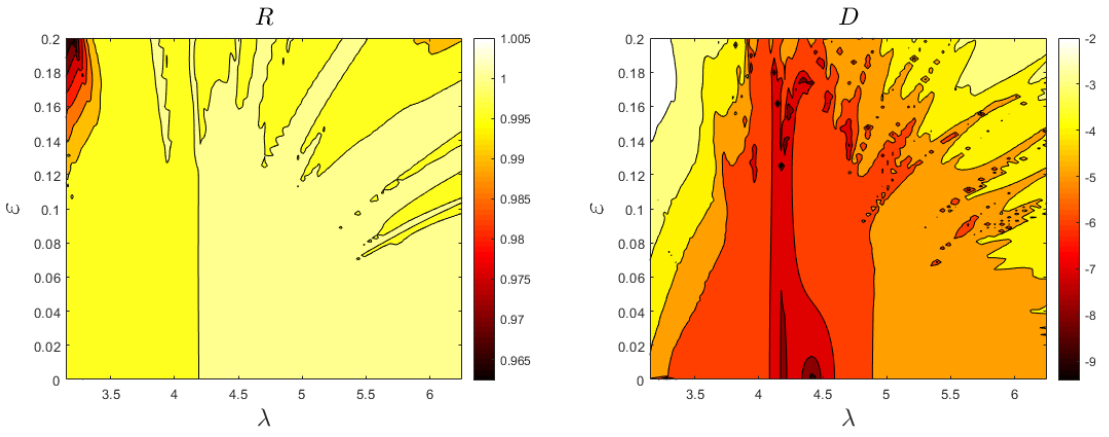
with the parameters

$$\alpha = 0.1, \quad \sigma = 0.99, \quad n^u = 10, \quad n^w = 25i, \quad N_x = N_z = 32, \quad N = M = 15. \tag{6.41}$$

In Figures 33(a) and 33(b) we plot the Reflectivity Map and Energy Defect for the sine profile on a single coordinate axis and compare this to an equivalent simulation for the cosine profile in Figures 33(c) and 33(d).



(a) Reflectivity Map (b) Energy Defect



(c) Reflectivity Map (d) Energy Defect

Figure 33: The Reflectivity Map, $R(\varepsilon, \delta)$, and Energy Defect D computed with Padé summation. We set $N = M = 15$ with a granularity of $N_\varepsilon = N_\delta = 100$ per invocation. (Top) The grating surface was (6.39) and physical parameters were (6.41). (Bottom) The grating surface was (6.40) and physical parameters were (6.41).

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This thesis establishes a novel HOPS/AWE algorithm that is particularly well suited to simulating scattering returns for periodic media problems. Our main contribution is that of Theorem 4.6.1 which guarantees the existence and uniqueness of solutions to a system of partial differential equations which model the interaction of linear waves in periodic layered structures with respect to multiple perturbation parameters. Through the introduction of DNOs and a change of variables based on the TFE methodology, we have shown that solutions to the Helmholtz problem are jointly analytic with respect to both interfacial and frequency perturbations. As a result, our HOPS/AWE algorithm is able to handle a variety of numerical simulations that are physically challenging in both the TE and TM polarization modes. Moreover, our extensive numerical results demonstrate the accuracy, speed, and robustness expected of all HOPS methods.

7.1 Future Directions

There are a wide range of improvements to both the HOPS/AWE algorithm and the proof of analyticity for linear waves in periodic layered media. Our main goals for future research are to expand the TFE method through a new proof of convergence, investigate expanding around singularities, evaluate analyticity theorems in multilayered configurations, add new parallel programming functionality, explore alternative methods to recover surface data without Dirichlet–Neumann Operators, and to reduce the execution time of the HOPS algorithm. We now summarize these six research goals and suggest predictions for future research.

Goal 1- Choice of Parameters: Does the geometry of the perturbation impact how large the size of the perturbation can be?

Goal 2- Rayleigh Singularities: Can we build a full HOPS algorithm based on points where the Taylor expansion is invalid?

Goal 3- Multiple Layers: Can we prove analyticity results when the number of layers is greater than three? Do the same theorems hold for ten or one hundred layers?

Goal 4- Parallel Programming: Can we implement parallel programming techniques so that our HOPS code runs on N processors?

Goal 5- Alternatives to DNOs: Do we need to use DNOs to recover surface data from information stored in the transformed field? Is there an alternative method which preserves the inversion of a single, sparse operator at the interface?

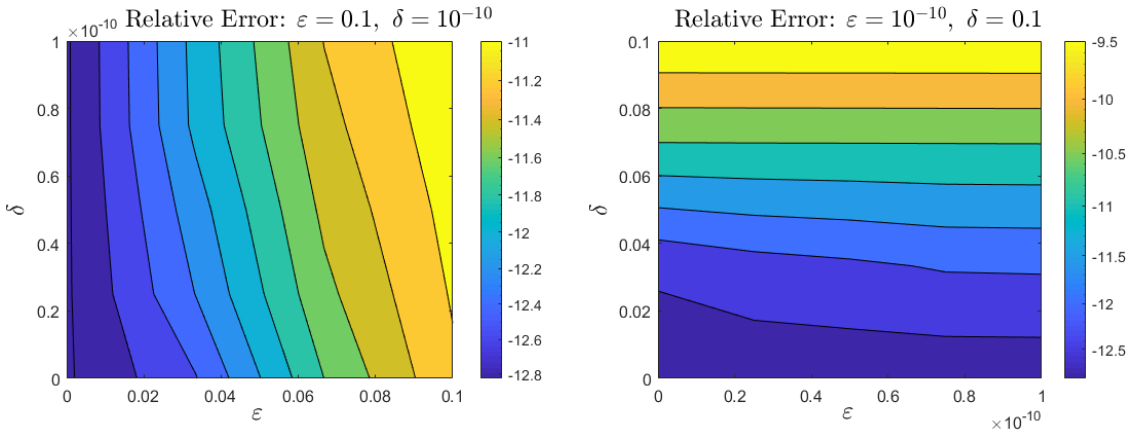
Goal 6- Computational Costs: Can we reduce the execution time per time step in our HOPS algorithm?

7.2 Choice of Parameters

Our HOPS/AWE algorithm is based on two smallness assumptions:

[1] Boundary Perturbation: $g(x) = \varepsilon f(x)$, $\varepsilon \in \mathbb{R}$, $\varepsilon \ll 1$,
[2] Frequency Perturbation: $\omega = (1+\delta)\underline{\omega} = \underline{\omega} + \delta\underline{\omega}$, $\omega \in \mathbb{R}$, $\delta \ll 1$,

with the additional assumption that f is sufficiently smooth ($f \in C^2$ (114; 68) or even Lipschitz (115)). Numerical simulations show that our HOPS/AWE algorithm can handle larger perturbations of ε (the height/slope) in comparison to δ (the frequency). With modest test parameters and a period of $d = 2\pi$, we are able to perturb the value of ε (to $\varepsilon = 0.1$ or even $\varepsilon = 0.2$) and still get reasonable convergence results. At a value around $\varepsilon = 10^{-4}$, our HOPS/AWE algorithm converges to machine precision provided that we sum to high enough Taylor orders.



(a) Large ε, Small δ (b) Small ε, Large δ

Figure 34: A contour plot of the relative error computed with our HOPS/AWE algorithm by holding $N = M = 8$ Taylor orders fixed. In Figure 34(a) we expand up to $\varepsilon = 0.1$ and $\delta = 10^{-10}$ simultaneously with $N = M = 8$ Taylor orders. In Figure 34(b) we expand up to $\varepsilon = 10^{-10}$ and $\delta = 0.1$ simultaneously with $N = M = 8$ Taylor orders.

Supplementary testing in both the upper and lower layers confirms that our HOPS/AWE algorithm is better suited towards larger ε.

Predictions: Our HOPS/AWE methodology takes advantage of exact enforcement of the OWC at an artificial boundary in order to truncate the computational domain to one of finite extent. After flattening the surface, the DNOs recover information through the solution stored at the interface. We suspect that this process mitigates large perturbations of the height/slope. By following techniques developed in (93; 116; 98; 117), we

intend to rigorously prove that the TFE method is analytic when ε is large. Additionally, we are interested in perturbing other physical parameters in the context of layered media problems. These are discussed in the engineering literature (118; 119).

7.3 Rayleigh Singularities

A fundamental equation in the HOPS/AWE algorithm is

$$\alpha_p^2 + (\gamma_p^q(\delta))^2 = (k^q)^2,$$

where k^q represents the wavenumber, $q \in \{u, w\}$, and $\alpha = k^q \sin(\theta), \gamma = k^q \cos(\theta)$, are parameters corresponding to refraction/reflection of the incidence angle θ. As shown in §5.4, a Rayleigh singularity (or Wood's anomaly) occurs when $\underline{\alpha}_p^2 = (\underline{k}^q)^2$ for any integer $p \neq 0$. That is, if $\underline{\gamma}_p^q(\delta) = 0$ for $p \neq 0$ then the Taylor series expansion of $\gamma_p^q(\delta)$ is invalid. In (75), the author investigated changing the Taylor expansion to a Puiseux expansion (120):

$$\gamma_p^q(\delta) = \sum_{m=0}^{\infty} \gamma_{p,m}^q \delta^{m+1/2} = \delta^{1/2} \sum_{m=0}^{\infty} \gamma_{p,m}^q \delta^m.$$

However, he found that this approach ran into external difficulties (§6 of (75)) simplifying explicit forms of the Dirichlet and Neumann trace operators.

Predictions: Rayleigh singularities are a central obstruction to the convergence of our HOPS/AWE algorithm. In all of our numerical tests, we select custom frequency ranges which maximize the radius of convergence of our algorithm by expanding away from the singularities (cf. §5.6). Alternative methods such as Padé summation also fail to be analytic in a neighborhood of a Rayleigh singularity. General perturbation theory provides a variety of known techniques (121; 122; 123; 124; 125) for expanding around divergent perturbation series. We suspect that adding these techniques to our HOPS/AWE algorithm will allow us perform a series expansion of $\underline{\gamma}_p^q(\delta)$ that does not diverge when $\underline{\gamma}_p^q(\delta) = 0$.

7.4 Multiple Layers

In (98), the author discusses how to apply our HOPS methodology in multilayered configurations. He considers a multilayered material with M (finite) interfaces at

$$z = a^{(m)} + g^{(m)}(x, y), \quad 1 \leq m \leq M,$$

which are $d_x \times d_y$ periodic

$$g^{(m)}(x + d_x, y + d_y) = g^{(m)}(x, y), \quad 1 \leq m \leq M,$$

separating $(M + 1)$–many layers.

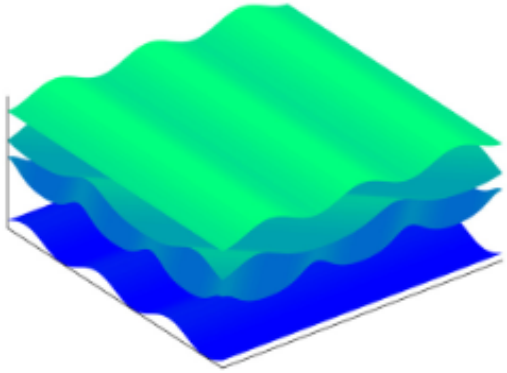


Figure 35: A five–layer problem configuration with layer interfaces $z = a^{(m)} + g^{(m)}(x)$.

A generalization of our analyticity theorems (cf. §4.5) up to M parameters is included as Theorem 3.2 in (98). For this, we consider quite general systems of linear equations of the form

$$\mathbf{A}(\tilde{\varepsilon})\mathbf{V}(\tilde{\varepsilon}) = \mathbf{R}(\tilde{\varepsilon}), \tag{7.1}$$

where

$$\mathbf{A}(\tilde{\varepsilon}) = \sum_{\tilde{n}=0}^{\infty} \mathbf{A}_{\tilde{n}} \tilde{\varepsilon}^{\tilde{n}}, \quad \mathbf{R}(\tilde{\varepsilon}) = \sum_{\tilde{n}=0}^{\infty} \mathbf{R}_{\tilde{n}} \tilde{\varepsilon}^{\tilde{n}}.$$

The tildes represent multi–index notation (126), in particular

$$\tilde{\varepsilon} := \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{pmatrix}, \quad \tilde{n} := \begin{pmatrix} n_1 \\ \vdots \\ n_M \end{pmatrix},$$

and the convention

$$\sum_{\tilde{n}=0}^{\infty} A_{\tilde{n}}\ \tilde{\varepsilon}^{\tilde{n}} = \sum_{n_1=0}^{\infty} \cdots \sum_{n_M=0}^{\infty} A_{n_1,\ldots,n_M} \varepsilon_1^{n_1} \cdots \varepsilon_M^{n_M}.$$

As in §4.5, we seek a solution of the form

$$\mathbf{V}(\tilde{\varepsilon}) = \sum_{\tilde{n}=0}^{\infty} \mathbf{V}_{\tilde{n}} \tilde{\varepsilon}^{\tilde{n}}, \tag{7.2}$$

and from (7.1) we find at order $\mathcal{O}(\tilde{\varepsilon}^{\tilde{n}})$

$$\mathbf{A}_0 \mathbf{V}_{\tilde{n}} = \mathbf{R}_{\tilde{n}} - \left(\sum_{\tilde{\ell}=0}^{\tilde{n}} \mathbf{A}_{\tilde{n}-\tilde{\ell}} \mathbf{V}_{\tilde{\ell}} - \mathbf{A}_0 \mathbf{V}_{\tilde{n}} \right),$$

or

$$\mathbf{V}_{\tilde{n}} = \mathbf{A}_0^{-1} \left\{ \mathbf{R}_{\tilde{n}} - \left(\sum_{\tilde{\ell}=0}^{\tilde{n}} \mathbf{A}_{\tilde{n}-\tilde{\ell}} \mathbf{V}_{\tilde{\ell}} - \mathbf{A}_0 \mathbf{V}_{\tilde{n}} \right) \right\}. \tag{7.3}$$

The above notation represents multi–indices in the form

$$\sum_{\tilde{\ell}=0}^{\tilde{n}} \mathbf{A}_{\tilde{n}-\tilde{\ell}} \mathbf{V}_{\tilde{\ell}} = \sum_{\ell_1=0}^{n_1} \cdots \sum_{\ell_M=0}^{n_M} \mathbf{A}_{n_1-\ell_1,\ldots,n_M-\ell_M} \mathbf{V}_{\ell_1,\ldots,\ell_m},$$

where $\tilde{n} = (n_1, \ldots, n_M)$, $\tilde{\ell} = (\ell_1, \ldots, \ell_M)$, and $0 = (0, \ldots, 0)$ with the convention

$$\tilde{n} \geq 0 \iff n_1 \geq 0, \ldots, n_M \geq 0, \quad \tilde{\ell} \geq 0 \iff \ell_1 \geq 0, \ldots, \ell_M \geq 0.$$

With these, we can extend our existence theorem (Theorem 4.5.1) to M parameters.

Theorem 7.4.1. *Given two Banach spaces, X and Y, suppose that:*

[1] $\mathbf{R}_{\tilde{n}} \in Y$ *for all* $\tilde{n} \geq 0$*, and there exist M–multi–indexed constants $C_R > 0$, $B_R > 0$,*

$$C_R = \begin{pmatrix} C_{R,1} \\ \vdots \\ C_{R,M} \end{pmatrix}, \quad B_R^{\tilde{n}} = \begin{pmatrix} B_{R,1}^{n_1} \\ \vdots \\ B_{R,M}^{n_M} \end{pmatrix},$$

such that

$$\|\mathbf{R}_{\tilde{n}}\|_Y \leq C_R B_R^{\tilde{n}},$$

[2] $\mathbf{A}_{\tilde{n}} : X \to Y$ *for all* $\tilde{n} \geq 0$*, and there exist M–multi–indexed constants $C_A > 0$, $B_A > 0$ such that*

$$\|\mathbf{A}_{\tilde{n}}\|_{X \to Y} \leq C_A B_A^{\tilde{n}},$$

[3] $\mathbf{A}_0^{-1} : Y \to X$*, and there exists a constant $C_e > 0$ such that*

$$\left\|\mathbf{A}_0^{-1}\right\|_{Y \to X} \leq C_e.$$

Then the equation (7.1) *has a unique solution,*

$$\mathbf{V}(\tilde{\varepsilon}) = \sum_{\tilde{n}=0}^{\infty} \mathbf{V}_{\tilde{n}} \tilde{\varepsilon}^{\tilde{n}}, \tag{7.4}$$

and there exist M–multi–indexed constants $C_V > 0$ and $B_V > 0$ such that

$$\|\mathbf{V}_{\tilde{n}}\|_X \leq C_V B_V^{\tilde{n}},$$

for all $\tilde{n} \geq 0$ *and any*

$$C_V \geq 2C_eC_R, \quad B_V \geq \max\left\{B_R, 2B_A, 2^{M+1}C_eC_AB_A\right\},$$

enforced componentwise. This implies that, for any M*–multi–indexed constant* $0 \leq \tilde{\rho} < 1$*,* (7.4)*, converges for all* $\tilde{\varepsilon}$ *such that* $B\tilde{\varepsilon} < \tilde{\rho}$*, i.e.,* $\tilde{\varepsilon} < \tilde{\rho}/B$*.*

Remark 7.4.1. Our proof strategy is a form of multidimensional induction where given a statement $\boldsymbol{P}(n_1, n_2, n_3, ..., n_M)$ for some $M \in \mathbb{N}$, we will show that $\forall n_1, n_2, \ldots, n_M \geq 0$, $\boldsymbol{P}(n_1, n_2, n_3, ..., n_M)$ is true by inducting on n_M. We will follow the steps outlined below.

[1] Establish $\boldsymbol{P}(0, \ldots, n_j, \ldots, 0)$ for all $1 \leq j < M$ and $n_1, \ldots, n_j \geq 0$.

[2] Given $\boldsymbol{P}(n_1, n_2, \ldots, n_j, \ldots, 0)$ for all $1 \leq j < M$ and $n_1, \ldots, n_j \geq 0$, establish $\boldsymbol{P}(n_1, n_2, \ldots, \bar{n}_j, \ldots, 0)$. This can be accomplished through the two steps below.

  (a) Establish $\boldsymbol{P}(0, \ldots, \bar{n}_j, \ldots, 0)$ for all $\bar{n}_j \geq 0$ (where the hypothesis in [2] gives the required case for $n_j < \bar{n}_j$).

  (b) Given $\boldsymbol{P}(n_1, n_2, \ldots, \bar{n}_j, \ldots, 0)$ for all $1 \leq j < M$ and $n_1 < \bar{n}_1, n_2 < \bar{n}_2, \ldots,$ $n_{j-1} < \bar{n}_{j-1}$ and $\bar{n}_j \geq 0$, establish $\boldsymbol{P}(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_j, \ldots, 0)$.

[3] Given $\boldsymbol{P}(n_1, n_2, \ldots, n_j, n_{j+1}, \ldots, 0)$ for all $1 \leq j + 1 < M$ and $n_1, \ldots, n_{j+1} \geq 0$, establish $\boldsymbol{P}(n_1, n_2, \ldots, n_j, \bar{n}_{j+1}, \ldots, 0)$. This can be accomplished by following the two steps outlined below.

  (a) Establish $\boldsymbol{P}(0, \ldots, \bar{n}_{j+1}, \ldots, 0)$ for all $\bar{n}_{j+1} \geq 0$ (where the hypothesis in [3] gives the required case for $n_{j+1} < \bar{n}_{j+1}$).

  (b) Given $\boldsymbol{P}(n_1, n_2, \ldots, n_j, \bar{n}_{j+1}, \ldots, 0)$ for all $1 \leq j + 1 < M$ and $n_1 < \bar{n}_1, n_2 < \bar{n}_2, \ldots, n_j < \bar{n}_j$ and $\bar{n}_{j+1} \geq 0$, establish $\boldsymbol{P}(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_j, \bar{n}_{j+1} \ldots, 0)$.

[4] Given $\boldsymbol{P}(n_1, n_2, \ldots, n_{M-1}, n_M)$ for all $n_1, n_2, \ldots, n_{M-1} \geq 0$ and $n_M < \bar{n}_M$, establish $\boldsymbol{P}(n_1, n_2, \ldots, n_{M-1}, \bar{n}_M)$. This can be accomplished by the two steps below (the base cases are handled through [2] and [3]).

  (a) Establish $\boldsymbol{P}(0, \ldots, \bar{n}_M)$ for $\bar{n}_M \geq 0$ (where the hypothesis in [4] handles the required case for $n_M < \bar{n}_M$).

  (b) Given $\boldsymbol{P}(n_1, n_2, \ldots, n_{M-1}, \bar{n}_M)$ for all $n_1 < \bar{n}_1, n_2 < \bar{n}_2, \ldots n_{M-1} < \bar{n}_{M-1}$ and $\bar{n}_M \geq 0$, establish $\boldsymbol{P}(\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_{M-1}, \bar{n}_M)$.

Proof. [Theorem 7.4.1] As with $\tilde{\varepsilon}$ and $\tilde{n}$, we represent $\tilde{\rho}$ by

$$\tilde{\rho} := \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_M \end{pmatrix}.$$

As before, we will work by induction and consider the general case for finite $M > 0$ where we want to establish

$$\|\mathbf{V}_{n_1,\ldots,n_M}\|_X \leq C_{V,1}\ldots C_{V,M}B_{V,1}^{n_1}\ldots B_{V,M}^{n_M}, \quad \forall n_1,\ldots,n_M \geq 0.$$

We prove this via an induction on n_M. The base case $n_1, n_2, \ldots, n_{j-1}, n_{j+1}, \ldots, n_M = 0$ and $1 \leq j < M$:

$$\left\|\mathbf{V}_{0,\ldots,n_j,\ldots,0}\right\|_X \leq C_{V,j}B_{V,j}^{n_j}, \quad \forall n_j \geq 0,$$

has previously been established by Theorem 4.5.1 where $\tilde{\varepsilon} = \varepsilon_j$ and $\delta = 0$. We now assume

$$\left\|\mathbf{V}_{n_1,\ldots,n_j,\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j}B_{V,1}^{n_1}\ldots B_{V,j}^{n_j}, \ \forall n_1,\ldots,n_{j-1} \geq 0, \ \forall n_j < \bar{n}_j, \ 1 \leq j < M,$$

and seek

$$\left\|\mathbf{V}_{n_1,\ldots,\bar{n}_j,\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j}B_{V,1}^{n_1}\ldots B_{V,j}^{\bar{n}_j}, \ \forall n_1,\ldots,n_{j-1} \geq 0.$$

This can be obtained through a chain of $(M-1)$ inductions on $n_1,\ldots,n_j$ where $1 \leq j < M$. For simplicity, we will show what happens in the arbitrary case n_j. The base case $n_1,\ldots,n_{j-1} = 0$:

$$\left\|\mathbf{V}_{0,\ldots,\bar{n}_j,\ldots,0}\right\|_X \leq C_{V,j}B_{V,j}^{\bar{n}_j}, \quad \forall \bar{n}_j \geq 0,$$

is established by Theorem 4.5.1 where $\tilde{\varepsilon} = \varepsilon_j$ and $\delta = 0$. Therefore, we assume

$$\left\|\mathbf{V}_{n_1,\ldots,\bar{n}_j,\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j}B_{V,1}^{n_1}\ldots B_{V,j}^{\bar{n}_j}, \ \forall n_1 < \bar{n}_1,\ldots,n_{j-1} < \bar{n}_{j-1}, \ \forall \bar{n}_j \geq 0, \ 1 \leq j < M,$$

and seek

$$\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_j,\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j}B_{V,1}^{\bar{n}_1}\ldots B_{V,j}^{\bar{n}_j}.$$

Recalling $\tilde{n} = (n_1,\ldots,n_j)$ and $\tilde{\ell} = (\ell_1,\ldots,\ell_j)$, we define

$$\sum_{\tilde{\ell}=0}^{\tilde{n}\ *}\mathbf{A}_{\tilde{n}-\tilde{\ell}}\mathbf{V}_{\tilde{\ell}} := \sum_{\tilde{\ell}=0}^{\tilde{n}}\mathbf{A}_{\tilde{n}-\tilde{\ell}}\mathbf{V}_{\tilde{\ell}} - \mathbf{A}_0\mathbf{V}_{\tilde{n}}, \tag{7.5}$$

and apply (7.3), (7.5) and the mapping properties of $\mathbf{A}_0^{-1}$ to find

$$\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_j,\ldots,0}\right\|_X \leq C_e\left\{\left\|\mathbf{R}_{\bar{n}_1,\ldots,\bar{n}_j}\right\|_Y + \sum_{\tilde{\ell}=0}^{\tilde{n}\ *}\left\|\mathbf{A}_{\tilde{n}-\tilde{\ell}}\mathbf{V}_{\tilde{\ell}}\right\|_Y\right\}.$$

Using the estimates on $\mathbf{R}_{n_1,\ldots,n_j}$ and $\mathbf{A}_{n_1,\ldots,n_j}$ (for all $n_1,\ldots,n_j$) and $\mathbf{V}_{n_1,\ldots,n_j}$ ($n_1 < \bar{n}_1,\ldots,n_j < \bar{n}_j$) we have

$$\begin{aligned}
\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_j,\ldots,0}\right\|_X &\leq C_e\Bigg\{C_{R,1}\ldots C_{R,j}B_{R,1}^{\bar{n}_1}\ldots B_{R,j}^{\bar{n}_j} + \sum_{\tilde{\ell}=0}^{\tilde{n}}{}^{*} C_{A,1}\ldots C_{A,j}B_{A,1}^{\bar{n}_1-\ell_1}\ldots B_{A,j}^{\bar{n}_j-\ell_j} \\
&\qquad\qquad \times C_{V,1}\ldots C_{V,j}B_{V,1}^{\ell_1}\ldots B_{V,j}^{\ell_j}\Bigg\} \\
&= C_eC_{R,1}\ldots C_{R,j}B_{R,1}^{\bar{n}_1}\ldots B_{R,j}^{\bar{n}_j} + C_eC_{A,1}\ldots C_{A,j}C_{V,1}\ldots C_{V,j} \\
&\quad\times\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1}\cdots\left(\frac{B_{A,j}}{B_{V,j}}\right)B_{V,j}^{\bar{n}_j}\sum_{\tilde{\ell}=0}^{\tilde{n}}{}^{*}\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1-\ell_1-1}\cdots \\
&\quad\times\left(\frac{B_{A,j}}{B_{V,j}}\right)B_{V,j}^{\bar{n}_j-\ell_j-1} \\
&\leq C_eC_{R,1}\ldots C_{R,j}B_{V,1}^{\bar{n}_1}\ldots B_{V,j}^{\bar{n}_j} + C_eC_{A,1}\ldots C_{A,j}C_{V,1}\ldots C_{V,j} \\
&\quad\times\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1}\cdots\left(\frac{B_{A,j}}{B_{V,j}}\right)B_{V,j}^{\bar{n}_j}\left(\frac{1}{1-1/2}\right)^j,
\end{aligned}$$

if $B_{A,k}/B_{V,k} \leq 1/2$, $k = 1,\ldots,j$ (implying $B_{V,k} \geq 2B_{A,k}$). We are done if we demand that

$$B_{V,k} \geq B_{R,k}, \quad C_eC_{R,k} \leq C_{V,k}/2, \quad 2^jC_eC_{A,k}C_{V,k}(B_{A,k}/B_{V,k}) \leq C_{V,k}/2.$$

This can be realized if

$$C_{V,k} \geq 2C_eC_{R,k}, \quad B_{V,k} \geq \max\left\{B_{R,k}, 2B_{A,k}, 2^{j+1}C_eC_{A,k}B_{A,k}\right\}.$$

We then assume

$$\left\|\mathbf{V}_{n_1,\ldots,n_{j+1},\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j+1}B_{V,1}^{n_1}\ldots B_{V,j+1}^{n_{j+1}}, \quad \forall n_1,\ldots,n_j \geq 0, \quad \forall n_{j+1} < \bar{n}_{j+1}, \quad 1 \leq j < M,$$

and seek

$$\left\|\mathbf{V}_{n_1,\ldots,\bar{n}_{j+1},\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j+1}B_{V,1}^{n_1}\ldots B_{V,j+1}^{\bar{n}_{j+1}}, \quad \forall n_1,\ldots,n_j \geq 0.$$

As before, this can be obtained through a chain of M inductions on $n_1,\ldots,n_{j+1}$ where $1 \leq j < M$. For simplicity, we will show what happens in the arbitrary case n_{j+1}. The base case $n_1,\ldots,n_j = 0$:

$$\left\|\mathbf{V}_{0,\ldots,\bar{n}_{j+1},\ldots,0}\right\|_X \leq C_{V,j+1}B_{V,j+1}^{\bar{n}_{j+1}}, \quad \forall \bar{n}_{j+1} \geq 0,$$

is established by Theorem 4.5.1 where $\tilde{\varepsilon} = \varepsilon_{j+1}$ and $\delta = 0$. Therefore, we assume

$$\left\|\mathbf{V}_{n_1,\ldots,\bar{n}_{j+1},\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j+1}B_{V,1}^{n_1}\ldots B_{V,j+1}^{\bar{n}_{j+1}},\quad \forall n_1<\bar{n}_1,\ldots,n_j<\bar{n}_j,\quad \forall \bar{n}_{j+1}\geq 0,\\ 1\leq j<M,$$

and seek

$$\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_{j+1},\ldots,0}\right\|_X \leq C_{V,1}\ldots C_{V,j+1}B_{V,1}^{\bar{n}_1}\ldots B_{V,j+1}^{\bar{n}_{j+1}}.$$

In this scenario, $\tilde{n} = (n_1,\ldots,n_{j+1})$ and $\tilde{\ell} = (\ell_1,\ldots,\ell_{j+1})$, so we apply (7.3), (7.5) and the mapping properties of $\mathbf{A}_0^{-1}$ to find

$$\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_{j+1},\ldots,0}\right\|_X \leq C_e\left\{\left\|\mathbf{R}_{\bar{n}_1,\ldots,\bar{n}_{j+1}}\right\|_Y + {\sum_{\tilde{\ell}=0}^{\tilde{n}}}^{*}\left\|\mathbf{A}_{\tilde{n}-\tilde{\ell}}\mathbf{V}_{\tilde{\ell}}\right\|_Y\right\}.$$

Using the estimates on $\mathbf{R}_{n_1,\ldots,n_{j+1}}$ and $\mathbf{A}_{n_1,\ldots,n_{j+1}}$ (for all $n_1,\ldots,n_{j+1}$) and $\mathbf{V}_{n_1,\ldots,n_{j+1}}$ ($n_1<\bar{n}_1,\ldots,n_{j+1}<\bar{n}_{j+1}$) we have

$$\begin{aligned}\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_{j+1},\ldots,0}\right\|_X &\leq C_e\Bigg\{C_{R,1}\ldots C_{R,j+1}B_{R,1}^{\bar{n}_1}\ldots B_{R,j+1}^{\bar{n}_{j+1}} + {\sum_{\tilde{\ell}=0}^{\tilde{n}}}^{*} C_{A,1}\ldots C_{A,j+1}B_{A,1}^{\bar{n}_1-\ell_1}\ldots\\ &\qquad\times B_{A,j+1}^{\bar{n}_{j+1}-\ell_{j+1}}C_{V,1}\ldots C_{V,j+1}B_{V,1}^{\ell_1}\ldots B_{V,j+1}^{\ell_{j+1}}\Bigg\}\\ &= C_eC_{R,1}\ldots C_{R,j+1}B_{R,1}^{\bar{n}_1}\ldots B_{R,j+1}^{\bar{n}_{j+1}} + C_eC_{A,1}\ldots C_{A,j+1}C_{V,1}\ldots C_{V,j+1}\\ &\quad\times\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1}\cdots\left(\frac{B_{A,j+1}}{B_{V,j+1}}\right)B_{V,j+1}^{\bar{n}_{j+1}}{\sum_{\tilde{\ell}=0}^{\tilde{n}}}^{*}\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1-\ell_1-1}\cdots\\ &\quad\times\left(\frac{B_{A,j+1}}{B_{V,j+1}}\right)B_{V,j+1}^{\bar{n}_{j+1}-\ell_{j+1}-1}\\ &\leq C_eC_{R,1}\ldots C_{R,j+1}B_{V,1}^{\bar{n}_1}\ldots B_{V,j+1}^{\bar{n}_{j+1}} + C_eC_{A,1}\ldots C_{A,j+1}C_{V,1}\ldots C_{V,j+1}\\ &\quad\times\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1}\cdots\left(\frac{B_{A,j+1}}{B_{V,j+1}}\right)B_{V,j+1}^{\bar{n}_{j+1}}\left(\frac{1}{1-1/2}\right)^{j+1},\end{aligned}$$

if $B_{A,t}/B_{V,t}\leq 1/2$, $t=1,\ldots,j+1$ (implying $B_{V,t}\geq 2B_{A,t}$). We are done if we demand that

$$B_{V,t}\geq B_{R,t},\quad C_eC_{R,t}\leq C_{V,t}/2,\quad 2^{j+1}C_eC_{A,t}C_{V,t}(B_{A,t}/B_{V,t})\leq C_{V,t}/2.$$

This can be realized if

$$C_{V,t}\geq 2C_eC_{R,t},\quad B_{V,t}\geq\max\left\{B_{R,t},2B_{A,t},2^{j+2}C_eC_{A,t}B_{A,t}\right\}.$$

To complete the general case for finite $M > 0$, we assume

$$\|\mathbf{V}_{n_1,\ldots,n_M}\|_X \leq C_{V,1}\ldots C_{V,M}B_{V,1}^{n_1}\ldots B_{V,M}^{n_M}, \quad \forall n_1,\ldots,n_{M-1} \geq 0, \quad \forall n_M < \bar{n}_M,$$

and seek

$$\|\mathbf{V}_{n_1,\ldots,\bar{n}_M}\|_X \leq C_{V,1}\ldots C_{V,M}B_{V,1}^{n_1}\ldots B_{V,M}^{\bar{n}_M}, \quad \forall n_1,\ldots,n_{M-1} \geq 0.$$

The base case $n_1, n_2, \ldots, n_{M-1} = 0$:

$$\|\mathbf{V}_{0,\ldots,\bar{n}_M}\|_X \leq C_{V,M}B_{V,M}^{\bar{n}_M}, \quad \forall \bar{n}_M \geq 0,$$

has previously been established by Theorem 4.5.1 where $\tilde{\varepsilon} = \varepsilon_M$ and $\delta = 0$. Finally, we assume

$$\left\|\mathbf{V}_{n_1,\ldots,n_{M-1},\bar{n}_M}\right\|_X \leq C_{V,1}\ldots C_{V,M}B_{V,1}^{n_1}\ldots B_{V,M}^{\bar{n}_M}, \quad \forall n_1 < \bar{n}_1,\ldots,n_{M-1} < \bar{n}_{M-1}, \quad \forall \bar{n}_M \geq 0,$$

and seek

$$\left\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_{M-1},\bar{n}_M}\right\|_X \leq C_{V,1}\ldots C_{V,M}B_{V,1}^{\bar{n}_1}\ldots B_{V,M}^{\bar{n}_M}.$$

In this case, $\tilde{n} = (n_1,\ldots,n_M)$ and $\tilde{\ell} = (\ell_1,\ldots,\ell_M)$, so we apply (7.3), (7.5) and the mapping properties of $\mathbf{A}_0^{-1}$ to find

$$\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_M}\|_X \leq C_e\left\{\|\mathbf{R}_{\bar{n}_1,\ldots,\bar{n}_M}\|_Y + \sum_{\tilde{\ell}=0}^{\tilde{n}\ *}\left\|\mathbf{A}_{\tilde{n}-\tilde{\ell}}\mathbf{V}_{\tilde{\ell}}\right\|_Y\right\}.$$

Using the estimates on $\mathbf{R}_{n_1,\ldots,n_M}$ and $\mathbf{A}_{n_1,\ldots,n_M}$ (for all $n_1,\ldots,n_M$) and $\mathbf{V}_{n_1,\ldots,n_M}$ ($n_1 < \bar{n}_1,\ldots,n_M < \bar{n}_M$) we have

$$\begin{aligned}\|\mathbf{V}_{\bar{n}_1,\ldots,\bar{n}_M}\|_X \leq C_e\Bigg\{&C_{R,1}\ldots C_{R,M}B_{R,1}^{\bar{n}_1}\ldots B_{R,M}^{\bar{n}_M} + \sum_{\tilde{\ell}=0}^{\tilde{n}\ *} C_{A,1}\ldots C_{A,M}B_{A,1}^{\bar{n}_1-\ell_1}\ldots B_{A,M}^{\bar{n}_M-\ell_M} \\ &\times C_{V,1}\ldots C_{V,M}B_{V,1}^{\ell_1}\ldots B_{V,M}^{\ell_M}\Bigg\} \\ = C_eC_{R,1}\ldots C_{R,M}B_{R,1}^{\bar{n}_1}\ldots B_{R,M}^{\bar{n}_M} &+ C_eC_{A,1}\ldots C_{A,M}C_{V,1}\ldots C_{V,M} \\ \times\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1}\cdots\left(\frac{B_{A,M}}{B_{V,M}}\right)&B_{V,M}^{\bar{n}_M}\sum_{\tilde{\ell}=0}^{\tilde{n}\ *}\left(\frac{B_{A,1}}{B_{V,1}}\right)B_{V,1}^{\bar{n}_1-\ell_1-1}\cdots \\ \times\left(\frac{B_{A,M}}{B_{V,M}}\right)&B_{V,M}^{\bar{n}_M-\ell_M-1}\end{aligned}$$

$$\leq C_e C_{R,1} \dots C_{R,M} B_{V,1}^{\bar{n}_1} \dots B_{V,M}^{\bar{n}_M} + C_e C_{A,1} \dots C_{A,M} C_{V,1} \dots C_{V,M}$$
$$\times \left(\frac{B_{A,1}}{B_{V,1}}\right) B_{V,1}^{\bar{n}_1} \cdots \left(\frac{B_{A,M}}{B_{V,M}}\right) B_{V,M}^{\bar{n}_M} \left(\frac{1}{1-1/2}\right)^M,$$

if $B_{A,i}/B_{V,i} \leq 1/2$, $i = 1, \dots, M$ (implying $B_{V,i} \geq 2B_{A,i}$). We are done if we demand that

$$B_{V,i} \geq B_{R,i}, \quad C_e C_{R,i} \leq C_{V,i}/2, \quad 2^M C_e C_{A,i} C_{V,i} (B_{A,i}/B_{V,i}) \leq C_{V,i}/2.$$

This can be realized if

$$C_{V,i} \geq 2C_e C_{R,i}, \quad B_{V,i} \geq \max\left\{B_{R,i}, 2B_{A,i}, 2^{M+1} C_e C_{A,i} B_{A,i}\right\}.$$

□

Using a similar approach in conjunction with the analysis in Chapters 2 and 3, we predict a more general form of Theorems 2.9.2 and 3.8.1 exists, which would establish the analyticity of the transformed field with respect to any finite $M > 0$ perturbation parameters.

Conjecture 7.4.2. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and $U_{\tilde{n}} \in H^{s+3/2}([0,d])$, $W_{\tilde{n}} \in H^{s+3/2}([0,d])$ such that*

$$\|U_{\tilde{n}}\|_{H^{s+3/2}} \leq K_U B_U^{\tilde{n}}, \quad \|W_{\tilde{n}}\|_{H^{s+3/2}} \leq K_W B_W^{\tilde{n}},$$

for constants $K_U, K_W > 0$ and M–multi–indexed constants $B_U, B_W > 0$, then $u_{\tilde{n}} \in H^{s+2}([0,d] \times [0,a])$, $w_{\tilde{n}} \in H^{s+2}([0,d] \times [-b,0])$ and

$$\|u_{\tilde{n}}\|_{H^{s+2}} \leq K B^{\tilde{n}}, \quad \|w_{\tilde{n}}\|_{H^{s+2}} \leq \tilde{K} \tilde{B}^{\tilde{n}},$$

for constants $K, \tilde{K} > 0$ and M–multi–indexed constants $B, \tilde{B} > 0$.

Analogously, a similar procedure would establish a more general form of Theorems 2.10.2 and 3.9.2 for the analyticity of the DNOs for any finite $M > 0$ perturbation parameters.

Conjecture 7.4.3. *Given any integer $s \geq 0$, if $f \in C^{s+2}([0,d])$ and $U_{\tilde{n}} \in H^{s+3/2}([0,d])$, $W_{\tilde{n}} \in H^{s+3/2}([0,d])$ such that*

$$\|U_{\tilde{n}}\|_{H^{s+3/2}} \leq K_U B_U^{\tilde{n}}, \quad \|W_{\tilde{n}}\|_{H^{s+3/2}} \leq K_W B_W^{\tilde{n}},$$

for constants $K_U, K_W > 0$ and M–multi–indexed constants $B_U, B_W > 0$, then $G_{\tilde{n}} \in H^{s+1/2}([0,d])$, $J_{\tilde{n}} \in H^{s+1/2}([0,d])$ and

$$\|G_{\tilde{n}}\|_{H^{s+1/2}} \leq \tilde{K} \tilde{B}^{\tilde{n}}, \quad \|J_{\tilde{n}}\|_{H^{s+1/2}} \leq \tilde{\tilde{K}} \tilde{\tilde{B}}^{\tilde{n}},$$

for constants $\tilde{K}, \tilde{\tilde{K}} > 0$ *and* M*–multi–indexed constants* $\tilde{B}, \tilde{\tilde{B}} > 0$.

Upon proving these, one has two key ingredients to the more general version of Theorem 4.6.1 which establishes the existence and uniqueness of solutions to a system of partial differential equations with respect to M perturbation parameters.

Conjecture 7.4.4. *Given an integer* $s \geq 0$*, if* $f \in C^{s+2}([0,d])$ *then the equation* (7.1) *has a unique solution,* (7.4)*, and there exist a constant* $C > 0$ *and* M*–multi–indexed constants* $B > 0$ *such that*

$$\|\mathbf{V}_{\tilde{n}}\|_{X^s} \leq CB^{\tilde{n}},$$

for all $\tilde{n} \geq 0$*. This implies that for any* M*–multi–indexed constant* $0 \leq \tilde{\rho} < 1$*,* (7.4)*, converges for all* $\tilde{\varepsilon}$ *such that* $B\tilde{\varepsilon} < \tilde{\rho}$*, i.e.,* $\tilde{\varepsilon} < \tilde{\rho}/B$.

Predictions: In application oriented fields such as signal processing or sea ice modeling, practitioners work with multiple frequencies (127; 128; 129; 130) at short or long wavelengths. Also, as depicted in Figure 35, the grating surface could have M different layers (131) with distinct values of $g_j(x) = \varepsilon f_j(x),\ j = 1, \ldots, M$. A proof of Conjecture 7.4.4 would enable the freedom to enforce any number of perturbation parameters and obtain an analytic solution. Given the widespread availability of parallel computing resources coupled with additional perturbation parameters associated with elastic media, we believe that future research will force hundreds or even thousands of distinct perturbation parameters, all of which should yield an analytic solution.

7.5 Parallel Programming

In the case of multiple layered interfaces, we need to compute intermediate DNOs for up to M layers. This will greatly increase the computational cost and execution time of our HOPS/AWE algorithm and we suspect that it will be necessary to introduce parallel programming techniques to offset the computational expense. In the context of the Operator Expansion (OE) method, preliminary work (132) has been completed in C++ to parallelize the computation of Navier's equations (83; 133). These techniques can be adapted to the TFE method through the choice of OpenMP (134), MPI (135), or CUDA (136).

Predictions: In two or three dimensions, our HOPS code is robust, efficient and has a runtime less than an hour. A local machine with an Intel Core i5 CPU, 8GB of RAM, and Windows 10 OS completed almost every simulation in this thesis in less than thirty minutes. However, with ten to one hundred layer configurations, we suspect that many simulations will take on the order of weeks or even months. As a result, it will be necessary to parallelize our Matlab code in a compiled programming language such as C or C++.

7.6 Alternatives to DNOs

In Chapter 4 we wrote our scattering problem as a linear system

$$\mathbf{AV} = \mathbf{R},$$

where, upon expanding $\{\mathbf{A}, \mathbf{V}, \mathbf{R}\}$ in both ε and δ, we arrived at the flat–interface solution $\mathbf{A}_{0,0}\mathbf{V}_{0,0} = \mathbf{R}_{0,0}$ at order $\mathcal{O}(\varepsilon^0, \delta^0)$. We then saw it was necessary to invert

$$\mathbf{A}_{0,0} = \begin{pmatrix} I & -I \\ -G_{0,0} & -\tau^2 J_{0,0} \end{pmatrix},$$

which features the two DNOs, $G_{0,0}$ and $J_{0,0}$, in order to show the existence and uniqueness of solutions. A primary feature of all HOPS schemes is the inversion of a single, sparse operator $\mathbf{A}_{0,0}$ through the use of DNOs. However, one may ponder if a different technique could produce a more competitive algorithm that is comparable to our HOPS/AWE algorithm (or even better). Is it absolutely necessary to pass in transformed field data in order to efficiently compute and recover internal information stored at the grating surface?

Predictions: A primary advantage of our HOPS/AWE scheme is that for every perturbation order, it is only necessary to invert a single sparse operator corresponding to a flat–interface, order–zero approximation. There are a number of competing approaches in general perturbation theory within the context of layered media problems. In regards to electromagnetic wave scattering, Galerkin and boundary element methods are discussed in (137; 138; 139; 140; 141) and a high–order perturbation approach based on boundary integral equations in (142). High–order schemes for linear waves can be computed using level set methods (143) and fast marching methods, as well as other methods involving domain decomposition (144; 145; 146; 147; 148; 149). A holistic evaluation of these competing methods could potentially improve our HOPS/AWE algorithm if we found a faster method of inverting linear operators without the use of DNOs.

7.7 Computational Complexity

One of the fundamental reasons for developing our HOPS/AWE algorithm is its advantageous computational complexity for problems within its domain of applicability. In comparison with other classical methods, our HOPS/AWE approach has several advantages for computing quantities such as the Reflectivity Map, $R = R(\varepsilon, \delta)$. To demonstrate this we begin by fixing the problem of computing R for N_ε many values of ε and N_δ many values of δ.

In the case of computing the DNOs G and J, we recall from §2.11 and §3.10 that our HOPS/AWE algorithm requires $N_x \times N_z$ unknowns at every perturbation order,

(n, m), corresponding to the N_x equally–spaced gridpoints in the lateral direction and the $N_z + 1$ collocation points in the vertical dimension. In §4.5 we saw that we could write our scattering problem as $\mathbf{A}(\varepsilon, \delta)\mathbf{V}(\varepsilon, \delta) = \mathbf{R}(\varepsilon, \delta)$ where

$$\mathbf{A}(\varepsilon, \delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbf{A}_{n,m}\varepsilon^n\delta^m, \quad \mathbf{R}(\varepsilon, \delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbf{R}_{n,m}\varepsilon^n\delta^m,$$

and

$$\mathbf{V}(\varepsilon, \delta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \mathbf{V}_{n,m}\varepsilon^n\delta^m.$$

At order $\mathcal{O}(\varepsilon^n, \delta^m)$ this becomes

$$\begin{aligned}\mathbf{A}_{0,0}\mathbf{V}_{n,m} = \mathbf{R}_{n,m} &- \sum_{\ell=0}^{n-1} \mathbf{A}_{n-\ell,0}\mathbf{V}_{\ell,m} - \sum_{r=0}^{m-1} \mathbf{A}_{0,m-r}\mathbf{V}_{n,r} \\ &- \sum_{\ell=0}^{n-1}\sum_{r=0}^{m-1} \mathbf{A}_{n-\ell,m-r}\mathbf{V}_{\ell,r}.\end{aligned} \tag{7.6}$$

A careful study of (7.6) reveals that the computational complexity of forming the right–hand side at order (n, m) is

$$\mathcal{O}\left(nmN_x \log(N_x)N_z \log(N_z)\right).$$

Inverting the operator $\mathbf{A}_{0,0}$ has complexity $\mathcal{O}\left(N_x \log(N_x)N_z \log(N_z)\right)$ so the full cost of computing the $\mathbf{V}_{n,m}$, $\{0 \le n \le N, 0 \le m \le M\}$, is

$$\mathcal{O}\left(N^2M^2N_x \log(N_x)N_z \log(N_z)\right).$$

Once these coefficients are recovered, the cost of summing the series in (ε, δ) is minimal, provided it is done in an efficient manner (e.g., by Horner's rule (150; 151)). Our algorithm then requires an additional $\mathcal{O}\left(N_\varepsilon N_\delta\right)$ steps to sum over every value of (ε, δ), therefore the full cost of computing the Reflectivity Map by our HOPS/AWE method is

$$\mathcal{O}\left(N^2M^2N_x \log(N_x)N_z \log(N_z) + N_\varepsilon N_\delta\right). \tag{7.7}$$

In contrast, for a single (ε, δ) pair, a Boundary Integral Method solver with N_x lateral gridpoints requires time proportional to $\mathcal{O}\left(N_x^3\right)$ for Gaussian elimination to solve the resulting dense system of N_x equations in N_x unknowns (150; 151; 56). Applying this $N_\varepsilon \times N_\delta$ times results in a total computational complexity of

$$\mathcal{O}\left(N_x^3 N_\varepsilon N_\delta\right). \tag{7.8}$$

Thus, once N_ε and N_δ become large, e.g.,

$$N_\varepsilon N_\delta > \frac{N^2 M^2 N_x \log(N_x) N_z \log(N_z)}{N_x^3},$$

our new algorithm becomes far more efficient. We speculate that the cost of (7.7) could be reduced to

$$\mathcal{O}\left(NM \log(NM) N_x \log(N_x) N_z \log(N_z) + N_\varepsilon N_\delta\right), \tag{7.9}$$

provided that we develop a more efficient method of computing the $\mathbf{V}_{n,m}$, $\{0 \leq n \leq N, 0 \leq m \leq M\}$, such as reducing the problem space at every step. Alternative approaches to layered media problems have also been proposed by other authors (152; 153), including interpolation (154) and Green's function (155).

Predictions: The combination of implementing parallel programming techniques (through, e.g., OpenMP or CUDA) and reducing the problem space at every step will greatly enhance the speed and fidelity of our HOPS/AWE algorithm. Considering the natural advantage surface methods have over conventional methods, such as finite difference, finite element, and spectral element methods, we expect that our HOPS/AWE algorithm will be among the most competitive available for periodic layered media problems.

APPENDICES

Appendix A

MATLAB CODE

We present a subset of the Matlab code used to simulate our numerical results. Three essential algorithms used to define the upper and lower fields are the computation of the flat–interface solution of $\mathbf{A}_{0,0}$, the upper and lower layer DNOs, $G_{n,m} = G_{n,m}(x;\varepsilon,\delta)$ and $J_{n,m} = J_{n,m}(x;\varepsilon,\delta)$, and the upper and lower transformed field solvers, $u_{n,m} = u_{n,m}(x,z;\varepsilon,\delta)$ and $w_{n,m} = w_{n,m}(x,z;\varepsilon,\delta)$. In Algorithm A.0.1 we show our technique for inverting the flat–interface solution of $\mathbf{A}_{0,0}$ (cf. §4.8).

Algorithm A.0.1 Inversion of the flat–interface operator $\mathbf{A}_{0,0}$

Set Nx: The number of discretization points
Set $i\gamma_p^u$: The Fourier multiplier for $G_{0,0}$
Set $i\gamma_p^w$: The Fourier multiplier for $J_{0,0}$
Set $\gamma^q = (1+\delta)\underline{\gamma}^q, q \in \{u, w\}$, where δ represents a small frequency perturbation
Set τ: Constant representing TE or TM polarization mode
$\zeta_{0,0} \in \mathbb{R}^{N_x},\ \psi_{0,0} \in \mathbb{R}^{N_x}$
$\widehat{\zeta}_{0,0} \in \mathbb{C}^{N_x},\ \widehat{\psi}_{0,0} \in \mathbb{C}^{N_x}$
for $j = 1 : Nx$ **do** $\rightarrow$ Entries of $\left[\widehat{\mathbf{A}}_{0,0}(p)\right]^{-1}$
    $\det_p = -\left\{i\underline{\gamma}_p^u(j) + \tau^2\left(i\underline{\gamma}_p^w(j)\right)\right\}$
    $\underline{a}(j) = \left[\tau^2\left(-i\underline{\gamma}_p^w(j)\right)\widehat{\zeta}_{0,0}(j) + \widehat{\psi}_{0,0}(j)\right] / \det_p$
    $\underline{b}(j) = \left[\left(i\underline{\gamma}_p^u(j)\right)\widehat{\zeta}_{0,0}(j) + \widehat{\psi}_{0,0}(j)\right] / \det_p$
end for
$U_{0,0} = \mathrm{IFFT}(\underline{a}),\ W_{0,0} = \mathrm{IFFT}(\underline{b})$
return $U_{0,0}, W_{0,0}$

Next, Algorithm A.0.2 demonstrates how we calculate the upper layer DNO, G (cf. §2.10 and §2.11).

Algorithm A.0.2 Computation of the upper layer DNO, G

Set Nx: The number of discretization points
Set Nz: The number of collocation points
Set N: The maximum number of Taylor orders for the interfacial perturbation
Set M: The maximum number of Taylor orders for the frequency perturbation
Set dx: The partial derivative with respect to the x component
Set dz: The partial derivative with respect to the z component
Set a: The artificial boundary imposed at the top of the upper layer

Set $\tilde{p} = (2\pi/d)p$ for an integer p where d is the periodicity of the grating interface
Set $f = \sin(x)$, or a similar test function representing the grating surface
Set f_x : The derivative of f with respect to the x component
Set $\ell_{\text{bottom}} = Nz + 1$, the bottom or last collocation point on the z–axis
$u_{n,m} \in \mathbb{C}^{N_x \times (N_z+1) \times (M+1) \times (N+1)}$, $G_{n,m} \in \mathbb{C}^{N_x \times (M+1) \times (N+1)}$
$u_x \in \mathbb{C}^{N_x \times (N_z+1)}$, $u_z \in \mathbb{C}^{N_x \times (N_z+1)}$
for $n = 1 : N$ **do**
    for $m = 1 : M$ **do** → Order (n, m) terms in equation (2.52) in §2.10
        $u_z = dz(u_{n,m}(:, :, m, n), a)$
        $G_{n,m}(:, m, n) = -u_z(:, \ell_{\text{bottom}})$
        if $n > 1$ **then** → Order $(n-1, m)$ terms in equation (2.53) in §2.10
            $u_x = dx(u_{n,m}(:, :, m, n-1), \tilde{p})$
            $G_{n,m}(:, m, n) = G_{n,m}(:, m, n) + f_x u_x(:, \ell_{\text{bottom}})$
            $G_{n,m}(:, m, n) = G_{n,m}(:, m, n) + (1/a) \cdot (f \cdot G_{n,m}(:, m, n-1))$
        end if
        if $n > 2$ **then** → Order $(n-2, m)$ terms in equation (2.53) in §2.10
            $u_x = dx(u_{n,m}(:, :, m, n-2), \tilde{p})$
            $G_{n,m}(:, m, n) = G_{n,m}(:, m, n) - (1/a) \cdot (f f_x \cdot u_x(:, \ell_{\text{bottom}}))$
            $u_z = dz(u_{n,m}(:, :, m, n-2), a)$
            $G_{n,m}(:, m, n) = G_{n,m}(:, m, n) - (f_x^2 \cdot u_z(:, \ell_{\text{bottom}}))$
        end if
    end for
end for
return $G_{n,m}$

Lastly, Algorithm A.0.3 summarizes how we compute the upper transformed field, u (cf. §2.4, §2.6, and §2.11). Due to its complexity and length, we leave out some details to the Matlab implementation in Listing A.1.

Algorithm A.0.3 Computation of the upper transformed field, u

Set Nx: The number of discretization points
Set Nz: The number of collocation points
Set N: The maximum number of Taylor orders for the interfacial perturbation
Set M: The maximum number of Taylor orders for the frequency perturbation
Set dx: The partial derivative with respect to the x component
Set dz: The partial derivative with respect to the z component
Set a: The artificial boundary imposed at the top of the upper layer
Set $\tilde{p} = (2\pi/d)p$ for an integer p where d is the periodicity of the grating interface
Set $f = \cos(x)$, or a similar test function representing the grating surface
Set f_x : The derivative of f with respect to the x component
Set $\ell_{\text{top}} = 0 + 1$, the top or first collocation point on the z–axis
Set T^u : Expansion of frequency operator, cf. §5.4
Set $g(x) = \varepsilon f(x)$, where ε represents a small interfacial perturbation
Set $\gamma^u = (1+\delta)\underline{\gamma}^u$, $\alpha = (1+\delta)\underline{\alpha}$, where δ is a small frequency perturbation

Set $z' = a(z - g(x))/(a - g(x))$, per transformation rules in §2.4. Relabel $z = z'$
$\xi_{n,m} \in \mathbb{R}^{N_x \times (M+1) \times (N+1)},\ \widehat{\xi}_{n,m} \in \mathbb{C}^{N_x \times (M+1) \times (N+1)}$
$U_{n,m} \in \mathbb{R}^{N_x \times (N_z+1)},\ \widehat{U}_{n,m} \in \mathbb{C}^{N_x \times (N_z+1)}$
$F_{n,m} \in \mathbb{R}^{N_x \times (N_z+1)},\ \widehat{F}_{n,m} \in \mathbb{C}^{N_x \times (N_z+1)}$
$u_{n,m} \in \mathbb{C}^{N_x \times (N_z+1) \times (M+1) \times (N+1)},\ J_{n,m} \in \mathbb{R}^{N_x},\ \widehat{J}_{n,m} \in \mathbb{C}^{N_x}$
Compute $A_1^{xx}, A_1^{xz}, A_1^{zx}, A_1^{zz}, A_2^{xx}, A_2^{xz}, A_2^{zx}, A_2^{zz}, B_1^x, B_1^z, B_2^x, B_2^z, S_0, S_1$, and
S_2 through equations (2.16) in §2.4
for $n = 1 : N$ **do**
    for $m = 1 : M$ **do**
        if $n > 1$ **then** → Order $(n-1, m)$ terms in equation (2.28) in §2.6
            $u_x = dx(u_{n,m}(:, :, m, n-1), \tilde{p})$
            $F_{n,m} = F_{n,m} - dx(A_1^{xx} \cdot u_x, \tilde{p})$
            $F_{n,m} = F_{n,m} - dz(A_1^{zx} \cdot u_x, a)$
            $F_{n,m} = F_{n,m} - B_1^x \cdot u_x$
            $u_z = dz(u_{n,m}(:, :, m, n-1), a)$
            $F_{n,m} = F_{n,m} - dx(A_1^{xz} \cdot u_z, \tilde{p})$
            $F_{n,m} = F_{n,m} - (2i\underline{\alpha}) \cdot S_1 \cdot u_x$
            $F_{n,m} = F_{n,m} - (\underline{\gamma}^u)^2 \cdot S_1 \cdot u_{n,m}(:, :, m, n-1)$
        end if
        if $m > 1$ **then** → Order $(n, m-1)$ terms in equation (2.28) in §2.6
            $u_x = dx(u_{n,m}(:, :, m-1, n), \tilde{p})$
            $F_{n,m} = F_{n,m} - (2i\underline{\alpha}) \cdot u_x$
            $F_{n,m} = F_{n,m} - (2(\underline{\gamma}^u)^2) \cdot u_{n,m}(:, :, m-1, n)$
        end if
        if $n > 1$ and $m > 1$ **then** → Order $(n-1, m-1)$ terms in equation (2.28)
            $u_x = dx(u_{n,m}(:, :, m-1, n-1), \tilde{p})$
            $F_{n,m} = F_{n,m} - (2i\underline{\alpha}) \cdot S_1 \cdot u_x$
            $F_{n,m} = F_{n,m} - (2(\underline{\gamma}^u)^2) \cdot S_1 \cdot u_{n,m}(:, :, m-1, n-1)$
        end if
        if $n > 2$ **then** → Order $(n-2, m)$ terms in equation (2.28) in §2.6
            $u_x = dx(u_{n,m}(:, :, m, n-2), \tilde{p})$
            $F_{n,m} = F_{n,m} - dx(A_2^{xx} \cdot u_x, \tilde{p})$
            $F_{n,m} = F_{n,m} - dz(A_2^{zx} \cdot u_x, a)$
            $F_{n,m} = F_{n,m} - B_2^x \cdot u_x$
            $u_z = dz(u_{n,m}(:, :, m, n-2), a)$
            $F_{n,m} = F_{n,m} - dx(A_2^{xz} \cdot u_z, \tilde{p})$
            $F_{n,m} = F_{n,m} - dz(A_2^{zz} \cdot u_z, a)$
            $F_{n,m} = F_{n,m} - B_2^z \cdot u_z - (2i\underline{\alpha}) \cdot S_2 \cdot u_x$
            $F_{n,m} = F_{n,m} - (\underline{\gamma}^u)^2 \cdot S_2 \cdot u_{n,m}(:, :, m, n-2)$
        end if

55: **if** $m > 2$ **then** $\rightarrow$ Order $(n, m-2)$ terms in equation (2.28) in §2.6
56: $F_{n,m} = F_{n,m} - (\underline{\gamma}^u)^2 \cdot u_{n,m}(:,:,m-2,n)$
57: **end if**
58: **if** $n > 1$ and $m > 2$ **then** $\rightarrow$ Order $(n-1, m-2)$ terms in equation (2.28)
59: $F_{n,m} = F_{n,m} - (\underline{\gamma}^u)^2 \cdot S_1 \cdot u_{n,m}(:,:,m-2,n-1)$
60: **end if**
61: **if** $n > 2$ and $m > 1$ **then** $\rightarrow$ Order $(n-2, m-1)$ terms in equation (2.28)
62: $u_x = dx(u_{n,m}(:,:,m-1,n-2), \tilde{p})$
63: $F_{n,m} = F_{n,m} - (2i\underline{\alpha}) \cdot S_2 \cdot u_x$
64: $F_{n,m} = F_{n,m} - (2(\underline{\gamma}^u)^2) \cdot S_2 \cdot u_{n,m}(:,:,m-1,n-2)$
65: **end if**
66: **if** $n > 2$ and $m > 2$ **then** $\rightarrow$ Order $(n-2, m-2)$ terms in equation (2.28)
67: $F_{n,m} = F_{n,m} - (\underline{\gamma}^u)^2 \cdot S_2 \cdot u_{n,m}(:,:,m-2,n-2)$
68: **end if**
69: **for** $r = 0 : m-1$ **do** $\rightarrow$ Transparent boundary condition, (2.29) in §2.6
70: $J_{n,m} = J_{n,m} + \text{IFFT}((T^u(:,m-r)) \cdot \text{FFT}(u_{n,m}(:,\ell_{\text{top}},r,n)))$
71: **end for**
72: **if** $n > 1$ **then**
73: **for** $r = 0 : m$ **do** $\rightarrow$ Transparent boundary condition, (2.29) in §2.6
74: $S_{n,m} = \text{IFFT}(T^u(:,m-r)) \cdot \text{FFT}(u_{n,m}(:,\ell_{\text{top}},r,n-1)))$
75: $J_{n,m} = J_{n,m} - (1.0/a) \cdot f \cdot S_{n,m}$
76: **end for**
77: **end if**
78: $\widehat{F}_{n,m} = \text{FFT}(F_{n,m}),\ \widehat{J}_{n,m} = \text{FFT}(J_{n,m})$
79: $\widehat{U}_{n,m} =$ Chebyshev collocation method of parameters in (2.63) of §2.11
80: **if** $n > 0$ or $m > 0$ **then**
81: $u_{n,m}(:,:,m,n) = \text{IFFT}(\widehat{U}_{n,m})$
82: **end if**
83: **end for**
84: **end for**
85: **return** $u_{n,m}$

We now turn to example Matlab implementations. Our first script is the code for the upper transformed field, $u = u(x, y; \varepsilon, \delta)$ (cf. Algorithm A.0.3). A computational novelty of our HOPS/AWE algorithm is the speed at which we can compute the flat–interface solution in Fourier space by inverting a sparse operator at every wavenumber. To do this, we apply the Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT) in Matlab. Because Matlab array indices start from 1 (linear indexing), we will add "+1" in all of the loop variables executed in our Matlab scripts.

Listing A.1: Upper Field Solver for the TFE Method

```
function [unm] = field_tfe_helmholtz_m_and_n(xi_n_m,f,p,gammap,alpha,...
        gamma,Dz,a,Nx,Nz,N,M,identy)

unm = zeros(Nx,Nz+1,M+1,N+1);

k2 = p(0+1)^2 + gammap(0+1)^2;

ell_top = 0 + 1;
xi_n_m_hat = 0*xi_n_m;
for n=0:N
  for m=0:M
    xi_n_m_hat(:,m+1,n+1) = fft(xi_n_m(:,m+1,n+1));
  end
end
f_x = real(ifft((1i*p).*fft(f)));

ll = [0:Nz]';
z_min = 0; z_max = a;
D = (2/(z_max-z_min))*Dz;
D2 = D*D;
D_start = D(1,:);
D_end = D(end,:);
tilde_z = cos(pi*ll/Nz);
z = ((z_max-z_min)/2.0)*(tilde_z - 1.0) + z_max;

f_full = repmat(f,1,Nz+1);
f_x_full = repmat(f_x,1,Nz+1);
a_minus_z_full = repmat(a - z.',Nx,1);

Uhat = zeros(Nx,Nz+1);

Tu = Tu_dno(alpha,p,gamma,gammap,k2,Nx,M);

% n=0 and m=0

for ell=0:Nz
  unm(:,ell+1,0+1,0+1) = ifft(exp(1i*gammap*z(ell+1)).*xi_n_m_hat(:,0+1,0+1));
end

A1_xx = -(2.0/a)*f_full;
A1_xz = -(1.0/a)*(a_minus_z_full).*f_x_full;
A1_zx = A1_xz;
%A1_zz = 0;

A2_xx = (1.0/a^2)*f_full.^2;
A2_xz = (1.0/a^2)*(a_minus_z_full).*(f_full.*f_x_full);
A2_zx = A2_xz;
A2_zz = (1.0/a^2)*((a_minus_z_full).^2).*(f_x_full.^2);

B1_x = (1.0/a)*f_x_full;
%B1_z = 0;

B2_x = -(1.0/a^2)*f_full.*f_x_full;
```

```
B2_z = -(1.0/a^2).*(a_minus_z_full).*(f_x_full.^2);

S1 = -(2.0/a)*f_full;
S2 = (1.0/a^2)*f_full.^2;

for n=0:N
  for m=0:M

    % Form Fnm, Jnm
    Fnm = zeros(Nx,Nz+1);
    Jnm = zeros(Nx,1);

    if(n>=1)
      u_x = dx(unm(:,:,m+1,n-1+1),p);
      temp = A1_xx.*u_x;
      Fnm = Fnm - dx(temp,p);
      temp = A1_zx.*u_x;
      Fnm = Fnm - dz(temp,Dz,a);
      temp = B1_x.*u_x;
      Fnm = Fnm - temp;

      u_z = dz(unm(:,:,m+1,n-1+1),Dz,a);
      temp = A1_xz.*u_z;
      Fnm = Fnm - dx(temp,p);
      %A1_zz = 0
      %B1_z = 0

      temp = 2*1i*alpha.*S1.*u_x;
      Fnm = Fnm - temp;
          temp = gamma^2.*S1.*unm(:,:,m+1,n-1+1);
          Fnm = Fnm - temp;
    end

        if(m>=1)
          u_x = dx(unm(:,:,m-1+1,n+1),p);
          temp = 2*1i*alpha.*u_x;
          Fnm  = Fnm - temp;
          temp = 2*gamma^2.*unm(:,:,m-1+1,n+1);
          Fnm  = Fnm - temp;
        end

        if(n>=1 && m>=1)
          u_x = dx(unm(:,:,m-1+1,n-1+1),p);
          temp = 2*1i*alpha.*S1.*u_x;
          Fnm  = Fnm - temp;
          temp = 2*gamma^2.*S1.*unm(:,:,m-1+1,n-1+1);
          Fnm  = Fnm - temp;
        end

    if(n>=2)
      u_x = dx(unm(:,:,m+1,n-2+1),p);
      temp = A2_xx.*u_x;
      Fnm = Fnm - dx(temp,p);
      temp = A2_zx.*u_x;
      Fnm = Fnm - dz(temp,Dz,a);
```

```
109         temp = B2_x.*u_x;
110         Fnm = Fnm - temp;
111
112         u_z = dz(unm(:,:,m+1,n-2+1),Dz,a);
113         temp = A2_xz.*u_z;
114         Fnm = Fnm - dx(temp,p);
115         temp = A2_zz.*u_z;
116         Fnm = Fnm - dz(temp,Dz,a);
117         temp = B2_z.*u_z;
118         Fnm = Fnm - temp;
119
120         temp = 2*1i*alpha.*S2.*u_x;
121         Fnm = Fnm - temp;
122             temp = gamma^2.*S2.*unm(:,:,m+1,n-2+1);
123             Fnm = Fnm - temp;
124     end
125
126     if(m>=2)
127             temp = gamma^2.*unm(:,:,m-2+1,n+1);
128             Fnm = Fnm - temp;
129     end
130
131     if(n>=1 && m>=2)
132           temp = gamma^2.*S1.*unm(:,:,m-2+1,n-1+1);
133           Fnm = Fnm - temp;
134     end
135
136     if(n>=2 && m>=1)
137           u_x = dx(unm(:,:,m-1+1,n-2+1),p);
138           temp = 2*1i*alpha.*S2.*u_x;
139           Fnm = Fnm - temp;
140           temp = 2*gamma^2.*S2.*unm(:,:,m-1+1,n-2+1);
141           Fnm = Fnm - temp;
142     end
143
144     if(n>=2 && m>=2)
145           temp = gamma^2.*S2.*unm(:,:,m-2+1,n-2+1);
146           Fnm = Fnm - temp;
147     end
148
149     for r=0:m-1
150       Jnm = Jnm + ifft((Tu(:,m-r+1)).*fft(unm(:,ell_top,r+1,n+1)));
151     end
152     if(n>=1)
153       for r=0:m
154         Snm = ifft((Tu(:,m-r+1)).*fft(unm(:,ell_top,r+1,n-1+1)) );
155         Jnm = Jnm - (1.0/a)*f.*Snm;
156       end
157     end
158
159     % Solve elliptic equation
160
161     Fnmhat = fft(Fnm);
162     Jnmhat = fft(Jnm);
163
```

```
164     b = Fnmhat.';
165     alphaalpha = 1.0;
166     betabeta = 0.0;
167     gammagamma = gamma*gamma - p.^2 - 2*alpha.*p;
168     d_min = 1.0;
169     n_min = 0.0;
170     r_min = xi_n_m_hat(:,m+1,n+1);
171     d_max = -1i*gammap;
172     n_max = 1.0;
173     r_max = Jnmhat;
174
175     % Solve BVP through the Chebyshev collocation method
176
177     Uhat = solvebvp_colloc(Uhat,b,alphaalpha,betabeta,gammagamma,...
178         d_min,n_min,r_min,d_max,n_max,r_max,Nx,identy,D,D2,D_start,D_end);
179
180     if((n>0)||(m>0))
181       unm(:,:,m+1,n+1)=ifft(Uhat);
182     end
183
184   end
185 end
186
187 return;
```

Our next script shows how we use the boundary data from the upper field solver to calculate the transformed field in Fourier space. We recover field data by inverting this operation for every perturbation order of ε and δ.

Listing A.2: BVP Solver for the Chebyshev Collocation Method

```
1  function [Uhat] = solvebvp_colloc(Uhat,b,alpha,beta,gamma,d_min,n_min,...
2      r_min,d_max,n_max,r_max,Nx,identy,D,D2,D_start,D_end)
3
4  A = alpha*D2 + beta*D + reshape(gamma,1,1,Nx).*identy;
5  A(end,:,:) = repmat(n_min*D_end,[1,1,Nx]);
6  b(end,:) = r_min;
7
8  A(1,:,:) = repmat(n_max*D_start,[1,1,Nx]);
9  A(end,end,:) = A(end,end,:) + d_min;
10 A(1,1,:) = A(1,1,:) + reshape(d_max,1,1,Nx);
11 b(1,:) = r_max;
12
13 for j=1:Nx
14    utilde = linsolve(A(:,:,j),b(:,j)); % A\b
15    Uhat(j,:) = utilde.';
16 end
17
18 return;
```

Linsolve (or, equivalently, the backslash operator) is the most computationally expensive part of our algorithm. We can increase the computational speed through making the following changes in the Parallel Computing Toolbox in Matlab. Further improvements

could be made by switching to a compiled programming language such as C++, Fortran, or Julia.

Listing A.3: Parallel Version of the BVP Solver for the Chebyshev Collocation Method

```
% Execute for-loop iterations in parallel on workers
parfor j=1:Nx
    utilde = linsolve(A(:,:,j),b(:,j)); % A\b
    Uhat(j,:) = utilde.';
end

% Remove the for loop by mldivide with GPU arrays
Uhat = permute( pagefun(@mldivide,A,reshape(b,[],1,Nx)),[2,1,3]);
```

We note that these changes are only necessary for large simulations (such as $N = M = 15$ or more Taylor orders and a granularity of $N_\delta = N_\varepsilon = 1000$ per invocation, cf. §6.3 and §6.5). The additional overhead and partitioning for parallel workers or the GPU is unwarranted for smaller simulations which can often be computed in a few minutes or less. Our next script shows how we calculate the upper layer DNO through our TFE methodology (cf. Algorithm A.0.2).

Listing A.4: Upper Layer DNO for the TFE Method

```
function [Gnm] = dno_tfe_helmholtz_m_and_n(unm,f,p,Dz,a,Nx,Nz,N,M)

Gnm = zeros(Nx,M+1,N+1);

ell_bottom = Nz + 1;
f_x = ifft((1i*p).*fft(f));

for n=0:N
  for m=0:M
    u_z = dz(unm(:,:,m+1,n+1),Dz,a);
    Gnm(:,m+1,n+1) = -u_z(:,ell_bottom);
    if(n>=1)
      u_x = dx(unm(:,:,m+1,n-1+1),p);
      Gnm(:,m+1,n+1) = Gnm(:,m+1,n+1) + f_x.*u_x(:,ell_bottom);

      Gnm(:,m+1,n+1) = Gnm(:,m+1,n+1) + (1.0/a)*(f.*Gnm(:,m+1,n-1+1));
    end
    if(n>=2)
      u_x = dx(unm(:,:,m+1,n-2+1),p);
      Gnm(:,m+1,n+1) = Gnm(:,m+1,n+1) - (1.0/a)*(f.*(f_x.*u_x(:,ell_bottom)));

      u_z = dz(unm(:,:,m+1,n-2+1),Dz,a);
      Gnm(:,m+1,n+1) = Gnm(:,m+1,n+1) - f_x.*(f_x.*u_z(:,ell_bottom));
    end
  end
end

return;
```

We then demonstrate how to invert $\mathbf{A}_{0,0}$ by Fourier inversion (cf. Algorithm A.0.1) where we multiply the numerator and denominator by i.

Listing A.5: Inversion of Flat–Interface $\mathbf{A}_{0,0}$

```
function [U,W] = AInverse(Q,R,gammap,gammapw,Nx,tau2)
% Q = Zeta_{0,0}, R = Psi_{0,0}

a = zeros(Nx,1);
b = zeros(Nx,1);
Q_hat = fft(Q);
R_hat = fft(R);

for j=1:Nx
  det_p = tau2*gammapw(j) + gammap(j);
  a(j) = ((tau2*gammapw(j))*Q_hat(j) + 1i*R_hat(j))/det_p;
  b(j) = ((-gammap(j))*Q_hat(j) + 1i*R_hat(j))/det_p;
end

U = ifft(a);
W = ifft(b);

return;
```

Finally, in Figures 36 and 37, we show how Spectral methods are implemented in Matlab. Recalling our strategy in §2.11, we enforce a Fourier spectral method in the x–axis with N_x equally–spaced gridpoints and a Chebyshev spectral method in the z–axis with N_z+1 collocation points where, for brevity, we demonstrate our methods with $N_x = N_z = 8$.

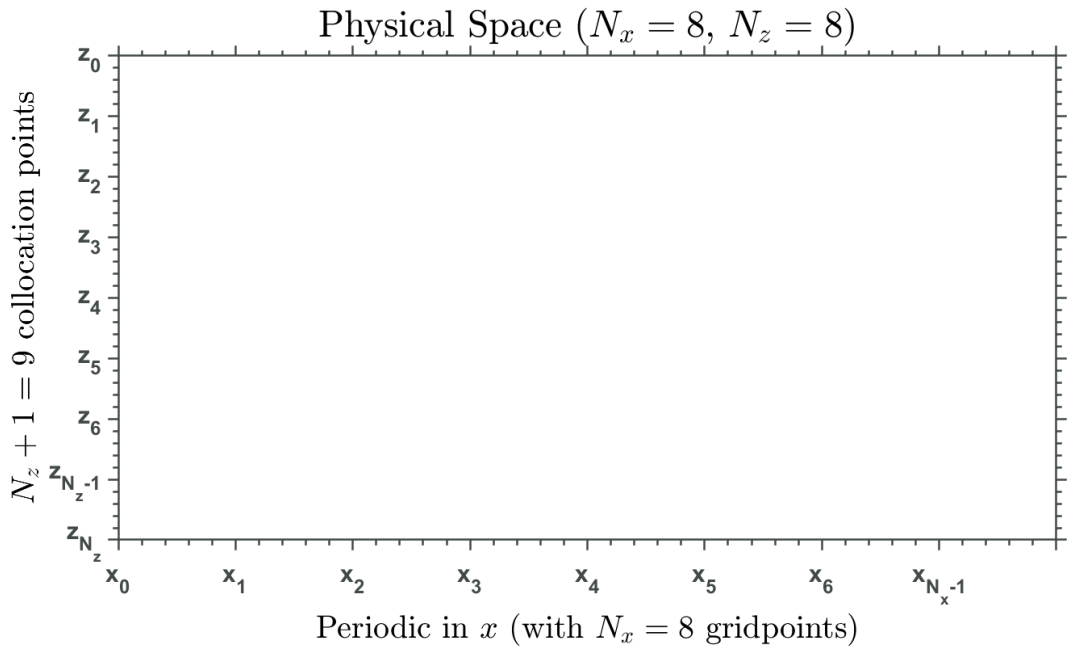


Figure 36: In Physical Space, we consider N_x discretization points on the x–axis and $N_z + 1$ collocation points on the z–axis.

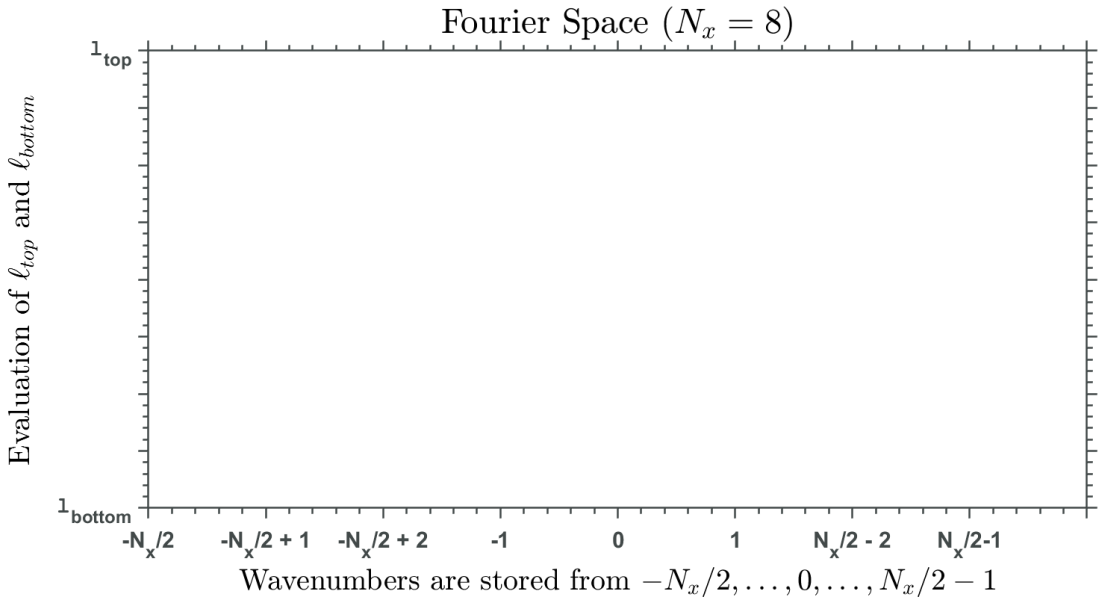


Figure 37: In Fourier Space, wavenumbers are stored in the order $-N_x/2$, $\ldots, 0, \ldots, N_x/2 - 1$ and ℓ_{top} and ℓ_{bottom} are evaluated at the upper boundary $z = a$ and the surface $z = 0$ (cf. Algorithm A.0.2 and A.0.3).

More generally, we consider the Fourier transform on the N_x–point grid with a period of d. For N_x discretization points and a step size of (d/N_x), we have

$$\begin{aligned} \text{Physical Space}: \quad & x \in \Big\{0,\ \frac{d}{Nx}, \frac{2d}{Nx}, \ldots, \frac{(N_x-2)d}{N_x}, \frac{(N_x-1)d}{N_x}\Big\} \\ \text{Fourier Space}: \quad & p \in \Big\{-\frac{N_x}{2}, -\frac{N_x}{2}+1, \ldots, \frac{N_x}{2}-2, \frac{N_x}{2}-1\Big\} \end{aligned}$$

where the Discrete Fourier Transform (DFT) is computed through several applications of the Fast Fourier Transform (FFT).

Appendix B

PROOF OF ALGEBRA PROPERTY, ELLIPTIC ESTIMATE, AND TRANSLATION PROPERTY

As discussed in §2.7, we present the proof of the three major tools used to show joint analyticity of the upper field in the appropriate Sobolev space. Our first property is the "Algebra Property" for estimating products of functions, the second property is a rigorous statement of the "Elliptic Estimate," and our final property shows how to bound translated elements in our function spaces. The same techniques will work for the lower field in §3.6 where the interval $[0, a]$ is translated to $[-b, 0]$.

Lemma B.0.1 (Algebra Property)**.** *Given an integer $s \geq 0$ and any $\sigma > 0$, there exists a constant $\mathcal{M} = \mathcal{M}(s)$ such that if $f \in C^s([0,d]), u \in H^s([0,d] \times [0,a])$ then*

$$\|fu\|_{H^s} \leq \mathcal{M}|f|_{C^s}\|u\|_{H^s}, \tag{B.1}$$

and if $\tilde{f} \in C^{s+1/2+\sigma}([0,d]), \tilde{u} \in H^{s+1/2}([0,d])$ then there exists a constant $\tilde{\mathcal{M}} = \tilde{\mathcal{M}}(s)$ such that

$$\|\tilde{f}\tilde{u}\|_{H^{s+1/2}} \leq \tilde{\mathcal{M}}|\tilde{f}|_{C^{s+1/2+\sigma}}\|\tilde{u}\|_{H^{s+1/2}}. \tag{B.2}$$

Proof. [Lemma B.0.1] Let $s \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $f \in C^s([0,d])$, and $u \in H^s([0,d] \times [0,a])$. We will first verify (B.1). For this, the definition of our Sobolev norms and Leibniz's formula delivers

$$\begin{aligned}
\|fu\|_{H^s}^2 &= \sum_{\ell=0}^{s}\sum_{m=0}^{\ell}\left\|\partial_z^{s-\ell}\partial_x^{\ell-m}(fu)\right\|_{L^2}^2 \\
&= \sum_{\ell=0}^{s}\sum_{m=0}^{\ell}\left\|\sum_{p=0}^{s-\ell}\sum_{q=0}^{\ell-m}\binom{s-\ell}{p}\binom{\ell-m}{q}\left[\partial_z^{s-\ell-p}\partial_x^{\ell-m-q}f\right]\left[\partial_z^p\partial_x^q u\right]\right\|_{L^2}^2
\end{aligned}$$

As $f \in C^s([0,d])$ only depends on the x–component, the expression inside the norm is zero unless $s - \ell = p$ and we deduce

$$\sum_{p=0}^{s-\ell}\sum_{q=0}^{\ell-m}\binom{s-\ell}{p}\binom{\ell-m}{q}\left[\partial_z^{s-\ell-p}\partial_x^{\ell-m-q}f\right]\left[\partial_z^p\partial_x^q u\right] = \sum_{q=0}^{\ell-m}\binom{\ell-m}{q}\left[\partial_x^{\ell-m-q}f\right]\left[\partial_z^{s-\ell}\partial_x^q u\right].$$

Therefore

$$\begin{aligned}\|fu\|_{H^s}^2 &= \sum_{\ell=0}^{s}\sum_{m=0}^{\ell}\left\|\sum_{q=0}^{\ell-m}\binom{\ell-m}{q}\left[\partial_x^{\ell-m-q}f\right]\left[\partial_z^{s-\ell}\partial_x^q u\right]\right\|_{L^2}^2 \\ &\le \sum_{\ell=0}^{s}\sum_{m=0}^{\ell}\sum_{q=0}^{\ell-m}\binom{\ell-m}{q}\left|\partial_x^{\ell-m-q}f\right|_{L^\infty}^2\left\|\partial_z^{s-\ell}\partial_x^q u\right\|_{L^2}^2 \\ &\le \sum_{\ell=0}^{s}\sum_{m=0}^{\ell}\sum_{q=0}^{\ell-m}\binom{\ell-m}{q}|f|_{C^s}^2\,\|u\|_{H^s}^2\,. \end{aligned} \tag{B.3}$$

By the binomial theorem we may observe

$$\sum_{q=0}^{\ell-m}\binom{\ell-m}{q} = 2^{\ell-m}.$$

Inserting the above expression into (B.3) and repeatedly applying the definition of the geometric series gives

$$\begin{aligned}\|fu\|_{H^s}^2 &\le \sum_{\ell=0}^{s}\sum_{m=0}^{\ell}2^{\ell-m}\,|f|_{C^s}^2\,\|u\|_{H^s}^2 \\ &= \sum_{\ell=0}^{s}2^{\ell}\sum_{m=0}^{\ell}2^{-m}\,|f|_{C^s}^2\,\|u\|_{H^s}^2 \\ &= \sum_{\ell=0}^{s}2^{\ell}\left(2-2^{-m}\right)|f|_{C^s}^2\,\|u\|_{H^s}^2 \\ &\le \sum_{\ell=0}^{s}2^{\ell+1}\,|f|_{C^s}^2\,\|u\|_{H^s}^2 \\ &= \left(2^{s+2}-2\right)|f|_{C^s}^2\,\|u\|_{H^s}^2\,,\end{aligned}$$

and the inequality (B.1) follows by taking the square root.

Next, we follow (114) to verify (B.2). Suppose $s=\rho$ for $0<\rho<1$ and $\Omega\subseteq\mathbb{R}^n$. Then a norm in $H^\rho(\Omega)$ that is equivalent to the usual Sobolev space norm is defined as

$$\|\tilde{u}\|_{H^\rho}^2 := \|\tilde{u}\|_{L^2}^2 + \int_\Omega\int_\Omega \frac{|\tilde{u}(x)-\tilde{u}(z)|^2}{|e^{ix}-e^{iz}|^{2\rho+n}}\,\mathrm{d}x\,\mathrm{d}z. \tag{B.4}$$

The above definition is a fractional order Sobolev space known as the Sobolev–Slobodeckij space. To establish (B.2) we start with the case $s=0$ and evaluate (B.4) with $\rho=1/2$

$$\begin{aligned}\|\tilde{f}\tilde{u}\|_{H^{1/2}}^2 &= \|\tilde{f}\tilde{u}\|_{L^2}^2 + \int_\Omega\int_\Omega \frac{|\tilde{f}(x)\tilde{u}(x)-\tilde{f}(z)\tilde{u}(z)|^2}{|e^{ix}-e^{iz}|^{n+1}}\,\mathrm{d}x\,\mathrm{d}z \\ &\le |\tilde{f}|_{L^\infty}^2\|\tilde{u}\|_{L^2}^2 + 2\int_\Omega\int_\Omega \frac{|\tilde{f}(x)-\tilde{f}(z)|^2}{|e^{ix}-e^{iz}|^{n+1}}|\tilde{u}(x)|^2\,\mathrm{d}x\,\mathrm{d}z \\ &\quad + 2\int_\Omega\int_\Omega |\tilde{f}(z)|^2\frac{|\tilde{u}(x)-\tilde{u}(z)|^2}{|e^{ix}-e^{iz}|^{n+1}}\,\mathrm{d}x\,\mathrm{d}z,\end{aligned} \tag{B.5}$$

where it is clear that the first and third terms in (B.5) can be grouped together and bounded

$$|\tilde{f}|^2_{L^\infty}\|\tilde{u}\|^2_{L^2} + 2\int_\Omega\int_\Omega |\tilde{f}(z)|^2 \frac{|\tilde{u}(x)-\tilde{u}(z)|^2}{|e^{ix}-e^{iz}|^{n+1}}\,\mathrm{d}x\,\mathrm{d}z \le C|\tilde{f}|^2_{L^\infty}\|\tilde{u}\|^2_{H^{1/2}}.$$

To bound the second term in (B.5) we observe

$$\begin{aligned} 2\int_\Omega\int_\Omega &\frac{|\tilde{f}(x)-\tilde{f}(z)|^2}{|e^{ix}-e^{iz}|^{n+1}}|\tilde{u}(x)|^2\,\mathrm{d}x\,\mathrm{d}z \\ &\le 2|\tilde{f}|^2_{C^{1/2+\sigma}}\int_\Omega\int_\Omega \frac{|x-z|^{1+2\sigma}}{|e^{ix}-e^{iz}|^{n+1}}|\tilde{u}(x)|^2\,\mathrm{d}x\,\mathrm{d}z \\ &\le C|\tilde{f}|^2_{C^{1/2+\sigma}}\|\tilde{u}\|^2_{L^2}, \end{aligned} \tag{B.6}$$

so that (B.5) and (B.6) establish the inequality (B.2) in the case $s = 0$

$$\|\tilde{f}\tilde{u}\|_{H^{1/2}} \le \tilde{\mathcal{M}}(s)|\tilde{f}|_{C^{1/2+\sigma}}\|\tilde{u}\|_{H^{1/2}}.$$

In general, for $s > 0$ we have

$$\|\tilde{u}\|^2_{H^{s+1/2}} = \|\tilde{u}\|^2_{H^s} + \|\partial_x^s\tilde{u}\|^2_{H^{1/2}}, \tag{B.7}$$

and from (B.1)

$$\|\tilde{f}\tilde{u}\|_{H^s} \le \tilde{\mathcal{M}}(s)|\tilde{f}|_{C^s}\|\tilde{u}\|_{H^s}. \tag{B.8}$$

For $s > 0$, the regularity of $\tilde{f} \in C^{s+1/2+\sigma}(\Omega)$ and the estimates (B.5) and (B.6) imply

$$\|\partial_x^s(\tilde{f}\tilde{u})\|_{H^{1/2}} \le \tilde{\mathcal{M}}(s)|\tilde{f}|_{C^{s+1/2+\sigma}}\|\tilde{u}\|_{H^{s+1/2}}. \tag{B.9}$$

Finally, the equation (B.7) and estimates (B.8) and (B.9) deliver

$$\|\tilde{f}\tilde{u}\|^2_{H^{s+1/2}} = \|\tilde{f}\tilde{u}\|^2_{H^s} + \|\partial_x^s(\tilde{f}\tilde{u})\|^2_{H^{1/2}} \le \tilde{\mathcal{M}}(s)|\tilde{f}|^2_{C^{s+1/2+\sigma}}\|\tilde{u}\|^2_{H^{s+1/2}},$$

which is the required estimate for $s > 0$. □

Theorem B.0.2 (Elliptic Estimate)**.** *Given an integer $s \ge 0$, if $F \in H^s([0,d]) \times [0,a])$, $\zeta^u \in H^{s+3/2}([0,d])$, $P \in H^{s+1/2}([0,d])$, then there exists a unique solution $u \in H^{s+2}([0,d]) \times [0,a])$ of*

$$\Delta u(x,z) + 2i\underline{\alpha}\partial_x u(x,z) + (\underline{\gamma}^u)^2 u(x,z) = F(x,z), \qquad 0 < z < a, \tag{B.10a}$$

$$u(x,0) = \zeta^u(x,0), \qquad \text{at } z = 0, \tag{B.10b}$$

$$u(x+d,z) = u(x,z), \tag{B.10c}$$

$$\partial_z u(x,a) - T_0^u[u(x,a)] = P(x), \qquad \text{at } z = a, \tag{B.10d}$$

satisfying

$$\|u\|_{H^{s+2}} \leq C_e\{\|F\|_{H^s} + \|\zeta^u\|_{H^{s+3/2}} + \|P\|_{H^{s+1/2}}\}, \tag{B.11}$$

for some constant $C_e = C_e(s) > 0$.

Proof. [Lemma B.0.2] Following (156), we let $\tilde{\zeta}^u = [-\partial_z u]_{z=0}$ where we define the DNO

$$G : (\zeta^u, P, F) \to \tilde{\zeta}^u, \quad G[\zeta^u, P, F] = G^{(0)}[\zeta^u] + G^{(a)}[P] + G^{([0,a])}[F].$$

With these, we will obtain the estimates

$$\left\|G^{(0)}[\zeta^u]\right\|_{H^{s+1/2}} \leq C_{G^{(0)}} \|\zeta^u\|_{H^{s+3/2}}, \tag{B.12a}$$

$$\left\|G^{(a)}[P]\right\|_{H^{s+1/2}} \leq C_{G^{(a)}} \|P\|_{H^{s+1/2}}, \tag{B.12b}$$

$$\left\|G^{([0,a])}[F]\right\|_{H^{s+1/2}} \leq C_{G^{([0,a])}} \|F\|_{H^s}. \tag{B.12c}$$

As in §2.11, we posit the expansions

$$\{u, F\}(x,z) = \sum_{p=-\infty}^{\infty} \{\hat{u}_p, \hat{F}_p\}(z) e^{i\tilde{p}x}, \quad \{\zeta^u, P\}(x) = \sum_{p=-\infty}^{\infty} \{\hat{\zeta}^u_p, \hat{P}_p\} e^{i\tilde{p}x},$$

into (B.10) which delivers the two–point boundary value problem

$$\begin{aligned} \partial_z^2 \hat{u}_p(z) + \left((\underline{\gamma}^u_p)^2 - \tilde{p}^2 - 2\underline{\alpha}\tilde{p}\right)\hat{u}_p(z) &= \hat{F}_p(z), && 0 < z < a, \\ \hat{u}_p(0) &= \hat{\zeta}^u_p, && \text{at } z = 0, \\ \partial_z\left[\hat{u}_p(a)\right] - (i\underline{\gamma}^u_p)[\hat{u}_p(a)] &= \hat{P}_p, && \text{at } z = a, \end{aligned}$$

where

$$\underline{\gamma}^u_p = \begin{cases} (\underline{\gamma}^u_p)' := \sqrt{(\underline{k}^u)^2 - \underline{\alpha}_p^2}, & \underline{\alpha}_p^2 < (\underline{k}^u)^2, \\ 0, & \underline{\alpha}_p^2 = (\underline{k}^u)^2, \quad (\underline{\gamma}^u_p)', (\underline{\gamma}^u_p)'' \in \mathbb{R}, \quad (\underline{\gamma}^u_p)', (\underline{\gamma}^u_p)'' > 0. \\ i(\underline{\gamma}^u_p)'' := i\sqrt{\underline{\alpha}_p^2 - (\underline{k}^u)^2}, & \underline{\alpha}_p^2 > (\underline{k}^u)^2, \end{cases}$$

The primed notation denotes $'$ as the real part and $''$ as the imaginary part. Observing

$$(\underline{\gamma}^u_p)^2 - \tilde{p}^2 - 2\underline{\alpha}\tilde{p} = \underline{\alpha}^2 + (\underline{\gamma}^u_p)^2 - (\underline{\alpha} + \tilde{p})^2 := (\underline{k}^u)^2 - \underline{\alpha}_p^2 = (\underline{\gamma}^u_p)^2,$$

delivers

$$\begin{aligned} \partial_z^2 \hat{u}_p(z) + (\underline{\gamma}_p^u)^2 \hat{u}_p(z) &= \hat{F}_p(z), && 0 < z < a, \\ \hat{u}_p(0) &= \hat{\zeta}_p^u, && \text{at } z = 0, \\ \partial_z \left[\hat{u}_p(a)\right] - (i\underline{\gamma}_p^u)[\hat{u}_p(a)] &= \hat{P}_p, && \text{at } z = a. \end{aligned}$$

We now consider a function $\Phi_0(z;p)$ satisfying

$$\begin{aligned} \partial_z^2 \Phi_0(z;p) + (\underline{\gamma}_p^u)^2 \Phi_0(z;p) &= 0, && 0 < z < a, \\ \Phi_0(0;p) &= 1, && \text{at } z = 0, \\ \partial_z \Phi_0(a;p) - (i\underline{\gamma}_p^u)\Phi_0(a;p) &= 0, && \text{at } z = a, \end{aligned}$$

so that the solution of

$$\begin{aligned} \partial_z^2 \hat{u}_p(z) + (\underline{\gamma}_p^u)^2 \hat{u}_p(z) &= 0, && 0 < z < a, \\ \hat{u}_p(0) &= \hat{\zeta}_p^u, && \text{at } z = 0, \\ \partial_z \left[\hat{u}_p(a)\right] - (i\underline{\gamma}_p^u)[\hat{u}_p(a)] &= 0, && \text{at } z = a, \end{aligned}$$

is

$$\hat{u}_p(z) = \hat{\zeta}_p^u \Phi_0(z;p).$$

Similarly, we consider a function $\Phi_a(z;p)$ satisfying

$$\begin{aligned} \partial_z^2 \Phi_a(z;p) + (\underline{\gamma}_p^u)^2 \Phi_a(z;p) &= 0, && 0 < z < a, \\ \Phi_a(0;p) &= 0, && \text{at } z = 0, \\ \partial_z \Phi_a(a;p) - (i\underline{\gamma}_p^u)\Phi_a(a;p) &= 1, && \text{at } z = a, \end{aligned}$$

so that the solution of

$$\begin{aligned} \partial_z^2 \hat{u}_p(z) + (\underline{\gamma}_p^u)^2 \hat{u}_p(z) &= 0, && 0 < z < a, \\ \hat{u}_p(0) &= 0, && \text{at } z = 0, \\ \partial_z \left[\hat{u}_p(a)\right] - (i\underline{\gamma}_p^u)[\hat{u}_p(a)] &= \hat{P}_p, && \text{at } z = a, \end{aligned}$$

is

$$\hat{u}_p(z) = \hat{P}_p \Phi_a(z;p).$$

With these, the unique solution of the two–point boundary value problem is given by

$$\hat{u}_p(z) = \hat{\zeta}_p^u \Phi_0(z;p) + \hat{P}_p e^{i\underline{\gamma}_p^u a} \Phi_a(z;p) - I_0[\hat{F}_p](z) - I_a[\hat{F}_p](z), \tag{B.13}$$

where one can readily verify that

$$\Phi_0(z;p) = e^{i\underline{\gamma}_p^u z} := \begin{cases} e^{i(\underline{\gamma}_p^u)' z}, & \underline{\alpha}_p^2 < (\underline{k}^u)^2, \\ 1, & \underline{\alpha}_p^2 = (\underline{k}^u)^2, \\ e^{-(\underline{\gamma}_p^u)'' z}, & \underline{\alpha}_p^2 > (\underline{k}^u)^2, \end{cases}$$

and

$$\Phi_a(z;p) = \frac{\sinh(\underline{\gamma}_p^u z)}{\underline{\gamma}_p^u} := \begin{cases} \frac{\sin\left((\underline{\gamma}_p^u)' z\right)}{(\underline{\gamma}_p^u)'}, & \underline{\alpha}_p^2 < (\underline{k}^u)^2, \\ z, & \underline{\alpha}_p^2 = (\underline{k}^u)^2, \\ \frac{\sinh\left((\underline{\gamma}_p^u)'' z\right)}{(\underline{\gamma}_p^u)''}, & \underline{\alpha}_p^2 > (\underline{k}^u)^2, \end{cases}$$

and

$$\begin{aligned} I_0[\hat{F}_p](z) &:= \int_0^z \Phi_0(z;p)\Phi_a(s;p)\hat{F}_p(s)\,\mathrm{d}s, \\ I_a[\hat{F}_p](z) &:= \int_z^a \Phi_0(s;p)\Phi_a(z;p)\hat{F}_p(s)\,\mathrm{d}s. \end{aligned}$$

By the Leibniz integral rule

$$\begin{aligned} \partial_z I_0[\hat{F}_p](z) &= \Phi_0(z;p)\Phi_a(z;p)\hat{F}_p(z) + \int_0^z \left(\partial_z \Phi_0(z;p)\right)\Phi_a(s;p)\hat{F}_p(s)\,\mathrm{d}s, \\ \partial_z I_a[\hat{F}_p](z) &= -\Phi_0(z;p)\Phi_a(z;p)\hat{F}_p(z) + \int_z^a \Phi_0(s;p)\left(\partial_z \Phi_a(z;p)\right)\hat{F}_p(s)\,\mathrm{d}s. \end{aligned}$$

Adding the two expressions above and substituting the result into (B.13) gives

$$\partial_z \hat{u}_p(z) = \hat{\zeta}_p^u \partial_z \Phi_0(z;p) + \hat{P}_p e^{i\underline{\gamma}_p^u a}\partial_z \Phi_a(z;p) - \tilde{I}_0[\hat{F}_p](z) - \tilde{I}_a[\hat{F}_p](z), \tag{B.14}$$

where

$$\begin{aligned} \tilde{I}_0[\hat{F}_p](z) &:= \int_0^z \left(\partial_z \Phi_0(z;p)\right)\Phi_a(s;p)\hat{F}_p(s)\,\mathrm{d}s, \\ \tilde{I}_a[\hat{F}_p](z) &:= \int_z^a \Phi_0(s;p)\left(\partial_z \Phi_a(z;p)\right)\hat{F}_p(s)\,\mathrm{d}s. \end{aligned}$$

Evaluating (B.14) at $z = 0$ and multiplying by negative one yields

$$\begin{aligned} -\partial_z \hat{u}_p(0) &= -\hat{\zeta}_p^u \partial_z \Phi_0(0;p) - \hat{P}_p e^{i\underline{\gamma}_p^u a}\partial_z \Phi_a(0;p) + \tilde{I}_0[\hat{F}_p](0) + \tilde{I}_a[\hat{F}_p](0) \\ &= -\hat{\zeta}_p^u (i\underline{\gamma}_p^u) - \hat{P}_p e^{i\underline{\gamma}_p^u a} + \int_0^a e^{i\underline{\gamma}_p^u s}\hat{F}_p(s)\,\mathrm{d}s. \end{aligned}$$

From this, we deduce

$$G^{(0)}[\zeta^u] = -\sum_{p=-\infty}^{\infty} \left[\partial_z \Phi_0(0;p)\right]\hat{\zeta}_p^u e^{i\tilde{p}x} = \sum_{p=-\infty}^{\infty} (-i\underline{\gamma}_p^u)\hat{\zeta}_p^u e^{i\tilde{p}x},$$

and

$$G^{(a)}[P] = -\sum_{p=-\infty}^{\infty} \left[e^{i\underline{\gamma}_p^u a}\partial_z \Phi_a(0;p)\right]\hat{P}_p e^{i\tilde{p}x} = \sum_{p=-\infty}^{\infty} (-e^{i\underline{\gamma}_p^u a})\hat{P}_p e^{i\tilde{p}x},$$

and

$$G^{([0,a])}[F] = \sum_{p=-\infty}^{\infty} \int_0^a \left(e^{i\underline{\gamma}_p^u s}\hat{F}_p(s)\,\mathrm{d}s\right) e^{i\tilde{p}x}.$$

With these, we use our Sobolev norms in §2.7 and follow the proof of Lemma 2.8.2 to estimate

$$\begin{aligned}\left\|G^{(0)}[\zeta^u]\right\|_{H^{s+1/2}}^2 &= \sum_{p=-\infty}^{\infty} \left|(i\underline{\gamma}_p^u)\hat{\zeta}_p^u\right|^2 \langle\tilde{p}\rangle^{2(s+1/2)} \\ &\le C_{G^{(0)}} \sum_{p=-\infty}^{\infty} \left|\hat{\zeta}_p^u\right|^2 \langle\tilde{p}\rangle^{2(s+3/2)} \\ &= C_{G^{(0)}} \left\|\zeta^u\right\|_{H^{s+3/2}}^2.\end{aligned}$$

and

$$\begin{aligned}\left\|G^{(a)}[P]\right\|_{H^{s+1/2}}^2 &= \sum_{p=-\infty}^{\infty} \left|\left(e^{i\underline{\gamma}_p^u a}\right)\hat{P}_p\right|^2 \langle\tilde{p}\rangle^{2(s+1/2)} \\ &\le C_{G^{(a)}} \sum_{p=-\infty}^{\infty} \left|\hat{P}_p\right|^2 \langle\tilde{p}\rangle^{2(s+1/2)} \\ &= C_{G^{(a)}} \left\|P\right\|_{H^{s+1/2}}^2.\end{aligned}$$

We then apply the Cauchy–Schwarz inequality to estimate

$$\begin{aligned}\left\|G^{([0,a])}[F]\right\|_{H^{s+1/2}}^2 &= \sum_{p=-\infty}^{\infty} \left|\int_0^a e^{i\underline{\gamma}_p^u s}\hat{F}_p(s)\,\mathrm{d}s\right|^2 \langle\tilde{p}\rangle^{2(s+1/2)} \\ &\le \sum_{p=-\infty}^{\infty} \int_0^a \left|e^{i\underline{\gamma}_p^u s}\right|^2 \mathrm{d}s \int_0^a \left|\hat{F}_p(s)\right|^2 \mathrm{d}s\ \langle\tilde{p}\rangle^{2(s+1/2)}.\end{aligned}$$

By the definition of $\Phi_0(s;p)$ the middle term becomes

$$\int_0^a \left|e^{i\underline{\gamma}_p^u s}\right|^2 \mathrm{d}s = \begin{cases} a, & \underline{\alpha}_p^2 < (\underline{k}^u)^2, \\ a, & \underline{\alpha}_p^2 = (\underline{k}^u)^2, \\ \int_0^a e^{-2(\underline{\gamma}_p^u)'' s}\,\mathrm{d}s, & \underline{\alpha}_p^2 > (\underline{k}^u)^2, \end{cases}$$

where

$$\int_0^a e^{-2(\underline{\gamma}_p^u)'' s}\,\mathrm{d}s = \frac{1}{2(\underline{\gamma}_p^u)''}\left(1 - e^{-2(\underline{\gamma}_p^u)'' a}\right) \le \frac{1}{2(\underline{\gamma}_p^u)''}.$$

By defining for $q \in \{u, w\}$

$$\underline{\mathcal{U}}^q := \{p \in \mathbb{Z} \mid \underline{\alpha}_p^2 \leq (\underline{k}^q)^2\}, \quad \underline{\gamma}_p^q := \begin{cases} \sqrt{(\underline{k}^q)^2 - \underline{\alpha}_p^2}, & p \in \underline{\mathcal{U}}^q, \\ i\sqrt{\underline{\alpha}_p^2 - (\underline{k}^q)^2}, & p \notin \underline{\mathcal{U}}^q, \end{cases} \tag{B.15}$$

the third estimate follows from the bounds

$$\begin{aligned} \left\| G^{([0,a])}[F] \right\|_{H^{s+1/2}}^2 &\leq \sum_{p=-\infty}^{\infty} \int_0^a \left| e^{i\underline{\gamma}_p^u s} \right|^2 \mathrm{d}s \int_0^a \left| \hat{F}_p(s) \right|^2 \mathrm{d}s \ \langle \tilde{p} \rangle^{2(s+1/2)} \\ &\leq \sum_{p \in \underline{\mathcal{U}}^u} a \langle \tilde{p} \rangle^{2(s+1/2)} \left\| \hat{F}_p \right\|_{L^2([0,a])}^2 + \sum_{p \notin \underline{\mathcal{U}}^u} \frac{\langle \tilde{p} \rangle^{2(s+1/2)}}{2(\underline{\gamma}_p^u)''} \left\| \hat{F}_p \right\|_{L^2([0,a])}^2 \\ &\leq C \sum_{p \in \underline{\mathcal{U}}^u} \langle \tilde{p} \rangle^{2s} \left\| \hat{F}_p \right\|_{L^2([0,a])}^2 + \tilde{C} \sum_{p \notin \underline{\mathcal{U}}^u} \langle \tilde{p} \rangle^{2s} \left\| \hat{F}_p \right\|_{L^2([0,a])}^2 \\ &\leq C_{G^{([0,a])}} \sum_{p=-\infty}^{\infty} \langle \tilde{p} \rangle^{2s} \left\| \hat{F}_p \right\|_{L^2([0,a])}^2, \quad C_{G^{([0,a])}} = \max\{a\langle \tilde{p} \rangle, 1/2\} \\ &= C_{G^{([0,a])}} \|F\|_{H^s}^2, \end{aligned}$$

which validates (B.12). These imply

$$\|u\|_{H^{s+2}} \leq C_e \{\|F\|_{H^s} + \|\zeta^u\|_{H^{s+3/2}} + \|P\|_{H^{s+1/2}}\},$$

where

$$C_e := \max\left\{C_{G^{(0)}}, C_{G^{(a)}}, C_{G^{([0,a])}}\right\}.$$

□

Lemma B.0.3 (Translation Property)**.** *Given an integer $s \geq 0$, if $F \in H^s([0,d]) \times [0,a])$, then $(a-z)F \in H^s([0,d]) \times [0,a])$ and there exists a positive constant $Z_a = Z_a(s)$ such that*

$$\|(a-z)F\|_{H^s} \leq Z_a \|F\|_{H^s}.$$

Proof. [Lemma B.0.3] As $(a-z)$ is a constant, it is clear that $(a-z)F \in H^s([0,d]) \times [0,a])$. The required estimate then follows from applying Lemma B.0.1.

□

Appendix C

CHANGE OF VARIABLES

This appendix covers a fundamental step in our Boundary Perturbation algorithm. One of the primary objectives in Chapters 2 and 3 is to show that both the upper/lower fields and the upper/lower layer DNOs are analytic with respect to two small perturbation parameters. In order to do this, we perform a domain–flattening change of variables (known as σ–coordinates in oceanography (157) and the C–method in the dynamical theory of gratings (158; 159)). We will present the theory in Cartesian coordinates and will later state the effects on the Helmholtz equation and the overall impact on our governing equations. The bulk of our analysis is based on Appendix E in (84).

We begin by considering the doubly–perturbed domain

$$S_{L,U} := \{L(x) < z < U(x)\} = \{\overline{\ell} + \ell(x) < z < \overline{u} + u(x)\}, \tag{C.1}$$

where the change of variables

$$x' = x, \quad z' = \overline{\ell}\left(\frac{U - z}{U - L}\right) + \overline{u}\left(\frac{z - L}{U - L}\right), \tag{C.2}$$

maps $S_{L,U}$ to $S_{\overline{\ell},\overline{u}}$. As discussed in Chapter 1, the variables u and U both refer to the upper boundary while ℓ and L reference the lower boundary. In the upper layer, the upper boundary is the artificial boundary at $\{z = a\}$ while the lower boundary is the surface $z = g(x)$. In the lower layer, the upper boundary is the surface $z = g(x)$ and the lower boundary is the artificial boundary at $\{z = -b\}$. Defining the height of the layer to be

$$\overline{h} := \overline{u} - \overline{\ell},$$

and using the formulas for L and U, we find

$$\left(1 + \frac{u(x) - \ell(x)}{\overline{h}}\right) z' = z - \left(\frac{\overline{u}\ell(x) - \overline{\ell}u(x)}{\overline{h}}\right),$$

or

$$C(x)z' = z - D(x),$$

where

$$C(x) := 1 + \frac{u(x) - \ell(x)}{\overline{h}}, \quad D(x) := \frac{\overline{u}\ell(x) - \overline{\ell}u(x)}{\overline{h}}. \tag{C.3}$$

In the upper layer we have

$$\overline{\ell} = 0, \quad \ell = g, \quad \overline{u} = a, \quad u = 0, \quad \overline{h} = \overline{u} - \overline{\ell} = a.$$

Similarly, in the lower layer we have

$$\overline{\ell} = -b, \quad \ell = 0, \quad \overline{u} = 0, \quad u = g, \quad \overline{h} = \overline{u} - \overline{\ell} = b.$$

For a function $v = v(x, z)$, $v \in \{u, w\}$, which is transformed to

$$v' = v'(x', z') = v(x(x', z'), z(x', z')), \quad v = v(x, z) = v'(x'(x, z), z'(x, z)),$$

and $v' \in \{u', w'\}$, we apply the chain rule

$$\frac{\partial v}{\partial x} = \frac{\partial v'}{\partial x'}\frac{\partial x'}{\partial x} + \frac{\partial v'}{\partial z'}\frac{\partial z'}{\partial x}, \quad \frac{\partial v}{\partial z} = \frac{\partial v'}{\partial x'}\frac{\partial x'}{\partial z} + \frac{\partial v'}{\partial z'}\frac{\partial z'}{\partial z}.$$

Then

$$\frac{\partial x'}{\partial x} = 1, \quad \frac{\partial x'}{\partial z} = 0, \quad \frac{\partial z'}{\partial z} = \frac{1}{C},$$

where differentiating $Cz' = z - D$ with respect to x yields

$$(\partial_x C)\, z' + C\left(\frac{\partial z'}{\partial x}\right) = -\left(\partial_x D\right),$$

and

$$\frac{\partial z'}{\partial x} = -\left(\frac{(\partial_x C)\, z' + (\partial_x D)}{C}\right) = -\frac{E}{C}.$$

We define

$$E(x, z') := (\partial_x C)\, z' + (\partial_x D)\,,$$

and observe that

$$\partial_x C = \frac{\partial_x u - \partial_x \ell}{\overline{h}}, \quad \partial_x D = \frac{\overline{u}\partial_x \ell - \overline{\ell}\partial_x u}{\overline{h}}.$$

This implies

$$E = \frac{(\partial_x u - \partial_x \ell)z' + \overline{u}\partial_x \ell - \overline{\ell}\partial_x u}{\overline{h}} = (\partial_x u)Z_L + (\partial_x \ell)Z_U, \tag{C.4}$$

for the definitions

$$Z_L := \frac{z' - \overline{\ell}}{\overline{h}}, \quad Z_U := \frac{\overline{u} - z'}{\overline{h}}.$$

We will later realize that it is more convenient to express our differentiation rules when premultiplied by C (either $C(x)$ or $C(x')$ as appropriate) by which we settle upon the following differentiation rules under the change of variables in (C.2)

$$C\partial_x = C\partial_{x'} - E\partial_{z'}, \quad C\partial_z = \partial_{z'}. \tag{C.5}$$

In §2.2 and §3.2 we showed that the Helmholtz equation in the upper and lower layers can be represented by

$$\Delta v + 2i\alpha\partial_x v + (\gamma^v)^2 v = 0. \tag{C.6}$$

We restate (C.6) as

$$\begin{aligned}0 &= C^2\left\{\Delta v + 2i\alpha\partial_x v + (\gamma^v)^2 v\right\}\\ &= C^2\left\{\partial_x[\partial_x v] + \partial_z[\partial_z v] + +2i\alpha\partial_x v + (\gamma^v)^2 v\right\}\\ &= C\partial_x[C\partial_x v] - C(\partial_x C)\partial_x v + C\partial_z[C\partial_z v] + 2C^2 i\alpha\partial_x v + C^2(\gamma^v)^2 v.\end{aligned}$$

By our transformation rules

$$\begin{aligned}0 &= [C\partial_{x'} - E\partial_{z'}][C\partial_{x'}v' - E\partial_{z'}v'] - (\partial_{x'}C)[C\partial_{x'}v' - E\partial_{z'}v'] + \partial_{z'}[\partial_{z'}v'] + 2C^2 i\alpha\partial_{x'}v'\\ &\quad + C^2(\gamma^{v'})^2 v'\\ &= C\partial_{x'}[C\partial_{x'}v'] - E\partial_{z'}[C\partial_{x'}v'] - C\partial_{x'}[E\partial_{z'}v'] + E\partial_{z'}[E\partial_{z'}v'] - (\partial_{x'}C)C\partial_{x'}v'\\ &\quad + (\partial_{x'}C)E\partial_{z'}v' + \partial_{z'}^2 v' + 2C^2 i\alpha\partial_{x'}v' + C^2(\gamma^{v'})^2 v'\\ &= \partial_{x'}[C^2\partial_{x'}v'] - (\partial_{x'}C)C\partial_{x'}v' - \partial_{z'}[EC\partial_{x'}v'] + (\partial_{z'}E)C\partial_{x'}v' - \partial_{x'}[CE\partial_{z'}v']\\ &\quad + (\partial_{x'}C)E\partial_{z'}v' + \partial_{z'}[E^2\partial_{z'}v'] - (\partial_{z'}E)E\partial_{z'}v' - (\partial_{x'}C)C\partial_{x'}v' + (\partial_{x'}C)E\partial_{z'}v'\\ &\quad + \partial_{z'}^2 v' + 2C^2 i\alpha\partial_{x'}v' + C^2(\gamma^{v'})^2 v',\end{aligned}$$

where

$$(\partial_{x'}C)E\partial_{z'}v' - (\partial_{z'}E)E\partial_{z'}v' - (\partial_{x'}C)C\partial_{x'}v' + (\partial_{x'}E)C\partial_{z'}v' = 0,$$

because

$$\partial_{z'}E = \partial_{x'}C = \partial_x C.$$

The second, forth, eight, and tenth terms cancel so that

$$\begin{aligned}0 &= \partial_{x'}[C^2\partial_{x'}v'] - \partial_{z'}[EC\partial_{x'}v'] - \partial_{x'}[CE\partial_{z'}v'] + (\partial_{x'}C)E\partial_{z'}v'\\ &+\partial_{z'}[E^2\partial_{z'}v'] - (\partial_{x'}C)C\partial_{x'}v' + \partial_{z'}^2 v' + 2C^2 i\alpha\partial_{x'}v' + C^2(\gamma^{v'})^2 v'.\end{aligned}$$

This may be written more compactly as

$$0 = \mathrm{div}'[A\nabla' v'] + B\cdot\nabla' v' + 2C^2 i\alpha\partial_{x'}v' + C^2(\gamma^{v'})^2 v',$$

where for $S = C^2$

$$A = \begin{pmatrix} S & -EC \\ -EC & 1+E^2\end{pmatrix}, \qquad B = (\partial_{x'}C)\begin{pmatrix}-C\\ E\end{pmatrix}.$$

By the definitions of C and E, (C.3) and (C.4), we have

$$S = 1 + \frac{2}{\overline{h}}u - \frac{2}{\overline{h}}\ell + \frac{1}{\overline{h}^2}u^2 + \frac{1}{\overline{h}^2}\ell^2 - \frac{2}{\overline{h}^2}\ell u,$$

$$CE = Z_L(\partial_x u) + Z_U(\partial_x \ell) + \frac{Z_L}{\overline{h}} u(\partial_x u) - \frac{Z_U}{\overline{h}} \ell(\partial_x \ell) - \frac{Z_L}{\overline{h}} \ell(\partial_x u) + \frac{Z_U}{\overline{h}} u(\partial_x \ell),$$
$$E^2 = Z_L^2(\partial_x u)^2 + Z_U^2(\partial_x \ell)^2 + 2Z_L Z_U(\partial_x \ell)(\partial_x u).$$

If $\ell = \delta\tilde{\ell}$ and $u = \varepsilon\tilde{u}$ then

$$A = A(\delta, \varepsilon) = A_{0,0} + A_{1,0}\delta + A_{0,1}\varepsilon + +A_{2,0}\delta^2 + A_{0,2}\varepsilon^2 + A_{1,1}\delta\varepsilon,$$
$$B = B(\delta, \varepsilon) = B_{1,0}\delta + B_{0,1}\varepsilon + +B_{2,0}\delta^2 + B_{0,2}\varepsilon^2 + B_{1,1}\delta\varepsilon,$$
$$S = S(\delta, \varepsilon) = S_{0,0} + S_{1,0}\delta + S_{0,1}\varepsilon + +S_{2,0}\delta^2 + S_{0,2}\varepsilon^2 + S_{1,1}\delta\varepsilon,$$

where

$$A_{0,0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_{1,0} = \frac{1}{\overline{h}} \begin{pmatrix} -2\tilde{\ell} & -\overline{h} Z_U(\partial_x \tilde{\ell}) \\ -\overline{h} Z_U(\partial_x \tilde{\ell}) & 0 \end{pmatrix},$$
$$A_{0,1} = \frac{1}{\overline{h}} \begin{pmatrix} -2\tilde{u} & -\overline{h} Z_L(\partial_x \tilde{u}) \\ -\overline{h} Z_L(\partial_x \tilde{u}) & 0 \end{pmatrix},$$
$$A_{2,0} = \frac{1}{\overline{h}^2} \begin{pmatrix} \tilde{\ell}^2 & \overline{h} Z_U \tilde{\ell}(\partial_x \tilde{\ell}) \\ \overline{h} Z_U \tilde{\ell}(\partial_x \tilde{\ell}) & \overline{h}^2 Z_U^2(\partial_x \tilde{\ell})^2 \end{pmatrix},$$
$$A_{0,2} = \frac{1}{\overline{h}^2} \begin{pmatrix} \tilde{u}^2 & -\overline{h} Z_L \tilde{u}(\partial_x \tilde{u}) \\ -\overline{h} Z_L \tilde{u}(\partial_x \tilde{u}) & \overline{h}^2 Z_L^2(\partial_x \tilde{u})^2 \end{pmatrix},$$
$$A_{1,1} = \frac{1}{\overline{h}^2} \begin{pmatrix} -2\tilde{\ell}\tilde{u} & \overline{h}\left(Z_L \tilde{\ell}(\partial_x \tilde{u}) - Z_U \tilde{u}(\partial_x \tilde{\ell})\right) \\ \overline{h}\left(Z_L \tilde{\ell}(\partial_x \tilde{u}) - Z_U \tilde{u}(\partial_x \tilde{\ell})\right) & 2\overline{h}^2 Z_U Z_L(\partial_x \tilde{\ell})(\partial_x \tilde{u}) \end{pmatrix},$$

and

$$B_{1,0} = \frac{1}{\overline{h}} \begin{pmatrix} (\partial_x \tilde{\ell}) \\ 0 \end{pmatrix}, \quad B_{0,1} = \frac{1}{\overline{h}} \begin{pmatrix} -(\partial_x \tilde{u}) \\ 0 \end{pmatrix},$$
$$B_{2,0} = \frac{1}{\overline{h}^2} \begin{pmatrix} -\tilde{\ell}(\partial_x \tilde{\ell}) \\ -\overline{h} Z_U(\partial_x \tilde{\ell})^2 \end{pmatrix},$$
$$B_{0,2} = \frac{1}{\overline{h}^2} \begin{pmatrix} -\tilde{u}(\partial_x \tilde{u}) \\ \overline{h} Z_L(\partial_x \tilde{u})^2 \end{pmatrix},$$
$$B_{1,1} = \frac{1}{\overline{h}^2} \begin{pmatrix} \tilde{u}(\partial_x \tilde{\ell}) + \tilde{\ell}(\partial_x \tilde{u}) \\ \overline{h}(Z_U - Z_L)(\partial_x \tilde{\ell})(\partial_x \tilde{u}) \end{pmatrix},$$

and

$$S_{0,0} = 1, \quad S_{1,0} = -\frac{2}{\overline{h}} \tilde{\ell}, \quad S_{0,1} = \frac{2}{\overline{h}} \tilde{u},$$
$$S_{2,0} = \frac{1}{\overline{h}^2} \tilde{\ell}^2, \quad S_{1,0} = \frac{1}{\overline{h}^2} \tilde{u}^2, \quad S_{0,1} = -\frac{2}{\overline{h}^2} \tilde{\ell}\tilde{u}.$$

Appendix D

PERMISSIONS FOR THE INCLUSION OF PUBLISHED WORKS

The proof of joint analyticity and computation of the Reflectivity Map, including the related algorithms and numerical experiments, are submitted to two separate SIAM journals, which allows authors to use their articles in their thesis. Their policy states "Figures or tables created by someone other than the author or borrowed from a previously published source, even those created by the author and published elsewhere, must carry an appropriate credit line at the end of the caption." The full policy is available at https://epubs.siam.org/journal-authors. Upon acceptance, the author will update the necessary chapters and give the appropriate credit to the respective journal.

CITED LITERATURE

1. Varadan, V. K.: Low and High Frequency Asymptotics: Acoustic, Electromagnetic and Elastic Wave Scattering. Elsevier, 2013.

2. Rayleigh, L.: X. on the electromagnetic theory of light. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 12(73):81–101, 1881.

3. Hulst, H. C. and van de Hulst, H. C.: Light scattering by small particles. Courier Corporation, 1981.

4. Twersky, V.: Rayleigh scattering. Applied Optics, 3(10):1150–1162, 1964.

5. Kerker, M.: The scattering of light and other electromagnetic radiation academic. 1969.

6. Petit, R.: Electromagnetic theory of gratings. Berlin, Springer-Verlag, 1980.

7. Wilcox, C. H.: Scattering Theory for Diffraction Gratings. Berlin, Springer, 1984.

8. Arens, T.: Scattering by Biperiodic Layered Media: The Integral Equation Approach. Habilitationsschrift, Karlsruhe Institute of Technology, 2009.

9. Choudhury, A. K. R.: Principles of colour and appearance measurement: Visual measurement of colour, colour comparison and management. Woodhead Publishing, 2014.

10. Rayleigh, J. W. S. B.: On the scattering of light by small particles. 1871.

11. Lord, R.: On the light from the sky, its polarization and colour. Phil Mag, 41:274, 1871.

12. Foldy, L. L.: The multiple scattering of waves. i. general theory of isotropic scattering by randomly distributed scatterers. Physical review, 67(3-4):107, 1945.

13. Lax, M.: Multiple scattering of waves. Reviews of Modern Physics, 23(4):287, 1951.

14. Fikioris, J. and Waterman, P.: Multiple scattering of waves. ii. "hole corrections" in the scalar case. Journal of Mathematical Physics, 5(10):1413–1420, 1964.

15. Waterman, P. C. and Truell, R.: Multiple scattering of waves. Journal of mathematical physics, 2(4):512–537, 1961.

16. Twersky, V.: Multiple scattering of sound by a periodic line of obstacles. The Journal of the Acoustical Society of America, 53(1):96–112, 1973.

17. Varadan, V. K., Varadan, V. V., and Pao, Y.-H.: Multiple scattering of elastic waves by cylinders of arbitrary cross section. i. sh waves. The Journal of the Acoustical Society of America, 63(5):1310–1319, 1978.

18. Sheng, P. and van Tiggelen, B.: Introduction to wave scattering, localization and mesoscopic phenomena., 2007.

19. Tsang, L. and Kong, J. A.: Scattering of electromagnetic waves: advanced topics, volume 26. John Wiley & Sons, 2004.

20. Twersky, V.: Multiple scattering of radiation by an arbitrary configuration of parallel cylinders. The Journal of the Acoustical Society of America, 24(1):42–46, 1952.

21. Twersky, V.: Multiple scattering of radiation by an arbitrary planar configuration of parallel cylinders and by two parallel cylinders. Journal of Applied Physics, 23(4):407–414, 1952.

22. Twersky, V.: On a multiple scattering theory of the finite grating and the wood anomalies. Journal of Applied Physics, 23(10):1099–1118, 1952.

23. Kavaklıoğlu, Ö. and Lang, R. H.: Exact matrix representation of the transverse magnetic multiple scattering of obliquely incident plane waves by the diffraction grating of penetrable cylinders. Journal of Applied Mathematics, 2012, 2012.

24. Twersky, V.: On the scatttering of waves by an infinite grating. IRE Transactions on Antennas and Propagation, 4(3):330–345, 1956.

25. Twersky, V.: Elementary function representations of schlömilch series. Archive for Rational Mechanics and Analysis, 8(1):323–332, 1961.

26. Brown, G. S.: Coherent wave propagation through a sparse concentration of particles. Radio Science, 15(3):705–710, 1980.

27. Brown, G. S.: An alternate approach to coherent wave propagation through sparsely populated media. In In: Multiple scattering and waves in random media; Proceedings of the Workshop, pages 77–87, 1981.

28. Keller, J. B.: Stochastic equations and wave propagation in. Stochastic processes in mathematical physics and engineering, 16:145, 1964.

29. De Nicola, S.: Stochastic description of wave propagation in random media. In Photon Propagation in Tissues III, volume 3194, pages 453–461. International Society for Optics and Photonics, 1998.

30. Frisch, U.: Wave Propagation in Random Media: A Theory of Multiple Scattering. 1965.

31. Varadan, V., Bringi, V., and Varadan, V.: Coherent electromagnetic wave propagation through randomly distributed dielectric scatterers. Physical Review D, 19(8):2480, 1979.

32. Varadan, V., Ma, Y., and Varadan, V.: Coherent electromagnetic wave propagation through randomly distributed and oriented pair-correlated dielectric scatterers. Radio science, 19(06):1445–1449, 1984.

33. Tsang, L. and Kong, J.: Multiple scattering of electromagnetic waves by random distributions of discrete scatterers with coherent potential and quantum mechanical formalism. Journal of Applied Physics, 51(7):3465–3485, 1980.

34. Tsang, L., Kong, J., and Habashy, T.: Multiple scattering of acoustic waves by random distribution of discrete spherical scatterers with the quasicrystalline and percus–yevick approximation. The Journal of the Acoustical Society of America, 71(3):552–558, 1982.

35. Virieux, J. and Operto, S.: An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6):WCC1–WCC26, 2009.

36. Bleibinhaus, F. and Rondenay, S.: Effects of surface scattering in full-waveform inversion. Geophysics, 74(6):WCC69–WCC77, 2009.

37. Natterer, F. and Wübbeling, F.: Mathematical methods in image reconstruction. SIAM Monographs on Mathematical Modeling and Computation. Philadelphia, PA, Society for Industrial and Applied Mathematics (SIAM), 2001.

38. ed. C. Godrèche Solids far from equilibrium. Cambridge, Cambridge University Press, 1992.

39. Raether, H.: Surface plasmons on smooth and rough surfaces and on gratings. Berlin, Springer, 1988.

40. Maier, S. A.: Plasmonics: Fundamentals and Applications. New York, Springer, 2007.

41. Enoch, S. and Bonod, N.: Plasmonics: From Basics to Advanced Topics. Springer Series in Optical Sciences. New York, Springer, 2012.

42. Brekhovskikh, L. M. and Lysanov, Y. P.: Fundamentals of Ocean Acoustics. Berlin, Springer-Verlag, 1982.

43. Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T., and Wolff, P. A.: Extraordinary optical transmission through sub-wavelength hole arrays. Nature, 391(6668):667–669, 1998.

44. Moskovits, M.: Surface–enhanced spectroscopy. Reviews of Modern Physics, 57(3):783–826, 1985.

45. Homola, J.: Surface plasmon resonance sensors for detection of chemical and biological species. Chemical Reviews, 108(2):462–493, 2008.

46. Im, H., Lee, S. H., Wittenberg, N. J., Johnson, T. W., Lindquist, N. C., Nagpal, P., Norris, D. J., and Oh, S.-H.: Template-stripped smooth Ag nanohole arrays with silica shells for surface plasmon resonance biosensing. ACS Nano, 5:6244–6253, 2011.

47. Lindquist, N. C., Johnson, T. W., Jose, J., Otto, L. M., and Oh, S.-H.: Ultrasmooth metallic films with buried nanostructures for backside reflection-mode plasmonic biosensing. Annalen der Physik, 524:687–696, 2012.

48. Jose, J., Jordan, L. R., Johnson, T. W., Lee, S. H., Wittenberg, N. J., and Oh, S.-H.: Topographically flat substrates with embedded nanoplasmonic devices for biosensing. Adv Funct Mater, 23:2812–2820, 2013.

49. Reitich, F., Johnson, T. W., Oh, S.-H., and Meyer, G.: A fast and high-order accurate boundary perturbation method for characterization and design in nanoplasmonics. Journal of the Optical Society of America, A, 30:2175–2187, 2013.

50. Nicholls, D. P., Reitich, F., Johnson, T. W., and Oh, S.-H.: Fast high–order perturbation of surfaces (HOPS) methods for simulation of multi–layer plasmonic devices and metamaterials. Journal of the Optical Society of America, A, 31(8):1820–1831, 2014.

51. Nicholls, D. P.: A method of field expansions for vector electromagnetic scattering by layered periodic crossed gratings. Journal of the Optical Society of America, A, 32(5):701–709, 2015.

52. Nicholls, D. P. and Tammali, V.: A high–order perturbation of surfaces (HOPS) approach to Fokas integral equations: Vector electromagnetic scattering by periodic crossed gratings. Applied Numerical Methods, 101:1–17, 2016.

53. Nicholls, D. P., Oh, S.-H., Johnson, T. W., and Reitich, F.: Launching surface plasmon waves via vanishingly small periodic gratings. Journal of the Optical Society of America, A, 33(3):276–285, 2016.

54. Ambrose, D. and Nicholls, D. P.: Fokas integral equations for three dimensional layered–media scattering. Journal of Computational Physics, 276:1–25, 2014.

55. Nicholls, D. P.: A high-order perturbation of surfaces (hops) approach to fokas integral equations: three-dimensional layered media scattering. Appl. Math, 74(1):61–87, 2016.

56. Colton, D. and Kress, R.: Inverse acoustic and electromagnetic scattering theory, volume 93 of Applied Mathematical Sciences. Springer, New York, third edition, 2013.

57. Greengard, L. and Rokhlin, V.: A fast algorithm for particle simulations. J. Comput. Phys., 73(2):325–348, 1987.

58. Barnett, A. and Greengard, L.: A new integral representation for quasi-periodic scattering problems in two dimensions. BIT Numerical Mathematics, 51:67–90, 2011.

59. Cho, M. H. and Barnett, A.: Robust fast direct integral equation solver for quasi–periodic scattering problems with a large number of layers. Optics Express, 23(2):1775–1799, 2015.

60. Lai, J., Kobayashi, M., and Barnett, A.: A fast and robust solver for the scattering from a layered periodic structure containing multi-particle inclusions. J. Comput. Phys., 298:194–208, 2015.

61. Reitich, F. and Tamma, K.: State–of–the–art, trends, and directions in computational electromagnetics. CMES Comput. Model. Eng. Sci., 5(4):287–294, 2004.

62. Rayleigh, L.: On the dynamical theory of gratings. Proc. Roy. Soc. London, A79:399–416, 1907.

63. Rice, S. O.: Reflection of electromagnetic waves from slightly rough surfaces. Comm. Pure Appl. Math., 4:351–378, 1951.

64. Bruno, O. and Reitich, F.: Numerical solution of diffraction problems: A method of variation of boundaries. J. Opt. Soc. Am. A, 10(6):1168–1175, 1993.

65. Bruno, O. and Reitich, F.: 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.

66. Bruno, O. and Reitich, F.: Numerical solution of diffraction problems: A method of variation of boundaries. III. Doubly periodic gratings. J. Opt. Soc. Am. A, 10(12):2551–2562, 1993.

67. Nicholls, D. P. and Reitich, F.: Shape deformations in rough surface scattering: Cancellations, conditioning, and convergence. J. Opt. Soc. Am. A, 21(4):590–605, 2004.

68. Nicholls, D. P. and Reitich, F.: Shape deformations in rough surface scattering: Improved algorithms. J. Opt. Soc. Am. A, 21(4):606–621, 2004.

69. Nicholls, D. P. and Reitich, F.: Boundary perturbation methods for high–frequency acoustic scattering: Shallow periodic gratings. J. Acoust. Soc. Amer., 123(5):2531–2541, 2008.

70. Malcolm, A. and Nicholls, D. P.: A field expansions method for scattering by periodic multilayered media. Journal of the Acoustical Society of America, 129(4):1783–1793, 2011.

71. Pillage, L. and Rohrer, R.: Asymptotic waveform evaluation for timing analysis. IEEE Transactions on Computer-Aided Design, 9(4):352–366, 1990.

72. Kolbehdar, M., Srinivasan, M., Nakhla, M., Q.-J., Z., and Achar, R.: Simultaneous time and frequency domain solutions of em problems using finite element and cfh techniques. IEEE Transactions on Microwave Theory and Techniques, 44(9):1526–1534, 1996.

73. Reddy, C. J., Deshpande, M. D., Cockrell, C. R., and Beck, F. B.: Fast rcs computation over a frequency band using method of moments in conjunction with asymptotic waveform evaluation technique. IEEE Transactions on Antennas and Propagation, 46(8):1229–1233, 1998.

74. Sloane, R., Lee, R., and Lee, J.-F.: Multipoint galerkin asymptotic waveform evaluation for model order reduction of frequency domain fem electromagnetic radiation problems. IEEE Transactions on Antennas and Propagation, 49(10):1504–1513, 2001.

75. Nicholls, D. P.: Numerical solution of diffraction problems: A high–order perturbation of surfaces/asymptotic waveform evaluation method. SIAM Journal on Numerical Analysis, 55(1):144–167, 2017.

76. Nicholls, D. P.: Three–dimensional acoustic scattering by layered media: A novel surface formulation with operator expansions implementation. Proceedings of the Royal Society of London, A, 468:731–758, 2012.

77. Nicholls, D. P.: Numerical simulation of grating structures incorporating two–dimensional materials: A high–order perturbation of surfaces framework. SIAM Journal on Applied Mathematics, 78(1):19–44, 2018.

78. Sanghera, P.: Quantum physics for scientists and technologists: Fundamental principles and applications for biologists, chemists, computer scientists, and nanotechnologists. John Wiley & Sons, 2011.

79. Snyder, A. W. and Love, J.: Optical waveguide theory. Springer Science & Business Media, 2012.

80. Okamoto, K.: Fundamentals of optical waveguides. Elsevier, 2021.

81. Jackson, J. D.: Classical electrodynamics. New York, John Wiley & Sons Inc., second edition, 1975.

82. Halliday, D., Resnick, R., and Walker, J.: Fundamentals of physics. John Wiley & Sons, 2013.

83. Billingham, J. and King, A. C.: Wave motion. Cambridge Texts in Applied Mathematics. Cambridge, Cambridge University Press, 2000.

84. Nicholls, D. P.: High–Order Perturbation of Surfaces Methods. 2022.

85. Maier, S. A. et al.: Plasmonics: fundamentals and applications, volume 1. Springer, 2007.

86. Gottlieb, D. and Orszag, S. A.: Numerical analysis of spectral methods: theory and applications. Philadelphia, Pa., Society for Industrial and Applied Mathematics, 1977. CBMS-NSF Regional Conference Series in Applied Mathematics, No. 26.

87. Canuto, C., Hussaini, M. Y., Quarteroni, A., and Zang, T. A.: Spectral methods in fluid dynamics. New York, Springer-Verlag, 1988.

88. Deville, M. O., Fischer, P. F., and Mund, E. H.: High-order methods for incompressible fluid flow, volume 9 of Cambridge Monographs on Applied and Computational Mathematics. Cambridge, Cambridge University Press, 2002.

89. Boyd, J. P.: Chebyshev and Fourier spectral methods. Mineola, NY, Dover Publications Inc., second edition, 2001.

90. Shen, J. and Tang, T.: Spectral and high-order methods with applications, volume 3 of Mathematics Monograph Series. Science Press Beijing, Beijing, 2006.

91. Shen, J., Tang, T., and Wang, L.-L.: Spectral methods, volume 41 of Springer Series in Computational Mathematics. Springer, Heidelberg, 2011. Algorithms, analysis and applications.

92. Baker, Jr., G. A. and Graves-Morris, P.: Padé approximants. Cambridge, Cambridge University Press, second edition, 1996.

93. Nicholls, D. P. and Reitich, F.: Analytic continuation of Dirichlet-Neumann operators. Numer. Math., 94(1):107–146, 2003.

94. Bender, C. M. and Orszag, S. A.: Advanced mathematical methods for scientists and engineers. New York, McGraw-Hill Book Co., 1978. International Series in Pure and Applied Mathematics.

95. Lions, P.-L.: On the Schwarz alternating method. III. A variant for nonoverlapping subdomains. In Third International Symposium on Domain Decomposition Methods for Partial Differential Equations (Houston, TX, 1989), pages 202–223. SIAM, Philadelphia, PA, 1990.

96. Després, B.: Méthodes de décomposition de domaine pour les problèmes de propagation d'ondes en régime harmonique. Le théorème de Borg pour l'équation de Hill vectorielle. Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, 1991. Thèse, Université de Paris IX (Dauphine), Paris, 1991.

97. Després, B.: Domain decomposition method and the Helmholtz problem. In Mathematical and numerical aspects of wave propagation phenomena (Strasbourg, 1991), pages 44–52. SIAM, Philadelphia, PA, 1991.

98. Nicholls, D. P.: On analyticity of linear waves scattered by a layered medium. Journal of Differential Equations, 263(8):5042–5089, 2017.

99. Krantz, S. G. and Parks, H. R.: A primer of real analytic functions. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston, Inc., Boston, MA, second edition, 2002.

100. Roache, P. J.: The method of manufactured solutions for code verification. In Computer Simulation Validation, pages 295–318. Springer, 2019.

101. Salari, K. and Knupp, P.: Code verification by the method of manufactured solutions. Technical report, Sandia National Labs., Albuquerque, NM (US); Sandia National Labs ..., 2000.

102. Burggraf, O. R.: Analytical and numerical studies of the structure of steady separated flows. J. Fluid Mech., 24:113–151, 1966.

103. Roache, P. J.: Code verification by the method of manufactured solutions. J. Fluids Eng., 124(1):4–10, 2002.

104. Roy, C. J.: Review of code and solution verification procedures for computational simulation. J. Comp. Phys., 205(1):131–156, 2005.

105. Pourahmadi, M.: Taylor expansion of $exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications. Amer. Math. Monthly, 91(5):303–307, 1984.

106. Roberts, A. J.: Highly nonlinear short-crested water waves. J. Fluid Mech., 135:301–321, 1983.

107. Marchant, T. R. and Roberts, A. J.: Properties of short-crested waves in water of finite depth. J. Austral. Math. Soc. Ser. B, 29(1):103–125, 1987.

108. Johnson, P. and Christy, R.: Optical constants of the noble metals. Physical Review B, 6:4370, 1972.

109. Ordal, M. A., Bell, R. J., Alexander, R. W., Newquist, L. A., and Querry, M. R.: Optical properties of al, fe, ti, ta, w, and mo at submillimeter wavelengths. Appl. Opt., 27(6):1203–1209, Mar 1988.

110. Das, S., Bhar, G. C., Gangopadhyay, S., and Ghosh, C.: Linear and nonlinear optical properties of zngep 2 crystal for infrared laser device applications: revisited. Applied optics, 42(21):4335–4340, 2003.

111. Bond, W.: Measurement of the refractive indices of several crystals. Journal of Applied Physics, 36(5):1674–1677, 1965.

112. Nicholls, D. P. and Reitich, F.: Stability of high-order perturbative methods for the computation of Dirichlet-Neumann operators. J. Comput. Phys., 170(1):276–298, 2001.

113. Johnson, P. and Christy, R.: Optical constants of transition metals: Ti, v, cr, mn, fe, co, ni, and pd. Physical review B, 9(12):5056, 1974.

114. Nicholls, D. P. and Reitich, F.: A new approach to analyticity of Dirichlet-Neumann operators. Proc. Roy. Soc. Edinburgh Sect. A, 131(6):1411–1433, 2001.

115. Hu, B. and Nicholls, D. P.: Analyticity of dirichlet–neumann operators on hölder and lipschitz domains. SIAM journal on mathematical analysis, 37(1):302–320, 2005.

116. Nicholls, D. P. and Taber, M.: Joint analyticity and analytic continuation for Dirichlet–Neumann operators on doubly perturbed domains. J. Math. Fluid Mech., 10(2):238–271, 2008.

117. Nicholls, D. P. and Shen, J.: A rigorous numerical analysis of the transformed field expansion method. SIAM Journal on Numerical Analysis, 47(4):2708–2734, 2009.

118. Mashayekh, H., Kallivokas, L. F., and Tassoulas, J. L.: Parameter estimation in layered media using dispersion-constrained inversion. Journal of Engineering Mechanics, 144(11):04018099, 2018.

119. Mashayekh, H. et al.: Parameter estimation in layered media using dispersion-constrained full waveform inversion. Doctoral dissertation, 2018.

120. Basu, S., Pollack, R., and Coste-Roy, M.: Algorithms in Real Algebraic Geometry. Algorithms and Computation in Mathematics. Springer Berlin Heidelberg, 2007.

121. Suslov, I.: Divergent perturbation series. Journal of Experimental and Theoretical Physics, 100(6):1188–1233, 2005.

122. Costin, O. and Dunne, G.: Convergence from divergence. Journal of Physics A: Mathematical and Theoretical, 51, 05 2017.

123. Heinz, M.: New resummation techniques of divergent series: the painlevé equation pii, 2020.

124. Dienes, P.: The Taylor series: an introduction to the theory of functions of a complex variable. Dover New York, 1957.

125. Arteca, G. A., Fernández, F. M., and Castro, E. A.: Summation of strongly divergent perturbation series. Journal of mathematical physics, 25(12):3492–3496, 1984.

126. Evans, L. C.: Partial differential equations. Providence, RI, American Mathematical Society, second edition, 2010.

127. Qiu, Y.: High-frequency modeling and analyses for buck and multiphase buck converters. Doctoral dissertation, Virginia Polytechnic Institute and State University, 2005.

128. Bosse, E., Turner, R. M., and Riseborough, E. S.: Model-based multifrequency array signal processing for low-angle tracking. IEEE Transactions on Aerospace and electronic systems, 31(1):194–210, 1995.

129. Zhao, S., Wang, F., Xu, H., and Zhu, J.: Multi-frequency identification method in signal processing. Digital Signal Processing, 19(4):555–566, 2009.

130. Blanchard-Wrigglesworth, E., Donohoe, A., Roach, L. A., DuVivier, A., and Bitz, C. M.: High-frequency sea ice variability in observations and models. Geophysical Research Letters, 48(14):e2020GL092356, 2021.

131. Imperatore, P., Iodice, A., and Riccio, D.: Perturbation theory for scattering from multilayers with randomly rough fractal interfaces: remote sensing applications. Sensors, 18(1):54, 2017.

132. Fang, Z.: Operator Expansions for Linear Waves: Parallel Implementation and Multilayer Inversion. Doctoral dissertation, University of Illinois at Chicago, 2015.

133. Achenbach, J.: Wave propagation in elastic solids. Elsevier, 2012.

134. Chandra, R., Dagum, L., Kohr, D., Menon, R., Maydan, D., and McDonald, J.: Parallel programming in OpenMP. Morgan kaufmann, 2001.

135. Snir, M., Gropp, W., Otto, S., Huss-Lederman, S., Dongarra, J., and Walker, D.: MPI–the Complete Reference: the MPI core, volume 1. MIT press, 1998.

136. Sanders, J. and Kandrot, E.: CUDA by example: an introduction to general-purpose GPU programming. Addison-Wesley Professional, 2010.

137. Escapil-Inchauspé, P. and Jerez-Hanckes, C.: Helmholtz scattering by random domains: first-order sparse boundary element approximation. SIAM Journal on Scientific Computing, 42(5):A2561–A2592, 2020.

138. Silva-Oelker, G., Aylwin, R., Jerez-Hanckes, C., and Fay, P.: Quantifying the impact of random surface perturbations on reflective gratings. IEEE Transactions on Antennas and Propagation, 66(2):838–847, 2017.

139. Nakata, Y. and Koshiba, M.: Boundary-element analysis of plane-wave diffraction from groove-type dielectric and metallic gratings. JOSA A, 7(8):1494–1502, 1990.

140. Elschner, J. and Hu, G.: An optimization method in inverse elastic scattering for one-dimensional grating profiles. Communications in Computational Physics, 12(5):1434–1460, 2012.

141. Rathsfeld, A., Schmidt, G., and Kleemann, B.: On a fast integral equation method for diffraction gratings. Communications in Computational Physics, 1:984–1009, 12 2006.

142. Escapil-Inchauspé, P. and Jerez-Hanckes, C.: Helmholtz scattering by random domains: first-order sparse boundary element approximation. SIAM Journal on Scientific Computing, 42(5):A2561–A2592, 2020.

143. Sethian, J. A.: Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science, volume 3. Cambridge university press, 1999.

144. El-Sayed, S. M. and Kaya, D.: Comparing numerical methods for helmholtz equation model problem. Applied Mathematics and Computation, 150(3):763–773, 2004.

145. Benamou, J.-D. and Després, B.: A domain decomposition method for the helmholtz equation and related optimal control problems. Journal of Computational Physics, 136(1):68–82, 1997.

146. Larsson, E.: A domain decomposition method for the helmholtz equation in a multilayer domain. SIAM Journal on Scientific Computing, 20(5):1713–1731, 1999.

147. Gong, S., Gander, M. J., Graham, I. G., Lafontaine, D., and Spence, E. A.: Convergence of parallel overlapping domain decomposition methods for the helmholtz equation. arXiv preprint arXiv:2106.05218, 2021.

148. Pérez-Arancibia, C., Shipman, S., Turc, C., and Venakides, S.: Domain decomposition for quasi-periodic scattering by layered media via robust boundary-integral equations at all frequencies. arXiv preprint arXiv:1801.09094, 2018.

149. Chan, T. F. and Mathew, T. P.: Domain decomposition algorithms. Acta numerica, 3:61–143, 1994.

150. Burden, R. and Faires, J. D.: Numerical analysis. Pacific Grove, CA, Brooks/Cole Publishing Co., sixth edition, 1997.

151. Atkinson, K. and Han, W.: Theoretical numerical analysis, volume 39 of Texts in Applied Mathematics. New York, Springer-Verlag, 2001. A functional analysis framework.

152. Bai, B. and Li, L.: Reduction of computation time for crossed-grating problems: a group-theoretic approach. JOSA A, 21(10):1886–1894, 2004.

153. Chew, W. C., Hu, B., Pan, Y., and Jiang, L.: Fast algorithms for layered media. Comptes Rendus Physique, 6(6):604–617, 2005.

154. Atkins, P. R. and Chew, W. C.: Fast computation of the dyadic green's function for layered media via interpolation. IEEE Antennas and Wireless Propagation Letters, 9:493–496, 2010.

155. Konno, K., Chen, Q., and Burkholder, R. J.: Fast computation of layered media green's function via recursive taylor expansion. IEEE Antennas and Wireless Propagation Letters, 16:1048–1051, 2016.

156. Hong, Y. and Nicholls, D. P.: A rigorous numerical analysis of the transformed field expansion method for diffraction by periodic, layered structures. SIAM Journal on Numerical Analysis, 59(1):456–476, 2021.

157. Phillips, N. A.: A coordinate system having some special advantages for numerical forecasting. Journal of the Atmospheric Sciences, 14(2):184–185, 1957.

158. Chandezon, J., Dupuis, M., Cornet, G., and Maystre, D.: Multicoated gratings: a differential formalism applicable in the entire optical region. J. Opt. Soc. Amer., 72(7):839, 1982.

159. Chandezon, J., Maystre, D., and Raoult, G.: A new theoretical method for diffraction gratings and its numerical application. J. Opt., 11(7):235–241, 1980.

VITA

NAME Matthew Shawn Kehoe

EDUCATION

Ph.D., Applied Mathematics, University of Illinois Chicago, IL, 2022

M.S., Computational Mathematics, University of Michigan at Dearborn, MI, 2015

B.A., Economics, Oakland University, MI, 2010

EXPERIENCE

Graduate Research and Teaching Assistant, University of Illinois Chicago, 2018 – 2022

NSF Mathematical Sciences Graduate Intern, Cold Regions Research and Engineering Laboratory (CRREL), Summer 2020

NSF Mathematical Sciences Graduate Intern, Argonne National Laboratory (ANL), Summer 2019

Software Consultant and Programmer, Workforce Software, 2010 – 2017

REU Observational Astronomy, CUREAU Program Physics, Summer 2013

Exchange Student, University of Otago, New Zealand, 2010

Web Developer, Oakland University, 2009 – 2010

TEACHING

MATH 180: Calculus 1 (4 semesters)
MCS 471: Numerical Analysis (2 semesters)
MATH 220: Differential Equations (1 semester)
MATH 419: Mathematical Biology (1 semester)
MATH 121: Precalculus (1 semester)

PUBLICATIONS

M. Kehoe and D. P. Nicholls,
"A Stable High-Order Perturbation of Surfaces/Asymptotic Waveform Evalutation Method for the Numerical Solution of Grating Scattering Problems", (submitted)

M. Kehoe and D. P. Nicholls,
"Joint Geometry/Frequency Analyticity of Fields Scattered by Periodic Layered Media", (submitted)