Spectral Optimization Problems Controlling Wave Phenomena

Braxton Osting

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Spectral Optimization Problems Controlling Wave Phenomena

Braxton Osting

Design problems seek a material arrangement or shape which fully harnesses the physical properties of the material(s) to create an environment in which a particular phenomena is most (or least) pronounced. Mathematically, design problems are formulated as PDE-constrained optimization problems to find the material arrangement that maximizes an objective function which expresses the desired behavior. The PDE constraint describes the relationship between the material and the phenomena of interest. The focus of this thesis is four design problems where the PDE constraint is a time-independent wave equation and the objective function governs some aspect of wave motion.

(1) We consider the shape optimization of functions of Dirichlet-Laplacian eigenvalues associated with the set of star-shaped, symmetric, bounded planar regions with smooth boundary. The boundary of such a region is represented using a Fourier-cosine series and the optimization problem is solved numerically using a quasi-Newton method. The method is applied to maximizing two particular nonsmooth functions of the eigenvalues: (a) the ratio of the n-th to first eigenvalues and (b) the ratio of the n-th eigenvalue gap to first eigenvalue. Both are generalizations of the Payne-Pólya-Weinberger ratio. The optimal values of these ratios and regions for which they are attained, for $n \leq 13$, are presented and interpreted as a study of the range of the Dirichlet-Laplacian eigenvalues. For both spectral functions and each n, the optimal region has multiplicity two n-th eigenvalue.

(2) We consider a system governed by the wave equation with index of refraction $n(x)$, taken to be variable within a bounded region $\Omega \subset \mathbb{R}^d$, and constant in $\mathbb{R}^d \backslash \Omega$. The solution of the time-dependent wave equation with initial data, which is localized in Ω, spreads and decays with advancing time; the spatially localized energy decays with time. This rate of decay can be measured in terms of the eigenvalues of the scattering resonance problem, a non-selfadjoint eigenvalue problem consisting of the time-harmonic wave (Helmholtz) equation with outgoing radiation condition at infinity. Specifically, the rate of energy escape from Ω is governed by the complex scattering eigenfrequency,

which is closest to the real axis. We study the structural design problem: Find a refractive index profile $n_\star$ within an admissible class which has a scattering frequency with minimal imaginary part. The admissible class is defined in terms of the compact support of $n(\mathbf{x}) - 1$ and pointwise upper and lower (material) bounds on $n(\mathbf{x})$, $0 < n_- \leq n(\mathbf{x}) \leq n_+ < \infty$. We formulate this problem as a constrained optimization problem and prove that an optimal structure, $n_\star(\mathbf{x})$ exists. Furthermore, $n_\star(\mathbf{x})$ is piecewise constant and achieves the material bounds, *i.e.* $n_\star(\mathbf{x}) \in \{n_-, n_+\}$. In one dimension, we establish a connection between $n_\star(x)$ and the well-known class of *Bragg structures*, where $n(x)$ is constant on intervals whose length is one-quarter of the effective wavelength.

(3) Consider a system governed by the time-dependent Schrödinger equation in its ground state. When subjected to weak parametric forcing by an "ionizing field" (time-varying), the state decays with advancing time due to coupling of the bound state to radiation modes. The decay-rate of this metastable state is governed by *Fermi's Golden Rule*, $\Gamma[V]$, which depends on the potential V and the details of the forcing. We pose the potential design problem: find V_{opt} which minimizes $\Gamma[V]$ (maximizes the lifetime of the state) over an admissible class of potentials with fixed spatial support. We formulate this problem as a constrained optimization problem and prove that an admissible optimal solution exists. Then, using quasi-Newton methods, we compute locally optimal potentials. These have the structure of a truncated periodic potential with a localized defect. In contrast to optimal structures for other spectral optimization problems, the optimizing potentials appear to be interior points of the constraint set and to be smooth. The multi-scale structures that emerge incorporate the physical mechanisms of energy confinement via material contrast and interference effects. An analysis of locally optimal potentials reveals local optimality is attained via two mechanisms: (i) decreasing the density of states near a resonant frequency in the continuum and (ii) tuning the oscillations of extended states to make $\Gamma[V]$, an oscillatory integral, small. Finally, we explore the performance of optimal potentials via simulations of the time-evolution.

(4) We consider a general class of two-dimensional passive propagation media, represented as a planar graph where nodes are capacitors connected to a common ground and edges are inductors. Capacitances and inductances are fixed in time but vary in space. Kirchhoff's laws give the time dynamics of voltage and current in the system. By harmonically forcing input nodes and collecting the resulting steady-state signal at output nodes, we obtain a linear, analog device that transforms the inputs to outputs. We pose the lattice synthesis problem: given a linear transformation, find the inductances and capacitances for an inductor-capacitor circuit that can perform this transfor-

mation. Formulating this as an optimization problem, we numerically demonstrate its solvability using gradient-based methods. By solving the lattice synthesis problem for various desired transformations, we design several devices that can be used for signal processing and filtering.

In addition to these spectral optimization problems, we study several problems on wave propagation, diffraction, and scattering. The focus is on the behavior of time-harmonic solutions to continuous and discrete wave equations.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, I'd like to thank the three advisors I had as a graduate student: Harish Bhat, who served as my advisor for my first two years and David Keyes and Michael Weinstein who were my co-advisors for my remaining four years. Their deep insight, good humor, and generosity have guided my research and positively influenced my life. Thank you.

I would also like to thank the other members of my committee, Cev Noyan and Chee Wei Wong, for their interest and assistance in the revision process of this thesis.

I was fortunate to study at the Advanced Computations Department (ACD) at the Stanford Linear Accelerator Center (SLAC) for two summer internships in 2008 and 2009. I'd like to thank Volkan Akçelik, Alex Candel, Andreas Kabel, Kwok Ko, Lie-Quan (Rich) Lee, and Cho Ng for their mentoring during this time.

I would like to acknowledge Andrew Ying, Sean Polvino, Cev Noyan, Conal Murray, Martin Holt, and Jörg Maser for collecting the data that is analyzed in Chapter 2. It was my pleasure to collaborate on a project with experimental data. I am also grateful to Prof. Noyan's group for the opportunity to visit the Advanced Photon Source at Argonne National Lab in July, 2009.

I would like to thank my undergraduate advisor, Nathan Kutz, for his encouragement to go to graduate school in applied mathematics and for his continued support.

I'd like to acknowledge valuable discussions with Guillaume Bal, Alex Barnett, David Bindel, Matias Courdurier, Percy Deift, Dirk Englund, Don Goldfarb, Tony Heinz, Bob Kohn, Chris Marianetti, Michael Overton, Lorenzo Polvani, Stephen Shipman, Chris Wiggins, and Maciej Zworski.

I have immense gratitude towards the organizations that funded this work. I was supported by the APAM department during my first year. During my second and third years, I was supported by an NSF Fellowship in the IGERT Joint Program in Applied Mathematics and Earth and Environmental Science at Columbia University. For the remaining three years, I was supported by US NSF Grant No. DMS06-02235, EMSW21- RTG: Numerical Mathematics for Scientific Computing.

I'd also like to generally thank the faculty, staff, and students who make APAM and Columbia University such an enriching atmosphere.

I would like to thank my parents for instilling in me a love of learning. I have always been able to count on them and my brother, Devin, for support and encouragement. Thank you.

Lastly, I'd like to thank Eileen, my wonderful wife and source of inspiration, to whom I'm deeply indebted.

to my parents,
Chris and Danielle Osting

Part I

Introduction

Chapter 1

Introduction

Many problems in the physical and engineering sciences involve finding a material arrangement or shape which fully harnesses the physical properties of the material(s), creating an environment in which a particular phenomena is most (or least) pronounced. In various contexts, these problems may be referred to as *design*, *synthesis*, or *control problems*. Examples abound and include finding

- the shape of an obstacle in fluid flow to achieve a desired effect *e.g.*, maximizing lift from a wing or minimizing drag on the hull of a ship [Pironneau, 1984],
- the design of a musical instrument to produce a desired sound [Kinsler *et al.*, 2000; Fletcher and Rossing, 1998],
- the shape of an antenna which has a desired far-field pattern performance measure such as directivity, gain, or signal-to-noise ratio [Angell and Kirsch, 2004],
- the shape of a column which has the least mass and will support a specified weight [Lewis and Overton, 1996],
- the design of a structure to be non-resonant to forcing, *e.g.*, preventing Tacoma Narrows-type aeroelastic flutter in bridges [Tisseur and Meerbergen, 2001], and
- the design of a material to guide, focus, or generally manipulate light, *e.g.*, an optical lens to focus light, a photomask for photolithographic printing, or a cavity to trap light for a maximal period of time [Joannopoulos *et al.*, 2008].

In each of these examples, the structure to be designed is described by a material coefficient or the shape of an object which, in turn, is related to a measurable quantity of interest through a physical law, typically expressed as a partial differential equation (PDE). Mathematically, these problems are formulated as PDE-constrained optimization problems. The ubiquity of such problems stimulates enormous interest and their size and complexity presents great challenge for the mathematical and computational science communities.

The main focus of this thesis is design problems in which the quantity of interest involves an aspect of wave motion, particularly, a spectral characteristic of wave motion, which is to say, a time-independent quantity which controls wave propagation. We refer to the design problems studied here as *spectral design problems* and, since the governing PDE constraints are spectral equations, their mathematical formulation as *spectral optimization problems*. This class of problems generalizes eigenvalue optimization problems to include spectral characteristics involving the continuous spectrum of operators and the associated solutions.

The spectral design problems studied here may be considered problems from *control theory*. However, control theory more often refers to the optimal choice of a set of (time-dependent or -independent) parameters which can be adjusted to control a finite-dimensional dynamical system [Lions, 1971]. Before proceeding, it may be helpful to distinguish design problems from inverse problems. Inverse problems are typically defined as the reconstruction of a physical system from (incomplete and noisy) data [Kirsch, 1996]. In fact, there is a large community devoted to the subject of inverse spectral problems, whose goal is to reconstruct a physical system from measured spectral data [Chu and Golub, 2005; Pöschel and Trubowitz, 1987]. Both inverse spectral problems and spectral design problems are formulated as spectral optimization problems. Typically, the objective function for an inverse spectral problem will be the mismatch between the measured data and that produced by a given model. However, the primary difference between these two types of problems is that the problems considered here are *not* data-driven and therefore the focus isn't on the error associated with measured data and the propagation of this error to the solution of the optimization problem. In this sense, it is sometimes convenient to consider design problems as a type of inverse problem with perfect data.

The spectral optimization problems considered in this thesis can be broadly cast into one of the following two forms:

Structural Optimization

$$\min_{d(x) \in \mathcal{A}_d} \quad J(\lambda_j, u_j)$$
$$\text{such that} \quad L(d)\, u_j = \lambda_j\, u_j \quad x \in \Omega$$

Shape Optimization

$$\min_{\Omega \in \mathcal{A}_\Omega} \quad J(\lambda_j, u_j)$$
$$\text{such that} \quad L(\Omega)\, u_j = \lambda_j\, u_j \quad x \in \Omega$$

In both cases, the *objective function* J to be minimized depends on one or more eigenpairs (λ_j, u_j) of the linear operator L. For the structural optimization problem, L depends on a spatially varying coefficient, $d(x)$, and for the shape optimization problem, L depends on the shape of the domain, Ω. The *design variables*, d and Ω, are constrained to the *admissible sets*, $\mathcal{A}_d$ and $\mathcal{A}_\Omega$, respectively. In both types of spectral optimization problems, there are equality constraints, which relate the design variables to the eigenpairs, and inequality constraints, which are encoded in the admissible set of the design variables. The admissible set is typically chosen to represent design constraints, such as point-wise constraints on a material in a structural optimization problem or a bound on the area of the domain for a shape optimization problem. A careful specification of the admissible set is crucial for obtaining a well-posed optimization problem. In the context of the optimization problem, the PDE constraint is sometimes referred to as a *forward problem*. In this thesis, we often suppress the PDE constraint when writing spectral optimization problems. This is, in part, notationally convenient, but also justified since we use reduced space methods in which the variation of the objective function J is computed assuming the quantities involved satisfy the underlying PDE constraint [Akcelik *et al.*, 2006].

Design problems are inherently interdisciplinary in nature, requiring

1. an understanding of the discipline from which the problem arises (here a physical or engineering science) so that the problem may be formulated as an optimization problem,

2. mathematical analysis to show that the optimization problem is well-posed and, perhaps, *a priori* information that may be obtained about the optimal solution,

3. sensitivity analysis to calculate how the objective function changes with respect to changes in the design variable(s),

4. scientific computational methods for the efficient computation of solutions to the underlying

PDE constraint,

5. optimization methods which utilize the available mathematical structure for the numerical solution of the optimization problem, and

6. mathematical analysis of the emergent optimal structure to describe why that particular design has small objective function value.

The dependency and interaction between these steps makes it difficult to discuss analytic methods or numerical algorithms for design problems in any generality; the solution to a design problem greatly depends on the mathematical structure of the underlying PDE, the constraint set, and the objective function.

In what follows, we outline our own small contribution to this rich and rapidly growing subject and then provide a brief historical background on the subject in more generality.

1.1 Thesis outline and statement of results

This thesis is structured into five parts comprising ten chapters, an extensive bibliography, and two appendixes. Essentially, Chapters 2-10 are each a peer-reviewed publication and could be read independently of one another, although we have made an attempt to connect chapters and provide cohesion throughout. Each chapter is further divided into sections and subsections for organization.

Part I, the current part, provides an introduction to this thesis and a brief background on spectral design problems. Part II presents a number of new results in wave propagation, diffraction, and scattering. The focus is on the behavior of time-harmonic solutions to continuous and discrete wave equations. Part III is the heart of the thesis and focuses on four spectral optimization problems controlling wave phenomena. References are given in Part IV. Finally, Part V consists of 2 appendixes which discuss a few properties of wave propagation in terms of spectral theory and the quasi-Newton optimization methods used in this thesis.

The objectives and contributions of each chapter may be summarized as follows. References are given for chapters where the intellectual content has been published elsewhere.

Part I, Introduction

Ch. 1 The current chapter presents a broad introduction to this thesis, including a historical context of spectral optimization problems.

Part II, Problems in Wave Propagation, Diffraction, and Scattering

Ch. 2 We use the Fresnel and Rayleigh-Sommerfeld diffraction formulas to model a scattering experiment where a coherent, monochromatic, x-ray beam is focused by a Fresnel zone plate onto a thin, perfect, single-crystal layer. The predictions of this model agree quite well with experimental data measured at the Center for Nanoscale Materials Nanoprobe instrument at Sector 26 of the Advanced Photon Source [Ying *et al.*, 2010].

Ch. 3 Beginning with Maxwell's equations in an inhomogeneous planar medium, we derive a finite volume method that we recognize as Kirchhoff's laws for a corresponding circuit consisting of inductors, capacitors, and resistors. This association automatically gives local charge and energy conservation. The method is implemented and used to find the steady-state solution for two test problems. By comparison with the exact solution for the homogeneous medium problem, the method is shown to be linearly convergent [Bhat and Osting, 2011b].

Ch. 4 We solve the thin-slit diffraction problem for the two-dimensional discrete wave equation. More precisely, for the discrete Helmholtz equation on the semi-infinite square lattice with data prescribed on the left boundary (the aperture), we use lattice Green's functions and a discrete Sommerfeld outgoing radiation condition to derive the exact solution everywhere in the lattice. The solution is a discrete convolution that can be evaluated in closed form for the wave number $k = 2$. For other wave numbers, we give an algorithm for computing the convolution kernel [Bhat and Osting, 2009a].

Ch. 5 The solution of the discrete wave equation developed in Ch. 4 is applied to study wave propagation in two-dimensional, transmission line model metamaterials (also referred to as composite and dual-composite right/left handed metamaterials). We find that for physically realizable parameters, wave propagation depends strongly on the temporal forcing frequency. [Bhat and Osting, 2008; Bhat and Osting, 2010].

Ch. 6 The discrete wave equation on a triangular lattice is analyzed by taking the quasi-continuum limit in the lattice spacing parameter $h > 0$. The resulting PDE is a wave equation with an $\mathcal{O}(h^2)$ isotropic, dispersive term which models the discreteness of the original equation. Using Green's function methods, dispersively-corrected, Rayleigh-Sommerfeld diffraction formulae are derived which are used to describe the qualitative effect of high-frequency wave propagation in discrete media [Osting and Bhat, 2008].

Part III, Spectral Optimization Problems Controlling Wave Phenomena

Ch. 7 We consider the shape optimization of functions of Dirichlet-Laplacian eigenvalues associated with the set of star-shaped, symmetric, bounded planar regions with smooth boundary. The boundary of such a region is represented using a Fourier-cosine series and the optimization problem is solved numerically using a quasi-Newton method. The method is applied to maximizing two particular nonsmooth functions of the eigenvalues: (a) the ratio of the n-th to first eigenvalues and (b) the ratio of the n-th eigenvalue gap to first eigenvalue. Both are generalizations of the Payne-Pólya-Weinberger ratio. The optimal values of these ratios and regions for which they are attained, for $n \leq 13$, are presented and interpreted as a study of the range of the Dirichlet-Laplacian eigenvalues. For both spectral functions and each n, the optimal region has multiplicity two n-th eigenvalue [Osting, 2010].

Ch. 8 We consider a system governed by the wave equation with index of refraction $n(x)$, taken to be variable within a bounded region $\Omega \subset \mathbb{R}^d$, and constant in $\mathbb{R}^d \setminus \Omega$. The solution of the time-dependent wave equation with initial data, which is localized in Ω, spreads and decays with advancing time; the spatially localized energy decays with time. This rate of decay can be measured in terms of the eigenvalues of the scattering resonance problem, a non-selfadjoint eigenvalue problem consisting of the time-harmonic wave (Helmholtz) equation with outgoing radiation condition at ∞. Specifically, the rate of energy escape from Ω is governed by the complex scattering eigenfrequency, which is closest to the real axis. We study the structural design problem: Find a refractive index profile $n_\star$ within an admissible class which has a scattering frequency with minimal imaginary part. The admissible class is defined in terms of the compact support of $n(\mathbf{x}) - 1$ and pointwise upper and lower (material) bounds on

$n(\mathbf{x})$, $0 < n_- \leq n(\mathbf{x}) \leq n_+ < \infty$. We formulate this problem as a constrained optimization problem and prove that an optimal structure, $n_\star(\mathbf{x})$ exists. Furthermore, $n_\star(\mathbf{x})$ is piecewise constant and achieves the material bounds, *i.e.* $n_\star(\mathbf{x}) \in \{ n_-, n_+ \}$. In one dimension, we establish a connection between $n_\star(x)$ and the well-known class of *Bragg structures*, where $n(x)$ is constant on intervals whose length is one-quarter of the effective wavelength [Osting and Weinstein, 2011b].

Ch. 9 Consider a system governed by the time-dependent Schrödinger equation in its ground state. When subjected to weak parametric forcing by an "ionizing field" (time-varying), the state decays with advancing time due to coupling of the bound state to radiation modes. The decay-rate of this metastable state is governed by *Fermi's Golden Rule*, $\Gamma[V]$, which depends on the potential V and the details of the forcing. We pose the potential design problem: find V_{opt} which minimizes $\Gamma[V]$ (maximizes the lifetime of the state) over an admissible class of potentials with fixed spatial support. We formulate this problem as a constrained optimization problem and prove that an admissible optimal solution exists. Then, using quasi-Newton methods, we compute locally optimal potentials. These have the structure of a truncated periodic potential with a localized defect. In contrast to optimal structures for other spectral optimization problems, such as the one studied in Ch. 8, the optimizing potentials appear to be interior points of the constraint set and to be smooth. The multi-scale structures that emerge incorporate the physical mechanisms of energy confinement via material contrast and interference effects. An analysis of locally optimal potentials reveals local optimality is attained via two mechanisms: (i) decreasing the density of states near a resonant frequency in the continuum and (ii) tuning the oscillations of extended states to make $\Gamma[V]$, an oscillatory integral, small. Finally, we explore the performance of optimal potentials via simulations of the time-evolution [Osting and Weinstein, 2011a].

Ch. 10 We consider a general class of two-dimensional passive propagation media, represented as a planar graph where nodes are capacitors connected to a common ground and edges are inductors. Capacitances and inductances are fixed in time but vary in space. Kirchhoff's laws give the time dynamics of voltage and current in the system. By harmonically forcing input nodes and collecting the resulting steady-state signal at output nodes, we obtain a

linear, analog device that transforms the inputs to outputs. We pose the lattice synthesis problem: given a linear transformation, find the inductances and capacitances for an inductor-capacitor circuit that can perform this transformation. Formulating this as an optimization problem, we numerically demonstrate its solvability using gradient-based methods. By solving the lattice synthesis problem for various desired transformations, we design several devices that can be used for signal processing and filtering [Bhat and Osting, 2011b].

Part IV, Bibliography

Part V, Appendices

App. A We review properties of the Schrödinger and wave equations using tools from spectral theory. In particular, we discuss wave propagation in a homogeneous media, diffraction, and scattering resonance expansions for inhomogeneous media.

App. B The BFGS method for approximating the Hessian of an objective function using gradient information is discussed.

1.2 Overview of spectral optimization problems

Although shape optimization has roots in ancient Greece and Carthage, PDE-constrained and, in particular, spectral optimization problems weren't studied until the calculus of variations was developed. And it wasn't until the last 50 years or so that we were able to compute non-trivial solutions to optimization problems and the precision in manufacturing processes progressed to a point where optimal designs could be built. With these developments, PDE-constrained and spectral optimization problems are now a prominent field in the mathematical, physical, and engineering communities. This brief overview is not intended to be comprehensive, but simply recognize a few of the works that have inspired the content of this thesis.

The isoperimetric problem. The oldest shape optimization problem is almost certainly the isoperimetric problem, stated

> Amongst all closed curves in the plane of fixed length, which curve encloses the largest region?

The answer to this question is, of course, the circle. A colorful account of the history of this problem and the generalization of it posed by Queen Dido, the first queen of Carthage, in ~ 814 BC can be found in [Kelvin, 1894] with more recent accounts given in [Osserman, 1978; Burago and Zalgaller, 1988; Ashbaugh and Benguria, 2010]. With regularity assumptions on the boundary, the isoperimetric problem was solved in 1744 by Leonhard Euler using the calculus of variations. It wasn't until the 1841 that Jakob Steiner and others relaxed these assumptions, using a geometric technique which became known as *Steiner symmetrization*. The existence proof for an optimal shape wasn't given until 1890 by Hermann Schwarz, whom some credit as giving the first complete solution of the isoperimetric problem.

In several places in this thesis, we use the terminology *isoperimetric* or *universal* inequality to be any inequality which relates two or more geometric and/or physical quantities associated with the same domain [Payne, 1967]. The solution to any shape optimization problem thus generates a sharp universal inequality, which for the isoperimetric problem can be stated: For any planar region, Ω,

$$|\Omega| \leq (4\pi)^{-1}|\partial\Omega|^2,$$

where $|\Omega|$ is the area of Ω and $|\partial\Omega|$ is the length of the boundary, $\partial\Omega$. Universal inequalities can be viewed as a-priori estimates for shape optimization problems.

Variational principles and the calculus of variations. In 1698, Johann Bernoulli posed the brachistochrone problem:

> Assuming constant gravity and no friction, find the path between two points which will be traversed by a body in least time.

The importance of this problem is apparent from Fermat's principle, which states that light travels along the path between two points which is traversed in least time. This question, and others posed soon thereafter such as finding the shape of a suspended rope, excited a fantastic competition amongst the greatest mathematicians of the day, resulting in the development of the calculus of variations [Gelfand and Fomin, 1991]. Contributors to this theory included, amongst others, Johann Bernoulli, Jakob Bernoulli, Leonhard Euler, Pierre de Fermat, Guillaume de l'Hôspital, Joseph-Louis Lagrange, Gottfried Leibniz, and Isaac Newton.

During this time, many *variational principles* were established which show the equivalence of solving an ordinary or partial differential equation with the minimization of a functional, typically an energy functional. We note that the problems which are considered in Part III of this thesis are not variational problems (although the underlying spectral constraints can be recast using a variational principle), yet the solution of both types of problems typically utilizes the calculus of variations.

Projectile shape minimizing air resistance. In 1687, Newton posed in *Philosophiæ Naturalis Principia Mathematica* the problem

> What is the shape of the projectile which minimizes air resistance (with constraints on the projectile's size)?

This is likely the first PDE-constrained shape optimization problem. The solution to this problem has been exceedingly elusive due to the fact that the optimal shape is neither radially symmetric nor smooth [Bucur and Buttazzo, 2005]. Today, the importance of this problem is paramount and the solution methods are used to design wings and propellers which generate maximal lift, wind turbines which harness maximal energy, and ship hulls which minimize drag [Pironneau, 1984].

Shape of the strongest column. The first spectral optimization problem was posed by Lagrange in 1773, when he asked:

> What is the shape of the strongest axial symmetric column with prescribed length, volume, and boundary conditions?

This problem has a fascinating history which is elegantly described in [Lewis and Overton, 1996]. The problem requires the maximization of the smallest eigenvalue of a self-adjoint fourth-order differential operator where the shape of the column appears as a coefficient of the operator.

Modern shape/structural optimization. As the list given at the beginning of this chapter indicates, shape and structural optimization are now vast fields of study and the technique used to approach a specific problem, depends strongly on the mathematical structure available in the problem. In each chapter of Part III, additional references are given for the specific problems studied. In the remainder of this section, we provide a few general references.

The theory of spectral optimization problems where the underlying operator is finite-dimensional is, of course, better developed and understood than that for problems with underlying differential operators. Discussions of finite dimensional eigenvalue optimization problems are given in [Lewis and Overton, 1996; Borwein and Lewis, 2000; Chu and Golub, 2005]. One obvious and important class of discrete eigenvalue optimization problems are those arising from the discretization of continuous ones. This approach to solving a PDE-constrained optimization problem is sometimes referred to as the discretize-then-optimize approach [Akcelik *et al.*, 2006].

A general reference for analytical and numerical methods for shape optimization problems with quite a few applications is [Haslinger and Mäkinen, 2003]. [Bucur and Buttazzo, 2005] and [Delfour and Zolésio, 2001] discuss mathematical tools for shape optimization, with a focus on variational methods and the description of shapes. [Cherkaev, 2000] discusses structural optimization, also emphasizing variational methods. [Henrot, 2006] is a beautifully written description of shape and structural optimization problems where the objective function is a function of the eigenvalues of an elliptic operator (see also Ch. 7). [Pironneau, 1984] discusses PDE-constrained shape optimization where the underlying equation is an elliptic system, especially for the shape optimization of an obstacle in fluid flow. [Laporte and Tallec, 2003] focuses on numerical methods for shape optimization, including both reduced-space methods and

full-space (one-shot) methods. The computational challenges involved in solving large scale PDE-constrained optimization problems are further discussed in [Akcelik *et al.*, 2006; Biegler *et al.*, 2007; Borzi and Schulz, 2009].

Level-set methods are often used in shape optimization to describe the shape of the domain being optimized and are advantageous over parameterizations of the domain if the topology of the optimal shape is unknown [Burger and Osher, 2005]. The term *topology optimization* is sometimes used to describe shape optimization when the topology of the domain is allowed to change. The term fictitious domain methods is sometimes used if a shape is embedded into a larger domain for application of the level-set method. Level-set methods can sometimes also be used for structural optimization problems. The solution to many structural optimization problems satisfy the *bang-bang principle*, a term from control theory introduced by Lev Semenovich Pontryagin, which refers to the solution attaining lower and upper bounds [Lions, 1971]. In this case, or if this constraint is imposed in the admissible set definition, a structural optimization problem reduces to a shape optimization problem in a lower dimensional space to find the interface between two materials. In this case, the level-set method can be used to find the optimal interface shape.

The mathematical theory of wave propagation. The spectral constraints that we consider in this thesis arise from the mathematical description of wave phenomena. In particular, we are concerned with acoustic, Schrödinger, and Maxwell wave equations and their frequency domain representations. Further discussion of topics in the theory of wave propagation is provided in appendix A.

Joe Keller published a review of mathematical methods for the study of wave motion [Keller, 1979]. Classical references which in part describe wave motion are [Courant and Hilbert, 1953] and [Morse and Feshbach, 1953] and one which emphasizes qualitative aspects of wave motion, including nonlinear and dispersive effects is [Whitham, 1974].

A nice introduction to spectral theory for unbounded self-adjoint operators as developed by John von Neumann is given in [Hislop and Sigal, 1996]. The spectral theory for perturbed operators is developed in [Kato, 1980; Rellich, 1969].

For PDE-constrained optimization problems, it is also necessary to compute the variation of the objective function with respect to a change in the design variable. For the spectral optimization

problems considered here, this invovles the way a spectral quantity changes with a change in the shape of a domain or material coefficient. A thorough description of the theory for the perturbation of the boundary in boundary value problems for partial differential equations is given in [Henry, 2005].

Computational methods for solving wave equations. For wave equations on bounded domains, the basic computational methods consist of finite element, finite difference, finite volume, and spectral methods [Larsson and Thomée, 2003; LeVeque, 2002; Trefethen, 2000]. These are referred to as volume methods, since the volume of the domain is discretized.

Boundary integral methods, sometimes also referred to as the method of moments (MOM), method of fundamental solutions (MFS), or boundary element methods (BEM), can be used when the Green's function of the equation is known, which is a severe restriction, requiring the media to be linear and also typically homogeneous [Hsiao and Wendland, 2008; Nedelec, 2001; Steinbach, 2008; Colton and Kress, 1983; Colton and Kress, 1998]. One of the advantages of these methods is that they reduce the dimensionality of the problem by one.

For problems posed on an infinite domain, outgoing boundary conditions must be enforced. Volume methods approximate outgoing boundary conditions by using a truncation of the Dirichlet-to-Neumann (DtN) map, examples of which include perfectly matched layers (PML) and absorbing boundary conditions (ABC). Another advantage to boundary integral methods is that the outgoing boundary conditions can be enforced exactly, by using Green's functions satisfying an outgoing boundary condition.

Hybrid methods, which combine two or more of these methods, are now often used for solving wave problems posed on infinite domains with inhomogeneous or nonlinear media.

Part II

Problems in Wave Propagation, Diffraction, and Scattering

Part II of this thesis, comprising Chapters 2-6, focuses on problems in wave propagation, diffraction, and scattering. Here, we provide a sketch and informal discussion of the problems considered. The reader may also wish to consult Appendix A for a more mathematically rigorous discussion of wave propagation in terms of spectral theory. We begin with several definitions which we use in the discussion below.

The spatially inhomogeneous wave equation is given by

$$\partial_t^2 v(x,t) = c^2(x)\ \Delta v(x,t) \tag{II.1}$$

where $c(x)$ is the wave speed and $\Delta = \sum_{j=1}^{d} \partial_{x_j}^2$ is the d-dimensional Laplacian. The spatial part of a time-periodic solution to the wave equation (II.1), $v(x,t) = u(x)e^{-\imath\omega t}$, satisfies the inhomogeneous Helmholtz equation

$$c^2(x)\ \Delta u(x,\omega) + \omega^2 u(x,\omega) = 0. \tag{II.2}$$

The Green's function $G(x,y,\omega)$ is defined to satisfy

$$c^2(x)\ \Delta G(x,y,\omega) + \omega^2 G(x,y,\omega) = \delta(x-y) \tag{II.3}$$

where $\delta(x-y)$ denotes the Dirac delta function. If the media is homogeneous, *i.e.* $c(x) \equiv c$, then the Green's function depends only on the difference $x-y$. In this case, we abbreviate the homogeneous Green's function $G(x,y,\omega)$ by $G(x-y,\omega)$. In Appendix A, we compute the homogeneous, outgoing Green's function to be

$$G(r,\omega) = \begin{cases} -(4\imath)^{-1} H_0^{(1)}(k|r|) & d=2 \\ -(4\pi|r|)^{-1}\exp(\imath k|r|) & d=3. \end{cases} \tag{II.4}$$

where $k = \omega/c$.

Diffraction occurs when a wave leaves an "aperture" Σ and propagates into free space. Here $\Sigma \subset V$ is defined to be a compact subset of a $d-1$ dimensional hyperplane, $V \subset \mathbb{R}^d$. If $u(x,\omega)$ vanishes on $V \setminus \Sigma$ and satisfies outgoing boundary conditions as $|x| \uparrow \infty$, then the (first) Rayleigh-Sommerfeld diffraction formula states that

$$u(x,\omega) \approx \int_{\Sigma} u(y,\omega)\ \nabla_x G^-(x,y,\omega)\cdot \nu\, \mathrm{d}y \tag{II.5}$$

where $\nu \in \mathbb{R}^d$ is normal to Σ,

$$G^-(x,y,\omega) = G(x,y,\omega) - G(x',y,\omega),$$

and x' is the reflection of x across Σ. The approximation made in Eq. (II.5) is that $|x - \Sigma| \gg \frac{2\pi}{\omega}$. The Fresnel diffraction formula may be obtained from Eq. (II.5) by using Taylor's theorem and assuming that the distance from the screen is large [Goodman, 2004].

In Chapter 2, we use the Rayleigh-Sommerfeld and Fresnel diffraction formulae to model a scattering experiment conduced at Argonne National Laboratory, where a coherent, monochromatic, x-ray beam is focused by a Fresnel zone plate onto a single-crystal layer of silicon.

In the remainder of Part II, we consider the discrete analogue of Eq. (II.1) on a graph, given by the inhomogeneous, discrete wave equation

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} v_j(t) = c_j \ \Delta_d v_j(t) \tag{II.6}$$

where c_j is the wave speed at node j,

$$\Delta_d v_j = \sum_{i \in \mathcal{N}(j)} (v_i - v_j)$$

is the discrete Laplacian, and $\mathcal{N}(j)$ denotes the neighbors of node j. The "spatial" part of a time-periodic solution to Eq. (II.6), $u_j(t) = v_j(\omega) e^{-\imath\omega t}$ satisfies the inhomogeneous, discrete Helmholtz equation

$$c_j \ \Delta_d v_j(\omega) + \omega^2 v_j(\omega) = 0. \tag{II.7}$$

In analogy to Eq. (II.3), the discrete lattice Green's function $G_{ij}(\omega)$ satisfies

$$c_j \ \Delta_d G_{ij}(\omega) + \omega^2 G_{ij}(\omega) = \delta_{ij}. \tag{II.8}$$

where δ_{ij} is the Kronecker delta function.

Discrete wave equations such as (II.6) arise in the discretization of PDEs, and also in various physical models such as mass-spring lattices and the tight binding approximation in solid-state physics (see *e.g.*, [Economou, 2006; Kevrekidis and Porter, 2009; Ablowitz and Zhu, 2010]). In Chapter 3 we show that the discrete wave equation (II.6) arises in a finite volume discretization of the planar Maxwell's equations. More concretely, we study Maxwell's equations for the (H_1, H_2, E) polarized mode in an inhomogeneous planar medium. For constant magnetic permeability μ, the E-field satisfies Eq. (II.1) where $c^{-2}(x) = \mu\epsilon(x)$ and ϵ is the electric permittivity. The planar Maxwell equations are discretized using a finite volume method to obtain a discrete set of equations which we recognize as Kirchhoff's laws for a corresponding inductor-capacitor circuit. If μ is constant,

this system of equations is precisely (II.6) where v_j is the voltage on a volume cell and the wave speed c_j is prescribed in terms of the capacitances and inductances of the circuit.

In Ch. 4, we use the lattice Green's function given in Eq. (II.8) to derive a discrete analogue of the classical Rayleigh-Sommerfeld diffraction formula given in Eq. (II.5) for the square lattice with homogeneous wave speed. More precisely, we show that there exists a modified lattice Green's function G_{ij}^{-} such that if $u_i(\omega)$ is specified on an aperture Σ and satisfies a discrete Sommerfeld outgoing radiation condition, then

$$u_j(\omega) = \sum_{i \in \Sigma} u_i(\omega) G_{ij}^{-}(\omega). \tag{II.9}$$

In contrast to Eq. (II.5), Eq. (II.9) is exact. We give a recursive algorithm for computing G_{ij}^{-}.

In Chapter 5, we use Eq. (II.9) to study wave propagation in two-dimensional, transmission-line model metamaterials. We demonstrate that for physically realizable parameters, the media is strongly dispersive.

As discussed above, the discrete wave equation (II.6) can be derived via a discretization of the wave equation (II.1). In Chapter 6, we study the "reverse direction" of deriving a PDE approximation to the discrete wave equation (II.6). For a regular triangular lattice, the continuum limit in the lattice spacing parameter $h > 0$, yields the wave equation (II.1) where $c_j \propto h$ and h is the lattice spacing. The quasi-continuum limit yields the wave equation with an additional $\mathcal{O}(h^2)$ isotropic, dispersive term which models the discreteness of the original equation. For this modified wave equation, we derive a dispersively-corrected, Rayleigh-Sommerfeld diffraction formula as in Eq. (II.5) where k^2 is replaced by a term that roughly looks like $\frac{k^2}{1-h^2k^2}$. This formula is used to describe the effect of discreteness on high-frequency wave propagation in a triangular lattice.

Chapter 2

Modeling of kinematic diffraction from a thin silicon film illuminated by a coherent, focused x-ray nanobeam

2.1 Introduction

Current X-ray nanodiffractometers use X-ray beams with diameters in the range of 25 to 50 nm [Chu *et al.*, 2008; Mimura *et al.*, 2007; Schroer *et al.*, 2005; Maser *et al.*, 2004a].[1] True nanometer sized ($<$ 10 nm) beams are expected in the near future [Maser *et al.*, 2004b; Kang *et al.*, 2008; Yan, 2009]. These beams are very divergent (on the order of $m$rad for wavelengths $\sim$ 1 Å) and necessarily have high spatial coherence, as coherent illumination is required to achieve diffraction-limited resolution from optical systems. When such a wavefront is used for diffraction analysis of crystalline materials, computation of the resulting scattering pattern requires an accurate representation of the spatial and angular incident beam distribution at the sample, a proper model of the scattering mechanism(s) within the sample, and propagation of the resulting coherent wavefront from the sample to detector [Yan *et al.*, 2008].

[1] Throughout this chapter, the terminology "diffraction" is used slightly differently than in other chapters, but consistent with the X-ray diffraction community. The diffracted wavefield from a silicon film is taken to mean the scattered wavefield.

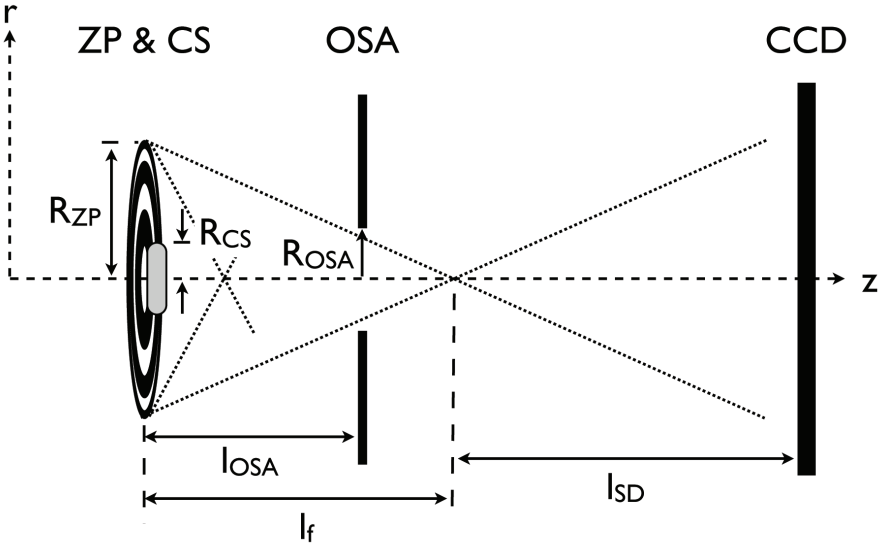


Figure 2.1: Schematic of a beamline with a Fresnel zone plate (ZP) focusing optic. A central stop (CS) is attached to the ZP to stop the zeroth-order beam. The dotted lines represent the focusing of the wavefield for the first and third order focal spots. An order sorting aperture (OSA) is placed close to the primary focal spot to choose the focus diffraction order. The sample is placed at the primary focal spot.

An accurate model of the incident beam on the sample surface must incorporate all optical components encountered by incident wavefront before impinging on the sample. In this chapter we model Fresnel phase zone plates (ZP), which to date have achieved the highest spatial resolution in the hard X-ray range. ZPs consist of concentric zones with alternating susceptibilities so that the amplitude and phase of the incident wave is modulated in these alternating zones. Diffraction of the incident wave produces multiple orders of focii along the axis of the ZP, and the desired focus is chosen by an order sorting aperture (OSA), as in Fig. 2.1. A significant fraction of the direct beam passes undiffracted through the ZP. This is prevented from reaching the sample by using a central stop (CS) and OSA. These additional optical components also modify the incident beam distribution.

Several recent studies [Yan *et al.*, 2008; Kohn and Kazimirov, 2007; Kazimirov *et al.*, 2009] have modeled the diffracted waves excited by coherently focused incident X-ray beams from single crystals. [Yan *et al.*, 2008] considered the incident beam as a point source emitting a spherical wave,

and modeled the evolution of the diffracted wave from a perfect or weakly deformed single crystal sample. [Kohn and Kazimirov, 2007] modeled a topographic technique in which an ideal Gaussian X-ray beam focused by a parabolic refractive lens was diffracted in the Bragg geometry from a single crystal substrate-epitaxial layer composite, and the one-dimensional (linear) distribution of intensity with position was recorded by a detector placed at the focus of the lens. In an extension of this technique, [Kazimirov *et al.*, 2009] used a Fresnel zone plate to focus the beam and a charge-coupled device (CCD) detector to record the two-dimensional diffraction image from silicon-on-insulator thin films. The cross-section of the measured data at the diffraction plane showed qualitative agreement with the model calculations of one-dimensional intensity variations with spatial position. Each of these models treat relatively simple incident beam profiles and use one-dimensional formulations to calculate the spatial variation of scattered intensity. To our knowledge, there have been no models that account for the full focused wavefield expected from a Fresnel zone plate defined by a set of vendor supplied specifications or the 2-D topographic images that would be observed when the diffracted beam from a single crystal sample illuminated by such an incident focused beam is recorded by a two-dimensional detector. In this chapter we address these issues.

In what follows, a diffraction experiment utilizing a phase zone plate focusing optic and a CCD detector is modeled. In this model the Fresnel diffraction formula is used to calculate the complete focused wavefield emanating from an ideal phase zone plate. The simulation incorporates the effects of the central stop and order sorting aperture. Using this incident beam and kinematic scattering theory, the diffracted wave from a thin, perfect, Si layer placed at the first order focal spot is calculated. This wave is then propagated to a two-dimensional detector, and the intensity pattern that would be recorded is constructed. The incident and diffracted beam computations are then compared to experimental data measured with the Center for Nanoscale Materials (CNM) Hard X-ray Nanoprobe at Sector 26-ID-C of the Advanced Photon Source (APS).

2.2 Theory

To simulate the diffracted wavefield from a thin film illuminated with a convergent, coherent nanobeam, we follow the method outlined by [Yan *et al.*, 2008]. This framework uses the theories of free-space wave propagation and crystal diffraction, both of which are well known [Goodman, 2004;

Authier, 2002]. They are briefly reviewed here to establish notation.

2.2.1 Fresnel wave propagation

We define the *Fresnel diffraction integral operator* by

$$\mathcal{R}_z[f](\mathbf{x}) = \begin{cases} (f * g_z)(\mathbf{x}) & z > 0 \\ f(\mathbf{x}) & z = 0 \end{cases}, \tag{2.1}$$

where $g_z(\mathbf{x})$ is the convolution kernel

$$g_z(\mathbf{x}) = \frac{\exp(2\pi\imath kz)}{\imath\lambda z} \exp\left(\frac{\imath\pi k}{z}|\mathbf{x}|^2\right).$$

If we consider $U_0(\mathbf{x})$ to be a monochromatic, scalar, electromagnetic field (with wavelength $\lambda = 1/k$) measured at an aperture, the wavefield at a distance z from the aperture, $U_z(\mathbf{x})$, can be described in the Fresnel approximation by $U_z(\mathbf{x}) = \mathcal{R}_z[U_0](\mathbf{x})$. Derivations of this can be found in, for example, Appendix A or [Goodman, 2004]. We interchangeably use *aperture function* and wavefield at $z = 0$ throughout. For a radially-symmetric aperture function $f(\mathbf{x}) = f(\rho)$, we rewrite Equation (2.1) in polar coordinates, and integrate over the angular coordinate to yield the simpler expression

$$\mathcal{R}_z[f](r) = 2\pi\frac{\exp(2\pi\imath kz)}{\imath\lambda z} \exp\left(\imath\frac{\pi k}{z}r^2\right) \int_0^\infty f(\rho) \exp\left(\imath\frac{\pi k}{z}\rho^2\right) J_0\left(\frac{2\pi}{\lambda z}r\rho\right) \rho \,\mathrm{d}\rho \tag{2.2}$$

where J_0 is the zeroth-order Bessel function of the first kind. The integral in Equation (2.2) may be referred to as the Hankel transform of order zero or the radial Fourier transform. In this chapter, we will use the uniqueness property of the Fresnel diffraction integral, which can be stated $\mathcal{R}_{z_1+z_2} f = \mathcal{R}_{z_1}\mathcal{R}_{z_2} f$ for all $z_1,\, z_2 \geq 0$.

In the Fraunhofer approximation, the wavefield a distance z from the aperture is written

$$U_z(\mathbf{x}) = \frac{1}{\imath\lambda z} \exp\left(2\pi\imath kz\right) \exp\left(\imath\frac{\pi k}{z}|\mathbf{x}|^2\right) \mathcal{F}[U_0]\left(\frac{\mathbf{x}}{\lambda z}\right), \tag{2.3}$$

where $\mathcal{F}[f](\mathbf{k}) = \int f(\mathbf{x}) e^{-2\pi\imath\mathbf{k}\cdot\mathbf{x}} \,\mathrm{d}\mathbf{x}$ is the Fourier Transform. This approximation is valid when the (dimensionless) Fresnel number

$$F = \frac{a^2}{\lambda z},$$

is small ($F \ll 1$), where a is the characteristic width of the aperture function [Goodman, 2004]. Following [Yan *et al.*, 2008], we can also define an angular Fresnel number

$$Y_A = \frac{\alpha^2 z}{\lambda},$$

where α is the characteristic width of the Fourier transform of the aperture function. The far field approximation is valid when $Y_A \gg 1$.

2.2.2 X-ray diffraction from a thin film

To model diffraction from a thin film, we consider a kinematically scattering sample with structure factor $F_{\mathbf{h}}$ and reciprocal lattice vector $\mathbf{h}$. The diffracted angular spectrum $A_h(\mathbf{k}_i, \mathbf{k}_d)$ for a unit plane wave with wave vector $\mathbf{k}_i$ and diffracted wave vector $\mathbf{k}_d$ is given by

$$A_h(\mathbf{k}_i, \mathbf{k}_d) = \frac{F_{\mathbf{h}}}{v_c} \int_V \exp\left(2\pi\imath \Delta\mathbf{k} \cdot \mathbf{x}\right) \mathrm{d}\mathbf{x} \tag{2.4}$$

where v_c is the unit cell volume, V is the crystal domain, and $\Delta\mathbf{k}_h \equiv \mathbf{k}_d - \mathbf{k}_i - \mathbf{h}$ [Authier, 2002]. When $\Delta\mathbf{k}_h = 0$, the Bragg condition is satisfied and diffracted intensity exhibits a maximum. We note that, for Equation (2.4) to be valid, the thickness of the (single crystal) thin film along the diffraction vector must be smaller than the extinction depth, t_e, of the material. For films appreciably thicker than t_e, dynamical diffraction formulations [Yan *et al.*, 2007] must be used.

Let us consider a thin film with thickness t_f in the $\hat{\mathbf{z}}$-direction, and infinite in the $\hat{\mathbf{x}} - \hat{\mathbf{y}}$) plane. Furthermore, let us assume that the crystal and sample coordinate systems are aligned such that $\mathbf{h}$ and $\hat{\mathbf{z}}$ are anti-parallel. In this case, the integrals in the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions in Equation (2.4) are Dirac-delta functions, yielding

$$\begin{aligned} A_h(\mathbf{k_i}, \mathbf{k_d}) &\propto \delta(\Delta k_{hx})\delta(\Delta k_{hy}) \int_0^{t_f} \exp\left(i2\pi\Delta k_{hz} z\right) \mathrm{d}z \\ &\propto \delta(\Delta k_{hx})\delta(\Delta k_{hy}) \operatorname{sinc}\left(\pi(k_{dz} + k_{iz} - h_z)t_f\right). \end{aligned} \tag{2.5}$$

In a traditional radial scan (*i.e.* scans along $\mathbf{h}$) of a symmetric reflection for a thin film, $\Delta k_{hx} = \Delta k_{hy} = 0$ and $k_{dz} + k_{iz} - h_z = 2(\sin\theta - \sin\theta_B)/\lambda$, where θ is the angle between the incident beam and the sample surface, and θ_B is the Bragg angle of the reflection. The diffracted intensity

is given by

$$I(\Delta 2\theta) \propto |A_h|^2 \propto \mathrm{sinc}^2\left(\frac{\pi t_f}{\lambda}\cos\theta_B \Delta 2\theta\right), \tag{2.6}$$

where $\Delta 2\theta = 2\theta - 2\theta_B$. Equation (2.6) predicts a central peak bracketed by fringe peaks[2]. The full-width at half-maximum-intensity (FWHM) of the diffraction peak, β_t, can be calculated from Equation (2.6), which yields the classical Scherrer equation [Ying *et al.*, 2009]

$$\beta_t \approx \frac{0.886\lambda}{t_f \cos(\theta_B)}. \tag{2.7}$$

The radial scan is also sensitive to the lattice strain ϵ, along $\mathbf{h}$. The strain is measured as a peak shift $\Delta 2\theta$ relative to θ_B and is given by

$$\epsilon = \frac{\sin\theta_B}{\sin(\theta_B + \Delta\theta)} - 1, \tag{2.8}$$

where $\Delta\theta = \Delta 2\theta/2$. It is important to note that the angular coordinate 2θ (thus β_t and $\Delta 2\theta$) is measured with respect to the incident wave vector $\mathbf{k}_i$.

2.3 Simulations

In this section, we compute the spatial intensity distributions using the equations from Section 2.2. Numerical values used in the calculations are tabulated in Table 2.1. These values were selected to match the experimental settings that are described in Section 2.4.

2.3.1 Incident beam simulation

We consider a monochromatic plane wave of unit amplitude in the hard X-ray spectrum incident on an ideal Fresnel ZP. The ZP consists of alternating concentric rings of material (zones) of decreasing width. Light passing through odd zones are given a phase shift $\phi_{ZP} = \frac{2\pi}{\lambda}(n-1)t_{ZP}$, where n is the index of refraction and t_{ZP} is the thickness of the ZP [Jones, 1969]. This is an idealization of the "staggered spoke" zone plate structure used in the experiment and described by [Feng *et al.*, 2007]. Each zone is bounded between radii given by

$$\sqrt{(m-1/2)\lambda l_f} \le r \le \sqrt{(m+1/2)\lambda l_f},$$

[2]This result is the manifestation of the Pendellösung effect in this formulation of the diffracted intensity.

	INSTRUMENT
E	11.2 keV
λ	1.107×10^{-10} m
k	9.033×10^{9} m^{-1}
	ZONE PLATE
N_{ZP}	1385
R_{ZP}	66.5×10^{-6} m
Δr	24×10^{-9} m
l_f	28.8346×10^{-3} m
n	$1 - (2.3193 \times 10^{-5} - 1.4665 \times 10^{-6}\ \imath)$
t_{ZP}	300×10^{-9} m
ϕ_{ZP}	$-(0.1257 - 0.0079\ \imath)\pi$
	CENTRAL STOP
R_{CS}	30×10^{-6} m
t_{CS}	80×10^{-6} m
ϕ_{CS}	$-(33.5219 - 2.1196\ \imath)\pi$
	ORDER SORTING APERTURE
R_{OSA}	15×10^{-6} m
l_{OSA}	25.8347×10^{-3} m
	CCD
l_{SD}	700 mm

Table 2.1: Numerical values used in the simulations. The zone plate and central stop are both gold; the order-sorting aperture is platinum-iridium.

for $0 < m \leq N_{ZP}$, where m is the integer order of the zone and N_{ZP} is the order of the outermost zone. For the $m = 0$ zone, $0 \leq r \leq \sqrt{\lambda l_f/2}$. Thus, the wavefield measured in the aperture of the ZP is

$$U_{ZP}(r) = \begin{cases} \exp(\imath\phi_{ZP}) & \text{for odd zones} \\ 1 & \text{for even zones} \\ 0 & \text{for } r \geq \sqrt{(N_{ZP}+1/2)\lambda l_f} \end{cases}. \tag{2.9}$$

A central stop with radius R_{CS} is attached to the ZP and the X-rays propagate through this region with an additional phase shift ϕ_{CS}. The composite aperture function is

$$U_0(r) = \begin{cases} U_{ZP}(r)\exp(\imath\phi_{CS}) & \text{for } r < R_{CS} \\ U_{ZP}(r) & \text{otherwise} \end{cases}, \tag{2.10}$$

with $\phi_{CS} = \frac{2\pi}{\lambda}(n-1)t_{CS}$, where t_{CS} is the thickness of the central stop.

To calculate the wavefield downstream of the OSA, we need to first calculate the wavefield at the OSA, U_{OSA}, by applying the radially symmetric Fresnel diffraction integral operator (Equation (2.2)) to U_0 in Equation (2.10). The one-dimensional integral in Equation (2.2) is piecewise smooth over each zone and computed using a recursive adaptive Simpson quadrature rule. Due to the thickness and material of the OSA, it effectively truncates the wavefield for $r > R_{OSA}$ so at the downstream side of the OSA, we have

$$U_{OSA}(r) = \begin{cases} \mathcal{R}_{l_{OSA}}[U_0](r) & r < R_{OSA} \\ 0 & \text{otherwise.} \end{cases}$$

We remind the reader that the subscript (OSA) on the field U is a label, while the subscript (l_{OSA}) on the Fresnel operator $\mathcal{R}$ is the distance of propagation. We now use the uniqueness property of the Fresnel integral operator to continue the beam from the OSA to the focal spot or past the focal spot to the CCD detector. This integral in Equation (2.2) was simply calculated using the trapezoid rule with a 0.05 nm discretization.

This two-step wave propagation process was used to generate a plot of the amplitude downstream of the ZP and the log of the amplitude is plotted in Fig. 2.2a. The simulation shows that the central portion of the straight through beam is blocked by the central stop. Most of the incident beam is seen to pass straight through the ZP, while the diffracted wavefield converges for

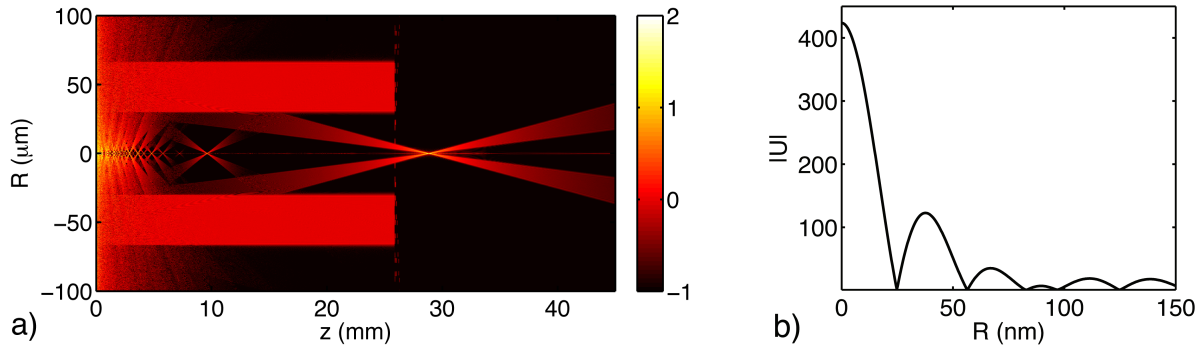


Figure 2.2: (a) Log of the amplitude of the wavefield generated from the ZP with CS and OSA considered. The OSA is at $z \approx 26$ mm. The primary focus of the ZP is seen at $z \approx 29$ mm, with the divergent cone of radiation propagating to large z. A central line of amplitude (2 orders of magnitude smaller than the focused amplitude) can be seen due to the zone plate used in the model. (b) The amplitude of the wavefield at the primary focal spot.

multiple orders of focii. The focal order is selected by the order sorting aperture, which is placed at $z = l_{OSA}$ for the primary focus spot. Here, only the focused wavefield within the radius of the OSA is allowed to propagate through. We note the Fresnel approximation used in this simulation is invalid for z near the ZP and OSA, and the scattering at the OSA is due to this approximation.

The field at the focal spot is given by

$$U_f(r) = \mathcal{R}_{l_f - l_{OSA}}[U_{OSA}](r), \tag{2.11}$$

with the amplitude plotted in Fig. 2.2b. The simulation shows an intensity FWHM of 22.4 nm, slightly smaller than the Rayleigh resolution $1.22\Delta r = 30$ nm, where Δr is the outermost zone width. This complex wavefield at the focal spot U_f is used in Section 2.3.2 to simulate diffraction from a kinematically scattering thin film.

As seen in Fig. 2.2a, the focused incident beam diverges from the focal spot as it propagates to the CCD detector. The wavefield at the CCD detector is

$$U_{CCD}(\mathbf{x}) = \mathcal{R}_z[U_{OSA}](r) \tag{2.12}$$

where $z = l_f - l_{OSA} + l_{SD}$ is the distance from the OSA to the detector. In Fig. 2.3a we plot the computed radial incident wavefield intensity, normalized to the integrated intensity in the non-

zero region. The calculation uses a 1 micron step size at the CCD and averages the intensity over 13 micron regions to emulate CCD pixels. Our simulations show that these oscillations are a consequence of the truncation of the wavefield by the OSA, a physical realization of Gibb's phenomena for truncated Fourier series. In Fig. 2.3b, we show the simulated CCD image based on this calculated radial distribution. The simulation shows a ring of intensity resulting from the incident wave being limited in the center by the CS.

2.3.2 Thin film diffraction simulation

A general wavefield incident on a sample, $U_i(\mathbf{x})$, may be decomposed into its angular spectrum by the Fourier Transform $A_i(\mathbf{k}) = \mathcal{F}[U_i](\mathbf{k})$. Each plane wave component of the incident beam excites a diffracted plane wave, which interferes coherently with all other diffracted waves resulting from the other incident plane wave components [Yan *et al.*, 2008]. Thus, the diffracted amplitude for a particular wave vector $\mathbf{k}'_d$ can be written as an integral over all components of the incident angular spectrum

$$A(\mathbf{k}'_d) = \int A_i(\mathbf{k}'_i) A_h(\mathbf{k}'_i, \mathbf{k}'_d)\, \mathrm{d}\mathbf{k}'_i, \tag{2.13}$$

where $\mathbf{k}'_d$ and $\mathbf{k}'_i$ are two dimensional wave vectors defined in the sample coordinate system. It should be noted here that while $A_h(\mathbf{k}_i, \mathbf{k}_d)$ is a three dimensional field, the third dimension ($\hat{\mathbf{z}}$) is not independent of the other two and its magnitude is given by $k_z = \sqrt{\lambda^{-2} - |\mathbf{k}|^2}$. From Equation (2.4), we see Equation (2.13) is just a convolution (in Fourier space) of the incident wavefield with the diffracted wavefield of a unit plane wave from the sample.

In Fig. 2.4, we introduce the coordinate systems used to simulate diffraction from the sample at the focal spot. To compute the angular spectrum of the diffracted field at the sample $A(\mathbf{k}'_i)$, we first interpolate the incident wavefield U_f (Equation (2.11)) into cartesian coordinates using a piecewise cubic hermite interpolating polynomial. The angular spectrum at the focal spot $A_i(\mathbf{k}_i)$ can then be calculated numerically using the fast Fourier transform. The wavefield $U_f(x, y)$ is a 2-D slice of the incident wavefield on the focal plane, and the resulting $A_i(k_{ix}, k_{iy})$ is defined strictly in two dimensions. The third component is given by $k_{iz} = \sqrt{\lambda^{-2} - (k_{ix}^2 + k_{iy}^2)}$. Finally, the coordinate

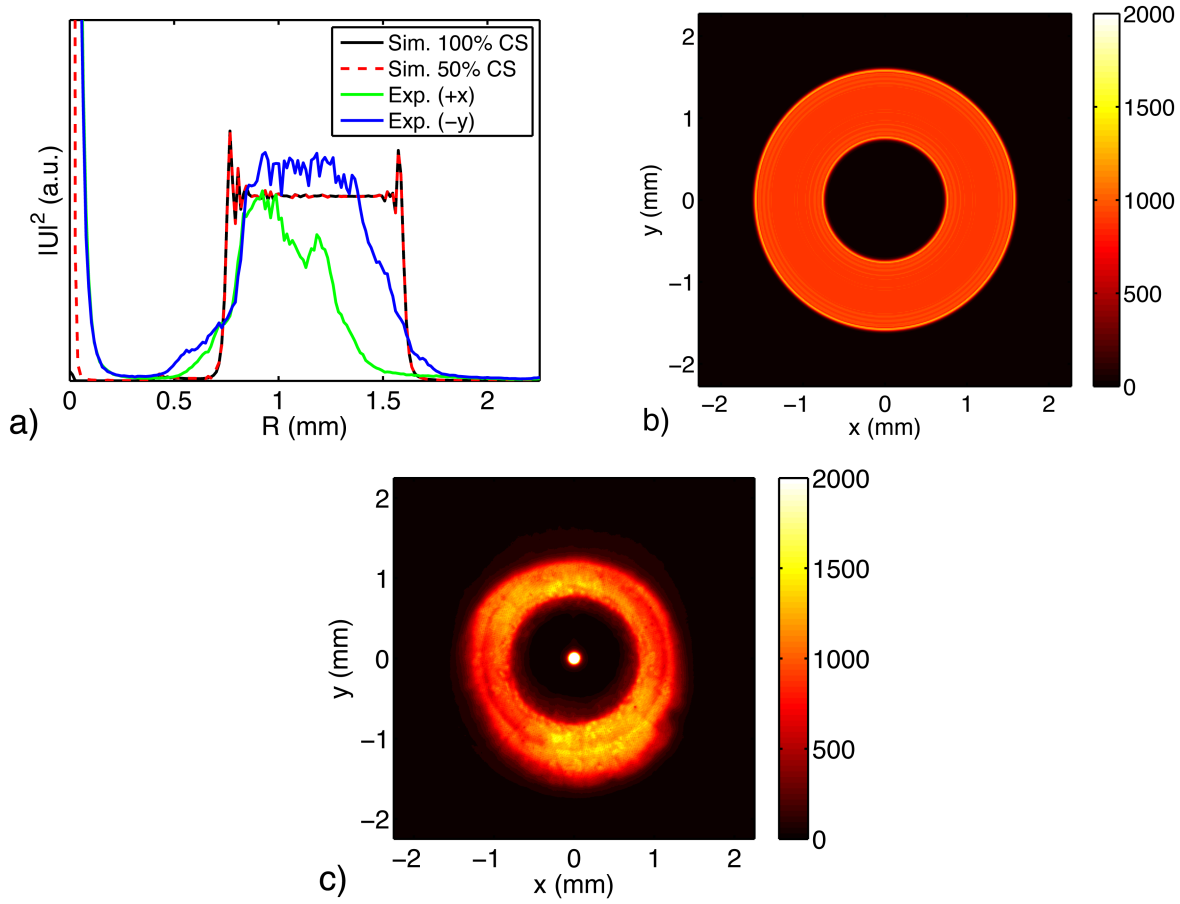


Figure 2.3: (a) The radial intensity distribution of the incident beam at the detector plane ($z = 700$ mm): computed (black and red), experimental (green and blue). Using a central stop with half the thickness of the ideal central stop (dashed red), a peak in intensity at $R = 0$ is reproduced in the simulation. (b) Simulated CCD image of the incident beam intensity. (c) The measured incident beam intensity at the CCD. Slices of the radial intensity distribution along $+\hat{\mathbf{x}}$ and $-\hat{\mathbf{y}}$ are plotted in (a) in green and blue, respectively. Leakage from the central stop is seen in the center of the annulus, and saturated the CCD detector. The simulated plots are normalized to the integrated intensity of the non-zero annulus. The experimental plots are normalized to the integrated intensity of the slice along $-\hat{\mathbf{y}}$.

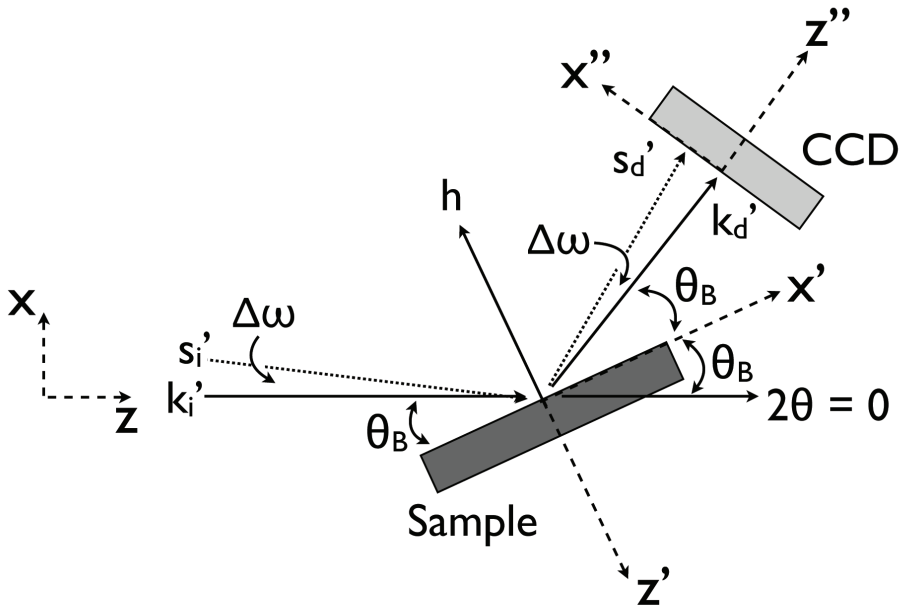


Figure 2.4: Sketch of coordinate systems used, with the sample and detector set at the Bragg condition. For a plane wave component of the incident beam at an angle $\theta_B + \Delta\omega$ to the sample surface, $\mathbf{s}_i'$, the symmetric condition is met for a diffracted wave vector, $\mathbf{s}_d'$, that has an angle $2\theta = 2\theta_B + \Delta\omega$ with respect to the incident beam. On the CCD, this (symmetric) diffracted wave vector is $\Delta\omega$ from the perfect Bragg condition.

system is rotated by the transformation matrix

$$T = \begin{bmatrix} \sin\theta_B & 0 & \cos\theta_B \\ 0 & 1 & 0 \\ -\cos\theta_B & 0 & \sin\theta_B \end{bmatrix}$$

to obtain $A_i(\mathbf{k}'_i)$. We can now apply Equation (2.13) to obtain the diffracted field at the sample. For a symmetric reflection from a thin film sample, A_h is given by Equation (2.5) and the Dirac-delta functions require $k'_{ix} = k'_{dx}$ and $k'_{iy} = k'_{dy}$ for non-zero amplitude. Therefore $k'_{iz} = k'_{dz}$, and Equation (2.13) simplifies to

$$A(\mathbf{k}'_d) \propto A_i(\mathbf{k}'_d)\ \mathrm{sinc}\left(\pi(2k'_{dz} - h_z)t_f\right). \tag{2.14}$$

It is seen that the diffracted angular spectrum for this sample can be computed by point-wise multiplication of the incident beam spectrum (for each diffracted beam coordinate) with the sample diffraction function from a parallel plane wave.

To simulate the real-space diffracted wavefield as measured by the CCD, $U_{SOI}(\mathbf{x}'')$, $A(\mathbf{k}'_d)$ is first rotated to the detector (double-primed) coordinate system to get $A(\mathbf{k}''_d)$. The real-space diffracted wavefield at the sample $U_d(\mathbf{x}'') = \mathcal{F}^{-1}[A](\mathbf{x}'')$, which can be propagated to the CCD by utilizing the Fraunhofer approximation since the angular Fresnel number is large. For a 124 nm thin film, we approximate the angular acceptance of the crystal α by the Scherrer Equation (2.7). The angular Fresnel number $Y_A \sim 10^5$ for $z = l_{SD}$, where l_{SD} is the sample-to-detector distance. Thus, we apply Equation (2.3) to yield the far-field image as measured by the CCD

$$U_{SOI}(\mathbf{x}'') \propto \mathcal{F}[U_d]\left(\frac{\mathbf{x}''}{\lambda l_{SD}}\right). \tag{2.15}$$

Here we drop the phase terms since only the intensity is measured at the detector. Unfolding Equation (2.15), it is clear that the real space diffracted wavefield in the Fraunhofer approximation is just the angular spectrum of the diffracted wavefield at the sample

$$U_{SOI}(\mathbf{x}'') \propto A\left(\frac{\mathbf{x}''}{\lambda l_{SD}}\right). \tag{2.16}$$

This computation is normalized to the maximum intensity and plotted on a log scale in Fig. 2.5a. In the figure, the two main vertical lobes of intensity correspond to the main diffraction peak, with

the central portion missing due to the central stop. The ancillary intensity maxima on either side of the primary peak, in x'', correspond to the thickness fringes predicted by Equation (2.6).

The approximate computation time for each part of the simulation based on code implemented in Matlab 7.3© and run on a 2.4 GHz intel processor are given as follows: one (coarse grid) vertical slice of the Fig. 2.2a takes 16 seconds, with the entire mesh plot taking approximately 1.5 hours. The fine grid vertical slice computed at the OSA that was used to continue the wavefield to the focal spot took 1 hour. The diffraction equation (2.14) computation for the thin film sample takes a few seconds. Thus within a few hours, the diffraction pattern from a kinematically scattering thin film sample can be accurately calculated. We note here that, since the simulations conducted for this manuscript were reasonably straightforward, extensions of this formalism to asymmetric incident beams and kinematically diffracting samples of finite size in three dimensions should not be formidable. These extensions are currently under investigation.

2.4 Experimental verification

To test the validity of the simulations presented in Section 2.3, we conducted experiments at the CNM nanodiffractometer at Sector 26-ID-C of the Advanced Photon Source. Further details about this instrument can be found in [Maser *et al.*, 2004a]. In our experiments the incident beam from the dual undulator source of this beamline was monochromated via a double crystal Si (111) monochromator, with an energy resolution ($\Delta E/E$) of 1.7×10^{-4}, tuned to 11.2 keV ($\lambda = 1.107$ Å). This monochromatic beam was focused to nominally 30 nm by a "staggered spoke" Xradia© Au Fresnel phase zone plate [Feng *et al.*, 2007] with a 24 nm outermost zone width, 66.5 micron radius (R_{ZP}) and a 30 micron radius central stop (R_{CS}) (Fig. 2.1). The focal distance is about 29 mm from the ZP, with a depth of focus of about 10 μm and an angular divergence of 2.4 mrad. A 250 μm thick platinum-iridium (Pt:Ir / 95:5) order sorting aperture, with a radius (R_{OSA}) of 15 microns, is located approximately 3 mm upstream of the focal spot.

A Princeton© PIXIS-XF 2-D CCD detector consisting of a 1024 × 1024 array of 13 μm square pixels was used to record X-ray intensities. At the sample-to-detector distance (l_{SD}) of 700 mm, the angular resolution is 18.6 μrad (3.83 arcsec), with total angular range of 19.05 mrad (1.09°). The angular resolution is probably slightly poorer due to the point spread function of the detector,

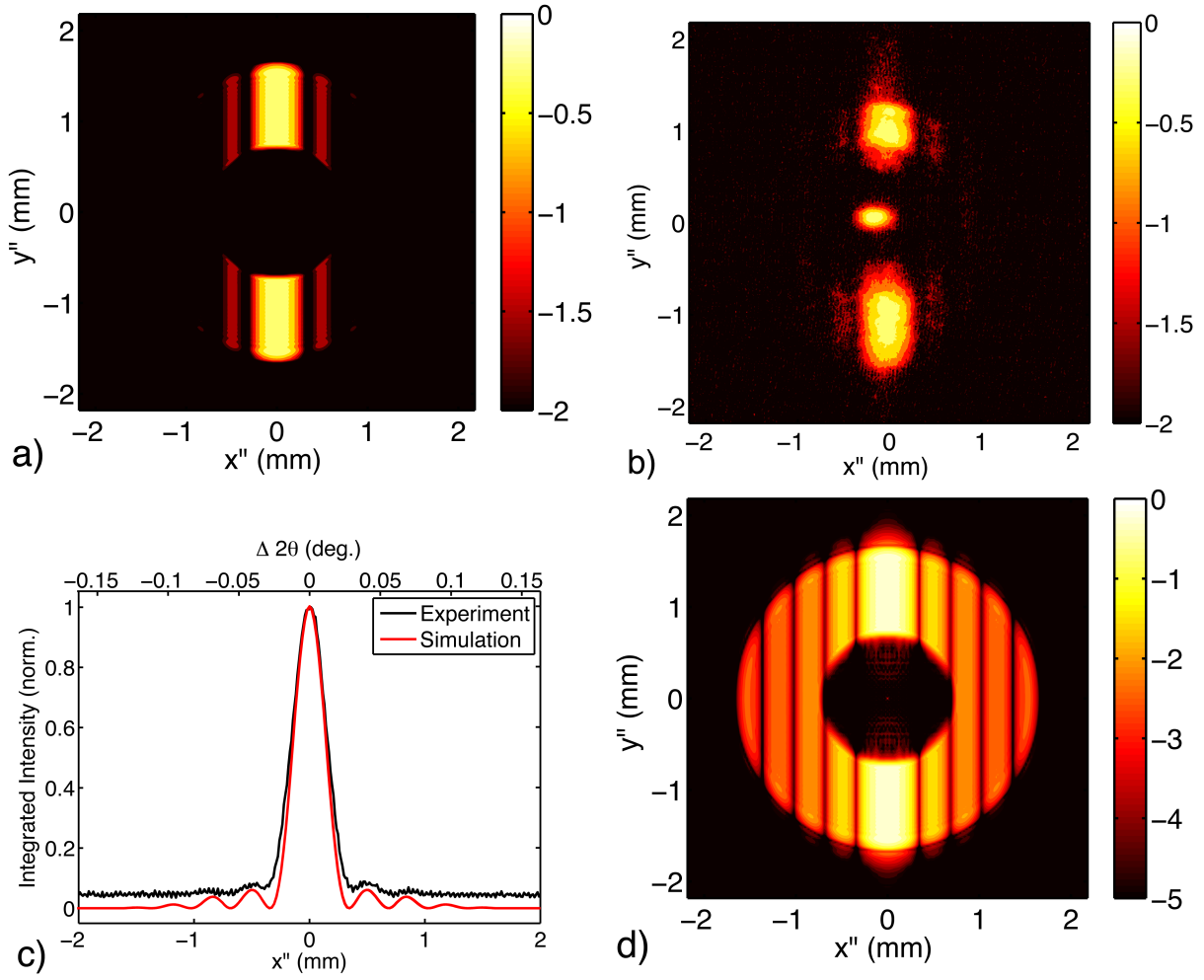


Figure 2.5: (a) Simulated thin film diffraction from a perfect zone plate with central stop and order sorting aperture, normalized to the maximum intensity. (b) The measured 004 diffraction peak from SOI, normalized to the maximum intensity. Leakage from the central stop is seen in the center of the diffracted intensity. Both (a) and (b) are plotted on a log scale to highlight thickness fringes on either side of the main peak. (c) A comparison between experimental and simulated wavefield intensity integrated along y''. (d) The simulated diffraction pattern plotted with a larger range of intensity.

which was not measured or available. However, it is unlikely that the spread is larger than ∼10 μm, as a short 1:1 fiber taper is used to image the phosphor screen. As we will show in the next section, ∼5 pixel resolution (∼ 0.005°) is adequate for the purposes of this experiment, which is significantly larger than the expected point spread of the detector.

The incident beam intensity distribution was measured at a distance of 700 mm from the focal point (Fig. 2.1) with the CCD normal to the zone plate axis. The measured distribution is shown in Fig. 2.3c.

For the diffraction experiments, we used a commercially available semiconductor grade Soitec© silicon-on-insulator sample [Ying *et al.*, 2009]. This sample consisted of a crystalline Si thin film layer (the SOI layer), stacked on a 140 nm SiO_2 layer, which was on a 0.7 mm thick Si 001-type substrate. The film thickness of the SOI layer was measured by cross-sectional TEM to be 124 nm ± 1 nm. This is an order of magnitude smaller than the extinction depth of the 004 reflection at this energy and ensured that the sample was scattering in the fully kinematical mode. Due to the bonding procedure used to make these samples, the [001] vector of the SOI layer is tilted by approximately ∼ 0.4° with respect to the substrate [001] direction. Consequently, it was possible to measure the diffraction peak of the thin surface layer without interference from the substrate peak. The SOI 004 reflection was found by varying the sample angle θ, and detector position 2θ, to maximize the intensity measured by the CCD; the peak was found at a sample angle $\theta = 24.02°$ with the CCD detector at $2\theta = 48.175°$. The (2-D) spatial distribution of the diffracted intensity is shown in Fig. 2.5b. To acquire this image the CCD was exposed for 15 seconds and background corrected using a 15 second dark count image.

2.5 Discussion / Conclusions

2.5.1 Incident beam profile and analysis

Both the simulated and measured incident wavefields at the CCD detector show an annulus of intensity (Fig. 2.3) resulting from a combination of the central stop and the divergence of the focused beam. Here both simulated and measured fields are normalized by the integrated intensity over the non-zero annulus, enforcing both signals have equal energy. While the spatial distribution of the intensities agree quite well, there are three differences that merit discussion: the experimentally

measured intensity is radially asymmetric, has less sharp edges, and features a bright spot in the center of the annulus.

To highlight the asymmetry in the measured incident beam (Fig. 2.3c), we plot two slices of the radial distribution along $+\hat{\mathbf{x}}$ and $-\hat{\mathbf{y}}$ in Fig. 2.3a. The experimental data is normalized to the integrated intensity of the latter. The slice along $-\hat{\mathbf{y}}$ exhibits higher and more evenly distributed intensity than the slice in $+\hat{\mathbf{x}}$, in much better agreement with the ideal ZP simulation.

The lack of symmetry in the measured incident beam may be caused by manufacturing defects, time-dependent radiation damage to the ZP, or slight misalignment of the beamline components. Our ability to model the first is impeded by the lack of information from the manufacturer of this focusing optic; the specification sheet does not include either a spatial roughness distribution or an average roughness number. A recent 1-D study of the roughness between zones in the ZP has shown that provided the zones are consistent with Equation (2.9) and the RMS roughness is sufficiently small compared to the outermost zone width, the ideal ZP approximation is valid [Yan, 2009]. Thus, we expect that small radially-symmetric RMS roughness would not qualitatively change the incident beam intensity distribution. However, the question remains whether a 2-D roughness model could, at least partially, explain the asymmetry in the experimentally measured incident beam intensity. Using the same reasoning, we did not include time-dependent radiation damage or beamline misalignment in the model since we have no data on these parameters. A detailed study of such effects is beyond the scope of this chapter and will be carried out separately. At this time we can only conclude that the measured profile indicates a non-ideal focusing optic.

The smoother edges of the measured intensity in Fig. 2.3a are mainly due to the ZP defects discussed above. Other contributing factors to the sharper edges of the model could be diffuse scattering in the experiment, edge effects and/or approximating the "staggered spoke" structure of the actual ZP by an ideal ZP in the simulation. While we have not accounted for these, there is still good agreement between portions of the measured incident beam and our model.

We attribute the bright spot in the center of the annulus in the experimentally measured incident beam CCD image (Fig. 2.3c) to be caused by a "thinner-than-expected" central stop. The central stop attenuates the amplitude of the un-diffracted (zeroth-order) wavefield and also adds an extra phase shift so the diffracted wavefield does not constructively interfere at the focal spot. For a thinner CS, more of the un-diffracted beam would pass and result in a brighter spot of intensity at

the CCD detector with a matching diameter (80 μm). To demonstrate this, we considered a CS with half of the thickness specified by the manufacturer. This led to a decrease of attenuation of the zeroth-order beam by a factor of 780 in the CS. The resulting wavefield at the CCD detector (dashed red in Fig. 2.3a) does feature a bright central spot, which would illuminate only the central few pixels of the CCD.

Another phenomenon that could result in a bright central spot is the presence of higher harmonics from the dual undulator source. This is unlikely for two reasons: first, the upstream mirrors at the 26-ID-C beamline have an energy cut-off of about 12 keV and should prevent these higher energy photons from reaching the zone plate. Second, layers of aluminum foil were used as attenuators to prevent damage to the CCD while measuring the incident beam. As the amount of attenuation was increased, the central spot intensity was observed to decrease at the same rate as the (focused) ring of intensity. We concluded that the central spot is primarily of the same energy as the focused X-rays, and any effect of the higher harmonics was secondary.

2.5.2 Diffracted beam profile and analysis

The simulated and experimentally measured diffraction patterns are shown in Figs. 2.5(a,b) respectively. In these images the x'' coordinates correspond to the 2θ direction, and the plotted intensities have been normalized by the maximum peak intensity. Qualitatively, the simulated and measured patterns look similar. In both, the central portion of the main diffraction peak is shadowed by the central stop and thickness fringes with lower intensities bracket the main peak (in x''). There is also reasonable agreement in the real-space position, breadth, and relative intensities of the peak features.

The differences between the simulated and measured diffraction patterns can be attributed to the incident beam spectrum. The bright spot at the center of the experimental pattern is due to leakage of the zeroth-order beam through the central stop. As with the incident beam distribution (Fig. 2.3c), the lower lobes of the diffracted signal ($y'' < 0$) are more intense than the upper lobes. These intensity artifacts are predicted by Equation (2.14): the diffracted amplitude for a particular wave vector is just the thin film diffracted amplitude (from a plane-parallel wave) modulated by the incident beam amplitude. Thus, any imperfections in the incident beam distribution will have a linear effect on the measured diffraction signal. Interpretation of such artifacts is much easier if

a test sample with a simple diffraction function is chosen.

The SOI thin film layer is such an ideal test sample. First, the small thickness gives an angular Fresnel number that is much greater than unity. For almost any placement of detector the Fraunhofer approximation is valid for the measured diffracted intensity. Thus, while the measured diffracted signal by the 2-D CCD is a real-space intensity distribution, it is simply the rescaling of the diffracted angular spectrum found at the sample. For films thicker than the extinction depth, dynamical diffraction simulations may be required. As shown by [Yan *et al.*, 2008], the Fraunhofer approximation becomes invalid in such cases for most detector positions. This leads to complex real-space diffracted intensity distributions, especially if strain gradients are present [Yan *et al.*, 2007].

In addition to the effects from the thickness, the thin film can be considered to be infinite in the plane of the film, leading to Equation (2.5), and subsequently Equation (2.14). The latter explains the intensity artifacts that are found in the measured diffracted intensity that result from the incident beam. The former states that the only appreciable intensity can be measured when the difference between a diffracted and an incident beam vector is co-linear with the reciprocal lattice vector (*i.e.* $\Delta k_{hx} = \Delta k_{hy} = 0$). For a symmetric reflection, this condition is satisfied when components of the incident and diffracted spectra make symmetric angles with the sample surface. Due to the large divergence of the incident beam and the large angular acceptance of the detector, a single CCD image thus contains information from the entire radial scan about the Bragg condition. However, it is not a true radial scan since each angular position on the detector does not maintain a 2:1 ratio with the (fixed) sample angle. For samples with finite lateral domains, the symmetric requirement for the incident and diffracted components is relaxed and each spatial position on the detector will contain scattered information from a range of incident beam angular components. In these cases, distinguishing between intensity distributions from the sample versus incident beam artifacts becomes non-trivial.

To obtain quantitative information from the two-dimensional diffraction patterns, the intensity along y'' can be integrated at each x'' position. The resultant intensity vs. x'' plot is equivalent to a traditional *detector scan* across the diffraction peak, where the slit width is given by the pixel size (or the effective pixel size corrected by the point spread function of a non-ideal CCD detector). In Fig. 2.5c the results of this integration, with the slit width set to 13 μm (1 pixel), is shown

for both experimental and simulated data. While the resolution of the experimental curve is lower than the theoretical one, there is good agreement between the two curves. We note that, in case of the experimental data, the spurious intensities within the shadow of the central stop are excluded from the integration.

For analysis of peak position and peak shape, the spatial coordinate x'' must be transformed into angular coordinates, with $\Delta 2\theta \approx x''/l_{SD}$. The angular coordinates obtained in this manner are marked on the top abscissa of Fig. 2.5c. For the experimental intensity vs. $\Delta 2\theta$ data, the Scherrer equation (2.7) yields a film thickness of 220 nm ± 6 nm, while the film thickness obtained from the period of the thickness fringes is 244 nm ± 8 nm. These values are almost double the actual film thickness measured by cross-sectional TEM. Applying these formalisms to the simulated diffraction profile yields a film thickness of 248 nm for both cases, exactly twice the input sample thickness.

The reason for this discrepancy can be understood by comparing the geometry of a traditional radial scan to the geometry of this experiment. Let us first consider the traditional radial scan with a perfect parallel plane wave incident on the sample (as used to derive Equations (2.7) and (2.8)). For a symmetric reflection at the perfect Bragg condition, both the incident and diffracted wave vectors are inclined to the sample surface by θ_B. If the sample is rotated by an amount $+\Delta\omega$ from the Bragg condition, the angle between the incident beam and the sample surface increases to $\theta_B + \Delta\omega$. At this point, the angle between the wave vector captured by point detector (which has not yet been moved) and the sample surface is $\theta_B - \Delta\omega$. To maintain the symmetric geometry of the radial scan, where the incident and scattered beams make equal angles with the sample surface, the detector angle must be increased by $\mathbf{2}\Delta\omega$. The deviation of the (symmetric) scattered vector from the perfect Bragg condition, with respect to the transmitted beam, is $\Delta 2\theta = 2\Delta\omega$. Now let us consider the geometry of this experiment. We have a divergent beam incident on a stationary sample, with the scattered intensity measured by a CCD (Fig. 2.4). The central axis of the incident beam makes an angle θ_B with respect to the sample surface. For an incident beam divergence $\alpha > \Delta\omega$, there exists a plane wave component of the incident beam $\mathbf{s}_i'$ that makes an angle $\theta_B + \Delta\omega$ with respect to the sample surface. The scattered wave vector $\mathbf{s}_d'$ that satisfies the symmetric scanning condition of the radial scan also makes this angle with respect to the sample surface. The deviation from the Bragg condition (in 2θ) is therefore $\Delta 2\theta = \Delta\omega$, half of the deviation observed in the traditional case using a sample rotation. Thus, any quantity that is

measured relative to a reference value in 2θ, such as peak shifts, peak breadths, or fringe periodicity, are only half as large as would be measured when sample rotations are required. Applying this coordinate correction factor of 2 to the experimental intensity vs. $\Delta 2\theta$ data, the Scherrer equation yields a film thickness of 110 nm $\pm$ 3 nm, while the period of thickness fringes gives 122 nm $\pm$ 4 nm. Analysis of the simulated diffraction profile yields the exact film thickness used in the simulation, 124 nm, for both methods.

In addition to this coordinate correction factor of 2, there is another difference between the traditional radial scan and the one used in this experiment: the incident wave vector recorded at different angular positions are not equivalent. As described above, a traditional radial scan rotates the sample so the *same* plane wave, with the *same* intensity, is incident on sample at different angles θ. However, in the case of this experiment, different angular components of the divergent incident beam are used. It is only because the integrated intensity of the incident beam (along y'') in the region within the FHWM of the diffraction peak is relatively constant that the Scherrer analysis gives a fairly accurate result. For incident beams with non-constant angular intensity distributions, different "Scherrer-like" equations can be derived.

The period of the thickness fringes, on the other hand, does not depend on the absolute intensity and yields the correct film thickness even when aberrations in peak intensity and shape occur. This technique should be preferred over the Scherrer analysis whenever possible. Thickness fringe analysis from small (perfect) domains would be enhanced by the availability of two-dimensional detectors with lower background and higher dynamic range. In our experimental data (Fig. 2.5c), the 2nd set of thickness fringes are barely measurable above background, while the third-order fringes are fully resolved in the simulated pattern. If we plot the simulated profile with an expanded intensity range, even more fringes become visible. This is shown in Fig. 2.5d. If higher resolution detectors enable the acquisition of such images, full-profile fitting of the 1-D compressed images could be employed for data analysis. Such analysis should take into account the presence of complete and incomplete (partially blocked) features.

We conclude that diffraction experiments using coherently focused X-ray beams via phase-retarding zone plates can generate complex scattering profiles even from simple samples. In particular, manufacturing defects or radiation damage in the focusing optic can add features into the diffracted signal that may be misinterpreted as originating from the sample. To obtain quantitative

structural information from such images requires comprehensive full-field physics-based modeling of the relevant wavefields. Simplified one-dimensional models may not be adequate for this purpose.

Chapter 3

Kirchhoff's Laws as a Finite Volume Method for the Planar Maxwell Equations

3.1 Introduction

Maxwell's equations for the (H_1, H_2, E) polarized mode in a planar, inhomogeneous medium with permittivity $\epsilon(x, y)$ and permeability $\mu(x, y)$ are

$$\mu \partial_t \Lambda = -\nabla E, \qquad \Lambda = (-H_2, H_1) \tag{3.1a}$$

$$\epsilon \partial_t E = -\operatorname{div} \Lambda. \tag{3.1b}$$

Let $\Omega = [0, M] \times [0, L]$ be the rectangular region occupied by the medium. Let $\overline{\Gamma} = \{0\} \times [0, L]$ be the left boundary of Ω. Suppose that on $\overline{\Gamma}$, we have harmonic forcing at frequency α:

$$E(0, y, t) = f(0, y) e^{2\pi \imath \alpha t}. \tag{3.1c}$$

On the remaining three sides of the boundary, we impose impedance or Leontovich boundary conditions:

$$\Lambda(x, y, t) \cdot \hat{\mathbf{n}} = \sigma(x, y) E(x, y, t), \quad (x, y) \in \partial\Omega \setminus \overline{\Gamma}, \tag{3.1d}$$

where $\partial\Omega$, $\hat{\mathbf{n}}$, and $\sigma(x, y)$ denote the boundary of Ω, the unit outward normal, and the conductance on the boundary.

In this chapter, we accomplish the following goals:

1. We show that a finite volume discretization of (3.1) results in Kirchhoff's laws of voltage and current for a particular circuit consisting of inductors, capacitors, and resistors.

2. By comparing finite volume solutions of (3.1) for constant ϵ and μ with exact solutions obtained via separation of variables, we numerically establish first-order convergence.

3.1.1 Relationship to previous work

The idea of demonstrating equivalent circuits whose continuum limit yields Maxwell's equations is quite old [Whinnery and Ramo, 1944; Kron, 1944; Whinnery *et al.*, 1944; Brillouin, 1946]. These early works predate the widespread use of digital computers to solve differential/integral equations.

Since then, when a new numerical method for Maxwell's equations has been introduced, the corresponding equivalent circuit has been explored, often as a way to gain physical insight useful for modeling purposes [Christopoulos, 2006, Chap. 1]. One of the first papers proposing an equivalent circuit for the FDTD discretization was [Gwarek, 1985]. Equivalent circuits for the Finite Element Method and the Method of Moments have been described in [Guillouard *et al.*, 1999] and [Felsen *et al.*, 2008, Chap. 5], respectively. For the Transmission Line Matrix method [Hoefer, 1985; Christopoulos, 1995; Christopoulos, 2006] and the Spatial Network Method [Ko *et al.*, 1990; Satoh *et al.*, 2006], equivalent circuits feature prominently.

The finite volume (FV) method appeared in computational electromagnetics in the late 1980s [Madsen and Ziolkowski, 1988; Shankar *et al.*, 1989; Madsen and Ziolkowski, 1990; Shankar *et al.*, 1990]. More recent work indicates that FV methods may hold an advantage over other methods for problems with large variations in the material parameters and sub-grid scale variations in the fields [Lager *et al.*, 2003; Fumeaux *et al.*, 2004; Baumann *et al.*, 2005; Krohne *et al.*, 2007]. Note that the convergence of at least two versions of the FV method has been proven rigorously [Chung *et al.*, 2003; Chung and Engquist, 2005; Hermeline, 2004; Hermeline *et al.*, 2008].

Despite the fact that the FV method has been employed successfully for over 20 years, and unlike the situation for any of the other popular methods for solving Maxwell's equations, there has to date been no discussion in the literature of an equivalent circuit for the FV discretization. We find two main benefits of carefully deriving an equivalent circuit formulation of the FV discretization.

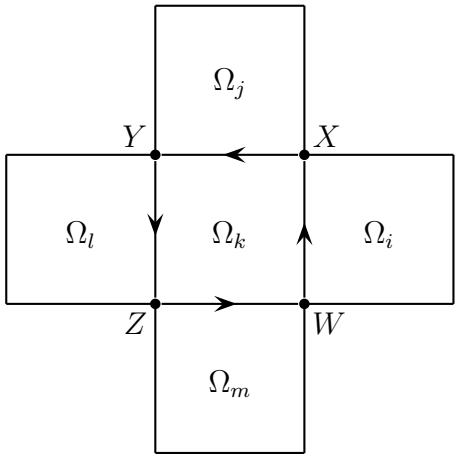


Figure 3.1: Cell diagram of finite volume discretization for an interior cell Ω_k.

First, we obtain precise formulas that relate the local inductance, capacitance, and boundary conductance of the circuit to *spatial averages* of their continuum counterparts: $\mu(x, y)$, $\epsilon(x, y)$ and $\sigma(x, y)$, respectively. Second, relating the FV discretization to Kirchhoff's laws for a circuit automatically yields local energy and charge conservation, in addition to global energy and charge functionals that are natural discretizations of the continuum energy and charge functionals for Maxwell's equations.

From the point of view of using FV to analyze a two-dimensional case of Maxwell's equations, our work is most similar to [Edelvik, 1999]. We solve (3.1) in steady-state for an arbitrary frequency $2\pi\alpha$ in (3.1c); this amounts to finding the frequency-domain solution, which is exactly the goal of the frequency-domain FV method proposed in [Krohne *et al.*, 2007]. In the present work, we are not concerned with issues related to unstructured, adaptive, or hybridized meshes [Gedney *et al.*, 1998; Wang *et al.*, 2002; Abenius *et al.*, 2002], though we note here that our derivation can be generalized in this direction.

3.2 Finite volume discretization of planar Maxwell's equations

In this section, we derive a finite volume method discretization for (3.1). We tile our rectangular domain Ω with small square cells, m in the vertical direction and n in the horizontal direction. We

define the *charge* Q_k in the cell Ω_k:

$$Q_k(t) := \int_{\Omega_k} \epsilon(x,y)E(x,y,t)\,\mathrm{d}\mathbf{x}. \tag{3.2}$$

We define the *capacitance* C_k to be the scaled permittivity in the cell Ω_k:

$$C_k := \frac{1}{\eta}\int_{\Omega_k} \epsilon(x,y)\,\mathrm{d}\mathbf{x}, \tag{3.3}$$

where $\eta > 0$ is a characteristic length scale in the out-of-plane direction. Next, we define the *voltage* $V_k(t) = Q_k(t)/C_k$, so

$$C_k\dot{V}_k = \int_{\Omega_k} \partial_t(\epsilon E)\,\mathrm{d}\mathbf{x} = -\int_{\Omega_k} \operatorname{div}\Lambda\,\mathrm{d}\mathbf{x} = -\oint_{\partial\Omega_k} \Lambda\cdot\mathbf{n}\,\mathrm{d}\ell. \tag{3.4}$$

Let us assume for a moment that Ω_k is an interior cell, so that $\partial\Omega_k$ has no intersection with $\partial\Omega$. As shown in Figure 3.1, Ω_k has four neighboring cells labeled Ω_i, Ω_j, Ω_l, and Ω_m, the right, up, left, and down neighbors, respectively. We label the four corners of Ω_k as W, X, Y, and Z. With this notation,

$$C_k\dot{V}_k = \int_{\overrightarrow{WX}} H_2\,\mathrm{d}\ell + \int_{\overrightarrow{XY}} (-H_1)\,\mathrm{d}\ell + \int_{\overrightarrow{YZ}} (-H_2)\,\mathrm{d}\ell + \int_{\overrightarrow{ZW}} H_1\,\mathrm{d}\ell. \tag{3.5}$$

Now let us define the currents. In general, when we have two neighboring cells Ω_k and Ω_i that are separated by a vertical segment γ, if Ω_k is to the left of Ω_i, then we define the *horizontal current*

$$I_{k,i} := \int_\gamma (-H_2)\,\mathrm{d}\ell. \tag{3.6}$$

In general, when we have two neighboring cells Ω_k and Ω_j that are separated by a horizontal segment γ, if Ω_k is below Ω_j, then we define the *vertical current*

$$I_{k,j} := \int_\gamma H_1\,\mathrm{d}\ell. \tag{3.7}$$

Note that the right-hand sides of (3.5), (3.6), and (3.7) all involve line integrals of scalar fields, which are all independent of parametrization.

With these conventions, we have

$$C_k\dot{V}_k = -I_{k,i} - I_{k,j} + I_{l,k} + I_{m,k}. \tag{3.8}$$

Note that this equation says that at each lattice node, the sum of incoming currents must equal the sum of outgoing currents, implying local charge conservation.

We define the *inductance* of the segment γ as

$$L_\gamma := \frac{\eta}{|\gamma|}\int_\gamma \mu \,\mathrm{d}\ell. \tag{3.9}$$

Let us check how the currents evolve in time. We compute

$$\dot{I}_{k,i} = -\int_\gamma \frac{\partial_x E}{\mu}\,\mathrm{d}\ell \approx -\frac{\eta}{L_\gamma}\int_\gamma \partial_x E\,\mathrm{d}\ell \tag{3.10a}$$

$$\approx -\frac{\eta}{L_\gamma}\left(|\Omega_i|^{-1}\int_{\Omega_i} E\,\mathrm{d}\mathbf{x} - |\Omega_k|^{-1}\int_{\Omega_k} E\,\mathrm{d}\mathbf{x}\right) \tag{3.10b}$$

$$\approx \frac{1}{L_\gamma}(V_k - V_i). \tag{3.10c}$$

Let us explain the sequence of approximations made above:

- In (3.10a), we replace μ by its segment average L_γ/η.
- To approximate the flux between cells in (3.10a), the integral $\int_\gamma \partial_x E\,\mathrm{d}\ell$ is approximated by the value of $\partial_x E$ evaluated at the midpoint of γ, a second-order accurate finite-difference formula is applied using the values of E at the center of the cells Ω_q and Ω_p, and then these values are replaced by the cell averages. This is the main finite volume approximation [LeVeque, 2002].
- To go from (3.10b) to (3.10c), we replace ϵ by its cell average, which gives us

$$V_k = \frac{Q_k}{C_k} = \frac{\int_{\Omega_k}\epsilon E\,\mathrm{d}\mathbf{x}}{\eta^{-1}\int_{\Omega_k}\epsilon\,\mathrm{d}\mathbf{x}} \approx \frac{\eta}{|\Omega_k|}\int_{\Omega_k} E\,\mathrm{d}\mathbf{x}. \tag{3.11}$$

Using analogous approximations, we compute

$$\dot{I}_{k,j} \approx \frac{1}{L_\gamma}(V_k - V_j). \tag{3.12}$$

For an interior cell Ω_k, (3.8), (3.10), and (3.12) are Kirchhoff's laws of voltage and current for a regular square lattice of inductors and capacitors (See chapters 4 and 5 and [Bhat and Osting, 2009a; Bhat and Osting, 2010]).

3.2.1 Boundary conditions

To handle boundary condition (3.1c) on $\overline{\Gamma}$, we use a column of ghost cells. Each ghost cell, where the electric field is prescribed, is directly to the left of a cell in the first column, where the electric

field is an unknown. Let $\gamma \subset \overline{\Gamma}$ be the right boundary of the ghost cell. We compute the voltage of the ghost cell using (3.11):

$$V_k \approx \frac{\eta}{|\Omega_k|}\int_{\Omega_k} E\,\mathrm{d}\mathbf{x} \approx \frac{\eta}{|\gamma|}\int_{\gamma} E\,\mathrm{d}\ell = \tilde{V}_k e^{2\pi\imath\alpha t}, \tag{3.13}$$

with

$$\tilde{V}_k = \eta\left(\frac{1}{|\gamma|}\int_{\gamma} f(y)\,dy\right).$$

The ghost cells yield n new horizontal currents $I_{p,q}$, each of which satisfies an equation of the form (3.10). Each such equation involves one unknown and one prescribed voltage.

If the top, right, or bottom boundary of Ω_k intersects $\partial\Omega$, then we apply the other boundary condition (3.1d). Let $\gamma = \partial\Omega_k \cap \partial\Omega$. Then going back to (3.4), we find

$$\begin{aligned} C_k\dot{V}_k &= -\int_{\gamma} \Lambda\cdot\mathbf{n}\,\mathrm{d}\ell - \int_{\partial\Omega_k\setminus\gamma} \Lambda\cdot\mathbf{n}\,\mathrm{d}\ell \\ &= -\int_{\gamma} \sigma E\,\mathrm{d}\ell - \int_{\partial\Omega_k\setminus\gamma} \Lambda\cdot\mathbf{n}\,\mathrm{d}\ell \end{aligned} \tag{3.14}$$

The second line integral in (3.14) can be evaluated in the same way as (3.5) above; we focus on the first line integral. We write

$$\int_{\gamma}\sigma E\,\mathrm{d}\ell \approx \left(\frac{1}{|\gamma|}\int_{\gamma} E\,\mathrm{d}\ell\right)\left(\int_{\gamma}\sigma\,\mathrm{d}\ell\right) \approx \left(\frac{\eta}{|\Omega_k|}\int_{\Omega_k} E\,\mathrm{d}\mathbf{x}\right)\left(\frac{1}{\eta}\int_{\gamma}\sigma\,\mathrm{d}\ell\right) \approx V_k G_k,$$

where G_k is the *conductance*

$$G_k := \frac{1}{\eta}\int_{\gamma}\sigma\,\mathrm{d}\ell. \tag{3.15}$$

Note that if $\sigma = 0$, then (3.1d) and (3.1a) imply $\nabla E\cdot\hat{\mathbf{n}} = 0$, a perfectly insulating boundary condition. On the other hand, if $\sigma = \infty$, then (3.1d) implies $E = 0$, a perfectly conducting boundary condition. In this chapter, we choose σ to approximate outgoing boundary conditions, which are obtained as follows.

Dotting (3.1a) with $\hat{\mathbf{n}}$ and then using (3.1d), we find that on $\partial\Omega\setminus\overline{\Gamma}$,

$$\partial_t E + \frac{1}{\mu\sigma}\nabla E\cdot\mathbf{n} = 0.$$

At each $(x,y)\in\partial\Omega\setminus\overline{\Gamma}$, the value of $\sigma(x,y)$ for which this equation is the Engquist-Majda outgoing condition [Engquist and Majda, 1977] is

$$\sigma(x,y) = \sqrt{\epsilon(x,y)/\mu(x,y)}. \tag{3.16}$$

Remark. In Appendix 3.A, we show that the (H_1, H_2, E) polarized mode described by (3.1) is an exact solution of the fully three-dimensional Maxwell's equations for a physical system described by two horizontally infinite parallel plates that are separated vertically by the distance $\eta > 0$. All the definitions made above (*e.g.*, charge, capacitance, resistance, etc.) can be derived in a physically consistent fashion using the setup in Appendix 3.A. One may also make these definitions on the grounds that the quantities being derived have the correct units.

3.2.2 Assembling the discretized system

Discretization gives us a two-dimensional rectangular lattice with m rows and n columns, which we represent as an oriented graph, c.f. [Foulds, 1992, Chap. 13]. This graph is the dual graph of the finite volume mesh as shown in Fig. 3.1. Nodes represent capacitors and edges represent inductors. The direction or orientation of the edge represents the direction of positive current flow through the associated inductor.

In a lattice of size $m \times n$, there are mn nodes and $(2m-1)n$ edges, mn horizontal ones and $(m-1)n$ vertical ones. We let $\mathfrak{N} = \{1, 2, \ldots, mn\}$ denote the set of all nodes, and $\mathfrak{E} = \{1, 2, \ldots, (2m-1)n\}$ denote the set of all edges. Let $\mathbf{C}$ be a vector of size mn such that C_j is the capacitance at node j. Let $\mathbf{L}$ be a vector of size $(2m-1)n$ such that L_j is the inductance at edge j. We partition $\mathbf{L}$ into horizontal and vertical inductances by writing $\mathbf{L} = [\mathbf{L}_h, \mathbf{L}_v]$. At time t, $V_j(t)$ and $I_k(t)$ are, respectively, the voltage across capacitor j and the current through inductor k. By $\mathbf{V}(t)$ and $\mathbf{I}(t)$ we denote the vectors of all voltages and currents, respectively.

Of the horizontal edges, there are m boundary edges that form a subset $\Gamma \subset \mathfrak{E}$, each of which is incident upon only one node and corresponds to a ghost cell to the left of the domain Ω. Specifically, Γ is the left-most column of horizontal edges. All other edges in the graph are incident upon two nodes. In general, we think of an edge as an ordered pair (i_1, i_2), where $i_k \in \mathfrak{N}$. The direction of the edge is given by the ordering of these numbers, so that i_1 is the tail and i_2 is the head of (i_1, i_2). For a boundary edge j that is incident only upon node i, we write $j = (\emptyset, i)$.

We let $\mathfrak{B}$ denote the $|\mathfrak{N}| \times |\mathfrak{E}| = mn \times (2m-1)n$ incidence matrix of the oriented graph for our

circuit. We have

$$\mathfrak{B}_{ij} = \begin{cases} 1 & \text{if } j = (i', i) \text{ for some } i' \in \mathfrak{N} \cup \{\emptyset\} \\ -1 & \text{if } j = (i, i') \text{ for some } i' \in \mathfrak{N} \\ 0 & \text{otherwise.} \end{cases}$$

In addition to the structure described already, the lattice also has resistors and forcing along the boundary. We represent the set of nodes connected to resistors by $\mathfrak{G} \subset \mathfrak{N}$, and let G_i be the conductance of node $i \in \mathfrak{G}$. We then extend G_i by defining $G_i \equiv 0$ for all $i \in \mathfrak{N} \setminus \mathfrak{G}$, so that $\mathbf{G} = (G_1, \ldots, G_{mn})$ is a vector of size $|\mathfrak{N}| = mn$.

Let $N = |\mathfrak{N}| + |\mathfrak{E}| = (3m-1)n$. Then we define the $|\Gamma| \times N = m \times (3m-1)n$ projection matrix P_Γ by $(P_\Gamma)_{ij} = 1$ if $\Gamma_i = j$ and $(P_\Gamma)_{ij} = 0$ otherwise. Note that because $\Gamma_i \in \mathfrak{E}$, the final mn columns of P_Γ are all zero. The forcing applied at edges Γ is

$$\mathbf{W}(t) = P_\Gamma^t \mathbf{f} e^{2\pi\imath\alpha t}.$$

The frequency α is the same α in the boundary condition (3.1c). The vector $\mathbf{f} = (f_1, \ldots, f_m) \in \mathbb{C}^{|\Gamma|}$ is arranged as follows: each edge $i \in \Gamma$ is of the form $(\emptyset, k')$ for some $k' \in \mathfrak{N}$. We set f_i equal to $\tilde{V}_k$ as defined in (3.13), where Ω_k is the ghost cell to the left of cell $\Omega_{k'}$.

The finite volume scheme from the previous section, which we have already noted is equivalent to Kirchhoff's Laws on an inductor-capacitor lattice, can now be written in the following matrix-vector form:

$$\operatorname{diag}(\mathbf{L})\frac{\mathrm{d}\mathbf{I}}{\mathrm{d}t} = -\mathfrak{B}^t\mathbf{V} + \mathbf{W} \tag{3.17a}$$

$$\operatorname{diag}(\mathbf{C})\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = \mathfrak{B}\mathbf{I} - \operatorname{diag}(\mathbf{G})\mathbf{V}. \tag{3.17b}$$

3.2.3 Steady-state solution of the discretized equation

Define $\mathbf{z}(t) = (\mathbf{I}(t), \mathbf{V}(t))$ so for each t, $\mathbf{z}(t) \in \mathbb{C}^N$. Define

$$M = \begin{bmatrix} 0 & -\mathfrak{B}^t \\ \mathfrak{B} & -\operatorname{diag}(\mathbf{G}) \end{bmatrix},$$

Then the system (10.1) can be written in the form

$$\operatorname{diag}(\mathbf{L}, \mathbf{C})\dot{\mathbf{z}}(t) = M\mathbf{z}(t) + P_\Gamma^t \mathbf{f} e^{2\pi\imath\alpha t}. \tag{3.18}$$

Consider the steady-state solution $\mathbf{z}(t) = \mathbf{u}e^{2\pi\imath\alpha t}$. Inserting this into (10.2), we derive

$$\mathbf{u} = (2\pi\imath\alpha\,\mathrm{diag}(\mathbf{L},\mathbf{C}) - M)^{-1}P_\Gamma^t\mathbf{f}. \tag{3.19}$$

3.2.4 Discussion

1. The matrix $(2\pi\imath\alpha\,\mathrm{diag}(\mathbf{L},\mathbf{C}) - M)$ will be invertible if and only if $2\pi\imath\alpha$ is not an eigenvalue of $\mathrm{diag}(\mathbf{L},\mathbf{C})^{-1}M$. Note that if all nodes are resistive, *i.e.*, if $G_k > 0$ for for all $i \in \mathfrak{N}$, then the spectrum of $\mathrm{diag}(\mathbf{L},\mathbf{C})^{-1}M$ has strictly negative real part, implying that (10.3) can be computed for all real α.

2. Using Matlab on a desktop computer with 4GB of RAM, (10.3) can easily be solved for $m, n \leq 400$.

3. We have formulated the circuit as an oriented graph in order to write the equations compactly and take advantage of the graph-theoretic interpretation of the incidence matrix $\mathfrak{B}$, which appears naturally in Kirchhoff's laws. Though we have formulated the problem for an $m \times n$ rectangular lattice, the graph-theoretic framework easily accommodates other topologies.

4. Inserting (3.16) into (3.15), we find that at a boundary node $i \subset \mathfrak{G}$, we have the impedance-matched value of the conductance,

$$G_i = \frac{|\gamma|}{|\Omega_i|^{1/2}}\sqrt{\frac{C_i}{L_j}}, \tag{3.20}$$

where $j \in \mathfrak{E}$ is the edge incident on node i that is normal to the boundary, and γ is the segment that is dual to edge j. In the case where all cells are identical squares, we have $|\gamma| = |\Omega_i|^{1/2}$.

3.3 Conservation properties of the continuous and discrete systems

It is instructive to calculate the time evolution of the total energy for the Maxwell system (3.1):

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\int_\Omega \epsilon|E|^2 + \mu\|\Lambda\|^2\,\mathrm{d}A &= -\Re\int_{\partial\Omega} E^*\Lambda\cdot\mathbf{n}\,\mathrm{d}\ell \\ &= -\Re\int_{\overline{\Gamma}} f(y)e^{-2\pi\imath\alpha t}H_2\,\mathrm{d}\ell - \int_{\partial\Omega\setminus\overline{\Gamma}}\sigma|E|^2\,\mathrm{d}\ell.\end{aligned} \tag{3.21}$$

This says that the rate of change of energy equals the power forced in through the left boundary minus the power dissipated through the other three sides of the medium. It is clear that power is dissipated at a rate proportional to σ.

We also compute the time-evolution of the total charge of the system

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\epsilon E\,\mathrm{d}x = -\int_{\overline{\Gamma}}\Lambda\cdot\mathbf{n}\,\mathrm{d}x - \int_{\partial\Omega\setminus\overline{\Gamma}}\sigma E\,\mathrm{d}x. \tag{3.22}$$

The interpretation of this equation is that the rate of change of charge equals the current entering the domain on the left boundary minus the current exiting the domain on the other three sides. Again, the outgoing current is proportional to σ.

The association of the finite volume discretized system as Kirchhoff's laws for a circuit allows for natural definitions of discrete energy and charge. The rate of change of the total energy of the discrete system can be calculated using (10.1):

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{V}^*\mathrm{diag}(\mathbf{C})\mathbf{V} + \mathbf{I}^*\mathrm{diag}(\mathbf{L})\mathbf{I}\right) = \Re\mathbf{I}^* P_\Gamma^t \mathbf{f} e^{2\pi\imath\alpha t} - \mathbf{V}^*\mathrm{diag}(\mathbf{G})\mathbf{V}.$$

The right hand side, which has the form of power in minus power out, corresponds perfectly with the right hand side of (3.21). The calculation shows that the dynamics of energy for the entire lattice can be understood by observing boundary phenomena only; this implies that, locally, in the interior of the lattice, energy is conserved.

We also compute the time evolution of the total charge of the discrete system (10.1):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{1}^t\mathrm{diag}(\mathbf{C})\mathbf{V} &= \mathbf{1}^t\mathfrak{B}\mathbf{I} - \mathbf{1}^t\mathrm{diag}(\mathbf{G})\mathbf{V}\\ &= \sum_{j\in\Gamma} I_j - \sum_{k\in\mathfrak{G}} G_k V_k.\end{aligned}$$

This has the form of current in minus current out, corresponding perfectly with the right hand side of (3.22).

3.4 Separation of variables solution

In this section, we use separation of variables to develop the exact, steady-state solution of (3.1) for constant ϵ and μ. We begin by assuming harmonic time-dependence of the fields,

$$E(x,y,t) = \tilde{E}(x,y)e^{2\pi\imath\alpha t}, \quad \Lambda(x,y,t) = \tilde{\Lambda}(x,y)e^{2\pi\imath\alpha t},$$

in which case system (3.1) reduces to

$$(\nabla^2 + k^2)\tilde{E} = 0 \tag{3.23a}$$

$$\tilde{E}(0, y) = f(y) \tag{3.23b}$$

$$\frac{\partial \tilde{E}}{\partial n} + \imath k z \tilde{E} = 0 \qquad \text{on } \partial\Omega \setminus \overline{\Gamma}, \tag{3.23c}$$

where $k^2 = \mu\epsilon(2\pi\alpha)^2$ and $z = \sigma\sqrt{\mu/\epsilon}$. We now assume a solution of the form $\tilde{E}(x, y) = \rho(x)\psi(y)$. Inserting this into the Helmholtz equation (3.23a), we split the problem as follows:

$$\frac{\rho''(x)}{\rho(x)} + k^2 = -\frac{\psi''(y)}{\psi(y)} = \lambda. \tag{3.24}$$

This yields a non-selfadjoint problem for a complex eigenfunction $\psi(y)$ and a complex eigenvalue λ:

$$\psi''(y) = -\lambda\psi(y) \tag{3.25a}$$

$$\psi'(L) + \imath k z \psi(L) = 0 \tag{3.25b}$$

$$-\psi'(0) + \imath k z \psi(0) = 0 \tag{3.25c}$$

We say that $\phi(y)$ solves the adjoint problem if it satisfies:

$$\phi''(y) = -\lambda\phi(y) \tag{3.26a}$$

$$\phi'(L) - \imath k z \phi(L) = 0 \tag{3.26b}$$

$$-\phi'(0) - \imath k z \phi(0) = 0 \tag{3.26c}$$

We list without proof the properties of the eigenvalue problem that are most relevant to developing a separation of variables solution. For details, refer to [Coddington and Levinson, 1955; Cohen, 1964].

1. (3.25) is not a Sturm-Liouville problem because the boundary conditions are not self-adjoint.

2. If the eigenpair (λ_1, ψ) solves (3.25), the eigenpair (λ_2, ϕ) solves (9.61), and $\lambda_1 \neq \overline{\lambda_2}$, then ψ and ϕ are orthogonal with respect to the L^2 inner product:

$$\langle \psi, \phi \rangle := \int_0^L \psi(y)\overline{\phi(y)}\, dy = 0.$$

3. If the eigenpair (λ, ψ) solves (3.25), then the eigenpair $(\overline{\lambda}, \overline{\psi})$ solves (9.61). In this case, $\langle \psi, \overline{\psi} \rangle \neq 0$.

4. The eigenvalues are discrete, simple, and live in the first quadrant of $\mathbb{C}$.

5. The set $\{\psi_n\}_{n=1}^{\infty}$ is a complete basis of $L^2([0, L])$.

6. As $n \uparrow \infty$, the eigenfunctions ψ_n are increasingly oscillatory and alternatingly even and odd about $y = L/2$.

Note that

$$\psi_n(y) = e^{\imath\sqrt{\lambda_n}y} + \frac{\sqrt{\lambda_n} - kz}{\sqrt{\lambda_n} + kz} e^{-\imath\sqrt{\lambda_n}y} \tag{3.27}$$

is an eigenfunction of (3.25) as long as λ_n solves the transcendental equation

$$e^{2\imath\sqrt{\lambda}L} = \left(\frac{\sqrt{\lambda} - kz}{\sqrt{\lambda} + kz} \right)^2. \tag{3.28}$$

Using this and the above properties, we can derive the solution of (3.23). We expand the left-hand side boundary condition via

$$f(y) = \sum_{n=1}^{\infty} c_n \psi_n(y).$$

Taking inner products, we find

$$c_m = \frac{\langle f, \phi_m \rangle}{\langle \psi_m, \phi_m \rangle}.$$

We return to (3.24) and see that $\rho(x)$ must satisfy

$$\rho_n''(x) = (\lambda_n - k^2)\rho_n(x) \tag{3.29a}$$

$$\rho_n(0) = 1 \tag{3.29b}$$

$$\rho_n'(M) + \imath kz \rho_n(M) = 0 \tag{3.29c}$$

The solution of (3.29) is

$$\rho_n(x) = \frac{q}{q+1} e^{\sqrt{\lambda_n - k^2}x} + \frac{1}{q+1} e^{-\sqrt{\lambda_n - k^2}x},$$

where

$$q = \frac{\sqrt{\lambda_n - k^2} - \imath kz}{\sqrt{\lambda_n - k^2} + \imath kz} e^{-2\sqrt{\lambda_n - k^2}M}.$$

The solution of (3.23) is then

$$\tilde{E}(x, y) = \sum_{n=1}^{\infty} c_n \rho_n(x) \psi_n(y). \tag{3.30}$$

3.4.1 Solving (3.28) for the eigenvalues

Let $\sqrt{\lambda} = a + \imath b$. Taking the square root of both sides of (3.28) and then splitting the resulting equation into its real and imaginary parts leads us to the fixed point iteration scheme

$$\begin{pmatrix} a_{j+1} \\ b_{j+1} \end{pmatrix} = F_n \begin{pmatrix} a_j \\ b_j \end{pmatrix},$$

where

$$F_n \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \dfrac{n\pi}{L} + \dfrac{1}{L}\tan^{-1}\dfrac{2bkz}{a^2+b^2-k^2z^2} \\ -\dfrac{1}{2L}\log\dfrac{(a^2+b^2-k^2z^2)^2+(2bkz)^2}{((a+kz)^2+b^2)^2} \end{pmatrix}.$$

Let D be the disc $\{w \in \mathbb{C} \mid |w-(kz)^2| < 2kz/L\}$. If the eigenvalue λ_n satisfies $\lambda_n \notin D$, it can be shown using the contraction mapping principle that F_n has a unique fixed point (a,b) where $aL/\pi \in (n-1/2, n+1/2)$. In practice, we find that this means that applying F_n for $1 \le n \le N$, one obtains all eigenvalues with real parts in the interval $(0, ((N+1/2)\pi/L)^2)$ except possibly for one eigenvalue that can be found by applying Newton's method in the disc D.

The eigenvalues found in this way constitute the full spectrum of (3.25). Note that as $n \uparrow \infty$, the eigenvalues have the asymptotic form

$$\sqrt{\lambda_n} \sim \frac{n\pi}{L} + \imath\frac{2kz}{n\pi}. \tag{3.31}$$

3.4.2 The transfer function on the rectangle

We define the transfer function $T(f)$ to be the mapping from the left boundary condition $f(y)$ to the solution $\tilde{E}(M,y)$ on the right boundary, *i.e.*,

$$T(f) = \sum_{n=1}^{\infty} \frac{\langle \overline{\psi}_n, f\rangle}{\langle \overline{\psi}_n, \psi_n\rangle} \rho_n(x=M)\psi_n(y). \tag{3.32}$$

Let $\mu_n = \sqrt{\lambda_n - k^2}$. Using (3.30), we derive

$$\rho_n(M) = \frac{2\mu_n e^{-\mu_n M}}{(\mu_n + \imath kz) + (\mu_n - \imath kz)e^{-2\mu_n M}}. \tag{3.33}$$

Combining this with (3.31), we see that for large n,

$$|\rho_n(M)| \sim e^{-n\pi M/L}. \tag{3.34}$$

Since $\langle \overline{\psi}_m, T(f)\rangle = \langle \overline{\psi}_m, f\rangle \rho_n(M)$, the upshot of (3.34) is that the transfer function (3.32) does not conserve energy, since the large n modes of $f(y)$ are severely damped. Also, the solution on the right boundary will be much smoother than $f(y)$.

3.5 Numerical implementation and convergence

In this section, we discuss the application of the finite volume method to two test problems. Throughout the finite volume solution, we set $\eta = 1$. We use V to denote the components of the finite volume solution u that represent voltages at lattice nodes.

3.5.1 Homogeneous medium

For the separation of variables solution, we use (3.30), truncated at $n = 50$ modes, to produce a function $\tilde{E}(x, y)$. When we compare $\tilde{E}(x, y)$ against V_k, we average $\tilde{E}(x, y)$ over the cell Ω_k, following (3.11)—we denote the the averaged separation of variables solution by $\overline{E}$.

We focus on Gaussian boundary data $f(y) = e^{-a(y-1/2)^2}$ with $a = 150$ on the square domain with $M = L = 1$. We set $\epsilon = 9$ and $\mu = 1$. The finite volume solution of this problem at $\alpha = 1.9$ on a 400×400 lattice is given in the upper-left panel of Fig. 3.2.

For four different values of α, we compare the separation of variables solution to the finite volume solution V_m on an $m \times m$ lattice where $m = 20$, 32, 40, 64, 80, 100, 160, 200, 320, 400, and 800. The lower-left panel of Fig. 3.2 shows a log-log plot of the L^2 error $\|V_m - \overline{E}\|_2$ versus m. When α equals 0.25, 0.5, 1.0, and 1.9, the least squares fit to the data gives slopes of, respectively, -1.10, -1.09, -1.27, and -1.36, indicating first-order convergence.

3.5.2 Periodic medium with a linear defect

We now consider a medium, modeled after a photonic crystal device [Joannopoulos *et al.*, 2008], that consists of a periodic array of low index circular inclusions with a linear defect. The permittivity outside the inclusions is $\epsilon = 9$ and inside, $\epsilon = 1$. The domain is the square with $M = L = 1$. The distance between the centers of the circles is $1/10$ and each circle has radius $1/40$. The linear defect is created by simply removing a row of inclusions. The finite volume solution of this problem at $\alpha = 1.9$ on a 400×400 lattice is given in the upper-right panel of Fig. 3.2. As expected, the mode

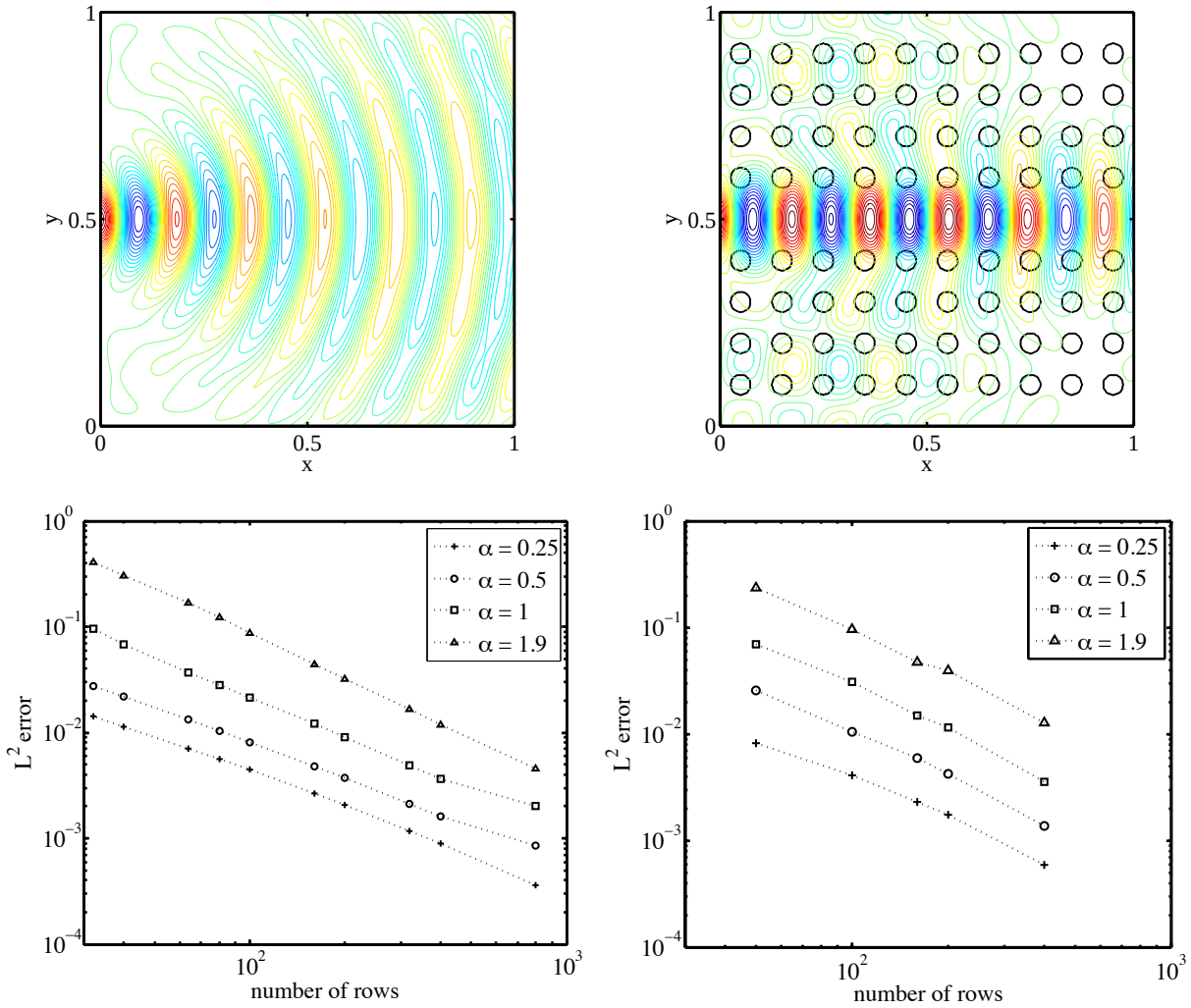


Figure 3.2: The left panels show results from simulations of a homogeneous medium with $\epsilon = 9$ and $\mu = 1$ as in Sec. 3.5.1. The right panels show results from simulations of an inhomogeneous medium as in Sec. 3.5.2. The black circles in the upper-right panel mark the parts of the medium within which $\epsilon = 1$; otherwise, $\epsilon = 9$. Again, $\mu = 1$ everywhere. The same Gaussian boundary forcing at angular frequency $\alpha = 1.9$ is applied in both problems. The numerically computed real part of the electric field is indicated by the color contour plots. Note that the effect of the periodic array with linear defect is to confine the electromagnetic field preventing it from diffracting into the rest of the domain as in the homogeneous medium. The lower panels show the results of a convergence study in the form of log-log plots of the L^2-error between numerically computed solutions and a reference solution, as a function of the number of lattice rows. In the lower-left panel, the reference solution is a 50-mode truncation of the exact solution, while in the lower-right panel, the reference solution is the finite volume solution on an 800×800 lattice. The lower panels both show four curves, one for each indicated value of the angular frequency α. All eight curves have best-fit slope less than -1, indicating first-order convergence.

is confined to the defect, rather than diffracting as in the homogeneous medium.

We study the convergence of the finite volume method for this inhomogeneous medium by first obtaining a fine-scale finite volume solution V_{fine} on an 800×800 lattice. For the four values of α mentioned above, we compare this solution to the finite volume solution V_m on an $m \times m$ lattice for $m = 50, 100, 200, 400$. A log-log plot of $\|V_m - V_{\text{fine}}\|_2$ versus m is given in the lower-right panel of Fig. 3.2. A least squares fit of the error gives slopes of, respectively, -1.25, -1.39, -1.42, and -1.39, indicating first-order convergence.

3.6 Conclusion

We have derived a physically motivated finite volume method for a planar Maxwell system. The method is easy to implement. Here we have done so to obtain the frequency domain solution of two problems with harmonic time-dependence. However, the system (10.1) obtained after spatial discretization could be used with a time-stepping scheme to solve an initial value problem in the time domain. Note that to find a steady-state solution through time stepping that is as accurate as the solutions we obtained, one would require a temporal discretization with first-order global error. This typically means that the time-stepping scheme must be at least second-order accurate.

To demonstrate convergence, we compared numerical solutions with a separation of variables solution for constant ϵ and μ. Note that it is possible to generalize the separation of variables solution in Sec. 3.4 to handle separable ϵ and μ.

The choice of discretization in Sec. 3.2 does not require smoothness of ϵ and μ. In other words, an advantage of the first-order method proposed here is that discontinuous material parameters can be handled readily. A goal for future work is to extensively test how roughness and/or short-wavelength oscillations in the coefficients ϵ and μ affect the performance of the finite volume method, and to compare the finite volume method to other frequency-domain methods for such problems.

3.A An idealized physical configuration

In this section, we describe an idealized physical configuration in which the (H_1, H_2, E) polarized mode is an exact solution of Maxwell's equations and interpret the finite volume method derived in Sec. 3.2 in this context. The idea of formulating more systematic relationships between circuit-

and field-theoretic concepts stems from [Adler *et al.*, 1960].

We consider two perfectly conducting plates that are infinite in extent in the $\hat{x}$ and $\hat{y}$ directions and separated by a distance $\eta > 0$ in the $\hat{z}$ direction. Between the plates is a medium with parameters ϵ and μ that may vary in the $\hat{x}$ and $\hat{y}$ directions, but are constant in the $\hat{z}$ direction. Between the plates, the electric and magnetic fields satisfy Maxwell's equations with no free charge or currents. The boundary conditions on the plates are:

$$\hat{n} \times \vec{E} = 0, \qquad \hat{n} \cdot \vec{B} = 0 \tag{3.35a}$$

$$\hat{n} \cdot \vec{D} = \rho_s, \qquad \hat{n} \times \vec{H} = \vec{j}_s. \tag{3.35b}$$

where ρ_s and $\vec{j}_s$ are the surface charge and surface current. Here, $\hat{n} = \hat{z}$ for the upper surface and $\hat{n} = -\hat{z}$ for the lower surface. The $\big(H_1(x,y,t), H_2(x,y,t), E(x,y,t)\big)$ polarized mode automatically satisfies (3.35a). The last two boundary conditions reduce to

$$\epsilon E = \rho_s, \qquad \Lambda = (j_s^1, j_s^2),$$

where, as before, $\Lambda = (-H_2, H_1)$. Evaluating the line integral of $\nabla \cdot (\epsilon E)$ connecting $(x, y, 0)$ and (x, y, η), we find that the charge density on the two plates at fixed (x, y) are equal in magnitude but have opposite signs. We now identify the charge $Q_k(t)$, defined by (3.2), with the area integral over Ω_k of the surface charge on the top plate. For constant ϵ, (3.3) agrees with the capacitance between two parallel plates of area $|\Omega_k|$ separated by a distance η: $\epsilon|\Omega_k|/\eta$. The electrostatic potential difference between the two plates at position (x, y) can be defined by

$$V(x,y,t) = \int_0^\eta E(x,y,t)\,\mathrm{d}z = \eta E(x,y,t).$$

The approximation of the quantity $V_k \equiv Q_k/C_k$ in (3.11) is precisely the average value $V(x, y, t)$ on Ω_k. Thus the approximation made in (3.11) can be interpreted as an electrostatic approximation.

Continuity of charge requires that for any rectangular region Ω_k on the top conducting plate, we must have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega_k} \rho_s(x,y)\,\mathrm{d}\vec{x} = -\oint_{\partial\Omega_k} j_s\,\mathrm{d}\ell = -\oint_{\partial\Omega_k} \Lambda\,\mathrm{d}\ell.$$

Thus the line integral of the surface current over one segment of $\partial\Omega_k$ is equivalent to the current defined by (3.7), and the continuity equation is equivalent to Kirchhoff's law (3.8).

Suppose there is a surface current between two cells in the $\hat{x}$ direction. This induces a magnetic field in the $\hat{y}$ direction just below the top plate. If the current increases (resp. decreases),

then the field will also increase (resp. decrease). By Faraday's law of induction, this increasing (resp. decreasing) field will induce an electromotive force in the $\hat{x}$ direction (resp. $-\hat{x}$ direction) proportional to μ. This is Kirchhoff's law (3.12).

Chapter 4

Diffraction on the two-dimensional square lattice

4.1 Introduction

Consider the discrete wave equation on the semi-infinite square lattice with a Dirichlet condition along the line $i = 0$:

$$\frac{d^2}{dt^2} u_{i,j} = c^2 \left(\Delta_d u\right)_{i,j}, \ i \geq 1 \tag{4.1a}$$

$$u_{0,j}(t) = f_j e^{\imath\omega t}. \tag{4.1b}$$

Here f_j is supported only for $j \in \Sigma = [-A, A]$ where $A > 0$, and Δ_d is the discrete Laplacian operator defined by

$$(\Delta_d \phi)_{ij} = \phi_{i,j+1} + \phi_{i,j-1} + \phi_{i+1,j} + \phi_{i-1,j} - 4\phi_{i,j}. \tag{4.2}$$

As a motivating experiment, we numerically solve (4.1) with initial conditions $u_{i,j}(0) = 0$ and $\frac{d}{dt} u_{i,j}(0) = 0$. We take a 400×400 lattice ($1 \leq i \leq 400$, $-199 \leq j \leq 200$), $A = 10$, constant data $f_j = 1$, speed $c = 1$, and the following successively larger frequencies: $\omega = \sqrt{1/2}$, $\omega = \sqrt{2}$, $\omega = \sqrt{11/4}$, and $\omega = \sqrt{7/2}$. In each case, we step forward in time from $t = 0$ until some $t = T > 0$, and then plot $|u_{100,j}(T)|$ for $-100 \leq j \leq 100$. Further details are given in Appendix 4.A. The results of the numerical experiment, plotted in red in Figure 4.1, show the diffraction of the spatially discrete waves that propagate from the aperture $\{0\} \times [-A, A]$ into the lattice.

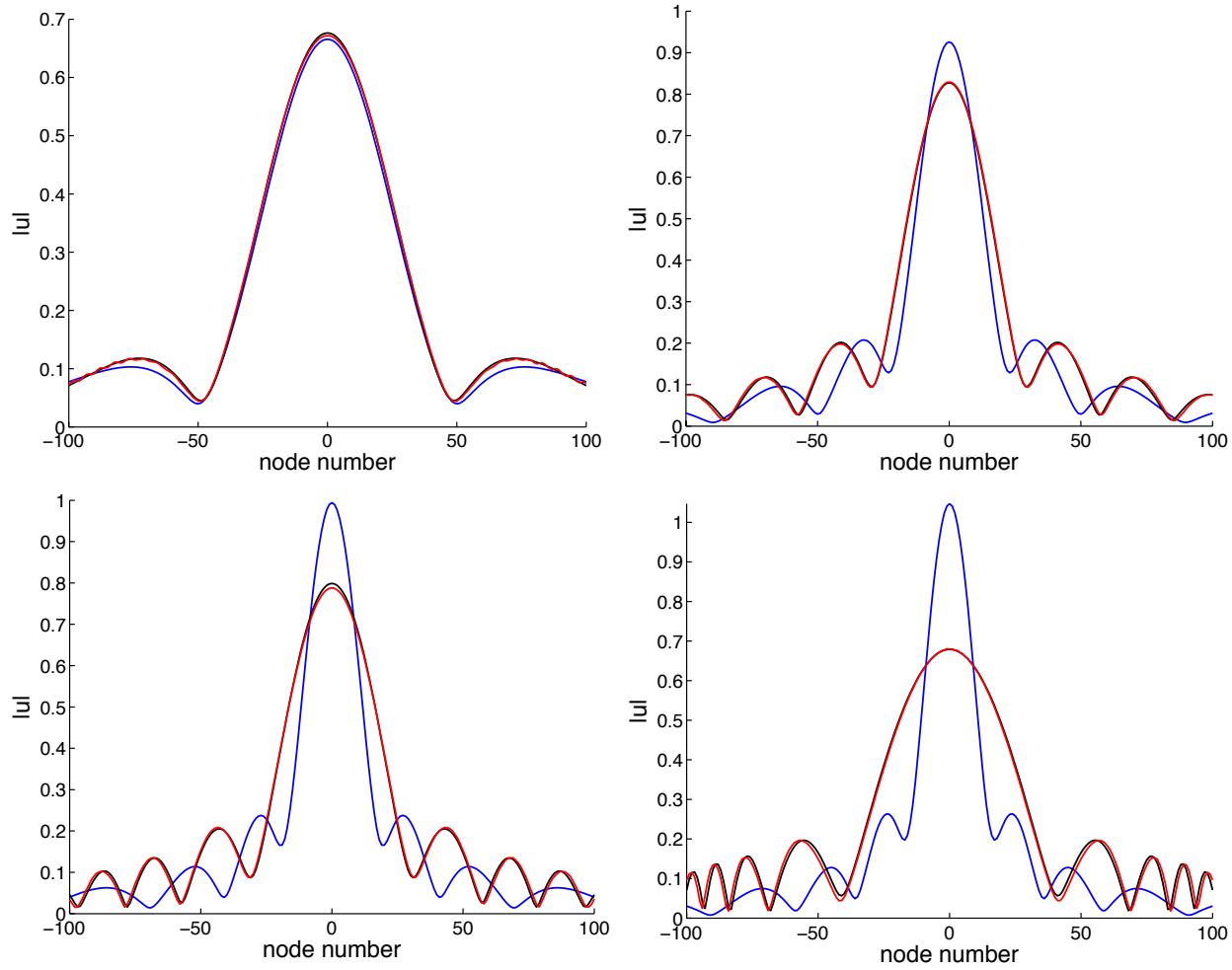


Figure 4.1: From left to right, top to bottom, we have diffraction patterns for frequencies $\omega = \sqrt{1/2}$, $\omega = \sqrt{2}$, $\omega = \sqrt{11/4}$, and $\omega = \sqrt{7/2}$. Results from the numerical experiment described in the text (time-stepping the discrete wave equation) are plotted in red. In blue we plot the results of continuum Rayleigh-Sommerfeld diffraction theory, and in black we plot the results of our discrete diffraction theory. Note that continuum diffraction theory serves as a useful approximation of the red curve only for the smallest value of ω, while the discrete diffraction theory closely tracks the red curve for all values of ω shown.

The main result of this chapter, which is what we used to plot the black curves in Figure 4.1, is a physical derivation of the exact solution—see (4.28) and Theorem 4.2.1 for details—of the discrete Helmholtz equation on the semi-infinite lattice with a Dirichlet condition on the left boundary:

$$\left(\Delta_d + k^2\right) U_{ij} = 0, \ i \geq 1, \tag{4.3a}$$

$$U_{0j} = f_j. \tag{4.3b}$$

Here f_j is supported only for $j \in \Sigma$ for a finite set of integers Σ, and U is required to satisfy an outgoing boundary condition specified in Section 4.2.7. System (4.3) is a discrete version of the classical thin-slit diffraction problem; we use discrete versions of classical arguments to derive the solution. As shown in Figure 4.1, the solution of this discrete diffraction problem closely captures the behavior of the numerical experiment for all four values of ω.

We have also plotted in blue the diffraction pattern predicted by standard Rayleigh-Sommerfeld (R-S) theory [Bouwkamp, 1954; Born and Wolf, 1980] for a two-dimensional continuum:

$$U(x,y) = -\frac{kx}{2\imath} \int_{-A}^{A} u_0(\xi) \frac{H_1(kr)}{r} \, d\xi,$$

where H_1 is the Hankel function of the first kind, and r denotes the magnitude of the vector $\mathbf{r}$ that joins (x,y) to $(0,\xi)$, a point on the aperture. As expected, the continuum theory diverges from the numerical experiment as ω increases.

Along the way to the solution, we apply recent asymptotic estimates of the lattice Green's function derived by P. A. Martin [Martin, 2006] to derive a discrete version of the Sommerfeld outgoing radiation condition. Having expressed the solution of (4.3) as a discrete convolution, we describe in detail a nontrivial computation of the convolution kernel. This involves applying different results from the literature to generate useable lattice Green's functions. For the particular wavenumber $k = 2$, we are able to evaluate the convolution kernel in closed form. For $k \neq 2$, the numerical evaluation of the convolution solution of (4.3) is far faster than finding the steady-state solution to (4.1) through time-stepping.

4.1.1 Motivation

We are motivated by two application areas, metamaterials and analog circuits, that are both of recent interest. Of course, other applications exist, including mass-spring lattices, numerical discretizations of the continuum wave equation, and applications in which the discrete Schrödinger

equation arises such as in the tight binding model (TBA) or linear combination of atomic orbitals (LCAO) [Kevrekidis and Porter, 2009; Ablowitz and Zhu, 2010].

4.1.1.1 Left-handed two-dimensional inductor-capacitor metamaterial

Starting from the integer lattice $\mathbb{Z}^2$, suppose there is an inductor connecting each node (i, j) to a common ground plane, and suppose that there is a capacitor connecting each node (i, j) to its four nearest neighbors $(i \pm 1, j \pm 1)$. Assume that all inductances equal L and all capacitances equal C, where both L and C are positive constants. In this case, Kirchhoff's Laws of voltage and current imply the following second-order equation for the voltage V_{ij} across the inductor at node (i, j):

$$LC\frac{d^2}{dt^2}\left(\Delta_d V\right)_{ij} = V_{ij}. \tag{4.4}$$

This equation admits plane wave solutions $V_{ij}(t) = \exp[\imath(\omega t - \phi_1 i - \phi_2 j)]$ as long as the following dispersion relation is satisfied:

$$\omega^2 = \left[4LC\left(\sin^2\frac{\phi_1}{2} + \sin^2\frac{\phi_2}{2}\right)\right]^{-1}. \tag{4.5}$$

For both $m = 1$ and $m = 2$, we see that $\omega/\phi_m > 0$ while $d\omega/d\phi_m < 0$; for this reason, this type of inductor-capacitor lattice is referred to as a left-handed metamaterial [Caloz and Itoh, 2006]. Metamaterials have been of great recent theoretical and experimental interest [Sarychev and Shalaev, 2007; Engheta and Ziolkowski, 2006; Marqués *et al.*, 2008]. Experimental groups have succeeded in fabricating artificial materials for which the dispersion relation is close to (4.5) for specific intervals of frequencies $\omega \in [\omega^-, \omega^+]$—see [Caloz and Nguyen, 2007], for example. This application will be further discussed in Ch. 5.

4.1.1.2 Standard two-dimensional inductor-capacitor lattice

In this system, there is a capacitor connecting each node (i, j) to ground, and inductors connecting each node to its nearest neighbors. Let u_{ij} denote the voltage across the capacitor. Assuming all inductances equal L and all capacitances equal C, Kirchhoff's Laws of voltage and current can be used to derive (4.1a) with $c^2 = (LC)^{-1}$:

$$LC\frac{d^2}{dt^2}u_{ij} = \left(\Delta_d u\right)_{ij}. \tag{4.6}$$

An inductor-capacitor lattice of this type was used to design a high-frequency power amplifier [Afshari *et al.*, 2006a] on chip. Experimental measurements of the chip show that it generates 125 mW of power at 85 GHz, one of the best reported results for any chip on a Silicon substrate [Afshari *et al.*, 2006b]. The same lattice can be used in a different mode of operation as a Fourier transform device [Afshari *et al.*, 2008] or as an electrical prism [Momeni and Afshari, 2008]. Nonlinear versions of this lattice, where the capacitors are voltage-dependent, have been shown to exhibit nonlinear constructive interference [Bhat and Afshari, 2008], which can be used to generate high-power, high-frequency harmonics of input signals [Lilis *et al.*, 2010].

4.1.2 Unified treatment via analysis of the discrete model

For the standard lattice, one is often interested in waves whose wavelength is large compared to the lattice spacing h. In this case, one may take the continuum limit of (4.6). The steps are standard: divide both sides of (4.6) by h^2, define the per-unit-length inductance and capacitance $\hat{L} = L/h$ and $\hat{C} = C/h$, and then take $h \to 0$. The discrete Laplacian Δ_d is replaced by the continuum Laplacian Δ.

Note, however, that one cannot take the continuum limit of (4.4) in the same way—dividing both sides by h^2, defining $\hat{L}$ and $\hat{C}$ in the same way, and taking $h \to 0$ leads to blow-up on the right-hand side. Of course, we could have determined this from the dispersion relation (4.5), because $\omega \to \infty$ in the long-wave limit $\phi_1, \phi_2 \to 0$.

With this in mind, we avoid continuum limits and instead directly analyze the discrete Helmholtz diffraction problem (4.3). This provides a unified treatment of (4.1), valid when (4.1a) is either the standard discrete wave equation (4.6) or the left-handed discrete wave equation (4.4), and also valid for all waves regardless of how their wavelengths compare to the lattice spacing h. Note that

- If we start from (4.1) as it is written, and search for solutions of the form $u_{ij}(t) = e^{\imath\omega t}U_{ij}$, we obtain (4.3) with $k^2 = \omega^2/c^2$, *i.e.*, $k^2 = \omega^2 LC$.
- If we start from (4.1) where (4.1a) is replaced by (4.4) for $i \geq 1$, and search for solutions of the form $V_{ij}(t) = e^{\imath\omega t}V_{ij}$, we obtain (4.3) with $k^2 = (\omega^2 LC)^{-1}$.

This shows that the solution of (4.3) can be used to solve propagation and diffraction problems for both standard and left-handed inductor-capacitor lattices. The solution can also be used to

solve problems for composite right/left-handed (CRLH) [Sanada *et al.*, 2004] and dual-composite right/left-handed (d-CRLH) [Caloz, 2006] lattices, where parallel and/or series LC blocks are used between nodes to bring about realistic dispersion relations that interpolate between (4.5) and the standard dispersion relation of (4.6). We have shown in past work [Bhat and Osting, 2008] that in CRLH and d-CRLH lattices, voltages are governed by fourth-order discrete wave equations that reduce to (4.3a) if one assumes time-harmonic solutions.

Just as continuum diffraction theories may be derived starting from the continuum Helmholtz equation $(\Delta + k^2)U = 0$, our diffraction theory proceeds from (4.3a). Before getting into the derivation, we review prior work on similar problems.

4.1.3 Prior work

The most relevant prior work is that of Shaban and Vainberg [Shaban and Vainberg, 2001]; this paper considers the general propagation problem on a d-dimensional lattice. A detailed analysis proving that the discrete Sommerfeld condition singles out a unique solution of the problem is given. Together with the paper of Islami and Vainberg [Islami and Vainberg, 2006], this paper explains that, for example, as $t \to \infty$, the solution of the time-dependent problem (4.1) approaches $Ue^{i\omega t}$ where U solves (4.3). We should also mention the work of Schultz [Schultz, 1998], who uses pointwise estimates of the Green's function the discrete wave equation to analyze the solution of both linear and nonlinear lattice wave equations in two and three spatial dimensions. Other papers from the numerical analysis literature [Bamberger *et al.*, 1988; Zemla, 1995] analyze problems similar to (4.3) to determine how closely solutions of the discrete problem approximate solutions of the continuum problem.

While analytical considerations are very important, our goals in this paper are different. We provide, firstly, a physical derivation of the solution of the discrete diffraction problem (4.3). The arguments we use are discrete versions of arguments originally put forth by Sommerfeld. Our arguments are constructive and can be generalized to other lattice topologies, including the two-dimensional triangular and hexagonal lattices [Bhat and Osting, 2009b]. Secondly, we write the solution of (4.3) in a way that enables fast and accurate computation of discrete diffraction patterns. Despite their importance for practitioners who use lattices in electromagnetic and circuit applications, these considerations are absent from prior work on this topic. Both [Shaban and Vain-

berg, 2001] and [Islami and Vainberg, 2006] do not describe the actual shape of discrete diffraction patterns and how they differ from continuum diffraction patterns with the same parameters.

4.2 Derivation of the discrete Rayleigh-Sommerfeld theory

We proceed in stages, building up from basic identities to a discrete diffraction formula. The summation by parts and discrete Green's identity have appeared before [Cheng and Lu, 1991; Yang and Albregtsen, 1994; Brlek, 2005]—we include our own derivations to keep this paper self-contained.

4.2.1 Summation by parts

The discrete version of integration by parts is

$$\sum_{k=m}^{n} f_k(g_{k+1} - g_k) = f_n g_{n+1} - f_{m-1} g_m - \sum_{k=m}^{n} g_k(f_k - f_{k-1}), \tag{4.7}$$

also referred to as Abel's Lemma. This also gives

$$\sum_{k=m}^{n} f_k(g_k - g_{k-1}) = f_{n+1} g_n - f_m g_{m-1} - \sum_{k=m}^{n} g_k(f_{k+1} - f_k). \tag{4.8}$$

Let $\partial_d^2 g_k = g_{k+1} - 2g_k + g_{k-1}$ be the discrete one-dimensional second derivative operator. Then subtracting (4.8) from (4.7) yields

$$\sum_{k=m}^{n} f_k \partial_d^2 g_k = f_n g_{n+1} - f_{n+1} g_n + f_m g_{m-1} - f_{m-1} g_m + \sum_{k=m}^{n} g_k \partial_d^2 f_k. \tag{4.9}$$

4.2.2 Green's second identity

We associate the indices i and j with the horizontal and vertical directions, respectively. Let Ω be the rectangular region of the discrete lattice defined by $\Omega = \{(i,j) \,|\, W \le i \le E,\ S \le j \le N\}$ where $W, E, S, N \in \mathbb{Z}$. Then we claim that

$$\begin{aligned} \sum_{ij \in \Omega} U_{ij} \Delta_d V_{ij} - V_{ij} \Delta_d U_{ij} = &\sum_{W \le i \le E} U_{iN} V_{iN+1} - U_{iN+1} V_{iN} + U_{iS} V_{iS-1} - U_{iS-1} V_{iS} \\ &+ \sum_{S \le j \le N} U_{Ej} V_{E+1j} - U_{E+1j} V_{Ej} + U_{Wj} V_{W-1j} - U_{W-1j} V_{Wj} \end{aligned} \tag{4.10}$$

Proof.

$$\begin{aligned}\sum_{ij\in\Omega} U_{ij}\Delta_d V_{ij} - V_{ij}\Delta_d U_{ij} &= \sum_{i=W}^{E}\sum_{j=S}^{N} U_{ij}\Delta_d V_{ij} - V_{ij}\Delta_d U_{ij} \\ &= \sum_{i=W}^{E}\left\{\sum_{j=S}^{N} U_{ij}(V_{ij+1} - 2V_{ij} + V_{ij-1}) - V_{ij}(U_{ij+1} - 2U_{ij} + U_{ij-1})\right\} \\ &\quad + \sum_{j=S}^{N}\left\{\sum_{i=W}^{E} U_{ij}(V_{i+1j} - 2V_{ij} + V_{i-1j}) - V_{ij}(U_{i+1j} - 2U_{ij} + U_{i-1j})\right\}\end{aligned}$$

Now apply (4.9) to each of the inner sums grouped inside curly braces. The result is precisely (4.10). □

4.2.3 Remark

For $(i,j) \in \partial\Omega$, let $\delta_{\hat{\mathbf{n}}}$ denote the discrete derivative in the outward normal direction, so that on the four sides of the rectangle we have:

$$\begin{aligned}\delta_{\hat{\mathbf{n}}}\phi_{ij} &= \phi_{iN+1} - \phi_{iN} && \text{top} \\ \delta_{\hat{\mathbf{n}}}\phi_{ij} &= \phi_{E+1j} - \phi_{Ej} && \text{right} \\ \delta_{\hat{\mathbf{n}}}\phi_{ij} &= \phi_{iS-1} - \phi_{iS} && \text{bottom} \\ \delta_{\hat{\mathbf{n}}}\phi_{ij} &= \phi_{W-1j} - \phi_{Wj} && \text{left}\end{aligned}$$

Let us return to (4.10) and consider the sum along the top side of the rectangle. It is easy to rewrite the sum using $\delta_{\hat{\mathbf{n}}}$:

$$\begin{aligned}\sum_{W\le i\le E} U_{iN}V_{iN+1} - U_{iN+1}V_{iN} &= \sum_{W\le i\le E} U_{iN}\left(V_{iN+1} - V_{iN}\right) - \left(U_{iN+1} - U_{iN}\right)V_{iN} \\ &= \sum_{W\le i\le E} U_{iN}\delta_{\hat{\mathbf{n}}}V_{iN} - V_{iN}\delta_{\hat{\mathbf{n}}}U_{iN}.\end{aligned}$$

Carrying out the same procedure on all four sides, (4.10) can be summarized as

$$\sum_{ij\in\Omega} U_{ij}\Delta_d V_{ij} - V_{ij}\Delta_d U_{ij} = \sum_{ij\in\partial\Omega} U_{ij}\delta_{\hat{\mathbf{n}}}V_{ij} - V_{ij}\delta_{\hat{\mathbf{n}}}U_{ij}, \tag{4.11}$$

which is of precisely the same form as the continuum version of Green's second identity.

4.2.4 Removing one point

Fix $(p,q) \in \Omega$. Let $\overline{U}_{pq}$ denote the average over the neighbors of U_{pq}:

$$\overline{U}_{pq} = \frac{1}{4}\big(U_{pq+1} + U_{pq-1} + U_{p+1q} + U_{p-1q}\big).$$

Then a direct calculation using (4.2) shows that

$$-\left(U_{pq}\Delta_d V_{pq} - V_{pq}\Delta_d U_{pq}\right) = 4\overline{U}_{pq}V_{pq} - 4\overline{V}_{pq}U_{pq}. \tag{4.12}$$

Let $\Omega_0 = \Omega - (p,q)$ be the rectangle with the point (p,q) removed. Then adding (4.12) to (4.11) gives

$$\sum_{ij\in\Omega_0} U_{ij}\Delta_d V_{ij} - V_{ij}\Delta_d U_{ij} = 4\overline{U}_{pq}V_{pq} - 4\overline{V}_{pq}U_{pq} + \sum_{ij\in\partial\Omega} U_{ij}\delta_{\hat{\mathbf{n}}}V_{ij} - V_{ij}\delta_{\hat{\mathbf{n}}}U_{ij}$$

Suppose U and V both satisfy (4.3a) in the region Ω_0. Then

$$\sum_{ij\in\Omega_0} U_{ij}\Delta_d V_{ij} - V_{ij}\Delta_d U_{ij} = 0,$$

so the previous equation reduces to

$$-4\overline{U}_{pq}V_{pq} + 4\overline{V}_{pq}U_{pq} = \sum_{ij\in\partial\Omega} U_{ij}\delta_{\hat{\mathbf{n}}}V_{ij} - V_{ij}\delta_{\hat{\mathbf{n}}}U_{ij}. \tag{4.13}$$

4.2.5 Lattice Green's function

Let $G_{ij;pq}$ be the Green's function for (4.3a) centered at the point (p,q), evaluated at (i,j). By definition, $G_{ij;pq}$ must satisfy

$$(\Delta_d + k^2)G_{ij;pq} = \delta_{ip}\delta_{jq}. \tag{4.14}$$

for all $(i,j) \in \mathbb{Z}^2$. The lattice Green's function $G_{ij;pq}$ that satisfies (4.14) is quite well-known [Katsura and Inawashiro, 1971; Economou, 2006]. Using trigonometric identities, one may write it in the form

$$G_{ij;pq} = \frac{1}{\pi^2}\int_0^\pi \int_0^\pi \frac{\cos[(i-p)\xi]\cos[(j-q)\eta]}{\sigma(\xi,\eta;k)}d\xi d\eta \tag{4.15}$$

with

$$\sigma(\xi,\eta;k) = k^2 - 4\sin^2\frac{1}{2}\xi - 4\sin^2\frac{1}{2}\eta$$

From (4.15), it is evident that the lattice Green's function centered at (p,q) evaluated at (i,j) is the same as the lattice Green's function centered at $(0,0)$ evaluated at $(i-p, j-q)$, *i.e.*,

$$G_{ij;pq} = G_{i-p,j-q;00}.$$

Henceforth we use $G_{i-p,j-q}$ to denote $G_{i-p,j-q;00}$, *i.e.*, if we do not specify otherwise, G denotes the lattice Green's function centered at $(0,0)$.

4.2.6 Diffraction

Assume that U solves (4.3a) in Ω. By definition, $G_{i-p,j-q}$ solves (4.3a) in the punctured rectangle Ω_0. Therefore, in (4.13), we can replace V_{ij} by $G_{i-p,j-q}$. Note that V_{pq} is then replaced by G_{00}. The left-hand side of (4.13) reduces to

$$4U_{pq}\overline{G}_{00} - 4G_{00}\overline{U}_{pq} = U_{pq}\big((4-k^2)G_{00}+1\big) - G_{00}(4-k^2)U_{pq} = U_{pq},$$

so we obtain

$$U_{pq} = \sum_{ij\in\partial\Omega} U_{ij}\delta_{\hat{\mathbf{n}}}G_{i-p,j-q} - G_{i-p,j-q}\delta_{\hat{\mathbf{n}}}U_{ij}.$$

We consider Ω defined by $W=1$, $E=M>0$, $N=M>0$, and $S=-M+1<0$. Then the previous equation can be written $U_{pq} = S_1 + S_2$ where

$$S_1 = \sum_{j=-M+1}^{M} U_{1j}\delta_{\hat{\mathbf{n}}}G_{1-p,j-q} - G_{1-p,j-q}\delta_{\hat{\mathbf{n}}}U_{1j} \tag{4.16}$$

and

$$\begin{aligned} S_2 = &\sum_{i=1}^{M} U_{iM}\delta_{\hat{\mathbf{n}}}G_{i-p,M-q} - G_{i-p,M-q}\delta_{\hat{\mathbf{n}}}U_{iM} \\ &+ \sum_{j=-M+1}^{M} U_{Mj}\delta_{\hat{\mathbf{n}}}G_{M-p,j-q} - G_{M-p,j-q}\delta_{\hat{\mathbf{n}}}U_{Mj} \\ &+ \sum_{i=1}^{M} U_{i,-M+1}\delta_{\hat{\mathbf{n}}}G_{i-p,-M+1-q} - G_{i-p,-M+1-q}\delta_{\hat{\mathbf{n}}}U_{i,-M+1} \end{aligned} \tag{4.17}$$

4.2.7 Discrete Sommerfeld outgoing radiation condition

Our goal now is to show that if U satisfies a discrete Sommerfeld outgoing radiation condition [Shaban and Vainberg, 2001], then $\lim_{M\to\infty} S_2 = 0$. First let us estimate the three sums on the

right-hand side of S_2 by

$$\begin{aligned} S_2 \leq & \max_{i\in[1,M]} M(U_{iM}\delta_{\hat{\mathbf{n}}}G_{i-p,M-q} - G_{i-p,M-q}\delta_{\hat{\mathbf{n}}}U_{iM}) \\ & + \max_{j\in[-M+1,M]} 2M(U_{Mj}\delta_{\hat{\mathbf{n}}}G_{M-p,j-q} - G_{M-p,j-q}\delta_{\hat{\mathbf{n}}}U_{Mj}) \\ & + \max_{i\in[1,M]} M(U_{i,-M+1}\delta_{\hat{\mathbf{n}}}G_{i-p,-M+1-q} - G_{i-p,-M+1-q}\delta_{\hat{\mathbf{n}}}U_{i,-M+1}) \end{aligned} \quad (4.18)$$

Consider a point (m,n) on any of the three sides (top, right, bottom) of the rectangle included in S_2. In polar coordinates centered at $(0,0)$, we have $(m,n) = (R\cos\alpha, R\sin\alpha)$. Along the sides of the rectangle, we see that $R \in [M, \sqrt{2}M]$ and $\alpha \in [-\pi/2, \pi/2]$. Then, using a stationary phase calculation, P. A. Martin obtains the asymptotic form of the Green's function,

$$G_{mn} \sim \frac{e^{\imath(m\xi_0(\alpha,k)+n\eta_0(\alpha,k))}}{\sqrt{2\pi R}} F(\alpha,k) \quad \text{as} \quad R\to\infty,$$

where there are two cases [Martin, 2006]. In the first case, $0 < k^2 < 4$,

$$\begin{aligned} F(\alpha,k) &= \frac{-e^{\imath\pi/4}}{\sqrt{k}} \frac{\left[4 - k^2(\cos^4\theta_0(\alpha,k) + \sin^4\theta_0(\alpha,k))\right]^{1/4}}{\sqrt{(4-k^2)(2-k^2\sin^2\theta_0(\alpha,k)\cos^2\theta_0(\alpha,k))}} \\ \theta_0(\alpha,k) &= \tan^{-1}\sqrt{-\lambda(\alpha,k) + \sqrt{\lambda(\alpha,k)^2 + \tan^2\alpha}} \\ \lambda(\alpha,k) &= \frac{2(1-\tan^2\alpha)}{4-k^2} \\ \xi_0(\alpha,k) &= 2\sin^{-1}\left[\frac{k}{2}\cos\theta_0(\alpha,k)\right] \\ \eta_0(\alpha,k) &= 2\sin^{-1}\left[\frac{k}{2}\sin\theta_0(\alpha,k)\right]. \end{aligned}$$

In the second case, $4 < k^2 < 8$,

$$\begin{aligned} F(\alpha,k) &= \frac{e^{-\imath\pi/4}}{(8-k^2)^{1/4}} \frac{\left[k^2 - 4 + 2(8-k^2)\sin^2\theta_0(\alpha,k)\cos^2\theta_0(\alpha,k)\right]^{1/4}}{\sqrt{(k^2-4)(2-(8-k^2)\sin^2\theta_0(\alpha,k)\cos^2\theta_0(\alpha,k))}} \\ \theta_0(\alpha,k) &= \tan^{-1}\sqrt{-\lambda(\alpha,k) + \sqrt{\lambda(\alpha,k)^2 + \tan^2\alpha}} \\ \lambda(\alpha,k) &= \frac{2(1-\tan^2\alpha)}{k^2-4} \\ \xi_0(\alpha,k) &= 2\cos^{-1}\left[\frac{1}{2}\sqrt{8-k^2}\cos\theta_0(\alpha,k)\right] \\ \eta_0(\alpha,k) &= 2\cos^{-1}\left[\frac{1}{2}\sqrt{8-k^2}\sin\theta_0(\alpha,k)\right]. \end{aligned}$$

Using these results, let us treat the three sides in turn.

4.2.7.1 Top side

Consider the quantity

$$S_i^T = M(U_{iM}\delta_{\hat{\mathbf{n}}}G_{iM} - G_{iM}\delta_{\hat{\mathbf{n}}}U_{iM})$$

where $i \in [1, M]$. In this section, when we use G_{ij}, we mean $G_{ij;pq}$, the lattice Green's function centered at (p, q). Note that S_i^T is associated with the point (i, M), and this point is associated with the angle $\alpha = \tan^{-1}(M/i)$. For the left endpoint $(1, M)$, the angle α approaches $\pi/2$ as $M \to \infty$. For the right endpoint (M, M), the angle α equals $\pi/4$ for all M.

For any $\alpha \in [\pi/4, \pi/2]$, let $r = \cot\alpha$. Let $\lfloor rM \rfloor$ denote the greatest integer less than rM. Then the sequence $(\lfloor rM \rfloor, M)$ has angle α in the limit where $M \to \infty$. Clearly the sequence of points $(\lfloor rM \rfloor, M + 1)$ also has angle α in the $M \to \infty$ limit. The reason we mention this is that the asymptotic form of $\delta_{\hat{\mathbf{n}}}G_{iM}$ will depend on G_{iM} as well as G_{iM+1}. We think of the sequence $(\lfloor rM \rfloor, M)$ as a "discrete ray" of asymptotic angle α.

Along this ray with $i = \lfloor rM \rfloor$, the asymptotic form of G gives, as $M \to \infty$,

$$S_i^T \sim \sqrt{M}\left[\left(e^{\imath\eta_0(\alpha,k)} - 1\right) - \delta_{\hat{\mathbf{n}}}\right] U_{iM} \frac{e^{\imath([i-p]\xi_0(\alpha,k)+[M-q]\eta_0(\alpha,k))}}{\sqrt{2\pi\csc\alpha}} F(\alpha, k).$$

To obtain this expression, we used

$$R = M\sqrt{\lfloor rM \rfloor^2/M^2 + 1} \to M\csc\alpha \quad \text{as} \quad M \to \infty.$$

So, along the top side, the discrete Sommerfeld outgoing radiation condition is

$$\sqrt{M}\left[U_{iM}\left(e^{\imath\eta_0(\alpha,k)} - 1\right) - \delta_{\hat{\mathbf{n}}}U_{iM}\right] \to 0 \quad \text{as} \quad M \to \infty. \tag{4.19}$$

Assuming this condition holds for all $\alpha \in [\pi/4, \pi/2]$ (or equivalently, for each $i \in [1, M]$) as $M \to \infty$, the first term on the right-hand side of (4.18) goes to zero as $M \to \infty$.

4.2.7.2 Right/bottom sides

Calculations completely analogous to those just presented lead us to discrete Sommerfeld outgoing radiation conditions on the right side of the rectangle of interest,

$$\sqrt{M}\left[U_{Mj}\left(e^{i\xi_0(\alpha,k)} - 1\right) - \delta_{\hat{\mathbf{n}}}U_{Mj}\right] \to 0 \quad \text{as} \quad M \to \infty, \tag{4.20}$$

as well as the bottom side of the rectangle of interest,

$$\sqrt{M}\left[U_{i,-M+1}\left(e^{-\imath\eta_0(\alpha,k)}-1\right)-\delta_{\hat{\mathbf{n}}}U_{i,-M+1}\right]\to 0 \quad \text{as} \quad M\to\infty. \tag{4.21}$$

We give the calculations leading to these conditions in Appendix 4.B.

4.2.8 Method of images

We have shown that if (4.19), (4.20), and (4.21) hold, then S_2—defined in (4.17)—vanishes. Then

$$\begin{aligned} U_{pq}=S_1 &= \sum_{j=-M+1}^{M} U_{1,j}G_{-p,j-q}-U_{1,j}G_{1-p,j-q}-G_{1-p,j-q}U_{0,j}+G_{1-p,j-q}U_{1,j} \\ &= \sum_{j=-M+1}^{M} U_{1,j}G_{0,j;p,q}-G_{1,j;p,q}U_{0,j} \end{aligned} \tag{4.22}$$

We now use the method of images to further simplify (4.22). Consider the lattice Green's function centered about the point $(-p,q)$; by definition, this function satisfies

$$(\Delta_d+k^2)G_{i,j;-p,q}=\delta_{-p,i}\delta_{q,j}.$$

for all $(i,j)\in\mathbb{Z}^2$. For fixed (p,q), define

$$G^{-}_{i,j;p,q}=G_{i,j;p,q}-G_{i,j;-p,q} \quad \forall\,(i,j)\in\Omega. \tag{4.23}$$

Because $G^{-}_{i,j;p,q}$ satisfies (4.14) in Ω and (4.3a) in Ω_0, we can replace G in (4.22) by G^{-}. Then, because $G^{-}_{i,j;p,q}$ vanishes for $i=0$, the second term on the right-hand side of (4.22) vanishes and we obtain

$$U_{pq}=\sum_{j=-M+1}^{M}-G^{-}_{1,j;p,q}U_{0j}.$$

Of course, by (4.3b), we know that $U_{0j}=f_j$ and that f_j is supported only for $j\in\Sigma$. Hence

$$U_{pq}=\sum_{j\in\Sigma}-G^{-}_{1,j;p,q}f_j. \tag{4.24}$$

Note from (4.24) that the desired Green's function is

$$G^{-}_{1,j;p,q}=G_{1-p,j-q}-G_{1+p,j-q}, \tag{4.25}$$

which by symmetry reduces to

$$G^{-}_{1,j;p,q}=G_{p-1,q-j}-G_{p+1,q-j}=G_{p,q-j;1,0}-G_{p,q-j;-1,0}. \tag{4.26}$$

Therefore, for each value of k, it is sufficient to find G such that $\left(\Delta_d+k^2\right)G$ gives $+1$ at $(1,0)$, -1 at $(-1,0)$, and zero everywhere else.

4.2.9 Discrete Rayleigh-Sommerfeld (R-S) formula

Let us examine the behavior of our solution (4.24) on $p = 0$. Using (4.25), we get

$$\begin{aligned} U_{0q} &= \sum_{j\in\Sigma} -G^{-}_{1,j;0,q} f_j \\ &= \sum_{j\in\Sigma} \left(-G_{1,j-q} + G_{1,j-q}\right) f_j = 0. \end{aligned} \tag{4.27}$$

Obviously, this does not match the true boundary condition (4.3b). To remedy the situation, we redefine U_{pq} on the boundary.

Theorem 4.2.1. *The discrete diffraction problem (4.3a) is solved exactly by*

$$U_{pq} = \begin{cases} \displaystyle\sum_{j\in\Sigma} -G^{-}_{1,j;p,q} f_j & p \geq 1 \\ f_q & p = 0. \end{cases} \tag{4.28}$$

Proof. There are two cases. First let us examine what happens for $p \geq 2$. Since $p - 1 \geq 1$, we may use the top branch of (4.28) to compute $U_{p-1,q}$ and the other four values of U upon which $(\Delta_d + k^2)U$ depends. This yields

$$\left(\Delta_d + k^2\right) U_{pq} = \sum_{j\in\Sigma} -\delta_{1,p}\delta_{j,q} f_j + \delta_{1,-p}\delta_{j,q} f_j = 0,$$

because both $\delta_{1,p}$ and $\delta_{1,-p}$ vanish for $p \geq 2$.

Now, when $p = 1$, we have to use both the $p \geq 1$ and the $p = 0$ branches of (4.28) to calculate $(\Delta_d + k^2)U$. With this in mind, we obtain

$$\begin{aligned} \left(\Delta_d + k^2\right) U_{1q} &= U_{0q} + U_{2q} + U_{1,q+1} + U_{1,q-1} - 4U_{1q} + k^2 U_{1q} \\ &= f_q + \left(0 + U_{2q} + U_{1,q+1} + U_{1,q-1} - 4U_{1q} + k^2 U_{1q}\right) && (4.29) \\ &= f_q + \left(\Delta_d + k^2\right) \left[\sum_{j\in\Sigma} -G_{1,j;p,q} f_j + G_{1,j;-p,q} f_j\right]_{p=1} && (4.30) \\ &= f_q + \sum_{j\in\Sigma} \left[-\delta_{1,p}\delta_{j,q} f_j + \delta_{1,-p}\delta_{j,q} f_j\right]_{p=1} \\ &= f_q + \sum_{j\in\Sigma} -\delta_{j,q} f_j \\ &= f_q - f_q = 0. && (4.31) \end{aligned}$$

In the above calculation, we used (4.27) to go from (4.29) to (4.30). □

Since (4.28) solves (4.3a) exactly and also satisfies the boundary condition (4.3b), it is an exact solution of the discrete diffraction problem (4.3). We refer to (4.28) as the discrete Rayleigh-Sommerfeld formula.

4.2.10 Convolution

We can now use the expression for G given in (4.15) to write

$$G^{-}_{i,j;p,q} = \frac{2}{\pi^2}\int_0^\pi \int_0^\pi \frac{\cos[(j-q)\eta]\sin(i\xi)\sin(p\xi)}{\sigma(\xi,\eta;k)} d\xi d\eta. \tag{4.32}$$

Note that (4.32) depends on j and q only through $(q-j)$. Slightly abusing notation, we write $G^{-}_{i,j;p,q} = G^{-}_{i;p}(q-j)$. Next, note that $U_{0,j}$ is supported only for $j \in \Sigma$. Putting everything together, we write (4.28) as a discrete convolution:

$$U_{pq} = \begin{cases} -\left(G^{-}_{1;p} * f\right)[q] & p \geq 1 \\ f_q & p = 0. \end{cases} \tag{4.33}$$

4.3 Computing the lattice Green's function

The main task in numerically evaluating (4.28) is to compute the lattice Green's function. For large values of m and n, the integrands in (4.15) and (4.32) both suffer from extremely rapid oscillations and a curve of singularities where $\sigma(\xi,\eta;k) = 0$. Fortunately, there is a way to compute G_{mn} that does not use numerical quadrature.

4.3.1 Diagonal elements

Recent work by P. A. Martin [Martin, 2006] provides expressions for the diagonal elements of the lattice Green's function in terms of Legendre functions:

$$G_{n,n} = \frac{(-1)^n}{2\pi\imath} \begin{cases} Q_{n-1/2}(z) - \frac{\pi\imath}{2} P_{n-1/2}(z) & k^2 < 4 \\ Q_{n-1/2}(z) + \frac{\pi\imath}{2} P_{n-1/2}(z) & k^2 > 4, \end{cases} \tag{4.34}$$

with $z = 1 - (4-k^2)^2/8$. Since $Q_{n-1/2}(z)$ blows up at $z = 1$, we cannot use the above expressions when $k^2 = 4$. We return to this point later.

4.3.2 Off-diagonal elements

Using 8-fold symmetry, we only need to compute the lattice Green's function $G(m,n)$ in one octant of the plane. To do this, we apply a set of recurrence relations due to Morita [Morita, 1971]. Other recursive approaches may be found in the physics literature [Buneman, 1971; Katsura and Inawashiro, 1971]. Morita's equations use the diagonal elements of the lattice Green's function to uniquely determine the remaining elements:

$$G_{1,0} = \frac{1-(k^2-4)G_{0,0}}{4} \tag{4.35a}$$

$$G_{m,n} = \begin{cases} (4-k^2)G_{m-1,0} - G_{m-2,0} - 2G_{m-1,1} & n = 0 \\ (4-k^2)G_{m-1,n} - G_{m-2,n} - G_{m-1,n+1} - G_{m-1,n-1} & 0 < n < m-1 \\ \frac{4-k^2}{2}G_{m-1,n} - G_{m-1,n-1} & n = m-1 \end{cases} \tag{4.35b}$$

For $k^2 \neq 4$, we use (5.12) and (5.13) in turn and compute all values of $G^{-}_{1,j;p,q}$ needed to evaluate (4.28). For $k^2 = 4$, a different approach is needed.

4.3.3 Green's function for $k^2 = 4$

In this case, we note that the discrete Helmholtz operator simplifies to a sum of the nearest neighbor elements of U:

$$(\Delta_d + 4)U_{pq} = U_{p+1,q} + U_{p-1,q} + U_{p,q+1} + U_{p,q-1} = 4\overline{U}_{pq}. \tag{4.36}$$

Using this representation of $\Delta_d + 4$, we find by inspection the lattice Green's function

$$G^{k^2=4}_{mn} = \frac{1}{4}(-1)^{1+\max\{|m|,|n|\}} = \begin{pmatrix} \ddots & & & \vdots & & & \iddots \\ & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \\ & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \\ & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \\ \cdots & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \cdots \\ & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \\ & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \\ & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \\ \iddots & & & \vdots & & & \ddots \end{pmatrix} \tag{4.37}$$

consisting of a central $-1/4$ surrounded by concentric alternating $1/4$ and $-1/4$ rings. Note that, unlike the lattice Green's functions for $k^2 \in (0,4)$ and $k^2 \in (4,8)$, the above function does not decay as $R = \sqrt{m^2+n^2} \to \infty$. Define

$$H_{ij} = \begin{cases} (-1)^i & (i+j) \text{ even} \\ 0 & (i+j) \text{ odd} \end{cases} = \begin{pmatrix} \ddots & & & & \vdots & & & & \iddots \\ & -1 & 0 & -1 & 0 & -1 & 0 & -1 & \\ & 0 & 1 & 0 & 1 & 0 & 1 & 0 & \\ & -1 & 0 & -1 & 0 & -1 & 0 & -1 & \\ \cdots & 0 & 1 & 0 & 1 & 0 & 1 & 0 & \cdots \\ & -1 & 0 & -1 & 0 & -1 & 0 & -1 & \\ & 0 & 1 & 0 & 1 & 0 & 1 & 0 & \\ & -1 & 0 & -1 & 0 & -1 & 0 & -1 & \\ \iddots & & & & \vdots & & & & \ddots \end{pmatrix}. \tag{4.38}$$

Both $H_{i,j}$ and $H_{i-1,j}$ are non-trivial, non-decaying solutions of the homogeneous discrete Helmholtz equation (4.3a) on all of $\mathbb{Z}^2$. For $k^2 = 4$ and for all $\alpha, \beta \in \mathbb{R}$, the functions $G^{k^2=4}_{ij} + \alpha H_{ij} + \beta H_{i-1,j}$ are valid lattice Green's functions.

The same can be said for the method of images Green's function G^-. Using the specific form of $G^{k^2=4}$ given by (4.37) and (4.26), we obtain

$$G^-_{-1,0;m,n} = G_{m,n;-1,0} - G_{m,n;1,0} = \begin{pmatrix} \ddots & & & & \vdots & & & & \iddots \\ & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & \\ & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & \\ & 0 & 0 & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ & 0 & 0 & \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & \\ & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & \\ & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & \\ \iddots & & & & \vdots & & & & \ddots \end{pmatrix}. \tag{4.39}$$

The diagonal and anti-diagonal elements alternate between $+1/2$ and $-1/2$, but do not decay to zero. Again, we could add arbitrary multiples of $H_{i,j}$ and/or $H_{i-1,j}$ to $G^-_{-1,0;i,j}$ without changing $(\Delta_d + 4)G^-_{-1,0;i,j}$. In computations, when $k^2 = 4$, we use (4.39).

4.3.4 Discrete Rayleigh-Sommerfeld solution: $k^2 = 4$

Using the $k^2 = 4$ Green's function (4.39) together with the discrete R-S formula (4.28), we obtain in closed form the *exact* steady-state solution of (4.1) with $u_{0,j}(t) = e^{2\imath t}$ for $|j| \le A$ and $u_{0,j}(t) = 0$ for $|j| > A$. For example, with a small aperture $A = 2$ we obtain $u_{m,n}(t) = U_{m,n}e^{2\imath t}$, where

$$U_{m,n} = \begin{pmatrix} & & & & & \iddots \\ & & & & 1/2 & \\ & & & -1/2 & 1/2 & \\ & & 1/2 & -1/2 & 1/2 & \\ & -1/2 & 1/2 & -1/2 & 1/2 & \\ 1 & -1/2 & 1/2 & -1/2 & 1/2 & \\ 1 & -1 & 1/2 & -1/2 & 0 & \\ 1 & -1 & 1 & 0 & 0 & \cdots \\ 1 & -1 & 1/2 & -1/2 & 0 & \\ 1 & -1/2 & 1/2 & -1/2 & 1/2 & \\ & -1/2 & 1/2 & -1/2 & 1/2 & \\ & & 1/2 & -1/2 & 1/2 & \\ & & & -1/2 & 1/2 & \\ & & & & 1/2 & \\ & & & & & \ddots \end{pmatrix}. \tag{4.40}$$

Let us label the left-most column of this solution as the $m = 0$ column. Then for all (m, n) with $m \ge 1$, we find that the sum over the nearest neighbors vanishes, *i.e.*, $4\overline{U}_{m,n} = 0$. Then, by (4.36), we obtain

$$\frac{d^2}{dt^2}U_{m,n}e^{2\imath t} = \Delta_d U_{m,n}e^{2\imath t},$$

so as claimed, $u_{m,n}(t)$ gives an exact solution. The solution is steady-state in the sense that it is assumed that the driving force has been in existence since $t = -\infty$.

The solution for all other aperture sizes is similar: the initial wave splits into two traveling waves of half the amplitude moving at angles $+\pi/4$ and $-\pi/4$. The traveling waves are multiplied by -1 each time they move one unit to the right.

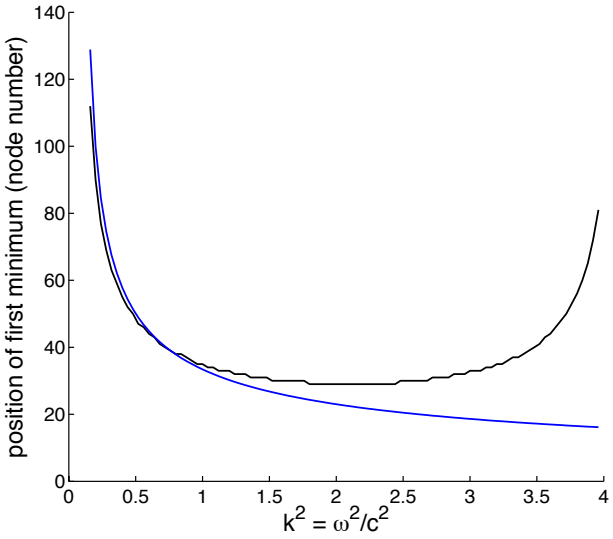


Figure 4.2: Position of first minimum for the continuous (blue) and discrete (black) R-S theories for various values of $k^2 = \omega^2/c^2$.

4.4 Conclusion

We opened this chapter by describing a time-domain numerical experiment and proceeded to develop a discrete theory that exactly solves the steady-state version of the problem, for all values of the driving frequency ω. The question remains: for which values of ω does the continuum theory usefully predict the discrete result?

To address this question, we plot in Figure 4.2 the position of the first minimum of the diffraction pattern as predicted by the continuous R-S theory (blue) and discrete R-S theory (black) for various values of $k^2 = \omega^2/c^2$ and constant amplitude input. For $k^2 \lesssim 0.14$, both diffraction patterns have no minima. Above this value, the diffraction pattern looks qualitatively like that in Figure 4.1. The two diffraction theories closely agree for $k^2 \lesssim 1$. The continuous R-S theory suggests that the diffraction pattern continues to condense about the y-axis as $k^2 \to 4$. However, the discrete R-S theory predicts the first minimum moves away from the y-axis as $k^2 \to 4$.

Since the discrete R-S formula (4.28) exactly solves (4.3) for all values of k^2, we can interpret Figure 4.2 as a measure of how well the continuous R-S theory agrees with the true diffraction pattern. Searching for solutions with wavevector $\vec{\mathbf{k}} = (k_x, k_y) = (\kappa, \kappa)$ in (4.1a) we obtain $\kappa =$

$\cos^{-1}[1-\omega^2/(4c^2)]/(2\pi)$. At $\omega^2/c^2 = 1$, $\kappa \approx (8.7)^{-1}$. According to Figure 4.2, $\omega^2/c^2 = 1$ is where the discrete and continuum theories begin to diverge. Putting everything together, we see that if waves occupy less than 9 lattice nodes, the continuous R-S theory no longer accurately predicts the output. In this case, the discrete R-S theory should be used.

Using similar lattice Green's function methods, discrete R-S theories can be derived for other regular lattices, such as the 2-D honeycomb, 2-D triangular, and 3-D cubic lattices. We expect that for all such media, discrete diffraction theories can answer questions regarding short-wavelength phenomena in situations where continuum diffraction theories cannot.

4.A Numerical details

In Section 4.1, when carrying out time-domain simulations of an $N \times N$ lattice, we began with

$$\frac{d^2}{dt^2}\mathbf{u} = \Delta_d \mathbf{u} - \mathbf{b}.\frac{d}{dt}\mathbf{u}. \tag{4.41}$$

The notation $\mathbf{v}.\mathbf{w}$ denotes element-wise multiplication of equal-length vectors. For each t, $\mathbf{u}(t) \in \mathbb{R}^{N^2}$ is an ordinary column vector, with one unknown at each lattice node. If A and D are the $N^2 \times N^2$ adjacency and degree matrices for the lattice, then $\Delta_d = A - D$. When $\Delta_d \mathbf{u}$ refers to a node on the left boundary, we simply use the value of u on the boundary given by (4.1b).

Boundary Conditions To mimic a semi-infinite domain, we use first-order absorbing boundary conditions of Engquist-Majda [Engquist and Majda, 1977] type on the top, right, and bottom boundaries. We accomplish this through the vector $\mathbf{b} \in \mathbb{R}^{N^2}$ in (4.41). For any node j that borders the left boundary, where there is a Dirichlet condition, we set $b_j = 0$. For all other nodes, we set $b_j = 1$ if node j is a non-corner boundary node, $b_j = 2$ if node j is a corner boundary node, and $b_j = 0$ otherwise.

To see the effect of the damping term represented by $\mathbf{b}$, consider a non-corner boundary node on the right boundary of the square lattice. Then, rewriting (4.41) using two-dimensional indices, we obtain

$$\frac{d^2}{dt^2}u_{m,n} = u_{m-1,n} + u_{m,n+1} + u_{m,n-1} - 3u_{m,n} - \frac{d}{dt}u_{m,n}. \tag{4.42}$$

What if we instead had an infinite domain? Then, at the same boundary node (m, n), we would

instead have the equation

$$\frac{d^2}{dt^2}u_{m,n} = u_{m+1,n} + u_{m-1,n} + u_{m,n+1} + u_{m,n-1} - 4u_{m,n}. \tag{4.43}$$

The difference between the two equations is

$$(u_{m,n} - u_{m+1,n}) - \frac{d}{dt}u_{m,n} = 0, \tag{4.44}$$

where the term in parentheses is a simple finite difference. Equation (4.44) is a first-order, spatially discretized version of the Engquist-Majda absorbing boundary condition (ABC). Our lattice equation (4.41) is equivalent to the infinite lattice equation (4.43) plus the boundary condition (4.44) for non-forced, non-corner boundary nodes.

To see what happens at the corners, let us examine (4.41) in the upper-right corner. Again using two-dimensional indices, we obtain

$$\frac{d^2}{dt^2}u_{m,n} = u_{m-1,n} + u_{m,n-1} - 2u_{m,n} - 2\frac{d}{dt}u_{m,n}. \tag{4.45}$$

This equation follows from the infinite lattice equation (4.43) together with the spatially discrete ABCs

$$\begin{aligned}(u_{m,n} - u_{m+1,n}) - \frac{d}{dt}u_{m,n} &= 0\\ (u_{m,n} - u_{m,n+1}) - \frac{d}{dt}u_{m,n} &= 0\end{aligned} \tag{4.46}$$

The reason for setting $b_j = 2$ for corner nodes should now be clear.

We close by stating that, given how we have defined $\mathbf{b}$, a discrete ABC similar to (4.44) holds on the top, right, and bottom boundaries. A discrete ABC similar to (4.46) holds at the lower-right corner.

Physical Interpretation We may interpret (4.41) as a second-order equation for voltage $\mathbf{u}$ that can be derived from Kirchhoff's Laws for an inductor-capacitor lattice as described in Section 4.1.1.2. In this case, the $\mathbf{b}$ term arises from connecting non-forced boundary nodes to grounded resistors whose resistance equals the local lattice impedance $\sqrt{L/C}$. Our ABC amounts to impedance matching, a well-known concept in circuit design.

Time-Stepping Starting from (4.41), we discretize in time using centered differences. Let $\mathbf{u}^k$ denote our numerical approximation to $\mathbf{u}(k\Delta t)$. Let B be a diagonal matrix with entries equal to the vector $\mathbf{b}$. Then our scheme is

$$\left(I + \frac{\Delta t}{2}B\right)\mathbf{u}^{k+1} = 2\mathbf{u}^k - \mathbf{u}^{k-1} + (\Delta t)^2\left[\Delta_d \mathbf{u}^k\right] + \frac{\Delta t}{2}B\mathbf{u}^{k-1}.$$

Since I and B are both diagonal, it is trivial to solve for $\mathbf{u}^{k+1}$ at each time step. To generate $\mathbf{u}^1$ given $\mathbf{u}^0$, we take one step using the standard semi-implicit Euler method applied to (4.41).

For $\omega = \sqrt{1/2}$, we choose $\Delta t = 2\pi/(128\omega)$. For $\omega = \sqrt{2}$, we choose $\Delta t = 2\pi/(256\omega)$. For $\omega = \sqrt{11/4}$, we choose $\Delta t = 2\pi/(320\omega)$. Finally, for $\omega = \sqrt{3/2}$, we choose $\Delta t = 2\pi/(384\omega)$.

As this is a linear problem, it is easy to analyze the stability of the scheme by writing it as a map from $(\mathbf{u}^{k-1}, \mathbf{u}^k)$ to $(\mathbf{u}^k, \mathbf{u}^{k+1})$. For all values of ω used in this paper, our time step Δt is chosen so that the eigenvalues of the mapping lie strictly inside the unit circle in the complex plane, ensuring stability.

4.B Discrete Sommerfeld conditions

Our purpose here is to give conditions on U under which the second and third terms on the right-hand side of (4.18) vanish. These terms correspond, respectively, to the right and bottom sides of the rectangle over which S_2 is summed in (4.17). The derivations here are completely analogous to that of Section 4.2.7.1. As before, when we use G_{ij}, we mean $G_{ij;pq}$, the lattice Green's function centered at (p, q).

4.B.1 Right side

Consider the quantity

$$S_j^R = 2M(U_{Mj}\delta_{\hat{\mathbf{n}}}G_{M-p,j-q} - G_{M-p,j-q}\delta_{\hat{\mathbf{n}}}U_{Mj})$$

where $j \in [-M+1, M]$. Note that S_j^R is associated with the point (M, j) and this point is associated with the angle $\alpha = \tan^{-1}[j/M]$. For any $\alpha \in [-\pi/4, \pi/4]$, let $r = \tan\alpha$. Let $\lfloor\beta\rfloor$ denote the greatest integer less than or equal to β if $\beta \geq 0$, or the smallest integer greater than β if $\beta < 0$. Then $-M+1 \leq \lfloor rM\rfloor \leq M$ and both sequences of points $(M, \lfloor rM\rfloor)$ and $(M+1, \lfloor rM\rfloor)$ have angle α in the $M \to \infty$ limit.

Along the ray with $j = \lfloor rM \rfloor$, the asymptotic form of G gives, as $M \to \infty$,

$$S_j^R \sim 2\sqrt{M}\left[\left(e^{i\xi_0(\alpha,k)} - 1\right) - \delta_{\hat{\mathbf{n}}}\right] U_{Mj} \frac{e^{\imath([M-p]\xi_0(\alpha,k)+[j-q]\eta_0(\alpha,k))}}{\sqrt{2\pi \sec\alpha}} F(\alpha,k).$$

To obtain this expression, we used $R = M\sqrt{1+\lfloor rM\rfloor^2/M^2} \to M\sec\alpha$, again, as $M \to \infty$. This shows that, along the right side of the rectangle, the discrete Sommerfeld outgoing radiation condition is (4.20). Assuming this condition holds for all $\alpha \in [-\pi/4, \pi/4]$ (or equivalently, for each $j \in [-M+1, M]$) as $M \to \infty$, the second term on the right-hand side of (4.18) goes to zero as $M \to \infty$.

4.B.2 Bottom side

The treatment is nearly identical to the top side. Define

$$S_i^B = M(U_{i,-M+1}\delta_{\hat{\mathbf{n}}} G_{i-p,-M+1-q} - G_{i-p,-M+1-q}\delta_{\hat{\mathbf{n}}} U_{i,-M+1}),$$

where again $i \in [1, M]$. For any $\alpha \in [-\pi/2, -\pi/4]$, let $r = -\cot\alpha$. Let $\lfloor rM \rfloor$ denote the greatest integer less than rM. Both sequences $(\lfloor rM\rfloor, -M+1)$ and $(\lfloor rM\rfloor, -M)$ have angle α in the $M \to \infty$ limit.

Along the ray with $i = \lfloor rM \rfloor$, the asymptotic form of G gives, as $M \to \infty$,

$$S_i^T \sim \sqrt{M}\left[\left(e^{-\imath\eta_0(\alpha,k)} - 1\right) - \delta_{\hat{\mathbf{n}}}\right] U_{i,-M+1} \frac{e^{\imath([i-p]\xi_0(\alpha,k)+[-M+1-q]\eta_0(\alpha,k))}}{\sqrt{2\pi \csc\alpha}} F(\alpha,k).$$

To obtain this expression, we used $R = M\sqrt{\lfloor rM\rfloor^2/M^2 + (-M+1)^2/M^2} \to M\csc\alpha$. Therefore, along the bottom side, the discrete Sommerfeld outgoing radiation condition is (4.21). Again, assuming this condition holds for all $\alpha \in [-\pi/2, -\pi/4]$ (or equivalently, for each $i \in [1, M]$) as $M \to \infty$, the third term on the right-hand side of (4.18) goes to zero as $M \to \infty$.

Chapter 5

Discrete wave propagation in two-dimensional transmission line metamaterials

5.1 Introduction

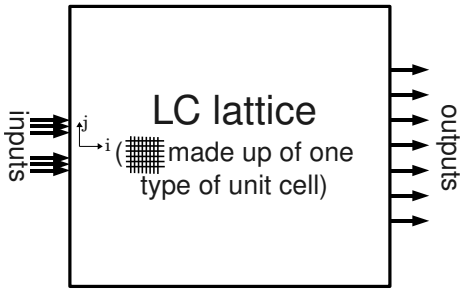


Figure 5.1: Given prescribed inputs and a rectangular slab with fine-scale structure consisting of unit cells of a single type, how can we efficiently solve for the outputs?

Consider a two-dimensional rectangular metamaterial which, at the fine scale, consists of a lattice of repeated cells of a single type. Fig. 5.1 shows a pictorial representation of the central problem: suppose monochromatic inputs of the form $f_j \exp(\imath\omega t)$ are connected to the left boundary. Assume that the number of unit cells in the rectangular slab is large enough to make it prohibitively

expensive to solve numerically for the voltage/current at every cell in the lattice until the system reaches steady state. In this case, how can we efficiently solve for the steady-state output amplitudes g_j at the right boundary? And what will the outputs look like? In this chapter, we use lattice Green's functions to answer these questions, under the assumption that the top/bottom boundaries of the lattice are resistively terminated in such a way as to simulate outgoing boundary conditions, *i.e.*, the lattice is effectively infinite in the top and bottom directions. See Figs. 5.2-5.3 for definitions of the composite and dual-composite right/left-handed (CRLH and d-CRLH, respectively) unit cells.

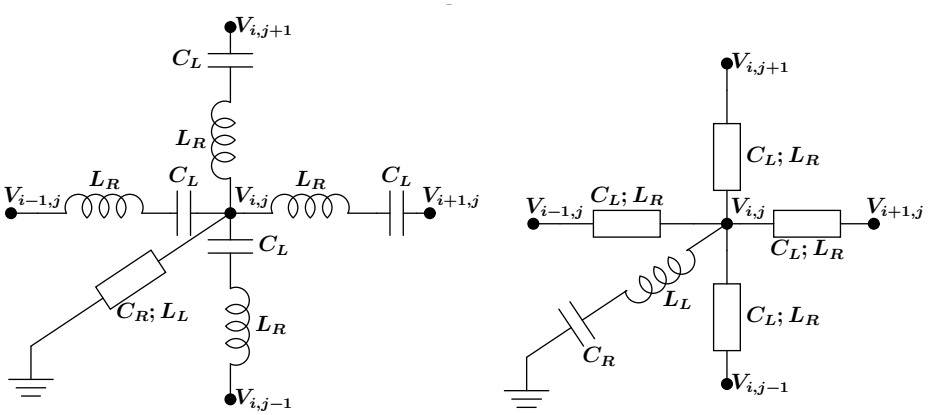


Figure 5.2: Unit cells for CRLH (left) and d-CRLH (right) metamaterials.

Figure 5.3: Schematic abbreviation for parallel LC block.

One key observation underpinning our method is that for the case where all unit cells are of one type, the problem reduces to solving a thin-slit diffraction problem for the two-dimensional discrete Helmholtz equation. Using the Green's function for the infinite square lattice, we can write down the exact solution of this diffraction problem. We apply known algorithms from the computational physics literature to compute the lattice Green's function. Once we have fixed the input frequency ω and computed the lattice Green's function for a given wave number k, we can then solve the

propagation problem and determine the outputs g_j for any set of inputs f_j. This final computation is a simple convolution that can be performed in a time proportional to the number of lattice nodes in the vertical direction.

In earlier work [Bhat and Osting, 2008], we approximated the lattice as a continuum and then used continuum diffraction theory to show that, for physically realizable material parameters, both CRLH and d-CRLH metamaterials show strong discrimination of temporal input frequencies in regimes where group velocity is small. In this letter, we show how to solve the problem without making any continuum assumptions/approximations, thus improving our earlier results and providing an alternative method for solving propagation problems in metamaterials.

5.2 Mathematical modeling and derivation

Let us define the discrete Laplacian Δ_d by

$$(\Delta_d V)_{i,j} = V_{i+1,j} + V_{i-1,j} + V_{i,j+1} + V_{i,j-1} - 4V_{i,j}. \tag{5.1}$$

Given the problem and assumptions described in the introduction, we begin by using Kirchhoff's laws for the CRLH unit cell (shown on the left of Fig. 5.2) to derive the following fourth-order equation for voltage $V_{i,j}$:

$$C_L L_R L_L C_R \frac{d^4}{dt^4} V_{i,j} = L_L C_L \frac{d^2}{dt^2} (\Delta_d V)_{i,j} - (L_L C_R + C_L L_R) \frac{d^2}{dt^2} V_{i,j} - V_{i,j}, \tag{5.2}$$

where i and j are integers that give, respectively, the horizontal/vertical positions in the lattice, as shown in Fig. 5.1. We stipulate that (5.2) holds for all $i \geq 1$, and that along the left boundary $i = 0$, we have the time-dependent boundary condition

$$V_{0,j}(t) = f_j e^{\imath\omega t}. \tag{5.3}$$

Here f_j is non-zero only for $-\sigma \leq j \leq \sigma$, a finite number of nodes along the left boundary. We assume that at time $t = 0$, $V_{i,j}$ and all its derivatives are zero for all $i \geq 1$ and all j. Then, as $t \to \infty$, the boundary term causes waves to propagate into the lattice. As t increases, the system approaches steady-state, at which point the solution is given by $V_{i,j}(t) = \psi_{i,j} \exp(\imath\omega t)$. Substituting this expression for V into (5.2) and (5.3), we find that the spatial part of the wave, ψ, must solve

the discrete Helmholtz system

$$(\Delta_d + k^2)\psi_{i,j} = 0, \quad i \geq 1 \tag{5.4}$$

$$\psi_{0,j} = f_j, \tag{5.5}$$

where k and ω are related through the dispersion relation

$$L_R C_R \omega^2 + \frac{1}{L_L C_L}\omega^{-2} - \frac{L_L C_R + C_L L_R}{L_L C_L} = k^2. \tag{5.6}$$

This derivation began by assuming that each cell in the lattice was a CRLH cell. If instead the lattice is made up of d-CRLH unit cells (shown on the right of Fig. 5.2) we only have to replace (5.2) by

$$C_L L_R L_L C_R \frac{d^4}{dt^4}(\Delta_d V)_{i,j} = L_R C_R \frac{d^2}{dt^2} V_{i,j} - (L_L C_R + C_L L_R)\frac{d^2}{dt^2}(\Delta_d V)_{i,j} - (\Delta_d V)_{i,j}. \tag{5.7}$$

Starting from this equation and following precisely the same steps as above, we may derive the discrete Helmholtz system (8.4-5.5) with the dispersion relation

$$\left(L_L C_L \omega^2 + \frac{1}{L_R C_R}\omega^{-2} - \frac{L_L C_R + C_L L_R}{L_R C_R}\right)^{-1} = k^2. \tag{5.8}$$

Note that the CRLH and d-CRLH dispersion relations include, as limiting cases, both purely right-handed and purely left-handed media. For instance, if we take $L_L, C_L \to \infty$ in (5.6), we obtain the dispersion relation for a purely right-handed (PRH) medium: $L_R C_R \omega^2 = k^2$. Similarly, if we take $L_R, C_R \to \infty$ in (5.8), we obtain the dispersion relation for a purely left-handed (PLH) medium: $(L_L C_L \omega^2)^{-1} = k^2$.

What the above arguments show is that, for a variety of media, the propagation problem depicted in Fig. 5.1 boils down to solving the discrete Helmholtz system (8.4-5.5) for a given dispersion relation. As we have shown in recent work [Bhat and Osting, 2009a], for $0 < k^2 < 8$, the exact solution to this system is

$$\psi_{pq} = \begin{cases} (G_p^- * f)[q] & p \geq 1 \\ f_q & p = 0 \end{cases} \tag{5.9}$$

where

$$G_p^-[q] = -\frac{2}{\pi}\int_0^\pi \int_0^\pi \frac{\cos(q\eta)\sin(\xi)\sin(p\xi)}{k^2 - 4\sin^2(\xi/2) - 4\sin^2(\eta/2)}\,\mathrm{d}\xi\,\mathrm{d}\eta. \tag{5.10}$$

The key to deriving this result is to view the system (8.4-5.5) as a thin-slit diffraction problem. Thinking of f as an aperture function, one uses discrete versions of classical Rayleigh-Sommerfeld arguments to derive (5.9). The function (5.10) is in fact a method of images Green's function that may be derived from the lattice Green's function G on the infinite square lattice. Let $G_{ij;pq}$ denote the lattice Green's function centered at the point (p, q), evaluated at (i, j). Denoting the Kronecker delta function by δ_{xy}, we have, by definition of the lattice Green's function,

$$(\Delta_d + k^2)G_{ij;pq} = \delta_{ip}\delta_{jq} \tag{5.11}$$

Note that if one replaces the Kronecker deltas by a two-dimensional Dirac delta and also replaces the discrete Laplacian by the continuous Laplacian, then (5.11) defines the Green's function for the continuous Helmholtz operator. With these definitions in place, one may write the method of images Green's function as $G_p^-[q] = G_{p,q;-1,0} - G_{p,q;1,0}$. Then, to derive the integral formula (5.10), one uses a standard integral representation of G—see [Katsura and Inawashiro, 1971; Economou, 2006]. For a derivation that includes details of all the above steps, please consult [Bhat and Osting, 2009a]. The behavior of the solution (5.9) is similar to the continuous case for $k^2 \lesssim 1.5$ and as k approaches 4 diverges signficantly. For $k^2 > 4$, waves may only enter the lattice diagonally and are totally different than the continuous case.

5.3 Numerical method

Since (5.9) solves the discrete Helmholtz system (8.4-5.5), it can be used to solve propagation problems for CRLH, d-CRLH, and PLH metamaterials, as well as for PRH materials. However, it is difficult to evaluate the Green's function using the integral representation (5.10), since the integrand suffers from extremely rapid oscillations and a curve of singularities where the denominator vanishes. Fortunately, there is a way to compute $G_p[q]$ that does not use numerical quadrature.

Recent work by P. A. Martin [Martin, 2006] provides expressions for the diagonal elements of the lattice Green's function in terms of Legendre functions:

$$G_{n,n} = \frac{(-1)^n}{2\pi\imath} \begin{cases} Q_{n-1/2}(z) - \frac{\pi\imath}{2} P_{n-1/2}(z) & k^2 < 4 \\ Q_{n-1/2}(z) + \frac{\pi\imath}{2} P_{n-1/2}(z) & k^2 > 4, \end{cases} \tag{5.12}$$

with $z = 1 - (4 - k^2)^2/8$.

Using 8-fold symmetry, we only need to compute the lattice Green's function $G(m,n)$ in one octant of the plane. To do this, we apply a set of recurrence relations due to Morita [Morita, 1971]. Other recursive approaches may be found in the computational physics literature [Buneman, 1971; Katsura and Inawashiro, 1971]. Morita's equations use the diagonal elements of the lattice Green's function to uniquely determine the remaining elements:

$$G_{1,0} = \frac{1-(k^2-4)G_{0,0}}{4} \tag{5.13a}$$

$$G_{m,n} = \begin{cases} (4-k^2)G_{m-1,0} - G_{m-2,0} - 2G_{m-1,1} & n = 0 \\ (4-k^2)G_{m-1,n} - G_{m-2,n} - G_{m-1,n+1} - G_{m-1,n-1} & 0 < n < m-1 \\ \frac{4-k^2}{2}G_{m-1,n} - G_{m-1,n-1} & n = m-1 \end{cases} \tag{5.13b}$$

For $k^2 \neq 4$, we use (5.12) and (5.13) in turn and compute all values of $G^{-}_{1,j;p,q}$ needed to evaluate (5.9). We have developed a Mathematica code that implements this method for calculating G and thereby determining the diffracted field U. The code may be downloaded at `http://www.cds.caltech.edu/~bhat/discreteRS.nb`.

Since $Q_{n-1/2}(z)$ blows up at $z = 1$, we cannot use the above approach when $k^2 = 4$; for details on how that case can be handled analytically, see [Bhat and Osting, 2009a].

5.4 Numerical results / Discussion

We present several numerical tests using the above theoretical results. Examining data for experimentally realized CRLH/d-CRLH metamaterials [Caloz and Nguyen, 2007], we chose the parameters $C_R = 1.33\text{pF}$, $C_L = 0.97\text{pF}$, $L_R = 0.96\text{nH}$, and $L_L = 0.29\text{nH}$. Both CRLH and d-CRLH dispersion relations for $0 < k^2 < 8$ are plotted in Fig. 5.4, following the sign conventions derived in earlier works [Caloz and Itoh, 2006; Caloz, 2006]. Vertical lines have been drawn at $k = \pm 2$ to emphasize the change in behavior for $k^2 < 4$ and $k^2 > 4$. Note that both materials have pass-bands for low and high frequencies, and a stop-band (or gap) for central frequencies.

For our first numerical test, we use a domain with 100×200 nodes and an aperture of size 20 centered on the left. We consider an input signal that equals one on the aperture and zero elsewhere. We choose two triads $\omega = \{3.4\text{GHz}, 3.5\text{GHz}, 3.6\text{GHz}\}$ and $\omega = \{4.8\text{GHz}, 4.9\text{GHz}, 5.0\text{GHz}\}$ and using the numerical method described above, evaluate the discrete diffraction formulae for both

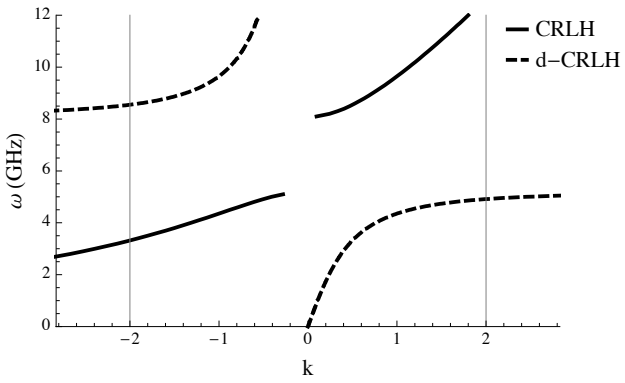


Figure 5.4: Dispersion relations for CRLH and d-CRLH metamaterials.

CRLH and d-CRLH metamaterials, and plot in Fig. 5.5 the magnitude of the diffracted image on the right side of the domain, 100 units from the aperture. For both metamaterials, we observe that when the frequency ω varies rapidly as a function of k, *i.e.*, regimes with large group velocity (GV) $|d\omega/dk|$, the diffracted image hardly depends on the input frequency as demonstrated by the dashed (resp. solid) curve on the left (resp. right) of Fig. 5.5. In small GV regimes, the diffracted image is extremely sensitive to input frequency; here one may be able to use diffraction to effect a large spatial separation of slightly different temporal frequencies - see especially the dashed curves on the right of Fig. 5.5. Though the spatial profile of the input is precisely the same in all cases, slight differences in the input frequency cause noticeably different spatial diffraction patterns. The $\omega = 5$GHz dashed curve on the right $k^2 \approx 6 > 4$ and exemplifies the behavior of the large k regime.

For our final numerical experiment, we consider two-slit interference on a domain with 100×200 nodes and an input signal which is one for $j = \pm 20$ and zero otherwise. We choose two values, $k^2 = 3.5$ and $k^2 = 6$ and plot the results in Fig. 5.6. The value $k^2 = 3.5$ is well into the regime where the continuous theory does not accurately predict the behavior of the discrete system. This is exemplified by the uneven period of the signal at the right hand side of the domain which is in contrast with the continuous theory. This effect is not seen for $k^2 > 4$.

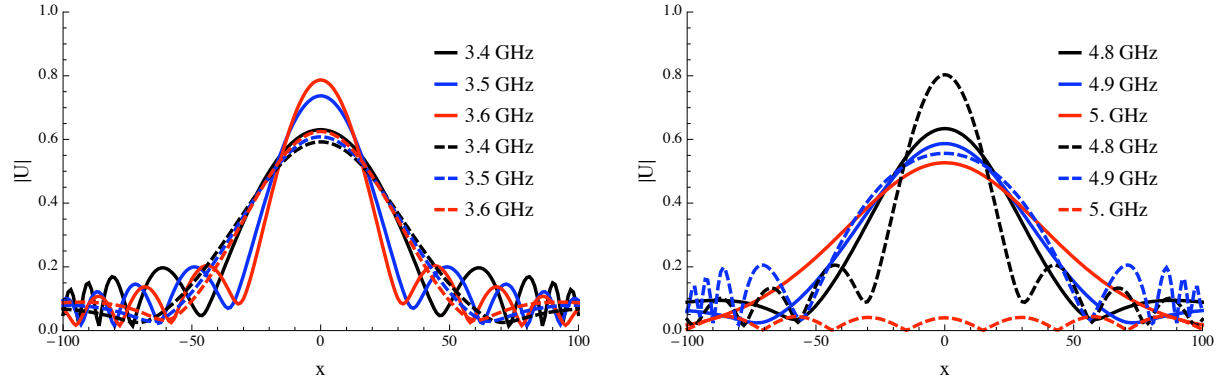


Figure 5.5: Diffracted wavefields for CRLH (solid) and d-CRLH (dashed) metamaterials in the low- and high-frequency regimes.

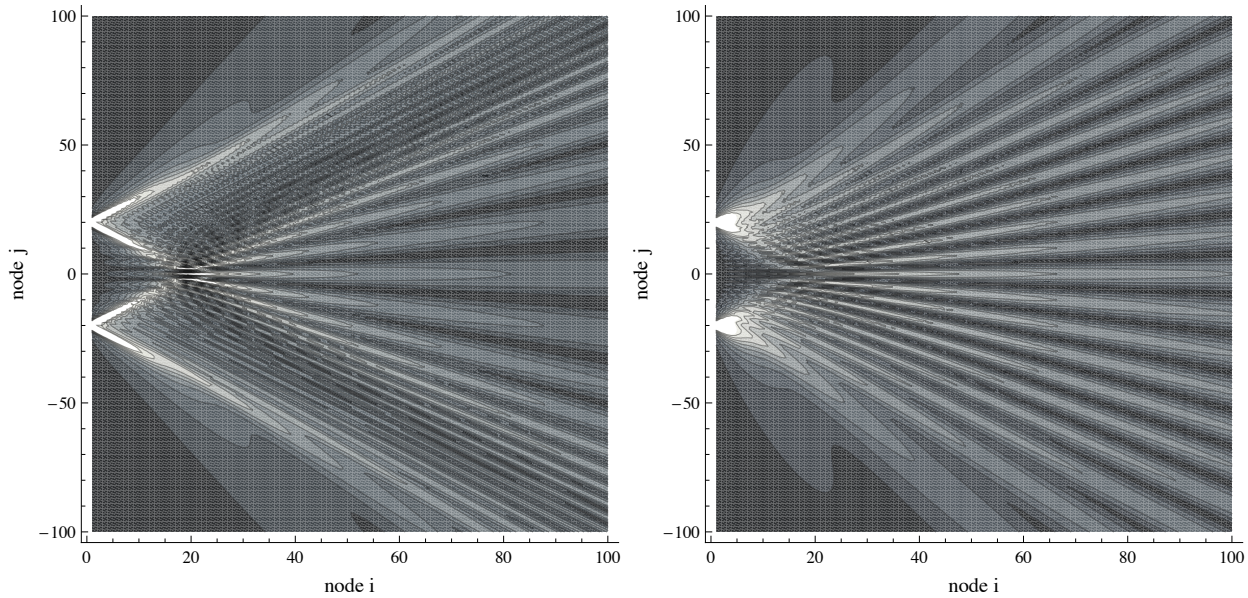


Figure 5.6: Two slit interference for $k^2 = 3.5$ (left) and $k^2 = 6$ (right).

5.5 Discussion

Suppose that we have a fundamentally discrete medium, *i.e.*, a medium that is literally made up of individual capacitors and inductors. If we take the medium to be purely right-handed (PRH), the effect of discreteness on diffraction is only important for extremely high-frequency (short-wavelength) waves. However, due to the dispersive properties of CRLH and d-CRLH materials, there may exist low- and medium-frequency bands where continuum diffraction theory gives grossly

incorrect results. In such cases, the discrete diffraction formula must be used. In fact, if we again consider our previous work [Bhat and Osting, 2008] where the continuum diffraction theory was used to approximate the solution, we find that we computed approximate solutions for large k^2 where no solutions actually exist!

The method can also be extended to solve propagation problems for nonlinear metamaterials.

Chapter 6

Diffraction on the two-dimensional triangular lattice

6.1 Introduction

In this chapter, we consider the analogous problem to that considered in Ch. 4 on a triangular lattice. That is, we consider the diffraction problem on a triangular inductor-capacitor (LC) lattice, as in Fig. 6.1(left). Each edge represents a conductor and each node represents a capacitor connected to a common ground. We take the inductances L and capacitances C to be identical at each edge and node. As in Ch. 4, the lattice is considered to be semi-infinite and harmonically-forced on a portion of the boundary, which we refer to as the aperture.

In Ch. 4, we use lattice Green's functions and a discrete Sommerfeld outgoing radiation condition to derive the exact solution everywhere in the lattice. This results in a solution which can be written as a discrete convolution, where the kernel is computed via a recursive algorithm.

In this chapter, we approximate the exact solution by taking the quasi-continuum limit of the discrete wave equation. While the continuum limit of the discrete wave equation yields the wave equation [Afshari *et al.*, 2008], the quasi-continuum limit yields the wave equation with an additional isotropic, dispersive term which, to second-order, models the discreteness of the lattice. For this modified wave equation, we derive a dispersively-corrected, Rayleigh-Sommerfeld diffraction formula. Remarkably, the convolution kernel for this formula is expressible in terms of the Green's function for the Helmholtz equation. This formula is used to describe the effect of discreteness on

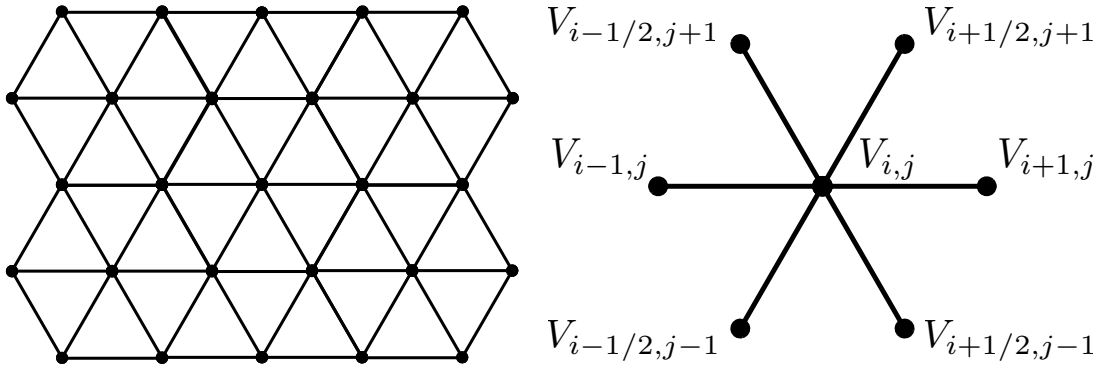


Figure 6.1: Regular triangular lattice.

high-frequency wave propagation in the lattice, especially at wavelengths comparable to the lattice spacing (*i.e.*, close to the Bragg cutoff for the lattice).

The motivation for deriving an approximate solution to this problem when an exact solution is available is as follows: (i) The evaluation of the discrete Green's function depends on a recursion formula and the analytical evaluation of some elements. These have been worked out for some lattices (square, triangular, honeycomb), but rely on finding symmetries. For other lattice topologies, this may not be possible. (ii) The evaluation of the recursive formula is expensive for large lattices and the amplification of small errors requires many digits of precision. See Ch. 4 or [Bhat and Osting, 2009a; Bhat and Osting, 2009b] for further discussion.

6.2 Derivation of the diffraction formula

Using the labeling of voltage nodes shown in Fig. 6.1, we use Kirchhoff's Laws to derive the discrete wave equation governing voltage within the lattice:

$$LC\frac{d^2}{dt^2}V_{i,j} = V_{i-1/2,j+1} + V_{i+1/2,j+1} + V_{i-1,j} + V_{i+1,j} + V_{i-1/2,j-1} + V_{i+1/2,j-1} - 6V_{i,j}. \tag{6.1}$$

To analyze diffraction governed by this model, we derive a continuum PDE. Assume that each edge in the lattice has length $h > 0$. Define distributed inductance and capacitance by $\ell = L/h$ and $c = C/h$. Let $V(x, y, t)$ denote the continuum approximation of $V_{i,j}(t)$ where $x = ih$ and $y = jh$. Then the Taylor expansion of the right-hand side of (6.1) about the central voltage $V_{i,j}(t)$ gives

$$\ell c\partial_t^2 V = \frac{3}{2}\Delta V + \frac{3}{32}h^2\Delta^2 V + \mathcal{O}(h^4), \tag{6.2}$$

where $\Delta = \partial_x^2 + \partial_y^2$ is the two-dimensional Laplacian. Note that Eq. (6.2) is the classical wave equation plus an $\mathcal{O}(h^2)$ dispersive correction. Ignoring the $\mathcal{O}(h^4)$ error term, we rearrange Eq. (6.2) to read $\Delta V = \frac{2}{3}\ell c\partial_t^2 V - \frac{1}{16}h^2\Delta^2 V$. Taking the Laplacian of both sides gives

$$\Delta\Delta V = \frac{2}{3}\ell c\Delta\partial_t^2 V - \frac{1}{16}h^2\Delta^3 V.$$

Substituting this expression in (6.2), we obtain, after again truncating the $\mathcal{O}(h^4)$ error term,

$$\ell c\partial_t^2 V = \frac{3}{2}\Delta V + \frac{\ell c}{16}h^2\partial_t^2\Delta V. \tag{6.3}$$

The reason we have replaced what is known as the "bad Bousinesq equation" (6.2) by the "good Bousinesq equation" (6.3) is that (6.2) has an unphysical blow-up at short wavelengths, while (6.3) does not [Rosenau, 1986; Whitham, 1974]. One may verify that the dispersion relation of (6.3) approximates the dispersion relation of the fully discrete equation (6.1) with an $\mathcal{O}(h^4)$ error.

Substituting $V(x, y, t) = e^{-\imath\omega t}U(x, y)$ into (6.3), we obtain the Helmholtz equation

$$\left(\nabla^2 + \gamma_h^2\right)U = 0, \qquad \text{where} \qquad \gamma_h^2 := \frac{16\ell c\omega^2}{24 - \ell c h^2\omega^2}. \tag{6.4}$$

Using standard techniques (see appendix A or [Bouwkamp, 1954; Born and Wolf, 1980; Goodman, 2004], we may then derive from (6.4) the dispersively corrected two-dimensional Rayleigh-Sommerfeld (RS) diffraction integral

$$U(x, y) = -\frac{\gamma_h x}{2\imath}\int_\Sigma U(\xi)\frac{H_1(\gamma_h|\mathbf{r}|)}{|\mathbf{r}|}d\xi. \tag{6.5}$$

Here the aperture Σ lies on the line $x = 0$. The quantity $|\mathbf{r}|$ is the magnitude of the vector $\mathbf{r}$ joining the point (x, y) to the point $(0, \xi)$ on the aperture. Equation(6.5) with γ_h replaced by $\gamma_0 = \sqrt{(2/3)\ell c}\omega$ is the traditional, non-dispersively corrected diffraction integral. Note that if γ_0 and h are both known, one can compute γ_h using $\gamma_h^{-2} = \gamma_0^{-2} - h^2/16$.

6.3 A comparison of diffraction integrals

In this section, we quantify the effect of dispersion on observed diffraction patterns. In other words, we investigate how the RS integral (6.5) changes when we use γ_h given by (6.4) instead of the non-dispersive quantity $\gamma_0 = (2/3)\ell c\omega^2$. First we take a 640×640 lattice with a 4 node

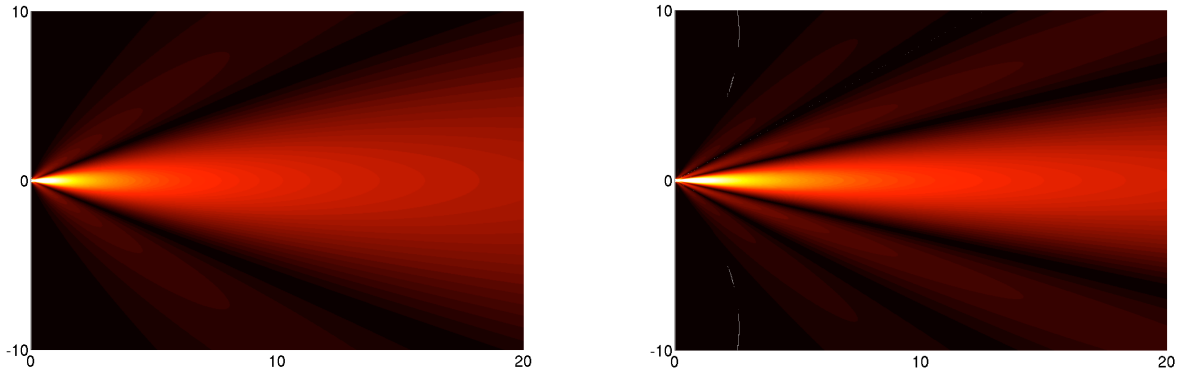


Figure 6.2: Comparison of traditional (left) and dispersively corrected (right) diffraction patterns, produced via numerical evaluation of (6.5) using γ_0 and γ_h, respectively. Note that dispersion changes the width and number of fringes in the diffraction pattern.

aperture centered along the left boundary. We set $h = 1/32$ and let $h\gamma_0 = \pi$ so that the effective wavelength is $\lambda_r = 2h$, close to the Bragg cutoff wavelength of $\sqrt{2}h$. We choose a constant input signal $U(\xi) = 1$ on the aperture. With these parameters, we evaluate (6.5) numerically, using both the dispersive γ_h and the non-dispersive γ_0.

Fig. 6.2 shows the intensity $|U|$ of the resulting non-dispersive and dispersive diffraction patterns. As shown, the net effect of dispersion is to compress the output in the y-direction. For example, it is evident that the non-dispersive pattern has three fringes, while the dispersive pattern has five. Though this result is for the constant input, extensive numerical tests have shown that if we choose a Gaussian, sinusoidal or other type of input, as long as $h\gamma_0 = \pi$, the qualitative effect of the dispersive correction is the same: the diffracted output is compressed in the y-direction.

For subsequent tests, we compute the output only along the right boundary of the domain. Consider first a 320×320 lattice with spacing $h = 1/16$ and an aperture width of 16 nodes. Again choosing a constant input signal $U(\xi) = 1$ on the aperture, we compute the traditional and dispersively-corrected diffracted outputs at $h\gamma_0 = \pi$—see the left panel of Fig. 6.3. For the right panel of Fig. 6.3, we sweep through values of $h\gamma_0$ and compute the full width at half max (FWHM) of the central peak for both types of diffracted outputs. The FWHM difference for $h\gamma_0 = \pi$ is $18h$, a substantial difference considering that the aperture width is only $16h$.

The results so far might lead one to believe that the effect of dispersion is important *only* when

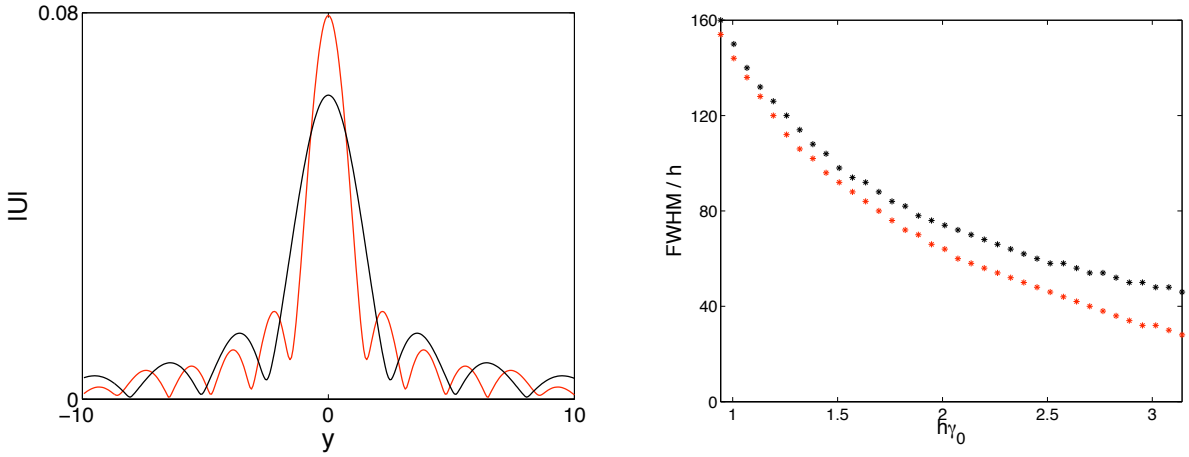


Figure 6.3: All plots are for a 320×320 lattice with a 16-node aperture and lattice spacing $h = 1/16$. On the left, we fix $h\gamma_0 = \pi$ and compute the dispersively corrected (red) and traditional (black) diffracted images at the far right of the domain. On the right, we show the full widths at half maximum (FWHM) of the central peak for dispersively corrected (red) and traditional (black) diffracted images, as the parameter $h\gamma_0 = hk_r$ is varied.

$h\gamma_0 = \pi$, *i.e.*, only when the effective wavelength is quite close to the Bragg cutoff. Our numerical tests have shown that this is true in the case when the input is a square wave. For our next test, we take a non-square wave input, *e.g.*,

$$U(\xi) = \begin{cases} (1/10)\sin^2(10\pi\xi) & \xi \in \Sigma \\ 0 & \xi \notin \Sigma \end{cases} \tag{6.6}$$

We set $h = 1/32$ and use an aperture width of 32 nodes on a 280×640 lattice. In this case, even when $h\gamma_0 = \pi/2$, the dispersive and non-dispersive diffraction patterns are significantly different in the tails. As shown in Fig. 6.4, the peaks of the red (dispersive) signal correspond to the zeros of the black (non-dispersive) signal. This is true even though the FWHM values for the two signals are indistinguishable.

6.4 Conclusion

By deriving a PDE model for lattice voltage dynamics that tracks the dispersive correction to $\mathcal{O}(h^2)$, we derived the dispersively corrected RS integral (6.5). Numerical experiments have confirmed that,

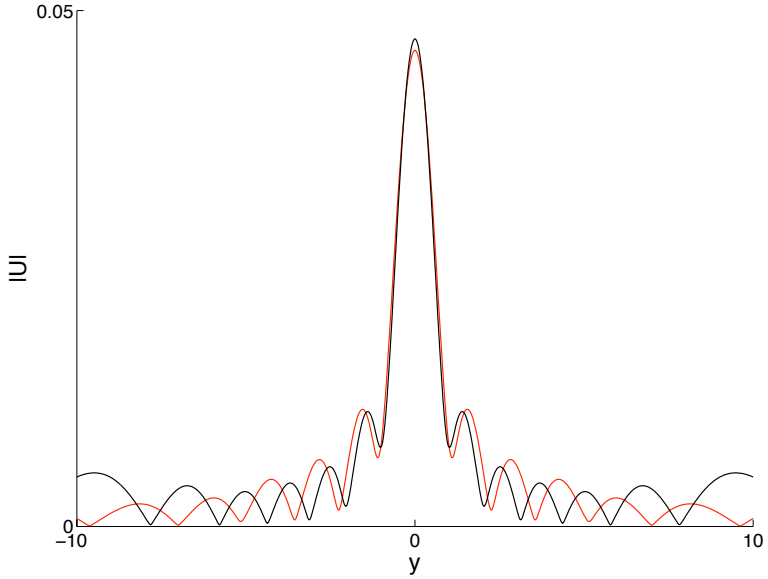


Figure 6.4: Traditional (black) and dispersively corrected (red) diffraction patterns at $h\gamma_0 = \pi/2$ for sinusoidal input given by (6.6). The output is taken at the right-hand side of a 280×640 lattice with spacing $h = 1/32$ and an aperture width of 32 nodes. Note that away from the central peak, the peaks of the red curve correspond to the zeros of the black curve.

when the effective wavelength is close to the Bragg cutoff, the dispersion of the medium causes a significant distortion in observed diffraction patterns: peak amplitudes, FWHM values, and fringe widths are all affected. Based on Fig. 6.4, we conjecture that for smaller values of $h\gamma_0$, there exist oscillatory input signals whose tail diffraction patterns can be substantially altered due to dispersion.

Part III

Spectral Optimization Problems Controlling Wave Phenomena

Part III of this thesis, comprising Chapters 7-10, focuses on spectral optimization problem controlling wave phenomena. Here, we summarize the mathematical structure of the problems considered. We recall from the introduction that the optimization problems considered here take one of the following forms:

Structural Optimization	Shape Optimization
$\min_{d(x) \in \mathcal{A}_d} \quad J(\lambda_j, u_j)$	$\min_{\Omega \in \mathcal{A}_\Omega} \quad J(\lambda_j, u_j)$
such that $\quad L(d)\ u_j = \lambda_j\ u_j \quad x \in \Omega$	such that $\quad L(\Omega)\ u_j = \lambda_j\ u_j \quad x \in \Omega$

In Ch. 7, we consider a shape optimization problem where L is the Dirichlet-Laplacian on $H^2(\Omega)$, $\mathcal{A}_\Omega$ is the set of star-shaped, symmetric, bounded planar regions with smooth boundary, and $J = J(\lambda_j)$ is taken to be (i) $J = -\frac{\lambda_n}{\lambda_1}$ and (ii) $J = -\frac{\lambda_n - \lambda_{n-1}}{\lambda_1}$ where λ_j denotes the j-th eigenvalue. The problem is studied numerically using quasi-Newton methods for $n \leq 13$.

In Ch. 8, we consider a structural optimization problem where $L = \frac{1}{d(x)-1}\Delta$ with outgoing boundary conditions defined on $H^1_{loc}(\mathbb{R}^d)$. The admissible set $\mathcal{A}_d$ is taken to be the set of $L^\infty(\mathbb{R}^d)$ functions with fixed compact support and pointwise upper and lower bounds. The eigenvalues of L are in the lower-half complex plane and the objective function is taken to be $J = \Im|\sqrt{\lambda}|$. We show analytically that an optimal $d \in \mathcal{A}_d$ exists and that it achieves the upper and lower bounds almost everywhere. The one-dimensional optimal structures are also studied numerically.

In Ch. 9, we consider a structural optimization problem where $L = \Delta + d$ defined on $H^1_{loc}(\mathbb{R}^d)$. The admissible set $\mathcal{A}_d$ is taken to be the set of compactly supported, H^1-bounded functions such that L has exactly one eigenpair (λ, ψ) such that $\lambda > -\mu$ where μ is a fixed positive constant. Let $f_\pm(x)$ be the outgoing solutions of $Lf_\pm = (\lambda + \mu)f_\pm$ at $x = \pm\infty$. Then for fixed, compactly supported $\beta(x)$ we consider the objective function $J = \sum_\pm |\langle \beta\psi, f_\pm \rangle|^2$. We show analytically that an optimal $d \in \mathcal{A}_d$ exists and study the properties of the one-dimensional optimal structures numerically. In contrast to the optimal structures in Ch. 8, the optimal structures appear to be interior points of the constraint set and to be smooth.

In Ch. 10, we consider a discretization of the following structural optimization problem. Define the operator $Lf = d_1\nabla \cdot (d_2 \nabla f)$ with dissipative boundary conditions on $H^1(\Omega)$ where $\Omega \subset \mathbb{R}^2$ is

compact. The admissible set $\mathcal{A}_d$ is taken to be the set of pairs (d_1, d_2) such that d_j for $j = 1, 2$ are pointwise bounded above and below. For a given pair $(d_1, d_2) \in \mathcal{A}_d$, the resolvent is defined in terms of the eigenpairs (λ_j, ψ_j) of L:

$$R_d(\gamma)[f] = \sum_j (\lambda_j - \gamma)^{-1} \langle \psi_j, f \rangle_{L^2(\Omega)} \psi_j.$$

Let Γ and Υ be disjoint segments of $\partial\Omega$ and let P_Γ and P_Υ be their respective trace operators. Define the operator $T_d \colon L^2(\Gamma) \to L^2(\Upsilon)$ to be the mapping $T_d := P_\Upsilon R_d P_\Gamma^t$. For a desired operator $\tilde{T}$ and fixed $\gamma \notin \sigma(L)$, we define the objective function $J = \|T_d(\gamma) - \tilde{T}\|_{HS}$, which depends on the eigenpairs of L through the resolvent. This structural optimization problem is discretized using the finite volume discretization developed in Ch. 3. Using numerical methods, we demonstrate that the discretized problem can be solved for a variety of desired operators $\tilde{T}$.

Chapter 7

Shape optimization of functions involving Dirichlet-Laplacian eigenvalues

7.1 Introduction

Denote by $\mathcal{D}$ the set of star-shaped, symmetric, bounded planar regions with smooth boundary. The Dirichlet-Laplace (D-L) eigenvalue problem for a region $D \in \mathcal{D}$ seeks eigenvalues $\lambda \in \mathbb{R}$ and eigenfunctions $u \in \mathcal{C}^2(D) \cap \mathcal{C}^0(\bar{D})$, nontrivial, such that

$$-\Delta u = \lambda u \qquad \mathbf{x} \in D \tag{7.1a}$$

$$u = 0 \qquad \mathbf{x} \in \partial D. \tag{7.1b}$$

There is a tremendous body of work studying the distribution of the D-L eigenvalues and the properties of D-L eigenfunctions—see, for example, [Kuttler and Sigillito, 1984; Trefethen and Betcke, 2006; Ashbaugh and Benguria, 2007; Henrot, 2006; Courant and Hilbert, 1953] and references within. Notably, there are a countable number of positive eigenvalues with no finite accumulation point. These eigenvalues are invariant under isometry of the domain (rotation and translation) and satisfy domain monotonicity (*i.e.*, larger regions have smaller eigenvalues: $D \subset D' \Rightarrow \lambda'_k \leq \lambda_k$).

Both of these facts are consequences of the max-min principle, stated

$$\lambda_k = \max_{\{v_j\}_{j=1}^{k-1}} \min_{\mathcal{Z}_{k-1}} \frac{\int_D |\nabla v|^2 \, \mathrm{d}\mathbf{x}}{\int_D v^2 \, \mathrm{d}\mathbf{x}} \tag{7.2}$$

where $\mathcal{Z}_{k-1} \equiv \{v \in H_0^1(\Omega) \setminus \{0\} \colon v \perp \{v_j\}_{j=1}^{k-1}\}$. The ratio in Eq. (7.2) is called the Rayleigh quotient. Low-lying eigenvalues satisfy numerous isoperimetric or universal inequalities, a few of which are discussed in §7.2. The distribution of D-L eigenvalues for large n satisfies Weyl's Law

$$\lambda_n(D) \sim 4\pi n A(D)^{-1} + o(n) \tag{7.3}$$

where $A(D)$ is the area of $D \in \mathcal{D}$ [Courant and Hilbert, 1953; Arendt *et al.*, 2009; Kuttler and Sigillito, 1984]. Each eigenfunction is smooth (C^∞) on D and zero on a set of C^∞ curves referred to as *nodal lines* with well-known properties. Closed form expressions for the eigenfunctions cannot generally be obtained unless the domain can be transformed into a separable coordinate system. If the domain has symmetry, the eigenfunctions are either even or odd with respect to the axis of symmetry, simplifying their computation. The D-L eigenvalue problem arises in a number of physical, engineering, and mathematical contexts including the study of vibrating membranes, electromagnetism, acoustic wave propagation, heat flow, the semi-classical approximation of quantum bound states, and number theory.

We denote by $\Lambda_n(D)$ the first n increasingly-ordered D-L eigenvalues of a domain $D \in \mathcal{D}$ counting multiplicity and refer to the mapping $\Lambda_n \colon \mathcal{D} \to \mathbb{R}^n$ as the *D-L eigenvalue operator*. In this article, we study optimization problems of the form

$$\max_{D \in \mathcal{D}} F \circ \Lambda_n(D) \tag{7.4}$$

where $F \colon \mathbb{R}^n \to \mathbb{R}$ is invariant to permutation of its arguments, *i.e.*, $F(\mathbf{x}) = F(\pi(\mathbf{x}))$ where $\pi(\mathbf{x})$ is a reordering of the components of $\mathbf{x} \in \mathbb{R}^n$. For such F, the composition $F \circ \Lambda_n$ is referred to as a *spectral function* [Borwein and Lewis, 2000, §5.2]. Eq. (7.4) is a constrained shape optimization problem where the constraints are given by n D-L eigenvalue equations. We generally refer to problem (7.4) as an *eigensystem-constrained shape optimization problem*.

One may interpret Eq. (7.4) as a method to study the range of the D-L eigenvalue operator.

To this end, we consider two particular nonsmooth functions for $F(\mathbf{x})$, given by

$$R_n(\mathbf{x}) = \frac{[\mathbf{x}]_n}{[\mathbf{x}]_1} \tag{7.5}$$

$$G_n(\mathbf{x}) = \frac{[\mathbf{x}]_n - [\mathbf{x}]_{n-1}}{[\mathbf{x}]_1} \tag{7.6}$$

where $[\mathbf{x}]_p$ denotes the p-th smallest component of $\mathbf{x} \in \mathbb{R}^n$. The spectral function $r_n(D) \equiv R_n \circ \Lambda_n(D)$ is the ratio of the n-th to first D-L eigenvalues of the domain $D \in \mathcal{D}$. The spectral function $g_n(D) \equiv G_n \circ \Lambda_n(D)$ measures the gap between $r_n(D)$ and $r_{n-1}(D)$. Both $r_n(D)$ and $g_n(D)$ are invariant to translation, rotation, and dilation of the region $D \in \mathcal{D}$, so no additional constraints need be imposed in Eq. (7.4).

Results and Outline Our findings can be summarized:

1. In §7.3-§7.5, a BFGS quasi-Newton method is developed to solve the general eigensystem-constrained optimization problem in Eq. (7.4). In §7.3, we discuss the representation of the domain $D \in \mathcal{D}$ by Fourier-cosine coefficients, $\{b_k\}_{k=1}^{\infty}$ and a finite-dimensional approximation to $\mathcal{D}$, denoted $\mathcal{D}_m$. In §7.4, we compute the gradient of the objective function with respect to the Fourier-cosine coefficients, b_k. Then in §7.5, we discuss a numerical implementation of the method.

2. In §7.6 and §7.7, the method is applied to the objective functions in Eqs. (7.5) and (7.6) for $n = 2, \ldots, 13$. The optimal values are given in Table 7.2 and the achieving regions are plotted in Figs. 7.3 and 7.4. For all n considered, the domain D_n^* maximizing either $r_n(D)$ or $g_n(D)$ has eigenvalues satisfying $\lambda_n(D_n^*) = \lambda_{n+1}(D_n^*)$. The results for both objective functions extend and support earlier work on the Payne-Pólya-Weinberger inequality and an eigenvalue multiplicity conjecture by Ashbaugh and Benguria [Ashbaugh and Benguria, 1992a].

7.2 Background and related work

Two well-written and extensive recent manuscripts on isoperimetric inequalities involving D-L eigenvalues can be found in [Ashbaugh and Benguria, 2007; Henrot, 2006]. The oldest and best-known such inequality is the Rayleigh-Faber-Krahn inequality, originally conjectured by Lord

Rayleigh in 1894, stating that $\min \lambda_1(D)$ over the set of all membranes of fixed area is attained only by the disk.

The problems of maximizing $r_n(D)$ for $n = 2, 3, 4$ have been considered by many authors [Ashbaugh and Benguria, 2007; Henrot, 2006]. In 1955, Payne, Pólya, and Weinberger (PPW) showed that $r_2(D) \leq 3$ for all smooth bounded domains and correctly conjectured that the optimal value is attained by the disk [Payne *et al.*, 1956]. This bound was studied numerically [Haeberly, 1991] and improved many times until finally being proved in 1992 by Ashbaugh and Benguria (AB) [Ashbaugh and Benguria, 1992b] and the corresponding inequality now bears the PPW name. With this proof, AB established that for the region D^* (=disk) attaining the optimal value $r_2^* = \max r_2(D) \approx 2.539$, we have $\lambda_2(D^*) = \lambda_3(D^*)$. Subsequently, the range of the first two D-L eigenvalues has been studied numerically [Wolf and Keller, 1994] and analytically (see [Henrot, 2006, §6.4]). In 2003, after numerically searching through 65,000 trial and error regions, Levitin and Yagudin (LY) conjectured that $r_3^* \lesssim 3.202$ [Levitin and Yagudin, 2003]. For the dumbell-shaped region D^* with largest value r_3, they found $\lambda_3(D^*) = \lambda_4(D^*)$, supporting an earlier conjecture of AB [Ashbaugh and Benguria, 1992a]. In 1993, AB gave a bound for $n = 4$ stated $r_4^* \leq (r_2^*)^2 \approx 6.445$ [Ashbaugh and Benguria, 1993].

Recently, there has been much work on the value of r_n for larger values of n. Cheng and Yang have shown that

$$r_{n+1} \leq \frac{\sqrt{41}}{3} n \approx 2.134\, n \tag{7.7}$$

for $n \geq 3$ [Cheng and Yang, 2007] and Harrell and Hermi have shown $r_{n+1} \leq \frac{21}{8} n = 2.625\, n$ [Harrell and Hermi, 2008]. Taking $n = 3$, we have: $r_4^* \lesssim 6.402$, a slight improvement over the bound given by AB. From Weyl's Law (7.3) and the Rayleigh-Faber-Krahn inequality, we expect that asymptotically

$$r_n^*(D) \sim \frac{4}{\alpha_{0,1}^2} n \approx .6916\, n \qquad \text{as } n \uparrow \infty \tag{7.8}$$

where $\alpha_{\ell,k}$ is the kth zero of the ℓth Bessel function of the first kind [Harrell and Hermi, 2008]. The factor of ≈ 3.085 that exists between the best isoperimetric bound (7.7) and the asymptotic bound (7.8) suggests that more careful analysis might yet result in better constants for large n. The present work indicates that the same is true for small to moderate values of n.

The difficulty in establishing analytic upper bounds for r_2 (PPW inequality) lies in the fact that at the optimal solution, $\lambda_2 = \lambda_3$ and also, $R_2(\mathbf{x})$ is not $\mathcal{C}^1$ at points $\mathbf{x} \in \mathbb{R}^n$ such that $[\mathbf{x}]_2 = [\mathbf{x}]_3$.

We will see in §7.6 and §7.7 that coalescing eigenvalues also has consequences for the computational treatment of Eq. (7.4).

There are also several isoperimetric inequalities relevant to the study of g_n. Of course, the problem of maximizing $g_n(D)$ for $n = 2$ is equivalent to maximizing $r_2(D) - 1$. The inequality

$$g_{n+1}(D) \leq \frac{2}{n} \sum_{j=1}^{n} r_j(D) \tag{7.9}$$

was proved for bounded domains and all n by PPW in 1956 [Payne *et al.*, 1956]. There are many generalizations of this inequality involving sums of ratios of eigenvalues and their differences which provide tighter bounds on g_n^* than Eq. (7.9) for certain values of n [Ashbaugh and Benguria, 2007]. There has also been much work on *minimizing* the eigenvalue gap $\lambda_2 - \lambda_1$ (fundamental gap) for bounded, convex domains [Ashbaugh and Benguria, 2007].

We note that there are also many results for the similar problem of minimizing $\lambda_k(D)$ over the set of domains with fixed area [Henrot, 2006, §5] [Bratus and Myshkis, 1992]. Of particular interest here is Édouard Oudet's work [Oudet, 2004] on the computation of optimal shapes for this problem for $k = 3, \ldots, 10$ using a level-set method for the representation and evolution of the domain, a projected-gradient method for handling the volume constraint, and a (relaxed-formulation) finite-element method for the computation of the eigenvalues.

In 1966, Mark Kac notoriously posed the inverse problem: Can one hear the shape of a drum? [Kac, 1966; Okada *et al.*, 2005]. It is noted in [Ashbaugh and Benguria, 2007] that in fact, one interpretation of the PPW inequality is that one can hear whether or not a drum is circular by just the first two frequencies. This prompts the following (much easier) existence question:

> Given a sequence of n real numbers, could these be the first n D-L eigenvalues for *some* planar region?

The isoperimetric inequalites above and the work in this article provide necessary conditions for such a sequence.

We briefly mention that the answer to this question is known for a few related operators. (1) The one-dimensional D-L spectrum is simply determined by the length of the domain, L (recall $\lambda_n = (n\pi/L)^2$) so that only squared arithmetic sequences are attained. (2) For any given sequence of n distinct numbers $\{\lambda_j\}_{j=1}^n$, there is an n^2-dimensional set of complex $n \times n$ matrices, given by

$\{U\mathrm{diag}(\lambda)U^T : U \text{ unitary}\}$, each having this as spectrum. A recent book on (finite-dimensional) inverse eigenvalue problems discusses extensions of this simple result [Chu and Golub, 2005]. (3) Lastly, there exists a family of potentials $q(x) \in L^2([0,1])$ such that the spectrum of the one-dimensional Schrödinger operator, $\partial_x^2 + q$, with Dirichlet boundary conditions is given by any sequence of n numbers [Pöschel and Trubowitz, 1987, §6].

Finally, we discuss three applications where this work has direct relevance. In accelerator physics, it is desirable to find the cavity shape that maximizes the quality factor of a fixed accelerating mode while minimizing the quality factor of (parasitic) higher-order modes [Akcelik *et al.*, 2005; Akcelik *et al.*, 2008]. This design problem can be formulated as an eigensystem-constrained shape optimization problem of the form in Eq. (7.4). Analogous to a Helmholtz resonating cavity, the boundary conditions for this problem are non-self-adjoint. The self-adjoint problem considered in this article is a model eigensystem-constrained optimization problem for this application.

Photonic crystals (PCs) are now being used for optical control and manipulation. As such, there has been much work designing the index of refraction in materials to create spectral gaps of maximal length or have resonances with maximum lifetime. These can be formulated as optimization problems with objective functions of the form in Eq. (7.4) [Heider *et al.*, 2008; Osher and Santosa, 2001; Dobson and Santosa, 2004; Kao and Santosa, 2008; Men *et al.*, 2010; Harrell and Svirsky, 1986; Svirsky, 1987].

Recently, ratios of D-L eigenvalues have been used as feature vectors in shape recognition and classification of binary images because these quantities are invariant to translation, rotation, and dilation [Khabou *et al.*, 2007]. Understanding the range and stability of Λ_n is necessary to study the tolerance to noise and uncertainty quantification within this feature space.

7.3 Representation of the domain by Fourier-cosine coefficients

Before discussing the representation of a domain $D \in \mathcal{D}$ by Fourier-cosine coefficients, we review a few definitions. A planar region D is said to be symmetric with respect to the $\hat{x}$-axis if $(x, y) \in D$ implies $(x, -y) \in D$ and *symmetric* if it is symmetric with respect to some line. A planar region D is *star-shaped* if there exists a point $\mathbf{x}_0$ such that for all $\mathbf{x}$ in D the line segment from $\mathbf{x}_0$ to $\mathbf{x}$ is contained in D. As above, we denote by $\mathcal{D}$ the set of star-shaped, symmetric, bounded planar

regions with smooth boundary. Two planar regions are *isometric* if they are related by a rotation and translation.

Every region $D \in \mathcal{D}$ is isometric to a domain D' which can be represented in polar coordinates by a Fourier-cosine expansion

$$D' = \{(r,\theta)\colon\ 0 < r < \sum_{k=0}^{\infty} b_k \cos(k\theta)\}.$$

This representation is not unique since the Fourier-cosine coefficients are not preserved by rotations by π and small translations along the x-axis. (Here "small" refers to the fact that star-shapeness with respect to the origin is lost for large translations.) In Appendix 7.A, the transformation of coefficients is explicitly given for these two isometries.

7.3.1 Representation of the domain by a truncated Fourier-cosine series

Let $\mathcal{D}_m \subset \mathcal{D}$ be the set of regions such that $D \in \mathcal{D}_m$ if and only if D is isometric to a domain D' which can be represented by the truncated Fourier-cosine expansion

$$D' = \{(r,\theta)\colon\ 0 < r < \sum_{k=0}^{m} b_k \cos(k\theta)\}. \tag{7.10}$$

We now approximate the admissible set $\mathcal{D}$ in Eq. (7.4) by $\mathcal{D}_m$ and consider

$$\max_{D\in\mathcal{D}_m} F \circ \Lambda_n(D). \tag{7.4'}$$

The following proposition due to Cox and Ross [Cox and Ross, 1995], gives an upper bound on the error of the n-th eigenvalue due to truncation of the Fourier-cosine series.

Proposition 7.3.1. *Let $\{b_k\}_{k=0}^{\infty}$ be the Fourier-cosine coefficients for a domain $D_\infty \in \mathcal{D}$ and define the partial sum and corresponding domain*

$$\rho_m(\theta) = \sum_{k=0}^{m} b_k \cos(k\theta)$$

$$D_m = \{(r,\theta)\colon\ 0 < r < \rho_m(\theta)\}.$$

Suppose r_* is such that $\rho_m(\theta) > r_*$ and $\rho_\infty(\theta) > r_*$ for all $\theta \in [0, 2\pi]$. Then,

$$|\lambda_n(D_m) - \lambda_n(D_\infty)| \le \frac{3}{r_*}\lambda_n(B_{r_*}) \sum_{k\ge m+1} |b_k|$$

where $\lambda_n(B_{r_*})$ is the n-th eigenvalue of the disk of radius r_*.

Suppose that for a given function F the maximum of Eq. (7.4) exists and is attained by D^* with r_* as in Prop. 7.3.1. Suppose further that D^* is well represented by Fourier-cosine coefficients b_k, *i.e.*, there exists an M and $\epsilon < r_*$ such that

$$\sum_{k \geq M+1}^{\infty} |b_k| < \epsilon.$$

If F is Lipschitz continuous in a neighborhood of $\Lambda_n(D^*)$, Prop. 7.3.1 implies that the solution of Eq. (7.4') for $m = M$ is a good approximation to the solution of (7.4). Thus, our strategy is to solve (7.4') for increasing m until further increasing m has no appreciable effect on the solution.

With the non-uniqueness of this representation (as described above and in App. 7.A) in mind, in what follows we nevertheless abusively identify $D \in \mathcal{D}_m$ with a sequence $\{b_k\}_{k=1}^m$ such that $\sum_{k=0}^m b_k \cos(k\theta) > 0$ for all $\theta \in [0, \pi)$.

7.4 Eigenvalue perturbation formulae

We compute the derivative of a D-L eigenvalue with respect to a change in the Fourier-cosine coefficient b_k, as illustrated in Fig. 7.1.

Proposition 7.4.1. *Let a domain $D \in \mathcal{D}_m$ be represented by Fourier-cosine coefficients $\{b_k\}_{k=0}^m$ as in Eq. (7.10) and let $\rho(\theta) = \sum_{k=0}^m b_k \cos(k\theta)$. If (λ, u) is a simple, unit-normalized eigenpair satisfying Eq. (7.1) then*

$$\frac{\partial \lambda}{\partial b_k} = -2 \int_0^\pi \rho(\theta) \cos(k\theta) |\nabla u(\rho, \theta)|^2 \, \mathrm{d}\theta. \tag{7.11}$$

Proof. Differentiating Eq. (7.1) with respect to a coefficient b_k results in the Hadamard variational formula

$$\frac{\partial \lambda}{\partial b_k} = \frac{-\int_{\partial D} c_k |\nabla u|^2 \, \mathrm{d}\mathbf{x}}{\int_D u^2 \, \mathrm{d}x}, \tag{7.12}$$

where $c_k \equiv \frac{\partial \mathbf{x}}{\partial b_k} \cdot \hat{\mathbf{n}}$ is the (outward) velocity of the boundary ∂D resulting from a perturbation in b_k [Rellich, 1969, §2.6]. Denoting $\rho(\theta) = \sum_{k=0}^m b_k \cos(k\theta)$ we compute the (outward) normal vector

$$\hat{\mathbf{n}} = \frac{\rho \hat{\mathbf{r}} - \rho' \hat{\theta}}{\sqrt{\rho^2 + (\rho')^2}}.$$

Then using $\frac{\partial \mathbf{x}(\theta)}{\partial b_k} = \cos(k\theta)\hat{\mathbf{r}}$, we find

$$c_k = \frac{\partial \mathbf{x}(\theta)}{\partial b_k} \cdot \hat{\mathbf{n}} = \frac{\rho \cos(k\theta)}{\sqrt{\rho^2 + (\rho')^2}}.$$

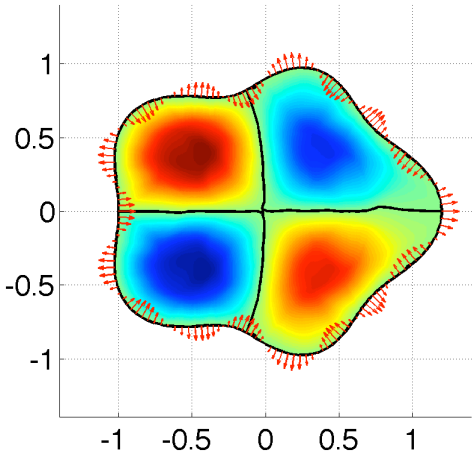


Figure 7.1: The 5th eigenfunction of the domain $r < 1 + .1\cos(2\theta) + .1\cos(5\theta)$ with corresponding eigenvalue $\lambda_5 = 26.87$. The black lines are nodal lines. The red arrows represent $c_9\hat{\mathbf{n}}$, the "velocity of the boundary" resulting from a perturbation in b_9. The rate of change of the 5th eigenvalue due to this perturbation is $\frac{\partial \lambda_5}{\partial b_9} = -12.74$.

Thus for the unit normalized eigenfunction $u(r,\theta)$, from Eq. (7.12) we obtain

$$\begin{aligned}\frac{\partial \lambda}{\partial b_k} &= -\int_0^{2\pi} \frac{\rho(\theta)\cos(k\theta)|\nabla u(\rho,\theta)|^2}{\sqrt{\rho^2 + (\rho')^2}} \left|\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\theta}\right| \mathrm{d}\theta \\ &= -2\int_0^{\pi} \rho(\theta)\cos(k\theta)|\nabla u(\rho,\theta)|^2 \,\mathrm{d}\theta\end{aligned}$$

as desired. This formula may also be obtained using the adjoint method. □

Eigenvalues of multiplicity greater than one are differentiable, but the computation of the derivative is impeded because one must identify the proper basis for the corresponding unperturbed eigenfunction subspace [Courant and Hilbert, 1953; Rellich, 1969; Henry, 2005]. However, the compositions $r_n = R_n \circ \Lambda_n$ and $g_n = G_n \circ \Lambda_n$ are not differentiable at regions with multiple eigenvalues. Thus, the computation for multiple eigenvalues is irrelevant for the two objective functions considered here. We'll further explain in §7.5 that since almost all domains have simple eigenvalues [Kuttler and Sigillito, 1984], it is possible to choose an optimization method that behaves well despite these non-differentiable points.

The gradient of the objective function $f = F \circ \Lambda_n$ in Eq. (7.4) can now be computed using the

chain rule

$$\frac{\partial f}{\partial b_k} = \sum_{j=1}^{n} \frac{\partial \lambda_j}{\partial b_k} \frac{\partial F(\mathbf{x})}{\partial x_j}\bigg|_{\mathbf{x}=\Lambda_n(D)}. \tag{7.13}$$

Using this calculation, necessary conditions for a region attaining the maximum of f can be given, including the case where the attaining region has multiple eigenvalues [Henrot, 2006; Bratus and Myshkis, 1992], but this will not be pursued here.

7.5 Computational method and implementation

To solve the optimization problem (7.4), we will utilize a quasi-Newton method which requires the computation of the cost function F and gradient $\frac{\partial F}{\partial b_k}$, found in Eq. (9.45b). These in turn involve the computation of Λ_n and the Jacobian $\frac{\partial \lambda_j}{\partial b_k}$.

Eigenvalues For a given domain $D \in \mathcal{D}$, we use the method of particular solutions (MPS) to compute the D-L eigenvalues, $\Lambda_n(D)$ and eigenfunctions [Fox *et al.*, 1967; Moler and Payne, 1968; Betcke and Trefethen, 2005; Barnett, 2009]. We choose two sets of particular solutions

$$u_1^{(k)}(r,\theta) = J_k(\sqrt{\lambda} r)\cos(k\theta) \tag{7.14a}$$

$$u_2^{(k)}(r,\theta) = J_k(\sqrt{\lambda} r)\sin(k\theta) \tag{7.14b}$$

defined on the upper half-plane where J_k is the k-th order Bessel function of the first kind. The first, extends evenly over the $\hat{x}$-axis and the second, oddly. For collocation points (r_i, θ_i), we use for $i = 1, \ldots, M_B$ uniformly distributed points on the boundary and for $i = M_B+1, \ldots, M_B+M_I \equiv M$, points distributed on the interior of the domain D. We discuss the location of the interior points below. Approximate eigenfunctions are sought by constructing a linear combination

$$u = \sum_{k=1}^{K} z_k u_{(\cdot)}^{(k)}(r,\theta), \tag{7.15}$$

forming the matrix $[A(\lambda)]_{ik} = u_{(\cdot)}^{(k)}(r_i, \theta_i)$, finding the QR decomposition

$$A(\lambda) = \begin{bmatrix} A_B(\lambda) \\ A_I(\lambda) \end{bmatrix} = \begin{bmatrix} Q_B(\lambda)R \\ Q_I(\lambda)R \end{bmatrix},$$

and considering $\sigma_1(\lambda)$, the smallest singular value of $Q_B(\lambda)$. Since the columns of Q_B are the range of A_B, when $\sigma_1(\lambda)$ is small, λ is an approximate D-L eigenvalue [Betcke and Trefethen, 2005].

For a region with a line of symmetry (which is assumed here by the definition of $\mathcal{D}_m$), all eigenfunctions within the eigenspace can be taken to be symmetric or antisymmetric with respect to that line [Trefethen and Betcke, 2006]. For an eigenvalue of multiplicity two, if one eigenfunction is even and one eigenfunction is odd with respect to the $\hat{x}$-axis, the use of the two sets of basis functions in Eq. (7.14) avoids the problem of distinguishing two eigenvalues as they coalesce.

Basis functions other than the Fourier-Bessel functions in Eq. (7.14) such as plane waves and fundamental solutions have also been used for the MPS. To distinguish eigenvalues of multiplicity two as above, one could consider both Dirichlet and Neumann boundary conditions separately on the $\hat{x}$-axis. For regions with smooth boundary, fundamental solutions have even proven to yield more accurate eigenvalues [Barnett, 2009; Barnett and Betcke, 2008], however placement of the charge points is challenging for an *a priori* unknown and evolving domain.

One of the advantages of the MPS is the following theorem due to Moler-Payne [Moler and Payne, 1968].

Theorem 7.5.1. *Let (u, λ) be an approximate normalized eigenpair satisfying Eq. (7.1a) but not necessarily (7.1b). Then there exists an eigenvalue $\bar{\lambda}$ such that*

$$\frac{|\lambda - \bar{\lambda}|}{\bar{\lambda}} \leq C_D \|u\|_{\partial D}$$

where C_D is a constant that may depend on the region D and $\|\cdot\|_{\partial D}$ is the L^2-norm on the boundary.

Thus the size of the smallest singular value σ_1 controls the relative error of a proposed eigenvalue λ. Monitoring the value of σ_1 in eigenvalue computations, we report 4 significant figures everywhere in this manuscript, all of which are believed to be accurate. To achieve this, we use approximately $M_B = 250$ boundary points and $M_I = 350$ interior points.

To verify our numerical implementation of the MPS method, we also compared computed D-L eigenvalues for the ellipse against well-known expressions involving the zeros of Mathieu functions.

Jacobian To compute the Jacobian $\frac{\partial \lambda_j}{\partial b_k}$ in Eq. (7.11), we require normalized eigenfunctions and the normal derivative of the eigenfunction on the boundary $\nabla u_j \cdot \hat{\mathbf{n}}$. Thus some of the M_I interior points are distributed for normalization which is done via quadrature and the remaining points are

used to evaluate either a first- or second-order finite difference formula

$$\nabla u \cdot \hat{\mathbf{n}}(\mathbf{x}) = h^{-1}[u(\mathbf{x}) - u(\mathbf{x} - h\hat{\mathbf{n}})] + \mathcal{O}(h^2) \tag{7.16a}$$

$$\nabla u \cdot \hat{\mathbf{n}}(\mathbf{x}) = (2h)^{-1}[-3u(\mathbf{x}) + 4u(\mathbf{x} - h\hat{\mathbf{n}}) - u(\mathbf{x} - 2h\hat{\mathbf{n}})] + \mathcal{O}(h^3). \tag{7.16b}$$

This method of computing the Jacobian is much faster than the alternative of evaluating an eigenvalue finite difference formula for $m + 1$ different small perturbations of the domain, each corresponding to a change in $\{b_k\}_{k=0}^m$. The pointwise errors in $u(\mathbf{x} - h\hat{\mathbf{n}})$ and $u(\mathbf{x} - 2h\hat{\mathbf{n}})$ in Eq. (7.16) can also be bounded, see Moler-Payne [Moler and Payne, 1968].

Optimization For the optimization problem in Eq. (7.4), a BFGS quasi-Newton method [Nocedal and Wright, 2006] is used with an inexact line search as described in [Lewis and Overton, 2009] and implemented in Matlab [Overton, 2010]. Since the objective function is non-convex and non-smooth (at regions with multiple eigenvalues – see §7.2), it is not guaranteed that the optimization method will converge to a local maximum. However, it has been conjectured that for random initial conditions, this method almost surely generates an infinite sequence of iterates converging to a local maximum [Lewis and Overton, 2009].

Convergence is difficult to establish if the optimal region has multiple eigenvalues. In addition, the translation and dilation symmetries of the Fourier coefficient representation discussed in §7.3 means that the BFGS Hessian approximation will only be negative semi-definite near an optimal solution. For these reasons, in our implementation we terminate the method if the line search cannot find a sufficient decrease in the objective function.

For each cost function in the following sections, we apply this method using many randomly initialized points. We find that we are generally unable to reduce the objective function after approximately 30-50 iterations. Using continuation on the number of Fourier-cosine coefficients improves convergence and may help avoid local maxima. All computations were performed using Matlab on a 2.4 GHz dual core processor with 2GB memory.

D-L Eigenvalues of the Unit Disk

$\ell\backslash k$		1		2		3
0	(1)	5.783	(6)	30.47	(15)	74.88
1	(2,3)	14.68	(9,10)	49.21	(22,23)	103.4
2	(4,5)	26.37	(13,14)	70.84	(28,29)	135.0
3	(7,8)	40.70	(18,19)	95.27	(35,36)	169.3
4	(11,12)	57.58	(24,25)	122.4	(43,44)	206.5

Table 7.1: Eigenvalues of the unit disk, $\alpha^2_{\ell,k}$. The numbers in parenthesis are the ordering $j = \mathcal{I}^\circ(\ell,k)$, counting multiplicity.

7.6 Dirichlet-Laplacian eigenvalue ratios

Let $r_n(D) = \frac{\lambda_n(D)}{\lambda_1(D)}$ be the ratio of the n-th smallest to the smallest D-L eigenvalue of the region D and consider the shape optimization problem

$$\max_{D\in\mathcal{D}_m} r_n(D). \tag{7.17}$$

Before presenting the numerical solution of this problem, for reference we discuss the value of r_n for disks, rectangles, and equilateral triangles.

Disks Let $\alpha_{\ell,k}$ be the kth zero of the ℓth Bessel function of the first kind, denoted $J_\ell(r)$. The D-L eigenvalues of the unit disk are given by $\alpha^2_{\ell,k}$ with corresponding eigenfunctions $J_\ell(\alpha_{\ell,k}r)\cos(\ell\theta)$ and $J_\ell(\alpha_{\ell,k}r)\sin(\ell\theta)$. For $\ell \geq 1$ the eigenvalues have multiplicity 2. Counting multiplicity, we order the eigenvalues and label them λ°_j where $j = \mathcal{I}^\circ(\ell,k)$ is the eigenvalue ordering. These values are given in Table 7.1. We then define $r^\circ_n \equiv r_n(\text{disk}) = \lambda^\circ_n/\lambda^\circ_1$ and list these values in the first column of Table 7.2(a).

Rectangles The eigenvalues of the $a\times 1$ rectangle denoted R_a where $a > 1$ are given by $\beta^a_n = \pi^2(\ell^2 + k^2/a^2)$ where $n = \mathcal{I}^a(\ell,k)$ gives the ordering. The ratio of eigenvalues is

$$r_n(R_a) = \beta^a_n/\beta^a_1 = \frac{\ell^2a^2 + k^2}{a^2+1}. \tag{7.18}$$

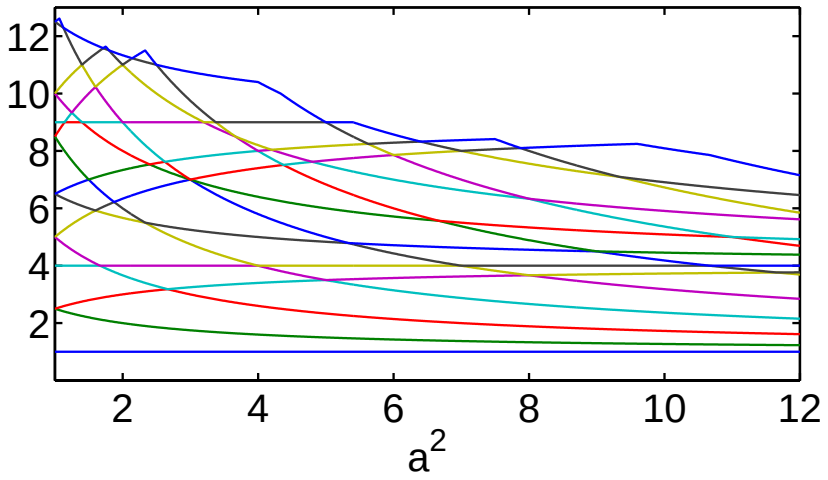


Figure 7.2: The squared aspect ratio, a^2 vs. $r_n(R^a)$ in Eq. (7.18) for $n = 1, \ldots, 15$.

We define $r_n^{\square} \equiv \max_{a \geq 1} r_n(R_a)$ and tabulate these values in Table 7.2(a). These values are given by the maximum over a of the n-th lowest curve in Fig. 7.2, which is precisely the eigenvalue avoidance plot in [Betcke and Trefethen, 2004; Trefethen and Betcke, 2006]. We note a few interesting features of this figure:

1. The optimal value $r_n^{\square}$ can be attained for a single value of a, a collection of values of a (*e.g.*, for $n = 8$, the optimal value is attained at both $a = \sqrt{3/2}$ and $a = \sqrt{3}$), or even an interval (*e.g.*, for $n = 4$ the optimal value is attained for $a = [1, \sqrt{5/3}]$).

2. Because the curves in Fig. 7.2 are monotonic between crossings, for all n the optimal value $r_n^{\square}$ is attained by a rectangle for which $\lambda_n = \lambda_{n+1}$.

3. Local optima are present in this figure. For example, the optimal value $r_4^{\square} = 4$ is attained by the interval $a = [1, \sqrt{5/3}]$ as noted above. However, there is a local optima at $a = \sqrt{5} \approx 2.2$ with value $r_4(R_{\sqrt{5}}) = 3.5$.

4. There is a triple eigenvalue crossing for $a = \sqrt{3}$ where $(\ell, k) = (3, 1)$, $(2, 4)$, and $(1, 5)$ each correspond to the eigenvalue $\beta_8^{\sqrt{3}} = 28\pi^2/3$ with multiplicity three. There exist D-L eigenvalues of a square with arbitrarily large multiplicity [Kuttler and Sigillito, 1984].

5. When a^2 is irrational, the eigenvalues of R_a are all simple. Thus, almost all rectangles have

(a) D-L eigenvalue ratios

n	r_n°	$r_n^\triangle$	$r_n^\square$	r_n^*
2	2.538	2.333	2.5	2.538
3	2.538	2.333	3.181	3.202
4	4.560	4	4	4.560
5	4.560	4.333	5	5.126
6	5.268	4.333	5.923	6.198
7	7.038	6.333	6.5	7.129
8	7.038	6.333	7	7.811
9	8.510	7	8.5	8.809
10	8.510	7	9	9.616
11	9.956	9	9.333	10.34
12	9.956	9.333	10.23	11.01
13	12.25	9.333	11	12.25

(b) D-L eigenvalue gaps

n	g_n°	$g_n^\triangle$	$g_n^\square$	g_n^*
2	1.538	1.333	1.5	1.538
3	0	0	1.363	1.449
4	2.021	1.666	1.5	2.087
5	0	.3333	1	1.814
6	0.708	0	1.875	2.221
7	1.769	2	1.5	2.652
8	0	0	1.75	1.862
9	1.471	.6666	2	2.579
10	0	0	1.8	2.431
11	1.446	2	1.173	3.125
12	0	.3333	1.379	2.637
13	2.294	0	2	2.557

Table 7.2: Optimal values of r_n and g_n attained for disks, equilateral triangles, all rectangles, and $\mathcal{D}_m$. All decimal values are truncated (not rounded up) at four significant digits.

simple eigenvalues [Courant and Hilbert, 1953].

6. It is clear from Fig. 7.2 that the optimal value of the (opposite) optimization problem: $\min_D r_n(D)$ is 1 for all n. This optimal value is approached by the rectangle with aspect ratio tending to infinity. In this limit, the spectrum becomes continuous.

Equilateral Triangles The eigenvalues of an equilateral triangle $\triangle$ were first computed by Gabriel Lamé and are given by $\lambda_n^\triangle = \frac{16\pi^2}{27}(\ell^2 + k^2 - \ell k)$ where $n = \mathcal{I}^\triangle(\ell, k)$ is the ordering (see, for example, [McCartin, 2003; Pinsky, 1980]). The integer pairs (ℓ, k) are required to satisfy $\mod(\ell + k, 3) = 0$, $\ell + k \neq 0$, $\ell \neq 2k$, and $k \neq 2\ell$. The first 15 eigenvalues are written

$$\Lambda_{15}(\triangle) = \frac{16\pi^2}{27}\,(9, 21, 21, 36, 39, 39, 57, 57, 63, 63, 81, 84, 84, 93, 93)\,.$$

We denote by $r_n^\triangle \equiv r_n(\triangle) = \lambda_n^\triangle/\lambda_1^\triangle$ and tabulate these values in Table 7.2(a).

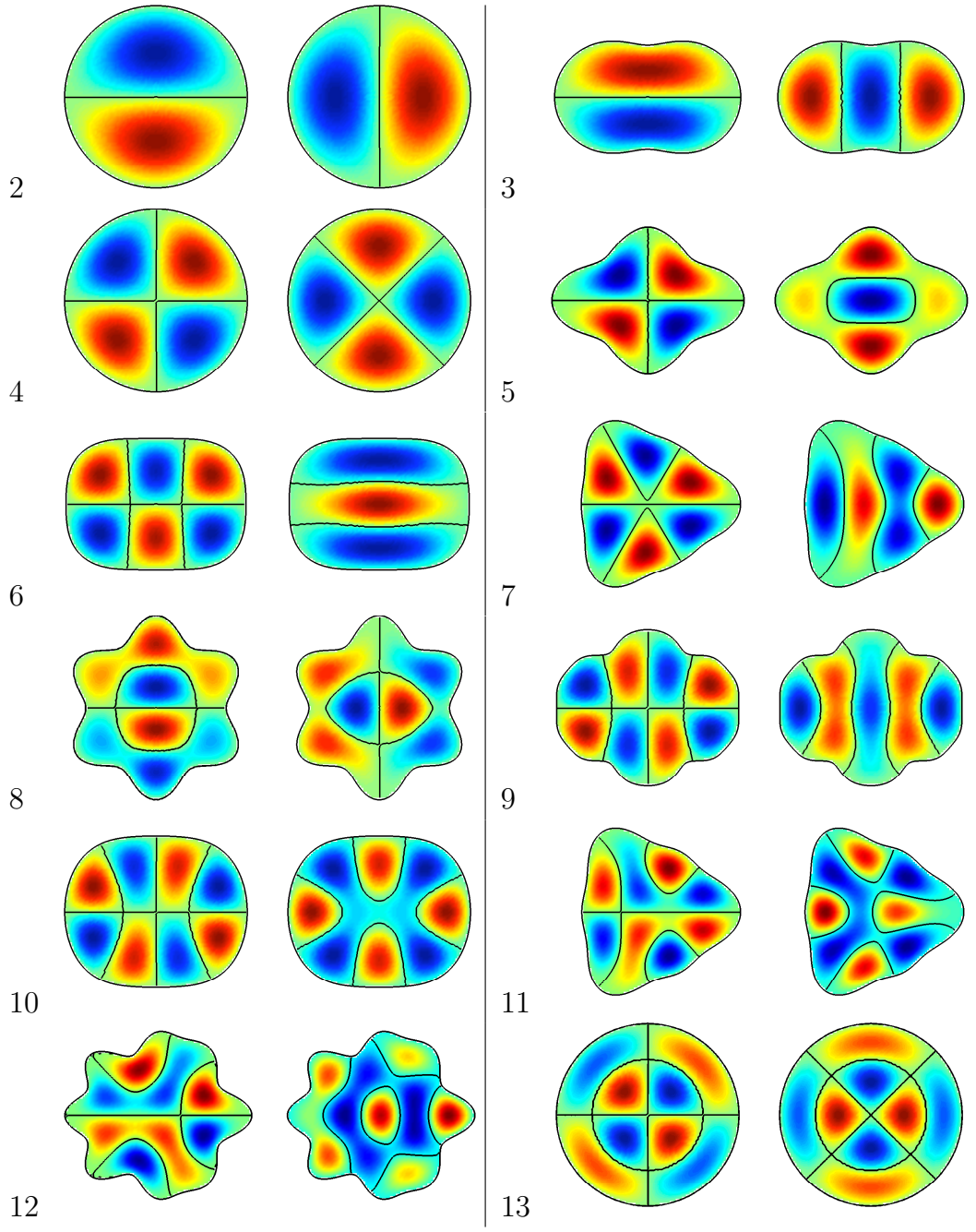


Figure 7.3: The regions which attain the maximum value of r_n for $n = 2, \ldots, 13$. For each region, we have plotted two eigenfunctions spanning the subspace corresponding to λ_n and $\lambda_{n+1} = \lambda_n$. The black lines are nodal lines.

Regions in $\mathcal{D}$ We now apply the framework developed in §7.5 to solve the constrained optimization problem (7.17) for $n = 2, \ldots, 13$. We tabulate the results in Table 7.2(a) and plot the achieving regions in Fig. 7.3. All digits given in Table 7.2 are believed to be accurate.

We find that for a region which achieves the optimal value r_n^*, we have $\lambda_n^* = \lambda_{n+1}^*$. Thus for each region in Fig. 7.3 we have also plotted two eigenfunctions corresponding to this eigenvalue of multiplicity two. In each case, one eigenfunction is even with respect to the $\hat{x}$-axis and one is odd.

For each value n, the number of Fourier-cosine coefficients $\{b_k\}_{k=0}^m$ needed to represent the regions varies, but is less than 20. In each case, we believe enough coefficients have been used so that the domains are well represented.

From Eq. (7.8) we know that the optimal value r_n^* is attained by the disk as $n \uparrow \infty$. On the other hand, in Fig. 7.3, there seems to be a subsequence of domains ($n = 3, 5, 8, 12$) with increasingly oscillatory boundary.

7.7 Dirichlet-Laplacian eigenvalue gaps

Let $g_n(D) = \frac{\lambda_n(D) - \lambda_{n-1}(D)}{\lambda_1(D)}$ be the ratio of the n-th eigenvalue gap to first eigenvalue and consider the shape optimization problem

$$\max_{D \in \mathcal{D}_m} g_n(D). \tag{7.19}$$

Weyl's asymptotic series (7.3) is insufficient to analyze the large n behavior of the optimal value g_n^*, so we might expect a wider variety of optimal solutions of Eq. (7.19) than of Eq. (7.17).

In the first three columns of Table 7.2(b), we give the values for $g_n^\circ \equiv g_n(\text{disk})$, $g_n^\triangle \equiv g_n(\triangle)$, and $g_n^\square \equiv \max_{a \geq 1} g_n(R_a)$. We use the method described in §7.5 to solve the constrained optimization problem in Eq. (7.19) for $n = 2, \ldots, 13$, collecting the results in Table 7.2(b) and Fig. 7.4. We again find that for a region which achieves the optimal value g_n^*, we have $\lambda_n = \lambda_{n+1}$. As in Fig. 7.3, in Fig. 7.4 we plot the two eigenfunctions corresponding to this multiplicity two eigenvalue.

7.8 Discussion

We have presented a general approach for finding (local) maxima of shape optimization problems over the set $\mathcal{D}_m$ where the objective function is a spectral function of the D-L eigenvalues. The numerical method used is an optimize-then-discretize approach, which involves evaluating the analyt-

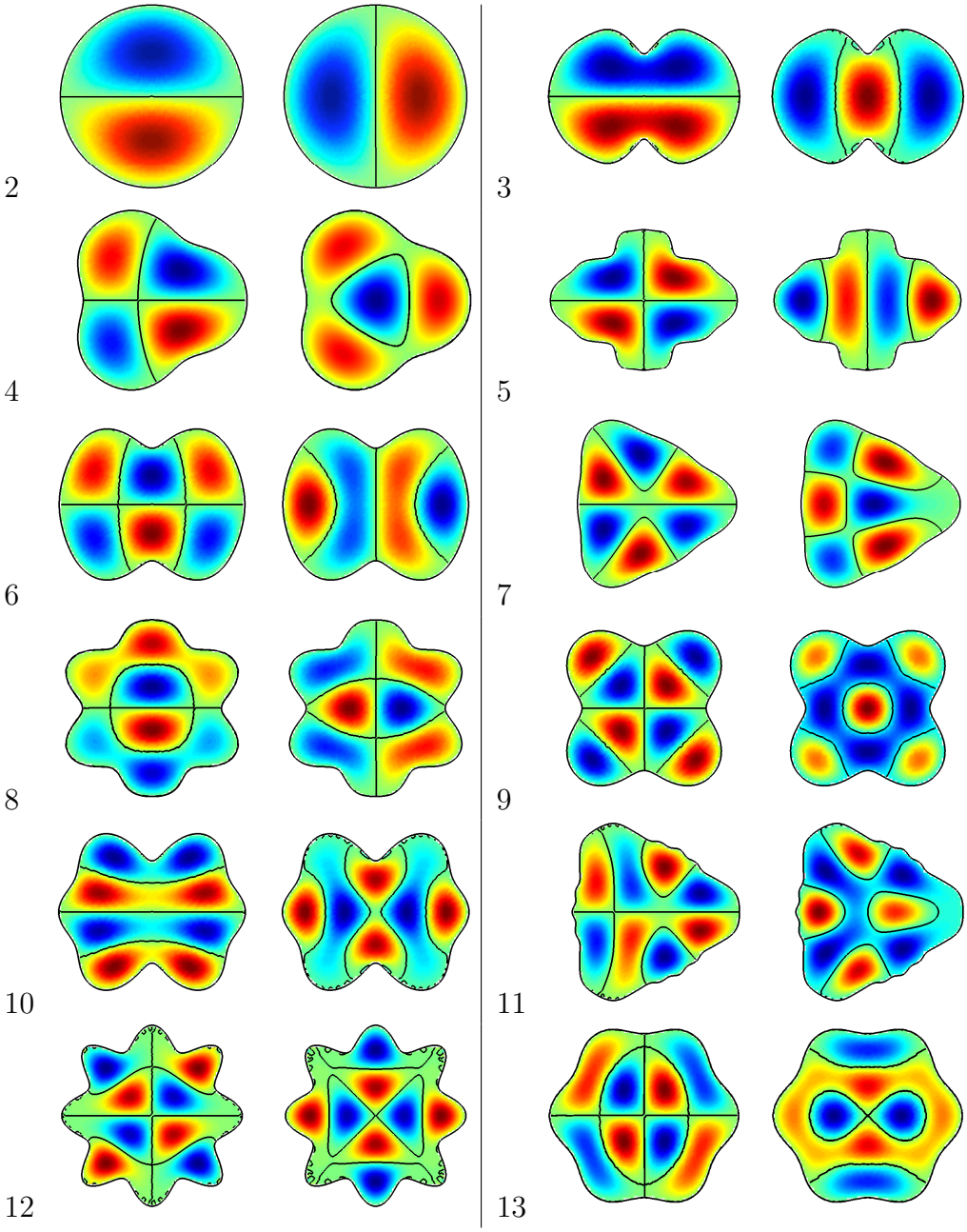


Figure 7.4: The regions which attain the maximum value of g_n for $n = 2, \ldots, 13$. For each region, we have plotted two eigenfunctions spanning the subspace corresponding to λ_n and $\lambda_{n+1} = \lambda_n$. The black lines are nodal lines.

ically computed gradient and Hessian at a numerically computed solution. This is in contrast with a discretize-then-optimize approach, which would require defining the D-L eigenvalues Λ_n in terms of a discretized equation and then differentiating this with respect to the decision variables b_k. The use of a spectrally accurate numerical method justifies this approach and since the method is mesh-free, avoids challenges such as mesh generation, differentiability, and movement. The representation of the boundary by Fourier coefficients rather than the level-set method [Osher and Santosa, 2001; Dobson and Santosa, 2004; Kao and Santosa, 2008] was chosen because (1) the representation by Fourier coefficients results in a comparatively low-dimensional space for optimization, (2) the method of particular solutions, which is more accurate than the finite element method, requires a set of points on the boundary of the domain and avoids the need for a relaxed formation of the problem as in [Oudet, 2004], and (3) the redundancy in representation as discussed in §7.3 is not explicit in the level-set framework. On the other hand, the parameterization chosen here does not allow for changes in topology of the region and requires star-shapedness.

The method developed is applied to maximizing r_n and g_n for $n = 2, \ldots, 13$, but is general and could be employed to find optimal solutions of other objective functions, including some of the open problems given in [Ashbaugh and Benguria, 2007; Henrot, 2006] and also r_n and g_n for larger values of n.

Although only one-fold symmetry was assumed, many of the optimal regions have two-fold symmetry. The lumpy triangular regions which maximize r_7, r_{11}, g_4, g_7, and g_{11} all have $\frac{2\pi}{3}$ rotational symmetry and have values which exceed that for the equilateral triangle. For regions with two-fold symmetry (D also symmetric about the $\hat{y}$-axis), from Eq. (7.11) we compute

$$\frac{\partial \lambda}{\partial b_k} = -2 \int_0^{\frac{\pi}{2}} \rho(\theta) |\nabla u(\rho, \theta)|^2 [\cos(k\theta) + (-1)^k \cos(k\theta))] \, \mathrm{d}\theta$$

where we have again used the trigonometric identity in Eq. (7.20). We see that for a domain with two-fold symmetry, $\frac{\partial \lambda}{\partial b_k} = 0$ for odd k, implying two-fold symmetry is preserved by gradient flow. Because of this property, the optimization method must be initialized with a region without two-fold symmetry.

For each value of n in Figs. 7.3 and 7.4, the optimal regions and corresponding eigenfunctions have similar characteristics but do not have obvious structure. For example, only for $n = 2, 3, 4,$ and 6 do the nodal lines divide the region into n subregions as in Courant's nodal line theorem [Kuttler and Sigillito, 1984; Courant and Hilbert, 1953]. Several of the eigenfunctions have

nodal lines which are closed.

Just as for r_2 (PPW inequality) and conjectured by Ashbaugh and Benguria for r_3 [Ashbaugh and Benguria, 1992a], we have found that at the optimally computed solutions for both r_n and g_n, we have $\lambda_n = \lambda_{n+1}$. This is a truly remarkable property and may continue to hold for larger values of n. In Fig. 7.2, we observed that this property also holds for the objective function r_n when the admissible class of regions is restricted to rectangles.

All of the numerically computed values in Table 7.2 satisfy the isoperimetric inequalities discussed in §7.2. Also, the result for r_3 agrees extremely well with the numerical result due to Levitin and Yagudin [Levitin and Yagudin, 2003]. If one assumes that the maximum value $r_4^* = 4.560$ attained for regions in $\mathcal{D}$ actually maximizes r_4 over all bounded smooth planar regions, then the inequality of Cheng and Yang [Cheng and Yang, 2007] stating $r_4 \lesssim 6.402$ could be improved by a considerable 29%. Similar statements can be made for the other n considered here. This work also supports the suggestion of Ashbaugh and Benguria that $r_4^* = 4.560$ [Ashbaugh and Benguria, 2007].

Engineers (*e.g.*, in aerodynamics and lithography) have long used optimization as a tool for design. Investigation into why optimal structures are actually optimal has often provided design principles for practice and a deeper understanding of the question at hand. The use of numerical optimization methods can also be used to further advance our understanding of isoperimetric inequalities and Kac's question.

7.A Domain preserving isometries

Rotations Using the trigonometric identity

$$\cos k(\theta - \pi) = (-1)^k \cos(k\theta), \qquad k \in \mathbb{N} \tag{7.20}$$

we find that a rotation by π corresponds to the transformation

$$b_k \mapsto \begin{cases} b_k & k \text{ even} \\ -b_k & k \text{ odd.} \end{cases}$$

Similarly, using the trigonometric identity

$$\cos k\left(\theta - \frac{\pi}{n}\right) = \cos(k\theta)\cos\left(\frac{k\pi}{n}\right) + \sin(k\theta)\sin\left(\frac{k\pi}{n}\right), \qquad k \in \mathbb{N}$$

we find that if $\sin(k\pi/n) = 0$, *i.e.*, $k \in n\mathbb{N}$, we have that

$$\cos k\left(\theta - \frac{\pi}{n}\right) = (-1)^{\frac{k}{n}} \cos(k\theta), \qquad k \in n\mathbb{N}.$$

Thus if $b_k = 0$ for all $k \notin n\mathbb{N}$, a rotation by π/n corresponds to the mapping

$$b_k \mapsto \begin{cases} b_k & k/n \text{ even} \\ -b_k & k/n \text{ odd.} \end{cases}$$

Small translations along the axis of symmetry We consider the change in coefficients $\{b_k\}_{k=1}^{\infty}$ for the small transformation $(x, y) \mapsto (x + \epsilon, y)$. Let $\rho(\theta) = \sum_{k=0}^{\infty} b_k \cos(k\theta)$. As long as the region remains star shaped from the new origin, we can transform variables

$$r \mapsto r + \epsilon \cos\theta + \mathcal{O}(\epsilon^2)$$
$$\theta \mapsto \theta - \epsilon\rho^{-1} \sin\theta + \mathcal{O}(\epsilon^2)$$

and write $b_k \mapsto b_k + \epsilon b_k' + \mathcal{O}(\epsilon^2)$. We then expand in a Taylor series about $\epsilon = 0$ to find

$$\begin{aligned} r + \epsilon\cos\theta &= \sum_{k=0}^{\infty} (b_k + \epsilon b_k' + \mathcal{O}(\epsilon^2)) \cos\left[k\left(\theta - \epsilon\rho^{-1}\sin\theta + \mathcal{O}(\epsilon^2)\right)\right] \\ &= \sum_{k=0}^{\infty} b_k \cos(k\theta) + \epsilon b_k' \cos(k\theta) + \epsilon b_k \rho^{-1} \sin(\theta)\sin(k\theta) + \mathcal{O}(\epsilon^2) \end{aligned}$$

so that at $\mathcal{O}(\epsilon)$ we have

$$\sum_{k=0}^{\infty} b_k' \cos(k\theta) = \cos\theta - \sum_{k=0}^{\infty} b_k \frac{1}{\rho(\theta)} \sin(\theta)\sin(k\theta).$$

Multiplying by $\cos(m\theta)$ and integrating we find that

$$\begin{aligned} b_0' &= -\frac{1}{2\pi} \sum_{k=0}^{\infty} b_k \int_0^{2\pi} \frac{1}{\rho(\theta)} \sin(\theta)\sin(k\theta)\,\mathrm{d}\theta \\ b_1' &= 1 - \frac{1}{\pi} \sum_{k=0}^{\infty} b_k \int_0^{2\pi} \frac{1}{\rho(\theta)} \sin(\theta)\sin(k\theta)\cos(\theta)\,\mathrm{d}\theta \\ b_m' &= -\frac{1}{\pi} \sum_{k=0}^{\infty} b_k \int_0^{2\pi} \frac{1}{\rho(\theta)} \sin(\theta)\sin(k\theta)\cos(m\theta)\,\mathrm{d}\theta \qquad m \geq 2. \end{aligned}$$

Note that if $\rho(\theta) = 1$ (*i.e.*, $b_k = \delta_{k,0}$) these formulas simplify to $b_k' = \delta_{k,1}$. Thus at first order, the disk is translated by varying b_1.

Chapter 8

Long-lived scattering resonances of the Helmholtz equation and the Bragg relation

8.1 Introduction and overview

Many device applications, ranging from photonic to micro-mechanical require the controlled localization of energy within a compact region of space or "cavity". In such settings, an important performance-limiting loss mechanism is *scattering loss*, leakage from or tunneling out of the structure. We have in mind applications to wave phenomena in non-dissipative media governed by time-dependent wave equations arising, for example, in (i) electromagnetic waves in dielectric media, (ii) acoustic waves, and (iii) elastic waves. An important class of motivating examples concerns the control of light via micro- and nano-scale photonic crystal devices; see, *e.g.*, [Joannopoulos *et al.*, 2008; Busch *et al.*, 2007].

Thus, the following optimization problem naturally arises:

Given constraints on material parameters and the size of the structure surrounding the cavity, how does one design a structure which maximizes the confinement time of energy?

We next explain how the confinement-time of energy in a cavity can be expressed in terms of the imaginary parts of complex eigenvalue of the non-selfadjoint scattering resonance problem

(SRP). We then formulate the optimization problem, summarize the results of this chapter and review related work.

8.1.1 Energy escape and the scattering resonance problem

Our point of departure is the time-dependent wave equation for an inhomogeneous medium:

$$n^2(\mathbf{x})\ \partial_t^2 v(\mathbf{x},t) = \Delta v(\mathbf{x},t) \qquad \mathbf{x}\in\mathbb{R}^d. \tag{8.1}$$

Here, $n(\mathbf{x})$ denotes a spatially varying index of refraction,[1] which we assume to satisfy upper and lower bounds:

$$0 \ < \ n_- \ \leq \ n(\mathbf{x}) \ \leq \ n_+ \ < \ \infty. \tag{8.2}$$

We consider structures, which are supported in a fixed compact set, , *i.e.*,

$$\text{supp}\,(n(\mathbf{x})-1) \ = \ \overline{\Omega},$$

where Ω is a bounded open subset of $\mathbb{R}^d$.

Solutions to the Cauchy problem for (8.1) with localized initial data: $u(\mathbf{x},0),\ \partial_t u(\mathbf{x},0)$, conserve the energy:

$$E\left[v(\cdot,t),\partial_t v(\cdot,t)\right] \equiv \int_{\mathbb{R}^d} n^2(\mathbf{x})|\partial_t v(\mathbf{x},t)|^2 + |\nabla v(\mathbf{x},t)|^2\,\mathrm{d}\mathbf{x} \ = \ E\left[v(\cdot,0),\partial_t v(\cdot,0)\right] \tag{8.3}$$

Yet, such solutions decay to zero as $t\to\infty$ in the pointwise or *local energy* sense:

$$\text{for any compact subset } K\subset\mathbb{R}^d, \quad \int_K |u(\mathbf{x},t)|^2\ d\mathbf{x} \ \to \ 0, \quad \text{as } \ t\to\infty.$$

The rate of local energy decay can be derived by studying the solution of the initial value problem, expressed as an inverse Laplace transform, *i.e.*,

$$u(\mathbf{x},t) \ \sim \ \int_{i\kappa-\infty}^{i\kappa+\infty} e^{-i\omega t}\ (-\Delta - n^2\omega^2)^{-1}\ d\omega \ \circ(\text{data}), \quad \kappa>0$$

The resolvent kernel, $(-\Delta - n^2\omega^2)^{-1}(x,y)$, has no poles in the upper half plane. In spatial dimensions $d=1$ and 3, it has a meromorphic continuation to the lower half plane, and in $d=2$, a meromorphic continuation to a logarithmic covering of the complex plane. In both cases, pole

[1]The index of refraction is the reciprocal of the relative wave speed of the propagation medium.

singularities exist in the lower half-plane, which are referred to as *scattering resonances*, *scattering frequencies*, or *scattering poles*.

Due to the time-dependence $e^{-i\omega t}$ in the inverse Laplace transform representation of the solution of the time-dependent initial value problem, time decay can be shown by deforming the contour into the lower half plane to a parallel contour along which the imaginary part is slightly larger than that of the scattering resonance which is closest to the real ω-axis [Lax and Phillips, 1989; Tang and Zworski, 2000], *i.e.*, the pole $\omega_\star[n]$ that is closest to the real ω-axis gives rise to the exponential decay rate $\sim \exp\left(-\left|\Im\omega_\star[n]\right| t\right)$. $\left|\Im\omega_\star[n]\right|$ is called the *width* of the resonance and $\tau := \left|\Im\omega_\star\right|^{-1}$ is called its *lifetime*.

For our purposes it is very useful that scattering resonances can also be characterized as eigenvalues of the non-selfadjoint spectral problem consisting of the Helmholtz equation with outgoing (Sommerfeld) radiation condition imposed at infinity [Colton and Kress, 1998]:

The Scattering Resonance Problem (SRP):

Find $(\omega, u(\mathbf{x},\omega) \neq 0)$ such that

$$\left(\Delta + \omega^2 n^2(\mathbf{x}) \right) u(\mathbf{x},\omega) = 0 \qquad \mathbf{x} \in \mathbb{R}^d \tag{8.4a}$$

$$\frac{\partial u(\mathbf{x},\omega)}{\partial |\mathbf{x}|} - \imath\, \omega\, n(\mathbf{x})\, u(\mathbf{x},\omega) = o\left(|\mathbf{x}|^{-\frac{d-1}{2}}\right) \qquad |\mathbf{x}| \to \infty \tag{8.4b}$$

A solution, $u(\mathbf{x},\omega)$, corresponding to a scattering frequency, ω is called a *scattering resonance mode*, *quasi-normal mode*, or *quasi-mode*.

The eigenvalues of (8.4) are complex and lie in the open lower half plane, $\Im\omega < 0$. Define

$$\mathrm{Res}_n := \{ \text{ the set of all (complex) eigenvalues, } \omega, \text{ of (8.4), for coefficient } n(\mathbf{x}) \} \tag{8.5}$$

If (ω, u) is a scattering resonance pair, then so is $(-\overline{\omega}, \overline{u})$. It follows that the set Res_n is discrete and symmetric about the imaginary axis. The set Res_n may be empty, as in the case where $n(\mathbf{x}) \equiv 1$ or may be non-empty, as in the explicit example of section 8.2.1.

In this chapter, we study the problem of designing a refractive index profile, $n(\mathbf{x})$, subject to physically motivated constraints, for which there are very long-lived resonances. By the previous

discussion, this corresponds to choosing $n(\mathbf{x})$ so that there are scattering resonances very close to the real axis. resonances with long lifetime, *i.e.*, small width, $|\Im\omega|$.

Roughly speaking, long-lived resonances can arise in the following ways:

(A) **Total internal reflection:** Confinement of energy can be achieved by the mechanism of (nearly) total internal reflection. In fact, arbitrarily long confinement can be achieved as follows. Consider a circular or spherical region in two or three space dimensions on which $n(\mathbf{x}) > 1$ is constant. If the angular momentum of the resonance mode is large, the mode will be strongly confined to the interface of the cavity. In the geometric optics approximation, the light rays have very shallow angle of incidence and therefore are nearly totally internally reflected. Such modes are referred to as *whispering gallery* or *glancing* modes and are the basis for spherical resonators; see figure 8.1 and section 8.2.1.

(B) **Interference effects:** The cavity can be surrounded by strongly reflective medium which is periodic of an appropriate period. In this case, wave interference effects provide the localizing mechanism. This is the basis for the Bragg resonator or Fabry-Pérot cavity [Joannopoulos *et al.*, 2008].

Figure 8.1 illustrates the difference between mechanisms (A) and (B). The left figure displays radial confinement via interference effects; the refractive index profile consists of concentric annular regions alternating between $n_- = 1$ and $n_+ = 2$. The right figure illustrates confinement via (nearly) total internal reflection; the refractive index is a constant $n_+ = 2$ inside the circular cavity, $\Omega = \{|x| < 1\}$, and $n = 1$ outside. Modes, $f_\ell(r)e^{\pm\imath\ell\theta}$, $\ell = 0, 1, 2, \ldots$ (in 2D) and $f_\ell(r)Y_\ell^m(\theta,\phi)$, $|m| \leq 2\ell+1$, $\ell = 0, 1, 2, \ldots$ (in 3D) of increasing angular momentum, ℓ, have longer and longer lifetimes; the imaginary parts of the corresponding scattering resonances tend to zero.

We now formulate the optimization problem considered in this article. We shall seek $n(x) \in \mathcal{A}$, a specified admissible set of structures, having a resonance ω closest to the real axis ($|\Im\omega| \to \min$) for which $|\Re\omega|$ no larger than a prescribed upper bound ($|\Re\omega| \leq \rho$) .

To obtain a precise formulation, we first introduce admissible sets of structures. Let $\Omega \subset \mathbb{R}^d$ denote a fixed open and bounded set. Also, let $0 < n_- < n_+$ be specified. Then, our first admissible

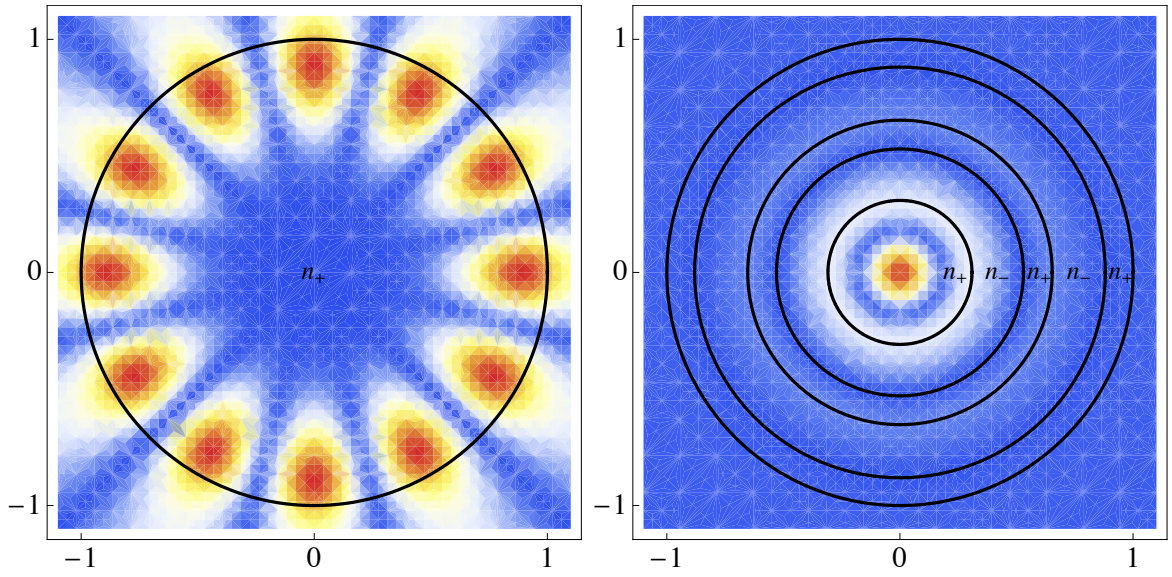


Figure 8.1: (left) A contour plot of the modulus of a mode with long lifetime ($\omega = 4.2 - 0.033\imath$) due to (A) total internal reflection because of large angular momentum ($\ell = 6$) as in a spherical resonator. (right) The modulus of a mode with $\omega = 6.5 - .039\imath$ confined by (B) interference effects, as in a Fabry-Pérot cavity.

set is given by:

$$\mathcal{A}(\Omega, n_-, n_+) \equiv \{n \colon \text{supp } (n(x) - 1) \subset \overline{\Omega}, \quad n_- \ \leq \ n(x) \ \leq \ n_+\} \tag{8.6}$$

In one space dimension, we shall take $\Omega = [0, L]$ and consider the set of admissible structures which are symmetric:

$$\mathcal{A}_{sym}(L, n_-, n_+) \equiv \{n \in \mathcal{A}([0, L], n_-, n_+) \colon \ n(x) = n(L - x)\}. \tag{8.7}$$

When the choice of Ω, n_-, n_+ are unambiguous, we shall simply write $\mathcal{A}$ and $\mathcal{A}_{sym}$.

For $\rho > 0$ define, for $n(\mathbf{x}) \in \mathcal{A}$, an admissible set of the above type $\text{Res}_n^\rho \subset \text{Res}_n$:

$$\text{Res}_n^\rho := \{\omega \in \text{Res}_n \colon |\Re\omega| \leq \rho\} \tag{8.8}$$

The minimal resonance width in the set Res_n^ρ is given by

$$\Gamma^\rho[n] := \inf_{\omega \in \text{Res}_n^\rho} |\Im\omega|. \tag{8.9}$$

If $\mathrm{Res}_n^\rho = \emptyset$, then we set $\Gamma^\rho[n] = \infty$. We study the following

Optimal Design Problem:

$$\Gamma_\star^\rho(\mathcal{A}) \;:=\; \inf_{n\in\mathcal{A}} \Gamma^\rho[n] \;=\; \inf_{n\in\mathcal{A}} \inf_{\omega\in\mathrm{Res}_n^\rho} |\Im\omega|. \tag{8.10}$$

By the explicit example in section 8.2.1, $\Gamma_\star^\rho(\mathcal{A}) < \infty$, for the above choices of $\mathcal{A}$. The corresponding lifetime of the optimal resonance mode is given by:

$$\tau_\star(\mathcal{A}) := \frac{1}{\Gamma_\star(\mathcal{A})}. \tag{8.11}$$

8.1.2 Outline and summary of results

1. In section 8.2, we compute the resonances for a simple example where $n(\mathbf{x}) = 1 + 1_\Omega$ where 1_Ω is the indicator function for the region Ω, which is taken to be a ball in $\mathbb{R}^d$.

   We also derive variational-type identities for resonances. In one dimension, we use these identities to show that for $\Omega = [0, L]$, there is a general lower bound on the resonance width $|\Im\omega_\star|$ of the following type:

   $$|\Im\omega_\star| \;\geq\; \alpha_{\mathcal{A}}\, e^{-\beta_{\mathcal{A}}\; L^2|\Re\omega|^2} \tag{8.12}$$

   where $\alpha_{\mathcal{A}}, \beta_{\mathcal{A}} > 0$ depends on the constraint set, $\mathcal{A}$; see Proposition 8.2.2. In particular, this implies that for L sufficiently large, the maximal lifetime resonance $\omega_\star$ has satisfies $|\Re\omega_\star| > 0$. We also use a maximum principle argument to show that there exists a triangular resonance-free region in the lower half-plane.

2. In section 8.3 we present results on the existence of a maximal lifetime resonance for dimensions $d \geq 1$. We prove, using direct methods, that there exists a structure $n_\star(\mathbf{x}) \in \mathcal{A}$ with a scattering resonance, $\omega_\star \in \mathrm{Res}_{n_\star}^\rho$, satisfying (8.4) with minimal width $|\Im\omega_\star| = \Gamma_\star^\rho(\mathcal{A})$.

3. In section 8.4 we show, in dimensions $d \geq 1$, that the locally optimal structures $n_\star(\mathbf{x}) \in \mathcal{A}$ satisfy bang-bang constraints, that is, $n_\star(\mathbf{x})$ is either n_+ or n_- for almost every $\mathbf{x} \in \Omega$.

4. In sections 8.5-8.6, we specialize to the case of one-dimensional structures where we can prove considerably more. In section 8.5, we compute optimal structures using quasi-Newton optimization methods. In section 8.6, we characterize one-dimensional optimal structures,

$n_\star \in \mathcal{A}_{sym}$ and establish a connection between $n_\star$ and the well-known class of Bragg structures, where $n(x)$ is constant on intervals whose length is one-quarter of the effective wavelength.

8.1.3 Brief review of related work

For a general and fairly recent review on wave propagation in locally periodic media, consult [Griffiths and Steinke, 2001].

Resonance mode expansions for the Helmholtz operator. While scattering resonance states or quasi-normal modes have long been used in quantum mechanics and the study of black hole dynamics, these tools have only been used to study optical systems and the Helmholtz operator in the last two decades. In [Leung *et al.*, 1994a; Leung *et al.*, 1994b; Leung *et al.*, 1994c; Ching *et al.*, 1998], the authors study leaky, one-dimensional optical cavities using scattering resonance states. Orthogonality of the resonance states is demonstrated under a generalized inner product and the modes are formally shown to be complete within a region Ω, provided $n(x)$ is discontinuous on $\partial\Omega$. In [Settimi *et al.*, 2003], this framework is applied to the study of one-dimensional photonic crystals where $n(x)$ is a piecewise constant function. The definition of density of states is also given in terms of the resonances. In [Iantchenko, 2006; Ramdani and Shipman, 2008], it is demonstrated that for a locally periodic medium which is repeated n times, the resonances converge to the continuous spectrum as $n \uparrow \infty$. In [Maksimovic, 2008], resonance states are used to numerically investigate photonic crystals with multiple defects. In [Settimi *et al.*, 2009], the transmission properties of a 1D photonic crystal structure is investigated in terms of resonance modes. A very recent paper has investigated some of the same problems that are studied here, although only in the one dimensional setting and with slightly different boundary conditions [Karabash, 2011]. Purely imaginary resonances of the Helmholtz operator have also been studied in greater detail [Ralston, 1972; Labreuche, 1998]. Lastly a rigorous study of the perturbation of resonances can be found in [Agmon, 1996].

Optical cavity design. Optical cavities have now been designed to perform many different tasks for a variety of applications. Such design problems are typically formulated as an optimization

problem for a particular figure of merit and solved using numerical optimization methods.

The problem of maximizing the lifetime of a state trapped within a leaky cavity can be framed in several ways. The figure of merit can be taken to be the minimization of energy flux through the boundary [Lipton *et al.*, 2003] or a measure of mode localization [Dobson and Santosa, 2004; Akcelik *et al.*, 2005]. In [Kao and Santosa, 2008], the problem of minimizing $|\Im\omega|$ for a chosen resonance was investigated computationally in both one- and two-dimensions. See Sec. 8.5 for discussion of how this compares to the minimization problem (8.10). The one-dimensional problem was also studied computationally in [Heider *et al.*, 2008]. In particular, the variation $\frac{\delta\omega}{\delta n}$ is formally computed. [Heider *et al.*, 2008] primarily focuses on the optimization of σ to minimize $|\Im\omega|$ which satisfies

$$\partial_x \sigma \partial_x u(x,\omega) + \omega^2 u(x,\omega) = 0$$

where u also satisfies outgoing boundary conditions. In [Scheuer *et al.*, 2006], transfer matrix methods were used to design low-loss 2D resonators with radial symmetry. In each of these papers, gradient-based optimization methods were used to solve the optimization problem.

Genetic algorithms have also been employed to minimize energy flux through the boundary [Gondarenko *et al.*, 2006; Gondarenko and Lipson, 2008]. In [Englund *et al.*, 2005; Geremia *et al.*, 2002; Felici *et al.*, 2010] the "inverse method" is employed, where a desired mode shape is chosen and then the material properties which produce that mode are found algebraically. In [Bauer *et al.*, 2008], the time-dependent problem is solved to steady state using a finite-difference method with perfectly matched layers to approximate the outgoing boundary conditions. The design problem is solved using a Nelder-Mead method.

Results on the existence of optimal scattering resonances and general bounds on the imaginary parts of scattering resonances for Schrödinger operators can be found in [Harrell, 1982; Harrell and Svirsky, 1986; Svirsky, 1987]. Our results for the Helmholtz equation make use of some of the arguments introduced in these papers.

An important, related class of problems are to find photonic structures with large spectral band gaps. In fact, an intuitive way of trapping a resonance in a cavity, and the basis for the Fabry-Pérot cavity, is to surround a cavity with a "reflective" material. One such reflective material is a periodic structure with a spectral band gap at a desired frequency. This is what is referred to a Bragg reflector, which we discuss in more detail below. Structures with optimally large band gaps have

been proven to exist [Cox and Dobson, 1999] and numerical methods have been applied to finding them [Cox and Dobson, 2000; Burger *et al.*, 2004; Kao *et al.*, 2005]. In [Sigmund and Jensen, 2003] topology optimization was used to find photonic crystals with optimally large bandgaps and also which optimally damp or guide waves. In [Sigmund and Hougaard, 2008], properties of photonic crystals with optimally large bandgaps are investigated.

One property of optimal structures for the class of problems (8.10) is that they are piecewise constant structures which achieve the material bounds, *i.e.*, they are *bang-bang* controls. This property is also realized in a number of optimization problems for eigenvalues of self-adjoint operators [Krein, 1955; Cox and McLaughlin, 1990a; Cox and McLaughlin, 1990b] as well as for Schrödinger resonances [Harrell and Svirsky, 1986]. In [Osting and Weinstein, 2011a] the authors consider the problem of maximizing the lifetime of a state coupled to radiation by an *ionizing* perturbation. For this class of problems, optimizers are interior points of the constraint set.

Lastly, we note that some attempts have already been made to find photonic structures which *robustly* achieve their desired figure of merit in the sense that small perturbations to the optimal structure are also good structures [Bertsimas *et al.*, 2007].

Bragg's relation. Consider a one-dimensional $n(x)$, infinite in extent, which is periodic with period d, *i.e.* $n(x+d) = n(x)$, and has alternating layers, *i.e.*

$$n(x) = \begin{cases} n_1 & 0 < x < b \\ n_2 & b < x < d. \end{cases}$$

It is shown in [Yeh, 1988] that the wave equation (8.1) has a solution of the form $v(x,t) = e^{\imath(\omega t - kx)} u_{per}(x)$ where $u_{per}(x+d) = u_{per}(x)$ if ω and k satisfy the dispersion relation

$$cos(kd) = cos(\omega n_1 b)\cos\omega n_2(d-b) - \frac{1}{2}\left(\frac{n_2}{n_1} + \frac{n_1}{n_2}\right)\sin(\omega n_1 b)\sin\omega n_2(d-b). \tag{8.13}$$

The *Bragg relation* is defined

$$n_1 b = n_2(d-b) = \frac{1}{4}\frac{2\pi}{\omega} \tag{8.14a}$$

$$\Rightarrow d = \frac{1}{2n_h}\frac{2\pi}{\omega}, \qquad b = \frac{n_h}{n_1}\frac{d}{2}, \qquad (d-b) = \frac{n_h}{n_2}\frac{d}{2} \tag{8.14b}$$

where $n_h = 2(n_1^{-1} + n_2^{-1})^{-1}$ is the harmonic mean of n_1 and n_2. An optical structure with this relation is referred to as a *quarter-wave stack* [Yeh, 1988; Joannopoulos *et al.*, 2008]. Note that

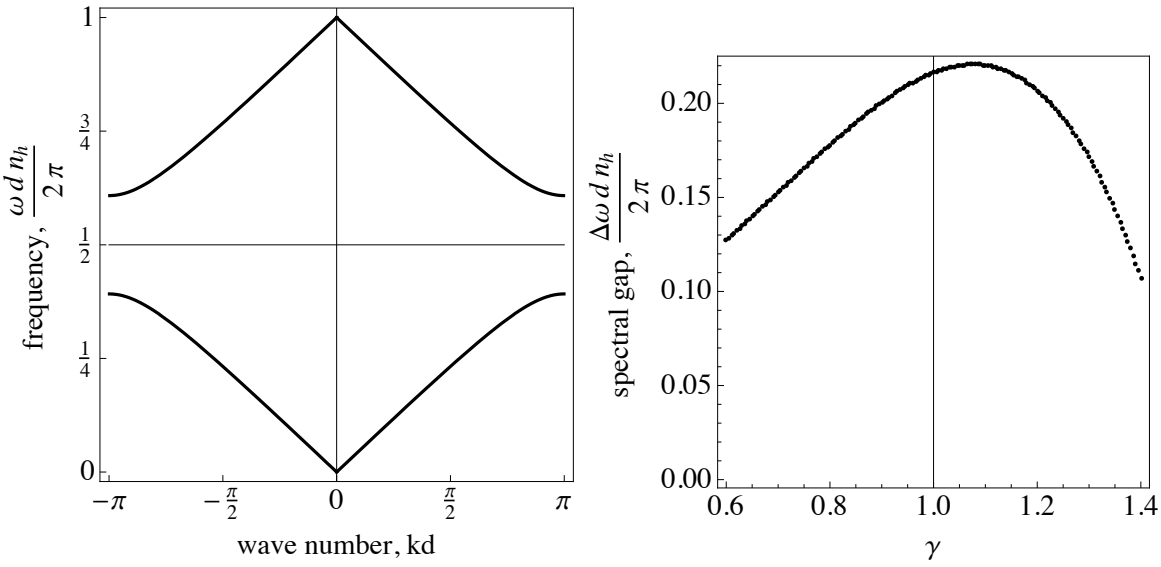


Figure 8.2: (left) The dispersion relation (8.13) for a quarter wave stack where $b = \frac{n_h}{n_1}\frac{d}{2}$. There is a spectral band gap centered at $\omega = \frac{\pi}{dn_h}$ of width $\frac{4}{dn_h}\sin^{-1}\frac{|n_1-n_2|}{n_1+n_2}$. (right) For $n_1 = 1$ and $n_2 = 2$, we let $b = \gamma\frac{n_h}{n_1}\frac{d}{2}$ and plot γ vs. the width of the spectral gap. We find that the gap width is maximized for a stack with $b_\star \approx 1.08\frac{n_h}{n_1}\frac{d}{2}$.

the width of each layer is a quarter wavelength, *i.e.*, the phase of the wave changes by $\pi/2$ in each layer. The dispersion relation (8.13) for the quarter-wave stack simplifies to

$$cos(kd) = 1 - \frac{1}{2}\frac{(n_1+n_2)^2}{n_1 n_2}\sin^2\left(\frac{\omega n_h d}{2}\right).$$

The dispersion relation with $n_1 = 1$ and $n_2 = 2$ is plotted in Fig. 8.2(left). Thus, the dispersion relation for the quarter-wave stack has a spectral band gap centered at $\omega = \frac{\pi}{dn_h}$ of width $\frac{4}{dn_h}\sin^{-1}\frac{|n_1-n_2|}{n_1+n_2}$.

The Bragg relation given in Eq. (8.14) can be interpreted to state that constructive interference occurs when the path length of a reflected wave is equal to (a multiple of) the wavelength. It may be intuitive that given n_1, n_2 and d, this choice of b would maximize the spectral band gap width (see [Joannopoulos *et al.*, 2008, p.51]). However, the following numerical experiment demonstrates otherwise. For fixed n_1, n_2 and d, define

$$b(\gamma) = \gamma\frac{n_h}{n_1}\frac{d}{2},$$

so that $b(1)$, gives the Bragg relation. In Fig. 8.2(right), we plot the width of the spectral gap for the periodic device with this choice of $b(\gamma)$ as a function of γ. The maximal width of the spectral gap occurs for $\gamma \approx 1.08 \neq 1$.

8.2 Scattering resonances: examples, variational identities, and bounds in one dimension

In this section we present simple examples of analytically solvable scattering resonance problems in one-, two-, and three-dimensional problems. We then derive some variational-type identities and use them to obtain universal inequalities on one-dimensional scattering resonances, in terms of the support of $n(x) - 1$ and the pointwise bounds on $n(x)$.

8.2.1 Examples: resonances for symmetric cavities in $\mathbb{R}^d$, $d = 1, 2, 3$

In this subsection, we discuss the resonances for a simple example in 1-, 2-, and 3-dimensions. We take $\Omega = \{\mathbf{x} \colon |\mathbf{x}| < a\}$ and $n(\mathbf{x})$ defined by:

$$n(\mathbf{x}) = \begin{cases} n_0 & |\mathbf{x}| < a \\ 1 & |\mathbf{x}| > a. \end{cases} \tag{8.15}$$

where $n_0 > 1$ and $a > 0$ are constants. We now consider the scattering resonance (8.4).

Dimension d=1. Imposing outgoing radiation conditions we find that

$$u(x) = \begin{cases} Ae^{-\imath\omega x} & x < -a \\ Be^{\imath\omega n_0 x} + Ce^{-\imath\omega n_0 x} & |x| < a \\ De^{\imath\omega x} & x > a \end{cases} \tag{8.16}$$

where A, B, C, and D are constants. Imposing continuity of u and $\partial_x u$ at $x = \pm a$ yields the 4×4 linear system of equations

$$\begin{pmatrix} e^{\imath\omega a} & -e^{-\imath\omega n_0 a} & -e^{\imath\omega n_0 a} & 0 \\ -e^{\imath\omega a} & -n_0 e^{-\imath\omega n_0 a} & n_0 e^{\imath\omega n_0 a} & 0 \\ 0 & e^{\imath\omega n_0 a} & e^{-\imath\omega n_0 a} & -e^{\imath\omega a} \\ 0 & n_0 e^{\imath\omega n_0 a} & -n_0 e^{-\imath\omega n_0 a} & -e^{\imath\omega a} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \mathbf{0}.$$

Resonances are the values $\omega \in \mathbb{C}$ for which a non-trivial solution exists. Setting the determinant of this matrix to zero yields the equation

$$e^{-2\imath\omega(n_0-1)a}\left[e^{4\imath\omega n_0 a}(n_0-1)^2-(n_0+1)^2\right]=0.$$

The solutions to this equation are given by

$$\omega_m=\frac{\pi m}{2n_0 a}-\imath\frac{1}{2n_0 a}\log\left|\frac{n_0+1}{n_0-1}\right| \qquad m\in\mathbb{N}. \tag{8.17}$$

The resonances in Eq. (8.17) for $n_0=2$, $a=1$ are plotted in Fig. 8.3(a).

Dimension d=2. Solutions that are bounded at the origin and outgoing at infinity are given by

$$u(r,\theta)=\begin{cases} AJ_m(n_0\omega r)e^{\imath m\theta} & r<a \\ BH_m^{(1)}(\omega r)e^{\imath m\theta} & r>a \end{cases}$$

where A and B are constants and $m\in\mathbb{Z}$. Imposing continuity of u and $\partial_r u$ at $r=a$ is equivalent to finding $\omega\in\mathbb{C}$ such that the system of equations

$$\begin{pmatrix} J_m(n_0\omega a) & -H_m^{(1)}(\omega a) \\ n_0\omega J_m'(n_0\omega a) & -\omega H_m^{(1)\prime}(\omega a)\end{pmatrix}\begin{pmatrix}A\\B\end{pmatrix}=\mathbf{0}$$

has a nontrivial solution. Taking the determinant of this matrix yields the transcendental equation

$$J_m(n_0\omega a)H_m^{(1)\prime}(\omega a)-n_0H_m^{(1)}(\omega a)J_m'(n_0\omega a)=0. \tag{8.18}$$

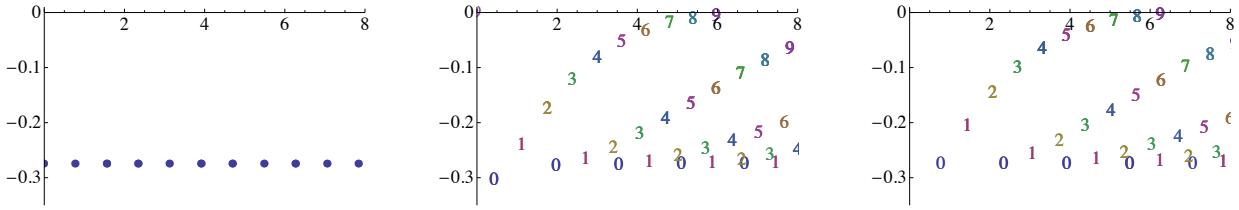
We numerically solve Eq. (8.18) for $n_0=2$, $a=1$, and $m=0,\dots,9$ and plot the resonances in Fig. 8.3(b).

Dimension d=3. Solutions that are bounded at the origin and outgoing at infinity are given by

$$u(r,\theta,\phi)=\begin{cases} Aj_\ell(n_0\omega r)Y_\ell^m(\theta,\phi) & r<a \\ Bh_\ell^{(1)}(\omega r)Y_\ell^m(\theta,\phi) & r>a \end{cases}$$

where A and B are undetermined coefficients and $|m|\le\ell\in\mathbb{N}$. Imposing continuity of u and $\partial_r u$ at $r=a$ is equivalent to finding $\omega\in\mathbb{C}$ such that the system of equations

$$\begin{pmatrix} j_\ell(n_0\omega a) & -h_\ell^{(1)}(\omega a) \\ n_0\omega j_\ell'(n_0\omega a) & -\omega h_\ell^{(1)\prime}(\omega a)\end{pmatrix}\begin{pmatrix}A\\B\end{pmatrix}=0$$



(a) 1D Helmholtz resonances. Each resonance has equal imaginary part and geometric multiplicity one.

(b) 2D resonances. The plotting markers indicate the m-value of each resonance. The resonances for $m = 0$ have multiplicity 1 and for $m \geq 1$ have multiplicity 2.

(c) 3D resonances. The plotting markers indicate the ℓ-value of each resonance. Each resonance has multiplicity $2\ell + 1$.

Figure 8.3: Resonances for the radially symmetric Helmholtz operator with index of refraction given in Eq. (8.15) in 1-, 2-, and 3-dimensions computed in Eqs. (8.17), (8.18), and (8.19). In all cases, $n_0 = 2$ and $a = 1$.

has a nontrivial solution. Taking the determinant of this matrix yields the transcendental equation

$$j_\ell(n_0\omega a)h_\ell^{(1)\prime}(\omega a) - n_0 h_\ell^{(1)}(\omega a)j_\ell'(n_0\omega a) = 0. \tag{8.19}$$

We numerically solve Eq. (8.19) for $n_0 = 2$, $a = 1$, and $\ell = 0, \ldots, 9$ and plot the resonances in Fig. 8.3(c).

Observations. From this example, we observe the following:

1. For $d = 1$, there are a countable number of geometrically simple resonances.[2] For $d = 2$ there are a countable number of solutions corresponding to each $m \in \mathbb{N}$. For each $m \geq 1$, there are two linearly independent resonance states associated with each resonance, *i.e.*, these resonances have geometric multiplicity 2. For $d = 3$ there are a countable number of solutions corresponding to each $\ell \in \mathbb{N}$. For each $\ell \geq 0$, there are $2\ell + 1$ linearly independent resonance states associated with each resonance, *i.e.*, each resonance has geometric multiplicity $2\ell + 1$.

2. In dimensions 2 and 3, the resonances accumulate on the real axis as m and $\ell \uparrow \infty$ respectively. As described as mechanism (A) in the introduction, these solutions with large angular mo-

[2]The geometric multiplicity of a resonance is the number of corresponding resonance states. One-dimensional resonances are always geometrically simple.

mentum (grazing modes) have high quality factor. The mode plotted in Fig. 8.1 corresponds to the 2-dimensional resonance with $m = 6$. In dimension 1, all resonances are bounded away from the real axis. For this $d = 1$ example, all resonances are simple and have equal imaginary part.

3. In one-dimension, one may show that the resonance states in this example are alternating even and odd. That is, for m even, the state written in Eq. (8.16) has $A = D$ and $B = C$ and for m odd, $A = -D$ and $B = -C$. We show in Prop. 8.6.1 that this holds for all $n(x) \in \mathcal{A}_{sym}$.

4. In 1D, it follows from Eq. (8.17) that the lifetime $\tau \propto a$, the size of the structure. In the high contrast limit $n_0 \uparrow \infty$,

$$\omega_m \sim \frac{\pi m}{2n_0 a} - \imath \frac{1}{n_0^2 a} \qquad m \in \mathbb{N}.$$

This demonstrates that there exist one-dimensional Helmholtz resonances with lifetimes $\tau \propto n_0^2$ as $n_0 \uparrow \infty$.

8.2.2 Variational-type identities

It is convenient to re-express the outgoing boundary condition in Eq. (8.4b) enforced at $|\mathbf{x}| = \infty$ on the boundary of Ω using the Dirichlet-to-Neumann (DtN) map. The DtN map, $\Lambda\colon H^{\frac{1}{2}}(\partial\Omega) \to H^{-\frac{1}{2}}(\partial\Omega)$ assigns to a function $f \in H^{\frac{1}{2}}$ the normal derivative $\partial_\nu u_f = \nabla u_f \cdot \nu$ on $\partial\Omega$, where u_f solves the Helmholtz equation in the exterior of Ω with outgoing radiation condition at infinity [Colton and Kress, 1998]. Here, ν is the outward normal to $\partial\Omega$. Thus Eq. (8.4b) may be re-expressed using the DtN map

$$\partial_\nu u = \Lambda u. \tag{8.20}$$

Proposition 8.2.1. *Let n be in the admissible set $\mathcal{A}$, defined in (8.6).*

1. *If $(\omega, u(\mathbf{x}, \omega))$ denotes a scattering resonance pair, then*

$$\Re(\omega^2) = \frac{\int_\Omega |\nabla u(\cdot,\omega)|^2 \; - \; \Re \int_{\partial\Omega} \overline{u(\cdot,\omega)} \; \Lambda[u(\cdot,\omega)]}{\int_\Omega n^2 |u(\cdot,\omega)|^2} \tag{8.21a}$$

$$\Im(\omega^2) = -\frac{\Im \int_{\partial\Omega} \overline{u(\cdot,\omega)} \; \Lambda[u(\cdot,\omega)]}{\int_\Omega \; n^2 \; |u(\cdot,\omega)|^2}. \tag{8.21b}$$

where Λ is the Dirichlet-to-Neumann map.

2. *In one spatial dimension, where $\Omega = [0, L]$, $\Lambda[u] = \imath\omega u$ and*

$$\Re(\omega^2) = \frac{\int_0^L |u'(\cdot,\omega)|^2 \ dx + \Im\omega \left(|u(0,\omega)|^2 + |u(L,\omega)|^2\right)}{\int_0^L n^2 \, |u(\cdot,\omega)|^2} \tag{8.22a}$$

$$\Im(\omega^2) = -\frac{\Re\omega \left(|u(0,\omega)|^2 + |u(L,\omega)\,|^2\right)}{\int_0^L n^2 \, |u(\cdot,\omega)|^2}. \tag{8.22b}$$

3. *Furthermore, Equation (8.22b) and $\Im\omega < 0$ imply:*

$$|\Im\omega| = \frac{|u(0,\omega)|^2 + |u(L,\omega)|^2}{2\int_0^L n^2 |u(\cdot,\omega)|^2}. \tag{8.23}$$

Proof. Multiply Eq. (8.4a) by $\overline{u(\mathbf{x},\omega)}$ and integrate over Ω to obtain

$$\begin{aligned}\left(\Re(\omega^2) + \imath\Im(\omega^2)\right)\int_\Omega n^2|u|^2 \,\mathrm{d}\mathbf{x} &= \int_\Omega |\nabla u|^2 \,\mathrm{d}\mathbf{x} - \int_{\partial\Omega} \overline{u}(\nabla u \cdot \mathbf{n}) \,\mathrm{d}\mathbf{x} \\ &= \int_\Omega |\nabla u|^2 \,\mathrm{d}\mathbf{x} - \int_{\partial\Omega} \overline{u}\Lambda[u] \,\mathrm{d}\mathbf{x}.\end{aligned}$$

Identifying real and imaginary parts yields Eq. (8.21). Eq. (8.23) follows from Eq. (8.22b) and the relationship $|\Im(\omega^2)| = 2|\Re\omega||\Im\omega|$. □

8.2.3 Lower bounds for resonances of the one-dimensional Helmholtz equation

We use the variational-type identities from Sec. 8.2.2 to show the following universal inequality for one-dimensional resonances.

Proposition 8.2.2. *Let $\Omega = [0, L] \subset \mathbb{R}^1$ and $n \in \mathcal{A}(\Omega, n_-, n_+)$. For any scattering resonance $\omega \in \mathrm{Res}_n$ and $\xi > 0$,*

$$|\Im\omega| \geq \min\left[\xi, \ \frac{3\exp\left(-\left(|\Re\omega|^2 + \xi^2\right) n_+^2 L^2\right)}{n_+^2 L\left(3 + L^2\left(|\Re\omega|^2 + \xi^2\right)\right)}\right].$$

In particular, if $n_+ > e^{-1}$

$$|\Im\omega| \geq \frac{3\exp\left(-n_+^2 L^2 |\Re\omega|^2\right)}{eL\left(1 + 3n_+^2 + n_+^2 L^2 |\Re\omega|^2\right)}.$$

Corollary 8.2.3. *In one-dimension, if ρ is sufficiently large, the resonance $\omega_\star$ attaining the infimum of the spectral optimization problem has $\Re\omega_\star > 0$.*

We follow the strategy of [Harrell, 1982] to prove Prop. 8.2.2. Proposition 8.2.2 relies on the following two lemmas, which we shall prove first. In this section only, we normalize the resonance state by assuming, without loss of generality,

$$1 = u(0) \leq |u(L)|^2 \tag{8.24}$$

(otherwise we make the substitution $x \mapsto L - x$).

Lemma 8.2.4. *For $0 \leq x \leq L$, we have the pointwise bound*

$$|u(x)| \leq \sqrt{1 + |\omega|^2 x^2} \exp\left(|\omega|^2 \int_0^x (x-y) n^2(y) \, \mathrm{d}y \right). \tag{8.25}$$

Proof. In one dimension, Eq. (8.4) with Eq. (8.24) is written

$$\begin{aligned} -\partial_x^2 u &= \omega^2 n^2 u && x \in (0, L) \\ u &= 1 && x = 0 \\ u_x &= -\imath\omega && x = 0 \\ u_x &= \imath\omega u && x = L \end{aligned}$$

Integrating twice, we obtain the integral equation

$$u(x) = 1 - \imath\omega x - \omega^2 \int_0^x \int_0^y n^2(z) u(z) \, \mathrm{d}z \, \mathrm{d}y.$$

We integrate the outer integral by parts to obtain

$$\left| \int_0^x \int_0^y n^2(z) u(z) \, \mathrm{d}z \, \mathrm{d}y \right| \leq \int_0^x (x-y) n^2(y) |u(y)| \, \mathrm{d}y$$

and thus

$$|u(x)| \leq \sqrt{1 + |\omega|^2 x^2} + |\omega|^2 \int_0^x (x-y) n^2(y) |u(y)| \, \mathrm{d}y.$$

Eq. (8.25) now follows from Gronwall's inequality. □

Lemma 8.2.5. *The following inequality holds*

$$\int_0^L n^2 |u|^2 \, \mathrm{d}x \leq n_+^2 L \exp\left(|\omega|^2 n_+^2 L^2 \right) \left(1 + |\omega|^2 L^2 / 3 \right) \tag{8.26}$$

where $n_+ = \max_{x \in (0,L)} n(x)$.

Proof. Using Lemma 8.2.4, we compute

$$\begin{aligned}\int_0^L n^2|u|^2\,\mathrm{d}x &\le \int_0^L n^2(x)(1+|\omega|^2x^2)\exp\left(2|\omega|^2\int_0^x (x-y)n^2(y)\,\mathrm{d}y\right)\,\mathrm{d}x\\ &\le n_+^2\exp\left(2|\omega|^2\int_0^L (L-y)n^2(y)\,\mathrm{d}y\right)\int_0^L 1+|\omega|^2x^2\,\mathrm{d}x\\ &\le n_+^2\exp\left(|\omega|^2n_+^2L^2\right)\left(L+|\omega|^2L^3/3\right)\end{aligned}$$

as desired. □

Proof of Theorem 8.2.2. Using Eqs. (8.23) and (8.24) and Lemma 8.2.5, we compute

$$\begin{aligned}|\Im\omega| &= \frac{|u(0)|^2+|u(L)|^2}{2\int_0^L n^2|u|^2\,\mathrm{d}x}\\ &\ge \frac{\exp\left(-|\omega|^2n_+^2L^2\right)}{n_+^2L\left(1+|\omega|^2L^2/3\right)}\\ &= \frac{\exp\left(-\left(|\Re\omega|^2+|\Im\omega|^2\right)n_+^2L^2\right)}{n_+^2L\left(1+\left(|\Re\omega|^2+|\Im\omega|^2\right)L^2/3\right)}\\ &\equiv f(|\Im\omega|)\end{aligned}$$

This is a nonlinear inequality for $|\Im\omega|$. Note that the function $f(x)$ is a monotonically decreasing function for $x \ge 0$ with $f \downarrow 0$ as $x \uparrow \infty$. Thus, for $\xi \ge |\Im\omega|$, $f(\xi) \le f(|\Im\omega|) \le |\Im\omega|$. Thus for all $\xi > 0$,

$$|\Im\omega| \ge \min[\xi, f(\,\xi\,)].$$

To obtain the optimal bound, one would choose $\xi = \xi_0$ such that $\xi_0 = f(\,\xi_0\,)$. For simplicity, we choose $\xi_0 = (n_+L)^{-1}$. If $n_+ > e^{-1}$, we find that $\min[\xi_0, f(\xi_0)] = f(\xi_0)$ for all $\Re\omega$ and thus

$$|\Im\omega| \ge \frac{3\exp\left(-n_+^2L^2|\Re\omega|^2\right)}{eL\left(1+3n_+^2+n_+^2L^2|\Re\omega|^2\right)}$$

as desired. □

Remark 8.2.6. Theorem 8.2.2 also gives an upper bound for the quality factor, defined $Q \equiv \frac{|\Re\omega|}{2|\Im\omega|}$. In particular this bound shows that $Q \downarrow 0$ as $\Re\omega \downarrow 0$.

The following proposition shows that there is a triangular resonance-free region in the lower-half complex plane.

Proposition 8.2.7. *Let $n \in \mathcal{A}_{sym}$, i.e., $n_- < n(x) < n_+$ and $n(L-x) = L(x)$. If $\omega \in Res_n$ is a one-dimensional Helmholtz resonance satisfying (8.4) with $d = 1$, then*

$$\omega \notin \{\omega \colon |\Im\omega| > |\Re\omega| \text{ and } |\Im\omega| \leq \frac{1}{n_+^2 L}\}.$$

Proof. The proof of this theorem follows [Harrell and Svirsky, 1986]. Let $|\Im\omega| > |\Re\omega|$ and we'll show that $|\Im\omega| > \frac{1}{n_+^2 L}$. Using Eq. (8.23) and $|u(0)| = |u(L)|$ (see Prop. 8.6.1), we have

$$|\Im\omega| \geq \frac{|u(0)|^2}{n_+^2 \int_0^L |u|^2 \, \mathrm{d}x}. \tag{8.28}$$

Kato's inequality [Reed and Simon, 1980] and $|\Im\omega| \geq |\Re\omega| \;\Rightarrow\; \Re(\omega^2) \leq 0$ then give

$$\Delta|u| \geq \Re\left(\frac{\overline{u}}{|u|}\Delta u\right) = -\Re(\omega^2) n^2 |u| \geq 0.$$

We now apply the maximum principle to the subharmonic function $|u(x)|$ to obtain $|u(x)| \leq |u(0)|$. It now follows from Eq. 8.28 that $|\Im\omega| > \frac{1}{n_+^2 L}$. □

8.3 Existence of a solution for the spectral optimization problem

In this section, we consider the spectral optimization problem in dimension $d = 1, 2, 3$ with admissible set $\mathcal{A}(\Omega, n_-, n_+)$, as defined in Eq. (8.6), the set of $n(\mathbf{x})$ satisfying upper and lower bounds on the compact set $\overline{\Omega}$ with $n(\mathbf{x}) \equiv 1$ for $\mathbf{x} \notin \overline{\Omega}$. Recall

$$\mathrm{Res}_n^\rho \;\equiv\; \{\, \omega \colon \omega \text{ is a scattering frequency for the structure } n(\mathbf{x}), \quad |\Re\omega| \leq \rho \,\}. \tag{8.29}$$

Theorem 8.3.1. *Consider the scattering resonance problem on $\mathbb{R}^d$, $d = 1, 2, 3$. Fix $\rho \geq 0$. Assume that there exists $n \in \mathcal{A}$ such that $\mathrm{Res}_n^\rho \neq \emptyset$. Then the double infimum, defined in (8.10),*

$$\Gamma_\star^\rho(\mathcal{A}) \;=\; \inf_{n \in \mathcal{A}} \; \inf_{\mathrm{Res}_n^\rho} \; |\Im\omega|$$

is strictly positive and is attained for an admissible structure. That is, there exists $n_\star \in \mathcal{A}$, with associated longest-lived resonance mode, $u_\star$, of frequency $\omega_\star \in \mathrm{Res}_n^\rho$ and such that $|\Im\omega_\star| = \Gamma_\star^\rho(\mathcal{A})$.

Proof. Since $\mathrm{Res}_n^\rho \neq \emptyset$ and $\mathrm{Res}_n^\rho \subset \{\omega \colon \Im\omega \leq 0\}$, we have $0 \leq |\Im\omega| < \infty$ and there is a minimizing sequence $\{n_m\}_{m=1}^\infty \subset \mathcal{A}_1$, such that

$$\inf\{\, |\Im\omega| : \omega \in \mathrm{Res}_{n_m}^\rho \,\} \downarrow \Gamma_\star^\rho \geq 0, \quad \text{as } m \uparrow \infty \,. \tag{8.30}$$

We first show that $\Gamma^\rho_\star$ is attained and then conclude the proof by showing $\Gamma^\rho_\star > 0$.

Let $(\ \omega_m, u_m(\mathbf{x}, \omega_m)\ )$ denote the resonance pair corresponding to this minimizing sequence of structures in $\mathcal{A}$. Since $[-\rho, \rho]$ is compact, there exists a convergent subsequence, which we continue to denote by $\{\omega_m\}$, with $\omega_m \to \omega_\star$.

By linearity and boundness of Ω we can impose the normalization $\|u_m\|_{L^2(\Omega)} = 1$. Squaring the differential equation for u_m, integrating over Ω and using the uniform boundedness of n_m yields that the sequence $\{u_m\}$ is uniformly bounded in $H^2(\Omega)$. By Rellich's Lemma, for any $s < 2$ there exists $u_\star \in H^s(\Omega)$ and a strongly convergent subsequence converging to $u_\star$. Moreover, u_m is uniformly Hölder continuous with exponent $\alpha \in (0, 1/2)$. Thus, $\{u_m(\mathbf{x})\}$ is uniformly bounded and equicontinuous. Therefore, there exists a subsequence, again denoted $\{u_m\}$, such that $u_m \to u_\star$ uniformly in $\overline{\Omega}$. It follows that $\|u_\star\|_{L^2(\Omega)} = 1$ and thus $u_\star$ is nonzero.

Note also that by the uniform bounds $n(x) \subset [n_-, n_+]$ imply that the sequence $\{n_m\}$, is uniformly bounded in $L^2(\Omega)$, and therefore along some subsequence converges weakly in $L^2(\Omega)$ to some $n_\star$, *i.e.*, $n_m \rightharpoonup n_*$ with $n_\star \in \mathcal{A}$.

It remains to show that $(u_\star(\cdot, \omega_\star), \omega_\star)$ is a resonance pair, *i.e.*, $u_\star$ is non-trivial (established just above) and satisfies

$$\begin{aligned} \Delta u_\star(\mathbf{x}) \ + \ n^2_\star(\mathbf{x})\ \omega^2_\star\ u_\star(\mathbf{x}) \ &= \ 0 \\ \frac{\partial u_\star}{\partial |\mathbf{x}|} - \imath \omega_\star u_\star &= o\left(|\mathbf{x}|^{-\frac{d-1}{2}}\right). \end{aligned} \tag{8.31}$$

It is well known (see *e.g.*, *[Colton and Kress, 1998])* that Eq. (8.31) *is equivalent to the Lippmann-Schwinger integral equation*

$$u_\star(\mathbf{x}) = \omega^2_\star \int_\Omega G(|\mathbf{x}-\mathbf{y}|, \omega_\star)\ [n^2_\star(\mathbf{y}) - 1]\ u_\star(\mathbf{y})\, \mathrm{d}\mathbf{y} \quad . \tag{8.32}$$

Here, $G(|\mathbf{x}-\mathbf{y}|, \omega) \ = \ \left(-\Delta - \omega^2\right)^{-1}(|\mathbf{x}-\mathbf{y}|, \omega)$ *denotes the free space d-dimensional outgoing Green's function:*

$$G(|\mathbf{x}-\mathbf{y}|, \omega) = \begin{cases} -(2\imath\omega)^{-1} \exp(\imath\omega|x-y|) & d = 1 \\ -(4\imath)^{-1} H^{(1)}_0(\omega|\mathbf{x}-\mathbf{y}|) & d = 2 \\ (4\pi|\mathbf{x}-\mathbf{y}|)^{-1} \exp(\imath\omega|\mathbf{x}-\mathbf{y}|) & d = 3. \end{cases} \tag{8.33}$$

For each $m \in \mathbb{N}$, we also have that

$$u_m(\mathbf{x}) = \omega_m^2 \int_\Omega G(|\mathbf{x}-\mathbf{y}|, \omega_m)[n_m^2(\mathbf{y}) - 1]u_m(\mathbf{y})\, \mathrm{d}\mathbf{y}. \tag{8.34}$$

As noted above, the right hand side of (8.34) converges to $u_\star(\mathbf{x})$ uniformly on $\overline{\Omega}$. Therefore to establish (8.32), it suffices to show that

$$\omega_m^2 \int_\Omega G(|\mathbf{x}-\mathbf{y}|, \omega_m)[n_m^2(\mathbf{y})-1]u_m(\mathbf{y})\, \mathrm{d}\mathbf{y} \; - \; \omega_\star^2 \int_\Omega G(|\mathbf{x}-\mathbf{y}|, \omega_\star)[n_\star^2(\mathbf{y})-1]u_\star(\mathbf{y})\, \mathrm{d}\mathbf{y} \; \longrightarrow \; 0 \tag{8.35}$$

for $\mathbf{x} \in \overline{\Omega}$. To show that the difference in (8.35) tends to zero, it suffices to show

$$\int_\Omega [G(|\mathbf{x}-\mathbf{y}|, \omega_m) - G(|\mathbf{x}-\mathbf{y}|, \omega_\star)] \; [n_m^2(\mathbf{y}) - 1] \; u_m(\mathbf{y})\, \mathrm{d}\mathbf{y} \to 0 \tag{8.36}$$

$$\int_\Omega G(|\mathbf{x}-\mathbf{y}|, \omega_\star) \; \left[n_m^2(\mathbf{y}) - n_\star^2(\mathbf{y})\right] \; u_m(\mathbf{y})\, \mathrm{d}\mathbf{y} \to 0 \tag{8.37}$$

$$\int_\Omega G(|\mathbf{x}-\mathbf{y}|, \omega_\star) \; [n_m^2(\mathbf{y}) - 1] \; \left(\; u_m(\mathbf{y}) - u_\star(\mathbf{y}) \; \right) \; \mathrm{d}\mathbf{y} \to 0 \tag{8.38}$$

Note that

$$\sup_{\omega \in \mathbb{R}, \mathbf{x} \in \overline{\Omega}} \|G(|\mathbf{x} - \cdot|, \omega)\|_{L^1(\Omega)} < \infty \text{ (integrable singularity).} \tag{8.39}$$

We have indeed that

1. *(8.36) holds by (8.39) and the dominated convergence theorem.*

2. *(8.38) holds by the uniform bound on n_m, (8.39), and uniform convergence of u_m to $u_\star$ on $\overline{\Omega}$.*

3. *Finally, (8.37) can be proved as follows. Let $\mathbf{x} \in \Omega$ and choose $\varepsilon > 0$ be arbitrary and sufficiently small. Divide the region of integration into*

$$\Omega = \Omega_{1,\varepsilon} \cup \Omega_{2,\varepsilon} \equiv \Omega \setminus \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| < \varepsilon\} \; \cup \; \{\mathbf{y} : |\mathbf{y} - \mathbf{x}| \geq \varepsilon\}.$$

By (8.39) and the boundness of both $\{n_m\}$ and $\{u_m\}$, the integral in (8.37), with Ω replaced by $\Omega_{2,\varepsilon}$ is $\mathcal{O}(\varepsilon^2)$. Consider now the integral (8.37) with Ω replaced by $\Omega_{1,\varepsilon}$. The integrand of this integral is uniformly bounded by $C_{\mathcal{A}} \times |G(\varepsilon; \omega_\star)|$ and tends to zero as $m \to \infty$ almost everywhere in Ω. Hence, by the dominated convergence theorem (8.37) holds.

Finally, we claim that $\Gamma_\star^\rho > 0$. Suppose not. Then $\Gamma_\star^\rho = 0$ and the scattering resonance problem (8.31) has a non-trivial solution with real frequency $\omega_\star$. By the unique continuation principle [Colton and Kress, 1998] $u_\star \equiv 0$, a contradiction.

This completes the proof of Theorem 8.3.1 . □

Remark 8.3.2. In the example of Sec. 8.2.1, we showed that in dimensions $d = 2, 3$ for $\Omega = \{|\mathbf{x}| < a\}$ and $n(\mathbf{x}) = 1 + n_0 \mathbf{1}_\Omega$, the resonances approach the real axis as the angular momentum $\ell \uparrow \infty$. In section 8.5, we show that in dimension $d = 1$, we the sequence $n_\star^k$ for increasing $\rho^k \uparrow \infty$ such that $|\Im\omega_\star^k| \downarrow 0$ with $\Re\omega_\star^k \approx \rho^k \uparrow \infty$. Thus for $d = 1, 2, 3$ the optimal solution for the limit $\rho \uparrow \infty$ does not exist. This result contrasts the analog setting for resonances of a Schrödinger operator [Harrell, 1982; Harrell and Svirsky, 1986; Svirsky, 1987].

8.4 Optimizers are piecewise constant structures which saturate the constraints

In this section, we focus on properties of optimal solutions of the design problem

$$\min_{n \in \mathcal{A}} \ \Gamma^\rho[n]. \tag{8.40}$$

We begin by computing the variation of ω with respect to changes in the index $n(\mathbf{x})$. Denote the $L^2(\Omega)$ inner product by

$$\langle f, g \rangle = \int_\Omega \overline{f(\mathbf{x})}\ g(\mathbf{x})\, \mathrm{d}\mathbf{x}\ .$$

The first variation of a Fréchet differentiable functional $\mathcal{J}[n] \colon L^2(\Omega) \to \mathbb{C}$ is defined by $\frac{\delta \mathcal{J}}{\delta n}$, satisfying, for all $\delta n \in L^2(\Omega)$,

$$\mathcal{J}[n + \delta n] = \mathcal{J}[n] + \left\langle \frac{\delta \mathcal{J}}{\delta n}, \delta n \right\rangle + o(\|\delta n\|_2).$$

Proposition 8.4.1. *Let $(\omega, u(\mathbf{x}, \omega))$ be a scattering resonance pair of the scattering resonance problem (8.4) for index of refraction, $n(\mathbf{x})$. Then*

1. *The first variation of $\omega[n]$ with respect to n is given by*

$$\frac{\delta \omega}{\delta n}(\mathbf{x}) = -2\overline{\alpha}\ \overline{\omega}^2\ n(\mathbf{x})\ \overline{u(\mathbf{x})}^2 \tag{8.41}$$

   where $\alpha \in \mathbb{C} \setminus \{0\}$.

2. *In one-dimension, on a domain $\Omega = [0, L]$, the first variation is given by (8.41) with*

$$\alpha^{-1} = 2\omega \int_0^L n^2 u^2 + \imath[u^2(0) + u^2(L)] \tag{8.42a}$$

$$= \frac{1}{\omega} \int_0^L u_x^2 + \omega^2 n^2 u^2 \tag{8.42b}$$

Proof. The proof of Proposition 8.4.1 is given in Appendix 8.A. □

Remark 8.4.2. The following are properties of $\frac{\delta\omega}{\delta n}(x)$:

1. $\frac{\delta\omega}{\delta n}(\mathbf{x})$ is invariant of any normalization of the mode $u(\mathbf{x})$, since α depends quadratically $u^2(\mathbf{x})$.
2. Suppose $d = 1$ and $\Omega = [0, L]$. If $n(x)$ is symmetric, *i.e.*, $n(x) = n(L - x)$, then $\frac{\delta\omega}{\delta n}(x)$ is also symmetric. This follows from Prop. 8.6.1.

Proposition 8.4.3. *Any $n_\star(\mathbf{x}) \in \mathcal{A}$ which is a local minimizer the optimal design problem (8.40) is piecewise constant and attains the material bounds, i.e.,*

$$[n_\star(\mathbf{x}) - n_-]\,[n_\star(\mathbf{x}) - n_+] = 0 \text{ almost everywhere in } \mathbf{x} \in \Omega.$$

Remark 8.4.4. Proposition 8.4.3 has been previously observed computationally [Kao and Santosa, 2008; Heider *et al.*, 2008]. A similar result in one-dimension with slightly different boundary conditions is given in [Karabash, 2011]. This phenomena has also been studied in self-adjoint systems [Krein, 1955; Cox and McLaughlin, 1990a; Cox and McLaughlin, 1990b] and for the analogous problem for Schrödinger resonances [Harrell and Svirsky, 1986]. In control theory, $n(\mathbf{x})$ might be referred to as a "bang-bang control." Proposition 8.4.3 is significant because

1. the computation of Helmholtz resonances for piecewise constant $n(\mathbf{x})$ can be performed efficiently using layer potential methods [Colton and Kress, 1998] and
2. the standard manufacturing technique for photonic crystals involves drilling air holes into a dielectric slab [Joannopoulos *et al.*, 2008]. Thus, in practice, $n(\mathbf{x})$ only takes two values.

Proof. Denote by $\mathcal{J}\colon L^2(\Omega) \to \mathbb{R}$ the functional $\mathcal{J}[n] = -\Im\omega[n]$. Thus Eq. (8.40) is equivalent to

$$\begin{aligned} \underset{n\in L^2(\Omega)}{\text{minimize}} \quad & \mathcal{J}[n] \\ \text{such that} \quad & n(\mathbf{x}) - n_+ \le 0 \qquad \mathbf{x} \in \Omega \\ & n_- - n(\mathbf{x}) \le 0 \qquad \mathbf{x} \in \Omega \\ & (\Re\omega)^2 - \rho^2 \le 0 \end{aligned} \tag{8.43}$$

Introducing the dual variables $\lambda_+(\mathbf{x})$, $\lambda_-(\mathbf{x}) \in L^2(\Omega)$ and $\xi \in \mathbb{R}$ with $\lambda_+(\mathbf{x})$, $\lambda_-(\mathbf{x}) \geq 0 \ \ \forall \mathbf{x} \in \Omega$ and $\xi \geq 0$, we define the Lagrangian

$$\mathcal{L}[n, \lambda_+, \lambda_-, \xi] = \mathcal{J}[n] + \langle \lambda_+, n - n_+ \rangle + \langle \lambda_-, n_- - n \rangle + \xi \left((\Re\omega)^2 - \rho^2 \right) \qquad \lambda_+,\, \lambda_-,\, \xi \geq 0.$$

The Karush-Kuhn-Tucker (KKT) conditions, which are necessary conditions for a local minima, require

$$\frac{\delta \mathcal{J}}{\delta n} + \lambda_+ - \lambda_- + 2\xi(\Re\omega)\Re\frac{\delta\omega}{\delta n} = 0 \qquad \lambda_+,\, \lambda_-,\, \xi \geq 0. \tag{8.44a}$$

$$\lambda_+(n - n_+) = 0 \tag{8.44b}$$

$$\lambda_-(n - n_-) = 0 \tag{8.44c}$$

$$\xi \left((\Re\omega)^2 - \rho^2 \right) = 0 \tag{8.44d}$$

Eqns. (8.44b), (8.44c), and (8.44d) are referred to as the complementarity conditions. Define the sets

$$A = \{\mathbf{x} \in \Omega \colon n(\mathbf{x}) = n_+\}$$
$$B = \{\mathbf{x} \in \Omega \colon n(\mathbf{x}) = n_-\}.$$

We now decompose $\Omega = A \cup B \cup [\Omega \setminus (A \cup B)]$ and consider the three cases in turn. Note that by Prop. 8.4.1,

$$\frac{\delta \mathcal{J}}{\delta n} = -\Im\frac{\delta\omega}{\delta n} = \Im 2\overline{\alpha\omega}^2 n \overline{u}^2$$

For $x \in A$, $n(\mathbf{x}) = n_+$ and the complementarity conditions give that $\overline{\lambda} \geq 0$ and $\underline{\lambda} = 0$. Thus, $2\Im\left(\overline{\alpha\omega}^2 n \overline{u}^2\right) + \overline{\lambda} + 2\xi\Re\left(\overline{\alpha\omega}^2 n \overline{u}^2\right) = 0$ which implies that

$$2\xi(\Re\omega)\Re\left(\alpha\omega^2 u^2(\mathbf{x})\right) - \Im\left(\alpha\omega^2 u^2(\mathbf{x})\right) \leq 0 \qquad \mathbf{x} \in A.$$

For $x \in B$, $n(\mathbf{x}) = n_-$ and the complementarity conditions give that $\underline{\lambda} \geq 0$ and $\overline{\lambda} = 0$. Thus, $2\Im\left(\overline{\alpha\omega}^2 n \overline{u}^2\right) - \underline{\lambda} + 2\xi\Re\left(\overline{\alpha\omega}^2 n \overline{u}^2\right) = 0$ which implies that

$$2\xi(\Re\omega)\Re\left(\alpha\omega^2 u^2(\mathbf{x})\right) - \Im\left(\alpha\omega^2 u^2(\mathbf{x})\right) \geq 0 \qquad \mathbf{x} \in B.$$

For $x \in \Omega \setminus (A \cup B)$, $n_- \leq n(\mathbf{x}) \leq n_+$ and $\overline{\lambda} = \underline{\lambda} = 0$. Thus

$$2\xi(\Re\omega)\Re\left(\alpha\omega^2 u^2(\mathbf{x})\right) - \Im\left(\alpha\omega^2 u^2(\mathbf{x})\right) = 0 \qquad \mathbf{x} \in \Omega \setminus (A \cup B).$$

Thus the function $\alpha\omega^2 u^2(\mathbf{x})$ takes values in the complex plane on a line at an angle $\theta = \arctan(2\xi\Re\omega)$ with the real axis. Thus, the function $\exp(-i\theta)\alpha\omega^2 u^2(\mathbf{x})$ is a real function. In the neighborhood of any $\mathbf{x}$ such that $u(\mathbf{x}) \neq 0$, we can choose the sign such that $v(\mathbf{x}) := \sqrt{\pm \exp(-i\theta)\alpha\omega^2 u^2(\mathbf{x})}$ is a real function which satisfies the complex Eq. (8.4a):

$$\Delta v + \Re(\omega^2) n^2 v = -\imath \Im(\omega^2) n^2 v$$

Since the left hand side of this equation is real and the right hand side is purely imaginary, we conclude that $v \equiv 0$. Thus $u \equiv 0$.

Since $u(\mathbf{x})$ satisfies Eq. (8.4a), we know that by the unique continuation principle [Colton and Kress, 1998], it cannot vanish on an open set of Ω, or else $u(\mathbf{x}) \equiv 0$, a contradiction. Thus $|\Omega \backslash (A \cup B)| = 0$ and we identify it with the zero level set of $2\xi(\Re\omega)\Re\left(\alpha\omega^2 u^2(\mathbf{x})\right) - \Im\left(\alpha\omega^2 u^2(\mathbf{x})\right)$. □

A corollary to the proof of Prop. 8.4.3 is the following

Corollary 8.4.5. *If $n_\star(\mathbf{x}) \in \mathcal{A}$ is a local minimizer the optimal design problem (8.40) and $(u_\star, \omega_\star)$ denotes the corresponding minimizing scattering resonance pair, then there exists a $\xi \geq 0$ such that*

$$n_\star(\mathbf{x}) = \begin{cases} n_+ & 2\xi(\Re\omega_\star)\Re\left(\alpha\omega_\star^2 u_\star^2(\mathbf{x})\right) - \Im\left(\alpha\omega_\star^2 u_\star^2(\mathbf{x})\right) < 0 \\ n_- & 2\xi(\Re\omega_\star)\Re\left(\alpha\omega_\star^2 u_\star^2(\mathbf{x})\right) - \Im\left(\alpha\omega_\star^2 u_\star^2(\mathbf{x})\right) > 0 \end{cases} \tag{8.45}$$

where $\alpha \in \mathbb{C}$ is defined as in Prop. 8.4.1. If $|\Re\omega_\star| < \rho$, then $\xi = 0$.

8.5 Computation of optimal one-dimensional $n(x)$

In this section we describe one-dimensional computations that will be used to motivate analytical results in section 8.6. Computationally, we solve a slightly different problem here than the double infimum problem given in Eq. (8.9). Following [Heider *et al.*, 2008; Kao and Santosa, 2008], we pick a resonance $\tilde{\omega}$ and improve the structure $n(x)$ to extend the lifetime of that particular resonance. Thus we solve the optimization problem

$$\min_{n \in \mathcal{A}} \mathcal{J}[n] := |\Im\tilde{\omega}[n]|. \tag{8.46}$$

The local minimizers of Eq. (8.46) which additionally satisfy $|\Re\tilde{\omega}_\star| < \rho$ are local minima of Eq. (8.40).

8.5.1 Computational method

Below, we refer to the forward problem as the computation of the resonances for a given $n(x)$ and the optimization problem as the solution of Eq. (8.46).

Forward problem. For a one-dimensional, piecewise-constant, refractive index $n(x)$, the resonances satisfying Eq. (8.4) are the roots of a nonlinear system of equations obtained by imposing transmission conditions at the material discontinuities. In [Heider *et al.*, 2008], this system of equations is derived and Newton's method is applied for finding the roots of this system. We find that this method works extremely well if initialized sufficiently near the desired resonance. We initialize Newton's method by either

(a) using a finite difference discretization of Eq. (8.4) to form a quadratic eigenvalue problem (QEP) which is solved using Matlab's `polyeig` command [Tisseur and Meerbergen, 2001], or

(b) using the ω computed at a previous optimization iteration.

Other references that derive the transmission conditions at material discontinuities, including for two-dimensional, radially symmetric $n(x)$ are [Yeh, 1988; Yeh *et al.*, 1978].

Optimization problem. In Prop. 8.4.1 and App. 8.A, we computed the variation $\frac{\delta \mathcal{J}}{\delta n}$. Thus, a number of gradient-based methods are available to solve Eq. (8.46). In [Heider *et al.*, 2008; Kao and Santosa, 2008], the authors use gradient descent methods. To solve Eq. (8.46), we have applied a BFGS interior point method and the fixed point method given in Algorithm 1, which makes use of Eq. (8.45). We find that the the fixed point method converges in just a few iterations when it converges, but this is sensitive to the initial guess $n_0(x)$. The BFGS interior point method is more reliable, but requires more iterations to converge to the same tolerance.

Algorithm 1: A fixed point algorithm for solving Eq. (8.46).

Data: Initial guess $n_0(x)$, chosen scattering resonance $\tilde{\omega}$, and convergence tolerance $\epsilon > 0$.

For $n = n_0$ compute the chosen scattering resonance pair $(u, \tilde{\omega})$ by solving Eq. (8.4).

Evaluate α defined in Eq. (8.42a).

while $\|n(x) - n_+ \mathbf{1}_{\{x:\Im\alpha\tilde{\omega}^2u^2(x)>0\}} - n_- \mathbf{1}_{\{x:\Im\alpha\tilde{\omega}^2u^2(x)<0\}}\| > \epsilon$, **do**

Use the formula given in Eq. (8.45) with $\xi = 0$ to define a new index $n(x)$.

Using $n(x)$ compute the chosen scattering resonance pair $(u, \tilde{\omega})$ by solving Eq. (8.4).

Evaluate α defined in Eq. (8.42a).

8.5.2 Computational results

Let $\Omega = [0, L]$, $L = 1$, $n_+ = 2$ and $n_- = 1$. We define ω^j to be the scattering resonance for which the corresponding mode modulus has j minima[3] (see Fig. 8.4). Using the method described in Sec. 8.5.1, we solve

$$n_\star^j := \arg\min_{n \in \mathcal{A}} \quad |\Im\omega^j[n]|. \tag{8.47}$$

for $j = 0, \ldots, 9$. For each j, we plot in Fig. 8.4 the optimal refractive index $n_\star^j(x)$ and the corresponding resonance pairs denoted, $(u_\star^j, \omega_\star^j)$. In Fig. 8.5, we plot the transmission coefficient modulus, $|t^j(\omega)|$, associated with $n_\star^j(x)$. As usual, the transmission coefficient $t(\omega)$ is defined by the solution of the form

$$u(x, \omega) = \begin{cases} e^{\imath\omega x} + r(\omega)e^{-\imath\omega x} & x < 0 \\ t(\omega)e^{\imath\omega x} & x > L \end{cases} \qquad \text{for } \omega \in \mathbb{R}$$

Remark 8.5.1. We observe the following:

1. For each j, the optimal $n_\star^j$ is symmetric, *i.e.*, $n_\star^j \in \mathcal{A}_{sym} \subset \mathcal{A}$.
2. For each j, $n_\star^j(0) = n_\star^j(L/2) = n_\star^j(L) = n_+$.
3. For each j, the optimal $n_\star^j(x)$ is roughly periodic with a defect near $x = L/2$. The number of repeated blocks increases as j increases. The width of the repeated structure approximately satisfies Bragg's relation as defined in Eq. (8.14). For even modes, the defect width is small and for odd modes, the defect width is approximately twice the Bragg width. These observations will be made more precise in Sec. 8.6.
4. The values $|\Im\omega_\star^j|$ are monotonically decreasing in j, *i.e.*,
$$\min_{n \in \mathcal{A}} |\Im\omega^{j+1}[n]| \; < \; \min_{n \in \mathcal{A}} |\Im\omega^j[n]|.$$
5. The values $\Re\omega_\star^j$ are monotonically increasing in j, *i.e.*, the optimal modes $u_\star(x)$ becomes increasingly oscillatory as $j \uparrow \infty$.

[3] except for mode 0, which has one minima and whose eigenvalue is purely imaginary

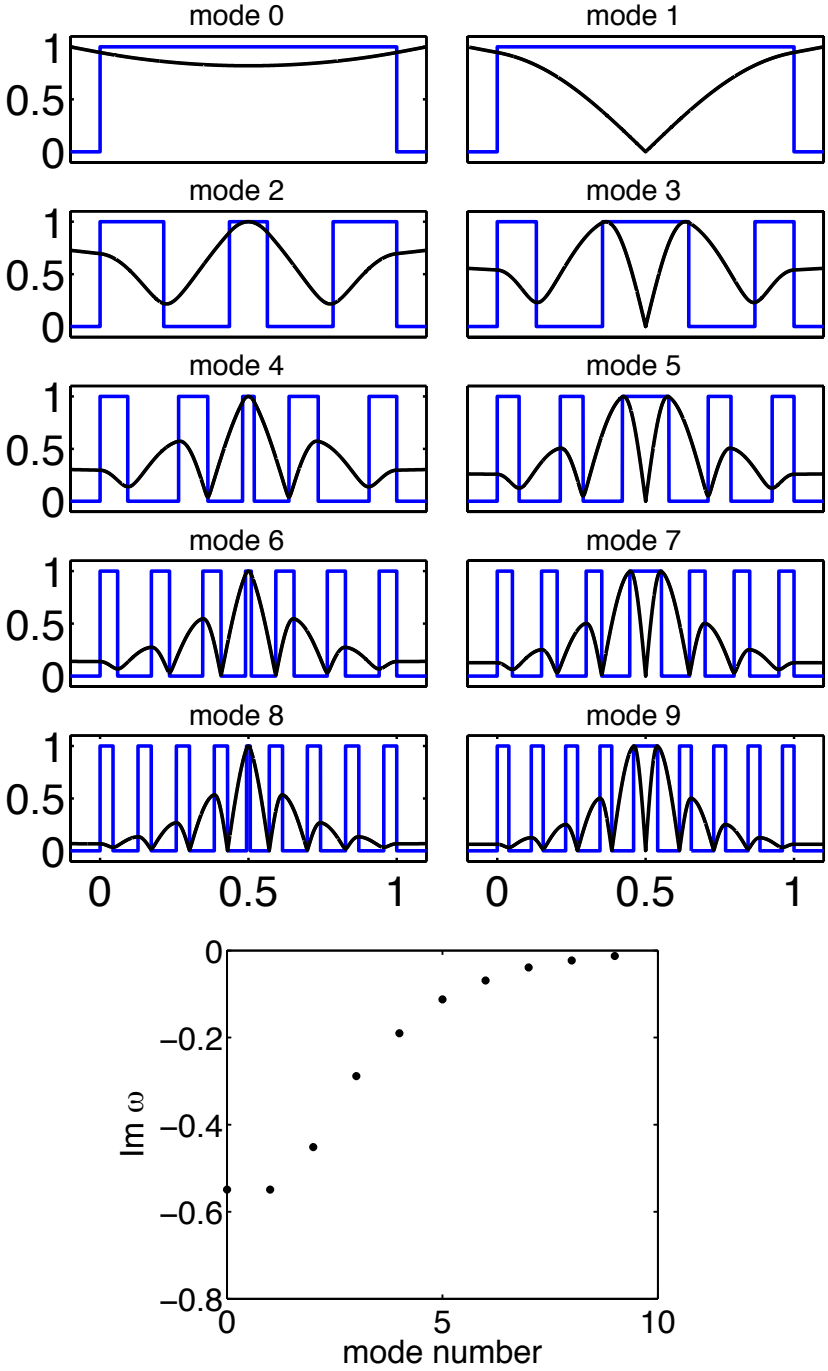


Figure 8.4: (top) For $j = 0, 1, \ldots, 9$, a plot of $n_\star^j$ (blue) minimizing Eq. (8.47) and the associated modulus modes $|u_\star^j|$ (black). (bottom) A plot of $\Im\omega_\star^j$ for $j = 0, 1, \ldots, 9$.

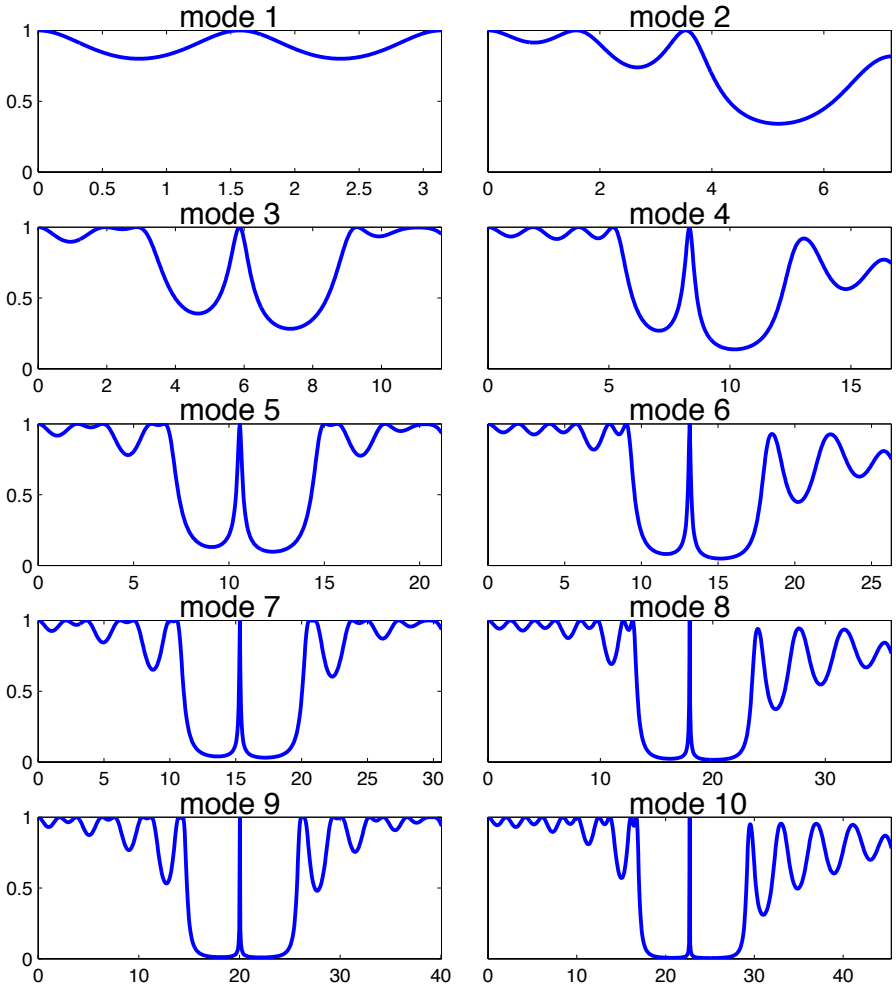


Figure 8.5: The transmission coefficients $|t^j(\omega)|$ corresponding to the refractive indices $n_\star^j(x)$ minimizing Eq. (8.47) for $j = 1, \ldots, 10$. The range of the ω-axis is taken to be $[0,\ 2\Re\omega_\star^j]$ to demonstrate that in the center, $|t^j(\Re\omega_\star^j)| \approx 1$, while the density of states near $\omega = \Re\omega_\star^j$ is very small as $j \uparrow \infty$.

8.6 Characterization of locally optimal $n(x)$ in dimension one / Bragg relation

In this section, we fix n_+, $n_- = 1$, ρ, and a domain $[0, L]$. We will use a combination of numerical observations from section 8.5 and analysis to characterize one-dimensional refractive indices $n_\star \in \mathcal{A}$ which are local minima of the optimization problem

$$\min_{n \in \mathcal{A}} \Gamma^\rho[n]. \tag{8.40}$$

Recall that local minimizers of Eq. (8.46) which additionally satisfy $|\Re\omega_\star| < \rho$ are local minima of Eq. (8.40). In this section, we only consider states such that $|\Re\omega_\star| < \rho$, that is $\omega_\star$ is an interior point on the constraint set and $\xi = 0$ in Corollary 8.4.5.

Even / oddness properties of $n_\star$ and $u_\star$. We begin the discussion with even/oddness properties of the optimal refractive indices $n_\star(x)$ satisfying Eq. (8.40) and the corresponding modes $u_\star(x)$. The following conjecture is supported by numerical experiments (see Sec. 8.5.2).

Conjecture 8.6.1. *There exists $n_\star \in \mathcal{A}_{sym}$ such that $\Gamma^\rho[n_\star] = \Gamma^\rho_\star(\mathcal{A})$. That is, the refractive index obtained by minimizing Γ^ρ over the set $\mathcal{A}$ can be taken to be symmetric.*

Due to this conjecture, in what follows we only discuss maximizing over $\mathcal{A}_{sym}$.

Proposition 8.6.1. *If $d = 1$ and $n \in \mathcal{A}_{sym}$, then any resonance state satisfying the Helmholtz Eq. (8.4) is either even or odd with respect to $x = L/2$.*

Proof. Let $(u(x), \omega)$ be a scattering resonance pair satisfying Eq. (8.4). Since $n(L-x) = n(x)$, the following are also solutions:

$$u_1(x) = u(x) + u(L - x)$$
$$u_2(x) = u(x) - u(L - x).$$

Computing the Wronskian

$$W(u_1, u_2) = u_1 u_2' - u_1' u_2,$$

and using Eq. (8.4a), we find that $W'(x) = 0$ and thus the Wronskian is constant in x. Then using the outgoing boundary condition, we find that $W(x) \equiv 0$. Thus u_1 and u_2 are linearly dependent which is possible only if either u_1 or u_2 vanishes identically. Thus, $u(x)$ is either even or odd. □

From the numerical experiments in Sec. 8.5.2, we observe that modes $u_\star$ corresponding to locally optimal $n_\star$ may be either even or odd (see Fig. 8.4). It follows that the solution of (8.40), will be either even or odd depending on the choice of ρ, L, and n_+.

Remark 8.6.2. For $n(x) \in \mathcal{A}_{sym}$, *i.e.*, symmetric about $L/2$, the modes being even or odd implies that $u(L/2) = 0$ or $u'(L/2) = 0$ respectively. In both cases, this implies that $u(x)$ and $u'(x)$ do not vanish for any other x-value because this would result in $u(x)$ being an eigenfunction of a self-adjoint operator with real eigenvalue, which is a contradiction.

We now prove the following proposition which describes the change in argument of a solution $u(x)$ satisfying the Helmholtz Eq. (8.4) with arbitrary index $n(x)$.

Proposition 8.6.3. *Let $(\omega, u(\cdot,\omega))$ denote a scattering resonance pair satisfying the Helmholtz Eq. (8.4) with $n \in \mathcal{A}$. There exists a unique value $x_\# \in [0, L]$ such that*

$$2|\Im\omega| \int_0^{x_\#} n^2(z)\,|u(z,\omega)|^2\,\mathrm{d}z = |u(0)|^2. \tag{8.48}$$

If $\Re\omega > 0$, the corresponding resonance state has increasing argument for $x > x_\#$ and decreasing argument for $x < x_\#$. The opposite statement holds true for the resonance state corresponding to $-\overline{\omega}$. Furthermore, if $n(x) \in \mathcal{A}_{sym}$, i.e., symmetric about $x = L/2$, then $x_\# = L/2$ and

$$\frac{\mathrm{d}}{\mathrm{d}x} \arg u(x) = 2(\Re\omega)|\Im\omega||u(x)|^{-2} \int_{L/2}^{x} n^2(z)|u(z)|^2\,\mathrm{d}z. \tag{8.49}$$

Proof. A similar proof may be found in [Harrell and Svirsky, 1986]. We use the Ricatti transformation $y = \frac{u'}{u}$ and note that $\Im y = (\Im(\log u))' = (\arg u)'$. We derive a differential equation for $\Im y$ and then show that the solution has the desired properties. Using Eq. (8.4a) we obtain

$$y' = -\omega^2 n^2 - y^2.$$

Taking the imaginary part we obtain

$$(\Im y)' = 2(\Re\omega)|\Im\omega|n^2 - 2(\Re y)(\Im y).$$

Then using the integrating factor $\exp(\log|u|^2) = |u|^2$, we have the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[|u|^2 \Im y\right] = 2(\Re\omega)|\Im\omega|n^2|u|^2.$$

Integrating from 0 to x we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\arg u(x) &= \Im y(x)\\ &= |u(x)|^{-2}\left(|u(0)|^2\Im y(0) + 2(\Re\omega)|\Im\omega|\int_0^x n^2(z)|u(z)|^2\,\mathrm{d}z\right)\\ &= |u(x)|^{-2}(\Re\omega)\left(-|u(0)|^2 + 2|\Im\omega|\int_0^x n^2(z)|u(z)|^2\,\mathrm{d}z\right)\end{aligned}$$

since $y(0) = u'(0)/u(0) = -\imath\omega$. Integrating from the other direction we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}\arg u(x) = |u(x)|^{-2}(\Re\omega)\left(|u(L)|^2 - 2|\Im\omega|\int_x^L n^2(z)|u(z)|^2\,\mathrm{d}z\right).$$

If $\Re\omega > 0$, then $[\arg u(x)]' < 0$ at $x = 0$ and $[\arg u(x)]' > 0$ at $x = L$. By continuity, there exists an $x_\#$ satisfying Eq. (8.48). □

Remark 8.6.4. If $n(x) \in \mathcal{A}_{sym}$, then for any mode $u(x)$ satisfying the Helmholtz Eq. (8.4), $|\frac{\mathrm{d}}{\mathrm{d}x}\arg u(x)|$ is bounded above on the interval $[0, L]$. If $u(x)$ is even about $x = L/2$, this follows from Eq. (8.49), Prop. 8.2.5 and Remark 8.6.2. If $u(x)$ is odd about $x = L/2$, then $|u(L/2)| = 0$ and one may use and L'Hôpital's rule to show that

$$\lim_{x\to L/2}|u(x)|^{-2}\int_{L/2}^x n^2(z)|u(z)|^2\,\mathrm{d}z = 0.$$

Proposition 8.6.3 can be used to significantly strengthen Prop. 8.4.3 and Cor. 8.4.5 for locally optimal $n_\star(x)$ in dimension one. Here, Corollary 8.4.5 implies that $\arg u_\star(x)$ cannot change by more than $\pi/2$ on any interval with constant n. By Remark 8.6.4, $|\frac{\mathrm{d}}{\mathrm{d}x}\arg u_\star(x)|$ is bounded above, implying that the length of any interval is bounded below. Thus, a locally optimal $n_\star(x)$ has a finite number of discontinuity points, say $N+1$. We denote these discontinuity points by $\{x_j\}_{j=0}^N$, such that[4]

$$n_\star(x) = \begin{cases} n_+ & x \in (x_j, x_{j+1}),\quad j \text{ even}\\ 1 & \text{otherwise.}\end{cases} \tag{8.50}$$

In Fig. 8.6, we illustrate Eqs. (8.50) and (8.45) by plotting an optimal refractive index $n_\star(x)$ and the quantity $\Im[\alpha\omega_\star^2 u_\star^2(x)]$. (This mode is the same as the one obtained in Sec. 8.5.2 for $j = 4$.)

Since the energy of a mode concentrates where n is large and we expect the energy to be concentrated in the center of the interval $\Omega = [0, L]$, we conjecture the following:

[4]Note: we have not ruled out the possibility that $x_0 > 0$ and $x_N < L$.

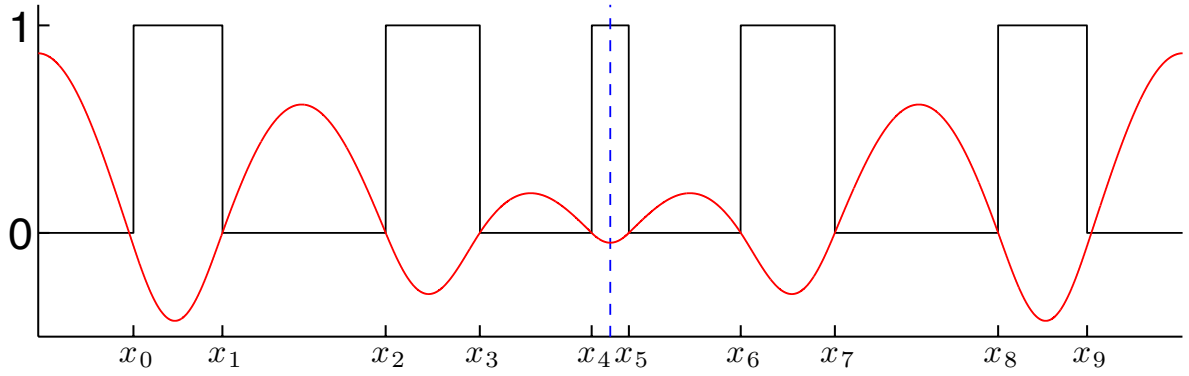


Figure 8.6: A plot of an optimal $n_\star(x) - 1$ in black and $\Im[\alpha\omega_\star^2 u_\star^2(x)]$ in red. Note that $x_0 = 0$ and $\Im(\alpha\omega_\star^2 u_\star^2(0)) \neq 0$, but the optimality condition (8.45) is satisfied.

Conjecture 8.6.2. $n_\star(L/2) = n_+$.

Conjectures 8.6.1 and 8.6.2 imply that the number of intervals with $n = n_+$ is odd. Equivalently, the total number of intervals, $N = 4M + 1$ for $M \in \mathbb{N}$. For example, for the refractive index in Fig. 8.6, $N = 9$ and $M = 2$. We refer to $[x_0, x_1]$ as the *left-most interval*, $[x_{N-1}, x_N]$ as the *right-most interval*, and $[x_{(N-1)/2}, x_{(N+1)/2}]$ as the *center interval*. We refer to all other intervals as *interior intervals*.

Proposition 8.6.5. *The length of an interior interval $I_j = [x_j, x_{j+1}]$ satisfies*

$$|x_{j+1} - x_j| \geq \frac{\min_{x\in I_j} |u_\star(x)|^2}{4|u_\star(0)|^2} \quad \frac{2\pi}{|\Re\omega_\star|}. \tag{8.51}$$

Proof. Corollary 8.4.5 implies that the argument of $u_\star(x)$ changes by exactly $\pi/2$ on an interior interval $I_j = [x_j, x_{j+1}]$. Using Eq. (8.49), we compute

$$\begin{aligned}
\frac{\pi}{2} &= 2|\Re\omega_\star|\ |\Im\omega_\star| \int_{x_j}^{x_{j+1}} |u_\star(x)|^{-2} \int_{L/2}^{x} n_\star^2(z)|u_\star(z)|^2\,\mathrm{d}z\,\mathrm{d}x \\
&\leq 2|\Re\omega_\star|\ |\Im\omega_\star| \int_{x_j}^{x_{j+1}} |u_\star(x)|^{-2} \int_{L/2}^{L} n_\star^2(z)|u_\star(z)|^2\,\mathrm{d}z\,\mathrm{d}x \\
&= |\Re\omega_\star|\ |u_\star(0)|^2 \int_{x_j}^{x_{j+1}} |u_\star(x)|^{-2}\,\mathrm{d}x \\
&\leq |\Re\omega_\star|\ |u_\star(0)|^2\ |x_{j+1} - x_j|\ \max_{x\in I_j} |u_\star(x)|^{-2}
\end{aligned}$$

where used Eq. (8.23) and the symmetry of $n(x)$ and $|u_\star(x)|$. Now by Remark 8.6.2, $|u_\star(x)| > 0$ for all $x \in I_j$, from which Eq. (8.51) follows. □

Locally optimal refractive indices and the Bragg relation. We now show that a sufficient condition for the length of all the non-center intervals, *i.e.*, $I_j = [x_j, x_{j+1}]$ for all $j \neq (N-1)/2$, to satisfy the Bragg relation is that the argument of the mode, $u_\star$ change by $\pi/2$ on the first interval:

$$\arg u_\star(x_1) - \arg u_\star(x_0) = \frac{\pi}{2}. \tag{8.52}$$

Denote $\sigma(x) := \Im(\alpha\omega_\star^2 u_\star^2(0))$. Clearly $\sigma(0) = 0 \Leftrightarrow$ Eq. (8.52) holds. Note that in Fig. 8.6, we observed that $\sigma(x) \neq 0$, but $\sigma(x) \approx 0$. Hence, we expect that Eq. (8.52) only approximately holds.

Proposition 8.6.6 (Bragg relation)**.** *Suppose Eq. (8.52) holds. Then the the length of all non-center intervals, i.e., $I_j = [x_j, x_{j+1}]$ for all $j \neq (N-1)/2$ is given by*

$$d_\pm = \frac{1}{4}\frac{2\pi}{n_\pm|\Re\omega_\star|} \;=\; \frac{1}{4}\,\lambda_\pm$$

where $\lambda_\pm$ is the "effective" wavelength of $u_\star$ in the medium $n_\pm$. Thus, a full "period" of the repeated structure is

$$\delta = d_+ \;+\; d_- \;=\; \frac{1}{2n_h}\frac{2\pi}{|\Re\omega_\star|},$$

where n_h denotes the harmonic mean, defined by

$$\frac{1}{n_h} \;=\; \frac{1}{2}\left(\frac{1}{n_+} \;+\; \frac{1}{n_-}\right) \tag{8.53}$$

Additionally, the length of the center interval is less than $2d_+$.

Proof. The solution on any interval $[x_j, x_{j+1}]$ where $n(x) \equiv n_\pm$ can be written

$$u(x) = \alpha\cos[\omega n_\pm(x - x_j)] - \imath\beta/n_\pm \sin[\omega n_\pm(x - x_j)]$$

where α and β are chosen such that

$$u(x_j) = \alpha$$
$$u'(x_j) = -\imath\omega\beta.$$

Write $\omega_\star = \omega_R + \imath\omega_I$. We compute the following formulas, which we use in the argument below:

$$\begin{aligned} u'(x) &= -\alpha\omega n_\pm \sin[\omega n_\pm (x - x_j)] - \imath\omega\beta \cos[\omega n_\pm (x - x_j)] \\ \cos\left(\frac{\pi\omega}{2\omega_R}\right) &= -\imath \sinh\left(\frac{\pi\omega_I}{2\omega_R}\right) \\ \sin\left(\frac{\pi\omega}{2\omega_R}\right) &= \cosh\left(\frac{\pi\omega_I}{2\omega_R}\right) \\ \Rightarrow \quad u'(x_j + d_\pm) &= -\alpha\omega n_\pm \cosh\left(\frac{\pi\omega_I}{2\omega_R}\right) - \omega\beta \sinh\left(\frac{\pi\omega_I}{2\omega_R}\right). \end{aligned}$$

Without loss of generality, we assume $\Re\omega > 0$ and that on the first interval, $I_0 = [x_0, x_1]$, $\alpha = \beta = 1$ which satisfies the outgoing boundary conditions.

Case 1. Suppose that on the interval $I_j = [x_j, x_{j+1}]$, $\alpha, \beta \in \mathbb{R}$ and $n = n_+$. This is the case for the first interval, I_0. The point x_{j+1} is defined to be that for which $\arg u_\star(x)$ has increased by $\pi/2$ because of the assumption in Eq. 8.52 on the first interval and because of Corollary 8.4.5 on an interior interval. But, $\arg u_\star(x)$ increasing by $\pi/2$ is equivalent to $\Re u(x_{j+1}) = 0$. The solution to this equation yields $x_{j+1} = x_j + d_+$. Since $u'(x_j + d_+) = \omega\gamma$ where $\gamma \in \mathbb{R}$ the coefficients α and β in the next interval (x_{j+1}, x_{j+2}) will be purely imaginary.

Case 2. Suppose $n = n_-$ and α and β are purely imaginary for an interior interval $[x_j, x_{j+1}]$, *i.e.*, $\alpha, \beta \in \imath\mathbb{R}$. Then x_{j+1} is defined to be the point satisfying $\Im u(x_{j+1}) = 0$. The solution to this equation yields $x_{j+1} = x_j + d_-$. Since $u'(x_j + d_-) = \imath\omega\gamma$ where $\gamma \in \mathbb{R}$, we have that the coefficients α and β in the next interval (x_{j+1}, x_{j+2}) will be purely real. This is precisely Case 1.

Thus, the intervals alternate between Case 1 and 2 until the center interval, $I_{(N-1)/2} = [x_{(N-1)/2}, x_{(N+1)/2}]$. is reached. Here, $\arg u(x)$ increases on $[x_{(N-1)/2}, L/2]$ and decreases on $[L/2, x_{(N+1)/2}]$. By Corollary 8.4.5, $\arg u(x)$ changes by less than $\pi/2$ on each of these half intervals, which implies that the width of each of these half intervals is less than d_+. □

8.7 Conclusions / Discussion

From the discussion given in Sec. 8.1.3, we do not expect that locally optimal structures will exactly satisfy the Bragg relation but we might expect that it will be approximately satisfied. Indeed, we have verified this in numerical experiments. Future work will include studying the

deviation of locally optimal $n_\star(x)$ from the Bragg relation and finding asymptotic limits for which the assumption in Prop. (8.6.6) is satisfied.

There are many motivations for trapping light in cavities, some of which involve modified figures-of-merit than a resonance lifetime [Krauss, 2008]. These include the quality factor, Q which is proportional to $\frac{\Re\omega}{|\Im\omega|}$ and the Purcell factor which is proportional to Q/V where V is the mode volume. The Purcell factor is particularly important in applications where a strong light-matter interaction is required and the maximization of this quantity would be an interesting extension of this work.

To conclude, we summarize our expectations for the dependence of $\Gamma^\rho_\star(\mathcal{A})$ as defined in Eq. (8.10) on the parameters ρ, n_+, L.

1. The optimal resonance $\omega_\star \in \mathcal{A}$ such that $\Gamma^\rho_\star(\mathcal{A}) = |\Im\omega_\star|$ has

$$\Re\omega_\star \approx \rho \qquad \text{and} \qquad \Im\omega_\star \sim A\exp(-B\rho)$$

   where A and B are constants dependent on $\mathcal{A}$.

2. The number of times that the optimal relative wave speed $n_\star(x)$ alternates between n_+ and n_- on $[0, L/2]$ is approximately

$$M = (N-1)/4 \approx \frac{L/2}{\delta} = \frac{n_h \Re\omega L}{2\pi} \approx \frac{n_h \rho L}{2\pi}.$$

3. From previous, unpublished work of M. I. Weinstein and P. Heider, we suspect that $|\Im\omega_\star| \sim n_+^{-M}$ and thus

$$\Gamma^\rho_\star \sim n_+^{-\rho n_h L/2\pi}.$$

8.A Calculation of variation of ω with respect to n

Let $R(\omega) := (-\Delta - \omega^2)^{-1}$ denote the free resolvent operator so that Eq. (8.4) may be written as in Eq. (8.32):

$$u = \omega^2 R(\omega) m u$$

where $m := n^2 - 1$ is non-negative and has support in Ω. For $\chi > 0$ and exponentially decaying, $\chi \sim e^{-\alpha|\mathbf{x}|}$ as $|x| \uparrow \infty$, we define

$$U = \chi u \tag{8.54a}$$

$$R_\chi(\omega) = \chi R(\omega)\chi \tag{8.54b}$$

$$M = \chi^{-1} m \chi^{-1} \tag{8.54c}$$

which satisfy

$$\left[\mathrm{Id} - \omega^2 R_\chi(\omega) M\right] U = 0. \tag{8.55}$$

For $|\Im\omega| < \alpha$, the operator $R_\chi(\omega)M \colon L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is Hilbert-Schmidt and therefore compact. Analytic Fredholm Theory then gives that the resolvent can be analytically continued to the lower-half ω-plane with poles which correspond to resonances.

Denoting variations with respect to n using a prime, $'$, we take variations of Eq. (8.55) to obtain

$$\left[\mathrm{Id} - \omega^2 R_\chi(\omega) M\right] U' = 2\omega\omega' R_\chi(\omega) M U + \omega^2 R_\chi'(\omega)\omega' M U + \omega^2 R_\chi(\omega) M' U.$$

We now take the inner product with a general $V \in L^2(\mathbb{R}^d)$ to obtain

$$\langle V, \left[\mathrm{Id} - \omega^2 R_\chi(\omega) M\right] U' \rangle = \omega' \langle V, \left(2\omega R_\chi(\omega) + \omega^2 R_\chi'(\omega)\right) M U \rangle + \omega^2 \langle V, R_\chi(\omega) M' U \rangle. \tag{8.56}$$

Using the adjoint operator $R_\chi^*(\omega) = R_\chi(\overline{\omega})$, the left hand side of Eq. (8.56) can be rewritten

$$\langle V, \left[\mathrm{Id} - \omega^2 R_\chi(\omega) M\right] U' \rangle = \langle \left[\mathrm{Id} - \omega^2 M R_\chi(\overline{\omega})\right] V, U' \rangle.$$

Setting $V = \chi^{-1}(-\Delta - \overline{\omega}^2)\chi^{-1}\overline{U}$, we find that

$$\left[\mathrm{Id} - \omega^2 M R_\chi(\overline{\omega})\right] V = \chi^{-1}(-\Delta - \overline{\omega}^2)\chi^{-1} \left[\mathrm{Id} - \omega^2 R_\chi(\overline{\omega}) M\right] \overline{U} = 0.$$

Thus for this choice of V, the right hand side of Eq. (8.56) vanishes and we use Eq. (8.54) to obtain

$$\begin{aligned}
\omega' &= -\frac{\omega^2 \langle \chi^{-1}(-\Delta - \overline{\omega}^2)\chi^{-1}\overline{U}, R_\chi(\omega) M' U \rangle}{\langle \chi^{-1}(-\Delta - \overline{\omega}^2)\chi^{-1}\overline{U}, \left(2\omega R_\chi(\omega) + \omega^2 R_\chi'(\omega)\right) M U \rangle} \\
&= -\frac{\omega^2 \langle \overline{u}, m' u \rangle}{2\omega \langle \overline{u}, m u \rangle + \omega^2 \langle \overline{u}, (-\Delta - \omega^2) R'(\omega) m u \rangle}.
\end{aligned}$$

Using $m' = 2n\delta n$, this can be written

$$\frac{\delta\omega}{\delta n}[\delta n] = \omega' = \langle -2\overline{\omega}^2 \overline{\alpha} n \overline{u}^2, \delta n \rangle.$$

where

$$\alpha^{-1} = 2\omega\langle \overline{u}, mu\rangle + \omega^2\langle \overline{u}, (-\Delta - \omega^2)R'(\omega)mu\rangle.$$

Equation (8.41) now follows.

In one dimension, one may take $\chi = \mathbf{1}_{[0,L]}$ and use the explicit formula for the Green's function given in Eq. (8.33) to obtain Eq. (8.42a). This formula agrees with one given in [Heider *et al.*, 2008]. Eq (8.42b) can be obtained using the identity

$$\omega^2 \int n^2 u^2 = \int u_x^2 - \imath\omega[u^2(0) + u^2(L)].$$

Chapter 9

Emergence of periodic structure from maximizing the lifetime of a Schrödinger bound state coupled to radiation

9.1 Introduction

In many problems of fundamental and applied science a scenario arises where, due to physical laws or engineering design, a state of the system is *metastable*; the state is long-lived but of finite lifetime due to coupling or leakage to an *environment*. In settings as diverse as linear and nonlinear optics, cavity QED, Bose-Einstein condensation (BEC), and quantum computation, one is interested in the manipulation of the lifetime of such metastable states. Our goal in this chapter is to explore the problem of maximizing the lifetime of a metastable state for a class of *ionization problems*. The approach we take is applicable to a wide variety of linear and nonlinear problems. Specific examples where metastable states exist include:

1. the excited state of an atom, *e.g.*, hydrogen, where due to coupling to a photon field, the atom in its excited state spontaneously undergoes a transition to its ground state, after some excited state *lifetime* [Cohen-Tannoudji *et al.*, 1992],

2. ionization of an atom by an external electric field [Costin *et al.*, 2000],

3. an approximate bound state (quasi-mode) of a quantum system, *e.g.*, atom in a cavity or BEC in a magnetic trap, which leaks ("tunnels") out of the cavity and whose wave function decays with advancing time [Pitaveskii and Stringari, 2003],

4. an approximate guided mode of an optical waveguide which, due to scattering, bends in the waveguide, or diffraction, leaks out of the structure, resulting in attenuation of the wave-field within the waveguide with increasing propagation distance [Marcuse, 1974], and

5. a "scatterer" which confines "rays" , but leaks energy to spatial infinity due to their wave nature, *e.g.*, Helmholtz resonator, traps rays between obstacles [Lax and Phillips, 1989].

These examples are representative of a class of extended (infinite spatial domain), yet energy-preserving (*closed*) systems, where the mechanism for energy loss is *scattering loss*, the escape of energy from a compact spatial domain to spatial infinity.

Such systems can often be viewed as two coupled subsystems, one with oscillator-like degrees of freedom characterized by discrete frequencies and the other a wave-field characterized by continuous spectrum. When (artificially) decoupled from the wave-field, the discrete system has infinitely long-lived time-periodic bound states. Coupling leads to energy transfer from the system with oscillator like degrees of freedom to the wave field. In many situations, a (typically approximate) reduced description, which is a closed equation for the oscillator amplitudes, can be derived. This reduction captures the view of the oscillator degrees of freedom as an *open* system with an effective (radiation) damping term. In the problems considered in this chapter, the reduced equation is of the simple form:

$$\imath\partial_t A^\epsilon(t) \;\sim\; \epsilon^2 \;(\; \Lambda - \;\imath\Gamma\;)\, A^\epsilon(t). \tag{9.1}$$

Here, ϵ is a real-valued small parameter, measuring the degree of coupling between oscillator and field degrees of freedom. $A^\epsilon(t)$ denotes the slowly-varying complex envelope amplitude of the perturbed bound state. Λ is a real frequency and $\Gamma > 0$ is an effective damping, governing the rate of transfer of energy from the oscillator to field degrees of freedom.

For example, consider the general linear or nonlinear Schrödinger equation

$$\imath\partial_t\phi^\epsilon \;=\; H_V\phi^\epsilon + \epsilon\, W(t,x,|\phi^\epsilon|)\;\phi^\epsilon, \qquad H_V \equiv -\Delta + V(x). \tag{9.2}$$

Here, $V(x)$ is a real-valued time-independent potential and $W(t,x,|\phi|)$ is a time-dependent potential (parametric forcing), $W=\tilde{\beta}(t,x)$, or nonlinear potential, *e.g.* $W=\pm|\phi|^2$. Equation (9.2) defines an evolution which is unitary in $L^2(\mathbb{R})$.

In this article we focus on the class of one-dimensional *ionization* problems, where

$$W(t,x) \;=\; \cos(\mu t)\;\beta(x)$$

where $\beta(x)$ is a spatially localized and real-valued function and $\mu>0$ is a parametric forcing frequency. Thus, our equation is a parametrically forced Schrödinger equation:

$$\imath\partial_t\phi^\epsilon \;=\; H_V\phi^\epsilon \;+\; \epsilon\;\cos(\mu t)\;\beta(x)\;\phi^\epsilon. \tag{9.3}$$

We focus on the case where the parameter ϵ is real-valued and assumed sufficiently small.

Assumptions for the unperturbed problem, $\epsilon=0$: Initially we assume that the potential $V(x)$, decays sufficiently rapidly as $|x|\to\infty$, although we shall later restrict to potentials with a fixed compact support. Furthermore, we assume that H_V has exactly one eigenvalue $\lambda_V<0$, with corresponding (bound state)eigenfunction, $\psi_V(x)$:

$$H_V\;\psi_V \;=\; \lambda_V\;\psi_V, \quad \|\psi_V\|_2=1. \tag{9.4}$$

Thus, $\phi^0(x,t)=e^{-\imath\lambda_V t}\psi_V(x)$ is a time-periodic and spatially localized solution of the unperturbed linear Schrödinger equation:

$$\imath\partial_t\phi \;=\; H_V\phi$$

We indicate an explicit dependence of λ_V and ψ_V on V, since we shall be varying V.

Fermi's Golden Rule: We cite consequences of the general theory of [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001; Kirr and Weinstein, 2003]. If μ, the forcing frequency, is such that $\lambda_V+\mu>0$, then for initial data, $\phi(x,0)=\psi_V(x)$ (or close to ψ_V), the solution decays to zero as $t\to\infty$. On a time scale of order ϵ^{-2} the decay is controlled by (9.1), *i.e.*,

$$|A(t)| \;\sim\; |A(0)|\,e^{-\epsilon^2\Gamma[V]t}, \quad 0<t<\mathcal{O}(\epsilon^{-2}) \tag{9.5}$$

where $|A(t)|=|\langle\psi_V,\phi^\epsilon(t)\rangle|$ and $\Gamma[V]$ is a positive constant. Thus, we say the bound state has a lifetime of order $\left(\epsilon^2\cdot\Gamma[V]\right)^{-1}$ and the perturbation *ionizes* the bound state.

The emergent damping coefficient, $\Gamma[V]$, is often called *Fermi's Golden Rule* [Visser, 2009], arising in the context of the spontaneous emission problem. However, the notion of effective radiation damping due to coupling of an oscillator to a field has a long history [Lamb, 1900]. In general, $\Gamma[V]$ is a sum of expressions of the form:

$$\left| \langle e_V(\cdot, k_{res}(\lambda_V)), \mathcal{G}_W(\psi_V)\rangle_{L^2(\mathbb{R}^d)} \right|^2 = |t_V(k_{res})|^2 \left| \langle f_V(\cdot, k_{res}(\lambda_V)), \mathcal{G}_W(\psi_V)\rangle_{L^2(\mathbb{R}^d)} \right|^2, \tag{9.6}$$

(see (9.33)) where $\mathcal{G}_W(\psi_V)$ depends on the coupling perturbation W in (9.2) and the unperturbed bound state, ψ_V. Here, $e_V(\cdot, k_{res}) = t_V(k_{res}) f_V(\cdot, k_{res})$ is the *distorted plane wave* (continuum radiation mode) associated with the Schrödinger operator, H_V, at a *resonant frequency* $k_{res} = k_{res}(\lambda_V)$, for which $k_{res}^2 \in \sigma_{cont}(H_V)$. $t_V(k)$ denotes the transmission coefficient and $f_V(x,k)$ a *Jost solution*. In Secs. 9.2 and 9.3 we present an outline of the background theory for scattering and the ionization problem, leading to (9.5), (9.6); see [Soffer and Weinstein, 1998].

We study the problem of maximizing the lifetime of a metastable state, or equivalently, minimizing the scattering loss of a state due to radiation by appropriate deformation of the potential, $V(x)$, within some admissible class, $\mathcal{A}_1(a,b,\mu)$:

$$\min_{V \in \mathcal{A}_1(a,b,\mu)} \Gamma[V]. \tag{9.7}$$

We refer to Eq. (9.7) as the potential design problem (PDP). Our admissible class, $\mathcal{A}_1(a,b,\mu)$, is defined as follows:

Definition 9.1.1. $V \in \mathcal{A}_1(a,b,\mu)$ if

1. V has support contained in the interval $[-a,a]$, *i.e.*, $V \equiv 0$ for $|x| > a$
2. $V \in H^1(\mathbb{R})$ and $\|V\|_{H^1} \leq b$
3. $H_V = -\partial_x^2 + V(x)$ has exactly one negative eigenvalue, λ_V, with corresponding eigenfunction $\psi_V \in L^2(\mathbb{R})$, which satisfies Eq. (9.4) : $H_V\psi_V = \lambda_V\psi_V$, $\|\psi_V\|_2 = 1$.
4. $k_{res} \equiv \sqrt{\lambda_V + \mu} > 0$ (formal coupling to continuous spectrum)

Remark 9.1.1. In restricting to systems with a single bound state, we have sought to formulate a simple optimization problem, which incorporates many essential features of arising in controlling

the coupling of an electromagnetic or quantum state to an environment. For systems with multiple bound states, a theory analogous to that which is summarized in section 3, has been developed [Kirr and Weinstein, 2003]. In this case, the Fermi Golden Rule $V \mapsto \Gamma[V]$ is a matrix valued function. Optimization and control problems can, for example, be formulated in terms of the eigenvalues of $\Gamma[V]$ and studied by the methods of this chapter.

Remark 9.1.2. Based on numerical simulations and analytical calculations with families of potentials, we conjecture that the hypothesis 2., imposing a bound of V, can be significantly weakened.

The idea of controlling the lifetime of states by varying the characteristics of a background potential goes back to the work of E. Purcell [Purcell, 1946; Purcell, 1952], who reasoned that the lifetime of a state can be influenced by manipulating the set of states to which it can couple, and through which it can radiate.

Remark 9.1.3. We discuss the potential design problem where

1. $\beta(x)$ is a fixed function, chosen independently of V, for example, $\beta(x) = \mathbf{1}_{[-2,2]}(x)$
2. $\beta(x) = V(x)$.

Remark 9.1.4. **How does one minimize an expression of the form (9.6)?**
We can think of two ways in which (9.6) can be made small:

Mechanism (A) Find a potential in $\mathcal{A}_1$ for which the first factor in (9.6), $|t_V(k_{res})|^2$ is small, corresponding to low *density of states* near k_{res}^2.
N.B. As proved in Proposition 9.2.6, $|t_V(k)| \geq \mathcal{O}(e^{-Ka})$ for V with support contained in $[-a, a]$.

Mechanism (B) Find a potential in $\mathcal{A}_1$ which may have significant density of states near k_{res}^2 (say $|t_V(k_{res})| \geq 1/2$) but such that the oscillations of $f_V(x, k_{res})$ are tuned to make the matrix element expression (inner product) in (9.6) small due to cancellation in the integral.

Indeed, we find that both mechanisms occur in our optimization study.

Remark 9.1.5. We are interested in the problem of deforming V within an admissible set in such a way as to maximize the lifetime of decaying (metastable) state. Intuitively, there are two physical

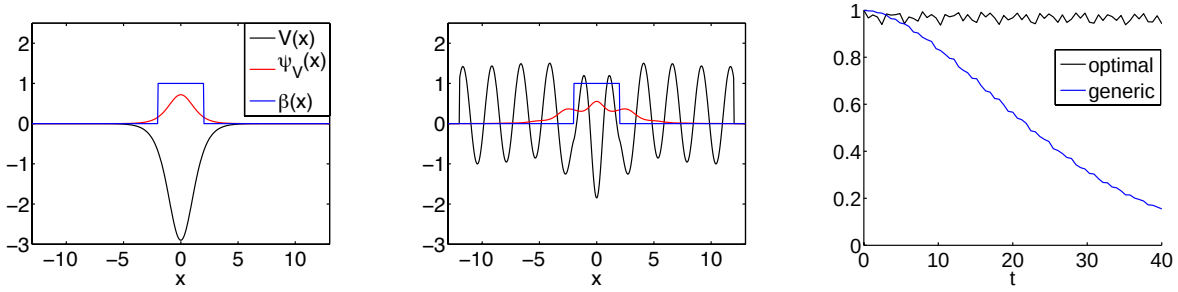


Figure 9.1: Numerical demonstration for Eq. (9.3) ($\epsilon = 1$) of bound state time-decay for a typical potential well (left) and bound state persistence for an optimized potential (center). We plot the potentials V_{init} and V_{opt} (black), corresponding ground states ψ_{init} and ψ_{opt} (red), and forcing function $\beta(x) = \mathbf{1}_{[-2,2]}(x)$ (blue). The rightmost figure displays the time evolution of the projection $|\langle \phi(t,\cdot), \psi_V(\cdot)\rangle|^2$ for each potential. Details are given in Sec. 9.6.6. $\Gamma[V_{init}] = 2.1 \times 10^{-2}$ and $\Gamma[V_{opt}] = 3.3 \times 10^{-9}$.

mechanisms with which one can confine wave-energy in a region: via the depth of the potential (material contrast) and via interference effects. We shall see that our (locally) optimal solutions, of types (A) and (B) find the proper balance of these mechanisms.

9.1.1 Overview of results:

1. In Proposition 9.4.12 we show that the optimal solution to Eq. (9.7) exists, for an admissible set, $\mathcal{A}_1^\delta(a,b,\mu)$, derived from $\mathcal{A}_1(a,b,\mu)$ by regularizing a discrete constraint; see Eqs. (9.42) and (9.53).

2. Fix the admissible set $\mathcal{A}_1^\delta(a,b,\mu)$, *i.e.*, parameters a, b, μ, δ. Select an initial potential, $V_{init} \in \mathcal{A}_1^\delta(a,b,\mu)$. For example, we have chosen a potential of the form $V_{init} = -A\text{sech}(Bx)\ \mathbf{1}_{|x|\le a}$ with the parameters A and B appropriately chosen. We use a quasi-Newton method within $\mathcal{A}_1^\delta(a,b,\mu)$ and, after about 50-100 iterations, find a potential $V_{opt} \in \mathcal{A}_1^\delta(a,b,\mu)$, for which $\Gamma[V]$ achieves a local minimum in $\mathcal{A}_1^\delta(a,b,\mu)$.

3. In a typical search $\Gamma[V_{init}] \approx 10^{-2}$ and $\Gamma[V_{opt}] \approx 10^{-9}$. Therefore, by Theorem 9.3.1 [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001; Kirr and Weinstein, 2003], the decay time for the solution of (9.3) with potential $V = V_{opt}$ and data $\phi^\epsilon(0) = \psi_{V_{opt}}$ is much, much longer

than that for the Schrödinger equation with potential $V = V_{init}$ and data $\phi^\epsilon(0) = \psi_{V_{init}}$. Thus our optimization procedure finds a potential which supports a metastable state which has a very long lifetime, in the presence of parametric forcing coupling to scattering states.

4. As an independent check on the performance of our optimal structures, we numerically solve the initial value problem for the time-dependent Schrödinger equation (9.3) with $\epsilon = 1$ for both $V = V_{init}$ with data $\phi^\epsilon(0) = \psi_{V_{init}}$ and $V = V_{opt}$ with data $\phi^\epsilon(0) = \psi_{V_{opt}}$. Figure 9.1 displays a representative comparison of these numerical experiments, revealing the decay of the bound state for V_{init} and a striking persistence (non-decay) of the bound state for V_{opt}. The details of this simulation are given in Sec. 9.6.6.

5. In Section 9.6.1, we investigate the optimization for classes of potentials with increasing support, *i.e.*, $\mathcal{A}_1^\delta(a, b, \mu)$ for an increasing sequence of a—values: $0 < a_1 < a_2 < \cdots < a_m$. Figure 9.2 shows local optima found in $\mathcal{A}_1^\delta(a_j, b, \mu)$. As a is taken larger, the sequence $V_{opt,a_1}, V_{opt,a_2}, \ldots V_{opt,a_m}$ appears to take on the character of a truncation to the interval $[-a, a]$ of a periodic structure with a localized defect. This suggests the following

   Conjecture 9.1.1. *$\{V_{opt,a}\}$ converges to $V_{opt,\infty}(x) = V_{per}(x) + V_{loc}(x)$, where V_{per} is periodic on $\mathbb{R}$ and $V_{loc}(x)$ is spatially localized.*

6. Our computational methods find locally optimal solutions which have small values of Γ due to either of the mechanisms discussed in Remark 9.1.4 above. In Sec. 9.6.3, we find the confinement properties of potentials, which are optimal due to the cancellation mechanism (mechanism (B)), are very sensitive to perturbations in the forcing frequency away from the forcing frequency, μ, for which the optimization is carried out.

7. In section 9.6.6 we study the stability or robustness of the state, $\psi_{V_{opt}}$, for a locally optimal potential, V_{opt}. Time-dependent simulations of the parametrically forced Schrödinger equation are performed for an un-optimized potential, V_{init}, and initial data $\psi_{V_{init}} + \mathit{noise}$ and for V_{opt}, and initial data $\psi_{V_{opt}} + \mathit{noise}$. Optimal structures effectively filter noise from a ground state, while a generic potential does not. The time scale of bound state radiation damping $\sim \left(\epsilon^2 \Gamma[V_{opt}]\right)^{-1}$ is $\gg$ the time scale for dispersion of noise.

8. Our computations show that the inequality constraints of the (regularized) admissible set, (9.42), are not active at optimal potentials. This is in contrast to studies of other spectral optimization problems, *i.e.*, scattering resonances [Harrell and Svirsky, 1986; Heider *et al.*, 2008] and band gaps [Cox and Dobson, 1999; Cox and Dobson, 2000; Kao *et al.*, 2005; Sigmund and Hougaard, 2008] and other problems [Krein, 1955; Cox and McLaughlin, 1990a; Osher and Santosa, 2001; Dobson and Santosa, 2004; Gondarenko *et al.*, 2006], where periodic patterns attaining material bounds are obtained.

9.1.2 Outline of the article

In section 9.2 we introduce the needed scattering theory background to explain resonant radiative time-decay and Fermi's golden rule, $\Gamma[V]$, which characterizes the lifetime of metastable states. In section 9.3, we summarize the theory of [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001; Kirr and Weinstein, 2003] in the context of the ionization problem (9.3). In section 9.4 we introduce an appropriate regularization, $\mathcal{A}_1^\delta$, of the admissible set of potentials, $\mathcal{A}_1$ (see Definition 9.1.1) which is advantageous for numerical computation, and prove the existence of a minimizer within this class. In section 9.5 we outline the numerical methods used to solve the optimization problem. In section 9.6 we present numerical results for optimal structures and, as an independent check, investigate the effectiveness and robustness of these structures for the time-dependent Schrödinger equation with optimized and un-optimized potentials. Section 9.7 contains discussion and conclusions and Appendix 9.A contains the detailed computations of functional derivatives and gradients used in the optimization.

9.1.3 Notation and conventions

1. $L^2(\mathbb{R})$ inner product: $\langle f, g\rangle = \int_{\mathbb{R}} \overline{f(x)} g(x)\,\mathrm{d}x$

2. $L^2_{\text{comp}}(\mathbb{R})$ is the space of compactly supported $L^2(\mathbb{R})$ functions and $L^2_{\text{loc}}(\mathbb{R})$ is the space of functions which are square-integrable on any compact subset of $\mathbb{R}$.

3. Weighted L^2 space:

$$L^{2,s}(\mathbb{R}) = \{f\colon (1+|x|^2)^{\frac{s}{2}} f \in L^2(\mathbb{R})\}, \qquad s \in \mathbb{R}$$

with norm $\|f\|^2_{L^{2,s}} = \int_{\mathbb{R}} (1+|x|^2)^s f^2\,\mathrm{d}x$

4. Weighted Sobolev space:

$$H^{k,s}(\mathbb{R}) = \{f\colon \partial_x^\alpha f \in L^{2,s}(\mathbb{R}),\, 0 \le \alpha \le k\}, \qquad s \in \mathbb{R}$$

with norm $\|f\|^2_{H^{k,s}} = \|(1+|x|^2)^{\frac{s}{2}} f\|_{H^k}$

5. $\mathcal{B}(X,Y)$ denotes the space of bounded linear operators from X to Y and $\mathcal{B}(X) = \mathcal{B}(X,X)$.

6. Summation notation: $\sum_\pm f_\pm \equiv f_+ + f_-$.

7. The letter C shall denote a generic constant.

9.2 Spectral theory for the one-dimensional Schrödinger operator with compact potential

In this section, we discuss relevant properties of the Schrödinger operator $H_V \equiv -\partial_x^2 + V$ for sufficiently regular and compactly supported potentials, *e.g.* $V \in \mathcal{A}_1(a,b,\mu)$. More general and complete treatments can be found, for example, in [Agmon, 1975; Tang and Zworski,].

9.2.1 The outgoing resolvent operator

Let $0 \neq k \in \mathbb{C}$. For $V = 0$, we introduce the outgoing free resolvent

$$\psi(x) = R_0[f](x,k) = \int_{\mathbb{R}} G_0(x,y,k) f(y)\,\mathrm{d}y, \qquad G_0(x,y,k) \equiv \imath(2k)^{-1}\exp(\imath k|x-y|) \tag{9.8}$$

defined for $f \in L^2_{\mathrm{comp}}(\mathbb{R})$. The function $\psi = R_0(k)f$ satisfies the free Schrödinger equation and outgoing boundary condition

$$(-\partial_x^2 - k^2)\psi = f, \quad \lim_{x\to\pm\infty}(\partial_x \mp \imath k)\psi = 0.$$

For $V \neq 0$ we introduce the outgoing resolvent, $R_V(k) \equiv (H_V - k^2)^{-1}$, satisfying

$$(H_V - k^2)\circ R_V(k) = \mathrm{Id} \tag{9.9}$$

and which, for $\Im k > 0$, is bounded on $L^2(\mathbb{R})$ except for a discrete set of the form, $k_l = \imath\kappa_l,\ \kappa_l > 0$, where $-\kappa_l^2$ are eigenvalues of H_V. We have the identity

$$R_0 = R_0 \circ (H_V - k^2)\circ R_V = (\mathrm{Id} + R_0 V)\circ R_V,$$

or equivalently, the Lipmann-Schwinger equation,

$$R_V = (\mathrm{Id} + R_0 V)^{-1} \circ R_0, \quad \Im k > 0, \quad k \neq \imath\kappa_l \tag{9.10}$$

Proposition 9.2.1. *The following are properties of the resolvent, R_V [Agmon, 1975; Tang and Zworski,].*

1. *The family of operators $R_V(k)\colon L^2_{comp}(\mathbb{R}) \to L^2_{loc}(\mathbb{R})$, given by Eq. (A.17), exists and has a meromorphic extension to $k \in \mathbb{C}$. It has no pole for $k \in \mathbb{R} \setminus \{0\}$.*

2. *For $k \in \mathbb{R}$ and arbitrary $f \in L^2_{comp}$, the function $\psi = R_V(k)f$ satisfies*

$$(H_V - k^2)\psi = f, \qquad \lim_{x\to\pm\infty} (\partial_x \mp \imath k)\psi = 0.$$

We denote by $G_V(x, y, k)$, the Green's function, defined as the kernel of the integral operator $R_V(k)$, in analogy with Eq. (9.8). In the upper half plane, $\Im k > 0$, $G_V(x, y, k)$ has a finite number of simple poles at $k_l = i\kappa_l$, $\kappa_l > 0$. In the lower half plane, $\Im k < 0$, $G_V(x, y, k)$ may have poles at *resonances*, values of k for which the scattering resonance spectral problem:

$$(H_V - k^2)\psi = 0, \qquad \lim_{x\to\pm\infty} (\partial_x \mp \imath k)\psi = 0.$$

has a non-trivial solution.

A consequence of Theorem 4.2 of [Agmon, 1975] is the following:

Theorem 9.2.2 (Limiting absorption principle)**.** *For $\Im k > 0$, the resolvent $R_V(k) = (H_V - k^2)^{-1}$ is a meromorphic function with values in $\mathcal{B}(L^2)$. For $s > \frac{1}{2}$, it can be extended to $\Im k \geq 0$ as an operator on $\mathcal{B}(L^{2,s}, H^{2,-s})$ with limit*

$$R_V(k_0) \;=\; \lim_{\substack{\Im k>0,\\ k\to k_0\in\mathbb{R}}} R_V(k). \tag{9.11}$$

Throughout this chapter, we shall understand $R_V(k_0)$, for $k_0 \in \mathbb{R}$, to be the limit taken in this way.

Since we are interested in how the properties of solutions change with the potential, we make use of the resolvent identity

$$R_V - R_U = R_V(U - V)R_U. \tag{9.12}$$

We now refine Thm. A.1.4 by showing that $R_V(k)\colon L^{2,s} \to H^{2,-s}$ is (locally) Lipschitz continuous with respect to V. To prove this, we shall use the following bounds, used in the proof of Theorem A.1.4 of [Agmon, 1975]:

$$\|R_0(k)\|_{L^{2,s}\to H^{2,-s}} \leq C \tag{9.13a}$$

$$\|(\mathrm{Id} + R_0(k)V)^{-1}\|_{H^{2,-s}\to H^{2,-s}} \leq C(V), \quad \Im k \geq 0, \quad s > \frac{1}{2} \tag{9.13b}$$

and the following

Lemma 9.2.3. *Suppose $f \in L^{2,-s}(\mathbb{R})$ has compact support with $\mathrm{supp}(f) \subset [-a,a]$. Then $f \in L^{2,s}(\mathbb{R})$ and*

$$\|f\|_{L^{2,s}} \leq C(a)a^{2s}\|f\|_{L^{2,-s}}.$$

Proof. $\|f\|^2_{L^{2,s}} \equiv \int f^2(1+|x|^2)^s \,\mathrm{d}x \leq C(a)a^{4s}\int f^2(1+|x|^2)^{-s}\,\mathrm{d}x = C(a)a^{4s}\|f\|^2_{L^{2,-s}}$. □

Proposition 9.2.4. *Fix $a,b,\mu \in \mathbb{R}$, $V \in \mathcal{A}_1(a,b,\mu)$ and for $\rho > 0$ denote by*

$$B^\infty(V,\rho) = \{U \in \mathcal{A}_1(a,b,\mu)\colon \|V-U\|_\infty < \rho\}. \tag{9.14}$$

There exists a $\rho_0 > 0$ such that if $U \in B^\infty(V,\rho_0)$, then for $s > \frac{1}{2}$,

$$\|R_V(k) - R_U(k)\|_{L^{2,s}\to H^{2,-s}} \leq C(V,\rho_0,a)\|V-U\|_\infty \tag{9.15}$$

uniformly for all $k \in \mathbb{R}$.

Proof. Let $f \in L^{2,s}(\mathbb{R})$. Using Eq. (9.12) and Thm. A.1.4, we compute

$$\|(R_V - R_U)f\|_{H^{2,-s}} \leq C(V)\|(U-V)R_U f\|_{L^{2,s}}.$$

Then using Lemma 9.2.3 we have

$$\begin{aligned}\|(U-V)R_U f\|_{L^{2,s}} &\leq C(a)a^{2s}\|U-V\|_\infty\|R_U f\|_{L^{2,-s}} \\ &\leq C(a)a^{2s}\|U-V\|_\infty\|R_U f\|_{H^{2,-s}}\end{aligned}$$

so that

$$\|(R_V - R_U)f\|_{H^{2,-s}} \leq C(V,a)\|U-V\|_\infty\|R_U\|_{L^{2,s}\to H^{2,-s}}\|f\|_{L^{2,s}}.$$

We now claim that there exists a $\rho_0 > 0$ and constants $C(V)$ and $C(V,a)$ such that for $U \in B^\infty(V,\rho_0)$

$$\|R_U\|_{L^{2,s}\to H^{2,-s}} \leq C(V)\frac{1}{1-C(V,a)\rho_0}. \tag{9.16}$$

Equation (9.15) now follows once we have shown Eq. (9.16). To show Eq. (9.16), we use the resolvent identity

$$R_U = \left(\mathrm{Id} + (\mathrm{Id}+R_0V)^{-1}R_0(U-V)\right)^{-1} R_V$$

and Thm. A.1.4 to obtain

$$\|R_U\|_{L^{2,s}\to H^{2,-s}} \leq C(V)\| \left(\mathrm{Id} + (\mathrm{Id}+R_0V)^{-1}R_0(U-V)\right)^{-1} \|_{H^{2,-s}\to H^{2,-s}} \tag{9.17}$$

Using Eqs. (9.13a) and (9.13b) we have

$$\|(\mathrm{Id}+R_0V)^{-1}R_0(U-V)\|_{H^{2,-s}\to H^{2,-s}} \leq C(V,a)\rho_0$$

and Eq. (9.16) follows from using the Neumann series in Eq. (9.17). □

Proposition 9.2.5. *Let $V \in \mathcal{A}_1(a,b,\mu)$, $k \in \mathbb{R}$, $k \neq 0$, $s > \frac{1}{2}$. There exists a $\rho_0 > 0$ such that if $k' \in B(k,\rho_0)$*

$$\|R_0(k) - R_0(k')\|_{L^{2,s}\to H^{2,-s}} \leq C(\rho_0,a)|k-k'| \tag{9.18a}$$

$$\|R_V(k) - R_V(k')\|_{L^{2,s}\to H^{2,-s}} \leq C(\rho_0,V,a)|k-k'| \tag{9.18b}$$

Proof. Eq. (9.18a) follows from Eq. (9.8). To show Eq. (9.18b), we compute

$$\begin{aligned}\|R_V(k) - R_V(k')\|_{L^{2,s}\to H^{2,-s}} \leq& \| \left[(\mathrm{Id}+R_0(k)V)^{-1} - (\mathrm{Id}+R_0(k')V)^{-1}\right] R_0(k)\|_{L^{2,s}\to H^{2,-s}} \\ &+ \|(\mathrm{Id}+R_0(k')V)^{-1}[R_0(k)-R_0(k')]\|_{L^{2,s}\to H^{2,-s}} \\ \leq& C\|(\mathrm{Id}+R_0(k)V)^{-1} - (\mathrm{Id}+R_0(k')V)^{-1}\|_{H^{2,-s}\to H^{2,-s}} \\ &+ C(V)\|[R_0(k)-R_0(k')]\|_{L^{2,s}\to H^{2,-s}}\end{aligned} \tag{9.19}$$

where we used Eq. (9.13). We now use the resolvent identity

$$(\mathrm{Id}+R_0(k)V)^{-1} - (\mathrm{Id}+R_0(k')V)^{-1} = (\mathrm{Id}+R_0(k)V)^{-1}\left[R_0(k)-R_0(k')\right]V(\mathrm{Id}+R_0(k')V)^{-1}$$

and Eq. (9.13) on the first term in Eq. (9.19) to obtain

$$\|(\mathrm{Id}+R_0(k)V)^{-1} - (\mathrm{Id}+R_0(k')V)^{-1}\|_{H^{2,-s}\to H^{2,-s}} \leq C(V)\| \left(R_0(k)-R_0(k')\right) \|_{H^{2,-s}\to H^{2,-s}}.$$

Now applying Eq. (9.18a) to Eq. (9.19) yields Eq. (9.18b) as desired. □

9.2.2 Distorted plane waves, $e_{V\pm}(x;k)$, and Jost solutions, $f_{V\pm}(x;k)$

Distorted plane waves, $e_{V\pm}(x;k)$, are states which explicitly encode the scattering experiment of a plane wave incident on a potential resulting in reflected and transmitted waves. The Jost solutions, $f_{V\pm}(x;k)$, can be thought of as the states to which $e^{\pm ikx}$ deform for nonzero $V(x)$ in the spectral decomposition of H_V. In this section, we introduce these states and give their basic properties.

The continuous spectrum of H_V is $\sigma_c(H_V) = [0,\infty)$. Corresponding to each point $k^2 \in \sigma_c(H_V)$ are two *distorted plane waves* $e_{V\pm}(x,k)$ satisfying

$$H_V e_{V\pm}(x,k) = k^2 e_{V\pm}(x,k) \tag{9.20a}$$

$$\lim_{x\to\pm\infty} (\partial_x \mp \imath k) e_{V\pm}(x,k) = 0. \tag{9.20b}$$

For $V = 0$ these are the plane wave solutions $e_{0\pm}(x,k) = e^{\pm \imath kx}$. For $V \neq 0$, the unique solution to Eq. (9.20) is given by

$$e_{V\pm}(x,k) = e^{\pm \imath kx} - R_V[V e_{0\pm}(\cdot,k)](x,k). \tag{9.21}$$

If V is compactly supported within $[-a,a]$, for $x \notin [-a,a]$, the solutions $e_{V\pm}(x,k)$ are given in terms of the transmission $t_V(k)$ and reflection $r_V(k)$ coefficients

$$e_{V+}(x,k) = \begin{cases} e^{\imath kx} + r_V(k) e^{-\imath kx}, & x < -a \\ t_V(k) e^{\imath kx}, & x > a \end{cases} \tag{9.22}$$

For $k \neq 0$, we have $|r_V(k)|^2 + |t_V(k)|^2 = 1$. If V is a symmetric, then $e_{V-}(x,k) = e_{V+}(-x,k)$.

The following proposition establishes that if V is compactly supported then $|t_V(k)|$ is bounded away from zero, uniformly in k. We shall use this result in the interpretation of our numerical computations in Section 9.6.3.

Proposition 9.2.6. *Suppose* $supp(V) \subset [-a,a]$, $k \neq 0$

$$|t_V(k)| \geq \exp\left(-\min\{1/|k|,\, 2a\} \int_{-a}^{a} |V(s)|\, \mathrm{d}s \right). \tag{9.23}$$

Proof. Consider the integral equation governing $e_{V+}(x,k)$:

$$e_{V+}(x,k) = t_V(k) e^{\imath kx} - \int_x^a \frac{\sin k(x-y)}{k} V(y) e_{V+}(y,k)\, \mathrm{d}y, \;\; x < a.$$

For $x \geq a$, $e_{V+}(x,k) = t_V(k)e^{\imath kx}$. Since $k^{-1}\sin(k(x-y))$ is bounded by $\min\{|k|^{-1}, |x-y|\}$ we have

$$|e_{V+}(x,k)| \leq |t_V(k)| + \int_x^a \min\{|k|^{-1}, |x-y|\}\ |V(y)|\ |e_{V+}(y,k)|\,\mathrm{d}y \tag{9.24}$$

Therefore, by Gronwall's inequality

$$\begin{aligned} |e_{V+}(x,k)| &\leq |t_V(k)|\ \exp\left(\int_x^a \min\{|k|^{-1}, |x-y|\}\ |V(y)|\,\mathrm{d}y\right) \\ &\leq\ |t_V(k)|\ \exp\left(\min\{|k|^{-1}, 2a\}\ \int_{-a}^a |V(y)|\,\mathrm{d}y\right), \quad x < a, \end{aligned} \tag{9.25}$$

and thus

$$|t(k)| \geq |e_{V+}(x,k)| \exp\left(-\ \min\{|k|^{-1}, 2a\}\ \int_{-a}^a |V(y)|\,\mathrm{d}y\right),\ x < a. \tag{9.26}$$

To bound $|e_{V+}(x,k)|$, observe that for fixed $k \neq 0$, we can choose $x^* < -a$ such that $\arg(r_V(k)) = 2kx^*$. Therefore

$$|e_{V+}(x^*,k)| = |e^{\imath kx} + r_V(k)e^{-\imath kx}| = |1 + r_V(k)e^{-2\imath kx^*}| = |1 + |r_V(k)|| \geq 1. \tag{9.27}$$

The bounds (9.26) and (9.27) imply (9.23). □

The following proposition states that we can choose a constant to bound the distorted plane waves for all potentials in a small L^∞-neighborhood of a $V \in \mathcal{A}_1$.

Proposition 9.2.7. *Fix $a, b, \mu \in \mathbb{R}$ and $V \in \mathcal{A}_1(a,b,\mu)$ and let $B^\infty(V,\rho)$ be as in Eq. (9.14). There exists a $\rho_0 > 0$ such that for $U \in B^\infty(V,\rho_0)$ the distorted plane waves $e_{U\pm}(x,k)$ satisfy*

$$\|e_{U\pm}(\cdot,k)\|_{L^\infty([-a,a])} \leq C(a,V,\rho_0).$$

Proof. Using Eq. (9.21), we compute

$$\begin{aligned} \|R_U[Ue^{\imath kx}]\|_{L^\infty([-a,a])} &\leq C(a)\|(1+|x|^2)^{-s}R_U[Ue^{\imath kx}]\|_{L^\infty} \\ &\leq C(a)\|(1+|x|^2)^{-s}R_U[Ue^{\imath kx}]\|_{H^2} \\ &= C(a)\|R_U[Ue^{\imath kx}]\|_{H^{2,-s}} \\ &\leq C(a,V,\rho_0) \end{aligned}$$

This last line follows from a Proposition 9.2.4. □

Definition 9.2.1. The *Jost solutions*, $f_{V\pm}(x,k)$, associated with the time-independent Schrödinger equation $(H_V - k^2)u = 0$ are defined by

$$e_{V+}(x;k) \;=\; t_V(k)\; f_{V+}(x;k), \qquad e_{V-}(x;k) \;=\; t_V(k)\; f_{V-}(x;k), \tag{9.28}$$

where $f_{V+}(x;k) \sim e^{ikx}$ as $x \to +\infty$ and $f_{V-}(x;k) \sim e^{-ikx}$ as $x \to -\infty$.

By results of [Deift and Trubowitz, 1979], for any $k \in \mathbb{R}$ and any compact subset, C, of $\mathbb{R}$

$$\max_{x\in C} |f_{V\pm}(x;k)| \le K_{k,C} < \infty \tag{9.29}$$

Note also that Propositions 9.2.7 and 9.2.6 imply a bound on $|f_{V\pm}|$ in the case where V has compact support.

9.2.3 Spectral decomposition of the 1D Schrödinger operator

We state the spectral theorem in terms of the distorted plane waves (see *e.g.* [Tang and Zworski,]):

Proposition 9.2.8 (Spectral Decomposition)**.** *Let $e_{V\pm}$ and $f_{V\pm}$ denote the distorted plane waves and Jost solutions given by (9.21) and (9.28). Let λ_j for $j = 1 \dots N$ be the eigenvalues of H_V with corresponding (normalized) eigenfunctions $\psi_j(x)$. Then, $h = P_d h + P_c h$ where P_d and P_c are, respectively, projections onto the discrete and continuous spectral parts of H_V given by*

$$\begin{aligned} P_c h &= \frac{1}{2\pi}\int_0^\infty \left[\, (e_{V+}(\cdot,k),h)\, e_{V+}(x,k) + (e_{V-}(\cdot,k),h)\, e_{V-}(x,k) \,\right] \,\mathrm{d}k \\ &= \frac{1}{2\pi}\int_0^\infty \left[\, (f_{V+}(\cdot,k),h)\, f_{V+}(x,k) + (f_{V-}(\cdot,k),h)\, f_{V-}(x,k) \,\right] \,|t_V(k)|^2 \,\mathrm{d}k \\ P_d h &= \sum_{j=1}^N \lambda_j (\psi_j,\, h)\psi_j(x) \end{aligned} \tag{9.30}$$

Moreover,

$$\begin{aligned} g(H_V)h = \frac{1}{2\pi}\int_0^\infty g(k^2) \,\left[\, (f_{V+}(\cdot,k),h)\, f_{V+}(x,k) + (f_{V-}(\cdot,k),h)\, f_{V-}(x,k) \,\right] \,|t_V(k)|^2 \,\mathrm{d}k \\ + \sum_{j=1}^N g(\lambda_j)(\psi_j,\, h)\psi_j(x), \end{aligned} \tag{9.31}$$

where g is any Borel function. Finally, by approximation we have that (A.6) holds with $g(\zeta) = \delta(\zeta)$, the Dirac delta distribution in the distributional sense.

9.3 Radiation damping and Fermi's Golden Rule

In this section, we explain how $\Gamma[V]$, given in Eq. (9.6), emerges as the key quantity controlling the lifetime of the metastable state. We now state a theorem on the ionization and decay of the bound state and then sketch the idea of a proof, which explains the mechanism of decay and (9.5). A detailed proof can be found in [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001; Kirr and Weinstein, 2003]. The following result holds for generic potentials with one bound state. In particular, these hypotheses are satisfied by $V \in \mathcal{A}_1^\delta(a,b,\mu)$.

Theorem 9.3.1. *Consider the parametrically forced Schrödinger equation*

$$\imath\partial_t\phi^\epsilon \;=\; H_V\phi^\epsilon \;+\; \epsilon\,\cos(\mu t)\;\beta(x)\;\phi^\epsilon. \tag{9.32}$$

Assume V and β satisfies the general conditions of [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001]. Consider the initial value problem for Eq. (9.3) with $\phi^\epsilon(x,0)=\phi_0\in L^{2,\sigma}(\mathbb{R})$, where $\sigma\geq 1$. Assume

1. *$k_V^2\equiv\lambda_V+\mu>0$ (resonance with the continuum at $\mathcal{O}(\epsilon^2)$)*
2. *$\Gamma[V]>0$, where $\Gamma[V]$ is the non-negative quantity defined by*

$$\Gamma[V] \equiv \frac{\pi}{4}\;\langle\beta\psi_V,\delta(H_V-k_V^2)P_c\beta\psi\rangle \tag{9.33a}$$

$$= \frac{1}{16\;k_V}\sum_{\pm}|\langle\beta\psi_V,e_{V\pm}(\cdot,k_V)\rangle|^2 \tag{9.33b}$$

$$= \frac{1}{16\;k_V}\;|t_V(k_V)|^2\;\sum_{\pm}|\langle\beta\psi_V,f_{V\pm}(\cdot,k_V)\rangle|^2, \tag{9.33c}$$

where $e_{V\pm}$ and $f_{V\pm}$ denote, respectively, the distorted plane wave and Jost solutions, and $t_V(k)$ denotes the transmission coefficient.

Then, there exists $\epsilon_0>0$ such that for $\epsilon<\epsilon_0$

$$|\langle\psi_V,\phi^\epsilon(\cdot,t)\rangle| \sim |\langle\psi_V,\phi_0\rangle|\,e^{-\epsilon^2\Gamma[V]t}\;+\;\mathcal{O}(\epsilon), \qquad 0\leq t\leq\mathcal{O}(\epsilon^{-2})$$

$$\|\phi^\epsilon(\cdot,t)\|_{L^{2,-\sigma}}\lesssim\; t^{-\frac{1}{2}}\|\phi_0^\epsilon\|_{L^{2,\sigma}}, \qquad t\gg 1.$$

Remark 9.3.1. For certain choices of potentials, the parametrically forced Schrödinger (ionization) problem is exactly solvable by Laplace transform methods and the time-behavior can be computed for *all* $\epsilon>0$. See, for example, [Costin *et al.*, 2000; Costin *et al.*, 2001].

A sketch of the proof. In this sketch, we drop the subscript on ψ_V and superscript on ϕ^ϵ. For small ϵ, it is natural to decompose the solution as

$$\phi(t,x) = a(t)\psi(x) + \phi_c(t,x) \tag{9.34}$$

where $a(t) = \langle \psi, \phi(\cdot,t)\rangle$ and $\phi_c = P_c[\phi]$ is the continuum projection; see (9.30). To simplify the discussion we take as initial data:

$$a(0) = a_0, \qquad \phi_c(0,x) \equiv 0. \tag{9.35}$$

Substitution of (9.34) into (9.32) and projecting onto the discrete and continuous spectral parts of H_V yields the following coupled system:

$$(\imath\partial_t - \lambda)a(t) = \epsilon\cos(\mu t)\langle\psi, \beta\psi\rangle a(t) + \epsilon\cos(\mu t)\langle\psi,\beta\phi_c\rangle \tag{9.36a}$$

$$(\imath\partial_t - H_V)\phi_c = \epsilon\cos(\mu t)P_c[\beta\psi]a(t) + \epsilon\cos(\mu t)P_c[\beta\phi_c]. \tag{9.36b}$$

Since ϵ has been assumed small, the coupling between $a(t)$ and $\phi_c(t,x)$ is weak. We now proceed to make a set of simplifications leading to a minimal model, in which the mechanism of radiation damping is fairly transparent. First, since the first term on the right hand side of (9.36a) contributes an order ϵ mean-zero frequency shift from λ, we neglect it. Second, from equation (9.36b) we formally have that $\phi_c = \mathcal{O}(\epsilon)$. Therefore, the last term on the right hand side of (9.36b) is $\mathcal{O}(\epsilon^2)$ and we therefore neglect it. Finally, the second equation evolves in the continuous spectral part of H_V and we formally replace H_V by $H_0 = -\Delta$.

The resulting system is the following Hamiltonian system of an oscillator of complex amplitude $a(t)$ coupled to a field $\phi_c(t,x)$:

$$(\imath\partial_t - \lambda)a(t) = \epsilon\cos(\mu t)\langle\psi,\beta\phi_c\rangle \tag{9.37a}$$

$$(\imath\partial_t + \Delta)\phi_c = \epsilon\cos(\mu t)\beta\psi a(t). \tag{9.37b}$$

We can exploit a separation of time-scales by extracting the fast phase from $a(t)$ via the substitution

$$a(t) = e^{-\imath\lambda t}A(t),$$

giving the following equation for the slowly varying amplitude, $A(t)$:

$$\imath\partial_t A(t) = \epsilon\cos(\mu t)\langle\psi,\beta\phi_c\rangle e^{\imath\lambda t} \tag{9.38}$$

Now, Duhamel's formula is used to rewrite Eq. (9.37b) as

$$\phi_c(t) = -\imath\epsilon \int_0^t e^{\imath\Delta(t-s)} \cos(\mu s)\beta\psi a(s)\,\mathrm{d}s$$

since $\phi_c(0) = 0$. We insert this back into Eq. (9.38) to obtain the closed equation for $A(t)$.

$$\partial_t A(t) = -\epsilon^2 \cos(\mu t) e^{-\imath\lambda t} A(t) \int_0^t \langle \beta\psi, e^{\imath\Delta(t-s)}\beta\psi\rangle \cos(\mu s) e^{-\imath\lambda s} A(s)\,\mathrm{d}s$$

Writing $\cos(\mu t) = \frac{1}{2}\left(e^{\imath\mu t} + e^{-\imath\mu t}\right)$, we find that if $k_{res}^2 \equiv \lambda + \mu > 0$, then it is a resonant frequency and

$$\begin{aligned}\partial_t A(t) &\approx -\frac{1}{4}\epsilon^2 e^{-\imath k_{res}^2 t} A(t) \int_0^t \langle \beta\psi, e^{\imath\Delta(t-s)}\beta\psi\rangle e^{-\imath k_{res}^2 s} A(s)\,\mathrm{d}s \\ &\approx -\frac{1}{4}\epsilon^2 \langle \beta\psi, (-\Delta - k_{res}^2 - \imath 0)^{-1}\beta\psi\rangle A(t) \end{aligned} \tag{9.39}$$

Here, $(-\Delta - E - \imath 0)^{-1} = \lim_{\delta\downarrow 0}(-\Delta - E^2 - \imath\delta)^{-1}$. The choice of regularization is dictated by the outgoing radiation condition for $t \to +\infty$; see [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001].

Returning to the original (un-approximated) equations (9.36), we have analogously

$$\partial_t A(t) \approx -\frac{1}{4}\epsilon^2 \langle \beta\psi, (H_V - k_{res}^2 - \imath 0)^{-1} P_c[\beta\psi]\rangle A(t) \;\equiv -\epsilon^2(\Lambda + i\Gamma)A(t). \tag{9.40}$$

Remark 9.3.2. In making the approximations appearing in (9.39) and (9.40) we have neglected terms which, for the special class of initial conditions (9.35), corresponding to the bound state of the unperturbed system, are negligible on the time scale of interest, $\mathcal{O}(\epsilon^{-2})$; see [Soffer and Weinstein, 1998; Kirr and Weinstein, 2001; Costin and Soffer, 2001; Kirr and Weinstein, 2003] for details.

The coefficient of $A(t)$ in (9.40) can be computed by applying the functional calculus identity (A.6) to the function $g(s) = (s - k_{res}^2 - \imath\tau)^{-1}$, together with the distributional identity

$$\lim_{\tau\downarrow 0}(s - k_{res}^2 - \imath\tau)^{-1} = P.V.\ (s - k_{res}^2)^{-1} + \imath\pi\ \delta(s - k_{res}^2)$$

and the identification $s \to H_V$. In particular,

$$\begin{aligned}\Gamma[V] &= \frac{1}{4}\cdot\frac{1}{2\pi}\langle\ \beta\psi_V, \delta(H_V - k_{res}^2)\ P_c\beta\psi\ \rangle \\ &= \frac{1}{8\pi}\int_0^\infty \delta(k^2 - k_{res}^2)\ \left[\ |\langle f_{V+}(\cdot,k), \beta\psi_V\rangle|^2 + \ |\langle f_{V-}(\cdot,k), \beta\psi_V\rangle|^2\ \right]\ |t_V(k)|^2\ dk,\end{aligned}$$

from which the expression (9.33) follows after setting $\nu = k^2$ and carrying out the integral.

9.4 A constrained optimization problem: design of a potential to minimize radiative loss

We now consider the Potential Design Problem (PDP) given in Eq. (9.7) with $\Gamma[V]$ defined in Eq. (9.33). We begin by discussing the set of admissible potentials $\mathcal{A}_1(a,b,\mu)$ defined in Def. 9.1.1. For the purpose of numerical computation we regularize the admissible set, $\mathcal{A}_1 \to \mathcal{A}_1^\delta$, by replacing the discrete constraint (H_V has exactly one eigenvalue) by an inequality constraint, in terms of a regularization parameter, δ. We then show that the objective function is locally Lipschitz and that a solution to the PDP exists in the modified admissible set, $\mathcal{A}_1^\delta(a,b,\mu)$.

9.4.1 The admissible set $\mathcal{A}_1$ and its regularization, $\mathcal{A}_1^\delta$

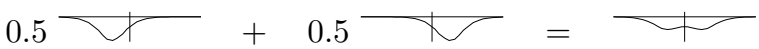
Denote the Wronskian of the distorted plane waves, $e_{V\pm}$, by

$$W_V(k) \equiv \mathrm{Wron}(e_{V+}(\cdot,k), e_{V-}(\cdot,k)), \qquad k \in \mathbb{C}. \tag{9.41}$$

Zeros of $W_V(k)$ correspond to poles of the Green's function $G_V(x,y,k)$ as introduced in Sec. 9.2.1. In particular, the zeros of $W_V(k)$, in the upper half plane, are eigenvalues. The number of eigenvalues is increased or decreased by one, typically through the crossing of a simple zero of $W_V(k)$ through $k=0$ as V varies.[1] Our strategy to fix the number of eigenvalues is then to start with a one bound state potential and deform V, keeping $W_V(0) \neq 0$. However, numerically it is advantageous to replace contraint $W_V(0) \neq 0$ by the inequality constraint $W_V(0)^2 \geq \delta$. For $\delta > 0$, we regularize $\mathcal{A}_1(a,b,\mu)$ by introducing

$$\mathcal{A}_1^\delta(a,b,\mu) \equiv \mathcal{A}_1(a,b,\mu) \cap \{V : W_V(0)^2 \geq \delta\}. \tag{9.42}$$

Remark 9.4.1. Note that the set of admissible potentials $\mathcal{A}_1^\delta(a,b,\mu)$ is not convex. Indeed, counter-examples can be explicitly generated and are illustrated by the following cartoon superposition of potential wells at sufficiently separated points [Harrell, 1980]:



[1] Potentials, V, for which $W_V(k=0)=0$ are called *exceptional*. The value $k=0$, corresponding to edge of the continuous spectrum is then called a zero energy resonance or a half-eigenvalue with half-bound state $e_{V\pm}(x,0)$ [Reed and Simon, 1980].

The two potentials on the left hand side of the equation support a single bound state while the convex combination on the right supports two.

Lemma 9.4.2. *For $\delta > 0$, let $\gamma\colon [0,1] \to \mathcal{A}_1^\delta(a,b,\mu)$ be a smooth function valued path. If $H_{\gamma(0)}$ supports a single bound state then so does $H_{\gamma(t)}$ for $t \in [0,1]$.*

Proof. We show that no bound states are lost along the path $\gamma(t)$. A similar argument shows that no bound states are gained.

Define $f(k,t) \equiv W_{\gamma(t)}(k)$ and consider the equation $f(k,t) = 0$. Let λ_0 denote the eigenvalue of $H_{\gamma(0)}$, *i.e.*, $f(\imath\sqrt{|\lambda_0|},0) = 0$. Since $\partial_k f(k,t)\Big|_{(\imath\sqrt{|\lambda_0|},0)} \neq 0$, *i.e.* λ_0 is a simple eigenvalue, by the implicit function theorem, there exists $T > 0$ such that for $|t| < T$, there is a parameterized family $t \mapsto \lambda_t$ with $\lambda_t < 0$ and $f(\imath\sqrt{|\lambda_t|},t) = 0$. Let

$$t^\# = \sup\{0 \le t \le 1\colon f(\imath\sqrt{|\lambda_t|},t) = 0 \text{ and } \lambda_t < 0\} \;>\; 0.$$

If $\lambda_{t^\#} = 0$, $t^\# < 1$ then $f(0,t^\#) = W_{\gamma(t^\#)}(0) = 0$. This contradicts $\gamma(t) \subset \mathcal{A}_1^\delta(a,b,\mu)$ Therefore, $t^\# = 1$. □

Let $\eta_V^\pm(x) = e_{V\pm}(x,0)$ be the *distorted plane waves* at $k = 0$ which satisfy

$$H_V\eta_V^\pm = 0, \qquad \lim_{x\to\pm\infty}\eta_V^\pm = 1; \tag{9.43}$$

see equation (9.22).

Our gradient-based optimization approach requires that we compute the variation of the Wronskian, $W_V(0)$, with respect to the potential V. This calculation will also be used to establish Lipschitz continuity of $W_V(0)$.

Proposition 9.4.3. *Let $\eta_V^\pm(x)$ satisfy Eq. (9.43). The Fréchet derivative of the Wronskian $W_V(0) = Wron(\eta_V^+,\eta_V^-)\colon L^2_{comp} \to \mathbb{R}$ with respect to the potential is given by*

$$\frac{\delta W_V(0)}{\delta V} = -\eta_V^+\eta_V^-.$$

Proof. See Appendix 9.A. □

Remark 9.4.4. If V is symmetric, then $\frac{\delta W_V(0)}{\delta V}$ is symmetric.

To prove that $W_V(0)$ is locally Lipschitz, we use the following lemma.

Lemma 9.4.5. *Let $f[V]\colon L^2([-a,a]) \to \mathbb{R}$ be a Fréchet differentiable functional with*

$$f[U] = f[V] + \left\langle \frac{\delta f}{\delta V}\Big|_V, U - V \right\rangle + o(\|U - V\|_2).$$

Suppose further that the variation $\frac{\delta f}{\delta V}$ is bounded in an L^∞-neighborhood of V. Then there exists a $\rho_0 > 0$ and a constant $C(\rho_0, V, a)$ such that for $U \in B^\infty(V, \rho_0)$

$$|f[U] - f[V]| \leq C(\rho_0, V, a)\|U - V\|_{L^\infty([-a,a])}$$

Proof. The Mean Value Theorem and Proposition 9.4.3 imply that there exists a $\rho_0 > 0$ such that for every $U \in B^\infty(V, \rho_0)$, there exists a potential $\tilde{V} = tV + (1-t)U$ for some $t \in [0,1]$ such that

$$f[U] = f[V] + \left\langle \frac{\delta f}{\delta V}\Big|_{\tilde{V}}, U - V \right\rangle .$$

This gives the estimate

$$|f[U] - f[V]| \leq \|U - V\|_\infty \left| \int_{-a}^{a} \frac{\delta f}{\delta V}\Big|_{\tilde{V}} \, \mathrm{d}x \right|.$$

The proof is completed by choosing $C(\rho_0, V, a) = \sup_{\tilde{V} \in B^\infty(V,\rho_0)} \left| \int_{-a}^{a} \frac{\delta f}{\delta V}\big|_{\tilde{V}} \, \mathrm{d}x \right|$. □

Proposition 9.4.6 (local Lipschitz continuity of $W_V(0)$)**.** *Fix $a, b, \mu \in \mathbb{R}$, $\delta > 0$, and let $V \in \mathcal{A}_1^\delta(a,b,\mu)$. For $\rho > 0$ denote by*

$$B^\infty(V, \rho) = \{U \in \mathcal{A}_1^\delta(a,b,\mu)\colon \|V - U\|_\infty < \rho\} \tag{9.44}$$

the $L^\infty(\mathbb{R})$ ball around V in $A_1^\delta(a,b,\mu)$. Let $W_V(0)$ be as defined in Eq. (9.41). There exists $\rho_0 > 0$ and a constant $C(\rho_0, V, a)$ such that if $U \in B^\infty(V, \rho_0)$ then

$$|W_U(0) - W_V(0)| \leq C(\rho_0, V, a)\|U - V\|_\infty.$$

Proof. Propositions 9.2.7 and 9.4.3 give that $\frac{\delta W_V}{\delta V} = -\eta_V^+ \eta_V^-$ is pointwise bounded in a neighborhood of V. The result now follows immediately from Lemma 9.4.5. □

9.4.2 Properties of the objective functional, $\Gamma[V]$

In this section, we begin with a formal calculation of the Fréchet derivative of $\Gamma[V]$, given by (9.33). We then show that $\Gamma[V]$ is (locally) Lipschitz with respect to V.

Proposition 9.4.7. *Let $L > a$. The Fréchet derivative of $\Gamma[V]\colon L^2_{comp}([-a,a]) \to \mathbb{R}$ given in Eq. (9.33) with respect to the potential V is given by*

$$\frac{\delta\Gamma}{\delta V} = \frac{\delta\Gamma}{\delta\psi_V}\left[\frac{\delta\psi_V}{\delta V}[\delta V]\right] + \sum_{\pm}\frac{\delta\Gamma}{\delta e_{V\pm}}\left[\frac{\delta e_{V\pm}(\cdot,k_V)}{\delta V}[\delta V]\right] - \frac{\Gamma}{k_V}\langle\frac{\delta k_V}{\delta V},\delta V\rangle \tag{9.45a}$$

$$= -\frac{1}{8k_V}\psi_V R_V(\sqrt{\lambda_V})P_c\left[\Re\sum_{\pm}\langle e_{V\pm},\beta\psi_V\rangle\beta e_{V\pm}\right] - \frac{\Gamma}{2k_V^2}\psi_V^2 \tag{9.45b}$$

$$+\sum_{\pm}\frac{1}{8k_V}\Re\langle e_{V\pm},\beta\psi_V\rangle\left(\frac{1}{2k_V}\langle\beta\psi_V,A_\pm\rangle\psi_V^2 - \overline{e_{V\pm}}R_V(k_V)\left[\beta\psi_V\right]\right)$$

where

$$A_\pm(x) = \pm\imath x e^{\imath k_V x} - R_V(k_V)\left[2k_V\phi_{V\pm} \pm \imath x V e^{\pm\imath k_V x}\right] \tag{9.46}$$

$$+\frac{e^{\imath k_V L}}{2k_V}\left(\phi_{V\pm}(-L)e_{V+}(x,k_V) + \phi_{V\pm}(L)e_{V-}(x,k_V)\right),$$

$e_{V\pm} = e_{V\pm}(\cdot,k_V)$, and $\phi_{V\pm}(x) \equiv e^{\pm\imath k_V x} - e_{V\pm}(x,k_V)$ satisfies Eq. (9.61).

Proof. Equation (9.45a) is obtained by the chain rule. A detailed computation of each term is given in Appendix 9.A. □

Remark 9.4.8. If the potential and β are symmetric, so is $\frac{\delta\Gamma}{\delta V}$.

Proposition 9.4.9 (local Lipschitz continuity of Γ)**.** *Fix $a,b,\mu \in \mathbb{R}$, $\delta > 0$ and $V \in \mathcal{A}_1^\delta(a,b,\mu)$ and define $B^\infty(V,\rho)$ as in Eq. (9.44). There exists $\rho_0 > 0$ and a constant $C(V,\rho_0,a)$ such that if $U \in B^\infty(V,\rho)$ then*

$$|\Gamma[U] - \Gamma[V]| \le C(V,\rho_0,a)\|U - V\|_\infty.$$

Proof. First we use the triangle inequality to obtain

$$\begin{aligned}
|\Gamma[U] - \Gamma[V]| &= \Big|\sum_{\pm}\frac{1}{16k_U}|\langle\beta\psi_U, e_{U\pm}(\cdot,k_U)\rangle|^2 - \frac{1}{16k_V}|\langle\beta\psi_V, e_{V\pm}(\cdot,k_U)\rangle|^2\Big| \\
&\le \frac{1}{16}\sum_{\pm}\Big|\frac{1}{k_U} - \frac{1}{k_V}\Big|\ \ |\langle\beta\psi_U, e_{U\pm}(\cdot,k_U)\rangle|^2 \\
&\quad + \frac{1}{16k_V}\sum_{\pm}\Big[\ \Big||\langle\beta\psi_U, e_{U\pm}(\cdot,k_U)\rangle|^2 - |\langle\beta\psi_V, e_{U\pm}(\cdot,k_U)\rangle|^2\Big| \\
&\quad + \Big||\langle\beta\psi_V, e_{U\pm}(\cdot,k_U)\rangle|^2 - |\langle\beta\psi_V, e_{V\pm}(\cdot,k_U)\rangle|^2\Big| \\
&\quad + \Big||\langle\beta\psi_V, e_{V\pm}(\cdot,k_U)\rangle|^2 - |\langle\beta\psi_V, e_{V\pm}(\cdot,k_V)\rangle|^2\Big|\ \Big] \\
&\equiv \sum_{\pm} A_\pm + B_\pm + C_\pm + D_\pm
\end{aligned} \tag{9.47}$$

We now treat the terms $A_\pm - D_\pm$ in Eq. (9.47) in turn. We'll repeatedly use the inequality $\big||a|^2 - |b|^2\big| \le |a+b||a-b|$.

A. We compute

$$A_\pm = \frac{|k_V - k_U|}{16 k_V k_U} |\langle \beta\psi_U, e_{U\pm}(\cdot, k_U)\rangle|^2.$$

Recalling $k_V = \sqrt{\lambda_V + \mu}$ and using Eq. (9.58) we have

$$\begin{aligned} |k_V - k_U| &\le \frac{1}{2k_V}|\lambda_V - \lambda_U| + o(|\lambda_V - \lambda_U|) \\ &\le \frac{1}{2k_V}\langle \psi_V^2, |U - V|\rangle + o(\|U - V\|_\infty) \\ &\le \frac{1}{2k_V}\|U - V\|_\infty + o(\|U - V\|_\infty). \end{aligned} \tag{9.48}$$

B. We compute

$$\begin{aligned} B_\pm &= \frac{1}{16k_V}\Big||\langle \beta\psi_U, e_{U\pm}(\cdot, k_U)\rangle|^2 - |\langle \beta\psi_V, e_{U\pm}(\cdot, k_U)\rangle|^2\Big| \\ &\le \frac{1}{16k_V}|\langle \beta(\psi_U + \psi_V), e_{U\pm}(\cdot, k_U)\rangle||\langle \beta(\psi_U - \psi_V), e_{U\pm}(\cdot, k_U)\rangle| \\ &\le \frac{1}{16k_V}|\langle \beta(\psi_U + \psi_V), e_{U\pm}(\cdot, k_U)\rangle|\|e_{U\pm}(\cdot, k_U)\|_{L^2(K)}\|\beta(\psi_U - \psi_V)\|_2. \end{aligned}$$

Now using Eq. (9.59) and Thm. A.1.4 we obtain

$$\begin{aligned} \|\psi_U - \psi_V\|_2 &\le \|R_V(\sqrt{\lambda})P_c[\psi_V(U - V)]\|_2 + o(\|U - V\|)_\infty \\ &\le C(V)\|U - V\|_\infty + o(\|U - V\|)_\infty. \end{aligned}$$

C. We compute

$$\begin{aligned} C_\pm &= \frac{1}{16k_V}\Big||\langle \beta\psi_V, e_{U\pm}(\cdot, k_U)\rangle|^2 - |\langle \beta\psi_V, e_{V\pm}(\cdot, k_U)\rangle|^2\Big| \\ &\le \frac{1}{16k_V}|\langle \beta\psi_V, e_{U\pm}(\cdot, k_U) + e_{V\pm}(\cdot, k_U)\rangle||\langle \beta\psi_V, e_{U\pm}(\cdot, k_U) - e_{V\pm}(\cdot, k_U)\rangle| \end{aligned}$$

Using Eq. (9.21) we have that

$$\begin{aligned} e_{U\pm}(\cdot, k_U) - e_{V\pm}(\cdot, k_U) &= -R_U(k_U)[Ue^{\pm \imath k_U x}] + R_V(k_U)[Ve^{\pm \imath k_U x}] \\ &= R_U(k_U)[(V - U)e^{\pm \imath k_U x}] + (R_V(k_U) - R_U(k_U))[Ve^{\pm \imath k_U x}] \end{aligned}$$

Fact: For $K \subset B(0,r)$ compact we have

$$\begin{aligned}\|f\|_{L^\infty(K)} &= \|(1+|x|^2)^{-\frac{s}{2}} f (1+|x|^2)^{\frac{s}{2}}\|_{L^\infty(K)} \\ &\leq C_r \|(1+|x|^2)^{-\frac{s}{2}} f\|_{L^\infty(\mathbb{R})} \\ &\leq C_r \|(1+|x|^2)^{-\frac{s}{2}} f\|_{H^1(\mathbb{R})} \\ &= C_r \|f\|_{H^{1,-s}(\mathbb{R})}. \end{aligned} \tag{9.49}$$

Thus, by Prop. (9.2.4) and Eq. (9.49) we have

$$\begin{aligned}&\|e_{U\pm}(\cdot,k_U) - e_{V\pm}(\cdot,k_U)\|_{L^\infty(K)} \\ &\leq C_r \|e_{U\pm}(\cdot,k_U) - e_{V\pm}(\cdot,k_U)\|_{H^{1,-s}(\mathbb{R}))} \\ &\leq C_r \left(\|R_V\|_{L^{2,s}\to H^{2,-s}} \|(V-U)e^{\pm \imath k_U x}\|_{L^{2,s}} + \|R_V - R_U\|_{L^{2,s}\to H^{2,-s}} \|V e^{\pm \imath k_U x}\|_{L^{2,s}} \right) \\ &\leq C_r \left(\|R_V\|_{L^{2,s}\to H^{2,-s}} \|e^{\pm \imath k_U x}\|_{L^{2,s}(K)} + C(V,\rho_0) \|V e^{\pm \imath k_U x}\|_{L^{2,s}(K)} \right) \|V-U\|_\infty \end{aligned}$$

D. We compute

$$\begin{aligned} D_\pm &= \frac{1}{16 k_V} \left| |\langle \beta \psi_V, e_{V\pm}(\cdot,k_U)\rangle|^2 - |\langle \beta\psi_V, e_{V\pm}(\cdot,k_V)\rangle|^2 \right| \\ &\leq \frac{1}{16 k_V} |\langle \beta\psi_V, e_{V\pm}(\cdot,k_U) + e_{V\pm}(\cdot,k_V)\rangle| |\langle \beta\psi_V, e_{V\pm}(\cdot,k_U) - e_{V\pm}(\cdot,k_V)\rangle| \end{aligned}$$

But,

$$\begin{aligned} e_{V\pm}(\cdot,k_U) - e_{V\pm}(\cdot,k_V) =& e^{\pm \imath k_V x} \left(e^{\pm \imath (k_U - k_V)x} - 1 \right) - R_V(k_U) \left[V e^{\pm \imath k_V x} \left(e^{\pm \imath (k_U - k_V)x} - 1 \right) \right] \\ &+ (R_V(k_V) - R_V(k_U)) \left[V e^{\pm \imath k_V x} \right] \end{aligned} \tag{9.50}$$

Insertion of (9.50) into the above bound on $D_\pm$ and use of Prop. 9.2.5, (9.48) and Proposition 9.2.7 implies $|D_\pm| \leq C\|V-U\|_\infty$.

Proposition 9.4.9 now follows from assembling the estimates **A**-**D**. □

Remark 9.4.10. Numerical investigations for a sample potential indicate that $\Gamma[V]$ is not locally convex.

Remark 9.4.11. As mentioned in Remark 9.1.3, we shall consider optimization problems where (i) $\beta(x)$ is a fixed specified function and where (ii) $\beta(x) = V(x)$. For type (ii) problems $\Gamma[V]$ is given

by

$$\Gamma[V] = \frac{1}{16k_V} \sum_{\pm} |\langle V\psi_V, e_{V\pm}(\cdot, k_V)\rangle|^2. \tag{9.51}$$

That $\Gamma[V]$ in this case is Lipschitz follows by the same arguments as above. Furthermore, it is Fréchet differentiable and the additional contribution of the Fréchet derivative of $\Gamma[V]$, to be added to the expression in Proposition 9.4.7, is given by:

$$\frac{1}{8k_V}\psi_V(x)\ \Re \sum_{\pm} e_{V\pm}(x, k_V)\langle e_{V\pm}(\cdot, k_V), V\psi_V\rangle.$$

9.4.3 Existence of a minimizer

We show that the potential design problem attains a minimum in the admissible class $\mathcal{A}_1^\delta(a, b, \mu)$. Define

$$\gamma_*^\delta(a, b, \mu) = \inf\{\Gamma[V] \colon V \in \mathcal{A}_1^\delta(a, b, \mu)\} \geq 0. \tag{9.52}$$

Proposition 9.4.12. *There exists $V_* \in \mathcal{A}_1^\delta(a, b, \mu)$ such that $\Gamma[V_*] = \gamma_*^\delta(a, b, \mu)$.*

Proof. Let $\{V_n\} \subset \mathcal{A}_1^\delta(a, b, \mu)$ be a minimizing sequence, *i.e.*, $\lim_{n\uparrow\infty} \Gamma[V_n] = \gamma_*$. Since $\|V_n\|_{H^1(\mathbb{R})} \leq b$, there is a weakly convergent subsequence converging to $V_* \in H^1$. Moreover, the family $\{V_n\}$ is uniformly bounded and equicontinuous on $[-a, a]$. By the Arzelà-Ascoli theorem, there is a subsequence (which we continue to denote by $\{V_n\}$) converging to $V_* \in H^1$ and such that $V_n \to V_*$ uniformly on $[-a, a]$. By Prop. 9.4.6, $W_V(0)$ is continuous with respect to V on $\mathcal{A}_1^\delta(a, b, \mu)$, implying $V_* \in \mathcal{A}_1^\delta(a, b, \mu)$. By Prop. 9.4.9, $\Gamma[V]$ is continuous on $\mathcal{A}_1^\delta(a, b, \mu)$ and

$$\Gamma[V_*] = \lim_{n\uparrow\infty} \Gamma[V_n] = \gamma_* \ .$$

□

Remark 9.4.13. Since $\mathcal{A}_1^\delta$ is not convex (Remark 9.4.1), uniqueness of the minimizer is not guaranteed.

Corollary 9.4.14. *By Remark 9.4.11, a minimizer also exists if we take Γ with $\beta = V$.*

9.5 Numerical solution of the optimization problem

In this section, we discuss a numerical solution of the potential design problem (PDP)

$$\min_{V \in \mathcal{A}_1^\delta(a,b,\mu)} \Gamma[V] \tag{9.53}$$

for fixed a, b, μ, δ, and $\beta(x)$, where $\Gamma[V]$ is given in Eq. (9.33) and $\mathcal{A}_1^\delta(a,b,\mu)$ in Eq. (9.42).

Forward Problem. We refer to the evaluation of $\Gamma[V]$ and $W_V(0)$ for a given $V \in \mathcal{A}_1^\delta(a,b,\mu)$ as the forward problem.

To evaluate the objective function $\Gamma[V]$, first the eigenpair (λ_V, ψ_V) satisfying Eq. (9.4) is computed using a three-point finite difference discretization and the Matlab `eigs` command. Next, the distorted plane waves $e_{V\pm}(x, \sqrt{\mu+\lambda_V})$ are computed using the decomposition as in Eq. (9.60) and then solving Eq. (9.61) using the same discretization. The integrals for $\Gamma[V]$ are then evaluated using the trapezoidal rule.

The evaluation of $W_V(0)$ requires the distorted plane waves at $k = 0$, which are computed using a Crank-Nicholson method. For numerical stability, the Wronskian, which is analytically a constant in x, is computed on a uniform grid and is averaged over the spatial domain. As a check on the discretization, we ensure that the variance of the Wronskian does not exceed a specified tolerance.

Optimization Problem. Local optima of Eq. (9.53) are found using a line-search based L-BFGS quasi-Newton interior-point method [Nocedal and Wright, 2006] as implemented in the Matlab command `fmincon`. We use the *optimize-then-discretize approach*, where gradients are computed as in Proposition 9.4.3 and 9.4.7 and evaluated using the discretized counterparts. The constraints,

$$\lambda + \mu \geq 0 \tag{9.54a}$$

$$W_V(0)^2 \geq \delta \tag{9.54b}$$

$$\|V\|_{H^1} \leq b \tag{9.54c}$$

are enforced using a logarithmic-barrier function. The method terminates when the line search cannot find a sufficient decrease in the objective function.

In the numerical experiments, presented in Section 9.6, we use a computational domain larger than the interval $[-a, a]$ defining the support of the potential V and depending on the magnitude of a, between 1000 and 3000 grid points. The method converges in less than 100 iterations and takes approximately 5-20 minutes using a 2.4 GHz dual processor machine with 2GB memory.

Time-dependent simulations. In Sections 9.6.6 and 9.6.7 we study time evolution for the initial value problem in Eq. (9.3). This is accomplished using the same discretization as above and the time stepping routine for stiff ordinary differential equations implemented in `ode15s` in Matlab. The outgoing boundary conditions are approximated using a large domain with a dissipative term localized at the boundary.

9.6 Results of numerical experiments

In this section, we present the results of many numerical experiments using the methods described in Section 9.5 to study locally optimal solutions of the potential design problem (9.53). The constraints in Eq. (9.53) depend on μ (forcing frequency), a (support width), b (H^1 bound on V), and δ (regularization parameter), while the objective function depends on the choice of spatial perturbation of the potential, $\beta(x)$. For δ sufficiently small and b sufficiently large, we find that in all numerically computed solutions of Eq. (9.53), a local optimum is achieved at an interior point of the constraint set, $\mathcal{A}_1^\delta(a, b, \mu)$, *i.e.*, the constraints (9.54) are not active at the optimal solution. This is in contrast to the structure of optimal solutions of other design problems studied in [Krein, 1955; Cox and McLaughlin, 1990a; Harrell and Svirsky, 1986; Cox and Dobson, 1999; Cox and Dobson, 2000; Osher and Santosa, 2001; Dobson and Santosa, 2004; Kao *et al.*, 2005; Heider *et al.*, 2008] where the optimal potentials always attain the bounds and are referred to as "bang bang" controls.

We conjecture that the H^1 bound on V can be regularized in Proposition 9.4.12 and that the constraints of a compactly supported potential with a finite number of bound states is sufficient for the minimization to be well posed, *i.e.*, there exists $b_0, \delta_0 > 0$ such that for $b \geq b_0$ and $\delta < \delta_0$:

$$\min_{V \in \mathcal{A}_1^\delta(a,b,\mu)} \Gamma[V] \;=\; \min_{V \in \mathcal{A}_1^0(a,\infty,\mu)} \Gamma[V].$$

Thus, we consider potential optimization problems for the two classes of $\beta(x)$ in Remark 9.1.3 and

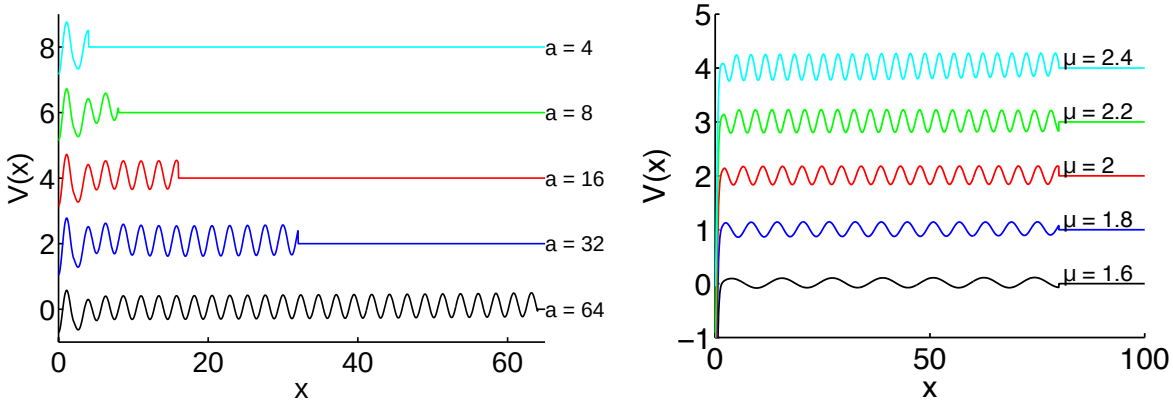


Figure 9.2: Locally optimal potentials for varying values of support a, with fixed frequency $\mu = 2$ (left) and varying forcing frequency μ, with fixed support $a = 80$ (right). The potentials are symmetric in x, only $x \geq 0$ is plotted.

vary μ and a.

9.6.1 Optimal potentials for varying support size, a, with forcing frequency $\mu = 2$ and $\beta(x) = \mathbf{1}_{[-2,2]}(x)$

In Fig. 9.2 (left) we plot locally optimal potentials for 5 different values of the support, a. (The potentials for different values of a are shifted vertically.) The potentials that emerge are symmetric in x (see remarks 9.4.4 and 9.4.8) and periodic on the interval $[-a, a]$ with a defect at the origin. Let V_a^* denote the optimal potential for support parameter a. We note that existing structure changes very little as we increase a. That is, V_a^* and V_b^* are nearly equivalent on the set $[c, c]$ where $c = \min(a, b)$. This is numerical support for Conjecture 9.1.1.

The following table gives the value of $\Gamma[V]$ for the sequence of potentials in Fig. 9.2(left).

a	4	8	16	32	64
Γ	3×10^{-10}	2×10^{-11}	5×10^{-10}	4×10^{-10}	8×10^{-10}

Certainly the infimum $\gamma_*^\delta(ab, b, \mu)$ defined in Eq. (9.52) is monotonically decreasing with increasing support parameter, a. The non-monotonicity of the computed values of Γ is due to the accuracy in which we are able to evaluate Γ (an oscillatory integral) on large domains and the fact that we are only able to compute local minima. The discretization itself may introduce these local minima [Burkardt *et al.*, 2002]. For small a, a relatively small number of grid points is needed for the

accurate computation of Γ and the discrete optimization problem is of relatively small dimension. However, as a increases, the numerical method degrades as we are forced to balance accuracy with the number of optimization variables.

9.6.2 Optimal potentials for varying forcing frequency, μ, with fixed support size, $a = 80$, and $\beta = \mathbf{1}_{[-2,2]}$

In Fig. 9.2(right) we plot locally optimal potentials for 5 different values of forcing frequency μ. The optimal potentials vary smoothly as we change μ with the period of the oscillation in the tails of the potentials decreasing with increasing μ.

The following table gives the value of $\Gamma[V]$ for the potentials in Fig. 9.2(right).

μ	1.6	1.8	2	2.2	2.4
Γ	2×10^{-9}	7×10^{-9}	2×10^{-8}	4×10^{-8}	2×10^{-8}

9.6.3 Two mechanisms for potentials attaining small Γ

The functional to be minimized, $\Gamma[V]$, is given by

$$\Gamma[V] = \frac{1}{16\ k_V}\ |t_V(k_V)|^2\ \sum_{\pm} |\langle \beta\psi_V, f_{V\pm}(\cdot, k_V)\rangle|^2;$$

see (9.33). As discussed in the introduction, two possible mechanisms can be used to decrease the values of $\Gamma[V]$; see Remark 9.1.4:

Mechanism (A) Find a potential in $\mathcal{A}_1$ for which the first factor in (9.6), $|t_V(k_{res})|^2$ is small, corresponding to low *density of states* near k_{res}^2, or

Mechanism (B) Find a potential in $\mathcal{A}_1$ which may have significant density of states near k_{res}^2 (say $|t_V(k_{res})| \geq 1/2$) but such that the oscillations of $f_V(x, k_{res})$ are tuned to make the matrix element expression (inner product) in (9.6) small due to cancellation in the integral.

In Fig. 9.3 we display the results of numerical simulations illustrating examples of both mechanisms at work. On the left is the potential, $V_{A,opt}(x)$, and diagnostics exhibiting mechanism (A) and on the right we exhibit mechanism (B) for the potential labeled $V_{B,opt}(x)$. For both examples we choose $\beta(x) = \mathbf{1}_{[-2,2]}(x)$.

The potential $V_{A,opt}(x)$ is obtained via optimization on the set $\mathcal{A}_1^\delta$ with $a = 64$ and $\mu_A = 2$ (same as in Fig. 9.2 (left)). The potential $V_{B,opt}(x)$ is obtained via optimization on the set $\mathcal{A}_1^\delta$ with

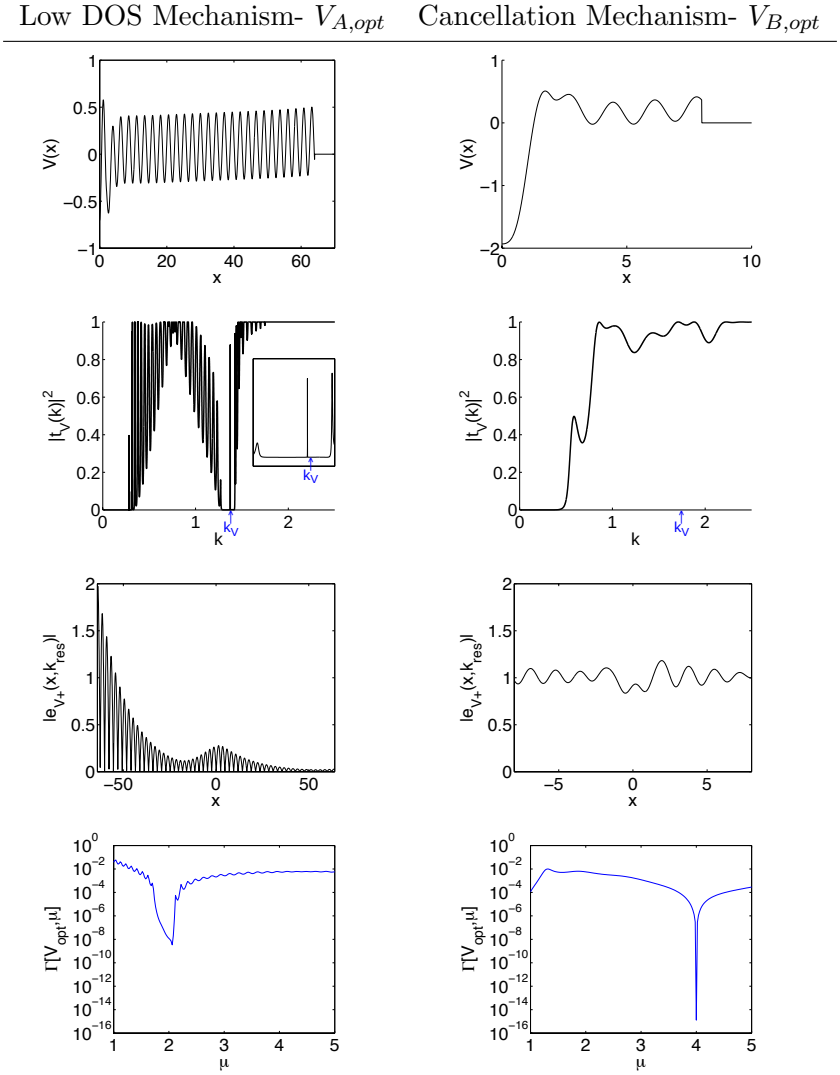


Figure 9.3: Comparison of two locally optimal potentials achieving small $\Gamma[V]$ due the low density of states mechanism (left) and the cancellation mechanism (right). First row: displays plots of potentials. Second row: transmission, $t_V(k)$, a measure of the density of states. At the resonant frequency, k_V, indicated by the arrow, $t_V(k)$ is very small on the left and approximately one on the right. The left figure inset shows that the resonant frequency is distinct from the resonant spike in the "gap". Third row: distorted plane waves, $|e_{V+}(x, k_V)|$. Fourth row: $\Gamma[V_{opt}; \mu]$ vs. μ for $V = V_{opt,A}$ optimized for forcing frequency $\mu_A = 2$ (left), and for $V = V_{opt,B}$ optimized for forcing frequency $\mu_B = 4$ (right). Note contrasting sensitivity to perturbations in μ away from $\mu_{A,B}$.

$a = 8$ and $\mu_B = 4$. The first row of figures displays the potentials as functions of x. The value of Γ for $V_{A,opt}(x)$ and $V_{B,opt}(x)$ are $\Gamma\left[V_{A,opt}\right] = 1.2 \times 10^{-8}$ and $\Gamma\left[V_{B,opt}\right] = 1.3 \times 10^{-15}$.

The second row of plots is of the transmission coefficients $|t_V(k)|^2$ (see Eq. (9.22)) of $V_{A,opt}(x)$ and $V_{B,opt}(x)$. The small vertical arrows along the $k-$ axes indicate the location of the resonant frequency $k = k_V$.

Remark 9.6.1. **Relevance of the transmission coefficient, $t_V(k)$, to the density of states:** Consider a periodic potential, $q(x)$, defined on $\mathbb{R}$. The spectrum of $-\partial_x^2+q(x)$ is equal to the union of closed intervals (bands) separated by open intervals (gaps). Now consider $q_M(x) = q(x)\mathbf{1}_{[-M,M]}(x)$. The decaying potential $q_M(x)$ has continuous spectrum extending from zero to infinity. We expect however the spectral measure, associated with the self-adjoint operator, $-\partial_x^2+q_M(x)$ for $M \gg 1$, to have little mass on those intervals corresponding to the gaps in the spectrum of the limit operator, $-\partial_x^2 + q(x)$. Related to this is the observation that the $t_{q\ \mathbf{1}_{[-M,M]}}(k)$, for $-\partial_x^2 + q(x)\mathbf{1}_{[-M,M]}(x)$, is uniformly small, for k^2 in the spectral gaps of the limit operator, and converge weakly to one for k^2 in the spectral bands; see, for example, [Barra and Gaspard, 1999; Iantchenko, 2006]. Thus, by plotting the amplitude of the transmission coefficient for our optimal potentials we can anticipate whether the density of states is small and a spectral gap is being opened around the resonant frequency, k_V. Thus, if k_V lies in an interval of very low transmission, $t_V(k)$, the Γ, given by (9.33) will be small.

The left plot in the second row shows that the transmission coefficient for $V_{A,opt}(x)$ is very close to zero very near the resonant frequency, $k_{V_{A,opt}}$. On the right we see that for $V_{B,opt}(x)$ the transmission coefficient very near $k_{V_{A,opt}}$ close to one.

In the third row of plots, for each potential, we plot the modulus of the distorted plane wave at the resonant frequency, $|e_{V+}(x, k_V)|$. (Recall that for a symmetric potential, $e_{V-}(x,k) = e_{V+}(-x,k)$.) The modulus of the distorted plane wave associated with $V_{A,opt}(x)$ decays rapidly as it enters the support of the potential, as expected since the transmission coefficient is nearly zero for this frequency (see Eq. (9.22)). The modulus of the distorted plane wave associated with $V_{B,opt}(x)$ is nearly unity over the support of the potential.

In the bottom row of plots of Fig. 9.3 we highlight an additional distinction between these two mechanisms. We fix the optimal potentials, $V_{A,opt}(x)$ and $V_{B,opt}(x)$, respectively optimized for

forcing fixed frequencies μ_A and μ_B. We then consider the variation of the function $\mu \mapsto \Gamma[V_{opt}; \mu]$, where

$$\Gamma[V_{opt};\mu] \equiv \frac{1}{16\sqrt{\lambda_{V_{opt}}+\mu}} \left| t_{V_{opt}}\left(\sqrt{\lambda_{V_{opt}}+\mu}\,\right)\right|^2 \sum_{\pm} \left|\left\langle \beta\psi_{V_{opt}}, f_{V_{opt}\pm}\left(\cdot, \sqrt{\lambda_{V_{opt}}+\mu}\,\right)\right\rangle\right|^2 .$$

Here, μ varies over a range of forcing frequencies above and below μ_A, respectively, μ_B.

We find that for $V_{A,opt}(x)$, the value of Γ is relatively insensitive to small changes in μ near μ_A. Indeed, this is expected. Small variations in μ, imply small variations in $\sqrt{\lambda_V+\mu}$. Therefore, if $k_{V_{A,opt}} = \sqrt{V_{A,opt}+\mu_A}$ is located in a spectral "gap", then for values of μ near μ_A, $k(\mu) \equiv \sqrt{V_{A,opt}+\mu}$ is also in this "gap" . Therefore, $t_{V_{A,opt}}(k(\mu))$ and therefore $\Gamma[V_{A,opt},\mu]$ is small.

In contrast, for $V_{B,opt}(x)$, the range of μ for which $\Gamma[V_{B,opt};\mu]$ remains small is extremely narrow; the smallness of the oscillatory integral, $\Gamma[V_{B,opt};\mu]$, is not preserved over a range of values of μ.

Remark 9.6.2. These observations on the sensitivity of $\Gamma[V_{opt},\mu]$ with respect to the forcing frequency, μ, for the two different kinds of optimizers, $A-$ type and $B-$ type, should have ramifications for applications.

By Proposition 9.2.6,

$$V \in \mathcal{A}_1^{\delta}(a,b,\mu) \implies |t_V(k)| \geq \exp\left(-4a^2 b\right). \tag{9.55}$$

Thus we find that $\Gamma > 0$ due to Mechanism (A). However, in principle, one could find a potential such that due to perfect cancellation, $\Gamma = 0$ by mechanism (B). Indeed, the potential V_B has an extremely small value of Γ.

9.6.4 Further discussion of mechanism (A); potentials which open a gap in the spectrum

We have observed that some locally optimal potentials, *e.g.* the potential associated with the left column of Fig. 9.3, have small values of Γ due to mechanism (A), creating a low density of states at the resonant frequency $k_{V_{opt}}$. We explore this phenomena further here and discuss the relation to Bragg resonance.

For the sequence of potentials given in Fig. 9.2 (left) corresponding to an increasing sequence of values for the support parameter a, we plot in Fig. 9.4 (left) the transmission coefficients (top)

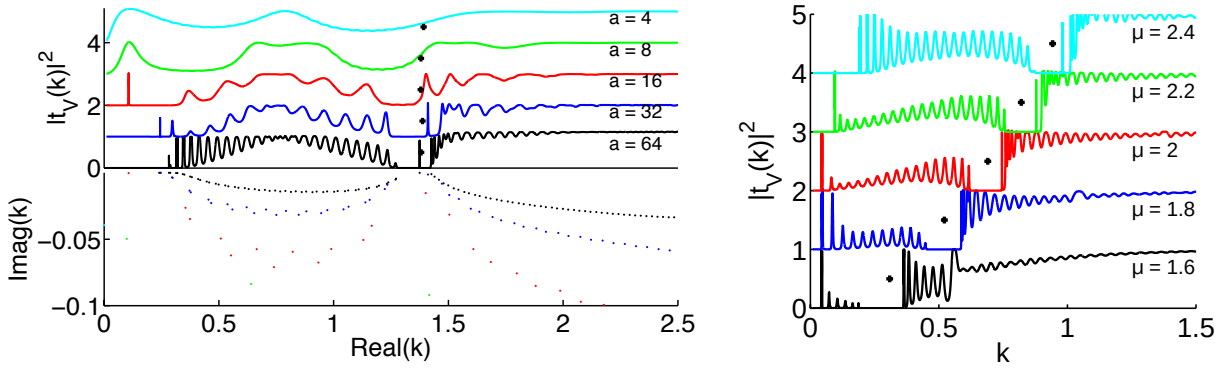


Figure 9.4: (Left) For the sequence of potentials in Fig. 9.2(left), a spectral gap forms as $a \uparrow \infty$. (Right) For each of the potentials in Fig. 9.2(right), the resonant frequency lies in a spectral gap.

and resonances in the lower complex plane (bottom) in corresponding colors. For each potential, the location of the resonant frequency, $k_{V_{opt}}$ is indicated by a black cross (+) in the transmission diagram. The resonances were computed by solving the associated quadratic eigenvalue problem using MatScat [Bindel, 2006].

Remark 9.6.3. As in Remark 9.6.1, let $q(x)$ be a periodic potential and $q_M(x) = q(x)\mathbf{1}_{[-M,M]}(x)$. As $M \uparrow \infty$, the resonances of $-\partial_x^2 + q_M$ will converge to the spectrum of $-\partial_x^2 + q_\infty$ [Barra and Gaspard, 1999; Iantchenko, 2006].

In Fig. 9.4 (left), we see from both the transmission coefficient and the resonances that a gap is opening in the spectrum as $a \uparrow \infty$, supporting Conjecture 9.1.1, that $q_\infty(x)$ is a periodic potential with a localized defect.

Remark 9.6.4. For large support parameter a, a narrow spike forms in the transmission coefficient for a value k within the spectral gap of the limiting operator and a resonance lies nearby. In the limit that $a \uparrow \infty$, this resonance converges to a point eigenvalue within the spectral gap. For periodic potentials with a localized defect, such defect eigenvalues exist [Hoefer and Weinstein, 2010; Figotin and Klein, 1997; Figotin and Klein, 1998; Parzygnat *et al.*, 2010]. Our V_{opt} are qualitatively similar to the class studied in [Hoefer and Weinstein, 2010]. Note that the spike in the transmission coefficient in Fig. 9.4 (left) appears to lie near the resonant frequency, but at a distinct value.

In Fig. 9.4 (right) we plot k vs. the transmission coefficient $|t_V(k)|^2$ for the color-corresponding potentials obtained by varying μ (the forcing frequency for which the optimization is performed)

in Fig. 9.2 (right). For each value of μ, the resonant frequency lies in a spectral gap for each value of μ and there appears only to be a single gap.

Remark 9.6.5. The Schrödinger operator $H_q = -\partial_x^2 + q$ with one specified spectral gap is unique and can be explicitly written in terms of Jacobi elliptic functions [Hochstadt, 1965]. These are called *one-gap potentials*. Using the transmission coefficient plots in Fig 9.4 (right) to estimate the position of the spectral gap, we find that the corresponding one-gap potential has period comparable to that of the periodic tail of the potentials given in Fig. 9.2(right).

This suggests a good heuristic for finding potentials with small values of $\Gamma[V]$: Start with a localized potential well supporting a single bound state. Then, create a low density of states at $k = \sqrt{\lambda_V + \mu}$ by adding a truncated one-gap potential with appropriate spectral gap. If the potential added has small amplitude, then this heuristic is equivalent to adding a cosine or Mathieu potential with frequency given by the Bragg relation.

9.6.5 Optimizing Γ with $\beta = V$ as in Eq. (9.51)

Here we study the case where $\beta = V$ as in Remark 9.1.3, Eq. (9.51), and Corollary 9.4.14.

In Fig 9.5 (left), we take $\mu = 2$ and plot locally optimal potentials for 4 different values of the support, a. The values of Γ are given in the following table.

a	4	8	16	32
Γ	8×10^{-13}	3×10^{-13}	2×10^{-13}	2×10^{-12}

In Fig. 9.5 (right), we take $a = 32$, and plot locally optimal potentials for 4 values of forcing frequency μ. The values of Γ are given in the following table.

μ	2	3	4	5
Γ	2×10^{-12}	1×10^{-12}	2×10^{-13}	2×10^{-15}

As noted in Remark 9.4.1, the solution of the potential design problem is not guaranteed to be unique, since the admissible set is non-convex. Regarding Conjecture 9.1.1 on the character of the limit of optimizers, $V_{opt,a}$ as a tends to infinity, since for $\beta(x) = V(x)$ and $a = \infty$, the functional $V \mapsto \Gamma[V]$ is invariant under the transformation $V(x) \mapsto V(x + x_0)$, we could expect convergence to $V_{opt,\infty}(x)$, a localized perturbation of a periodic potential, only modulo translations.

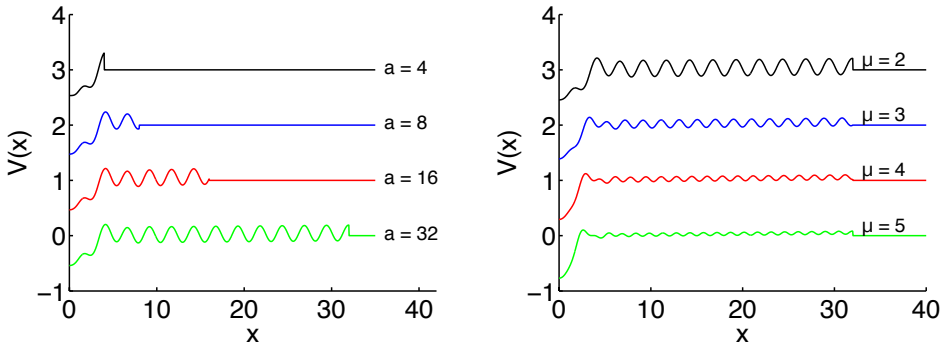


Figure 9.5: With $\beta = V$ as in Eq. (9.51), we plot locally optimal potentials for varying values of support a and forcing frequency μ.

9.6.6 Time dependent simulations

For a locally optimal potential of the potential design problem, (9.53), we independently verify that the potential supports a very long-lived metastable state by conducting time-dependent simulations. See Section 9.5 for a discussion of the numerical method. We set $V_{init} = -A\text{sech}^2(Bx)$ for suitably chosen A, B and take V_{opt} to be a locally optimal solution to the PDP (9.53) with $\beta = \mathbf{1}_{[-2,2]}$, $\mu = 2$, and $a = 12$ (same parameter choice as in Section 9.6.1). We then solve the parametrically forced Schrödinger Eq. (9.3) with $\epsilon = 1$ until $t = 40$ with initial conditions given by the ground state of H_V for the two potentials, *i.e.*, $\phi^\epsilon(0) = \psi_{V_{opt}}$ and $\phi^\epsilon(0) = \psi_{V_{init}}$. In Figs. 9.1 (left) and 9.1 (center) we plot V, β, and ψ_V for the two potentials. In Figure 9.1 (right), we plot t vs. $|\langle \phi^\epsilon(t,\cdot), \psi_V(\cdot)\rangle|^2$, the square modulus of the projection of the wave function onto the bound state for the two potentials.

9.6.7 Filtering study

For the same potentials studied in Sec. 9.6.6 and Fig. 9.1 plus the one studied in Fig. 9.3(right), we conduct the following experiment. We consider the time evolution of Eq. (9.3) until time $t = 50$ with initial condition taken to be $\psi_V +$noise. The noise is taken to be normally distributed random numbers generated using Matlab's `randn` function for each point in the interval $[-a, a]$. The initial condition is then normalized so that $\langle \phi^\epsilon(0), \psi_V \rangle = 1$. The results are plotted in Fig. 9.6. We find that for a non-optimized potential, the final state of the system is nearly zero. While for the locally optimal potential, the bound state emerges as the final state. In the central panel of Figure 9.6 we

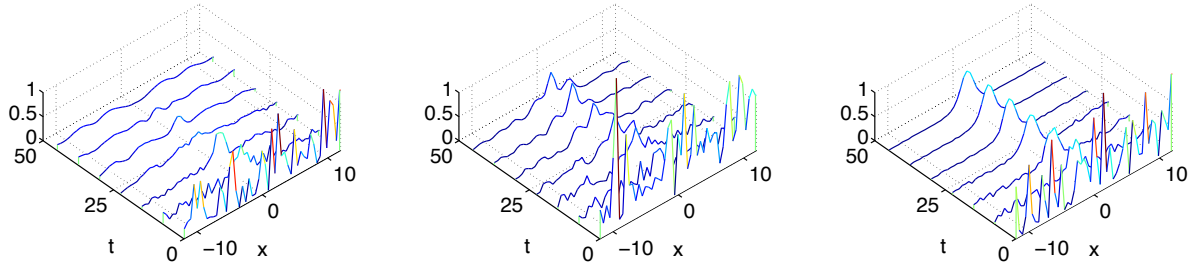


Figure 9.6: For the two potentials in Sec. 9.6.6 and Fig. 9.1, and cancellation potential in Fig. 9.3(right), we plot the time evolution $\phi^\epsilon(t,x)$, governed by Eq. (9.3) with initial condition taken to be ψ_V + noise. The simulation was performed on a spatial domain $[-60,60]$ with absorbing boundary conditions.

see convergence to the projection of the initial condition onto the bound state of H_{V_A}; see central panel of Figure 9.1. In the right panel of Figure 9.6 we see convergence to the projection of the initial condition onto the bound state of H_{V_B}.

This study suggests that such a device could be used as a filter to select a particular spatial mode profile. For these potentials, the system behaves as a mode-selecting waveguide, preserving the discrete components of the initial condition, while radiating the continuous components. Alternatively, this study demonstrates the robustness of $\psi_{V_{opt}}$ to large fluctuations in the data.

9.7 Discussion / Conclusions

Scattering loss, a conservative loss mechanism, is often a limiting factor in the performance of many engineered devices. Therefore, there is great interest in finding structures with low scattering loss-rate. Loss can occur due to parametric or nonlinear time-dependent perturbations which couple an ideally isolated state to an environment. We consider a model of a bound state supported by a potential, V, subject to a time-periodic and spatially localized "ionizing" perturbation. The rate of scattering loss, $\Gamma[V]$, due to coupling of the bound state to radiation modes is given by Fermi's Golden Rule, which depends on the potential V. Using gradient-based optimization methods we find locally optimal structures with much longer-lived bound states. These potentials appear to be truncations of smooth periodic structures with localized defects. This approach can be extended to the wide class of problems presented in the introduction.

9.A Computation of gradients / functional derivatives

9.A.1 Proof of Prop. 9.45b, gradient of $\Gamma[V]$

Proof. Here we compute the individual terms in Eq. (9.45a) and then assemble below.

Computation of $\frac{\delta\Gamma}{\delta\psi_V}$ and $\frac{\delta\Gamma}{\delta e_{V\pm}}$. We compute

$$\frac{\delta\Gamma}{\delta\psi_V}[\delta\psi] = \frac{1}{16k_V}\sum_{\pm}\langle\beta\delta\psi, e_{V\pm}\rangle\overline{\langle\beta\psi_V, e_{V\pm}\rangle} + \text{c.c.} \tag{9.56a}$$

$$= \frac{1}{8k_V}\Re\sum_{\pm}\langle\beta\psi_V, e_{V\pm}\rangle\langle\beta e_{V\pm}, \delta\psi\rangle. \tag{9.56b}$$

Similarly,

$$\frac{\delta\Gamma}{\delta e_{V\pm}}[\delta e_V] = \frac{1}{8k_V}\Re\langle e_{V\pm}, \beta\psi_V\rangle\langle\beta\psi_V, \delta e_V\rangle. \tag{9.57}$$

Computation of $\frac{\delta\lambda_V}{\delta V}$ and $\frac{\delta\psi_V}{\delta V}$. Taking variations of $H_V\psi_V = \lambda_V\psi_V$ we find that

$$(H_V - \lambda_V)\delta\psi_V = -(\delta V\psi_V - \delta\lambda_V\psi_V).$$

Multiplying by ψ and integrating, we obtain $\delta\lambda_V = \langle\psi_V, \delta V\psi_V\rangle$, *i.e.*,

$$\frac{\delta\lambda_V}{\delta V} = \psi_V^2 \tag{9.58}$$

and

$$\begin{aligned}(H_V - \lambda_V)\delta\psi_V &= -(\delta V\psi_V - \langle\psi_V, \delta V\psi_V\rangle\psi_V)\\ &\equiv -P_{\lambda_V}^{\perp}\left[\psi_V\delta V\right].\end{aligned}$$

where $P_{\lambda_V}^{\perp}$ is the orthogonal projection onto the space spanned by ψ_V. Since H_V supports only a single bound state, $P_{\lambda_V}^{\perp} = P_c$. The solution of this equation can be written in terms of the resolvent operator

$$\frac{\delta\psi_V}{\delta V}[\delta V] = \delta\psi_V = -R_V(\sqrt{\lambda_V})P_c[\psi_V\delta V] \tag{9.59}$$

Computation of $\frac{\delta e_{V\pm}}{\delta V}$ and $\frac{\delta k_V}{\delta V}$. We note that the distorted plane waves $e_{V\pm}$ can be expressed

$$e_{V\pm}(x,k_V) = e^{\pm \imath k_V x} - \phi_{V\pm}(x,k_V) \tag{9.60}$$

where $\phi_{V\pm}$ satisfies the following equation with outgoing boundary conditions

$$(H_V - k_V^2)\phi_{V\pm} = Ve^{\pm \imath k_V x} \qquad x \in \Omega = [-L,L] \tag{9.61a}$$

$$\nabla \phi_{V\pm} \cdot \hat{\mathbf{n}} = \imath k_V \phi_{V\pm} \qquad x \in \partial\Omega. \tag{9.61b}$$

Taking variations of Eq. (9.61), we obtain

$$(H_V - k_V^2)\delta\phi_{V\pm} = \delta V e_{V\pm} + \left(2k_V\phi_{V\pm} \pm \imath x V e^{\pm \imath k_V x}\right)\delta k_V \tag{9.62a}$$

$$\nabla \delta\phi_{V\pm} \cdot \hat{\mathbf{n}} - \imath k_V \delta\phi_{V\pm} = \imath \delta k_V \phi_{V\pm}. \tag{9.62b}$$

Recalling $k_V^2 = \lambda_V + \mu$ and Eq. (9.58) we find that $\delta k_V = \frac{\delta k_V}{\delta V}[\delta V] = \langle \frac{\psi_V^2}{2k_V}, \delta V\rangle$ or equivalently

$$\frac{\delta k_V}{\delta V} = \frac{\psi_V^2}{2k_V}. \tag{9.63}$$

Equation (9.62) is a forced equation for $\delta\phi_{V\pm}$ with a unique solution since there is no nontrivial, outgoing solution to the homogenous equation [Tang and Zworski,]. The general solution of Eq. (9.62a) is

$$\frac{\delta\phi_{V\pm}}{\delta V}[\delta V] = \alpha_\pm e_{V+} + \beta_\pm e_{V-} + R_V(k_V)\left[\delta V e_{V\pm} + \left(2k_V\phi_{V\pm} \pm \imath x V e^{\pm \imath k_V x}\right)\delta k_V\right]$$

where α, β are constants. Matching boundary conditions in Eq. (9.62b) and recalling that R_V is the outgoing resolvent, we obtain

$$\alpha_\pm = -\frac{\delta k_V}{2k_V}\phi_{V\pm}(-L)e^{\imath k_V L}$$
$$\beta_\pm = -\frac{\delta k_V}{2k_V}\phi_{V\pm}(L)e^{\imath k_V L}$$

so that

$$\frac{\delta\phi_{V\pm}}{\delta V}[\delta V] = \left(R_V(k_V)\left[2k_V\phi_{V\pm} \pm \imath x V e^{\pm \imath k_V x}\right] - \frac{e^{\imath k_V L}}{2k_V}\left(\phi_\pm(-L)e_{V+} + \phi_\pm(L)e_{V-}\right)\right)\frac{\delta k_V}{\delta V}[\delta V]$$
$$+ R_V(k_V)[\delta V e_{V\pm}]. \tag{9.64}$$

Now using Eq. (9.60) we find that

$$\delta e_{V\pm} = \pm \imath x \delta k_V e^{\pm \imath k_V x} - \delta\phi_{V\pm}. \tag{9.65}$$

Computation of Terms in Eq. (9.45a). Using Eqs. (9.56) and (9.59) we obtain for the first term in Eq. (9.45a)

$$\frac{\delta\Gamma}{\delta\psi_V}\left[\frac{\delta\psi_V}{\delta V}[\delta V]\right] = \langle -\frac{1}{8k_V}\psi_V R_V(\sqrt{\lambda_V})P_c\left[\Re\sum_{\pm}\langle e_{V\pm},\beta\psi_V\rangle\beta e_{V\pm}\right],\delta V\rangle \tag{9.66}$$

where we have used the fact that the operator $R_V(\sqrt{\lambda})P_c\colon L^2\to L^2$ is symmetric.

The second term of Eq. (9.45a) can be written using Eqs. (9.57), (9.63), (9.64), and (9.65)

$$\begin{aligned}\frac{\delta\Gamma}{\delta e_{V\pm}}\left[\frac{\delta e_{V\pm}(\cdot,k_V)}{\delta V}[\delta V]\right] &= \frac{1}{8k_V}\Re\langle\beta\psi_V,e_{V\pm}\rangle\langle\beta\psi_V,A_\pm\delta k_V - R_V(k_V)\left[\delta V e_{V\pm}\right]\rangle \qquad (9.67a)\\ &= \frac{1}{8k_V}\Re\langle\beta\psi_V,e_{V\pm}\rangle\left(\langle\beta\psi_V,A_\pm\rangle\langle\frac{\delta k_V}{\delta V},\delta V\rangle - \langle\overline{e_{V\pm}}R_V(k_V)\left[\beta\psi_V\right],\delta V\rangle\right)\\ &= \langle\frac{1}{8k_V}\Re\langle\beta\psi_V,e_{V\pm}\rangle\left(\frac{1}{2k_V}\langle\beta\psi_V,A_\pm\rangle\psi_V^2 - \overline{e_{V\pm}}R_V(k_V)\left[\beta\psi_V\right]\right),\delta V\rangle\end{aligned}$$

where $A_\pm$ is given in Eq. (9.46) and we have again used the fact that R_V is a symmetric operator.

Using Eq. (9.63), the third term of Eq. (9.45a) is given by

$$-\frac{\Gamma}{k_V}\langle\frac{\delta k_V}{\delta V},\delta V\rangle = -\frac{\Gamma}{2k_V^2}\langle\psi_V^2,\delta V\rangle \tag{9.68}$$

From Eqs. (9.45a), (9.66), (9.67), and (9.68) and the Riesz representation theorem we obtain Eq. (9.45b) as desired. □

9.A.2 Proof of Prop. 9.4.3, gradient of $W_V(0)$

Proof. Denoting $\dot f(x) \equiv \frac{\delta f(x)}{\delta V}[\delta V(y)]$, we fix x and compute

$$\dot W(x) = \dot\eta_+\eta_-' + \eta_+\dot\eta_-' - \dot\eta_+'\eta_- - \eta_+'\dot\eta_-. \tag{9.69}$$

To compute $\dot\eta_\pm$, we take variations of Eq. (9.43) to obtain

$$H_V\delta\eta_\pm = -\delta V\eta_\pm$$
$$\lim_{x\to\pm\infty}\partial_x\delta\eta_\pm = 0.$$

Using the variation of parameters formula, we find

$$\dot\eta_\pm(x) \equiv \frac{\delta\eta_\pm(x)}{\delta V}[\delta V] = -\int q(x,y)\delta V(y)\eta_\pm(y)\,\mathrm{d}y$$

where

$$q(x,y) = \frac{1}{W}\begin{cases}\eta_-(x)\eta_+(y) & x \le y \\ \eta_+(x)\eta_-(y) & x \ge y\end{cases}$$

is the Green's function. Differentiating we find

$$\dot\eta'_\pm(x) = -\int \partial_x q(x,y)\delta V(y)\eta_\pm(y)\,\mathrm{d}y.$$

We now break Eq. (9.69) into 2 parts: $\dot W = \dot W_1 + \dot W_2$ where $\dot W_1 = \int_{-\infty}^x \star\,\mathrm{d}y$, $\dot W_2 = \int_x^\infty \star\,\mathrm{d}y$, and the integrand is given by

$$\star = -\delta V(y)\left(q(x,y)\eta_+(y)\eta'_-(y) + \eta_+(x)\partial_x q(x,y)\eta_-(y) - \partial_x q(x,y)\eta_+(y)\eta_-(x) + \eta'_+(x)q(x,y)\eta_-(y)\right).$$

We then evaluate

$$\begin{aligned}\dot W_1 &= -\frac{1}{W}\int_{-\infty}^x \Big(\eta_+(x)\eta_-(y)\eta_+(y)\eta'_-(x) + \underline{\eta_+(x)\eta'_+(x)\eta_-(y)\eta_-(y)} \\ &\qquad - \eta'_+(x)\eta_-(y)\eta_+(y)\eta_-(y) - \underline{\eta'_+(x)\eta_+(x)\eta_-(y)\eta_-(y)}\Big)\delta V(y)\,\mathrm{d}y \\ &= -\int_{-\infty}^x \eta_+(y)\eta_-(y)\delta V(y)\,\mathrm{d}y\end{aligned}$$

where the underlined terms cancel and

$$\begin{aligned}\dot W_2 &= -\frac{1}{W}\int_x^\infty \Big(\underline{\eta_-(x)\eta_+(y)\eta_+(y)\eta'_-(x)} + \eta_+(x)\eta'_-(x)\eta_+(y)\eta_-(y) \\ &\qquad - \underline{\eta'_-(x)\eta_+(y)\eta_+(y)\eta_-(x)} - \eta'_+(x)\eta_-(x)\eta_+(y)\eta_-(y)\Big)\delta V(y)\,\mathrm{d}y \\ &= -\int_x^\infty \eta_+(y)\eta_-(y)\delta V(y)\,\mathrm{d}y.\end{aligned}$$

Thus

$$\dot W = -\int \eta_+(y)\eta_-(y)\delta V(y)\,\mathrm{d}y$$

and the result follows. Note that $\dot W(x,y)$ is constant in x as expected. □

Chapter 10

Two-dimensional inductor-capacitor lattice synthesis

10.1 Introduction

We investigate a general class of two-dimensional passive propagation media that can be used for signal processing and filtering. These media consist of two-dimensional (2-D) inductor-capacitor (LC) lattices, an example of which is shown in Fig. 10.1, with spatially varying inductance and capacitance. The lattice is a natural generalization of the one-dimensional transmission line. The 2-D LC lattice was first explored by Léon Brillouin [Brillouin, 1946], who showed its equivalence to 2-D mass-spring lattices used to model crystals.

In this chapter, the input $f_j e^{2\pi\imath\alpha t}$ is applied to node j on the left boundary of the lattice and the steady-state output $g_j e^{2\pi\imath\alpha t}$ is tapped from node j on the right boundary. The choice of inductance $\mathbf{L}$ and capacitance $\mathbf{C}$ vectors defines a transfer function from the inputs to the outputs. If there are m rows in the lattice, then for a fixed basis in $\mathbb{C}^m$, the transfer function can be represented by an $m \times m$ complex matrix, denoted $T = T(\mathbf{L}, \mathbf{C})$. Note that T is a linear transformation from $\mathbf{f}$ to $\mathbf{g}$, but T depends nonlinearly on $\mathbf{L}$ and $\mathbf{C}$.

The central result of this chapter is the derivation and demonstration of an algorithm that accepts as input a desired transfer matrix T_d and produces as output a 2-D LC lattice whose

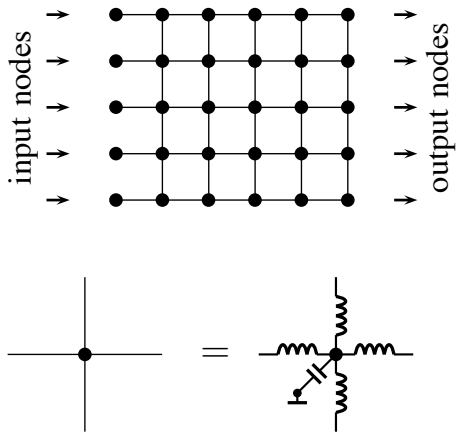


Figure 10.1: (top) A graph which represents a 2-D inductor-capacitor (LC) lattice. (bottom) Each node represents a capacitor connected to ground. Each edge represents an inductor. The capacitances and inductances vary throughout the lattice.

transfer matrix is very close to T_d. We formulate this as the following optimization problem:

$$(\mathbf{L}^*, \mathbf{C}^*) = \arg\min_{(\mathbf{L},\mathbf{C})} \|T(\mathbf{L}, \mathbf{C}) - T_d\|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm. We cannot expect this optimization problem to be solvable for all possible matrices T_d; however, we demonstrate that a large class of transfer matrices can be attained, with the norm difference between the true and desired transfer matrices on the order of 10^{-5}. Our approach to solving the design problem can be generalized to lattice topologies other than the one chosen here.

The general outline of this chapter is as follows. In Section 10.2, the synthesis problem is formulated as an optimization problem. The objective function makes use of the transfer matrix steady-state solution of Kirchhoff's laws on the lattice. The gradient and Hessian of the objective function are calculated analytically in Section 10.3. In Section 10.4, we define design variables that reduce the dimensionality of the problem. In Section 10.5, we present and discuss numerical solutions of the optimization problem formulated in this chapter. We solve the design problem for four different transfer functions: (A) a diagonal transfer matrix, (B) a rank-one projection, (C) a low-pass filter, and (D) a power combiner/funnel. For the low-pass filter, we present results on the robustness of the optimal solution. Finally, we present two results on ill-posedness.

10.1.1 Motivation and Context

The motivation for this work stems from a number of analog devices that operate in the 30-400 GHz range. Earlier work [Afshari *et al.*, 2006a] demonstrated that an inhomogeneous 2-D LC lattice could be used as a power combiner, which was used to implement a power amplifier that generates 125mW at 85 GHz [Afshari *et al.*, 2006b], more than three times the maximum reported power output for an amplifier in the same frequency range on a silicon substrate. Electrical prisms [Momeni and Afshari, 2009], filters that spatially separate the frequency content of an input signal, have been designed using 2-D LC lattices, implemented on chip, tested using 30-50 GHz inputs, and shown to have quality factors from 8 to 12. Simulations show that these filters should scale up to 200-400 GHz. Other work shows that 2-D LC lattices can be used to design a 4-bit quantizer that can process 2×10^{10} samples/sec consuming 194 mW [Tousi and Afshari, 2010], as well as a device that performs discrete Fourier transforms in space [Afshari *et al.*, 2008].

Because the 2-D LC lattice consists only of passive components, it has the desirable properties of high cut-off frequency, low latency, and high throughput, especially as compared with active-device solutions on the same substrate [Afshari *et al.*, 2006a].

This chapter represents a first step towards automatic synthesis of 2-D LC lattices that can be used in high-frequency analog devices. We develop a framework to study and design these lattices, potentially including all applications listed above. Of course, framing the synthesis problem in the language of optimization does not guarantee its solvability. In this chapter, we give computational evidence that, for a large class of desired transfer matrices, the synthesis problem is solvable using gradient-based algorithms.

We now place our problem in the context of problems that have appeared in the literature.

Analog Circuit Design / Device Sizing. The idea of using optimization to synthesize analog circuits has been explored by many authors [Gielen and Rutenbar, 2000; Rutenbar *et al.*, 2007; Martens and Gielen, 2008; Ali, 2009; Zou, 2009; Hägglund, 2006] for a variety of figures of merit. One popular approach wraps an optimization method, either gradient- or stochastic-based, around existing circuit simulation software, such as HSPICE or Spectre. There are several tools that employ this strategy, such as SPICE OPUS [Olenšek *et al.*, 2009] and DELIGHT.SPICE [Nye *et al.*, 1988], which can be used for sizing up to ≈ 100 components. There have also been numerous

efforts to use genetic algorithms and neural networks for analog device synthesis—see, *e.g.*, [Koza *et al.*, 1997] and [Hägglund, 2006, Chap 3.3.3].

Another approach is to mathematically model a circuit and then apply optimization to the model. Examples include [Hershenson *et al.*, 2001], where convex programming is applied to a posynomial model of an op-amp; [Chechurin *et al.*, 2007; Wing, 2008], where Newton and quasi-Newton methods are applied to Kirchhoff's law models of small analog circuits; and [Li *et al.*, 2007], where transistor-level simulations are used to fit quadratic models that are then optimized using geometric programming. For larger problems, hierarchical methods, which build large devices from smaller ones, may be applied [Rutenbar *et al.*, 2007; Ali, 2009; Zou, 2009]; a key step is the use of device-level simulations to extract macromodels that can be used for synthesis.

Our focus in this chapter on the design of 2-D LC lattices has important ramifications for the structure and size of the resulting optimization problem and leads to several differences from the works cited above.

First, the 2-D LC lattice is, by definition, a multiple-input, multiple-output (MIMO) device. A vector of inputs applied to the left boundary is transformed spatially into a vector of outputs at the right boundary. While [Chechurin *et al.*, 2007; Wing, 2008] do use Kirchhoff's law models and gradient-based optimization in much the same way we do, these works synthesize single-input, single-output (SISO) devices. Such devices operate in the time domain, and a typical application is pulse shaping.

One way that the difference between SISO and MIMO design manifests itself is that the optimization problem framed in this chapter involves more degrees of freedom than considered in the above works. For a 2-D LC lattice of size $m \times n$, there are $N = (3m-1)n$ unknown lattice components; note that in Section 10.5.4, we design a 31×31 lattice where $N = 2,852$.

Second, the structure of our optimization problem allows us to fruitfully derive and apply analytical expressions for (i) the solution of the forward problem and (ii) the derivatives of the objective function. Using these analytical expressions in conjunction with quasi-Newton methods is what makes the design problem tractable, especially at large lattice sizes. Other analytical approaches for the forward problem have been explored [Osting and Bhat, 2008; Bhat and Osting, 2009a; Bhat and Osting, 2010] and may, in future works, be applied to the optimization problem as well.

Finally, we seek to synthesize a 2-D LC lattice from scratch, rather than improve upon an existing design, in contrast to some of the above papers and also, *e.g.*, [Hachtel *et al.*, 1973; Brayton *et al.*, 1979; Brayton *et al.*, 1981; Bandler and Chen, 1988].

Inverse Problems. We first mention *transmission line synthesis*: given a finite 1-D LC lattice, an input $f(t)$, and an output $g(t)$, solve for $(\mathbf{L}, \mathbf{C})$ such that when we apply $f(t)$ to one side of the 1-D LC lattice, we obtain $g(t)$ at the other side. This problem was solved 30 years ago using inverse scattering [Caflisch, 1981; Dickinson, 1984; Bruckstein and Kailath, 1987]—here, $f(t)$ and $g(t)$ are prescribed for all t, including both transient and steady-state responses. In contrast, for 2-D LC lattice synthesis, we assume time-harmonic inputs and consider only the steady-state output.

Two-dimensional electromagnetic inverse problems have been considered by numerous authors, *e.g.*, [Colton and Kress, 1998; Frolik and Yagle, 1996; Frolik and Yagle, 1997; Yagle and Frolik, 1996]. These problems are posed on infinite, continuous domains. Far-field scattering data is used to reconstruct unknown parameters $\varepsilon(x, y)$ and/or $\mu(x, y)$, assumed to be inhomogeneous within a compact region. Related work [Angell and Kirsch, 2004; Balanis, 2005; Stutzman and Thiele, 1998] seeks to design electromagnetic devices that either have prescribed radiative behavior in the far field, or that have optimal values of various far-field figures of merit, *e.g.*, directivity, gain, and signal-to-noise ratio. In 2-D LC lattice synthesis, the domain is discrete and finite, and the output signal is collected immediately adjacent to the scattered obstacle, a completely different regime.

Inverse problems on lattices of resistors have been extensively studied by, *e.g.*, [Curtis and Morrow, 2000; Borcea *et al.*, 2010]. Like 2-D LC lattice synthesis, these problems are discrete inverse problems on finite domains. The goal is to reconstruct the conductivity in the interior of the lattice using measurements made using DC sources on the boundary. The resistor lattice is fundamentally different from the LC lattice: the forward problem for a resistor lattice is a discretization of the heat equation, and its steady-state solution is a smooth distribution. For 2-D LC lattices, on the other hand, the forward problem is a discretization of Maxwell's equations for spatially varying ϵ and μ [Bhat and Osting, 2011a], and the steady-state solution is a superposition of standing waves.

10.2 Formulation of the synthesis / design problem

The notation and formulation developed in this section is similar to that in [Bhat and Osting, 2011a], where we discuss the continuum limit of Kirchhoff's laws on a lattice.

We consider a 2-D rectangular LC lattice, as shown in Fig. 10.1, which we represent as an oriented, planar graph, c.f. [Foulds, 1992, Chap. 13]. Nodes represent capacitors and edges represent inductors. The orientation of the edge represents the direction of positive current flow through the associated inductor.

In a lattice of size $m \times n$, there are mn nodes and $(2m-1)n$ edges, mn horizontal ones and $(m-1)n$ vertical ones. Let $\mathfrak{N} = \{1, 2, \ldots, mn\}$ denote the set of all nodes, and $\mathfrak{E} = \{1, 2, \ldots, (2m-1)n\}$ the set of all edges. Let $\mathbf{C}$ be a vector of size mn such that C_j is the capacitance at node j. Let $\mathbf{L}$ be a vector of size $(2m-1)n$ such that L_j is the inductance at edge j. We decompose $\mathbf{L} = [\mathbf{L}_h, \mathbf{L}_v]$ into the horizontal and vertical inductors, respectively. We denote by $V_j(t)$ the voltage across capacitor j and by $I_k(t)$ the current across inductor k at time t. By $\mathbf{V}(t)$ and $\mathbf{I}(t)$ we denote the vectors of all voltages and currents, respectively.

Of the horizontal edges, there are m boundary edges that form a subset $\Gamma \subset \mathfrak{E}$, each of which is incident upon only one node. In Fig. 10.1, Γ is the left-most column of horizontal edges. All other edges in the graph are incident upon two nodes. In general, an edge is an ordered pair (i_1, i_2), where $i_k \in \mathfrak{N}$. The direction of the edge is given by the ordering of these numbers, so that i_1 is the tail and i_2 is the head. For a boundary edge j that is incident only upon node i, we write $j = (\emptyset, i)$.

Let $\mathfrak{B}$ denote the $|\mathfrak{N}| \times |\mathfrak{E}| = mn \times (2m-1)n$ incidence matrix for the oriented graph that represents our circuit. Then

$$\mathfrak{B}_{ij} = \begin{cases} 1 & \text{if } j = (i', i) \text{ for some } i' \in \mathfrak{N} \cup \{\emptyset\} \\ -1 & \text{if } j = (i, i') \text{ for some } i' \in \mathfrak{N} \\ 0 & \text{otherwise.} \end{cases}$$

The matrix $\mathfrak{B}$ will be used shortly to write Kirchhoff's laws in a compact form.

In addition to the structure described already, the 2-D rectangular LC lattice also has resistors and forcing along the boundary. We represent the set of nodes connected to resistors by $\mathfrak{G} \subset \mathfrak{N}$, and let G_i be the conductance of node $i \in \mathfrak{G}$. We then extend G_i by defining $G_i \equiv 0$ for all $i \in \mathfrak{N} \setminus \mathfrak{G}$, so that $\mathbf{G} = (G_1, \ldots, G_{mn})$ is a vector in $\mathbb{R}^{|\mathfrak{N}|}$.

Let $N = |\mathfrak{N}| + |\mathfrak{E}| = (3m-1)n$. Then we define the $|\Gamma| \times N = m \times (3m-1)n$ projection matrix P_Γ by $(P_\Gamma)_{ij} = 1$ if $\Gamma_i = j$ and $(P_\Gamma)_{ij} = 0$ otherwise. Note that because $\Gamma_i \in \mathfrak{E}$, the final mn columns of P_Γ are all zero. The forcing applied at edges Γ is given by $\mathbf{W}(t) = P_\Gamma^t \mathbf{f} e^{2\pi\imath\alpha t}$, where $\mathbf{f} \in \mathbb{C}^{|\Gamma|}$.

Kirchhoff's Laws on this inductor-capacitor lattice can now be written in the following matrix-vector form:

$$\operatorname{diag}(\mathbf{L})\frac{\mathrm{d}\mathbf{I}}{\mathrm{d}t} = -\mathfrak{B}^t\mathbf{V} + \mathbf{W} \tag{10.1a}$$

$$\operatorname{diag}(\mathbf{C})\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = \mathfrak{B}\mathbf{I} - \operatorname{diag}(\mathbf{G})\mathbf{V} \tag{10.1b}$$

Define $\mathbf{z}(t) = (\mathbf{I}(t), \mathbf{V}(t))$ so for each t, $\mathbf{z}(t) \in \mathbb{C}^N$. Define

$$M(\mathbf{G}) = \begin{bmatrix} 0 & -\mathfrak{B}^t \\ \mathfrak{B} & -\operatorname{diag}(\mathbf{G}) \end{bmatrix}.$$

Then the system (10.1) can be written in the form

$$\operatorname{diag}(\mathbf{L}, \mathbf{C})\dot{\mathbf{z}}(t) = M(\mathbf{G})\mathbf{z}(t) + P_\Gamma^t \mathbf{f} e^{2\pi\imath\alpha t}. \tag{10.2}$$

Let $\Upsilon \subset \mathfrak{G}$ denote the vector of right boundary nodes. Let P_Υ be the $|\Upsilon| \times N$ projection matrix defined by $(P_\Upsilon)_{ij} = 1$ if $\Upsilon_i = j$ and $(P_\Upsilon)_{ij} = 0$ otherwise. Note that because $\Upsilon_i \in \mathfrak{N}$, columns 1 to $|\mathfrak{E}| = (2m-1)n$ of P_Υ are all zero.

Forward Problem. Let $\mathbf{z}(t) = \mathbf{u}e^{2\pi\imath\alpha t}$. Then the forward problem is to find $\mathbf{g} = P_\Upsilon \mathbf{u}$ given $\mathbf{f}$, $\mathbf{L}$, $\mathbf{C}$, and $\mathbf{G}$. Using the Fourier transform, one can show that the solution of the forward problem is

$$\mathbf{f} \mapsto \mathbf{g} = P_\Upsilon \Big(2\pi\imath\alpha \operatorname{diag}(\mathbf{L}, \mathbf{C}) - M(\mathbf{G})\Big)^{-1} P_\Gamma^t \mathbf{f}. \tag{10.3}$$

Given $\operatorname{diag}(\mathbf{L}, \mathbf{C}, \mathbf{G})$, we define the *transfer matrix* to be:

$$T(\mathbf{L}, \mathbf{C}, \mathbf{G}) := P_\Upsilon \Big(2\pi\imath\alpha \operatorname{diag}(\mathbf{L}, \mathbf{C}) - M(\mathbf{G})\Big)^{-1} P_\Gamma^t. \tag{10.4}$$

We have formulated the circuit as an oriented graph in order to write the equations compactly and take advantage of the graph-theoretic interpretation of the incidence matrix $\mathfrak{B}$, which appears naturally in Kirchhoff's laws. Though we have formulated the problem for an $m \times n$ rectangular lattice, the beauty of the graph-theoretic framework outlined above is that it easily accommodates other lattice topologies.

Note that since (10.3) is invariant under the transformation

$$\alpha \mapsto \tau\alpha \quad \text{and} \quad (\mathbf{L}, \mathbf{C}) \mapsto \tau^{-1}(\mathbf{L}, \mathbf{C}), \tag{10.5}$$

a lattice with values $(\mathbf{L}, \mathbf{C})$ which performs a transfer function at frequency α can be rescaled by a factor of α'/α to create a lattice that performs the same function at frequency α'.

Design / Synthesis Problem. We define the admissible set

$$\begin{aligned} \mathfrak{A} := \{(\mathbf{L}, \mathbf{C}, \mathbf{G})\colon\ & \underline{L} < L_i < \overline{L} && \text{for all } i \in \mathfrak{E}, \\ & \underline{C} < C_j < \overline{C} && \text{for all } j \in \mathfrak{N}, \text{ and} \\ & \underline{G} < G_j < \overline{G} && \text{for all } j \in \mathfrak{G} \subset \mathfrak{N}\} \end{aligned}$$

where $\underline{L}$, $\overline{L}$, $\underline{C}$, $\overline{C}$, $\underline{G}$, and $\overline{G}$ are constants. Let

$$\{(\mathbf{f}^i, \mathbf{g}^i) \,|\, 1 \le i \le p\}$$

be a collection of desired input-output pairs. The design problem is: find $(\mathbf{L}, \mathbf{C}, \mathbf{G}) \in \mathfrak{A}$ such that for each i, the steady-state output $T\mathbf{f}^i$ generated by input $\mathbf{f}^i$ is equal to $\mathbf{g}^i$. We formulate this as the constrained optimization problem:

$$\min_{(\mathbf{L},\mathbf{C},\mathbf{G})\in\mathfrak{A}} \quad \mathcal{J}(\mathbf{u}^i) := \frac{1}{2}\sum_{i=1}^{p} \left\| P_\Upsilon \mathbf{u}^i - \mathbf{g}^i \right\|^2 \tag{10.6a}$$

$$\text{s.t.} \quad \Big(2\pi\imath\alpha\, \mathrm{diag}(\mathbf{L}, \mathbf{C}) - M(\mathbf{G})\Big)\mathbf{u}^i = P_\Gamma^t \mathbf{f}^i, \quad 1 \le i \le p. \tag{10.6b}$$

It is convenient to set $p = m$ and choose the input basis vectors to be $(f^i)_j = \delta_{ij}$. The *desired transfer matrix* is then

$$T_d = [\mathbf{g}^1 | \mathbf{g}^2 | \cdots | \mathbf{g}^m].$$

We can then write the solution of (10.6b) using (10.4) and rewrite the optimization problem (10.6) in the following compact form:

$$\min_{(\mathbf{L},\mathbf{C},\mathbf{G})\in\mathfrak{A}} \quad \tilde{\mathcal{J}}(\mathbf{L}, \mathbf{C}, \mathbf{G}) := \frac{1}{2}\|T(\mathbf{L}, \mathbf{C}, \mathbf{G}) - T_d\|_F^2. \tag{10.7}$$

As written, the objective function $\mathcal{J}(\mathbf{u}^i)$ in (10.6a) does not depend explicitly on $(\mathbf{L}, \mathbf{C}, \mathbf{G})$, only implicitly through the constraint (10.6b). We use the notation $\tilde{\mathcal{J}}(\mathbf{L}, \mathbf{C}, \mathbf{G}) = \mathcal{J}(\mathbf{u}^i(\mathbf{L}, \mathbf{C}, \mathbf{G}))$ to refer to the composition that explicitly depends on $(\mathbf{L}, \mathbf{C}, \mathbf{G})$.

We consider two different choices of boundary conditions:

(BC1) The resistive boundary $\mathfrak{G}$ consists of all nodes on the top, right, and bottom boundaries of the lattice. For each $i \in \mathfrak{G}$, we prescribe the locally impedance-matched conductance

$$G_i = \sqrt{C_i/L_j}, \tag{10.8}$$

where $j \in \mathfrak{E}$ is the edge incident on node i that is normal to the boundary. This impedance boundary condition can be viewed as a first-order discretization of the Silver-Müller outgoing boundary condition for Maxwell's equations, as described in [Bhat and Osting, 2011a].

(BC2) The resistive boundary $\mathfrak{G}$ consists only of Υ, *i.e.*, the nodes on the right boundary of the lattice. For each $i \in \mathfrak{G}$, we set G_i according to (10.8), as before. Unlike the previous case, $G_i = 0$ along top/bottom boundaries.

Slightly abusing notation, we take $\tilde{\mathcal{J}}(\mathbf{L}, \mathbf{C})$ to be the composition of $\tilde{\mathcal{J}}(\mathbf{L}, \mathbf{C}, \mathbf{G})$ with (10.8). We thus arrive at the following N-dimensional optimization problem:

$$\min_{(\mathbf{L},\mathbf{C})\in\mathfrak{A}} \quad \tilde{\mathcal{J}}(\mathbf{L}, \mathbf{C}) \tag{10.9}$$

where $\mathfrak{A}$ is also modified to reflect (10.8) by letting $\underline{G}_j = 0$ and $\overline{G}_j = \infty$ for all $j \in \mathfrak{G}$. Thus the only constraints in (10.9) are box constraints on the design variables $\mathbf{L}$ and $\mathbf{C}$.

Numerical tests show (10.9) is not convex, which implies that the solution to (10.9) is not guaranteed to be unique.

10.3 Computation of the gradient and Hessian

In this section, we compute the gradient and Hessian of $\tilde{\mathcal{J}}(\epsilon)$ in preparation for quasi-Newton and Newton numerical solutions of the optimization problem (10.6).

10.3.1 Computation of the gradient via the adjoint method

Here we set $\mathbf{s} = (\mathbf{L}, \mathbf{C})$ and $A = 2\pi\imath\alpha\,\mathrm{diag}(\mathbf{s}) - M$. We introduce the dual variables $\mathbf{v}^i \in \mathbb{C}^p$ and the Lagrangian

$$\mathcal{L}(\mathbf{u}^i, \mathbf{v}^i, \mathbf{s}) = \mathcal{J}(\mathbf{u}^i) + \sum_{i=1}^{p} \Re\langle \mathbf{v}^i, A\mathbf{u}^i - P_\Gamma^t \mathbf{f}^i\rangle. \tag{10.10}$$

The state equations (10.6b) are obtained by setting the derivative of (10.10) with respect to $\mathbf{v}^{i*}$ equal to zero. The adjoint equations are obtained by setting the derivative of (10.10) with respect to the state variables $\mathbf{u}^i$ equal to zero:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}^i} = \frac{\partial \mathcal{J}}{\partial \mathbf{u}^i} + \frac{1}{2}\mathbf{v}_i^* A = 0. \tag{10.11}$$

Here we use

$$\frac{\partial \mathcal{J}}{\partial \mathbf{u}^i} = \frac{1}{2}(P_\Upsilon \mathbf{u}^i - \mathbf{g}^i)^* P_\Upsilon. \tag{10.12}$$

The decision equations are obtained by setting the derivative of (10.10) with respect to the design variables $\mathbf{s}$ equal to zero and recalling $\dfrac{\partial \mathcal{J}}{\partial s_k} = 0$ for all k.

$$\frac{\partial \mathcal{L}}{\partial s_k} = \sum_{i=1}^{p} \Re \left\langle \mathbf{v}^i, \frac{\partial A}{\partial s_k}\mathbf{u}^i \right\rangle = 0 \tag{10.13}$$

To compute $\partial A/\partial s_k$, we must compute

$$\frac{\partial \mathrm{diag}(\mathbf{s})_{ij}}{\partial s_k} = \delta_{ij}\delta_{ik} \quad \text{and} \quad \frac{\partial M}{\partial s_k} = \begin{bmatrix} 0 & 0 \\ 0 & -\dfrac{\partial G}{\partial s_k} \end{bmatrix}.$$

It is easy to show that $\partial G_{ij}/\partial s_k = 0$ unless the node k is a top, right, or bottom boundary node. There are three cases for the non-zero entries: non-corner top/bottom, non-corner right, and corners, each of which can be computed using (10.8).

The KKT equations consist of (10.6b), (10.11), and (10.13). A full space method involves the simultaneous solution of these three nonlinear equations. Alternatively, the reduced space method consists of taking $\tilde{\mathcal{J}}(\mathbf{s}) = \mathcal{J}(\mathbf{u}^i(\mathbf{s}))$. Then we have

$$\frac{\partial \tilde{\mathcal{J}}}{\partial s_k} = \Re \sum_{i=1}^{p} \left\langle \mathbf{v}^i, \frac{\partial A}{\partial s_k}\mathbf{u}^i \right\rangle, \tag{10.14}$$

where $\mathbf{v}^i$ and $\mathbf{u}^i$ are solutions of (10.6b) and (10.11).

10.3.2 Direct computation of the gradient

Here we compute

$$\begin{aligned} \frac{\partial \tilde{\mathcal{J}}}{\partial s_k} &= \frac{\partial \mathcal{J}}{\partial s_k} + \sum_{i=1}^{p} \frac{\partial \mathcal{J}}{\partial \mathbf{u}^i}\frac{\partial \mathbf{u}^i}{\partial s_k} + \frac{\partial \mathcal{J}}{\partial \mathbf{u}^{i*}}\frac{\partial \mathbf{u}^{i*}}{\partial s_k} \\ &= 2\Re \sum_{i=1}^{p} \frac{\partial \mathcal{J}}{\partial \mathbf{u}^i}\left(-A^{-1}\frac{\partial A}{\partial s_k}\mathbf{u}^i\right) \end{aligned} \tag{10.15}$$

where we have used

$$\frac{\partial A}{\partial s_k}\mathbf{u}^i + A\frac{\partial \mathbf{u}^i}{\partial s_k} = 0, \tag{10.16}$$

obtained from differentiating (10.6b). We now see that (10.14) and (10.15) are the same by (10.11). The advantage to computing $\mathbf{v}^i$ first and then computing the gradient via (10.14) is that only p adjoint solves are required (one for each input-output pair). Computing (10.15) literally (*i.e.*, computing the expression in parentheses first and then computing the vector-matrix product) would require $N \cdot p$ state solves [Strang, 2007].

10.3.3 Computation of the Hessian

Differentiating (10.14) enables us to write the Hessian

$$\frac{\partial^2 \tilde{\mathcal{J}}}{\partial s_j \partial s_k} = \Re \sum_{i=1}^{p} \mathbf{v}^{i*} \left[\frac{\partial^2 A}{\partial s_j \partial s_k}\right] \mathbf{u}^i + \frac{\partial \mathbf{v}^{i*}}{\partial s_j}\left[\frac{\partial A}{\partial s_k}\right]\mathbf{u}^i + \mathbf{v}^*\left[\frac{\partial A}{\partial s_k}\right]\frac{\partial \mathbf{u}^i}{\partial s_j}.$$

Differentiating the adjoint eq. (10.11) with respect to s_j, gives

$$\begin{aligned}\frac{\partial \mathbf{v}^{i*}}{\partial s_j} &= -\left(2\frac{\partial}{\partial s_j}\frac{\partial \mathcal{J}}{\partial \mathbf{u}^i} + \mathbf{v}^{i*}\left[\frac{\partial A}{\partial s_j}\right]\right)A^{-1} \\ &= -\left(\frac{\partial \mathbf{u}^{i*}}{\partial s_j}P_\Upsilon^t P_\Upsilon + \mathbf{v}^{i*}\left[\frac{\partial A}{\partial s_j}\right]\right)A^{-1}.\end{aligned}$$

Combining the previous two equations with (10.16), and defining $h_{ji} = P_\Upsilon A^{-1}\frac{\partial A}{\partial s_j}\mathbf{u}^i$, we have the Hessian

$$\frac{\partial^2 \tilde{\mathcal{J}}}{\partial s_j \partial s_k} = \Re \sum_{i=1}^{p} h_{ji}^* h_{ki} + \mathbf{v}^{i*}\left[\frac{\partial^2 A}{\partial s_j \partial s_k} - \frac{\partial A}{\partial s_j}A^{-1}\frac{\partial A}{\partial s_k} - \frac{\partial A}{\partial s_k}A^{-1}\frac{\partial A}{\partial s_j}\right]\mathbf{u}^i. \tag{10.17}$$

As expected, the Hessian is symmetric.

10.4 Design variables

To reduce the size of the optimization problem (10.9), we introduce *design variables*, a reduced representation for $\mathbf{L}$ and $\mathbf{C}$. There are many natural choices for the design variables $\mathbf{r}$. The following choices are labeled for future reference.

(D1) If $\mathbf{L}$ and $\mathbf{C}$ are symmetric in the sense that

$$L^h_{i,j} = L^h_{m+1-i,j}, \qquad L^v_{i,j} = L^v_{m+1-i,j},$$
$$C_{i,j} = C_{m+1-i,j},$$

then the transfer matrix satisfies $T_{i,j} = T_{m+1-i,m+1-j}$. Thus, if the desired transfer matrix has this property, $\mathbf{r}$ can be chosen to enforce this symmetry on $\mathbf{L}$ and $\mathbf{C}$. This reduces the dimension of the design variable space by a factor of approximately two.

(D2) The vectors $\mathbf{L}_h$ and $\mathbf{L}_v$ can be chosen as a discretization of a single continuous function $\mu(\mathbf{x})$ as in [Bhat and Osting, 2011a]. This imposes a compatibility condition on $\mathbf{L}_h$ and $\mathbf{L}_v$, reducing the dimension of the design space by approximately three. Specifically, we let $\boldsymbol{\mu}$ be a $m+1 \times n+1$ matrix and set

$$L^h_{ij} = \frac{1}{2}\left(\mu_{ij} + \mu_{i+1,j}\right),\ 1 \le i \le m,\ 1 \le j \le n \tag{10.18a}$$
$$L^v_{ij} = \frac{1}{2}\left(\mu_{ij} + \mu_{i,j+1}\right),\ 2 \le i \le m,\ 1 \le j \le n. \tag{10.18b}$$

The design variables then consist of $\mathbf{C}$ and $\boldsymbol{\mu}$.

(D3) Restricting to lattices with $\mathbf{L} = 1$ reduces the dimension of the design space by a factor of three. This is analogous to considering media with constant permeability [Bhat and Osting, 2011a].

(D4) Combining the ideas in (D1) and (D3), we take $\mathbf{L} = 1$ and force C to have symmetry. This reduces the design variable space by a factor of six.

(D5) The vectors $\mathbf{L}$ and $\mathbf{C}$ can also be represented in terms of a truncated basis, such as the Fourier, wavelet, or block bases, but we do not pursue this here.

For a (BC1) lattice, energy leaks out of the top/bottom boundaries, so the total energy collected at the output is less than the input energy. Since we are primarily interested in the shape of the output $g(y)$, we include an extra design variable δ in the objective function (10.9), replacing T_d by δT_d. For all design variable choices, we let $r_1 = \delta$.

Let $\mathbf{s} = \mathbf{s}(\mathbf{r})$ denote the dependence of $\mathbf{s}$ on a set of design variables $\mathbf{r}$. Then the gradient and Hessian can be computed

$$g \equiv \nabla_{\mathbf{r}} \tilde{\mathcal{J}}(\mathbf{s}(\mathbf{r})) = \mathbf{s}_{\mathbf{r}} \nabla_{\mathbf{s}} \tilde{\mathcal{J}}$$
$$H \equiv \nabla_{\mathbf{r}} \nabla_{\mathbf{r}} \tilde{\mathcal{J}}(\mathbf{s}(\mathbf{r})) = \mathbf{s}_{\mathbf{r}} \nabla_{\mathbf{s}} \nabla_{\mathbf{s}} \tilde{\mathcal{J}} \mathbf{s}_{\mathbf{r}}^t,$$

where $\mathbf{s}_{\mathbf{r}}$ denotes the Jacobian and $\nabla_{\mathbf{s}} \tilde{\mathcal{J}}$ and $\nabla_{\mathbf{s}} \nabla_{\mathbf{s}} \tilde{\mathcal{J}}$ were computed in (10.14) and (10.17) respectively.

Once the design variables are chosen, the optimization problem (10.9) can be written

$$\min_{\mathbf{r} \in \mathfrak{A}_r} \quad \tilde{\mathcal{J}}(\mathbf{r}) := \frac{1}{2} \|T(\mathbf{r}) - r_1 T_d\|_F^2 \tag{10.20}$$

where $\mathfrak{A}_r$ is an admissible set for the design variables $\mathbf{r}$,

$$\mathfrak{A}_r := \{\mathbf{r} : \quad \underline{r} \leq r_j \leq \overline{r} \quad \text{for all } j\}.$$

10.5 Computational results

In Sections 10.5.1 through 10.5.4, we apply gradient-based optimization tools [Nocedal and Wright, 2006] to solve the lattice synthesis problem (10.20) for four desired transfer matrices. In Section 10.5.1, we also compare the performance of several different optimization methods. In Sec. 10.5.3 we compare the two choices of boundary conditions given in Sec. 10.2. In all other sections, we use (BC1). In Section 10.5.5, we discuss the sensitivity of the transfer matrix of an inductor-capacitor lattice to small perturbations in $\mathbf{L}$ or $\mathbf{C}$. Finally, in Sections 10.5.6 and 10.5.7, we study numerically the well-posedness of the synthesis problem.

10.5.1 Diagonal transfer matrix

In this section, we define the desired transfer matrix to be the diagonal matrix $T_d = \text{diag}(\mathbf{t})$. For a lattice with m rows, let $j_c = (m+1)/2$ and $t_j = \exp(-2(j - j_c)^2)$, $j = 1, \ldots, m$. We set $\alpha = .08$ and choose (D1) design variables with lower and upper bounds 0.05 and 5.

We now solve the synthesis problem (10.20) for an $m \times m$ lattice for $m = 8$ ($N = 184$) and $m = 16$ ($N = 752$) using several different numerical methods. For each m and numerical method used, in Fig. 10.2, we plot both iteration number and wall time vs. the objective function value.

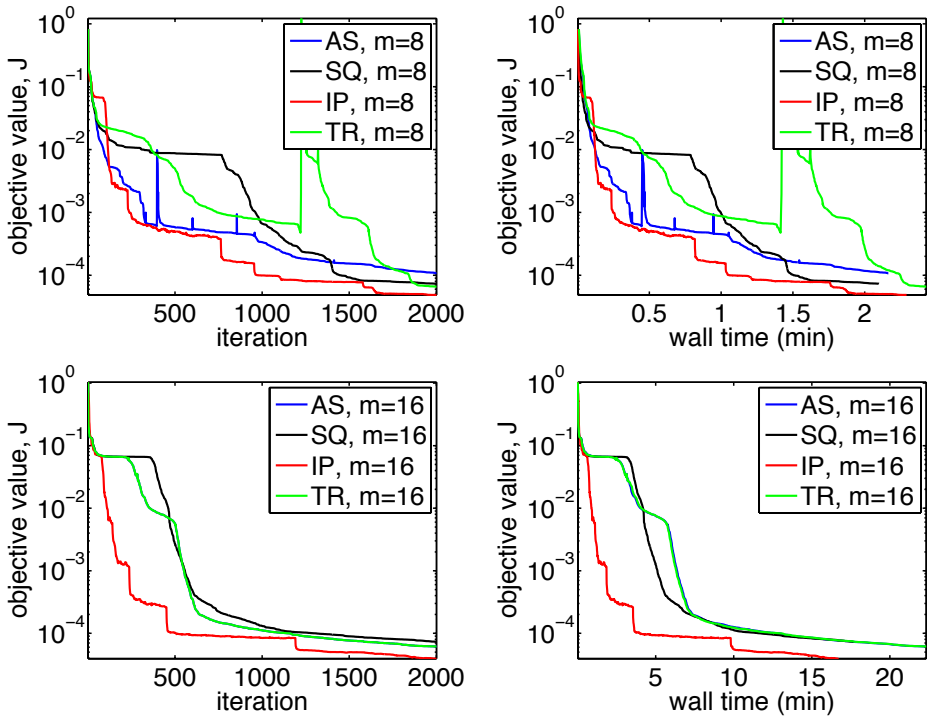


Figure 10.2: We plot (left) iteration number vs. objective function value and (right) wall time vs. objective function value for the solution of (10.20) on an $m \times m$ lattice for $m = 8$ (top) and $m = 16$ (bottom) and various optimization methods (see Sec. 10.5.1 for method abbreviation definitions).

In what follows, we describe the methods compared in Fig. 10.2. All computations were done using Matlab 7.11 on a 2.4 GHz Intel Core 2 Duo desktop computer with 2GB of RAM. In each case, the convergence criteria was set using the Matlab options: `MaxIter` $= 2000$, `TolX` $= 10^{-14}$, and `TolFun` $= 10^{-13}$. In all examples here and below, the optimization method is initialized with constant design variables, $\mathbf{r}$. We compare Matlab's `fmincon` implementation of the following nonlinear constrained optimization algorithms:

(SQ) `sqp`: The sequential quadratic programming (SQP) approach is to approximate (10.20) by a quadratic minimization problem at each iteration. This quadratic form involves the Hessian of the objective function, which is approximated using the BFGS method [Nocedal and Wright, 2006, Ch. 18].

(AS) `active-set`: The active set method solves a sequence of unconstrained optimization prob-

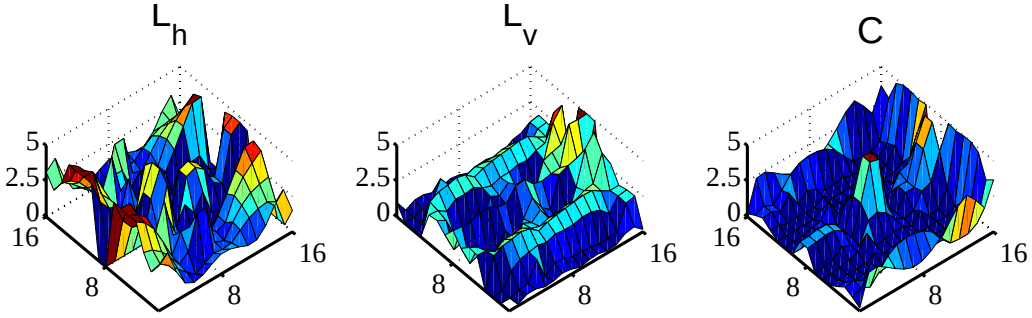


Figure 10.3: The $(\mathbf{L}, \mathbf{C})$ matrices for the 16×16 diagonal transfer lattice found in Section 10.5.1 with objective value $J = 7.3 \times 10^{-5}$.

lems. The optimization variables do not necessarily satisfy the bounds at each iteration.

(IP) `interior-point`: This line-search based quasi-Newton method uses the BFGS method to update the approximate Hessian at each iteration. The constraints are enforced using a logarithmic barrier function.

(TR) `trust-region-reflective`: We use this subspace trust-region method with `large-scale` = `off`.

From Fig. 10.2, we conclude that all tested methods are able to find solutions with low objective values. The other methods perform approximately the same in both iteration count and wall time. The interior point method (IP) performs best; however, the solution obtained tends to be less smooth than that obtained via the other methods. In what follows, we primarily use the (AS) method. In addition to the four methods described above, we also tried Newton's method, but found that the cost of computing the Hessian (10.17) was prohibitively large for lattice sizes of interest.

Let us return to the design problem for the diagonal transfer matrix T_d. The optimal solution $(\mathbf{L}^*, \mathbf{C}^*)$ for $m = 16$ obtained using (SQ) is plotted in Fig. 10.3 and has objective value $J = 7.3 \times 10^{-5}$. The method terminated when the maximum number of iterations, `MaxIter` = 2000, was reached.

For this transfer function and *all* transfer functions considered in the subsequent sections, the design variable $r_1 = \delta$ attains the lower bound constraint of $\underline{r}_1 = .6$. This indicates it is easier to

synthesize energy-dissipative lattices.

10.5.2 Waveguide filter / Rank-one projection

In this section, we define

$$T_d = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 1 & 0 & \cdots \\ \cdots & 0 & 1 & 1 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

the discrete analogue of a waveguide transfer function $f \mapsto \langle \psi, f \rangle \psi$, where ψ is a desired bound state.

With $\alpha = .32$, we use (D1) design variables with lower and upper bounds given by 0.05 and 50. For a 24×24 lattice, we use the active set method (AS) to obtain the optimal solution $(\mathbf{L}^*, \mathbf{C}^*)$ plotted in Fig. 10.4 with objective value $J = 6 \times 10^{-6}$. The method terminated after 547 iterations because the predicted change in the objective function was less than $\texttt{TolFun} = 10^{-13}$.

The optimal solution, plotted in Fig. 10.4, has horizontal inductors $\mathbf{L}^h$ and capacitors $\mathbf{C}$ which take large values in a strip from the center inputs to the center outputs. Outside of this strip, the C matrix has periodic structure arranged to impede an incoming wave. The fact that we can recognize structure in the solution to an optimization problem in $\mathbb{R}^{1704}$ is remarkable, and suggests rigidity in the synthesis problem.

10.5.3 Low-pass filter / Smoothing convolution

In [Bhat and Osting, 2011a], we used separation of variables to obtain the exact solution for the continuous analogue of the forward problem (10.3) for a homogeneous lattice. We concluded that a homogenous lattice strongly damps oscillatory input, which suggests that this type of lattice is well-suited for performing low-pass filtering functions. We investigate this intuition here by constructing a circuit that behaves as a low-pass filter.

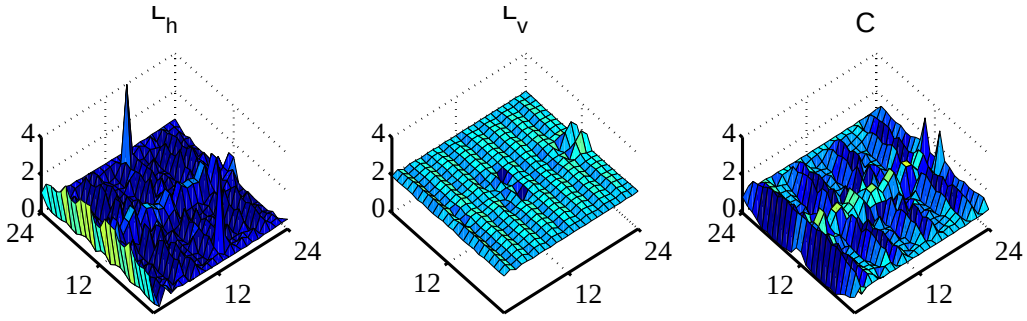


Figure 10.4: The $(\mathbf{L}, \mathbf{C})$ matrices for the 24×24 waveguide in Section 10.5.2 with objective value $J = 6 \times 10^{-6}$.

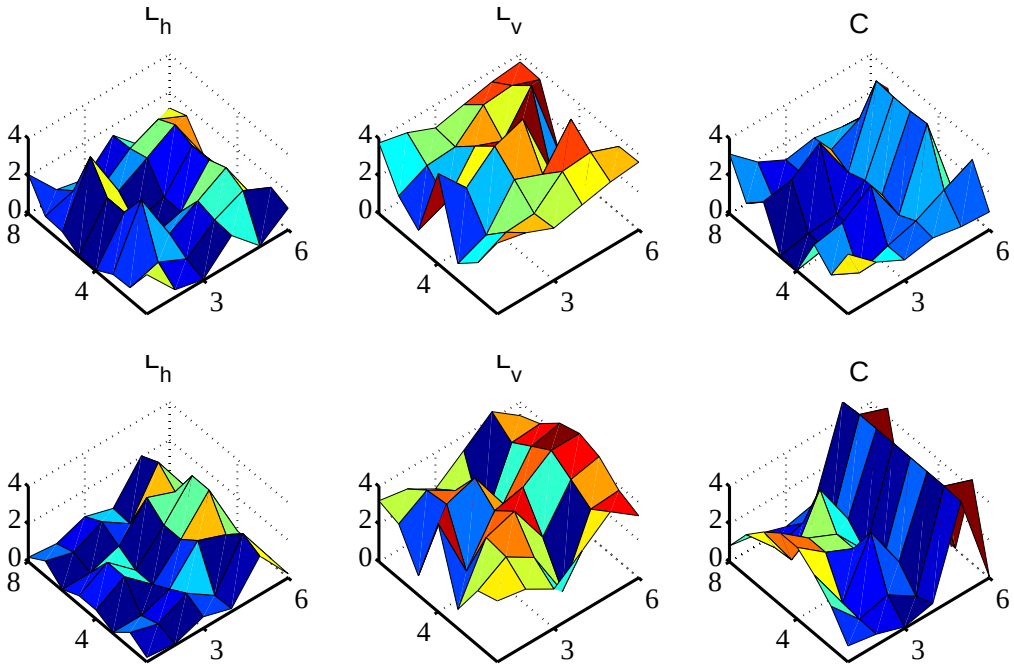


Figure 10.5: The $(\mathbf{L}, \mathbf{C})$ matrices for the low-pass filter in Section 10.5.3 for the 8×6 lattice for boundary conditions as described by (BC1) in the top panel and (BC2) in the lower panel with resp. objective values $J = 6.24 \times 10^{-7}$ and $J = 2.98 \times 10^{-5}$.

For an 8×6 lattice, we define the transfer matrix:

$$T_d = \frac{1}{44} \begin{pmatrix} 2 & 1 & 0 & 0 \\ 4 & 2 & 1 & 0 \\ 8 & 4 & 2 & 1 \\ 4 & 8 & 4 & 2 \\ 2 & 4 & 8 & 4 \\ 1 & 2 & 4 & 8 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 2 \end{pmatrix}. \tag{10.21}$$

The matrix T_d can be obtained by removing the first two and last two columns from an 8×8 Toeplitz matrix. We also remove the first and last two columns of the transfer matrix T in (10.20). With $\alpha = 0.16$ and (D1) design variables with lower and upper bounds given by .05 and 50, we use the active set method (AS) for each of the boundary conditions given in Sec. 10.2. For (BC1), the final objective function value is $J = 6.24 \times 10^{-7}$ and for (BC2), the final objective value is $J = 2.98 \times 10^{-5}$. In both cases, the method terminated because the predicted change in the objective function was less than `TolFun` $= 10^{-13}$. In Fig. 10.5, we plot the optimal solution $(\mathbf{L}^*, \mathbf{C}^*)$ for both choices of boundary conditions.

10.5.4 Power combiner / Funnel

Motivated by the power combiner introduced in [Afshari *et al.*, 2006a; Afshari *et al.*, 2006b], we consider the transfer matrix that maps all inputs to the center output. The desired transfer matrix T_d of size $m \times m$ (where $m = 2j + 1$ is odd) consists of a matrix where row $j + 1$ has a 1 in each column, and all other rows are identically zero.

We set $\alpha = 0.08$ and choose (D2) design variables. The upper and lower bounds were .05 and 20. In Fig. 10.6, we plot the optimal solution $(\mathbf{L}^*, \mathbf{C}^*)$ for the synthesis problem attained using the active set method (AS). The solution is plotted for $m \times m$ lattices where $m = 11$, 21, and 31 with respective objective function values 2×10^{-5}, 3×10^{-5}, and 3×10^{-5}. In each case, the method terminated because the maximum number of iterations, `MaxIter` $= 3000$, was reached.

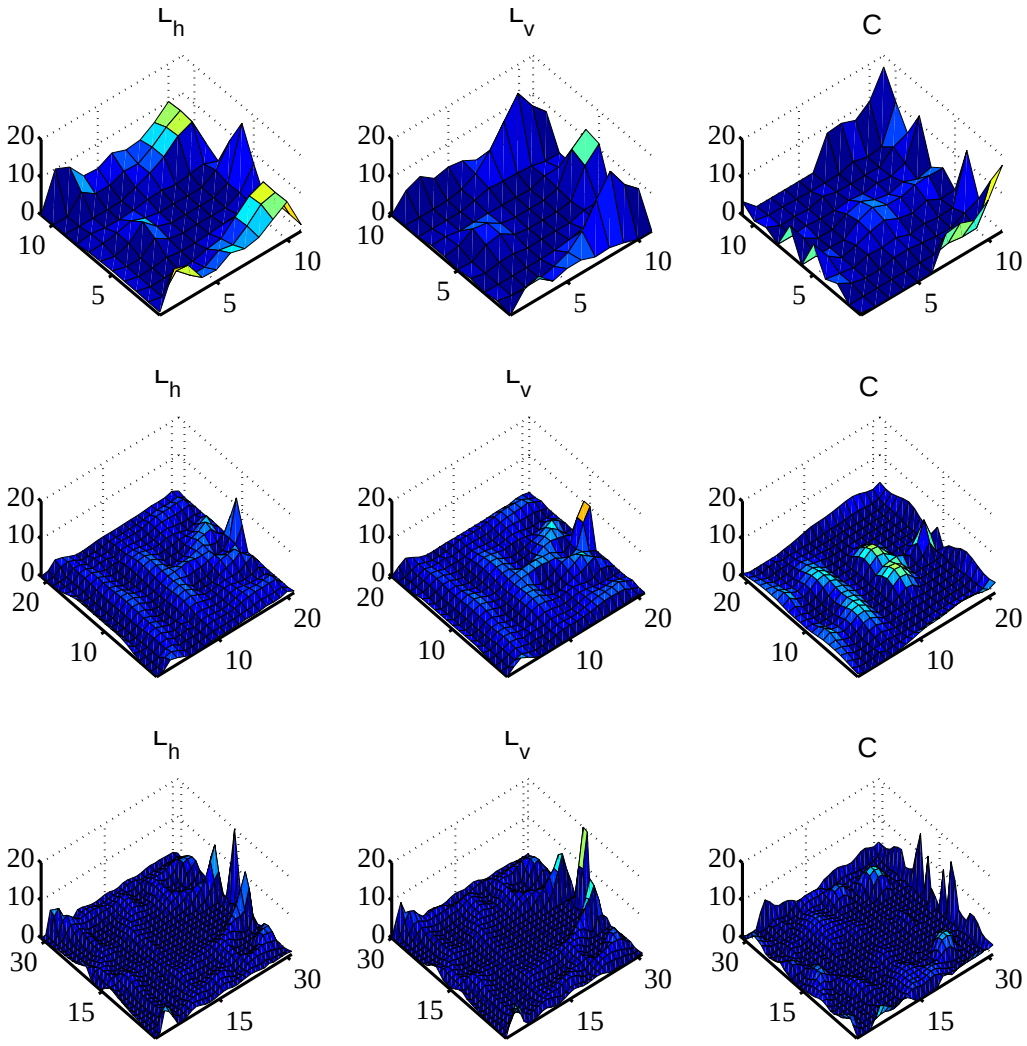


Figure 10.6: The $(\mathbf{L}, \mathbf{C})$ surfaces for the funnel in Sec. 10.5.4 for the $m \times m$ lattice for $m = 11$, 21, 31 with objective values 2×10^{-5}, 3×10^{-5}, and 3×10^{-5}.

10.5.5 Robustness / Sensitivity of optimal devices

In this section, we consider the sensitivity of optimal devices to small changes in $(\mathbf{L}, \mathbf{C}, \mathbf{G})$.

Proposition 10.5.1. *Let $T_j = P_\Upsilon A_j^{-1} P_\Gamma^t$, $j = 1, 2$ be the transfer matrices for two different circuits with capacitances, inductances, and conductances given by $(\mathbf{C}_j, \mathbf{L}_j, \mathbf{G}_j)$, where*

$$A_j := 2\pi\imath\alpha\, diag(\mathbf{L}_j, \mathbf{C}_j) - M(\mathbf{G}_j).$$

Assume $\rho := \|A_1^{-1}(A_2 - A_1)\|_2 < 1$, and define $\gamma = 1/\sigma_1(A_1)$ where $\sigma_1(A_1) > 0$ is the smallest singular value of A_1. Then

$$\|T_1 - T_2\|_F \leq \frac{mN\gamma^2}{1-\rho}\Big[2\pi\alpha\left(\|\mathbf{L}_2 - \mathbf{L}_1\|_2 + \|\mathbf{C}_2 - \mathbf{C}_1\|_2\right) + \|\mathbf{G}_2 - \mathbf{G}_1\|_2\Big].$$

Proof. The matrix identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ gives

$$T_1 - T_2 = P_\Upsilon A_1^{-1}(A_2 - A_1)A_2^{-1}P_\Gamma^t.$$

Taking the Frobenius norm of both sides and using the the sub-multiplicative property of the Frobenius norm we obtain

$$\|T_1 - T_2\|_F \leq \|P_\Upsilon A_1^{-1}\|_F \|A_2 - A_1\|_F \|A_2^{-1}P_\Gamma^t\|_F. \tag{10.22}$$

We treat the 3 pieces on the right hand side of (10.22) in turn. First note $\|P_\Upsilon\|_F = |\Upsilon|^{1/2} = \sqrt{m}$ and $\|P_\Gamma\|_F = |\Gamma|^{1/2} = \sqrt{m}$.

1. We compute

$$\|P_\Upsilon A_1^{-1}\|_F \leq \|P_\Upsilon\|_F \|A_1^{-1}\|_F \leq \sqrt{m}\sqrt{N}\|A_1^{-1}\|_2 = \sqrt{mN}\gamma.$$

Here we used the norm relation: $\|A\|_F \leq \sqrt{r}\|A\|_2$ where r is the rank of A and $\|A_1^{-1}\|_2 = \sigma_N(A_1^{-1}) = 1/\sigma_1(A_1) = \gamma$.

2. We compute

$$\begin{aligned}\|A_2 - A_1\|_F &\leq 2\pi\alpha\|\mathrm{diag}(\mathbf{L}_2, \mathbf{C}_2) - \mathrm{diag}(\mathbf{L}_1, \mathbf{C}_1)\|_F + \|M(\mathbf{G}_2) - M(\mathbf{G}_1))\|_F \\ &= 2\pi\alpha\left(\|\mathbf{L}_2 - \mathbf{L}_1\|_2 + \|\mathbf{C}_2 - \mathbf{C}_1\|_2\right) + \|\mathbf{G}_2 - \mathbf{G}_1\|_2.\end{aligned}$$

3. As above, we compute

$$\|A_2^{-1}P_\Gamma^t\|_F \leq \|A_2^{-1}\|_F \|P_\Gamma^t\|_F \leq \sqrt{mN}\|A_2^{-1}\|_2.$$

Our goal now is to estimate $\|A_2^{-1}\|_2$ in terms of γ. We compute

$$\begin{aligned}\|A_2^{-1}\|_2 &= \| \left[A_1 \left(\mathrm{Id} + A_1^{-1}(A_2 - A_1)\right)\right]^{-1} \|_2 \\ &= \| \left(\mathrm{Id} + A_1^{-1}(A_2 - A_1)\right)^{-1} A_1^{-1} \|_2 \\ &\leq \gamma \| \left(\mathrm{Id} + A_1^{-1}(A_2 - A_1)\right)^{-1} \|_2\end{aligned}$$

Note that $\left(\mathrm{Id} + A_1^{-1}(A_2 - A_1)\right)^{-1}$ exists by the assumption $\rho < 1$. Summing the Neumann series for this expression gives

$$\| \left(\mathrm{Id} + A_1^{-1}(A_2 - A_1)\right)^{-1} \|_2 \leq \sum_{j=0}^{\infty} \|A_1^{-1}(A_2 - A_1)\|_2^j = \sum_{j=0}^{\infty} \rho^j = \frac{1}{1-\rho}.$$

Putting these 3 pieces together yields the desired result. □

The upshot of this proposition is that if a circuit is perturbed by modifying $(\mathbf{L}, \mathbf{C}, \mathbf{G})$, then the change in the transfer matrix for the circuit is bounded by the size of the perturbation. However, the bounding constant could be large and increases with increasing circuit size.

We conduct a numerical experiment to further investigate this dependence for the low-pass filtering device introduced in Section 10.5.3. Let $(\mathbf{L}^*, \mathbf{C}^*)$ denote the 8×6 device with (BC1) boundary conditions plotted in Fig. 10.5(top panel) that minimizes $\mathcal{J}$ for the desired transfer matrix in (10.21) with objective value $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*) = 6.24 \times 10^{-7}$. We now evaluate $\mathcal{J}$ for a distribution of perturbations to $(\mathbf{L}^*, \mathbf{C}^*)$. Specifically, we consider multiplicative noise and evaluate $\mathcal{J}(\mathbf{L}.\mathbf{u}, \mathbf{C}.\mathbf{v})$, where $\mathbf{a}.\mathbf{b}$ denotes entry-wise multiplication of the vectors $\mathbf{a}$ and $\mathbf{b}$, and $(\mathbf{u}, \mathbf{v})$ have entries which are normally distributed with mean 1 and standard deviation 0.02. We interpret a structure $(\mathbf{L}.\mathbf{u}, \mathbf{C}.\mathbf{v})$ to be a low-pass filtering device manufactured with 2% tolerance. In Fig. 10.7, we plot a histogram of the objective function value evaluated on a sample size of $100,000$ drawn from this distribution. The 10th, 50th and 90th quantiles are 1.8×10^{-3}, 6.9×10^{-3}, and 3.9×10^{-2}.

We might also consider the sensitivity of optimal devices to small changes in α. However, since (10.6) is invariant under the transformation in (10.5), perturbing α is equivalent to choosing a multiplicative perturbation $(\mathbf{u}, \mathbf{v})$ from a skewed distribution.

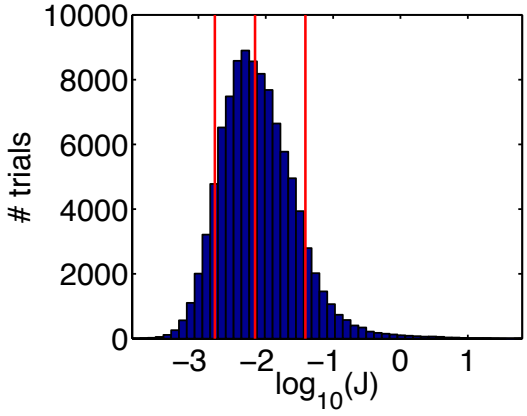


Figure 10.7: A histogram of the objective function evaluated for 100,000 low-pass filters (see Fig. 10.5) with 2% normally-distributed, multiplicative noise. The vertical lines indicate the 10th, 50th and 90th quantiles. See Sec. 10.5.5.

10.5.6 Known lattice recovery / Inverse crime study

In the preceding sections, our goal was to obtain useful circuits. Here and in the next section, we conduct numerical experiments to quantify the ill-posedness of the problem.

In this first numerical experiment, we commit a so-called "inverse crime." We take $p = m$ and generate a transfer matrix T_d by solving the forward problem for *known* values of $(\mathbf{L}^0, \mathbf{C}^0)$. We then set aside $(\mathbf{L}^0, \mathbf{C}^0)$ and solve the design problem for the transfer matrix T_d, giving us a computed solution $(\mathbf{L}^*, \mathbf{C}^*)$.

This is refered to as an "inverse crime" since the model of the forward problem used to generate the transfer matrix is precisely the same model assumed in the solution of the design problem. For these transfer matrices, we happen to know that there is a point—namely $(\mathbf{L}^0, \mathbf{C}^0)$—where the objective function is zero. We can then measure how well our algorithm does by comparing $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*)$ with zero. We can also characterize the solution space by checking how $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*)$ depends on the known solution $(\mathbf{L}^0, \mathbf{C}^0)$.

Let us describe how we generate a random matrix $\mathbf{C}^0$. We fix integer parameters $\nu > 0$ and σ as well as real parameters $\rho_{\min}$ and $\rho_{\max}$. We choose two random vectors of Fourier sine coefficients $\mathbf{k}^x$ and $\mathbf{k}^y$, both of size $\nu \times 1$. The j-th entry $k_j^{(\cdot)}$ is sampled from a $U(0,1)$ distribution and then

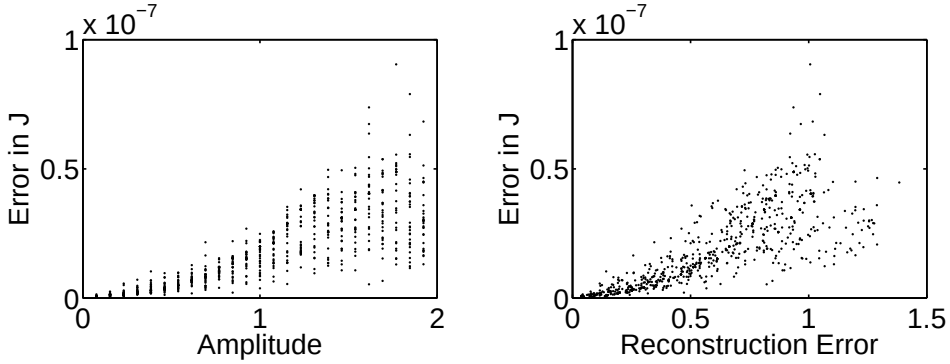


Figure 10.8: Objective function value $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*)$ versus ρ (left panel) and versus $\|(\mathbf{L}^*, \mathbf{C}^*) - (\mathbf{L}^0, \mathbf{C}^0)\|_\infty$ (right panel) for 750 runs, all on 8×8 lattices. See Section 10.5.6.

multiplied by $j^{-\sigma}$. We sample $P(x, y) = \sum_{i=1}^{\nu} \sum_{j=1}^{\nu} k_i^x k_j^y \sin(ix) \sin(jy)$ to create an $m \times n$ matrix $\mathbf{C}^0$ that is then scaled and translated so its max/min values are, respectively, $\rho_{\max}$ and $\rho_{\min}$.

For $\mathbf{L}^h$ and $\mathbf{L}^v$, we follow (10.18) after generating an $(m+1) \times (n+1)$ matrix $\boldsymbol{\mu}$ by sampling $P(x, y)$. We scale and translate the matrices $\mathbf{L}^h$ and $\mathbf{L}^{\mathbf{v}}$ so their max/min values are, respectively, $\rho_{\max}$ and $\rho_{\min}$. In all cases, sampling of $P(x, y)$ is performed on a regular grid in the square $[0, 2\pi]^2$.

Using the above approach for generating random pairs $(\mathbf{L}^0, \mathbf{C}^0)$, we solved the design problem 750 times on an 8×8 lattice. We used the active set method (AS) with `TolX` $= 10^{-14}$, `TolFun` $= 10^{-13}$, and (D2) design variables with lower and upper bounds of .05 and 50. For all 750 runs, the code terminated because the magnitude of the directional derivative in the search direction was less than $2\,$`TolFun`.

We stepped ν from 0 to 5 and σ from 1 to 5. We stepped ρ through 25 equispaced values in the interior of $(0, 2)$, and set $\rho_{\max} = 1 + \rho/2$, $\rho_{\min} = 1 - \rho/2$.

In the left panel of Fig. 10.8, we plot the objective function value $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*)$ versus amplitude ρ for all 750 runs. The plot shows that the code performed very well across all runs, with the maximum value of $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*)$ less than 10^{-7}. The plot also reflects a correlation coefficient of 0.82, which indicates that the larger the amplitude of spatial oscillations in $\mathbf{L}^0$ and $\mathbf{C}^0$, the poorer the quality of the local optimum reached.

In the right panel of Fig. 10.8, we plot the objective function value $\mathcal{J}(\mathbf{L}^*, \mathbf{C}^*)$ versus reconstruction error $\|(\mathbf{L}^*, \mathbf{C}^*) - (\mathbf{L}^0, \mathbf{C}^0)\|_\infty$ for all 750 runs. The plot reflects that, as we move further

away from the global minimum $(\mathbf{L}^*, \mathbf{C}^*)$, we are still able to achieve transfer matrices that are very close to what is desired. However, the correlation coefficient of 0.80 indicates a small degradation in the quality of the local optima as a function of distance from a global optimum.

10.5.7 Lattice refinement and coarsening

For a lattice with homogeneous $(\mathbf{L}, \mathbf{C})$ the Nyquist principle states that $\alpha\sqrt{LC} < \sqrt{2}/\pi$. In [Bhat and Osting, 2009a], we found that Kirchhoff's laws (10.3) behave like their continuum limit if $\alpha\sqrt{LC} < 1/(2\pi)$, which is roughly one-third of the Nyquist frequency. In [Bhat and Osting, 2011a] we showed that the continuum limit is precisely the system of equations for the (H_1, H_2, E) polarized mode for Maxwell's equations in a planar medium. Thus we expect that for α sufficiently small, even if $(\mathbf{L}, \mathbf{C})$ is inhomogeneous, one may increase the size of the lattice and rescale $(\mathbf{L}, \mathbf{C})$ so that both problems are a discretization of the same continuum problem.

In this section, we use this principle to provide quantitative estimates on the ill-posedness of the synthesis problem. Throughout, we set $\mathbf{L} = 1$ and $\alpha = 1$.

On a 40×40 lattice, we set $C_{ij} = 1 + \operatorname{sech}^2\gamma\left((i-20.5)^2 + (j-20.5)^2\right)$ for $\gamma = 25/39^2$. Using this $\mathbf{C}_{40}$, we solve (10.3) for the transfer matrix T_{40}.

We then average 2×2 subblocks of both T_{40} and $\mathbf{C}_{40}$ to obtain a transfer function T_{20} and capacitances $\mathbf{C}_{20}$ on a 20×20 lattice. The vector $\mathbf{C}_{20}$ is divided by 4 based on the finite volume derivation in [Bhat and Osting, 2011a]. The synthesis problem with desired transfer function T_{20} is then initialized using $\mathbf{C}_{20}$ and solved using (D4) decision variables. We denote this solution $\tilde{\mathbf{C}}_{20}$ and note that the objective function value is $\mathcal{J} = 3.7 \times 10^{-3}$.

We now refine $\tilde{\mathbf{C}}_{20}$ to a 40×40 lattice by repeating 2×2 blocks of $\tilde{\mathbf{C}}_{20}$. We denote these capacitances by $\tilde{\mathbf{C}}_{40}$. The synthesis problem with transfer function T_{40} is initialized using $\tilde{\mathbf{C}}_{40}$ and solved to obtain $\overline{\mathbf{C}}_{40}$. Both surfaces are plotted in Fig. 10.9. The final value of the objective function is 9.5×10^{-8} and $\|\overline{\mathbf{C}}_{40} - \mathbf{C}_{40}\|_F = 11.1$. Thus, $\mathbf{C}_{40}$ and $\overline{\mathbf{C}}_{40}$ are far apart in the Frobenius norm but achieve almost the same transfer function. Although the problem is ill-posed and the solution obtained is different than $\mathbf{C}_{40}$, we emphasize that we view $\overline{\mathbf{C}}_{40}$ as a excellent solution to the synthesis problem since it achieves a phenomenally low objective function value.

In inverse problems, one applies regularization methods to enforce *a priori* known information such as smoothness. Similarly, in the design problem considered here, where the "data" (*i.e.*,

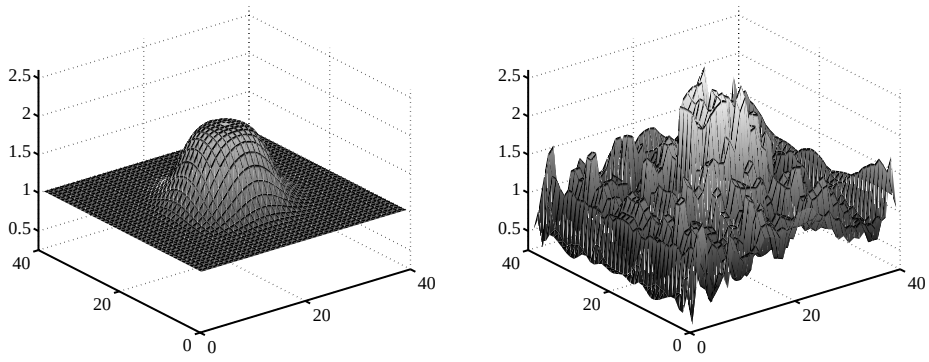


Figure 10.9: A plot of the capacitance matrices $\mathbf{C}_{40}$ (left) and $\overline{\mathbf{C}}_{40}$ (right) as defined in Section 10.5.7.

desired transfer matrix) is known perfectly, one could apply regularization methods to force $\mathbf{L}$ and $\mathbf{C}$ to have desired properties. We do not pursue this direction here.

10.6 Conclusion / Discussion

We have formulated the two-dimensional transmission lattice synthesis problem as an optimization problem, the solution of which yields inductor-capacitor lattices that can be fabricated for custom/novel applications in analog signal processing and filtering. For several chosen transfer functions, we have demonstrated that gradient-based optimization methods can be used to obtain excellent solutions to the synthesis problem.

In other contexts, the ideas presented in this chapter are familiar: one can engineer the permittivity ε and permeability μ of a medium to control the propagation of EM waves [Burger *et al.*, 2004; Joannopoulos *et al.*, 2008], and in quantum mechanics, one may engineer a potential to have desired scattering properties [Osting and Weinstein, 2011a]. As the frequency of analog circuits marches into the THz range, it is increasingly important that the circuit model be related, both qualitatively and quantitatively, to Maxwell's equations. Ultimately, if one is interested in designing a microwave frequency device, one performs a direct numerical simulation of Maxwell's equations to confirm that the circuit model accurately predicts the device's behavior. Based on our findings in [Bhat and Osting, 2011a], these connections can be made more precise. Kirchhoff's laws for the

2-D LC lattice (10.1) can be viewed as a finite volume discretization of Maxwell's equations for a planar, inhomogeneous medium. For large circuits with smoothly varying $(\mathbf{C}, \mathbf{L})$, this discretization is accurate, and one can interpret the present work as a discretize-then-optimize approach to solving the (ε, μ) synthesis problem for Maxwell's equations. This is the subject of forthcoming work.

Part IV

Bibliography

Bibliography

[Aarao, 2007] J. Aarao. Fundamental solutions for some partial differential operators from fluid dynamics and statistical physics. *SIAM Review*, 49(2):303–314, 2007.

[Abenius *et al.*, 2002] E. Abenius, U. Andersson, F. Edelvik, L. Eriksson, and G. Ledfelt. Hybrid time domain solvers for the Maxwell equations in 2D. *Internat. J. for Num. Meth. in Eng.*, 53(9):2185–2199, 2002.

[Ablowitz and Zhu, 2010] M. J. Ablowitz and Y. Zhu. Evolution of Bloch-mode envelopes in two-dimensional generalized honeycomb lattices. *Phys. Rev. A*, 82(1):013840, 2010.

[Adler *et al.*, 1960] R. B. Adler, L. J. Chu, and R. M. Fano. *Electromagnetic Energy Transmission.* John Wiley & Sons, 1960.

[Afshari *et al.*, 2006a] E. Afshari, H. S. Bhat, A. Hajimiri, and J. E. Marsden. Extremely wideband signal shaping using one- and two-dimensional nonuniform nonlinear transmission lines. *J. Appl. Phys.*, 99(5):054901, 2006.

[Afshari *et al.*, 2006b] E. Afshari, H. S. Bhat, X. Li, and A. Hajimiri. Electrical funnel: a new signal combining method. In *Proc. of the IEEE International Solid-State Circuits Conference (ISSCC'06)*, pages 206–208, San Francisco, CA, 2006.

[Afshari *et al.*, 2008] E. Afshari, H. S. Bhat, and A. Hajimiri. Ultrafast analog Fourier transform using two-dimensional LC lattice. *IEEE Trans. on Circuits and Systems I*, 55(8):2332–2343, 2008.

[Agmon, 1975] S. Agmon. Spectral properties of Schrödinger operators and scattering theory. *Annali della Scuola Normale Superiore di Pisa*, 2(2):151–218, 1975.

[Agmon, 1996] S. Agmon. A perturbation theory of resonances. *Journées Équations aux Dérivées Partielles*, pages 1–11, 1996.

[Akcelik *et al.*, 2005] V. Akcelik, G. Biros, O. Ghattas, D. Keyes, K. Ko, L.-Q. Lee, and E. G. Ng. Adjoint methods for electromagnetic shape optimization of the low-loss cavity for the international linear collider. *J. Phys: Conference Series*, 16:435–445, 2005.

[Akcelik *et al.*, 2006] V. Akcelik, G. Biros, O. Ghattas, J. Hill, D. Keyes, and B. van Bloemen Waanders. Parallel algorithms for PDE-constrained optimization. *SIAM Frontiers of Parallel Computing*, 2006.

[Akcelik *et al.*, 2008] V. Akcelik, K. Ko, L.-Q. Lee, Z. Li, C. Ng, and L. Xiao. Shape determination for deformed electromagnetic cavities. *J. Comp. Phys.*, 227:1722–1738, 2008.

[Ali, 2009] S. H. M. Ali. *System Level Performance and Yield Optimisation for Analogue Integrated Circuits*. PhD thesis, University of Southampton, 2009.

[Angell and Kirsch, 2004] T. S. Angell and A. Kirsch. *Optimization Methods in Electromagnetic Radiation*. Springer-Verlag, 2004.

[Arendt *et al.*, 2009] W. Arendt, R. Nittka, W. Peter, and F. Steiner. Weyl's law: spectral properties of the Laplacian in mathematics and physics. In *Mathematical Analysis of Evolution, Information, and Complexity*, pages 1–71. Wiley, 2009.

[Ashbaugh and Benguria, 1992a] M. S. Ashbaugh and R. D. Benguria. Isoperimetric inequalities for eigenvalue ratios. *Partial Differential Equations of Elliptic Type, Cortona*, pages 1–36, 1992.

[Ashbaugh and Benguria, 1992b] M. S. Ashbaugh and R. D. Benguria. A sharp bound for the ratio of the first two eigenvalues of Dirichlet-Laplacians and extensions. *Annals of Mathematics*, 135(3):601–628, 1992.

[Ashbaugh and Benguria, 1993] M. S. Ashbaugh and R. D. Benguria. Isoperimetric bounds for higher eigenvalue ratios for the n-dimensional fixed membrane problem. *Proc. Royal Soc. Edinburgh A*, 123:977–985, 1993.

[Ashbaugh and Benguria, 2007] M. S. Ashbaugh and R. D. Benguria. Isoperimetric inequalities for eigenvalues of the Laplacian. *Proc. of Symposia in Pure Math.*, 76(1):105–139, 2007.

[Ashbaugh and Benguria, 2010] M. S. Ashbaugh and R. D. Benguria. The problem of queen Dido. www.math.uiuc.edu/l̃augesen/dido-isoperimetry-history.pdf, 2010.

[Authier, 2002] A. Authier. *Dynamical Theory of X-ray Diffraction.* IUCr Monographs on Crystallography. Oxford University Press, 2002.

[Balanis, 2005] C. A. Balanis. *Antenna Theory: Analysis and Design.* John Wiley, 2005.

[Bamberger *et al.*, 1988] A. Bamberger, J. C. Guillot, and P. Joly. Numerical diffraction by a uniform grid. *SIAM J. Numer. Anal.*, 25(4):753–783, 1988.

[Bandler and Chen, 1988] J. W. Bandler and S. H. Chen. Circuit optimization: the state of the art. *IEEE Trans. on Microwave Theory and Techniques*, 36(2):424–443, 1988.

[Barnett and Betcke, 2008] A. H. Barnett and T. Betcke. Stability and convergence of the method of fundamental solutions for Helmholtz problems on analytic domains. *J. Comp. Phys.*, 227:7003–7026, 2008.

[Barnett, 2009] A. H. Barnett. Perturbative analysis of the method of particular solutions for improved inclusion of high-lying Dirichlet eigenvalues. *SIAM J. Numer. Anal.*, 47(3):1952–1970, 2009.

[Barra and Gaspard, 1999] F. Barra and P. Gaspard. Scattering in periodic systems: from resonances to band structure. *J. Phys. A: Math. Gen.*, 32:3357–3375, 1999.

[Bauer *et al.*, 2008] C. A. Bauer, G. R. Werner, and J. R. Cary. Truncated photonic crystal cavities with optimized mode confinement. *J. Appl. Phys.*, 104(5):053107, 2008.

[Baumann *et al.*, 2005] D. Baumann, M. Gimersky, C. Fumeaux, and R. Vahldieck. Accuracy considerations of the FVTD method for radiating structures. In *European Microwave Conference*, volume 2, 2005.

[Beals, 1999] R. Beals. A note on fundamental solutions. *Commun. PDE*, 24(1):369–376, 1999.

[Beder and Orszag, 1999] C. M. Beder and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory.* Springer, 1999.

[Bertsimas *et al.*, 2007] D. Bertsimas, O. Nohadani, and K. M. Teo. Robust optimization in electromagnetic scattering problems. *J. Appl. Phys.*, 101(7):074507, 2007.

[Betcke and Trefethen, 2004] T. Betcke and L. N. Trefethen. Computations of eigenvalue avoidance in planar domains. *Proc. Appl. Math. Mech.*, 4:634–635, 2004.

[Betcke and Trefethen, 2005] T. Betcke and L. N. Trefethen. Reviving the method of particular solutions. *SIAM Review*, 47(3):469–491, 2005.

[Bhat and Afshari, 2008] H. S. Bhat and E. Afshari. Nonlinear constructive interference in electrical lattices. *Phys. Rev. E*, 77(1):066602, 2008.

[Bhat and Osting, 2008] H. S. Bhat and B. Osting. Thin slit diffraction in conventional and dual composite right/left-handed transmission line metamaterials. In *APMC Conf. Proc.*, Hong Kong, 2008.

[Bhat and Osting, 2009a] H. S. Bhat and B. Osting. Diffraction on the two-dimensional square lattice. *SIAM J. Appl. Math.*, 70(5):1389–1406, 2009.

[Bhat and Osting, 2009b] H. S. Bhat and B. Osting. Diffraction on the two-dimensional triangular lattice. preprint, 2009.

[Bhat and Osting, 2009c] H. S. Bhat and B. Osting. The zone boundary mode in periodic nonlinear electrical lattices. *Physica D*, 238:1216–1228, 2009.

[Bhat and Osting, 2010] H. S. Bhat and B. Osting. Discrete diffraction in two-dimensional transmission line metamaterials. *Microwave and Optical Technology Lett.*, 52(3):721–725, 2010.

[Bhat and Osting, 2011a] H. S. Bhat and B. Osting. Kirchhoff's laws as a finite volume method for the planar Maxwell equations. *IEEE Trans. on Antennas and Propagation*, 2011. accepted.

[Bhat and Osting, 2011b] H. S. Bhat and B. Osting. Two-dimensional lattice synthesis. submitted, 2011.

[Biegler *et al.*, 2007] L. T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders. *Real-Time PDE-Constrained Optimization*. SIAM, 2007.

[Bindel, 2006] D. Bindel. Theory and computation of resonances in 1d scattering. Personal website accessed 1/9/09, 2006.

[Borcea *et al.*, 2010] L. Borcea, V. Druskin, A. V. Mamonov, and F. G. Vasquez. Pyramidal resistor networks for electrical impedance tomography with partial boundary measurements. *Inverse Problems*, 26(10):105009, 2010.

[Born and Wolf, 1980] M. Born and E. Wolf. *Principles of Optics*. Pergamon Press, 6th (corrected) edition, 1980.

[Borwein and Lewis, 2000] J. M. Borwein and A. S. Lewis. *Convex Analysis and Nonlinear Optimization*. Springer-Verlag, 2000.

[Borzi and Schulz, 2009] A. Borzi and V. Schulz. Multigrid methods for PDE optimization. *SIAM Review*, 51(2):361–395, 2009.

[Bouwkamp, 1954] C. J. Bouwkamp. Diffraction theory. *Rep. Prog. Phys.*, 17:35–100, 1954.

[Boyd and Vandenberghe, 2004] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[Bratus and Myshkis, 1992] A. S. Bratus and A. D. Myshkis. Extremum problems for Laplacian eigenvalues with free boundary. *Nonlinear Analysis, Theory, Methods & Applications*, 19(9):815–831, 1992.

[Brayton *et al.*, 1979] R. K. Brayton, S. W. Director, G. D. Hachtel, and L. M. Vidigal. A new algorithm for statistical circuit design based on quasi-Newton methods and function splitting. *IEEE Trans. on Circuits and Systems*, 26(9):784–794, 1979.

[Brayton *et al.*, 1981] R. K. Brayton, G. D. Hachtel, and A. L. Sangiovanni-Vincentelli. A survey of optimization techniques for integrated-circuit design. *Proc. IEEE*, 69(10):1334–1362, 1981.

[Brillouin, 1946] L. Brillouin. *Wave Propagation in Periodic Structures*. McGraw-Hill, 1946.

[Brlek, 2005] S. Brlek. The discrete Green theorem and some applications in discrete geometry. *Theoretical Computer Science*, 346:200–225, 2005.

[Bruckstein and Kailath, 1987] A. M. Bruckstein and T. Kailath. Inverse scattering for discrete transmission-line models. *SIAM Review*, 29(3):359–389, 1987.

[Bucur and Buttazzo, 2005] D. Bucur and G. Buttazzo. *Variatonal Methods in Shape Optimization Problems*. Birkhäuser, 2005.

[Buneman, 1971] O. Buneman. Analytic inversion of the five-point Poisson operator. *J. Comp. Phys.*, 8:500–505, 1971.

[Burago and Zalgaller, 1988] Y. D. Burago and V. A. Zalgaller. *Geometric Inequalities*. Springer-Verlag, 1988.

[Burger and Osher, 2005] M. Burger and S. J. Osher. A survey on level set methods for inverse problems and optimal design. *Eur. J. Appl. Math.*, 16(02):263–301, 2005.

[Burger *et al.*, 2004] M. Burger, S. Osher, and E. Yablonovitch. Inverse problem techniques for the design of photonic crystals. *IEICE Trans. Electron.*, 87:258–265, 2004.

[Burkardt *et al.*, 2002] J. Burkardt, M. Gunzburger, and J. Peterson. Insensitive functionals, inconsistent gradients, spurious minima, and regularized functionals in flow optimization problems. *Internat. J. Comp. Fluid Dynamics*, 16(3):171–185, 2002.

[Busch *et al.*, 2007] K. Busch, G. von Freymann, S. Linden, S. F. Mingaleev, L. Tkeshelashvili, and M. Wegener. Periodic nanostructures for photonics. *Physics Reports*, 444:101–202, 2007.

[Caflisch, 1981] R. E. Caflisch. An inverse problem for Toeplitz matrices and the synthesis of discrete transmission lines. *Linear Algebra and its Applications*, 38:207–225, 1981.

[Caloz and Itoh, 2006] C. Caloz and T. Itoh. *Electromagnetic Metamaterials, Transmission Line Theory and Microwave Applications*. Wiley, 2006.

[Caloz and Nguyen, 2007] C. Caloz and H. V. Nguyen. Novel broadband conventional- and dual-composite right/left-handed (C/D-CRLH) metamaterials: properties, implementation and double-band coupler application. *Appl. Phys. A*, 87:309–316, 2007.

[Caloz, 2006] C. Caloz. Dual composite right/left-handed (D-CRLH) transmission line metamaterial. *IEEE Microwave and Wireless Components Lett.*, 16(11):585–587, 2006.

[Chechurin *et al.*, 2007] V. L. Chechurin, N. V. Korovkin, and M. Hayakawa. *Inverse Problems in Electric Circuits and Electromagnetics*. Springer, 2007.

[Cheng and Lu, 1991] S. S. Cheng and R. F. Lu. Discrete Wirtinger's inequalities and conditions for partial difference equations. *Fasciculi Mathematici*, 23:9–24, 1991.

[Cheng and Yang, 2007] Q.-M. Cheng and H. Yang. Bounds on eigenvalues of Dirichlet Laplacian. *Mathematische Annalen*, 337:159–175, 2007.

[Cherkaev, 2000] A. Cherkaev. *Variational Methods for Structural Optimization*. Springer-Verlag, 2000.

[Chicone, 2006] C. Chicone. *Ordinary Differential Equations with Applicaitons*. Springer, 2006.

[Ching *et al.*, 1998] E. S. C. Ching, P. T. Leung, A. Maassen van den Brink, W. M. Suen, S. S. Tong, and K. Young. Quasinormal-mode expansion for waves in open systems. *Rev. Mod. Phys.*, 70(4):1545–1554, 1998.

[Christopoulos, 1995] C. Christopoulos. *The Transmission-Line Modeling Method (TLM)*. IEEE Press, 1995.

[Christopoulos, 2006] C. Christopoulos. *The Transmission-Line Modeling (TLM) Method in Electromagnetics*. Morgan and Claypool Publishers, 2006.

[Chu and Golub, 2005] M. T. Chu and G. H. Golub. *Inverse Eigenvalue Problems: Theory, Algorithms, and Applications*. Oxford University Press, 2005.

[Chu *et al.*, 2008] Y. S. Chu, J. M. Yi, F. De Carlo, Q. Shen, W.-K. Lee, H. J. Wu, C. L. Wang, J. Y. Wang, C. J. Liu, C. H. Wang, S. R. Wu, C. C. Chien, Y. Hwu, A. Tkachuk, W. Yun, M. Feser, K. S. Liang, C. S. Yang, J. H. Je, and G. Margaritondo. Hard-x-ray microscopy with Fresnel zone plates reaches 40 nm rayleigh resolution. *Appl. Phys. Lett.*, 92(10):103119, 2008.

[Chung and Engquist, 2005] E. T. Chung and B. Engquist. Convergence analysis of fully discrete finite volume methods for Maxwell's equations in nonhomogeneous media. *SIAM J. Numer. Anal.*, 43(1):303–317, 2005.

[Chung *et al.*, 2003] E. T. Chung, Q. Du, and J. Zou. Convergence analysis of a finite volume method for Maxwell's equations in nonhomogeneous media. *SIAM J. Numer. Anal.*, 41(1):37–63, 2003.

[Coddington and Levinson, 1955] E. A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

[Cohen-Tannoudji *et al.*, 1992] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg. *Atom-Photon Interactions*. Wiley-Interscience, 1992.

[Cohen, 1964] D. S. Cohen. Separation of variables and alternative representations for non-selfadjoint boundary value problems. *Commun. Pure Appl. Math.*, 17:1–22, 1964.

[Colton and Kress, 1983] D. Colton and R. Kress. *Integral Equation Methods in Scattering Theory*. Wiley-Interscience, 1983.

[Colton and Kress, 1998] D. Colton and R. Kress. *Inverse Acoustic and Elecromagnetic Scattering Theory*. Springer, second edition, 1998.

[Costin and Soffer, 2001] O. Costin and A. Soffer. Resonance theory for Schrödinger operators. *Commun. Math. Phys.*, 224(1):133–152, 2001.

[Costin *et al.*, 2000] O. Costin, J. L. Lebowitz, and A. Rokhlenko. Exact reults for the ionization of model quantum system. *J. Phys. A: Math. Gen.*, 33:6311–6319, 2000.

[Costin *et al.*, 2001] O. Costin, R. D. Costin, J. L. Lebowitz, and A. Rokhlenko. Evolution of a model quantum system under time periodic forcing: condition for complete ionization. *Commun. Math. Phys.*, 221:1–26, 2001.

[Courant and Hilbert, 1953] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Interscience Publishers, Inc, 1953.

[Cox and Dobson, 1999] S. J. Cox and D. C. Dobson. Maximizing band gaps in two-dimensional photonic crystals. *SIAM J. Appl. Math.*, 59(6):2108–2120, 1999.

[Cox and Dobson, 2000] S. J. Cox and D. C. Dobson. Band structure optimization of two-dimensional photonic crystals in H-polarization. *J. Comp. Phys.*, 158(2):214–224, 2000.

[Cox and McLaughlin, 1990a] S. J. Cox and J. R. McLaughlin. Extremal eigenvalue problems for composite membranes, I. *Appl. Math. Optim.*, 22:153–167, 1990.

[Cox and McLaughlin, 1990b] S. J. Cox and J. R. McLaughlin. Extremal eigenvalue problems for composite membranes, II. *Appl. Math. Optim.*, 22:169–187, 1990.

[Cox and Ross, 1995] S. J. Cox and M. Ross. Extremal eigenvalue problems for starlike planar domains. *J. Diff. Eq.*, 120:174–197, 1995.

[Curtis and Morrow, 2000] E. B. Curtis and J. A. Morrow. *Inverse Problems for Electrical Networks*. World Scientific, 2000.

[Deconinck and Kutz, 2006] B. Deconinck and J. N. Kutz. Computing spectra of linear operators using the Floquet-Fourier-Hill method. *J. Comp. Phys.*, 219:296–321, 2006.

[Deift and Trubowitz, 1979] P. Deift and E. Trubowitz. Inverse scattering on the line. *Commun. Pure Appl. Math.*, 32:121–251, 1979.

[Delfour and Zolésio, 2001] M. C. Delfour and J. P. Zolésio. *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*. SIAM, 2001.

[Dickinson, 1984] B. W. Dickinson. An inverse problem for Toeplitz matrices. *Linear Algebra and its Applications*, 59:79–83, 1984.

[Dobson and Santosa, 2004] D. C. Dobson and F. Santosa. Optimal localization of eigenfunctions in an inhomogeneous medium. *SIAM J. Appl. Math.*, 64(3):762–774, 2004.

[Dolph *et al.*, 1966] C. L. Dolph, J. B. McLeod, and D. Thoe. The analytic continuation of the resolvent kernel and scattering operator associated with the Schroedinger operator. *J. Mathematical Analysis and Applications*, 16:311–332, 1966.

[Eastham, 1973] M. S. P. Eastham. *The Spectral Theory of Periodic Differential Equations*. Scottish Academic Press Ltd., 1973.

[Economou, 2006] E. N. Economou. *Green's Functions in Quantum Physics*. Springer-Verlag, 2006.

[Edelvik, 1999] F. Edelvik. Analysis of a finite volume solver for Maxwell's equations. In *Finite volumes for complex applications II*, pages 141–148. Hermes Sci. Publ., 1999.

[Engheta and Ziolkowski, 2006] N. Engheta and R. W. Ziolkowski. *Metamaterials: Physics and Engineering Explorations*. IEEE Press, 2006.

[Englund *et al.*, 2005] D. Englund, I. Fushman, and J. Vuckovic. General recipe for designing photonic crystal cavities. *Opt. Express*, 13(16):5961–5975, 2005.

[Engquist and Majda, 1977] B. Engquist and A. Majda. Absorbing boundary conditions for numerical simulation of waves. *Proc. Natl. Acad. Sci. USA*, 74(5):1765–1766, 1977.

[Evans, 2000] L. C. Evans. *Partial Differential Equations*. AMS, 2000.

[Felici *et al.*, 2010] M. Felici, K. A. Atlasov, A. Surrente, and E. Kapon. Semianalytical approach to the design of photonic crystal cavities. *Phys. Rev. B*, 82(11):115118, 2010.

[Felsen *et al.*, 2008] L. B. Felsen, M. Mongiardo, and P. Russer. *Electromagnetic Field Computation by Network Methods*. Springer-Verlag, 2008.

[Feng *et al.*, 2007] Y. Feng, M. Feser, A. Lyon, S. Rishton, X. Zeng, S. Chen, S. Sassolini, and W. Yun. Nanofabrication of high aspect ratio 24 nm x-ray zone plates for x-ray imaging applications. *J. Vac. Sci. B*, 25:2004–2007, 2007.

[Figotin and Klein, 1997] A. Figotin and A. Klein. Localized classical waves created by defects. *J. Stat. Phys.*, 86:165–177, 1997.

[Figotin and Klein, 1998] A. Figotin and A. Klein. Midgap defect modes in dielectric and acoustic media. *SIAM J. Appl. Math.*, 58(6):1748–1773, 1998.

[Fletcher and Rossing, 1998] N. H. Fletcher and T. D. Rossing. *The Physics of Musical Instruments*. Springer, second edition, 1998.

[Foulds, 1992] L. R. Foulds. *Graph Theory Applications*. Springer-Verlag, 1992.

[Fox *et al.*, 1967] L. Fox, P. Henrici, and C. B. Moler. Approximations and bounds for eigenvalues of elliptic operators. *SIAM J. Numer. Anal.*, 4:89–102, 1967.

[Frolik and Yagle, 1996] J. L. Frolik and A. E. Yagle. A discrete-time formulation for the variable wave speed scattering problem in two dimensions. *Inverse Problems*, 12:909–924, 1996.

[Frolik and Yagle, 1997] J. L. Frolik and A. E. Yagle. Forward and inverse scattering for discrete layered lossy and absorbing media. *IEEE Trans. on Circuits and Systems II*, 44:710–722, 1997.

[Fumeaux *et al.*, 2004] C. Fumeaux, D. Baumann, and R. Vahldieck. Advanced FVTD simulation of dielectric resonator antennas and feed structures. *ACES Journal*, 19:155–164, 2004.

[Gedney *et al.*, 1998] S. D. Gedney, J. A. Roden, N. K. Madsen, A. H. Mohammadian, W. F. Hall, V. Shankar, and C. Rowell. Explicit time-domain solutions of Maxwell's equations via generalized grids. In *Advances in Computational Electrodynamics*, pages 163–262. Artech House, 1998.

[Gelfand and Fomin, 1991] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover, 1991.

[Geremia *et al.*, 2002] J. M. Geremia, J. Williams, and H. Mabuchi. Inverse-problem approach to designing photonic crystals for cavity QED experiments. *Phys. Rev. E*, 66(6):066606, 2002.

[Gielen and Rutenbar, 2000] G. G. E. Gielen and R. A. Rutenbar. Computer-aided design of analog and mixed-signal integrated circuits. *Proc. IEEE*, 88(12):1825–1852, 2000.

[Golowich and Weinstein, 2005] S. E. Golowich and M. I. Weinstein. Scattering resonances of microstructures and homogenization theory. *SIAM Multiscale Model Simul.*, 3(3):477–521, 2005.

[Gondarenko and Lipson, 2008] A. Gondarenko and M. Lipson. Low modal volume dipole-like dielectric slab. *Optics Express*, 16(11):17689, 2008.

[Gondarenko *et al.*, 2006] A. Gondarenko, S. Preble, J. Robinson, L. Chen, H. Lipson, and M. Lipson. Spontaneous emergence of periodic patterns in a biologically inspired simulation of photonic structures. *Phys. Rev. Lett.*, 96(143904), 2006.

[Goodman, 2004] J. W. Goodman. *Introduction to Fourier Optics*. Roberts and Company, third edition, 2004.

[Griffiths and Steinke, 2001] D. J. Griffiths and C. A. Steinke. Waves in locally periodic media. *Am. J. Phys.*, 69(2):137–154, 2001.

[Guillouard *et al.*, 1999] K. Guillouard, M. F. Wong, V. F. Hanna, and J. Citerne. A new global time-domain electromagnetic simulator of microwave circuits including lumped elements based

on finite-element method. *IEEE Trans. on Microwave Theory and Techniques*, 47:2045–2049, 1999.

[Gwarek, 1985] W. K. Gwarek. Analysis of an arbitrarily-shaped planar circuit a time-domain approach. *IEEE Trans. on Microwave Theory and Techniques*, 33:1067–1072, 1985.

[Hachtel *et al.*, 1973] G. D. Hachtel, R. K. Brayton, and F. G. Gustavson. The sparse tableau approach to network analysis and design. *IEEE Trans. on Circuit Theory*, 18(1):101–113, 1973.

[Haeberly, 1991] J. P. Haeberly. On shape optimizing the ratio of the first two eigenvalues of the Laplacian. Computer Science Technical Report 586, Courant Institute, New York Univ., 1991.

[Hägglund, 2006] R. Hägglund. *An Optimization-Based Approach to Efficient Design of Analog Circuits*. PhD thesis, Linköping University Institute of Technology, 2006.

[Harrell and Hermi, 2008] E. M. Harrell and L. Hermi. Differential inequalities for Riesz means and Weyl-type bounds for eigenvalues. *J. Funct. Anal.*, 254(12):3171–3191, 2008.

[Harrell and Svirsky, 1986] E. M. Harrell and R. Svirsky. Potentials producing maximally sharp resonances. *Trans. of the AMS*, 293(2):723–736, 1986.

[Harrell, 1980] E. M. Harrell. Double wells. *Commun. Math. Phys.*, 75(3):239–261, 1980.

[Harrell, 1982] E. M. Harrell. General lower bounds for resonances in one dimension. *Commun. Math. Phys*, 86:221–225, 1982.

[Haslinger and Mäkinen, 2003] J. Haslinger and R. A. E. Mäkinen. *Introduction to Shape Optimization*. SIAM, 2003.

[Heider *et al.*, 2008] P. Heider, D. Berebichez, R. V. Kohn, and M. I. Weinstein. Optimization of scattering resonances. *Structural and Multidisciplinary Optimization*, 36:443–456, 2008.

[Henrot, 2006] A. Henrot. *Extremum Problems for Eigenvalues of Elliptic Operators*. Birkhäuser Verlag, 2006.

[Henry, 2005] D. Henry. *Perturbation of the Boundary in Boundary-Value Problems of Partial Differential Equations*. Cambridge University Press, 2005.

[Hermeline *et al.*, 2008] F. Hermeline, S. Layouni, and P. Omnes. A finite volume method for the approximation of Maxwell's equations in two space dimensions on arbitrary meshes. *J. Comp. Phys.*, 227(22):9365–9388, 2008.

[Hermeline, 2004] F. Hermeline. A finite volume method for solving Maxwell equations in inhomogeneous media on arbitrary meshes. *Comptes Rendus Mathématique. Académie des Sciences*, 339(12):893–898, 2004.

[Hershenson *et al.*, 2001] M. D. Hershenson, S. P. Boyd, and T. H. Lee. Optimal design of a CMOS op-amp via geometric programming. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 20(1):1–21, 2001.

[Hislop and Sigal, 1996] P. D. Hislop and I. M. Sigal. *Introduction to Spectral Theory.* Springer, 1996.

[Hochstadt, 1965] H. Hochstadt. On the determination of a Hill's equation from its spectrum. *Arch. Rational Mech. Anal.*, 19:353–362, 1965.

[Hoefer and Weinstein, 2010] M. A. Hoefer and M. I. Weinstein. Defect modes and homogenization of periodic Schrödinger operators. *SIAM J. Math. Anal.*, 2010. in press.

[Hoefer, 1985] W. J. R. Hoefer. The transmission-line matrix method—theory and applications. *IEEE Trans. on Microwave Theory and Techniques*, 33:882–893, 1985.

[Hsiao and Wendland, 2008] G. C. Hsiao and W. L. Wendland. *Boundary Integral Equations.* Springer, 2008.

[Iantchenko, 2006] A. Iantchenko. Resonance spectrum for one-dimenional layered media. *Applicable Analysis*, 85(11):1383–1410, 2006.

[Islami and Vainberg, 2006] H. Islami and B. Vainberg. Large time behavior of solutions to difference wave operators. *Commun. PDE*, 31(3):397–416, 2006.

[Joannopoulos *et al.*, 2008] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade. *Photonic Crystals: Molding the Flow of Light.* Princeton University Press, second edition, 2008.

[Jones, 1969] A. R. Jones. The focal properties of phase zone plates. *Br. J. App. Phys.*, 2(12):1789–91, 1969.

[Kac, 1966] M. Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23, 1966.

[Kang *et al.*, 2008] H. C. Kang, H. Yan, R. P. Winarski, M. V. Holt, J. Maser, C. Liu, R. Conley, S. Vogt, A. T. Macrander, and G. B. Stephenson. Focusing of hard x-rays to 16 nanometers with a multilayer Laue lens. *Appl. Phys. Lett.*, 92:221114, 2008.

[Kao and Santosa, 2008] C.-Y. Kao and F. Santosa. Maximization of the quality factor of an optical resonator. *Wave Motion*, 45(4):412–427, 2008.

[Kao *et al.*, 2005] C.-Y. Kao, S. Osher, and E. Yablonovitch. Maximizing band gaps in two-dimensional photonic crystals using level set methods. *Appl. Phys. B*, 81:235–244, 2005.

[Karabash, 2011] I. M. Karabash. Optimization of quasi-normal eigenvalues for 1-d wave equations in inhomogeneous media; description of optimal structures. http://arxiv.org/abs/1103.4117, 2011.

[Kato, 1980] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, second edition, 1980.

[Katsura and Inawashiro, 1971] S. Katsura and S. Inawashiro. Lattice Green's functions for the rectangular and the square lattices at arbitrary points. *J. Math. Phys.*, 12(8):1622–1630, 1971.

[Kazimirov *et al.*, 2009] A. Kazimirov, V. G. Kohn, and Z.-H. Cai. New imaging technique based on diffraction of a focused x-ray beam. *J. Phys. D*, 42(1):012005, 2009.

[Keller, 1979] J. B. Keller. Progress and prospects in the theory of linear wave propagation. *SIAM Review*, 21(2):229–245, 1979.

[Kelvin, 1894] W. T. Kelvin. Isoperimetrical problems. In *Popular Lectures and Addresses*, volume 2. Macmillan and Co., 1894.

[Kevrekidis and Porter, 2009] P. G. Kevrekidis and M. A. Porter. Experimental results related to discrete nonlinear Schrödinger equations. In *The Discrete Nonlinear Schrödinger Equation*, volume 232, pages 175–189. Springer Berlin / Heidelberg, 2009.

[Khabou *et al.*, 2007] M. A. Khabou, L. Hermi, and M. B. H. Rhouma. Shape recognition using eigenvalues of the Dirichlet Laplacian. *Pattern Recognition*, 40(1):141–153, 2007.

[Kinsler *et al.*, 2000] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders. *Fundamentals of Acoustics*. John Wiley & Sons, fourth edition, 2000.

[Kirr and Weinstein, 2001] E. Kirr and M. I. Weinstein. Parametrically excited Hamiltonian partial differential equations. *SIAM J. Appl. Math.*, 33(1):16–52, 2001.

[Kirr and Weinstein, 2003] E. Kirr and M. I. Weinstein. Metastable states in parametrically excited multimode Hamiltonian systems. *Commun. Comp. Phys.*, 236(2):335–372, 2003.

[Kirsch, 1996] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, 1996.

[Ko *et al.*, 1990] Y. Ko, N. Yoshida, and I. Fukai. Three-dimensional analysis of a cylindrical waveguide converter for circular polarization by the spatial network method. *IEEE Trans. on Microwave Theory and Techniques*, 38:912–918, 1990.

[Kohn and Kazimirov, 2007] V. G. Kohn and A. Kazimirov. Simulations of Bragg diffraction of a focused x-ray beam by a single crystal with an epitaxial layer. *Phys. Rev. B*, 75(22):224119, 2007.

[Koza *et al.*, 1997] J. R. Koza, F. H. Bennett, D. Andre, M. A. Keane, and F. Dunlap. Automated synthesis of analog electrical circuits by means of genetic programming. *IEEE Trans. Evol. Comput.*, 1(2):109–128, 1997.

[Krauss, 2008] T. F. Krauss. Why do we need slow light? *Nat. Photon.*, 2(8):448–450, 2008.

[Krein, 1955] M. G. Krein. On certain problems on the maximum and minimum of characteristic valuesand on the lyapunov zones of stability. *AMS Translations Ser.*, 2(1):163–187, 1955.

[Krohne *et al.*, 2007] K. Krohne, D. Baumann, C. Fumeaux, E.-P. Li, and R. Vahldieck. Frequency-domain finite-volume simulations. In *European Microwave Conference*, pages 158–161, 2007.

[Kron, 1944] G. Kron. Equivalent circuit of the field equations of Maxwell—I. *Proc. IRE*, 32(5):289–299, 1944.

[Kuttler and Sigillito, 1984] J. R. Kuttler and V. G. Sigillito. Eigenvalues of the Laplacian in two dimensions. *SIAM Review*, 26(2):163–193, 1984.

[Labreuche, 1998] C. Labreuche. Purely imaginary resonant frequencies for a lossy inhomogeneous medium. *SIAM J. Appl. Math.*, 59(2):725–742, 1998.

[Lager *et al.*, 2003] I. E. Lager, E. Tonti, A. T. de Hoop, G. Mur, and M. Marrone. Finite formulation and domain-integrated field relations in electromagnetics—a synthesis. *IEEE Trans. on Magnetics*, 39:1199–1202, 2003.

[Lamb, 1900] H. Lamb. On a peculiarity of the wave-system due to the free vibrations of a nucleus in an extended medium. *Proc. London Math. Soc.*, 32:208–211, 1900.

[Laporte and Tallec, 2003] E. Laporte and P. Le Tallec. *Numerical Methods in Sensitivity Analysis and Shape Optimization.* Birkhäuser, 2003.

[Larsson and Thomée, 2003] S. Larsson and V. Thomée. *Partial Differential Equations with Numerical Methods.* Texts in Applied Mathematics. Springer, 2003.

[Lax and Phillips, 1989] P. D. Lax and R. S. Phillips. *Scattering Theory.* Academic Press, 1989.

[Leung *et al.*, 1994a] P. T. Leung, S. Y. Liu, S. S. Tong, and K. Young. Time-independent perturbation theory for quasinormal modes in leaky optical cavities. *Phys. Rev. A*, 49(4):3068–3073, 1994.

[Leung *et al.*, 1994b] P. T. Leung, S. Y. Liu, and K. Young. Completeness and orthogonality of quasinormal modes in leaky optical cavities. *Phys. Rev. A*, 49(4):3057–3067, 1994.

[Leung *et al.*, 1994c] P. T. Leung, S. Y. Liu, and K. Young. Completeness and time-independent perturbation of the quasinormal modes of an absorptive and leaky cavity. *Phys. Rev. A*, 49(5):3982–3989, 1994.

[LeVeque, 2002] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.

[Levitin and Yagudin, 2003] M. Levitin and R. Yagudin. Range of the first three eigenvalues of the planar Dirichlet Laplacian. *LMS J. Comput. Math.*, 6:1–17, 2003.

[Lewis and Overton, 1996] A. S. Lewis and M. L. Overton. Eigenvalue optimization. *Acta Numerica*, 5:149–190, 1996.

[Lewis and Overton, 2009] A. S. Lewis and M. L. Overton. Nonsmooth optimization via BFGS. submitted to SIAM J. Optim., 2009.

[Li *et al.*, 2007] X. Li, P. Gopalakrishnan, Y. Xu, and L. T. Pileggi. Robust analog/RF circuit design with projection-based performance modeling. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, 26(1):2–15, 2007.

[Lieb and Loss, 2001] E. H. Lieb and M. Loss. *Analysis*. AMS, second edition, 2001.

[Lilis *et al.*, 2010] G. N. Lilis, J. Park, W. Lee, G. Li, H. S. Bhat, and E. Afshari. Harmonic generation using nonlinear LC lattices. *IEEE Trans. on Microwave Theory and Techniques*, 58(7):1713–1723, 2010.

[Lions, 1971] J. L. Lions. *Optimal Control of Systems Goerned by Partial Differential Equations*. Springer-Verlag, 1971.

[Lipton *et al.*, 2003] R. P. Lipton, S. P. Shipman, and S. Venakides. Optimization of resonances in photonic crystal slabs. In *Proc. SPIE 5184*, pages 168–177, 2003.

[Madsen and Ziolkowski, 1988] N. K. Madsen and R. W. Ziolkowski. Numerical solution of Maxwell's equations in the time domain using irregular nonorthogonal grids. *Wave Motion*, 10(6):583–596, 1988.

[Madsen and Ziolkowski, 1990] N. K. Madsen and R. W. Ziolkowski. A three-dimensional modified finite volume technique for Maxwell's equations. *Electromagnetics*, 10(1):147–161, 1990.

[Magnus and Winkler, 1966] W. Magnus and S. Winkler. *Hill's Equation*. John Wiley & Sons, 1966.

[Maksimovic, 2008] M. Maksimovic. Coupled optical defect microcavities in one-dimensional photonic crystals and quasi-normal modes. *Optical Engineering*, 47, 2008.

[Marcuse, 1974] D. Marcuse. *Theory of Dielectric Optical Waveguides.* Academic Press, 1974.

[Marqués *et al.*, 2008] R. Marqués, F. Martín, and M. Sorolla. *Metamaterials with Negative Parameters: Theory, Design, and Microwave Applications.* Wiley, 2008.

[Martens and Gielen, 2008] E. S. J. Martens and G. G. E. Gielen. *High-Level Modeling and Synthesis of Analog Integrated Systems.* Springer, 2008.

[Martin, 2006] P. A. Martin. Discrete scattering theory: Green's function for a square lattice. *Wave Motion*, 43(7):619–629, 2006.

[Maser *et al.*, 2004a] J. Maser, G. B. Stephenson, D. Shu, B. Lai, S. Vogt, A. Khounsary, Y. Li, C. Benson, and G. Schneider. Conceptual design for a beamline for a hard x-ray nanoprobe with 30 nm spatial resolution. In *Synchrotron Radiation Instrumentation: Eighth International Conference on Synchrotron Radiation Instrumentation*, volume 705, pages 470–473, 2004.

[Maser *et al.*, 2004b] J. Maser, G. B. Stephenson, S. Vogt, W. Yun, A. Macrander, H. C. Kang, C. Liu, and R. Conley. Multilayer Laue lenses as high-resolution x-ray optics. *Proc. SPIE 5539*, page 185, 2004.

[McCartin, 2003] B. J. McCartin. Eigenstructure of the equilateral triangle, part 1: The Dirichlet problem. *SIAM Review*, 45(2):267–287, 2003.

[Melrose, 1995] R. B. Melrose. *Geometric Scattering Theory.* Cambridge University Press, 1995.

[Men *et al.*, 2010] H. Men, N. C. Nguyen, R. M. Freund, P. A. Parrilo, and J. Peraire. Band gap optimization of two-dimensional photonic crystals using semidefinite programming and subspace methods. *J. Comp. Phys.*, 229(10):3706–3725, 2010.

[Mimura *et al.*, 2007] H. Mimura, H. Yumoto, S. Matsuyama, Y. Sano, K. Yamamura, Y. Mori, M. Yabashi, Y. Nishino, K. Tamasaku, T. Ishikawa, and K. Yamauchi. Efficient focusing of hard x rays to 25 nm by a total reflection mirror. *Appl. Phys. Lett.*, 90(5):051903, 2007.

[Moler and Payne, 1968] C. B. Moler and L. E. Payne. Bounds for eigenvalues and eigenfunctions of symmetric operators. *SIAM J. Numer. Anal.*, 5:64–70, 1968.

[Momeni and Afshari, 2008] O. Momeni and E. Afshari. Electrical prism: a high quality factor filter for mm wave and terahertz frequencies. In *Proc. of the Asia-Pacific Microwaves Conference*, Hong Kong, 2008.

[Momeni and Afshari, 2009] O. Momeni and E. Afshari. Electrical prism: A high quality factor filter for millimeter-wave and terahertz frequencies. *IEEE Trans. Microwwave Theory and Techniques*, 57(11):2790–2799, 2009.

[Morita, 1971] T. Morita. Useful procedure for computing the lattice Green's function—square, tetragonal, and bcc lattices. *J. Math. Phys.*, 12(8):1744–1747, 1971.

[Morse and Feshbach, 1953] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics*. McGraw-Hill, 1953.

[Nedelec, 2001] J.-C. Nedelec. *Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems*. Springer, 2001.

[Nocedal and Wright, 2006] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, second edition, 2006.

[Nye *et al.*, 1988] W. Nye, D. C. Riley, A. Sangiovanni-Vincentelli, and A. L. Tits. DELIGHT.SPICE: an optimization-based system for the design of integrated circuits. *IEEE Trans. Comput.-Aided Design*, 7(4):501–519, 1988.

[Okada *et al.*, 2005] Y. Okada, A. Shudo, S. Tasaki, and T. Harayama. 'can one hear the shape of a drum?': revisited. *J. Phys. A*, 38:163–170, 2005.

[Olenšek *et al.*, 2009] J. Olenšek, Á. Bűrmen, J. Puhan, and T. Tuma. Automated analog electronic circuits sizing. In A. Qing, editor, *Differential Evolution*, pages 353–367. Wiley, 2009.

[Osher and Santosa, 2001] S. J. Osher and F. Santosa. Level set methods for optimization problems involving geometry and constraints 1. frequencies of a two-density inhomogeneous drum. *J. Comp. Phys.*, 171:272–288, 2001.

[Osserman, 1978] R. Osserman. The isoperimetric inequality. *Proc. Amer. Math. Soc.*, 84(6):1182–1238, 1978.

[Osting and Bhat, 2008] B. Osting and H. S. Bhat. Dispersive diffraction in a two-dimensional hexagonal transmission lattice. In *Proc. International Symposium on Antennas and Propagation*, Taipei, Taiwan, 2008.

[Osting and Weinstein, 2011a] B. Osting and M. I. Weinstein. Emergence of periodic structure from maximizing the lifetime of a bound state coupled to radiation. *SIAM Multiscale Model Simul.*, 2011. in press.

[Osting and Weinstein, 2011b] B. Osting and M. I. Weinstein. Long-lived resonant states of the Helmholtz equation. in preparation, 2011.

[Osting, 2010] B. Osting. Optimization of spectral functions of Dirichlet-Laplacian eigenvalues. *J. Comp. Phys.*, 229(22):8578–8590, 2010.

[Oudet, 2004] E. Oudet. Numerical minimization of eigenmodes of a membrane with respect to the domain. *ESAIM COCV*, 10:315–335, 2004.

[Overton, 2010] M. L. Overton. BFGS quasi-Newton minimization algorithm, version 2.0. http://www.cs.nyu.edu/overton/software/hanso, 2010.

[Parzygnat *et al.*, 2010] A. Parzygnat, K. K. Y. Lee, Y. Avniel, and S. G. Johnson. Sufficient conditions for two-dimensional localization by arbitrarily weak defects in periodic potentials with band gaps. *Phys. Rev. B*, 81:155324, 2010.

[Payne *et al.*, 1956] L. E. Payne, G. Pólya, and H. F. Weinberger. On the ratio of consecutive eigenvalues. *J. Math. and Phys.*, 35:289–298, 1956.

[Payne, 1967] L. E. Payne. Isoperimetric inequalities and their applications. *SIAM Review*, 9(3):453–488, 1967.

[Perry, 2002] P. A. Perry. Introduction to scattering theory. Lecutre Notes, 2002.

[Pinsky, 1980] M. A. Pinsky. The eigenvalues of an equilateral triangle. *SIAM J. Math. Anal.*, 11(5):819–827, 1980.

[Pironneau, 1984] O. Pironneau. *Optimal Shape Design for Elliptic Systems*. Springer-Verlag, 1984.

[Pitaveskii and Stringari, 2003] L. P. Pitaveskii and S. Stringari. *Bose Einstein Condensation*. Oxford University Press, 2003.

[Pöschel and Trubowitz, 1987] J. Pöschel and E. Trubowitz. *Inverse Spectral Theory*. Academic Press, 1987.

[Purcell, 1946] E. M. Purcell. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.*, 69:681, 1946.

[Purcell, 1952] E. M. Purcell. Research in nuclear magnetism. Nobel Lecture, 1952.

[Ralston, 1972] J. V. Ralston. Variation of the transmission coefficient and comparison theorems for the purely imaginary poles of the scattering matrix. *Commun. Pure Appl. Math.*, 25(1):45–61, 1972.

[Ramdani and Shipman, 2008] K. Ramdani and S. Shipman. Trnamission through a thick periodic slab. *Math. Models and Meth. in Appl. Sci.*, 18(4):543–572, 2008.

[Reed and Simon, 1980] M. Reed and B. Simon. *Methods of Modern Mathematical Physics*. Academic Press, 1980.

[Rellich, 1969] F. Rellich. *Perturbation Theory of Eigenvalue Problems*. Gordon and Breach Science Publishers, 1969.

[Rosenau, 1986] P. Rosenau. Dynamics of nonlinear mass-spring chains near the continuum limit. *Phys. Lett. A*, 118:222–227, 1986.

[Rutenbar *et al.*, 2007] R. A. Rutenbar, G. G. E. Gielen, and J. Roychowdhury. Hierarchical modeling, optimization, and synthesis for system-level analog and RF designs. *Proc. IEEE*, 95(3):640–669, 2007.

[Sanada *et al.*, 2004] A. Sanada, C. Caloz, and T. Itoh. Characteristics of the composite right/left-handed transmission lines. *IEEE Microwave and Wireless Components Lett.*, 14(2):68–70, 2004.

[Sarychev and Shalaev, 2007] A. K. Sarychev and V. M. Shalaev. *Electrodynamics of Metamaterials*. World Scientific, Singapore, 2007.

[Satoh *et al.*, 2006] H. Satoh, N. Yoshida, S. Kitayama, and S. Konaka. Analysis of 2-D frequency converter utilizing compound nonlinear photonic-crystal structure by condensed node spatial network method. *IEEE Trans. on Microwave Theory and Techniques*, 54:210–215, 2006.

[Scheuer *et al.*, 2006] J. Scheuer, W. M. J. Green, and A. Yariv. Annular Bragg resonators (ABR) - the ideal tool for biochemical sensing, nonlinear optics, and cavity QED. *Proc. SPIE 6123*, 2006.

[Schmalz *et al.*, 2010] J. A. Schmalz, G. Schmalz, T. E. Gureyev, and K. M. Pavlov. On the derivation of the Green's function for the Helmholtz equation using generalized functions. *Am. J. Phys.*, 78(2):181–186, 2010.

[Schroer *et al.*, 2005] C. G. Schroer, O. Kurapova, J. Patommel, P. Boye, J. Feldkamp, B. Lengeler, M. Burghammer, C. Riekel, L. Vincze, A. van der Hart, and M. Kuchler. Hard x-ray nanoprobe based on refractive x-ray lenses. *Appl. Phys. Lett.*, 87(12):124103, 2005.

[Schultz, 1998] P. Schultz. The wave equation on the lattice in two and three dimensions. *Commun. Pure Appl. Math.*, 51(6):663–695, 1998.

[Settimi *et al.*, 2003] A. Settimi, S. Severini, N. Mattiucci, C. Sibilia, M. Centini, G. D'Aguanno, M. Bertolotti, M. Scalora, M. Bloemer, and C. M. Bowden. Quasinormal-mode description of waves in one-dimensional photonic crystals. *Phys. Rev. E*, 68(2):026614, 2003.

[Settimi *et al.*, 2009] A. Settimi, S. Severini, and B. J. Hoenders. Quasi-normal-modes description of transmission properties for photonic bandgap structures. *J. Opt. Soc. Am. B*, 26(4):876–891, 2009.

[Shaban and Vainberg, 2001] W. Shaban and B. Vainberg. Radiation conditions for the difference Schrödinger operators. *Applicable Analysis*, 80(3):525–556, 2001.

[Shankar *et al.*, 1989] V. Shankar, W. F. Hall, and A. H. Mohammadian. A time-domain differential solver for electromagnetic scattering problems. *Proc. IEEE*, 77(5):709–721, 1989.

[Shankar *et al.*, 1990] V. Shankar, A. H. Mohammadian, and W. F. Hall. A time-domain, finite-volume treatment for the Maxwell equations. *Electromagnetics*, 10:127–145, 1990.

[Shenk and Thoe, 1972] N. Shenk and D. Thoe. Resonant states and poles of the scattering matrix for perturbations of -Δ. *J. Math. Anal. Appl.*, 37(2):467–491, 1972.

[Sigmund and Hougaard, 2008] O. Sigmund and K. Hougaard. Geometric properties of optimal photonic crystals. *Phys. Rev. Lett.*, 100:153904, 2008.

[Sigmund and Jensen, 2003] O. Sigmund and J. S. Jensen. Systematic design of phononic band-gap materials and structures by topology optimization. *Phil. Trans. R. Soc. Lond. A*, 361(1806):1001–1019, 2003.

[Simon, 1991] B. Simon. Fifty years of eigenvalue perturbation theory. *Bulletin of the AMS*, 24(2):303–319, 1991.

[Soffer and Weinstein, 1998] A. Soffer and M. I. Weinstein. Nonautonomous Hamiltonians. *J. Stat. Phys.*, 93:359–391, 1998.

[Stein and Weiss, 1971] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, 1971.

[Steinbach, 2008] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems*. Springer, 2008.

[Strang, 2007] G. Strang. *Computational Science and Engineering*. Wellesley-Cambridge Press, 2007.

[Stutzman and Thiele, 1998] W. L. Stutzman and G. A. Thiele. *Antenna Theory and Design*. John Wiley, 1998.

[Svirsky, 1987] R. Svirsky. Maximally resonant potentials subject to p-norm constraints. *Pacific J. Math.*, 129(2):357–374, 1987.

[Tang and Zworski,] S.-H. Tang and M. Zworski. Potential scattering on the real line. Lecture notes. See http://math.berkeley.edu/$\tilde{z}$worski/tz1.pdf.

[Tang and Zworski, 2000] S.-H. Tang and M. Zworski. Resonance expansions of scattered waves. *Commun. Pure Appl. Math.*, 53(10):1305–1334, 2000.

[Taylor, 1996] M. E. Taylor. *Partial Differential Equations 1.* Springer, 1996.

[Teschl, 2009] G. Teschl. *Mathematical Methods in Quantum Mechanics With Application to Schrodinger Operators.* AMS, 2009.

[Tisseur and Meerbergen, 2001] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Review*, 43(2):235–286, 2001.

[Tousi and Afshari, 2010] Y. M. Tousi and E. Afshari. 2-d electrical interferometer: A novel high-speed quantizer. *IEEE Trans. Microwwave Theory and Techniques*, 58(10):2549–2561, 2010.

[Trefethen and Betcke, 2006] L. N. Trefethen and T. Betcke. Computed eigenmodes of planar regions. *Contemporary Mathematics*, 412:297–314, 2006.

[Trefethen and Embree, 2005] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators.* Princeton University Press, 2005.

[Trefethen, 2000] L. N. Trefethen. *Spectral Methods in Matlab.* SIAM, 2000.

[Visser, 2009] T. D. Visser. Whose golden rule is it anyway? *Am. J. Phys.*, 77(6):487–488, 2009.

[Wang *et al.*, 2002] Z. J. Wang, A. J. Przekwas, and Y. Liu. A FV-TD electromagnetic solver using adaptive Cartesian grids. *Computer Phys. Commun.*, 148(1):17–29, 2002.

[Whinnery and Ramo, 1944] J. R. Whinnery and S. Ramo. A new approach to the solution of high-frequency field problems. *Proc. IRE*, 32:284–288, 1944.

[Whinnery *et al.*, 1944] J. R. Whinnery, C. Concordia, W. Ridgway, and G. Kron. Network analyzer studies of electromagnetic cavity resonators. *Proc. IRE*, 32:360–367, 1944.

[Whitham, 1974] G. B. Whitham. *Linear and Nonlinear Waves.* Pure and Applied Mathematics. Wiley-Interscience, 1974.

[Wing, 2008] O. Wing. *Classical Circuit Theory.* Springer-Verlag, 2008.

[Wolf and Keller, 1994] S. A. Wolf and J. B. Keller. Range of the first two eigenvalues of the Laplacian. *Proc. R. Soc. Lond. A*, 447:397–412, 1994.

[Yagle and Frolik, 1996] A. E. Yagle and J. L. Frolik. On the feasibility of impulse reflection response data for the two-dimensional inverse scattering problem. *IEEE Trans. on Antennas and Propagation*, 44:1551–1564, 1996.

[Yan *et al.*, 2007] H. Yan, O. Kalenci, and I. C. Noyan. Diffraction profiles of elastically bent single crystals with constant strain gradients. *J. Appl. Cryst.*, 40(2):322–331, 2007.

[Yan *et al.*, 2008] H. Yan, O. Kalenci, C. I. Noyan, and J. Maser. Coherency effects in nanobeam x-ray diffraction analysis. *J. Appl. Phys.*, 104(2):023506, 2008.

[Yan, 2009] H. Yan. X-ray dynamical diffraction from multilayer Laue lenses with rough interfaces. *Phys. Rev. B*, 79(16):165410, 2009.

[Yang and Albregtsen, 1994] L. Yang and F. Albregtsen. Fast and exact computation of moments using discrete Green's theorem. In *Norwegian Image Processing and Pattern Recognition*, pages 82–90, 1994.

[Yeh *et al.*, 1978] P. Yeh, A. Yariv, and E. Marom. Theory of Bragg fiber. *J. Opt. Soc. Am.*, 68(9):1196–1201, 1978.

[Yeh, 1988] P. Yeh. *Optical Waves in Layered Media.* John Wiley & Sons, 1988.

[Ying *et al.*, 2009] A. J. Ying, C. E. Murray, and I. C. Noyan. A rigorous comparison of X-ray diffraction thickness measurement techniques using silicon-on-insulator thin films. *J. Appl. Cryst.*, 42(3):401–410, 2009.

[Ying *et al.*, 2010] A. Ying, B. Osting., I. C. Noyan, C. E. Murray, M. Holt, and J. Maser. Modeling of kinematical diffraction from a thin silicon film illuminated by a coherent, focused x-ray nanobeam. *J. Appl. Cryst.*, 43:587–595, 2010.

[Zemla, 1995] A. Zemla. On the fundamental solutions for the difference Helmholtz operator. *SIAM J. Numer. Anal.*, 32(2):560–570, 1995.

[Zou, 2009] J. Zou. *Hierarchical Optimization of Large-Scale Analog/Mixed-Signal Circuits Based-on Pareto-Optimal Fronts.* PhD thesis, Technische Universität München, 2009.

[Zworski, 1999] M. Zworski. Resonance in phyiscs and geometry. *Notices of the AMS*, 1999.

Part V

Appendices

Appendix A

Spectral theory and wave propagation

Since the subject of much of this thesis is the propagation of waves through inhomogeneous and discrete materials, one of the objectives of this appendix is to discuss the solution of the Schrödinger equation and wave equation in the simplest of settings: continuous, homogeneous media in $\mathbb{R}^d$ for $d = 1, 2, 3$. The Cauchy initial value problem for each of these quintessential equations is written

$$\begin{aligned} \imath\partial_t\phi &= H\phi \qquad \text{(Schrödinger Eq.)} \\ \phi(0) &= \phi_0 \end{aligned}$$

and

$$\begin{aligned} \partial_t^2 u &= -Hu \qquad \text{(Wave Eq.)} \\ u(0) &= f \\ u_t(0) &= g. \end{aligned}$$

Here, $H = -\Delta = -\sum_{j=1}^d \partial_{x_j}^2$ is the Laplacian and the data ϕ_0, f, and g are taken to be smooth, localized functions.

The solutions to the Schrödinger and wave equations may be obtained by the Fourier transform and its inverse, for which we use the conventions

$$\begin{aligned} \hat{f}(\xi) = \mathfrak{F}[f](\xi) &:= \int_{\mathbb{R}^d} f(x) e^{-\imath\xi\cdot x}\,\mathrm{d}x \\ f(x) = \mathfrak{F}^{-1}[\hat{f}](x) &:= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(\xi) e^{\imath\xi\cdot x}\,\mathrm{d}\xi. \end{aligned}$$

Formally, the Fourier transform diagonalizes H, i.e. $\mathfrak{F}[Hu] = \|\xi\|^2\mathfrak{F}[u]$ so that the solutions to the Schrödinger and Wave Eqs. are written

$$\phi(t,x) = \mathfrak{F}^{-1}\left[\exp(-\imath|\xi|^2 t)\hat{\phi}_0(\xi)\right](x) \tag{A.1a}$$

$$u(t,x) = \mathfrak{F}^{-1}\left[\cos(|\xi|t)\hat{f}(\xi) + \frac{1}{|\xi|}\sin(|\xi|t)\hat{g}(\xi)\right](x). \tag{A.1b}$$

This derivation may be made rigorous (see e.g. [Stein and Weiss, 1971]). However, since we are interested in analyzing operators which are not diagonalized by the Fourier transform, in Sec. A.1, we discuss tools from spectral theory which may be used to achieve this result in greater generality. In Sec. A.2, we discuss the solutions to the Schrödinger and wave equations in a homogeneous media. Finally, in Sec. A.3 we discuss the inhomogeneous Schrödinger and wave equations and introduce the scattering resonance expansion.

A.1 The outgoing resolvent and spectral decomposition of the Laplacian

We begin with a brief review of spectral theory, summarizing (without proofs) results necessary to describe the outgoing resolvent and spectral decomposition of the Laplacian. More complete presentations of the content here may be found in [Agmon, 1975; Hislop and Sigal, 1996; Teschl, 2009; Melrose, 1995] and in one-dimension, [Tang and Zworski,].

A.1.1 A brief review of spectral theory

Let X be a Hilbert space with inner product $\langle\cdot,\cdot\rangle\colon X\times X\to\mathbb{C}$ and $A\colon D(A)\subset X\to X$ be a linear operator. We define the *resolvent set*, denoted $\rho(A)$, to be the set of values $\lambda\in\mathbb{C}$ such that the *resolvent operator*

$$R_A(\lambda) := (A-\lambda)^{-1}$$

exists as a function on X and the *spectrum* of A to be

$$\sigma(A) := \mathbb{C}\setminus\rho(A).$$

Thus, $R_A(\lambda)$ is an analytic, operator-valued function on $\rho(A)$. Generally speaking, there are three ways in which $A-\lambda$ can fail to be invertible:

1. $\ker(A-\lambda) \neq \{0\}$. In this case, λ is called an *eigenvalue* of A and any $x \in \ker(A-\lambda)$ are called *eigenfunctions*. The *geometric multiplicity* of λ is the dimension of $\ker(A-\lambda)$. The *algebraic multiplicity* of λ is the largest n such that $(A-\lambda)^n\psi = 0$ for some generalized eigenvector ψ and $(A-\lambda)^{n-1}\psi \neq 0$. The geometric multiplicity of an eigenvalue is always less than or equal to its algebraic multiplicity. We denote by $\sigma_d(A)$ the set of all eigenvalues with finite algebraic multiplicity which are isolated points of $\sigma(A)$. The *essential spectrum* is defined $\sigma_{ess}(A) := \sigma(A) \setminus \sigma_d(A)$.

2. $\ker(A-\lambda) = \{0\}$, $\mathrm{Ran}(A-\lambda)$ is dense so that $(A-\lambda)^{-1}$ is defined, $(A-\lambda)^{-1}$ is an unbounded operator on X.

3. $\ker(A-\lambda) = \{0\}$, $\mathrm{Ran}(A-\lambda)$ is not dense so $(A-\lambda)^{-1}$ cannot be uniquely defined on X. This part of the spectrum is called the *residual spectrum* and denoted $\sigma_{res}(A)$.

Self-Adjointness. The *adjoint* of an operator A, denoted A^*, satisfies $\langle x, Ay\rangle = \langle A^*x, y\rangle$ and can be shown to exist by the Riesz representation theorem. An operator A with domain $D(A)$ is *self-adjoint* if (1) A is *symmetric*, i.e. $(y, Ax) = (Ay, x)$ and (2) $D(A) = D(A^*)$. In general, this second criteria is a difficult property to establish and, in practice, the perturbation theories of Kato and Rellich are typically used to show that a Schrödinger operator with a particular potential is self-adjoint. The consequences of self-adjointness are significant; if A is a self-adjoint operator then

1. A is a *closed* operator on $D(A)$, i.e the graph of A, $\Gamma(A) = (D(A), \mathrm{Ran}(A))$, is a closed subset of $X \oplus X$.

2. $\sigma(A) \subset \mathbb{R}$.

3. $\sigma_{res} = \emptyset$.

4. The algebraic multiplicity of an eigenvalue is equal to the geometric multiplicity.

5. The eigenvectors corresponding to distinct eigenvalues are orthogonal.

6. If $\lambda \in \rho(A)$, then $R_A(\lambda) \in \mathcal{B}(X)$ and $\|R_A(\lambda)\| \leq \mathrm{dist}(\lambda, \sigma(A))^{-1}$.

7. if $\lambda \in \rho(A)$, then $R_A(\lambda)^* = R_A(\overline{\lambda})$.

8. The spectral theorem decomposes A into projections, which independently are often easier to analyze than all of A. The spectral theorem states that there exists a unique projection-valued measure $\mathrm{d}P_A$ such that $A = \int_{\mathbb{R}} \lambda \, \mathrm{d}P_A(\lambda)$.

9. Stone's theorem states that for a bounded and continuous function of A,

$$f(A) = \frac{1}{2\pi\imath} \lim_{\epsilon \downarrow 0} \int_{-\infty}^{\infty} f(\lambda) \left[R_A(\lambda + \imath\epsilon) - R_A(\lambda - \imath\epsilon) \right] \mathrm{d}\lambda. \tag{A.2}$$

An operator A with domain $D(A)$ is referred to as *essentially self-adjoint* if it closure is self-adjoint.

Compactness. A bounded linear operator, $A \in \mathcal{B}(X)$, on a Hilbert space, X, is *compact* if it maps any weakly convergent sequence into a strongly convergent sequence. If A is a compact operator, then it has the following spectral properties:

1. (Riesz-Schauder theorem) $\sigma_{ess}(A) \subset \{0\}$. That is, $\sigma(A)$ consists of nonzero isolated eigenvalues of finite multiplicity with the only possible accumulation point at zero and, possibly, the point zero (which may have infinite multiplicity) [Hislop and Sigal, 1996, p. 93].

2. (Fredholm Alternative theorem) For any $\lambda \in \mathbb{C} \setminus \{0\}$, either

   (i) $\lambda \notin \sigma(A)$ and the equation $(A - \lambda)f = g$ has a unique solution for every $g \in H$ or

   (ii) $\lambda \in \sigma(A)$ and the equation $(A - \lambda)f = 0$ has a nonzero solution.

An important example of compact operators are the class of Hilbert-Schmidt integral operators, which are defined as integral operators with Schwartz kernel $K(x, y)$ satisfying

$$\int \int |K(x, y)|^2 \, \mathrm{d}x \, \mathrm{d}y < \infty.$$

We will use the following theorem from [Reed and Simon, 1980, Vol. 1, p. 201].

Theorem A.1.1 (analytic Fredholm theorem)**.** *Let D be an open connected subset of $\mathbb{C}$. Let $A\colon D \to \mathcal{B}(\mathcal{H})$ be an analytic operator-valued function such that $A(k)$ is compact for each $k \in D$. Then either*

(a) $(Id - A(k))^{-1}$ exists for no $k \in D$

or

(b) $(Id - A(k))^{-1}$ *exists for all* $k \in D \setminus S$ *where* S *is a discrete subset of* D*. In this case,* $(Id - A(k))^{-1}$ *is meromorphic in* D*, analytic in* $D \setminus S$*, the residues at the poles are finite rank operators, and if* $k \in S$ *then* $A(k)\psi = \psi$ *has a nonzero solution in* $\mathcal{H}$*.*

A.1.2 Outgoing resolvent and spectral decomposition of the Laplacian

In this section, we express properties of the Laplacian H in terms of the general theory. In what follows, we consider the Sommerfeld outgoing boundary condition

$$\frac{\partial u}{\partial |x|} - \imath\omega u = o\left(|x|^{-\frac{d-1}{2}}\right) \qquad |x| \to \infty.$$

The resolvent of the Laplacian, $(H - \lambda)^{-1}$, with this boundary condition is sometimes referred to as the *outgoing resolvent*.

Proposition A.1.2. *The following are properties of* $H = -\Delta$*.*

1. H *is self-adjoint on* $H^2(\mathbb{R}^d)$ *and essentially self-adjoint on* $L^2(\mathbb{R}^d)$ *and* $\mathcal{S}(\mathbb{R}^d)$*.*
2. $\sigma_d(H) = \emptyset$ *and* $\sigma_{ess}(H) = [0, \infty)$*.*
3. $R_H(\lambda) \in \mathcal{B}(L^2)$ *for* $\lambda \in \mathbb{C} \setminus \mathbb{R}$*.*
4. *For a smooth function* χ *with exponential decay, i.e.* $\chi(x) \sim \exp(-\alpha|x|)$ *for* $\alpha > 0$*, the operator* $\chi R_H \chi$ *is a Hilbert-Schmidt integral operator, hence compact.*

We identify an element $\lambda \in \mathbb{C}$ with a spectral parameter, $k \in \mathbb{C}^+$, the upper-half complex plane, such that $\lambda = k^2$. Note that for $k \in \mathbb{R}$, $k > 0$, we have the property

$$\lim_{\epsilon \downarrow 0} \sqrt{k^2 \pm \imath\epsilon} = \pm k.$$

Then for $k^2 = \lambda \in \rho(H)$, it is convenient to (abusively) define the outgoing resolvent operator in terms of the spectral parameter, k,

$$R(k) := (H - k^2)^{-1},$$

rather than λ. The Schwartz kernel of the resolvent is referred to as the *Green's function*, denoted $G(x, y, k)$ and satisfies

$$R[f](x, k) = \int G(x, y, k) f(y) \, \mathrm{d}y$$

Since the medium is homogeneous, the Green's function depends only on the difference $x-y$. Thus we abbreviate $G(x,y,k)$ by $G(x-y,k)$ and we note that the resolvent is expressible by convolution. $R[f](x,k) = G * f(x,k)$.

Proposition A.1.3. *For $\Im k > 0$, $R(k) \in \mathcal{B}(L^2)$ with norm bounded above by $|\Im k^2|^{-1}$. The Green's function $G(r,k)$ is given by*

$$G(r,k) = \begin{cases} -(2\imath k)^{-1} \exp(\imath k|r|) & d=1 \\ -(4\imath)^{-1} H_0^{(1)}(k|r|) & d=2 \\ -(4\pi|r|)^{-1} \exp(\imath k|r|) & d=3 \\ -(4\imath)^{-1} \left(\frac{k}{2\pi|r|}\right)^{\frac{d-2}{2}} H^{(1)}_{\frac{d-2}{2}}(k|r|) & d \text{ general.} \end{cases} \tag{A.3}$$

There are several techniques for deriving Eq. (A.3). Typically one uses the fact that the Green's function satisfies the equation

$$(H-k^2)G(r,k) = \delta(r) \tag{A.4a}$$

$$\frac{\partial G}{\partial |r|} - \imath k G = o\left(|r|^{-\frac{d-1}{2}}\right) \qquad |r| \to \infty. \tag{A.4b}$$

In one dimension, this can be solved for $r>0$ and $r<0$ and the continuity of G and its derivative can be used for to derive Eq. (A.3). In terms of the *Jost solutions* satisfying the boundary conditions $\lim_{x\to\pm\infty}(\partial_x \mp \imath k) f_\pm = 0$,

$$G(x-y,k) = W(k)^{-1} \begin{cases} f_+(x,k) f_-(y,k) & x > y \\ f_-(x,k) f_+(y,k) & x < y \end{cases}$$

where $W(k) = \text{Wronskian}(f_+(\cdot,k),\, f_-(\cdot,k))$. For $d \geq 2$, the second Green's theorem is applied to G and a solution of the Helmholtz Eq. on the set $\{r \in \mathbb{R}^d \colon \epsilon \leq |r| \leq R\}$. Then one obtains Eq. (A.3) by taking the limit $\epsilon \downarrow 0$ and $R \uparrow \infty$. In 2D, this calculation is discussed in [Afshari *et al.*, 2008], [Economou, 2006, p.12], and [Colton and Kress, 1983, p.106]. In 3D, this approach is taken in [Goodman, 2004, p.38] and [Colton and Kress, 1983, p.46]. One may also evaluate the integral obtained by taking the Fourier transform of Eq. (A.4) to obtain

$$G(r,k) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{e^{\imath \xi \cdot r}}{|\xi|^2 - k^2} \, \mathrm{d}\xi \tag{A.5a}$$

$$= \frac{1}{(2\pi)^d} \int_0^\infty \int_{S^{d-1}} \frac{e^{\imath \zeta \omega \cdot r}}{\zeta^2 - k^2} \, \mathrm{d}\omega \zeta^{d-1} \, \mathrm{d}\zeta \tag{A.5b}$$

where $\xi = \zeta\omega$, $\zeta > 0$, $\omega \in S^{d-1}$ are polar coordinates. The ζ integral is treated by closing a contour in the upper-half plane which for $\Im k > 0$, encloses exactly one pole with residue yielding Eq. (A.3). This method is employed by [Economou, 2006, p.10] and [Hislop and Sigal, 1996, p.46]. This method can be used to derive Eq. (A.3) in general dimension [Hislop and Sigal, 1996, p.165]. [Schmalz *et al.*, 2010] further discusses the dependence of the Green's function on the boundary conditions. For some equations, it is convenient to apply parabolic methods to derive Green's functions [Beals, 1999; Aarao, 2007].

With the construction of the resolvent in place, Stone's theorem (see Eq. (A.2)) now provides a method for defining functions of the Laplacian. In terms of the spectral parameter, k, Stone's theorem states

$$f(H) = \frac{1}{2\pi\imath} \lim_{\epsilon \downarrow 0} \int_0^\infty f(k^2) \left[R(k + \imath\epsilon) - R(-k + \imath\epsilon)\right] 2k \, \mathrm{d}k.$$

We are thus interested in the extension of $R(k)$ to $k \in \mathbb{R}$, which is given by the following

Theorem A.1.4 (Limiting Absorption Principle)**.** *For $\Im k > 0$, the resolvent $R(k) = (H - k^2)^{-1}$ is an analytic function with values in $\mathcal{B}(L^2)$. For χ a compactly supported function, $\chi R(k)\chi \in \mathcal{B}(L^2)$ for $k \in \mathbb{R}$.*

Proof. See [Agmon, 1975] and in one-dimension, [Tang and Zworski,]. □

The spectral decomposition of the Laplacian may now be given.

Proposition A.1.5 (Spectral Decomposition)**.** *The spectral decomposition of H on $H^2(\mathbb{R}^d)$ is given by*

$$Hg(x) = \int_0^\infty \lambda d\mu_\lambda[g(x)]$$

where the projection valued measure is defined

$$d\mu_\lambda[g(x)] := \frac{1}{(2\pi)^d} \frac{1}{2} \lambda^{\frac{d}{2}-1} \int_{S^{d-1}} e^{\imath\sqrt{\lambda}\omega\cdot x} \hat{g}(\sqrt{\lambda}\omega) \, \mathrm{d}\omega \, \mathrm{d}\lambda$$

Moreover,

$$f(H)g = \int_0^\infty f(\lambda) d\mu_\lambda[g] \tag{A.6}$$

where f is any Borel function. Finally, by approximation we have that (A.6) holds with $f(\zeta) = \delta(\zeta)$, the Dirac delta distribution in the distributional sense.

Proof. By Stone's Theorem, we have

$$\mathrm{d}\mu_\lambda[g] = \frac{1}{2\pi\imath}\lim_{\epsilon\downarrow 0}\left[R(\sqrt{\lambda+\imath\epsilon}) - R(\sqrt{\lambda-\imath\epsilon})\right] g\,\mathrm{d}\lambda \tag{A.7}$$

Using Eq. (A.5), we compute

$$G(r,\sqrt{\lambda+\imath\epsilon}) - G(r,\sqrt{\lambda-\imath\epsilon}) = \frac{1}{(2\pi)^d}\int_0^\infty \int_{S^{d-1}} e^{\imath\zeta\omega\cdot r}\left[\frac{1}{\zeta^2-\lambda-\imath\epsilon} - \frac{1}{\zeta^2-\lambda+\imath\epsilon}\right]\mathrm{d}\omega\zeta^{d-1}\,\mathrm{d}\zeta.$$

The Sokhotskyi-Plemelj formula, $\lim_{\epsilon\downarrow 0}\frac{\epsilon}{a^2+\epsilon^2} \to \pi\delta(a)$, implies that

$$\lim_{\epsilon\downarrow 0}\frac{1}{\zeta^2-\lambda-\imath\epsilon} - \frac{1}{\zeta^2-\lambda+\imath\epsilon} = \lim_{\epsilon\downarrow 0}\frac{2\imath\epsilon}{(\zeta^2-\lambda)+\epsilon^2} = 2\pi\imath\delta(\zeta^2-\lambda) = 2\pi\imath\frac{\delta(\zeta-\sqrt{\lambda})}{2\sqrt{\lambda}}$$

for $\zeta > 0$. Inserting these expressions into Eq. (A.7), we obtain

$$\begin{aligned}\mathrm{d}\mu_\lambda[g] &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{S^{d-1}} e^{\imath\sqrt{\lambda}\omega\cdot(x-y)} g(y)\frac{1}{2}\lambda^{\frac{d}{2}-1}\,\mathrm{d}\omega\,\mathrm{d}y\,\mathrm{d}\lambda \\ &= \frac{1}{(2\pi)^d}\frac{1}{2}\lambda^{\frac{d}{2}-1}\int_{S^{d-1}} e^{\imath\sqrt{\lambda}\omega\cdot x}\hat{g}(\sqrt{\lambda}\omega)\,\mathrm{d}\omega\,\mathrm{d}\lambda\end{aligned}$$

as desired. □

The spectral theorem, which defines functions of H, allows us to concisely write the solution of the Schrödinger and Wave equations as

$$\phi(t) = \exp(-\imath Ht)\phi_0 \tag{A.8a}$$

$$u(t) = \cos(H^{\frac{1}{2}}t)f + H^{-\frac{1}{2}}\sin(H^{\frac{1}{2}}t)g \tag{A.8b}$$

respectively. In particular, these agree with Eq. (A.1).

We now briefly discuss the meromorphic continuation of the resolvent $\chi R(k)\chi$ to the lower-half k-plane.[1] Our motivation for this will be apparent in Sec. A.3 when we consider the scattering resonance expansion for the inhomogeneous wave equation. The meromorphic continuation of $\chi R(k)\chi$ to the lower-half k-plane depends on the dimension d. Using (A.3), in one-dimension, we find that the resolvent $R(k)$ can be expanded in a Laurent series about $k = 0$:

$$R(k) = \frac{1}{k}P + Q(k)$$

[1] When the spectral parameter k is in the lower k-plane, $\lambda = k^2$ is in the second sheet of the λ-plane.

where $P[f] = \frac{\imath}{2}\int f$ and $Q[f](x,k) = -\frac{1}{2}\int |x-y| e^{\imath k|x-y|} f(y)\,\mathrm{d}y$. Thus $\chi R(k)\chi \in \mathcal{B}(L^2)$ has a meromorphic extension from $\Im k > 0$ to $\mathbb{C}$ with a simple pole at $k = 0$. In odd dimensions $d \geq 3$, the Green's function given in (A.3) is an entire function of k. In even dimensions it is entire on a logarithmic covering of $\mathbb{C}$ as a function of the variable $log(k)$ [Dolph *et al.*, 1966; Shenk and Thoe, 1972; Golowich and Weinstein, 2005; Melrose, 1995]. We define

$$\Lambda_d = \{k \in \mathbb{C} \colon R_0(k) \text{ has an analytic continuation in dimension } d\}. \tag{A.9}$$

A.2 Wave propagation in a homogeneous media

In this section, we discuss two examples of wave propagation in homogeneous media, namely the solutions to the Cauchy problem for the Schrödinger and wave equations as given in Eq. (A.8). We also discuss the phenomena of coherent diffraction in the wave equation. Let us begin with a few qualitative definitions.

Although wave phenomena is familiar and ubiquitous in our everyday lives, it is rather difficult to define. In [Keller, 1979], Joseph B. Keller defines wave propagation as follows: "Waves are motions or phenomena which are more or less oscillatory in time and in space. Propagation is the process of travel or movement from one place to another. Thus wave propagation is another name for the movement of an oscillatory phenomenon." Wave *interference* occurs when multiple waves, typically from different sources, interact with one another. The waves are said to be *coherent* if they have the same (or nearly the same) frequency. These concepts are beautifully illustrated in Fig. A.1 by the waves generated by the periodic, synchronous beating of the wings of a bee trapped in a pool of water.

The term *diffraction* typically describes the exit of a wave from a thin aperture[2]. The best-known example of diffraction is the double-slit experiment performed by Thomas Young in 1801, in which light diffracting from two closely-spaced slits, interfered with one and another to demonstrate the wave-like properties of light.

It is important to differentiate the aforementioned wave properties from those which depend

[2]Although this language is not completely standard. Namely, in material science the word "diffraction" is also applied to the scattering of light from a sample, e.g. kinematic or dynamic diffraction. In Chapter 2, we adopt this language.

Figure A.1: The wings of a bee trapped in a pool of water generate waves which are seen to coherently interfere with one another. [source: photo submitted to the 2009 National Geographic photography contest by Michael Johnson]

on an obstacle or inhomogeneity of the propagation medium, such as reflection, refraction, and scattering. Reflection and refraction occur when a plane wave encounters an interface between two media. The *reflected* wave refers to the part which, remaining in the same medium as the incident wave, propagates in a new direction with equal angle (with respect to the incident normal) as the incident wave. The *refracted* part of the wave enters the second media at an angle given by Snell's law. *Scattering* occurs when a wave encounters an obstacle or a localized inhomogeneity in a medium. The scattered wave describes the deviation of a wave from the free case. Scattering will be further discussed in Sec. A.3.

A.2.1 The Cauchy problem for the Schrödinger equation

One may rewrite write Eq. (A.8a) in the following form:

$$\phi(t,x) = e^{-\imath Ht} f = K_t * f(x) \tag{A.10a}$$

$$K_t(x) := (4\pi\imath t)^{-\frac{d}{2}} e^{\imath \frac{|x|^2}{4t}}. \tag{A.10b}$$

The evolution operator $e^{-\imath Ht}$: $\phi(0) \to \phi(t)$ is a one-parameter group of unitary operators since

$$\|e^{-\imath Ht} f\|_{L^2} = \|f\|_{L^2} \tag{A.11}$$

and $e^{-\imath Ht} e^{-\imath Hs} = e^{-\imath H(t+s)}$ $\forall s,t \in \mathbb{R}$. The Schrödinger equation is a dispersive equation with dispersion relation $\omega = |k|^2$. Thus components of the wave with different k values travel at different group velocity [Whitham, 1974]. The solution spreads and decays as $t \uparrow \infty$. Using (A.10), it is straightforward to show that if $f \in L^1(\mathbb{R}^d)$, then $e^{-\imath Ht} f \in L^\infty(\mathbb{R}^d)$ for $t \neq 0$ and

$$\|e^{-\imath Ht} f\|_{L^\infty} \leq |4\pi t|^{-\frac{d}{2}} \|f\|_{L^1}. \tag{A.12}$$

Using an interpolation argument [Stein and Weiss, 1971, Ch.5], it is possible to extend Equations (A.11) and (A.12) to the following

Proposition A.2.1. *Let $1 \leq p \leq 2$ and $2 \leq q \leq \infty$, where p and q are Sobolev conjugate, i.e. $p^{-1} + q^{-1} = 1$. If $f \in L^p(\mathbb{R}^d)$, then for $t \neq 0$, $e^{-\imath Ht} f \in L^q(\mathbb{R}^d)$ and*

$$\|e^{-\imath Ht} f\|_{L^q} \leq |4\pi t|^{-\left(\frac{d}{2} - \frac{d}{q}\right)} \|f\|_{L^p}.$$

In particular, Equations (A.11) and (A.12) are obtained for $q = 2$ and $q = \infty$ respectively.

A.2.2 Wave equation

Unlike the Schrödinger Eq, the wave equation is not dispersive (the dispersion relation is $\omega = |k|$). The equation is said to be *hyperbolic*; the solution travels along "characteristics" of the equation at finite propagation speeds [Whitham, 1974]. Thus for finite times, the solutions satisfy the principle of causality, that is, a point in space-time has both a finite domain of dependence and influence. In some contexts, these are also referred to as forward and backward light-cones. It is easily verified that solutions of the wave equation with localized initial data conserve the energy

$$E[u(\cdot,t), \partial_t u(\cdot,t)] := \int_{\mathbb{R}^d} (\partial_t u(x,t))^2 + |\nabla u(x,t)|^2 \, \mathrm{d}x.$$

One of the striking features that arises in the analysis of the wave equation is that the qualitative behavior of solutions is dimension dependent. Consider initial data

$$u(0) = 0$$
$$u_t(0) = g.$$

In one-dimension, d'Alembert's formula,

$$u(x,t) = \frac{1}{2}\int_{x-t}^{x+t} g(\sigma)\,\mathrm{d}\sigma,$$

gives the solution to the wave equation. In higher dimensions, one may average the wave equation over the surface of a sphere in $\mathbb{R}^d$ to obtain the Euler-Poisson-Darboux equation [Evans, 2000]. In odd dimensions, a transformation can be found which reduces this equation for the spherical mean to the one-dimensional wave equation! In particular, one finds that the domain of influence for a point $y \in \mathbb{R}^d$ at time t is precisely

$$\{x \in \mathbb{R}^d \colon |x-y| = |t|\}, \qquad d \text{ odd}.$$

This is sometimes referred to as the (strong) *Huygen's principle*: If d is odd and g has bounded support, then for any $y \in \mathbb{R}^d$, there exists a time $\tau_y = max\{\|x-y\| \colon x \in \mathrm{supp}(g)\}$ such that $u(y,t) = 0$ for $t > t_y$. The region in the forward light cone where the solution vanishes is sometimes referred to as the *Petrovsky lacuna*.

In even dimensions however, the Euler-Poisson-Darboux formula governing the spherical mean of the solution cannot be transformed into the one-dimensional wave equation. In this case, Hadamard's method of descent is used to show that the domain of influence for a point $y \in \mathbb{R}^d$ at time t is

$$\{x \in \mathbb{R}^d \colon |x-y| \leq |t|\} \qquad d \text{ even}.$$

Thus in even dimensions, the strong Huygen's principle does not hold; the wave does not identically vanish on the interior of an outward propagating wave front.[3]

These facts can also be recovered from the solution of the wave equation given in Eq. (A.8b). Equation (A.8b) gives the solution to the wave equation as

$$u(t,x) = H^{-\frac{1}{2}}\sin(H^{\frac{1}{2}}t)g = L_t * g$$

[3]Interestingly the solution of the discrete wave equation, studied in Ch. 4, does not satisfy the strong Huygen's principle in even or odd dimension [Schultz, 1998].

where the kernel L_t depends on the dimension, d, as follows [Taylor, 1996; Perry, 2002]:

d=1

$$L_t = \frac{1}{2}\text{Heavyside}(|t| - |x|)$$

d=2

$$L_t = \begin{cases} \frac{1}{2\pi}(t^2 - |x|^2)^{-\frac{1}{2}} & |x| < |t| \\ 0 & |x| > |t| \end{cases}$$

d=3

$$L_t = \frac{1}{4\pi|x|}\delta(|x| - |t|)$$

A.2.2.1 Coherent diffraction and the Fresnel diffraction integral

We consider the wave equation on the semi-infinite domain (z, x), $z > 0$ and $x \in \mathbb{R}^d$ with a spatially varying, harmonic forcing along the line $z = 0$:

$$\partial_t^2 u = \partial_z^2 + \Delta_x^2 u \tag{A.13a}$$

$$u(t, z = 0, x) = f(x)e^{-\imath\omega t}. \tag{A.13b}$$

The function f is taken to be supported on Σ, a compact region which we refer to as the aperture. Anticipating that the wave is going to primarily move in the z-direction, we make the ansatz

$$u(z, x, t) = e^{-\imath\omega(t-z)}v(z, x)$$

yielding the equation

$$2\imath\omega\partial_z v = -\Delta_x v - \partial_z^2 v.$$

We now make the paraxial approximation

$$|\partial_z^2 v| \ll |2\omega\partial_z v|$$

which states that v does not vary much in the z-direction on the length-scale $\lambda = 2\pi/\omega$. Writing $H = -\Delta_x$ and $\tau = (2\omega)^{-1}z$ we find that v satisfies the Schrödinger equation $\imath\partial_\tau v = Hv$. Using Eq. (A.10), the solution to Eq. (A.13) can thus be written

$$u(t, z, x) = e^{-\imath\omega(t-z)}\left(\frac{\omega}{2\pi\imath z}\right)^{\frac{d}{2}} \int_\Sigma e^{\frac{\imath\omega|x-y|^2}{2z}} f(y)\,\mathrm{d}y. \tag{A.14}$$

Equation (A.14) is referred to as the *Fresnel diffraction integral* and can also be obtained using the Green's formula and the method of images [Goodman, 2004; Bouwkamp, 1954; Afshari *et al.*, 2008]. Remarkably, this equation is qualitatively dimension independent (aside from the $z^{-\frac{d}{2}}$ L^∞-decay). Furthermore, while the Cauchy problem for the wave equation is non-dispersive in time, the spatial part of the steady-state solution of Eq. (A.13) is dispersive in z-direction and Proposition A.2.1 applies.

A.3 Wave propagation in media with compactly supported inhomogeneity

In this section we will consider the inhomogeneous Schrödinger and wave equations

$$\imath\partial_t\phi = H_V\phi \qquad H_V := H + V \tag{A.15}$$

$$\partial_t^2 u = -H_m u \qquad H_m := (1+m^2)^{-1}H \tag{A.16}$$

where $V, m \in L^\infty_{comp}(\mathbb{R}^d)$, the space of point-wise bounded, compactly supported functions. We additionally assume that V and m are everywhere non-negative. Note that the index of refraction for the wave equation is given by $n = \sqrt{1+m^2}$. We denote the resolvents of these operators

$$R_V(k) := (H_V - k^2)^{-1} \tag{A.17a}$$

$$= R(k)[\mathrm{Id} + VR(k)]^{-1} \tag{A.17b}$$

$$= [\mathrm{Id} + R(k)V]^{-1}R(k) \tag{A.17c}$$

and

$$R_m(k) := \left(H_m - k^2\right)^{-1} \tag{A.18a}$$

$$= [\mathrm{Id} - k^2R(k)m^2]^{-1}R(k)(1+m^2) \tag{A.18b}$$

$$= R(k)[\mathrm{Id} - k^2m^2R(k)]^{-1}(1+m^2) \tag{A.18c}$$

where, as before, we abuse notation by using the spectral parameter k rather than k^2.

The following proposition gives the limiting absorption principle for the perturbed resolvents [Agmon, 1996; Tang and Zworski,].

Proposition A.3.1. *Let χ be a compactly supported, $C^\infty(\mathbb{R}^d)$ function such that $\chi \equiv 1$ on the support of V or m. Then $\chi R_V(k)\chi$ and $\chi R_m(k)\chi$ satisfying Eqs. (A.17) and (A.18) are meromorphic families of $\mathcal{B}(L^2)$ operators for $k \in \Lambda_d$ as defined in Eq. (A.9). Furthermore, $\chi R_V(k)\chi$ and $\chi R_m(k)\chi$ have no pole for $k \in \mathbb{R} \setminus \{0\}$.*

The perturbation theories of Kato and Rellich can be used to show that H_V and H_m are self-adjoint operators. The spectrum of H_V and H_m agree with that of H:

$$\sigma_d(H_V) = \sigma_d(H_m) = \emptyset$$
$$\sigma_{ess}(H_V) = \sigma_{ess}(H_m) = [0, \infty)$$

which implies that $R_V(k)$, $R_m(k) \in \mathcal{B}(L^2)$ for $\Im k > 0$. In particular, there exist spectral decompositions for both H_V and H_m.

A.3.1 Scattering resonance expansion for the wave equation

We now discuss the scattering resonance expansion for wave equations with d odd and compactly supported inhomogeneity and localized data.

$$\partial_t^2 u = -H_m u \tag{A.19a}$$
$$u(0, x) = 0 \tag{A.19b}$$
$$\partial_t u(0, x) = g(x) \tag{A.19c}$$

Taking the Laplace transform[4] of Eq. (A.19a) we obtain

$$[H_m + p^2]\tilde{u} = g.$$

Let $\chi \in \mathcal{C}^\infty(\mathbb{R}^d)$ be a compactly supported function with $\chi \equiv 1$ on the $\operatorname{supp}(g) \cup \operatorname{supp}(m)$. We then have

$$\chi\tilde{u}(p) = \chi[H_m + p^2]^{-1}\chi g$$

where the operator $[H_m + p^2]^{-1}$ is defined by the spectral theorem. We solve for $u(t, x)$ using the inverse Laplace transform integral, taken over the vertical Bromwich contour. For $\alpha > 0$ we have

$$\chi U(t)\chi g := \chi u(t, x) = \frac{1}{2\pi\imath} \lim_{M \uparrow \infty} \int_{\alpha - \imath M}^{\alpha + \imath M} e^{pt}\chi[H_m + p^2]^{-1}\chi g \, \mathrm{d}p.$$

[4] $\tilde{f}(p) = \mathcal{L}[f](p) = \int_0^\infty e^{-pt} f(t) \, \mathrm{d}t$

Now let $p = -\imath k$ to obtain

$$\chi U(t)\chi g = \frac{1}{2\pi} \lim_{M\uparrow\infty} \int_{-M+\imath\alpha}^{M+\imath\alpha} e^{-\imath kt} \chi R_m(k) \chi g \, \mathrm{d}k.$$

This integral can be computed exactly by closing the contour in the lower half-plane along a line at $k = -\imath A$. If we continue to deform this contour by letting $A \uparrow \infty$, the contour will encloses all of the resonances. Thus, Cauchy's residue theorem gives the full resonance expansion

$$\chi U(t)\chi g(x) = \sum_{k_j \in Res} \sum_{\ell=1}^{M_j} e^{-\imath k_j t} t^{\ell-1} \langle \psi_{j,\ell}\chi, g\rangle \chi(x)\psi_{j,\ell}(x)$$

where M_j is the multiplicity of the resonance k_j with corresponding resonant states $\{\psi_{j,\ell}\}_{\ell=1}^{M_j}$. We are particularly interested in the partial resonance expansion, where A remains finite and the contour only encloses some of the resonances in the lower half-plane. In this case, as the following proposition shows, the error term is given by the integral on the contour $k = -\imath A$, which can be shown to be exponentially decreasing as $A \uparrow \infty$ [Tang and Zworski, 2000].

Theorem A.3.2. *Let d be odd and $\chi \equiv 1$ on $supp(g) \cup supp(m)$. Fix $A > 0$ arbitrarily large and small $\epsilon > 0$. Then for $t > 0$ sufficiently large, for all $h \in H^2(\mathbb{R}^d)$*

$$\|\chi U(t)\chi g(x) - \sum_{\substack{k_j \in Res \\ |\Im k_j| < A}} \sum_{\ell=1}^{M_j} e^{-\imath k_j t} t^{\ell-1} \langle \psi_{j,\ell}\chi, g\rangle \chi(x)\psi_{j,\ell}(x)\|_{L^2} \sim C(\epsilon) e^{-t|A-\epsilon|}$$

where M_j is the multiplicity of the resonance k_j with corresponding resonant states $\{\psi_{j,\ell}\}_{\ell=1}^{M_j}$.

Appendix B

BFGS approximation of the Hessian

"I'm a Hessian without no aggression. If you can't beat 'em, join 'em."

—Yosemite Sam in Bunker Hill Bunny, 1950

In the chapters of Part III this thesis, nonlinear optimization methods are used which, in part, rely on an approximation strategy for the Hessian of the objective function $f(x)$ being minimized. In this appendix, we discuss one particular strategy of approximating the Hessian which is referred to as the BFGS method, named after its discoverers: J. Broyden, R. Fletcher, D. Goldfarb, and D. F. Shanno. Our discussion largely follows [Nocedal and Wright, 2006].

Iterative methods for unconstrained, nonlinear optimization can be generally viewed as follows: Let $f(x)\colon \mathbb{R}^n \to \mathbb{R}$ be the objective function to be minimized. At iteration k, a model of f is introduced based on information obtained about the objective function at the current iterate x_k and the iteration history $\{x_j\}_{j=1}^{k-1}$. The model optimization problem is then (approximately) solved to generate the next iterate x_{k+1}. This process is then repeated until the iterates have converged, which for $f \in \mathcal{C}^1$, typically requires $\|\nabla f(x_k)\|$ to be less than a specified tolerance. If the nonlinear optimization problem also includes constraints, then an iterative method will construct a model optimization problem at each iteration, but each of these might also involve constraints.

If f is smooth, a natural candidate for such a model is based on the Taylor expansion of $f(x)$ at the current iterate $x = x_k$:

$$f(x) = f(x_k) + \langle g_k, x\rangle + \frac{1}{2}\langle x, H_k x\rangle + o(\|x - x_k\|^2)$$

where $g_k := \nabla f(x_k)$ is the gradient and $H_k := \nabla^2 f(x_k)$ is the Hessian. The following methods for unconstrained optimization problems are based on this approach:

1. The method of steepest descent with an exact line-search generates a model of $f(x)$ at iteration n by truncating the Taylor expansion at first order. The first local minima of the function $f(x) \approx f(x_k) + \langle g_k, x \rangle$ generates the next iterate.

2. Newton's method generates a model of $f(x)$ by truncating the Taylor expansion at second order. The next iterate is generated by solving the linear equation $H_k x = -g_k$.

Note that while these two methods are easy to describe and analyze, modifications of these methods which include an inexact line-searchs or trust-regions are generally recommended for application. If the optimization problem also includes constraints, natural candidates for the model objective function at iteration k will still involve g_k and H_k (and perhaps barrier functions of the constraints).

In both the unconstrained and constrained problem iterations, a model optimization problem involving both the gradient and Hessian can better approximate the original optimization problem than one involving only the gradient. Thus one naïvely expects that Hessian-based methods will converge in fewer iterates than gradient-based methods. Indeed for unconstrained problems where f is sufficiently regular, as the following propositions make precise, the method of steepest descent converges linearly while Newton's method converges quadratically [Nocedal and Wright, 2006].

Proposition B.0.3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ function and suppose that iterates $\{x_k\}$ generated by the steepest descent method with an exact line-search converge to x_* with $\nabla^2 f(x_*) \succ 0$. Then*

$$\frac{|f(x_{k+1}) - f(x_*)|}{|f(x_k) - f(x_*)|} \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2$$

where λ_1 and λ_n are the smallest and largest eigenvalues of $\nabla^2 f(x_*)$ respectively.

Proposition B.0.4. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ function and suppose that iterates $\{x_k\}$ generated by Newton's method converge to x_*. Assume $\nabla^2 f(x_*) \succ 0$ and that $\nabla^2 f(x)$ is Lipschitz continuous in a neighborhood of the solution x_* with Lipschitz constant L.*[1] *Then*

$$\frac{|f(x_{k+1}) - f(x_*)|}{|f(x_k) - f(x_*)|^2} \leq L.$$

[1] $\nabla^2 f(x)$ is Lipschitz continuous in a neighborhood of the solution x_* with Lipschitz constant L if $\|\nabla^2 f(x) - \nabla^2 f(x_*)\| \leq L\|x - x_*\|$ for all x near x_*.

It is often the case that the Hessian is an expensive quantity to compute and the cost of its computation overwhelms the increased convergence rate of a Hessian-based method. If instead, one tries to learn information about the curvature of $f(x)$ along the sequence of iterates without explicitly computing the Hessian, a natural question arises:

> Is it possible to approximate the Hessian at iterate k, H_k, using the gradient computation at prior iterates: $\{g_j\}_{j=1}^{k-1}$?

Such a method would be gradient-based, yet by emulating a Hessian-based method, hopefully yield a better convergence rate than a more rudimentary gradient-based method, such as the method of steepest descent. Such methods are generally referred to as *quasi-Newton methods*. Indeed for unconstrained optimization problems, quasi-Newton methods generally converge superlinearly, a rate which is quicker than the steepest descent method, yet slower than Newton's method.

There are several strategies of approximating the Hessian at iterate k using the gradient computation at prior iterates. The BFGS method is a popular choice.[2]

The BFGS method. Let $f(x)$ be a $\mathcal{C}^2$ objective function to be minimized in an unconstrained or constrained optimization problem. Suppose that at the prior iteration of an iterative method, we had iterate x_{k-1}, gradient g_{k-1}, and approximate Hessian $B_{k-1} \approx H_{k-1}$, which we assume to be symmetric and positive definite. At iteration k, the current iteration, the optimization method has generated x_k, and $g_k = \nabla f(x_k)$ has been computed. We seek to update B_{k-1} to obtain an approximate Hessian for the current iterate, B_k, retaining symmetry and positive definiteness.

Defining

$$s_k = x_k - x_{k-1}$$

$$y_k = g_k - g_{k-1},$$

[2]Other strategies include the Davidon-Fletcher-Powell formula, the secant-rank-1 (SR1) formula, and the Broyden class. For misfit-type objective functions of the form $\min f(x) = \frac{1}{2}\|g(x) - b\|_2^2$, the Gauss-Newton method approximates $H \approx J^t J$ where J is the Jacobian of g.

it is instructive to compute

$$\begin{aligned} y_k &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \nabla f(x_{k-1} + t s_k)\, \mathrm{d}t \\ &= \int_0^1 \nabla^2 f(x_{k-1} + t s_k) s_k \, \mathrm{d}t \\ &= \overline{H} s_k \end{aligned}$$

where $\overline{H} = \int_0^1 \nabla^2 f(x_{k-1} + t s_k)\, \mathrm{d}t$ is the average Hessian along the line segment from x_{k-1} to x_k. Thus, a natural requirement is that the approximate Hessian B_k satisfies the *secant condition*

$$B_k s_k = y_k. \tag{B.1}$$

Note that the secant condition only provides n constraints on B_k which due to symmetry has $n(n+1)/2$ components. Let $A_k = B_k^{-1}$ be the inverse approximate Hessian at iterate k. The BFGS method chooses $B_k = A_k^{-1}$ where A_k is defined to be the minimizer of the following optimization problem:

$$\begin{aligned} \min_A \quad & \|A - A_{k-1}\|_{\mathrm{W}} \\ \text{s.t.} \quad & A = A^T \\ & A y_k = s_k. \end{aligned} \tag{B.2}$$

Here, $\|\cdot\|_W$ is the weighted Frobenius norm, defined $\|A\|_{\mathrm{W}} = \|W^{\frac{1}{2}} A W^{\frac{1}{2}}\|_{\mathrm{F}}$, where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm and W can be chosen to be any matrix satisfying the secant condition. The solution to (B.2) is given by

$$A_k = (Id - \rho_k s_k y_k^t) A_{k-1} (Id - \rho_k s_k y_k^t) + \rho_k s_k s_k^t$$

where $\rho_k = (y_k^t s_k)^{-1}$. The BFGS update formula for $B_k = A_k^{-1}$ can be found by applying the Sherman-Morrison-Woodbury formula to obtain

$$B_k = B_{k-1} - \frac{B_{k-1} s_k s_k^t B_{k-1}}{s_k^t B_{k-1} s_k} + \frac{y_k y_k^t}{y_k^t s_k}. \tag{B.3}$$

One may verify that the BFGS update preserves symmetry and, provided the *curvature condition*: $\langle y_k, s_k \rangle > 0$ holds, the BFGS update preserves positive definiteness.[3]

The following proposition states that if the BFGS method converges, it does so superlinearly.

[3] If an inexact line search is used, the curvature condition is enforced by the Wolfe condition.

Proposition B.0.5. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ function and suppose that iterates $\{x_k\}$ generated by the BFGS method with an exact line-search converge to x_*. Assume $\nabla f(x_*)$ is Lipschitz continuous in a neighborhood of x_*. Then*

$$\lim_{k\uparrow\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \to 0.$$