Learned infinite elements for helioseismology

Dissertation

zur Erlangung des

mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsstudiengang *Mathematical Sciences*

der Georg-August University School of Science (GAUSS)

vorgelegt von:

Janosch Preuß

aus Oldenburg

Göttingen, 2021

Betreuungsausschuss

Prof. Dr. Christoph Lehrenfeld
Institut für Numerische und Angewandte Mathematik
Georg-August-Universität Göttingen

Prof. Dr. Thorsten Hohage
Institut für Numerische und Angewandte Mathematik
Georg-August-Universität Göttingen und
Max-Planck-Institut für Sonnensystemforschung, Göttingen

Dr. Damien Fournier
Max-Planck-Institut für Sonnensystemforschung, Göttingen

Mitglieder der Prüfungskommission

Referent: **Prof. Dr. Christoph Lehrenfeld**

Koreferent: **Prof. Dr. Thorsten Hohage**

Weitere Mitglieder der Prüfungskommission

Prof. Dr. Martin J. Gander
Section de Mathématiques
Université de Genève

Prof. Dr. Laurent Gizon
Max-Planck-Institut für Sonnensystemforschung, Göttingen und
Institut für Astrophysik, Georg-August-Universität Göttingen

Jun.-Prof. Dr. Anne Wald
Institut für Numerische und Angewandte Mathematik
Georg-August-Universität Göttingen

Prof. Dr. Ingo Witt
Mathematisches Institut
Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 02.12.2021

Abstract

Time-harmonic wave equations are often formulated on unbounded domains including additional radiation conditions at infinity. To treat such problems using the finite element method (FEM), an artifical boundary has to be introduced rendering the surrounded computational domain bounded. It is crucial to impose an appropriate transparent boundary condition on this artificial boundary to guarantee that the solution computed on the bounded domain is an accurate representation of the true solution of the original problem posed on the infinite domain. The exact transparent boundary condition is given by the Dirichlet-to-Neumann (DtN) map which is a non-local operator with dense matrix representation whose use should be avoided in implementations lest the computational costs become unaccceptable.

This thesis presents efficient techniques for integrating the information contained in the DtN map of time-harmonic waves propagating in a stratified medium into FEM discretizations. This task arises also more generally in the context of domain decomposition methods. If the stratified medium under consideration is homogeneous, a wealth of suitable approaches, e.g. perfectly matched layers (PML), exists. However, methods that are applicable for media like the Sun, that allow for strong reflection of waves, are lacking. This thesis aims to fill this gap with a new method called learned infinite elements.

We start by introducing a general framework to describe DtN maps of stratified exterior domains with separable geometry. It is well-known that the DtN map is separable in the eigenbasis of a suitable self-adjoint differential operator on the artificial truncation boundary. It is less well-known that its diagonal entries can be interpreted as values of a function called dtn which often admits a natural meromorphic extension from the spectrum to the complex plane and fulfills a common set of properties across a wide range of different stratified media. We use these properties to study the well-posedness of problems with DtN map as transparent boundary condition.

The thesis proceeds to the discrete level by considering generic transparent boundary conditions of tensor-product type which are determined by small matrices in stratification direction. Many popular methods like classical infinite elements or tensor-product PMLs can be described in this framework. Analogously to the continuous case we show that the approximate DtN map obtained with such an ansatz diagonalizes in a discrete eigenbasis and can be described by a discrete dtn function which depends only on these small matrices. A semi-discrete error analysis shows that the error introduced by the transparent boundary condition is determined by how accurate this discrete dtn function approximates the dtn function of the continuous DtN operator. This turns out to be a rational approximation problem. By appealing to results of rational approximation theory and capitalizing on the meromorphic structure of the continuous dtn function, we show that extremely accurate approximations on bounded subsets of the spectrum can be achieved. Assuming sufficient smoothness of the data this suffices to obtain highly accurate approximate solutions.

Whereas the small matrices in stratification direction are specified analytically in conventional transparent boundary conditions, the main idea of learned infinite elements (learned IEs) is to determine these matrices by solving a small optimization problem. The objective function of this problem penalizes the misfit between the continuous and discrete dtn functions which leads to an optimal DtN approximation. Thanks to the tensor-product structure, the arising linear systems automatically remain sparse. The sparsity can be improved further by a reduction step after which the number of unknowns grows only linearly with respect to the dimension of the small matrices while the accuracy improves exponentially fast. This renders learned IEs very competitive with and often even superior to established transparent boundary conditions for the Helmholtz equation. Since the complexity for setting up learned IEs is independent of the dimension of the space in which the waves propagate, their computational overhead will be negligible in three-dimensional simulations compared to other costs which arise irrespective of the transparent boundary condition. The optimization step also makes them extremely flexible as illustrated in numerical experiments involving inhomogeneous exterior domains, elliptical truncation boundaries or waveguides. We also investigate an extension of learned IEs that is applicable for computing resonances.

A substantial part of this thesis is devoted to the task of modelling the solar atmosphere using learned IEs. We carefully study the influence of the transparent boundary conditions on the accuracy of helioseismic observables for two different atmospheric models. Learned IEs are observed to deliver high-accuracy at small computational costs and are moreover the first transparent boundary condtions that can be used in conjunction with realistic models of the solar chromosphere.

In the second half of the thesis the use of learned IEs as transmission conditions in sweeping preconditioners, which are a special class of domain decomposition methods, is investigated. Usually, sweeping preconditioners rely on PMLs to approximate the DtN maps at subdomain interfaces, which leads to severe issues if the medium allows for strong reflection of waves. Numerical experiments indicate that learned IEs are able to fix this issue provided that the underlying equation and the geometry are separable. This assumption is crucial because strong reflections render DtN maps extremely sensitive to perturbations. Since learned IEs currently rely on separability, they can only be applied to non-separable equations (allowing for strong reflections) if these can be regarded as nanoscopic perturbations of separable equations. For helioseismology this entails a restriction to sound speed perturbations of smaller than half a percent. Further research is needed to remove the stringent separability requirement and turn sweeping preconditioners based on learned IEs into robust and efficient solvers for this application.

Acknowledgements

The first one and a half years of this thesis project were tough. It was a time of many failed attempts and knitted brows. After arduous drilling we eventually struck gold (learned infinite elements) and the research started to get rolling. I am grateful to everyone who helped to overcome this initial draught and urged me towards the end to put the hammer down to finish this work.

I was privileged to receive equal support from two excellent supervisors, Prof. Christoph Lehrenfeld and Prof. Thorsten Hohage, who contributed to this work with their complementary sets of expertise. Christoph supported this project in particular with his knowledge on finite elements and was at all times available for questions regarding implementational aspects or matters of reproducible research. It never ceases to surprise and inspire me when he pulls some new implementational trick or useful tool out of his hat. In addition to his involvement in this thesis, I would also like to thank Christoph for mentoring me since I joined his group as a master's student in 2017 and for always remaining down-to-earth and easily approachable despite his increasing set of obligations in recent years. Thorsten supported this project with his broad knowledge of wave propagation and initiated the main research pursuits of this thesis - learned infinite elements and sweeping preconditioners for problems involving strong reflections. Despite his commitments in various administrative institutions and research projects, I was always able to count on his constructive feedback and advice. For instance, his insights and hints on technical matters helped me to improve initial versions of the analysis presented in the first two chapters of this thesis significantly.

I would also like to express my gratitude to Dr. Damien Fournier who supported me as a third advisor and was in particular available for questions on computational helioseismology. In this regard, I do not want to miss the opportunity to thank the Inria team Magique-3D for all the interesting discussions during our joint workshops on computational helioseismology and for their kind hospitality during our visit in Bordeaux.

Concerning sweeping preconditioners, I would like to thank Prof. Martin Gander for coming to Göttingen in 2018 to give a colloquium talk on the topic. Our ensuing discussion regarding the applicability of this technique for helioseismology was very helpful. Additionally, he coorganized a minisymposium on sweeping preconditioners together with Thorsten at the WAVES 2019 conference which gave me the opportunity to present and discuss my work with the experts in the field. I am also very grateful that Martin joined the examination board for this thesis.

Further, I would like to express my gratitude to Prof. Laurent Gizon, Jun.-Prof. Anne Wald and Prof. Ingo Witt for serving in the committee for my thesis defense. Additionally, I would like to thank Laurent for inviting me to join his lecture on helioseismology which helped me to pick up some physical background about the problems towards which the methods presented in this thesis are geared.

Besides this more specialized knowledge, this thesis builds upon a solid foundation in

Contents

Nomenclature

Geometrical objects

Notation	Description	Introduced on page
d	Dimension of the space $\mathbb{R}^d$	12
Ω	An open, connected subset (a domain) $\Omega \subset \mathbb{R}^d$ with decomposition $\overline{\Omega} = \overline{\Omega}_{\text{int}} \cup \overline{\Omega}_{\text{ext}}$	12
Ω_{int}	Bounded interior domain	12
Ω_{ext}	Exterior domain (usually unbounded)	12
Γ	Coupling boundary between interior and exterior domain, $\Gamma := \overline{\Omega}_{\text{int}} \cap \overline{\Omega}_{\text{ext}}$	12
Ψ	Diffeomorphism between $[a, \infty) \times \Gamma$ and $\overline{\Omega}_{\text{ext}}$	12
$(r, \widehat{x})$	Coordintes on $[a, \infty) \times \Gamma$ We often refer to r as the radial coordinate	13
Γ_{D}	Dirichlet boundary of Ω_{int}, possibly empty	14
Γ_{N}	Neumann boundary of Ω_{int}, possibly empty	14
$\mathbb{S}^{d-1}$	Euclidian unit sphere: $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$	-
$E(\lambda, \eta)$	Ellipse with foci $(0,0)$ and $(\lambda, 0)$ through $(0, \eta)$	58
$D(\lambda, \eta)$	Open interior of $E(\lambda, \eta)$	57
S_η	Part of second quadrant for $y \geq \eta$	57

Special functions

Notation	Description	Introduced on page
$J_\nu(x)$	Bessel function of the first kind of order ν	19
$Y_\nu(x)$	Bessel function of the second kind of order ν	19
$H_\nu^{(1)}(x)$	Hankel function of the first kind of order ν $H_\nu^{(1)}(x) = J_\nu(x) + iY_\nu(x)$	19
$K_\nu(x)$	Modified Bessel function of the second kind of order ν.	19
$h_\ell^{(1)}(x)$	Spherical Hankel function of the first kind of order ℓ	24
$W_{\kappa,\mu}(z)$	Whittaker function	35

Operators and sesquilinear forms

Notation	Description	Page
$\mathcal{L}$	Elliptic partial differential operator	12,14
$\mathcal{K}_\Gamma$	Second order, self-adjoint, elliptic (partial) differential operator on Γ, e.g. $\mathcal{K}_\Gamma = -\Delta_\Gamma$	12
$\mathcal{M}_\Gamma$	Multiplication operator on Γ, e.g. $\mathcal{M}_\Gamma = \mathrm{Id}_\Gamma$	12
$\tilde{\mathcal{K}}_\Gamma$	Weighted version of $\mathcal{K}_\Gamma$, i.e. $\tilde{\mathcal{K}}_\Gamma := \mathcal{M}_\Gamma^{-1/2}\mathcal{K}_\Gamma\mathcal{M}_\Gamma^{-1/2}$	13
$\mathcal{A}$	Second-order differential operator in r	12
$\mathcal{B}$	Multiplication operator on $L^2([a,\infty))$	12
$\mathcal{D}t\mathcal{N}$	Dirichlet-to-Neumann (DtN) map	14
$b_{\mathrm{int}}(\cdot,\cdot)$	Sesquilinear form on Ω_{int} (without contributions from Γ)	15
$b(\cdot,\cdot)$	Sesquilinear form $b_{\mathrm{int}}(\cdot,\cdot)$ plus coupling to exterior problem via $\mathcal{D}t\mathcal{N}$ on Γ.	38
$\mathcal{D}t\mathcal{N}_N$	Semi-discrete DtN map based on tensor-product transparent b.c. using radial matrices of dimension $(N+1)^2$	50
$b_N(\cdot,\cdot)$	Sesquilinear form $b_{\mathrm{int}}(\cdot,\cdot)$ plus coupling to exterior problem via $\mathcal{D}t\mathcal{N}_N$ on Γ.	52
tr_Γ	Trace operator from interior domain to $H^{1/2}(\Gamma)$	15,186
Z_Γ	Continuous right inverse of trace operator, $\mathrm{tr}_\Gamma \circ Z_\Gamma = \mathrm{Id}_\Gamma$	15,186
Π_L	Projector onto first L eigenmodes on Γ	53
$\mathscr{L}(X,Y)$	Set of bounded linear operators mapping between Hilbert spaces X and Y	185

Matrices

Notation	Description	Page
K	Finite element discretization of $\mathcal{K}_\Gamma$	49
M	Finite element discretization of $\mathcal{M}_\Gamma$	49
A	Radial matrix $A \in \mathbb{C}^{(N+1)\times(N+1)}$ in ansatz (3.7) for transparent b.c. of tensor-product type	49
B	Radial matrix $B \in \mathbb{C}^{(N+1)\times(N+1)}$ in ansatz (3.7) for transparent b.c. of tensor-product type	49

Function spaces

Notation	Description	Page
V	Subspace of $H^1(\Omega_{\text{int}})$, used in variational formulation for $b(\cdot,\cdot)$	15
V_+	Subspace containing those $v \in V$ with $\operatorname{tr}_\Gamma v \in H^1(\Gamma)$, used in variational formulation for $b_N(\cdot,\cdot)$	51
$H^s(\Gamma)$	Sobolev space on Γ	17,187

Spectral calculus

Notation	Description	Page
(λ_ℓ, v_ℓ)	Generalized eigenpairs at continuous level: $\mathcal{K}_\Gamma v_\ell = \lambda_\ell \mathcal{M}_\Gamma v_\ell$	13
$\{w_\ell\}$	Corresponding orthonormal system on $L^2(\Gamma)$, i.e. $\tilde{\mathcal{K}}_\Gamma w_\ell = \lambda_\ell w_\ell$ and $\langle w_\ell, w_{\ell'} \rangle_\Gamma = \delta_{\ell\ell'}$	13
$(\underline{\lambda}_\ell, \underline{v}_\ell)$	Generalized eigenpairs at discrete level: $K\underline{v}_\ell = \underline{\lambda}_\ell M\underline{v}_\ell$	51

Named constants

Notation	Description	Page
θ	Ellipticity constant of differential operator $\mathcal{L}$	14
C_{III}	Involved in growth bound for $\lvert dtn \rvert$	16
$C_{\text{III}}(N)$	Involved in growth bound for $\lvert \operatorname{dtn}_N \rvert$	64
$C_{\text{II}}(N)$	Involved in (asymptotic) lower bound for $\operatorname{Re} \operatorname{dtn}_N$	64
c_Ψ	Under assumption (C-V) the constant reciprocal length of the normal vector on Γ	38
γ_N	Inf-sup constant for sesquilinear form $b_N(\cdot,\cdot)$	53
C_V	Involved in Gårding inequality for $b_{\text{int}}(\cdot,\cdot)$	16
$C_{L^2(\Omega_{\text{int}})}$	Involved in Gårding inequality for $b_{\text{int}}(\cdot,\cdot)$	16
$C_{L^2(\Gamma)}$	Involved in Gårding inequality for $b(\cdot,\cdot)$	42
$C_{L^2(\Gamma)}(N)$	Involved in Gårding inequality for $b_N(\cdot,\cdot)$	64
$C_{V_+}(N)$	Involved in Gårding inequality for $b_N(\cdot,\cdot)$	64

Assumptions

Notation	Description	Page
(C-I)	Strong ellipticity of $\mathcal{L}$	14
(C-II)	Flow is divergence-free with vanishing normal component on boundary of Ω_{int}	14
(C-III)	Attenuation term has appropriate sign	14
(C-IV)	Diffusion matrix is proportional to the identity on Γ	15
(C-V)	Length of normal vector independent of position on Γ	38
($\textit{dtn}$-I$_\leq$) / ($\textit{dtn}$-I$_<$)	Im $\textit{dtn}(\lambda_\ell)$ is smaller or strictly smaller zero	16
($\textit{dtn}$-II)	Re $\textit{dtn}(\lambda_\ell)$ asymptotically non-negative	16
($\textit{dtn}$-III)	Growth bound on $|\textit{dtn}(\lambda_\ell)|$	16
(dtn$_N$-I$_\leq$) / (dtn$_N$-I$_<$)	Im dtn$_N(\lambda_\ell)$ is smaller or strictly smaller zero	63
(dtn$_N$-II)	Asymptotic lower bound on Re dtn$_N(\lambda_\ell)$	64
(dtn$_N$-III)	Growth bound on $|\text{dtn}_N(\lambda_\ell)|$	64

$\textit{dtn}$ functions

Notation	Medium	Page
$\textit{dtn}^{\text{hom,2d}}$	Homogeneous medium exterior to a disk	19
$\textit{dtn}^{\text{hom,3d}}$	Homogeneous medium exterior to a ball in $d=3$	24
$\textit{dtn}^{\text{jump}}$	Exterior domain of a disk in which the wavenumber jumps at radius R_{J} from k_I to k_∞	26
$\textit{dtn}^{\text{ellipse}}$	Homogeneous medium exterior to an ellipse	30
$\textit{dtn}^{\text{guide}}$	Homogeneous semi-infinite waveguide	32
$\textit{dtn}^{\text{Atmo}}$	Solar atmosphere according to `Atmo` model	35
$\textit{dtn}^{\text{VAL}-\text{C}}$	Solar atmosphere according to VAL-C model	35
q	Potential involved in ODE for $\textit{dtn}^{\text{Atmo}}$ and $\textit{dtn}^{\text{VAL}-\text{C}}$	34
dtn$_N$	Approximation of $\textit{dtn}$ by tensor-product transparent b.c.	51
err$_{\textit{dtn}}(N,\ell)$	Modal error $|\textit{dtn}(\lambda_\ell) - \text{dtn}_N(\lambda_\ell)|$	53
(σ_ℓ^{err})	Weighted modal error	53
$\textit{dtn}_{N_p}^{\text{reg}}$	Regularized version of $\textit{dtn}$ with first N_p simples poles lying in $\mathbb{C} \setminus S_\eta$ removed	60

Chapter 1

Introduction

1.1   Time-harmonic wave equations

Understanding wave phenomena has been crucial for many scientific discoveries. Sometimes, waves are the only source of information we receive from distant or inaccesible objects. For instance, the solar interior is inaccesible to direct observation. Its complex dynamics can only be probed by observing and interpreting oscillations at the visible solar surface (called the photosphere). This challenging task is pursued in the field of helioseismology. For drawing inferences from the oscillation data about the solar interior a thorough understanding of wave propagation in the Sun is indispensable.

Wave equations can usually only be solved analytically in simplified special cases. Even though such simplified models are sometimes useful to gain basic insights about the physics, eventually more complex problems need to be solved for which numerical simulations are required. These can be divided into two main approaches. Either one tries to solve the equations directly in the time domain or one transforms to the frequency domain by means of a time-harmonic ansatz and solves the equations there. The latter approach is considered in this thesis.

The numerical solution of time-harmonic wave equations by means of the finite element method (FEM) entails three main challenges.

- The accuracy of the finite element solution deteriorates drastically with increasing wavenumber. This phenomenon, known as the pollution effect, is unavoidable under reasonable assumptions on the FEM in dimensions larger than one [BS97]. In practice, each wavelength needs to be discretized with a fixed number of degrees of freedom (DOFs) to obtain a reliable solution, which leads to extremely large linear systems in the high-frequency regime.

- Additionally, the arising linear systems are also severely ill-conditioned. Classical preconditioning strategies, which work well for Poisson-like problems, are not applicable as discussed in reference [EG12]. Recently, significant progress has been made by the invention of so called sweeping preconditioners [EY11b, GZ19]. Even though sweeping preconditioners can be very effective for some problems, they encounter severe difficulties in the presence of strong reflections [GZ18, PHL20]. Since strong reflections occur in many problems of practical interest, e.g. in computational helioseismology which is the main application of interest in this thesis, the preconditioning issue for time-harmonic waves is far from resolved.

- Wave equations are often posed on unbounded domains with certain radiation conditions at infinity. For a numerical treatment with the FEM the equation has to be reduced to a bounded computational domain in such a way that the solution computed on the bounded domain is a good approximation to true solution of the problem posed on the unbounded domain. This can be achieved by imposing a so called transparent boundary condition on the artificial truncation boundary. The challenge consists of constructing a transparent boundary condition which is accurate and can be realized efficiently at the same time. Transparent boundary conditions also appear as crucial ingredients of domain decompositioning methods [TW05] for general boundary value problems. Hence, their applicability is not limited to the truncation of unbounded domains.

This thesis offers contributions to the last two mentioned issues. While helioseismology fueled this research, the techniques presented here are applicable to a much broader class of problems and are therefore described in a generalized framework for stratified media. Chapter 2-Chapter 6 introduce novel transparent boundary conditions which are then utilized in the remainder of the thesis in the context of sweeping preconditioners. Special care is taken to address the requirements of computational helioseismology.

1.2 The Dirichlet-to-Neumann map

The basic principle of domain decompositioning can be described as follows. Consider a PDE

$$\mathcal{L}u = f \text{ in } \Omega, \tag{1.1}$$

where $\Omega \subset \mathbb{R}^d$ is some domain and $\mathcal{L}$ an elliptic second order partial differential operator. To keep this introduction simple, assume that the principal part of $\mathcal{L}$ is given by the Laplacian $-\Delta$. If Ω is unbounded, then (1.1) is usually complemented by a radiation condition at infinity which will likewise be ignored in this introduction for simplicity. Assume that $\overline{\Omega} = \overline{\Omega}_{\text{int}} \cup \overline{\Omega}_{\text{ext}}$ is a disjoint partition of the domain with common interface $\Gamma := \overline{\Omega}_{\text{int}} \cap \overline{\Omega}_{\text{ext}}$, see Fig. 1.1 for a sketch, and that $\text{supp} f \subset \Omega_{\text{int}}$. The PDE can be split
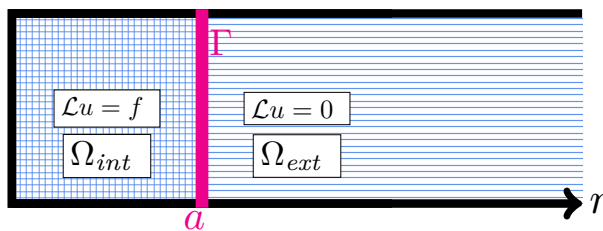
Figure 1.1: Simple domain decomposition.

into two subproblems posed in the subdomains by solving

$$\begin{aligned} \mathcal{L}u^{\text{int}} = f \text{ in } \Omega_{\text{int}}, && \mathcal{L}u^{\text{ext}} = 0 \text{ in } \Omega_{\text{ext}}, \\ -\nabla u^{\text{int}} \cdot \mathbf{n}_\Gamma = \mathcal{DtN} u^{\text{int}} \text{ on } \Gamma, && u^{\text{ext}} = u^{\text{int}} \text{ on } \Gamma, \end{aligned} \tag{1.2}$$

where $\mathcal{DtN} u^{\text{int}} = -\nabla u^{\text{ext}} \cdot \mathbf{n}_\Gamma$ is the Dirichlet-to-Neumann map and $\mathbf{n}_\Gamma$ the normal vector on Γ. By construction, the composite function

$$u(x) = \begin{cases} u^{\text{int}}(x) \text{ for } x \in \Omega_{\text{int}}, \\ u^{\text{ext}}(x) \text{ for } x \in \Omega_{\text{ext}}, \end{cases}$$

2

solves the equation in the subdomains and u as well as its normal flux are continuous across Γ. If $\mathcal{L}$ and the right hand side are regular enough, this implies that u is a strong solution of (1.1). Consequently, the PDE can be reduced to solving problems on the individual subdomains coupled via boundary data.

Suppose that we are primarily interested in obtaining the solution in Ω_{int} and would like to avoid computing the solution in Ω_{ext}. This situation occurs naturally when the domain Ω_{ext} is infinite and the problem in Ω_{int} should be discretized, e.g. using the FEM. In view of (1.2), this boils down to the task of finding a simple expression for the $\mathcal{DtN}$-operator. For instance, if $\mathcal{DtN}u^{\text{int}} = cu^{\text{int}}$ for some constant c were valid, then $\mathcal{DtN}$ could be implemented as a Robin boundary condition and the problem on Ω_{int} would be easy to treat using FEM. Unfortunately, $\mathcal{DtN}$ is in general far more complicated which originates from its nature as a non-local operator.

At this point, the main assumption of the thesis comes into play, which allows to simplify $\mathcal{DtN}$ considerably. The medium in Ω_{ext} is assumed to be stratified in the direction orthogonal to Γ. This means that there are suitable coordinates $(r, \hat{x})$ in Ω_{ext} (see Fig. 1.1) so that the medium varies only in the r direction. This assumption is fulfilled in many problems of practical interest. For instance, the solar atmosphere can be assumed to be spherically symmetric for helioseismic purposes. However, it is important to mention that solar density and sound speed vary drastically with increasing height above the photosphere and that atmospheric models only provide tabulated values for these parameters which leads to non-analytic functions. A significantly simpler example is given by Helmholtz scattering problems with homogeneous exterior domains. Further examples follow in Chapter 2.

A simple form of $\mathcal{DtN}$ is now obtained by separation of variables. Suppose that the PDE in Ω_{ext} in coordinates $(r, \hat{x})$ is given by

$$[\mathcal{A} \otimes \text{Id}_\Gamma + \mathcal{B} \otimes (-\Delta_\Gamma)]\, u^{\text{ext}} = 0, \tag{1.3}$$

where $\mathcal{A}$ and $\mathcal{B}$ are differential operators in r and $-\Delta_\Gamma$ is the Laplace-Beltrami[1] operator on Γ. The solution u^{ext} can then be developed into eigenfunctions of $-\Delta_\Gamma$, i.e. functions w_ℓ solving $-\Delta_\Gamma w_\ell = \lambda_\ell w_\ell$ and which are orthonormal with respect to the inner product $\langle \cdot, \cdot \rangle_\Gamma$ on $L^2(\Gamma)$. The coefficients in these expansions depend on r and can be determined by solving an ODE for each mode ℓ which arises after replacing $-\Delta_\Gamma$, in (1.3) by λ_ℓ. The initial data at $r = a$ is provided by the coefficients of the Dirichlet data u^{int} in the eigenbasis. The $\mathcal{DtN}$ operator then becomes

$$\mathcal{DtN}u^{\text{int}} := \sum_{\ell=0}^{\infty} \textit{dtn}(\lambda_\ell) w_\ell \langle u^{\text{int}}, w_\ell \rangle_\Gamma, \tag{1.4}$$

where the numbers $\textit{dtn}(\lambda_\ell)$ are obtained by evaluating the derivative of the ODE solutions at the coupling boundary $r = a$. This exposes $\mathcal{DtN}$ as a spectral calculus of $-\Delta_\Gamma$.

The non-locality of $\mathcal{DtN}$ is immediately visible from (1.4) since the eigenfunctions w_ℓ have global support. If (1.4) is used directly to implement $\mathcal{DtN}$ in the context of FEM, as in the well-known DtN-FE method [KG89], this results in a dense matrix representation. In the interest of computational efficiency such a dense matrix block in the linear system should be avoided. For a three dimensional helioseismology simulation it might be impossible to even assemble the matrix of $\mathcal{DtN}$ in the FEM basis due to lack of memory. Hence, computational techniques have to be found which incorporate the

---

[1]In the thesis also separation with respect to more general operators on Γ will be treated.

information contained in $\mathcal{DtN}$ accurately into the discrete solution, yet remain within the realm of sparse linear algebra. Futhermore, rigorous error estimates are desirable to guarantee the accuracy of the computed solutions. For the case of homogeneous media a variety of such techniques exists, yet reliable and accurate methods for stratified media like the Sun, which exhibit strong variations in the direction of stratification, are lacking. This thesis sets out to propose such techniques which include as much information from $\mathcal{DtN}$ as desired in a sparse manner and are at the same time flexible enough to adapt themselves to almost any given stratified medium.

1.3   Learned infinite elements

Many successful transparent boundary conditions are based on a tensor-product discretization of (1.3). They lead to linear systems of the form

$$\left(\begin{bmatrix} A_{\Gamma\Gamma} & A_{\Gamma E} \\ A_{E\Gamma} & A_{EE} \end{bmatrix} \otimes M + \begin{bmatrix} B_{\Gamma\Gamma} & B_{\Gamma E} \\ B_{E\Gamma} & B_{EE} \end{bmatrix} \otimes K \right) \begin{bmatrix} u_\Gamma \\ u_E \end{bmatrix} = \begin{bmatrix} M\,\mathrm{DtN}_N\,u_\Gamma \\ 0 \end{bmatrix}, \qquad (1.5)$$

where

- M and K are finite element discretizations of Id_Γ and $-\Delta_\Gamma$, respectively.

- The matrices $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ are derived analytically, for example by utilizing expansions of solutions of (1.3) valid in asymptotic regimes.

- In the formula above the degrees of freedom (DOFs) have been partitioned into two subsets: those associated with Γ and the rest associated with the exterior E.

- The operator DtN_N represents the discrete approximation of $\mathcal{DtN}$ obtained with this particular discretization.

The exterior DOFs could be eliminated from (1.5) by means of the Schur complement. Even though this is not a reasonable computational approach since the Schur complement is dense, it leads to a concrete formula for DtN_N in terms of A, B, M and K. This representation is still a bit unwieldy because the matrices M and K are large. One of the main observations of this thesis is that in a suitable basis DtN_N can be described in terms of the small matrices A and B only. This basis is simply the discrete analogue of the Laplace-Beltrami eigenbasis used in the previous section. If $(\underline{\lambda}_\ell, \underline{w}_\ell)$ solve the generalized eigenvalue problem $K\underline{w}_\ell = \underline{\lambda}_\ell M \underline{w}_\ell$, then

$$\mathrm{DtN}_N\,\underline{u}^{\mathrm{int}} := \sum_{\ell=0}^{\infty} \mathrm{dtn}_N(\underline{\lambda}_\ell)\underline{w}_\ell \langle \underline{u}^{\mathrm{int}}, \underline{w}_\ell \rangle_\Gamma, \qquad (1.6)$$

with dtn_N given by

$$\mathrm{dtn}_N(\lambda) := A_{\Gamma\Gamma} + \lambda B_{\Gamma\Gamma} - (A_{\Gamma E} + \lambda B_{\Gamma E})(A_{EE} + \lambda B_{EE})^{-1}(A_{E\Gamma} + \lambda B_{E\Gamma}). \qquad (1.7)$$

This representation, derived in Proposition 3.1 of this thesis, is completely analogous to formula (1.4) for the continuous $\mathcal{DtN}$. This observation gives rise to some of the thesis' main conclusions.

- The continuous $\mathcal{D}t\mathcal{N}$ and its approximation DtN_N with the tensor-product discretization (1.5) are completely determined by the functions $\textit{dtn}$ and dtn_N respectively. Consequently, the approximation quality of DtN_N can be assessed by measuring how accurate the scalar function dtn_N approximates $\textit{dtn}$ on the spectrum of $-\Delta_\Gamma$. Since dtn_N can be described in terms of the matrices A, B, which characterize the transparent boundary condition, this paves the way for a unified error analysis of tensor product discretizations of $\mathcal{D}t\mathcal{N}$.

- The matrices A and B are usually derived analytically in conventional transparent boundary conditions. In this thesis a completely different approach is proposed based on the observation that the approximation quality of DtN_N is essentially determined by the size of the misfit

$$\sum_\ell |\,\textit{dtn}(\lambda_\ell) - \mathrm{dtn}_N(\lambda_\ell)|^2 |\langle u, w_\ell \rangle_\Gamma|^2. \tag{1.8}$$

In practice, only a finite number of terms are relevant in this sum, since $|\langle u, w_\ell \rangle_\Gamma|$ is expected to decay rapidly for sufficiently smooth solutions. Hence, dtn_N has to provide a good approximation of $\textit{dtn}$ on bounded subsets of the spectrum of the Laplace-Beltrami operator. In view of (1.7), this will give rise to a rational approximation problem when A and B are considered as free parameters:

$$\min_{A,B \in \mathbb{C}^{(N+1)\times(N+1)}} \|\,\textit{dtn} - \mathrm{dtn}_N\{A, B\}\|_*, \tag{1.9}$$

where $\|\cdot\|_*$ is an appropriate norm on the space of sequences $(\textit{dtn}(\lambda_\ell))$. By extending $\textit{dtn}$ to a meromorphic function on the complex plane and appealing to results of rational approximation theory, we can show that for a finite number of modes exponential convergence can be achieved, i.e. there exist A and B such that

$$|\,\textit{dtn}(\lambda) - \mathrm{dtn}_N(\lambda)| \lesssim \exp(-cN) \tag{1.10}$$

for some $c > 0$ and all λ in a compact interval of $\mathbb{R}_+$. The minimizers A and B obtained from (1.9) determine the local element matrices of the learned infinite elements (learned IEs).

The first half of this thesis is devoted to a study of learned IEs as transparent boundary conditions. First, it shall be investigated theoretically whether the exponential convergence rates (1.10) for $\textit{dtn}$ carry over to the finite element solution of the interior problem. This can be answered in the affirmative provided that $\textit{dtn}$ and dtn_N fulfill certain natural conditions. Furthermore, numerical experiments will be presented in which the efficiency of learned IEs is compared to established transparent boundary conditions. Particular emphasis shall be placed on applications in helioseismology as the solar atmophere provides a prime example for a strongly varying, stratified medium.

1.4   Application to sweeping preconditioners

Time-harmonic wave equations give rise to large-scale linear systems which are difficult to solve. Domain decomposition (DD) methods are one approach to tackle this issue. They partition the domain into a set of smaller subdomains. The linear systems in

these subdomains can be solved more efficiently (often even using direct solvers) since the subdomains contain substantially fewer DOFs than the entire domain. In order to ensure that the solution obtained from these algorithms represents the solution of the problem posed on the entire domain accurately, appropriate boundary conditions (so called transmission conditions) have to be imposed on the subdomain interfaces. This is analogous to the case of two subdomains treated in (1.2). Some DD algorithms can be used as direct solvers provided that exact transmission conditions, e.g. $\mathcal{DtN}$, are imposed. As transmission operators are typically non-local, they must be replaced by approximations for the sake of computational efficiency. DD algorithms with approximate transmission conditions only deliver approximate solutions and are then usually applied as preconditioners for Krylov iterative solvers.

Sweeping preconditioners can be seen as a special kind of DD method which relies on a sequential partition of the domain into layers. In its original [EY11b] form $\mathcal{DtN}$ transmission conditions approximated by a moving perfectly matched layer (PML) are used on subdomain interfaces, see [GZ19, Section 7]. Even though impressive results have been achieved with sweeping-type preconditioners for many problems of practical interest (see e.g. [TEY12, TPEY14, Sto17, ZNSHD19] to name a few), their range of applicability is limited. In situations where approximation of $\mathcal{DtN}$ by means of PML fails the performance of sweeping preconditioners deteriorates dramatically. This occurs for instance in media that allow for strong reflection of waves, e.g. caused by strongly varying wavespeeds or reflective boundary conditions [GZ18, GZ19, PHL20]. The Sun provides a prominent example of such a medium as waves propagating at frequencies below the acoustic cut-off are reflected by steep density and sound speed gradients in the surface layers. In order to apply sweeping preconditioners for problems involving strong reflections, alternative approximations of $\mathcal{DtN}$ have to be found.

The second half of the thesis investigates whether sweeping preconditioners with learned IEs as transmission conditions can overcome the limitations of moving PML sweeping preconditioners for problems with strong reflections. In this regard, the separability condition on which learned IEs are based could pose an obstacle as can be readily explained for the case of the Sun. The solar atmosphere can be regarded as a stratified medium which allows for a perfect modelling by means of learned IEs. However, in the context of sweeping preconditioners also the solar interior would need to be modelled by learned IEs, which is a problem since the presence of sound speed and density perturbations, background flows or rotation may break spherical symmetry. Therefore, it has to be carefully investigated how learned IEs perform once the separability assumption is violated. It turns out that in the presence of strong reflections only very small perturbations from the stratified background medium can be tolerated. This is a physical obstruction caused by high-sensitivity of $\mathcal{DtN}$ with respect to perturbations.

1.5   Outline

The remainder of the thesis is structured as follows.

- Chapter 2 specifies the class of PDEs (1.1) which will be considered in this thesis. A general framework is introduced which allows to cover Helmholtz scattering problems, waveguides and wave propagation in the Sun at the same time. For each of these examples the continuous $\mathcal{DtN}$ map and its dtn function are identified. The dtn functions in these examples share some common properties which

can be utilized to derive well-posedness of the reduced problem in Ω_{int} with $\mathcal{DtN}$ as transparent boundary condition.

- Chapter 3 can be seen as the semi-discrete counterpart of Chapter 2. Tensor-product discretizations of $\mathcal{DtN}$ of the form (1.5) are considered. We carry out a semi-discrete error analysis to investigate the influence of replacing the exact dtn function in (1.4) by its approximation dtn_N stemming from the tensor-product discretization. By utilizing results of rational approximation theory, it is shown that exponential approximation rates (1.10) on finite intervals can be achieved. For sufficiently smooth data and dtn_N functions fulfilling certain assumptions, convergence of the corresponding solutions on Ω_{int} follows. This means that the accuracy of transparent boundary conditions of tensor-product type is fully determined by approximation properties of their dtn_N function. This motivates us to identify dtn_N for several popular transparent boundary conditions, including tensor product PMLs and classical infinite elements, and compare their accuracy based on the previously mentioned observation.

- Chapter 4 introduces learned IEs as the main innovation of this thesis. Using a weighted ℓ^2-norm, the minimization problem (1.9) transforms into a non-linear least squares problem which can be solved with the Levenberg–Marquardt algorithm. Various numerical experiments illustrate the astonishing accuracy and efficiency of learned IEs. Comparisons with established transparent boundary conditions like PMLs or Hardy Space infinite elements for a circular truncation boundary and homogeneous exterior domains are presented. Furthermore, first applications of learned IEs to inhomogeneous exterior domains which feature strong reflections are investigated. We also apply learned IEs to waveguides and elliptical truncation boundaries. The latter allows for an efficient treatment of elongated obstacles.

- The transparent boundary conditions obtained by solving the minimization problem (1.9) are only valid for a single wavenumber. However, Chapter 5 shows that a uniform approximation for all wavenumbers in a bounded subset of the complex plane can be obtained by modifying the objective function in (1.9). This allows to apply learned IEs for the computation of resonance frequencies.

- Chapter 6 presents an application of learned IEs to time-distance helioseismology. The main task here is to investigate the effect of the transparent boundary conditions on the accuracy of computed helioseismic observables like power spectra or travel times of acoustic waves. On the one hand, we compare our results with observations. On the other hand, the accuracy and efficiency of learned IEs is assessed by a comparsion with so called Atmospheric Radiation Boundary Conditions, which have been specifically designed for a particular model of the solar atmosphere. We also go one step further and demonstrate that learned IEs can be used for another, more realistic model of the solar chromosphere for which no other transparent boundary conditions are known to date.

- In Chapter 7, learned IEs are utilized as transmission conditions in sweeping preconditioners. Before proceeding to numerical experiments, a discussion on the literature is given and the sweeping algorithm is briefly reviewed in the framework of the Double Sweep Optimized Schwarz method (DOSM). The performance of sweeping with moving PML and learned IE based transmission conditions is carefully studied

in numerical experiments especially in the presence of reflections. Our investigations show that most of the arising issues and difficulties can be understood by an analysis of the corresponding *dtn* functions.

- Chapter 8 presents an application of sweeping preconditioners based on learned IEs to a realistic problem from computational helioseismology. This study complements and extends the results of the toy problems considered in Chapter 7. The chapter concludes with an evaluation of the current applicability of learned IEs in sweeping algorithms.

- The thesis finishes with a summary and outlook in Chapter 9. Open problems and ideas for further research are discussed.

- Some appendices which cover technical aspects or provide additional information complement the thesis. Some basics tools from functional analysis and theory of PDEs are collected in Appendix A. Appendix B gives a spectral characterization of Sobolev spaces on the transparent boundary Γ which is utilized in the error analysis of Chapter 2 and Chapter 3. Popular transparent boundary conditions of tensor-product form (1.5) are reviewed in Appendix C. Appendix D describes how the optimization problem (1.9) is solved.

1.6 Genesis of the thesis project

The idea for this project was born out of the desire to construct efficient preconditioners for solving the time-harmonic wave equations from helioseismology. At the beginning of 2018 when the project started, sweeping preconditioners had already proven their worth as efficient solvers in many different applications. Hence, it was natural to make an attempt at using sweeping preconditioners for computing solar oscillations. Right from the start of the project it was suspected that standard sweeping preconditioners based on moving PMLs would run into severe difficulties since the Sun supports resonance modes. In particular, there are complex wavenumbers with small imaginary part for which the homogeneous PDE admits a non-trivial solution. In the best case, the moving PML approximation of $\mathcal{DtN}$ would break down only in close vicinity of such a resonance frequency and would need to be replaced there by a different approach that should be developed within this project.

Unfortunately, the actual situation turned out to be even more problematic. We discovered that the breakdown of moving PMLs was not limited to an isolated and manageable number of particular frequencies. Sweeping preconditioner based on moving PMLs failed across the whole range of frequencies which are of primary interest in helioseismology. The explanation for this phenomenon is indeed very simple. Moving PML-type sweeping preconditioners based on a concentric decomposition of the Sun could only work if the waves were able to escape freely into the solar atmosphere. Unfortunately, this could not be further from the truth as is well-known in helioseismology. Essentially all waves propagating at frequencies lower than the acoustic cut-off are reflected back from the strong stratification in the surface layers into the solar interior. Hence, moving PMLs are not suitable for the Sun at all.

To propagate the observation that moving PMLs fail in the presence of strong reflections, we began to prepare our first paper [PHL20]. The solar application of course also pressed us to start the development of an alternative to moving PMLs which could deal

with the presence of strong reflections. The problem with moving PMLs in this setting is that its associated dtn_N function provides an extremely poor approximation of dtn (see e.g. Section 7.3.2). The approach presented in [PHL20] can be seen as a precursor to learned IEs, which in some sense provides the perfect approximation $\lim_{N \to \infty} \text{dtn}_N = dtn$, yet resorts in the end to solving dense linear systems on subdomain interfaces which learned IEs manage to avoid entirely. Since this method is superseded by learned IEs, it will not be presented in this thesis. In particular, we directly use learned IEs in the sweeping preconditioners presented in Chapter 7 and Chapter 8. Apart from the employed methods which are deprecated by now, the paper [PHL20] offers the important insight that $\mathcal{D}t\mathcal{N}$ maps of media which allow for strong reflections are highly sensitive to perturbations. This implies that no sweeping preconditioner based on the separable background model, in particular neither the method from [PHL20] nor separable learned IEs as presented in this thesis, can tolerate strong perturbations from the background model.

Admittedly, this insight, which we reached in early 2020 when submitting the paper, was a major setback. During this time the development of learned IEs was at an early stage and looked not particularly promising in the context of sweeping preconditioners for helioseismology. Fortunately, there is also a strong need in computational helioseismology for transparent boundary conditions which can model the solar atmosphere (see Section 2.4 and Chapter 6). In this setting the issue of having to treat perturbations from the background model does not occur. Consequently, the primary focus in the development of learned IEs switched from sweeping preconditioners to transparent boundary conditions. Despite originally being intended only for helioseismology, the flexibility and high-accuracy of learned IEs soon came to the fore when applied to various other problems. When we were unable to find any other method that can match the performance of learned IEs for the Helmholtz equation, it was clear that a paper had to be prepared in which learned IEs are presented in a more general context than helioseismology.

Learned IEs are introduced in [HLP21] as transparent boundary conditions for stratified media which allow for a separation of variables with respect to the Laplace-Betrami operator on Γ. This setting covers for example also certain waveguides. Although the paper [HLP21], accepted in July 2021, provides the main foundation for this thesis, the latter extends the results of [HLP21] significantly. A detailed discussion of these extensions is provided in Section 9.1. Here we only mention two of the most important results. Firstly, the application to helioseismology given in [HLP21] only scratches the surface. This thesis devotes its entire Chapter 6 to helioseismology providing extensive investigations on the accuracy of helioseismic observables. Secondly, in [HLP21] we only provided an argument of about one page to argue that an exponential approximation result of the form (1.10) on finite intervals should hold true. In Chapter 2 and Chapter 3 we set out to extend this to a complete convergence theory in a semi-discrete setting. Admittedly, this aim has not been entirely completed yet, see e.g. Section 9.2, albeit compared to [HLP21] a major leap forward has been taken.

In the final Chapter 8 of this thesis we return to the original preconditioning problem for helioseismology in an axisymmetric setting. Although no sudden twist of fate is observed and the GMRES iteration numbers are high, the results are nevertheless slightly better than expected. At least for the sound speed, perturbations from the separable background model could be tolerated which are within a realistic range that has to be expected in helioseismology. Perhaps, even an implementation and benchmarking of learned IEs for a fully three-dimensional discretization then appear as realistic prospects

for the future. Ultimately though, the development of learned IEs which are applicable to non-separable problems seems inevitable in order to treat stronger perturbations. Future extensions of learned IEs are discussed further in Section 9.2.

1.7 Reproducibility

A complete environment for reproducing the numerical experiments of this thesis is provided in the form of a docker image at [Pre21]. In particular, this image contains all software dependencies, code and the necessary instructions to replicate our results. The docker image has been built based on the gitlab repository `https://gitlab.gwdg.de/learned_infinite_elements/diss_jp_repro`, which also contains the code and a clean presentation of the reproduction instructions.

1.7.1 Software used in this thesis

This subsection provides a short overview on the software involved in the numerical experiments of this thesis. The general approach is to rely on established and fast modules written mostly in C++ and extend them for our purposes where necessary. These modules are combined by surrounding them with a common `Python` layer provided by `pybind11` [JRM17].

- The numerical experiments involving finite elements have been realized using the software `Netgen/NGSolve`, see [Sch97, Sch14],

- The experiments in the chapters on sweeping preconditioners make use of `ngsxfem` [LHPvW21], which is an Add-on to `NGSolve` for unfitted discretization methods. Even though this thesis deals exclusively with fitted discretization approaches, some of the tools offered in this Add-on are useful to improve efficiency in a domain-decomposition framework.

- Learned IEs are obtained by solving a non-linear least squares problem. For this purpose we use `ceres-solver` [AMO], which is a C++ library geared towards modelling and solution of exactly these kind of problems.

- The poles of the meromorphic extension of *dtn* play a crucial role in this thesis. Due to the lack of analytic formulas, these poles have to be computed numerically. To this end, we use a mesh-based technique introduced in references [Kow15, Kow18].

- Special functions are evaluated using the libarries `mpmath` [J$^+$21] and `arb` [Joh17].

Chapter 2

DtN maps for time-harmonic waves in separable exterior domains

This chapter considers PDEs posed in domains which allow for a decomposition into a bounded interior part and a possibly unbounded, stratified exterior part whose geometry is separable. The $\mathcal{DtN}$ map obtained by separation of variables is imposed on the truncation boundary to reduce the problem to the bounded interior domain. In its separable form, all information about $\mathcal{DtN}$ is contained in a scalar function called dtn which describes the action of $\mathcal{DtN}$ in direction of stratification. An abstract framework which allows to treat a broad class of scalar, second-order elliptic PDEs in the above setting is introduced in Section 2.1. Moreover, Section 2.1.2 also introduces certain assumptions on dtn which later serve as crucial ingredients for deriving well-posedness of the reduced, interior problem.

Before proceeding to the analysis, it makes sense to show that the class of problems covered by the framework introduced in Section 2.1 is not empty. Therefore, Section 2.2-Section 2.4 present several examples which fit into this framework and will recur frequently in the course of this thesis. Here, one of the main aims is to identify the corresponding dtn functions of these problems and demonstrate that they fulfill a common set of properties which justifies the assumptions postulated in Section 2.1.2. A reasonably broad range of problem settings will be treated. Section 2.2 deals with scattering problems in general, which may be truncated using spherical or elliptical coupling boundaries as considered in Section 2.2.1 and Section 2.2.2, respectively. The exterior domain may contain inhomogeneities, as the example of a jumping wavenumber introduced in Section 2.2.1 shows. Section 2.3 then considers a different class of problems called waveguides. The scalar wave equation of helioseismology is introduced in Section 2.4. Two different models for the solar atmosphere are considered: the so called `Atmo` model, which is based on a simplified chromosphere and the presumably more realistic, semi-empirical VAL-C model introduced in reference [VAL81].

Section 2.5 then investigates well-posedness of the interior problem with $\mathcal{DtN}$ as transparent boundary condition. Our proposed analysis proceeds along conventional lines. The identified properties of dtn turn out to be sufficient for establishing a Gårding inequality. Well-posedness then follows by means of the Fredholm alternative provided that uniqueness has been shown. To this end, assumptions on the imaginary part of dtn are required, which hold for the considered examples with the exception of certain waveguides which are known to be a bit peculiar. Whereas this type of analysis is sufficient to cover the helioseismology problem or scattering with spherical coupling boundaries, it

rests on an assumption on $\mathcal{DtN}$ which is not fulfilled for elliptical truncation boundaries. Generalizations of the presented analysis could thus be the subject of further research.

2.1 Abstract framework

We consider a generalization of the setting introduced in [HLP21]. Let $\Omega \subset \mathbb{R}^d$ be a domain which can be decomposed as follows. We assume that there exists a bounded interior domain Ω_{int} and a usually unbounded exterior domain Ω_{ext} such that

$$\overline{\Omega} = \overline{\Omega}_{\mathrm{int}} \cup \overline{\Omega}_{\mathrm{ext}} \quad \text{and} \quad \Omega_{\mathrm{int}} \cap \Omega_{\mathrm{ext}} = \emptyset.$$

Moreover, the boundary $\Gamma := \overline{\Omega}_{\mathrm{int}} \cap \overline{\Omega}_{\mathrm{ext}}$, which couples interior and exterior domain, is required to be smooth. Additionally, Γ is compact since Ω_{int} is assumed to be bounded. Denote the usual L^2 scalar products on $L^2(\Omega_{\mathrm{int}})$ and $L^2(\Gamma)$ by

$$(u, v)_{\Omega_{\mathrm{int}}} := \int_{\Omega_{\mathrm{int}}} u\bar{v} \, \mathrm{d}x; \quad \langle u, v \rangle_\Gamma := \int_\Gamma u\bar{v} \, \mathrm{d}S. \tag{2.1}$$

In the domain Ω a linear and second-order elliptic PDE

$$\mathcal{L}u = f \qquad \text{in } \Omega \qquad + \qquad \text{radiation condition}, \tag{2.2}$$

is considered. The source f is assumed to be compactly supported in $\overline{\Omega}_{\mathrm{int}}$, i.e. $\mathrm{supp}\, f \subset \overline{\Omega}_{\mathrm{int}}$. This implies that the PDE in the exterior Ω_{ext} is homogeneous. Since the radiation condition at infinity depends on the geometrical setup and the considered PDE, we postpone its discussion to Section 2.2-Section 2.4 in which concrete examples are discussed. Apart from the artificially introduced boundary Γ, the domain Ω may have other boundaries (possibly $\partial\Omega \neq \emptyset$). In this case, equation (2.2) has to be complemented by suitable boundary conditions on $\partial\Omega$, which are discussed further below.

To describe the $\mathcal{DtN}$ map in a convenient form, the Dirichlet problem posed in the exterior

$$\begin{aligned} \mathcal{L}u &= 0 \quad \text{in } \Omega_{\mathrm{ext}} \qquad + \qquad \text{radiation condition}, \\ u &= u_0 \quad \text{on } \Gamma, \end{aligned} \tag{2.3}$$

where u_0 represents the given data, has to be simplified. This rests on separability of the PDE and the geometry of Ω_{ext}.

To make the latter notion precise, we will suppose that a diffeomorphism

$$\begin{aligned} \Psi : &[a, \infty) \times \Gamma \to \overline{\Omega}_{\mathrm{ext}} \text{ such that} \\ &\Psi(\{a\} \times \Gamma) = \Gamma, \end{aligned} \tag{2.4}$$

exists. Moreover, the transformed operator is required to take the following form[1]

$$(\mathcal{L}u) \circ \Psi = 0 \Leftrightarrow [\mathcal{A} \otimes \mathcal{M}_\Gamma + \mathcal{B} \otimes \mathcal{K}_\Gamma] (u \circ \Psi) = 0. \tag{2.5}$$

This is nothing but separation of variables by a transformation to suitable coordinates (see e.g. Section 2.2.1 for the case of spherical coordinates). In the general setting considered here,

---

[1]We do not demand that $(\mathcal{L}u) \circ \Psi = [\mathcal{A} \otimes \mathcal{M}_\Gamma + \mathcal{B} \otimes \mathcal{K}_\Gamma] (u \circ \Psi)$ here as in the paper [HLP21] to gain a bit more flexibility for the setting of elliptic geometry in which $(\mathcal{L}u) \circ \Psi = \frac{-1}{c^2(\sinh(r)^2 + \sin(\varphi)^2)}(\partial_r^2 + \partial_\varphi^2)u - k^2u$ and we actually treat $(\partial_r^2 + \partial_\varphi^2)u - c^2(\sinh(r)^2 + \sin(\varphi)^2)k^2u = 0$ to separate variables.

- $\mathcal{A}$ is a second-order differential operator in r,

- $\mathcal{B}$ represents multiplication on $L^2([a, \infty))$,

- $\mathcal{K}_\Gamma$ is a second order, self-adjoint, elliptic differential operator on Γ,

- and $\mathcal{M}_\Gamma = 1/|\partial_r \Psi(a, \widehat{x})|_2$ is a multiplication operator on Γ. Since Ψ is assumed to be a diffeomorphism and Γ is compact there exist positive constants $c_l^\psi > 0$ and $c_u^\psi > 0$ such that

$$c_l^\psi \leq \frac{1}{|\partial_r \Psi(a, \widehat{x})|_2} \leq c_u^\psi. \tag{2.6}$$

In particular, $\mathcal{M}_\Gamma$ is positive definite and $(u, v) \mapsto \langle \mathcal{M}_\Gamma^{1/2} u, \mathcal{M}_\Gamma^{1/2} v \rangle_\Gamma$ gives rise to an equivalent scalar product on $L^2(\Gamma)$.

The tensor-product notation '$\otimes$' used above should be understood as follows. If u and v are L^2-functions, then $(u \otimes v)(x, y) = u(x)v(y)$. For linear operators A and B one has $(A \otimes B)(u \otimes v) = Au \otimes Bv$. This means that the operator A acts on the first variable, while B operates on the second variable. Hence, in (2.5) the operators $\mathcal{A}$ and $\mathcal{B}$ act on r in the direction of stratification while $\mathcal{M}_\Gamma$ and $\mathcal{K}_\Gamma$ operate on Γ in the tangential direction.

By the rule $(AB)^* = B^* A^*$ the operator

$$\tilde{\mathcal{K}}_\Gamma := \mathcal{M}_\Gamma^{-1/2} \mathcal{K}_\Gamma \mathcal{M}_\Gamma^{-1/2} \tag{2.7}$$

is self-adjoint. Under the assumption that $\tilde{\mathcal{K}}_\Gamma$ has compact resolvent, the spectral theorem yields the existence of an orthonormal basis $\{w_\ell : \ell \in \mathbb{N}_0\}$ of $L^2(\Gamma)$ consisting of eigenfunctions of $\tilde{\mathcal{K}}_\Gamma$, i.e. there exists $\lambda_\ell \in \mathbb{R}$ such that

$$\tilde{\mathcal{K}}_\Gamma w_\ell = \lambda_\ell w_\ell. \tag{2.8}$$

The functions $v_\ell := \mathcal{M}_\Gamma^{-1/2} w_\ell$ are then generalized eigenfunctions of the operator stencil $(\mathcal{M}_\Gamma, \mathcal{K}_\Gamma)$, i.e.

$$\mathcal{K}_\Gamma v_\ell = \lambda_\ell \mathcal{M}_\Gamma v_\ell. \tag{2.9}$$

The right equation in (2.5) may be rewritten as

$$\left[\mathcal{A} \otimes \mathrm{Id}_\Gamma + \mathcal{B} \otimes \tilde{\mathcal{K}}_\Gamma \right] (u \circ \Psi) = 0. \tag{2.10}$$

This equation can be solved by separation of variables in the coordinates $(r, \widehat{x})$. To determine the solution of $\mathcal{L}u = 0$ in Ω_{ext} which fulfills the radiation condition at infinity, we make the ansatz

$$(u \circ \Psi)(r, \cdot) = \sum_{\ell=0}^{\infty} \Lambda_r(\lambda_\ell) \langle u_0, w_\ell \rangle_\Gamma w_\ell \tag{2.11}$$

for given Dirichlet data $u|_\Gamma = u_0$. Here, the functions $r \mapsto \Lambda_r(\lambda_\ell)$ have to fulfill $\Lambda_a(\lambda_\ell) = 1$ for the Dirichlet condition on Γ to hold. Plugging this ansatz into (2.10) leads to the separated set of equations

$$[\mathcal{A} + \lambda_\ell \mathcal{B}] \Lambda_r(\lambda_\ell) = 0 \qquad \text{on } [a, \infty), \tag{2.12a}$$

$$\Lambda_a(\lambda_\ell) = 1, \qquad \Lambda_r \text{ satisfies the radiation condition.} \tag{2.12b}$$

This fully determines the radiating solution. In the sense of spectral calculus we may write $(u \circ \Psi)(r, \cdot) = \Lambda_r(\tilde{\mathcal{K}}_\Gamma) u_0$. Based on the obtained representation of the solution, the Dirichlet-to-Neumann map can be defined by

$$\mathcal{D}t\mathcal{N}u_0 := -\partial_r(u \circ \Psi)|_{r=a} = \sum_{\ell=0}^{\infty} \mathit{dtn}(\lambda_\ell)\langle u_0, w_\ell\rangle_\Gamma w_\ell, \qquad (2.13)$$

where the function

$$\mathit{dtn}(\lambda) := -\partial_r \Lambda_r(\lambda)|_{r=a}. \qquad (2.14)$$

is initially defined on the spectrum of the operator $\tilde{\mathcal{K}}_\Gamma$.

2.1.1 PDE in bounded interior

In the bounded interior domain Ω_{int}, a linear elliptic PDE of second order is considered:

$$\mathcal{L}u = -\sum_{i,j=1}^{d} \partial_{x_i}\left(a_{ij}\partial_{x_j}u\right) + i\beta\mathbf{b}\cdot\nabla(\beta u) + cu = f \qquad \text{in } \Omega_{\mathrm{int}}. \qquad (2.15)$$

The principal part of $\mathcal{L}$ is assumed to be smooth and real-valued, i.e. $a_{ij} \in C^\infty(\bar{\Omega}_{\mathrm{int}}, \mathbb{R})$. For the lower order terms $\beta \in C^1(\bar{\Omega}_{\mathrm{int}}, \mathbb{R})$ and $\mathbf{b}_j \in L^\infty(\Omega_{\mathrm{int}}, \mathbb{R})$ for $j = 1, \ldots, d$ and $c \in L^\infty(\Omega_{\mathrm{int}}, \mathbb{C})$ will be required. The right hand side should be square integrable, i.e. $f \in L^2(\Omega_{\mathrm{int}}, \mathbb{C})$. The matrix $(a_{ij})_{i,j=1}^d$ is assumed to be symmetric, which implies that $\sum_{i,j=1}^{d} a_{ij}\xi_i\bar{\xi}_j$ is real for any $\xi \in \mathbb{C}^d$. Additional assumptions are introduced below:

- $\mathcal{L}$ is strongly elliptic, i.e. there exists $\theta > 0$ such that

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\bar{\xi}_j \geq \theta\|\xi\|^2, \text{ for all } \xi \in \mathbb{C}^d \text{ and } x \in \Omega_{\mathrm{int}}. \qquad \text{(C-I)}$$

- The flow is divergence-free and its normal component vanishes on the boundary of the computational domain:

$$\mathrm{div}(\mathbf{b}) = 0 \text{ in } \Omega_{\mathrm{int}}, \quad \mathbf{b}\cdot\mathbf{n} = 0 \text{ on } \partial\Omega_{\mathrm{int}}. \qquad \text{(C-II)}$$

- If damping is present, it should have the appropriate sign to ensure uniqueness:

$$\mathrm{Im}\, c \leq 0 \text{ in } \Omega_{\mathrm{int}}. \qquad \text{(C-III)}$$

- The boundary of the interior domain can be decomposed into disjoint parts $\Gamma, \Gamma_{\mathrm{D}}$ and Γ_{N} so that

$$\partial\Omega_{\mathrm{int}} = \Gamma \cup \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}}, \qquad (2.16)$$

where the $\mathcal{D}t\mathcal{N}$ condition is placed on $\Gamma \neq \emptyset$. Here, Γ_{D} and Γ_{N} are additional (possibly empty) Lipschitz continuous boundaries on which Dirichlet and Neumann boundary conditions are posed, respectively. For simplicity, it will be assumed here that the boundary conditions on Γ_{D} are homogeneous.

- The diffusion matrix is proportional to the identity on Γ, i.e. there exist a nonzero constant a_0 such that $a_{ij}(\hat{x}) = a_0\delta_{ij}$ for $\hat{x} \in \Gamma$. Upon multiplying the PDE (2.2) by $1/a_0$ it can therefore be assumed that

$$a_{ij}(\hat{x}) = \delta_{ij}, \quad \hat{x} \in \Gamma. \tag{C-IV}$$

Apart from normalizing the PDE, the scaling with a_0 could also be incorporated in various alternative ways. For example, the factor a_0 may absorbed into the definition of $\mathcal{DtN}$ via $dtn \mapsto a_0 \, dtn$.

Remark 2.1. *Since the normal vector n_Γ on Γ is given by*

$$\frac{(\partial_r \Psi)(a, \widehat{x})}{|\partial_r \Psi(a, \widehat{x})|_2} = \mathbf{n}_\Gamma \circ \Psi(a, \hat{x}), \tag{2.17}$$

an application of the chain rule gives

$$\mathcal{M}_\Gamma \partial_r (u \circ \Psi)(a, \hat{x}) = (\nabla u)\left(\Psi(a, \hat{x})\right) \cdot \frac{(\partial_r \Psi)(a, \widehat{x})}{|\partial_r \Psi(a, \widehat{x})|_2} = (\nabla u)\left(\Psi(a, \hat{x})\right) \cdot \left(\mathbf{n}_\Gamma \circ \Psi(a, \hat{x})\right).$$

Inserting the definition of $\mathcal{DtN}$ then yields

$$\mathcal{M}_\Gamma \mathcal{DtN}(u) = -\nabla u \cdot \mathbf{n}_\Gamma \; \text{ on } \Gamma. \tag{2.18}$$

This explains the name of the $\mathcal{DtN}$ map: it maps Dirichlet to Neumann data.

For $\tilde{\Gamma} \in \{\Gamma, \Gamma_\mathrm{D}, \Gamma_\mathrm{N}\}$ and $v \in H^1(\Omega_\mathrm{int})$ denote by $\mathrm{tr}_{\tilde{\Gamma}} v \in H^{1/2}(\tilde{\Gamma})$ the trace of v on $\tilde{\Gamma}$. Further, let $Z_{\tilde{\Gamma}} : H^{1/2}(\tilde{\Gamma}) \to H^1(\Omega_\mathrm{int})$ denote a continuous linear right inverse of $\mathrm{tr}_{\tilde{\Gamma}}$, see Theorem A.3. If the boundary $\tilde{\Gamma}$ in question is clear from context the subscript $\tilde{\Gamma}$ will be omitted.

The function space for the variational formulation in the interior is given by

$$V := \{v \in H^1(\Omega_\mathrm{int}) \mid \mathrm{tr}_{\Gamma_\mathrm{D}} v = 0\}. \tag{2.19}$$

It would also be possible to allow for periodic boundary conditions on part of $\partial\Omega_\mathrm{int}$. To this end, the space $H^1(\Omega_\mathrm{int})$ in (2.19) should be replaced by an appropriate periodic Sobolev space $H^1_\mathrm{per}(\Omega_\mathrm{int})$ which incorporates the periodicity requirement. Multiplying the equation (2.15) by a test function and integrating by parts over Ω_int leads to

$$b_\mathrm{int}(u, v) + \langle -\nabla u \cdot \mathbf{n}_\Gamma, v \rangle_\Gamma = l(v),$$

where assumption (C-IV) has been used and the sesquilinear form on Ω_int is defined by

$$b_\mathrm{int}(u, v) := \sum_{i,j=1}^d (a_{ij}\partial x_j u, \partial x_i v)_{\Omega_\mathrm{int}} + i\left(\beta \mathbf{b} \cdot \nabla\left(\beta u\right), v\right)_{\Omega_\mathrm{int}} + (cu, v)_{\Omega_\mathrm{int}} \tag{2.20}$$

and $l : V \to \mathbb{C}$ is a continuous (anti-)linear functional. The unknown normal derivative $-\nabla u \cdot \mathbf{n}_\Gamma$ is replaced by the Dirichlet-to-Neumann map according to equation (2.18). The variational formulation takes the form: Find $u \in V$ such that

$$b_\mathrm{int}(u, v) + \langle \mathcal{M}_\Gamma \mathcal{DtN} u, v \rangle_\Gamma = l(v), \text{ for all } v \in V.$$

The required assumptions are already sufficient to establish a Gårding inequality for $b_\mathrm{int}(\cdot, \cdot)$. On the other hand, gaining control over the part involving $\mathcal{DtN}$ requires assumptions on dtn and will have postponed to Section 2.5.

Lemma 2.2 (Gårding inequality for $b_{\mathrm{int}}(\cdot, \cdot)$). *The sesquilinear form $b_{int}(\cdot, \cdot)$ is continuous on $V \times V$ and fulfills the lower bound*

$$\operatorname{Re} b_{\mathrm{int}}(u, u) \geq C_V \|u\|_V^2 - C_{L^2(\Omega_{\mathrm{int}})} \|u\|_{L^2(\Omega_{\mathrm{int}})}^2, \qquad (2.21)$$

with

$$C_V := \theta/2, \text{ and } C_{L^2(\Omega_{\mathrm{int}})} := \frac{\|\beta\|_{L^\infty}^4 \|\mathbf{b}\|_{L^\infty}^2}{2\theta} + \|\beta\|_{L^\infty} \|\mathbf{b} \cdot \nabla \beta\|_{L^\infty} + \|c\|_{L^\infty} + \frac{\theta}{2}. \quad (2.22)$$

Proof. Continuity follows by the regularity respectively boundedness assumptions on the coefficients of $\mathcal{L}$. To show the lower bound the ellipticity assumption (C-I) can be used to obtain

$$\operatorname{Re} b_{\mathrm{int}}(u, u) \geq \theta \|\nabla u\|_{L^2(\Omega_{\mathrm{int}})}^2 - \|\beta\|_{L^\infty(\Omega_{\mathrm{int}})}^2 \|\mathbf{b}\|_{L^\infty(\Omega_{\mathrm{int}})} \|\nabla u\|_{L^2(\Omega_{\mathrm{int}})} \|u\|_{L^2(\Omega_{\mathrm{int}})}$$
$$- \left(\|\beta\|_{L^\infty(\Omega_{\mathrm{int}})} \|\mathbf{b} \cdot \nabla \beta\|_{L^\infty(\Omega_{\mathrm{int}})} + \|c\|_{L^\infty(\Omega_{\mathrm{int}})} \right) \|u\|_{L^2(\Omega_{\mathrm{int}})}^2$$
$$\geq \left(\theta - \frac{\varepsilon}{2} \right) \|\nabla u\|_{L^2(\Omega_{\mathrm{int}})}^2 - \left(\frac{\|\beta\|_{L^\infty}^4 \|\mathbf{b}\|_{L^\infty}^2}{2\varepsilon} + \|\beta\|_{L^\infty} \|\mathbf{b} \cdot \nabla \beta\|_{L^\infty} + \|c\|_{L^\infty} \right) \|u\|_{L^2(\Omega_{\mathrm{int}})}^2,$$

where Young's inequality

$$\|\beta\|_{L^\infty}^2 \|\mathbf{b}\|_{L^\infty} \|\nabla u\|_{L^2(\Omega_{\mathrm{int}})} \|u\|_{L^2(\Omega_{\mathrm{int}})} \leq \frac{\varepsilon}{2} \|\nabla u\|_{L^2(\Omega_{\mathrm{int}})}^2 + \frac{\|\beta\|_{L^\infty}^4 \|\mathbf{b}\|_{L^\infty}^2}{2\varepsilon} \|u\|_{L^2(\Omega_{\mathrm{int}})}^2$$

for $\varepsilon > 0$ has been employed. Choosing $\varepsilon = \theta$ and adding an additional zero $(\pm \theta/2) \|u\|_{L^2(\Omega_{\mathrm{int}})}^2$ gives (2.21) with C_V and $C_{L^2(\Omega_{\mathrm{int}})}$ as defined in (2.22). □

2.1.2 Assumptions on *dtn*

In order to have a chance of deriving well-posedness for the interior problem with $\mathcal{DtN}$ as boundary condition, some assumptions on *dtn* need to be imposed. They essentially consist of growth conditions and positive or negative definiteness of the real and imaginary parts, respectively.

- For all $\ell \in \mathbb{N}_0$:

$$\operatorname{Im} dtn(\lambda_\ell) < 0, \qquad\qquad (\textit{dtn-}\mathrm{I}_<)$$
$$\operatorname{Im} dtn(\lambda_\ell) \leq 0. \qquad\qquad (\textit{dtn-}\mathrm{I}_\leq)$$

- There exists a finite (possibly empty) set $L_- \subset \mathbb{N}_0$ with complement $L_+ := \mathbb{N}_0 \setminus L_-$ such that

$$\operatorname{Re} dtn(\lambda_\ell) \geq 0 \text{ for } \ell \in L_+. \qquad\qquad (\textit{dtn-}\mathrm{II})$$

- There exists a constant $C_{\mathrm{III}} \geq 0$ such that

$$| dtn(\lambda_\ell) | \leq C_{\mathrm{III}} (1 + |\lambda_\ell|)^{1/2} \text{ for all } \ell \in \mathbb{N}_0. \qquad\qquad (\textit{dtn-}\mathrm{III})$$

At this point of the presentation, these assumptions seem to appear from nowhere. The remainder of this chapter aims to demonstrate that these are the appropriate properties to work with from an application and mathematical analysis point of view. Firstly, Section 2.2-Section 2.4 identify *dtn* for several examples of stratified media for all of

16

which said assumptions are observed to be satisfied using numerical investigations. For our most important examples we also present analytical proofs of these assumptions. Secondly, Section 2.5 sets out to investigate well-posedness of the reduced, interior problem for the class of *dtn* functions fulfilling these properties. In this vein, the precise mathematical role of each individual assumption will be identified.

- The condition (*dtn*-I$_<$)/ (*dtn*-I$_\leq$) is related to uniqueness.

- The requirement (*dtn*-II) on the real part will be needed for establishing a Gårding inequality. To give a physical interpretation of this condition, note that

$$\partial_r \left(|\Lambda_r(\lambda_\ell)|^2 \right) = \Lambda_r'(\lambda_\ell)\bar{\Lambda}_r(\lambda_\ell) + \Lambda_r(\lambda_\ell)\bar{\Lambda}_r'(\lambda_\ell). \tag{2.23}$$

  Evaluating this expression at $r = a$ leads to

$$\partial_r \left(|\Lambda_r(\lambda_\ell)|^2 \right)\big|_{r=a} = -2\,\mathrm{Re}\,\textit{dtn}(\lambda_\ell),$$

  since $\Lambda_a(\lambda_\ell) = 1$ and $\textit{dtn}(\lambda_\ell) = -\Lambda_a'(\lambda_\ell)$ by definition. Hence, the requirement (*dtn*-II) means that for $\ell \in L_+$ the amplitude of the modes decreases towards infinity.

- For Γ compact without boundary the norm in the Sobolev spaces on Γ is characterized in Appendix B as

$$\|u\|_{H^s(\Gamma)}^2 \simeq \sum_{\ell=0}^{\infty} (1 + |\lambda_\ell|)^s \, |\langle u, w_\ell \rangle_\Gamma|^2, \quad s \in \mathbb{R}. \tag{2.24}$$

  Therefore, the growth condition (*dtn*-III) is equivalent to the requirement that $\mathcal{DtN}$ maps $H^{1/2}(\Gamma)$ boundedly into $H^{-1/2}(\Gamma)$.
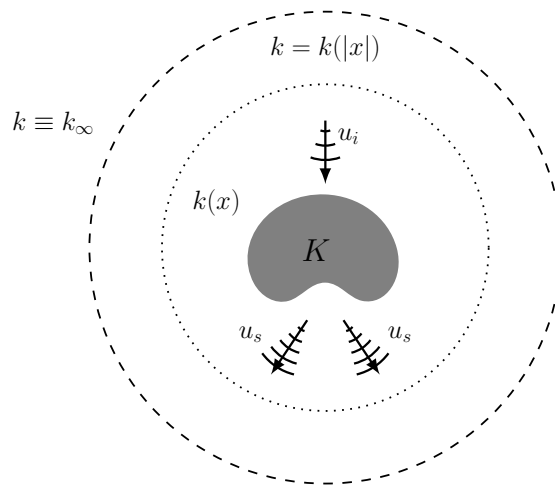
2.2 Scattering problems

Figure 2.1: Basic setup of direct scattering problem.

We consider acoustic waves that are scattered by an inhomogeneous medium $k(x)$ or a compact obstacle $K \subset \mathbb{R}^d$, see Fig. 2.1 for a sketch and e.g. [CK13] for more

17

information on scattering problems. It is assumed that the wavenumber is constant outside a possibly large ball, i.e. there exists $R_J > 0$ such that $k(x) = k_\infty$ for $|x| > R_J$. Let $u^i(x) = \exp(ik_\infty x \cdot v)$ for some $v \in \mathbb{R}^d$ be an incident wave. The sound-hard, direct scattering problem consists of finding the total field $u(x) = u^i(x) + u^s(x)$ in $\Omega := \mathbb{R}^d \setminus K$ such that

$$-\Delta u - k^2(x)u = 0 \qquad \text{in } \Omega, \qquad (2.25a)$$

$$\frac{\partial u}{\partial \mathbf{n}_K} = 0 \qquad \text{on } \partial K, \qquad (2.25b)$$

$$\lim_{|x| \to \infty} |x|^{(d-1)/2} \left(\frac{\partial u}{\partial |x|} - ik_\infty u \right) = 0, \qquad (2.25c)$$

where $\mathbf{n}_K$ denotes the exterior normal vector of K. Equation (2.25c) is the Sommerfeld radiation condition which characterizes outgoing waves for $\text{Im}(k_\infty) \geq 0$ as is henceforth assumed. The Neumann boundary condition (2.25b) can alternatively be replaced by a Dirichlet boundary condition at the scatterer. This is called sound-soft scattering. This problem fits into the framework introduced in Section 2.1 with $a_{ij} = \delta_{ij}, \beta = 0$ and $c = -k^2$. It is naturally posed on an infinite domain which needs to be truncated for computational purposes. To this end, truncation boundaries of two different geometrical shapes will be considered. Section 2.2.1 deals with the spherical or circular case and Section 2.2.2 with ellipses.

2.2.1 Spherical truncation boundary

Transparent boundary conditions for homogeneous media would need to be constrained to set the truncation boundary at $|x| \geq R_J$ where the wavenumber is constant. However, the introduced framework allows for more generality. It is possible to treat a wavenumber which is spherically symmetric, i.e. $k(x) = k(|x|)$ outside a smaller ball $|x| \geq a$ for $a \leq R_J$ as sketched in Fig. 2.1 and set the truncation boundary at $|x| = a$. The diffeomorphism (2.4) for spherical truncation boundaries $\Gamma = a\mathbb{S}^{d-1}$ is given by

$$\Psi(r, \widehat{x}) := \frac{r}{a}\widehat{x} \qquad \text{for } r \in [a, \infty), \, \widehat{x} \in \Gamma. \qquad (2.26)$$

It fulfills $|\partial_r \Psi(a, \widehat{x})|_2 = 1$. The equation transforms in these coordinates to

$$\left(-\Delta u - k^2 u \right)(\Psi(r, \widehat{x})) = \left(-r^{1-d}\partial_r(r^{d-1}\partial_r) - \frac{a^2}{r^2}\Delta_\Gamma - k^2(r)\,\text{Id} \right)(u \circ \Psi)(r, \widehat{x}). \qquad (2.27)$$

The operators in (2.5) can then be identified as

$$\mathcal{K}_\Gamma = -\Delta_\Gamma, \qquad\qquad \mathcal{M}_\Gamma = \text{Id}_\Gamma,$$

$$\mathcal{A} = -r^{1-d}\partial_r(r^{d-1}\partial_r) - k^2(r)\,\text{Id}, \qquad\qquad \mathcal{B} = \frac{a^2}{r^2}.$$

Here, $\mathcal{K}_\Gamma$ is the Laplace-Beltrami operator on Γ, which is self-adjoint with compact resolvent. In two dimensions its eigenfunctions are given by trigonometric polynomials and in three dimensions by spherical harmonics. The corresponding eigenvalues are $\lambda_\ell = (\ell/a)^2$ on the circle and $\lambda_\ell = \ell(\ell+1)/a^2$ on the two-dimensional sphere. In the next sections the *dtn* functions are identified for important special cases.

18

Homogeneous medium $d = 2$

Consider the two dimensional case. If the medium in the exterior is homogeneous, i.e. $k(r) = k = k_\infty$ for $r \geq a$, then the solution satisfying (2.12b) is given by $\Lambda_r(\lambda) = H^{(1)}_{a\sqrt{\lambda}}(kr)/H^{(1)}_{a\sqrt{\lambda}}(ka)$, where $H^{(1)}_{a\sqrt{\lambda}}$ are the Hankel functions of the first kind of order $a\sqrt{\lambda}$. This leads to the *dtn*-function

$$dtn^{\text{hom,2d}}(\lambda) = \frac{-k}{H^{(1)}_{a\sqrt{\lambda}}(ka)}(H^{(1)}_{a\sqrt{\lambda}})'(ka). \tag{2.28}$$

The prime denotes the derivative with respect to the argument of the Hankel function. A plot of $dtn^{\text{hom,2d}}$ for $k = 16$ and $a = 1$ is shown in Figure 2.2. It can be clearly seen that the assumptions (*dtn*-I$_<$), (*dtn*-II) and (*dtn*-III) are fulfilled. The next lemma indeed proves this analytically for general $k > 0$ and $a > 0$.

Lemma 2.3. *Let $k > 0$ and $a > 0$. Then*

(a) Im $dtn^{\text{hom,2d}}(\lambda) < 0$ *for all $\lambda \geq 0$ and* Im $dtn^{\text{hom,2d}}(\lambda) \to -0$ *as $\lambda \to +\infty$.*

(b) *There exists a constant $C(k, a) > 0$ such that*

$$\text{Re } dtn^{\text{hom,2d}}(\lambda = 0) = C(k, a) > 0.$$

(c) *For $\lambda \geq 0$ it holds that*

$$\text{Re } dtn^{\text{hom,2d}}(\lambda) = \sqrt{\lambda} - k \, r_{a\sqrt{\lambda}}(ka),$$

with $|r_{a\sqrt{\lambda}}(ka)| \leq 1$ for $a\sqrt{\lambda} \geq 1$.

Proof. For ease of notation set $\nu = a\sqrt{\lambda}$ and $x = ka$. Then

$$dtn^{\text{hom,2d}}(\lambda) = -\frac{k}{|H^{(1)}_\nu(x)|^2}\left[(J_\nu(x)J'_\nu(x) + Y_\nu(x)Y'_\nu(x)) + iW\{J_\nu(x), Y_\nu(x)\}\right], \tag{2.29}$$

where J_ν and Y_ν denote the Bessel functions of the first and second kind, respectively, and $W\{J_\nu(x), Y_\nu(x)\} := J_\nu(x)Y'_\nu(x) - J'_\nu(x)Y_\nu(x)$ is the Wronskian.

(a) Using [AS64, equation (9.1.16)], i.e.

$$W\{J_\nu(x), Y_\nu(x)\} = \frac{2}{\pi x},$$

yields

$$\text{Im } dtn^{\text{hom,2d}}(\lambda) = -\frac{2k}{\pi x}\frac{1}{|H^{(1)}_\nu(x)|^2}.$$

According to Nicholson's integral formula (see [Wat22, chapter 13-73] for a proof), we have

$$|H^{(1)}_\nu(x)|^2 = \frac{8}{\pi^2}\int_0^\infty K_0(2x\sinh(t))\cosh(2\nu t)\,\mathrm{d}t, \tag{2.30}$$

where K_μ is the modified Bessel function of the second kind of order μ. K_0 is a positive function on $\mathbb{R}_+$. Hence, the integrand is strictly positive, i.e. $|H^{(1)}_\nu(x)|^2 > 0$, so that Im $dtn^{\text{hom,2d}}(\lambda) < 0$ for all $\lambda \geq 0$ follows. Furthermore, $\cosh(2\nu t) \to \infty$ as $\nu \to \infty$ implies Im $dtn^{\text{hom,2d}}(\lambda) \to -0$ as $\lambda \to +\infty$.

19

(b) Using $J_0'(x) = -J_1(x)$ and $Y_0'(x) = -Y_1(x)$ (see [AS64, equation (9.1.28)]) gives

$$\operatorname{Re} \mathit{dtn}^{\mathrm{hom,2d}}(\lambda = 0) = k\frac{(J_1(x)J_0(x) + Y_1(x)Y_0(x))}{|H_\nu^{(1)}(x)|^2}.$$

According to [Fre18, Section 4.1], the nominator can be written as

$$J_1(x)J_0(x) + Y_1(x)Y_0(x) = \frac{8}{\pi^2}\int_0^\infty K_1(2x\sinh(t))\sinh(t)\,\mathrm{d}t.$$

The integrand is positive on $\mathbb{R}_+$ and so is the right hand side. Combining this with $|H_\nu^{(1)}(x)|^2 > 0$ as seen in (a) yields the claim.

(c) Using recurrence relations [AS64, equation (9.1.27)] for the derivatives gives

$$\operatorname{Re} \mathit{dtn}^{\mathrm{hom,2d}}(\lambda) = \frac{k\nu}{x} - kr_\nu(x),$$

with

$$r_\nu(x) := \frac{J_\nu(x)J_{\nu-1}(x) + Y_\nu(x)Y_{\nu-1}(x)}{|H_\nu^{(1)}(x)|^2}.$$

For θ_ν being the argument of $H_\nu^{(1)}(x)$, i.e. $\theta_\nu = \arg H_\nu^{(1)}(x)$, it holds according to [AS64, equation (9.2.19)] that

$$J_\nu(x) = |H_\nu^{(1)}(x)|\cos(\theta_\nu), \qquad Y_\nu(x) = |H_\nu^{(1)}(x)|\sin(\theta_\nu).$$

Therefore,

$$r_\nu(x) = \frac{|H_{\nu-1}^{(1)}(x)|}{|H_\nu^{(1)}(x)|}\cos(\theta_\nu - \theta_{\nu-1}).$$

From Nicholson's integral formula (2.30) it follows that $|H_{\nu-1}^{(1)}(x)| \leq |H_\nu^{(1)}(x)|$ for $\nu \geq 1$, which implies $|r_\nu(x)| \leq 1$ for $\nu \geq 1$ and concludes the proof.

$\square$

To begin with, the domain of the function $\mathit{dtn}^{\mathrm{hom,2d}}$ coincides with the spectrum of the Laplace-Beltrami operator on Γ. The next proposition first shows that $\mathit{dtn}^{\mathrm{hom,2d}}$ admits an extension to a meromorphic function on $\mathbb{C}$ and then investigates its pole structure. The proposition is essentially obtained as a corollary from the results of Magnus and Kotin [MK60] and Cochran [Coc65] about zeros of Hankel functions as functions of their order.

Proposition 2.4 (Meromorphic structure of $\mathit{dtn}^{\mathrm{hom,2d}}$). *Let $z := ka$ such that $\operatorname{Re}(z) > 0$ and $\operatorname{Im}(z) \geq 0$.*

(a) *We can extent $\lambda \mapsto \mathit{dtn}^{\mathrm{hom,2d}}(\lambda)$ to a meromorphic function on $\mathbb{C}$.*

(b) *The poles of $\mathit{dtn}^{\mathrm{hom,2d}}$ are all simple.*
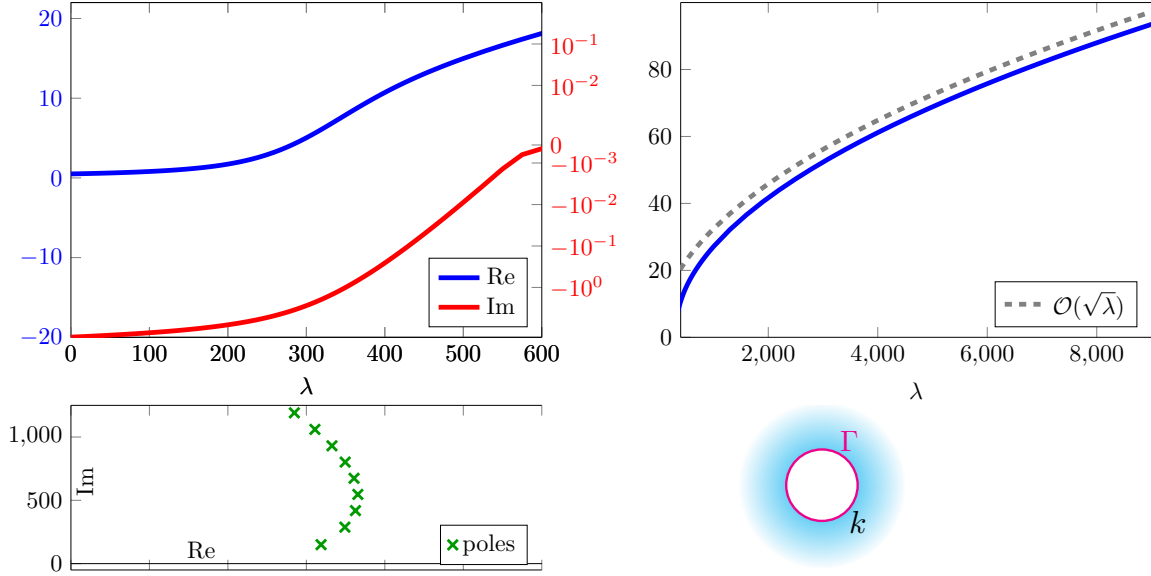
Figure 2.2: The $dtn^{\text{hom,2d}}$ function and its poles for $k = 16$ and $a = 1$.

(c) Let λ_{p}^j and λ_{r}^j for $j = 1, 2, \ldots$ denote the roots of $\lambda \mapsto H^{(1)}_{a\sqrt{\lambda}}(z)$ and $\lambda \mapsto (H^{(1)})'_{a\sqrt{\lambda}}(z)$, respectively. The following expansion into a sum of simple poles holds:

$$dtn^{\text{hom,2d}}(\lambda) = \lim_{N \to \infty} S_N(\lambda) \qquad (2.31)$$

for

$$S_N(\lambda) := -k \frac{(H_0^{(1)})'(z)}{H_0^{(1)}(z)} \left[\sum_{n=1}^{N} \frac{P_N(\lambda_{\mathrm{p}}^n)}{Q_N(\lambda_{\mathrm{p}}^n)} + \sum_{n=1}^{N} \frac{P_N(\lambda_{\mathrm{p}}^n)(\lambda_{\mathrm{p}}^n - \lambda_{\mathrm{r}}^N)}{Q_N(\lambda_{\mathrm{p}}^n)} \frac{1}{\lambda - \lambda_{\mathrm{p}}^n} \right], \qquad (2.32)$$

where

$$P_N(\lambda) := \frac{\lambda_{\mathrm{p}}^N}{\lambda_{\mathrm{r}}^N} \prod_{n=1}^{N-1} (\lambda - \lambda_{\mathrm{r}}^n) \frac{\lambda_{\mathrm{p}}^n}{\lambda_{\mathrm{r}}^n}, \qquad Q_N(\lambda_{\mathrm{p}}^n) := \prod_{j=1, j \neq n}^{N} (\lambda_{\mathrm{p}}^n - \lambda_{\mathrm{p}}^j).$$

Proof. (a) For any $z \neq 0$ the function $H_\nu^{(1)}(z)$ is an entire function of ν, see e.g. [Olv74, Chapter 7, Section 4.1]. For the derivative with respect to z the recurrence relation $2(H_\nu^{(1)})'(z) = H_{\nu-1}^{(1)}(z) - H_{\nu+1}^{(1)}(z)$ holds. Hence, it is also an entire function of ν. Since $\nu = a\sqrt{\lambda}$ in the definition (2.28) of $dtn^{\text{hom,2d}}$, we do not obtain meromorphicity directly on all of $\mathbb{C}$. If the negative real axis $\mathbb{R}_- := \{-x : x \geq 0\}$ is chosen as branch cut of the square root, then it first follows that $dtn^{\text{hom,2d}}$ is meromorphic on $\mathbb{C}$ except for $\mathbb{R}_-$. The following argument shows that meromorphicity even extends to $\mathbb{R}_-$.

Let $\zeta(\nu) := (H_\nu^{(1)})'(z)/H_\nu^{(1)}(z)$ for $z = ka$. According to [AS64, equation 9.1.6] the identity $H_{-\nu}^{(1)}(z) = \exp(\nu\pi i) H_\nu^{(1)}(z)$ holds. The recurrence relation implies that the same identity is valid for the derivative, i.e. $(H_{-\nu}^{(1)})'(z) = \exp(\nu\pi i)(H_\nu^{(1)})'(z)$. Hence, $\zeta(-\nu) = \zeta(\nu)$, i.e. ζ is an even function. Outside the roots of the denominator $\nu \mapsto H_\nu^{(1)}(z)$ (which do not cross $\mathbb{R}_-$), the function ζ is analytic and all odd coefficients in its local series expansion must vanish because ζ is even. Therefore, evaluating $\zeta(\nu)$ at $\nu = a\sqrt{\lambda}$ is unproblematic and $dtn^{\text{hom,2d}}(\lambda) = -k\zeta(a\sqrt{\lambda})$ is in fact meromorphic on all of $\mathbb{C}$.

21

(b) By part (a), the poles of $\mathit{dtn}^{\mathrm{hom,2d}}$ coincide with the roots of $\lambda \mapsto H^{(1)}_{a\sqrt{\lambda}}(z)$. According to [MK60, Thm. 2.3.], these roots are all simple, which yields the claim.

(c) Let

$$g(\nu) := \frac{\pi}{2} e^{(\nu+1)\pi i/2} H^{(1)}_\nu(z) = \int\limits_0^\infty e^{i\omega \cosh(t)} \cosh(\nu t) \, \mathrm{d}t. \qquad (2.33)$$

Magnus and Kotin [MK60] have shown that $f(\lambda) = g(\sqrt{\lambda})$ is an entire function of order $\leq 1/2$ by estimating the integral in (2.33). Hadamard's factorization theorem for entire functions then yields

$$f(\lambda) = f(0) \prod_{n=1}^\infty \left(1 - \frac{\lambda}{a^2 \lambda_\mathrm{p}^n} \right),$$

where $a^2 \lambda_\mathrm{p}^n$ are the zeros of $f(\lambda)$, i.e. the zeros of $\lambda \mapsto H^{(1)}_{\sqrt{\lambda}}(z)$. This means that

$$H^{(1)}_{\sqrt{\lambda}a}(z) = e^{-\frac{1}{2}\sqrt{\lambda}a\pi i} H^{(1)}_0(z) \prod_{n=1}^\infty \left(1 - \frac{\lambda}{\lambda_\mathrm{p}^n} \right) \qquad (2.34)$$

holds true. In reference [Coc65] it has been shown that a similar formula is valid for the derivative $(H^{(1)}_{\sqrt{\lambda}})'(z)$, which follows again essentially by means of the recurrence relation. Hence,

$$(H^{(1)}_{\sqrt{\lambda}a})'(z) = e^{-\frac{1}{2}\sqrt{\lambda}a\pi i} (H^{(1)}_0)'(z) \prod_{n=1}^\infty \left(1 - \frac{\lambda}{\lambda_\mathrm{r}^n} \right). \qquad (2.35)$$

The representation formula (2.31) follows by combining equations (2.34) and (2.35) and performing partial fractions decomposition. We have:

$$\mathit{dtn}^{\mathrm{hom,2d}}(\lambda) = \frac{-k}{H^{(1)}_{a\sqrt{\lambda}}(ka)} (H^{(1)}_{a\sqrt{\lambda}})'(ka) = -k \frac{(H^{(1)}_0)'(z)}{H^{(1)}_0(z)} \lim_{N\to\infty} \prod_{n=1}^N \frac{(1 - \lambda/\lambda_\mathrm{r}^n)}{(1 - \lambda/\lambda_\mathrm{p}^n)}.$$

A calculation gives

$$\prod_{n=1}^N \frac{(1 - \lambda/\lambda_\mathrm{r}^n)}{(1 - \lambda/\lambda_\mathrm{p}^n)} = \prod_{n=1}^N \frac{(\lambda - \lambda_\mathrm{r}^n)\lambda_\mathrm{p}^n}{(\lambda - \lambda_\mathrm{p}^n)\lambda_\mathrm{r}^n} = (\lambda - \lambda_\mathrm{r}^N) \frac{P_N(\lambda)}{R_N(\lambda)},$$

with $R_N(\lambda) := \prod_{j=1}^N (\lambda - \lambda_\mathrm{p}^j)$. Evaluating

$$R'_N(\lambda) = \sum_{n=1}^N \prod_{j=1, j\neq n}^N (\lambda - \lambda_\mathrm{p}^j).$$

at $\lambda = \lambda_\mathrm{p}^n$ gives

$$R'_N(\lambda_\mathrm{p}^n) = \prod_{j=1, j\neq n}^N (\lambda_\mathrm{p}^n - \lambda_\mathrm{p}^j) := Q_N(\lambda_\mathrm{p}^n)$$

22

It follows that

$$(\lambda - \lambda_{\mathrm{r}}^N)\frac{P_N(\lambda)}{R_N(\lambda)} = (\lambda - \lambda_{\mathrm{r}}^N)\sum_{n=1}^{N}\frac{P_N(\lambda_{\mathrm{p}}^n)}{R_N'(\lambda_{\mathrm{p}}^n)}\frac{1}{\lambda - \lambda_{\mathrm{p}}^n} = \sum_{n=1}^{N}\frac{P_N(\lambda_{\mathrm{p}}^n)}{Q_N(\lambda_{\mathrm{p}}^n)}\left[1 + \frac{\lambda_{\mathrm{p}}^n - \lambda_{\mathrm{r}}^N}{\lambda - \lambda_{\mathrm{p}}^n}\right]$$

$$= \sum_{n=1}^{N}\frac{P_N(\lambda_{\mathrm{p}}^n)}{Q_N(\lambda_{\mathrm{p}}^n)} + \sum_{n=1}^{N}\frac{P_N(\lambda_{\mathrm{p}}^n)(\lambda_{\mathrm{p}}^n - \lambda_{\mathrm{r}}^N)}{Q_N(\lambda_{\mathrm{p}}^n)}\frac{1}{\lambda - \lambda_{\mathrm{p}}^n},$$

which shows the claim.

$\square$
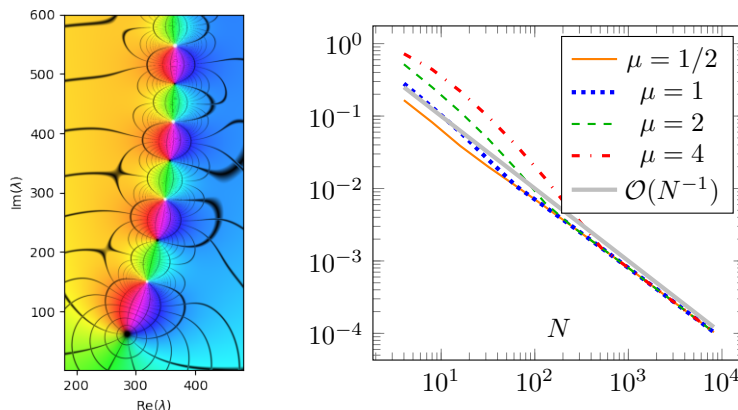
Figure 2.3: The plot on the right displays the relative error (2.37) for the approximation of $dtn^{\mathrm{hom,2d}}$ by its expansion into a sum of N simple poles on the interval $[0, \mu a k^2]$ for $\mu \in [1/2, 1, 2, 4]$. On the left the meromorphic function $\lambda \mapsto dtn^{\mathrm{hom,2d}}(\lambda)$ is visualized using domain coloring. Poles and roots are displayed in white and black respectively. The phase is visualized by means of a color hue. Grid lines of polar coordinates are shown in black.

It is interesting to investigate how fast the expansion in (2.31)-(2.32) converges. To get a rough idea, asymptotic expansions for the poles

$$a\sqrt{\lambda_{\mathrm{p}}^n} = \frac{i\pi n}{i(\pi/2 - \arg(ka)) + \ln\left(\frac{3\pi n}{e|ka|}\right)}\left[1 + \frac{\ln\ln n}{\ln n}\mathcal{O}(1)\right], \qquad (2.36)$$

obtained by Cochran [Coc65], which are valid for large n, are useful. For $\lambda \geq 0$

$$\sum_{n=N+1}^{\infty}\left|\frac{1}{\lambda - \lambda_{\mathrm{p}}^n}\right| \lesssim \sum_{n=N+1}^{\infty}\frac{1}{|\operatorname{Im}\lambda_{\mathrm{p}}^n|} \sim \sum_{n=N+1}^{\infty}\frac{g(n)}{n^2} \lesssim \int_{N}^{\infty}\frac{g(x)}{x^2} \lesssim \frac{\tilde{g}(N)}{N},$$

where the functions g and $\tilde{g}$ grow polylogarithmically. Hence, asymptotically one might expect at best linear convergence in the supremum norm.

To check this, we computed the first ten thousand poles λ_{p}^n and roots λ_{r}^n numerically and evaluated the relative supremum error

$$\sup_{\lambda \in [0, \mu a k^2]}\frac{|dtn^{\mathrm{hom,2d}}(\lambda) - S_N(\lambda)|}{|dtn^{\mathrm{hom,2d}}(\lambda)|} \qquad (2.37)$$

23

for $k = 16, a = 1$ and $\mu \in [1/2, 1, 2, 4]$. The results displayed in Fig. 2.3 clearly show a linear convergence in N. As the length of the interval grows, i.e. when μ increases, it takes slightly longer before the asymptotic regime of convergence of $\mathcal{O}(N^{-1})$ becomes apparent.

The poles λ_{p}^n are also of independent interest as the following remark shows.

Remark 2.5 (Poles of $\textit{dtn}^{\mathrm{hom,2d}}$ as poles of scattering amplitudes). *The poles $a\sqrt{\lambda_{\mathrm{p}}^n}$ have received considerable attention in the literature, see e.g. [MK60, Coc65, KRG63]. As for example discussed in the introduction of [KRG63], this is partly motivated by their occurence as poles of certain scattering amplitudes in scattering phenomena. One approach to see this connection proceeds by applying a certain integral transform to the eigenfunction expansion of the solution, see e.g. (2.11), and evaluate this by means of the residue theorem. For the intricate computational details we refer to the paper [Nus65] by Nussenzveig on high-frequency scattering by a totally reflecting sphere. More modern treatments of the topic in the context of microlocal analysis can be found in references [Tay79, MT85]. The integral transformation which is employed here is named after Watson, who carried out pioneering research in this direction [Wat33]. He also seems to be the first who discovered [Wat18] for a problem from electromagnetism that a certain scattering amplitude had poles at roots of $\lambda \mapsto H_\lambda^{(1)}(ka)$. Later, it has been found for many other problems that poles of scattering amplitudes are connected to roots of certain transcendental functions. Regge [Reg59] investigated this in the context of quantum mechanics. Explicit derivations may be found in Newton's monograph on scattering theory [New82, Chapter 13].*

It is an interesting question to investigate whether there exists a relation between poles of $\textit{dtn}$ and resonances, which are those complex wavenumbers for which the exterior problem with homogeneous boundary conditions admits non-trivial outgoing solutions. In the context of the Helmholtz equation in the exterior of a disk the resonances are given by the roots of $k \mapsto H_\ell^{(1)}(ka)$ for $\ell = 0, 1, \ldots$, see Section 5.3. It is not immediately clear how to relate roots of the Hankel functions with respect to their order (poles of $\textit{dtn}^{\mathrm{hom,2d}}$ and scattering amplitudes) to roots with respect to their argument (resonances). This matter if left open for future investigations.

Homogeneous medium $d = 3$

In three dimensions the $\textit{dtn}$ function for a homogeneous medium is given by

$$\textit{dtn}^{\mathrm{hom,3d}}(\lambda) = -k \frac{\left(h_{\frac{1}{2}\left(-1+\sqrt{1+4a^2\lambda}\right)}^{(1)}\right)'(ka)}{\left(h_{\frac{1}{2}\left(-1+\sqrt{1+4a^2\lambda}\right)}^{(1)}\right)(ka)}, \quad \textit{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) = -k \frac{(h_\ell^{(1)})'(ka)}{h_\ell^{(1)}(ka)}, \quad (2.38)$$

where $h_\ell^{(1)}$ denotes the spherical Hankel function of the first kind of order ℓ. It behaves very similar to the two-dimensional case. A plot for $k = 16$ and $a = 1$ is shown in Fig. 3.6. The following result from the literature shows that the assumptions ($\textit{dtn}$-I$_<$), ($\textit{dtn}$-II) and ($\textit{dtn}$-III) are fulfilled.

Lemma 2.6. *Let $k > 0$ and $a > 0$. Then*

(a) $\mathrm{Im}\,\textit{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) < 0$ for all $\ell \in \mathbb{N}_0$ and $\mathrm{Im}\,\textit{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) \to -0$ as $\ell \to +\infty$.

24

(b) The real part of $\boldsymbol{dtn}^{\mathrm{hom,3d}}(\lambda_\ell)$ is positive and bounded from above:

$$\mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) \geq \frac{1}{a} > 0.$$

(c) For $\ell \in \mathbb{N}_0$ it holds that

$$\sqrt{\frac{1}{a^2} + 4\lambda_\ell} + \frac{1}{2a} - k \leq \mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) \leq \sqrt{\frac{1}{a^2} + 4\lambda_\ell} + \frac{1}{2a} + k.$$

Proof. See [DI01, Lemma 3.2]. $\qquad\qquad\square$

The spherical Hankel functions can be regarded as Hankel functions of fractional order in view of the relation $h_\ell^{(1)}(x) = \sqrt{\pi/(2x)}H_{\ell+1/2}^{(1)}(x)$, see [AS64, equation (10.1.1)]. Therefore, it follows from the results of the previous subsection that $\boldsymbol{dtn}^{\mathrm{hom,3d}}$ may also be extended to a meromorphic function. Naturally, its pole structure, which can be seen in Fig. 3.6, is similar to that of $\boldsymbol{dtn}^{\mathrm{hom,2d}}$.

Remark 2.7. *The proof in reference [DI01] relies on the property*

$$|h_\ell^{(1)}(x)|^2 = \frac{\pi}{2x}|H_{\ell+1/2}^{(1)}(x)|^2 = \frac{1}{x^2}\sum_{j=0}^{\ell}\frac{(2\ell - j)!(2\ell - 2j)!}{j!\,[(\ell - j)!]^2}(2x)^{2j-2\ell}, \qquad (2.39)$$

see [AS64, equation (10.1.27)]. To the best of our knowledge, such a series expansion of $|H_\nu^{(1)}(x)|^2$ is only possible when ν is an half-integer. This means that it is not possible to infer results from Lemma 2.6 for the two-dimensional $\boldsymbol{dtn}$ function

$$\boldsymbol{dtn}^{\mathrm{hom,2d}}(\lambda_\ell) = -k\frac{(H_\ell^{(1)})'(ka)}{H_\ell^{(1)}(ka)}$$

by using the relation $h_\ell^{(1)}(x) = \sqrt{\pi/(2x)}H_{\ell+1/2}^{(1)}(x)$, because this would require results about the right hand side of (2.38) at non-integer ℓ's. On the other hand, the advantage of property (2.39) is that it leads to stronger results in the case of $d = 3$. Indeed, $\mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) \geq \frac{1}{a} > 0$ was shown in [DI01], whereas Lemma 2.3 established in this thesis only proves that $\mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,2d}}(\lambda)$ is bounded from below for $\sqrt{\lambda} \in (0, \xi]$ and positive for $\sqrt{\lambda} \in \{0\} \cup (\xi, \infty)$ for $\xi = \max\{k, 1/a\}$. However, numerical experiments (see e.g. Fig. 2.2 lead to the conjecture that $\mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,2d}}(\lambda)$ is a non-decreasing function on the positive real line so that $\mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,2d}}(\lambda) \geq \mathrm{Re}\,\boldsymbol{dtn}^{\mathrm{hom,2d}}(\lambda = 0) = C(a, k) > 0$ should hold true.

Jumping coefficient $d = 2$

Let us now drop the assumption of a uniformly homogeneous exterior domain. Consider a discontinuous wavenumber

$$k(r) = \begin{cases} k_I & r < R_{\mathrm{J}}, \\ k_\infty & r > R_{\mathrm{J}}, \end{cases} \qquad (2.40)$$

for some $k_I, k_\infty > 0$. The $\boldsymbol{dtn}$ function corresponding to this example will be referred to as $\boldsymbol{dtn}^{\mathrm{jump}}$. It can be described by an analytic formula derived in Appendix E. The derivation
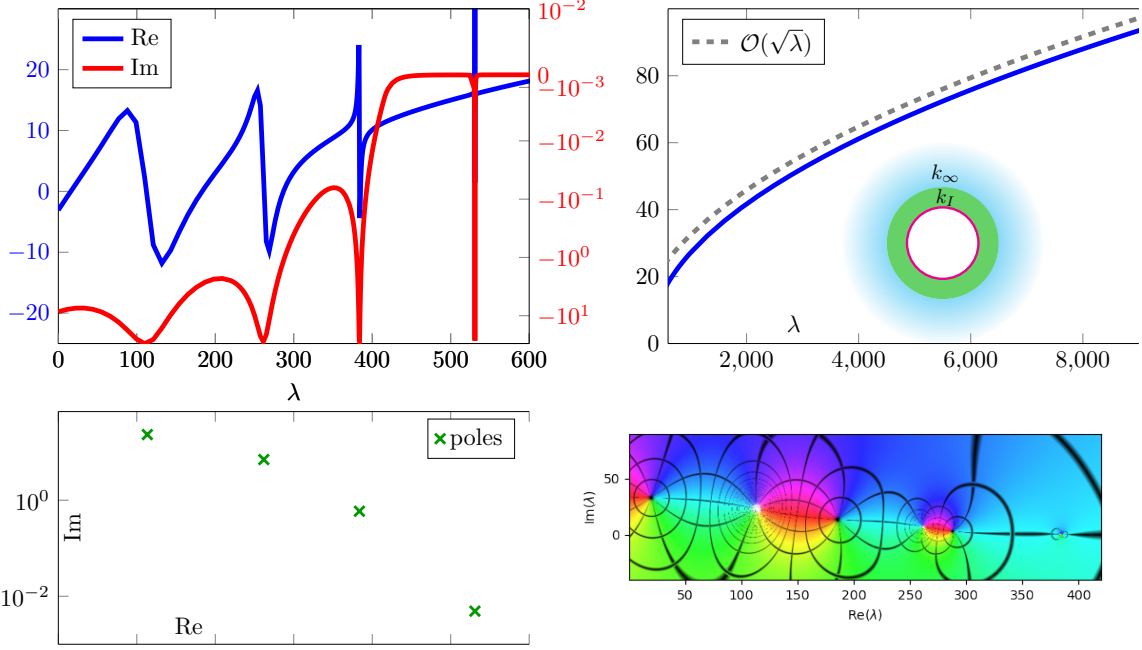
Figure 2.4: The dtn^{jump} function and its pole structure for $k_I = 16, k_\infty = 8, a = 1$ and $R_{\text{J}} = 2$.

determines the exact solution of (2.12a)-(2.12b) by imposing smoothness constraints at $r = R_{\text{J}}$ and the boundary condition at $r = a$. In the end, one arrives at

$$dtn^{\text{jump}}(\lambda) = \zeta(a\sqrt{\lambda}), \quad \zeta(\nu) = -k_I \left[A_\nu J_\nu'(k_I a) + B_\nu Y_\nu'(k_I a) \right], \qquad (2.41)$$

with $A_\nu = 1 - B_\nu Y_\nu(k_I a)/J_\nu(k_I a)$ and

$$B_\nu = \frac{1}{\det(M_\nu^{\text{J}})} \left[\frac{k_\infty}{k_I} (H_\nu^{(1)})'(k_\infty R_{\text{J}}) \frac{J_\nu(k_I R_{\text{J}})}{J_\nu(k_I a)} - H_\nu^{(1)}(k_\infty R_{\text{J}}) \frac{J_\nu'(k_I R_{\text{J}})}{J_\nu(k_I a)} \right],$$

for

$$\det(M_\nu^{\text{J}}) = -\frac{k_\infty}{k_I} \frac{(H_\nu^{(1)})'(k_\infty R_{\text{J}})}{J_\nu(k_I a)} \left[J_\nu(k_I a) Y_\nu(k_I R_{\text{J}}) - Y_\nu(k_I a) J_\nu(k_I R_{\text{J}}) \right]$$
$$+ \frac{H_\nu^{(1)}(k_\infty R_{\text{J}})}{J_\nu(k_I a)} \left[Y_\nu'(k_I R_{\text{J}}) J_\nu(k_I a) - Y_\nu(k_I a) J_\nu'(k_I R_{\text{J}}) \right].$$

A plot of dtn^{jump} for $a = 1, R = 2, k_I = 16$ and $k_\infty = 8$ is given in Fig. 2.4. The dtn^{jump} function exhibits a series of spikes for small λ_ℓ which are associated with poles of its meromorphic extension that are located close to the real axis. In the vicinity of the poles the real part $\text{Re}\, dtn^{\text{jump}}(\lambda_\ell)$ can become negative. These exceptional modes ℓ need to be absorbed into the set L_- defined in assumption (*dtn*-II). The plot shows that only a finite number of these outliers seem to occur and that the real part at the eigenvalues is eventually positive. Hence assumption (*dtn*-II) is fulfilled. The other assumptions (*dtn*-I$_<$) and (*dtn*-III) hold as well according to the plot.

It is also possible to prove the properties (*dtn*-I$_\leq$) and (*dtn*-III) analytically as shown in the following lemmas.

26

Lemma 2.8. *The function $\mathit{dtn}^{\mathrm{jump}}$ fulfills (dtn-I$_\leq$). If the constant $C^{\mathrm{J}}_{a\sqrt{\lambda_\ell}}$ defined in (E.6) is nonzero for all $\ell \in \mathbb{N}_0$, then even the sharp bound (dtn-I$_<$) holds.*

Proof. Integration by parts for some $R > a$ yields

$$\int\limits_{[a,R]} \left([\mathcal{A} + \lambda_\ell \mathcal{B}]\, u\right) v \left(\frac{r}{a}\right)^{d-1} \mathrm{d}r = u'(a)v(a) - \left(\frac{R}{a}\right)^{d-1} u'(R)v(R)$$

$$+ \int\limits_{[a,R_{\mathrm{V}}]} \left(u'v' + \left(-k^2(r) + \frac{\lambda_\ell a^2}{r^2}\right) uv\right) \left(\frac{r}{a}\right)^{d-1} \mathrm{d}r.$$

Now taking $u = \Lambda_r(\lambda_\ell), v = \bar{\Lambda}_r(\lambda_\ell)$ as solutions of (2.12a)-(2.12b) and using that $v(a) = \bar{\Lambda}_a(\lambda_\ell) = 1$ and $u'(a) = \Lambda'_a(\lambda_\ell) = -\mathit{dtn}(\lambda_\ell) =$ by definition yields the representation

$$\mathit{dtn}^{\mathrm{jump}}(\lambda_\ell) = \int\limits_{[a,R]} \left(|\Lambda'_r(\lambda_\ell)|^2 + \left(\frac{\lambda_\ell a^2}{r^2} - k^2(r)\right)|\Lambda_r(\lambda_\ell)|^2\right) \left(\frac{r}{a}\right)^{d-1}\mathrm{d}r \tag{2.42}$$

$$- \left(\frac{R}{a}\right)^{d-1} \Lambda'_R(\lambda_\ell)\bar{\Lambda}_R(\lambda_\ell).$$

Taking the imaginary part gives

$$\mathrm{Im}\,\mathit{dtn}^{\mathrm{jump}}(\lambda_\ell) = -\left(\frac{R}{a}\right)^{d-1} \mathrm{Im}(\Lambda'_R(\lambda_\ell)\bar{\Lambda}_R(\lambda_\ell))$$

for any $R > a$. By Sommerfeld's radiation condition

$$R^{d-1}\left|\Lambda'_R - ik_\infty \Lambda_R\right|^2 = R^{d-1}\left(|\Lambda'_R|^2 - 2k_\infty \mathrm{Im}(\Lambda'_R\bar{\Lambda}_R) + k_\infty^2 |\Lambda_R|^2\right)$$

goes to zero as $R \to \infty$. Hence,

$$\mathrm{Im}\,\mathit{dtn}^{\mathrm{jump}}(\lambda_\ell) = -\frac{1}{2k_\infty a^{d-1}} \lim_{R\to\infty} R^{d-1}\underbrace{(|\Lambda'_R|^2 + k_\infty^2 |\Lambda_R|^2)}_{\geq 0} \leq 0, \tag{2.43}$$

which already implies (dtn-I$_\leq$). To prove the sharp bound (dtn-I$_<$), the limit on the right hand side of (2.43) has to be calculated explicitly. From Appendix E it is known that for $R \geq R_{\mathrm{J}}$

$$\Lambda_R(\lambda_\ell) = C^{\mathrm{J}}_{a\sqrt{\lambda_\ell}} H^{(1)}_{a\sqrt{\lambda_\ell}}(k_\infty R),$$

with $C^{\mathrm{J}}_{a\sqrt{\lambda_\ell}}$ defined in (E.6). From [AS64, equation (9.2.3)] the asymptotic behavior

$$|H^{(1)}_\nu(z)| = \sqrt{\frac{2}{\pi z}} \quad \text{as } |z| \to \infty$$

is known. Thus, for $d = 2$

$$\mathrm{Im}\,\mathit{dtn}^{\mathrm{jump}}(\lambda_\ell) \leq -\frac{1}{\pi a^{d-1}}|C^{\mathrm{J}}_{a\sqrt{\lambda_\ell}}|^2,$$

which is strictly smaller zero provided $C^{\mathrm{J}}_{a\sqrt{\lambda_\ell}}$ does not vanish. $\qquad\square$

In view of the complexity of the formula for dtn^{jump}, it seems challenging to prove (dtn-III) based on equation (2.41) as was possible for $dtn^{\text{hom},2d}$. A better strategy is to prove directly that $\mathcal{DtN} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is well-defined. This can be inferred from the already established well-definedness of the $\mathcal{DtN}$ map for the homogeneous medium by extending Ω_{int} up to $r = R_{\text{J}}$ beyond which the exterior medium becomes homogeneous. A formal derivation is given in the next lemma.

Lemma 2.9. *The function dtn^{jump} fulfills (dtn-III).*

Proof. Let $R_{\infty} \geq R_{\text{J}}$ such that $k(r) = k_{\infty}$ for $r \geq R_{\infty}$ holds and define the annulus $\tilde{\Omega}_{\text{int}} := \{x \in \mathbb{R}^d \mid a < \|x\| < R_{\infty}\}$ with outer boundary $\Gamma_{\infty} := \{x \in \mathbb{R}^d \mid \|x\| = R_{\infty}\}$. Since the medium for $r \geq R_{\infty}$ is homogeneous, the $\mathcal{DtN}_{\infty} : H^{1/2}(\Gamma_{\infty}) \to H^{-1/2}(\Gamma_{\infty})$ map is continuous[2] and its corresponding dtn_{∞} function fulfills the properties (dtn-I$_<$), (dtn-II) and (dtn-III). From Corollary 2.20 it will follow[3] that the problem: Given boundary data $u_{\Gamma} \in H^{1/2}(\Gamma)$ find $u \in \tilde{V} := \{v \in H^1(\tilde{\Omega}_{\text{int}}) \mid \text{tr}_{\Gamma} u = 0\}$ such that

$$\tilde{b}_{\text{int}}(u, v) = -\tilde{b}_{\text{int}}(Z u_{\Gamma}, v) \quad \text{for all } v \in H^1_0(\tilde{\Omega}_{\text{int}}), \tag{2.44}$$

with

$$\tilde{b}_{\text{int}}(u, v) := \int\limits_{\tilde{\Omega}_{\text{int}}} \nabla u \nabla \bar{v} - k(x)^2 u \bar{v} \, dx + \langle \mathcal{DtN}_{\infty} u, v \rangle_{\Gamma_{\infty}}$$

is well-posed. Here Z denotes a right inverse for the trace operator on Γ, see Appendix A. Elliptic regularity implies that $u_{\text{ext}} := u + Z u_{\Gamma} \in H^2_{\text{loc}}(\tilde{\Omega}_{\text{int}})$ and that the PDE

$$-\Delta u_{\text{ext}} = k^2 u_{\text{ext}}$$

is fulfilled almost everywhere in $\tilde{\Omega}_{\text{int}}$. That is, $\|\Delta u_{\text{ext}}\|_{L^2(\tilde{\Omega}_{\text{int}})} \leq \|k\|^2_{L^\infty(\tilde{\Omega}_{\text{int}})} \|u_{\text{ext}}\|_{L^2(\tilde{\Omega}_{\text{int}})}$ so that in fact $\Delta u_{\text{ext}} \in L^2(\tilde{\Omega}_{\text{int}})$. Therefore, Proposition A.4 from Appendix A allows to define a weak normal derivative $\nabla u_{\text{ext}} \cdot \mathbf{n}_{\Gamma} \in H^{-1/2}(\Gamma)$, which fulfills the estimate

$$\|\nabla u_{\text{ext}} \cdot \mathbf{n}_{\Gamma}\|_{H^{-1/2}(\Gamma)} \leq \|Z\| \left(1 + \|k\|^2_{L^\infty(\tilde{\Omega}_{\text{int}})}\right) \|u_{\text{ext}}\|_{H^1(\tilde{\Omega}_{\text{int}})}$$

and agrees with the strong normal derivative on sufficiently smooth functions. Since $\|u_{\text{ext}}\|_{H^1(\tilde{\Omega}_{\text{int}})} \leq C \|u_{\Gamma}\|_{H^{1/2}(\Gamma)}$ by well-posedness of (2.44), the map $\mathcal{DtN} u_{\Gamma} = -\nabla u_{\text{ext}} \cdot \mathbf{n}_{\Gamma}$ is continuous from $H^{1/2}(\Gamma)$ to $H^{-1/2}(\Gamma)$, which establishes the claim. $\square$

2.2.2 Elliptical truncation boundary

Elongated obstacles may be enveloped more efficiently using elliptical than spherical coupling boundaries, see Fig. 2.5 for an illustration. Let $a_{\Gamma} > 0$ be the semi-major axis of Γ and $b_{\Gamma} > 0$ its semi-minor axis. Using the level set function

$$\phi(x, y) = \left(\frac{x}{a_{\Gamma}}\right)^2 + \left(\frac{y}{b_{\Gamma}}\right)^2 - 1,$$

the geometry can be described as

$$\Omega_{\text{int}} = \{\phi < 0\}, \quad \Gamma = \{\phi = 0\}, \quad \Omega_{\text{ext}} = \{\phi > 0\}.$$

(a) Geometry.　　　　　　　　　　　　　　(b) $dtn^{\text{ellipse}}(\lambda(q))$
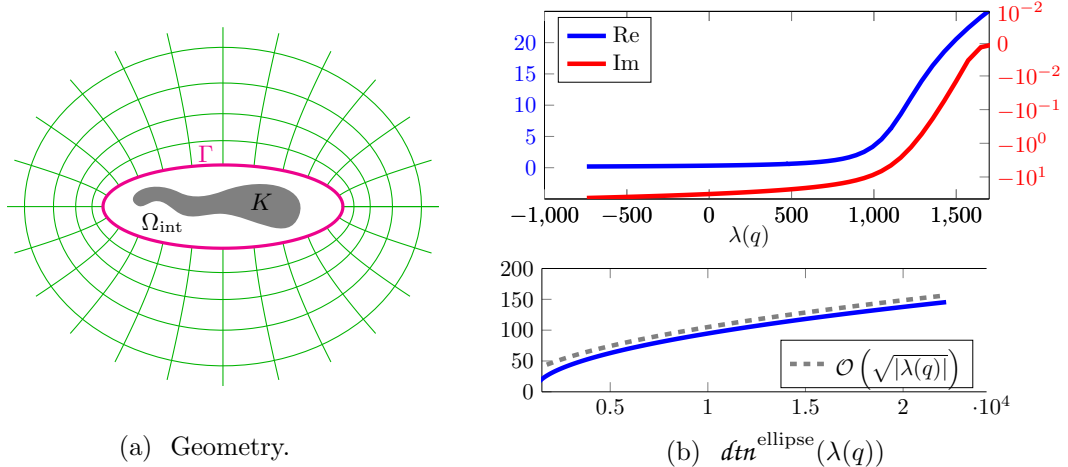
Figure 2.5: Scattering problem with elliptical coupling boundary. (a) Sketch of geometrical setup and coordinate lines of elliptical coordinate system. (b) The dtn^{ellipse} function for semi-major axis $a_\Gamma = 4/3$ and semi-minor axis $b_\Gamma = 1/2$.

Elliptical coordinates in Ω_{ext} are defined by

$$\begin{pmatrix} x \\ y \end{pmatrix} = c \begin{pmatrix} \cosh(r)\cos(\varphi) \\ \sinh(r)\sin(\varphi) \end{pmatrix}, \quad r \in [a, \infty), \ \varphi \in [0, 2\pi],$$

with $c := \sqrt{a_\Gamma^2 - b_\Gamma^2}$ and $a := \operatorname{arccosh}(a_\Gamma/c)$. The coordinates lines as sketched in Fig. 2.5 are ellipses in φ and hyperbolas in r. Assuming that $k(r) \equiv k$ in Ω_{ext}, the Helmholtz equation in these coordinates is given by

$$\left(-\frac{\partial^2}{\partial r^2} - 2q\cosh(2r) \right) u + \left(-\frac{\partial^2}{\partial \varphi^2} + 2q\cos(2\varphi) \right) u = 0, \tag{2.45}$$

with $q := c^2 k^2/4$, complemented by a Sommerfeld radiation condition at infinity. It is well-known that $\mathcal{DtN}$ is separable in these coordinates, see e.g. [BPG95]. The eigenvalues or separation constants $\lambda(q)$ are defined by

$$\left(-\frac{\partial^2}{\partial \varphi^2} + 2q\cos(2\varphi) \right) v_\lambda(q) = \lambda(q) v_\lambda(q). \tag{2.46}$$

If the eigenfunctions $v_\lambda(q)$ are required to be π or 2π-periodic, they can be developed into a Fourier series. Plugging such an ansatz into (2.46) leads to a three term recurrence relation for the Fourier coefficients (see [AS64, section 20.2]), which can be solved by different approaches, e.g. by transformation to a matrix eigenvalue problem or by using a continued fraction expansion. The eigenvalues $\lambda(q)$, known as characteristic numbers, are usually distinguished into two categories.

- Eigenvalues a_j are associated with the even periodic Mathieu functions of the first kind $ce_j(\varphi, q)$.

---

[2] As we have established the growth bound (dtn-III) for $dtn^{\text{hom,2d}}$ which implies that $\mathcal{DtN}_\infty$ maps $H^{1/2}(\Gamma_\infty)$ boundedly into $H^{-1/2}(\Gamma_\infty)$.

[3] The corollary applies since k is in $L^\infty(\tilde{\Omega}_{\text{int}})$. We should also remark that multiplication with $\mathcal{M}_\Gamma$ on Γ_∞ is omitted here since this map is just the identity for spherical coupling boundaries.

- Eigenvalues b_j are associated with the odd periodic Mathieu functions of the first kind $se_j(\varphi, q)$.

In contrast to the case of the Laplace-Beltrami operator on spherical coupling boundaries, (a finite number of) these eigenvalues can be negative. Let

$$\tilde{\lambda}_{2j} := a_j, \qquad v_{2j} := ce_j,$$
$$\tilde{\lambda}_{2j+1} := b_j, \qquad v_{2j+1} := se_j.$$

The corresponding solutions of the radial equation

$$\left(-\frac{\partial^2}{\partial r^2} - 2q \cosh(2r) + \tilde{\lambda}_j \right) \Lambda_j = 0$$

fulfilling the radiation condition at infinity are given by the radial Mathieu functions of the third kind

$$\Lambda_{2j}(r) := Mc_j(r), \quad \Lambda_{2j+1} := Ms_j(r).$$

For convenience, we let $\{\lambda_\ell\}$ be a permutation of the set $\{\tilde{\lambda}_j\}$ such that $\lambda_\ell \leq \lambda_{\ell+1}$ for all $\ell \in \mathbb{N}_0$. Then

$$dtn^{\text{ellipse}}(\lambda_\ell) = -\Lambda'_\ell(a)/\Lambda_\ell(a) \tag{2.47}$$

as usual.

To cast (2.45) into the form of equation (2.5) used in the general framework, define the diffeomorphism Ψ by

$$\Psi\left(r, \begin{pmatrix} \widehat{x} \\ \widehat{y} \end{pmatrix} \right) := \begin{pmatrix} \frac{\cosh(r)}{\cosh(a)} \widehat{x} \\ \frac{\sinh(r)}{\sinh(a)} \widehat{y} \end{pmatrix}, \quad \begin{pmatrix} \widehat{x} \\ \widehat{y} \end{pmatrix} = c \begin{pmatrix} \cosh(a) \sin(\varphi) \\ \sinh(a) \cos(\varphi) \end{pmatrix} \in \Gamma,$$

The Laplace-Beltrami operator and tangential gradient are given by

$$\nabla_\Gamma = \frac{1}{|\partial_r \Psi(a, \widehat{x})|_2} \partial_\varphi, \quad \Delta_\Gamma = \frac{1}{|\partial_r \Psi(a, \widehat{x})|_2} \partial_\varphi \left(\frac{1}{|\partial_r \Psi(a, \widehat{x})|_2} \partial_\varphi \right),$$

with $|\partial_r \Psi(a, \widehat{x})|_2 = c\sqrt{\sin(\varphi)^2 + \sinh(a)^2}$ describing the length of the normal vector on Γ. With the operators

$$\mathcal{M}_\Gamma := 1/|\partial_r \Psi(a, \widehat{x})|_2, \quad \mathcal{K}_\Gamma := -|\partial_r \Psi(a, \widehat{x})|_2 \Delta_\Gamma - \nabla_\Gamma \left(|\partial_r \Psi(a, \widehat{x})|_2 \right) \nabla_\Gamma + \frac{2q \cos(2\varphi)}{|\partial_r \Psi(a, \widehat{x})|_2}$$

the eigenvalue problem (2.46) can be written as

$$\mathcal{K}_\Gamma v_\lambda(q) = \lambda(q) \mathcal{M}_\Gamma v_\lambda(q).$$

With

$$\mathcal{A} = -\frac{\partial^2}{\partial r^2} - 2q \cosh(2r), \quad \mathcal{B} = 1,$$

equation (2.5) holds true in the transformed coordinates.

Remark 2.10 (Variational formulation). *Note that integration by parts leads to*

$$
\int_\Gamma -|\partial_r\Psi(a,\widehat{x})|_2 \,(\Delta_\Gamma u)\,\bar{v}\,\mathrm{d}\Gamma = \int_\Gamma \nabla_\Gamma u \nabla_\Gamma \left(|\partial_r\Psi(a,\widehat{x})|_2\bar{v}\right)\,\mathrm{d}\Gamma
$$

$$
= \int_\Gamma |\partial_r\Psi(a,\widehat{x})|_2 \nabla_\Gamma u \nabla_\Gamma \bar{v} + (\nabla_\Gamma u)\,\bar{v}\nabla_\Gamma\left(|\partial_r\Psi(a,\widehat{x})|_2\right)\,\mathrm{d}\Gamma.
$$

Hence, the non-symmetric terms cancel in the weak form of the operator $\mathcal{K}_\Gamma$ and we obtain that the bilinear forms $\mathcal{K}_\Gamma(\cdot,\cdot)$ and $\mathcal{M}_\Gamma(\cdot,\cdot)$ associated with the operators $\mathcal{K}_\Gamma$ and $\mathcal{M}_\Gamma$ fulfilling the eigenvalue equation are given by

$$
\mathcal{K}_\Gamma(u,v) = \int_\Gamma \left[|\partial_r\Psi(a,\widehat{x})|_2 \nabla_\Gamma u \nabla_\Gamma \bar{v} + \frac{c^2 k^2}{2}\frac{1}{|\partial_r\Psi(a,\widehat{x})|_2}\cos(2\varphi)u\bar{v}\right]\mathrm{d}\Gamma,
$$

$$
\mathcal{M}_\Gamma(u,v) = \int_\Gamma \frac{1}{|\partial_r\Psi(a,\widehat{x})|_2}u\bar{v}\mathrm{d}\Gamma.
$$

Compared to the case of Bessel functions, relatively few identities are known for Mathieu functions. This may partly stem from the complexity of determining the characteristic values $\lambda_\ell(q)$. As a result, we have not been able to establish a counterpart of Lemma 2.3, i.e. a verification of (*dtn*-I$_<$), (*dtn*-II) and (*dtn*-III) based on the properties of Mathieu functions. Nevertheless, numerical studies suggest that these assumptions are fulfilled as can also be seen in Fig. 2.5. Moreover, property (*dtn*-III) could be established with a similar technique as utilized in Lemma 2.9 for *dtn*$^{\mathrm{jump}}$. To this end, the interior domain could be extended beyond Γ and eventually truncated by a large enough circle at which the $\mathcal{DtN}$ map for spherical geometries is placed. Continuity of $\mathcal{DtN} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ follows then along the lines of Lemma 2.9.

2.3 Waveguides

Next we would like to arrange a suitable geometry for a waveguide. To this end, consider a smooth and bounded domain $\tilde{\Gamma} \subset \mathbb{R}^{d-1}$. Let $\Gamma := \{a\} \times \tilde{\Gamma}$ for some $a \in \mathbb{R}$ be the artificial truncation boundary between the interior $\Omega_{\mathrm{int}} := (0,a) \times \tilde{\Gamma}$ and the exterior $\Omega_{\mathrm{ext}} := (a,\infty) \times \tilde{\Gamma}$ of the waveguide. Define the diffeomorphism

$$
\Psi(r,\widehat{x}) := \begin{pmatrix} r-a \\ 0 \end{pmatrix} + \widehat{x} \qquad \text{for } r \in [a,\infty),\ \widehat{x} \in \Gamma.
$$

In the exterior we consider the Helmholtz equation

$$
\begin{aligned}
(-\Delta - k^2)u &= 0 \quad &&\text{in } \Omega_{\mathrm{ext}}, \\
u &= 0 \quad &&\text{on } (a,\infty) \times \partial\tilde{\Gamma},
\end{aligned} \tag{2.48}
$$

for $k > 0$ together with a radiation condition that will be discussed below. The PDE in the interior Ω_{int} of the waveguide may be of the general form (2.15). Let Δ_Γ denote the Laplace-Beltrami operator on Γ. Boundary conditions are assumed so that Δ_Γ is self-adjoint. In the coodinates given by Ψ, the exterior equation (2.48) takes the form

$$
(-\Delta u - k^2 u)(\Psi(r,\widehat{x})) = \left(-\partial_r^2 - \Delta_\Gamma - k^2\right)(u \circ \Psi)(r,\widehat{x}).
$$

Hence, $\mathcal{A} = -\partial_r^2 - k^2 \operatorname{Id}$ and $\mathcal{B} = \operatorname{Id}$. It follows that (2.12a) has two independent solutions given by $\exp(\pm i\sqrt{k^2 - \lambda_\ell}(\cdot - a))$. As before, the branch cut of the square root function is assumed to be the negative imaginary axis. Moreover,

$$k^2 \notin \sigma(-\Delta_\Gamma) \tag{2.49}$$

will be assumed, where $\sigma(-\Delta_\Gamma)$ denotes the spectrum of the Laplace-Beltrami operator. Then a solution to (2.48) is said to satisfy the radiation condition, if it admits an expansion (2.11) with functions $\Lambda_r(\lambda_\ell)$ that decay for $\lambda_\ell > k^2$ as $r \to \infty$. This requirement characterizes $\Lambda_r(\lambda_\ell) := \exp(i\sqrt{k^2 - \lambda_\ell}(r - a))$ as the outgoing solution. Hence, the *dtn* function is given by

(a) Geometry.   (b) dtn^{guide}
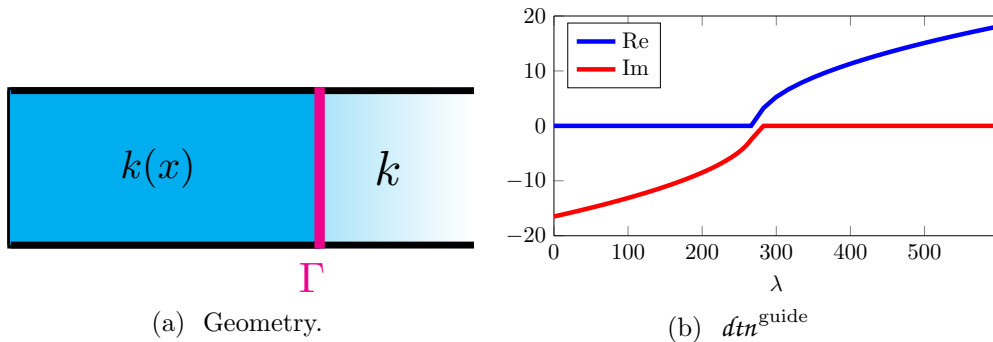
Figure 2.6: Geometry and dtn^{guide} function for a waveguide with $k = 16.5$.

$$dtn^{\mathrm{guide}}(\lambda) = -i\sqrt{k^2 - \lambda}, \tag{2.50}$$

which is shown in Fig. 2.6b for $k = 16.5$. Even though dtn^{guide} can be extended holomorphically to a neighborhood of $\sigma(-\Delta_\Gamma)$ provided that assumption (2.49) is fulfilled, the branch cut singularity prevents a natural meromorphic extension to the entire complex domain. Since $|dtn^{\mathrm{guide}}(\lambda)| \leq \max\{1, |k|^2\}(1 + |\lambda|)^{1/2}$ and $\operatorname{Re} dtn^{\mathrm{guide}}(\lambda) = \operatorname{Re}\sqrt{\lambda - k^2} > 0$ for $\lambda > k^2$, the assumptions (*dtn*-II) and (*dtn*-III) are fulfilled. However, in contrast to the previously investigated examples which fulfilled (*dtn*-I$_<$) because $\operatorname{Im} dtn(\lambda)$ only approached zero from below as $\lambda \to +\infty$, we have $\operatorname{Im} dtn^{\mathrm{guide}}(\lambda) = 0$ for $\lambda > k^2$. Hence, only the weaker assumption (*dtn*-I$_\leq$) holds for the waveguide.

2.4   Helioseismology

The Sun is filled with various types of waves which can be utilized for its study. Solar or more generally stellar oscillations can be modelled by a generalization of Galbrun's equation derived in [LBO67], which includes differential rotation and gravity. In this thesis a simpler model [GBD+17] is considered, which only accounts for pressure waves. This model has lead to significant scientific insights, see e.g. [GCP+20], and is therefore the natural starting point for our investigations. Eventually, the techniques presented in this thesis should be extended to the general equations of stellar oscillations.

2.4.1   Original formulation

The outer 28% of the solar interior form the solar convection zone. This region is convectively unstable and highly dynamic. Turbulent processes in the convection zone excite

pressure waves which propagate through the entire solar interior and can be observed at the visible solar solar surface by satellites or ground-based telescopes. The following time-harmonic convected wave equation describes these waves to a reasonable extent:

$$-\frac{\sigma^2}{\rho c^2}\tilde{u} - \frac{2i\omega}{\rho c}(\rho\mathbf{w})\cdot\nabla\left(\frac{\tilde{u}}{\rho c}\right) - \nabla\cdot\left(\frac{1}{\rho}\nabla\tilde{u}\right) = \frac{f}{c}. \tag{2.51}$$

The scalar unknown $\tilde{u}$ is associated with the divergence of the wave displacement. The coefficients ρ and c describe solar density and sound speed, respectively, and are provided by a given background model for the Sun. For the solar interior the Model S of [CD$^+$96] will be used throughout this thesis. A positive damping $\gamma(r,\omega) > 0$ is included in $\sigma^2 = \omega^2 + 2i\omega\gamma$, see [GBD$^+$17, Section 7.3] or Section 6.1 and Section 6.4 for details. Here, $\omega > 0$ denotes the frequency. The quantity $\mathbf{w}$ represents a background flow which can e.g. be used to model differential rotation or meridional circulation.

(a) Solar interior.　　　　(b) Atmosphere.
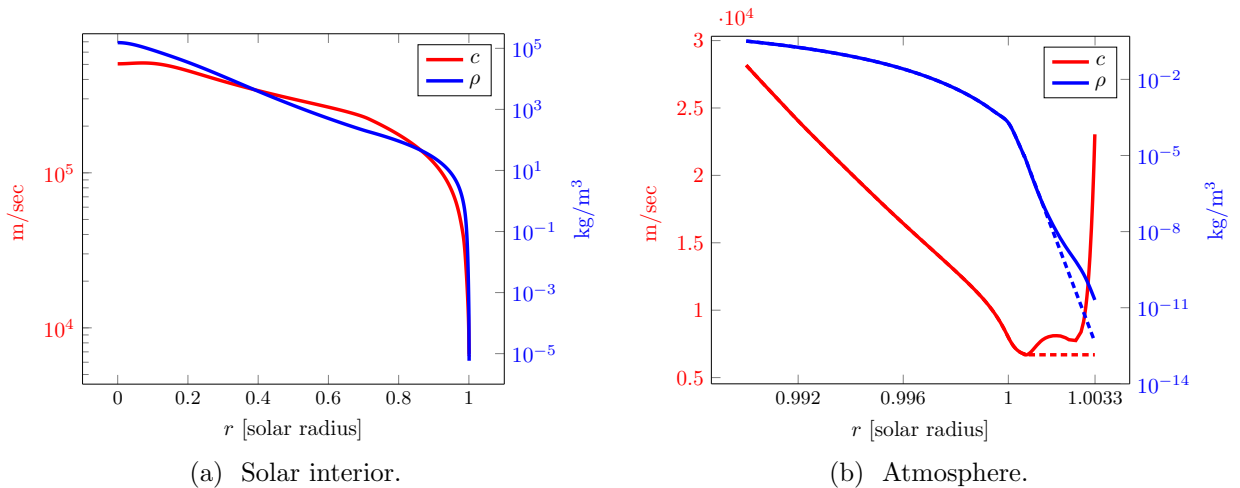
Figure 2.7: Sound speed c and density ρ for solar models. The dashed lines in the right figure display the `Atmo` extension, while solid lines represent the VAL-C model [VAL81].

Sound speed and density of the solar background model are functions of the radial coordinate $r = |x|$ which vary drastically in upper layers of the solar interior. Fig. 2.7a displays these coefficients for Model S. From the visible surface (photosphere) towards the core the density varies by nine orders of magnitude, while the sound speed increases by a factor of about 50. The latter reflects the fact that the core is a few thousand times hotter than the photosphere as temperature is proportional to the square of the sound speed. As a result, inward propagating waves are refracted by the rise in sound speed towards the core while outward propagating, low-frequency waves cannot pass the steep density gradient in the surface layers and are reflected back into the interior. High frequency waves, however, can escape into the atmosphere and may be partially reflected back from the higher atmospheric layers into the interior. This offers the potential of probing part of the solar atmosphere with high frequency waves and neccessitates the need for complementing the interior model of the Sun by an appropriate atmospheric extension.

Equation (2.51) is posed in a domain Ω which includes the solar interior and a part of the solar atmosphere, which will be specified more precisely below. Standard models of the solar atmosphere [VAL81, AL08] describe sound speed and density in terms of radially

symmetric functions. Moreover, the flow $\mathbf{w}$ and the source f are assumed to vanish in the solar atmosphere, which ensures separability of (2.51) above the photosphere. This already suggests to position the truncation boundary at the solar surface or slightly above, yet there is an even more compelling argument in favor of this approach. Compared to the deep solar interior, the wavenumber in the atmosphere is extremely small, which would lead to a significant increase of the computational costs, if the atmosphere was meshed. Therefore, the interior domain Ω_{int} is a ball with radius a with a being equal to or slightly larger than the solar radius.

The equation in Ω_{int} is of the form (2.15) with

$$(a_{ij}) = \frac{1}{\rho}\,\mathrm{Id}, \qquad\qquad c = -\frac{\sigma^2}{\rho \mathrm{c}^2},$$

$$\mathbf{b} = -2\omega(\rho\mathbf{w}), \qquad\qquad \beta = \frac{1}{\rho \mathrm{c}}.$$

Ellipticity (C-I) holds since the density is positive and bounded from below in the solar interior. Due to $\gamma > 0$ one has $\operatorname{Im} c < 0$ so that even a strict inequality in (C-III) is fulfilled. Assumption (C-II) corresponds to conservation of mass and to the constraint that the flow is confined to the solar interior. Both are standard requirements in helioseismology, see [GBD⁺17, Section 2.4].

In the exterior domain the equation is separable with Ψ defined by (2.26) and the operators in the separation condition (2.5) given as

$$\mathcal{A} = -r^{1-d}\partial_r(r^{d-1}\frac{1}{\rho}\partial_r) - \frac{\sigma^2}{\rho \mathrm{c}^2}\,\mathrm{Id}, \qquad (\mathcal{B}v)(r) = \frac{a^2}{r^2}v(r).$$

The operators on Γ are simply the identity and the Laplace-Beltrami operator, i.e. $\mathcal{M}_\Gamma = \mathrm{Id}_\Gamma$ and $\mathcal{K}_\Gamma = -\Delta_\Gamma$.

2.4.2 Transformation to Schrödinger equation

Recently, it has become popular [AHN18, GFY⁺18, BFP20] to work with an equivalent formulation of (2.51). By means of the substitution $u = \rho^{-1/2}\tilde{u}$ the equation transforms to a Schrödinger-like equation

$$\mathrm{q}u - \frac{2i\omega}{\rho^{1/2}\mathrm{c}}(\rho\mathbf{w})\cdot\nabla\left(\frac{u}{\rho^{1/2}\mathrm{c}}\right) - \Delta u = \rho^{1/2}\frac{f}{\mathrm{c}}, \qquad (2.52)$$

with an effective potential

$$\mathrm{q} = \rho^{1/2}\Delta\left(\rho^{-1/2}\right) - \frac{\sigma^2}{\mathrm{c}^2}. \qquad (2.53)$$

This also fits into the framework of Section 2.1 with

$$\begin{aligned}(a_{ij}) &= \mathrm{Id}, & c &= \mathrm{q}, \\ \mathbf{b} &= -2\omega(\rho\mathbf{w}), & \beta &= \frac{1}{\rho^{1/2}\mathrm{c}},\end{aligned} \qquad (2.54)$$

and the assumptions are fulfilled as for the original model. In the exterior

$$\mathcal{A} = -r^{1-d}\partial_r(r^{d-1}\partial_r) + q, \qquad \mathcal{B} = \frac{a^2}{r^2}$$

and the tangential differential operators on Γ are unchanged. The ODE (2.12a), which determines the *dtn* function of the Sun, has to be carefully dicussed as it is neither clear how to model sound speed and density in the atmosphere nor which radiation condition to impose to specify an outgoing solution. Two options will be discussed here.

Atmo extension

A first approach [SCGM11] is to extend the sound speed at the end a of the standard solar model (for Model S about 0.5 Mm above the photosphere) by a constant and let density decay exponentially:

$$\rho(r) = \rho(a)\exp(-\alpha(r-a)), \qquad \mathrm{c}(r) = \mathrm{c}(a), \;\; \text{for } r \geq a, \tag{2.55}$$

where $\alpha = -\rho'(a)/\rho(a)$ is the inverse density scale height. This extension does not represent a realistic solar chromosphere as the authors of [SCGM11] point out. In fact, sound speed increases sharply in the upper chromosphere as a result of a steep temperature gradient between the photosphere and the corona. The latter is more than a million degrees hot while the temperature of the former is only at about $5500°C$. The motivation for a constant sound speed extension rather stems from the demand to reduce artificial reflection and promote wave attenuation.

Model (2.55) later became known at the `Atmo` extension [BCD$^+$18]. It has been subject of a series of publications in which local transparent boundary conditions for this model have been developed, see e.g. [BCD$^+$18, FLH$^+$17]. In reference [BFP20] it has been shown that the exact *dtn* function corresponding to the outgoing solution (also for $\gamma = 0$) can in fact be described analytically by

$$dtn^{\mathrm{Atmo}}(\lambda_\ell) = \frac{1}{a} + 2ik\frac{W'_{-i\eta,\ell+\frac{1}{2}}(-2ika)}{W_{-i\eta,\ell+\frac{1}{2}}(-2ika)}, \tag{2.56}$$

with

$$k^2 := \frac{\sigma^2}{\mathrm{c}(a)^2} - \frac{\alpha^2}{4}, \qquad \eta := \frac{\alpha}{2k}.$$

Here, $W_{\kappa,\mu}(z)$ denotes the Whittaker function [Whi03], which also appears in quantum mechanical scattering problems with Coulomb potentials [New82]. The Whittaker function for $z \in \mathbb{C} \setminus \mathbb{R}_-$ is an entire function of μ, see[4] [Olv74, Chapter 7, Section 10.3 and Section 11.1], which allows for a meromorphic extension of dtn^{Atmo}. A plot of dtn^{Atmo} is shown in Fig. 6.9 from which it can be seen that assumptions (*dtn*-I$_<$), (*dtn*-II) and (*dtn*-III) are fulfilled. Since equation (2.56) gives an analytic formula for dtn^{Atmo}, it might even be possible to verify these properties analytically similar to Lemma 2.3, yet so far our attempts have been stalled by a shortage of useful identities or asymptotic estimates valid for the range of parameters μ, κ and z occuring in the definition of dtn^{Atmo}.

VAL-C atmospheric model

The main shortcoming of the `Atmo` model is its failure to account for the drastic rise in sound speed in the upper chromosphere featured in realistic models, see e.g.[VAL81, AL08]. A comparison between the `Atmo` model and the more realistic VAL-C model of [VAL81] is shown in Fig. 2.7b. To determine the $dtn^{\mathrm{VAL\text{-}C}}(\lambda)$ function for a solar atmosphere based on the VAL-C model, the ODE (2.12a) is solved numerically on a finite interval $[a, R_{\mathrm{V}}]$. A sophisticated radiation condition is not needed here as the

---

[4]In [Olv74, Chapter 7, Section 10.3] the confluent hypergeometric function $U(b,c,z)$ is characterized as entire in b and c for $z \neq 0$ by appealing to [Olv74, Chapter 7, Theorem 3.1]. This implies our statement about $W_{\kappa,\mu}(z)$ since according to [Olv74, Chapter 7, Section 11.1] we have $W_{\kappa,\mu}(z) = e^{-z/2}z^{\mu+1/2}U(\mu - \kappa + 1/2, 2\mu + 1, z)$.

damping $\gamma = \operatorname{Im} \sigma$ is required to be positive. Simply imposing a homogeneous Neumann boundary condition at R_{V}, which marks the end of the VAL-C model, then suffices to guarantee uniqueness of the solution.
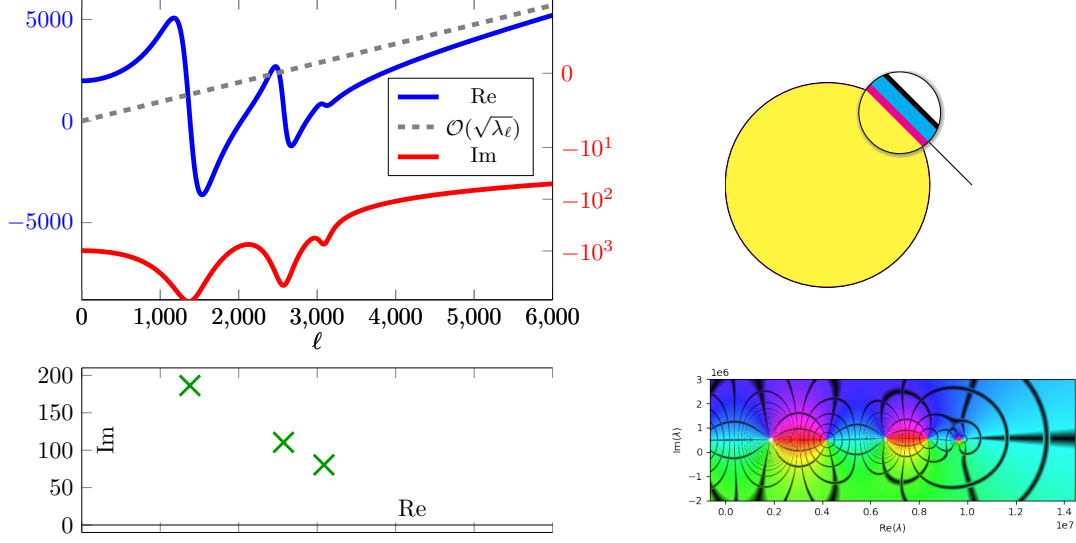

Figure 2.8: The $dtn^{\mathrm{VAL-C}}$ function and its pole structure at 7.0 mHz. The photosphere is taken as the coupling boundary.

A plot of $dtn^{\mathrm{VAL-C}}$ is show in Fig. 2.8. The distinguished feature of $dtn^{\mathrm{VAL-C}}(\lambda)$ compared to $dtn^{\mathrm{Atmo}}(\lambda)$ is the existence of poles extremely close to the real axis in the preasymptotic regime (small to moderate λ). As already observed for dtn^{jump}, this results in negative values of the real part at a few eigenvalues $\lambda = \lambda_\ell$ lying close to these poles. These outliers have to be absorbed into the set L_- introduced in assumption (dtn-II). For the analysis presented in Section 2.5 it is crucial that the set L_- is finite, which is justified according to Fig. 2.8. In particular, the domain coloring plot shows that the number of poles in the positive half-plane is finite. There seems to be an infinite tail of alternating poles and roots which starts in the positive half-plane and extends towards negative infinity. According to the plot, the other assumptions (dtn-$\mathrm{I}_<$) and (dtn-III) are also fulfilled. The following lemma indeed proves that properties (dtn-$\mathrm{I}_<$) and (dtn-II) always hold.

Lemma 2.11. *Let* $dtn^{\mathrm{VAL-C}}(\lambda)$ *be the* dtn *function of the VAL-C model based on a homogeneous Neumann boundary conditions at* R_{V}, *i.e.* $\Lambda'_{R_{\mathrm{V}}}(\lambda_\ell) = 0$ *is imposed as boundary condition in* (2.12a). *Then* $dtn^{\mathrm{VAL-C}}$ *fulfills* (dtn-$\mathrm{I}_<$) *and* (dtn-II).

Proof. Proceeding as in Lemma 2.8, an integration by parts on $[a, R_{\mathrm{V}}]$ yields

$$\int\limits_{[a,R_{\mathrm{V}}]} \left([\mathcal{A} + \lambda_\ell \mathcal{B}]\, u \right) v \left(\frac{r}{a} \right)^{d-1}\, \mathrm{d}r = u'(a)v(a) - \left(\frac{R_{\mathrm{V}}}{a} \right)^{d-1} u'(R_{\mathrm{V}})v(R_{\mathrm{V}})$$

$$+ \int\limits_{[a,R_{\mathrm{V}}]} \left(u'v' + \left(\mathrm{q} + \frac{\lambda_\ell a^2}{r^2} \right) uv \right) \left(\frac{r}{a} \right)^{d-1}\, \mathrm{d}r.$$

Now taking $u = \Lambda_r(\lambda_\ell)$ which solves (2.12a) and $v = \bar{\Lambda}_r(\lambda_\ell)$ and using the identities

$$\Lambda_a(\lambda_\ell) = 1, \quad \Lambda'_{R_{\mathrm{V}}}(\lambda_\ell) = 0, \quad dtn(\lambda_\ell) = -\Lambda'_a(\lambda_\ell)$$

36

which hold by definition, yields the representation

$$\text{\textit{dtn}}^{\text{VAL}-\text{C}}(\lambda_\ell) = \int\limits_{[a,R_{\text{V}}]} \left(|\Lambda_r'(\lambda_\ell)|^2 + \left(\mathsf{q} + \frac{\lambda_\ell a^2}{r^2} \right) |\Lambda_r(\lambda_\ell)|^2 \right) \left(\frac{r}{a} \right)^{d-1} \, \mathrm{d}r. \tag{2.57}$$

It follows that $\operatorname{Re} \text{\textit{dtn}}^{\text{VAL}-\text{C}}(\lambda_\ell) > 0$ for $\lambda_\ell > \| \operatorname{Re} \mathsf{q} \|_{[a,R_{\text{V}}],\infty}(R_{\text{V}}/a)^2$, which gives ($\text{\textit{dtn}}$-II)
with $L_- := \{\ell \in \mathbb{N}_0 \mid \lambda_\ell < \| \operatorname{Re} \mathsf{q} \|_{[a,R_{\text{V}}],\infty}(R_{\text{V}}/a)^2\}$. Moreover,

$$\operatorname{Im} \text{\textit{dtn}}^{\text{VAL}-\text{C}}(\lambda_\ell) = - \int\limits_{[a,R_{\text{V}}]} \frac{2\gamma\omega}{\mathsf{c}^2} |\Lambda_r|^2 \left(\frac{r}{a} \right)^{d-1} \, \mathrm{d}r < 0$$

because of the positive damping γ. Hence, ($\text{\textit{dtn}}$-I$_<$) is fulfilled. $\qquad\square$

Remark 2.12. *The representation formula* (2.57) *extends to general* $\lambda \in \mathbb{C}$ *provided that the ODE* (2.12a) *with the respective boundary conditions has a unique solution in* $H^1([a, R_{\text{V}}])$ *and then* $| \text{\textit{dtn}}^{\text{VAL}-\text{C}}(\lambda)| < \infty$ *holds. This implies that poles of* $\text{\textit{dtn}}^{\text{VAL}-\text{C}}$ *are associated with those complex numbers* λ *for which this ODE problem is not well-posed.*

An argument similar to Lemma 2.9 for $\text{\textit{dtn}}^{\text{jump}}$ shows that ($\text{\textit{dtn}}$-III) is satisfied.

Lemma 2.13. *The function* $\text{\textit{dtn}}^{\text{VAL}-\text{C}}$ *fulfills* ($\text{\textit{dtn}}$-III).

Proof. It suffices to show that $\mathcal{D}t\mathcal{N} : H^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is continuous. Define the annulus $\tilde{\Omega}_{\text{int}} := \{x \in \mathbb{R}^d \mid a < \|x\| < R_{\text{V}}\}$ and let $u_\Gamma \in H^{1/2}(\Gamma)$ be given. Let the sesquilinearform $\tilde{b}_{\text{int}}(\cdot, \cdot)$ be defined as in (2.20) with the coefficients from (2.54) except that the domain of integration is replaced by $\tilde{\Omega}_{\text{int}}$, and let $\tilde{V} := \{v \in H^1(\tilde{\Omega}_{\text{int}}) \mid \operatorname{tr}_\Gamma u = 0\}$. To define $\mathcal{D}t\mathcal{N}$, we have to solve the problem: Given $u_\Gamma \in H^{1/2}(\Gamma)$ find $u \in \tilde{V}$ such that

$$\tilde{b}_{\text{int}}(u, v) = -\tilde{b}_{\text{int}}(Zu_\Gamma, v) \quad \text{for all } v \in H_0^1(\tilde{\Omega}_{\text{int}}). \tag{2.58}$$

The Gårding inequality established in Lemma 2.2 remains valid for $\tilde{b}_{\text{int}}(\cdot, \cdot)$ on $\tilde{V} \times \tilde{V}$. Furthermore, solutions of (2.58) are unique if they exist which can be inferred by considering the imaginary part of $\tilde{b}_{\text{int}}(u, u) = 0$ and using that $\gamma > 0$. A standard application of the Fredholm alternative using compactness of the embedding $\tilde{V} \hookrightarrow L^2(\tilde{\Omega}_{\text{int}})$ then shows that (2.58) is well-posed (see the proof of Theorem 2.19 for a detailed argument in a more complicated case). By interior elliptic regularity $u_{\text{ext}} := u + Zu_\Gamma \in H^2_{\text{loc}}(\tilde{\Omega}_{\text{int}})$ and the strong form of the PDE, which is

$$\Delta u_{\text{ext}} = \mathsf{q} u_{\text{ext}}$$

because the flow and the source were assumed to vanish exterior to Ω_{int}, is fulfilled almost everywhere in $\tilde{\Omega}_{\text{int}}$. Since $\mathsf{q} \in L^\infty(\tilde{\Omega}_{\text{int}})$, this implies $\Delta u_{\text{ext}} \in L^2(\tilde{\Omega}_{\text{int}})$. This allows to define a weak normal derivative $\nabla u_{\text{ext}} \cdot \mathbf{n}_\Gamma \in H^{-1/2}(\Gamma)$, see Proposition A.4, and $\mathcal{D}t\mathcal{N}u_\Gamma = -\nabla u_{\text{ext}} \cdot \mathbf{n}_\Gamma$ is continuous from $H^{1/2}(\Gamma)$ to $H^{-1/2}(\Gamma)$ which can be inferred as in Lemma 2.9. $\qquad\square$

2.5 Well-posedness of the reduced interior problem

The aim of this section is to establish well-posedness of the reduced problem on Ω_{int} with $\mathcal{D}t\mathcal{N}$ as transparent boundary condition as introduced in Section 2.1. Let us briefly recall

the setup. For a continuous (anti)-linear functional $l : V \to \mathbb{C}$ consider the variational problem: Find $u \in V$ such that

$$b(u, v) = l(v) \quad \text{for all } v \in V, \tag{2.59}$$

where

$$b(u, v) := b_{\text{int}}(u, v) + \langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}u, v \rangle_\Gamma. \tag{2.60}$$

Let the assumptions introduced in Section 2.1 hold. Additionally, the following simplifying assumption on the diffeomorphism Ψ will be made from now on: There exists a constant $c_\Psi > 0$ such that

$$\frac{1}{|\partial_r \Psi(a, \widehat{x})|_2} = c_\Psi. \tag{C-V}$$

This implies $\mathcal{M}_\Gamma = c_\Psi \text{Id}_\Gamma$. This assumption is fulfilled for all the considered examples with the exception of the ellipse. The geometrical intuition behind this assumption is that the reciprocal length of the normal vector (represented by the left hand side of (C-V)) for the case of an ellipse depends on the position on the boundary, i.e. it is a non-constant function of the coordinates $\hat{x} \in \Gamma$. Since the full Dirichlet-to-Neumann map has to represent differentiation in direction of the unit normal vector, cp. equation (2.18), yet the separable part $\partial_r (u \circ \Psi)(a, \hat{x})$ only represents differentiation with respect to the unnormalized normal vector, an additional multiplication with $\mathcal{M}_\Gamma$ is necessary. This multiplication by $\mathcal{M}_\Gamma$ would give rise to a non-separable part in $\langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}u, v \rangle_\Gamma$ if the left hand side of (C-V) were not constant, which would complicate the analysis. Perhaps assumption (C-V) could be eliminated by a refinement of the analysis presented here. However, as this thesis is mainly concerned with the helioseismology problem for which (C-V) holds true, this matter will be left for future investigations.

The argument presented here proceeds in a classical manner by establishing well-posedness via the Fredholm alternative. This consists of three main steps:

1. Showing that solutions of (2.59) must be unique if they exist.

2. Proving a Gårding inequality for the sesquilinear-form $b(\cdot, \cdot)$.

3. Rewriting (2.59) as an operator equation which is invertible up to a compact perturbation.

The argument starts by investigating properties of the convection term. The idea is to show that it can be written in an equivalent form from which it is easy to infer that it does not disturb in proving uniqueness.

Lemma 2.14 (Properties of convection term). *Let* $b_{\mathbf{b}}(u, v) := i\, (\beta \mathbf{b} \cdot \nabla\, (\beta u)\, , v)_{\Omega_{\text{int}}}$ *for* $u, v \in V$.

(a) *An equivalent form is given by*

$$b_{\mathbf{b}}(u, v) = \frac{i}{2} \int_{\Omega_{\text{int}}} \beta \mathbf{b} \cdot (\nabla(\beta u)\bar{v} - \nabla(\beta \bar{v})u)\ dx$$

for $u, v \in V$.

(b) *It holds that* $b_{\mathbf{b}}(u, u)$ *is real for any* $u \in V$.

38

Proof. (a) Since the flow is divergence free with vanishing normal component on $\partial\Omega_{\text{int}}$, an application of Gauss' theorem yields for any $\phi, \psi \in V$:

$$0 = \int_{\partial\Omega_{\text{int}}} \mathbf{b} \cdot \mathbf{n} \, \phi\bar{\psi} \, \mathrm{d}S = \int_{\Omega_{\text{int}}} \operatorname{div}\left(\mathbf{b}\phi\bar{\psi}\right) \, \mathrm{d}x = \int_{\Omega_{\text{int}}} \mathbf{b} \cdot \left(\nabla\phi\bar{\psi} + \phi\nabla\bar{\psi}\right) \, \mathrm{d}x.$$

With $\phi = \beta u$ and $\psi = \beta v$ for $u, v \in V$ one obtains

$$\int_{\Omega_{\text{int}}} \beta\mathbf{b} \cdot \nabla(\beta u)\bar{v} \, \mathrm{d}x = -\int_{\Omega_{\text{int}}} \beta\mathbf{b} \cdot \nabla(\beta\bar{v})u \, \mathrm{d}x.$$

It follows that

$$b_{\mathbf{b}}(u, v) = \frac{i}{2}\int_{\Omega_{\text{int}}} \beta\mathbf{b} \cdot \nabla(\beta u)\bar{v} \, \mathrm{d}x + \frac{i}{2}\int_{\Omega_{\text{int}}} \beta\mathbf{b} \cdot \nabla(\beta u)\bar{v} \, \mathrm{d}x$$

$$= \frac{i}{2}\int_{\Omega_{\text{int}}} \beta\mathbf{b} \cdot \left(\nabla(\beta u)\bar{v} - \nabla(\beta\bar{v})u\right) \, \mathrm{d}x.$$

(b) Using the equivalent expression established in part (a) gives

$$b_{\mathbf{b}}(u, u) = \frac{i}{2}\int_{\Omega_{\text{int}}} \beta\mathbf{b} \cdot \left(\psi - \bar{\psi}\right) \, \mathrm{d}x$$

with $\psi := \nabla(\beta u)\bar{u} = \operatorname{Re}(\psi) + i\operatorname{Im}(\psi)$. Since $\psi - \bar{\psi} = 2i\operatorname{Im}(\psi)$ and β as well as $\mathbf{b}$ are real-valued, it follows that

$$b_{\mathbf{b}}(u, u) = -\int_{\Omega_{\text{int}}} \operatorname{Re}\left(\beta\mathbf{b}\right)\operatorname{Im}(\psi) \, \mathrm{d}x,$$

which is real-valued.

$\square$

The next step is to show uniqueness. This can be established by an assumption on the imaginary part of *dtn*, as has been noted in one way or another by many authors who proved similar uniqueness results for the constant coefficient Helmholtz equation. See for instance, [GK95, Theorem 3.1], [Ihl98, Theorem 3.2], [DI01, section 3.2] or [Nan08, Satz 3.9 (iii)].

Proposition 2.15 (Uniqueness of continuous problem)**.** *Assume one of the following*

(i) (*dtn-I$_\leq$*) *and* $\operatorname{Im} c \leq c_0 < 0$ *in* Ω_{int}*, or*

(ii) (*dtn-I$_<$*)*.*

Then (2.59) *admits at most one solution.*

Proof. Assume that there exists

$$u \in V \text{ such that } b(u, v) = 0 \text{ for all } v \in V. \tag{2.61}$$

39

Choose $u = v$. When taking the imaginary part of $b(u, u) = 0$, the convection term drops out due to Lemma 2.14 (b). The diffusion term disappears as well since $A = (a_{ij})$ is assumed to be symmetric and real valued (thus hermitian). This results in

$$(\text{Im}(c)u, u)_{\Omega_{\text{int}}} + c_\Psi \sum_{\ell=0}^{\infty} \text{Im}\left(\textit{dtn}(\lambda_\ell)\right) |\langle u, w_\ell \rangle_\Gamma|^2 = 0.$$

Two cases according to the different assumptions have to be distinguished.

(i) In this case the assumptions yield that

$$0 \leq c_0(u, u)_{\Omega_{\text{int}}}.$$

Since $c_0 < 0$, this implies $u = 0$ and we are done.

(ii) This case is slightly more involved. From the general assumption (C-III), i.e. $\text{Im}(c) \leq 0$ and ($\textit{dtn}$-I$_<$) we conclude that

$$|\langle u, w_\ell \rangle_\Gamma|^2 = 0 \text{ for all } \ell \in \mathbb{N}_0.$$

This means that u vanishes on Γ. The normal derivative $\nabla u \cdot \mathbf{n}_\Gamma$ has to vanish as well, which can be inferred as follows. Interior elliptic regularity implies that $u \in H^2_{\text{loc}}(\Omega_{\text{int}})$. Using (2.61) for test functions vanishing on $\partial \Omega_{\text{int}}$ yields

$$0 = b_{\text{int}}(u, v) = (\mathcal{L}u, v)_{\Omega_{\text{int}}}.$$

Hence,

$$\mathcal{L}u = 0 \text{ a.e. in } \Omega_{\text{int}}, \quad u = 0 \text{ on } \Gamma \cup \Gamma_D.$$

Due to the homogeneous boundary condition the H^2-regularity extends up to Γ which implies that the normal derivative $\nabla u \cdot \mathbf{n} \in H^{1/2}(\Gamma)$ is well-defined. Let $v_\Gamma \in H^{1/2}(\Gamma)$ be arbitrary. Define $v_\chi := \chi Z v_\Gamma$ where $\chi \in C^\infty(\bar{\Omega}_{\text{int}})$ is a bump function such that $\chi \equiv 1$ near Γ and χ vanishes on $\Gamma_N \cup \Gamma_D$. Then integration by parts yields

$$0 = b_{\text{int}}(u, \bar{v}_\chi) = -\int_\Gamma \nabla u \cdot \mathbf{n}_\Gamma \, \text{tr} \, v_\chi \, dS,$$

where assumption (C-IV) has been employed. Since $\text{tr} \, v_\chi = \text{tr} \, Z v_\Gamma = v_\Gamma$, which was arbitrary it follows that $\nabla u \cdot \mathbf{n}_\Gamma$ vanishes on Γ. By extending u by zero into Ω_{ext} one obtains a function $\hat{u} \in H^1(\Omega)$. This solves $\mathcal{L}\hat{u} = 0$ in Ω_{ext} since the equation is homogeneous. As u and its normal derivative[5] vanish on Γ, the function $\hat{u}$ is a weak solution of $\mathcal{L}\hat{u} = 0$ in any compact set of $K \subset \Omega$ (in particular $u \in H^1(K)$). Similar as above elliptic regularity implies $\mathcal{L}\hat{u} = 0$ almost everywhere in K. This allows to apply[6] the unique continuation principle for elliptic equations as stated in Theorem A.5. Since $\hat{u}$ vanishes on Ω_{ext}, which contains an open subset of $\mathbb{R}^d$, it follows that $\hat{u}$ vanishes on all of K. Because K was arbitrary, $\hat{u}$ and therefore u vanish on all of Ω_{int}. Hence, the solution is unique if it exists.

$\square$

---

[5]This implies that boundary terms on Γ stemming from integration by parts vanish.
[6]Note that the principal part of $\mathcal{L}$ is real-valued and smooth by assumption.

Conditions (i) and (ii) in Proposition 2.15 are sufficient, yet not necessary for uniqueness to hold. For the waveguide one has $\operatorname{Im} \mathbf{dtn}^{\mathrm{guide}}(\lambda_\ell) = 0$ for $\lambda_\ell > k^2$ so that condition (ii) is violated. This suggests to investigate whether the solution for the undamped waveguide, i.e. $\operatorname{Im} c = 0$ so that (i) does not hold either, is still unique with $\mathcal{DtN}$ as transparent boundary condition. Example 2.16 shows uniqueness under assumption (2.49) for a very specific geometrical configuration, while Remark 2.17 comments on the more general case.

Example 2.16 (Uniqueness for a special waveguide under assumption (2.49))**.** *Consider a semi-infinite waveguide with $\Omega = [0, \infty) \times \tilde{\Gamma}$ for $\tilde{\Gamma} = [0, \pi]$. Let $\Omega_{\mathrm{int}} = [0, a] \times \tilde{\Gamma}$ and assume $\beta = 0, c = 1$ and $\mathbf{b} = 0$ in (2.15) so that the PDE in coordinates $(r, \hat{x}) \in \Omega_{\mathrm{int}}$ is separable. For homogeneous Dirichlet boundary conditions at the top and bottom of the waveguide the general solution in Ω_{int} is given by*

$$u(r, \hat{x}) = \sum_{\ell=0}^{\infty} \sin(\sqrt{\lambda_\ell}\hat{x}) \left[A_\ell \exp(+i\sqrt{k^2 - \lambda_\ell}(r - a)) + B_\ell \exp(-i\sqrt{k^2 - \lambda_\ell}(r - a)) \right]$$

for complex constants A_ℓ and B_ℓ. The question is, whether the $\mathcal{DtN}$ boundary condition at $r = a$ and another homogeneous boundary condition at $r = 0$ guarantee that these constants must vanish. By applying the definition (2.13) of $\mathcal{DtN}$ with $u_0 = u(a, \hat{x}) = \sum_\ell \sin(\sqrt{\lambda_\ell}\hat{x})(A_\ell + B_\ell)$, it follows that

$$\mathcal{DtN}u(a, \hat{x}) = \sum_{\ell=0}^{\infty} \sin(\sqrt{\lambda_\ell}\hat{x}) \, \mathbf{dtn}^{\mathrm{guide}}(\lambda_\ell)(A_\ell + B_\ell).$$

On the other hand, by direct differentiation of the solution formula

$$-(\partial_r u)(a, \hat{x}) = \sum_{\ell=0}^{\infty} \sin(\sqrt{\lambda_\ell}\hat{x}) \left(-i\sqrt{k^2 - \lambda_\ell} \right) (A_\ell - B_\ell).$$

The $\mathcal{DtN}$ condition requires that $\mathcal{DtN}u(a, \hat{x}) = -(\partial_r u)(a, \hat{x})$ which in view of $\mathbf{dtn}^{\mathrm{guide}}(\lambda_\ell) = -i\sqrt{k^2 - \lambda_\ell}$ leads to

$$\mathbf{dtn}^{\mathrm{guide}}(\lambda_\ell)\left[A_\ell + B_\ell\right] = \mathbf{dtn}^{\mathrm{guide}}(\lambda_\ell)\left[A_\ell - B_\ell\right].$$

Hence, $B_\ell = 0$ as long as $\mathbf{dtn}^{\mathrm{guide}}(\lambda_\ell) \neq 0$ is satisfied. The latter is ensured by (2.49) ,i.e. that k^2 is not in the spectrum of $-\Delta_\Gamma$. Homogeneous Neumann or Dirichlet boundary conditions at $r = 0$ will then imply that $A_\ell = 0$ so that the solution is unique, although conditions (i) and (ii) of Proposition 2.15 are violated.

Uniqueness with $\mathcal{DtN}$ boundary conditions has also been shown for the convected, constant coefficient Helmholtz equation in the more general case of a uniform flow in an infinite duct [BDL04], which served as inspiration for this example. A similar condition as (2.49), involving the magnitude of the flow, is necessary and sufficient for well-posedness.

Remark 2.17 (Trapped modes)**.** *It is important to point out that Example 2.16 only proved uniquenss of the solution for a very specific geometrical setup. In more general situations uniqueness can fail. It is well-known that some waveguides are capable of supporting so called 'trapped modes', which are solution of the homogeneous equation subject to certain boundary conditions that have finite energy and decay towards infinity. For example, Evans et al. [ELV94] proved that an infinite Neumann waveguide bounded by two parallel lines which contains a symmetric obstacle at its centerline supports at least one trapped mode. A comprehensive review of the literature on trapped modes occuring in water waves and acoustics was provided by Linton and McIver [LM07].*

According to the introduced plan, properties of the sesquilinear form $b(\cdot, \cdot)$ should now be investigated. While treatment of the interior part $b_{\text{int}}(\cdot, \cdot)$ is fairly standard, the term $\langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N} u, v \rangle_\Gamma$, representing the transparent boundary condition, requires special care. Formally, one has

$$V \times V : (u, v) \mapsto \langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N} u, v \rangle_\Gamma = c_\Psi \sum_{\ell=0}^\infty dt\!n(\lambda_\ell) \langle u, w_\ell \rangle_\Gamma \overline{\langle v, w_\ell \rangle_\Gamma}.$$

As mentioned in Section 2.1.2, continuity of this operation on $V \times V$ requires that $|dt\!n(\lambda_\ell)|$ asymptotically has to behave like $\sim (1 + |\lambda_\ell|)^{1/2}$ This is ensured by assumption ($dt\!n$-III).

Another crucial ingredient needed for application of Fredholm theory in the considered setting is a Gårding inequality. For the contribution stemming from the transparent boundary condition this requires to show a lower bound of the form

$$\operatorname{Re}\langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N} u, u \rangle_\Gamma \geq -c_W \|u\|_W^2,$$

where W is a space such that the embedding $V \hookrightarrow W$ is compact and $c_W \geq 0$. The natural candidate is $W = L^2(\Gamma)$. If $\operatorname{Re} dt\!n(\lambda_\ell) \geq 0$ for all $\ell \in \mathbb{N}_0$, then such an upper bound is valid for $c_W = 0$. However, from the plot of $dt\!n^{\text{jump}}$ or $dt\!n^{\text{VAL}-\text{C}}$ it can be see that this assumption is in general not valid for all $\ell \in \mathbb{N}_0$. Asymptotically, $\operatorname{Re} dt\!n(\lambda_\ell)$ converges towards infinity as $\ell \to \infty$ and so is eventually positive, yet a finite number of outliers for small ℓ with $\operatorname{Re} dt\!n(\lambda_\ell) < 0$ might be expected to arise from poles of $dt\!n(\lambda)$ lying in close proximity of the real axis. Assumption ($dt\!n$-II) formalizes this observation. Since the number of outliers is assumed to be bounded, they give rise to a compact perturbation in $L^2(\Gamma)$. According to the proof below, one might expect that the magnitude of the constant c_W is determined by the strength of these poles.

Lemma 2.18 (Gårding inequality). *Let ($dt\!n$-III) hold. Then the sesquilinear form $b(\cdot, \cdot)$ from (2.60) is continuous. Moreover, if ($dt\!n$-II) holds, then the lower bound*

$$\operatorname{Re} b(u, u) \geq C_V \|u\|_V^2 - C_{L^2(\Omega_{\text{int}})} \|u\|_{L^2(\Omega_{\text{int}})}^2 - C_{L^2(\Gamma)} \|u\|_{L^2(\Gamma)}^2, \tag{2.62}$$

for constants C_V and $C_{L^2(\Omega_{\text{int}})}$ as defined in (2.22) and $C_{L^2(\Gamma)} := c_\Psi \max_{\ell \in L_-} |\operatorname{Re} dt\!n(\lambda_\ell)|$ is fulfilled.

Proof. Continuity requires assumption ($dt\!n$-III) while the lower bound relies on ($dt\!n$-II).

- Continuity of the interior bilinear form $b_{\text{int}}(\cdot, \cdot)$ is known from Lemma 2.2. To check continuity of $\langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N} u, v \rangle_\Gamma$ it is useful to utilize the spectral characterization of the norm on $H^{1/2}(\Gamma)$ derived in Appendix B:

$$\|u\|_{H^{1/2}(\Gamma)}^2 = \sum_{\ell=0}^\infty (1 + |\lambda_\ell|)^{1/2} |\langle u, w_\ell \rangle_\Gamma|^2.$$

Then ($dt\!n$-III) yields

$$|\langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N} u, v \rangle_\Gamma| = c_\Psi \left| \sum_{\ell=0}^\infty dt\!n(\lambda_\ell) \langle u, w_\ell \rangle_\Gamma \overline{\langle v, w_\ell \rangle_\Gamma} \right|$$

$$\leq C_{\text{III}} c_\Psi \sum_{\ell=0}^\infty (1 + |\lambda_\ell|)^{1/4} |\langle u, w_\ell \rangle_\Gamma| (1 + |\lambda_\ell|)^{1/4} |\langle v, w_\ell \rangle_\Gamma|$$

$$\leq C_{\text{III}} c_\Psi \|u\|_{H^{1/2}(\Gamma)} \|v\|_{H^{1/2}(\Gamma)},$$

by the Cauchy-Schwarz inequality.

- As the Gårding inequality for $b_{\text{int}}(\cdot,\cdot)$ has already been established in Lemma 2.2, it remains to treat the contribution from $\mathcal{DtN}$ on Γ. We have

$$\operatorname{Re}\langle \mathcal{M}_\Gamma \mathcal{DtN} u, u\rangle_\Gamma \geq c_\Psi \sum_{\ell\in L_+} \operatorname{Re} \mathit{dtn}(\lambda_\ell)|\langle u, w_\ell\rangle_\Gamma|^2 - c_\Psi \max_{\ell\in L_-} |\operatorname{Re} \mathit{dtn}(\lambda_\ell)| \sum_{\ell\in L_-} |\langle u, w_\ell\rangle_\Gamma|^2.$$

Since $\operatorname{Re} \mathit{dtn}(\lambda_\ell) \geq 0$ for $\ell \in L_+$ by assumption (dtn-II), this yields the claim with $C_{L^2(\Gamma)} := c_\Psi \max_{\ell\in L_-} |\operatorname{Re} \mathit{dtn}(\lambda_\ell)|$, which is finite because L_- is assumed to be a finite set.

$\square$

Let $B \in \mathscr{L}(V,V)$ bet the continuous linear operator associated with the sesquilinear form $b(\cdot,\cdot)$, i.e.

$$b(u,v) = (Bu,v)_V, \quad u,v \in V.$$

Problem (2.59) can be recast as an operator equation:

$$\text{Find } U \in V \text{ such that } BU = L, \tag{2.63}$$

for $L \in V \cong V^*$ such that $l(v) = (L,v)_V$ for any $v \in V$. The next theorem shows that the assumptions (dtn-II) and (dtn-III) imply that B is a Fredholm operator.

Theorem 2.19 (Fredholm property). *Assume that (dtn-II) and (dtn-III) hold. Then $B = \tilde{B} - K$, where $\tilde{B}$ is continously invertible and K compact. Moreover,*

- *either the equation (2.59) is uniquely solvable for any $l \in V^*$,*

- *or the homogeneous equation posesses a finite-dimensional null space, i.e. $0 < \dim(\ker(B)) = \dim(\ker(B^*)) < \infty$ and (2.63) has a solution if and only if L fulfills the compatability condition $(L,X)_V = 0$ for all $X \in \ker(B^*)$.*

Proof. By Lemma 2.18 the operator $\tilde{B}$ associated with the sesquilinear form

$$\tilde{b}(u,v) := b(u,v) + C_{L^2(\Omega_{\text{int}})}(u,v)_{\Omega_{\text{int}}} + C_{L^2(\Gamma)}\langle u,v\rangle_\Gamma$$

is continuously invertible as $\tilde{b}(\cdot,\cdot)$ is coercive. The embedding $\iota_{\text{int}} : V \hookrightarrow L^2(\Omega_{\text{int}})$ is compact by the Rellich-Kondrachov theorem. Likewise, $\iota_\Gamma : V \xrightarrow{\text{tr}} H^{1/2}(\Gamma) \hookrightarrow L^2(\Gamma)$ is compact, since the trace operator is continuous and $H^{1/2}(\Gamma) \hookrightarrow L^2(\Gamma)$ is compact. Then (denoting the adjoint maps by a star) we obtain

$$\begin{aligned}
(\tilde{B}u,v)_V &= (Bu,v)_V + C_{L^2(\Omega_{\text{int}})}(\iota_{\text{int}}u, \iota_{\text{int}}v)_{\Omega_{\text{int}}} + C_{L^2(\Gamma)}\langle \iota_\Gamma u, \iota_\Gamma v\rangle_\Gamma \\
&= (Bu,v)_V + C_{L^2(\Omega_{\text{int}})}(\iota_{\text{int}}^* \iota_{\text{int}}u, v)_V + C_{L^2(\Gamma)}(\iota_\Gamma^* \iota_\Gamma u, v)_V \\
&= ((B+K)u,v)_V,
\end{aligned}$$

where $K = C_{L^2(\Omega_{\text{int}})}\iota_{\text{int}}^* \iota_{\text{int}} + C_{L^2(\Gamma)}\iota_\Gamma^* \iota_\Gamma$ is compact. Hence, $B = \tilde{B} - K$, where $\tilde{B}$ is invertible and K compact. The claim follows by applying the Fredholm alternative as stated in Theorem A.2. $\square$

Corollary 2.20. *In addition to the assumptions of Theorem 2.19 let*

(a) either (dtn-I$_<$)

(b) *or* (dtn-I$_\leq$) *and* $\operatorname{Im} c \leq c_0 < 0$ *in* Ω_{int} *be fulfilled.*

Then the problem (2.59) *admits a unique solution for any right hand side* $l \in V^*$.

Proof. By Proposition 2.15 the assumptions (a) or (b) imply uniqueness, i.e. $\ker(B) = 0$. Hence, the first part of the Fredholm alternative stated in Theorem 2.19 applies. The Fredholm alternative also yields continuous dependence on the data, so that the problem is indeed well-posed. $\square$

dtn	(dtn-I$_<$)	(dtn-I$_\leq$)	(dtn-II)	(dtn-III)
$dtn^{\mathrm{hom,2d}}$	✔	✔	✔	✔
$dtn^{\mathrm{hom,3d}}$	✔	✔	✔	✔
dtn^{jump}	(✔)	✔	(✔)	✔
dtn^{ellipse}	(✔)	(✔)	(✔)	(✔)
dtn^{Atmo}	(✔)	(✔)	(✔)	(✔)
$dtn^{\mathrm{VAL\text{-}C}}$	✔	✔	✔	✔
dtn^{guide}	✘	✔	✔	✔

Table 2.1: Check of the assumptions for different dtn functions. A '✔' means that these properties have been proven analytically, while a '(✔)' signifies a numerical verification. Violated assumptions are marked with '✘'.

It is interesting to conclude with a discussion to what extent the presented analysis applies to the example problems introduced in Section 2.2-Section 2.4. A check of the assumptions on dtn for the specific examples is provided in Table 2.1.

- For *scattering* problems with *spherical* coupling boundaries, as introduced in Section 2.2.1, the theory fully applies. For homogeneous exterior domains the assumptions on dtn have also been verified analytically. Whereas a proof of (dtn-II) remains an open problem for the case of a jumping wavenumber, the other properties (dtn-I$_\leq$) and (dtn-III) have been established rigorously.

- Currently, the only obstruction to applying the theory to *elliptical* coupling boundaries is a violation of assumption (C-V). A more sophisticated analysis may perhaps remove this assumption by utilizing that $\mathcal{M}_\Gamma : H^s(\Gamma) \to H^s(\Gamma)$ is an isomorphism.

- The *helioseismology* problem of Section 2.4 is covered in both formulations and for both atmospheric models. Proofs for all the required properties of $dtn^{\mathrm{VAL\text{-}C}}$ have been provided in Section 2.4.2.

- The approach presented here has some limitations for *waveguides*. Firstly, the spectral characterization of Sobolev spaces on Γ derived in Appendix B and utilized in the analysis has only been derived for Γ compact without boundary. Therefore, the analysis only applies rigorously to periodic waveguides, e.g. if one can identify $\tilde{\Gamma} \simeq \mathbb{S}^{d-1}$. To this end, periodicity conditions should be incorporated into the space V defined in (2.19).

While the first limitation is a shortcoming of the analysis that can be overcome, the second limitation concerning our inability to establish a general uniqueness result in the absence of damping is rooted in the nature of wave propagation in waveguides. As explained in Remark 2.17, waveguides may support trapped modes in which case uniqueness fails. However, as dtn^{guide} fulfills properties (dtn-II) and (dtn-III) the solution operator is still Fredholm according to Theorem 2.19. This implies that only a finite number of trapped modes can propagate in the waveguide and that the equation is still solvable under certain compatability conditions on the data.

The well-posedness result established here provides the foundation for the analysis of tensor product discretizations of $\mathcal{DtN}$ as presented in Chapter 3.

Chapter 3

Tensor-product discretizations of DtN

Chapter 2 introduced time-harmonic wave equations posed in stratified media on possibly unbounded domains at the continuous level. Such problems can be reduced to a finite computational domain by imposing the $\mathcal{DtN}$ map as a transparent boundary condition on the artificial truncation boundary Γ. If the geometry and the differential operator in the exterior are separable, then $\mathcal{DtN}$ may be diagonalized in the eigenbasis of a suitable self-adjoint differential operator on Γ. In particular, all information about $\mathcal{DtN}$ is contained in a scalar function dtn, which describes the diagonal entries of $\mathcal{DtN}$ in the eigenbasis.

This chapter deals with tensor-product discretizations DtN_N of $\mathcal{DtN}$ for which essentially all of the above statements carry over to the discrete level. The notion of such tensor-product discretizations is made precise in Section 3.1. The main insight is that all information about DtN_N is also contained in a single scalar function called dtn_N whose ability to approximate the continuous dtn function determines the accuracy of any (approximate) transparent boundary condition of tensor-product type. The problem of approximating dtn by dtn_N is analyzed in Section 3.2. It turns out that highly accurate approximations can be achieved with negligible effort on bounded subsets of the spectrum of $-\Delta_\Gamma$ (or a more general operator).

In Section 3.3, these result are extended to a semi-discrete error analysis which bounds the H^1-error between discrete and approximate solution for the PDE in terms of the approximation of dtn by dtn_N. This estimate does not include any finite element discretization error because the discretization of the interior problem is assumed to be perfect in this setting (explaining the specification 'semi-discrete'). Firstly, the focus is entirely on understanding the error introduced by the approximation of $\mathcal{DtN}$. Later, a basic convergence result in the fully-discrete setting is established, see Corollary 3.23.

The semi-discrete convergence analysis presented in this chapter is inspired by the paper [DI01] of Demkowicz and Ihlenburg in which a similar analysis has been carried out for a specific type of classical infinite elements for the Helmholtz equation in the exterior of a sphere. The analysis presented in this thesis aims to achieve a higher level of generality in the following aspects:

- In view of Chapter 2, it should be applicable to a larger class of PDEs and geometrical configurations. Our main objective is to cover the case of helioseismology, which requires extensions of the results of [DI01]. For example, the analysis in this reference is based on positive definiteness of $\mathrm{Re}\,\mathit{dtn}(\lambda_\ell)$, which can be shown for the Helmholtz equation, see Lemma 2.6. However, for helioseismology $\mathrm{Re}\,\mathit{dtn}^{\mathrm{VAL-C}}$ can

become negative due to the occurence of poles close the real axis, see Fig. 2.8.

- It should be applicable to any (approximate) transparent boundary condition of tensor-product type provided the corresponding *dtn* and dtn_N functions have been identified and fulfill the required assumptions.

Whereas many instances of continuous *dtn* functions have been presented in Chapter 2, specific examples for discrete dtn_N functions will not appear until Section 3.4. In this section several dtn_N functions arising from tensor-product transparent boundary conditions for the Helmholtz equation in the exterior of a sphere will be compared based on their ability to approximate *dtn*. The corresponding dtn_N functions are explicitly derived in Appendix C. The range of examples includes classical infinite elements, Hardy space infinite elements, tensor-product PMLs and some local absorbing conditions.

3.1 Describing discrete DtN_N operators through dtn_N

This section deals with tensor-product discretizations of exterior problems. Section 3.1.1 introduces these type of discretizations in detail and describes their associated DtN_N operators. Similar to the continuous $\mathcal{DtN}$ map, its discrete approximation DtN_N can be diagonalized in a discrete eigenbasis and described through a scalar function called dtn_N. This is shown in Section 3.1.2. Most of the results of Section 3.1.1 and Section 3.1.2 appeared previously in our publication [HLP21]. Here we go a step further and present a first a priori error estimate in terms of the approximation error of *dtn* by dtn_N in Section 3.1.3.

3.1.1 From a dense to a sparse system involving additional DOFs

In the previous chapter, a PDE posed on an infinite domain with a radiation condition at infinity has been reduced to a bounded computational domain by introducing the $\mathcal{DtN}$ map as transparent boundary condition. This leads to the variational formulation: Find $u \in V$ such that

$$b_{\text{int}}(u, v) + \langle \mathcal{M}_\Gamma \mathcal{DtN} u, v \rangle_\Gamma = l(v), \quad \text{for all } v \in V. \tag{3.1}$$

To obtain an approximate solution, this variational formulation will be discretized using a conforming finite element discretization. Let $\{\phi_j\}$ denote the basis functions of a finite dimensional subspace $V_h \subset V$. The entries of the matrix L corresponding to the differential operator $\mathcal{L}$ operating only in Ω_{int} are obtained as usual from $b_{\text{int}}(\phi_j, \phi_i)$. This matrix is sparse since the basis functions have local support and the differential operator $\mathcal{L}$ is local, i.e. if $\text{supp}(u)$ denotes the support of a function u, then $\text{supp}(\mathcal{L}u) \subseteq \text{supp}(u)$ holds true. Unfortunately, $\mathcal{DtN}$ is a non-local operator. So, if $\{\phi_i\}$ for $i = 1, \ldots, n_\Gamma$ denote the basis functions with support on Γ, then even though $\text{supp}(\phi_j)$ commonly only occupies a tiny fraction of Γ, after applying the $\mathcal{DtN}$ map $\text{supp}(\mathcal{DtN}\phi_j)$ has usually expanded to all of Γ. Hence,

$$\langle \mathcal{DtN}\phi_j, \phi_i \rangle_\Gamma, \neq 0$$

for most (i, j). This implies that the matrix DtN^{ext} that expresses $\mathcal{DtN}$ in the finite element basis is dense.

Partitioning the degrees of freedom (DOFs) as $\underline{u}_I$ belonging to the interior of the domain Ω_{int} (including $\Gamma_N \cup \Gamma_D$) and a vector $\underline{u}_\Gamma$ of DOFs on the transparent boundary $\Gamma \subset \partial\Omega_{\mathrm{int}}$ the linear system obtained from (3.1) takes the form

$$\begin{bmatrix} L_{II} & L_{I\Gamma} \\ L_{\Gamma I} & L_{\Gamma\Gamma}^{\mathrm{int}} + M\,\mathrm{DtN}^{\mathrm{ext}} \end{bmatrix} \begin{bmatrix} \underline{u}_I \\ \underline{u}_\Gamma \end{bmatrix} = \begin{bmatrix} \underline{f}_I \\ \underline{f}_\Gamma^{\mathrm{int}} \end{bmatrix}.$$

Here M corresponds to the finite element discretization of $\mathcal{M}_\Gamma$ on Γ, i.e.

$$M_{ij} = \int_\Gamma 1/|\partial_r \Psi|_2 \phi_j \phi_i \, \mathrm{d}\hat{x}. \tag{3.2}$$

The block $L_{\Gamma\Gamma}^{\mathrm{int}}$ contains contributions from Ω_{int}, especially volume terms close to Γ.

Since the full matrix $\mathrm{DtN}^{\mathrm{ext}}$ is dense, its use may be computationally inefficient. Particularly for helioseismology in which a drastic refinement in the surface layers of the Sun is required, using $\mathrm{DtN}^{\mathrm{ext}}$ for a fully-three dimensional discretization is computationally infeasible. Therefore, we would like to approximate system (3.3) by a sparse system involving additional degrees of freedom $\underline{u}_E$:

$$\begin{bmatrix} L_{II} & L_{I\Gamma} & 0 \\ L_{\Gamma I} & L_{\Gamma\Gamma}^{\mathrm{int}} + L_{\Gamma\Gamma} & L_{\Gamma E} \\ 0 & L_{E\Gamma} & L_{EE} \end{bmatrix} \begin{bmatrix} \underline{u}_I \\ \underline{u}_\Gamma \\ \underline{u}_E \end{bmatrix} = \begin{bmatrix} \underline{f}_I \\ \underline{f}_\Gamma^{\mathrm{int}} \\ 0 \end{bmatrix}. \tag{3.3}$$

To this end, sparse matrices $L_{\Gamma\Gamma}$, $L_{\Gamma E}$, $L_{E\Gamma}$ and L_{EE} of small size need to be found such that the Schur complement of L_{EE} approximates $M\,\mathrm{DtN}^{\mathrm{ext}}$:

$$\begin{bmatrix} L_{II} & L_{I\Gamma} \\ L_{\Gamma I} & L_{\Gamma\Gamma}^{\mathrm{int}} + M\,\mathrm{DtN} \end{bmatrix} \begin{bmatrix} \underline{u}_I \\ \underline{u}_\Gamma \end{bmatrix} = \begin{bmatrix} \underline{f}_I \\ \underline{f}_\Gamma^{\mathrm{int}} \end{bmatrix}, \tag{3.4}$$

where

$$\mathrm{DtN} \overset{!}{\approx} \mathrm{DtN}^{\mathrm{ext}} \qquad \text{with} \qquad \mathrm{DtN} := M^{-1}\left(L_{\Gamma\Gamma} - L_{\Gamma E} L_{EE}^{-1} L_{E\Gamma}\right). \tag{3.5}$$

Although (3.3) is the preferred form for implementation, equation (3.5) is very informative since it provides an explicit formula for the $\mathcal{DtN}$-approximation in terms of the matrix elements of

$$\begin{bmatrix} L_{\Gamma\Gamma} & L_{\Gamma E} \\ L_{E\Gamma} & L_{EE} \end{bmatrix}. \tag{3.6}$$

Many successful transparent boundary conditions lead to linear systems of the form (3.3). They can be distinguished by the choice of the matrix in (3.6). In this thesis, separable PDEs in separable exterior domains are considered. Therefore, it is natural (although not necessary) to impose that (3.6) has a tensor-product structure, i.e.

$$\begin{bmatrix} L_{\Gamma\Gamma} & L_{\Gamma E} \\ L_{E\Gamma} & L_{EE} \end{bmatrix} = A \otimes M + B \otimes K = \begin{bmatrix} A_{\Gamma\Gamma} & A_{\Gamma E} \\ A_{E\Gamma} & A_{EE} \end{bmatrix} \otimes M + \begin{bmatrix} B_{\Gamma\Gamma} & B_{\Gamma E} \\ B_{E\Gamma} & B_{EE} \end{bmatrix} \otimes K, \tag{3.7}$$

where K is the stiffness matrix associated with the discretization of $\mathcal{K}_\Gamma$ and M is as introduced in (3.2). In view of equation (2.5), this is a very natural choice. A conceptual sketch of the DOFs of such tensor-product discretizations and the couplings in the matrices $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ is presented in Fig. 3.1. From now on DtN_N will appear with the subscript $N \in \mathbb{N}_0$ to keep track of the dimension of the radial matrices A and B.

49

$$\begin{bmatrix} A_{\Gamma\Gamma} & A_{\Gamma E} \\ A_{E\Gamma} & A_{EE} \end{bmatrix} \otimes M + \begin{bmatrix} B_{\Gamma\Gamma} & B_{\Gamma E} \\ B_{E\Gamma} & B_{EE} \end{bmatrix} \otimes K$$
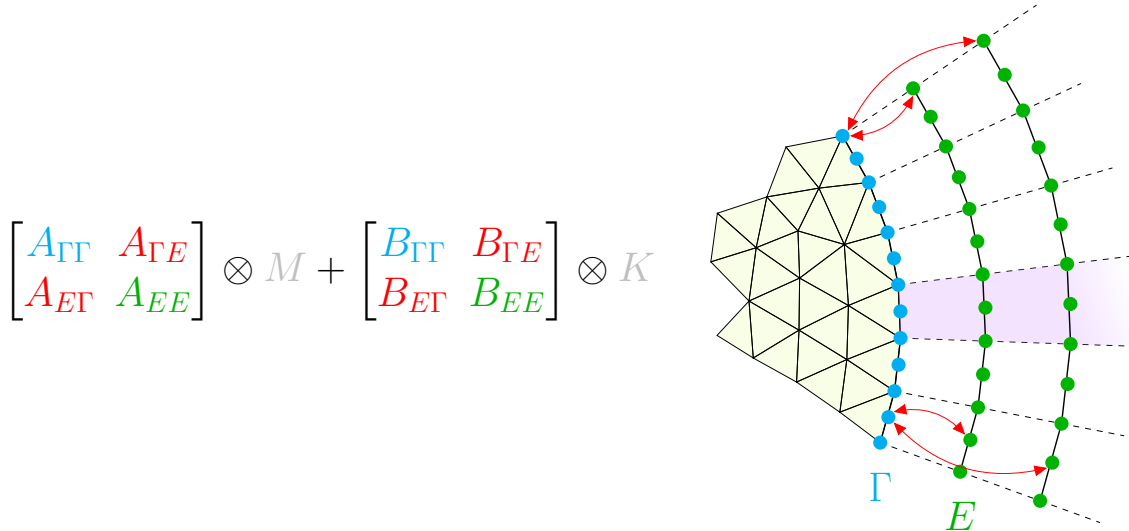
Figure 3.1: Schematic illustration of tensor-product discretization of separable exterior problems. The violet region indicates one such tensor-product element. The first DOFs lie on the transparent boundary Γ and are associated with the matrices $A_{\Gamma\Gamma}$ and $B_{\Gamma\Gamma}$. Additional DOFs in the exterior E are tensorized with copies of the FEM discretization of $(\mathcal{M}_\Gamma, \mathcal{K}_\Gamma)$ on Γ. The exterior E couples to Γ (and thereby to Ω_{int}) by means of $A_{\Gamma E}, B_{\Gamma E}, A_{E\Gamma}$ and $B_{E\Gamma}$.

For investigative purposes it is useful to consider an intermediate step between (2.5) and the fully discrete equation (3.7). A semi-discrete ansatz is obtained by retaining the continuous differential operators $\mathcal{K}_\Gamma$ and $\mathcal{M}_\Gamma$, i.e.

$$\begin{bmatrix} \mathcal{L}_{\Gamma\Gamma} & \mathcal{L}_{\Gamma E} \\ \mathcal{L}_{E\Gamma} & \mathcal{L}_{EE} \end{bmatrix} = \begin{bmatrix} A_{\Gamma\Gamma} & A_{\Gamma E} \\ A_{E\Gamma} & A_{EE} \end{bmatrix} \otimes \mathcal{M}_\Gamma + \begin{bmatrix} B_{\Gamma\Gamma} & B_{\Gamma E} \\ B_{E\Gamma} & B_{EE} \end{bmatrix} \otimes \mathcal{K}_\Gamma. \tag{3.8}$$

Roughly speaking, this ansatz corresponds to a perfect finite element discretization on Γ. Correspondingly, denote

$$\mathcal{D}t\mathcal{N}_N := \mathcal{M}_\Gamma^{-1} \left(\mathcal{L}_{\Gamma\Gamma} - \mathcal{L}_{\Gamma E} \mathcal{L}_{EE}^{-1} \mathcal{L}_{E\Gamma} \right). \tag{3.9}$$

3.1.2 Diagonalization of the (semi-)discrete DtN map

Plenty of well-known transparent boundary conditions are of the form (3.7) as shown in Appendix C. They can be distinguished by the choice of the matrices A and B. To assess their accuracy and efficiency, it is in principle possible to assemble the matrices DtN_N and DtN^{ext} and compute the error via some matrix norm $\| \text{DtN}^{\text{ext}} - \text{DtN}_N \|$ in terms of the number of DOFs. However, this is a rather crude measure which does not promise to offer many insights. Besides, the previous chapter revealed that at the continuous level all information[1] about $\mathcal{D}t\mathcal{N}$ is already contained in the scalar function dtn. The next proposition shows that an analogous characterization holds true for the discrete DtN_N and semi-discrete $\mathcal{D}t\mathcal{N}_N$ maps associated with the tensor-product discretization (3.7), respectively semi-discretization (3.8). This insight paves the way for a unified error analysis of $\mathcal{D}t\mathcal{N}_N$ operators associated with tensor-product discretizations.

---

[1]in the separable setting as considered in this thesis.

Proposition 3.1 (Diagonalization of DtN_N and $\mathcal{DtN}_N$). *Consider the fully-discrete and semi-discrete cases:*

- *If $(\underline{\lambda}_\ell, \underline{v}_\ell)$ solves the generalized eigenvalue problem $K\underline{v}_\ell = \underline{\lambda}_\ell M \underline{v}_\ell$, then*

$$\mathrm{DtN}_N \, \underline{v}_\ell = \mathrm{dtn}_N(\underline{\lambda}_\ell)\underline{v}_\ell,$$

with the rational function dtn_N defined, for $\lambda \in \mathbb{C}$ by

$$\mathrm{dtn}_N(\lambda) := A_{\Gamma\Gamma} + \lambda B_{\Gamma\Gamma} - (A_{\Gamma E} + \lambda B_{\Gamma E})(A_{EE} + \lambda B_{EE})^{-1}(A_{E\Gamma} + \lambda B_{E\Gamma}). \quad (3.10)$$

- *If (λ_ℓ, v_ℓ) solves the generalized eigenvalue problem $\mathcal{K}_\Gamma v_\ell = \lambda_\ell \mathcal{M}_\Gamma v_\ell$, then*

$$\mathcal{DtN}_N v_\ell = \mathrm{dtn}_N(\lambda_\ell)v_\ell,$$

for the same dtn_N function as in (3.10).

Proof. For the first term in the definition of DtN_N, cf. (3.5), we have

$$M^{-1}L_{\Gamma\Gamma}\underline{v}_\ell = M^{-1}(A_{\Gamma\Gamma} \otimes M + B_{\Gamma\Gamma} \otimes K)\underline{v}_\ell = M^{-1}(A_{\Gamma\Gamma}M\underline{v}_\ell + B_{\Gamma\Gamma}K\underline{v}_\ell)$$
$$= M^{-1}(A_{\Gamma\Gamma}M\underline{w}_\ell + B_{\Gamma\Gamma}\underline{\lambda}_\ell M\underline{v}_\ell) = (A_{\Gamma\Gamma} + \underline{\lambda}_\ell B_{\Gamma\Gamma})\underline{v}_\ell.$$

Using a similar computation and the rule $(C \otimes D)(E \otimes F) = (CE \otimes DF)$, we obtain

$$M^{-1}L_{\Gamma E}L_{EE}^{-1}L_{E\Gamma}\underline{v}_\ell$$
$$= M^{-1}[(A_{\Gamma E} + \underline{\lambda}_\ell B_{\Gamma E}) \otimes M][(A_{EE} + \underline{\lambda}_\ell B_{EE}) \otimes M]^{-1}[(A_{E\Gamma} + \underline{\lambda}_\ell B_{E\Gamma}) \otimes M]\underline{v}_\ell$$
$$= (A_{\Gamma E} + \underline{\lambda}_\ell B_{\Gamma E})(A_{EE} + \underline{\lambda}_\ell B_{EE})^{-1}(A_{E\Gamma} + \underline{\lambda}_\ell B_{E\Gamma})(M^{-1}MM^{-1}M)\underline{v}_\ell$$

for the second term. Putting both together yields the result. The proof for the semi-discrete version is analogous. $\qquad\square$

In the following, $\mathcal{DtN}_N$ will be analyzed instead of DtN_N. Since both operators are described by the same dtn_N function, the difference between them stems purely from the eigenvalue, respectively, eigenfunction approximation, which is determined by the finite element discretization error on Γ. This may be taken into account in a more elaborate analysis by using appropriate tools of finite element theory. Working with $\mathcal{DtN}_N$ allows to focus exclusively on the error introduced by the transparent boundary condition, which is the main objective of this thesis.

3.1.3 A first bound for the discretization error

In view of Proposition 3.1, the action of the operator $\mathcal{DtN}_N$ can be described as

$$\mathcal{DtN}_N u = \sum_{\ell=0}^{\infty} \mathrm{dtn}_N(\lambda_\ell)\langle u, w_\ell \rangle_\Gamma w_\ell.$$

Hence, it behaves exactly like the continuous $\mathcal{DtN}$ operator, cp. equation (2.13), except that $\mathcal{dtn}$ is replaced by dtn_N. However, since dtn_N grows usually linearly as $\lambda \to \infty$, see equation (3.10), while $\mathcal{dtn}$ behaves like $\mathcal{O}(\sqrt{\lambda})$, continuity of the operator $\mathcal{DtN}_N$ on the space V introduced in (2.19) cannot be ensured. To circumvent this issue, the space

$$V_+ := \{v \in H^1(\Omega_{\mathrm{int}}) \mid \mathrm{tr}_{\Gamma_\mathrm{D}} v = 0, \mathrm{tr}_\Gamma v \in H^1(\Gamma)\}, \quad (3.11)$$

with the stronger norm

$$\|v\|_{V_+}^2 := \|v\|_{H^1(\Omega_{\text{int}})}^2 + \|\operatorname{tr}_\Gamma v\|_{H^1(\Gamma)}^2, \tag{3.12}$$

is introduced. The variational formulation for the function u_N with a semi-discrete radiation condition provided by $\mathcal{D}t\mathcal{N}_N$ is given by: Find $u_N \in V_+$ such that

$$b_N(u_N, v) = l(v), \quad \text{for all } v \in V_+, \tag{3.13}$$

with the sesquilinear form

$$b_N(u, v) := b_{\text{int}}(u, v) + \langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}_N u, v \rangle_\Gamma. \tag{3.14}$$

We hold on to assumption (C-V) in this chapter so that $\mathcal{M}_\Gamma = c_\Psi \operatorname{Id}_\Gamma$. Let $u \in V$ denote the solution of (2.59), i.e. with the exact transparent boundary condition $\mathcal{D}t\mathcal{N}$. It will later be shown, see Lemma 3.19, that u usually enjoys a much higher regularity on Γ than merely $H^{1/2}(\Gamma)$ as guaranteed by $u \in V$. Therefore, it is not a restriction to assume at this point that $u \in V_+$. A minor generalization of [DI01, Theorem 2.1] given in Proposition 3.5 then shows that the error $\|u - u_N\|_{V_+}$ between the exact solution u and the approximate solution u_N of (3.13) can be bounded by the $\mathcal{D}t\mathcal{N}$-approximation achieved with $\mathcal{D}t\mathcal{N}_N$ in the operator norm. To estimate the latter, it is useful to introduce weighted Sobolev spaces on Γ.

Definition 3.2. *For a sequence (σ_ℓ) of positive weights with $\inf_\ell \sigma_\ell > 0$ define the space $H_\sigma(\Gamma) := \{u \in L^2(\Gamma) \mid \|u\|_{H_\sigma} < \infty\}$, where*

$$\|u\|_{H_\sigma}^2 := \sum_{\ell=0}^\infty \sigma_\ell^2 |\langle u, w_\ell \rangle_\Gamma|^2.$$

The next lemma shows that $H_\sigma(\Gamma)$ are Hilbert spaces which generalize the usual Sobolev spaces $H^s(\Gamma)$ on Γ.

Lemma 3.3. *Let (σ_ℓ) be a sequence of positive weights with $\inf_\ell \sigma_\ell > 0$. Then*

(a) $H_\sigma(\Gamma)$ is a Hilbert space.

(b) For $\sigma_\ell = (1 + |\lambda_\ell|)^{s/2}$ with $s \in \mathbb{R}$ it holds that $H_\sigma(\Gamma) = H^s(\Gamma)$.

Proof. (a) Positive definiteness of $\|\cdot\|_{H_\sigma}$ follows from $\inf_\ell \sigma_\ell > 0$. Only completeness needs to be shown as the other properties of a Hilbert space are clear. To this end, recall the isometric isomophism

$$L^2(\Gamma) \to \ell^2(\mathbb{N}_0) \quad \ell^2(\mathbb{N}_0) \to L^2(\Gamma)$$

$$u \mapsto (\langle u, w_\ell \rangle_\Gamma), \quad (a_\ell) \mapsto \sum_{\ell=0}^\infty a_\ell w_\ell.$$

Let (u^n) be a Cauchy sequence in $H_\sigma(\Gamma)$. This means that (a^n) with $a^n := (\sigma_\ell \langle u, w_\ell \rangle_\Gamma)$ is a Cauchy sequence in $\ell^2(\mathbb{N}_0)$. By completeness of $\ell^2(\mathbb{N}_0)$ it has a limit $a = \lim_{n\to\infty} a^n$ in $\ell^2(\mathbb{N}_0)$. Define $u = \sum_\ell \sigma_\ell^{-1} a_\ell w_\ell$. Then $\|u\|_{H_\sigma}^2 = \|a\|_{\ell^2(\mathbb{N}_0)}^2$, so $u \in H_\sigma(\Gamma)$ and

$$\|u - u^n\|_{H_\sigma}^2 = \sum_\ell |a_\ell - \sigma_\ell \langle u^n, w_\ell \rangle_\Gamma|^2 = \sum_\ell |a_\ell - a_\ell^n|^2 = \|a - a^n\|_{\ell^2(\mathbb{N}_0)}^2,$$

which goes to zero as $n \to \infty$.

(b) This is immediate from the characterization (2.24) of $H^s(\Gamma)$.

$\square$

The weights are often related to the modal error of the dtn_N approximation which is described in the next definition.

Definition 3.4. *For a dtn_N function as in (3.10) define the modal error:*

$$\text{err}_{dtn}(N, \ell) := |\boldsymbol{dtn}(\lambda_\ell) - \text{dtn}_N(\lambda_\ell)| \tag{3.15}$$

for $\ell \in \mathbb{N}_0$.

For convenience let us also introduce the projector

$$\Pi_L v := \sum_{\ell=0}^{L-1} \langle v, w_\ell \rangle_\Gamma w_\ell \tag{3.16}$$

onto the first L modes and note that

$$\|u\|_{H_\sigma}^2 = \|\Pi_L u\|_{H_\sigma}^2 + \|(\text{Id} - \Pi_L)u\|_{H_\sigma}^2 \tag{3.17}$$

holds true. Now the stage is set for deriving the first abstract error estimate.

Proposition 3.5. *Suppose that the variational problem (2.59) for given data l has a unique solution $u \in V_+$. Assume further that the semi-discrete problem (3.13) satisfies the inf-sup condition:*

$$\text{For all } u \in V_+: \quad \sup_{v \in V_+ \setminus \{0\}} \frac{|b_N(u, v)|}{\|v\|_{V_+}} \geq \gamma_N \|u\|_{V_+}, \tag{3.18}$$

with constant $\gamma_N > 0$. Moreover, suppose that the solution $u_N \in V_+$ of the semi-discrete problem (3.13) exists.

(i) *Then the error $\|u - u_N\|_{V_+}$ can be bounded as*

$$\|u - u_N\|_{V_+} \leq \gamma_N^{-1} \sup_{v \in V_+ \setminus \{0\}} \frac{|\langle \mathcal{M}_\Gamma (\boldsymbol{DtN} - \boldsymbol{DtN}_N)u, v \rangle_\Gamma|}{\|v\|_{V_+}}. \tag{3.19}$$

(ii) *Assume further that the solution u of (2.59) fulfills $\text{tr}_\Gamma u \in H_{\sigma^{\text{err}}}(\Gamma)$ with $\sigma_\ell^{\text{err}} := \text{err}_{dtn}(N, \ell)(1 + |\lambda_\ell|)^{-1/2}$. Then the error can be bounded by*

$$\|u - u_N\|_{V_+} \leq \gamma_N^{-1} c_\Psi \|u\|_{H_{\sigma^{\text{err}}}(\Gamma)} \leq \gamma_N^{-1} c_\Psi \left(\|\Pi_L u\|_{H_{\sigma^{\text{err}}}(\Gamma)} + \|(\text{Id} - \Pi_L)u\|_{H_{\sigma^{\text{err}}}(\Gamma)} \right). \tag{3.20}$$

Proof. The argument follows the one given by Demkowicz and Ihlenburg [DI01, Theorem 2.1] for the Helmholtz scattering problem who used standard Sobolev spaces instead of H_σ and V_+.

53

(i) Galerkin orthogonality and the definitions of the sesquilinear forms yield

$$
\begin{aligned}
b_N(u - u_N, v) &= b_N(u, v) - b(u, v) + b(u, v) - b_N(u_N, v) \\
&= b_N(u, v) - b(u, v) + l(v) - l(v) \\
&= b_N(u, v) - b(u, v) \\
&= b_{\mathrm{int}}(u, v) + \langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}_N u, v \rangle_\Gamma - b_{\mathrm{int}}(u, v) - \langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N} u, v \rangle_\Gamma \\
&= \langle \mathcal{M}_\Gamma (\mathcal{D}t\mathcal{N}_N - \mathcal{D}t\mathcal{N}) u, v \rangle_\Gamma.
\end{aligned}
$$

Since b_N was assumed to fulfill an inf-sup condition,

$$
\gamma_N \| u - u_N \|_{V_+} \le \sup_{v \in V_+ \setminus \{0\}} \frac{|b_N(u - u_N, v)|}{\|v\|_{V_+}} = \sup_{v \in V_+ \setminus \{0\}} \frac{|\langle \mathcal{M}_\Gamma (\mathcal{D}t\mathcal{N} - \mathcal{D}t\mathcal{N}_N) u, v \rangle_\Gamma|}{\|v\|_{V_+}}
$$

holds, which establishes the first part of the claim.

(ii) A straightforward calculation using the Cauchy-Schwarz inequality gives

$$
\sup_{v \in V_+ \setminus \{0\}} \frac{|\langle \mathcal{M}_\Gamma (\mathcal{D}t\mathcal{N} - \mathcal{D}t\mathcal{N}_N) u, v \rangle_\Gamma|}{\|v\|_{V_+}} = c_\Psi \sup_{v \in V_+ \setminus \{0\}} \frac{|\langle (\mathcal{D}t\mathcal{N} - \mathcal{D}t\mathcal{N}_N) u, v \rangle_\Gamma|}{\|v\|_{V_+}}
$$

$$
= \sup_{v \in V_+ \setminus \{0\}} \frac{c_\Psi}{\|v\|_{V_+}} \left| \sum_{\ell=0}^{\infty} \overline{\langle v, w_\ell \rangle_\Gamma} (1 + |\lambda_\ell|)^{1/2} (dtn(\lambda_\ell) - \mathrm{dtn}_N(\lambda_\ell)) \langle u, w_\ell \rangle_\Gamma (1 + |\lambda_\ell|)^{-1/2} \right|
$$

$$
\le \sup_{v \in V_+ \setminus \{0\}} \frac{c_\Psi}{\|v\|_{V_+}} \sqrt{\sum_{\ell=0}^{\infty} (1 + |\lambda_\ell|) |\langle v, w_\ell \rangle_\Gamma|^2} \sqrt{\sum_{\ell=0}^{\infty} \mathrm{err}_{dtn}(N, \ell)^2 (1 + |\lambda_\ell|)^{-1} |\langle u, w_\ell \rangle_\Gamma|^2}
$$

$$
\le c_\Psi \sup_{v \in V_+ \setminus \{0\}} \frac{\|v\|_{V_+}}{\|v\|_{V_+}} \|u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)} \le c_\Psi \left(\|\Pi_L u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)} + \|(\mathrm{Id} - \Pi_L) u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)} \right),
$$

where the last inequality follows from (3.17).

□

To bound the inf-sup constants γ_N from below uniformly in N, information on dtn_N is required. For a specific classical infinite elements for the Helmholtz equation with a spherical coupling boundary this has been achieved in reference [DI01]. In Section 3.3 we will seek to extend the ideas of this reference to cover the more general setting considered in this thesis. Ideally, the analysis should yield an error estimate of the form

$$
\|u - u_N\|_{V_+} \le \chi(N),
$$

where $\chi(N)$ converges rapidly to zero as N increases, which is valid for any tensor-product discretization of the form (3.8) for separable exterior domains introduced in Chapter 2. According to Proposition 3.5 (ii), the error can be split into the two contributions $\|\Pi_L u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)}$ and $\|(\mathrm{Id} - \Pi_L) u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)}$ for some freely chosen $L \in \mathbb{N}_0$. For sufficiently smooth solutions the coefficients $|\langle u, w_\ell \rangle_\Gamma|$ in the eigenbasis decay rapidly as will be seen in Lemma 3.19. This characterizes $\|(\mathrm{Id} - \Pi_L) u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)}$ as an asymptotic contribution that can made arbitrarily small for L large enough. It then remains to estimate $\|\Pi_L u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)}$. This will require an accurate approximation of $dtn(\lambda_\ell)$ by $\mathrm{dtn}_N(\lambda_\ell)$ at the first $\ell < L$ eigenvalues λ_ℓ. Given the complexity of the dtn functions introduced in Chapter 2, see e.g. $dtn^{\mathrm{VAL-C}}$ shown in Fig. 2.8, it is not clear at all whether convergence can be achieved with the ansatz (3.10) for dtn_N. Giving a positive answer to this question under reasonable assumptions on dtn and characterizing the rate of convergence is subject of the following section.

3.2 Approximation properties of dtn_N on finite intervals

Proposition 3.5 (ii) shows that the error $\|u - u_N\|_{V_+}$ between the exact solution u and the approximate solution u_N obtained with a transparent boundary condition of tensor-product type is determined by the approximation of *dtn* by dtn_N on the spectrum of a self-adjoint operator on the truncation boundary. In view of equation (3.10), this is a rational approximation problem. This section investigates whether dtn_N converges to *dtn* if the variables in the ansatz (3.8) are chosen appropriately. If this were not the case, then tackling the arduous task of bounding the inf-sup constants γ_N, which appear in Proposition 3.5, would be in vain. Fortunately, convergence holds true, yet only on finite intervals, that is at a finite number of eigenvalues λ_ℓ. More cannot be expected since typical *dtn* functions behave as $\mathcal{O}(\sqrt{\lambda})$ at infinity (see e.g. Lemma 2.3 (iii)), which cannot be reproduced with the ansatz (3.10).

The convergence analysis on finite intervals presented in this section proceeds as follows. Firstly, we note that the approximation space spanned by the ansatz (3.10) contains in particular rational functions with simple poles, which are known as simple partial fractions (SPFs). Section 3.2.1 recalls a theorem from the theory of SPFs which establishes an equivalence between the approximation error of a function on a compact set using SPFs and the approximation error of a transformed function using polynomials. The proof of this theorem is constructive and allows to reduce the rational to a polynomial approximation problem. To obtain fast convergence rates with polynomial approximation, it is useful to recall from Chapter 2 that *dtn* is a meromorphic function. Section 3.2.2 cites convergence results for such functions from the literature. These results are applied in Section 3.2.3 to obtain rates of convergence for the approximation of *dtn* by dtn_N. Some of the abstract theoretical concepts are illustrated in Section 3.2.4 by means of a concrete numerical example.

3.2.1 Reduction to polynomial approximation problem

By choosing $A_{\Gamma\Gamma} = B_{\Gamma\Gamma} = 0$ and A_{EE} and B_{EE} as diagonal matrices it follows that the ansatz (3.10) includes in particular functions of the form

$$r_N(\lambda) = \sum_{j=1}^{N} \frac{1}{\lambda - a_j}, \quad \{a_j\} \subset \mathbb{C}. \tag{3.21}$$

In the literature these functions are known as simple partial fractions (SPFs). Fortunately, various deep results on their approximation properties have already been established, see reference [DKC18] for a discussion of known results and recent developments. For the purpose of this thesis, Kosukhin's result [Kos01] on the weak-equivalence between uniform approximation with SPFs and polynomials is of major importance. To state this relation, some preparatory definitions are necessary.

Definition 3.6. *Let $K(b, D)$ denote compact subsets of $\mathbb{C}$ such that there exists for any $z \in K(b, D)$ a path contained in K starting at b and ending at z of finite length $\leq D$. This means that $K \in K(b, D)$ is a compact rectifiable set.*

Definition 3.7. *Denote by $AC(K)$ the set of functions that are analytic at all interior points of K and continuous on the closure.*

Antiderivatives of analytic functions defined on compact, rectifiable sets are one of the main ingredients for establishing the results of reference [Kos01].

Definition 3.8. *For $f \in AC(K)$ with $K \in K(b, D)$ let*

$$\theta(f; b, z) := \int_b^z f(t) \; dt \tag{3.22}$$

be an antiderivative for f taken along a path contained in K with length smaller than D.

Furthermore, denote by SR_N the set of SPFs with degree at most N and the best uniform approximation on K by

$$\rho_N(f, K) := \inf\{\|f - r_N\|_{K,\infty} \mid r_N \in SR_N\},$$

where $\|f - r_N\|_{K,\infty} := \sup_{x \in K} |f(x) - r_N(x)|$. Denote the same quantity for polynomials $\mathbb{P}_{\mathbb{C}}^N$ of degree at most N by

$$e_N(f, K) := \inf\{\|f - p_N\|_{K,\infty} \mid p_N \in \mathbb{P}_{\mathbb{C}}^N\}.$$

With these definitions at hand the weak equivalence result of Kosukhin [Kos01] can be stated.

Theorem 3.9. *Let $K \in K(b, D)$, $f \in AC(K)$ and $\theta(z) = \theta(f; b, z)$. Then*

$$C_1(D\|f\|_{K,\infty})\rho_{N+1}(f, K) \leq e_N(fe^\theta, K) \leq C_2(D\|f\|_{K,\infty})\rho_{N+1}(f, K), \tag{3.23}$$

with $C_1(r) = 1/(2(1 + r)e^r)$ and $C_2(r) = (1 + 2re^r)e^r$ for $N = 1, 2, \ldots$ holds true.

Some remarks about this theorem are given below.

- The proof of Theorem 3.9 is in fact constructive. If the polynomial p_N is a good approximation of fe^θ, then

$$r_{N+1} := \log(q_{N+1})' = \frac{q'_{N+1}}{q_{N+1}}, \quad \text{with } q_{N+1}(z) = 1 + \int_b^z p_N(t) \; dt, \tag{3.24}$$

  is the corresponding approximation for f. Note that $\log(q_{N+1})'$ is of the form (3.21). Further details on the transformation between polynomials and SPFs are provided in Section 3.2.4.

- The constants $C_1(r)$ and $C_2(r)$ depend exponentially on the supremum norm of the function and the parameter D, which is essentially determined by the diameter of K. When applying this theorem to approximation of *dtn*, this will result in an exponential dependence of the approximation error on the length of the interval. This has to be countered by sufficiently fast convergence rates in N. Establishing such rates is subject of the next section.

3.2.2 Polynomial approximation of holomorphic functions

The rate of poynomial approximation of an analytic function on a bounded subset of the real line is closely related to how far this function can be extended holomorphically into certain ellipses in the complex plane. A detailed discussion on polynomial approximation in the complex domain can be found in the monograph [Wal60, §4] by Walsh. For our purpose, the following result established in [Kre14, Theorem 11.13] is a suitable starting point.

Theorem 3.10. *Let $\hat{g} : [-1, 1] \to \mathbb{R}$ be analytic. Then there exists an ellipse $\hat{E}$ with foci $(\pm 1, 0) \in \mathbb{R}^2$ such that $\hat{g}$ can be extended to a holomorphic and bounded function $\hat{g} : \hat{D} \to \mathbb{C}$, where $\hat{D}$ denotes the open interior of $\hat{E}$. Moreover, there exists a sequence of polynomials $\hat{p}_n : [-1, 1] \to \mathbb{R}$ such that*

$$\|\hat{g} - \hat{p}_n\|_{[-1,1],\infty} \le \frac{2\|\hat{g}\|_{\hat{D},\infty}}{R - 1} R^{-n}. \tag{3.25}$$

Here, R is given through the semi-axis $\hat{a}_E$ and $\hat{b}_E$ of $\hat{E}$ by $R = \hat{a}_E + \hat{b}_E$.

Remark 3.11. *(a) Theorem 3.10 extends to complex-valued $\hat{g} : [-1, 1] \to \mathbb{C}$ by splitting $\hat{g}$ into real and imaginary parts.*

(b) The polynomials $\hat{p}_n$ can be given explicitly by orthonormal expansion of $\hat{g}$ with respect to Tschebyscheff polynomials

$$T_n(z) := \cos(n \arccos(z)), \quad -1 \le z \le 1, \ n = 0, 1, 2, \dots$$

That is,

$$p_n(z) = \frac{c_0}{2} T_0(z) + \sum_{m=1}^{n} c_m T_m(z), \ \ for \ \ c_m = \frac{2}{\pi} \int_{-1}^{1} \frac{\hat{g}(\hat{x}) T_n(\hat{x})}{\sqrt{1 - \hat{x}^2}} \, \mathrm{d}\hat{x}.$$

We would like to use this theorem for approximating a transformed version of dtn (see Section 3.2.1) on a bounded interval $K = [\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}$. To keep the formulas simple, it is covenient to restrict to the case where the spectrum is positive[2], i.e. $\lambda_{\min} = 0$ and $\lambda_{\max} = \lambda > 0$. Firstly, a suitable ellipse into which dtn extends holomorphically has to be constructed. In view of the fact that some dtn functions, e.g. $\mathit{dtn}^{\text{VAL-C}}$, admit poles whose imaginary parts are tiny, such an ellipse would need to be extremely flat. As Theorem 3.10 shows, this is detrimental for the approximation rate. To avoid this problem, polynomial approximation will actually be applied to a regularized version of dtn, denoted later by $\mathit{dtn}_{N_p}^{\text{reg}}$, whose poles are confined to the sector

$$S_\eta := \{z \in \mathbb{C} \mid \operatorname{Re} z < 0, \operatorname{Im} z > \eta\}, \tag{3.26}$$

for some $\eta > 0$. This is made possible by the observation[3] that typical dtn functions only possess a finite number $N_p \in \mathbb{N}_0$ of simple poles located in the first quadrant of the plane, which can be captured perfectly by the rational approximant dtn_{N_p}. In the construction of a suitable ellipse $E(\lambda, \eta)$ in whose open interior $D(\lambda, \eta)$ the function $\mathit{dtn}_{N_p}^{\text{reg}}$ is holomorphic, we just have to ensure that

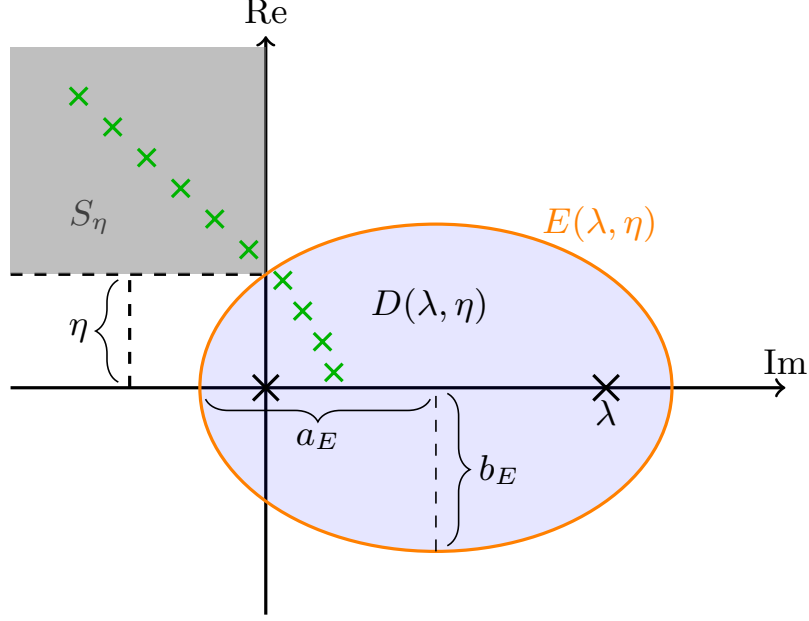$$D(\lambda, \eta) \subset \mathbb{C} \setminus S_\eta. \tag{3.27}$$

Figure 3.2: Sketch of the geometrical setup for Proposition 3.12 and Proposition 3.13. The green crosses illustrate the poles of a typical *dtn* function.

The geometrical setup we aim for is sketched in Fig. 3.2. Analytically, it can be achieved as follows. Recall that an ellipse can be described as the set of all points such that the distance to the focal points, which we choose as $(0,0)$ and $(\lambda, 0)$, is equal to a fixed value $2a_E$:

$$E(\lambda, \eta) = \{(x, y) \in \mathbb{R}^2 \mid \sqrt{(x-\lambda)^2 + y^2} + \sqrt{x^2 + y^2} = 2a_E\}, \qquad (3.28)$$

for $a_E = a_E(\lambda, \eta)$ determined as follows. The requirement (3.27) is fulfilled provided that $(0, \eta) \in E(\lambda, \eta)$, that is if

$$\sqrt{\lambda^2 + \eta^2} + \eta = 2a_E. \qquad (3.29)$$

Since the ellipse is centered at $(\lambda/2, 0)$, its covertices are given by $(\lambda/2, \pm b_E)$, where b_E is the semi-minor axis. Plugging $(x, y) = (\lambda/2, b_E)$ into (3.28) yields

$$2\sqrt{\left(\frac{\lambda}{2}\right)^2 + (b_E)^2} = 2a_E \qquad (3.30)$$

Squaring this equality and inserting (3.29) leads to

$$(b_E)^2 = (a_E)^2 - \frac{\lambda^2}{4} = \frac{1}{4}\left(\eta + \sqrt{\lambda^2 + \eta^2}\right)^2 - \frac{\lambda^2}{4}$$
$$= \frac{1}{2}\left(\eta^2 + \eta\sqrt{\lambda^2 + \eta^2}\right).$$

It follows that

$$2b_E = \sqrt{2}\sqrt{\eta^2 + \eta\sqrt{\lambda^2 + \eta^2}}. \qquad (3.31)$$

Now everything is in place to derive the required approximation result from Theorem 3.10.

---

[2]This excludes only the case of elliptical truncation boundaries.

[3]See the plots of the pole structure of *dtn* functions shown in Chapter 2. Moreover, in Section 3.3.3 an analytic argument for *dtn*$^{\text{hom,2d}}$ will be given.

Proposition 3.12 (Polynomial approximation of holomorphic functions)**.** *For $\lambda > 0$ and $\eta > 0$ let the ellipse $E(\lambda, \eta)$ be defined as in (3.28) and (3.29), and let $D(\lambda, \eta)$ denote the open interior of $E(\lambda, \eta)$. Suppose that $g : [0, \lambda] \to \mathbb{C}$ is an analytic function that extends to a holomorphic and bounded function on $D(\lambda, \eta)$. Then there exists a sequence of polynomials $p_n : [0, \lambda] \to \mathbb{C}$ such that*

$$\|g - p_n\|_{[0,\lambda],\infty} \leq \frac{2\|g\|_{D(\lambda,\eta),\infty}}{R - 1} R^{-n}, \text{ for some } R(\lambda, \eta) \geq 1 + \frac{\eta}{\lambda}.$$

Proof. Consider the affine linear transformation

$$\mathbb{R}^2 \to \mathbb{R}^2, \quad (\hat{x}, \hat{y}) \mapsto \left(\frac{\lambda}{2}(\hat{x} + 1), \frac{\lambda}{2}\hat{y} \right) := (x, y),$$

with inverse given by $(x, y) \mapsto (\frac{2}{\lambda}x - 1, \frac{2}{\lambda}y) = (\hat{x}, \hat{y})$. Since

$$\sqrt{(x - \lambda)^2 + y^2} + \sqrt{x^2 + y^2} = 2a_E$$

$$\Leftrightarrow \sqrt{\left(\frac{2x}{\lambda} - 1 - 1 \right)^2 + \left(\frac{2y}{\lambda} \right)^2} + \sqrt{\left(\frac{2x}{\lambda} - 1 + 1 \right)^2 + \left(\frac{2y}{\lambda} \right)^2} = 2a_E \left(\frac{2}{\lambda} \right)$$

$$\Leftrightarrow \sqrt{(\hat{x} - 1)^2 + \hat{y}^2} + \sqrt{(\hat{x} + 1)^2 + \hat{y}^2} = 2a_E \left(\frac{2}{\lambda} \right),$$

the ellipse $E(\lambda, \eta)$ given in coordinates (x, y) transforms in the coordinates $(\hat{x}, \hat{y})$ to an ellipse $\hat{E}$ with foci $(\pm 1, 0)$ and semi-major axis $\hat{a}_E = 2a_E/\lambda$. By using (3.30) the semi-minor axis $\hat{b}_E$ can be characterized as

$$\left(\hat{b}_E \right)^2 = (\hat{a}_E)^2 - 1^2 = 4 \left(\frac{a_E}{\lambda} \right)^2 - 1 = \frac{4}{\lambda^2} \left(b_E^2 + \frac{\lambda^2}{4} \right) - 1 = \left(\frac{2b_E}{\lambda} \right)^2.$$

That is, $\hat{b}_E = 2b_E/\lambda$. Now define

$$\hat{g}(\hat{x} + i\hat{y}) := g \left(\frac{\lambda}{2}(\hat{x} + 1) + i\frac{\lambda}{2}\hat{y} \right) = g(x + iy).$$

By construction, $\hat{g}$ is holomorphic in the ellipse $\hat{E}$. From Theorem 3.10 we obtain the existence of polynomials $\hat{p}_n : [-1, 1] \to \mathbb{R}$ fulfilling the approximation estimate (3.25) for $R = 2(a_E + b_E)/\lambda$. Clearly, $\|\hat{g}\|_{\hat{D},\infty} = \|g\|_{D(\lambda,\eta),\infty}$ and the polynomials $p_n : [0, \lambda] \to \mathbb{C}$ defined by

$$p_n(x) := \hat{p}_n \left(\frac{2}{\lambda}x - 1 \right) = \hat{p}_n(\hat{x})$$

deliver the approximation estimate

$$\|g - p_n\|_{[0,\lambda],\infty} = \|\hat{g} - \hat{p}_n\|_{[-1,1],\infty} \leq \frac{2\|g\|_{D(\lambda,\eta),\infty}}{R - 1} R^{-n}.$$

Using (3.29) and (3.31) allows to express R in terms of λ and η as

$$R = \frac{2(a_E + b_E)}{\lambda} = \frac{1}{\lambda} \left(\eta + \sqrt{\lambda^2 + \eta^2} + \sqrt{2}\sqrt{\eta^2 + \eta\sqrt{\lambda^2 + \eta^2}} \right)$$

$$\geq \frac{\eta}{\lambda} + \sqrt{1 + \left(\frac{\eta}{\lambda} \right)^2} \geq \frac{\eta}{\lambda} + 1.$$

$\square$

3.2.3 Application to approximation of *dtn* by dtn_N

To obtain convergence rates for approximation of *dtn* by dtn_N, it remains to recall from Chapter 2 that *dtn* is a meromorphic function and combine the results of Section 3.2.1 and Section 3.2.2.

Proposition 3.13. *For $\lambda > 0$ let $K = [0, \lambda]$. Suppose that $dtn : \mathbb{C} \to \mathbb{C}$ is a meromorphic function such that only a finite number $N_p \in \mathbb{N}_0$ of its poles, say λ_{p}^n for $n = 1, 2, \ldots, N_p$, are contained in the complement $\mathbb{C} \setminus S_\eta$ for S_η as defined in (3.26) and some $\eta > 0$, and that these poles are simple. Let $\mathrm{Res}(dtn, \lambda_{\mathrm{p}}^n)$ denote the residue of dtn at λ_{p}^n for $n = 1, 2, \ldots, N_p$ and define*

$$dtn_{N_p}^{\mathrm{reg}}(\lambda) := dtn(\lambda) - \sum_{n=1}^{N_p} \frac{\mathrm{Res}(dtn, \lambda_{\mathrm{p}}^n)}{\lambda - \lambda_{\mathrm{p}}^n}. \tag{3.32}$$

Then the dtn_{N+1+N_p} function (3.10) associated with a tensor-product discretization can approximate dtn on K at least with the rate

$$\| dtn - \mathrm{dtn}_{N+1+N_p} \|_{K,\infty} \le \frac{4\lambda}{\eta} \frac{\left(1 + \lambda \| dtn_{N_p}^{\mathrm{reg}} \|_{K,\infty}\right)}{R^N} e^{\lambda \| dtn_{N_p}^{\mathrm{reg}} \|_{K,\infty}} \| dtn_{N_p}^{\mathrm{reg}} e^{\theta(dtn_{N_p}^{\mathrm{reg}})} \|_{D(\lambda,\eta),\infty}, \tag{3.33}$$

where $\theta(dtn_{N_p}^{\mathrm{reg}})$ is an antiderivative for $dtn_{N_p}^{\mathrm{reg}}$ as in (3.22) and $R = R(\lambda, \eta) \ge 1 + \frac{\eta}{\lambda}$. Moreover, under the assumption that $(\mathrm{dtn}_N\text{-III})$ holds on the entire positive real axis (so not only at eigenvalues), the following growth bound is valid:

$$\| dtn_{N_p}^{\mathrm{reg}} \|_{K,\infty} \le (C_{\mathrm{III}} + C_{N_p}^{\mathrm{res}})(1 + |\lambda|)^{1/2}, \quad \text{for} \quad C_{N_p}^{\mathrm{res}} := \sum_{n=1}^{N_p} \frac{|\mathrm{Res}(dtn, \lambda_{\mathrm{p}}^n)|}{|\mathrm{Im}\,\lambda_{\mathrm{p}}^n|}. \tag{3.34}$$

Proof. We can[4] choose $A, B \in \mathbb{C}^{(N+2+N_p) \times (N+2+N_p)}$ such that

$$\mathrm{dtn}_{N+1+N_p}(\lambda) = \sum_{n=1}^{N_p} \frac{\mathrm{Res}(dtn, \lambda_{\mathrm{p}}^n)}{\lambda - \lambda_{\mathrm{p}}^n} + r_{N+1}(\lambda),$$

where r_{N+1} is any SPF as in (3.21). It follows that

$$\| dtn - \mathrm{dtn}_{N+1+N_p} \|_{K,\infty} = \| dtn_{N_p}^{\mathrm{reg}} - r_{N+1} \|_{K,\infty} \le \rho_{N+1}(dtn_{N_p}^{\mathrm{reg}}, K)$$

since r_{N+1} was arbitrary. An application of Theorem 3.9 then allows to trade rational for polynomial approximation, i.e.

$$\rho_{N+1}(dtn_{N_p}^{\mathrm{reg}}, K) \le 2(1 + \lambda \| dtn_{N_p}^{\mathrm{reg}} \|_{K,\infty}) e^{\lambda \| dtn_{N_p}^{\mathrm{reg}} \|_{K,\infty}} e_N\left(dtn_{N_p}^{\mathrm{reg}} e^{\theta(dtn_{N_p}^{\mathrm{reg}})}, K\right). \tag{3.35}$$

By construction of the ellipses $E(\lambda, \eta)$, the regularized $dtn_{N_p}^{\mathrm{reg}}$ function is holomorphic in the open interior $D(\lambda, \eta)$ of $E(\lambda, \eta)$. Hence, Proposition 3.12 yields

$$e_N\left(dtn_{N_p}^{\mathrm{reg}} e^{\theta(dtn_{N_p}^{\mathrm{reg}})}, K\right) \le \frac{2 \| dtn_{N_p}^{\mathrm{reg}} e^{\theta(dtn_{N_p}^{\mathrm{reg}})} \|_{D(\lambda,\eta),\infty}}{R - 1} R^{-N}, \quad \text{for } R \ge 1 + \frac{\eta}{\lambda}. \tag{3.36}$$

Combining these estimates and using $(R-1)^{-1} \le \lambda/\eta$ yields (3.33). The bound (3.34) follows from the triangle inequality and $(\mathrm{dtn}_N\text{-III})$ extended to $\mathbb{R}_+$. $\qquad\square$

---

[4] Set $A_{EE} = -\mathrm{diag}(\lambda_{\mathrm{p}}^1, \ldots, \lambda_{\mathrm{p}}^{N_p}, a_1, \ldots, a_{N+1})$ as diagonal, $B_{EE} = \mathrm{Id}$, $A_{\Gamma E} = -(\mathrm{Res}(dtn, \lambda_{\mathrm{p}}^1), \ldots, \mathrm{Res}(dtn, \lambda_{\mathrm{p}}^{N_p}), 1, \ldots, 1)$ and the remaining entries to zero.

3.2.4 Numerical illustration

This section illustrates the abstract concepts of the previous sections using concrete examples. Let f denote the function to be approximated by means of SPFs on the bounded interval $K = [a, b]$. Firstly, the simplest case in which the best approximation of f by SPFs and the corresponding approximation of the transformed function by polynomials can be determined analytically is considered. Then, as a more challenging example, a prototypical *dtn* function is constructed, which has to be approximated numerically. Finally, some conclusions about the implications of these experiments for dealing with actual *dtn* functions from Chapter 2 are drawn.

Transformation between SPFs and polynomials

If the function to be approximated is an SPF, i.e.

$$f = \sum_{j=1}^{N+1} \frac{1}{z - a_j}, \quad a_j \subset \mathbb{C},$$

then clearly $\rho_{N+1}(f, K) = 0$. It is slightly less obvious that the transformed function fe^θ can then also be approximated perfectly with polynomials, i.e. $e_N(fe^\theta, K) = 0$, as implied by Theorem 3.9. To see this, note that $F' = f$ for

$$F(z) := \log(\prod_{k=1}^{N+1} (z - a_k))$$

holds. Therefore, $\theta = F(z) - F(a)$ and

$$fe^\theta = f(z)e^{(F(z) - F(a))} = e^{-F(a)} \sum_{j=1}^{N+1} \prod_{k \neq j} (z - a_k).$$

Hence, the transformation turns SPFs of degree $N + 1$ into polynomials of degree N. In particular, $p_N = fe^\theta$ is a polynomial and $e_N(fe^\theta, K) = 0$.

As mentioned in Section 3.2.1, it is possible to reconstruct the SPF that best approximates the original function from p_N, by applying equation (3.24). Using

$$\left[\prod_{k=1}^{N+1} (z - a_k)\right]' = \sum_{j=1}^{N+1} \prod_{k \neq j} (z - a_k)$$

yields that $q_{N+1} = e^{-F(a)} \prod_{k=1}^{N+1} (z - a_k)$. Thus,

$$r_{N+1} = \frac{q'_{N+1}}{q_{N+1}} = \frac{\sum_{j=1}^{N+1} \prod_{k \neq j} (z - a_k)}{\prod_{k=1}^{N+1} (z - a_k)} = f,$$

so that the backtransformation consistently recovers the original SPF.

In the general case, f is neither an SPF nor is the transformed function a polynomial. However, the transformation formula (3.24) to construct the best approximation with SPFs given a polynomial of best approximation for the transformed function remains valid. This is illustrated in a numerical experiment presented in the next section.

Approximation of a prototypical *dtn* function

Now the setting in which the function to be approximated is not an SPF anymore is treated. To this end, let $K = [0, 1]$ and consider the following function

$$f(z) = \varepsilon(2\pi i \exp(2\pi i z)) + (3/2)(1 + z)^{1/2} + \sum_{j=1}^{3} \frac{1}{z - a_j},$$

which mimicks the behavior of a typical *dtn* function. It consists of a number of simple poles at $z = a_j$, $j = \{1, 2, 3\}$, a term which grows as $\sqrt{z}$ and a perturbation for which $\varepsilon = 1/10$ will be used. Fig. 3.3b shows a plot of $\operatorname{Re} f$, which resembles the behavior of dtn^{jump} displayed in Fig. 2.4. The real part of the transformed function fe^θ is shown in Fig. 3.3a. Already on the unit interval it can be seen that the magnitude of the transformed function grows noticeably with increasing length of the interval due to the exponential involved in the transformation. Since the antiderivative of f behaves asymptotically as $\mathcal{O}(z^{3/2})$, it is not practical to consider significantly larger intervals as would be required to treat actual *dtn* functions arising in the applications of Chapter 2.
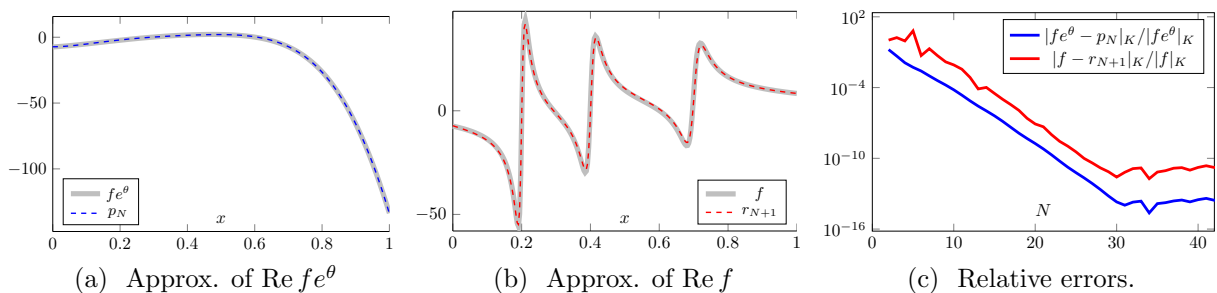

(a) Approx. of $\operatorname{Re} fe^\theta$ (b) Approx. of $\operatorname{Re} f$ (c) Relative errors.

Figure 3.3: Approximation problem with SPFs for a prototypical *dtn* function on the unit interval K. Figs. (a) and (b) show the approximation results for $N = 6$.

A polynomial p_N of degree N which approximates the transformed function has been determined by using interpolation at Tschebyscheff nodes. Already the reconstruction with $N = 6$ is very accurate as shown in Fig. 3.3a. The corresponding SPF r_{N+1} determined by formula (3.24), which approximates the original function, is shown in Fig. 3.3b. The general behavior of f is also already well-captured. The development of the relative errors for the original and transformed function in terms of N is compared in Fig. 3.3c. In agreement with the theoretical results an exponentially fast convergence in terms of N is observed for both quantities. Moreover, it can be seen that the relative error for approximation of the original function is about two orders of magnitude larger than the one for the transformed function. Additional numerical experiments not shown here suggest that the ratio of the relative errors also appears to increase with the length of the interval. To investigate the sharpness of Theorem 3.9, it would be interesting to check whether an exponential growth in terms of the supremum norm of f and the length of the interval can in fact be observed.

Conclusions

As already touched on, the above procedure is not well-suited for application to realistic *dtn* functions. Some of the obstructions are summarized below.

- Since $\mathrm{Re}\,\textit{dtn}(\lambda)$ grows like $\mathcal{O}(\sqrt{\lambda})$ and $\textit{dtn}$ has to be approximated on large intervals, the values of the exponential in the transformed function quickly exceed all reasonable bounds. Perhaps one could try to remove the asymptotic behavior from $\textit{dtn}$ in a preprocessing step although it is unclear how this can be realized in practice.

- According to Proposition 3.5 (ii), an approximation of $\textit{dtn}$ in the supremum norm at all points of K is not required and may be a waste of computational effort. Only the accuracy of the approximation at the eigenvalues λ_ℓ counts. Unfortunately, these points are typically unsuited for polynomial approximation.

- SPFs contain no information about the strength of the poles. Every pole is weighted equally, which might be suboptimal. The ansatz (3.10) is more general and may allow for faster convergence rates.

Hence, although Proposition 3.13 is theoretically highly valuable because it shows exponential convergence on finite intervals, the method of its proof cannot be exploited readily for the construction of transparent boundary conditions. The main problem is posed by the recourse to polynomial approximation in the supremum norm. Trying to fit f directly at eigenvalues using rational approximation seems much more appropriate than taking the detour via polynomial approximation of fe^θ. Such an approach will be presented in Chapter 4.

3.3 Semi-discrete error analysis

This section extends the results from Section 3.1.3 to an a priori error estimate in the semi-discrete setting. Here, the main task is to bound the inf-sup constants γ_N appearing in Proposition 3.5 from below independently of N. This requires assumptions on dtn_N which are introduced in Section 3.3.1. The main technical lemma for controlling γ_N is given in Section 3.3.2. Section 3.3.3 then presents an a priori error estimate, which allows to conclude exponential convergence on finite intervals and a global convergence without an explicit rate, i.e. $\|u - u_N\|_{V_+} \to 0$, as $N \to \infty$. For the Helmholtz equation we also present further efforts to characterize this global convergence as exponentially fast.

3.3.1 Assumptions on dtn_N and resulting properties of $b_N(\cdot, \cdot)$

Before proceeding with the error analysis, it is necessary to impose some assumptions on $\textit{dtn}_N$ to ensure that the semi-discrete problem (3.13) is uniquely solvable. The first condition, required to ensure uniqueness, is the same as for $\textit{dtn}$. The second condition is stricter than for $\textit{dtn}$ because a Gårding inequality now has to hold on the space V_+ which has a stronger norm on Γ than V. The assumptions are as follows:

- For all $\ell \in \mathbb{N}_0$:

$$\mathrm{Im}\,\mathrm{dtn}_N(\lambda_\ell) < 0, \qquad\qquad (\mathrm{dtn}_N\text{-}\mathrm{I}_<)$$

$$\mathrm{Im}\,\mathrm{dtn}_N(\lambda_\ell) \leq 0. \qquad\qquad (\mathrm{dtn}_N\text{-}\mathrm{I}_\leq)$$

- There exists a finite (possibly empty) set $L_-^N \subset \mathbb{N}_0, L_+^N := \mathbb{N}_0 \setminus L_-$ and a positive constant $C_{\mathrm{II}}(N)$ such that

$$\mathrm{Re}\, \mathrm{dtn}_N(\lambda_\ell) \geq C_{\mathrm{II}}(N)(1 + |\lambda_\ell|) \text{ for } \ell \in L_+^N. \qquad (\text{dtn}_N\text{-II})$$

Further, it will be assumed that the sets $L_\pm^N$ and the constant $C_{\mathrm{II}}(N)$ can be chosen independently of N for N large enough, i.e. $L_\pm^N = L_\pm^{N_0}$ and $C_{\mathrm{II}}(N) = C_{\mathrm{II}}(N_0)$ for all $N \geq N_0$ holds true.

Let us record the announced Gårding inequality for later use.

Lemma 3.14 (Gårding inequality)**.** *Let* (dtn_N-II) *hold. Then the sesquilinear form* $b_N(\cdot, \cdot)$ *from* (3.14) *fulfills the lower bound*

$$\mathrm{Re}\, b_N(u, u) \geq C_{V_+}(N)\|u\|_{V_+}^2 - C_{L^2(\Omega_{\mathrm{int}})}\|u\|_{L^2(\Omega_{\mathrm{int}})}^2 - C_{L^2(\Gamma)}(N)\|u\|_{L^2(\Gamma)}^2, \qquad (3.37)$$

for all $u \in V_+$ *with constants*

$$C_{V_+}(N) := \min\{C_V, c_\Psi C_{\mathrm{II}}(N)\}, \qquad (3.38)$$

where C_V *and* $C_{L^2(\Omega_{\mathrm{int}})}$ *are defined as in* (2.22) *and*

$$C_{L^2(\Gamma)}(N) := c_\Psi \max_{\ell \in L_-^N} |\mathrm{Re}\, \boldsymbol{dtn}(\lambda_\ell) - C_{\mathrm{II}}(N)(1 + |\lambda_\ell|)|. \qquad (3.39)$$

Proof. By definition

$$\mathrm{Re}\langle \mathcal{M}_\Gamma \boldsymbol{Dt}\mathcal{N}_N u, u\rangle_\Gamma = c_\Psi \left(\sum_{\ell \in L_+^N} \mathrm{Re}\, \mathrm{dtn}_N(\lambda_\ell)|\langle u, w_\ell\rangle_\Gamma|^2 + \sum_{\ell \in L_-^N} C_{\mathrm{II}}(N)(1 + |\lambda_\ell|)|\langle u, w_\ell\rangle_\Gamma|^2 \right.$$
$$\left. + \sum_{\ell \in L_-^N} [\mathrm{Re}\, \mathrm{dtn}_N(\lambda_\ell) - C_{\mathrm{II}}(N)(1 + |\lambda_\ell|)] |\langle u, w_\ell\rangle_\Gamma|^2 \right).$$

For the sum running over L_+^N the assumption (dtn_N-II) can be used to obtain a lower bound. This leads to

$$\mathrm{Re}\langle \mathcal{M}_\Gamma \boldsymbol{Dt}\mathcal{N}_N u, u\rangle_\Gamma \geq c_\Psi C_{\mathrm{II}}(N)\|\, \mathrm{tr}\, u\|_{H^1(\Gamma)}^2 - C_{L^2(\Gamma)}(N)\|u\|_{L^2(\Gamma)}^2$$

for $C_{L^2(\Gamma)}(N)$ as given in the statement of the lemma. Combining this with the lower bound for $\mathrm{Re}\, b_{\mathrm{int}}(u, u)$ known from Lemma 2.2 yields the claim. $\qquad \square$

For $\boldsymbol{dtn}$ it was also necessary to introduce the boundedness assumption ($\boldsymbol{dtn}$-III) to ensure continuity of the sesquilinear form $b(\cdot, \cdot)$ on $V \times V$. Since the ansatz (3.10) grows linearly as $\lambda \to \infty$ provided that B_{EE} is nonsingular as is from now on assumed, the inequality

$$|\mathrm{dtn}_N(\lambda_\ell)| \leq C_{\mathrm{III}}(N)(1 + |\lambda_\ell|) \text{ for all } \ell \in \mathbb{N}_0, \qquad (\text{dtn}_N\text{-III})$$

for some constant $C_{\mathrm{III}}(N)$ is automatically fulfilled for dtn_N. Continuity of b_N follows.

Lemma 3.15 (Continuity of $b_N(\cdot, \cdot)$)**.** *Assume that* B_{EE} *is nonsingular so that* (dtn_N-III) *holds. Then the sesquilinear form* $b_N(\cdot, \cdot)$ *from* (3.14) *is continuous on* $V_+ \times V_+$.

Proof. Since the norm on V_+ is stronger than the norm on V, continuity of $b_{\text{int}}(\cdot, \cdot)$ on $V_+ \times V_+$ follows from Lemma 2.18. It remains to estimate

$$|\langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}_N u, v\rangle_\Gamma| = c_\Psi \left| \sum_{\ell \in \mathbb{N}_0} \text{dtn}_N(\lambda_\ell) \langle u, w_\ell\rangle_\Gamma \overline{\langle v, w_\ell\rangle_\Gamma} \right|$$

$$\leq c_\Psi C_{\text{III}}(N) \left| \sum_{\ell \in \mathbb{N}_0} (1 + |\lambda_\ell|)^{1/2} |\langle u, w_\ell\rangle_\Gamma| (1 + |\lambda_\ell|)^{1/2} |\langle v, w_\ell\rangle_\Gamma| \right|$$

$$\leq c_\Psi C_{\text{III}}(N) \| \operatorname{tr} u \|_{H^1(\Gamma)} \| \operatorname{tr} v \|_{H^1(\Gamma)},$$

by the Cauchy-Schwarz inequality. $\qquad \square$

Well-posedness of the semi-discrete problem (3.13) follows now similarly as for the continuous problem treated in Section 2.5.

Proposition 3.16 (Well-posedness of semi-discrete problem)**.** *Assume that*

(a) either (dtn_N-$\text{I}_\leq$) *and* $\operatorname{Im} c \leq c_0 < 0$ *in* Ω_{int}

(b) or (dtn_N-$\text{I}_<$)

holds. Then the semi-discrete problem (3.13) *admits at most one solution. If additionally* (dtn_N-II) *holds, then the problem is well-posed on* V_+.

Proof. Assumption (dtn_N-II) implies that the sesquilinear form

$$\tilde{b}_N(u, v) := b_N(u, v) + C_{L^2(\Omega_{\text{int}})}(u, v)_{\Omega_{\text{int}}} + C_{L^2(\Gamma)}(N)\langle u, v\rangle_\Gamma \qquad (3.40)$$

is coercive on $V_+ \times V_+$ by Lemma 3.14. Moreover, the embeddings $\iota : V_+ \hookrightarrow L^2(\Omega_{\text{int}})$ and $\iota_\Gamma : V_+ \to H^1(\Gamma) \hookrightarrow L^2(\Gamma)$ are compact. Hence, the Fredholm property of the corresponding operator $B_N \in \mathscr{L}(V_+, V_+)$, which fulfills

$$b_N(u, v) = (B_N u, v)_{V_+}, \quad u, v \in V_+,$$

follows as in Theorem 2.19. Assumptions (a) or (b) imply that B_N is injective, i.e. uniqueness, which yields well-posedness as in Corollary 2.20. $\qquad \square$

In addition to properties of $\mathbf{dtn}_N$, a certain convergence of dtn_N to $\mathbf{dtn}$ has to be assumed for the argument in Section 3.3.2. For the specific classical infinite elements considered in reference [DI01] the property $\text{dtn}_N(\lambda_\ell) = \mathbf{dtn}(\lambda_\ell)$ for $\ell = 0, \dots, N$ holds true. In order to generalize this to other transparent boundary conditions, it will be assumed that for any fixed $L \in \mathbb{N}_0$:

$$\lim_{N \to \infty} \sup_{\ell \leq L} \text{err}_{\mathbf{dtn}}(N, \ell) \to 0 \qquad (\text{CFI-w})$$

is valid. The acronym (CFI-w) means 'Convergence For Intervals - weak'. A stronger assumption, which is suggested by the results of Section 3.2, albeit not needed in Section 3.3.2, would be to require for any fixed $L \in \mathbb{N}_0$:

$$\sup_{\ell \leq L} \text{err}_{\mathbf{dtn}}(N, \ell) \leq C_L \exp(-\kappa_L N), \qquad (\text{CFI-s})$$

with constants $C_L > 0$ and $\kappa_L > 0$ which may depend on $L \in \mathbb{N}_0$, yet not on N.

3.3.2 Bounding the inf-sup constants from below

For convenience, recall the sesquilinear form associated with the semi-discrete $\mathcal{D}t\mathcal{N}_N$ operator

$$b_N(u, v) := b_{\text{int}}(u, v) + \langle \mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}_N u, v \rangle_\Gamma$$

and the inf-sup constants γ_N defined by

$$\gamma_N := \inf_{\|u\|_{V_+}=1} \sup_{\|v\|_{V_+}=1} |b_N(u, v)|.$$

The aim of this section is to show that $\inf_N \gamma_N > 0$ using the assumptions on dtn_N introduced in Section 3.3.1 and the resulting properties of $b_N(\cdot, \cdot)$. The main step is the following technical lemma.

Lemma 3.17 (Lower bound for inf-sup constants)**.** *Assume that the continuous problem* (2.59) *admits at most one solution and that the bilinear form $b(\cdot, \cdot)$ is continuous. Let* (dtn$_N$-II) *and* (CFI-w) *hold. Then* $\liminf_N \gamma_N > 0$ *follows.*

Proof. Following [DI01, Lemma 3.7], the proof proceeds by contradiction. Assume that there exists a subsequence

$$\gamma_{N_k} := \inf_{\|u\|_{V_+}=1} \sup_{\|v\|_{V_+}=1} |b_{N_k}(u, v)|$$

converging to zero. Then there is a corresponding sequence (u_k) with $\|u_k\|_{V_+} = 1$ such that

$$\sup_{\|v\|_{V_+}=1} |b_{N_k}(u_k, v)| \to 0 \text{ as } k \to \infty. \tag{3.41}$$

Since V_+ is a Hilbert space, there exists a weakly convergent subsequence again denoted by (u_k) for ease of notation. Let $u \in V_+$ with $\|u\|_{V_+} \leq 1$ denote the weak limit. Due to compactness of the embeddings $V_+ \hookrightarrow L^2(\Omega_{\text{int}})$ and $V_+ \to H^1(\Gamma) \hookrightarrow L^2(\Gamma)$, the sequence (u_k) converges strongly to u in $L^2(\Omega_{\text{int}})$ and $L^2(\Gamma)$. The argument given below proceeds in three steps to show that

$$b(u, v) = 0 \text{ for all } v \in V_+ \text{ and } u \neq 0, \tag{3.42}$$

which contradicts uniqueness[5] of the continuous problem (2.59).

Step 1: We show the following auxiliary result: Let $v \in V_+$ be a test function such that there exists $L \in \mathbb{N}_0$ with $\langle v, w_\ell \rangle_\Gamma = 0$ for all $\ell > L$, i.e. its trace on Γ can be represented as a linear combination of a finite number of eigenfunctions. This implies $b(u, v) = 0$.

Indeed,

$$|b(u, v)| \leq |b(u - u_k, v)| + |b_{N_k}(u_k, v)| + |b(u_k, v) - b_{N_k}(u_k, v)|. \tag{3.43}$$

Since (u_k) converges weakly to u the first terms goes to zero as $k \to \infty$. The second term goes to zero by (3.41). It remains to treat the last term:

$$|b(u_k, v) - b_{N_k}(u_k, v)| = c_\Psi \left| \sum_{\ell=0}^{L} [\mathbf{dtn}(\lambda_\ell) - \text{dtn}_{N_k}(\lambda_\ell)] \langle u_k, w_\ell \rangle_\Gamma \overline{\langle v, w_\ell \rangle_\Gamma} \right|$$

$$\leq c_\Psi \sup_{\ell \leq L} \text{err}_{\mathbf{dtn}}(N_k, \ell) \|u_k\|_{V_+} \|v\|_{L^2(\Gamma)} = c_\Psi \sup_{\ell \leq L} \text{err}_{\mathbf{dtn}}(N_k, \ell) \|v\|_{L^2(\Gamma)}.$$

---

[5]Uniqueness for (2.59) holds on the space V while the argument in this proof requires uniqueness on V_+. This is not a problem since if $u \in V_+$ then in particular $u \in V$. Moreover, equation (3.42) implies $b(u, v) = 0$ for all $v \in V$ by density of V_+ in V.

The supremum goes to zero as $N_k \to \infty$ by assumption (CFI-w). It follows that $b(u, v) = 0$ since the left hand side of (3.43) is independent of k while the right hand side converges to zero.

Step 2: Next it will be shown that u is a weak solution of the homogeneous problem, i.e. that the left equation of (3.42) holds true. In the final step it then remains to show that $u \neq 0$.

For an arbitrary $v \in V_+$ we have $\mathrm{tr}(v) \in H^1(\Gamma)$ so that

$$\mathrm{tr}(v) = \sum_{\ell=0}^{\infty} \langle \mathrm{tr}(v), w_\ell \rangle_\Gamma w_\ell$$

and for the partial sums

$$\mathrm{tr}(v)_L := \sum_{\ell=0}^{L} \langle \mathrm{tr}(v), w_\ell \rangle_\Gamma w_\ell$$

it holds that
$$\| \mathrm{tr}(v) - \mathrm{tr}(v)_L \|_{H^1(\Gamma)} \to 0, \text{ as } L \to \infty.$$

Recall that the trace operator admits a continuous linear right inverse

$$Z : H^{1/2}(\Gamma) \to V, \quad \mathrm{tr} \circ Z = \mathrm{Id}_{H^{1/2}(\Gamma)}.$$

For $\tilde{v} := v - Z \circ \mathrm{tr}(v) \in V$ with $\mathrm{tr}(\tilde{v}) = 0$ on Γ define $v_L := \tilde{v} + Z \circ \mathrm{tr}(v)_L \in V$. Actually, $\mathrm{tr}(v_L) = \mathrm{tr}(\tilde{v}) + \mathrm{tr} \circ Z \circ \mathrm{tr}(v)_L = \mathrm{tr}(v)_L \in H^1(\Gamma)$ since $v \in V_+$, so that $v_L \in V_+$. Then

$$|b(u, v)| \leq |b(u, v_L)| + |b(u, v - v_L)| \to 0 \text{ as } L \to \infty.$$

This is because the first terms is zero by Step 1 for $\mathrm{tr}(v_L) = \mathrm{tr}(v)_L$ admits a finite eigenfunction representation. For the second term we use continuity of $b(\cdot, \cdot)$ and continuity of Z to conclude

$$
\begin{aligned}
|b(u, v - v_L)| &\leq \|b\| \|u\|_V \|v - v_L\|_V \\
&= \|b\| \|u\|_V \|\tilde{v} + Z \circ \mathrm{tr}(v) - (\tilde{v} + Z \circ \mathrm{tr}(v)_L)\|_V \\
&= \|b\| \|u\|_V \|Z(\mathrm{tr}(v) - \mathrm{tr}(v)_L)\|_V \\
&\leq \|b\| \|Z\| \|u\|_V \|\mathrm{tr}(v) - \mathrm{tr}(v)_L\|_{H^{1/2}(\Gamma)},
\end{aligned}
$$

which goes to zero as $L \to \infty$. Hence, $b(u, v) = 0$ for all $v \in V_+$.

Step 3: To obtain a contradiction, it remains to show that u cannot be zero.

To this end, we use the Gårding inequality for $b_N(\cdot, \cdot)$ established in Lemma 3.14 and assume that N_k is already sufficiently large so that the involved constants can be chosen independently of N_k, i.e. $C_{\mathrm{II}}(N_k) = C_{\mathrm{II}}(N_0)$ and $C_{L^2(\Gamma)}(N_k) = C_{L^2(\Gamma)}(N_0)$ for N_k large holds. This yields

$$\mathrm{Re}\, b_{N_k}(u_k, u_k) \geq C_{V_+}(N_0) \|u_k\|_{V_+}^2 - C_{L^2(\Omega_{\mathrm{int}})} \|u_k\|_{L^2(\Omega_{\mathrm{int}})}^2 - C_{L^2(\Gamma)}(N_0) \|u_k\|_{L^2(\Gamma)}^2 \tag{3.44}$$

$$= C_{V_+}(N_0) - C_{L^2(\Omega_{\mathrm{int}})} \|u_k\|_{L^2(\Omega_{\mathrm{int}})}^2 - C_{L^2(\Gamma)}(N_0) \|u_k\|_{L^2(\Gamma)}^2.$$

As $\operatorname{Re} b_{N_k}(u_k, u_k)$ converges to zero by (3.41) and (u_k) converges strongly to u in $L^2(\Omega_{\text{int}})$ as well as in $L^2(\Gamma)$ by compactness of the embeddings, this yields

$$C_{L^2(\Omega_{\text{int}})}\|u\|_{L^2(\Omega_{\text{int}})}^2 + C_{L^2(\Gamma)}(N_0)\|u\|_{L^2(\Gamma)}^2 \geq C_{V_+}(N_0) > 0.$$

This implies $u \neq 0$.

$\square$

Corollary 3.18. *In addition to the assumptions of Lemma 3.17 let one of the following hold*

(i) *(dtn$_N$-$I_\leq$) and $\operatorname{Im} c \leq c_0 < 0$ in Ω_{int}*

(ii) *(dtn$_N$-$I_<$).*

Then $\inf_N \gamma_N > 0$.

Proof. According to Proposition 3.16, the assumptions imply uniqueness for the discrete problem. That is, $b_N(u, v) = 0$ for all $v \in V_+$ implies $u = 0$. Since $\liminf_N \gamma_N > 0$ holds, γ_N could only vanish for finitely many N. However, $\gamma_N \neq 0$, else there would exist $u \neq 0$ with $\sup_{\|v\|_{V_+}=1} |b_N(u, v)| = 0$, which contradicts uniqueness. $\square$

3.3.3 A priori error estimate

Now the results from the previous investigations are combined to derive an a priori error estimate. Before this can be obtained, one remaining difficulty still needs to be overcome. Section 3.2 showed that only convergence of dtn$_N$ to *dtn* on finite intervals can be expected. Therefore, the asymptotic contribution $\|(\operatorname{Id} - \Pi_L)u\|_{H^{\text{err}}_\sigma(\Gamma)}$ in the error estimate of Proposition 3.5 (ii) is not controlled yet. To resolve this issue, smoothness properties of the exact solution u to problem (2.2) close to the truncation boundary Γ have to be converted into asymptotic decay of the coefficients $|\langle u, w_\ell \rangle_\Gamma|$. The following lemma makes these notions precise.

Lemma 3.19 (Regularity on Γ). *Let u denote the solution to (2.2), i.e. $\mathcal{L}u = f$.*

(a) *Assume that the coefficients of $\mathcal{L}$ and the right hand side f are C^∞ in a neighborhood of Γ. Then $\operatorname{tr} u \in H^s(\Gamma)$ for any $s \geq 0$.*

(b) *Consider the specific case of the Helmholtz equation of a homogeneous medium exterior to a disk, see Section 2.2.1. Assume that the separability assumption (2.5) extends to an annulus $b \leq r < a$. Then*

$$|\langle u, w_\ell \rangle_\Gamma| \leq C \exp\left(-a \log(a/b)\sqrt{|\lambda_\ell|}\right),$$

where C is independent of ℓ.

Proof. (a) Since the coefficients of $\mathcal{L}$ and the right hand side are smooth in a neighborhood of Γ and $\mathcal{L}$ is strongly elliptic, the solution is smooth in a neighboorhood of Γ. Therefore, $\operatorname{tr} u = u|_\Gamma \in C^\infty(\Gamma) \subset H^s(\Gamma)$ for any $s > 0$.

68

(b) From (2.11) the solution is known to allow for an expansion of the form

$$u = \sum_{\ell=0}^{\infty} \Lambda_r(\lambda_\ell) w_\ell \langle \operatorname{tr}_\Gamma u, w_\ell \rangle_\Gamma, \quad \text{with } \Lambda_r(\lambda_\ell) = H_\ell^{(1)}(kr)/H_\ell^{(1)}(ka),$$

where the diffeomorphism Ψ has been omitted for ease of notation. Since the separability assumption extends to $b \leq r < a$, the same formula can be applied at $\Gamma_b := \{r = b\}$, which gives

$$u = C \sum_{\ell=0}^{\infty} \frac{H_\ell^{(1)}(kr)}{H_\ell^{(1)}(kb)} w_\ell \langle \operatorname{tr}_{\Gamma_b} u, w_\ell \rangle_{\Gamma_b}.$$

The constant C is independent of ℓ and simply scales the eigenfunctions w_ℓ (trigonometric polynomials) to unit length on Γ_b. By evaluating at $r = a$ and comparing coefficients it follows that

$$|\langle \operatorname{tr}_\Gamma u, w_\ell \rangle_\Gamma| = \left| C \frac{H_\ell^{(1)}(ka)}{H_\ell^{(1)}(kb)} \langle \operatorname{tr}_{\Gamma_b} u, w_\ell \rangle_{\Gamma_b} \right| \leq C \left| \frac{H_\ell^{(1)}(ka)}{H_\ell^{(1)}(kb)} \right| \|u\|_{L^2(\Gamma_b)}$$

since $\operatorname{tr}_{\Gamma_b} u \in L^2(\Gamma_b)$. From [AS64, equation (9.3.1)] the asymptotic behavior

$$|H_\nu^{(1)}(z)| \sim \sqrt{\frac{2}{\pi\nu}} \left(\frac{ez}{2\nu} \right)^{-\nu}, \quad \text{as } \nu \to +\infty$$

is known. Hence,

$$\left| \frac{H_\ell^{(1)}(ka)}{H_\ell^{(1)}(kb)} \right| \sim \exp\left(-\ell \log(a/b) \right), \quad \text{as } \ell \to +\infty.$$

As $\lambda_\ell = (\ell/a)^2$, the claim follows.

$\square$

The fundamental semi-discrete error estimate can now be derived. Based on this theorem, further results can be obtained immediately by specifying the smoothness of the exact solution in the weighted Sobolev spaces $H_\sigma(\Gamma)$ and the approximation rate $\operatorname{err}_{dtn}(N, \ell)$. The latter can be characterized using the results of Section 3.2.3.

Theorem 3.20 (Semi-discrete error estimate). *Assume that* (*dtn*-II) / (dtn$_N$-II) *and* (*dtn*-III) *hold. Further assume that*

1. *either* (*dtn*-I$_<$) / (dtn$_N$-I$_<$)

2. *or* (*dtn*-I$_\leq$) / (dtn$_N$-I$_\leq$) *and* $\operatorname{Im} c \leq c_0 < 0$ *in* Ω_{int} *is fulfilled.*

Let the weak approximation condition (CFI-w) *hold. Let u be the solution of* (2.59) *and assume that $u \in V_+ \cap H_\sigma(\Gamma)$ for some sequence (σ_ℓ) of positive weights with $\inf_\ell \sigma_\ell > 0$. Let u_N be the solution of* (3.13) *and $L \in \mathbb{N}_0$. Moreover, let the eigenvalues be ordered in non-decreasing magnitude, i.e. $|\lambda_\ell| \leq |\lambda_{\ell+1}|$ holds true for all $\ell \in \mathbb{N}_0$. Then*

$$\|u - u_N\|_{V_+} \leq C' \left(\|\Pi_L u\|_{H_{\sigma^{\text{err}}}(\Gamma)} + (C_{\text{III}} + C_{\text{III}}(N)) \sup_{\ell \geq L} \frac{(1 + |\lambda_\ell|)^{1/2}}{\sigma_\ell} \|u\|_{H_\sigma(\Gamma)} \right)$$

$$\leq C' \left(\sup_{\ell < L} \frac{\operatorname{err}_{dtn}(N, \ell)}{\sigma_\ell (1 + |\lambda_\ell|)^{1/2}} + (C_{\text{III}} + C_{\text{III}}(N)) \sup_{\ell \geq L} \frac{(1 + |\lambda_\ell|)^{1/2}}{\sigma_\ell} \right) \|u\|_{H_\sigma(\Gamma)}$$

(3.45)

with $\|\Pi_L u\|_{H_{\sigma^{\text{err}}}(\Gamma)}$ as defined in Proposition 3.5 (ii) and $C' := \sqrt{2} \inf_N \gamma_N^{-1} c_\Psi < \infty$.

Proof. According to Proposition 3.5 (ii), the error can be bounded by

$$\|u - u_N\|_{V_+} \leq C' \left(\|\Pi_L u\|_{H_{\sigma\mathrm{err}}(\Gamma)} + \|(\mathrm{Id} - \Pi_L)u\|_{H_{\sigma\mathrm{err}}(\Gamma)} \right),$$

with $C' := \inf_N \gamma_N^{-1} c_\Psi < \infty$ thanks to Corollary 3.18. The asymptotic contribution, i.e. the second term, needs to be bounded by using growth conditions on *dtn*, respectively dtn_N, at infinity and smoothness assumptions on u. The assumption (*dtn*-III) and the automatically fulfilled condition (dtn_N-III) yield

$$\begin{aligned}
\|(\mathrm{Id} - \Pi_L)u\|^2_{H_{\sigma\mathrm{err}}(\Gamma)} &= \sum_{\ell=L}^{\infty} \mathrm{err}_{dtn}(N, \ell)^2 (1 + |\lambda_\ell|)^{-1} |\langle u, w_\ell \rangle_\Gamma|^2 \\
&\leq 2 \left(C_{\mathrm{III}}^2 + C_{\mathrm{III}}(N)^2 \right) \sum_{\ell=L}^{\infty} \left(\frac{(1 + |\lambda_\ell|)^{1/2}}{\sigma_\ell} \right)^2 \sigma_\ell^2 |\langle u, w_\ell \rangle_\Gamma|^2 \\
&\leq 2 \left(C_{\mathrm{III}}^2 + C_{\mathrm{III}}(N)^2 \right) \left(\sup_{\ell \geq L} \frac{(1 + |\lambda_\ell|)^{1/2}}{\sigma_\ell} \right)^2 \|u\|^2_{H_\sigma(\Gamma)}.
\end{aligned}$$

On the first $L - 1$ modes

$$\begin{aligned}
\|\Pi_L u\|^2_{H_{\sigma\mathrm{err}}(\Gamma)} &= \sum_{\ell=0}^{L-1} \mathrm{err}_{dtn}(N, \ell)^2 \sigma_\ell^2 |\langle u, w_\ell \rangle_\Gamma|^2 \frac{1}{\sigma_\ell^2 (1 + |\lambda_\ell|)} \\
&\leq \left(\sup_{\ell < L} \frac{\mathrm{err}_{dtn}(N, \ell)}{\sigma_\ell (1 + |\lambda_\ell|)^{1/2}} \right)^2 \|u\|^2_{H_\sigma(\Gamma)}.
\end{aligned}$$

Combining the two contributions yields the claim. $\qquad\square$

Refined estimates now readily follow by specifying the degree of smoothness of the exact solution on Γ. While Corollary 3.21 (a) only requires smoothness in terms of Sobolev scales, part (b) makes further requirements which yield a stronger result.

Corollary 3.21. *Let the assumptions of Theorem 3.20 be fulfilled.*

(a) If $u \in H^s(\Gamma)$ for some $s \geq 1$, which e.g. follows if the assumptions of Lemma 3.19 (a) are fulfilled, then

$$\|u - u_N\|_{V_+} \leq C' \left(\sup_{\ell < L} \frac{\mathrm{err}_{dtn}(N, \ell)}{(1 + |\lambda_\ell|)^{\frac{s+1}{2}}} + (C_{\mathrm{III}} + C_{\mathrm{III}}(N)) \frac{1}{(1 + |\lambda_L|)^{\frac{s-1}{2}}} \right) \|u\|_{H^s(\Gamma)},$$

with C' from Theorem 3.20. Further, there exist constants $C_L > 0, \kappa_L > 0$ depending on L and a fixed $N_p \in \mathbb{N}_0$ such that

$$\sup_{\ell < L} \frac{\mathrm{err}_{dtn}(N + 1 + N_p, \ell)}{(1 + |\lambda_\ell|)^{\frac{s+1}{2}}} \leq C_L \exp(-\kappa_L N), \tag{3.46}$$

provided that $A, B \in \mathbb{C}^{(N+2+N_p) \times (N+2+N_p)}$ are suitably chosen and $\lambda_\ell \geq 0$ for all ℓ.

(b) If

$$|\langle u, w_\ell \rangle_\Gamma| \leq C \exp(-\kappa \sqrt{|\lambda_\ell|}), \tag{3.47}$$

for some $C, \kappa > 0$, which follows e.g. if the assumptions of Lemma 3.19 (b) are satisfied, then

$$\|u - u_N\|_{V_+} \leq \tilde{C} \left(\sup_{\ell < L} \frac{\mathrm{err}_{dtn}(N, \ell)}{(1 + |\lambda_\ell|)^{1/2}} \exp\left(-\frac{\kappa}{2}\sqrt{|\lambda_\ell|}\right)\right.$$
$$\left. + (C_{\mathrm{III}} + C_{\mathrm{III}}(N)) \exp\left(-\frac{\kappa}{4}\sqrt{|\lambda_L|}\right) \right) \|u\|_{H_\sigma(\Gamma)},$$

with $\sigma_\ell = \exp(\frac{\kappa}{2}\sqrt{|\lambda_\ell|})$ and $\tilde{C} = \max\{C', \sup_{\ell \geq 0}(1 + |\lambda_\ell|)^{1/2}\exp(-\frac{\kappa}{4}|\lambda_\ell|)\}$. Further, there exist constants $C_L > 0, \kappa_L > 0$ depending on L and a fixed $N_p \in \mathbb{N}_0$ such that

$$\sup_{\ell < L} \frac{\mathrm{err}_{dtn}(N + 1 + N_p, \ell)}{(1 + |\lambda_\ell|)^{1/2}} \exp\left(-\frac{\kappa}{2}\sqrt{|\lambda_\ell|}\right) \leq C_L \exp(-\kappa_L N),$$

provided that $A, B \in \mathbb{C}^{(N+2+N_p)\times(N+2+N_p)}$ are suitably chosen and $\lambda_\ell \geq 0$ for all ℓ.

Proof. (a) The first part of the result immediately follows from (3.45) by taking $\sigma_\ell = (1 + |\lambda_\ell|)^{s/2}$ and applying Lemma 3.3 (b). The bound (3.46) follows from Proposition 3.13 since

$$\sup_{\ell \leq L-1} \mathrm{err}_{dtn}(N + 1 + N_p, \ell) \leq \sup_{\lambda \in [\lambda_0, \lambda_{L-1}]} |\mathit{dtn}(\lambda) - \mathrm{dtn}_{N+1+N_p}(\lambda)|$$

and $\min_{\ell < L}(1 + |\lambda_\ell|)^{\frac{s+1}{2}} \geq 1$.

(b) The decay (3.47) ensures that $u \in H_\sigma(\Gamma)$ with $\sigma_\ell = \exp(\frac{\kappa}{2}\sqrt{|\lambda_\ell|})$. Here we use that $\sum_\ell \exp(-\kappa\sqrt{|\lambda_\ell|}) < \infty$, which follows in the general case when the eigenvalues are not explicitly known from Weyl's asymptotic law. Then plugging σ_ℓ as defined above into (3.45) yields the claim upon noting that

$$\sup_{\ell \geq L} \frac{(1 + |\lambda_\ell|)^{1/2}}{\sigma_\ell} \leq \left(\sup_{\ell \geq L}(1 + |\lambda_\ell|)^{1/2}\exp(-\frac{\kappa}{4}|\lambda_\ell|)\right) \sup_{\ell \geq L}\exp\left(-\frac{\kappa}{4}\sqrt{|\lambda_\ell|}\right)$$
$$\leq \tilde{C}\exp\left(-\frac{\kappa}{4}\sqrt{|\lambda_L|}\right).$$

The second part of the statement follows as in part (a). $\qquad\square$

As another corollary a basic convergence result as $N \to \infty$ is easily obtained.

Corollary 3.22 (Convergence without rate). *Let the assumptions of Corollary 3.21 (a) for some $s > 1$ be fulfilled. Further assume that the constant $C_{\mathrm{III}}(N)$ in $(\mathrm{dtn}_N\text{-III})$ can be chosen independently of N for N sufficiently large. Then*

$$\|u - u_N\|_{V_+} \to 0 \text{ as } N \to \infty$$

can be achieved with transparent boundary conditions of tensor-product type.

Proof. Let $\varepsilon > 0$ be given. By assumption there exists an $N_0 \in \mathbb{N}_0$ such that for all $N \geq N_0$ one can choose $C_{\mathrm{III}}(N) = C_{\mathrm{III}}(N_0)$. Since $s > 1$ and $|\lambda_L| \to \infty$ as $L \to \infty$,

the asymptotic contribution in the error estimate of Corollary 3.21 (a) goes to zero as $L \to \infty$. Fix L_0 large enough such that

$$C'(C_{\mathrm{III}} + C_{\mathrm{III}}(N_0)) \frac{1}{(1 + |\lambda_{L_0}|)^{\frac{s-1}{2}}} \|u\|_{H_s(\Gamma)} < \frac{\varepsilon}{2}.$$

Then (3.46) guarantees that for fixed L_0 there exists an $N_1 \geq N_0$ such that for all $N \geq N_1$

$$C' \sup_{\ell < L_0} \frac{\mathrm{err}_{dtn}(N, \ell)}{(1 + |\lambda_\ell|)^{\frac{s+1}{2}}} \|u\|_{H_s(\Gamma)} < \frac{\varepsilon}{2}.$$

Thus, $\|u - u_N\|_{V_+} < \varepsilon$ for $N \geq N_1$ follows from Corollary 3.21 (a) by combining the two estimates. $\qquad\square$

So far the convergence analysis proceeded purely at the semi-discrete level. In particular, the discretization error of the differential operators on Γ was neglected. However, the next corollary shows that a basic convergence result for a fully-discrete scheme is easily obtained based on the already established results.

Corollary 3.23 (Convergence for fully-discrete problem). *Let the assumptions of Corollary 3.22 be fulfilled. For a finite dimensional subspace $V_h \subset V_+$ consider the Galerkin discretization of* (3.13): *Find $u_{N,h} \in V_h$ such that*

$$b_N(u_{N,h}, v_h) = l(v_h), \quad \text{for all } v_h \in V_h.$$

Then there exists a sequence (h_N) such that

$$\lim_{N \to +\infty} \|u - u_{N,h_N}\|_{V_+} \to 0.$$

Proof. Let $\varepsilon > 0$ be given. By Corollary 3.22 there exists $N \in \mathbb{N}$ such that $\|u - u_N\|_{V_+} < \varepsilon/2$, where u_N is the solution of the semi-discrete problem (3.13). By the triangle inequality

$$\|u - u_{N,h}\|_{V_+} \leq \|u - u_N\|_{V_+} + \|u_N - u_{N,h}\|_{V_+} < \frac{\varepsilon}{2} + \|u_N - u_{N,h}\|_{V_+}.$$

Hence, it only remains to show that the Galerkin discretization for the semi-discrete problem converges for fixed N, i.e. that $\|u_N - u_{N,h}\|_{V_+} < \varepsilon/2$ for $h = h(N)$ sufficiently small. To this end, recall from the proof of Proposition 3.16 that

$$b_N(w, v) = (B_N w, v)_{V_+}, \quad w, v \in V_+,$$

where $B_N \in \mathscr{L}(V_+, V_+)$ is an injective Fredholm operator. More precisely, we have:

- The operator B_N is of the form $B_N = \tilde{B}_N - K$, where the operator $\tilde{B}_N$ associated with the sesquilinearform $\tilde{b}_N$ from (3.40) is coercive on V_+ and K is compact thanks to compactness of the embeddings $\iota : V_+ \hookrightarrow L^2(\Omega_{\mathrm{int}})$ and $\iota_\Gamma : V_+ \to H^1(\Gamma) \hookrightarrow L^2(\Gamma)$.

- Since $\tilde{B}_N$ is coercive, the Galerkin discretization converges for $\tilde{B}_N$, see e.g. [Kre14, Thm. 13.30] for a proof.

Now we are exactly in the setting of [Kre14, Thm. 13.7], which states that under the above conditions the Galerkin discretization also converges for the perturbation of $\tilde{B}_N$ by the compact operator K, i.e. for B_N. It follows that $\|u_N - u_{N,h}\|_{V_+} < \varepsilon/2$ for h sufficiently small, which concludes the argument. $\qquad\square$

Towards a global convergence rate for the Helmholtz equation

Although convergence for $N \to \infty$ has been shown in Corollary 3.22, this result does not provide any information on the rate of convergence. If we fix L so that the asymptotic part of the error in the estimates of Corollary 3.21 is neglegible, then exponential convergence in N for the first $L-1$ modes follows by Proposition 3.13. However, to obtain global rates of convergence also L has to go to infinity in order to guarantee that the asymptotic part of the error converges to zero as well. Since the constants in Proposition 3.13 depend on the length of the interval and therefore on L, this is a delicate matter. Moreover, to apply Proposition 3.13 rigorously, we need to be able to actually prove results about the pole structure of dtn. This is one of the reasons why the considerations in this section are restricted to $dtn^{\mathrm{hom,2d}}$, whose pole structure is reasonably well-understood. According to (2.36), the poles λ_{p}^n of $dtn^{\mathrm{hom,2d}}$ behave as

$$\lambda_{\mathrm{p}}^n \sim \left(\frac{\pi n}{a}\right)^2 \frac{\left[\left(\frac{\pi}{2} - \arg(ka)\right)^2 - \ln\left(\frac{3\pi n}{e|ka|}\right)^2 + 2i\left(\frac{\pi}{2} - \arg(ka)\right)\ln\left(\frac{3\pi n}{e|ka|}\right)\right]}{\left[\ln\left(\frac{3\pi n}{e|ka|}\right)^2 + \left(\frac{\pi}{2} - \arg(ka)\right)^2\right]^2}$$

for $n \to \infty$. In particular, $\operatorname{Re}\lambda_{\mathrm{p}}^n \to -\infty$ and $\operatorname{Im}\lambda_{\mathrm{p}}^n \to +\infty$ as $n \to \infty$. Thus, there exists an $N_p = N_p(a,k) \in \mathbb{N}_0$ and $\eta > 0$ such that $\mathbb{C} \setminus S_\eta$ only contains the poles $\lambda_{\mathrm{p}}^1, \ldots, \lambda_{\mathrm{p}}^{N_p}$. Moreover, these poles are simple as shown Proposition 2.4 (b) and the bound (dtn-III) extends to $\mathbb{R}_+$ owing to Lemma 2.3. This allows to apply Proposition 3.13. In combination with the results from the previous section, the following estimate of the convergence rate can be established.

Proposition 3.24 (Global convergence rate for the Helmholtz equation). *Consider the setting in which the PDE problem (2.2) in two dimensions is truncated by a circle with radius a and the exterior medium is described by the homogeneous Helmholtz equation, i.e. $dtn^{\mathrm{hom,2d}}$, see Section 2.2.1. Suppose that the separability assumption (2.5) extends to an annulus $b \le r < a$. Let (dtn_N-I$_<$) and (dtn_N-II) be fulfilled. Let $N_p(a,k) \in \mathbb{N}_0$ such that the first N_p poles of $dtn^{\mathrm{hom,2d}}$ are contained in $\mathbb{C} \setminus S_\eta$. Then for any $N \in \mathbb{N}_0$ with $\tilde{N} := N + 1 + N_p$ the estimate*

$$\|u - u_{\tilde{N}}\|_{V_+} \le \tilde{C}\left(C_\alpha \| dtn_{N_p}^{\mathrm{reg}} e^{\theta(dtn_{N_p}^{\mathrm{reg}})}\|_{D(N^{2\alpha},\eta),\infty} \exp\left(-\frac{N}{2}\log\left(1 + \frac{\eta}{N^{2\alpha}}\right)\right)\right.$$

$$\left. + (C_{\mathrm{III}} + C_{\mathrm{III}}(\tilde{N})) \exp\left(-\log\left(\frac{a}{b}\right)\frac{\lfloor aN^\alpha \rfloor}{4}\right)\right)\|u\|_{H_\sigma(\Gamma)}, \quad (3.48)$$

for any $0 < \alpha < 1/5$ with $C_\alpha < \infty$ depending on α holds. Here, $\sigma_\ell = \exp\left(\frac{a}{2}\log(\frac{a}{b})\sqrt{|\lambda_\ell|}\right)$, and $\lfloor x \rfloor := \max\{m \in \mathbb{Z} \mid m \le x\}$.

Proof. According to Lemma 2.3, the function $dtn^{\mathrm{hom,2d}}$ fulfills assumptions (dtn-I$_<$), (dtn-II) and (dtn-III). Hence, the continuous problem is well-posed by Corollary 2.20. Since the separability assumption has been extended to $b \le r < a$, the Fourier coefficients of the solution on Γ decay exponentially fast according to Lemma 3.19 (b). As (dtn_N-I$_<$) and (dtn_N-II) have been assumed, Corollary 3.21 (b) with $\kappa = a\log(a/b)$ then implies the error bound

$$\|u - u_{\tilde{N}}\|_{V_+} \leq \tilde{C} \left(\sup_{\ell < L} \frac{\mathrm{err}_{dtn}(\tilde{N}, \ell)}{(1 + |\lambda_\ell|)^{1/2}} \exp\left(-\frac{a}{2} \log\left(\frac{a}{b}\right) \sqrt{|\lambda_\ell|} \right) \right.$$

$$\left. + (C_{\mathrm{III}} + C_{\mathrm{III}}(\tilde{N})) \exp\left(-\frac{a}{4} \log\left(\frac{a}{b}\right) \sqrt{|\lambda_L|} \right) \right) \|u\|_{H_\sigma(\Gamma)} \qquad (3.49)$$

for $\sigma_\ell = \exp\left(\frac{a}{2} \log(\frac{a}{b}) \sqrt{|\lambda_\ell|} \right)$. By appealing to Proposition 3.13 we obtain

$$\sup_{\ell < L} \mathrm{err}_{dtn}(\tilde{N}, \ell) \leq f(\lambda_L, N) \| \boldsymbol{dtn}_{N_p}^{\mathrm{reg}} e^{\theta(\boldsymbol{dtn}_{N_p}^{\mathrm{reg}})} \|_{D(\lambda_L, \eta), \infty} \exp\left(-\frac{N}{2} \log\left(1 + \frac{\eta}{\lambda_L} \right) \right),$$

for

$$f(\lambda, N) := \frac{4\lambda}{\eta} \left(1 + \lambda(1 + \lambda)^{1/2} \left(C_{\mathrm{III}} + C_{N_p}^{\mathrm{res}} \right) \right)$$

$$\times \exp\left(\lambda(1 + \lambda)^{1/2} \left(C_{\mathrm{III}} + C_{N_p}^{\mathrm{res}} \right) - \frac{N}{2} \log\left(1 + \frac{\eta}{\lambda} \right) \right) \qquad (3.50)$$

and

$$\boldsymbol{dtn}_{N_p}^{\mathrm{reg}}(\lambda) = \boldsymbol{dtn}^{\mathrm{hom,2d}}(\lambda) - \sum_{n=1}^{N_p} \frac{\mathrm{Res}(\boldsymbol{dtn}^{\mathrm{hom,2d}}, \lambda_{\mathrm{p}}^n)}{\lambda - \lambda_{\mathrm{p}}^n}.$$

Now one has to choose $L = L(N)$ in $\lambda_L = (L/a)^2$ to balance the two contributions in (3.49). Set $L = \lfloor aN^\alpha \rfloor$, i.e. $\lambda_L \leq N^{2\alpha}$, for $\alpha > 0$ to be determined such that

$$C_\alpha := \sup_{N \in \mathbb{N}_p} f(N^{2\alpha}, N) < \infty$$

holds true. To ensure this condition, we have to find α such that the argument of the exponential in (3.50) for $\lambda = N^{2\alpha}$ converges to $-\infty$ as $N \to \infty$. Using $\log(1 + x) = x + \mathcal{O}(x^2)$ for $|x| \ll 1$, it follows that the argument of the exponential is given by

$$N^{2\alpha} \left(1 + N^{2\alpha} \right)^{1/2} \left(C_{\mathrm{III}} + C_{N_p}^{\mathrm{res}} \right) - \frac{N}{2} \left(\eta N^{-2\alpha} + \mathcal{O}(N^{-4\alpha}) \right),$$

which converges to $-\infty$ as $N \to \infty$ provided that $3\alpha < 1 - 2\alpha$, that is $0 < \alpha < 1/5$. This leads to

$$\sup_{\ell < L} \mathrm{err}_{dtn}(\tilde{N}, \ell) \leq C_\alpha \| \boldsymbol{dtn}_{N_p}^{\mathrm{reg}} e^{\theta(\boldsymbol{dtn}_{N_p}^{\mathrm{reg}})} \|_{D(N^{2\alpha}, \eta), \infty} \exp\left(-\frac{N}{2} \log\left(1 + \frac{\eta}{N^{2\alpha}} \right) \right).$$

The claimed estimate (3.48) follows upon noting that

$$\frac{\exp\left(-\frac{a}{2} \log\left(\frac{a}{b}\right) \sqrt{|\lambda_\ell|} \right)}{(1 + |\lambda_\ell|)^{1/2}} \leq 1 \text{ and } \exp\left(-\frac{a}{4} \log\left(\frac{a}{b}\right) \sqrt{|\lambda_L|} \right) = \exp\left(-\log\left(\frac{a}{b}\right) \frac{\lfloor aN^\alpha \rfloor}{4} \right).$$

$\square$

Remark 3.25 (Open question). *The result of Proposition 3.24 is not entirely satisfactory yet due to occurence of the term $\| \boldsymbol{dtn}_{N_p}^{\mathrm{reg}} e^{\theta(\boldsymbol{dtn}_{N_p}^{\mathrm{reg}})} \|_{D(N^{2\alpha}, \eta), \infty}$ in (3.48) whose dependence*

on N is not clearly understood. This term stems from applying polynomial appproximation to the regularized and transformed $\textit{dtn}^{\text{hom,2d}}$ function in the ellipse $D(N^{2\alpha}, \eta)$, see Proposition 3.12 and Proposition 3.13. Currently, it is not possible to estimate this term rigorously since the bound $|\textit{dtn}_{N_p}^{\text{reg}}(\lambda)| \leq (C_{\text{III}} + C_{N_p}^{\text{res}})(1 + |\lambda|)^{1/2}$ has only been shown for $\lambda \geq 0$, whereas $\lambda \in D(N^{2\alpha}, \eta) \subset \mathbb{C}$ has to be treated here. This would require an extension of Lemma 2.3 in which $\textit{dtn}^{\text{hom,2d}}$ is replaced by $\textit{dtn}_{N_p}^{\text{reg}}$ and complex $\lambda \in D(N^{2\alpha}, \eta)$ are admitted. If one assumes nonetheless that such an estimate was valid also in $D(N^{2\alpha}, \eta)$, then the antiderivative $|\theta(\textit{dtn}_{N_p}^{\text{reg}}, \lambda)|$ behaves like $\sim (1 + |\lambda|)^{3/2}$ and $\|\textit{dtn}_{N_p}^{\text{reg}} \, e^{\theta(\textit{dtn}_{N_p}^{\text{reg}})}\|_{D(N^{2\alpha}, \eta), \infty} \sim N^\alpha \exp(CN^{3\alpha})$. As seen in the proof of Proposition 3.24,

$$\exp\left(-\frac{N}{2}\log\left(1 + \frac{\eta}{N^{2\alpha}}\right)\right) \sim \exp\left(-\frac{\eta N^{1-2\alpha}}{2}\right)$$

for $N \to \infty$. Hence, an additional prefactor of $N^\alpha \exp(CN^{3\alpha})$ would leave the rate unaffected as long as $\alpha < 1/5$, which is already assumed anyway.

This proposition concludes the theoretical part of this chapter. The following Section 3.4 presents a case study which investigates the performance of many popular transparent boundary conditions that fit into the tensor-product framework as introduced in Section 3.1.

3.4 Comparison of popular transparent boundary conditions based on approximation of $\textit{dtn}$

In the following, the error $\text{err}_{\textit{dtn}}(N, \ell)$ will be numerically investigated for different popular transparent boundary conditions recalled in Appendix C which fit into the framework introduced in Section 2.2.1. The problem under study is the constant coefficient Helmholtz equation (2.27) in the exterior of a three dimensional ball with radius a using a Sommerfeld radiation condition at infinity as introduced in Section 2.2.1. In Appendix C the matrices A and B in the ansatz (3.7) are identified for tensor-product PMLs ('TP-PML'), two classical infinite element formulations ('U-Burnett' and 'Astley-Leis'), Hardy-Space-Infinite elements ('HSIE') and the first and second order local absorbing boundary conditions of Bayliss, Gunzburger and Turkel ('BGT'). Readers interested in the details of this derivation are invited to read through Appendix C at this point. Once the matrices A and B in (3.7) for a specific transparent boundary condition have been determined, formula (3.10) yields the corresponding dtn_N function so that the error $\text{err}_{\textit{dtn}}(N, \ell)$ can be measured.

It is natural to start the investigation with the lowest order condition $N = 0$. From a practical perspective this condition is very attractive since it can be implemented directly on the coupling boundary without introducing additional degrees of freedom. In particular, the matrices A and B consist entirely of the contribution from the coupling interface: $A = A_{\Gamma\Gamma} \in \mathbb{C}$ and $B = B_{\Gamma\Gamma} \in \mathbb{C}$. According to (3.10), this implies that the lowest order conditions try to approximate $\lambda \mapsto \textit{dtn}(\lambda)$ by the affine linear function

$$\lambda \mapsto A + \lambda B.$$

Since $\textit{dtn}$ is usually not affine linear, such an approximation can only be accurate on a sufficiently small interval.

The affine linear $\mathrm{dtn}_{N=0}$ approximations of dtn achieved with different transparent boundary conditions are shown in Figure 3.4 for $a = 4$. It is interesting that the HSIE, unconjugated Burnett formulation and BGT condition (using $m = 2$, see Appendix C.4) seem to arrive at the same linear fit. That the initial value $\mathrm{dtn}_{N=0}(\lambda = 0) = A = -ik + 1/a$ actually coincides for all conditions can be inferred from the definitions introduced in Appendix C (upon choosing $\kappa_0 = ak$ for the HSIE). However, that the HSIE, unconjugated Burnett and BGT conditions also lead to linear approximations with very similar slope (i.e. B) is not obvious from the definitions[6]. For example, $B = 1/2(-ik + 1/a)^{-1}$ for the BGT condition whereas B is defined in terms of special functions for the unconjugated Burnett condition (C.14). Overall, these results demonstrate that a linear approximation is only appropriate for small λ corresponding to the lowest order modes.
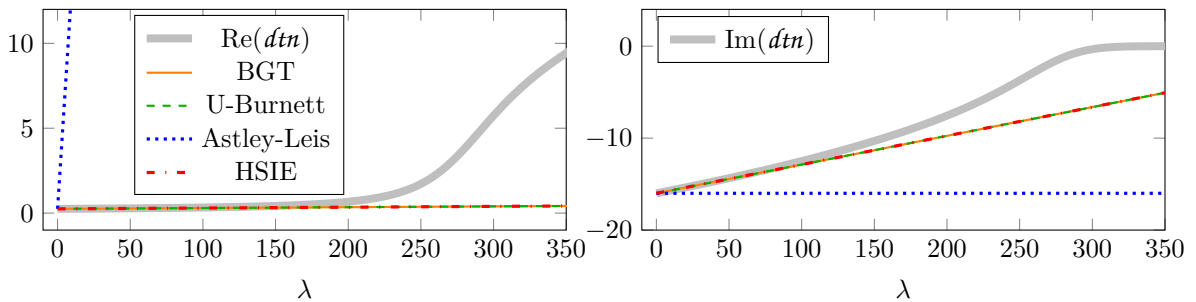

Figure 3.4: Comparison of $\mathit{dtn}^{\mathrm{hom,3d}}$ approximation for different transparent boundary conditions for $N = 0, k = 16, a = 4$ in dimension $d = 3$.

To improve the approximation of dtn, additional degrees of freedom can be introduced. According to (3.10), this amounts to enriching the linear approximation by the rational function

$$\lambda \mapsto -(A_{\Gamma E} + \lambda B_{\Gamma E})(A_{EE} + \lambda B_{EE})^{-1}(A_{E\Gamma} + \lambda B_{E\Gamma}),$$

which admits poles at the negative eigenvalues of $(B_{EE})^{-1}A_{EE}$. Fig. 3.6 displays these poles in the lower panel, the approximation of dtn by dtn_N in the upper panel and the relative error $|\mathit{dtn}^{\mathrm{hom,3d}}(\lambda_\ell) - \mathrm{dtn}_N(\lambda_\ell)|/|\mathit{dtn}^{\mathrm{hom,3d}}(\lambda_\ell)|$ for different transparent boundary conditions using $N \in \{3, 6, 9\}$. Note that the matrices for the infinite elements methods are dense, i.e. they have $(N+1)^2$ non-trivial entries, while the PML matrices are sparse, see Table 3.1 for a detailed comparison between the matrix dimension and the number of nonzero entries. Therefore, a direct comparison in terms of the matrix dimension would put the PML at an unfair disadvantage. It is more appropriate to compare the results achieved with infinite element matrices of dimension $(N+1) \times (N+1)$ to those obtained with TP-PML matrices containing about $(N+1)^2$ nonzero entries. Consequently, Fig. 3.6 is based on TP-PML matrices with $\{17, 49, 97\}$ nonzero entries for comparing with the other methods using $(N+1)^2 \in \{16, 49, 100\}$.

Before proceeding to a discussion of these figures, it has to be mentioned that reliable results for the classical infinite elements presented here with large N are difficult to achieve within finite precision arithmetic. This problem stems from the well-known ill-conditioning of the infinite element matrices. Computing $\mathrm{dtn}_N(\lambda)$ for $N > 0$ requires the solution of a linear system involving the exterior degrees of freedom, i.e. the matrix $A_{EE} + \lambda B_{EE}$. Fig. 3.5 displays the condition number of this matrix for increasing N.

---

[6]The agreement seems to improve further as a increases.

$N+1$	$(N+1)^2$	nze	ne
3	9	9	1
5	25	17	2
9	81	33	4
13	169	49	6
17	289	65	8
21	441	81	10

Table 3.1: Number of nonzero matrix entries 'nze' of tensor-product PML matrices $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ based on 'ne' quadratic finite elements in radial direction.

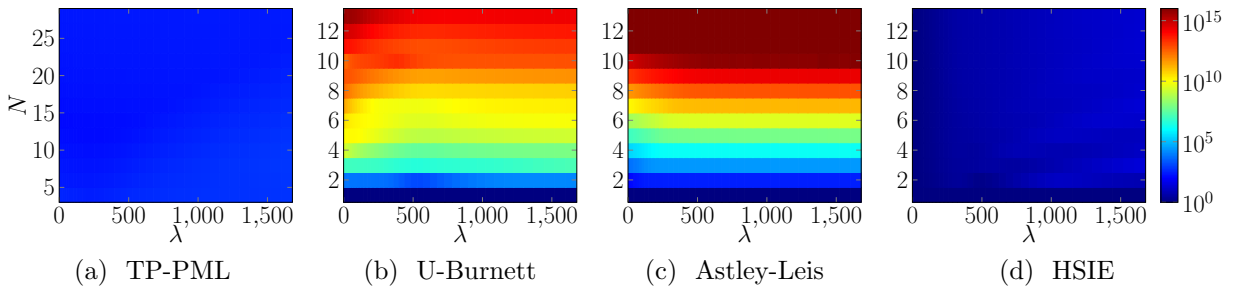(a) TP-PML   (b) U-Burnett   (c) Astley-Leis   (d) HSIE

Figure 3.5: Condition number of the matrix $A_{EE} + \lambda B_{EE}$ for different transparent boundary conditions for $k = 16, a = 1$ in dimension $d = 3$.

Whereas the condition number for tensor-product PML and HSIE is essentially independent of N, it grows exponentially for both classical infinite element discretizations. For $N = 9$ the condition number is already so large that the accuracy of dtn_N (right panel of Fig. 3.6) could be seriously affected. Although this might be remedied by using symbolic computation as done in [DI01] or multiple-precision arithmetic, such measures are most certainly not in the interest of computational efficiency. Hence, the computations for Fig. 3.5 have been performed adhering to conventional double precision. Another, more feasible approach to improve the conditioning of the infinite element matrices is to change the basis in the direction of radiation. Improvements for some infinite element formulations have been achieved based e.g. on shifted Legendre polynomials [AC01] or certain Jacobi polynomials [DE03]. Unfortunately, [AC01] concluded that the condition number for the unconjugated Burnett formulation increased rapidly no matter which radial basis was used.

With this being said, Fig. 3.5 can be discussed:

- The dtn_N approximation using TP-PML, Astley-Leis infinite elements or HSIE converges exponentially fast to *dtn* as N increases. This is consistent with the convergence proofs which have been derived for these methods in the literature, see [LS98, HSZ03b] for the PML, [DG98, DI01] for conjugated Astley-Leis infinite elements and [HN09] for the HSIE. In contrast, there is no convergence proof available in the literature for the unconjugated Burnett formulation. In fact, several sources [SB98, Ihl00] report that the this formulation might fail to converge in the far field. The middle panel of Fig. 3.5 indeed shows that the dtn_N approximation for the unconjugated Burnett formulation makes almost no progress for $\lambda > 600$ from $N = 6$ to $N = 9$. Achieving reliable results for larger N is difficult due to the
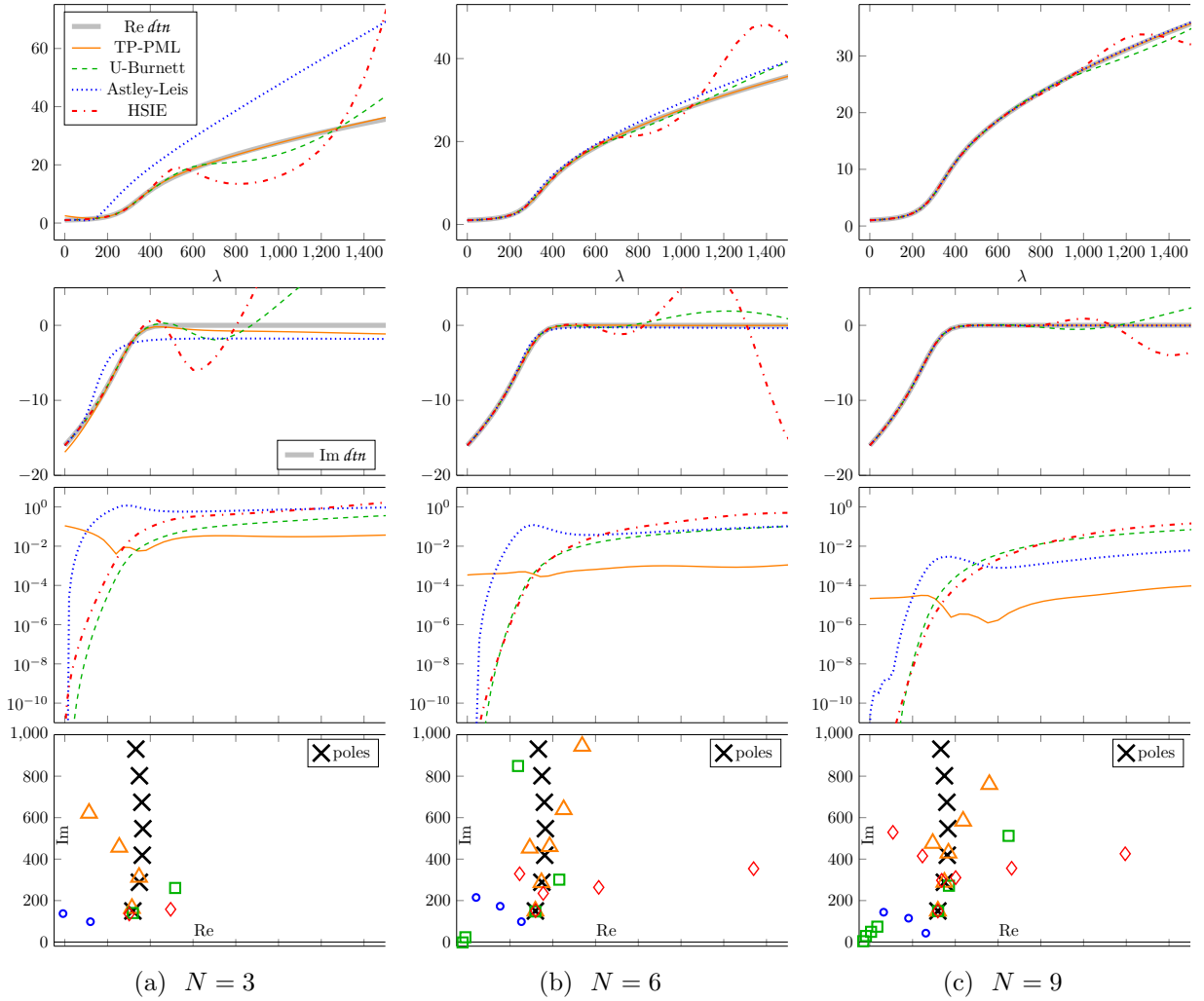
Figure 3.6: From top to bottom: approximation of $\text{Re}(\mathit{dtn}^{\text{hom,3d}})$, approximation of $\text{Im}(\mathit{dtn}^{\text{hom,3d}})$, relative error $|\mathit{dtn}^{\text{hom,3d}}(\lambda) - \text{dtn}_N(\lambda)|/|\mathit{dtn}^{\text{hom,3d}}(\lambda)|$ and poles for different transparent boundary conditions for $k = 16, a = 1$ in dimension $d = 3$. The TP-PML uses matrices with $\{17, 49, 97\}$ nonzero elements for $N \in \{3, 6, 9\}$ respectively.

increasing ill-conditioning of the matrices.

- It appears to be generally acknowledged in the literature on classical infinite elements [Ger98, SB98] that the unconjugated Burnett formulations exhibits the fastest convergence in the vicinity of a scatterer and at its boundary, whereas a conjugated formulation has to be used if convergence of the far field solution is required. The results shown in Fig. 3.5 are consistent with this statement. For small λ, corresponding to the low frequency components of the solution, $\mathrm{dtn}_N(\lambda)$ converges much faster to $\boldsymbol{dtn}^{\mathrm{hom,3d}}(\lambda)$ for the unconjugated Burnett than for the conjugated Astley-Leis formulation. On the other hand, the latter formulation performs well for large λ (see e.g. $N = 9$). For scattering problems with homogeneous exterior domains fast convergence in the near field as exhibited by the unconjugated Burnett formulation is often the preferred property though since the far field solution can be recovered from the boundary data by means of Green's formula if required (see e.g. [CK13] for basic relations and [GD05] for computationally efficient approaches).

- The HSIE method converges very slowly for large λ. This may also be expected since transforming the basis of the Hardy space used for the radial discretization back to the space domain results into polynomial basis functions in r multiplied by a plane wave component [HN09]. Such a basis is inappropriate for representing the far field behavior of the solution as the Atkinson-Wilcox expansion (C.18) shows. For small λ the HSIE method converges as fast as the unconjugated Burnett formulation and distinguishes itself from the latter by its well-conditioned matrices.

- The tensor-product PML provides an approximation of $\boldsymbol{dtn}^{\mathrm{hom,3d}}$ which is almost uniform in λ. As a result, it drastically outperforms the infinite element methods for large λ. This can be advantageous for problems whose solutions are rough in the sense that they involve a large number of highly oscillatory components. Such situations may arise in scattering from non-smooth obstacles or in the presence of sources located in close vicinity of the truncation boundary, see e.g. Section 4.2.2. Additionally, this can be an important factor in the context of domain decompositioning methods where the boundary data on interfaces coupling different subdomains is usually not smooth. The usage of PMLs in this setting has been popularized by the emergence of sweeping preconditioners - a special class of domain decomposition methods considered in Chapter 7 and Chapter 8.

- It is interesting to compare the poles of dtn_N with the exact poles of $\boldsymbol{dtn}$. Already for $N = 3$ the TP-PML, unconjugated Burnett formulation and HSIE place one of their poles very close to the first exact pole of $\boldsymbol{dtn}^{\mathrm{hom,3d}}$. The agreement improves further as N increases. Comparing the figures for $N \in \{3, 6, 9\}$ also a convergence towards the second exact pole can be seen. The PML method also fits the third pole relatively well for $N = 9$ and appears to be the fastest to converge towards the exact poles. The conjugated Astley-Leis formulation is the only method for which no clear convergence towards the exact poles can be detected although some of its poles are placed in close proximity to the first exact pole. In this regard, it is interesting to note that the relative error of the approximation of $\boldsymbol{dtn}^{\mathrm{hom,3d}}$ by dtn_N for this particular method is largest when λ is close to the real part of the first exact pole (the maximum of the blue dotted curve in the panel showing the relative

error is located exactly above the vertical line formed by the exact poles). The poor approximation quality of $\mathrm{dtn}_N(\lambda)$ on this interval of λ could therefore be related to a failure of the method to fit the exact poles of dtn. Overall, these results indicate that the poles could be a key component in achieving an accurate approximation of $dtn^{\mathrm{hom,3d}}$. Further evidence for this claim will be presented in Chapter 4.

Remark 3.26 (Extension to inhomogeneous exterior domains). *The $dtn^{\mathrm{hom,3d}}$ function considered in this example models a homogeneous exterior domain. In view of the application to helioseismology, the primary interest of this thesis is situated in strongly inhomogeneous exterior domains exemplified by the $dtn^{\mathrm{VAL-C}}$ function of the solar atmosphere or its simplified mockup dtn^{jump}. None of the methods considered here seems to be readily applicable to such a setting:*

- *The coordinate stretching on which PMLs are based requires that the coefficients of the PDE and its solution allow for an analytic extension into the complex plane. This requirement is violated for a discontinuous wavenumber as in (2.40) or if sound speed and density are given by BSpline approximations of tabulated values like for the VAL-C model.*

- *The author is not aware that classical infinite elements as introduced in Appendix C.2 have ever been applied successfully to solve problems involving inhomogeneous exterior domains.*

- *In [NS11] Nannen and Schädle introduced variants of the HSIE which can deal with certain types of inhomogeneous exterior domains. According to [NS11, Remark 7] the method appears to be limited to a very particular dependence of the medium coefficients on the variable in propagation direction (in this thesis r) which would not cover the discontinuous wavenumber as in (2.40) not to mention sound speed and density of the VAL-C model.*

- *Local absorbing boundary conditions have been developed for the simplified* `Atmo` *model of the solar atmosphere as mentioned in Section 2.4.2. An extensive comparison of these conditions with the newly proposed method of this thesis will be given in Section 6.2. For the more complicated VAL-C model local absorbing conditions are currently not known.*

Chapter 4

Learned infinite elements for individual wavenumbers

In Chapter 3 an error analysis for generic transparent boundary conditions of tensor-product type (3.8) has been presented. According to Theorem 3.20, the error $\|u - u_N\|_{V_+}$ between the exact solution u and the approximate solution u_N is essentially determined by the approximation of *dtn* by $\mathrm{dtn}_N\{A, B\}$. If the exact solution is sufficiently smooth in a neighborhood of the truncation boundary, it suffices to achieve an accurate fit at the first $\ell = 0, \ldots, L - 1$ eigenvalues λ_ℓ since the remaining modes decay rapidly as utilized e.g. in Corollary 3.21 (b) (see also Lemma 3.19). Therefore, we have

$$\|u - u_N\|_{V_+}^2 \lesssim \|\Pi_L u\|_{H_{\sigma^{\mathrm{err}}}(\Gamma)}^2 \lesssim \sum_{\ell=0}^{L-1} \mathrm{err}_{\textit{dtn}}(N, \ell)^2 |\langle u, w_\ell \rangle_\Gamma|^2, \tag{4.1}$$

up to a remainder which is negligible if L is chosen sufficiently large.

The challenge of constructing efficient and accurate transparent boundary conditions for stratified media therefore reduces to the task of finding matrices $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ which minimize the right hand side of equation (4.1). To maximize efficiency, it would be highly desirable to find a sequence of matrices such that the right hand side converges exponentially fast to zero as N is increased. Is this task feasible and how can suitable matrices be determined?

The results from Chapter 3 already provide some clues.

- At least for the case of a spherical coupling boundary with a homogeneous exterior several transparent boundary conditions providing exponential approximation rates exist according to the investigation in Section 3.4. However, none of these methods is flexible enough to be applied in the setting of helioseismology, which features a highly heterogeneous exterior domain. More generally, as shown in Appendix C, in these methods the matrices A, B are derived analytically without explicitly aiming to obtain an optimal approximation of *dtn*. Therefore, it seems likely that faster convergence rates should be possible for homogeneous exterior domains by optimizing the choice of A and B.

- According to the theoretical foundation etablished in Section 3.2, achieving exponential convergence rates should be possible for almost any stratified medium, even for the Sun. However, neither the analysis in Section 3.2 nor the methods presented in Appendix C provide a practical way of finding the matrices A and B for solar-like media. This chapter suggests an approach to close this gap.

The main idea is to obtain these matrices as minimzers of an optimization problem of the form

$$A, B \in \underset{A,B \in \mathbb{C}^{(N+1)\times(N+1)}}{\operatorname{argmin}} \| \mathbf{dtn} - \operatorname{dtn}_N\{A, B\}\|_*^2,$$

where $\| \cdot \|_*$ denotes a suitable norm. According to equation (4.1), the straightforward choice would be a weighted ℓ^2-norm. The optimization problem is introduced and studied in detail in Section 4.1. It will be seen that A and B can even be chosen to be sparse without impairing the approximation quality of dtn_N. Several numerical experiments presented in Section 4.2-Section 4.3 demonstrate that exponential convergence rates carry over to $\|u-u_N\|_{L^2(\Omega_{\text{int}})}$. The full range of different problem settings presented in Chapter 2 will be covered, with the exception of the helioseismology problem, which deserves its own dedicated chapter.

- Section 4.2 considers the Helmholtz equation for circular truncation boundaries. Firstly, a homogeneous exterior domain will be assumed which allows for a comparison with several popular transparent boundary conditions like PMLs or HSIEs (see Appendix C for an introduction to these methods). This assumption is relaxed in Section 4.2.3, which treats exterior wavespeeds with a step discontinuity.

- Elliptical trunctation boundaries are considered in Section 4.3.

- Section 4.4 presents numerical results for a semi-infinite waveguide.

Most of the results of this chapter already appeared in our publication [HLP21] from which some passages have been taken over verbatim. However, the extension to elliptical trunctation boundaries presented in Section 4.3 is new and we also used this opportunity to introduce a few additional improvements which are discussed in more detail in Section 9.1.

Before starting with the actual investigations it should be pointed out that the dtn_N approximation computed in this chapter, respectively in [HLP21], is valid for a single, fixed wavenumber k only. In particular, the matrices A and B depend on k in some unknown fashion and can therefore not be reused for another wavenumber, say $\tilde{k}$. Computations for $\tilde{k}$ would require a fresh solution of the corresponding minimization problem. Extensions to approximations which work uniformly for all wavenumbers in some bounded set, as required to solve resonance problems, will be investigated in Chapter 5.

4.1   The minimization problem

We start this section by introducing the minimization problem which serves as the foundation of learned infinite elements.

Definition 4.1. *Let λ_ℓ be the generalized eigenvaluess of equation (2.9). Denote by $\mathbf{dtn}(\lambda_\ell)$ the DtN numbers obtained from solving the ODEs (2.12a)-(2.12b). Further, let $\operatorname{dtn}_N(\lambda_\ell)$ be the approximate DtN_N numbers of the learned IEs as in equation (3.10). Recall that $\operatorname{err}_{\mathbf{dtn}}(N, \ell)$ was defined in (3.15). For positive weights ω_ℓ define the misfit function*

$$J(A, B) = \frac{1}{2} \sum_{\ell < L} \omega_\ell^2 \, \operatorname{err}_{\mathbf{dtn}}(N, \ell)^2. \tag{4.2}$$

The minimization problem is to find $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ so that

$$A, B \in \underset{A,B \in \mathbb{C}^{(N+1)\times(N+1)}}{\operatorname{argmin}} J(A, B). \tag{4.3}$$

It is appropriate to give several remarks to explain Definition 4.1 and draw relations to the analysis of Chapter 3.

- The sum in the objective function (4.2) includes only a finite number of terms $\ell < L$. Hence, we abstain from trying to obtain any asymptotic control of *dtn* by dtn_N. This is reasonable since from Chapter 2 it is well known that typical *dtn* functions behave like $\textit{dtn}(\lambda) \sim \sqrt{\lambda}$ as $\lambda \to +\infty$, see e.g. Lemma 2.3 (c), while $\text{dtn}_N(\lambda) = \mathcal{O}(\lambda)$, see ($\text{dtn}_N$-III). This implies that resolving the asymptotic behavior of *dtn* with tensor-product discretizations is impossible anyway. However, inspection of the proof of Corollary 3.21 reveals that asymptotic control is also not necessary provided that the solution is sufficiently regular in a neighborhood of the truncation boundary. In this case the asymptotic contribution to the error can be bounded utilizing asymptotic decay of the coefficients of the solution in the eigenbasis.

- Disregarding the asymptotics it remains to obtain a good approximation for $\ell < L$. According to Theorem 3.20, the weighted sum

$$\sum_{\ell=0}^{L-1} \text{err}_{\textit{dtn}}(N, \ell)^2 |\langle u, w_\ell \rangle_\Gamma|^2$$

needs to be minimized. This explains the choice of the objective function (4.2). In practice, the coefficients $\langle u, w_\ell \rangle_\Gamma$ for $\ell = 0, \ldots, \infty$ of the solution in the eigenbasis and therefore[1] $\text{tr}_\Gamma u$ are not known. Otherwise, one could impose $\text{tr}_\Gamma u$ as Dirichlet data on Γ for the problem on Ω_int which obviates the need for transparent boundary conditions. Hence, choosing the weights as $\omega_\ell = |\langle u, w_\ell \rangle_\Gamma|$ is not directly possible. However, Lemma 3.19 showed that smoothness properties of u are related to decay of $|\langle u, w_\ell \rangle_\Gamma|$. For example, for the constant coefficient Helmholtz equation in two dimensions it was shown in Lemma 3.19 that

$$|\langle u, w_\ell \rangle_\Gamma| \sim \left| \frac{H_\ell^{(1)}(ka)}{H_\ell^{(1)}(kb)} \right| \sim \exp\left(-\ell \log(a/b)\right), \text{ as } \ell \to +\infty, \tag{4.4}$$

where $b < a$ is limited by the requirement that waves in the annulus $b \leq r < a$ still propagate as in free space, which e.g. excludes the existence of sources within this region. Similar considerations as above often allow to derive reasonable proxies for $|\langle u, w_\ell \rangle_\Gamma|$ which can serve as weights ω_ℓ. More details will be presented in Section 4.2-Section 4.4 when dicussion numerical experiments.

- The error analysis presented in Chapter 3 relies on some assumptions on dtn_N given in Section 3.3.1. It is appropriate to discuss whether the dtn_N function obtained from solving the minimization problem fulfills these assumptions. As mentioned previously, the growth condition (dtn_N-III) is automatically fulfilled for any transparent boundary condition of the form (3.10) provided that B_{EE} is nonsingular. The inequality constraint (dtn_N-I$_<$) / (dtn_N-I$_\leq$) could be enforced by means of a Lagrange multiplier, albeit this seems unnecessary for N sufficiently large since dtn_N will naturally inherit the identical constraint (*dtn*-I$_<$) / (*dtn*-I$_\leq$) from *dtn* on finite intervals when converging to the latter. It remains to discuss the asymptotic

---

[1]Since $\{w_\ell\}$ is on orthonormal basis.

constraint (dtn_N-II) on $\operatorname{Re} \text{dtn}_N$. By construction the ansatz (3.10) behaves linearly at infinity, i.e. $\operatorname{Re} \text{dtn}(\lambda) \sim C(1 + \lambda)$ as $\lambda \to +\infty$ for a constant C which could be negative. Therefore, the demand of (dtn_N-II) is essentially that C is positive, which the optimization should inherently strive to comply with because *dtn* fulfills (*dtn*-II) or usually even the stronger property that $\operatorname{Re} \textbf{\textit{dtn}}(\lambda)$ behaves like $\sqrt{\lambda}$ as $\lambda \to +\infty$, cp. Lemma 2.3. In practice, we do not enforce any of these constraints and let the optimization simply run freely because this works just fine.

- The objective function (4.2) is defined in terms of the continuous eigenvalues λ_ℓ even though the discrete DtN_N operator diagonalizes with respect to the discrete eigenvalues $(\underline{\lambda}_\ell, \underline{w}_\ell)$ (see Proposition 3.1). This is justified since the discrete eigenvalues converge towards the continuous ones as the finite element discretization is refined. Moreover, using the continuous eigenvalues, which are usually known analytically, is computationally more efficient than using the discrete eigenvalues, which would need to be computed numerically.

- Obtaining the matrices A and B requires to solve the nonlinear least-squares problem (4.2)-(4.3) for which e.g. trust-region methods are suitable. We had good success in solving this problem using the Levenberg–Marquardt algorithm as implemented in `ceres-solver` [AMO]. A detailed description of the implementation is available in Appendix D.

The next subsections are concerned with different aspects of this minimization problem. While the minimization in (4.3) is performed over dense matrices $A, B \in \mathbb{C}^{(N+1) \times (N+1)}$, Section 4.1.1 demonstrates that it suffices to restrict to sparse matrices without loss of accuracy. This reduction step is motivated by the fact that (4.2)-(4.3) represents a rational approximation problem. Section 4.1.2 starts to investigate the poles of this rational approximation in a bit more detail. Section 4.1.3 presents some results concerning the condition number of the minimizers A and B. Finally, aspects of computational complexity are discussed in Section 4.1.4.

4.1.1 Reduction step

For dense matrices A and B the number of free parameters in the minimization problem (4.2)-(4.3) and the number of nonzero matrix entries of the tensor product system $A \otimes M + B \otimes K$ grows quadratically with N. To improve efficiency it would be desirable to achieve a linear growth in both quantities. Since finite element libraries and associated linear system solvers often benefit from symmetric linear systems, the feasibility of imposing symmetry of the matrices A and B should also be investigated. These goals are obtained in the following two steps.

Sparsification by diagonalization of exterior part

In the first step we aim to set as many entries of A and B to zero as possible. To this end, the dispensable matrix entries have to be identified for which it will be assumed that

$$(B_{EE})^{-1} A_{EE} \text{ is diagonalizable.} \tag{4.5}$$

That is, there exists a diagonal matrix D and an invertible matrix P such that $(B_{EE})^{-1} A_{EE} = PDP^{-1}$. Then

$$(A_{EE} + \lambda B_{EE})^{-1} = P(D + \lambda I)^{-1}(B_{EE}P)^{-1}$$

84

holds. We insert this identity into (3.10) and redefine

$$A_{\Gamma E} \mapsto A_{\Gamma E}P, \; B_{\Gamma E} \mapsto B_{\Gamma E}P, \; A_{E\Gamma} \mapsto (B_{EE}P)^{-1}A_{E\Gamma}, \; B_{E\Gamma} \mapsto (B_{EE}P)^{-1}B_{E\Gamma}$$

and

$$A_{EE} \mapsto D, \quad B_{EE} \mapsto \mathrm{Id}\,. \tag{4.6}$$

The last equation means that under assumption (4.5) we can suppose that the matrix A_{EE} in (3.7) is diagonal and that B_{EE} is equal to the identity. The corresponding dtn_N function is consequently of the form

$$\mathrm{dtn}_N(\lambda) = A_{00} + \lambda B_{00} - \sum_{j=1}^{N} \frac{(A_{0j} + \lambda B_{0j})(A_{j0} + \lambda B_{j0})}{A_{jj} + \lambda}. \tag{4.7}$$

Symmetrization by removing redundant variables

To symmetrize A and B, it has to be justified that the constraints $A_{0j} = A_{j0}$ and $B_{0j} = B_{j0}$ for $j = 1, \ldots, N$ can be imposed without impairing the approximation quality of (4.7) as a rational function. To see this, we use the identity

$$\frac{(\alpha + \lambda)(\beta + \lambda)}{\gamma + \lambda} = \lambda + \alpha + \beta - \gamma + \frac{(\beta - \gamma)(\alpha - \gamma)}{\gamma + \lambda}$$

for $\alpha, \beta, \gamma \in \mathbb{C}$ to rewrite (4.7) in the simplified form

$$\mathrm{dtn}_N(\lambda) = \alpha + \beta\lambda - \sum_{j=1}^{N} \frac{\gamma_j}{\lambda + \delta_j}, \tag{4.8}$$

with

$$\alpha = A_{00} - \sum_{j=1}^{N} \left(A_{0j}B_{j0} + A_{j0}B_{0j} - B_{0j}B_{j0}A_{jj} \right), \tag{4.9}$$

$$\beta = B_{00} - \sum_{j=1}^{N} B_{0j}B_{j0}, \tag{4.10}$$

$$\gamma_j = \left(A_{0j} - A_{jj}B_{0j} \right)\left(A_{j0} - A_{jj}B_{j0} \right), \; j = 1, \ldots, N, \tag{4.11}$$

$$\delta_j = A_{jj}, \; j = 1, \ldots, N. \tag{4.12}$$

The ansatz (4.7) provides the $5N + 2$ parameters

$$\{A_{j0}\}_{j=1}^{N} \cup \{A_{0j}\}_{j=1}^{N} \cup \{A_{jj}\}_{j=0}^{N} \cup \{B_{j0}\}_{j=0}^{N} \cup \{B_{0j}\}_{j=1}^{N}$$

to determine merely the $2N + 2$ constants

$$\{\alpha\} \cup \{\beta\} \cup \{\gamma_j\}_{j=1}^{N} \cup \{\delta_j\}_{j=1}^{N}, \tag{4.13}$$

which indicates that some of these parameters should be redundant. Setting $A_{0j} = A_{j0}$ and $B_{0j} = B_{j0}$ for $j = 1, \ldots, N$ removes $2N$ parameters and still allows to fulfill (4.9)-(4.12) for given constants (4.13). Indeed, (4.12) can clearly be fulfilled with the reduced set of parameters since the diagonal entries are left untouched. Equation (4.11) then becomes

$$\gamma_j = \left(A_{0j} - \delta_j B_{0j} \right)^2,$$

which can e.g. be satisfied by setting $B_{0j} = 1$ and $A_{0j} = \delta_j + \sqrt{\gamma_j}$ for $j = 1, \ldots, N$, where $\sqrt{\gamma_j}$ is a square root for γ_j. We are left with two free parameters A_{00} and B_{00} which suffices to fulfill the remaining two equations (4.9) and (4.10).

This argument[2] justifies to solve the minimization problem (4.3) over a space of matrices of the form

$$
A = \begin{bmatrix} A_{00} & A_{01} & \cdots & \cdots & A_{0N} \\ A_{01} & \ddots & & \mathbf{0} & \\ \vdots & & \ddots & & \\ \vdots & \mathbf{0} & & \ddots & \\ A_{0N} & & & & A_{NN} \end{bmatrix}, \quad
B = \begin{bmatrix} B_{00} & B_{01} & \cdots & \cdots & B_{0N} \\ B_{01} & 1 & & \mathbf{0} & \\ \vdots & & \ddots & & \\ \vdots & \mathbf{0} & & \ddots & \\ B_{0N} & & & & 1 \end{bmatrix}.
$$

This has numerous benefits as already shortly touched on above. Firstly, reducing the number of free parameters from $2(N + 1)^2$ to $3N + 2$ reduces the cost of each iteration within the minimization routine. Secondly, the linear system $A \otimes M + B \otimes K$ will be much sparser compared to using dense matrices A and B. Additionally, the matrices A and B in the reduced ansatz are symmetric, which implies that the entire exterior system will be symmetric provided M and K are (as is usually the case). This is beneficial for computational efficiency since it allows the use of symmetric storage formats and specialized solvers for symmetric linear systems. Moreover, the variational formulations at the continuous level are often symmetric, and it is desirable to preserve such properties in their corresponding discretizations.

Numerical investigation

To justify the previously introduced reduction numerically, the following experiment is considered. The analytic DtN numbers for the Helmholtz equation in the homogeneous exterior of a ball with $d = 2$ are obtained by evaluating the function given in (2.28) at $\lambda = \lambda_\ell = (\ell/a)^2$. These numbers $\mathit{dtn}^{\text{hom,2D}}(\lambda_\ell)$ with $a = 1$ and $k = 16$ are used as reference values for solving the minimization problem (4.2)-(4.3). If the diagonalization is justified, then the reduced ansatz should lead the to same results as the dense approach. For this test case exponentially decaying weights $\omega_\ell \sim \exp(-2\ell/3)$ are chosen. In Fig. 4.1a the relative errors

$$
\frac{|\mathit{dtn}^{\text{hom,2D}}(\lambda_\ell) - \text{dtn}_N(\lambda_\ell)|}{|\mathit{dtn}^{\text{hom,2D}}(\lambda_\ell)|}
$$

for the dense and reduced ansatz are compared for $N \in \{0, 2, 4, 6\}$. Apparently, both approaches yield the same results. Additionally, Table 4.1 compares the performance of the Levenberg–Marquardt algorithm for both approaches. The final cost, i.e. the value of the objective function (4.2) at the minimizers, turns out to be the same for both approaches up to machine precision. The total time spent in the solver for the reduced ansatz is at least one order of magnitude lower than for the full ansatz. Due to its numerous benefits, the reduced symmetric ansatz is in practice the method of choice. It is used in all numerical experiments in this thesis unless explicitly stated otherwise.

---

[2] The argument also suggests to set $B_{0j} = 1$ for $j = 1, \ldots, N$. However, this does not promise any additional benefit with respect to the solution of the linear systems. Therefore, we prefer to leave the parameters B_{0j} for $j = 1, \ldots, N$ free instead of fixing them because in our experience the optimization routine tends to perform better when allowed to use slightly more parameters than the bare minimum.
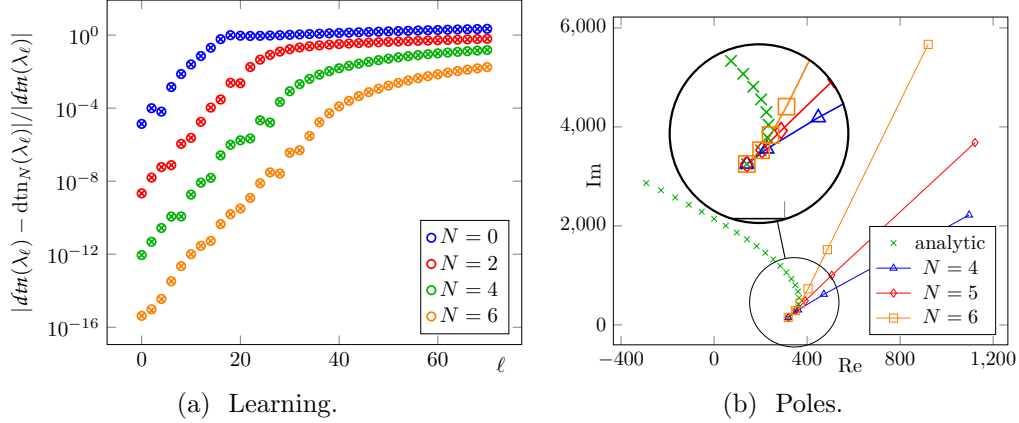
(a) Learning. (b) Poles.

Figure 4.1: Left: Comparison of the relative error in terms of N for solving the minimization problem for the analytic $dtn(\lambda_\ell)$ numbers from (2.28). The circles 'o' display the results for the dense ansatz of A and B while the crosses 'x' belong to the reduced ansatz. Note that all crosses lie perfectly inside the circles. For better illustration only results for even ℓ are shown. Right: Poles for the diagonal ansatz when learning the $dtn(\lambda_\ell)$ numbers from (2.28).

4.1.2 Pole structure

In the reduced symmetric ansatz dtn_N is a rational function which is asymptotically linear and has a finite number of simple poles. This can be seen from the formula

$$\text{dtn}_N(\lambda) = A_{00} + \lambda B_{00} - \sum_{j=1}^{N} \frac{(A_{0j} + \lambda B_{0j})^2}{A_{jj} + \lambda}. \tag{4.14}$$

The second term is a rational function with simple poles at

$$\lambda_j^*(N) = -A_{jj}, \tag{4.15}$$

provided $A_{0j} \neq B_{0j}A_{jj}$ at $j = 1, \ldots, N$ holds. From the results of Section 3.2 it is known that functions of the form (4.14) are sufficient to guarantee exponential convergence rates on finite intervals. However, we cannot exclude the possibility that the ansatz using dense matrices A and B could yield better approximation rates in some cases because it allows for poles with higher multiplicities. For such situations the diagonalizability assumption (4.5) would break down and the equivalence between both approaches is lost. Naturally, one would expect that the ability of dtn_N to represent poles with higher multiplicities could be an advantage if dtn itself features such poles. So far, all dtn functions we analyzed had exclusively simple poles, though[3]

The exact poles of the meromorphic extension of $dtn^{\text{hom,2D}}$ from the experiment of Section 4.1.1 are displayed in Fig. 4.1b. They are located in the quadrant $\{z \in \mathbb{C} \mid \text{Re}(z) > 0, \text{Im}(z) > 0\}$. Additionally, the learned poles (4.15) for $N \in \{4, 5, 6\}$ are shown. It can be observed that some of the learned poles seem to converge to the exact poles, particularly to the ones with the smallest imaginary part, which will surely exert the most influence on the restriction of $dtn^{\text{hom,2D}}$ to the positive real line. This phenomenon has

---

[3]Note also that even for dtn function with multiple poles the results of Section 3.2 guarantee exponential convergence rates on finite intervals with the ansatz (4.14) based on simple poles.

	Reduced symmetric			Full		
N	# iter	cost	time (s)	# iter	cost	time (s)
0	4	$8.26 \cdot 10^5$	$2.15 \cdot 10^{-4}$	4	$8.26 \cdot 10^5$	$5.12 \cdot 10^{-4}$
1	40	$1.31 \cdot 10^2$	$2.41 \cdot 10^{-3}$	36	$1.31 \cdot 10^2$	$2.67 \cdot 10^{-2}$
2	16	$6.14 \cdot 10^{-2}$	$1.58 \cdot 10^{-3}$	12	$6.14 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$
3	22	$2.95 \cdot 10^{-5}$	$3.23 \cdot 10^{-3}$	12	$2.95 \cdot 10^{-5}$	$5.34 \cdot 10^{-2}$
4	171	$1.44 \cdot 10^{-8}$	$4.13 \cdot 10^{-2}$	13	$1.44 \cdot 10^{-8}$	$1.37 \cdot 10^{-1}$
5	256	$7.22 \cdot 10^{-12}$	$6.87 \cdot 10^{-2}$	246	$7.22 \cdot 10^{-12}$	$4.81 \cdot 10^0$
6	469	$3.74 \cdot 10^{-15}$	$1.71 \cdot 10^{-1}$	184	$3.73 \cdot 10^{-15}$	$6.41 \cdot 10^0$

Table 4.1: Performance of the optimization routine for the two different ansatzes. Here, '#iter' denotes the number of performed Levenberg-Marquardt iterations, 'cost' the value of the objective function at the minimizer and 'time' the total time spent in the solver in seconds. The tolerances, which determine the stopping criterion for the iteration, were all set to machine precision.

already been observed in Section 3.4 for several other transparent boundary conditions of tensor-product structure. It also seems natural in view of Proposition 2.4 (iii), which has shown that $dtn^{\mathrm{hom,2D}}$ indeed admits an expansion into a sum of simple poles. However, simply approximating $dtn^{\mathrm{hom,2D}}$ by a sum of the first N simple poles merely leads to a rate of $\mathcal{O}(N^{-1})$ as shown Fig. 2.3 while learned IEs shown in Fig. 4.1a achieve exponential convergence. This is accomplished by only fitting the poles closest to the imaginary axis and placing the other free poles elsewhere. In the process of this chapter other dtn functions will be encountered which possess poles extremely close to the real axis. Fitting these poles will then turn out to be the key for obtaining a good approximation of dtn.

Remark 4.2. *It is interesting to discuss whether the matrices A and B in the reduced ansatz could also be chosen as hermitian. However, this is not a good idea because all poles of the correponding dtn_N function would lie on the real axis. This follows immediately from formula (4.15) since the diagonal entries of hermitian matrices are real.*

4.1.3 Condition number

Classical infinite elements suffer from poor condition numbers as N increases. In fact, the results from Section 3.4 have shown that exponential growth of the condition number can already be observed in the matrix $A_{EE} + \lambda B_{EE}$. This may lead to numerical instabilities and impair the solution of the arising linear systems. Therefore, it should be investigated whether learned IEs run into similar issues.

The condition number of the matrix $A_{EE} + \lambda B_{EE}$ for the learned IEs obtained from the numerical experiment of Section 4.1.1 is shown in Fig. 4.2. Apparently, the condition number for the reduced symmetric ansatz behaves more favourably than for the full ansatz. Since most exterior DOFs in the former ansatz have been eliminated in the reduction step, the condition number is simply given by

$$\mathrm{cond}(A_{EE} + \lambda B_{EE}) = \frac{\max_j |A_{jj} + \lambda|}{\min_j |A_{jj} + \lambda|}. \tag{4.16}$$

This should stay bounded unless one of the learned poles (4.15) lies exactly on one of the eigenvalues $\lambda = \lambda_\ell$. Since the learned poles tend to converge to those poles of dtn

that lie closest to the real axis, this case is expected to occur only in situations when *dtn* exhibits poles with an extremely small imaginary part. Further condition number studies shall be presented in the course of this chapter when such *dtn* functions are encountered. As an intermediate result we may conclude that the results of this subsection add to the merits of the reduced ansatz in the sense that the decreased number of DOFs may be beneficial for conditioning of the resulting linear systems. In any case it should render possible ill-conditioning problems easier to analyze.
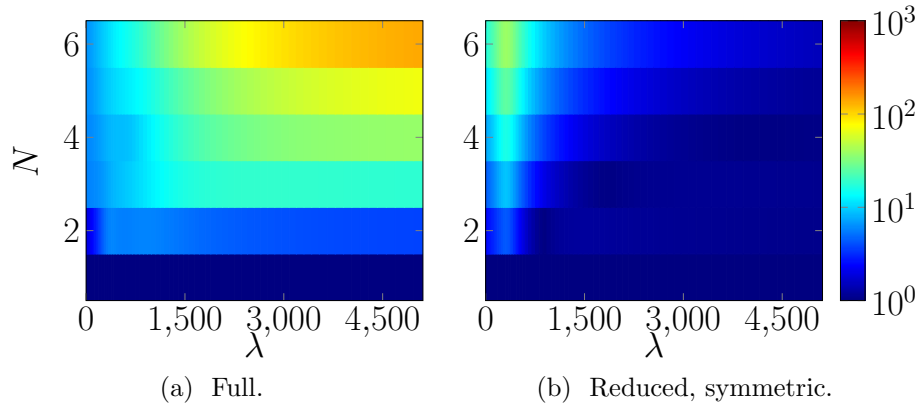

(a) Full. (b) Reduced, symmetric.

Figure 4.2: Condition number of the matrix $A_{EE} + \lambda B_{EE}$ for the two different ansatzes for learned IEs for $k = 16, a = 1$.

4.1.4   Computational costs

Learned IEs follow the same algebraic structure (3.7) as many other transparent boundary conditions of tensor-product type, e.g. tensor-product PMLs, HSIEs and classical infinite elements, see Section 3.4. Therefore, the same techniques for solving the arising linear systems can be applied. Notice in this regard that the exterior system (3.7) always remains sparse even if A and B are dense thanks to sparsity of the finite element mass and stiffness matrices M and K. Consequently, sparse direct solvers are readily applicable. The reduction step from Section 4.1.1 increases the sparsity even further and yields an additional performance boost. In applications learned IEs are usually combined with a FEM discretization of the interior problem, represented by the sesquilinear form $b_{\text{int}}(\cdot, \cdot)$, which suffers from the usual issues arising in time-harmonic wave propagation, see Section 7.1, and leads to extremely large linear systems. Direct solvers may reach their limits as the problem size increases so that iterative solution approaches combined with preconditioning techniques have to be considered. The benefits of using learned IEs in this context are extensively investigated in Chapter 7 and Chapter 8. To sum up, the solution of the linear systems arising from learned IEs as transparent boundary conditions does not pose a special obstacle. On the contrary, we will see in Section 4.2.2 that learned IEs lead to significantly sparser system matrices than many other popular transparent boundary conditions.

Learned IEs differ from other transparent boundary conditions of tensor-product type only in how the matrices A and B in the ansatz (3.7) are obtained. Whereas the matrix entries in conventional methods are described by analytic formulas, see e.g. (C.6) for tensor-product PMLs or (C.15)-(C.16) for certain classical infinite elements, learned IEs determine A and B by solving the minimization problem (4.2)-(4.3). The minimization

procedure allows for a seamless adaptation to different problem settings as the results of this chapter will demonstrate, yet comes at the expense of a slight increase in computational complexity compared to conventional approaches. These additional costs split into the following two components.

- To set up the objective function (4.2), the $dtn(\lambda_\ell)$ numbers for $\ell < L$ have to be determined. Often, analytic formulas for dtn are known, see e.g. (2.28) for $dtn^{\text{hom,2d}}$, in which case this entails effectively no additional costs except for the evaluation of some special functions. In the more general setting, the numbers $dtn(\lambda_\ell)$ for $\ell < L$ are obtained by solving the ODEs (2.12a)-(2.12b). This is an embarassingly parallel task since the values for different ℓ's are completely independent. Usually it is sufficient to take L on the order of k^2 so that more than a thousand ODE solutions are seldom required. Finally, we would like to emphasize that regardless of the dimension d of the space in which the PDE problem (2.2) is posed the computation of $dtn(\lambda_\ell)$ always comes down to solving merely ODEs, i.e. $d = 1$ dimensional problems. Therefore, the associated costs will remain small in any dimension.

- Likewise, the effort for solving the minimization problem (4.2)-(4.3) is independent of the problem dimension d. It depends on how many of the entries of the matrices $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ appear as free parameters in the minimization problem. Owing to the reduction step of Section 4.1.1, this number grows linearly with N whereas the approximation error $\text{err}_{dtn}(N, \ell)$ decreases exponentially fast as has already been seen in Fig. 4.1a and will be confirmed in many other examples presented in this chapter. Hence, $N < 10$ is usually sufficient and the time spent for solving the optimization problem to achieve accuracies required in practical simulations should remain on the order of a few seconds (see Table 4.1 for an example of computation times).

Both contributions to the computational overhead are independent of the problem dimension d. In contrast, the final complexity of finite element simulations in dimension $d = 3$ is usually completely dominated by the effort for solving the linear systems which scales[4] harshly with d. Therefore, the additional overhead for solving the minimization problem should usually be redeemed by the advantage that learned IEs lead to smaller and sparser system matrices compared to other transparent boundary conditions, see for example Section 4.2.2.

4.2 Helmholtz equation for spherical geometries in two dimensions

This section presents numerical experiments for circular truncation boundaries. The simplest setting considered in Section 4.2.1 and Section 4.2.2 is represented by wave propagation in a homogeneous medium. More precisely, Section 4.2.1 deals with scattering of a plane wave from a circular obstacle and Section 4.2.2 with waves emanating from a point source. For the latter example, the accuracy and efficiency of learned IEs will be evaluated by a comparison with various other popular transparent boundary conditions.

---

[4]Here we are referring mainly to time-harmonic wave equations which are notoriously challenging to so solve efficiently, see the discussion in Section 7.1.

In Section 4.2.3 the assumption of a homogeneous exterior medium is dropped so that the wavenumber in the exterior domain is allowed to have a jump, see the example introduced in Section 2.2.1.

4.2.1 Scattering of plane wave from a disk

Consider a plane wave $g(x) = -\exp(ikx)$ which is incident on a disk with radius $R_s = 1/2$. For sound-soft scattering, i.e. $u + g = 0$ at $r = R_s$, the solution of the Helmholtz equation is given by (see the Appendix E for a derivation)

$$u(r, \varphi) = \frac{H_0^{(1)}(kr)}{H_0^{(1)}(kR_s)} J_0(kR_s) + \sum_{\ell=1}^{\infty} \frac{H_\ell^{(1)}(kr)}{H_\ell^{(1)}(kR_s)} 2i^\ell J_\ell(kR_s) \cos(\ell\varphi). \tag{4.17}$$

The problem is discretized on an annulus $\Omega_{\text{int}} = \{x \in \mathbb{R}^2 \mid R_s \leq \|x\| \leq a\}$ using the Dirichlet boundary condition at $r = R_s$ and learned IEs at $r = a$. The resolution is increased by raising the polynomial degree while the mesh remains fixed. In virtue of the disussion following Definition 4.1, the weights should be chosen as in (4.4). Since the medium for $R_s < r < a$ is free of sources, one should set $b = R_s$ in (4.4) so that $\omega_\ell \sim |H_\ell^{(1)}(ka)/H_\ell^{(1)}(kR_s)|$.

The relative error on Ω_{int} using $a = 1$ for increasing number N of infinite element DOFs is shown in Fig. 4.3. The convergence in N is extremely fast. For $p = 6$ and $k = 16$ the spatial accuracy is reached with only $N = 3$ DOFs as Fig. 4.3a demonstrates. Figure 4.3b shows that high accuracy is obtained for a wide range of wavenumbers[5].
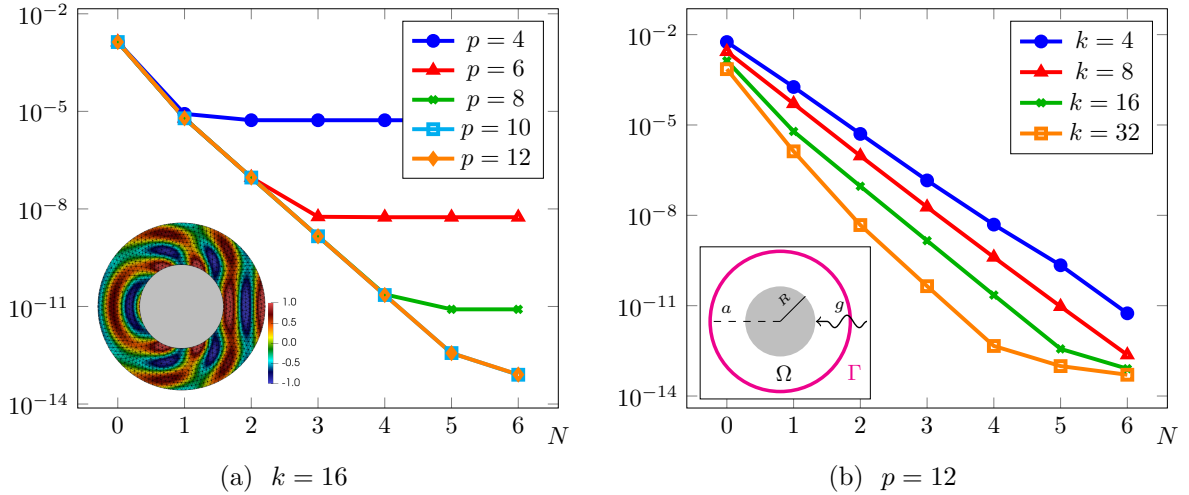

(a) $k = 16$ (b) $p = 12$

Figure 4.3: Comparison of relative error $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$ for the scattering of a plane wave from a disk of radius $R_s = 1/2$. The truncation boundary is placed at $a = 1$. Additionally, the real part of the reference solution for $k = 16$ is shown in the left figure.

The main parameter which determines the performance of most transparent boundary conditions is not the wavenumber alone but rather the number of wavelengths that fit between scatterer and the coupling boundary. The further away the coupling boundary, the more the highly oscillatory components of the solution have already decayed when

---

[5]Note that every single wavenumber k leads to its own pair of learned matrices A and B.

reaching it. This behavior is clearly visible in the factor

$$\exp\left(-\log\left(\frac{a}{b}\right)\frac{\lfloor aN^\alpha\rfloor}{4}\right)\|u\|_{H_\sigma(\Gamma)}, \quad b = R_s,$$

which describes the asymptotic contribution in the error estimate (3.48). Hence, when $|a - R_s|$ is large, it suffices to focus attention only on a couple of dtn numbers associated with slowly propagating modes, i.e. to fit $|dtn^{\text{hom}}(\lambda_\ell) - \text{dtn}(\lambda_\ell)|$ for small ℓ well, to obtain an accurate solution[6]. In this regard, the choice $|a - R_s| = 1/2$ considered for the previous experiment is rather generous. Therefore, an additional experiment is performed in which the coupling boundary is moved progressively closer to the scatterer. To this end, $R_s = 1/2$ is fixed and a is decreased so that $|a - R_s| \in [1/2, 1/4, 1/8, 1/16]$. It is interesting to compare the realistic choice of weights motivated above with the optimal choice $\omega_\ell = |\langle u, w_\ell \rangle_\Gamma|$ for which we have to assume that some oracle informed us in advance that the exact solution is given by (4.17).

(a) $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$ (b) $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$ (c) $\sqrt{2J(A,B)}$
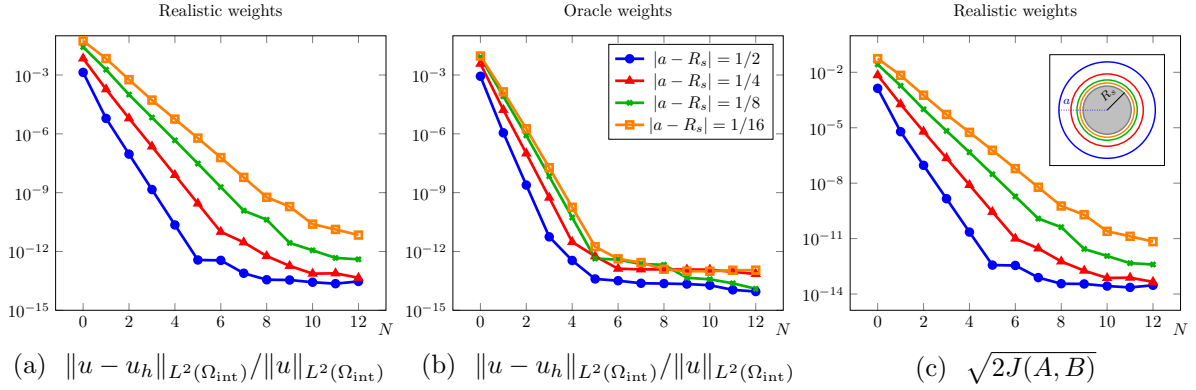
Figure 4.4: Influence of the distance $|a - R|$ between coupling boundary and scatterer on the performance of learned IEs for $k = 16$ and $p = 12$. Relative L^2-error on Ω_{int} on the left and center and value of the objective function on the right.

The results for $k = 16$ are shown in Fig. 4.4. Figure 4.4a, which displays the relative L^2-error for the realistically chosen weights, shows that the convergence rate slows down (even though it remains exponential) as the coupling boundary approaches the scatterer. This effect can be eliminated almost entirely by choosing the 'oracle weights' as Fig. 4.4b shows. A relative accuracy of at least 10^{-11} can then be achieved with $N = 5$ even for the case $|a - R_s| = 1/16$ in which the trunation boundary is extremely close to the scatterer. According to Theorem 3.20 one would also expect that the weighted ℓ^2-error of the dtn approximation given by $\sqrt{2J(A,B)}$ should be proportional to the L^2-error of the obtained solution. Figure 4.4c demonstrates that this is indeed the case. This provides an a priori estimate on the accuracy of learned IEs in the sense that the error stemming from the DtN approximation can be measured directly after the minimization problem (4.2)-(4.3) has been solved. Future work could aim to combine this estimate with an error estimator for the interior problem to derive a minimal choice of N which balances errors from the interior FEM discretization and approximation of the transparent boundary condition in cases where no exact solution is known.

---

[6]To account for this in the estimate (3.48), one would take α close to zero.

4.2.2 Point source inside unit disk

This subsection evaluates the performance of learned IEs by a comparison with other commonly used transparent boundary conditions. All of these conditions, except for an adaptive PML, fit into the tensor-product framework of Chapter 3 and are described in more detail in Appendix C.

Besides the use of plane waves, another convenient option to create an analytic reference solution fulfilling the radiation condition is to rely on the fundamental solution $\Phi(x, y) = (i/4)H_0^{(1)}(k\|x - y\|)$ of the Helmholtz equation. It fulfills

$$-\Delta_x\Phi(x, y) - k^2\Phi(x, y) = \delta(x - y) := \delta_y(x),$$

where δ_y is the Dirac distribution at $y \in \mathbb{R}^2$. For this experiment the point source y is placed inside a disk with radius $a = 1$. Here, we use $\nabla_x\Phi(x, y) \cdot \mathbf{n}(x)$ directly as Neumann data for the exterior problem realized e.g. by the learned IEs. In particular, there is no interior discretization. The quality of the solution is assessed by measuring the relative L^2-error in the Dirichlet data on Γ. The difficulty of the problem increases with shrinking distance of the source to the boundary as this adds more significant modes to the solution. Since the medium is free of source for $\|y\| < r \leq a$, the weights are chosen as $\omega_\ell \sim |H_\ell^{(1)}(ka)/H_\ell^{(1)}(k\|y\|)|$. To investigate the influence on the transparent boundary condition, the source positions $y = (0.5, 0.0)$ and $y = (0.95, 0.0)$ are considered in the experiment. Besides the reduced symmetric ansatz for the learned IEs, also results using the ansatz with fully populated matrices will be presented.

The following other transparent boundary conditions will serve as competitors for learned IEs.

- A tensor-product PML as described in Appendix C.1. For the experiments below a quadratic absorption coefficient $\sigma(t) = (C/k)(1/(\eta - a))((t - a)/(\eta - a))^2$ was used in (C.2). By trial and error the parameters $C = 40$ and $\eta - a = 0.02$ have been found to yield good results. The implementation uses finite elements of order four with uniform mesh refinements.

- Tuning of the PML and discretization parameters in the above approach is tedious and may lead to suboptimal results. A better strategy has been proposed in [CL06]. Following [CM98] the medium parameter σ_0 in $\sigma(t) = \sigma_0((r - a)/(\eta - a))^m, m \in \mathbb{N}$ and the thickness $\eta - a$ in the complex stretching of the PML are determined through an a posteriori error analysis. This is achieved by splitting the error into a finite element discretization error and a term describing the modeling error introduced by the PML (see [CL06, Theorem 3.1]). The contribution from the PML decreases with the exponentially decaying factor

$$\exp\left(-k\,\mathrm{Im}(\tilde{\eta})\left(1 - \frac{a^2}{|\tilde{\eta}|^2}\right)^{1/2}\right), \quad \tilde{\eta} = \eta + i\frac{\sigma_0}{m + 1}(\eta - a).$$

By choosing $m = 2$, $\sigma_0 = 4.5$, and $\eta = 2.5$ the PML error is of the order of the machine precision for the considered example. With the PML parameters being fixed, an adaptive mesh refinement strategy based on a standard residual error estimator is employed to reduce the discretization error. Note that this leads to an unstructured mesh. In particular, this method is not of the tensor product form (3.7). Whereas the mesh refinement is driven adaptively, the finite element order

still has to be chosen manually. After some experiments we ended up with elements of order six.

- As a last candidate, the HSIE as described in Appendix C.3, is considered. Some tuning has been carried out to determine a good choice of the parameter κ_0. For the experiments below $\kappa_0 = ak$ was used for $y = (0.5, 0)$ while a larger value of $\kappa_0 = 3ak$ turned out to be beneficial for $y = (0.95, 0)$.
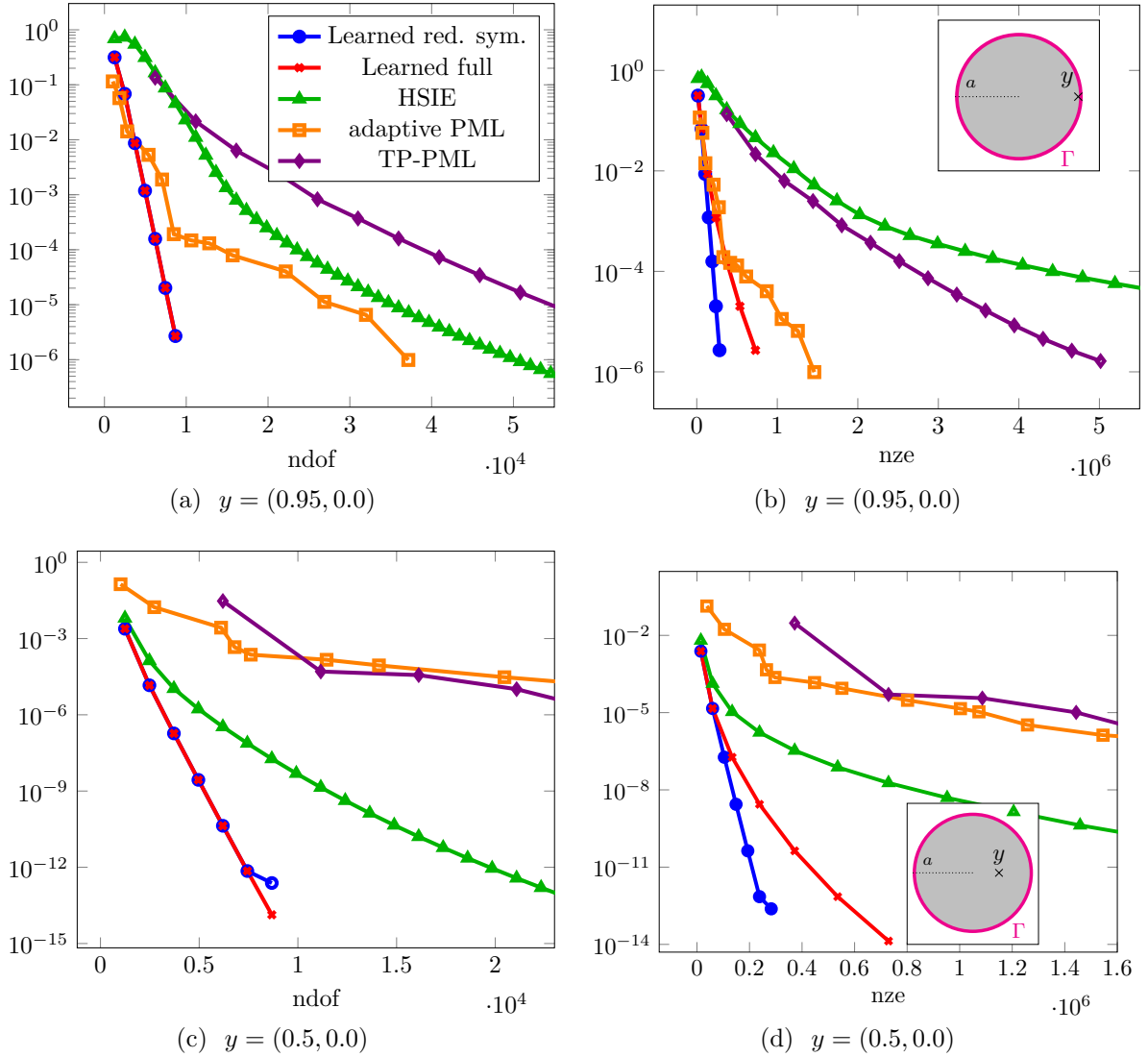

Figure 4.5: Comparison of relative error $\|u - u_h\|_{L^2(\Gamma)} / \|u\|_{L^2(\Gamma)}$ for the point source inside the unit disk at position y with $k = 16$.

In Fig. 4.5 the different methods are compared in terms of the number of DOFs ('ndof') and number of nonzero entries ('nze') of the resulting linear system.

- The lower panel displays the results for a source far away from the boundary. Here both methods based on infinite elements outperform the PML approaches. Although the HSIE already works very well for this problem, the learned approach yields a further improvement. The reduced symmetric ansatz for the learned IEs is noticeably more efficient than the one using full matrices in terms of the nonzero

matrix entries. However, in Fig. 4.5c it can be seen that both approaches can lead to different results in the region of extremely high accuracy (the blue line which respresents the reduced ansatz diverges from the red line line at the last data point). This does not stem from a violation of assumption (4.5). We checked that the eigenvalues of $(B_{EE})^{-1}A_{EE}$ for A and B from the full ansatz are all distinct. This observation should rather be attributed to the nature of local optimization routines which are always at risk of getting stuck in a local minimum. So far we observed such behavior in our numerical experiments only rarely when penetrating into the extremely high accuracy regime (relative error around 10^{-12} or less) which is seldom required in applications.

- The upper panel gives the results for the source close to the boundary. Here, the performance of the HSIE degrades strongly. Both PML approaches perform better in terms of nonzero entries of the linear system. Among the PML methods, the adaptive discretization achieves substantially better results than the tensor-product PML. In particular, it leads to very sparse matrices. Still, the performance of learned IEs is unmatched.

The other transparent boundary conditions featured in this experiment are among the most powerful methods currently known. A fair amount of parameter tuning has been performed to present all the competitors in their best shape. Despite these efforts, none of the other competitors was able to reach the performance of learned IEs. Therefore, the results of this subsection highlight the great potential of the new approach.

4.2.3 Jump in exterior wavespeed

In the previous examples a homogeneous exterior domain has been assumed. In view of the solar atmosphere, which is highly inhomogeneous, such an assumption is not realistic. Therefore, the following experiment allows for an exterior wavenumber which jumps from k_I to k_∞ at a radius of R_{J}. A full description of this example has been provided in Section 2.2.1. For the numerical experiments $a = 1$, $R_{\mathrm{J}} = 2$ and $k_I = 16$ are used. The weights are chosen in analogy to (4.4) as $\omega_\ell \sim |u_\ell(a)/u_\ell(R_{\mathrm{s}}k_I)|$, where $u_\ell(r) = A_\ell J_\ell(k_I r) + B_\ell Y_\ell(k_I r)$ is the solution of the radial equation for $r < R_{\mathrm{J}}$ (see Appendix E).

The computational results for this experiment are collected in Fig. 4.6. As explained in Remark 3.26, the other methods considered in the previous experiment are not readily applicable to this problem. Hence, only results obtained with learned IEs are presented. First the approximation of dtn^{jump} by dtn_N should be discussed. The upper left panel of the figure displays the case $k_\infty = k_I = 16$ in which there is no jump and the dtn^{hom} function of the homogeneous medium is recovered. Since dtn^{hom} is smooth, the fast convergence of the approximation by dtn_N shown in colors coded by N is as expected (cp. also Fig. 4.1a). For the case $k_\infty = 8$ displayed in the right panel dtn^{jump} is not smooth. It exhibits sharp peaks caused by poles located in very close proximity to the real axis as already discussed in Section 2.2.1. Despite the complicated behavior of dtn^{jump}, the approximation by dtn_N works extremely well. This can be explained by noticing that the learned poles as defined in (4.15) are placed successively at the positions of the exact poles. The middle panel of Fig. 4.6b visualizes this process. For $N = 1$ the first pole is reasonably approximated so that the first peak of dtn^{jump} can be represented by $\mathrm{dtn}_{N=1}$. Increasing N to three improves approximation of the first pole and covers the second pole as well. Consequently, the second peak of dtn^{jump} is well-represented by
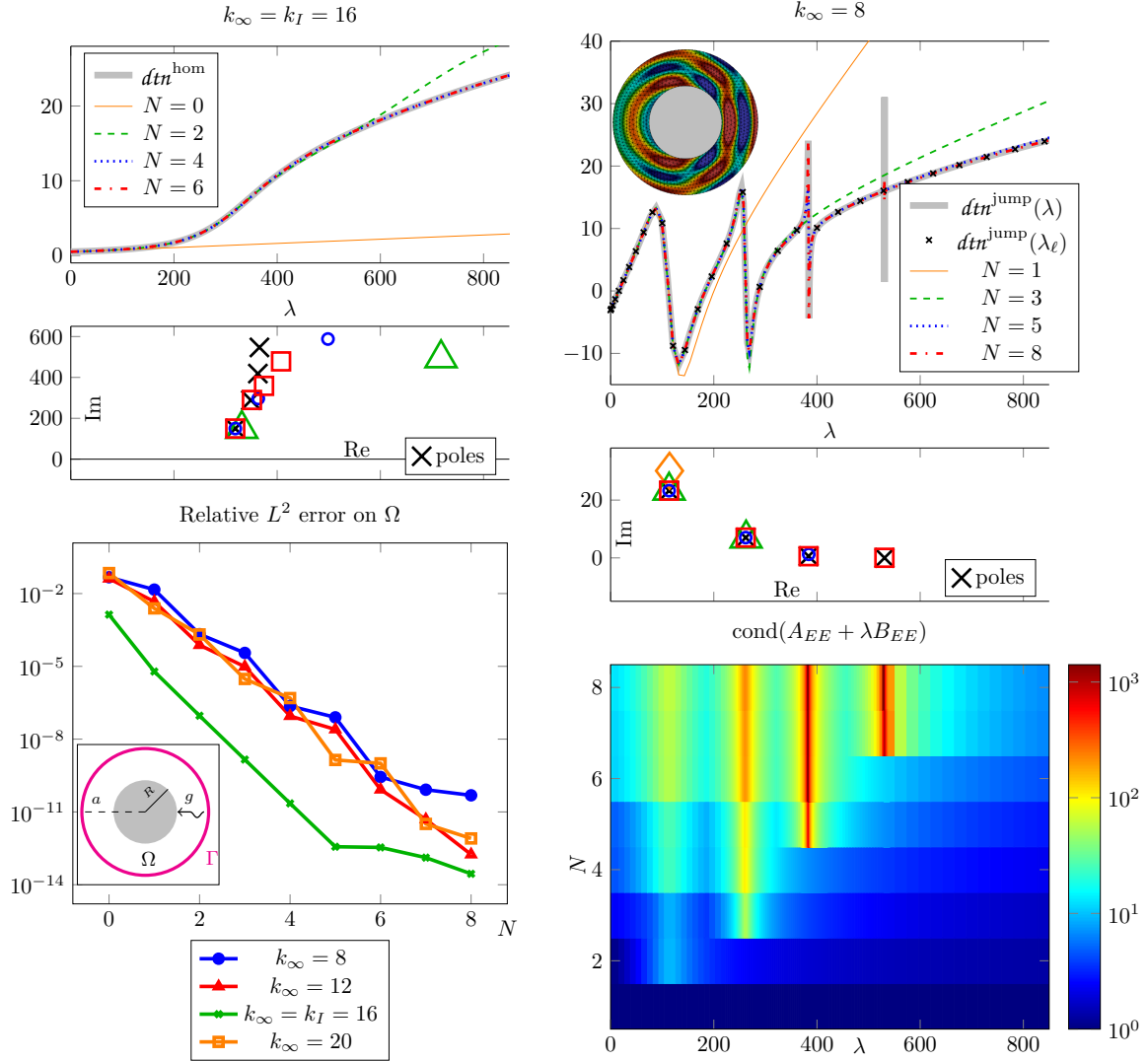
(a) Upper panel: $k_\infty = k_I = 16$, lower panel: relative error for different k_∞.

(b) $k_\infty = 8$.

Figure 4.6: Results for a discontinuous exterior wavespeed which jumps from $k_I = 16$ to k_∞. The upper panel shows the approximation results obtained with dtn_N for $k_\infty = k_I = 16$ on the left and $k_\infty = 8$ on the right. In the middle panel the poles of $\boldsymbol{dtn}$ (black crosses) are compared with the poles of dtn_N (color according to legend). Only the poles of dtn_N close to the real axis are displayed. The lower panel on the left shows the relative L^2 error for scattering of a plane wave from a disk for different k_∞. The lower panel on the right displays the condition number of the matrix $A_{EE} + \lambda B_{EE}$ for $k_\infty = 8$ in terms of N and λ.

the approximant. That the next two poles can be covered as well when N is increased further is astonishing because these are only very weakly visible at the sample points $\lambda = \lambda_\ell$ which are included into the objective function J. The magnitude of $\mathit{dtn}^{\text{jump}}$ at the last pole close to $\text{Re}(\lambda) = 530$ is not well-represented by the approximant with $N = 8$ though. It is reasonable to expect that this could be improved by including a finer sampling of the λ-axis into the objective function.

In Section 4.1.3 it was proposed to investigate the condition number of the exterior part $A_{EE} + \lambda B_{EE}$ for situations in which dtn admits poles with very small imaginary part. The $\mathit{dtn}^{\text{jump}}$ function for $k_\infty = 8$ is suited for such a study since its last two poles have imaginary part of about 0.59 and 0.0049 respectively. The condition number of the exterior part of the learned matrices is shown in the middle panel of Fig. 4.6b. The traces of the poles in the condition number are clearly visible. This is consistent with formula (4.16) and the dicussion in Section 4.1.3. As long as the exact poles have a non-vanishing imaginary part numerical instabilities should not occur since the condition numbers will stay bounded. However, in cases where dtn admits poles directly on the real axis or with extremely small imaginary part, say 10^{-13}, one has to be careful.

Finally, learned IEs are applied as a transparent boundary condition to compute the solution of a scattering problem. The setup is as in Section 4.2.1. A plane wave $g = -\exp(ik_\infty x)$ is incident on a disk with radius $R_{\text{s}} = 1/2$ so that the scattered field with sound-soft boundary conditions should be computed in the annulus $R_{\text{s}} \leq r \leq a$. An analytic reference solution for the problem, shown in the inset of Fig. 4.6b for $k_\infty = 8$, is derived in Appendix E. In case of no jump, i.e. $k_\infty = k_I := k$, it reduces to the solution of the homogeneous scattering problem considered in Section 4.2.1. The relative L^2-errors $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$ for different k_∞ are compared in the right lower panel of Fig. 4.6. Although the error is about three orders of magnitude larger when a jump actually occurs, i.e $k_\infty \neq k_I$, the convergence is nevertheless exponential. This further indicates the suitability of learned IEs for inhomogeneous exterior problems.

4.3 Helmholtz equation for elliptical geometries

This section continues the study of scattering problems. Now elliptical truncation boundaries and scatterers are considered. The foundations for these investigations at the continuous level have been provided in Section 2.2.2. Recall the notation a_Γ and b_Γ for the semi-major and semi-minor axis of the elliptical coupling boundary, respectively. For ease of implementation the scatterer will also be represented by an ellipse with semi-major axis a_S and semi-minor axis b_S. A sketch of the geometrical setup is provided as a small inset in Fig. 4.7d.

An analytic reference solution for convergence studies is created similarly as in Section 4.2.2. Two point sources are placed inside the scatterer at positions $y_1 = (-0.1, -0.13)$ and $y_2 = (0.1, 0.15)$. The reference solution is then given by

$$u(x) = \sum_{j=1}^{2} (i/4) H_0^{(1)}(k\|x - y_j\|).$$

The Neumann trace of this reference solution is used as data for a sound-hard scattering problem at wavenumber $k = 32$. Plots of the real part of the reference solution are shown in the insets of Fig. 4.7a-Fig. 4.7c for different geometrical configurations.
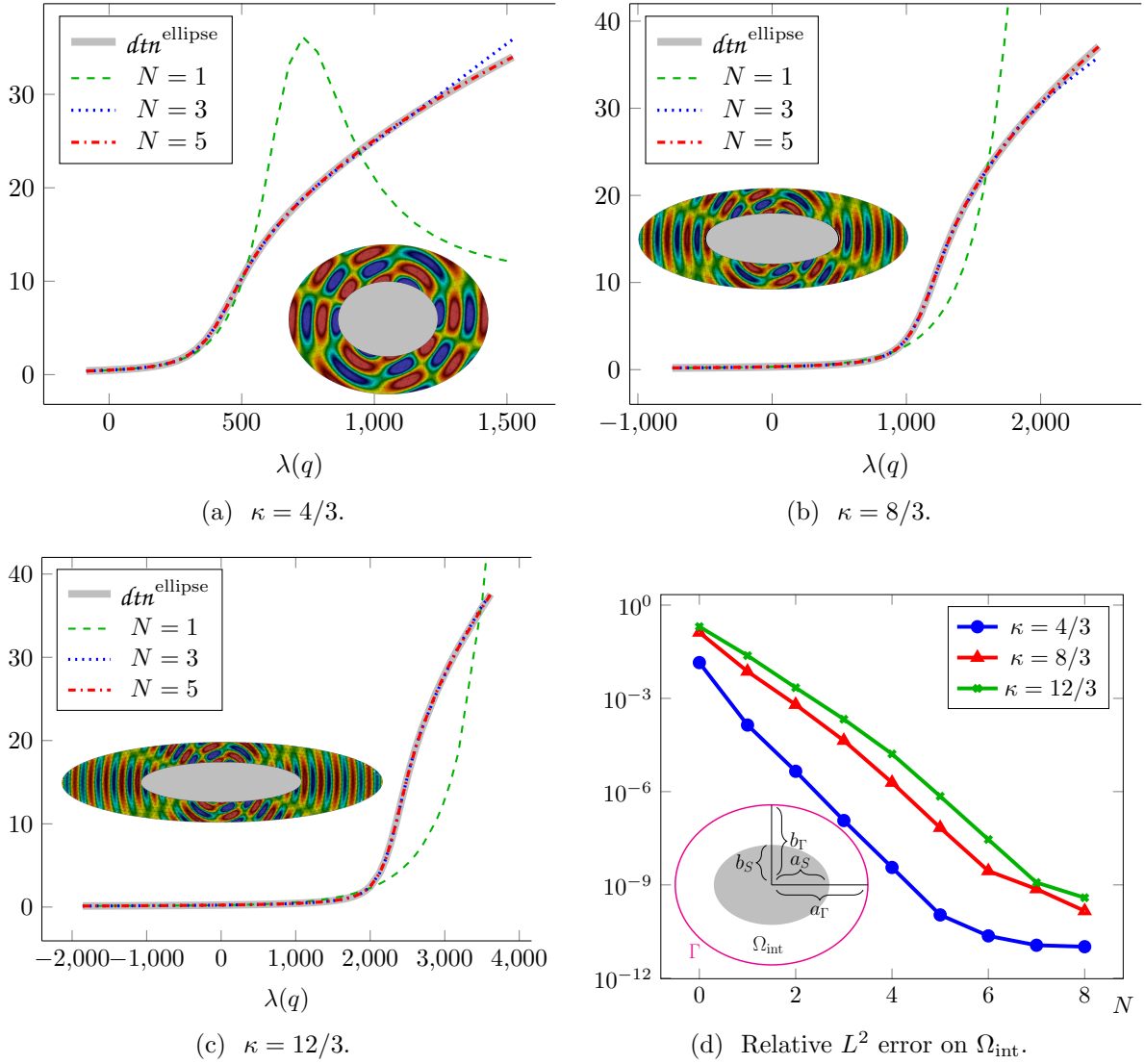
(a) $\kappa = 4/3$.

(b) $\kappa = 8/3$.

(c) $\kappa = 12/3$.

(d) Relative L^2 error on Ω_{int}.

Figure 4.7: Results for the ellipse with $k = 32$. The parameter κ denotes the ratio between semi-major and semi-minor axis.

The extension of learned IEs to elliptical truncation boundaries is mainly motivated by the desire to treat very elongated scattering objects efficiently. Otherwise, one could simply resort to using circular truncation boundaries. Therefore, it makes sense to investigate if situations in which the ratio

$$\kappa = a_\Gamma/b_\Gamma = a_S/b_S, \tag{4.18}$$

between semi-major and semi-minor axis is large can be treated reliably. For the experiments $b_s = 1/4$ and $b_\Gamma = 1/2$ will be fixed while the semi-major axis are streched by increasing κ. The weights are simply chosen as $\omega_\ell \sim \exp(-\ell/(3\kappa))$.

The computational results for $\kappa \in \{4/3, 8/3, 12/3\}$ are shown in Fig. 4.7. Fig. 4.7a-Fig. 4.7c demonstrate that the approximation of dtn^{ellipse} by dtn_N converges rapidly. It is interesting to note that dtn^{ellipse} appears to develop a kink as κ increases. This does not seem to impair the approximation by dtn_N though, which is promising in view of the waveguide which will be treated in the Section 4.4. The relative L^2 error $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$ in terms of N is shown in Fig. 4.7d. Fast convergence is observed for all considered ratios κ. Admittedly, the rate seems to suffer slightly for very elongated objects, yet is nevertheless exponential. Future studies could investigate whether the dependence on κ could be removed by consulting an oracle for optimal weights as was done in Section 4.2.1.

4.4 Waveguides

In this section the waveguide introduced in Section 2.3 will be considered. Setting $\tilde{\Gamma} = [0, \pi]$ and letting the waveguide start at the origin the solution shall be computed in $\Omega_{\text{int}} = [0, a) \times \tilde{\Gamma}$. A sketch of the geometry is given in Fig. 2.6b. Using Dirichlet boundary conditions on $\partial\Omega_{\text{int}} \setminus (\{a\} \times \tilde{\Gamma})$ the exact solution is given by

$$u(x, y) = \sum_{\ell=0}^{L_0} \sin\left(y\sqrt{\lambda_\ell}\right) \exp\left(ix\sqrt{k^2 - \lambda_\ell}\right),$$

for $\lambda_\ell = \ell^2$. For the numerical experiment the parameters $a = 2\pi$, $k = 16.5$ and $L_0 = 33$ are set. Note that (2.49) is fulfilled. The weights are chosen to cover all propagating waveguide modes equally well and emulate the decay of the evanescent modes, i.e. $\omega_\ell \sim 1$ for $\lambda_\ell \leq k^2$ and $\omega_\ell \sim |\exp(ia\sqrt{k^2 - \lambda_\ell})|$ else.

In Fig. 4.8a results for the approximation of $\text{Im}(dtn^{\text{guide}})$ are shown. Note that for the waveguide the imaginary part is of primary interest since $\text{Re}(dtn^{\text{guide}})$ is zero for the most relevant modes. As one might expect, resolving the kink of dtn^{guide} is the most challenging aspect. Interestingly, the learned poles are observed to accumulate in the part of the complex plane lying to the right of the kink, which is probably beneficial for approximation. Since the approximation at the sample points $dtn^{\text{guide}}(\lambda_\ell)$ is extremely accurate for large N, exponentially fast convergence of the relative L^2-error as seen in Fig. 4.8b follows. According to the paper [New64] by Newman such fast convergence rates for rational approximation of functions involving kinks also seem to be reasonable. He proved in this paper that the function $|x|$ can be approximated uniformly on the interval $[-1, 1]$ by rational functions of order N with an error $\leq 3\exp(-\sqrt{N})$.
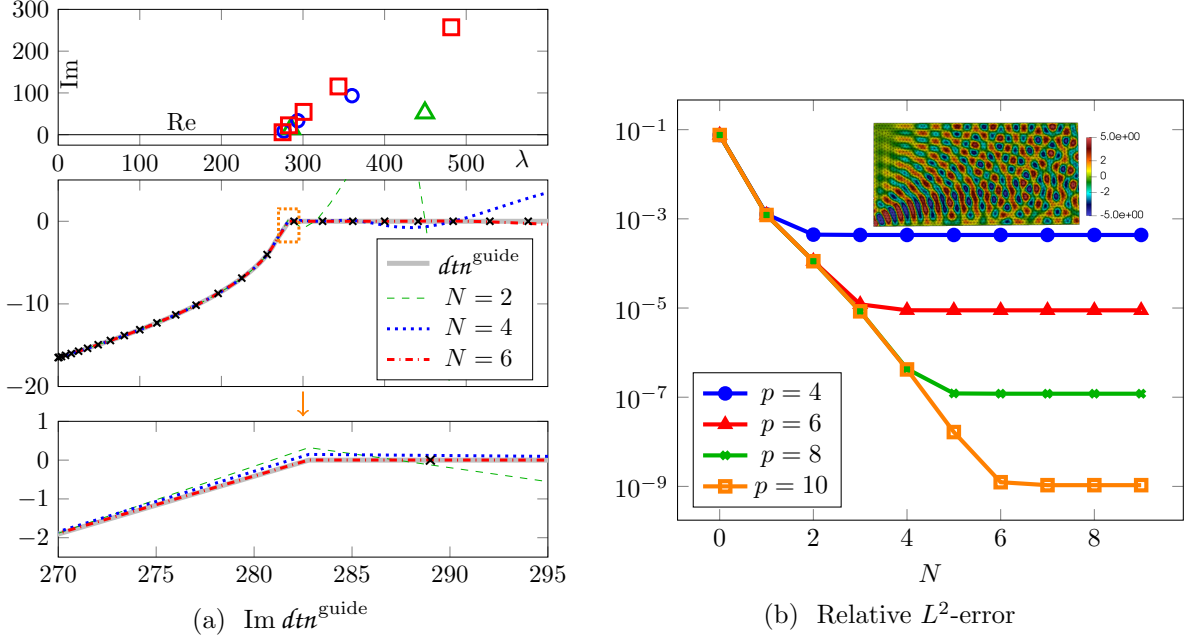
Figure 4.8: Left: Approximation results for imaginary part of dtn^{guide} with $k = 16.5$. The sample points $dtn^{\text{guide}}(\lambda_\ell)$ are displayed as black crosses. In the topmost plot also some poles of dtn_N are displayed. Right: Relative L^2-error on Ω_{int}. Additionally, the real part of the reference solution is shown.

4.5 Discussion and outlook on further chapters

In this chapter learned IEs have been introduced and tested for a variety of different problems. They deliver extremely accurate results for scattering problems involving circular and elliptical truncation boundaries. Also very good performance for a simple waveguide configuration has been achieved. A very attractive feature of learned IEs is their flexibility which allows them to adapt easily to inhomogeneities in the exterior wavespeed. As discussed in Section 4.1.4, the computational costs for setting up learned IEs are very reasonable. These results are promising and motivate to conduct further studies of learned IEs in the remainder of this thesis.

The following extensions and applications of learned IEs will be investigated.

- As already touched on in the introduction to this chapter, the matrices A and B obtained from solving the optimization problem (4.2)-(4.3) only work reliably for one specific wavenumber. For various applications, e.g. for solving resonance problems, it is necessary to obtain a uniform approximation in the wavenumber. This objective is pursued in Chapter 5.

- While scattering problems and waveguides are surely interesting for a broad audience, the main motivation for developing learned IEs stems from helioseismology. Chapter 6 seeks to find out whether learned IEs are appropriate for modeling the solar atmosphere, whose $dtn^{\text{VAL}-\text{C}}$ function may be regarded as the real world realization of dtn^{jump} from Section 4.2.3.

- As discussed in the introductory Section 1.4, the need for efficient approximation of $\mathcal{D}t\mathcal{N}$ maps also arises in sweeping preconditioners. Currently, it is customary to use PMLs for this purpose. However, this is known to lead to poor performance

if the medium contains an excessive amount of reflections e.g. caused by strong inhomogeneities of the wavespeed, see the discussion in Remark 3.26 and the numerical experiments in Section 7.3.1 and Section 7.3.2. As the results from this chapter indicate that learned IEs might be able to cope much better with reflections, it is natural to investigate their use in sweeping preconditioners. This is carried out in Chapter 7 for toy problems and in Chapter 8 for a realistic problem from helioseismology.

Chapter 5

Learned IEs providing uniform approximation in the wavenumber

In Chapter 4, learned IEs have been constructed which can be used for one specific wavenumber k. In various applications, for example full waveform inversion, the solution of many time-harmonic wave equations at different wavenumbers is requested, e.g. in order to probe the medium. To eliminate the necessity of rerunning the optimization routine for each wavenumber, an approximation of $\mathcal{DtN}$ which can be used uniformly for a fixed band of wavenumbers would be desirable. To this end, dtn needs to be considered as a function of k, e.g. for the homogeneous medium

$$\mathit{dtn}^{\mathrm{hom,2d}}(\lambda, k) = \frac{-k}{H^{(1)}_{a\sqrt{\lambda}}(ka)}(H^{(1)}_{a\sqrt{\lambda}})'(ka) \tag{5.1}$$

and a discrete approximation $\mathrm{dtn}_N(\lambda, k)$ has to be determined which delivers uniform approximation quality in λ *and* k. Clearly, the objective function (4.2) is unsuitable for this task because it only includes information from a single wavenumber.

Therefore, the essential step towards obtaining a uniform approximation in k is to find an appropriate generalization of the objective function. In Section 5.1 such a generalization is proposed. To check its performance, the scattering problem from Section 4.2.1 will be reconsidered in Section 5.2. This time the objective is to obtain accurate solutions for all wavenumbers in a bounded interval using a fixed set of learned IE matrices. This will already give a good indication on the effectiveness of the proposed approach.

On top of this, Section 5.3 presents first results for so called resonance problems. In a typical resonance problem one is not given a specific wavenumber but rather seeks to determine the wavenumbers k solving

$$-\Delta u = k^2 u \quad \text{in } \mathbb{R}^d \setminus K, \tag{5.2}$$

where K is a compact domain and $u \neq 0$ fulfills a homogeneous boundary condition on ∂K and a radiation condition at infinity. Wavenumbers k solving such problems are called resonances and are complex numbers with nonvanishing imaginary part[1]. In order to locate such resonances, the solution operator of the equation has to be well-approximated for all wavenumbers in a subset of the complex plane. This is even more

---

[1]Since for real wavenumbers the corresponding exterior problems with homogeneous Neumann or Dirichlet boundary conditions on the scatterer are uniquely solvable.

challenging than the case of intervals considered in Section 5.2. Promising results for a simple resonance problem will be presented.

The chapter concludes with a brief outlook towards further research in the direction of uniform approximation with learned IEs.

5.1 The extended optimization problem

It is useful to keep applications to resonance problems or time-domain simulations in mind when designing learned IEs with uniform approximation properties. At first sight, equation (5.2) represents a linear eigenvalue problem in k^2. It is desirable to preserve this linearity at the discrete level because linear eigenvalue problems are usually easier to solve than nonlinear ones. In this regard, the main obstacle is posed by the radiation condition. How can it be incorporated while preserving linearity of the eigenvalue problem?

To answer this question, it is helpful to consult the review of transparent boundary conditions of tensor-product type presented in Appendix C. Two among the considered approaches lead to linear eigenvalue problems, namely PMLs and HSIEs. Both generate exterior systems of the form

$$L = \left[A^{(1)} + g(k)A^{(2)} \right] \otimes M + B \otimes K, \tag{5.3}$$

with $g(k) = k^2$, where the matrices themselves are independent[2] of k (see Appendix C for explicit formulas for $A^{(1)}, A^{(2)}$ and B). Since PMLs and HSIEs are among the most powerful methods for solving resonance problems, it is natural to adopt the ansatz (5.3) for generalization of learned IEs. The corresponding dtn_N function is given by

$$\mathrm{dtn}_N(\lambda, k) := A_{\Gamma\Gamma}^{(1)} + g(k)A_{\Gamma\Gamma}^{(2)} + \lambda B_{\Gamma\Gamma} \tag{5.4}$$
$$- (A_{\Gamma E}^{(1)} + g(k)A_{\Gamma E}^{(2)} + \lambda B_{\Gamma E})(A_{EE}^{(1)} + g(k)A_{EE}^{(2)} + \lambda B_{EE})^{-1}(A_{E\Gamma}^{(1)} + g(k)A_{E\Gamma}^{(2)} + \lambda B_{E\Gamma}).$$

The idea is now to optimize $A^{(1)}, A^{(2)}, B$ so that $\mathrm{dtn}(\underline{\lambda}_\ell, k) \approx \mathit{dtn}(\lambda_\ell, k)$ for all wavenumbers k in a bounded set which has to be specified by the specific application.

Definition 5.1. *Let λ_ℓ be the generalized eigenvalues of equation (2.9). Denote by $\mathit{dtn}(\lambda_\ell, k)$ the complex numbers obtained from solving the ODEs (2.12a). Further, let $\mathrm{dtn}_N(\lambda_\ell, k)$ be the corresponding approximations provided by learned IEs as in equation (5.4). For positive weights $\omega_{\ell,k}$ we define the misfit function*

$$J(A^{(1)}, A^{(2)}, B) = \frac{1}{2} \sum_\ell \sum_k |\omega_{\ell,k}(\mathit{dtn}(\lambda_\ell, k) - \mathrm{dtn}(\lambda_\ell, k))|^2, \tag{5.5}$$

where the sum in k ranges over samples of k in a bounded region of the complex plane. The minimization problem is to find $A^{(1)}, A^{(2)}, B \in \mathbb{C}^{(N+1) \times (N+1)}$ so that

$$A^{(1)}, A^{(2)}, B \in \underset{A^{(1)}, A^{(2)}, B \in \mathbb{C}^{(N+1) \times (N+1)}}{\mathrm{argmin}} J(A^{(1)}, A^{(2)}, B). \tag{5.6}$$

For dense matrices $A^{(1)}, A^{(2)}$ and B the optimization problem contains $6(N+1)^2$ free[3] variables. Hence, already for $N = 12$ the treshold of over one thousand variables is exceeded. Moreover, the Jacobians are in general dense because most variables couple

---

[2]Provided that the scaling profile of the PML is independent of the frequency.

[3]Real and imaginary parts are considered as separate variables here.

with each other (see Appendix D.2 for details). Therefore, strategies to reduce the number of free variables similar to Section 4.1.1 should be discussed.

Numerical experiments in the spirit of Section 4.1.1 show that the accuracy is not impaired if $A^{(1)}, A^{(2)}$ and B are taken to be symmetric matrices. This nearly halves the number of free variables, yet the asymptotic complexity is still $\mathcal{O}(N^2)$. Complex symmetric matrices can also be beneficial for resonance problems because they allow for the use of adapted methods for solving the corresponding large scale eigenvalue problems, see e.g. [AH04].

To obtain linear scalings, a drastic reduction step as in Section 4.1.1 would be necessary. To this end, it is tempting to transfer the ansatz (4.14) to the present setting and postulate that discrete dtn_N functions of the form

$$\mathrm{dtn}_N(\lambda) = A_{00}^{(1)} + g(k)A_{00}^{(2)} + \lambda B_{00} - \sum_{j=1}^{N} \frac{(A_{0j}^{(1)} + g(k)A_{0j}^{(2)} + \lambda B_{0j})^2}{A_{jj}^{(1)} + g(k)A_{jj}^{(2)} + \lambda}, \qquad (5.7)$$

were already suitable to approximate $\mathit{dtn}(\lambda, k)$. However, with such an ansatz it is very difficult to fit the poles of $\mathit{dtn}(\lambda, k)$, whose positions depend on k, accurately. From the previous investigations it is well-known that this is essential for obtaining good approximations. The poles of dtn_N as in equation (5.7) are located at position $\lambda_j^*(N) = -A_{jj}^{(1)} - g(k)A_{jj}^{(2)}$ for $j = 1, \ldots, N$. In order to track the dependence of the poles on k, the function $g(k)$, which is technically already fixed, would need to be chosen correctly. To this end, analytic formulas for the trajectories of the poles w.r.t. k would be required. For $\mathit{dtn}^{\mathrm{hom},2D}$ asymptotic formulas for the poles have been derived in references [MK60] and [KRG63]. The latter also contains plots of their trajectories in the complex plane. These results could serve as a starting point for finding an appropriate choice for g. This could be an interesting research direction to follow. However, here we want to fix $g(k) = k^2$ to obtain linear eigenvalue problems for computing resonances.

The increased number of parameters in the more general ansatz (5.4) may allow the optimization to adjust better for the dependence of the poles on k. To illustrate this, let the generalization of assumption (4.5) be valid, i.e. there exists an invertible matrix $P(k)$ and a diagonal matrix $D(k)$ such that

$$(B_{EE})^{-1}\left(A_{EE}^{(1)} + g(k)A_{EE}^{(2)}\right) = P(k)D(k)P(k)^{-1}. \qquad (5.8)$$

Here, the dependence of the matrices $P(k)$ and $D(k)$ on k is not explicitly known. Plugging this assumption into (5.4) implies that the rational part of dtn_N is equivalent to

$$-\sum_{j=1}^{N} \frac{q_j\left(\lambda, g(k), P(k), A^{(1)}, A^{(2)}, B\right)}{D_{jj}(k) + \lambda},$$

where q_j are second order polynomials in λ. This means that even under assumption (5.8) the ansatzes (5.4) and (5.7) are not equivalent in general because the former allows for a more complex dependence of the poles on k. Therefore, the general ansatz (5.4) with symmetric matrices will be used for the numerical experiments of this chapter.

5.2   Scattering of plane wave from a disk

We revisit the example of a plane wave scattered by a disk which has already been considered in Section 4.2.1 for individual wavenumbers. This time a uniform approximation
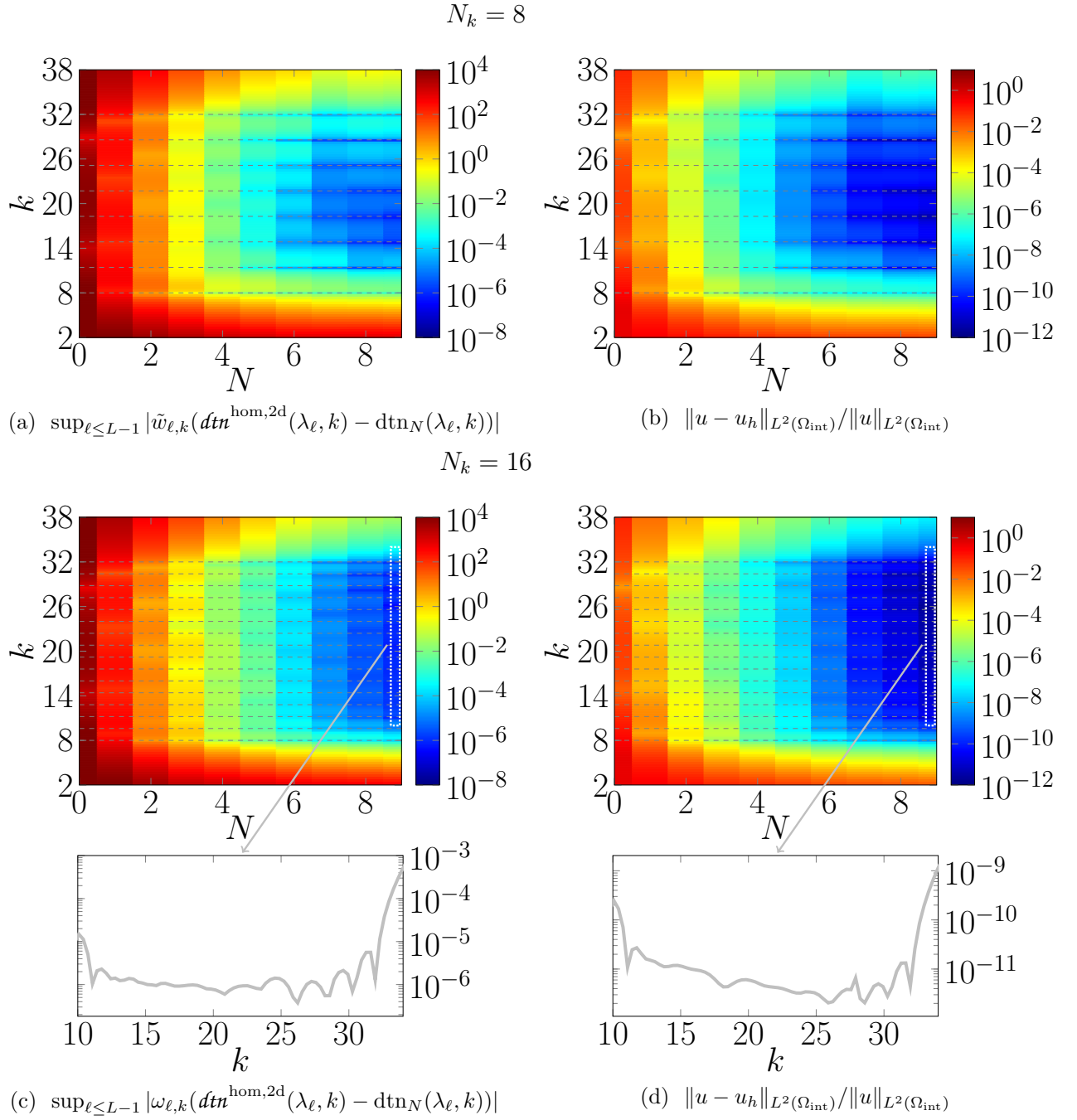
$N_k = 8$

(a) $\sup_{\ell \leq L-1} |\tilde{w}_{\ell,k}(\boldsymbol{dtn}^{\text{hom,2d}}(\lambda_\ell, k) - \text{dtn}_N(\lambda_\ell, k))|$

(b) $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$

$N_k = 16$

(c) $\sup_{\ell \leq L-1} |\omega_{\ell,k}(\boldsymbol{dtn}^{\text{hom,2d}}(\lambda_\ell, k) - \text{dtn}_N(\lambda_\ell, k))|$

(d) $\|u - u_h\|_{L^2(\Omega_{\text{int}})}/\|u\|_{L^2(\Omega_{\text{int}})}$

Figure 5.1: Results for scattering of a plane wave from a disk with learned IEs for intervals of wavenumbers. The upper panel was obtained with an equidistant sampling of 8 wavenumbers in $[8, 32]$. The lower panel uses a finer sampling of 16 wavenumbers. The wavenumbers which served as sample points have been indicated with dashed lines in gray. The lower panel in (c) and (d) displays the respective errors for $N = 9$.

for all wavenumbers $k \in [8, 32]$ is desired. To this end, N_k uniformly-spaced samples from this interval are included into the objective function (5.5). The weights are chosen as in Section 4.2.1, i.e. $\omega_{\ell,k} \sim |H_\ell^{(1)}(ka)/H_\ell^{(1)}(kR_s)|$. The disk has a radius of $R_s = 1/2$ and the truncation boundary is positioned at $a = 1$. Finite elements of order $p = 10$ are used for discretizing the interior problem. The linear systems are solved by means of a sparse Cholesky factorization.

The relative error on the interior domain and the supremum error on the modes $\sup_\ell |\omega_{\ell,k}(\mathit{dtn}^{\mathrm{hom},2d}(\lambda_\ell, k) - \mathrm{dtn}_N(\lambda_\ell, k))|$ are shown in Fig. 5.1. The upper panel displays the results for $N_k = 8$ while the sampling has been doubled for the lower panel. For the latter case an accuracy in the relative error of at least $2 \cdot 10^{-11}$ is achieved for all wavenumbers in the interval $[12, 32]$ with $N = 9$. With the optimization at individual wavenumbers (cp. Fig. 4.4a) a comparable accuracy has been achieved with $N = 4$ to $N = 5$ using the reduced symmetric ansatz. This shows that the number of additional DOFs required to achieve an approximation which can be used for intervals of wavenumbers is acceptable[4]. The left column of the figure suggests that the convergence of the L^2-errror may also be predicted from the supremum error of the dtn modes after the optimization problem has been solved. This is analogous to the optimization at single wavenumbers and suggests that the objective function is chosen appropriately. Note also that for $N_k = 8$ the position of the sample points can clearly be inferred from the position of the minima in the supremum error shown in Fig. 5.1a. Apparently, decreasing the sample size to $N_k = 8$ results in a loss of accuracy in the L^2-error. However, for coarser finite element discretizations this may already be sufficient to decrease the error stemming from truncation of the domain below the discretization error of the finite elements for the interior problem (cp. Fig. 4.3a).

5.3 Resonances of a disk

Since the ansatz (5.3) with $g(k) = k^2$ is linear in k^2, it can be utilized readily for the computation of resonances. To illustrate this, an example from the thesis [Nan08, Section 6.2] for which the exact resonances are analytically known will be considered. Let $K = \{x \in \mathbb{R}^2 \mid |x|_2 \leq R\}$ be a disk with radius R. The resonances k fulfilling

$$
\begin{aligned}
-\Delta u &= k^2 u && \text{in } \mathbb{R}^2 \setminus K, && \text{(5.9a)} \\
u &= 0 && \text{on } \partial K, && \text{(5.9b)} \\
\lim_{|x| \to \infty} |x|^{(d-1)/2} &\left(\frac{\partial u}{\partial |x|} - iku \right) = 0, &&&& \text{(5.9c)}
\end{aligned}
$$

are given by roots of the Hankel functions

$$
k \mapsto H_\ell^{(1)}(kR), \quad \ell = 0, 1, \ldots.
$$

The domain $\Omega = \mathbb{R}^2 \setminus K$ is decomposed as usual into $\overline{\Omega} = \overline{\Omega}_{\mathrm{int}} \cup \overline{\Omega}_{\mathrm{ext}}$, with $\Omega_{\mathrm{int}} := \{x \in \mathbb{R}^2 \mid R \leq |x|_2 \leq a\}$ and $\Omega_{\mathrm{ext}} := \{x \in \mathbb{R}^2 \mid |x|_2 > a\}$. The exterior domain Ω_{ext} is modelled using learned IEs. For the experiment the parameters $R = 1/2$ and $a = 6/10$ have been used. In contrast to the previous experiments, the sample points $k \mapsto \mathit{dtn}(\lambda_\ell, k)$ included into the optimization problem are now chosen from a box in the

---

[4]The number of additional DOFs can also be reduced by decreasing the size of the interval.

complex plane as displayed in Fig. 5.2. This generalization is necessary in order to obtain a good approximation of $dtn^{\mathrm{hom,2d}}(\lambda_\ell, k)$ in the vicinity of the resonance frequencies, which are located in the lower half plane. The weights $\omega_{\ell,k}$ are all chosen to be one. Thanks to the fact that the eigenvalue problem is linear in k^2, the resonances can be computed by the Arnoldi method for which a complex shift of $(10 - 5i)^2$ has been used.

Fig. 5.2 shows the results for the computation of the desired resonances using learned IEs. The panel in the upper left corner displays the value of the objective function in terms of N whereas the other panels show the numerical resonances computed with learned IEs using $N \in \{7, 13, 19\}$. We make the following observations.
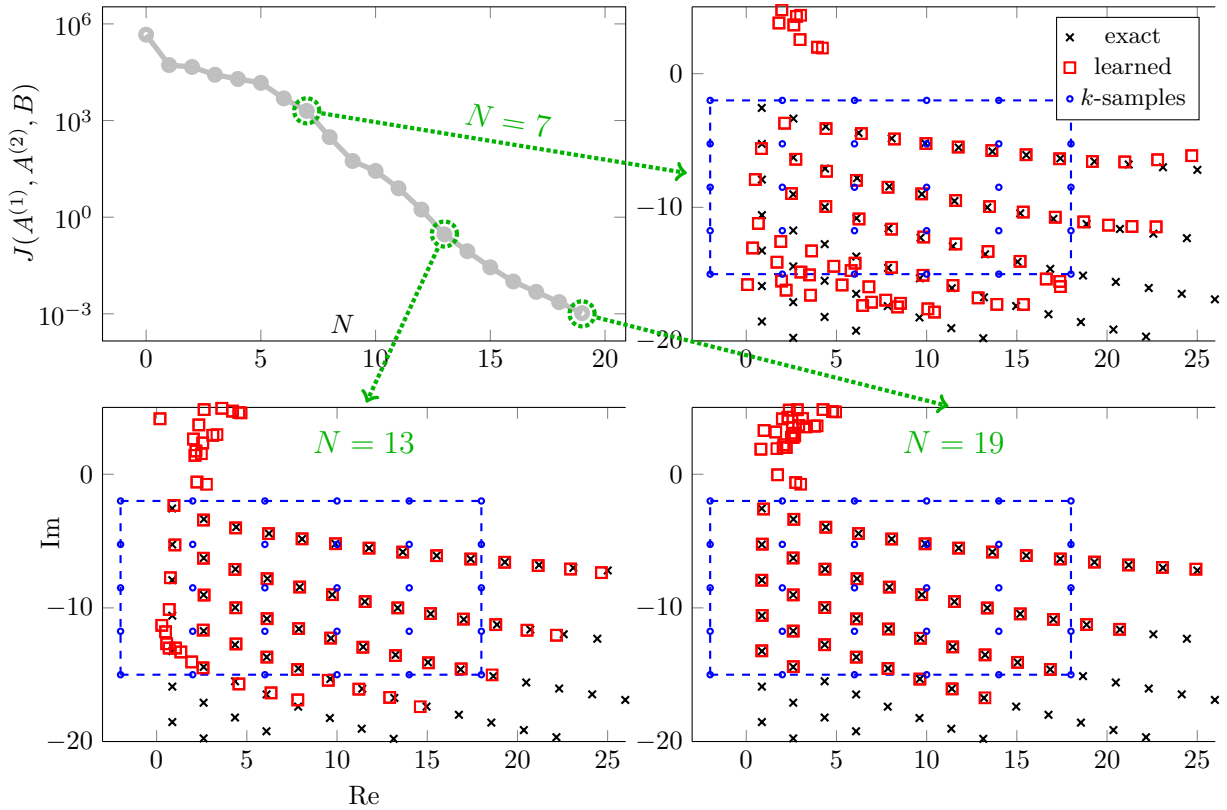

Figure 5.2: Resonances for a disk with radius $R = 1/2$ and Dirichlet boundary conditions. The sample points in k used for approximation of $\mathrm{dtn}(\underline{\lambda}_\ell, k) \approx dtn(\lambda_\ell, k)$ are indicated with blue circles. Exact resonances are shown as black crosses while approximate resonances computed with learned IEs are indicated by red boxes.

- For $N = 7$ only some of the resonances in the box are well-approximated. Apparently, the lower left corner of the box represents the trickiest area. Here, the computed resonances are observed to cluster without actually resolving the exact resonances accurately.

- For $N = 13$ the value of the objective function has decreased nearly by another four orders of magnitude and the approximation of the resonances has improved noticeably. Still a cluster of spurious resonances remains in the lower left corner of the box. Perhaps it could be beneficial to increase the number of sample points in this area.

- For $N = 19$ the value of the objective funcion has decreased by an additional two orders of magnitude. Now the blue box is free of spurious resonances and all exact resonances contained in this box have been found. Even some exact resonances outside the box have been detected. Moreover, most of the numerical resonances outside the box are easy to distinguish as numerical artifacts because they are located in the upper half plane. Therefore, in total only three numerical resonances remain (those located between the blue box and the real axis) which cannot be unambiguously identified as spurious without performing additional computations.

The good correspondence between the value of the objective function and the quality of the resonance approximation indicates that the choice of the objective function is reasonable. Certainly, the choice is not optimal yet as the clustering of spurious resonances in the lower left corner for $N = 13$ suggests. Nevertheless, this experiment demonstrates that computing resonances using learned IEs is possible and could be an interesting research direction to pursue.

5.4   Discussion and outlook

This chapter introduced an approach to compute learned IEs which work uniformly for all wavenumbers in some bounded subset of the complex plane. It is based on the idea to consider *dtn* additionally as a function of k and search for a discrete dtn$_N$ function which provides an accurate approximation of *dtn* for a whole range of k. The numerical results demonstrate that this can be achieved by a suitable modification of the objective function. Highly accurate results have been obtained for the scattering of a plane wave across a respectably large interval of wavenumbers. Moreover, an application of learned IEs to the computation of resonances has been presented. These results are promising and call for additional research in this direction.

- *Reduction step / efficiency:* When dense matrices $A^{(1)}, A^{(2)}, B$ are used in the objective function (5.5), the number of free parameters in the minimization problem grows quadratically, which becomes prohibitively expensive as N increases. This is particularly relevant for resonances problems which require large N and a sampling across many different wavenumbers which increases the number of residual blocks in the objective function. For this reason, learned IEs are not competive yet in terms of computational effort with more established methods like PMLs or HSIEs for computing resonances. To reduce the computational effort it will be necessary to sparsify the learned matrices. As disccued in Section 5.1, this reduction step is more intricate than for the case of individual frequencies. One approach would be to include a priori knowledge of the trajectories of the poles of *dtn* w.r.t. the frequency into the ansatz. Another more generic option is to investigate the suitability of learned matrices with a fixed bandwidth.

- *Reusability / accessibility:* An important aspect to consider in the design of a new numerical method is surely its ease of implementation and use by other researchers. In this regard, learned IEs as presented in Chapter 4 are not particularly attractive because they would require the implementation or installation of an optimization solver which can treat the corresponding minimization problem. However, for learned IEs as discussed in this chapter, libraries of learned matrices could be set

up which work for a specific problem across a whole range of wavenumbers, e.g. the Helmholtz equation in the exterior of a ball. Other researches could then simply download the corresponding matrices which would obliterate the necessity for a standalone optimization procedure.

- *Resonance problems:* The numerical experiment presented in Section 5.3 is of course only a first step into the considerably more complicated topic of computing resonances. More numerical experiments especially for problems involving inhomogeneous, yet separable exterior domains should be performed to explore the potential of learned IEs in this field.

- *Time-domain simulations:* In some applications it is requested to solve wave equations directly in the time domain. In this context, it would be desirable to transfer transparent boundary conditions developed in the frequency domain to the time domain. A prerequisite for this would be to have transparent boundary conditions at hand which cover intervals of wavenumbers and whose dependence on the wavenumber is explicitly known and not too complicated (preferably polynomial). In this regard, the ansatz (5.3) looks promising since its sole dependence on k is described by the function $g(k)$. Time-domain simulations will not be addressed here since this is beyond the scope of this thesis although the results of this chapter may encourage research in this direction.

Chapter 6

Modelling the solar atmosphere with learned IEs

Local helioseismology seeks to understand the two and three dimensional structure of the physical quantities in the interior of the Sun. To interpret the data provided by satellite or ground-based measurements, a forward problem has to be solved, which consists of computing synthetic data for a given solar model. This involves the solution of a (time-harmonic) wave equation posed in the solar interior. A suitable transparent boundary condition has to be imposed at the boundary of the computational domain to incorporate knowledge about the solar atmosphere into the simulation. Classical approaches for homogeneous media are not suitable for this purpose since the solar atmosphere is strongly inhomogeneous. This chapter investigates whether learned IEs are flexible enough to adapt to the harsh conditions in the Sun and could serve as accurate and reliable transparent boundary conditions for helioseismology.

For applications in time-distance helioseismology it is particularly important to understand the effect which truncation of the atmosphere by a transparent boundary condition exerts on the accuracy of helioseismic observables. This is easiest to analyze in a setting in which the separability assumption is extended to the entire Sun because this allows to compute the Green's function with the *exact* transparent boundary condition on a mode by mode basis. The latter can then serve as a reference for evaluating the accuracy of different (approximate) transparent boundary conditions which could be used also in the non-separable setting. Before proceeding to an evaluation of transparent boundary conditions, the question of the degree of accuracy that is a reasonable target to achieve for applications in helioseismology should first be addressed. Notice in this regard that even the Green's function computed with an exact transparent boundary condition provides an imperfect description of wave propagation in the Sun, e.g. because equation (2.52) is based on various approximations and the quantities of the solar background model do not capture the Sun perfectly. Therefore, we will start in Section 6.1 by comparing synthetic observables based on exact transparent boundary conditions for the `Atmo` and VAL-C model, respectively, with observables based on data from observations. The modelling errors observed there will set the scale for the accuracy transparent boundary conditions need to achieve. Afterwards, the accuracy of transparent boundary conditions for these two atmospheric models is assessed. Firstly, a comparison of learned IEs with established local transparent boundary conditions for the `Atmo` model is carried out in Section 6.2. In Section 6.3 the more involved case of the VAL-C model, for which no other transparent boundary conditions are available to date, is considered. The chapter concludes with a

short discussion of open questions and an outlook towards further research.

6.1 Observables in time-distance helioseismology

This section introduces the main observables of time-distance helioseismology in a specialized setting. The assumption of spherical symmetry already required for the atmosphere is extended to the entire Sun, which implies that equations (2.51) or (2.52) are fully separable. In particular, there is no background flow, i.e. $\mathbf{w} = 0$, and the coefficients are spherically symmetric functions. This will lead to a very natural framework for evaluating the accuracy of transparent boundary conditions for this application. The presentation of the observables in the fully separable setting considered in this chapter is mostly adopted from references [BFF+20b] and [BFF+20a]. For the case of a general or axisymmetric medium the reader is referred to [GBD+17]. Although this chapter considers exclusively the transformed equation (2.52), all results naturally extend to the original formulation since the corresponding *dtn* functions differ merely by a constant factor.

6.1.1 Green's function based on exact transparent b.c.

The main object to be computed is the Green's function of (2.52). To this end, it is useful for the implementation to scale the Sun to the unit ball by means of the coordinate transformation $\hat{x} = x/R_\odot$, respectively $\hat{r} = \|x\|/R_\odot$. Here, $R_\odot$ denotes the solar radius. The sound speed has to be scaled accordingly by $1/R_\odot$ and the source by $R_\odot^2$, see [BFF+20b, Section 3] for details. Transformed versions of the quantities ρ, c, γ and σ, which were introduced in Section 2.4, are defined by

$$\hat{\rho}(\hat{r}) = \rho(\hat{r}R_\odot), \quad \hat{c}(\hat{r}) = \frac{c(\hat{r}R_\odot)}{R_\odot}, \quad \hat{\gamma}(\hat{r}, \omega) = \gamma(\hat{r}R_\odot, \omega), \quad \hat{\sigma}^2(\hat{r}) = \omega^2 + 2i\omega\hat{\gamma}(\hat{r}, \omega).$$

The corresponding potential is given by

$$\hat{q} = \hat{\rho}^{1/2}\Delta_{\hat{x}}\left(\hat{\rho}^{-1/2}\right) - \frac{\hat{\sigma}^2}{\hat{c}^2}. \tag{6.1}$$

Utilizing spherical symmetry and picking a convenient source of excitation allows to compute the Green's function separately for each mode. A detailed derivation can be found in [BFF+20b, Section 4]. The modes $G_\ell(\hat{r}, \hat{s}, \omega)$ for $\hat{r}, \hat{s} \geq 0$ are obtained by solving the following ODE for $\tilde{G}_\ell = G_\ell/\hat{r}$:

$$-\frac{1}{\hat{r}^2}\frac{\partial}{\partial \hat{r}}\left(\hat{r}^2\frac{\partial \tilde{G}_\ell}{\partial \hat{r}}\right) + \left(\frac{\hat{\lambda}_\ell \hat{a}^2}{\hat{r}^2} + \hat{q}\right)\tilde{G}_\ell = \frac{1}{\hat{r}}\delta(\hat{r} - \hat{s}) \quad \text{on } [0, \hat{R}], \tag{6.2}$$

where $\hat{\lambda}_\ell = \ell(\ell+1)/\hat{a}^2$ for $\hat{a} = a/R_\odot$. Here, a denotes the radius of the truncation boundary in the original coordinates as introduced in (2.52) and $\hat{R} = R/R_\odot \in (0, \infty]$ marks the end of the atmospheric model.

- For the `Atmo` model $R = \infty$ holds, and the outgoing solution on a finite computational domain $[0, \hat{a}]$ can be obtained by means of the exact radiation condition

$$-(\partial_{\hat{r}}G_\ell)(\hat{r} = a) = dtn^{\texttt{Atmo}}(\hat{\lambda}_\ell) \tag{6.3}$$

provided by the Whittaker function as defined in (2.56) evaluated at the scaled quantities. More precisely,

$$dtn^{\texttt{Atmo}}(\hat{\lambda}_\ell) = \frac{1}{\hat{a}} + 2i\hat{k}\frac{W'_{-i\hat{\eta},\ell+\frac{1}{2}}(-2i\hat{k}\hat{a})}{W_{-i\hat{\eta},\ell+\frac{1}{2}}(-2i\hat{k}\hat{a})}, \tag{6.4}$$

with

$$\hat{c}_\infty = \hat{c}(\hat{a}), \quad \hat{\alpha}_\infty = -\hat{\rho}'(\hat{a})/\hat{\rho}(\hat{a}), \quad \hat{k}^2 := \frac{\hat{\sigma}(\hat{a})^2}{\hat{c}_\infty^2} - \frac{\hat{\alpha}_\infty^2}{4} \text{ and } \hat{\eta} := \frac{\hat{\alpha}_\infty}{2\hat{k}}.$$

As shown in [BFF+20b], the boundary condition $\lim_{\hat{r}\to 0} \hat{r}\partial_{\hat{r}}\tilde{G}_\ell(\hat{r},\hat{s},\omega) = 0$ ensures regularity of the solution at $\hat{r} = 0$. The truncation radius a is placed at the end of the background Model S, which is located at about 500 km above the photosphere.

- For the VAL-C model the physical radius $R = R_V$ is located at about 2.5 Mm height above the photosphere where a homogeneous Neumann boundary condition is implemented. The solution is unique since a positive damping is assumed, see Lemma 2.11 and Proposition 2.15.
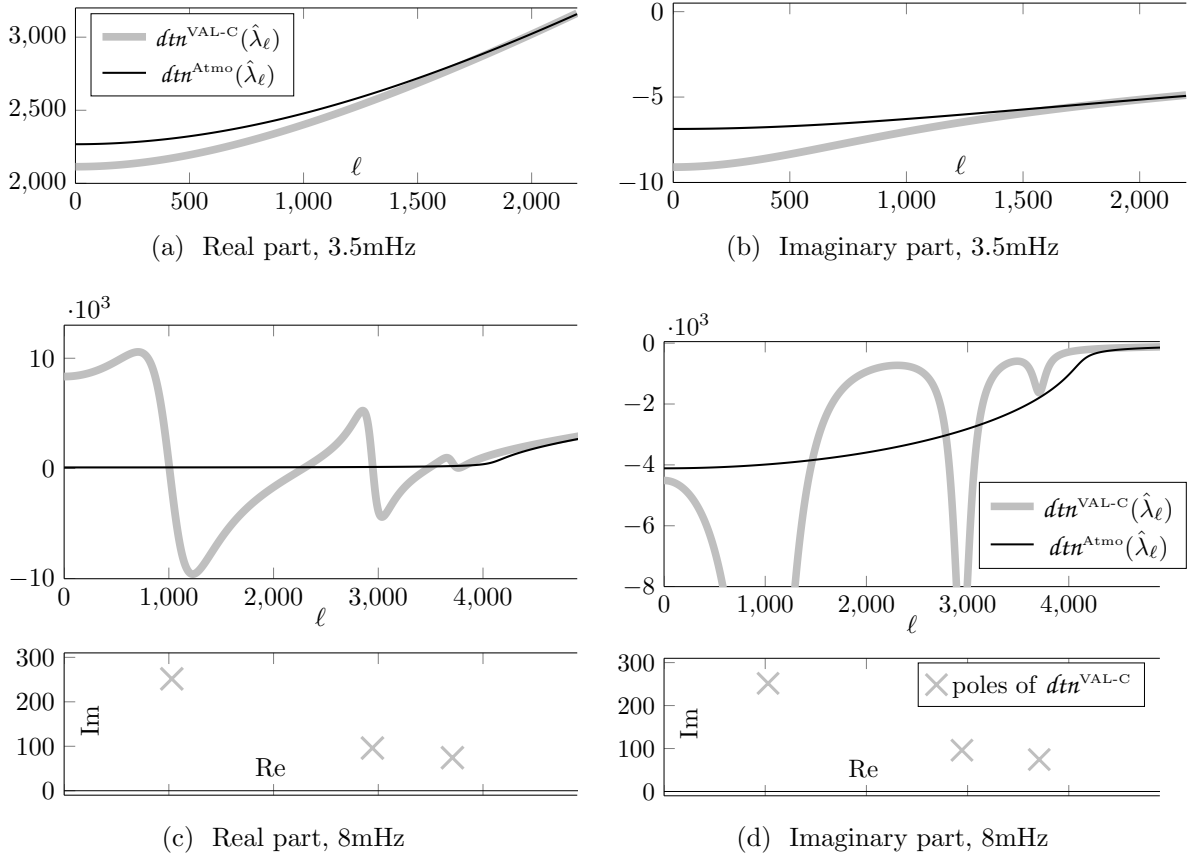

Figure 6.1: Comparison between $dtn^{\text{VAL-C}}$ and dtn^{Atmo} for a truncation boundary located at 0.5 Mm above the photosphere. The lower panel displays the exact poles of $dtn^{\text{VAL-C}}$.

All additional quantities required for this chapter can be computed from the imaginary part of the Green's function. Before proceeding to the definition of these derived

113

quantities, it makes sense to carry out a direct comparison of the *dtn* functions for the `Atmo` and VAL-C model. This will indicate for which frequencies and how the corresponding Green's functions are expected to differ. The upper panel of Fig. 6.1 compares the *dtn* functions of both models for 3.5 mHz while the lower panel considers a higher frequency of 8 mHz. For the sake of this comparison, the truncation boundary a has been placed at 0.5 Mm above the photosphere for both models. First it can be observed that the function values of $dtn^{\text{VAL-C}}$ and dtn^{Atmo} differ significantly even for low frequencies. While for low frequencies at least the qualitative behavior of both functions is still relatively similar, this changes completely for frequencies above the atmospheric cut-off (~ 5.2 mHz). In contrast to the smooth dtn^{Atmo} function, $dtn^{\text{VAL-C}}$ exhibits a series of pronounced peaks in the preasymptotic regime. These are associated with poles of the meromorphic continuation of $dtn^{\text{VAL-C}}$ lying in close proximity to the positive real axis as already discussed in Section 2.4.2. The poles of dtn^{Atmo} are located much higher up in the first quadrant and exert only minor influence on dtn^{Atmo} evaluated at eigenvalues $\hat{\lambda}_\ell$. As a consequence of the diverging nature of $dtn^{\text{VAL-C}}$ and dtn^{Atmo} at high-frequencies, one would expect to see significant differences in the quantities derived from the corresponding Green's functions in the high-frequency regime.

6.1.2 Power spectrum

Leighton et. al. [LNS62] were the first to observe that surface elements of the Sun undergo a vertical oscillation with a period of about five minutes. This phenomenon puzzled researchers for about a decade until these motions were eventually explained to originate from standing acoustic waves resonating in spherical cavities of the Sun [Ulr70, LS71]. The excitation mechanism of the waves is related to turbulent random motions in the solar convection zone, albeit not fully understood. To investigate the properties of solar oscillations, it is useful to analyze their power spectrum. To this end, the surface displacement is decomposed into spherical harmonics, and the amplitude is plotted as a function of frequency versus harmonic degree, see [GB05] for further discussion.

The power spectrum from simulations is obtained analogously by projecting the Green's function into spherical harmonics:

$$P^\ell(\omega) = \frac{\Pi(\omega)}{2\omega} \hat{\rho}(1) \, \text{Im} \left[G_\ell(\hat{r}_{\text{obs}}, \hat{r}_{\text{src}}, \omega) \right]. \tag{6.5}$$

The function $\Pi(\omega)$ controls the source power and is chosen in accordance with [GBD$^+$17] as

$$\Pi(\omega) = \left(1 + \left(\frac{\omega - \omega_0}{\omega_1} \right)^2 \right)^{-1} \quad \text{with } \frac{\omega_0}{2\pi} = 3.3 \text{ mHz} , \frac{\omega_1}{2\pi} = 0.6 \text{ mHz} . \tag{6.6}$$

Thinking in terms of Earth's seismology, the positions $\hat{r}_{\text{obs}}$ and $\hat{r}_{\text{src}}$ would represent the locations of the source of excitation and the receiver, respectively. Since solar oscillations are caused by random oscillations, they cannot be pinpointed to a single location like earthquakes so that the analogy above has to be treated with some caution. By comparing theoretical power spectra with MDI observations Nigam and Kosovichev [NK99] concluded that the region in which acoustic modes are excited can be narrowed down to a thin layer at depth of 75 ± 50 km below the photosphere. The location of the receiver, i.e. the observation height, depends on the particular instrument. In recent research [BFF$^+$20a], based on the same scalar wave equation as considered in this thesis,

it has been found that different source positions and observation heights can introduce significant differences in the computed power spectra. Currently, there seems to be no consensus yet on where source and receiver should be placed. Therefore, we simply opt for the established approach [GBD+17] to position both directly at the photosphere, i.e.

$$(\hat{r}_{\mathrm{obs}}, \hat{r}_{\mathrm{src}}) = (1, 1)$$

holds throughout the whole chapter. Learned IEs could also be applied for other choices as long as source and receiver are located inside the computational domain.

The power spectra obtained with the `Atmo` and the VAL-C models are compared in Fig. 6.2a and Fig. 6.2b. For this result Green's functions for harmonic degrees up to $\ell = 1000$ and 7200 equidistant frequencies ranging from 0 to 8.3 mHz have been computed. The attenuation has been chosen according to the power law damping model from [GBD+17], see Fig. 6.3b. The power is distributed along ridges which can be associated with the number of nodes of a corresponding radial resonant mode. The dispersion relation between frequency and ℓ (proportional to the horizontal wavenumber) is approximately given by a square root. The ridges reported in [KRSSL13] obtained from SOHO/MDI [SBB+95] observations are overplotted in the figure as white crosses. Due to the lack of gravity in the simulation the lowest ridge seen in the observations, which is associated with surface gravity waves, is not present in the computed power spectrum. The agreement below the acoustic cut-off frequency for both models is encouraging and appears to be consistent with results previously reported in the literature, see e.g. [GBD+17].

For high frequency the results for the two atmospheric models differ as anticipated from the comparison of the corresponding *dtn* functions, see Section 6.1.1. For the VAL-C model lines of reduced power are observed which have been reported previously in [BFF+20a], yet seem to be unexplained until now in the literature. Considering Fig. 6.2b very closely, it even possible to detect a slight disturbance already at around 4.2-4.5 mHz. Using our framework of *dtn* functions allows to shed some light on these observations (at least from a mathematical point of view). To this end, we proceed as in Chapter 5 and regard *dtn* additionally as a function of the frequency ω, which enters into equation (6.2) in terms of $\hat{\sigma}(\omega)$. In Fig. 6.2c and Fig. 6.2d the absolute values of the functions

$$(\ell, \omega) \mapsto \mathit{dtn}^{\mathtt{Atmo}}(\hat{\lambda}_\ell, \hat{\sigma}(\omega)) \text{ and } (\ell, \omega) \mapsto \mathit{dtn}^{\mathrm{VAL\text{-}C}}(\hat{\lambda}_\ell, \hat{\sigma}(\omega)) \tag{6.7}$$

are displayed. Apparently, these functions have a completely different structure at high-frequency. By comparing Fig. 6.2d with Fig. 6.2b we infer that the lines of reduced power in the spectrum are associated with lines at which $\mathit{dtn}^{\mathrm{VAL\text{-}C}}(\hat{\lambda}_\ell, \hat{\sigma}(\omega))$ is large in magnitude. Note that even the slight disturbance at around 4.2-4.5 mHz is cleary visible in Fig. 6.2d. It is tempting to conjecture that this drastic increase in magnitude around certain frequencies could be associated with poles $\mathit{dtn}^{\mathrm{VAL\text{-}C}}$ regarded as a function of frequency. Fig. 6.2e and Fig. 6.2f prove that this intuition is indeed correct. In these figures the functions (6.7) are displayed at fixed $\ell = 200$ as functions of ω. Their pole structure (also w.r.t. the frequency) is displayed in the lower panel of the figure. Please note the drastically different scalings of the imaginary axis. While $\mathit{dtn}^{\mathtt{Atmo}}$ has a tail of poles around 5.2 mHz which stretches away from the real axis into the fourth quadrant, $\mathit{dtn}^{\mathrm{VAL\text{-}C}}$ admits poles in extremely close proximity to the real axis around the whole high-frequency range. By definition, close to these poles the magnitude of $\mathit{dtn}^{\mathrm{VAL\text{-}C}}$ is extremely large. Hence, the poles of $\mathit{dtn}^{\mathrm{VAL\text{-}C}}$ close to the real frequency axis are responsible for

(a) Atmo

(b) VAL-C

(c) $|dtn^{\text{Atmo}}(\hat{\lambda}_\ell, \hat{\sigma}(\omega))|$

(d) $|dtn^{\text{VAL-C}}(\hat{\lambda}_\ell, \hat{\sigma}(\omega))|$

(e) $\omega \mapsto dtn^{\text{Atmo}}(\hat{\lambda}_{\ell=200}, \hat{\sigma}(\omega))$

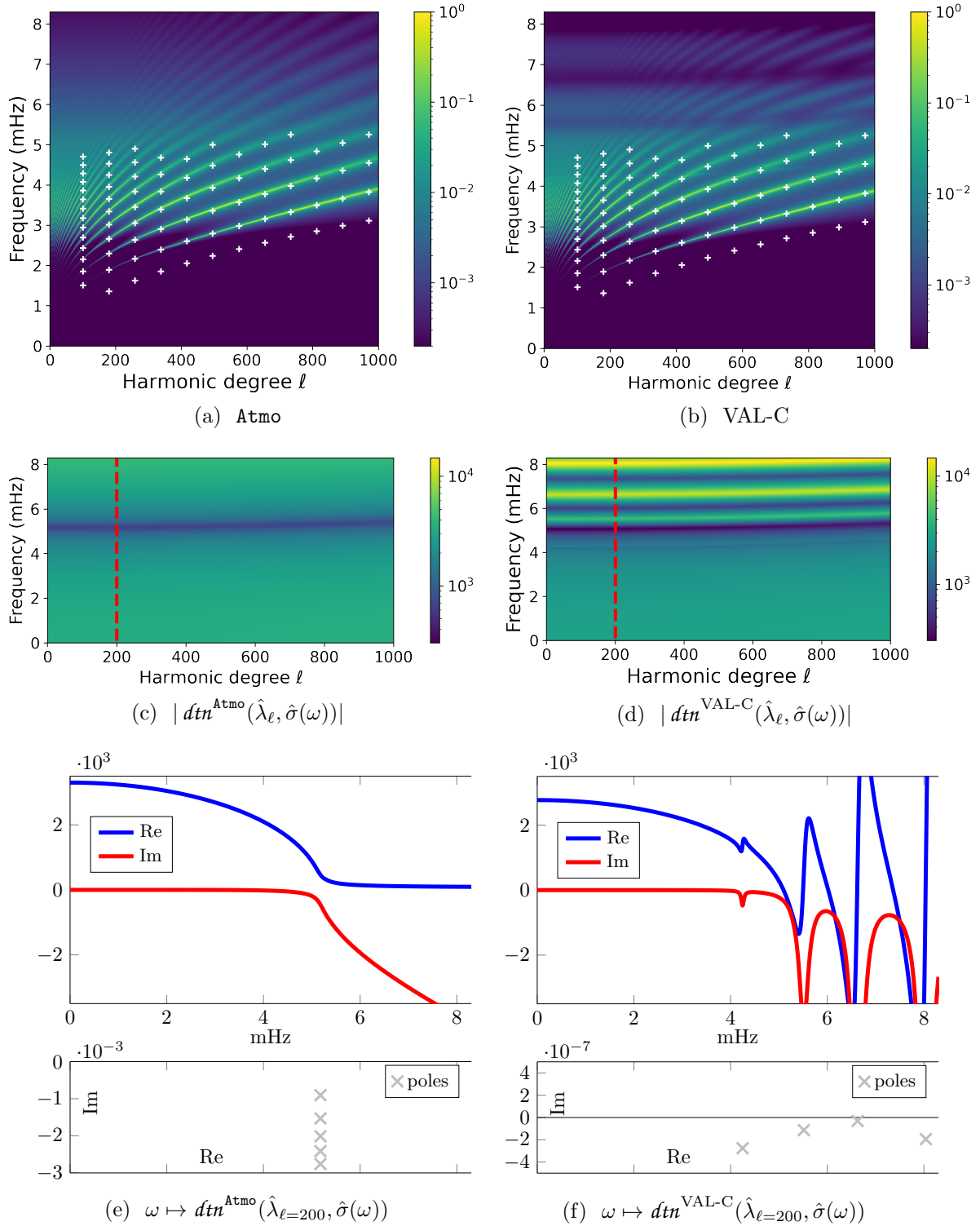(f) $\omega \mapsto dtn^{\text{VAL-C}}(\hat{\lambda}_{\ell=200}, \hat{\sigma}(\omega))$

Figure 6.2: In (a) and (b) the power spectrum obtained with different models of the atmosphere is compared. The white crosses mark the ridges reported from MDI observations [KRSSL13]. In (c) and (d) the absolute value of dtn as a function of frequency (y-axis) and harmonic degree ℓ (x-axis) is shown. A slice through $\ell = 200$, marked in (c) and (d) by the dashed red line, is displayed in figures (e) and (f). Here, we regard dtn as a function of frequency. Real and imaginary parts are shown in the upper panel while the lower panel displays the pole structure.

116

the lines of reduced power observed in the spectrum of the VAL-C model. They seem to be caused by the drastic increase in sound speed in the upper chromosphere featured in this model of the solar atmosphere, see Fig. 2.7b. In this regard, it is interesting to note that Remark 2.12 extends to the setting in which $dtn^{\text{VAL-C}}$ is considered as a function of frequency. This implies that frequencies at which $dtn^{\text{VAL-C}}$ has a pole are exactly those frequencies for which the exterior problem does not admit a unique solution in $H^1([a, R_V])$. Furthermore, for the case of the homogeneous medium it is shown in [Tay11, Chapter 9.7, following equation (7.64)] that poles of $\mathcal{DtN}$ with respect to the frequency coincide with resonances[1] of the exterior problem. Assuming that these results can be transferred to the inhomogeneous solar atmosphere would allow to conclude that the lines of reduced power in the spectrum of the VAL-C model are caused by resonances introduced by reflections in the upper chromosphere.
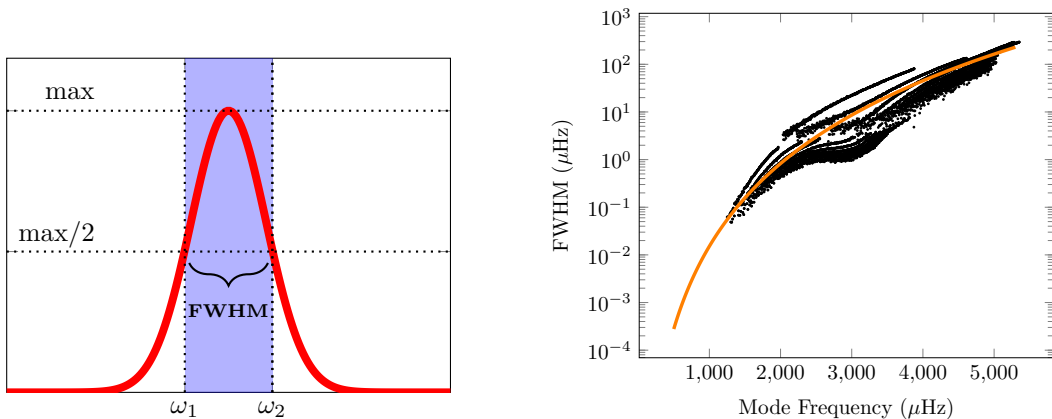
(a) The FWHM of a peak in the power spectrum (e.g. Fig. 6.4) is the width $|\omega_2 - \omega_1|$ at half the maximum amplitude.

(b) The FWHM of acoustic modes with orders $1 \leq n \leq 12$ (black dots) reported in [KRSSL13, LS15] based on MDI observations. The orange line shows the power law approximation $2\gamma(\omega)$.

Figure 6.3: Relation between full width at half maximum (FWHM) of acoustic modes and attenuation in the computational framework of [GBD+17]. In (b) we recreated [GBD+17, Fig. 8] using the same data sets as in this reference.

Let us now continue with the discussion of the power spectrum. A cut through the power spectrum at $\ell = 200$ shown in Fig. 6.4 allows for a closer comparison with observations. Here, frequencies below and above 5.2 mHz have been normalized independently and are discussed separately below.

- For low frequencies the Atmo and VAL-C models yield similar results. The simulated ridges with both atmospheric models display a small misalignment with respect to the observations which is well-known in the literature, see e.g. [BFF+20b, Fig. 18] or [GBD+17, Fig.10]. This is expected as the background Model S does not capture the surface layers perfectly [RCN+99]. Note that this misalignment is much larger than the difference between the two competing atmospheric models, hence the modelling error stemming from the surface layers apparently dominates here. Additionally, the attenuation model used in this study is oversimplified: In the framework of [GBD+17] the full width at half maximum of a peak (FWHM) in

---

[1]Please see also [Tay11, Proposition 7.8] in which scattering poles and resonances are identified.
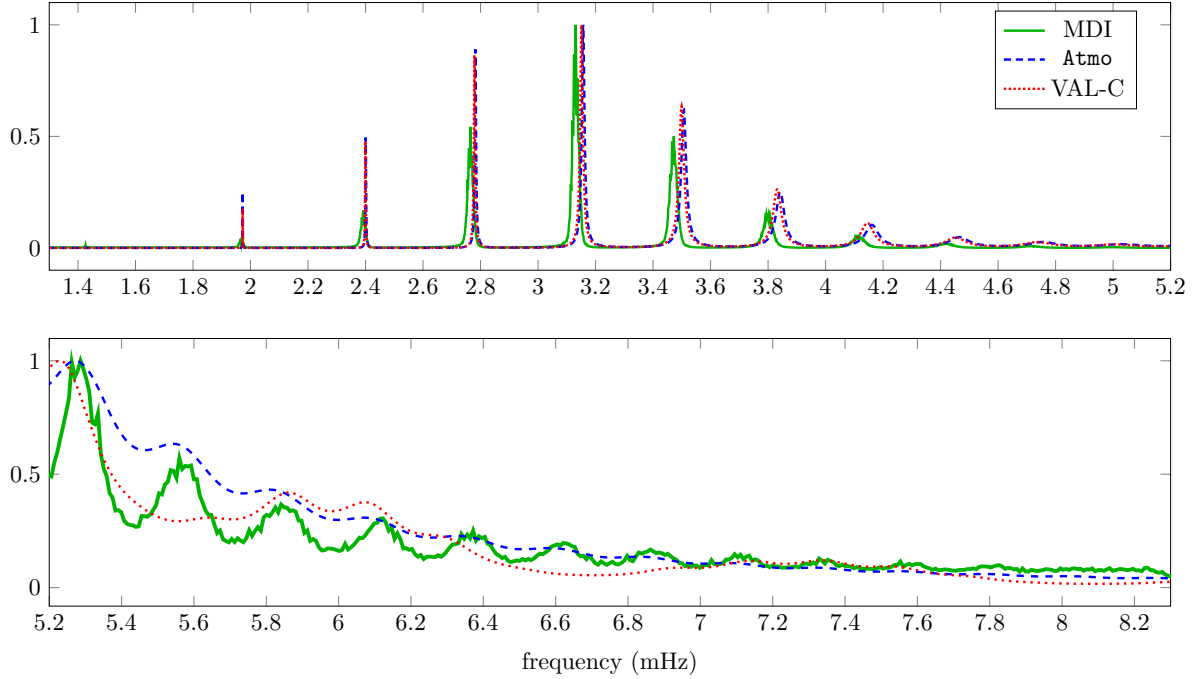
Figure 6.4: Cut through the power spectrum at $\ell = 200$. Low and high frequencies have been normalized independently.

the power spectrum is proportional to the attenuation γ. The power law damping model only accounts for the dependence of the FWHM on the frequency (see Fig. 6.3b), while observations, see e.g. [KRS04], show that the FWHM is a strongly increasing function of the harmonic degree ℓ as well. This point is discussed further in Section 6.4. Overall these results indicate that the influence of the atmospheric model for frequencies below 5.2 mHz appears to be relatively small.

- For frequencies above the acoustic cut-off both atmospheric models yield significantly different results. Unfortunately, neither of the models shows a satisfactory agreement with the MDI observations. This has also been reported in [BFF+20b]. For the VAL-C model the lines of reduced power are also clearly visible. For example, at around 5.5 mHz where a pole of $dtn^{\text{VAL-C}}$ is located (see Fig. 6.2f) the power based on the VAL-C model attains a local minimum, which is in disagreement with the MDI data. An outlook on further research towards resolving the discussed discrepancies will be given in Section 6.4.

6.1.3 Cross covariance and time distance diagram

Time-distance helioseismology [DJHP93] studies the solar interior by analyzing the time it requires waves to travel between two locations on the solar surface which is encoded in the so called cross-covariance function. Plotting the cross-covariance as a time-distance diagram allows for a visual intepretation of travel times. Fig. 6.5b displays such a time-distance diagram computed from MDI data provided in [RKS+97]. This plot displays the cross-covariance as a function of temporal correlation time lag between any two points which have the same angular distance on the solar surface. The distinguished feature are the ridges which are related to different paths the wave can travel between both locations. This is illustrated in Fig. 6.5a using ray theory [Gou84], which is an

(a) Wave paths in ray approximation.

(b) MDI observations.
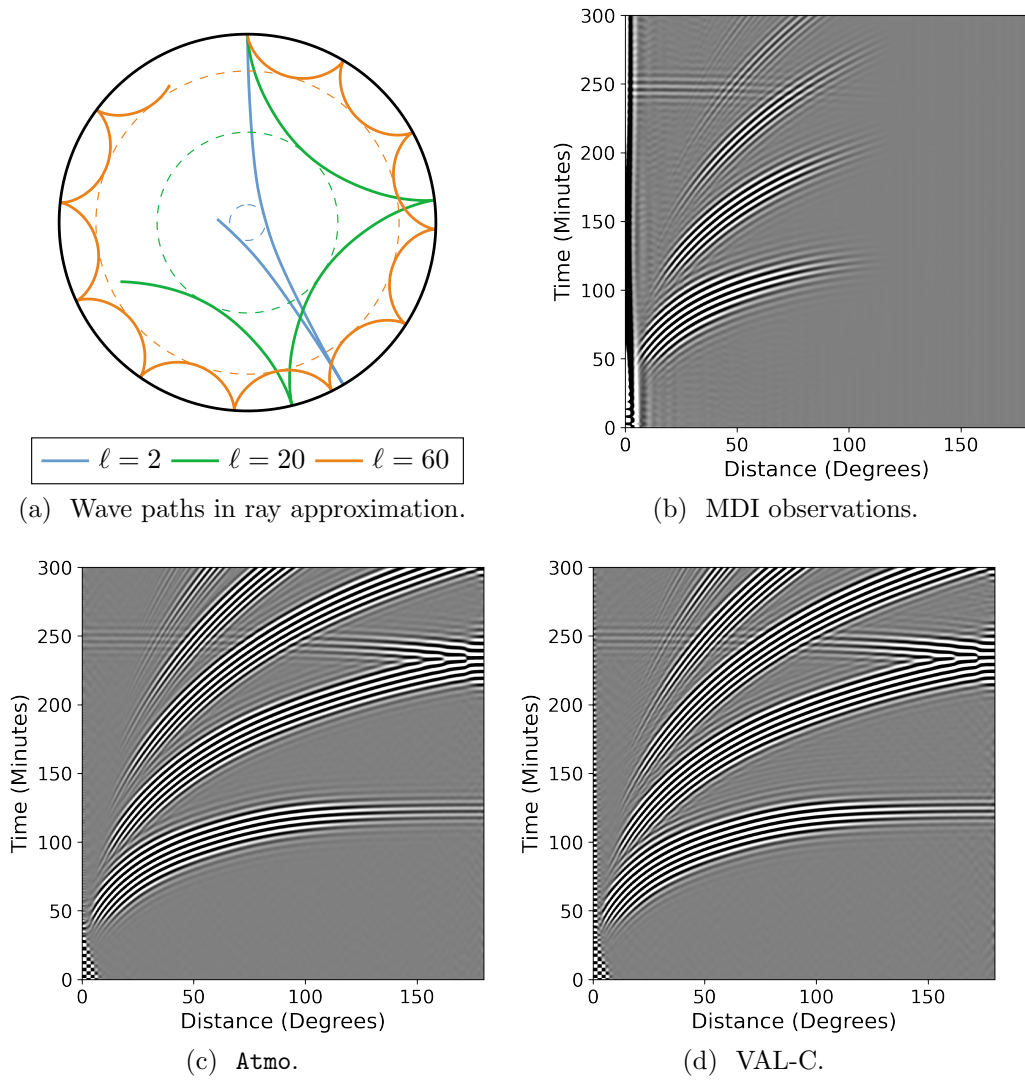
(c) `Atmo.`

(d) VAL-C.

Figure 6.5: Time distance diagram obtained from the cross-covariance. White for positive values and black for negative values.

approximation justified in the high-frequency regime. Here, the energy of the wave is visualized as a ray path which is orthogonal to the wavefront. Under the assumption that a simple local dispersion relation is valid[2], it is possible to compute the ray path for each mode ℓ simply by integrating a quantity which involves the sound speed, see [Kos11, Section 4.5] for details. As the waves travel deeper into the Sun they are refracted by the increase in sound speed towards the core. After reaching a lower turning point $\hat{r}_t$, which in the setting of [Kos11] is located at $\hat{c}(\hat{r}_t) = \hat{r}_t \omega / \sqrt{\ell(\ell+1)}$, they return to the surface. The first skip in the time-distance diagram then corresponds to the ray which connects two points on the surface directly, i.e. the first arrival. For the second skip the wave is reflected once at the surface and so on. For further interpretation we refer to [GBS10, Section 3.2] and references therein.
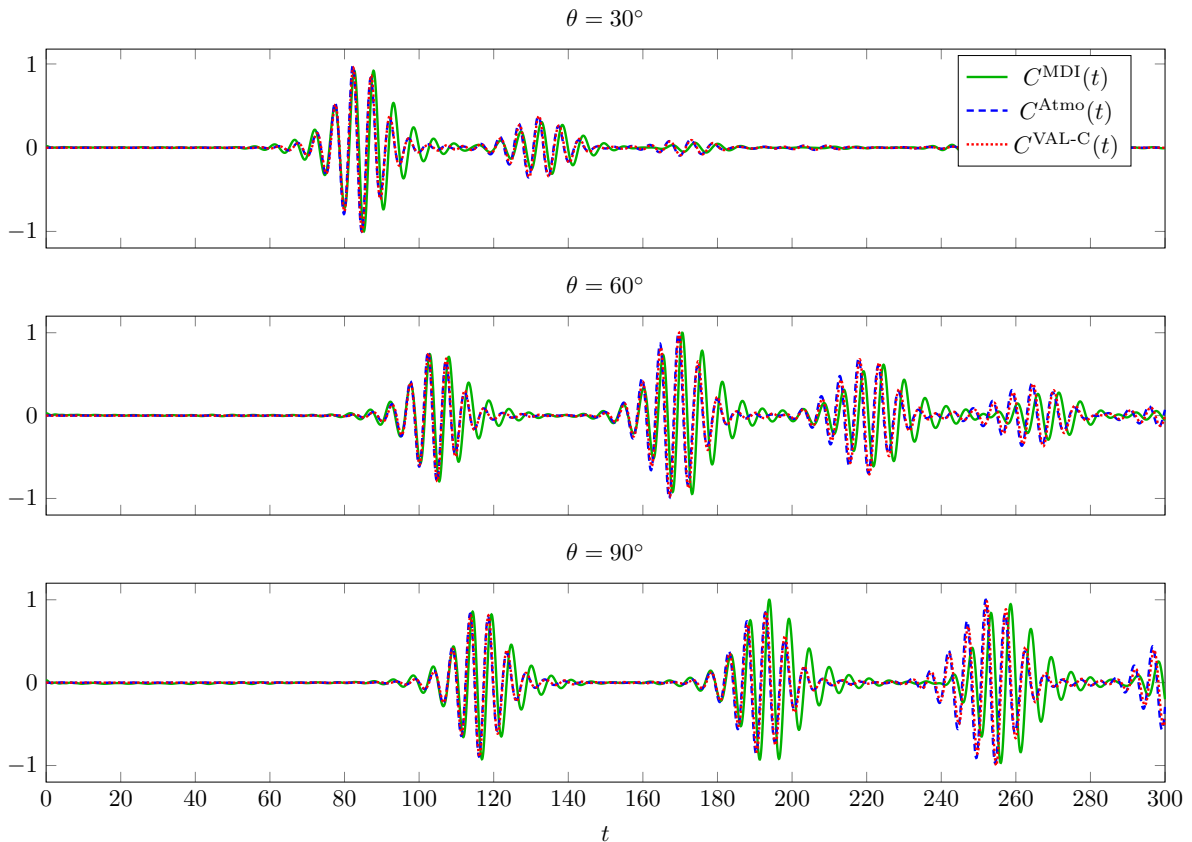

Figure 6.6: Comparison of the expectation value of the cross-covariance with observations.

Assuming a convenient source of excitation, the expectation value of the cross covariance $C(\theta, \omega)$ can be obtained from the imaginary part of the Green's function [GBD+17]. For the special setting in this chapter it is given by (see [BFF+20b] for a derivation):

$$C(\theta, \omega) = \frac{\Pi(\omega)}{8\pi\omega} \hat{\rho}(1) \sum_{\ell} (2\ell + 1) F_\ell(\omega) \operatorname{Im}\left[G_\ell(1, 1, \omega)\right] P_\ell(\cos\theta). \qquad (6.8)$$

Here, $\theta \in [0, \pi]$ is the angular distance between the two points on the surface, P_ℓ are the Legendre polynomials of degree ℓ and F_ℓ is a filter function to be specified below. The expression $C(\theta, t)$ in the time domain is obtained by applying the Fourier transform.

---

[2]This assumptions breaks down near the upper turning point of the ray.

The time-distance diagrams based on the `Atmo` and VAL-C models are displayed in Fig. 6.5c and Fig. 6.5d, respectively. To remove high-degree modes the filter $F_\ell = 1/2(1 - \tanh(0.03 - \ell))$ for $\ell < 300$ and zero else was used. Additionally, a low-frequency filter has been applied to the observations to filter out noise from granulation[3] and supergranulation. The most striking difference to the observations, c.p. Fig. 6.5b, is that the ridges in the observed time-distance diagram fade at large distances. This is caused by an observational effect known as foreshortening where spatial resolution on the Sun decreases towards the limb. Apart from this, the qualitative agreement between observations and simulations is reasonable and comparable to the results in the literature, see e.g. [GBD+17, Fig. 12].

Fig. 6.6 displays cuts through the cross-covariance at $\theta \in \{30°, 60°, 90°\}$. Such a plot has also been presented in [GBD+17, Fig. 13]. For the first few skips the relative amplitudes and widths of the simulations match the observations reasonably well. However, we record that small deviations can already be detected easily with the unaided eye. In view of the inaccuracies already observed in the power spectrum (see Fig. 6.4), one can also not expect to see a perfect match with the observations. The most noticeable defect might be that the widths of the skips are slightly underestimated in our simulations. Concerning the atmospheric models, no significant differences between `Atmo` and VAL-C can be observed in terms of how well the observations are reproduced. This will change in the next experiment which focusses on waves propagating at high-frequencies.

Double ridge structure at high-frequency

Jefferies et. al. [JOS+97] noticed that the time-distance diagram of high-frequency waves based on observations taken from the South pole exhibits a double-ridge structure. The ridges are split into a lower main ridge and an upper satellite ridge separated by gaps. The authors of [JOS+97] argued that this structure may arise due to reflection of high-frequency acoustic waves in the chromosphere, that is due to resonance effects. However, also alternative interpretations exist [SSJ04] attributing this phenomenon to an interference of waves with frequencies above and below the acoustic cut-off, which may have been introduced by a special choice of the filter. The filter in question given by

$$F_\ell(\omega) = \exp\left(-\frac{(\omega - \omega_0)^2}{2s^2}\right) \exp\left(-\frac{(\ell - \ell_0)^2}{(\delta\ell)^2}\right) \tag{6.9}$$

with $\omega_0/2\pi = 6.75$ mHz, $s/2\pi = 0.75$ mHz, $\ell_0 = 125$ and $\delta\ell = 33$. The frequency and spatial components of the filter are displayed in Fig. 6.7d. The resulting time-distance diagram computed with MDI data [RKS+97] is shown in Fig. 6.7a. The gaps in the first two ridges are clearly visible.

The double-ridge structure has been successfully reproduced in simulations by Fournier et. al. [FLH+17] using equation (2.51) with an atmospheric extension based on the VAL-C model [VAL81]. Additionally, it has been demonstrated that the gaps in the ridges are missing if the `Atmo` extension, which the existing atmospheric radiation boundary conditions [BCD+18, FLH+17, BFP20] for helioseismology rely on, is employed to model the atmosphere. It is interesting to check whether these results can be reproduced here.

---

[3]Granulation and supergranulation are convective phenomena which can be observed at the solar surface. Their traces are superimposed on the oscillation signal of the ascoustic modes and should therefore be filtered out for this study.

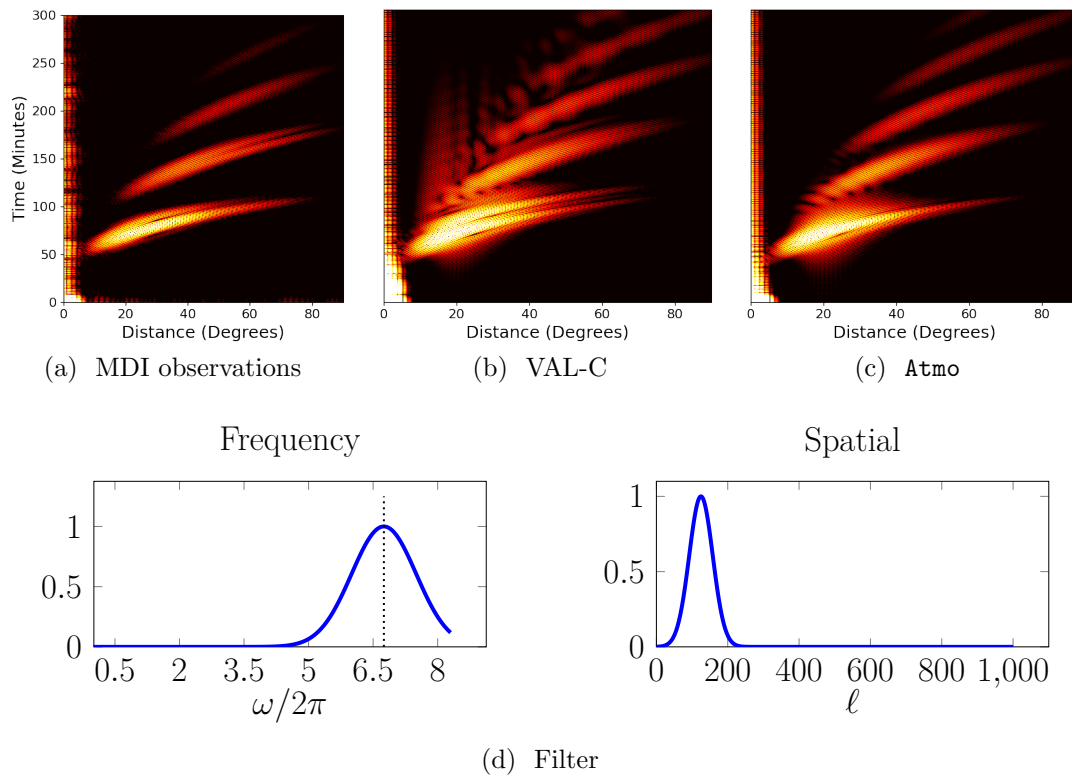(a) MDI observations   (b) VAL-C   (c) `Atmo`

(d) Filter

Figure 6.7: Double ridge structure of high-frequency time-distance diagram. Shown is the absolute value of the expected cross-covariance on a log scale with large values being yellow and small values black.

The time-distance diagram with the exact transparent boundary condition for the `Atmo` model is shown in Fig. 6.7c. No significant gaps in the ridges are visible. On the other hand, the gaps appear in the corresponding time-distance diagram based on the VAL-C model shown in Fig. 6.7b. This may be interpreted as an indication that the VAL-C model could be physically more appropriate than the `Atmo` model. To be fair, it should be pointed out that the agreement of the VAL-C results with the MDI observations is obviously not perfect either.

6.1.4 Conclusion on accuracy and purpose of approximate transparent b.c.

The experiments of this section were based on exact transparent boundary conditions which can only be realized in a one-dimensional setting, that is if the PDE at the continuous level is fully-separable. For axisymmetric or fully three-dimensional discretizations approximate transparent boundary conditions like learned IEs have to be used instead. The main objective of the following two sections is to investigate the influence of the approximate transparent boundary condition on the accuracy of the power spectrum and the expectation value of the cross-covariance. The one-dimensional setting assumed in this chapter allows for a direct comparison with the reference result computed with the exact transparent boundary condition. This raises the question on the level of accuracy the approximate transparent boundary conditions should be able to achieve. From the results of this section we infer the following conclusions.

- In the low-frequency regime the influence of the transparent boundary conditions seems to be small. As argued in the discussion of Fig. 6.4, the deviations from the observations appear to stem from other dominant sources not directly related to the atmospheric model. Therefore, it should be sufficient if the power spectra and cross-covariances (Fig. 6.6) based on the exact transparent boundary conditions can be reproduced with a relative accuracy in the region of $10^{-2} - 10^{-5}$.

- In the high-frequency regime the observables computed based on the `Atmo` extension and the VAL-C model display a fundamentally different behavior as expected from the comparison of *dtn* functions shown in Fig. 6.1 and Fig. 6.2c-Fig. 6.2f. We do not claim to know which of these models - if any - is the correct one to incorporate the solar atmosphere into the scalar equation of helioseismology. With respect to transparent boundary conditions we merely want draw the following two conclusions. Firstly, the VAL-C model is certainly interesting for helioseismology (see e.g. Section 6.1.3) so that transparent boundary conditions for its efficient implementation in axisymmetric or three-dimension discretizations should be investigated. Currently, no such conditions are known in the literature. We will fill this gap in Section 6.3. Secondly, the uncertainty on the appropriate atmospheric model calls for flexible transparent boundary conditions which can adapt in case the background model changes. Their natural flexibility also commends learned IEs in this context.

A final remark should be given to put the investigations following in Section 6.2 and Section 6.3 into perspective. In these experiments we will push for extremely high accuracy with learned IEs even though this is probably excessive in view of the application as argued above. This is in order to examine the accuracy and flexibility of learned IEs for

a nontrivial example. Certainly, helioseismology provides a significantly more interesting proving ground than the toy examples considered in Chapter 4.

6.2 Comparison with ARBC for the `Atmo` model

It is natural to consider the `Atmo` model first for which local transparent boundary conditions to compare with are available. These so called 'Atmospheric Radiation Boundary Conditions' have been introduced in [BCD+18] and were then further developed in a series of papers [FLH+17, BFP20, BFF+20b, BFF+20a]. Most of them are based on local approximations of the non-local condition

$$dtn^{\text{nonlocal}}(\lambda_\ell) = \frac{1}{\hat{a}} - i\hat{k}\sqrt{1 - \frac{\hat{\alpha}_\infty}{\hat{a}\hat{k}^2} - \frac{\hat{\lambda}_\ell}{\hat{k}^2}}. \tag{6.10}$$

The non-locality here stems from the square root of the Laplace-Beltrami operator (represented by $\hat{\lambda}_\ell$ in the separable setting). In fact, there is a zoo of different approximations available which can be described by the tensor-product ansatz (3.10) using $N = 0$. i.e. $A_{\Gamma\Gamma}$ and $B_{\Gamma\Gamma}$ are the only nonvanishing matrix elements. This justifies to employ the notation dtn for these transparent boundary conditions. It is beyond the scope of this thesis to consider all of these conditions. Therefore, we focus here on

$$dtn^{\text{S-HF-1a}}(\lambda_\ell) = \frac{1}{\hat{a}} - i\hat{k} + \frac{i}{2\hat{k}}\frac{\hat{\alpha}_\infty}{\hat{a}}, \tag{6.11}$$

which the researchers have singled out in their latest report [BFF+20a] as their favourite condition because of its high accuracy and independence of $\hat{\lambda}_\ell$ (that is $B_{\Gamma\Gamma} = 0$) which the authors argue is more convenient to implement in three dimensions. Additionally, also the condition

$$dtn^{\text{A-RBC-1}}(\lambda_\ell) = -i\hat{k} \tag{6.12}$$

will be considered, which is of the form of a Sommerfeld radiation condition for the wavenumber $\hat{k}$.

To assess the accuracy of these conditions, equation (6.2) is solved on the interval $[0, \hat{a}]$ with a being the end of Model S where the exact transparent boundary condition (6.3) at $\hat{r} = \hat{a}$ is replaced by one of the conditions (6.10), (6.11) or (6.12). Source and receiver are positioned at the solar surface. An analogous experiment has been carried out in [BFF+20a, Section 5.3.] and we should obtain qualitatively similar results here. A perfect agreement cannot be expected though as details of the implementation may vary. For example, different approaches to compute the contribution $\Delta_{\hat{x}}\hat{\rho}^{-1/2}$ appearing in the potential (6.1) or the parameter $\hat{\alpha}_\infty = -\hat{\rho}'(\hat{a})/\hat{\rho}(\hat{a})$ from the data lead to different results, e.g. this work uses $\hat{\alpha}_\infty \approx 6588.55$ while [BFF+20a] employs $\hat{\alpha}_\infty = 6663.62$. We use a constant attenuation of 20μHz as in the reference. Additionally, a comparison with learned IEs shall be performed here which can easily be realized. To this end, it suffices to solve the optimization problem (4.2)-(4.3) to obtain the learned $dtn_N(\hat{\lambda}_\ell)$ approximation to the exact $dtn^{\text{Atmo}}(\hat{\lambda}_\ell)$ numbers and then proceed as above. All experiments in this chapter are based on the reduced symmetric ansatz for the learned IE matrices, as introduced in Section 4.1.1.

6.2.1 Power spectrum

Firstly, the influence of the transparent boundary condition on the computed power spectrum will be investigated. Let $P^\ell(\omega)$ denote the reference power spectrum computed based on the exact transparent boundary condition (6.3), and let $P^\ell_{\mathrm{app}}(\omega)$ denote the power spectrum computed using any of the other (approximate) transparent boundary conditions from above. Fig. 6.8 displays the relative error

$$|P^\ell(\omega) - P^\ell_{\mathrm{app}}(\omega)| / |P^\ell(\omega)| \tag{6.13}$$

for different transparent boundary conditions. The upper panel of the figure shows the results for the atmospheric radiation boundary conditions which are in good agreement with the literature, c.p. [BFF$^+$20a, Fig.11]. Computations using learned IEs are displayed in the lower panel. Some of the results are as expected:

- The nonlocal condition is the most accurate among the atmospheric radiation boundary conditions.

- The accuracy of learned IEs improves exponentially fast with increasing N.

Other results are less obvious or even surprising:

- For $\ell \leq 200$ learned IEs with $N = 0$ match the accuracy of the nonlocal condition while for $N = 4$ they even surpass the latter in accuracy for all modes. This is astonishing since in contrast to learned IEs, it is not feasible to implement the nonlocal condition in dimensions larger than one.

- For all conditions the error is largest around the acoustic cut-off frequency.

To shed light on these observations, it is useful to compare the approximation of $dtn^{\mathtt{Atmo}}(\hat{\lambda}_\ell)$ achieved by the different transparent boundary conditions. Results for 3.0, 5.3 and 6.5 mHz are displayed in Fig. 6.9. For 3.0 and 6.5 mHz the behavior of $dtn^{\mathtt{Atmo}}(\lambda_\ell)$ for $\ell \leq 1000$ is approximately given by a quadratic function in ℓ, respectively an affine linear function in $\hat{\lambda}_\ell$. This extends to other frequencies away from the acoustic-cutoff and explains the good performance of lowest order learned IEs. Around the acoustic cut-off frequency the behavior of $dtn^{\mathtt{Atmo}}$ is more complex so that a few additional degrees of freedom are required for its accurate resolution. This is also clearly visible in the plot of the relative error provided in the lower panel of the figure. The increase in complexity of $dtn^{\mathtt{Atmo}}$ around the acoustic cut-off is related to a 'transition of phase': For small frequency $|\operatorname{Re} dtn^{\mathtt{Atmo}}| \gg |\operatorname{Im} dtn^{\mathtt{Atmo}}|$ while $|\operatorname{Re} dtn^{\mathtt{Atmo}}| \ll |\operatorname{Im} dtn^{\mathtt{Atmo}}|$ for high frequencies. This reflects the fact that as the frequency inreases the surface layers of the Sun become increasingly transparent and waves can propagate out into the atmosphere. Additionally, the plots of the relative error clearly demonstrate that learned IEs with $N = 4$ and away from the cut-off even for smaller N are more accurate than the nonlocal condition. It is apparently more appropriate to approximate $dtn^{\mathtt{Atmo}}$ by a rational function rather than by a square root.

6.2.2 Expectation value of cross-covariance

Next the influence of the transparent boundary condition on the expectation value of the cross-covariance will be investigated. Following common practice in time-distance
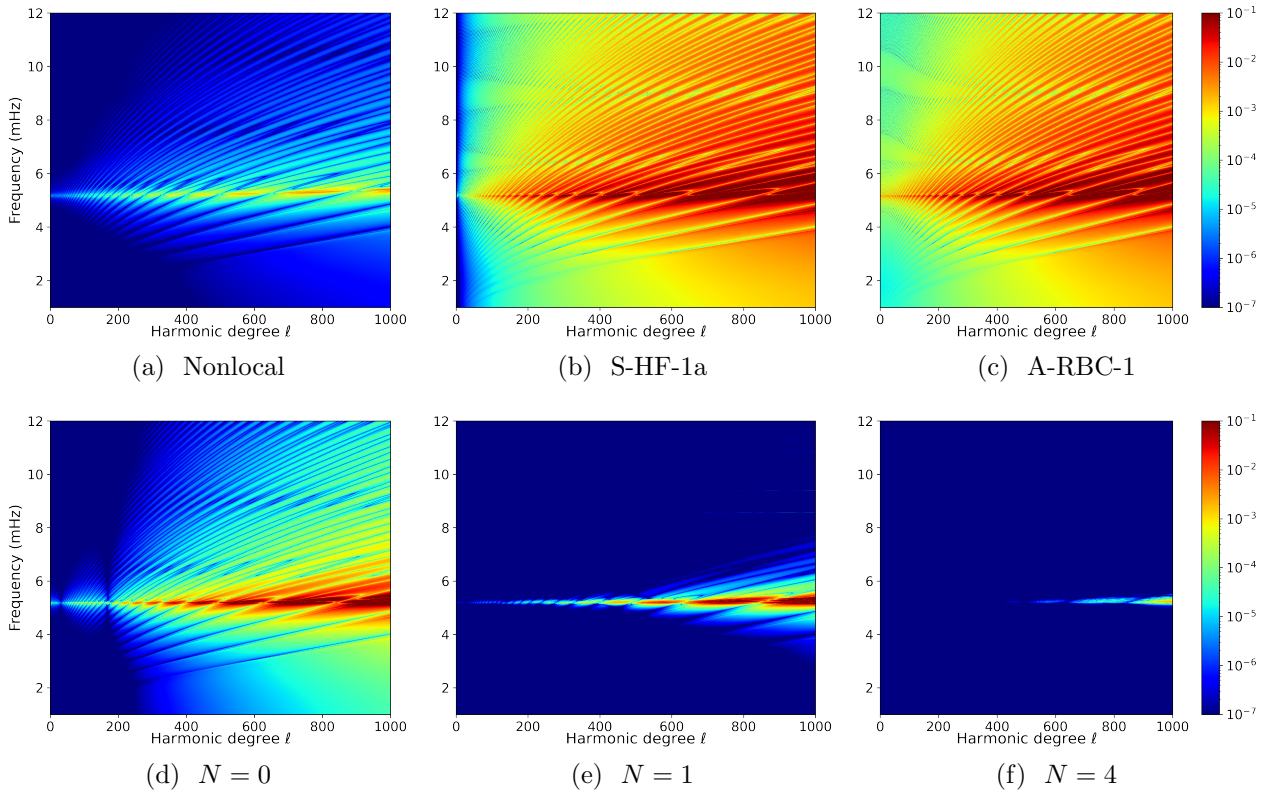
Figure 6.8: Relative error for power spectra obtained with different transparent boundary conditions. Upper panel: Atmospheric radiation boundary conditions, lower panel: Learned IEs.
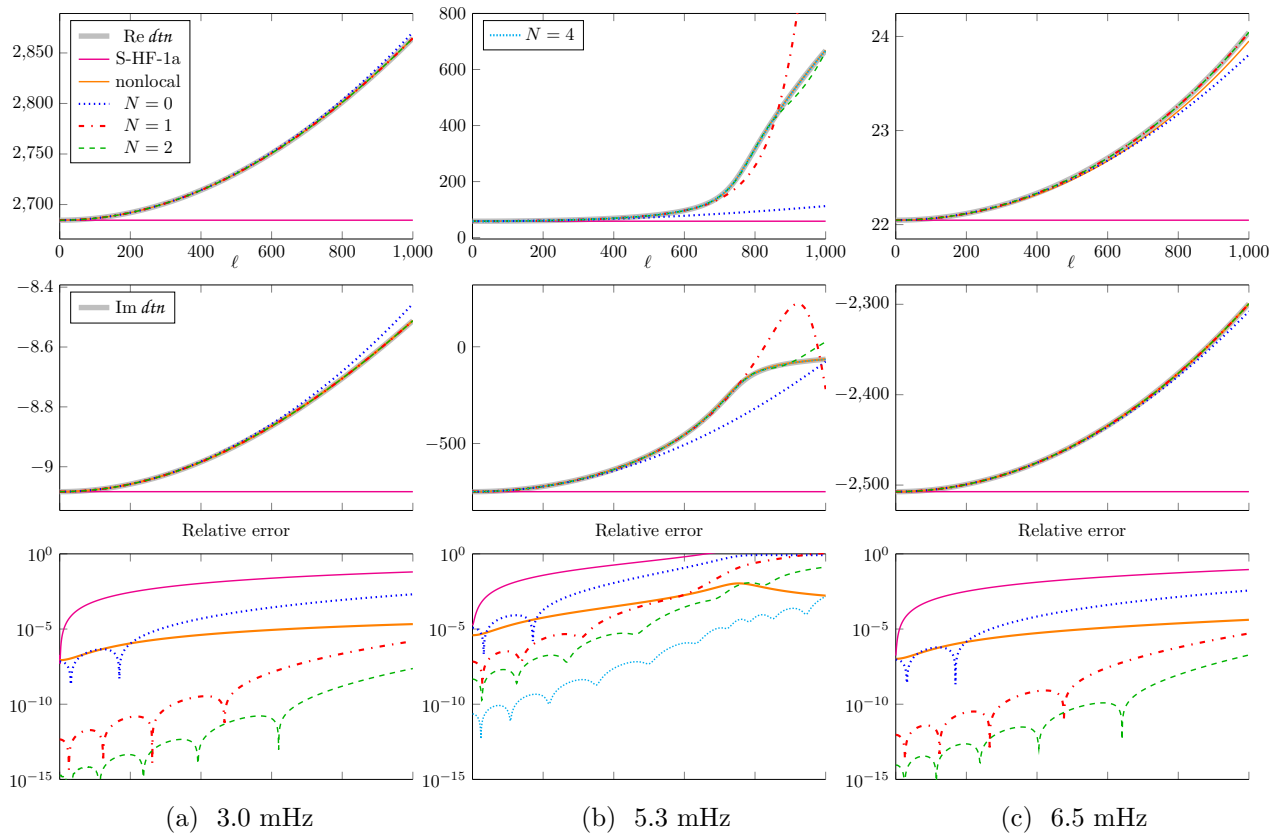
Figure 6.9: Comparison of *dtn* approximations for the `Atmo` model. Top: Real part, Middle: Imaginary part and bottom: relative error $|dtn(\hat{\lambda}_\ell) - \mathrm{dtn}_{\mathrm{app}}(\hat{\lambda}_\ell)|/|dtn(\hat{\lambda}_\ell)|$.

helioseismology, see [GB05, Section 4.3.1.], certain parts of the time-distance diagram will be isolated by applying appropriate filters. A combination of two Gaussian filters

$$F_\ell(\omega) = \exp\left(-\frac{(\omega - \omega_0)^2}{s^2}\right) \exp\left(-\frac{(\omega R_\odot/\ell - v_0)^2}{\delta v^2}\right), \qquad (6.14)$$

with $s/2\pi = 0.3$ mHz and $\delta v = 12.3$ km/sec, can be used to isolate waves with frequency around ω_0 and speed v_0. Following [BFF$^+$20b, Section 9.3], low-frequency $(\omega_0/2\pi, v_0) = (3 \text{ mHz}, 125.2 \text{ km/s})$ and high-frequency $(\omega_0/2\pi, v_0) = (6.5 \text{ mHz}, 250.4 \text{ km/s})$ waves will be considered. The corresponding filters are displayed in Fig. 6.10a and Fig. 6.11a, respectively.

Fig. 6.10 shows the results for waves with frequencies around 3 mHz and speeds of 125.2 km/s. The reference cross-covariance $C^{\text{ref}}(\theta, t)$ obtained with the exact transparent boundary condition is shown in Fig. 6.10b. The results obtained with the approximate transparent boundary conditions along cuts through $\theta \approx 14°$ and $\theta \approx 28°$ are displayed in the two lower panels. For the unaided eye no difference between reference cross-covariance $C^{\text{ref}}(\theta, t)$ in gray and any of the respective approximations $C^{\text{app}}(\theta, t)$ in colors are visible. However, in the absolute error $|C^{\text{ref}}(\theta, t) - C^{\text{app}}(\theta, t)|$ significant differences between the different transparent boundary conditions emerge. An exponentially fast convergence for the learned IEs with increasing N is observed. At $N = 0$ the accuracy of the nonlocal condition is matched and at $N = 2$ machine accuracy is reached. In view of the shape of the filter, these results are consistent with the relative error for the *dtn* approximation at 3.0 mHz shown in Fig. 6.9a. The magnitude of the absolute error for the S-HF-1a condition is in line with the results from [BFF$^+$20b, Section 9.3]. Although this condition is expectedly the least accurate among the considered transparent boundary conditions as it approximates $dtn^{\texttt{Atmo}}$ merely by a constant, the achieved absolute error of order $\sim 10^{-6}$ is nevertheless impressive and as discussed in Section 6.1.4 probably more than sufficient for the application.

The results for waves with frequences around 6.5 mHz and speeds of 250.4 km/s are displayed in Fig. 6.11. The absolute errors are slightly larger since the boundary condition exerts more influence on higher frequency waves. Consequently, $N = 4$ was necessary to achieve near machine precision accuracy with learned IEs. Apart from this, similar results as for the previous experiment are obtained.

Overall, the presented experiments demonstrate that learned IEs are well-suited for modelling a solar atmosphere based on the `Atmo` model. Thanks to the exponentially fast convergence, very few additional DOFs are sufficient to obtain highly accurate helioseismic observables. Although the atmospheric radiation boundary conditions cannot compete with learned IEs in terms of accuracy, they provide an attractive alternative because of their computational simplicity (e.g. no optimization step is required). As mentioned previously, the accuracy that these conditions provide is presumably already sufficient if one assumes that the solar atmosphere is well-represented by the `Atmo` model. However, as seen in Section 6.1.3 there seem to be indications that this is not the case. Since the sound speed in the `Atmo` model is extended by a constant starting from about 500 km above the photosphere, while the actual sound speed in the Sun increases sharply in the upper chromosphere towards the corona such discrepancies also do not come as a surprise. Therefore, it is crucial to develop transparent boundary conditions for more realistic models of the solar atmosphere like the VAL-C model. This task is tackled in the next section by leveraging the flexibility of learned IEs.
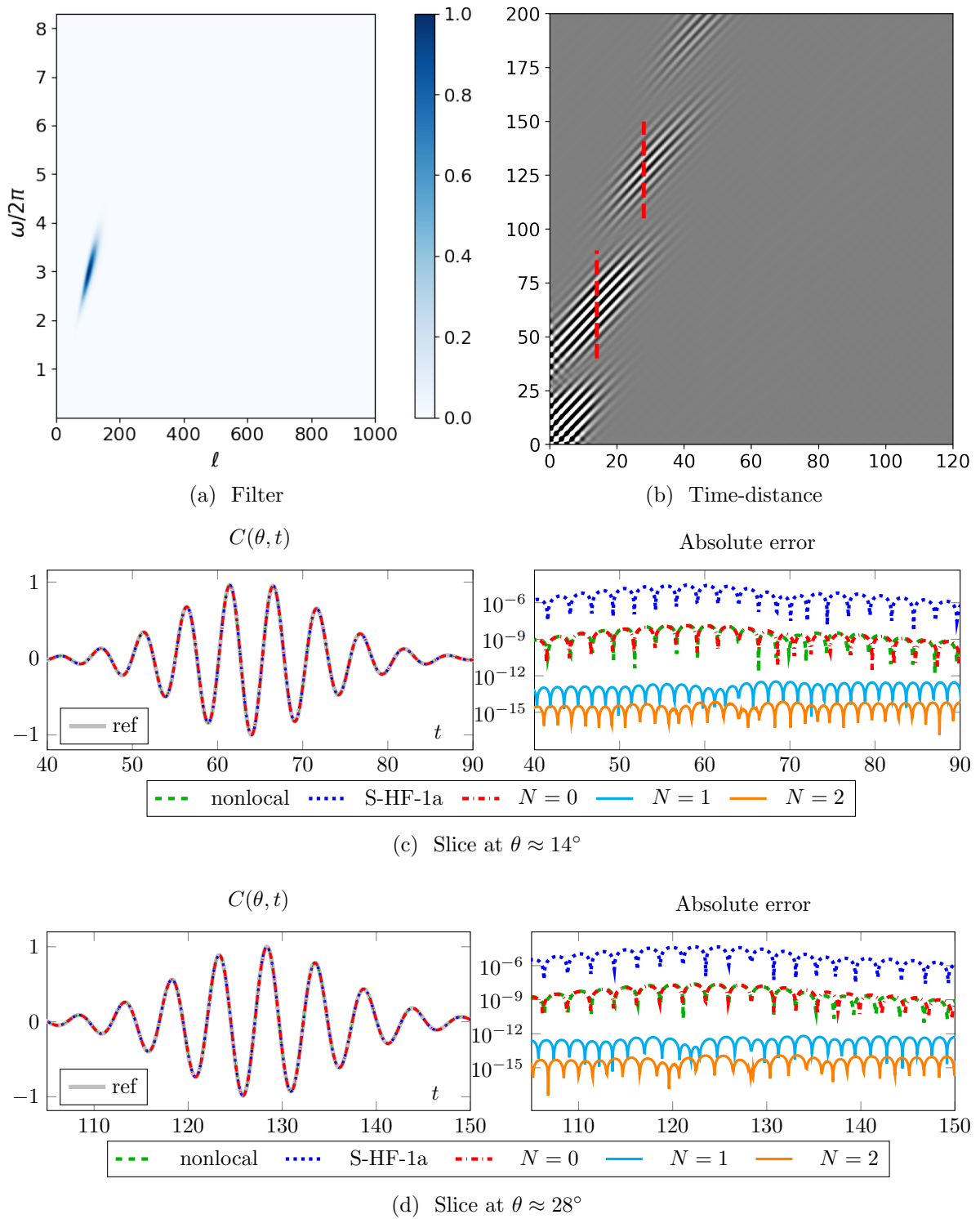
(a) Filter

(b) Time-distance

$C(\theta, t)$

Absolute error

- - - nonlocal   ⋯⋯ S-HF-1a   -·-·- $N = 0$ —— $N = 1$ —— $N = 2$

(c) Slice at $\theta \approx 14°$

$C(\theta, t)$

Absolute error

- - - nonlocal   ⋯⋯ S-HF-1a   -·-·- $N = 0$ —— $N = 1$ —— $N = 2$

(d) Slice at $\theta \approx 28°$

Figure 6.10: Cross-covariance for waves filtered around 3.0 mHz and speed 125.2 km/sec.

(a) Filter

(b) Time-distance

$C(\theta, t)$

Absolute error

ref

nonlocal ...... S-HF-1a ...... $N = 0$ —— $N = 1$ —— $N = 4$

(c) Slice at $\theta \approx 36°$

$C(\theta, t)$

Absolute error

ref

nonlocal ...... S-HF-1a ...... $N = 0$ —— $N = 1$ —— $N = 4$
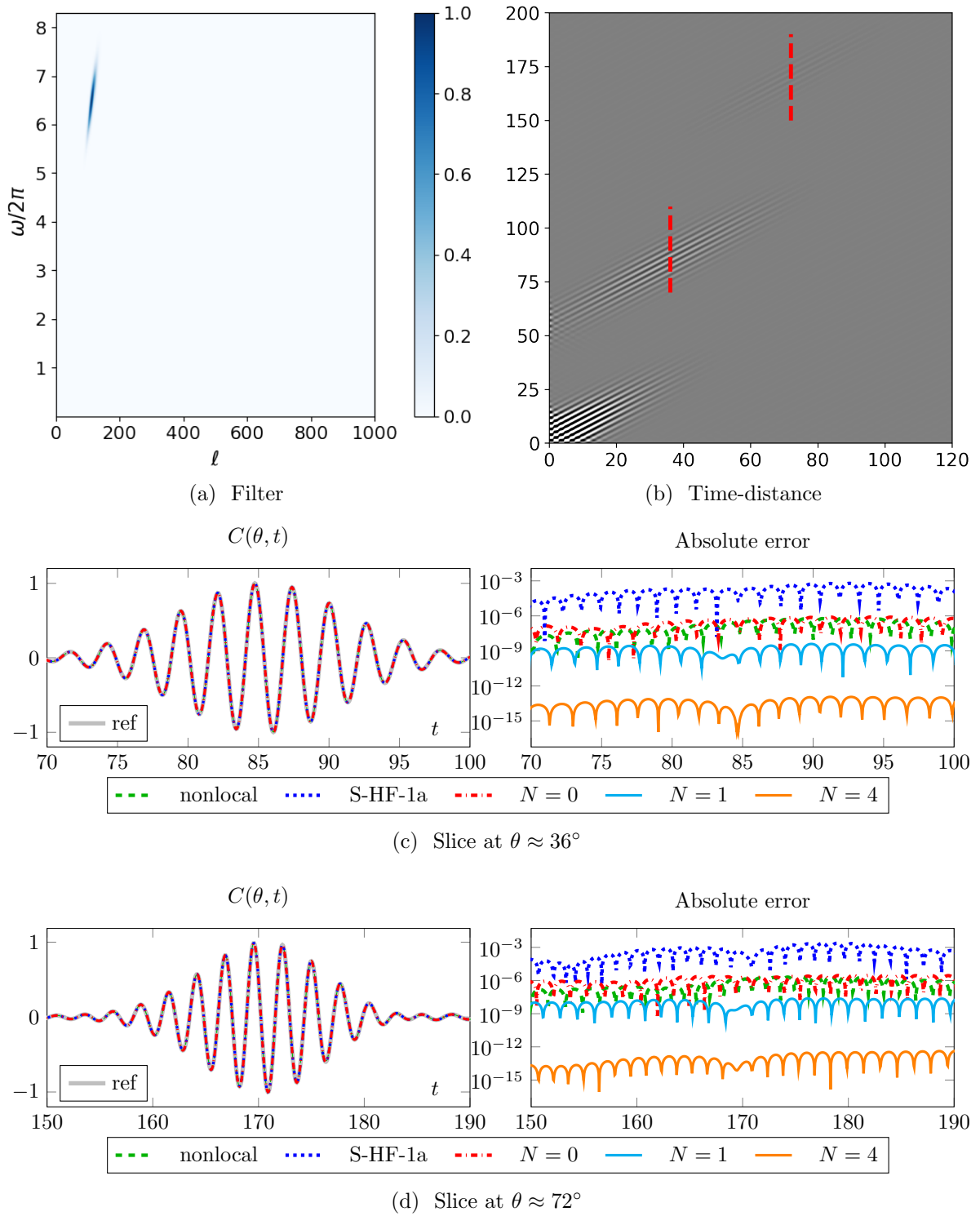
(d) Slice at $\theta \approx 72°$

Figure 6.11: Cross-covariance for waves filtered around 6.5 mHz and speed 250.4 km/sec.

6.3 Evaluation of learned IEs for VAL-C model

This section represents the analogue of the previous section for the VAL-C model. In contrast to the `Atmo` model, an analytic formula for the $dtn^{\text{VAL}-\text{C}}$ function of the VAL-C model is not known. Therefore, equation (6.2) is solved numerically on the interval $[\hat{a}, \hat{R}]$ with $\hat{a} < \hat{R}$ and boundary condition $\tilde{G}_\ell(\hat{a}) = 1$ to obtain $dtn^{\text{VAL-C}}(\lambda_\ell) = -(\partial_{\hat{r}}\tilde{G}_\ell)(\hat{a})$. Here, in contrast to the experiments for the `Atmo` model, a is located directly at the height of the photosphere.[4] Then the minimization problem (4.2)-(4.3) is solved with the reduced, symmetric ansatz to obtain the learned dtn numbers $\text{dtn}_N(\lambda_\ell)$ as before. In comparison with the toy examples considered in Chapter 4, the weights should decay slower for the purpose of helioseismology to resolve modes up to $\ell = 1000$. This motivates the choice $\omega_\ell \sim \exp(-\ell/45)$. The modal Green's function corresponding to an atmosphere modelled by learned IEs instead of meshing it can then be obtained by solving the equation (6.2) on $[0, \hat{a}]$ with boundary condition $(\partial_{\hat{r}}\tilde{G}_\ell^N)(\hat{a}) = -\text{dtn}_N(\lambda_\ell)$.

6.3.1 Power spectrum

The power spectrum is again a reasonable starting point for the investigations. The relative error (6.13) for increasing N is displayed in Fig. 6.12. As for the `Atmo` model, the error decreases exponentially fast as N increases. However, opposed to the `Atmo` model the error does not accumulate around 5.3 mHz. This might be expected since the atmospheric cut-off frequency in the VAL-C atmosphere cannot be described by a constant anymore as for the `Atmo` model. Rather, the error peaks at those frequencies at which poles of $dtn^{\text{VAL-C}}$ occur, cp. Fig. 6.2f. Nevertheless, as N increases these poles can be captured by learned IEs so that very accurate power spectra are obtained.

6.3.2 Expectation value of cross-covariance

Next the influence of the transparent boundary condition on the expectation value of the cross-covariance will be investigated. Firstly, the experiments from Section 6.2.2 are repeated for the VAL-C model. The results for waves with frequencies around 3 mHz and speeds of 125.2 km/s shown in Fig. 6.13 are similar to the `Atmo` model and do not require further discussion. The results for waves with higher frequencies around 6.5 mHz displayed in Fig. 6.14 are more interesting since the cross-covariance exhibits a different structure compared to the `Atmo` model (cp. Fig. 6.11 (c)). The signal splits from a single into a double skip. A similar phenomenon has already been observed in Section 6.1.3 and will be investigated in the context of learned IEs further below. Concerning the accuracy of the transparent boundary condition, the performance remains unaffected.

---

[4]This is computationally more efficient than placing the truncation boundary at 500 km above the photosphere which was only done in Section 6.2 because the atmospheric radiation boundary conditions we compared with rely on this setup.
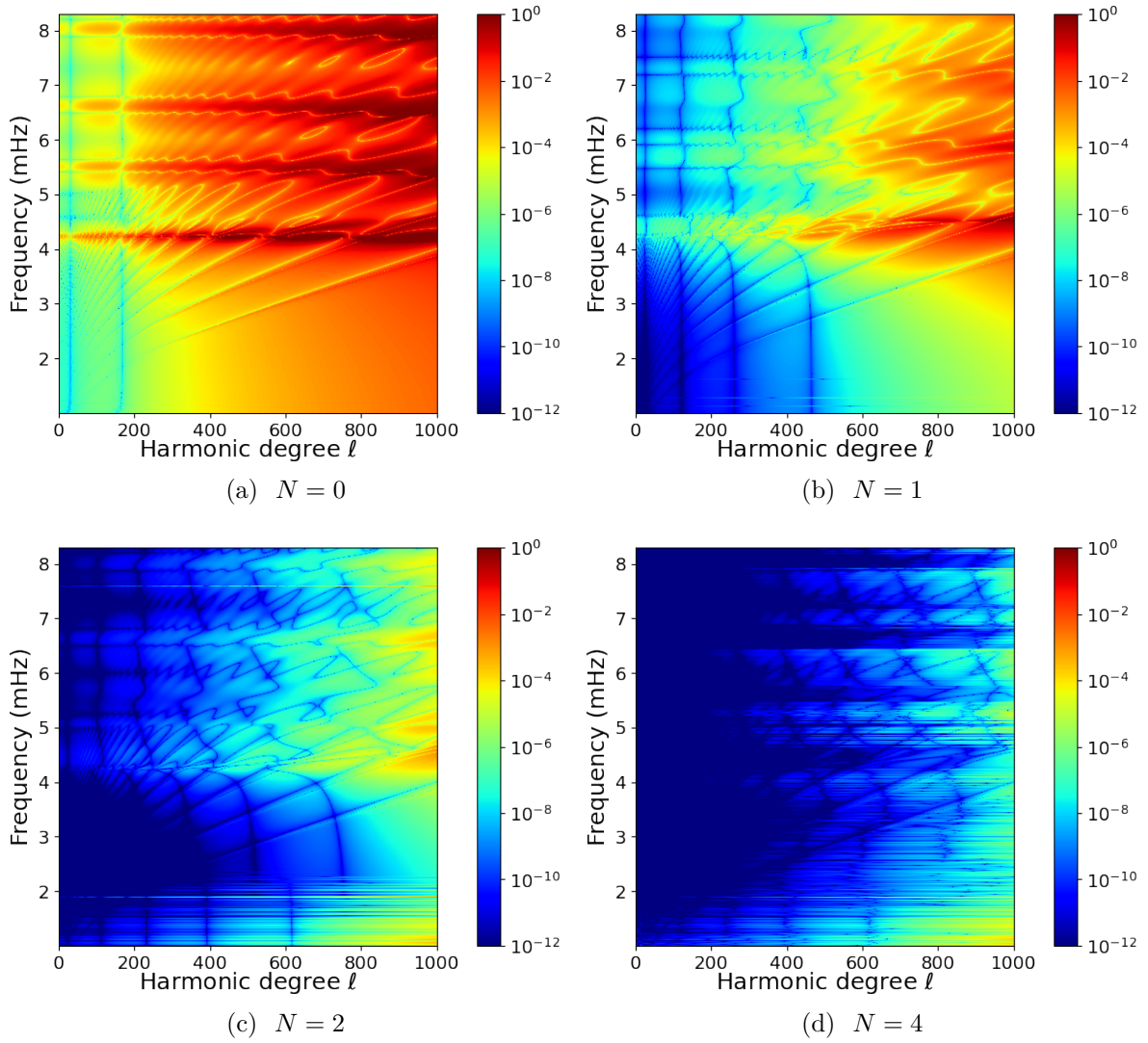
(a) $N = 0$

(b) $N = 1$

(c) $N = 2$

(d) $N = 4$

Figure 6.12: Relative error for power spectra obtained from learned IEs with respect to meshed VAL-C atmosphere.

(a) Filter

(b) Time-distance
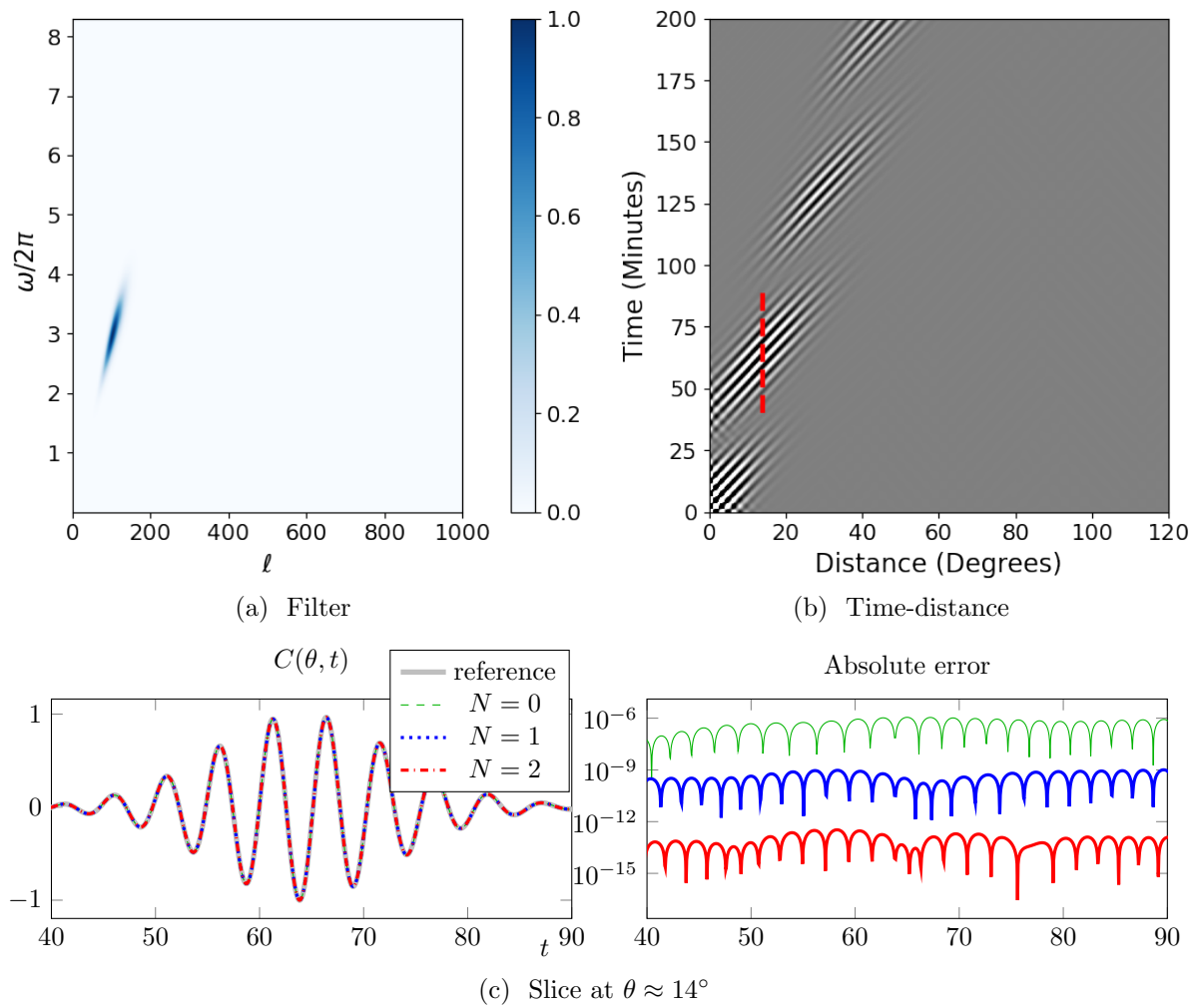
(c) Slice at $\theta \approx 14°$

Figure 6.13: Cross covariance for waves filtered around 3.0 mHz and speed 125.2 km/sec.

(a) Filter

(b) Time-distance

$C(\theta, t)$

ref --- $N = 0$ ······ $N = 1$ ---·--- $N = 3$
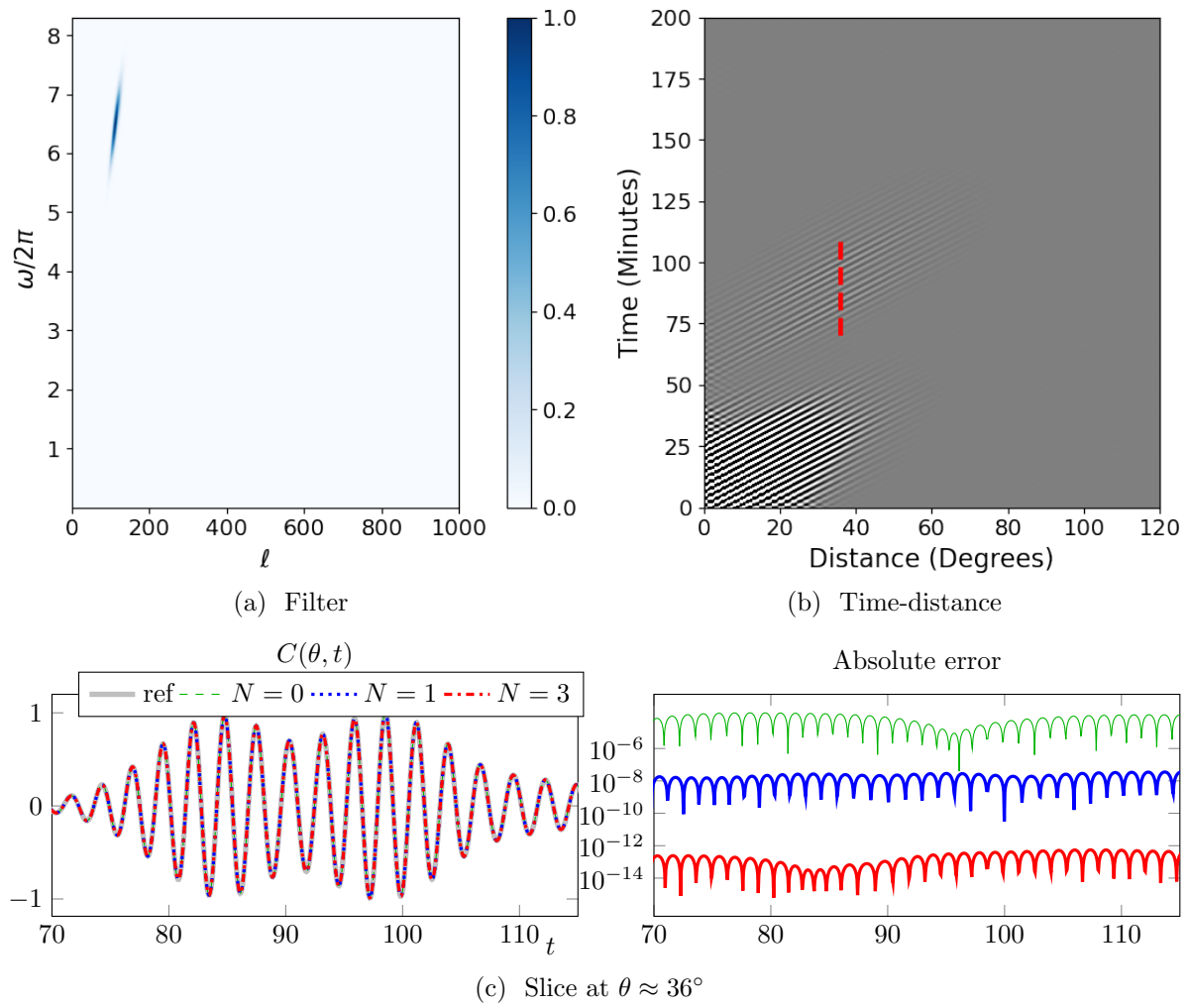
(c) Slice at $\theta \approx 36°$

Figure 6.14: Cross covariance for waves filtered around 6.5 mHz and speed 250.4 km/sec.

Travel-time accuracy

As an additional test it is also interesting to calculate the travel time perturbation defined by (see [GBD$^+$17, Section 8.2])):

$$\delta\tau^N := \int_{-\infty}^{\infty} \bar{W}(C^{\mathrm{ref}} - C^N)\,\mathrm{d}\omega, \qquad\qquad W(\omega) \sim \int_{-\infty}^{\infty} w(t)\partial_t C(\theta, t)e^{i\omega t}\,\mathrm{d}t.$$

Here, $w(t)$ is a temporal window function and $\bar{W}$ denotes the complex conjugate of W.

(a) Filter

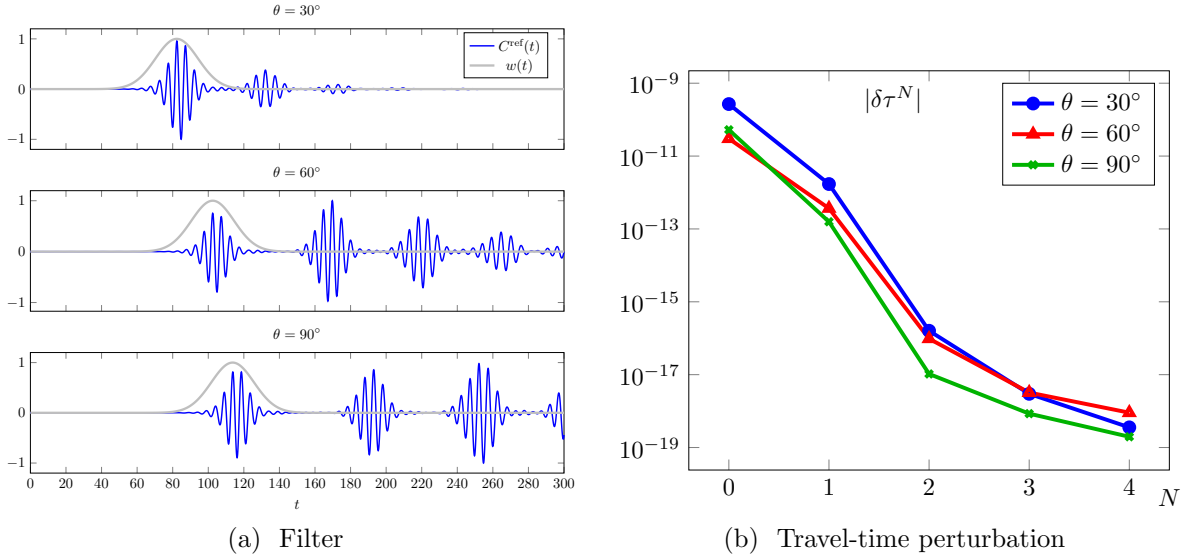(b) Travel-time perturbation

Figure 6.15: Travel-time perturbation $|\delta\tau^N|$ with respect to meshed atmosphere.

For this experiment w was simply chosen as a Gaussian which isolates the first skip of the cross-covariance, see Fig. 6.15 (a). Further, C^{ref} is the reference cross-covariance based on a meshed atmosphere and C^N the cross-covariance obtained from learned IEs using N infinite element DOFs. The travel-time perturbation is shown in Fig. 6.15 (b). An exponentially fast convergence with increasing N is observed.

Double ridge structure for high-frequency

In Section 6.1.3 it was seen that the high-frequency time-distance diagram based on the VAL-C model exhibits a double ridge structure which is in qualitative agreement with observations. It is important to investigate whether this feature can be reproduced using learned IEs. Indeed, Fig. 6.16c demonstrates that the double ridge already appears in the time-distance diagram obtained with learned IEs of the lowest order $N = 0$. In view of the previous experiments, it is not surprising that differences to the reference result with a meshed atmosphere (Fig. 6.16b) can scarcely be discerned by the unaided eye. These findings demonstrate that learned IEs are able to reproduce an important physical feature observed in observations at high-frequency. This is a considerable improvement over the established methods which are less flexible with respect to the atmospheric model and consequently would need to resort to meshing the atmosphere for reproducing the double ridge feature.

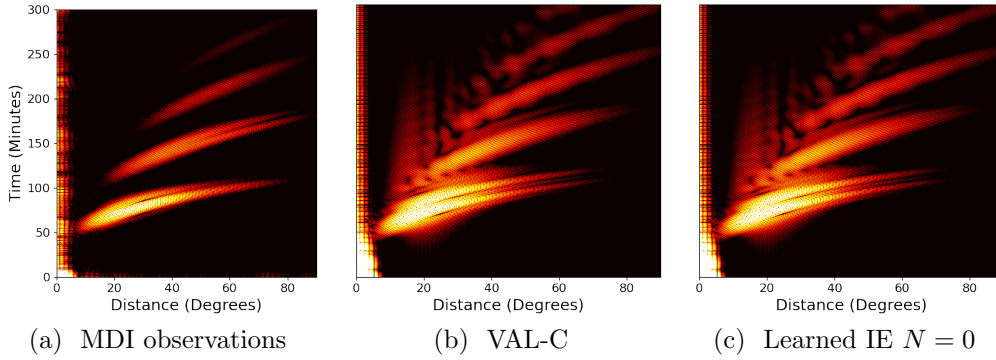(a) MDI observations   (b) VAL-C   (c) Learned IE $N = 0$

Figure 6.16: Double ridge structure of high-frequency time-distance diagram. Shown is the absolute value of the expected cross covariance on a log scale with large values in yellow and small values in black.

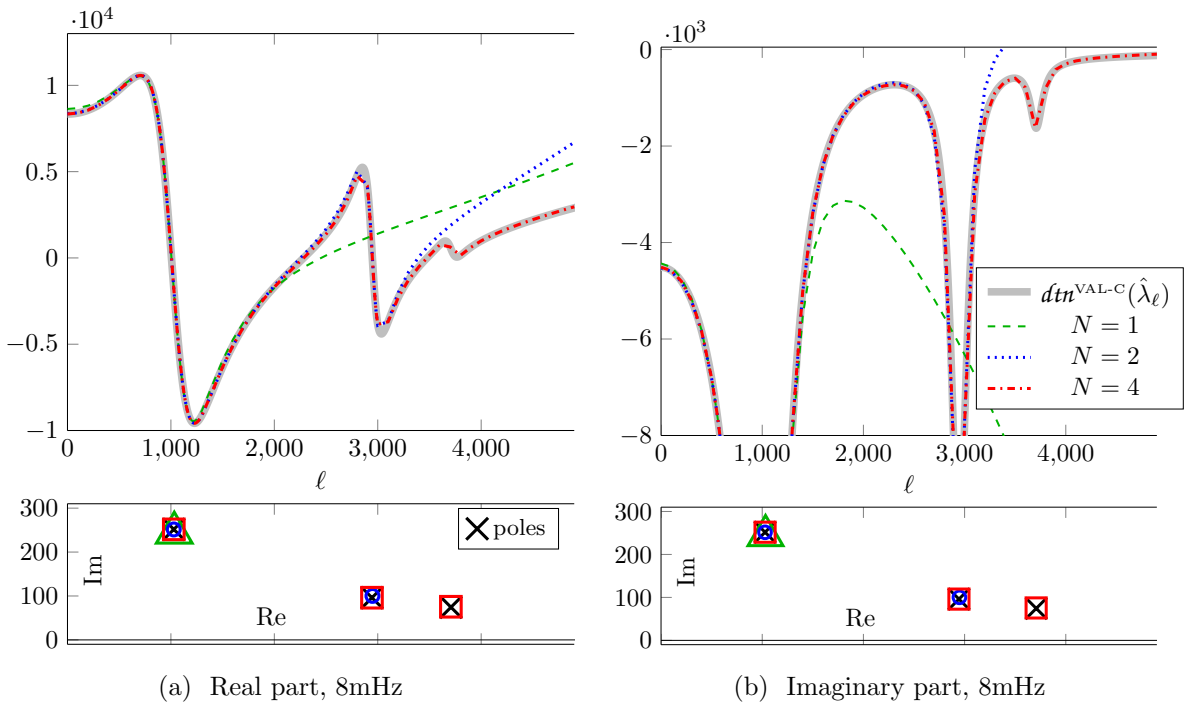(a) Real part, 8mHz   (b) Imaginary part, 8mHz

Figure 6.17: Approximation of $dtn^{\text{VAL-C}}$ by dtn_N from learned IEs for a truncation boundary located at 0.5 Mm above the photosphere. The lower panel displays the exact poles of $dtn^{\text{VAL-C}}$ and their numerical approximation with learned IEs.

Pole structure of the chromosphere

The shortcoming of the `Atmo` model to reproduce the double-ridge feature is clearly related to the fundamentally different structure of $dtn^{\text{VAL-C}}$ and dtn^{Atmo} in the high-frequency regime, see Fig. 6.1. The resonance effects caused by the sharp increase of the sound speed in the upper chromosphere featured in the VAL-C model are reflected in the pole structure of $dtn^{\text{VAL-C}}$. It is an important question for the application whether learned IEs can resolve this complex behavior. The learned dtn_N functions using N infinite element DOFs are displayed in Fig. 6.17a and Fig. 6.17b. Additionally, the poles of the rational function dtn_N are shown. It can be observed that the learned poles are placed exactly at the location of the exact poles, which turn out to be simple. As N increases one pole after another is covered and an excellent approximation of $dtn^{\text{VAL-C}}$ is already achieved with $N = 4$. The situation is reminiscent of the toy example of a discontinuous exterior wave speed treated in Section 4.2.3. This experiment shows that learned IEs are ideally suited for dealing with complicated exterior models as represented by the VAL-C atmosphere.

6.4 Discussion and outlook

The results in this chapter characterize learned IEs as a promising tool for computational helioseismology. A comparison with atmospheric radiation boundary conditions for the `Atmo` model presented in Section 6.2 showed that a significant gain in accuracy could be obtained by replacing these conditions with learned IEs. However, the more interesting domain of application for learned IEs are realistic atmospheric models exemplified here by the VAL-C model for which no other transparent boundary conditions are available to date. In Section 6.3 we demonstrated the great potential of learned IEs in this setting. Nevertheless, our experiments also show that much remains to be done to leverage the full potential of learned IEs in helioseismology. If neither the `Atmo` nor the VAL-C model reproduce the observed power spectra above the atmospheric cut-off accurately (see Fig. 6.4), then the extremely high-accuracy achievable with learned IEs seems currently mostly an overkill. This calls for urgent efforts to improve the model of the solar atmosphere to achieve a better agreement with observations. Some comments on this task are given in the section below.

6.4.1 Tuning the power spectrum

To improve the agreement of the computational results with observations, the coefficients of the background model and the source have to be tuned so that the differences between the computed and observed power spectrum are minimized. We discuss this endeavor separately for frequencies below and above the acoustic cut-off frequency.

Below the acoustic cut-off frequency

As argued previously, for low frequencies the atmospheric model exerts only minor influence on the power spectrum. This is related to the fact that low-frequency waves cannot penetrate past the steep gradients in the surface layers. So this part is about tuning the model of the solar interior.

Sound speed and density in the interior of the Sun are known with high accuracy. For example, the maximum difference between the square of the sound speed of Model S

and the Sun is merely 0.4 % [KSS$^+$97, BCCD$^+$97]. The surface layers pose some special challenge as mentioned earlier. Consequently, concerning sound speed and density one is looking for small and localized corrections to an already accurate background model.

Wave attenuation is arguably less well modelled so far. This is illustrated by an experiment shown in Fig. 6.18. In the left part of the Fig. 6.18a a damping model $2\gamma(\omega, \ell)$ is plotted, which has been determined from the measured FWHM available at the ridges shown in black. The power law damping model from Fig. 6.3b is also shown in the right part of Fig. 6.18a for comparison. Particularly for small ℓ these models differ significantly, which suggests that the power-law damping model may overestimate the attenuation of low degree modes. This point is investigated further in Fig. 6.18 where cuts through the power spectrum computed using the different attenuation models are displayed. While the linewidths are very similar for $\ell = 300$, the ℓ-dependent attenuation model determined from the data leads to considerably smaller linewidths at $\ell = 50$. Implementing such an ℓ-dependent attenuation in a axisymmetric or fully-three dimensional simulations might be challenging though. To complicate[5] matters further the attenuation should also be a function of depth which has e.g. been considered in [BKD04]. Additionally, in the framework of [GBD$^+$17] source covariance amplitude and attenuation are coupled which means that these quantities cannot be updated independently in inversions. All these issues seem to render tuning and implementation of the damping model as one of the outstanding challenges for the scalar problem of helioseismology.

**Above the acoustic cut-off frequency**

High-frequency waves are interesting because they hold the potential of probing the solar chromosphere. The inferences that can be drawn from the presented experiments about the suitability of the atmospheric models for representing the chromosphere are at best fragmentary. On the one hand, the VAL-C model allows to qualitatively reproduce the double-ridge feature seen in observations. On the other hand, the simulated power spectrum Fig. 6.2b shows lines of reduced power which are inconsistent with observations. The Atmo model does not run into the latter problem, yet the simulated power spectrum does not match the observations either. Therefore, it seems to be necessary to solve an inverse problem to invert for the atmosphere. This is an interesting task for future research, which could however be complicated by the increasing influence of the magnetic field in the chromosphere. For example, [FJC$^+$04] propose that high-frequency acoustic waves could be reflected from the magnetic canopy[6]. This is a potential problem since the magnetic field is currently neglected in the scalar equation of helioseismology. Despite these challenges, the inverse problem for the atmosphere will have to be addressed in forthcoming research.

---

[5]A dependence on depth could however be implemented very easily.

[6]The magnetic canopy is a horizontal layer of the magnetic field located in the low chromosphere above field-free regions in the photosphere.
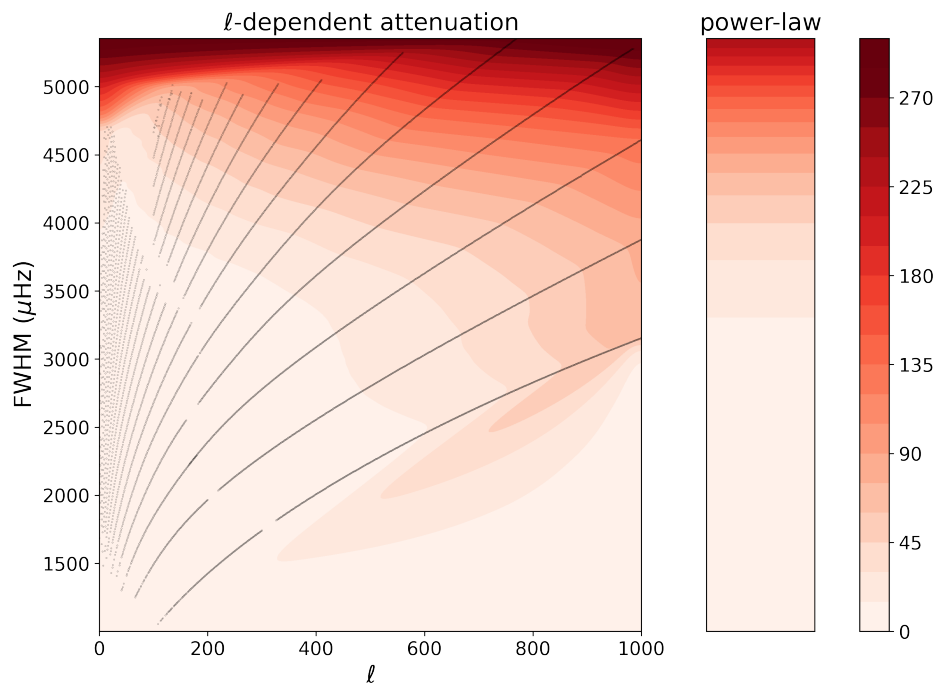
(a) Attenuation model.
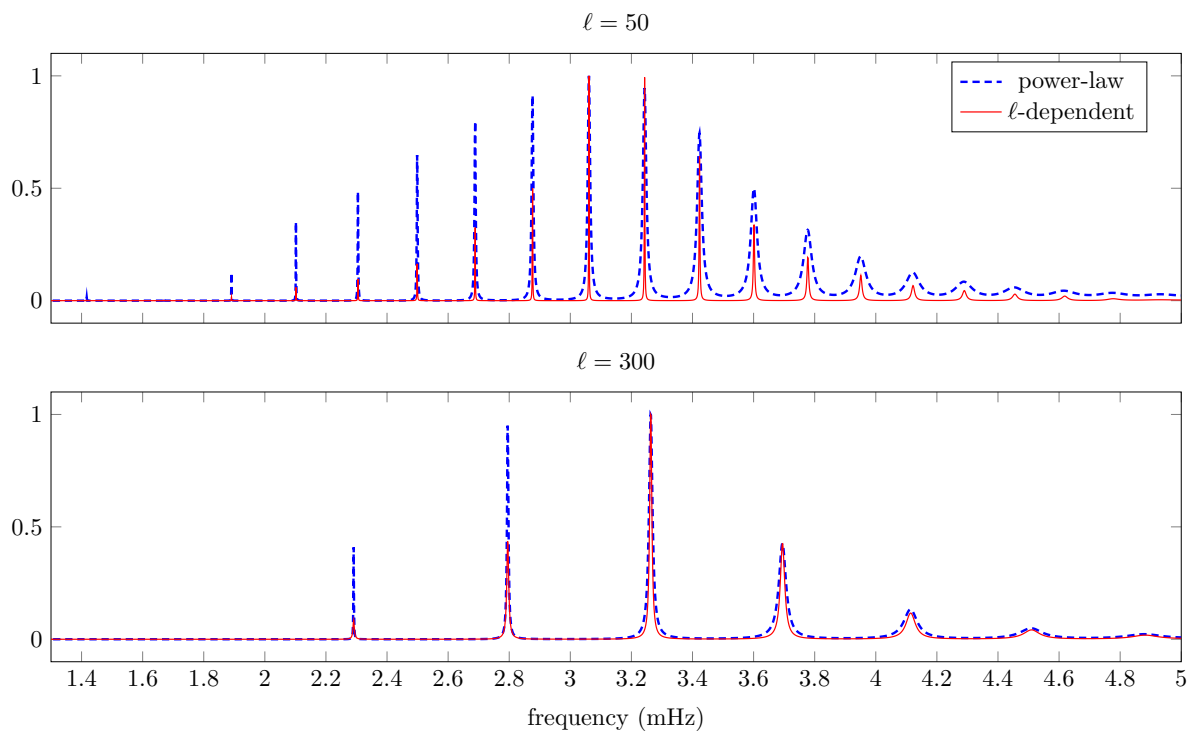
(b) Slices of power spectrum.

Figure 6.18: Damping model according to FWHM from data [KRSSL13, LS15] shown as the black ridges in the figure.

Chapter 7

Learned IEs as transmission conditions in sweeping preconditioners

The first half of this thesis was concerned with the application of learned IEs as transparent boundary conditions. In particular, a single truncation boundary or interface was considered on which learned IEs were applied to implement the $\mathcal{DtN}$ map of an unmeshed exterior domain. This chapter provides two generalizations. Firstly, any finite number of artifical interfaces can be treated. Secondly, these interfaces can be placed inside the computational domain so that part of the exterior domain whose $\mathcal{DtN}$ map needs to be represented by learned IEs is covered by a mesh. This paves the way for applying learned IEs as transmission conditions on subdomain interfaces of domain decomposition methods.

In this chapter a special partition of the domain into sequential layers will be considered which has become very popular in the context of sweeping preconditioners. Recently, these types of preconditioners have achieved impressive results when applied to certain kinds of time-harmonic wave equations. Even though a zoo of different sweeping-type preconditioners exists by now (see [GZ19] for an overview), these methods seem to share as key ingredient that the $\mathcal{DtN}$ operator at subdomain interfaces is approximated by a PML. Unfortunately, this turns out to be the Achilles heel of these approaches because for media containing strong reflections PML does not deliver an accurate approximation of $\mathcal{DtN}$ [GZ18, GZ19, PHL20]. Since the first part of this thesis suggests that learned IEs are very suitable for treating such problems, the objective of this chapter is clear: Replace the PML-based transmission conditions in sweeping preconditioners by learned IEs and investigate whether this can remedy the observed issues.

The outline of this chapter is as follows. Section 7.1 provides a brief and non-technical discussion on why time-harmonic wave equations are difficult to solve numerically, especially in the presence of reflections. Furthermore, recent strategies for solving the arising linear systems are outlined. Section 7.2 introduces sweeping preconditioners in the setting of the double sweep optimized Schwarz method (DOSM) [GZ19]. This framework is advantageous because it exposes the role of $\mathcal{DtN}$ in the sweeping algorithm very clearly and facilitates the use of learned IEs in this setting. Numerical experiments in which sweeping preconditioners with moving PML and learned IE based transmission conditions are compared are presented in Section 7.3. The main aim of these experiments is to investigate to what extent learned IEs can overcome the limitations of moving PMLs

for problems including strong reflections. Even though the examples considered in Section 7.3 are toy problems, they are very useful for understanding the results for the more complicated application to helioseismology presented in Chapter 8.

7.1 Difficulties in solving time-harmonic wave equations

This section provides a brief and necessarily incomplete introduction to the difficulties of solving time-harmonic wave equations. More detailed and thorough expositions may e.g. be found in [EG12, GZ19]. Here, the focus is on problems exhibiting excessive reflection of waves. Reflections may for example be caused by certain types of boundary conditions or drastically varying coefficients and pose the main obstacle for applying sweeping preconditioners to helioseismology.

The pollution effect is one of the main issues in the numerical discretization of time-harmonic wave equations and can already be observed for the constant coefficient Helmholtz equation. It describes the observation that the FEM solution starts to differ significantly from its best approximation in the trial space as the wavenumber increases. Babuška and Sauter [BS97] showed that this effect is in fact unavoidable under reasonable assumptions on the Galerkin FEM in dimensions larger than one. The quest for discretization methods that mitigate the pollution effect is an active area of research. However, to the best of our knowledge, discretization methods that eliminate this problem entirely and could also be applied to helioseismology do not exist. The usual remedy to preserve accuracy of the solution is to scale the number of DOFs in proportion to the wavenumber. In the high-frequency regime this gives rise to extremely large linear systems.

The linear systems arising from the standard variational formulation of the Helmholtz equation are severely ill-conditioned. This can partly be attributed to a lack of coercivity, which is one of the crucial ingredients for establishing condition number bounds for Poisson-like problems. If one is willing to consider non-standard variational formulations and discretization methods, it is possible to obtain positive (semi-)definite system matrices. For example, Moiola and Spence [MS14] introduced a coercive variational formulation for the Helmholtz equation in the interior of a star-shaped domain with impedance boundary conditions or in a star-shaped exterior domain with Dirichlet boundary conditions. This formulation can be obtained in a standard manner by multiplying the equation by a test function and integration by parts, yet conforming discretizations would necessitate the use of finite elements with $\mathcal{C}^1$-regularity. Another approach is to work with least squares formulations which are naturally coercive. For example, in the paper [LMMR00] the Helmholtz equation is reformulated as a first-order system which is then treated by an adapted multigrid solver. Related is the discontinuous Petrov Galerkin (DPG) method, which may be interpreted as a special kind of least squares approach (see [DG11, Section 2] for this interpretation). By breaking the tie between trial and test space, one is at liberty to adapt the latter for obtaining good stability properties. A wavenumber explicit analysis of a DPG method for the Helmholtz equation has been given in [DGMZ12]. Adaptive multigrid solvers for DPG discretizations of acoustic and electromagnetic problems have been presented very recently [PD21]. The DPG approach is potentially also interesting for helioseismology. However, trying to develop an efficient DPG solver for the solar application would be a new research project on its own and is certainly beyond the

scope of this work. Here we want to investigate whether the ill-conditioning inherent in the standard variational formulation can be fixed by developing suitable preconditioners.

When choosing a solver for large sparse linear systems one is faced with the decision between iterative and direct solution approaches. While direct solvers are very attractive due to their natural robustness, their complexity in time and memory consumption scales suboptimally[1] for applications in three dimensions. Many iterative solution approaches instead, e.g. Krylov subspace methods, only require the application of matrix-vector products and are therefore potentially much better suited for large-scale applications. The catch is that the rate of convergence of Krylov methods depends on the condition number, respectively spectrum of the linear system, so that preconditioning is almost always required to arrive at an accurate solution in an acceptable number of iterations. Unfortunately, good preconditioners for time-harmonic wave equations are difficult to find.

For Poisson-like problems the efficiency of multigrid methods as preconditioners or even as stand-alone solvers is hard to compete with. Unfortunately, standard multigrid methods are not suitable for solving time-harmonic wave equations, see e.g. [EG12, Section 3.3] for explanations. Strategies to overcome this problem seem to go in two different directions. Either, one tries to fix the multigrid method by suitable adaptations, see e.g. the wave-ray multigrid approach of [BL97], or one applies the multigrid method for solving a closely related, easier problem whose solution operator may be used for preconditioning the original problem. A popular approach for the latter option is the so called shifted Laplacian preconditioner [EVO04]. Here one considers a Helmholtz problem with a wavenumber that is shifted into the complex plane, i.e. $k^2 \mapsto k^2 + i\varepsilon(k)$. If $\varepsilon = \mathcal{O}(k^2)$ the shifted problem can be solved efficiently using standard multigrid [CG17]. On the other hand, in order to guarantee that GMRES converges independently of k when preconditioned by means of the inverse of the shifted operator only a shift of $\varepsilon = \mathcal{O}(k)$ can be tolerated [GGS15]. Obviously, there is a significant gap between how much one is allowed to shift to maintain a good preconditioner and how much one needs to shift to obtain an operator which can be inverted efficiently. In numerical experiments not shown here we tried to apply this preconditioning strategy to helioseismology but the mentioned gap was considerably too large to overcome.

In 2011 Engquist and Ying [EY11b] proposed sweeping preconditioners as a new approach to tackle high-frequency Helmholtz problems. The method has been quickly extended to other time-harmonic wave equations like the Maxwell system [TEY12] and sparked a wave of further research articles presenting variants or extensions (see e.g. [CX13a, CX13b, Sto13, TPEY14, ZND16] only to name a few). The main idea is to partition the domain into a sequence of layers and sweep through this layered partition in a sequential fashion. During the sweep the subdomain problems are solved one after another and information is transmitted via appropriate transmission conditions at subdomain interfaces. At the continuous level this follows the principle of optimized Schwarz methods [Gan06], while at the discrete level it becomes a block LU factorization [GN00] (see [GZ19] for a proof of this equivalence). Despite the underlying algorithmic structure being known for quite a while, the sweeping algorithm was not overly successful until Engquist and Ying [EY11b] added a missing ingredient - a highly accurate approximation of $\mathcal{DtN}$ provided

---

[1]Here we are referring to the vanilla version of multifrontal solvers based on nested dissection which yield the exact solution apart from round-off errors. New developments which aim to improve the complexity, e.g. by applying compression algorithms, will be mentioned later.

by a moving PML[2]. From then on sweeping preconditioners emerged as one of the most efficient strategies for solving time-harmonic wave equations in the high-frequency regime.

Unfortunately, PMLs only deliver accurate approximations of $\mathcal{DtN}$ for media that are free of reflections. This may lead to deteriorating performance or even complete breakdown of the preconditioner if reflections occur in abundance. To the best of our knowledge, this has been pointed out first by Gander and Zhang in [GZ18] who showed in a numerical experiment featuring a wavenumber with strong contrast[3] between layers that the performance of sweeping preconditioners deteriorates drastically as the contrast increases. They concluded with the final statement that "One must therefore investigate an approximation different from PML for the Dirichlet to Neumann operator in the case of non-constant wave numbers." Chapter 7 and Chapter 8 of this thesis aim to investigate the question whether learned IEs could provide such an approximation. Our answer will reflect the current limitation of learned IEs which require separability of the PDE (and tensor-product discretizations) to deliver accurate results. Namely, if the equation is separable, then learned IEs are able to provide accurate and efficient approximations of $\mathcal{DtN}$ even if an excessive number of reflections is present. However, once separability is slightly violated learned IEs cease to be efficient in presence of reflections since they are only able to represent the $\mathcal{DtN}$ maps of separable equations, which can differ significantly from the true $\mathcal{DtN}$ maps of the perturbed non-separable equation whose solution is sought. Therefore, further research is needed to develop learned IEs which can be applied to non-separable equations.

On the other hand, also alternative solution strategies should be considered. Recently, progress has been achieved in controllability methods [GNTT20, AGR20] which transform the Helmholtz equation into the time-domain and seek the corresponding time-harmonic solution there. These methods rely on an efficient solver for wave equations in the time-domain. This seems almost impossible to realize for the helioseismology problem because the majority of the mesh elements are extremely small as a consequence of the strong refinement in the surface layers (see e.g. Fig. 8.2). Hence, extremely restrictive CFL conditions (in the best case local) are to be expected which would render time-domain solutions very expensive. This would also limit the applicability of the related time-domain preconditioner recently proposed by Stolk [Sto21]. Nevertheless, time-domain methods may certainly be of interest for other type of applications.

In contrast to most iterative solution approaches, direct solvers are usually very robust with respect to problem and discretization parameters. They deliver the exact solution of the linear system up to round-off errors caused by finite precision arithmetic. Direct solver are also advantageous for multiple right hand sides since once a factorization of the matrix has been computed solving the linear system is comparatively cheap. Popular examples of direct solvers include `PARDISO` [SGFS01], `MUMPS` [ADLK01] and `SuperLU` [Li05]. The inherent robustness of these solvers comes at the expense of superlinear complexity in time and storage requirements. A recent approach to improve on these drawbacks is motivated by the observation that in applications it is often not necessary to solve the linear systems down to machine accuracy. This allows to introduce certain approximations during the factorization stage which can reduce storage costs and speed up computations. A potential target for these approximations are off-diagonal blocks of dense Schur complement matrices which arise during the factorization process. Depending on the PDE in question

---

[2]Alternatively, sweeping preconditioners can be implemented using the framework of hierarchical matrices [EY11a].

[3]Similar to the jumping wavenumber (2.40) introduced in Section 2.2.1.

these off-diagonal blocks may exhibit a low-rank structure which can be utilized for their efficient approximation and storage by compression techniques, e.g. using $\mathcal{H}$-matrices [Hac99], Block Low-Rank (BLR) format [AAB$^+$15] or hierarchically semiseparable matrices (HSS) [XCGL10]. The success of this approach is then determined by the ability of these techniques to efficiently represent the off-diagonal blocks for the time-harmonic wave equation to be solved. We refer to references [BH07, EZ18, GS17] for theoretical and numerical studies for the Helmholtz equation. Applications to seismic wave propagation suggest that significant improvements over plain direct solvers can be obtained [WdHX11, WAB$^+$13]. Further improvements may be expected by utilizing special compression techniques which are adapted to the highly oscillatory nature of time-harmonic wave equations. For instance, directional $\mathcal{H}^2$-matrices [Bï7, BM17] are a promising development, albeit the corresponding efficient matrix arithmetic is still incomplete. Very recently, Liu et al. [LGCL20] proposed a solver which leverages the butterfly algorithm [MB96] during the compression stage. This algorithm has been successfully applied to represent e.g. Fourier transforms [Yin09] and Fourier integral operators [CDY09] which demonstrates its potential for highly-oscillatory problems. The authors conclude that the resulting solver is able to attain quasi linear complexity in time and memory consumption for Helmholtz and Maxwell problems in the high-frequency regime. As the code is publicly available, the performance of this solver for applications in helioseismology should be investigated in future studies.

7.2 Double sweep optimized Schwarz method

The landmark paper [EY11b] by Engquist and Ying triggered a surge of variations on the sweeping algorithm. These algorithms often differ significantly in their derivations and formulations which renders it extremely difficult to grasp their underlying similarities and differences. In their review paper [GZ19] Gander and Zhang showed that many of these methods can be understood in the framework of the double sweep optimized Schwarz method (DOSM). Here, as in [PHL20], only a special case of DOSM will be considered which gives rise to the original sweeping preconditioner of Engquist and Ying. This allows to simplify the presentation considerably and is sufficient for the purpose of this thesis. Readers interested in the more general case (overlapping partitions, different transmission operators ...) are referred to the extensive review article [GZ19].

For $J \in \mathbb{N}$ let

$$\bar{\Omega}_{\text{int}} = \bigcup_{j=0}^{J-1} \bar{\Omega}_j, \quad \Omega_j \cap \Omega_k = \emptyset \text{ for } j \neq k,$$

be a non-overlapping decomposition of the interior domain into sequential layers Ω_j for $j = 0, \ldots, J-1$, see Fig. 7.1 or Fig. 7.2 for examples. The layering of the partition implies a distinguished direction into which the sweep is performed. The forward sweep starts at Ω_0 and proceeds along the direction of increasing j to subdomain Ω_{J-2}. The backward sweep returns from Ω_{J-1} in reverse order to Ω_0. Fig. 7.1 shows a sketch in which the subdomains have been arranged from left to right, i.e. starting with Ω_0 on the far left and ending with Ω_{J-1} on the far right. In this case, the sweep operates along the horizontal direction starting from the left with the forward sweep. Proceeding with the notation, we define the subdomain interfaces $\Gamma_j := \partial\Omega_{j-1} \cap \partial\Omega_j$ for $j = 1, \ldots, J-2$. Additionally, it is convenient to introduce two empty interfaces $\Gamma_0 = \Gamma_J = \emptyset$.
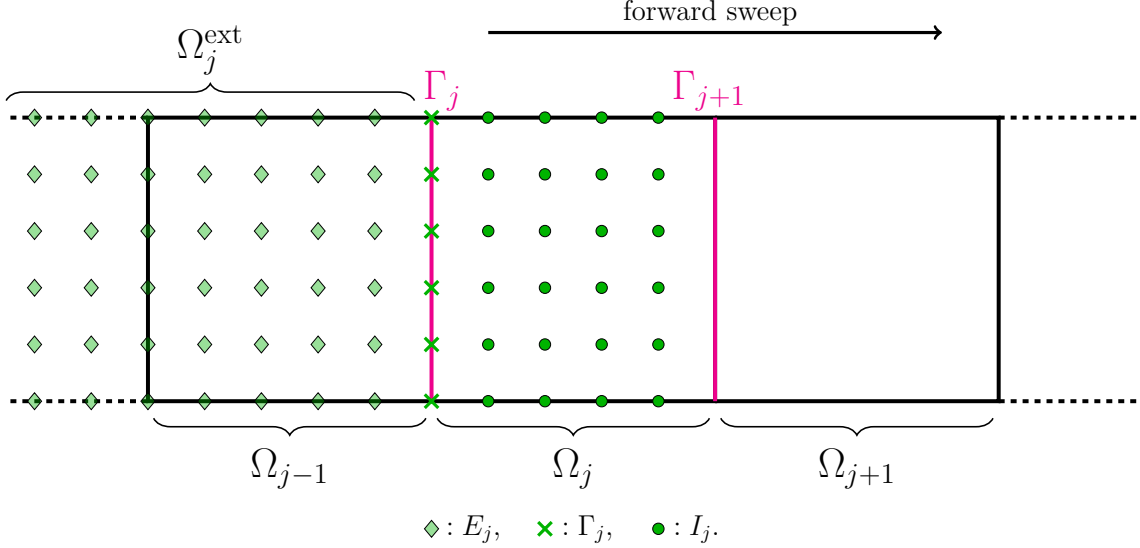
Figure 7.1: Decomposition of domain into layers and corresponding partition of the DOFs.

Let $\mathcal{L}$ denote the generic elliptic differential operator of equation (2.15). From (2.16) recall the disjoint decomposition of $\partial\Omega_{\text{int}}$ into a Dirichlet part, a Neumann part and a transparent boundary. To define the sweeping algorithm, it is convenient to summarize all these boundary conditions as $\mathcal{B}_{\partial\Omega_{\text{int}}} u = g$ on $\partial\Omega_{\text{int}}$, where the operator $\mathcal{B}_{\partial\Omega_{\text{int}}}$ is defined piecewise as

$$
\begin{aligned}
u &= g_D && \text{on } \Gamma_D, \\
\nabla u \cdot \mathbf{n}_{\Gamma_N} &= g_N && \text{on } \Gamma_N, \\
\mathcal{M}_\Gamma \mathcal{D}t\mathcal{N}(u) + \nabla u \cdot \mathbf{n}_\Gamma &= 0 && \text{on } \Gamma,
\end{aligned}
$$

for data $g_D \in H^{1/2}(\Gamma_D)$ and $g_N \in H^{-1/2}(\Gamma_N)$. Contrary to Chapter 2, we will allow for $\Gamma = \emptyset$ here. In the numerical experiments presented in Section 7.3 also periodic boundary conditions on part of $\partial\Omega_{\text{int}}$ will be used which can be incorporated into the corresponding finite element space.

In Section 7.2.1 the sweeping algorithm will first be stated at the continuous level for solving the equation $\mathcal{L}u = f$ in Ω_{int}. A corresponding matrix version at the discrete level follows in Section 7.2.2. In this context we also explain how learned IE based transmission conditions can be realized. Aspects of computational complexity are dicussed in Section 7.2.3.

7.2.1 PDE level

Let u_j denote the restriction of u to subdomain Ω_j. To formulate the (exact) transmission conditions, the $\mathcal{D}t\mathcal{N}_j$ map associated with the interface of Γ_j is required. To this end, denote by $\overline{\Omega_j^{\text{ext}}} = \cup_{i=0}^{j-1} \bar{\Omega}_i$ the exterior domain which couples via the interface Γ_j to Ω_j. Then $\mathcal{D}t\mathcal{N}_j$ is given by

$$
\mathcal{D}t\mathcal{N}_j : \ w \mapsto -\left(a_{ij}\right) \nabla v \cdot \mathbf{n}_{\Gamma_j}, \tag{7.1}
$$

146

where v solves

$$
\begin{aligned}
\mathcal{L}v &= 0 && \text{in } \Omega_j^{\text{ext}}, \\
\mathcal{B}_{\partial \Omega_{\text{int}}} v &= g && \text{on } \partial \Omega_j^{\text{ext}} \setminus \partial \Omega_j, \\
v &= w && \text{on } \Gamma_j.
\end{aligned}
\tag{7.2}
$$

Now all definitions are available to understand DOSM at the continuous level as presented in Algorithm 1. This algorithm starts from an initial guess $(u_j^{(0)})_{j=0}^{J-1}$ and iteratively computes an approximate solution. For completeness, one should mention that the transmission conditions are omitted if $\Gamma_j = \emptyset$, i.e. for $j = 0$ and $j = J$. A convergence analysis of DOSM is available in [GZ19]. Under reasonable assumptions, the solution $(u_j^{(1)})_{j=0}^{J-1}$ obtained after one double sweep across the domain already coincides with the exact solution provided that exact $\mathcal{DtN}_j$ maps are used as transmission operators. Of course, this is infeasible in practice due to the cost associated with computing the exact $\mathcal{DtN}_j$ operators. Therefore, accurate and efficient approximations of $\mathcal{DtN}_j$ are required on all interfaces Γ_j between subdomains. To this end, similar strategies as in the case of transparent boundary conditions can be applied. More detail is given in the next section which introduces DOSM at the matrix level.

Algorithm 1: Double sweep optimized Schwarz method at the PDE level

---

Forward sweep: Given the last iterate $u_j^{(n-1)}$ in $\Omega_j, j = 0, \ldots, J-1$ solve successively for $j = 0, \ldots, J-2$

$$
\begin{aligned}
\mathcal{L}u_j^{(n-\frac{1}{2})} &= f && \text{in } \Omega_j, \\
\mathcal{B}_{\partial \Omega_{\text{int}}} u_j^{(n-\frac{1}{2})} &= g && \text{on } \partial \Omega_{\text{int}} \cap \partial \Omega_j, \\
-(a_{ij}) \nabla u_j^{(n-\frac{1}{2})} \cdot \mathbf{n}_{\Gamma_j} + \mathcal{DtN}_j u_j^{(n-\frac{1}{2})} &= -(a_{ij}) \nabla u_{j-1}^{(n-\frac{1}{2})} \cdot \mathbf{n}_{\Gamma_j} + \mathcal{DtN}_j u_{j-1}^{(n-\frac{1}{2})} && \text{on } \Gamma_j, \\
u_j^{(n-\frac{1}{2})} &= u_{j+1}^{(n-1)} && \text{on } \Gamma_{j+1}.
\end{aligned}
$$

Backward sweep: Solve successively for $j = J-1, \ldots, 0$

$$
\begin{aligned}
\mathcal{L}u_j^{(n)} &= f && \text{in } \Omega_j, \\
\mathcal{B}_{\partial \Omega_{\text{int}}} u_j^{(n)} &= g && \text{on } \partial \Omega_{\text{int}} \cap \partial \Omega_j, \\
-(a_{ij}) \nabla u_j^{(n)} \cdot \mathbf{n}_{\Gamma_j} + \mathcal{DtN}_j u_j^{(n)} &= -(a_{ij}) \nabla u_{j-1}^{(n-\frac{1}{2})} \cdot \mathbf{n}_{\Gamma_j} + \mathcal{DtN}_j u_{j-1}^{(n-\frac{1}{2})} && \text{on } \Gamma_j, \\
u_j^{(n)} &= u_{j+1}^{(n)} && \text{on } \Gamma_{j+1}.
\end{aligned}
$$

7.2.2 Matrix level

Let L be a matrix representation of the differential operator $\mathcal{L}$ obtained from a suitable discretization. Although this thesis only considers conforming finite element discretizations, various other approaches would likewise be possible. To state DOSM at the matrix level, a suitable partition of the DOFs is required. Let $\underline{v}$ denote a discrete vector consisting of all the DOFs. Restrictions to a subset of DOFs will be denoted by attaching suitable subscripts. The following definitions are illustrated in Fig. 7.1.

- Let $\underline{v}_{\Gamma_j}$ denote the DOFs associated with the interface Γ_j,

- Let $\underline{v}_{I_j}$ denote the interior DOFs of subdomain Ω_j. These are obtained by taking all DOFs in $\bar{\Omega}_j$ and removing the contributions from the interfaces Γ_j and Γ_{j+1}.

- Let $\underline{v}_{E_j}$ denote the DOFs of the exterior domain Ω_j^{ext}. This does not include the DOFs on the interface Γ_j, yet the DOFs on all other interfaces contained in Ω_j^{ext}.

According to these definitions, the discrete solution on $\bar{\Omega}_j$ can be written as $\underline{u}_j = (\underline{u}_{\Gamma_j}, \underline{u}_{I_j}, \underline{u}_{\Gamma_{j+1}})$. Moreover, the partition above induces a corresponding partition of the matrix L. For example, $L_{I_j I_j}$ denotes the submatrix of L associated with the contribution from the interior of Ω_j. On an interface Γ_j the matrix will be split into $L_{\Gamma_j \Gamma_j}^{\text{int}} + L_{\Gamma_j \Gamma_j}$, where $L_{\Gamma_j \Gamma_j}$ contains contributions from the interface only and $L_{\Gamma_j \Gamma_j}^{\text{int}}$ contains coupling terms from the interior of Ω_j or Ω_{j-1} especially volume terms close to Γ_j. Furthermore, let $R^{\Gamma_j}\underline{v} = v_{\Gamma_j}$ denote the restriction of $\underline{v}$ to Γ_j. Using these definitions Algorithm 2 states DOSM at the matrix level.

Algorithm 2: DOSM at matrix level

Forward sweep: Given the last iterate $\underline{u}_j^{(n-1)}$ in Ω_j, $j = 0, \ldots, J-1$ do

successively for $j = 0, \ldots, J-2$: Set $\underline{u}_{\Gamma_{j+1}}^{(n-\frac{1}{2})} = R^{\Gamma_{j+1}}\underline{u}_{j+1}^{(n-1)}$ and solve

$$\begin{bmatrix} L_{I_j I_j} & L_{I_j \Gamma_j} \\ L_{\Gamma_j I_j} & L_{\Gamma_j \Gamma_j}^{\text{int}} + M_j \, \text{DtN}_j^{\text{ext}} \end{bmatrix} \begin{bmatrix} \underline{u}_{I_j}^{(n-\frac{1}{2})} \\ \underline{u}_{\Gamma_j}^{(n-\frac{1}{2})} \end{bmatrix} = \begin{bmatrix} \underline{f}_{I_j} - L_{I_j \Gamma_{j+1}}\underline{u}_{\Gamma_{j+1}}^{(n-\frac{1}{2})} \\ \underline{\tilde{f}}_{\Gamma_j}^{(n-\frac{1}{2})} \end{bmatrix}, \qquad (7.3)$$

with

$$\underline{\tilde{f}}_{\Gamma_j}^{(n-\frac{1}{2})} = \underline{f}_{\Gamma_j} - L_{\Gamma_j I_{j-1}}\underline{u}_{I_{j-1}}^{(n-\frac{1}{2})} + \left(M_j \text{DtN}_j^{\text{ext}} - L_{\Gamma_j \Gamma_j} \right) R^{\Gamma_j}\underline{u}_{j-1}^{(n-\frac{1}{2})}.$$

Backward sweep: Do successively for $j = J-1, \ldots, 0$: Set $\underline{u}_{\Gamma_{j+1}}^{(n)} = R^{\Gamma_{j+1}}\underline{u}_{j+1}^{(n)}$ and solve

$$\begin{bmatrix} L_{I_j I_j} & L_{I_j \Gamma_j} \\ L_{\Gamma_j I_j} & L_{\Gamma_j \Gamma_j}^{\text{int}} + M_j \, \text{DtN}_j^{\text{ext}} \end{bmatrix} \begin{bmatrix} \underline{u}_{I_j}^{(n)} \\ \underline{u}_{\Gamma_j}^{(n)} \end{bmatrix} = \begin{bmatrix} \underline{f}_{I_j} - L_{I_j \Gamma_{j+1}}\underline{u}_{\Gamma_{j+1}}^{(n)} \\ \underline{\tilde{f}}_{\Gamma_j}^{(n-\frac{1}{2})} \end{bmatrix}.$$

To complete the description of Algorithm 2, the matrices M_j and $\text{DtN}_j^{\text{ext}}$ still need to be specified. Here, M_j is simply the mass matrix of the FEM discretization on Γ_j. The matrix $\text{DtN}_j^{\text{ext}}$ is the discrete representation of $\mathcal{DtN}_j$ associated with the particular discretization L of the differential operator $\mathcal{L}$ including boundary conditions[4]. It is obtained by taking the Schur complement with respect to all DOFs exterior to Ω_j, i.e.

$$M_j \, \text{DtN}_j^{\text{ext}} = \left(L_{\Gamma_j \Gamma_j} - L_{\Gamma_j E_j} L_{E_j E_j}^{-1} L_{E_j \Gamma_j} \right). \qquad (7.4)$$

Practical realization of DOSM

DOSM at the matrix level is usually implemented differently than described in Algorithm 2. This is related to the fact that the matrix $\text{DtN}_j^{\text{ext}}$ is dense. Hence, to solve the linear system (7.3) directly, one would first need to invest the effort to assemble this matrix, which would require inversion of $L_{E_j E_j}$ and then deal with the resulting dense block associated with the DOFs on Γ_j. A better approach already known from Section 3.1.1,

---

[4]This could also include radiation conditions imposed on the artificial truncation boundary Γ.

which preserves the sparsity structure of L and also paves the way for introducing approximations of $\text{DtN}_j^{\text{ext}}$, is to write the linear system in Algorithm 2 in the equivalent form

$$
\begin{bmatrix}
L_{I_j I_j} & L_{I_j \Gamma_j} & 0 \\
L_{\Gamma_j I_j} & L_{\Gamma_j \Gamma_j}^{\text{int}} + L_{\Gamma_j \Gamma_j} & L_{\Gamma_j E_j} \\
0 & L_{E_j \Gamma_j} & L_{E_j E_j}
\end{bmatrix}
\begin{bmatrix}
\underline{u}_{I_j}^{(n-\frac{1}{2})} \\
\underline{u}_{\Gamma_j}^{(n-\frac{1}{2})} \\
\underline{u}_{E_j}^{(n-\frac{1}{2})}
\end{bmatrix}
=
\begin{bmatrix}
\underline{f}_{I_j} - L_{I_j \Gamma_{j+1}} \underline{u}_{\Gamma_{j+1}}^{(n-\frac{1}{2})} \\
\tilde{\underline{f}}_{\Gamma_j}^{(n-\frac{1}{2})} \\
0
\end{bmatrix}.
\tag{7.5}
$$

Here, the dense block on Γ_j has been exchanged for a sparse linear system which, however, now additionally involves all DOFs in Ω_j^{ext}. It is easy to see that this does not result in an efficient numerical method yet. For example, on the last subdomain $j = J - 1$ the complete original linear system would need to be solved. Therefore, to obtain an efficient algorithm, approximations of $\text{DtN}_j^{\text{ext}}$ have to be introduced which allow to compress the exterior system in Ω_j^{ext}.

From Section 3.1 it is already known how to derive and even assess the accuracy of a variety of such approximation for discretizations L which are of tensor-product form. Hence, to utilize this framework it will be assumed as in equation (3.7) that

$$
\begin{bmatrix}
L_{\Gamma_j \Gamma_j} & L_{\Gamma_j E_j} \\
L_{E_j \Gamma_j} & L_{E_j E_j}
\end{bmatrix}
=
\begin{bmatrix}
A_{\Gamma_j \Gamma_j} & A_{\Gamma_j E_j} \\
A_{E_j \Gamma_j} & A_{E_j E_j}
\end{bmatrix}
\otimes M_j
+
\begin{bmatrix}
B_{\Gamma_j \Gamma_j} & B_{\Gamma_j E_j} \\
B_{E_j \Gamma_j} & B_{E_j E_j}
\end{bmatrix}
\otimes K_j,
\tag{7.6}
$$

where K_j represents the stiffness matrix on Γ_j as usual. As shown in Proposition 3.1, the matrix $\text{DtN}_j^{\text{ext}}$ diagonalizes in the basis of generalized eigenfunctions $\{\underline{v}_\ell\}$ of the matrix stencil (M_j, K_j), i.e.

$$
\text{DtN}_j^{\text{ext}} \underline{v}_\ell = \text{dtn}^{\text{ext}}(\underline{\lambda}_\ell) \underline{v}_\ell,
$$

so that all information about $\text{DtN}_j^{\text{ext}}$ is contained in the scalar function

$$
\text{dtn}^{\text{ext}}(\lambda) := A_{\Gamma_j \Gamma_j} + \lambda B_{\Gamma_j \Gamma_j} - (A_{\Gamma_j E_j} + \lambda B_{\Gamma_j E_j})(A_{E_j E_j} + \lambda B_{E_j E_j})^{-1}(A_{E_j \Gamma_j} + \lambda B_{E_j \Gamma_j}).
\tag{7.7}
$$

According to the results of Chapter 3, the quest for efficient approximations of $\text{DtN}_j^{\text{ext}}$ now boils down to the following question: Is it possible to find small matrices $\tilde{A}, \tilde{B} \in \mathbb{C}^{(N+1) \times (N+1)}$ such that

$$
\text{dtn}_N(\lambda) := \tilde{A}_{\Gamma_j \Gamma_j} + \lambda \tilde{B}_{\Gamma_j \Gamma_j} - (\tilde{A}_{\Gamma_j \tilde{E}_j} + \lambda \tilde{B}_{\Gamma_j \tilde{E}_j})(\tilde{A}_{\tilde{E}_j \tilde{E}_j} + \lambda \tilde{B}_{\tilde{E}_j \tilde{E}_j})^{-1}(A_{\tilde{E}_j \Gamma_j} + \lambda \tilde{B}_{\tilde{E}_j \Gamma_j})
\tag{7.8}
$$

provides a approximation of $\text{dtn}^{\text{ext}}(\lambda)$? Here, $\tilde{E}_j$ is a set of artificial exterior DOFs which is supposed to be much smaller than E_j. Therefore, if this question can be answered in the affirmative, the matrix

$$
\begin{bmatrix}
\tilde{L}_{\Gamma_j \Gamma_j} & \tilde{L}_{\Gamma_j \tilde{E}_j} \\
\tilde{L}_{\tilde{E}_j \Gamma_j} & \tilde{L}_{\tilde{E}_j \tilde{E}_j}
\end{bmatrix}
=
\begin{bmatrix}
\tilde{A}_{\Gamma_j \Gamma_j} & \tilde{A}_{\Gamma_j \tilde{E}_j} \\
\tilde{A}_{\tilde{E}_j \Gamma_j} & \tilde{A}_{\tilde{E}_j \tilde{E}_j}
\end{bmatrix}
\otimes M_j
+
\begin{bmatrix}
\tilde{B}_{\Gamma_j \Gamma_j} & \tilde{B}_{\Gamma_j \tilde{E}_j} \\
\tilde{B}_{\tilde{E}_j \Gamma_j} & \tilde{B}_{\tilde{E}_j \tilde{E}_j}
\end{bmatrix}
\otimes K_j,
\tag{7.9}
$$

represents a compressed version of the full exterior system (7.6) and can be used to replace

the latter in equation (7.5). More precisely, the linear system (7.5) is replaced[5] by

$$
\begin{bmatrix}
L_{I_j I_j} & L_{I_j \Gamma_j} & 0 \\
L_{\Gamma_j I_j} & L_{\Gamma_j \Gamma_j}^{\text{int}} + \tilde{L}_{\Gamma_j \Gamma_j} & \tilde{L}_{\Gamma_j \tilde{E}_j} \\
0 & \tilde{L}_{\tilde{E}_j \Gamma_j} & \tilde{L}_{\tilde{E}_j \tilde{E}_j}
\end{bmatrix}
\begin{bmatrix}
\underline{u}_{I_j}^{(n-\frac{1}{2})} \\
\underline{u}_{\Gamma_j}^{(n-\frac{1}{2})} \\
\underline{u}_{\tilde{E}_j}^{(n-\frac{1}{2})}
\end{bmatrix}
=
\begin{bmatrix}
\underline{f}_{I_j} - L_{I_j \Gamma_{j+1}} \underline{u}_{\Gamma_{j+1}}^{(n-\frac{1}{2})} \\
\underline{\tilde{f}}_{\Gamma_j}^{(n-\frac{1}{2})} \\
0
\end{bmatrix}, \qquad (7.10)
$$

which is much cheaper to solve because $\tilde{E}_j$ contains significantly fewer DOFs than E_j. Typically, it suffices to include the DOFs from the neighbouring subdomain Ω_{j-1} into $\tilde{E}_j$ to obtain a good approximation of $\text{DtN}_j^{\text{ext}}$, while E_j would contain all DOFs in $\cup_{i=0}^{j-1}\Omega_i$.

Thanks to the extensive studies in the previous chapters, it is by now clear how the question raised above can be answered and how $\tilde{A}$ and $\tilde{B}$ can be obtained. To apply learned IEs, the reference *dtn* function in the objective function (4.2) simply has to be replaced by dtn^{ext}. Apart from this, the optimization process runs as asual. For the numerical experiments in this and the following chapter the reduced symmetric ansatz will be employed as default.

Alternatively, it is possible to obtain the matrices $\tilde{A}$ and $\tilde{B}$ by any[6] other tensor-product discretization of $\mathcal{DtN}$ discussed in Appendix C. The approximation by a PML is by far the most commonly[7] employed in sweeping preconditioners. To this end, the PML is often started right at the interface Γ_j and spreads over the neighbouring subdomain Ω_{j-1}. For a tensor-product PML the matrices $\tilde{A}$ and $\tilde{B}$ can then be obtained similarly as discussed in Appendix C.1, where the complex stretching terms need to be adapted to the specific equation and geometry.

At this point it is important to mention that in contrast to learned IEs sweeping preconditioners using PMLs can be derived without recourse to a particular type of discretization. This proceeds as follows. It is assumed that on subdomain Ω_0 a PML is present. To represent $\mathcal{DtN}_j$ the equation (7.1) would need to be solved on Ω_j^{ext}, which is too expensive as discussed previously. Therefore, the PML on Ω_0 is moved somewhere close to Ω_j. This attenuates the waves and justifies to replace the large domain Ω_j^{ext} by a much smaller truncated exterior domain Ω_j^{PML}. In practice, often the choice $\Omega_j^{\text{PML}} = \Omega_{j-1}$ is made. The compressed linear system in (7.10) is then obtained directly by assembling the linear system of the complex stretched PDE on $\Omega_j^{\text{PML}} \cup \Omega_j$. More information on this approach can be found in the literature, see e.g. [GZ19, Remark 13 and Remark 14] in the setting of DOSM or [EY11b] for the original sweeping preconditioners.

If the $\text{DtN}_j^{\text{ext}}$ matrices are replaced by one of the approximations discussed above, DOSM in general no longer converges in one double sweep to the exact solution. In practice, the algorithm is then usually applied as a preconditioner for GMRES using zero initial guess $\underline{u}_j^{(0)} = 0$ for $j = 0, \dots, J-1$.

Remark 7.1 (Relation between $\text{dtn}_j^{\text{ext}}$ and *dtn$_j$*)**.** *The presentation in this subsection was concerned with the discrete level at which the $\text{dtn}_j^{\text{ext}}$ function of a given discretization needs to be resolved. The latter in turn is an approximation of a continuous* **dtn$_j$** *function which is characterized as in Chapter 2 of this thesis. Assuming that the differential operator*

---

[5]Computing the right hand side $\underline{\tilde{f}}_{\Gamma_j}^{(n-\frac{1}{2})}$ involves an application of $\text{DtN}_j^{\text{ext}}$ as well (see the definition given in Algorithm 7.3), which is likewise replaced by the efficient approximation of $\text{DtN}_j^{\text{ext}}$ obtained by swapping L in equation (7.4) with $\tilde{L}$.

[6]Of course, this discretization has to be suitable for the equation and the geometrical setup.

[7]Actually, it is the only one the author is aware of.

$\mathcal{L} = \mathcal{A} \otimes \mathcal{M}_\Gamma + \mathcal{B} \otimes \mathcal{K}_\Gamma$ *has tensor product structure and that the interfaces* Γ_j *are positioned at locations* $\{a_j\}$ *along the direction of stratification the ODE for determining* dtn_j *is given in analogy to Section 2.1 as*

$$[\mathcal{A} + \lambda_\ell \mathcal{B}] \, \Lambda_r(\lambda_\ell) = 0 \qquad on \ (a_0, a_j),$$

with $\Lambda_{a_j}(\lambda_\ell) = 1$ *and an appropriate[8] boundary condition at* $r = a_0$. *Then* $\mathit{dtn}_j(\lambda) := -\partial_r \Lambda_r(\lambda)|_{r=a_j}$. *If this ODE is discretized following the ansatz of equation (7.6), i.e. the differential operators* $\mathcal{A}$ *and* $\mathcal{B}$ *on* $[a_0, a_j]$ *are represented by the matrices* A *and* B *appearing in equation (7.6), then the discrete approximation to* dtn_j *obtained from this discretization takes the form of* $\mathrm{dtn}_j^{\mathrm{ext}}$ *given in (7.7). In the numerical experiments below the numerical approximation* $\mathrm{dtn}_j^{\mathrm{ext}}$ *of* dtn_j *is often so accurate that both functions may be used interchangeably.*

7.2.3 Computational costs of transmission conditions

In the setting of transparent boundary conditions treated in the first half of this thesis learned IEs need to be implemented merely at a single interface which realizes the coupling between the bounded interior domain and the usually unbounded exterior domain. The associated computational costs have been discussed in Section 4.1.4. Let us recall here from this section that learned IEs differ from tensor-product PMLs solely in how the learned IE matrices $\tilde{A}$ and $\tilde{B}$ in (7.9) are obtained. Once these matrices have been determined, the solution of the linear systems proceeds as for tensor-product PMLs and is usually even more affordable since the learned IE matrices manage to achieve the same accuracy using less DOFs and sparser matrices, see Section 4.2.2. Hence, solving the optimization problem to determine the learned IE matrices poses the only potentially critical point.

In the context of sweeping preconditioners learned IEs are required on all subdomain interfaces Γ_j for $j = 1, \dots, J-1$. Hence, $J-1$ optimization problems (4.2)-(4.3) have to be solved. In this regard, the following observations about the optimization problems should be kept in mind.

- They only have to be solved once during the offline stage in which the preconditioner is set up.

- The problems for different interfaces are completely independent which allows for simple and efficient parallelization.

- The costs for their solution is independent of the space dimension d.

Based on these considerations it is reasonable to expect that the additional overhead stemming from the optimization procedure will be negligible - at least asymptotically with increasing complexity of the problem. The majority of the time will be spent in the online phase in which the preconditioner is applied since this requires the solution of d-dimensional linear systems[9] on the subdomains Ω_j for $j = 0, \dots, J-1$. Consequently, the number of required GMRES iterations, which determines how often the preconditioner needs to be applied, is the critical factor that will be monitored as a measure of performance in the numerical experiments presented in Section 7.3.

---

[8]This is obtained from $\mathcal{B}_{\partial\Omega_{\mathrm{int}}} u = g$.

[9]As the subdomains are thin, these can be regarded as quasi $d-1$-dimensional problems. Still, it remains the dominating cost.

7.3 Numerical experiments

Before applying the introduced techniques to helioseismology, it is useful to consider simple problems on the unit square. This allows to isolate the key issues which are expected to occur for the application in helioseismology and investigate them one after another. To this end, we consider the Helmholtz equation

$$-\Delta u - k^2(x,y)u = f$$

in $\Omega = [0,1]^2$. For all experiments to follow periodic boundary conditions are imposed on the left and right boundary of the square. At the bottom a homogeneous Neumann boundary condition is set. Two different boundary condition at the top will be studied: either a PML or a homogeneous Neumann boundary condition. All experiments use a random source (set to zero in the topmost layer where PML could be present) as the right hand side and a tolerance of 10^{-7} for GMRES. Finite elements of order $p = 4$ will be used throughout this section. Sweeping is performed in the vertical direction, i.e. from top to bottom for the forward sweep. To this end, the unit square is partitioned into vertical layers of equal thickness as sketched in Fig. 7.2. To maintain accuracy for high-frequency the number of layers J has to grow proportional to the wavenumber. In the experiments below the relation between J and k is chosen such that each wavelength is resolved by at least twelve DOFs. For some geophysical applications this might seem a little excessive, yet in view of the forthcoming results of Section 8.1 this requirement is reasonable for helioseismology.

The moving PML will always extend over one layer. Since one layer is two elements thick, this implies for $p = 4$ that there are nine DOFs in the sweeping direction within this layer. As a tensor-product PML is used, the exterior system can be written in the form $A \otimes M + B \otimes K$ as shown in Appendix C.1. In the considered case A and B are 9×9 matrices and A contains 49 nonzero entries. On the other hand, the symmetric ansatz for the learned IEs introduced in Section 4.1.1 results in $(N+1) \times (N+1)$ matrices with $3N + 1$ nonzero entries. Hence, in terms of the number of nonzero elements the moving PML is comparable to learned IEs with $N = 16$.

In all experiments the objective function (4.2) will be set up in terms of the continuous eigenvalues λ_ℓ even though the generalized eigenvalue equation $K\underline{w}_\ell = \underline{\lambda}_\ell M\underline{w}_\ell$ holds in terms of the discrete eigenvalues $\underline{\lambda}_\ell$. This is done for the sake of efficiency since the continuous eigenvalues are analytically known while the discrete eigenvalues would need to be computed numerically. The discrete eigenvalues converge towards the continuous ones as the resolution is increased, yet, for a fixed discretization the above approach always introduces a certain error. Therefore, it can in general not be expected that the iteration numbers converge to one as N is increased when the discretization remains fixed but will start to stagnate at a couple of iterations (usually 3-4 for 12 DOFs per wavelength).

7.3.1 Constant sound speed

Consider the simplest case of constant $k(x,y) = k$ first. In the top layer of the decomposition either a PML or a homogeneous Neumann boundary condition is implemented. Wavenumbers $k \in \{12, 24, 48, 96\}$ are considered. Note that these are not multiples of 2π, i.e. not resonances. The GMRES iteration numbers for moving PML and learned IEs for this configuration are compared in Table 7.1. The two different boundary conditions are discussed separately:
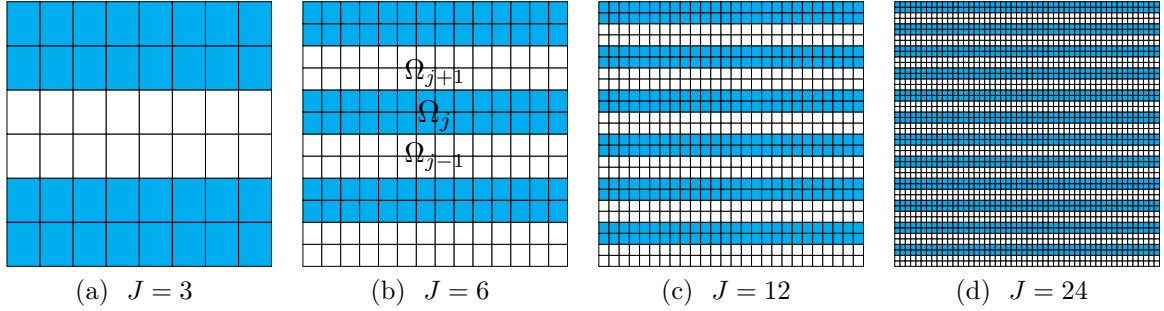
| (a) $J = 3$ | (b) $J = 6$ | (c) $J = 12$ | (d) $J = 24$ |

Figure 7.2: Meshes for experiments on unit square, which is partitioned into J layers. Each such layer has a thickness of two elements in vertical direction.

- If a PML is present at the top the moving PML ('mPML') achieves small iteration numbers which are almost stable with respect to the frequency. The iteration numbers for learned IEs decrease as N increases since the approximation of $\mathrm{dtn^{ext}}$ improves. As mentioned above, the tensor-product PML is based on $N = 8$. The approach using learned IEs with $N = 8$ converges for $k \in \{12, 24, 48\}$ at least twice as fast as the moving PML. According to the discussion above, the corresponding linear systems are also sparser than those obtained from usage of moving PML.

- The stability of the moving PML approach is lost when the PML at the top of the domain is removed. Even though a complete breakdown is not observed, the iteration numbers increase noticeably with the wavenumber. In contrast, the results for learned IEs with $N = 10$ are comparable to those obtained with a PML at the top of the domain.

PML at top

k	J	$N = 0$	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	mPML
12	3	16	7	4	3	3	3	7
24	6	39	10	5	3	3	3	7
48	12	119	20	6	5	3	3	8
96	24	-	50	12	7	6	5	9

Homogeneous Neumann b.c. at top

k	J	$N = 0$	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	mPML
12	3	48	9	6	4	4	3	9
24	6	46	16	8	5	5	5	11
48	12	-	162	26	7	6	4	21
96	24	-	-	-	225	38	8	34

Table 7.1: GMRES iteration numbers for constant wavenumber and two different boundary condition at the top of the domain. A dash '-' means that the desired tolerance was not achieved after 250 iterations.

The difference between the two boundary conditions can be explained by considering the corresponding $\mathrm{dtn^{ext}}$ functions. Fig. 7.3 displays these functions and their approximation by moving PML and learned IEs respectively for $k = 96$ on an interface of a typical layer. While the $\mathrm{dtn^{ext}}$ function for a PML at the top is smooth and can be

well-approximated by the moving PML, it exhibits poles on the real axis for a Neumann boundary condition. Since the moving PML only includes information from the neighbouring layer, it is oblivious[10] of the Neumann boundary condition at the top and consequently of the poles of dtn$^{\text{ext}}$. Learned IEs on the other hand are naturally aware of the boundary condition since they are based on approximation of dtn$^{\text{ext}}$ which contains full information about the complete half-space problem. Hence, they can adapt to the specific problem and obtain small iteration numbers irrespective of the boundary condition.
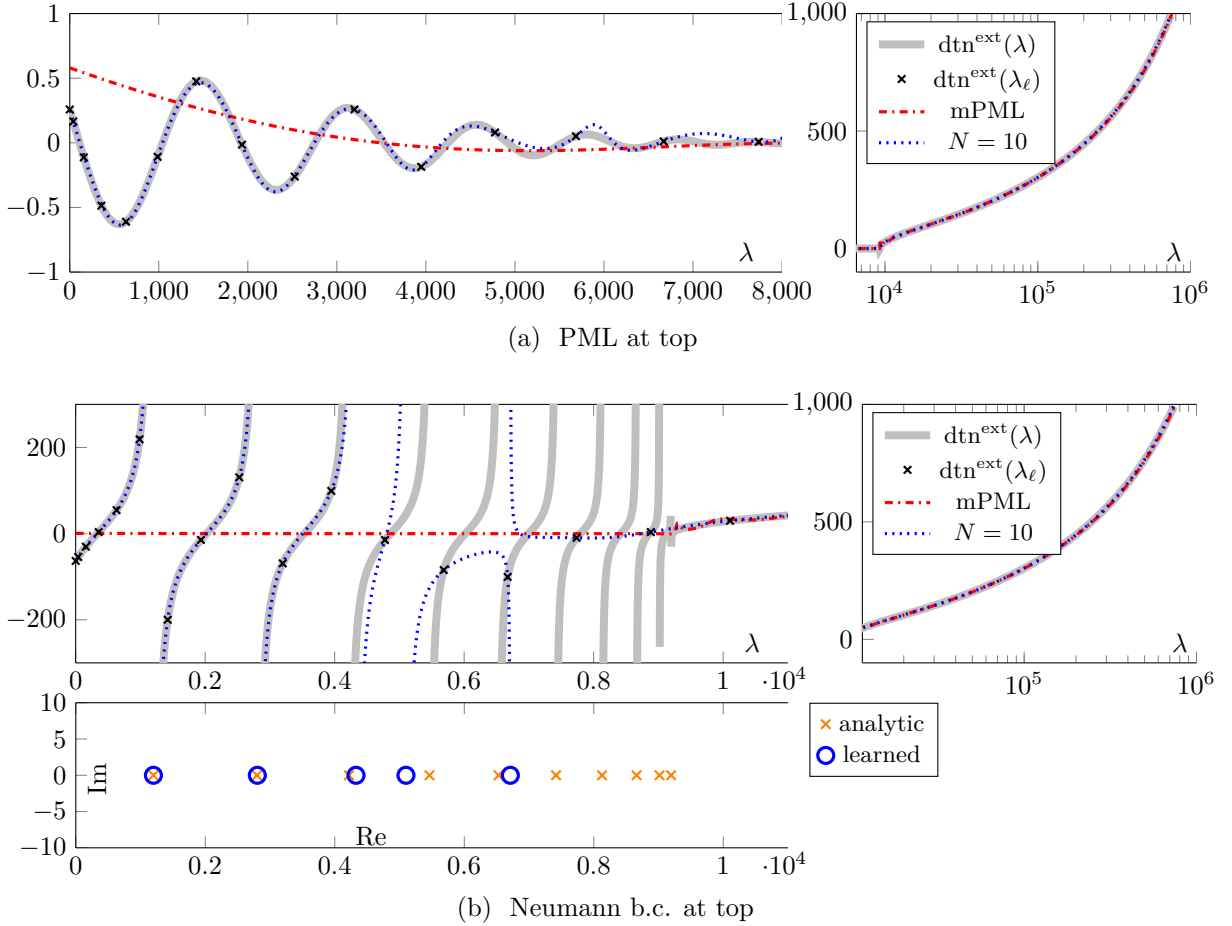

(a) PML at top

(b) Neumann b.c. at top

Figure 7.3: The dtn$^{\text{ext}}$ function (real part) for $k = 96$ on one of the layer interfaces for different boundary conditions at the top of the domain. The black crosses dtn$^{\text{ext}}(\lambda_\ell)$ display the function values only at the eigenvalues $\lambda_\ell = \ell^2$ while the gray solid line dtn$^{\text{ext}}(\lambda)$ represents a much finer sampling. The discrete approximations dtn$_N$ obtained with moving PML and learned IEs using $N = 10$ are shown in red and blue colors, respectively. The lowest panel in figure (b) displays the analytic poles of $d\!t\!n$ in orange while the poles of dtn$_N$ for learned IEs are shown in blue.

For the simple example at hand it can easily be understood analytically why the dtn$^{\text{ext}}$ function for the reflective boundary condition admits poles for $\lambda \leq k^2$ while it is smooth when an absorbing boundary condition is imposed. The $d\!t\!n = -\partial_y \Lambda_a(\lambda)$ function for the

---

[10]By construction the moving PML approximation (red dots in the figure) is exactly the same for cases (a) and (b).

Neumann problem on a layer interface $y = a$ is determined by the solution of the ODE

$$-\frac{\partial^2 \Lambda_y(\lambda)}{\partial^2 y} - k_\lambda^2 \Lambda_y(\lambda) = 0, \tag{7.11}$$

$$\Lambda_a(\lambda) = 1, \quad \partial_y \Lambda_b(\lambda) = 0,$$

for $k_\lambda^2 = k^2 - \lambda$ and $b = 1$. By inspection the solution is given by

$$\Lambda_y(\lambda) = \cos\left(k_\lambda(b - y)\right)/\cos\left(k_\lambda(b - a)\right),$$

so that $\mathit{dtn}(\lambda) = -\sigma_\lambda \tan\left(k_\lambda(b - a)\right)$. This function admits poles and roots lying exactly on the real line given by

$$\lambda_{\text{pole}}^j = k^2 - \left(\frac{\pi(2j+1)}{2(b-a)}\right)^2, \quad \lambda_{\text{root}}^j = k^2 - \left(\frac{\pi(2j)}{2(b-a)}\right)^2, \tag{7.12}$$

for $j \in \mathbb{N}_0$. These poles and their numerical approximation by the poles of dtn_N using learned IEs are displayed in the lowest panel of Figure 7.3b. The poles for large λ are not detected by the learned IEs because they are insufficiently covered by the sample points λ_ℓ. Of course one could try to approximate these poles as well by including samples of dtn^{ext} also at non-eigenvalues into the objective function. Future studies could investigate whether this might lead to a further improvement of the iteration numbers or a faster convergence in N.

Note that formula (7.12) implies that $\lambda_{\text{pole}}^j \leq k^2$, i.e. the poles are limited to the propagating part of the spectrum. This has direct consequences for the approximation of dtn^{ext}:

- As the number of poles of dtn on the positive real line grows linearly with k and dtn_N can only cover N simples poles at a time, we need to increase N linearly with k if all poles of dtn should be well-approximated. This is consistent with the results of Table 7.1 where for higher wavenumbers larger N were required to achieve small iteration numbers.

- Even though the moving PML yields a good approximation of dtn for $\lambda > k^2$, which explains its respectable performance for small wavenumbers for the Neumann boundary condition, it will eventually break down as k and thereby the number of poles increases since the approximation for the propagating modes is completely wrong.

If the Neumann boundary condition is replaced by an impedance condition of the form $\partial_y \Lambda_b(\lambda) = i\omega \Lambda_b(\lambda)$ the poles for $\lambda \leq k^2$ vanish[11]. Indeed, in this case the solution of the ODE is given by

$$\Lambda_y(\lambda) = \frac{(k_\lambda - k)e^{(b-y)ik_\lambda} + (k_\lambda + k)e^{-(b-y)ik_\lambda}}{(k_\lambda - k)e^{(b-a)ik_\lambda} + (k_\lambda + k)e^{-(b-a)ik_\lambda}}.$$

The corresponding dtn function

$$\mathit{dtn}(\lambda) = -ik_\lambda \left(\frac{ik_\lambda \sin\left((b-a)k_\lambda\right) - k\cos\left((b-a)k_\lambda\right)}{k_\lambda \cos\left((b-a)k_\lambda\right) - ik\sin\left((b-a)k_\lambda\right)}\right)$$

has no real poles for $\lambda \leq k^2$ since the denominator does not vanish. This statement holds for a constant sound speed and may extend to the case where the sound speed is continuous. However, the next Section 7.3.2 demonstrates that poles can occur near the real axis once the latter assumption is dropped.

---

[11] For the case of constant wavenumber.

7.3.2 Jumping wavenumber between layers

Besides boundary conditions, reflection of waves can also be triggered by local inhomogeneities of the wavenumber. To demonstrate this, we consider a wavenumber which jumps between layers according to a parameter $\alpha \in [0,1]$. On one layer the wavenumber will be given by $k_+ = (1 + \alpha/2)k$ while on the neighbouring layers it jumps to $k_- = (1 - \alpha/2)k$ and so on. The contrast between layers $|k_+ - k_-| = \alpha k$ increases with the k and α. It is well-known from similar experiments considered in the literature [GZ18, GZ19, PHL20] that the moving PML approach breaks down as the constrast between layers increases.

The iteration numbers with moving PML and learned IEs for PML boundary conditions at the top of the domain are shown in Table 7.2. The performance of the moving PML indeed deteriorates drastically as the contrast increases. For $\alpha = 1/4$ and $k = 96$ the desired tolerance cannot be achieved anymore within the set limit of 250 iterations. On the other hand, with learned IEs as transmission conditions convergence can be achieved within a few iterations (4-5) throughout the entire range of considered frequencies and contrasts.

<div align="center">

Moving PML

k	J	$\alpha = 0$	$\alpha = 1/16$	$\alpha = 1/8$	$\alpha = 1/4$	$\alpha = 1/2$	$\alpha = 1$
12	3	7	7	8	7	7	8
24	6	7	10	13	17	20	26
48	12	8	15	24	50	71	98
96	24	9	42	83	-	-	-

Learned infinite elements: $N = 12$

k	J	$\alpha = 0$	$\alpha = 1/16$	$\alpha = 1/8$	$\alpha = 1/4$	$\alpha = 1/2$	$\alpha = 1$
12	3	3	3	3	3	3	3
24	6	3	3	3	3	3	3
48	12	3	3	3	3	3	3
96	24	4	5	5	4	4	4

Learned infinite elements: $\alpha = 1/4$

k	J	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	$N = 12$
12	3	6	4	3	3	3	3
24	6	15	5	4	3	3	3
48	12	37	19	7	4	4	3
96	24	-	208	55	25	15	4

</div>

Table 7.2: GMRES iteration numbers for wavenumber which alternates $|k_+ - k_-| = \alpha k$ between consecutive layers. PML boundary condition implemented at the top. A dash '-' means that the desired tolerance was not achieved after 250 iterations.

To understand these results it is again illuminating to examine the corresponding $\mathrm{dtn}^{\mathrm{ext}}$ functions and their approximation by moving PML and learned IEs, respectively. Plots of these fuctions for $\alpha = 1$ and $k = 96$ on an interface of a typical layer are shown in Fig. 7.4. The $\mathrm{dtn}^{\mathrm{ext}}$ function again exhibits poles in close proximity to the real axis, this
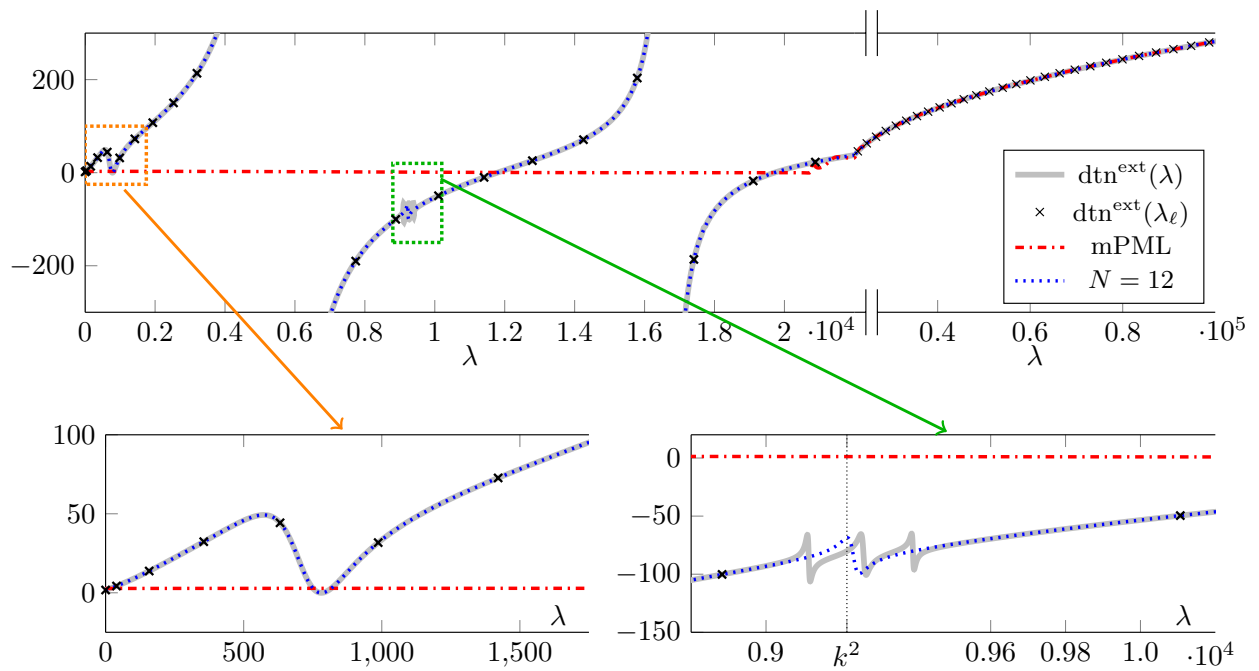
Figure 7.4: The dtn$^{\text{ext}}$ function (real part) for a jumping wavenumber between layers. The contrast is equal to αk where the plot displays the case $\alpha = 1$ and $k = 96$. The black crosses dtn$^{\text{ext}}(\lambda_\ell)$ show the function values only at the eigenvalues $\lambda_\ell = \ell^2$ while the gray solid line dtn$^{\text{ext}}(\lambda)$ represents a much finer sampling. The discrete approximations dtn$_N$ obtained with moving PML and learned IEs using $N = 12$ are shown in red and blue colors, respectively. The lower panel displays zooms of the regions marked in the upper panel.

time induced by the sharp contrast between layers. Again the moving PML is unsuited for this situation while learned IEs can adapt. It is astonishing that the approximation by learned IEs is seemingly able to detect also some minor variations of dtn$^{\text{ext}}$ which on first sight appear not to be covered by the samples dtn$^{\text{ext}}(\lambda_\ell)$. This can for instance be seen in the lower right panel of Fig. 7.4. Close to $\lambda = k^2$ the dtn$^{\text{ext}}$ function has some local inhomogeneities which seem to be noticed by the learned IEs even though the optimization has only access to information[12] at the sample points dtn$^{\text{ext}}(\lambda_\ell)$. Apparently, tiny variations in the input are sufficient to trigger a response of the optimization routine.

7.3.3 Perturbations from the background model

The previous experiments demonstrated that learned IEs perform well for perfectly stratified media even in the presence of strong reflections. However, many media of practical interest are only approximately stratified in the sense that the properties of the medium can be regarded as small perturbations from a perfectly stratified background. Since the equations for the approximately stratified medium may not be separable anymore, learned IEs (in their current form) cannot be applied directly to build asymptotically exact sweeping preconditioners for such problems as they only provide transmission conditions for the background model. Nevertheless, as long as the perturbations from the background model are small we can try to precondition the problem using learned IEs based on the background model. The following experiment investigates the performance of this approach for the simple model problem

$$-\Delta u - (1 + \varepsilon)k^2 u = f$$

on the unit square with boundary conditions set up as before. Here, $\varepsilon \geq 0$ is a small parameter which represents the strength of the perturbation from the background model $\varepsilon = 0$. For the sake of the experiment we will pretend in this section that we only have access to the background model to build the transmission conditions for the learned IEs. This is to mimic the situation for realistic problems in which the perturbation would be spatially dependent $\varepsilon = \varepsilon(x)$ so that the perturbed equation ceases to be separable. Here ε is chosen to be constant because this allows for a much more accessible analysis and visualization of the arising issues.

The GMRES iteration numbers for Neumann and PML boundary conditions at the top of the domain are shown in Table 7.3. Four different preconditioning strategies are compared: the first table displays the results for the moving PML approach. The second contains the results for preconditioning the perturbed system with the direct solver inverse of the background model ($\varepsilon = 0$). This is of course not very practical since a direct solution of the linear system for $\varepsilon = 0$ is not cheaper than for $\varepsilon \neq 0$. Therefore, the third table presents the iteration numbers for the case when the inverse of the background model is approximated using sweeping preconditioners with learned IE transmission conditions. The final table gives the result for applying sweeping preconditioners with learned IEs directly[13] to the system for $\varepsilon \neq 0$ with transmission conditions still based on the background model.

The results can be summarized as follows:

---

[12]Neither the imaginary part of dtn$^{\text{ext}}$ at the sample points λ_ℓ, which is not shown in the figure, seems to give any obvious indication that special behavior should occur around $\lambda = k^2$.

[13]This means that the matrices $L_{I_j I_j}, L_{I_j \Gamma_j}, L_{\Gamma_j I_j}$ and $L^{\text{int}}_{\Gamma_j \Gamma_j}$ in equation (7.5) are assembled based on the perturbed model.

<div align="center">Moving PML</div>

k	$\downarrow J/\varepsilon \rightarrow$	0.0%	0.0625%	0.125%	0.25%	0.5%	1.0%	2.0%	4.0%
12	3	9/7	9/7	9/7	9/7	9/7	9/7	9/7	9/7
24	6	11/7	11/7	11/7	11/7	12/7	12/7	12/8	13/8
48	12	21/8	21/8	20/8	20/8	20/8	20/8	19/8	21/9
96	24	34/9	34/9	33/9	35/8	32/9	32/9	36/9	33/10

<div align="center">Direct solver inverse of background model</div>

k	$\downarrow J/\varepsilon \rightarrow$	0.0%	0.0625%	0.125%	0.25%	0.5%	1.0%	2.0%	4.0%
12	3	1/1	5/5	5/5	5/6	6/6	7/7	8/9	10/11
24	6	1/1	5/5	6/5	6/6	7/7	8/8	10/10	14/13
48	12	1/1	7/6	8/6	10/7	11/8	15/9	18/12	30/17
96	24	1/1	11/7	14/8	19/10	22/12	33/16	50/22	85/37

<div align="center">Learned IE $N = 10$ for background model on layer</div>

k	$\downarrow J/\varepsilon \rightarrow$	0.0%	0.0625%	0.125%	0.25%	0.5%	1.0%	2.0%	4.0%
12	3	3/3	5/5	5/5	5/6	6/6	7/7	8/9	9/11
24	6	5/3	5/5	5/5	6/6	7/7	8/8	10/10	14/13
48	12	4/3	7/6	9/6	10/7	11/8	15/9	20/12	31/17
96	24	8/5	12/7	15/8	21/10	25/12	35/16	53/22	88/37

<div align="center">Learned IE $N = 10$ for perturbed model on layer</div>

k	$\downarrow J/\varepsilon \rightarrow$	0.0%	0.0625%	0.125%	0.25%	0.5%	1.0%	2.0%	4.0%
12	3	3/3	5/4	5/4	6/5	7/5	7/5	8/6	9/7
24	6	5/3	6/4	6/5	7/5	8/5	10/6	13/7	19/9
48	12	4/3	8/5	10/6	13/6	17/7	25/8	36/10	56/13
96	24	8/5	14/7	20/8	36/10	59/12	138/15	350/23	439/121

Table 7.3: GMRES iteration numbers for perturbations of strength ε from the background model. Results are given in the form x/y where x are the numbers for hom. Neumann boundary conditions at the top and y the numbers for a PML boundary condition.

- The iteration numbers for the moving PML are essentially independent of the strength of the perturbation. This was to be expected since the dtn$^{\text{ext}}$ approximation provided by PML does not adhere[14] to the background model. Even if this advantage renders PMLs very attractive for applications they are still unable to cope with problems involving strong reflections.

- Based on the experiments of the last sections one would try to resolve this issue by replacing the moving PML in the sweeping preconditioners by learned IEs as was done in the final table of Table 7.3. Unfortunately, this does not work well for Neumann boundary conditions as the iteration numbers increase drastically with the strength of the perturbation and the contrast. Only tiny perturbations $\varepsilon < 0.1\%$ can be tolerated. This leads to the conjecture (see also [PHL20]), which will be investigated in more detail below, that the *dtn* function in presence of reflections must be extremely sensitive to perturbations.

---

[14]Since we allow the PML to be aware of the perturbation.

- Taken the last observation for granted leads to the conclusion that in the presence of reflections learned IEs should only be used to approximate the inverse of the background model. The question whether the background model yields a good preconditioner for the perturbed system is investigated in the second table of Table 7.3. The iteration numbers for Neumann boundary conditions are significantly lower than in table four, although still larger than for PML boundary conditions, so that somewhat larger perturbations may be tolerated. The iteration numbers in the third table are almost the same as in the second which demonstrates that the direct solver inverse of the background model can be realized (nearly) efficiently with learned IEs.
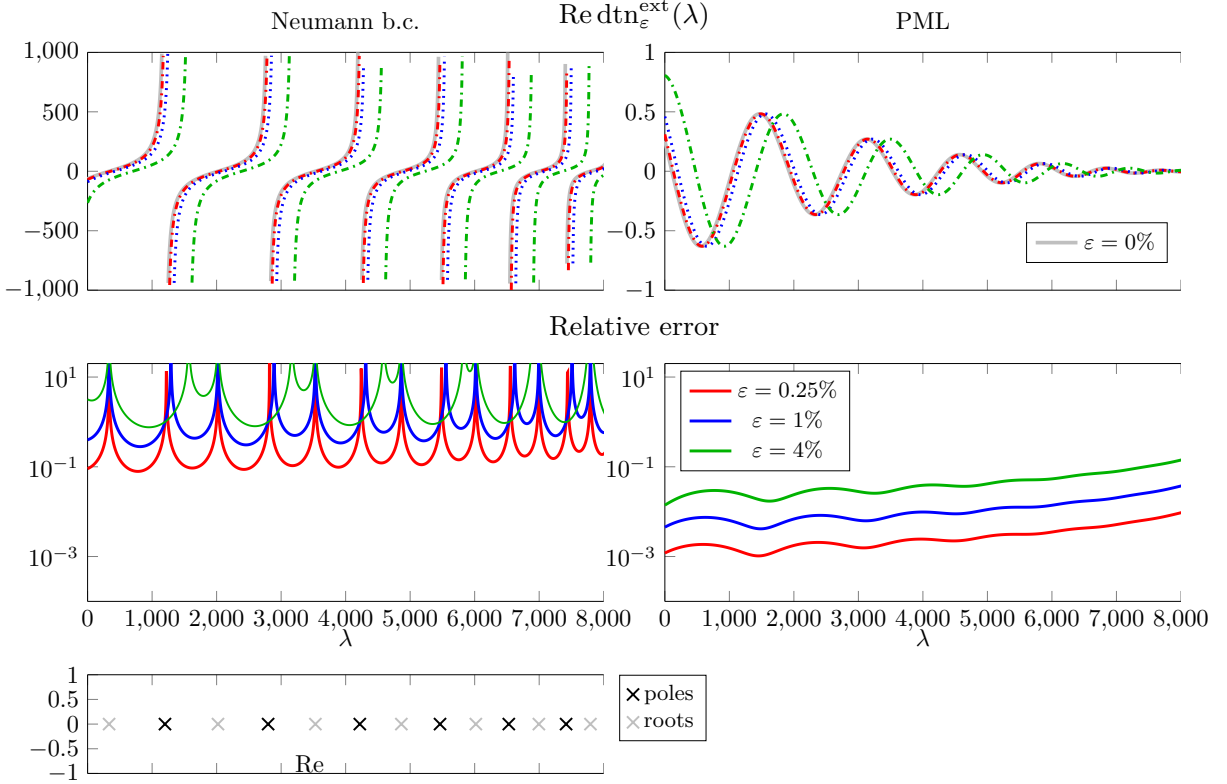
Figure 7.5: The upper panel shows $\operatorname{Re} \operatorname{dtn}_\varepsilon^{\mathrm{ext}}(\lambda)$ corresponding to the perturbed wavenumber $(1 + \varepsilon)k^2$. The lower panel displays the relative error $|\operatorname{dtn}_{\varepsilon=0}^{\mathrm{ext}}(\lambda) - \operatorname{dtn}_\varepsilon^{\mathrm{ext}}(\lambda)|/|\operatorname{dtn}_{\varepsilon=0}^{\mathrm{ext}}(\lambda)|$ w.r.t. to the $\operatorname{dtn}_{\varepsilon=0}^{\mathrm{ext}}$ function corresponding to the background sound speed. The panel is vertically divided into two parts, where the results on the left side are for a homogeneous Neumann boundary condition at the top of the domain and the results on the right for a PML at the top. The lowest panel shows roots and poles of $\mathit{dtn}_{\varepsilon=0}$.

Above it was conjectured that applying learned IEs directly to the perturbed problem for Neumann boundary conditions fails because of high sensitivity of dtn with respect to perturbations. Fig. 7.5 investigates this claim. The first panel shows the $\operatorname{dtn}_\varepsilon^{\mathrm{ext}}$ function of the perturbed problem for Neumann and PML boundary conditions. This is obtained by solving equation (7.11) with k_λ^2 replaced by $k_{\lambda,\varepsilon}^2 = (1 + \varepsilon)k^2 - \lambda$. The second panel displays the relative error

$$|\operatorname{dtn}_{\varepsilon=0}^{\mathrm{ext}}(\lambda) - \operatorname{dtn}_\varepsilon^{\mathrm{ext}}(\lambda)|/|\operatorname{dtn}_{\varepsilon=0}^{\mathrm{ext}}(\lambda)|$$

with respect to the background model. While the error for PML boundary conditions is smooth and remains reasonably small even for strong perturbations, it is very large with unbounded peaks for Neumann boundary conditions. For the simple problem under considerations this behavior can easily be explained analytically. The dtn_ε function of the perturbed problem is given by

$$dtn_\varepsilon(\lambda) = -k_{\lambda,\varepsilon} \tan\left(k_{\lambda,\varepsilon}(b-a)\right).$$

Since

$$\partial_\varepsilon \, dtn_{\varepsilon=0}(\lambda) = -\frac{k^2}{2k_\lambda} \tan\left(k_\lambda(b-a)\right) - k_\lambda(1 + \tan^2\left(k_\lambda(b-a)\right))$$

where $k_{\lambda,\varepsilon=0} = k_\lambda$, the relative error to first order is given by

$$\frac{|\, dtn_{\varepsilon=0}(\lambda) - dtn_\varepsilon(\lambda)|}{|\, dtn_{\varepsilon=0}(\lambda)|} = \varepsilon \left| \frac{k^2}{2k_\lambda^2} + \frac{1}{\tan\left(k_\lambda(b-a)\right)} + \tan\left(k_\lambda(b-a)\right) \right| + \mathcal{O}(\varepsilon^2).$$

The first order term has singularities at the roots and poles of $\lambda \mapsto \tan\left(k_\lambda(b-a)\right)$ which were given in equation (7.12) and are marked in the lowest panel of Fig. 7.5 as gray and black crosses, respectively. This accounts for the occurence of the peaks in the middle panel of Fig. 7.5.

7.3.4 Non tensor-product meshes

In their current form learned IEs are based on the assumption of a tensor-product discretization, i.e. they assume that the linear system exterior to the transmission interface has the form (7.6). In this case it has been shown in Proposition 3.1 that the $\mathrm{DtN}_j^{\mathrm{ext}} := M_j^{-1}\left(L_{\Gamma_j\Gamma_j} - L_{\Gamma_j E_j} L_{E_j E_j}^{-1} L_{E_j\Gamma_j}\right)$ map diagonalizes in the basis $V = \{\underline{v}_0, \underline{v}_1, \ldots\}$ of the generalized eigenvectors of the matrix stencil (M_j, K_j), i.e. the matrix

$$V^{-1} \, \mathrm{DtN}_j^{\mathrm{ext}} \, V \tag{7.13}$$

is diagonal. For discretizations which are not of tensor-product form this statement no longer holds true. From an algorithmic perspective there is no obstruction to applying learned IEs for preconditioning linear systems stemming from non-tensor product discretizations though. However, as theoretical justification is lacking the consequences of violating the assumption of tensor-product structure need to be studied.
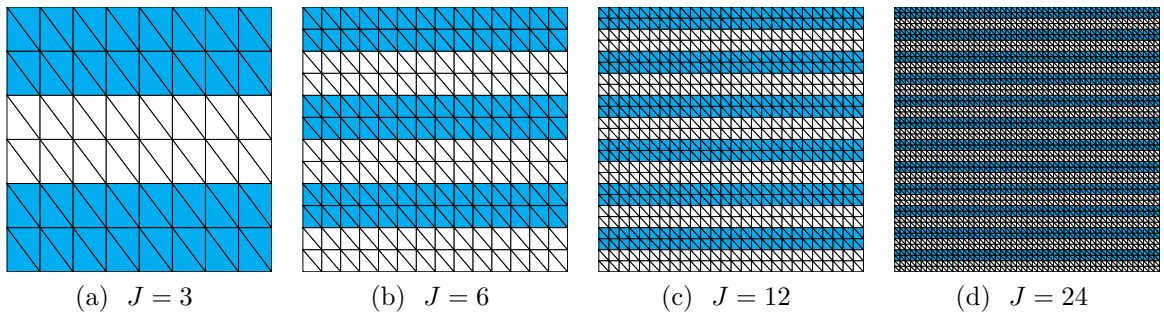

(a) $J = 3$ (b) $J = 6$ (c) $J = 12$ (d) $J = 24$

Figure 7.6: Simplicial meshes for experiments on unit square, which is partitioned into J layers. Each such layer has a thickness of four simplices in vertical direction.

This section investigates this issue for the simple case of a uniform simplicial discretization of the unit square for the constant coefficient Helmholtz equation as considered in Section 7.3.1. The series of meshes shown in Fig. 7.6 is the simplicial counterpart of the quadrilateral meshes displayed in Fig. 7.2. The GMRES iteration numbers on these simplicial meshes are displayed in Table 7.4. These results can be directly compared to the ones on tensor-product meshes from Table 7.1 since the computational setup differs only in the choice of the mesh. On the simplicial meshes the iteration numbers stagnate around 9-11 iterations no matter how large N is taken. As a results, the moving PML is now able to achieve lower iteration numbers than learned IEs in the presence of a PML boundary condition at the top of the domain. Although the iteration numbers for learned IEs are overall higher than on tensor-product meshes, the behavior with respect to the frequency is not impaired. As for tensor-product meshes a very mild growth of the iteration numbers with respect to the frequency is observed which would need to be countered by increasing N.

PML at top

k	J	$N = 0$	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	mPML
12	3	18	10	10	10	10	10	6
24	6	45	12	10	10	10	10	7
48	12	134	21	11	11	11	11	7
96	24	-	52	14	11	11	11	8

Homogeneous Neumann b.c. at top

k	J	$N = 0$	$N = 2$	$N = 4$	$N = 6$	$N = 8$	$N = 10$	mPML
12	3	42	8	9	9	9	9	8
24	6	40	13	10	10	10	10	12
48	12	-	161	24	12	10	10	20
96	24	-	-	-	242	43	12	34

Table 7.4: GMRES iteration numbers for a constant wavenumber on the simplicial meshes shown in Fig. 7.6. Polynomials of order $p = 4$ weres used for the FEM. A dash '-' means that the desired tolerance was not achieved after 250 iterations.

The increased number of iterations required on simplicial meshes stems from off-diagonal entries of $V^{-1} \mathrm{DtN}_j^{\mathrm{ext}} V$. Indeed, Fig. 7.7 displays this matrix on the mesh shown in Fig. 7.6b for different polynomial degrees p. For $p > 1$ off-diagonal blocks of size 2×2 are observed to appear in a regular pattern, which is not yet understood. The size of the blocks could be related to the fact that each layer is composed of two layers of two simplices in the vertical direction. The majority of these off-diagonal blocks are of far smaller magnitude than the diagonal elements, which explains why sweeping with learned IEs is still effective. Note also that the eigenvectors in the matrix V have been ordered according to increasing magnitude of the corresponding eigenvalue and that the magnitude of the off-diagonal blocks in $V^{-1} \mathrm{DtN}_j^{\mathrm{ext}} V$ is very small for the small eigenvalues yet increases for the larger ones. This makes sense since the continuous $\mathcal{DtN}_j$ operator is diagonal in the basis of the continuous eigenfunctions and the discrete $\mathrm{DtN}_j^{\mathrm{ext}}$ map converges to the continuous operator as the discretization is refined. While the small eigenvalues are already well resolved by the discretization, the large ones are not which explains the large magnitude of their corresponding off-diagonal blocks.
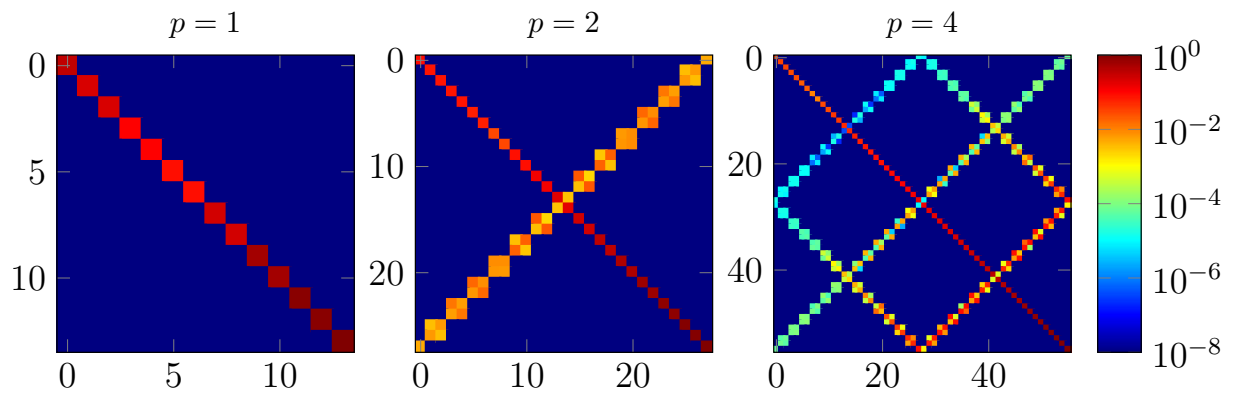
Figure 7.7: DtN map $V^{-1} \operatorname{DtN}_j^{\mathrm{ext}} V$ in the eigenbasis on the simplicial mesh shown in Fig. 7.6b for different polynomials degrees p. The eigenvectors in V have been ordered according to increasing magnitude of the corresponding eigenvalues. For clarity of presentation the largest entry of the matrix has been normalized to one.

Chapter 8

Sweeping for helioseismology in axisymmetric setting

In this chapter, the techniques developed in previous parts of this thesis are combined and applied for preconditioning the scalar equation of helioseismology in an axisymmetric setting. Firstly, the axisymmetric discretization itself and the implementation of learned IEs for this configuration will be introduced in Section 8.1. This constitutes an extension of the results from Chapter 6 where only one-dimensional discretizations, which are possible in a fully-separable setting, have been considered. The adaptations in the axisymmetric case are minor since the same learned IE matrices as in Section 6.3 can be used for representing the VAL-C atmosphere. However, in order to sweep as in Chapter 7 one has to go one step further. Being able to represent the $\mathcal{DtN}$ operator of the atmosphere with learned IEs is not sufficient for this purpose. Now also accurate and efficient realizations of $\mathcal{DtN}$ operators at concentric layers inside the Sun (see Fig. 8.2) are required as transmission conditions on subdomain interfaces. In view of the results of Chapter 7, this is challenging in at least two aspects.

- The $\mathcal{DtN}$ operator at the discrete level only diagonalizes for tensor-product discretizations as explained in Section 7.3.4. This reduces the effectiveness of learned IE on general meshes. Unfortunately, obtaining accurate synthetic observables for helioseismology (see Section 8.1) requires meshes which feature a strong refinement in the surface layers and this is incompatible with the tensor-product structure. Therefore, it has to be investigated whether learned IEs are appropriate for the realistic meshes introduced in Section 8.1.

- In contrast to the solar atmosphere, the interior of the Sun cannot be assumed to be spherically symmetric. This would defy the purpose of local helioseismology which aims at probing the three-dimensional structure of the solar interior. Hence, a good preconditioner also has to be able to deal with (small) perturbations from the separable background model. Ideally, it should be robust with respect to these perturbations. In light of the results of Section 7.3.3, this poses a serious challenge for learned IEs.

To avoid having to face all problems at once, Section 8.2 starts with sweeping for the separable background model. Section 8.3 then proceeds to consider sound speed perturbations. Finally, Section 8.4 draws conclusions from the presented experiments on the benefits and limitations of using learned IEs as transmission conditions in sweeping preconditioners.

8.1 Implementation of learned IEs in axisymmetric setting

The experiments in Chapter 6 were limited to the case in which sound speed and density are radially symmetric and there is no background flow. If one of these assumptions breaks down, the equations cannot be reduced to a set of ODEs anymore and solving PDEs in two or three dimensionsal geometries becomes necessary. Therefore, this section illustrates how to implement learned IEs in an axisymmetric setting. The extension to a fully three dimensional discretization is from a theoretical point of view even simpler, yet computationally more demanding.

The framework for treating solar oscillations under axial symmetry has been introduced in [GBD^{+}17] and is briefly recalled here for the special case of a vanishing background flow which is sufficient for the purpose of this section. Let (r, θ, ϕ) denote standard spherical coordinates with z representing the rotation axis of the Sun. Under the assumption that the background model is symmetric around the z-axis the solution u of equation (2.52) can be decomposed into a series of longitudinal Fourier modes $u = \sum_m u_m(r, \theta)e^{im\phi}$. To obtain the coefficients $u_m(r, \theta)$, a set of independent PDEs in the domain $\Sigma = \{(r, \theta) \mid 0 \leq r \leq a;\ 0 \leq \theta \leq \pi\}$ sketched in Fig. 8.1 has to be solved.
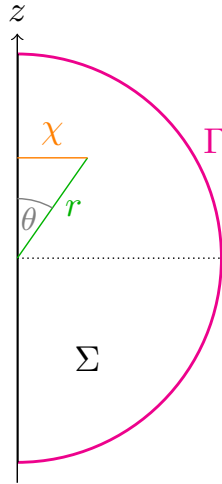
Figure 8.1: Generating section Σ of the geometry for axisymmetric computations. A point in Σ can be described in spherical coordinates (r, θ) with radius r and colatitude θ. Alternatively, coordinates (χ, z) can be used, where $\chi = r\cos(\theta)$ is the distance to the z-axis.

For implementational purposes, it is useful to employ the alternative coordinates (χ, z) in Σ, where $\chi = r\cos(\theta)$ is the distance to the z-axis. Equation (2.52) for each mode is then given by

$$\mathrm{q}_m u_m - \tilde{\nabla} \cdot \tilde{\nabla} u_m = \rho^{1/2}\frac{f_m}{c}, \tag{8.1}$$

with $\mathrm{q}_m = \mathrm{q} - m^2/\chi^2$ and the two dimensional gradient and divergence operators

$$\tilde{\nabla} F = e_\chi \partial_\chi F + e_z \partial_z, \quad \tilde{\nabla} \cdot F = \frac{1}{\chi}\partial_\chi(\chi F_\chi) + \partial_z F_z.$$

Here, e_χ and e_z represent unit vectors in χ and z-direction, respectively. For deriving the variational formulation, integration in Σ is perfomed with respect to the measure $\chi d\Sigma = \chi d\chi dz$ and a homogeneous Neumann boundary condition is assumed on the z axis. The interior part of the sesquilinear form is then given by

$$\int_\Sigma \left(\tilde{\nabla} u_m \tilde{\nabla} \bar{v}_m + q_m u_m \bar{v}_m \right) \chi d\Sigma.$$

To model the solar atmosphere, learned IEs will be imposed on the truncation boundary Γ at $r = a$ as a transparent boundary condition. From the equation above it follows that the sesquilinear forms associated with the operators $\mathcal{M}_\Gamma$ and $\mathcal{K}_\Gamma$ are given by

$$\mathcal{K}_\Gamma(u,v) = \int_\Gamma \tilde{\nabla} u_m \tilde{\nabla} \bar{v}_m \, \chi d\Gamma, \quad \mathcal{M}_\Gamma(u,v) = \int_\Gamma u_m \bar{v}_m \, \chi d\Gamma.$$

The weighting with χ and the homogeneous Neumann boundary condition on the axis ensure that the eigenvalues λ_ℓ in $\mathcal{K}_\Gamma v_\ell = \lambda_\ell \mathcal{M}_\Gamma v_\ell$ are given by $\lambda_\ell = \ell(\ell + 1)/a^2$. The operators $\mathcal{A}$ and $\mathcal{B}$ are the same as introduced in Section 2.4.2, which implies that the *dtn* function for the axisymmetric computation coincides[1] with the one used in the one-dimensional setting considered so far. This in turn coincides with the *dtn* function for a three-dimensional discretization. In particular, the same learned IE matrices A and B can be used in a one-dimensional, axisymmetric or fully three-dimensional discretization of the equations.

(a) 1.5 mHz   (b) 3.0 mHz   (c) 6.0 mHz
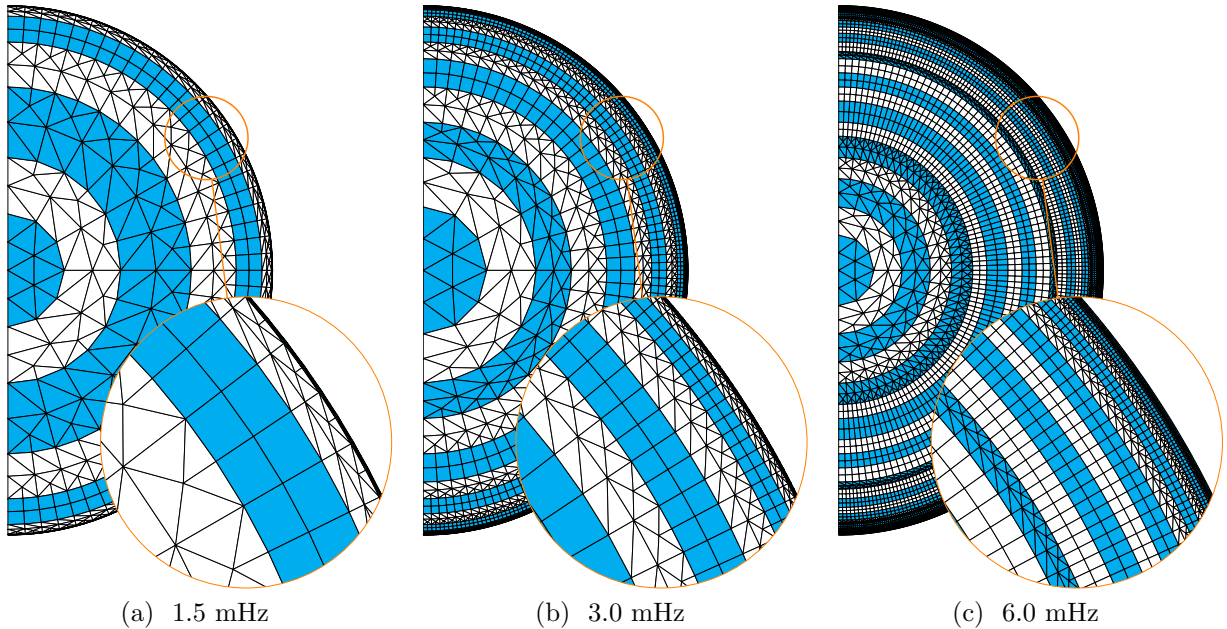
Figure 8.2: Meshes for axisymmetric computation.

The last ingredient required for the implementation is a mesh which resolves the waves properly. To this end, it is well-known that a constant number of DOFs per wavelength

---

[1]In the actual implementation the Sun is again scaled to the unit ball as in Chapter 6. To avoid cluttering of notation, this subtlety is suppressed here.

are necessary to maintain accuracy of the computational solution as the frequency increases. Hence, mesh construction should be based on the wavelength in radial direction given by $\lambda_r = 2\pi\omega/c(r)$. Our strategy for mesh generation is to start at the solar core and successively add concentric layers with thickness equal to the local wavelength to the mesh until reaching the truncation boundary. Since each such layer contains two elements in radial direction, a resolution of $2p$ DOFs per radial wavelength is ensured. Because the oscillations in the angular direction are expected to be of similar order as in the radial direction, we allow the grid cells in θ direction to be at most one and a half times longer than in r direction. This results in a very large number of mesh elements close to the surface as the example meshes for $1.5, 3.0$ and 6.0 mHz displayed in Fig. 8.2 show. The strategy employed here is motivated by [GBD$^+$17], yet differs from the approach considered in this publication to some extent, most notably in the fact that these researchers used a fixed mesh constructed for 9.0 mHz to carry out the computations for all the frequencies in the range $[0, 9.0]$ mHz. Thereby the issue of how many points per wavelength are actually needed to obtain an accurate solution for lower frequencies, e.g. 3.0 mHz, is avoided, since the problem is overresolved. While this approach might be resonable for an axisymmetric simulation, it is clearly infeasible for three-dimensional computations. Therefore, the question of how many points per wavelength are necessary to obtain accurate helioseismic quantities has to be studied. The following experiment may give an indication based on the computed power spectrum.
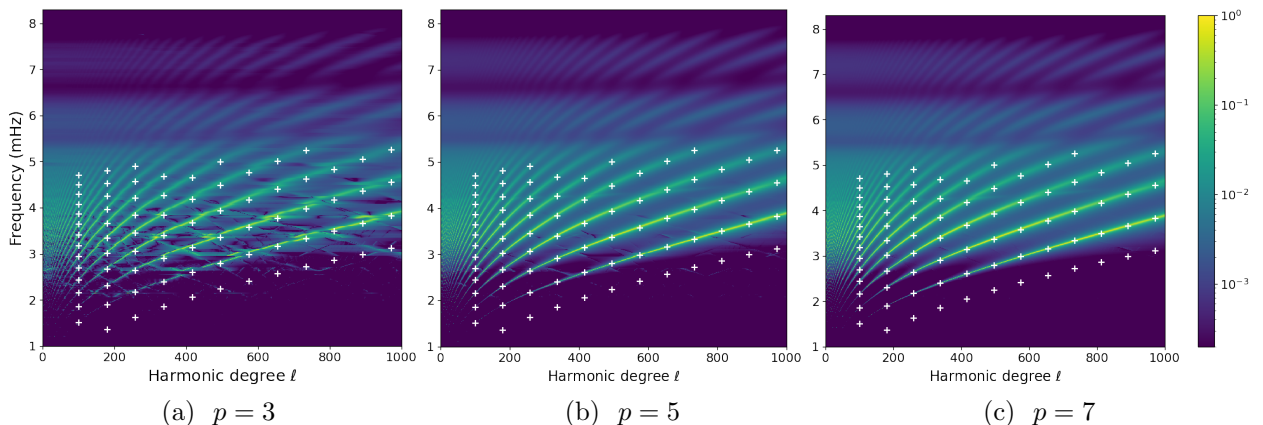
Figure 8.3: Power spectrum computed using elements of order p, which results into $2p$ DOFs per radial wavelength.

The power spectrum is computed based on the VAL-C model for the solar atmosphere. Learned IEs are implemented as transparent boundary conditions directly at the photosphere (i.e. $a = R_\odot$ or $\hat{a} = 1$ in scaled coordinates). In particular, the computational domain, respectively the mesh, only extends up to the photosphere. The Green's function G_m is obtained by solving equation (8.1) for a source $f_m \sim \delta(x - x_\mathrm{s})$. The source x_s is placed at the north pole so that only the $m = 0$ mode is present in the Green's function $G_m := G$. The power spectrum is obtained as usual by projecting the Green's function into spherical harmonics

$$P^\ell(\omega) = \frac{\Pi(\omega)}{\omega} \int\limits_0^\pi \operatorname{Im} G(\theta, \omega) P_\ell(\cos(\theta)) \sin(\theta) \, \mathrm{d}\theta,$$

see [GBD$^+$17, Section 7.5] for a detailed derivation of this formula.

168

The power spectra for learned IEs with $N = 3$ using $6, 10$ and 14 DOFs per radial wavelength are shown in Fig. 8.3. Firstly, the right plot, which displays the results using 14 DOFs per wavelength, demonstrates that learned IEs seem to work well in the axisymmetric setting since the power spectrum is in good agreement with the observational results (white crosses) as well as with the result obtained using the one-dimensional discretization (Fig. 6.2b). The left plot demonstrates that 6 DOFs per wavelength are insufficient to obtain a reasonable accuracy since unphysical artifacts are clearly visible in the power spectrum. Note that the most severe problems occur for the low frequencies on which the transparent boundary condition has little influence. Hence, these problems are not caused by the latter but stem from insufficient spatial resolution. Some artifacts even remain when 10 DOFs per wavelength are employed. Based on this experiment it seems reasonable to recommend a resolution of 12 DOFs per wavelength or higher. It should be investigated in future studies whether this demand can be relaxed by constructing meshes which feature a local refinement around the Dirac source.

The presented experiment demonstrates that discretizations in dimension larger than one lead to extremely large linear systems. This serves as motivation for the second part of this chapter in which the application of sweeping preconditioner for the iterative solution of these systems will be studied.

8.2   Sweeping for the background model

This section presents the results for sweeping for the background Model S in the solar interior and model VAL-C for the atmosphere. Wave attenuation is modelled according to the power law model as shown in Fig. 6.3b. A Dirac source placed at the surface provides the right hand side. The tolerance for GMRES is set to 10^{-6}. Sweeping is performed on the meshes described in the previous Section 8.1 (see Fig. 8.2 for examples) and the number of layers J scales accordingly[2]. As in Chapter 7 only the continuous eigenvalues are used in the construction of learned IEs. To improve the accuracy of the discrete representation of the Laplace-Beltrami operator the subdomain interfaces are curved by applying an appropriate deformation to the mesh.

The GMRES iteration numbers for sweeping with learned IEs are shown in Table 8.1. A comparison with moving PMLs is not presented here since this approach fails to converge in an acceptable number of iterations. Using learned IEs GMRES converges despite violation of the tensor-product structure, albeit in a significantly larger number of iterations than for the simple model problems considered in Chapter 7. The iteration numbers for $p = 4$ tend to be a bit larger than for $p = 6$. This may be related to the fact that the problem with $p = 4$ is still underresolved as dicussed in Section 8.1. Moreover, it is interesting to note that the iteration numbers decrease up to $N = 4$ or $N = 5$ and then start to stagnate. This is reminiscent of the behavior observed for the model problem on simplicial meshes as considered in Section 7.3.4. A plausible explanation could

---

[2]The scaling of J (see e.g. Table 8.1) is not perfectly linear in ω since the mesh construction mechanism introduced in Section 8.1 incorporates some additional special constraints. As explained before, the radii of the layer interfaces are determined according to the radial wavelength. However, there are some exceptions. For example, to prevent the layer which includes the solar core from becoming thicker than $0.3R_\odot$ (which would lead to unacceptable resolution) we break this rule and enforce a refinement. Another exception may occur when the marching algorithm reaches the radius of the truncation boundary a. To preserve the mesh quality, we take care here that the last and second to last layer are of comparable thickness.

| | | | | | $p = 4$ | | | | |
mHz	J	$N = 0$	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$
1	5	61	23	12	10	9	9	9	9
2	8	-	137	61	33	30	29	29	29
3	11	-	-	109	48	19	19	19	19
4	14	-	-	89	64	16	16	16	16
5	18	-	208	60	35	36	37	38	38
6	21	-	129	27	25	72	15	14	13
7	25	-	-	51	42	41	62	193	97

| | | | | | $p = 6$ | | | | |
mHz	J	$N = 0$	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$
1	5	65	20	14	9	9	9	9	9
2	8	-	155	75	31	25	23	23	23
3	11	-	-	124	58	16	13	12	12
4	14	-	-	88	80	15	13	12	12
5	18	-	203	52	28	32	34	35	35
6	21	-	181	32	25	16	14	14	14
7	25	-	-	59	43	49	41	39	39

Table 8.1: GMRES iteration numbers for solar background model using VAL-C as atmospheric model. A dash '-' means that the desired tolerance was not achieved after 250 iterations.

be as follows. Recall that deviation from the tensor-product structure introduces off-diagonal blocks into the representation of $\mathrm{DtN}_j^{\mathrm{ext}}$ in the discrete eigenbasis as illustrated in Fig. 7.7. However, the elements in blocks associated with the most significant part of the spectrum have a substantially smaller magnitude than the diagonal entries. By construction, increasing N only improves the approximation of these diagonal entries[3] via learned IEs. The off-diagonal blocks remain completely untreated and will at some point start to dominate the overall error even though they have comparatively small magnitude. Treating this issue would require the development of learned IEs which are not tied to tensor-product discretizations.

From the results of Table 8.1 no clear behavior of the iteration numbers on the frequency can be detected even though in light of the experiments presented in Chapter 7 a mild growth for fixed N could have been expected. Different factors are at work here which may account for this behavior. Firstly, the error introduced by violation of the tensor-product structure may simply dominate the overall error, i.e. the iteration numbers are already so large that the growth with frequency cannot be detected. Secondly, the damping in the power law model (see Fig. 6.3b) increases with the frequency. This is a significant aid for keeping the iteration numbers under control as the frequency grows. Thirdly, the medium properties of the Sun are not constant as for the majority of experiments considered in Chapter 7. Therefore, waves propagating at different frequencies inside the Sun behave differently. For example, the steep density gradient in the surface layers effectively represents a Dirichlet boundary condition for low-frequency waves which leads to reflections while high-frequency waves may escape into the atmosphere. This will

---

[3]Actually, even the diagonal entries of $\mathrm{DtN}^{\mathrm{ext}}$ are also only approximately represented by $\mathrm{dtn}^{\mathrm{ext}}(\lambda_\ell)$.

in turn leave traces in the corresponding dtn$^{\text{ext}}$ functions as will be investigated further in Section 8.3. Overall, the usual rule that the level of difficulty increases proportionally to the frequency should be considered with some caution for the helioseismology problem.

8.3 Sweeping for sound speed perturbations

For applications in local helioseismology one would like to be able to compute solutions also for perturbations of the background model. In this section the performance of sweeping with learned IEs for a very simple model of sound speed perturbations is considered. Let $c_S(r)$ denote the sound speed of the background model, which serves as the reference in the construction of learned IEs. Then the perturbed equation whose solution is desired is formulated in terms of the perturbed sound speed

$$c := c_S(r)\sqrt{1 - \varepsilon}, \tag{8.2}$$

where ε is a constant. This gives rise to a sound speed perturbation of strength $\varepsilon = |c_S^2 - c^2|/c_S^2$. The perturbation is set to zero for r larger than the solar radius so that the atmosphere remains unperturbed.

The advantages of considering such simple perturbations from the background model are twofold. Firstly, as the perturbed model remains separable a similar investigation as in Section 7.3.3 can be carried out to examine the sensitivity of the corresponding dtn$^{\text{ext}}$ functions. This will allow to gain insights into the arising issues, which is the main objective of this work. Secondly, as only the sound speed is perturbed while the other parameters remain fixed, i.e. no density perturbations or background flows are introduced, a clear correspondence between a single parameter (the strength of the perturbation ε) and the iteration numbers will emerge. Hence, a statement can be made about the strongest perturbations for which sweeping with learned IEs is still effective. From helioseismic inversions it is known that perturbations of strength $\varepsilon = 0.4$ % [KSS^{+}97, BCCD^{+}97] have to be expected.

mHz	$\downarrow J/\varepsilon \rightarrow$	0.0 %	0.0625 %	0.125 %	0.25 %	0.5 %	1 %	2 %
1	5	9 (1)	10 (6)	11 (7)	12 (9)	14 (11)	16 (14)	21 (20)
2	8	23 (1)	28 (12)	33 (16)	43 (24)	59 (37)	80 (52)	115 (80)
3	11	12 (1)	14 (11)	17 (15)	25 (23)	45 (44)	87 (85)	169 (166)
4	14	12 (1)	12 (7)	13 (9)	15 (11)	18 (15)	26 (24)	52 (50)
5	18	35 (1)	36 (6)	36 (7)	37 (8)	39 (9)	43 (12)	53 (17)
6	21	14 (1)	14 (6)	14 (6)	14 (7)	14 (9)	15 (11)	18 (15)

Table 8.2: GMRES iteration numbers for sweeping with learned IEs using $N = 7$ for a sound speed perturbation of strength ε. The numbers given in round brackets were obtained for preconditioning with the direct solver inverse of the background model.

The GMRES iteration numbers in terms of the perturbation are displayed in Table 8.2. Here, learned IEs were used to realize the inverse of the background model since this is expected[4] to be more robust than other approaches (see the discussion in Section 7.3.3).

---

[4] Other experiments not shown here prove that this intuition derived from the simple model problem treated in Section 7.3.3 indeed applies to the setting of helioseismology.

For comparison the iteration numbers for preconditioning with the direct solver inverse of the background model are displayed in round brackets.

For small perturbations significantly lower iterations numbers are obtained with the direct solver since the approximation of DtN$^{\text{ext}}$ cannot be improved beyond a certain treshold due to violation of the tensor-product structure. However, the differences tend to level out as ε increases as the error caused by deviation from the background model starts to dominate.

The strength of the perturbations which can be tolerated is observed to depend on the frequency in a manner which may run counter to the intuition of most non-helioseismologists. For instance, at 3 mHz the iteration numbers are extremely sensitive to the perturbation while for the higher frequency of 6 mHz perturbations of even 2 % can easily be tolerated. As in Section 7.3.3 an analysis of the $\text{dtn}_{\varepsilon}^{\text{ext}}$ functions belonging to the perturbed problem sheds light on this matter. These functions and their relative error

$$| \text{dtn}_{\varepsilon=0}^{\text{ext}}(\lambda) - \text{dtn}_{\varepsilon}^{\text{ext}}(\lambda)|/|\text{dtn}_{\varepsilon=0}^{\text{ext}}(\lambda)|$$

with respect to the background model are compared for 3.0 and 6.0 mHz in Fig. 8.4. For both frequencies an interface close to $0.86R_{\odot}$ has been considered. While dtn^{ext} admits poles close to the real axis for 3.0 mHz which render it extremely sensitive to perturbations, it is smooth for 6.0 mHz and thus relatively stable against perturbations. From previous investigations it is known that the appearance of poles close to the real axis is associated with the reflection of waves. As mentioned before, the reflection of low-frequency waves is caused by the steep gradients in the surface layers. High frequency waves on the other hand can penetrate through this obstacle and escape into the atmosphere. Therefore, their corresponding dtn^{ext} functions remain unaffected by the strong stratification of the surface layers.

8.4 Discussion and assessment of learned IEs for sweeping

Finally, it is reasonable to draw some conclusion from the experiments presented in Chapter 7 and Chapter 8. The aim here is to provide guidelines in which kind of situations learned IEs in their current form can serve as suitable transmission conditions in sweeping preconditioners. To this end, the discussion is arranged in terms of properties of the medium in which the waves propagate.

Smooth media. The simplest case of a smooth medium is represented by the constant coefficient Helmholtz equation with PML boundary conditions as considered in Section 7.3.1. Notwithstanding our result that learned IEs can achieve lower iteration numbers than moving PMLs, we would still recommend to stick to moving PMLs to solve such problems. This is not due to efficiency concerns (see Section 7.2.3) but based on the following other reasons.

- Learned IEs derive their extremely high accuracy from a preprocessing step in which the $\text{dtn}(\lambda_{\ell})$ numbers are computed by solving ODEs and the learned matrices A and B are obtained by running an optimization algorithm. Moving PMLs instead operate without invoking additional optimization routines and are consequently easier to realize from an implementational perspective.
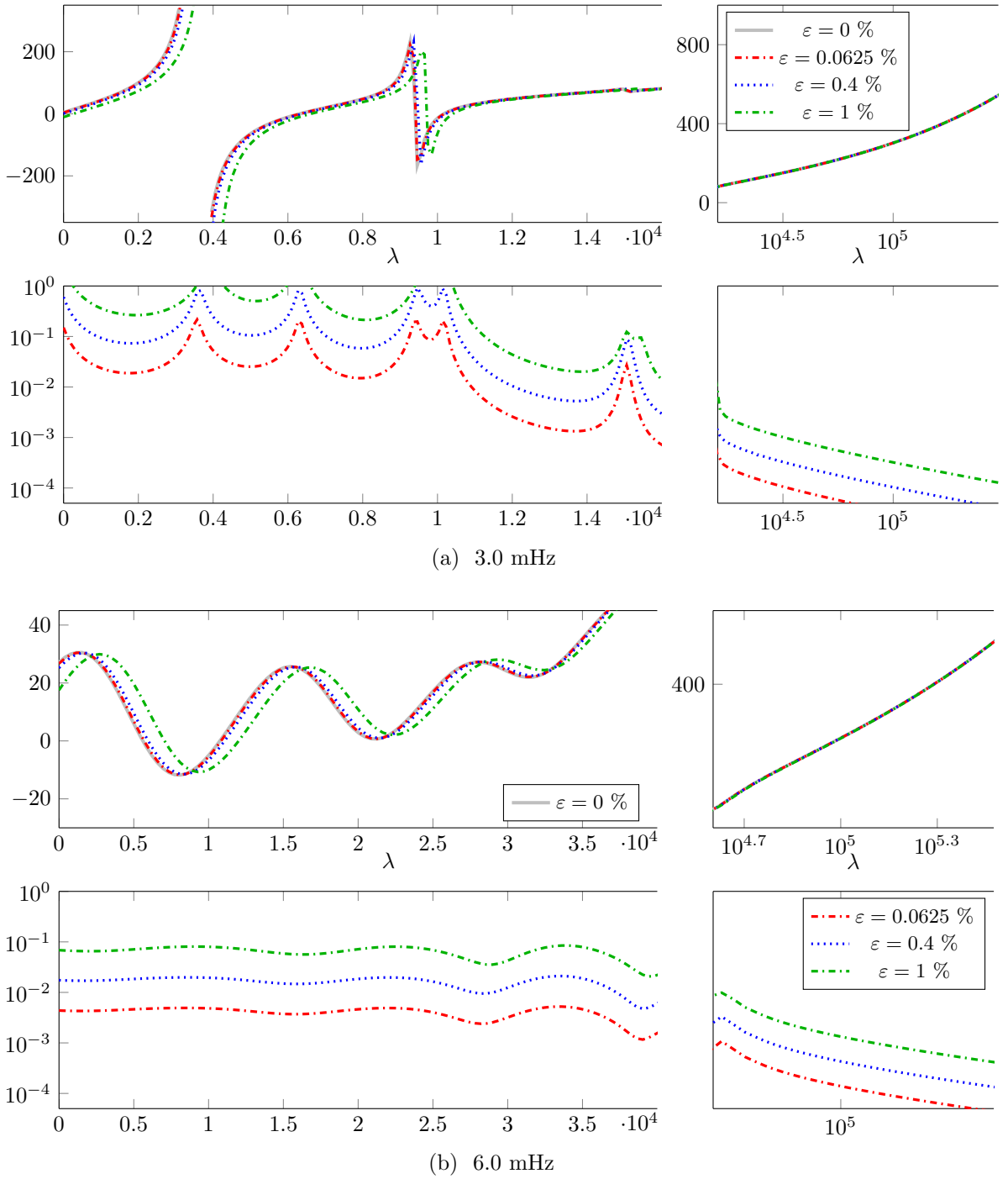
Figure 8.4: The upper panel of figures (a) and (b) shows $\operatorname{Re} \operatorname{dtn}_{\varepsilon}^{\text{ext}}(\lambda)$ corresponding to the perturbed sound speed (8.2) at the respective frequency. The lower panel displays the relative error $|\operatorname{dtn}_{\varepsilon=0}^{\text{ext}}(\lambda) - \operatorname{dtn}_{\varepsilon}^{\text{ext}}(\lambda)|/|\operatorname{dtn}_{\varepsilon=0}^{\text{ext}}(\lambda)|$ w.r.t. to the $\operatorname{dtn}_{\varepsilon=0}^{\text{ext}}$ function corresponding to the background sound speed.

- Moving PMLs do not require separability of the PDE at the continuous level. As long as the medium is free of reflections they seem to work well. On the other hand, for learned IEs in their current form the separability is a crucial assumption. At the moment, only small perturbations from separable background media can be treated.

- Moving PMLs are not tied to a particular type of discretization. For example, Table 7.1 and Table 7.4 demonstrate that the same performance is achieved on tensor-product and simplicial meshes. Even though learned IEs still work on more general meshes (see Section 7.3.4 and Section 8.2), they require tensor-product structure to reach peak performance.

Separable equations with strong reflections. The dtn or dtn$^{\text{ext}}$ functions of these kinds of media are characterized by the occurence of poles close to the positive real axis. The experiments in Section 7.3.1 and Section 7.3.2 demonstrate that learned IEs are ideally suited to resolve this behavior. As the number of poles close to the positive real axis appears to grow linearly with k, the number of infinite element DOFs N would need to be increased accordingly to cover these poles. This does not pose a major issue for the overall complexity of the method since the number of nonzero entries of the learned matrices A and B likewise only grows linearly with N. In addition, irrespective of the dimension d of the problem only $d = 1$ dimensional problems need to be solved to obtain the dtn$^{\text{ext}}$ functions for setting up the corresponding optimization problems for the learned IEs. Therefore, learned IEs appear to be the optimal choice to tackle these problems. The only remaining issue here seems to be the reduced performance once the tensor-product structure is violated (cp. Section 7.3.4 and Section 8.2). Hence, potential extensions of learned IEs which would allow to remove this assumption should be investigated in future studies.

Perturbations from separable equations. At present, learned IEs are solely able to represent DtN maps of separable media. Therefore, when trying to apply them to small perturbations from those separable media the crucial question becomes whether the DtN maps of the perturbed medium remain close to the separable background. If the medium is free of reflections, the results of Section 7.3.3 indicate that this can be answered in the affirmative and consequently learned IEs operate efficiently.

However, in presence of reflections DtN respectively dtn is very sensitive to perturbations. For helioseismology the results of Section 8.3 show that only perturbations up to 0.4 % may be tolerated by the preconditioner. Larger perturbations would cause severe problems especially around the physically very important range of 3 mHz. Even though these results are not entirely satisfactory, we think that an implementation in three dimensions and an application to the iterative solution of inverse problems could be attempted based on the current setup. In this regard, it is an advantage that the preconditioner in the present form only depends on the background model, which implies that it only has to be set up once and can then be used throughout the whole iterative process. On the other hand, this could also turn out to be a limitation since the preconditioner cannot[5] be updated should the iteratively determined model start to deviate significantly from the background. Moreover, the high-sensitivity of dtn to perturbations

---

[5]Perhaps one could define a new background model though by suitably averaging the model obtained from the iterative process.

caused by the presence of reflections implies that sweeping for a partition of the Sun into spherical shells with tensor-product learned IEs can never be stable against perturbations. Therefore, in order to develop a learned IE based preconditioner which is robust w.r.t. perturbations and scales up to the vector-valued equations of helioseismology the assumption of tensor product structure, respectively separability at the continuous level, needs to be overcome.

Additionally, it would be interesting to consider different decompositions of the Sun that are not concentric. For example, instead of sweeping from the core to the surface one could think about sweeping along the angular direction. However, this would require new kinds of learned IEs which incorporate periodicity requirements of the corresponding $\mathcal{DtN}$ maps. In this regard, the work of Joly, Li and Fliss [JLF06] in which $\mathcal{DtN}$ maps for periodic waveguides which contain local perturbations have been constructed could be of particular relevance.

The preceding discussion shows that the introduction of learned IEs as transmission conditions in sweeping algorithms is certainly an interesting development, particularly for problems involving reflections. This provides motivation to pursue further extensions of learned IEs, e.g. to non-separable equations. Some comments on this plan will be given in the outlook.

Chapter 9

Conclusion

On the face of it, the scope of this thesis seems rather broad. We have dealt with seemingly diverse topics such as travel times of waves in the Sun, poles of meromorphic functions and sweeping preconditioners for time-harmonic wave equations. On a closer look, however, all these topics are intimately connected with the $\mathcal{DtN}$ operator of a stratified medium and its representation through a meromorphic dtn function. Whether it be the accuracy of transparent boundary conditions of tensor-product type, the structure of high-frequency time-distance diagrams of the Sun, or the convergence rate of sweeping preconditioners, in the end, it all boils down to understanding and approximating dtn. From a philosophical point of view, this is the main conclusion of this work. A more detailed account on the practical contributions of this thesis is given below.

9.1   Summary of results

Chapter 2: DtN maps for time-harmonic waves in separable exterior domains. At the beginning of this thesis, a generic framework to describe $\mathcal{DtN}$ maps of separable, stratified media has been introduced. In its separable form, $\mathcal{DtN}$ is determined by a scalar function dtn, which is initially defined on the spectrum of a suitable self-adjoint differential operator on the truncation boundary and usually admits a natural meromorphic extension to the complex plane. We identified dtn for several problems arising in applications, e.g. obstacle scattering or helioseismology, and demonstrated that the corresponding dtn functions fulfill a common set of properties. Indeed, for the $dtn^{\mathrm{VAL-C}}$ function of the solar atmosphere and the $dtn^{\mathrm{hom,2d}}$ function of the homogeneous medium we managed to verify these properties analytically. Based on features of dtn functions and some natural assumptions, for example ellipticity, a well-posedness analysis for the problem on the truncated domain with $\mathcal{DtN}$ as transparent boundary condition has been presented. While such a well-posedness result for the constant coefficient Helmholtz equation is surely not new, see e.g. reference [DI01], the main novelty of our approach lies in its generality, accomplished by identifying universal properties of dtn functions.

Chapter 3: Tensor-product discretizations of DtN.   The thesis continued at the discrete level by considering transparent boundary conditions of tensor product form. Such conditions lead to DtN_N approximations which can be described in the same way as the continuous $\mathcal{DtN}$ operator. The DtN_N matrix diagonalizes in the discrete eigenbasis of a suitable self-adjoint operator on the truncation boundary and is fully determined

by a rational function denoted by dtn_N. Motivated by similar studies for specific infinite elements in the literature [DI01], we carried out a semi-discrete error analysis, which shows that the H^1-error for the approximate solution is determined by the ability of dtn_N to approximate dtn on the spectrum of $-\Delta_\Gamma$ or a more general self-adjoint operator on the truncation boundary. This result is valid for any transparent boundary condition of tensor-product form provided that dtn_N fulfills certain properties. By appealing to results of rational approximation theory and exploiting the meromorphic structure of dtn, we showed that exponential convergence rates on compact subsets of the spectrum are possible. We also extended this easily to a global convergence result on the complete spectrum and in the fully-discrete setting, albeit without an explicit rate and under restrictive assumptions on the mesh width. Our approach via rational approximation theory is novel and holds the potential of unifying and extending the exponential convergence results which have been derived for various transparent boundary conditions of tensor-product type in the literature, see e.g. [LS98, HSZ03b] for tensor-product PMLs, [DG98, DI01] for conjugated Astley-Leis infinite elements and [HN09] for Hardy space infinite elements. Note that the cited references do not cover[1] highly inhomogeneous exterior media like the solar atmosphere towards which our approach is geared.

Chapter 4: Learned infinite elements for individual wavenumbers. This chapter introduced learned IEs, which are novel transparent boundary conditions for stratified media. They are obtained by solving a non-linear least squares problem in which the misfit between dtn and dtn_N is penalized. After a reduction step, the number of additional DOFs in the linear systems, which remain sparse, only grows linearly with N. This makes learned IEs extremely competetive with other popular transparent boundary conditions. Several numerical examples, e.g. treating inhomogeneous exterior domains and waveguides, illustrate the great flexibility of our approach.

The results of this chapter are to a large extent based on the paper [HLP21], yet this thesis offered the following extensions. The results for elliptical truncation boundaries presented in this thesis are new. In [HLP21] only the setting in which separation with respect to the Laplace-Beltrami operator is possible has been treated, which does not cover the case of elliptical coordinates. The other extensions are more subtle. Firstly, the reduced ansatz for the learned IE matrices introduced in Section 4.1.1 differs from the one used in the paper. Here, we imposed symmetry as an additional assumption, which leads to $N-1$ less free complex parameters in the optimization problem and is beneficial for efficient treatment of the arising linear systems. This additional assumption does not impair the accuracy of learned IEs with respect to the paper. Secondly, we studied the condition number of the learned IE matrices and found that the distance of the poles of dtn to the spectrum of the self-adjoint operator on the truncation boundary is an important factor.

Chapter 5: Learned IEs providing uniform approximation in the wavenumber. The learned IEs of Chapter 4 can only be used for one specific wavenumber. By modifying the objective function, this chapter introduced learned IEs which can be used for all wavenumbers in a bounded region of the complex plane. An application to the computation of resonances for a simple model problem has been presented. If the objective

---

[1]Among the cited references only [HSZ03b] deals with the inhomogeneous case. However, the medium is requested to allow for a description by an analytic function which does not cover the case of jumping wavenumbers or the VAL-C atmosphere of the Sun.

function is suitably chosen, then its value seems to predict the validity of the computed resonances reliably and only linear eigenvalue problems need to be solved. The results of this chapter are new and unpublished.

Chapter 6: Modelling the solar atmosphere with learned IEs. The complexity of the solar atmosphere requires flexible transparent boundary conditions. The results of this chapter show that learned IEs seem to be ideally suited for this purpose. In particular, the accuracy of helioseismic observables, like the expectation value of the cross-covariance, improves exponentially fast with increasing N. This renders learned IEs very efficient and accurate. They also represent the first transparent boundary conditions that are able to incorporate the drastic temperature gradient present in realistic models of the solar chromosphere. So far only the approximation of $dtn^{\mathrm{VAL-C}}$ by learned IEs, cp. Fig. 6.17, has been published in the literature [HLP21]. The remainder of this chapter is new.

Chapter 7: Learned IEs as transmission conditions in sweeping preconditioners. The transmission conditions between subdomain interfaces in sweeping algorithms are usually approximated using a moving PML. In the presence of strong reflections, approximation of dtn by PML fails, which may lead to complete breakdown of the sweeping preconditioner. Our results show that for separable media which allow for tensor-product discretizations this problem may be overcome by exchanging the moving PML with learned IEs. It should be stressed though that the assumption of separability is crucial. The presence of reflections renders dtn functions extremely sensitive to perturbations. Therefore, learned IEs in their current form can only be used for non-separable equations if these can be regarded as nanoscopic perturbations of separable equations.

While the use of learned IEs in sweeping preconditioners is new, the insight that reflections render $\mathcal{DtN}$ very sensitive to perturbations has been published previously in [PHL20]. Here we extended the discussion from this paper by linking high sensitivity of $\mathcal{DtN}$ to the pole structure of the corresponding dtn functions.

Chapter 8: Sweeping for helioseismology in axisymmetric setting. While the experiments considered in Chapter 7 are toy problems, which were cooked up to isolate specific difficulties, this chapter presented an application of sweeping with learned IEs to a realistic problem from helioseismology. Notwithstanding that violating the tensor-product structure of the mesh leads to a small loss of performance, the iteration numbers for the background model are still acceptable. In consistency with the results of Chapter 7 only small perturbations from the separable background can be tolerated. In particular, the sensitivity peaks around the physically important range of 3.0 mHz, which can be explained by the presence of strong reflections exemplified in the corresponding dtn functions.

9.2 Outlook

Learned IEs have been developed over the course of the last two years. Neither this time period nor the scope of this thesis is sufficient to explore their full potential. Below some open problems and directions for further research are suggested. This list is surely not exhaustive - various other extensions are conceivable.

Learned IEs for non-separable problems and efficient linear solvers for helioseismology. As discussed previously, sweeping with learned IEs can currently only be applied to the solar background model or tiny perturbations thereof. This is because learned IEs can merely represent the $\mathcal{DtN}$ operator of the background model, which, in presence of reflections, differs significantly from the non-separable $\mathcal{DtN}$ operator of the perturbed problem that would actually be required in the sweeping algorithm. To overcome this issue, learned IEs for non-separable equations need to be developed. In this case neither $\mathcal{DtN}$ nor DtN$^{\text{ext}}$ can be (fully) diagonalized, respectively fully described by a scalar dtn or dtn$_N$ function. Therefore, we expect that a significantly larger amount of parameters will have to be included in the optimization problem to obtain a learned DtN matrix which delivers an accurate approximation of $\mathcal{DtN}$. To keep the costs for solving the optimization problems feasible, it is desirable to enforce some sparsity structure between the unknowns. Despite our expectation that the development of learned IEs for non-separable problems will require considerable efforts, in view of their success for separable problems involving strong reflections, we regard such an endeavor as justified.

Vector-valued problems and systems: Galbrun's equation. The scope of this thesis is limited to scalar equations. Many applications involve several fields which are vector-valued, e.g. the propagation of electromagnetic waves described by the Maxwell equations. Since the radiation condition may couple different fields, as for instance the Silver-Müller radiation condition, the corresponding $\mathcal{DtN}$ maps will most likely not allow for a description through a single (scalar) dtn function. The treatment of such equations requires new ideas and is an interesting direction for further research.

In the context of helioseismology the most pressing extension would be the development of learned IEs for the vector-valued Galbrun's equation. Recently, well-posedness of this equation [HH21, Hal21] has been established. Moreover, Halla [Hal21] showed that the system in the solar atmopshere under the assumption of spherical symmetry and the Cowling approximation can be reduced to a fourth-order scalar equation for a certain potential. Although this may allow recourse to the scalar case, non-separable learned IEs could be required to utilize this reduction step.

Replacing the optimization solver by rational approximation techniques. According to the results of Chapter 3, transparent boundary conditions of tensor-product type lead to a rational approximation problem (see equation (3.10)) for determining dtn$_N$. Learned IEs, as introduced in Chapter 4, solve this problem by relying on a non-linear least squares method, which does not make notable use of the rational nature of dtn$_N$. Moreover, non-linear least squares solvers are usually not incorporated as a standard component of finite element packages, which could render learned IEs unattractive for other researchers. Therefore, it would be beneficial if the learned IE matrices could be obtained more directly by a rational approximation algorithm. A candidate for this job could be the recently proposed AAA (adaptive Antoulas–Anderson) algorithm [NST18], which already received a lot of attention. However, it is not immediately clear how to apply this algorithm to our problem since it is formulated for rational functions in barycentric representation which are quotients of two polynomials with exactly the same degrees. It would be a significant advantage, if we could start directly with the more

efficient reduced ansatz from (4.14) which setting $A_{00} = B_{0j} = 0, j = 0, \ldots, N$ results in

$$\text{dtn}_N(\lambda) = -\sum_{j=1}^{N} \frac{A_{0j}^2}{A_{jj} + \lambda} \tag{9.1}$$

and infer the matrix entires directly from the algorithm. Note that this requires an algorithm which is applicable for rational functions which are quotients of polynomials with degrees $N - 1$ and N, respectively. A modification of the AAA algorithm which allows this has been proposed recently by Derevianko and Plonka [DP21]. It would be very interesting to compare its performance to our current approach based on non-linear least squares.

Numerical experiments for three-dimensional problems. In this thesis only two dimensional finite element simulations have been presented. From a theoretical point of view, the extension of learned IEs to three-dimensional interior domains is straight-forward. Also, the *dtn* function for the solar atmosphere is the same in a fully-separable, axisymmetric or three-dimensional setting. Therefore, for helioseismology one can use the same learned IE matrices in any case and they deliver always the same accuracy. The difficulty of three-dimensional simulations rather lies in their increased computational effort which requires a careful implementation and sufficient computational resources.

Separation beyond elliptical coordinates. A natural generalization of elliptical co-ordinates exists in three dimensions. Prolate or oblate spheroidal coordinates are ob-tained by rotating the elliptical coordinate system around a focal or non-focal axis of the ellipse. The Helmholtz equation in these coordinates is separable into so called spheroidal wave functions. This may be of interest for the efficient treatment of e.g. cigar-shaped obstacles.

Extension of well-posedness analysis at continuous level. The well-posedness analysis for the interior problem with $\mathcal{DtN}$ as transparent boundary condition does not cover the case of elliptical truncation boundaries in which $\mathcal{DtN}$ is only separable up to an isomorphism, which represents a scaling with the length of the normal vector. Likewise, it was assumed in (C-IV) that the diffusion matrix is proportional to the identity on Γ. Violation of this assumption would lead to a similar issue. In order to generalize the analysis it would clearly be desirable to eliminate these assumptions. To this end, the application of more sophisticasted analytical tools than employed in this thesis, e.g. the concept of T-coercivity [BCZ10], should be investigated.

Extension of error analysis for transparent boundary conditions of tensor-product type. The error analysis for tensor-product discretizations of exterior prob-lems presented in Chapter 3 should be extended in various ways. First and foremost, the exponential convergence rates have to be extended from compact subsets to the en-tire spectrum. We presented a corresponding attempt for the Helmholtz equation in Section 3.3.3, which still contains one factor whose dependence on N is not explicitly controlled yet, see Remark 3.25. It should be investigated whether the growth bounds on $|\textit{dtn}(\lambda)|$ valid for $\lambda \in \mathbb{R}_+$, which follow from Lemma 2.3, can be extended to the regular-ized $\textit{dtn}_{N_p}^{\text{reg}}(\lambda)$ function for λ within certain ellipses $D(N^{2\alpha}, \eta) \subset \mathbb{C}$. This would allow to

conclude global exponential convergence from Proposition 3.24. It should also be mentioned that the whole endeavor would be much simpler if stronger rational approximation results than in Proposition 3.13 could be derived. The shortcoming with the result of this proposition is its exponential dependence on the length of the interval, respectively supremum norm of $dtn_{N_p}^{\text{reg}}$. Further exploration of rational approximation theory may help to overcome these deficits and facilitate the derivation of exponentially fast convergence rates on a global scale. Our final aim is to extend the fully-discrete convergence result of Corollary 3.23 to an explicit error estimate in N and the mesh width h.

Relation between poles of dtn and resonances. Several indications for a relation between the poles of dtn and resonances phenomena can be found in this thesis which should be explored further. As pointed out in Section 6.1.2, by generalizing the results of [Tay11, Chapter 9.7] it should be possible to identify poles of $dtn^{\text{VAL-C}}(\hat{\lambda}_\ell, \hat{\sigma}(\omega))$ with respect to the frequency ω as resonances. This would be an interesting theoretical contribution to the area of high-frequency helioseismology. Then again, in the context of this thesis the poles of $dtn^{\text{VAL-C}}$ with respect to the first argument $\hat{\lambda}_\ell$ have been the primary focus of interest. In Remark 2.12 these poles have been characterized as those complex numbers for which the corresponding ODE problem is not well-posed. Moreover, in several numerical experiments it was observed also for other dtn functions that the presence of strong reflections in the medium, which allow for the occurence of resonance phenomena, pushes the poles closer to the real axis. Adding attenuation to the problem will in contrast increase the imaginary part of the poles. At present, a sound theoretical explanation of these phenomena is missing.

Reduction step for learned IEs working uniformly in the wavenumber. Learned IEs which work for all wavenumbers in a bounded set as introduced in Chapter 5 do not yet offer the same level of efficiency as their single-wavenumber compatriots from Chapter 4. This stems mainly from adhering to dense learned IE matrices and the large number of residual blocks which arise from having to fit dtn now also as a function of the wavenumber. A sparsification step for the learned IE matrices seems inevitable. This could e.g. be based on a generic ansatz using matrices with a fixed bandwith or by including knowledge about the trajectories of poles of dtn into the objective function. If these attempts to improve efficiency turn out to be successful, then applications to large-scale resonance problems seem realistic. Another interesting research direction would be the use of learned IEs in time-domain simulations.

Approximation of dtn in the supremum norm. The objective function for learned IEs is currently based on a weighted ℓ^2-Norm, which is justified according to Theorem 3.20 (ii) and (iii), whereas the approximation error estimate for dtn_N functions established in Section 3.2.3 holds in the supremum norm. To eliminate this discrepancy[2] it would be interesting to solve the minimization problem for learned IEs in the supremum norm. In this regard, we may benefit from techniques developed in the field of control theory since the minimization of analytic functions in the supremum norm is a central part of so called H_∞ methods used in this field, see e.g. [ZD98]. The role of the dtn_N function is played here by the so-called transfer function which relates inputs and outputs of the system which should be controlled. Often the transfer function seems to have a similar

---

[2]On finite subsets of the spectrum this is not really an issue due to equivalence of norms.

Schur complement structure as dtn_N which could facilitate the use of these techniques for our purpose.

Learned perfectly matched layers. A considerable amount of research has been carried out on improving the PML method, for example, on optimizing the profile of the absorption function [BHNPR07, BHNPR08]. In light of the results from Chapter 3, the accuracy of a tensor-product PML can be assessed in terms of how accurate the $dtn_N\{A, B\}$ function approximates dtn. The matrices A and B have been explicitly derived in Appendix C.1 in terms of the PML parameters, e.g. the absorption profile and the thickness of the extension layer. To optimize the PML parameters it therefore seems natural to proceed analogously as for learned IEs. One could set up a minimization problem as in (4.2)-(4.3), yet, instead of optimizing for the matrix entries of A and B, one would optimize for the PML parameters that minimize the misfit. It should be kept in mind though that this approach is still limited by the requirement that the underlying medium allows for an analytic continuation of the solution into the complex plane. Therefore, it will not be able to reach the flexibility of learned IEs.

Appendix A

Tools from functional analysis and theory of PDEs

This appendix recalls some basic analytical tools needed in the analysis presented in Chapter 2 and Chapter 3.

A.1 Functional analysis

Definition A.1. *For two Hilbert spaces X and Y denote the set of bounded linear operators mapping from X to Y by $\mathscr{L}(X,Y)$.*

For $T \in \mathscr{L}(X,Y)$ and $y \in Y$ consider the problem of finding $x \in X$ such that

$$Tx = y. \tag{A.1}$$

If T is invertible with $T^{-1} \in \mathscr{L}(Y,X)$, then clearly $x = T^{-1}y$ and (A.1) is well-posed. The Fredholm alternative covers the more general case when T is only invertible up to a compact perturbation.

Theorem A.2 (Fredholm alternative)**.** *Let $T = S - K$, where $K \in \mathscr{L}(X,Y)$ is compact and $S \in \mathscr{L}(X,Y)$ bijective with $S^{-1} \in \mathscr{L}(Y,X)$. Then one of the following holds.*

- *Either the equation (A.1) is uniquely solvable for any $y \in Y$.*

- *Or, it holds that $0 < \dim(\ker(T)) = \dim(\ker(T^*)) < \infty$ and (A.1) has a solution if and only if the right hand side $y \in Y$ fulfills the compatability condition*

$$(y, x)_Y = 0 \text{ for all } x \in \ker(T^*).$$

Proof. Upon multiplying (A.1) by S^{-1} and using the ideal property of compact operators we can assume that $S = \mathrm{Id}$. The statement for this case is shown in [Bre11, Theorem 6.6] or [Eva98, Appendix D, Theorem 5]. □

A.2 Tools from PDE theory

First we introduce a right inverse for the trace operator.

Theorem A.3 (Right inverse of trace operator)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. There exists a bounded linear operator $Z : H^{1/2}(\partial\Omega) \to H^1(\Omega)$ such that $\mathrm{tr} \circ Z g = g$ for all $g \in H^{1/2}(\partial\Omega)$.*

Proof. See [Wlo87, Theorem 8.8] or [McL00, Thm. 3.37]. $\qquad\square$

Functions in $H^2(\Omega)$ have a normal trace in $H^{1/2}(\Omega)$ if $\partial\Omega$ is sufficiently regular. Green's first identity also allows to define the normal trace of a function $u \in H^1(\Omega)$ under the additional requirement that $\Delta u \in L^2(\Omega)$.

Proposition A.4 (Weak normal derivative)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and let $\mathbf{n}$ denote the exterior normal vector on $\partial\Omega$. If $u \in H^1(\Omega)$ and $\Delta u \in L^2(\Omega)$, then*

$$\left\langle \frac{\partial u}{\partial \mathbf{n}}, v \right\rangle_{\partial\Omega} := \int_\Omega \nabla u \nabla Z v + \Delta u Z v \, \mathrm{d}x \tag{A.2}$$

defines a weak normal derivative in $H^{-1/2}(\partial\Omega)$ and

$$\left\| \frac{\partial u}{\partial \mathbf{n}} \right\|_{H^{-1/2}(\partial\Omega)} \leq \|Z\| \left(\|u\|_{H^1(\Omega)} + \|\Delta u\|_{L^2(\Omega)} \right). \tag{A.3}$$

For $u \in C^2(\bar\Omega)$ this definition coincides with the usual normal derivative.

Proof. The Green's formula (A.2) and the estimate[1] (A.3) are proven in [McL00, Lemma 4.3] (in fact, for general second order operators in divergence form and under slightly weaker regularity assumptions). Additionally, the statement that definition (A.2) coincides for $u \in C^2(\bar\Omega)$ with the usual normal derivative follows from Green's first identity and density of $C^1(\bar\Omega)$ in $H^1(\Omega)$. $\qquad\square$

The uniqueness proof given in Proposition 2.15 appeals to the unique continuation principle for elliptic PDEs in the following form.

Theorem A.5 (Unique continuation principle)**.** *Let $\mathcal{L}$ be an elliptic partial differential operator of order two with C^∞ real-valued coefficients in the principal part and L^∞_{loc} complex-valued lower-order terms in some open connected subset Ω of $\mathbb{R}^d$. If Ω_0 is a non-empty open connected subset of Ω, if u is an $H^1_{\mathrm{loc}}(\Omega)$ function such that $\mathcal{L}u$ belongs to $L^2_{\mathrm{loc}}(\Omega)$ and such that*

$$|\mathcal{L}u| \leq C(|u| + |\nabla u|) \text{ in } \Omega, \quad u|_{\Omega_0} = 0,$$

then $u = 0$ in Ω.

Proof. See [Ler19, Thm. 3.8.]. $\qquad\square$

---

[1] In the reference there is a generic constant C on the right hand side of (A.3). However, in our case it is easy to see that $C = \|Z\|$.

Appendix B

Spectral characterization of Sobolev spaces on Γ

Let Γ be the outer boundary of some bounded domain $\Omega_{\text{int}} \subset \mathbb{R}^d$. We assume that Γ admits the structure of a smooth $d-1$-dimensional closed manifold. The $\mathcal{D}t\mathcal{N}$ operators, which are the central object of this thesis, operate on certain function spaces defined on Γ. To describe the smoothness properties of $\mathcal{D}t\mathcal{N}$ and connect to the interior PDE posed on Ω_{int}, it is natural to work with traces of functions defined in Sobolev spaces on the interior domain. Different options to define such trace Sobolev spaces exist. A common approach, e.g. presented in [LM72, Chapter 1, Section 7.3], is to cover Γ by a partition of unity and define the trace space based on $H^s(\mathbb{R}^{d-1})$ by employing local coordinates. The resulting spaces are denoted $H^s(\Gamma)$ in the following.

Even though this definition might be convenient to work with in many situations, it is not well-suited for describing the smoothness properties of $\mathcal{D}t\mathcal{N}$. Recall that $\mathcal{D}t\mathcal{N}$ is defined in terms of an elliptic, self-adjoint, second order differential operator $\widetilde{\mathcal{K}_\Gamma} = \mathcal{M}_\Gamma^{-1/2} \mathcal{K}_\Gamma \mathcal{M}_\Gamma^{-1/2}$ with compact resolvent. Let (λ_ℓ, w_ℓ) be the corresponding eigenbasis. In particular,

$$\widetilde{\mathcal{K}_\Gamma} w_\ell = \lambda_\ell w_\ell, \quad \ell \in \mathbb{N}_0.$$

The eigenfunctions $\{w_\ell\}$ are smooth, form a complete orthonormal system of $L^2(\Gamma)$ and the eigenvalues grow towards infinity

$$|\lambda_0| \leq |\lambda_1| \leq \ldots \to \infty,$$

see e.g. [Gil95, Lemma 1.6.3]. Moreover, the ellipticity condition ensures that the spectrum is bounded from below (see [Gil95, Lemma 1.6.4]). If $\widetilde{\mathcal{K}_\Gamma} = -\Delta_\Gamma$ is the Laplace-Beltrami operator, then the lower bound is in fact zero, yet in view of a separation with respect to elliptical coordinates in which part of the spectrum can be negative (see Fig. 2.5b) we should treat the general case here. In the following, we want to show that an equivalent norm on $H^s(\Gamma)$ is given by

$$u \mapsto \left(\sum_{\ell=0}^{\infty} (1 + |\lambda_\ell|)^s \, |\langle u, w_\ell \rangle_\Gamma|^2 \right)^{1/2}. \tag{B.1}$$

This definition is significantly more convenient to work with for analyzing $\mathcal{D}t\mathcal{N}$. Equation (B.1) is a generalization of [LM72, Remark 7.6], which provides an intrinsic definition of $H^s(\Gamma)$ in terms of the Laplace-Beltrami operator on Γ. Its derivation requires some preliminaries.

B.1 Hilbert space interpolation

This section collects some basic facts about Hilbert space interpolation extracted from [LM72, Chapter 1, Section 2.1].

- Let X, Y be two separable Hilbert spaces with inner products $(\cdot, \cdot)_X$ and $(\cdot, \cdot)_Y$ respectively such that

$$X \subset Y, \quad X \text{ dense in } Y \text{ with continuous injection.} \tag{B.2}$$

- We want to define a self-adjoint, positive (spectrum in $[0, \infty)$) unbounded operator

$$\Lambda : \mathrm{Dom}(\Lambda) \to Y, \qquad \mathrm{Dom}(\Lambda) := \{u \in Y \mid \Lambda u \in Y\}.$$

- This can be achieved by the following construction. Let $\mathrm{Dom}(S)$ denote the set of u's such that

$$v \mapsto (u, v)_X, \ v \in X,$$

is continuous with respect to the topology induced by Y. Then an unbounded operator S on Y with domain $\mathrm{Dom}(S)$ can be defined by

$$(Su, v)_Y := (u, v)_X.$$

The operator S is self-adjoint and strictly positive

$$(Sv, v) = \|v\|_X^2 \gtrsim \|v\|_Y^2$$

owing to the continuous injection $X \hookrightarrow Y$. Then one can set

$$\Lambda = S^{1/2}, \tag{B.3}$$

which is self-adjoint, positive in Y and has domain X.

Definition B.1. *Under assumption* (B.2) *and with definition* (B.3) *define*

$$[X, Y]_\theta := \mathrm{Dom}(\Lambda^{1-\theta}), \quad 0 \leq \theta \leq 1.$$

The space $[X, Y]_\theta$ is endowed with the graph norm of $\Lambda^{1-\theta}$, i.e

$$u \mapsto \left(\|u\|_Y^2 + \|\Lambda^{1-\theta} u\|_Y^2 \right)^{1/2}.$$

The following result will be used later: If Λ_1 and Λ_2 are two positive, self-adjoint operators in Y with $\mathrm{Dom}(\Lambda_1) = \mathrm{Dom}(\Lambda_2) = X$, then

$$\mathrm{Dom}(\Lambda_1^{1-\theta}) = \mathrm{Dom}(\Lambda_2^{1-\theta})$$

and the induced graph norms are equivalent. This means that the space $[X, Y]_\theta$ does not depend on the concrete operator Λ used for its definition, yet is linked intrinsically to X and Y.

The following result describes interpolation between the trace spaces:

Theorem B.2. *For all s_1, s_2 in $\mathbb{R}$ with $s_1 > s_2$ it holds that*

$$[H^{s_1}(\Gamma), H^{s_2}(\Gamma)]_\theta = H^{(1-\theta)s_1 + \theta s_2}(\Gamma), \quad 0 \leq \theta \leq 1,$$

with equivalent norms.

Proof. See [LM72, Theorem 7.7 of Chapter 1]. $\qquad\square$

B.2 Spectral characterization of $H^s(\Gamma)$

We apply the interpolation theory from the previous section to $X = H^s(\Gamma)$ for some $s > 0$ and $Y = L^2(\Gamma)$. Note that assumption (B.2) is fulfilled.

- An appropriate operator Λ can be defined as a regularized version of $\widetilde{\mathcal{K}_\Gamma}$, i.e.

$$\Lambda = \widetilde{\mathcal{K}_\Gamma} + c\mathrm{Id},$$

  where $c > 0$ is large enough so that Λ is strictly positive, i.e. $\|\Lambda u\|_{L^2(\Gamma)} \gtrsim \|u\|_{L^2(\Gamma)}$. This is possible since the spectrum of $\widetilde{\mathcal{K}_\Gamma}$ is bounded from below.

- Let $m \in \mathbb{N}$. Then Λ^m is an elliptic differential operator of order $2m$. According to[1] [Gil95, Lemma 1.3.6 (c)], an equivalent norm on $H^{2m}(\Gamma)$ is given by

$$\|u\|_{H^{2m}(\Gamma)} \simeq \|u\|^2_{L^2(\Gamma)} + \|\Lambda^m u\|^2_{L^2(\Gamma)}. \tag{B.4}$$

  Since Λ^m is strictly positive and continuous from $H^{2m}(\Gamma) \to L^2(\Gamma)$, it also holds that

$$\|u\|^2_{L^2(\Gamma)} \lesssim \|\Lambda^m u\|^2_{L^2(\Gamma)} \lesssim \|u\|^2_{H^{2m}(\Gamma)}.$$

  Therefore, (B.4) implies that

$$H^{2m}(\Gamma) = \mathrm{Dom}(\Lambda^m).$$

- Applying the interpolation theory introduced in the previous section to $X = H^{2m}(\Gamma)$ and $Y = L^2(\Gamma)$ then yields

$$[H^{2m}(\Gamma), L^2(\Gamma)]_\theta = \mathrm{Dom}(\Lambda^{(1-\theta)m}), \quad 0 \le \theta \le 1.$$

- On the other hand,

$$[H^{2m}(\Gamma), L^2(\Gamma)]_\theta = H^{2m(1-\theta)}(\Gamma), \quad 0 \le \theta \le 1,$$

  according to Theorem B.2.

- Combining the previous two equations leads to

$$H^{2m(1-\theta)}(\Gamma) = \mathrm{Dom}(\Lambda^{(1-\theta)m}), \quad \text{for } m \in \mathbb{N}, \ 0 \le \theta \le 1.$$

  Since m can be any arbitrary positive integer

$$H^s(\Gamma) = \mathrm{Dom}(\Lambda^{s/2})$$

  for any $s > 0$ follows. As

$$\Lambda^{s/2} u = \sum_{\ell=0}^{\infty} (c_0 + \lambda_\ell)^{s/2} \langle u, w_\ell \rangle_\Gamma w_\ell,$$

  an equivalent norm on $H^s(\Gamma)$ is given by

$$\|u\|^2_{H^s(\Gamma)} \simeq \sum_{\ell=0}^{\infty} (c_0 + \lambda_\ell)^s |\langle u, w_\ell \rangle_\Gamma|^2 \simeq \sum_{\ell=0}^{\infty} (1 + |\lambda_\ell|)^s |\langle u, w_\ell \rangle_\Gamma|^2, \tag{B.5}$$

  where the second equivalence follows by a short calculation.

---

[1] This result is valid for pseudo-differential operators on closed Riemannian manifolds.

- We can extend the norm (B.5) to $s \in \mathbb{R}$ and identify the resulting spaces $H^{-s}(\Gamma)$ for $s > 0$ with the dual space $[H^s(\Gamma)]^*$ in the usual manner via the scalar product on $L^2(\Gamma)$. Indeed, let $l \in [H^s(\Gamma)]^*$ be given. By the Riesz representation theorem there exists $u \in H^s(\Gamma)$ such that $l(v) = \langle u, v \rangle_{H^s(\Gamma)}$ for all $v \in H^s(\Gamma)$. Define

$$\tilde{u} := \sum_{\ell=0}^{\infty} \langle u, w_\ell \rangle_\Gamma \left(1 + |\lambda_\ell|\right)^s w_\ell.$$

Then

$$\|\tilde{u}\|^2_{H^{-s}(\Gamma)} = \sum_{\ell=0}^{\infty} \left(1 + |\lambda_\ell|\right)^{2s} \left(1 + |\lambda_\ell|\right)^{-s} |\langle u, w_\ell \rangle_\Gamma|^2 = \|u\|^2_{H^s(\Gamma)},$$

so $\tilde{u} \in H^{-s}(\Gamma)$ and

$$l(v) = \langle u, v \rangle_{H^s(\Gamma)} = \sum_{\ell=0}^{\infty} \langle u, w_\ell \rangle_\Gamma \left(1 + |\lambda_\ell|\right)^s \overline{\langle v, w_\ell \rangle}_\Gamma = \langle \tilde{u}, v \rangle_\Gamma. \qquad \text{(B.6)}$$

That is, any $l \in [H^s(\Gamma)]^*$ is of the form (B.6) for some $\tilde{u} \in H^{-s}(\Gamma)$ and the estimate

$$|l(v)| \leq \|\tilde{u}\|_{H^{-s}(\Gamma)} \|v\|_{H^s(\Gamma)}$$

holds.

Appendix C

Popular transparent boundary conditions of tensor-product type

This appendix presents a survey on transparent boundary conditions for the constant coefficient Helmholtz equation (2.27) in the exterior of a ball with Sommerfeld radiation condition at infinity as introduced in Section 2.2.1. For ease of notation, the mapping Ψ is omitted here, so that the equation to be treated is

$$\left(-r^{1-d}\partial_r(r^{d-1}\partial_r) - \frac{a^2}{r^2}\Delta_\Gamma - k^2\,\mathrm{Id}\right)u(r,\widehat{x}) = 0, \quad r \geq a, \tag{C.1}$$

$$\lim_{r\to\infty} r^{(d-1)/2}\left(\frac{\partial u}{\partial r} - iku\right) = 0.$$

The wavenumber k is assumed to be a positive constant. The aim is to demonstrate that several popular transparent boundary conditions (e.g. tensor-product PMLs and classical infinite elements) can be described in terms of the tensor-product framework introduced in Section 3.1. To this end, the matrices A and B in the ansatz (3.7) have to be identified. By means of (3.10) this immediately yields their corresponding dtn_N function. According to the results of Section 3.3, the approximation of *dtn* achieved with dtn_N is the decisive measure of accuracy for any transparent boundary condition of tensor-product type. This serves as motivation for Section 3.4 in which a comparative study of the considered conditions based on this criterion is conducted.

C.1   Tensor-product PML

The technique of Perfectly Matched Layers (PML) has been introduced by Berenger [Ber94]. The basic idea is to truncate the unbounded domain by an artificial layer in which the waves are gradually attenuated. In the considered setting of a separable geometry, instead of the exterior domain $[a,\infty) \times \Gamma$ a finite domain $[a,\eta) \times \Gamma$ is considered for $\eta > a$. Within the PML layer, the following transformation is applied to the radial coordinate:

$$\tilde{r} = r + \int_a^r i\sigma(t). \tag{C.2}$$

Here, σ is the absorption coefficient (usually depending on the wavenumber). Apparently, this procedure can be interpreted as a complex coodinate stretching. It is based on the

requirement that the coefficients involved in $\mathcal{A}$ and $\mathcal{B}$ admit an analytic extension into the complex domain. The complex coordinate stretching only affects the operators $\mathcal{A}$ and $\mathcal{B}$ while the tangential operators $\mathcal{K}_\Gamma$ and $\mathcal{M}_\Gamma$ remain unchanged. In particular, the tensor-product structure at the continuous level is preserved. The corresponding matrices A and B can be derived as follows, see also reference [CM98] in which the PML formulation in curvilinear coordinates has been introduced originally.

Define the auxiliary functions

$$s(r) := (1 + i\sigma(r))^{-1}, \quad \tilde{s}(r) := 1 + \frac{i}{r}\int_a^r i\sigma(t). \tag{C.3}$$

Note that $r\tilde{s} = \tilde{r}$ and

$$\frac{\partial \tilde{r}}{\partial r} = 1 + i\sigma(r) = s^{-1}.$$

Therefore,

$$-\tilde{r}^{1-d}\partial_{\tilde{r}}(\tilde{r}^{d-1}\partial_{\tilde{r}}) = r^{1-d}\tilde{s}^{1-d}s\partial_r(r^{d-1}\tilde{s}^{d-1}s\partial_r)$$

and (C.1) in transformed coordinates is given by

$$\left(-r^{1-d}\partial_r(r^{d-1}\tilde{s}^{d-1}s\partial_r) - \frac{a^2}{\tilde{s}^{3-d}r^2}\Delta_\Gamma - \frac{k^2}{\tilde{s}^{1-d}s}\right)u = 0.$$

To derive a weak formulation, this equation is multiplied with a test function $\bar{v}$ which vanishes at $r = \eta$. Using that the measure is given by $(r/a)^{d-1}\mathrm{d}r\mathrm{d}\hat{x}$, where $\mathrm{d}\hat{x} = a^{d-1}\mathrm{d}\mathbb{S}^{d-1}$ in terms of the measure on $\mathbb{S}^{d-1}$, integration by parts on Γ yields:

$$\int_a^\eta \int_\Gamma \left\{ -\partial_r(r^{d-1}\tilde{s}^{d-1}s\partial_r u)\bar{v} + \frac{a^2 r^{d-1}}{\tilde{s}^{3-d}sr^2}\nabla_\Gamma u\nabla_\Gamma \bar{v} - \frac{k^2 r^{d-1}}{\tilde{s}^{1-d}s}u\bar{v} \right\} \frac{1}{a^{d-1}}\mathrm{d}r\mathrm{d}\hat{x} = 0.$$

Integration by parts with respect to r leads to

$$-\int_a^\eta \int_\Gamma \partial_r(r^{d-1}\tilde{s}^{d-1}s\partial_r u)\bar{v}\frac{1}{a^{d-1}}\mathrm{d}r\mathrm{d}\hat{x}$$

$$= \int_a^\eta \int_\Gamma \tilde{s}^{d-1}s\partial_r u\partial_r \bar{v}\left(\frac{r}{a}\right)^{d-1}\mathrm{d}r\mathrm{d}\hat{x} - \int_\Gamma \mathcal{D}t\mathcal{N}(u)\bar{v}_{|r=a}\mathrm{d}\hat{x},$$

with $\mathcal{D}t\mathcal{N}(u) = -(\partial_r u)_{|r=a}$. Here $s(a) = \tilde{s}(a) = 1$ has been used. It follows that

$$a(u, v) = \int_\Gamma \mathcal{D}t\mathcal{N}(u)\bar{v}_{|r=a}\mathrm{d}\hat{x} \tag{C.4}$$

with

$$a(u, v) = \int_a^\eta \int_\Gamma \left\{ \tilde{s}^{d-1}s\partial_r u\partial_r \bar{v} + \frac{a^2}{\tilde{s}^{3-d}sr^2}\nabla_\Gamma u\nabla_\Gamma \bar{v} - \frac{k^2}{\tilde{s}^{1-d}s}u\bar{v} \right\}\left(\frac{r}{a}\right)^{d-1}\mathrm{d}r\mathrm{d}\hat{x}.$$

In principle, the full spectrum of finite element techniques (hp-FEM, adaptivity, etc.) could be applied for discretizing this sesquilinearform. Here, we use a tensor-product discretization (with real-valued shaped functions)

$$\psi_\alpha(r, \hat{x}) = g_\mu(r)\phi_i(\hat{x}), \quad \alpha = (i, \mu) \in \{1, \dots, n_\Gamma\} \times \{0, \dots, N\}, \tag{C.5}$$

to fit the PML formulation into the framework (3.7). Later in this thesis, cf. Section 4.2.2, also adaptive discretization techniques will be considered which cannot be described in the tensor-product form (3.7). The choice (C.5) allows to write the system matrix as

$$\begin{aligned}
L_{\alpha\beta} &= a(\psi_\beta, \psi_\alpha) = a(\psi_\alpha, \psi_\beta) \\
&= \left(\int_a^\eta \tilde{s}^{d-1} s \partial_r g_\mu \partial_r g_\nu \left(\frac{r}{a}\right)^{d-1} \mathrm{d}r - k^2 \int_a^\eta \frac{1}{\tilde{s}^{1-d}s} g_\mu g_\nu \left(\frac{r}{a}\right)^{d-1} \mathrm{d}r \right) \int_\Gamma \phi_i \phi_j \, \mathrm{d}\hat{x} \\
&\quad + \left(\int_a^\eta \frac{a^2}{\tilde{s}^{3-d}s r^2} g_\mu g_\nu \left(\frac{r}{a}\right)^{d-1} \mathrm{d}r \right) \int_\Gamma \nabla_\Gamma \phi_i \nabla_\Gamma \phi_j \, \mathrm{d}\hat{x} \tag{C.6} \\
&= A(k, \sigma)_{\mu,\nu} M_{ij} + B(\sigma)_{\mu\nu} K_{ij},
\end{aligned}$$

with $A(k, \sigma)_{\mu,\nu} = A^{(1)}(\sigma)_{\mu,\nu} + k^2 A^{(2)}(\sigma)_{\mu,\nu}$. For wavenumber independent scaling profiles σ, the matrices are independent of k.

It follows that (C.4) transforms into the linear system

$$\begin{bmatrix} L_{\Gamma\Gamma} & L_{\Gamma E} \\ L_{E\Gamma} & L_{EE} \end{bmatrix} \begin{bmatrix} \underline{u}_\Gamma \\ \underline{u}_E \end{bmatrix} = \begin{bmatrix} M \, \mathrm{DtN}_N \, \underline{u}_\Gamma \\ 0 \end{bmatrix}, \tag{C.7}$$

with the left hand side being of the tensor-product form (3.7) and DtN_N representing the discrete approximation of $\mathcal{DtN}$. By eliminating the exterior degrees of freedom, we obtain that DtN_N is given by the Schur complement

$$\mathrm{DtN}_N \, u_\Gamma = M^{-1} \left(L_{\Gamma\Gamma} - L_{\Gamma E} L_{EE}^{-1} L_{E\Gamma} \right) u_\Gamma,$$

which is in agreement with formula (3.5).

C.2 Classical infinite elements

Whereas PMLs work with standard finite element shape functions and realize the radiation condition by means of a coordinate stretching, classical infinite elements incorporate the desired asymptotic behavior into the trial space by using specialized shape functions, see [Ast00] for a review. The variational formulation of (C.1) is usually first derived in a bounded exterior domain, e.g. in $a \leq r \leq R$ where $R > a$ is finite. A radiation condition $\partial_r u - iku = w$ is imposed at $r = R$.

The bilinear and linear form of the variational formulation are given by

$$a(u, v) = \lim_{R \to \infty} \left[\int_a^R \int_\Gamma \left\{ \partial_r u \partial_r v + \frac{a^2}{r^2} \nabla_\Gamma u \nabla_\Gamma v - k^2 uv \right\} \left(\frac{r}{a}\right)^{d-1} \mathrm{d}r \mathrm{d}\hat{x} - \int_{\Gamma_R} ikuv \, \mathrm{d}\Gamma_R \right], \tag{C.8}$$

$$f(v) = \int_\Gamma \mathcal{DtN}(u)v \, \mathrm{d}\hat{x} + \lim_{R \to \infty} \int_{\Gamma_R} wv \, \mathrm{d}\Gamma_R. \tag{C.9}$$

To compute the matrix of the linear system, the trial and test functions (C.5) are inserted and the limit $R \to \infty$ is taken. The system matrix can be written as before in the form (3.7), i.e. $L_{\alpha\beta} = a(\psi_\beta, \psi_\alpha) = A_{\mu,\nu} M_{ij} + B_{\mu\nu} K_{ij}$, with

$$A_{\mu\nu} = \lim_{R \to \infty} \left[\int_a^R \left(\partial_r g_\nu \partial_r g_\mu - k^2 g_\nu g_\mu \right) \left(\frac{r}{a} \right)^{d-1} \mathrm{d}r - i k g_\nu(R) g_\mu(R) \left(\frac{R}{a} \right)^{d-1} \right] \quad \text{(C.10)}$$

$$B_{\mu\nu} = \lim_{R \to \infty} \int_a^R \frac{a^2}{r^2} g_\nu g_\mu \left(\frac{r}{a} \right)^{d-1} \mathrm{d}r \quad \text{(C.11)}$$

for $\nu = 0, \ldots, N$ and $\mu = 0, \ldots, N$. To calculate the matrix entries, the shape functions have to be fixed. Different choices are possible and give rise to different infinite elements. Two of the most popular variants for $d = 3$ are presented here. In a comparative study of different formulations presented in [Ger98], these two approaches have emerged as the most suitable for exterior problems.

The trial functions in both formulations are given by

$$g_0(r) = \left(\frac{a}{r} \right) e^{ik(r-a)}, \quad g_\nu(r) = \left(\frac{a}{r} \right)^{\nu+1} e^{ik(r-a)} - g_0(r), \ \nu \geq 1,$$

to represent the correct asymptotic decay as inferred from the Atkinson-Wilcox expansion (C.18). Two possibilities of choosing the test functions are discussed separately.

C.2.1   Unconjugated Burnett formulation

In the unconjugated Burnett formulation the test function is chosen equal to the trial function, i.e.

$$g_0(r) = \left(\frac{a}{r} \right) e^{ik(r-a)}, \quad g_\mu(r) = \left(\frac{a}{r} \right)^{\mu+1} e^{ik(r-a)} - g_0(r), \ \mu \geq 1.$$

The matrix entries of $A_{\mu,\nu}$ and $B_{\mu\nu}$ can be described in terms of the integrals

$$I_j := \int_a^\infty \left(\frac{a}{r} \right)^j e^{2ik(r-a)} \frac{1}{a} \mathrm{d}r, \quad \text{(C.12)}$$

for $j = 1, 2, \ldots$. Special care is required to treat the case $\nu = \mu = 0$ since some of the integrals are undefined as $R \to \infty$. However, after calculating the integrals on $[a, R]$ the oscillatory components, which would diverge as $R \to \infty$, actually cancel each other out, and a finite value is obtained. Finally, one obtains

$$A_{\mu,\nu} = \begin{cases} J_{00} & (\mu = 0, \nu = 0) \\ J_{0\nu} - J_{00} & (\mu = 0, \nu \geq 1) \\ J_{\mu 0} - J_{00} & (\mu \geq 1, \nu = 0) \\ J_{\mu\nu} - J_{0\nu} - J_{\mu 0} + J_{00} & (\mu \geq 1, \nu \geq 1) \end{cases} \quad \text{(C.13)}$$

$$B_{\mu\nu} = a \begin{cases} I_2 & (\mu = 0, \nu = 0) \\ I_{\nu+2} - I_2 & (\mu = 0, \nu \geq 1) \\ I_{\mu+2} - I_2 & (\mu \geq 1, \nu = 0) \\ I_{\nu+\mu+2} - I_{\nu+2} - I_{\mu+2} + I_2 & (\mu \geq 1, \nu \geq 1) \end{cases} \quad \text{(C.14)}$$

with

$$J_{\mu\nu} = \begin{cases} \frac{1}{a} - ik & (\mu = 0, \nu = 0) \\ \frac{(\mu+1)(\nu+1)}{a} I_{\mu+\nu+2} - ik(\mu + \nu + 2)I_{\mu+\nu+1} - 2k^2 a I_{\mu+\nu} & (\mu \neq 0, \nu \neq 0). \end{cases}$$

The integrals I_j can be computed by using a recursion relation [Bur94]:

$$I_j = \frac{2ika}{j-1} I_{j-1} + \frac{1}{j-1}, \quad j \geq 2,$$
$$I_1 = e^{-2ika}\left[- \operatorname{Ci}(2ka) - i\operatorname{Si}(2ka) + i\frac{\pi}{2} \right],$$

where Si and Ci are the well-known sine and cosine integrals, see [AS64, Chapter 5].

C.2.2 Conjugated Astley-Leis formulation

The test function in the conjugated Astley-Leis formulation is chosen as

$$g_0(r) = \left(\frac{a}{r}\right)^3 e^{-ik(r-a)}, \quad g_\mu(r) = \left(\frac{a}{r}\right)^{\mu+3} e^{-ik(r-a)} - g_0(r), \ \mu \geq 1.$$

With this choice the oscillatory parts of the test and trial functions cancel each other out. Thanks to the weighting factor $1/r^2$, all integrals are well-defined. The matrix entries can be written in terms of elementary functions:

$$A_{\mu\nu} = \begin{cases} R_{00} & (\mu = 0, \nu = 0) \\ R_{0\nu} - R_{00} & (\mu = 0, \nu \geq 1) \\ R_{\mu 0} - R_{00} & (\mu \geq 1, \nu = 0) \\ R_{\mu\nu} - R_{0\nu} - R_{\mu 0} + R_{00} & (\mu \geq 1, \nu \geq 1) \end{cases} \tag{C.15}$$

$$B_{\mu\nu} = a^2 \begin{cases} Q_{00}^{l=2} & (\mu = 0, \nu = 0) \\ Q_{0\nu}^{l=2} - Q_{00}^{l=2} & (\mu = 0, \nu \geq 1) \\ Q_{\mu 0}^{l=2} - Q_{00}^{l=2} & (\mu \geq 1, \nu = 0) \\ Q_{\mu\nu}^{l=2} - Q_{0\nu}^{l=2} - Q_{\mu 0}^{l=2} + Q_{00}^{l=2} & (\mu \geq 1, \nu \geq 1) \end{cases} \tag{C.16}$$

for

$$Q_{\mu\nu}^l := \frac{a^{-l+1}}{\nu + \mu + l + 1}, \quad R_{\mu\nu} := -ik(\mu + 2 - \nu)Q_{\mu\nu}^{l=1} + (\nu + 1)(\mu + 3)Q_{\mu\nu}^{l=2}.$$

Since this ansatz is linear in ik, it can be extended to solve the wave equation in the time domain [Ast96, ACC98].

In both approaches the term on the right hand side in (C.9) containing the unknown function $w = \mathcal{O}(R^{-2})$ vanishes as $R \to \infty$, so that the familiar relation between DtN_N and Schur complement (3.5) is recovered at the discrete level.

C.3 Hardy Space Infinite Elements

One of the main difficulties in designing methods for wave equations posed in unbounded domains is the appropriate integration of a radiation condition which selects outgoing

waves. Classical infinite element methods proceed by matching the series expansion of outgoing waves. Hardy Space Infinite Elements (HSIEs) introduced by Hohage and Nannen [HN09] choose a different approach. They are based on the pole condition [SD95, HSZ03a, HSZ03b], which characterizes outgoing waves by means of the Laplace transform. The basic idea can be illustrated in the one-dimensional example

$$-u''(r) - k^2 u(r) = 0 \quad r \geq a.$$

The general solution is of the form

$$u(r) = C_+ e^{ikr} + C_- e^{-ikr},$$

for constants C_+ and C_-. Here $C_+ e^{ikr}$ corresponds to the outgoing and $C_- e^{-ikr}$ to the incoming solution. Taking the Laplace transform

$$(\mathscr{L}f)(s) := \int\limits_0^\infty e^{-sr} f(r) \, \mathrm{d}r, \quad \operatorname{Re}(s) > 0,$$

of this equation yields

$$(\mathscr{L}u)(s) = \frac{C_+}{s - ik} + \frac{C_-}{s + ik}, \quad \operatorname{Re}(s) > 0.$$

This means that u is outgoing if and only if $\mathscr{L}u$ has no pole in the lower complex half-plane. Actually, $\mathscr{L}u$ can be extended holomorphically to $\mathbb{C}\backslash\{+ik, -ik\}$ and u is outgoing if and only if $\mathscr{L}u$ belongs to the Hardy space $H^-(\mathbb{R})$ consisting of those square integrable functions which are boundary values of holomorphic functions in the lower half-plane.

The transformed equation lends itself to a Galerkin discretization by using finite dimensional subspaces of $H^-(\mathbb{R})$. In practice, an additional transformation to the unit disk is first applied because the corresponding space on the disk admits a convenient orthonormal basis. The HSIE method is thus a Galerkin method, albeit using a rather exotic variational formulation and unfamiliar spaces.

While using the HSIE for problem (C.1), the Laplace transform is only applied in the radial coordinate. To simplify the derivation of the variational formulation in the exterior domain, some rescaling is helpful. If the solution is sought in the form

$$\tilde{u}(\tilde{r}, \hat{x}) = (\tilde{r} + 1)^{(d-1)/2} u((\tilde{r} + 1)\hat{x}),$$

for the radial coordinate $\tilde{r} := |x|/a$ and $\hat{x} \in \Gamma$, the PDE transforms to

$$-\frac{1}{a}\partial_{\tilde{r}}^2 \tilde{u} - \left(ak^2 + \frac{C_d}{a} + \frac{a}{(\tilde{r} + 1)^2}\Delta_\Gamma\right)\tilde{u} = 0 \tag{C.17}$$

for $C_d = (d-1)(3-d)/4$.

It holds that:

$$(\partial_{\tilde{r}}\tilde{u})_{|\tilde{r}=0} = \left(\frac{d-1}{2}\right)u_{|\Gamma} - a\mathcal{D}t\mathcal{N}(u).$$

Therefore, multiplying (C.17) by a suitable test function $\tilde{v}$ and integrating by parts leads (cf. [HN09]) to:

$$\int_\Gamma \mathcal{D}t\mathcal{N}(u)v\mathrm{d}\hat{x} = \frac{(d-1)}{2a}\int_\Gamma uv\ \mathrm{d}\hat{x} + \int_0^\infty \int_\Gamma \frac{1}{a}\partial_{\tilde{r}}\tilde{u}\partial_{\tilde{r}}\tilde{v}\ \mathrm{d}r\mathrm{d}\hat{x}$$

$$+ a\int_0^\infty \int_\Gamma \left(\frac{\nabla_\Gamma\tilde{u}\nabla_\Gamma\tilde{v}}{(\tilde{r}+1)^2} - k^2\tilde{u}\tilde{v} - \frac{C_d}{a^2}\frac{\tilde{u}\tilde{v}}{(\tilde{r}+1)^2}\right)\ \mathrm{d}r\mathrm{d}\hat{x}.$$

The bilinear form on the right hand side is then transformed to the corresponding Hardy space on the unit disk and discretized in the radial coordinate with respect to a basis thereof. In the tangential coordinate a standard finite element discretization on Γ is employed. Such a tensor-product discretization then again leads to a linear system of the form (3.7). According to reference [HN09], the radial matrices are given by

$$A = A^{(1)} - k^2 A^{(2)},$$

$$A^{(1)} = \frac{d-1}{2a}\begin{pmatrix} 1 & \\ & \mathbf{0} \end{pmatrix} - \frac{2i\kappa_0}{a}\mathcal{T}_{N,+}^T\mathcal{T}_{N,+} - \frac{2C_d i}{\kappa_0 a}\mathcal{T}_{N,-}^T D_N^{-2}\mathcal{T}_{N,-},$$

$$A^{(2)} = \frac{2ai}{\kappa_0}\mathcal{T}_{N,-}^T\mathcal{T}_{N,-}, \quad B = \frac{2ai}{\kappa_0}\mathcal{T}_{N,-}^T D_N^{-2}\mathcal{T}_{N,-},$$

with the tridiagonal matrices

$$D_N = \mathrm{id}_{(N+1)\times(N+1)} + \frac{1}{2i\kappa_0}\begin{pmatrix} -1 & 1 & & & & \\ 1 & -3 & 2 & & & \\ & 2 & -5 & 3 & & \\ & & & \ddots & \ddots & \ddots \\ & & & & N & -2N-1 \end{pmatrix}$$

and

$$\mathcal{T}_{N,\pm} = \frac{1}{2}\begin{pmatrix} 1 & \pm 1 & & & \\ & 1 & \pm 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & \pm 1 \end{pmatrix}$$

Here, κ_0 represents a tuning parameter. For scattering problems it is usually chosen proportional to the wavenumber.

C.4 Local absorbing boundary conditions

The previously discussed techniques involve some artificial extension of the computational domain beyond the truncation boundary. Local absorbing boundary conditions, in contrast, are implementable directly on the coupling boundary. To this end, an approximate relation between the $\mathcal{D}t\mathcal{N}$ map and the solution on the coupling boundary, which may include derivatives, is employed. Such relations are often derived from exact or asymptotic expansions of the solution. As a popular representative of the class of local absorbing boundary conditions the first and second order conditions from Bayliss, Gunzburger and Turkel [BGT82] will be discussed here.

The derivation of these conditions in the exterior of a sphere can be sketched as follows (see [BGT82] for details) . The solution of the Helmholtz equation fulfilling the Sommerfeld radiation condition admits the Atkinson-Wilcox expansion

$$u(r, \hat{x}) = \frac{e^{ikr}}{kr} \sum_{j=0}^{\infty} \frac{F_j(\hat{x})}{(kr)^j}, \tag{C.18}$$

with smooth functions F_j. The series and its term by term derivatives converge absolutely and uniformly in the exterior of the sphere [Wil56]. The idea is to derive a differential relation on the coupling boundary which fits the computed solution to the first m terms of (C.18). The sequence of operators defined by

$$B_1 u = \left(\frac{\partial}{\partial r} - ik + \frac{1}{r} \right) u, \quad B_m u = \left(\frac{\partial}{\partial r} - ik + \frac{2m-1}{r} \right) B_{m-1} u \tag{C.19}$$

annihilates the first m terms of the expansion (C.18). This implies that

$$(B_m u) \!\restriction_{r=a} = \mathcal{O} \left(\frac{1}{a^{2m+1}} \right). \tag{C.20}$$

For computational efficiency it would be desirable to choose m large. Otherwise, the coupling boundary has to be placed very far away from the scatterer to obtain an accurate solution which requires a large computational domain. At first sight, however, only the lowest order condition $B_1 u = 0$, which can be written as $\partial_r u = (ik - 1/r)u$, seems straightforward to implement as the other conditions involve higher order derivates in r. Nevertheless, the second order condition can be written as

$$\mathscr{D}t\mathscr{N}(u) = - \left(\frac{\partial u}{\partial r} \right) \!\restriction_{r=a} = A(a, k)u - B(a, k)\Delta_\Gamma u \tag{C.21}$$

by using the Helmholtz equation in spherical coordinates to replace the second derivatives with respect to r by tangential derivates. The functions $A(a, k), B(a, k)$ are given in Table C.1. The conditions for $m > 2$ are more complicated to realize since stan-

	$d = 2$		$d = 3$	
	$m = 1$	$m = 2$	$m = 1$	$m = 2$
$A(a, \omega)$	$\alpha - \frac{1}{2a}$	$\left(\frac{3}{8a^2} - \frac{3ik}{2a} - k^2 \right)/\alpha$	α	α
$B(a, \omega)$	0	$\frac{a^2}{2a^2\alpha}$	0	$\frac{a^2}{2a^2\alpha}$

Table C.1:   Coefficients of the absorbing boundary conditions of [BGT82] in formula (C.21) for dimension d and order m. Here $\alpha(k, r) = -ik + 1/r$.

dard H^1-conforming finite elements lack the required smoothness on Γ to allow for an implementation of the higher-order tangential derivatives. Although higher-order local absorbing boundary conditions have been realized based on special finite elements providing more regularity on the coupling boundary [GPK97] or introducing auxiliary variables [Giv01], they seem to be rarely used in practice.

The conditions of [BGT82] for the two-dimensional case are derived analogously except that the exact expansion (C.18) is replaced by an asymptotic expansion in polar

coordinates. They can also be written in the form (C.21) with coefficients provided in Table C.1.

In the finite element method equation (C.21) is implemented as

$$\int_\Gamma \mathcal{D}t\mathcal{N}(u)v_{\restriction r=a}\, \mathrm{d}\hat{x} = \int_\Gamma A(a,\omega)uv + B(a,k)a^2\nabla_\Gamma u\nabla_\Gamma v\, \mathrm{d}\hat{x}.$$

Hence, local aborbing boundary conditions also lead to a linear system of the form (C.7) with

$$L_{\Gamma\Gamma} = A(a,k)M + B(a,k)a^2K, \quad L_{\Gamma E} = L_{E\Gamma} = L_{EE} = 0.$$

Appendix D

Solving the minimization problem for learned IEs

This appendix provides some details on the solution of the optimization problem for learned IEs. Both the minimization problem for fixed wavenumber (4.2)-(4.3) and the one including several wavenumbers (5.5)-(5.6) are of non-linear least squares type. The Levenberg-Marquardt algorithm is one of the most popular approaches for solving such problems. It is described in detail in standard texts on numerical optimization, see e.g. [NW06, Kel99]. Various open source implementations are readily available of which we use the `ceres-solver` [AMO]. These implementations usually require from the user the implementation of the cost function and its Jacobian, which will be calculated in this appendix. For the reduced symmetric ansatz involving a fixed wavenumber this is relatively simple and will be covered in Appendix D.1. Such a description has already been provided in [HLP21] for the slightly different ansatz we used in this publication. The general case of several wavenumbers presented in Appendix D.2 is somewhat more invovled, yet still manageable.

D.1 Reduced symmetric ansatz for single wavenumbers

Appendix D.1.1 describes the objective function and its Jacobian for $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ with N being fixed. In practice, one usually solves the optimization problem successively for increasing $N = 0, 1, \ldots$. The result from iteration N should then be used to construct a good initial guess for iteration $N + 1$ as described in subsection Appendix D.1.1.

D.1.1 Solving the optimization problem for a fixed N

The objective function is given by

$$J(A, B) = \frac{1}{2} \sum_{\ell} \omega_\ell^2 |f_\ell|^2, \tag{D.1}$$

with $f_\ell(A, B) := \mathit{dtn}(\lambda_\ell) - \mathrm{dtn}_N(\lambda_\ell)$ and $\mathrm{dtn}_N(\lambda_\ell)$ as defined in (4.14). The cost functions f_ℓ and their gradients depend on the parameters

$$\{A_{0j}\}_{j=0}^N \cup \{A_{jj}\}_{j=1}^N \cup \{B_{0j}\}_{j=0}^N.$$

201

A potential difficulty could arise from the fact that A and B are complex matrices while the available implementation of the Levenberg-Marquardt algorithm might be limited to real parameters. However, this is easily resolved by splitting into real and imaginary parts

$$A_{nm} = \operatorname{Re} A_{nm} + i \operatorname{Im} A_{nm}$$

and treating $\operatorname{Re} A_{nm}$ and $\operatorname{Im} A_{nm}$ as two separate real parameters.

The gradients of the cost functions can easily be calculated analytically:

$$\frac{\partial f_\ell}{\partial \operatorname{Re} A_{00}} = -1, \quad \frac{\partial f_\ell}{\partial \operatorname{Re} B_{00}} = -\lambda_\ell.$$

For $n \geq 1$:

$$\frac{\partial f_\ell}{\partial \operatorname{Re} A_{0n}} = 2\frac{(A_{0n} + \lambda_\ell B_{0n})}{(A_{nn} + \lambda_\ell)}, \quad \frac{\partial f_\ell}{\partial \operatorname{Re} B_{0n}} = 2\lambda_\ell \frac{(A_{0n} + \lambda_\ell B_{0n})}{(A_{nn} + \lambda_\ell)},$$

and

$$\frac{\partial f_\ell}{\partial \operatorname{Re} A_{nn}} = -\frac{(A_{0n} + \lambda_\ell B_{0n})^2}{(A_{nn} + \lambda_\ell)^2}.$$

The derivatives with respect to the imaginary part can be obtained from

$$\frac{\partial f_\ell}{\partial \operatorname{Im} A_{nm}} = i\frac{\partial f_\ell}{\partial \operatorname{Re} A_{nm}}, \quad \frac{\partial f_\ell}{\partial \operatorname{Im} B_{nm}} = i\frac{\partial f_\ell}{\partial \operatorname{Re} B_{nm}}.$$

Please note that the cost functions are complex-valued, so depending on the implementation they may have to be split into real and imaginary parts as well.

Successive learning

A description of the whole pipeline starting with the computation of $dtn(\lambda_\ell)$ is given in Algorithm 3.

Algorithm 3: Successive learning of matrices A and B

1: Define $N_{\max}$, $L_{\max}$
2: **for** $\ell = 0 : L_{\max}$ **do**
3:     Compute $dtn(\lambda_\ell)$ by solving (2.12a)-(2.12b).
4: **end for**
5: Initialize $\tilde{A} \in \mathbb{C}, \tilde{B} \in \mathbb{C}$ randomly.
6: **for** $N = 0 : N_{\max}$ **do**
7:     Obtain $A, B \in \mathbb{C}^{(N+1)\times(N+1)}$ by minimizing (D.1) with Levenberg-Marquardt using $\tilde{A}, \tilde{B}$ as initial guess. The gradients for the cost function have been described in Appendix D.1.1.
8:     Prepare new initial guess: $\tilde{A}, \tilde{B} \in \mathbb{C}^{(N+2)\times(N+2)}$.
9:     Initialize upper block of $\tilde{A}, \tilde{B}$ by A, B.
10:     Set $\tilde{A}_{N+1,N+1}$ to one or a good guess for the poles if available.
11:     Fill $\tilde{A}_{0,N+1}$ and $\tilde{B}_{0,N+1}$ with small (nonzero) random numbers.
12: **end for**
13: **return** Learned matrices A, B for $N = 0 : N_{\max}$.

A few additional remarks are given below.

- Often there is an analytic reference solution for $dtn(\lambda_\ell)$ available so that the solution of the ODEs can be skipped.

- The purpose of the loop over N in Algorithm 3 is to provide a good initial guess for the optimization. Basically, the learned matrices from step N are reused as an initial guess for step $N+1$. With this technique the learned $\mathrm{dtn}_{N+1}(\lambda_\ell)$-function in step $N+1$ starts at the value of the minimizer of step N plus the additional term

$$\frac{(\tilde{A}_{0,N+1} + \lambda_\ell \tilde{B}_{0,N+1})(\tilde{A}_{0,N+1} + \lambda_\ell \tilde{B}_{0,N+1})}{\tilde{A}_{N+1,N+1} + \lambda_\ell}.$$

In order to prevent the cost function from starting too far away from the minimizer of the previous step, this contribution should be small. This is the motivation for filling $\tilde{A}_{0,N+1}$ and $\tilde{B}_{0,N+1}$ with small random numbers. These should be nonzero to avoid getting stuck at the minimizer of step N. If a good guess for the poles is available then setting $-\tilde{A}_{N+1,N+1}$ equal to one of these poles can be beneficial.

D.2 Full ansatz for uniform approximation in wavenumber

Consider misfit functions of the form

$$J(A^{(1)}, A^{(2)}, B) = \frac{1}{2} \sum_\ell \sum_k \omega_{\ell,k} |f_{\ell,k}|^2,$$

with

$$\begin{aligned}
f_{\ell,k} &= dtn(\lambda_\ell, k) - S_{\Gamma\Gamma}(\ell, k) + S_{\Gamma E}(\ell, k) S_{EE}(\ell, k) S_{E\Gamma}(\ell, k) \\
&= dtn(\lambda_\ell, k) - S_{\Gamma\Gamma} + \sum_{j'=1}^{N} \sum_{j=1}^{N} S_{0j'} \left[S_{EE}^{-1} \right]_{j'-1,j-1} S_{j0}
\end{aligned}$$

for

$$S(\ell, k) = A^{(1)} + g(k) A^{(2)} + \lambda_\ell B.$$

Such objective functions arise in the minimization problem (5.5)-(5.6). The problem (4.2)-(4.3) for a fixed wavenumber is covered as a special case by setting $g(k) = 0$ and $A^{(1)} = A$. The derivatives of $f_{\ell,k}$ with repect to the matrix entries of $\mathrm{Re}\, A^{(1)}$, $\mathrm{Re}\, A^{(2)}$, $\mathrm{Re}\, B$, $\mathrm{Im}\, A^{(1)}$, $\mathrm{Im}\, A^{(2)}$ and $\mathrm{Im}\, B$ have to be computed.

Let us start with some preliminary observations to simplify differentiation of the inverse S_{EE}^{-1}.

- The derivative of a matrix Y with respect to a parameter α is given by

$$\frac{\partial Y^{-1}}{\partial \alpha} = -Y^{-1} \frac{\partial Y}{\partial \alpha} Y^{-1}. \tag{D.2}$$

- An application of (D.2) yields:

$$\frac{\partial S_{EE}^{-1}}{\partial \mathrm{Re}\, A_{nm}^{(1)}} = -S_{EE}^{-1} \frac{\partial S_{EE}^{-1}}{\partial \mathrm{Re}\, A_{nm}^{(1)}} S_{EE}^{-1}.$$

For $n = 0$ or $m = 0$ the derivative vanishes since in this case S_{EE} is independent of $\operatorname{Re} A_{nm}^{(1)}$. For $n \geq 1$ and $m \geq 1$ let E_{nm} denote the $N \times N$ matrix with 1 at position (n, m) and zero elsewhere and

$$T^{nm} := -S_{EE}^{-1} E_{n-1,m-1} S_{EE}^{-1}. \tag{D.3}$$

Then

$$\frac{\partial S_{EE}^{-1}}{\partial \operatorname{Re} A_{nm}^{(1)}} = T^{nm}, \qquad \frac{\partial S_{EE}^{-1}}{\partial \operatorname{Re} A_{nm}^{(2)}} = g(k) T^{nm}, \qquad \frac{\partial S_{EE}^{-1}}{\partial \operatorname{Re} B_{nm}} = \lambda_\ell T^{nm},$$

$$\frac{\partial S_{EE}^{-1}}{\partial \operatorname{Im} A_{nm}^{(1)}} = i T^{nm}, \qquad \frac{\partial S_{EE}^{-1}}{\partial \operatorname{Im} A_{nm}^{(2)}} = i g(k) T^{nm}, \qquad \frac{\partial S_{EE}^{-1}}{\partial \operatorname{Im} B_{nm}} = i \lambda_\ell T^{nm}.$$

The following cases of (n, m) have to be destinguished:

- $n = m = 0$:
$$\frac{\partial f_{\ell,k}}{\partial \operatorname{Re} A_{00}^{(1)}} = -1.$$

- For $n \geq 1$ and $m = 0$:
$$\frac{\partial f_{\ell,k}}{\partial \operatorname{Re} A_{n0}^{(1)}} = \sum_{j'=1}^{N} S_{0j'}(\ell, k) \left[S_{EE}(\ell, k)^{-1} \right]_{j'-1, n-1}.$$

- For $n = 0$ and $m \geq 1$:
$$\frac{\partial f_{\ell,k}}{\partial \operatorname{Re} A_{0m}^{(1)}} = \sum_{j=1}^{N} S_{j0}(\ell, k) \left[S_{EE}(\ell, k)^{-1} \right]_{m-1, j-1}.$$

- For $n \geq 1, m \geq 1$:
$$\frac{\partial f_{\ell,k}}{\partial \operatorname{Re} A_{nm}^{(1)}} = \sum_{j'=1}^{N} \sum_{j=1}^{N} S_{0j'}(\ell, k) T_{j'-1, j-1}^{nm}(\ell, k) S_{j0}(\ell, k).$$

The derivatives with respect to $\operatorname{Re} A^{(2)}, \operatorname{Re} B, \operatorname{Im} A^{(1)}, \operatorname{Im} A^{(2)}$ and $\operatorname{Im} B$ are obtained by multiplying the derivate with respect to $\operatorname{Re} A^{(1)}$ by $g(k), \lambda_\ell, i, ig(k)$ and $i\lambda_\ell$, respectively.

D.2.1 Symmetric ansatz for uniform approximation in wavenumber

For a symmetric ansatz only the entries of $A^{(1)} A^{(2)}$ and B below the diagonal have to be considered as free variables. For the derivative the following modifications of the full ansatz described above have to be made:

- For $n = 0$ and $m \geq 1$:
$$\frac{\partial f_{\ell,k}}{\partial \operatorname{Re} A_{0m}^{(1)}} = \sum_{j'=1}^{N} S_{0j'}(\ell, k) \left[S_{EE}(\ell, k)^{-1} \right]_{j'-1, m-1} + \sum_{j=1}^{N} S_{j0}(\ell, k) \left[S_{EE}(\ell, k)^{-1} \right]_{m-1, j-1}.$$

- For $n = 1$ and $m \geq 1$ with $n \neq m$ the definition of T^{nm} in (D.3) has to be modified to
$$T^{nm} := -S_{EE}^{-1} E_{n-1,m-1} S_{EE}^{-1} - S_{EE}^{-1} E_{m-1,n-1} S_{EE}^{-1}.$$

Appendix E

Derivation of dtn^{jump}

This appendix provides auxiliary computations for the case of a jumping wavenumber in the exterior domain introduced in Section 2.2.1 and featured in the numerical experiments of Section 4.2.3. These derivations are also included in the supplementary material of our publication [HLP21].

For $\nu, a > 0$ consider the problem

$$-\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \left(k^2(r) - \frac{\nu^2}{r^2}\right)u = 0 \quad r > a,$$
$$u(a) = 1,$$

with radiation condition at infinity. For some $a \le R_{\mathrm{J}} < \infty$ let the wavenumber be given by

$$k(r) = \begin{cases} k_I & r < R_{\mathrm{J}}, \\ k_\infty & r > R_{\mathrm{J}}, \end{cases}$$

for some $k_I, k_\infty > 0$.

The solution is given by

$$u_\nu(r) = \begin{cases} A_\nu^{\mathrm{J}} J_\nu(k_I r) + B_\nu^{\mathrm{J}} Y_\nu(k_I r) & r < R_{\mathrm{J}}, \\ C_\nu^{\mathrm{J}} H_\nu^{(1)}(k_\infty r) & r > R_{\mathrm{J}}, \end{cases}$$

with constants $A_\nu^{\mathrm{J}}, B_\nu^{\mathrm{J}}, C_\nu^{\mathrm{J}} \in \mathbb{C}$ to be determined by the following three constraints:

- Boundary condition at $r = a$:

$$A_\nu^{\mathrm{J}} J_\nu(k_I a) + B_\nu^{\mathrm{J}} Y_\nu(k_I a) = 1. \tag{E.1}$$

- Continuity at $r = R_{\mathrm{J}}$:

$$A_\nu^{\mathrm{J}} J_\nu(k_I R_{\mathrm{J}}) + B_\nu^{\mathrm{J}} Y_\nu(k_I R_{\mathrm{J}}) - C_\nu^{\mathrm{J}} H_\nu^{(1)}(k_\infty R_{\mathrm{J}}) = 0. \tag{E.2}$$

- Continuity of derivative at $r = R_{\mathrm{J}}$:

$$A_\nu^{\mathrm{J}} k_I J_\nu'(k_I R_{\mathrm{J}}) + B_\nu^{\mathrm{J}} k_I Y_\nu'(k_I R_{\mathrm{J}}) - C_\nu^{\mathrm{J}} k_\infty (H_\nu^{(1)})'(k_\infty R_{\mathrm{J}}) = 0. \tag{E.3}$$

Solving (E.1) for A_ν^{J} gives

$$A_\nu^{\mathrm{J}} = \frac{1 - B_\nu^{\mathrm{J}} Y_\nu(k_I a)}{J_\nu(k_I a)}. \tag{E.4}$$

Using this equation to eliminate A_ν^{J} from (E.2) and (E.3) leads to the linear system

$$\begin{bmatrix} \frac{J_\nu(k_I a) Y_\nu(k_I R_{\mathrm{J}}) - Y_\nu(k_I a) J_\nu(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} & -H_\nu^{(1)}(k_\infty R_{\mathrm{J}}) \\ \frac{Y_\nu'(k_I R_{\mathrm{J}}) J_\nu(k_I a) - Y_\nu(k_I a) J_\nu'(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} & -\frac{k_\infty}{k_I}(H_\nu^{(1)})'(k_\infty R_{\mathrm{J}}) \end{bmatrix} \begin{bmatrix} B_\nu^{\mathrm{J}} \\ C_\nu^{\mathrm{J}} \end{bmatrix} = \begin{bmatrix} -\frac{J_\nu(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} \\ -\frac{J_\nu'(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} \end{bmatrix}.$$

Denote the matrix in this equation as M_ν^{J}. We have

$$\det(M_\nu^{\mathrm{J}}) = -\frac{k_\infty}{k_I} \frac{(H_\nu^{(1)})'(k_\infty R_{\mathrm{J}})}{J_\nu(k_I a)} \left[J_\nu(k_I a) Y_\nu(k_I R_{\mathrm{J}}) - Y_\nu(k_I a) J_\nu(k_I , R_{\mathrm{J}}) \right]$$

$$+ \frac{H_\nu^{(1)}(k_\infty R_{\mathrm{J}})}{J_\nu(k_I a)} \left[Y_\nu'(k_I R_{\mathrm{J}}) J_\nu(k_I a) - Y_\nu(k_I a) J_\nu'(k_I R_{\mathrm{J}}) \right].$$

Hence,

$$(M^{\mathrm{J}})_\nu^{-1} = \frac{1}{\det(M_\nu^{\mathrm{J}})} \begin{bmatrix} -\frac{k_\infty}{k_I}(H_\nu^{(1)})'(k_\infty R_{\mathrm{J}}) & H_\nu^{(1)}(k_\infty R_{\mathrm{J}}) \\ \frac{-[Y_\nu'(k_I R_{\mathrm{J}}) J_\nu(k_I a) - Y_\nu(k_I a) J_\nu'(k_I R_{\mathrm{J}})]}{J_\nu(k_I a)} & \frac{J_\nu(k_I a) Y_\nu(k_I R_{\mathrm{J}}) - Y_\nu(k_I a) J_\nu(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} \end{bmatrix}.$$

The solution of the linear system is given by

$$B_\nu^{\mathrm{J}} = \frac{1}{\det(M_\nu)} \left[\frac{k_\infty}{k_I}(H_\nu^{(1)})'(k_\infty R_{\mathrm{J}}) \frac{J_\nu(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} - H_\nu^{(1)}(k_\infty R_{\mathrm{J}}) \frac{J_\nu'(k_I R_{\mathrm{J}})}{J_\nu(k_I a)} \right] \tag{E.5}$$

and

$$C_\nu^{\mathrm{J}} = \frac{1}{\det(M_\nu)} \Bigg[[Y_\nu'(k_I R_{\mathrm{J}}) J_\nu(k_I a) - Y_\nu(k_I a) J_\nu'(k_I R_{\mathrm{J}})] \frac{J_\nu(k_I R_{\mathrm{J}})}{J_\nu(k_I a)^2} \tag{E.6}$$

$$- [J_\nu(k_I a) Y_\nu(k_I R_{\mathrm{J}}) - Y_\nu(k_I a) J_\nu(k_I R_{\mathrm{J}})] \frac{J_\nu'(k_I R_{\mathrm{J}})}{J_\nu(k_I a)^2} \Bigg].$$

This yields the *dtn* function

$$\zeta(\nu) = -\frac{\partial u_\nu(a)}{\partial r} = -k_I \left[A_\nu^{\mathrm{J}} J_\nu'(k_I a) + B_\nu^{\mathrm{J}} Y_\nu'(k_I a) \right]. \tag{E.7}$$

A reference solution for a sound-soft scattering problem in the exterior of a disk with radius R_{s} can easily be obtained from the previous computations. Firstly, the radius of the truncation boundary has to be replaced by the radius of the scatterer, i.e. set $a = R_{\mathrm{s}}$. Assume that a plane wave $g = -\exp(ik_\infty x)$ is incident on the disk. It is well-known that g admits an expansion into the orthonormal functions

$$v_0 = \frac{1}{\sqrt{2\pi R_{\mathrm{s}}}}, \qquad v_\ell = \frac{1}{\sqrt{\pi R_{\mathrm{s}}}} \cos(\ell\varphi), \quad \text{for } \ell \geq 1$$

given by

$$g = -J_0(k_\infty r)\sqrt{2\pi R_{\mathrm{s}}} v_0 - \sum_{\ell=1} 2i^\ell J_\ell(k_\infty r)\sqrt{\pi R_{\mathrm{s}}} v_\ell.$$

Hence, by separation of variables the solution in polar coodinates is

$$u(r,\varphi) = u_0(r) J_0(k_\infty R_{\mathrm{s}})\sqrt{2\pi R_{\mathrm{s}}} v_0(\varphi) + \sum_{\ell=1}^{\infty} u_\ell(r) 2i^\ell J_\ell(k_\infty R_{\mathrm{s}})\sqrt{\pi R_{\mathrm{s}}} v_\ell(\varphi),$$

or simply

$$u(r,\varphi) = u_0(r) J_0(k_\infty R_{\mathrm{s}}) + \sum_{\ell=1}^{\infty} u_\ell(r) 2i^\ell J_\ell(k_\infty R_{\mathrm{s}}) \cos(\ell\varphi). \tag{E.8}$$

Case of no jump: In the special case

$$k_\infty = k_I := k \tag{E.9}$$

there is no jump and the following computation shows that the dtn^{hom} function of the homogeneous medium is recovered from (E.7). A straight-forward calculation using (E.9) yields:

$$
\begin{aligned}
\det(M_\nu^{\mathrm{J}}) ={}& -\frac{(H_\nu^{(1)})'(kR_{\mathrm{J}})}{J_\nu(ka)}\left[J_\nu(ka)Y_\nu(kR_{\mathrm{J}}) - Y_\nu(ka)J_\nu(kR_{\mathrm{J}})\right] \\
&+ \frac{H_\nu^{(1)}(kR_{\mathrm{J}})}{J_\nu(ka)}\left[Y_\nu'(kR_{\mathrm{J}})J_\nu(ka) - Y_\nu(ka)J_\nu'(kR_{\mathrm{J}})\right] \\
={}& \frac{1}{J_\nu(ka)}\Big[-J_\nu'(kR_{\mathrm{J}})J_\nu(ka)Y_\nu(kR_{\mathrm{J}}) - iY_\nu'(kR_{\mathrm{J}})J_\nu(ka)Y_\nu(kR_{\mathrm{J}}) \\
&+ J_\nu'(kR_{\mathrm{J}})Y_\nu(ka)J_\nu(kR_{\mathrm{J}}) + iY_\nu'(kR_{\mathrm{J}})Y_\nu(ka)J_\nu(kR_{\mathrm{J}}) \\
&+ J_\nu(kR_{\mathrm{J}})Y_\nu'(kR_{\mathrm{J}})J_\nu(ka) + iY_\nu(kR_{\mathrm{J}})Y_\nu'(kR_{\mathrm{J}})J_\nu(ka) \\
&- J_\nu(kR_{\mathrm{J}})Y_\nu(ka)J_\nu'(kR_{\mathrm{J}}) - iY_\nu(kR_{\mathrm{J}})Y_\nu(ka)J_\nu'(kR_{\mathrm{J}})\Big]
\end{aligned}
$$

Since four of the terms cancel out, one obtains

$$
\begin{aligned}
\det(M_\nu^{\mathrm{J}}) &= J_\nu(kR_{\mathrm{J}})Y_\nu'(kR_{\mathrm{J}}) - J_\nu'(kR_{\mathrm{J}})Y_\nu(kR_{\mathrm{J}}) + i\frac{Y_\nu(ka)}{J_\nu(ka)}\left[J_\nu(kR_{\mathrm{J}})Y_\nu'(kR_{\mathrm{J}}) - J_\nu'(kR_{\mathrm{J}})Y_\nu(kR_{\mathrm{J}})\right] \\
&= \left(1 + i\frac{Y_\nu(ka)}{J_\nu(ka)}\right)W\{J_\nu(kR_{\mathrm{J}}), Y_\nu(kR_{\mathrm{J}})\} \\
&= \frac{H_\nu^{(1)}(ka)}{J_\nu(ka)}W\{J_\nu(kR_{\mathrm{J}}), Y_\nu(kR_{\mathrm{J}})\},
\end{aligned}
$$

where $W\{J_\nu(kR_{\mathrm{J}}), Y_\nu(kR_{\mathrm{J}})\} = J_\nu(kR_{\mathrm{J}})Y_\nu'(kR_{\mathrm{J}}) - J_\nu'(kR_{\mathrm{J}})Y_\nu(kR_{\mathrm{J}})$ is the Wronskian. A similar calculation yields that

$$
\begin{aligned}
B_\nu^{\mathrm{J}} &= \frac{1}{\det(M_\nu)}\left[(H_\nu^{(1)})'(kR_{\mathrm{J}})\frac{J_\nu(kR_{\mathrm{J}})}{J_\nu(ka)} - H_\nu^{(1)}(kR_{\mathrm{J}})\frac{J_\nu'(kR_{\mathrm{J}})}{J_\nu(ka)}\right] \\
&= \frac{1}{\det(M_\nu)}\frac{i}{J_\nu(ka)}W\{J_\nu(kR_{\mathrm{J}}), Y_\nu(kR_{\mathrm{J}})\} = \frac{i}{H_\nu^{(1)}(ka)}.
\end{aligned}
$$

As

$$1 - B_\nu^{\mathrm{J}}Y_\nu(k_I a) = \frac{J_\nu(ka)}{H_\nu^{(1)}(ka)},$$

it follows from (E.4) that $A_\nu^{\mathrm{J}} = 1/H_\nu^{(1)}(ka)$. Inserting this into the ansatz for the solution yields $u_\nu(r) = H_\nu^{(1)}(kr)/H_\nu^{(1)}(ka)$ for $r < R_{\mathrm{J}}$ and hence $dtn^{\mathrm{jump}}(\lambda) = dtn^{\mathrm{hom},2d}(\lambda)$. The reference solution (E.8) then takes the explicit form

$$u(r,\varphi) = \frac{H_0^{(1)}(k_I r)}{H_0^{(1)}(k_I R_{\mathrm{s}})}J_0(k_I R_{\mathrm{s}}) + \sum_{\ell=1}^\infty \frac{H_\ell^{(1)}(k_I r)}{H_\ell^{(1)}(k_I R_{\mathrm{s}})}2i^\ell J_\ell(k_I R_{\mathrm{s}})\cos(\ell\varphi).$$

Bibliography

[AAB+15]   P. Amestoy, C. Ashcraft, O. Boiteau, A. Buttari, J.-Y. L'Excellent, and C. Weisbecker. Improving Multifrontal Methods by Means of Block Low-Rank Representations. *SIAM J. Sci. Comput.*, 37(3):A1451–A1474, 2015. doi:10.1137/120903476.

[AC01]   R. J. Astley and J.-P. Coyette. Conditioning of infinite element schemes for wave problems. *Commun. numer. methods eng.*, 17(1):31–41, 2001. doi:10.1002/1099-0887(200101)17:1<31::AID-CNM386>3.0.CO;2-A.

[ACC98]   R. J. Astley, J.-P. Coyette, and L. Cremers. Three-dimensional wave-envelope elements of variable order for acoustic radiation and scattering. Part II. Formulation in the time domain. *J. Acoust. Soc. Am.*, 103:64–72, 01 1998. doi:10.1121/1.421107.

[ADLK01]   P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster. A Fully Asynchronous Multifrontal Solver Using Distributed Dynamic Scheduling. *SIAM J. Matrix Anal. Appl.*, 23(1):15–41, January 2001. doi:10.1137/S0895479899358194.

[AGR20]   D. Appelö, F. Garcia, and O. Runborg. WaveHoltz: Iterative Solution of the Helmholtz Equation via the Wave Equation. *SIAM J. Sci. Comput.*, 42(4):A1950–A1983, 2020. doi:10.1137/19M1299062.

[AH04]   P. Arbenz and M. E. Hochstenbach. A Jacobi–Davidson method for Solving Complex Symmetric Eigenvalue Problems. *SIAM J. Sci. Comput.*, 25(5):1655–1673, 2004. doi:10.1137/S1064827502410992.

[AHN18]   A. D. Agaltsov, T. Hohage, and R. G. Novikov. Monochromatic identities for the Green function and Uniqueness Results for Passive Imaging. *SIAM J. Appl. Math.*, 78:2865–2890, 2018. doi:10.1137/18M1182218.

[AL08]   E. H. Avrett and R. Loeser. Models of the Solar Chromosphere and Transition Region from SUMER and HRTS Observations: Formation of the Extreme-Ultraviolet Spectrum of Hydrogen, Carbon, and Oxygen. *ApJS*, 175(1):229–276, March 2008. doi:10.1086/523671.

[AMO]   S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org.

[AS64]   M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards, Tenth printing, December 1972 edition, 1964.

[Ast96]   R. J. Astley. Transient wave envelope elements for wave problems. *J. Sound Vib.*, 192(1):245 – 261, 1996. doi:10.1006/jsvi.1996.0186.

[Ast00]   R. J. Astley. Infinite elements for wave problems: a review of current formulations and an assessment of accuracy. *Internat. J. Numer. Methods Engrg.*, 49(7):951–976, 2000. doi:10.1002/1097-0207(20001110)49:7<951::AID-NME989>3.0.CO;2-T.

[Bï7]   S. Börm. Directional -matrix compression for high-frequency problems. *Numer. Linear Algebra Appl.*, 24(6):e2112, 2017. doi:10.1002/nla.2112.

[BCCD+97]   S. Basu, W. J. Chaplin, J. Christensen-Dalsgaard, Y. Elsworth, G. R. Isaak, R. New, J. Schou, M. J. Thompson, and S. Tomczyk. Solar internal sound speed as inferred from combined BiSON and LOWL oscillation frequencies. *Mon. Notices Royal Astron. Soc.*, 292(2):243–251, 12 1997. doi:10.1093/mnras/292.2.243.

[BCD+18]   H. Barucq, J. Chabassier, M. Duruflé, L. Gizon, and M. Leguèbe. Atmospheric radiation boundary conditions for the Helmholtz equation. *ESAIM: M2AN*, 52(3):945–964, 2018. doi:10.1051/m2an/2017059.

[BCZ10]   A.S. Bonnet-Ben Dhia, P. Ciarlet, and C.M. Zwölf. Time harmonic wave diffraction problems in materials with sign-shifting coefficients. *J. Comput. Appl. Math.*, 234(6):1912–1919, 2010. doi:10.1016/j.cam.2009.08.041.

[BDL04]   E. Bécache, A. S. Bonnet-Ben Dhia, and G. Legendre. Perfectly Matched Layers for the Convected Helmholtz Equation. *SIAM J. Numer. Anal.*, 42(1):409–433, 2004. doi:10.1137/S0036142903420984.

[Ber94]   J.-P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 114(2):185 – 200, 1994. doi:10.1006/jcph.1994.1159.

[BFF+20a]   H. Barucq, F. Faucher, D. Fournier, L. Gizon, and H. Pham. Efficient and Accurate Algorithm for the Full Modal Green's Kernel of the Scalar Wave Equation in Helioseismology. *SIAM J. Appl. Math.*, 80(6):2657–2683, 2020. doi:10.1137/20M1336709.

[BFF+20b]   H. Barucq, F. Faucher, D. Fournier, L. Gizon, and H. Pham. Efficient computation of the modal outgoing Green's kernel for the scalar wave equation in helioseismology. Research Report RR-9338, Inria Bordeaux Sud-Ouest ; Magique 3D ; Max-Planck Institute for Solar System Research, April 2020. URL https://hal.archives-ouvertes.fr/hal-02544701.

[BFP20]   H. Barucq, F. Faucher, and H. Pham. Outgoing solutions and radiation boundary conditions for the ideal atmospheric scalar wave equation in helioseismology. *ESAIM Math. Model. Numer. Anal.*, pages 1111–1138, 2020. doi:10.1051/m2an/2019088.

[BGT82]   A. Bayliss, M. Gunzburger, and E. Turkel. Boundary Conditions for the Numerical Solution of Elliptic Equations in Exterior Regions. *SIAM J. Appl. Math.*, 42(2):430–451, 1982. doi:10.1137/0142032.

[BH07]   L. Banjai and W. Hackbusch. Hierarchical matrix techniques for low- and high-frequency Helmholtz problems. *IMA J. Numer. Anal.*, 28(1):46–79, 04 2007. doi:10.1093/imanum/drm001.

[BHNPR07] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An optimal perfectly matched layer with unbounded absorbing function for time-harmonic acoustic scattering problems. *J. Comput. Phys.*, 223(2):469–488, 2007. doi:https://doi.org/10.1016/j.jcp.2006.09.018.

[BHNPR08] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. An Exact Bounded Perfectly Matched Layer for Time-Harmonic Scattering Problems. *SIAM J. Sci. Comput.*, 30(1):312–338, 2008. doi:10.1137/060670912.

[BKD04]   A. C. Birch, A. G. Kosovichev, and Jr. Duvall, T. L. Sensitivity of Acoustic Wave Travel Times to Sound-Speed Perturbations in the Solar Interior. *ApJ*, 608(1):580–600, June 2004. doi:10.1086/386361.

[BL97]   A. Brandt and I. Livshits. Wave-ray multigrid method for standing wave equations. *Electron. Trans. Numer. Anal.*, 6:162–181, 1997. URL `http://eudml.org/doc/119506`.

[BM17]   S. Börm and J. M. Melenk. Approximation of the high-frequency Helmholtz kernel by nested directional interpolation: error analysis. *Numer. Math.*, 137(1):1–34, Sep 2017. doi:10.1007/s00211-017-0873-y.

[BPG95]   G. Ben-Porat and D. Givoli. Solution of unbounded domain problems using elliptic artificial boundaries. *Commun. numer. methods eng.*, 11(9):735–741, 1995. doi:10.1002/cnm.1640110904.

[Bre11]   H. Brezis. *Function Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext. Springer-Verlag, New York, 2011. doi:10.1007/978-0-387-70914-7.

[BS97]   I. M. Babuška and S. A. Sauter. Is the Pollution Effect of the FEM Avoidable for the Helmholtz Equation Considering High Wave Numbers? *SIAM J. Numer. Anal.*, 34(6):2392–2423, 1997. doi:10.1137/S0036142994269186.

[Bur94]   D. S. Burnett. A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion. *J. Acoust. Soc. Am.*, 96(5):2798–2816, 1994. doi:10.1121/1.411286.

[CD⁺96]   J. Christensen-Dalsgaard et al. The Current State of Solar Modeling. *Science*, 272(5266):1286–1292, 1996. doi:10.1126/science.272.5266.1286.

[CDY09]   E. Candès, L. Demanet, and L. Ying. A Fast Butterfly Algorithm for the Computation of Fourier Integral Operators. *Multiscale Model. Simul.*, 7(4):1727–24, 2009. doi:10.1137/080734339.

[CG17]   P.-H. Cocquet and M. Gander. How Large a Shift is Needed in the Shifted Helmholtz Preconditioner for its Effective Inversion by Multigrid? *SIAM J. Sci. Comput.*, 39:A438–A478, 03 2017. doi:10.1137/15M102085X.

[CK13]   D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory.* Springer, New York, third edition, 2013. doi:10.1007/978-1-4614-4942-3.

[CL06]   Z. Chen and X. Liu. An Adaptive Perfectly Matched Layer Technique for Time-Harmonic Scattering Problems. *SIAM J. Numer. Anal.*, 43(2):645–671, 2006. doi:10.1137/040610337.

[CM98]   F. Collino and P. Monk. The Perfectly Matched Layer in Curvilinear Coordinates. *SIAM J. Sci. Comput.*, 19(6):2061–2090, November 1998. doi:10.1137/S1064827596301406.

[Coc65]   J. A. Cochran. The Zeros of Hankel Functions as Functions of Their Order. *Numer. Math.*, 7(3):238–250, June 1965. doi:10.1007/BF01436080.

[CX13a]   Z. Chen and X. Xiang. A Source Transfer Domain Decomposition Method for Helmholtz Equations in Unbounded Domain. *SIAM J. Numer. Anal.*, 51(4):2331–2356, 2013. doi:10.1137/130917144.

[CX13b]   Zhiming Chen and Xueshuang Xiang. A Source Transfer Domain Decomposition Method for Helmholtz Equations in Unbounded Domain Part II: Extensions. *Numerical Mathematics: Theory, Methods and Applications*, 6(3):538–555, 2013. doi:10.1017/S1004897900001021.

[DE03]   D. P. Dreyer and O. Estorff. Improved conditioning of infinite elements for exterior acoustics. *Int. J. Numer. Methods Eng.*, 58:933–953, 2003. doi:10.1002/nme.804.

[DG98]   L. Demkowicz and K. Gerdes. Convergence of the infinite element methods for the Helmholtz equation in separable domains. *Numer. Math.*, 79:11–42, 1998. doi:10.1007/s002110050330.

[DG11]   L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov–Galerkin methods. II. Optimal test functions. *Numer. Methods Partial Differ. Equ.*, 27(1):70–105, 2011. doi:10.1002/num.20640.

[DGMZ12]   L. Demkowicz, J. Gopalakrishnan, I. Muga, and J. Zitelli. Wavenumber explicit analysis of a DPG method for the multidimensional Helmholtz equation. *Comput. Methods Appl. Mech. Eng.*, 213-216:126–138, 2012. doi:10.1016/j.cma.2011.11.024.

[DI01]   L. Demkowicz and F. Ihlenburg. Analysis of a coupled finite-infinite element method for exterior helmholtz problems. *Numer. Math.*, 88:43–73, 03 2001. doi:10.1007/PL00005440.

[DJHP93]   Jr. Duvall, T. L., S. M. Jefferies, J. W. Harvey, and M. A. Pomerantz. Time-distance helioseismology. *Nature*, 362(6419):430–432, April 1993. doi:10.1038/362430a0.

[DKC18]   V. Danchenko, M. Komarov, and P. Chunaev. Extremal and Approximative Properties of Simple Partial Fractions. *Russian Math.(Iz. VUZ)*, 62:6–41, 2018. doi:10.3103/S1066369X18120022.

[DP21]   N. Derevianko and G. Plonka. Exact Reconstruction of Extended Exponential Sums using Rational Approximation of their Fourier Coefficients. *arXiv:2103.07743*, 2021. URL https://arxiv.org/abs/2103.07743.

[EG12]   O. G. Ernst and M. J. Gander. Why it is difficult to solve Helmholtz problems with classical iterative methods. In *Numerical analysis of multiscale problems*, volume 83 of *Lect. Notes Comput. Sci. Eng.*, pages 325–363. Springer, Heidelberg, 2012. doi:10.1007/978-3-642-22061-6_10.

[ELV94]   D. V. Evans, M. Levitin, and D. Vassiliev. Existence theorems for trapped modes. *J. Fluid Mech.*, 261:21–31, 1994. doi:10.1017/S0022112094000236.

[Eva98]   L. C. Evans. *Partial differential equations*, volume 19 of *Graduate studies in mathematics*. American Mathematical Society, Providence, Rhode Island, 1998.

[EVO04]   Y.A Erlangga, C. Vuik, and C. W. Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Appl. Numer. Math.*, 50(3):409–425, 2004. doi:10.1016/j.apnum.2004.01.009.

[EY11a]   B. Engquist and L. Ying. Sweeping preconditioner for the Helmholtz equation: Hierarchical matrix representation. *Commun. Pure Appl. Math.*, 64(5):697–735, 2011. doi:10.1002/cpa.20358.

[EY11b]   B. Engquist and L. Ying. Sweeping Preconditioner for the Helmholtz Equation: Moving Perfectly Matched Layers. *Multiscale Model. Simul.*, 9(2):686–710, 2011. doi:10.1137/100804644.

[EZ18]   B. Engquist and H. Zhao. Approximate Separability of the Green's Function of the Helmholtz Equation in the High Frequency Limit. *Commun. Pure Appl. Math.*, 71(11):2220–2274, 2018. doi:10.1002/cpa.21755.

[FJC+04]   W. Finsterle, S. M. Jefferies, A. Cacciani, P. Rapex, and S. W. McIntosh. Helioseismic Mapping of the Magnetic Canopy in the Solar Chromosphere. *The Astrophysical Journal*, 613(2):L185–L188, October 2004. doi:10.1086/424996.

[FLH+17]   D. Fournier, M. Leguèbe, C. S. Hanson, L. Gizon, H. Barucq, J. Chabassier, and M. Duruflé. Atmospheric-radiation boundary conditions for high-frequency waves in time-distance helioseismology. *A&A*, 608:A109, December 2017. doi:10.1051/0004-6361/201731283.

[Fre18]   Pedro Freitas. Sharp bounds for the modulus and phase of Hankel functions with applications to Jaeger integrals. *Math. Comput.*, 87:289–308, 2018.

[Gan06]   M. J. Gander. Optimized Schwarz Methods. *SIAM J. Numer. Anal.*, 44(2):699–731, 2006. doi:10.1137/S0036142903425409.

[GB05]   L. Gizon and A. C. Birch. Local Helioseismology. *Living Rev. Sol. Phys.*, 2(1), December 2005. doi:10.12942/lrsp-2005-6.

[GBD⁺17] L. Gizon, H. Barucq, M. Duruflé, C. S. Hanson, M. Leguèbe, A. C. Birch, J. Chabassier, D. Fournier, T. Hohage, and E. Papini. Computational helioseismology in the frequency domain: acoustic waves in axisymmetric solar models with flows. *A&A*, 600:A35, 2017. doi:10.1051/0004-6361/201629470.

[GBS10] L. Gizon, A. C. Birch, and H. C. Spruit. Local Helioseismology: Three-Dimensional Imaging of the Solar Interior. *Annu. Rev. Astron. Astrophys.*, 48(1):289–338, 2010. doi:10.1146/annurev-astro-082708-101722.

[GCP⁺20] L. Gizon, R. H. Cameron, M. Pourabdian, Z.-C. Liang, D. Fournier, A. C. Birch, and C. S. Hanson. Meridional flow in the sun's convection zone is a single cell in each hemisphere. *Science*, 368(6498):1469–1472, 2020. doi:10.1126/science.aaz7119.

[GD05] N. A Gumerov and R. Duraiswami. *Fast multipole methods for the Helmholtz equation in three dimensions.* Elsevier, 2005. doi:10.1016/B978-0-08-044371-3.X5000-5.

[Ger98] K. Gerdes. The conjugated vs. the unconjugated infinite element method for the Helmholtz equation in exterior domains. *Comput. Methods Appl. Mech. Eng.*, 152(1):125 – 145, 1998. doi:10.1016/S0045-7825(97)00186-2. Containing papers presented at the Symposium on Advances in Computational Mechanics.

[GFY⁺18] L. Gizon, D. Fournier, D. Yang, A. C. Birch, and H. Barucq. Signal and noise in helioseismic holography. *A&A*, 620:A136, December 2018, 1810.00402. doi:10.1051/0004-6361/201833825.

[GGS15] M. J. Gander, I. G. Graham, and E. A. Spence. Applying GMRES to the Helmholtz Equation with Shifted Laplacian Preconditioning: What is the Largest Shift for Which Wavenumber-Independent Convergence is Guaranteed? *Numer. Math.*, 131(3):567–614, November 2015. doi:10.1007/s00211-015-0700-2.

[Gil95] P. B. Gilkey. *Invariance Theory, the Heat Equation and the Atiyah-Singer Index Theorem, Second Edition.* CRC-Press, 1995.

[Giv01] Dan Givoli. High-Order Nonreflecting Boundary Conditions without High-Order Derivatives. *J. Comput. Phys.*, 170(2):849 – 870, 2001. doi:10.1006/jcph.2001.6766.

[GK95] M. J. Grote and J. B. Keller. On Nonreflecting Boundary Conditions. *J. Comput. Phys.*, 122(2):231–243, 1995. doi:10.1006/jcph.1995.1210.

[GN00] M. J. Gander and F. Nataf. AILU: a preconditioner based on the analytic factorization of the elliptic operator. *Numer. Linear Algebra Appl.*, 7(7-8):505–526, 2000. doi:10.1002/1099-1506(200010/12)7:7/8<505::AID-NLA210>3.0.CO;2-Z.

[GNTT20] M. J. Grote, F. Nataf, J. H. Tang, and P.-H. Tournier. Parallel controllability methods for the Helmholtz equation. *Comput. Methods Appl. Mech. Eng.*, 362:112846, 2020. doi:10.1016/j.cma.2020.112846.

[Gou84]   D. O. Gough. On the Rotation of the Sun. *Philos. Trans. Royal Soc. A*, 313(1524):27–38, 1984. doi:10.1098/rsta.1984.0080.

[GPK97]   D. Givoli, I. Patlashenko, and J. B. Keller. High-order boundary conditions and finite elements for infinite domains. *Comput. Methods Appl. Mech. Eng.*, 143(1):13 – 39, 1997. doi:10.1016/S0045-7825(96)01150-4.

[GS17]   M. J. Gander and S. Solovyev. A Numerical Study on the Compressibility of Subblocks of Schur Complement Matrices Obtained from Discretized Helmholtz Equations. In I. Dimov, I. Faragó, and L. Vulkov, editors, *Numerical Analysis and Its Applications*, pages 70–81, Cham, 2017. Springer International Publishing. doi:10.1007/978-3-319-57099-0_7.

[GZ18]   M. J. Gander and H. Zhang. Restrictions on the Use of Sweeping Type Preconditioners for Helmholtz Problems. In P. E. Bjørstad, S. C. Brenner, L. Halpern, H. H. Kim, R. Kornhuber, T. Rahman, and O. B. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XXIV*, pages 321–332, Cham, 2018. Springer International Publishing.

[GZ19]   M. J. Gander and H. Zhang. A Class of Iterative Solvers for the Helmholtz Equation: Factorizations, Sweeping Preconditioners, Source Transfer, Single Layer Potentials, Polarized Traces, and Optimized Schwarz Methods. *SIAM Review*, 61(1):3–76, 2019. doi:10.1137/16M109781X.

[Hac99]   W. Hackbusch. A Sparse Matrix Arithmetic Based on H-Matrices. Part i: Introduction to H-Matrices. *Computing*, 62:89–108, 04 1999. doi:10.1007/s006070050015.

[Hal21]   M. Halla. On the treatment of exterior domains for the time-harmonic equations of stellar oscillations. *arXiv:2105.04161*, 2021. URL http://arxiv.org/abs/2105.04161.

[HH21]   M. Halla and T. Hohage. On the Well-posedness of the Damped Time-harmonic Galbrun Equation and the Equations of Stellar Oscillations. *SIAM J. Math. Anal.*, 53(4):4068–4095, 2021. doi:10.1137/20M1348558.

[HLP21]   T. Hohage, C. Lehrenfeld, and J. Preuß. Learned infinite elements. *SIAM J. Sci. Comput.*, 43(5):A3552–A3579, 2021. doi:10.1137/20M1381757.

[HN09]   T. Hohage and L. Nannen. Hardy Space Infinite Elements for Scattering and Resonance Problems. *SIAM J. Numer. Anal.*, 47(2):972–996, 2009. doi:10.1137/070708044.

[HSZ03a]   T. Hohage, F. Schmidt, and L. Zschiedrich. Solving Time-Harmonic Scattering Problems Based on the Pole Condition I: Theory. *SIAM J. Math. Anal.*, 35(1):183–210, 2003. doi:10.1137/S0036141002406473.

[HSZ03b]   T. Hohage, F. Schmidt, and L. Zschiedrich. Solving Time-Harmonic Scattering Problems Based on the Pole Condition II: Convergence of the PML Method. *SIAM J. Math. Anal.*, 35(3):547–560, 2003. doi:10.1137/S0036141002406485.

[Ihl98]   F. Ihlenburg. *Finite Element Analysis of Acoustic Scattering, Appl. Math. Sci. 132*. Springer, New York, 1998. doi:10.1007/b98828.

[Ihl00]   F. Ihlenburg. On fundamental aspects of exterior approximations with infinite elements. *J. Comput. Acoust.*, 08(01):63–80, 2000. doi:10.1142/S0218396X00000054.

[J⁺21]   F. Johansson et al. *mpmath: a Python library for arbitrary-precision floating-point arithmetic*, February 2021. http://mpmath.org/.

[JLF06]   P. Joly, J.-R. Li, and S. Fliss. Exact boundary conditions for periodic waveguides containing a local perturbation. *Commun. Comput. Phys.*, 1:945–973, 2006.

[Joh17]   F. Johansson. Arb: efficient arbitrary-precision midpoint-radius interval arithmetic. *IEEE Trans. Comput.*, 66:1281–1292, 2017. doi:10.1109/TC.2017.2690633.

[JOS⁺97]   S. M. Jefferies, Y. Osaki, H. Shibahashi, J. W. Harvey, S. D'Silva, and Jr. T. L. Duvall. Sounding the Sun's chromosphere. *ApJL*, 485(1):L49–L52, August 1997. doi:10.1086/310805.

[JRM17]   W. Jakob, J. Rhinelander, and D. Moldovan. `pybind11` – seamless operability between C++ and Python, 2017. https://github.com/pybind/pybind11.

[Kel99]   C. T. Kelley. *Iterative Methods for Optimization*. Society for Industrial and Applied Mathematics, 1999. doi:10.1137/1.9781611970920.

[KG89]   J. B. Keller and D. Givoli. Exact non-reflecting boundary conditions. *J. Comput. Phys.*, 82(1):172–192, 1989. doi:10.1016/0021-9991(89)90041-7.

[Kos01]   O. N. Kosukhin. Approximation properties of the most simple fractions. *Vestnik Moskov. Univ. Ser. 1. Mat. Mekh.*, (4):54–59, 2001.

[Kos11]   A. G. Kosovichev. *Advances in Global and Local Helioseismology: An Introductory Review*, pages 3–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-19928-8_1.

[Kow15]   P. Kowalczyk. Complex Root Finding Algorithm Based on Delaunay Triangulation. *ACM Trans. Math. Softw.*, 41(3), June 2015. doi:10.1145/2699457.

[Kow18]   P. Kowalczyk. Global Complex Roots and Poles Finding Algorithm Based on Phase Analysis for Propagation and Radiation Problems. *IEEE Trans. Antennas and Propagation*, 66(12):7198–7205, Dec 2018. doi:10.1109/TAP.2018.2869213.

[Kre14]   R. Kress. *Linear integral equations*. Springer, third edition, 2014. doi:10.1007/978-1-4614-9593-2.

[KRG63]   J. B. Keller, S. I. Rubinow, and M. Goldstein. Zeros of Hankel Functions and Poles of Scattering Amplitudes. *J. Math. Phys.*, 4(6):829–832, 1963. doi:10.1063/1.1724325.

[KRS04]   S. G. Korzennik, M. C. Rabello-Soares, and J. Schou. On the Determi-
             nation of Michelson Doppler Imager High-Degree Mode Frequencies. *ApJ*,
             602(1):481–516, February 2004, astro-ph/0207371. doi:10.1086/381021.

[KRSSL13]   S. G. Korzennik, M. C. Rabello-Soares, J. Schou, and T. P. Larson. Accu-
             rate Characterization of High-Degree Modes Using MDI Observations. *ApJ*,
             772(2):87, jul 2013. doi:10.1088/0004-637x/772/2/87.

[KSS⁺97]   A. Kosovichev, J. Schou, P. Scherrer, B. Richard, R. Bush, T. Hoeksema,
             J. Aloise, L. Bacon, A. Burnette, C. Forest, P. Giles, K. Leibrand, R. Nigam,
             M. Rubin, K. Scott, S. Williams, Sarbani Basu, J. Christensen-Dalsgaard,
             W. Dappen, and P. Milford. Internal structure and rotation of the Sun:
             First results from MDI data. *Proc. Int. Astron. Union*, 181:203, 04 1997.
             doi:10.1007/978-94-011-5163-4_18.

[LBO67]   D. Lynden-Bell and J. P. Ostriker. On the Stability of Differentially Ro-
             tating Bodies. *Mon. Notices Royal Astron. Soc.*, 136(3):293–310, 07 1967.
             doi:10.1093/mnras/136.3.293.

[Ler19]   N. Lerner. *Carleman Inequalities: An Introduction and More.* Springer Inter-
             national Publishing, Cham, Switzerland, first edition, 2019. doi:10.1007/978-
             3-030-15993-1.

[LGCL20]   Y. Liu, P. Ghysels, L. Claus, and X. S. Li. Sparse Approximate Multifrontal
             Factorization with Butterfly Compression for High Frequency Wave Equa-
             tions. *arXiv:2007.00202*, 2020. URL http://arxiv.org/abs/2007.00202.

[LHPvW21]   C. Lehrenfeld, F. Heimann, J. Preuß, and H. von Wahl. `ngsxfem`: Add-on
             to NGSolve for geometrically unfitted finite element discretizations. *J. Open
             Source Softw.*, 6(64):3237, 2021. doi:10.21105/joss.03237.

[Li05]   X. S. Li. An Overview of SuperLU: Algorithms, Implementation, and
             User Interface. *ACM Trans. Math. Softw.*, 31(3):302–325, September 2005.
             doi:10.1145/1089014.1089017.

[LM72]   J. L. Lions and E. Magenes. *Non-homogeneous boundary value problems and
             applications. Translated from the French by P. Kenneth .* Springer-Verlag
             Berlin, New York, 1972.

[LM07]   C.M. Linton and P. McIver. Embedded trapped modes in wa-
             ter waves and acoustics. *Wave Motion*, 45(1):16–29, 2007.
             doi:10.1016/j.wavemoti.2007.04.009. Special Issue on Localization of
             Wave Motion.

[LMMR00]   B. Lee, T. A. Manteuffel, S. F. McCormick, and J. Ruge. First-Order Sys-
             tem Least-Squares for the Helmholtz Equation. *SIAM J. Sci. Comput.*,
             21(5):1927–1949, 2000. doi:10.1137/S1064827598339773.

[LNS62]   R. B. Leighton, R. W. Noyes, and G. W. Simon. Velocity Fields in
             the Solar Atmosphere. I. Preliminary Report. *ApJ*, 135:474, March 1962.
             doi:10.1086/147285.

[LS71]   J. W. Leibacher and R. F. Stein. A New Description of the Solar Five-Minute Oscillation. *Astrophys. Lett.*, 7:191–192, January 1971.

[LS98]   M. Lassas and E. Somersalo. On the existence and convergence of the solution of PML equations. *Computing*, 60(3):229–241, 1998. doi:10.1007/BF02684334.

[LS15]   T. Larson and J. Schou. Improved Helioseismic Analysis of Medium-ℓ Data from the Michelson Doppler Imager. *Solar Physics*, 290, 11 2015. doi:10.1007/s11207-015-0792-y.

[MB96]   E. Michielssen and A. Boag. A multilevel matrix decomposition algorithm for analyzing scattering from large structures. *IEEE Trans. Antennas Propag.*, 44(8):1086–1093, 1996. doi:10.1109/8.511816.

[McL00]   W. McLean. *Strongly elliptic systems and boundary integral equations.* Cambridge University Press, Cambridge, United Kingdom, 2000.

[MK60]   W. Magnus and L. Kotin. The Zeros of the Hankel Function as a Function of Its Order. *Numer. Math.*, 2(1):228–244, December 1960. doi:10.1007/BF01386226.

[MS14]   A. Moiola and E. A. Spence. Is the Helmholtz Equation Really Sign-Indefinite? *SIAM Review*, 56(2):274–312, 2014. doi:10.1137/120901301.

[MT85]   R. Melrose and M. Taylor. Near peak scattering and the corrected Kirchhoff approximation for a convex obstacle. *Adv. Math.*, 55(3):242–315, 1985. doi:10.1016/0001-8708(85)90093-3.

[Nan08]   L. Nannen. *Hardy-Raum Methoden zur numerischen Lösung von Streu-und Resonanzproblemen auf unbeschränkten Gebieten.* PhD thesis, University of Göttingen, Der Andere Verlag, Tönning, 2008.

[New64]   D. J. Newman. Rational approximation to $|x|$. *Michigan Math. J.*, 11:11–14, 1964. doi:10.1307/mmj/1028999029.

[New82]   R. G. Newton. *Scattering theory of waves and particles; 2nd ed.* Springer, Berlin, 1982. doi:10.1007/978-3-642-88128-2. Also a reprint ed.: Mineola, NY, Dover, 2002.

[NK99]   R. Nigam and A. G. Kosovichev. Source of Solar Acoustic Modes. *ApJ*, 514(1):L53–L56, mar 1999. doi:10.1086/311939.

[NS11]   L. Nannen and A. Schädle. Hardy space infinite elements for Helmholtz-type problems with unbounded inhomogeneities. *Wave Motion*, 48(2):116–129, 2011. doi:10.1016/j.wavemoti.2010.09.004.

[NST18]   Y. Nakatsukasa, O. Sète, and L. N. Trefethen. The AAA Algorithm for Rational Approximation. *SIAM J. Sci. Comput.*, 40(3):A1494–A1522, 2018. doi:10.1137/16M1106122.

[Nus65]   H.M Nussenzveig. High-frequency scattering by an impenetrable sphere. *Ann. Physics*, 34(1):23 – 95, 1965. doi:10.1016/0003-4916(65)90041-2.

[NW06]   J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer, New York, NY, USA, second edition, 2006.

[Olv74]   F. W. J. Olver. *Asymptotics and special functions.* Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York (i.a.), 1974.

[PD21]   S. Petrides and L. Demkowicz. An adaptive multigrid solver for DPG methods with applications in linear acoustics and electromagnetics. *Comput. Math*, 87:12–26, 2021. doi:10.1016/j.camwa.2021.01.017.

[PHL20]   J. Preuß, T. Hohage, and C. Lehrenfeld. Sweeping preconditioners for stratified media in the presence of reflections. *SN PDE*, 1, 2020. doi:10.1007/s42985-020-00019-x.

[Pre21]   J. Preuss. Learned IE thesis: Docker image, October 2021. doi:10.25625/BDJD7J.

[RCN+99]   C. S. Rosenthal, J. Christensen-Dalsgaard, Å. Nordlund, R. F. Stein, and R. Trampedach. Convective contributions to the frequencies of solar oscillations. *A&A*, 351:689–700, November 1999, astro-ph/9803206.

[Reg59]   T. Regge. Introduction to complex orbital momenta. *Il Nuovo Cimento (1955-1965)*, 14:951–976, 1959. doi:10.1007/BF02728177.

[RKS+97]   Jr. Rhodes, E. J., A. G. Kosovichev, J. Schou, P. H. Scherrer, and J. Reiter. Measurements of Frequencies of Solar Oscillations from the MDI Medium-l Program. *Solar Physics*, 175(2):287–310, October 1997. doi:10.1023/A:1004963425123.

[SB98]   J. J. Shirron and I. Babuška. A comparison of approximate boundary conditions and infinite element methods for exterior helmholtz problems. *Comput. Methods Appl. Mech. Eng.*, 164(1):121 – 139, 1998. doi:10.1016/S0045-7825(98)00050-4. Exterior Problems of Wave Propagation.

[SBB+95]   P. H. Scherrer, R. S. Bogart, R. I. Bush, J. T. Hoeksema, A. G. Kosovichev, J. Schou, W. Rosenberg, L. Springer, T. D. Tarbell, A. Title, C. J. Wolfson, I. Zayer, and MDI Engineering Team. The Solar Oscillations Investigation - Michelson Doppler Imager. *Solar Physics*, 162(1-2):129–188, December 1995. doi:10.1007/BF00733429.

[SCGM11]   H. Schunker, R. Cameron, L. Gizon, and H. Moradi. Constructing and Characterising Solar Structure Models for Computational Helioseismology. *Solar Physics*, 271:1–26, 07 2011. doi:10.1007/s11207-011-9790-x.

[Sch97]   J. Schöberl. NETGEN An advancing front 2D/3D-mesh generator based on abstract rules. *Comput. Vis. Sci.*, 1(1):41–52, 1997. doi:10.1007/s007910050004.

[Sch14]   Joachim Schöberl. C++ Implementation of Finite Elements in NGSolve. Technical report, ASC-2014-30, Institute for Analysis and Scientific Computing; Karlsplatz 13, 1040 Vienna, Austria, September 2014.

[SD95]   F. Schmidt and P. Deuflhard. Discrete transparent boundary conditions for the numerical solution of Fresnel's equation. *Comput. Math.*, 29(9):53 – 76, 1995. doi:10.1016/0898-1221(95)00037-Y.

[SGFS01]   O. Schenk, K. Gärtner, W. Fichtner, and A. Stricker. PARDISO: a high-performance serial and parallel sparse linear solver in semiconductor device simulation. *Future Gener. Comput. Syst.*, 18(1):69–78, 2001. doi:10.1016/S0167-739X(00)00076-5. I. High Performance Numerical Methods and Applications. II. Performance Data Mining: Automated Diagnosis, Adaption, and Optimization.

[SSJ04]   T. Sekii, H. Shibahashi, and S. M. Jefferies. An Alternative Interpretation of the Double-Ridge Structure in the - Frequency Time-Distance Autocorrelation Function? In D. Danesy, editor, *SOHO 14 Helio- and Asteroseismology: Towards a Golden Future*, volume 559 of *ESA Special Publication*, page 619, October 2004.

[Sto13]   C. C. Stolk. A rapidly converging domain decomposition method for the Helmholtz equation. *J. Comput. Phys.*, 241:240–252, 2013. doi:10.1016/j.jcp.2013.01.039.

[Sto17]   C. C. Stolk. An improved sweeping domain decomposition preconditioner for the Helmholtz equation. *Adv. Comput. Math.*, 43, 02 2017. doi:10.1007/s10444-016-9475-y.

[Sto21]   C. C. Stolk. A Time-Domain Preconditioner for the Helmholtz Equation. *SIAM J. Sci. Comput.*, 43(5):A3469–A3502, 2021. doi:10.1137/20M1359997.

[Tay79]   M. Taylor. Fourier integral operators and harmonic analysis on compact manifolds. In *Proc. Symp. Pure Math.*, volume 35, pages 115–136, 1979. URL https://mtaylor.web.unc.edu/wp-content/uploads/sites/16915/2018/04/FIHA.pdf.

[Tay11]   M. Taylor. *Partial differential equations II: Qualitative studies of linear equations*, volume 116 of *Applied Mathematical Sciences*. Springer-Verlag New York, second edition, 2011. doi:10.1007/978-1-4419-7052-7.

[TEY12]   P. Tsuji, B. Engquist, and L. Ying. A sweeping preconditioner for time-harmonic Maxwell's equations with finite elements. *J. Comput. Phys.*, 231(9):3770–3783, 2012. doi:10.1016/j.jcp.2012.01.025.

[TPEY14]   P. Tsuji, J. Poulson, B. Engquist, and L. Ying. Sweeping preconditioners for elastic wave propagation with spectral element methods. *ESAIM: M2AN*, 48(2):433–447, 2014. doi:10.1051/m2an/2013114.

[TW05]   A. Toselli and O. Widlund. *Domain Decomposition Methods – Algorithms and Theory*, volume 34. Springer Series in Computational Mathematics, 2005. doi:doi.org/10.1007/b137868.

[Ulr70]   R. K. Ulrich. The Five-Minute Oscillations on the Solar Surface. *ApJ*, 162:993, December 1970. doi:10.1086/150731.

[VAL81]   J. E. Vernazza, E. H. Avrett, and R. Loeser. Structure of the solar chromosphere. III. Models of the EUV brightness components of the quiet sun. *ApJ*, 45:635–725, April 1981. doi:10.1086/190731.

[WAB+13]   C. Weisbecker, P. Amestoy, O. Boiteau, R. Brossier, A. Buttari, J. Y. L'Excellent, S. Operto, and J. Virieux. *3D frequency-domain seismic modeling with a block low-rank algebraic multifrontal direct solver*, pages 3411–3416. Society of Exploration Geophysicists, 2013. doi:10.1190/segam2013-0603.1.

[Wal60]   J. L. Walsh. *Interpolation and approximation by rational functions in the complex domain; 3rd ed.* Colloquium publications. American Mathematical Society, Providence, RI, 1960.

[Wat18]   G. N. Watson. The diffraction of electric waves by the earth. *Proc. Roy. Soc. London Ser.A*, 95(666):83–99, 1918. doi:10.1098/rspa.1918.0050.

[Wat22]   G. N. Watson. *A Treatise on the Theory of Bessel Functions.* University Press, Cambridge, 1922.

[Wat33]   G. N. Watson. General Transforms. *Proc. London Math. Soc.*, s2-35(1):156–199, 1933. doi:10.1112/plms/s2-35.1.156.

[WdHX11]   S. Wang, M. V. de Hoop, and J. Xia. On 3d modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver. *Geophys. Prospect.*, 59(5):857–873, 2011. doi:10.1111/j.1365-2478.2011.00982.x.

[Whi03]   E. T. Whittaker. An expression of certain known functions as generalized hypergeometric functions. *Bull. Amer. Math. Soc.*, 10(3):125–134, 12 1903. URL https://projecteuclid.org:443/euclid.bams/1183417759.

[Wil56]   C. H. Wilcox. A Generalization of Theorems of Rellich and Atkinson. *Proc. Am. Math. Soc.*, 7(2):271–276, 1956. doi:10.2307/2033184.

[Wlo87]   J. Wloka. *Partial Differential Equations.* Cambridge University Press, 1987. doi:10.1017/CBO9781139171755.

[XCGL10]   J. Xia, S. Chandrasekaran, M. Gu, and X. S. Li. Fast algorithms for hierarchically semiseparable matrices. *Numer. Linear Algebra Appl.*, 17(6):953–976, 2010. doi:10.1002/nla.691.

[Yin09]   L. Ying. Sparse Fourier Transform via Butterfly Algorithm. *SIAM J. Sci. Comput.*, 31(3):1678–1694, 2009. doi:10.1137/08071291X.

[ZD98]   K. Zhou and J. C. Doyle. *Essentials of robust control*, volume 104. Prentice hall Upper Saddle River, NJ, 1998.

[ZND16]   L. Zepeda-Núñez and L. Demanet. The method of polarized traces for the 2D Helmholtz equation. *J. Comput. Phys.*, 308:347 – 388, 2016. doi:10.1016/j.jcp.2015.11.040.

[ZNSHD19]   L. Zepeda-Núñez, A. Scheuer, R. J. Hewett, and L. Demanet. The method of polarized traces for the 3D Helmholtz equation. *GEOPHYSICS*, 84(4):T313–T333, 2019. doi:10.1190/geo2018-0153.1.