On the numerical computation of photonic crystal waveguide band structures

vorgelegt von
Diplom-Technomathematiker
Dirk Klindworth
aus Zeven

von der Fakultät II — Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(Dr. rer. nat.)

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Wilhelm Stannat
Gutachter:	Dr. Kersten Schmidt
Gutachter:	Prof. Dr. Volker Mehrmann
Gutachter:	Prof. Dr. Patrick Joly (INRIA Saclay / ENSTA ParisTech)

Tag der wissenschaftlichen Aussprache: 30.11.2015

Berlin 2015

Acknowledgements

It is a pleasure for me to thank all the persons that supported me throughout the time of writing this thesis.

First of all I would like to express my sincere gratitude to my supervisor Kersten Schmidt. He introduced me to the fascinating mathematical topic of photonic crystals and the calculation of their band structures. I thank him for giving me the possibility to start a PhD in his junior research group at Technische Universität Berlin. From the very first day in his group I enjoyed a motivating atmosphere, countless fruitful discussions and an intensive supervision. Kersten has always encouraged me to follow my own interests while providing guidance whenever it was needed.

Moreover, I would like to thank Volker Mehrmann and Patrick Joly for examining my thesis and for fruitful discussions that helped to improve the manuscript.

Many thanks go to Sonia Fliss, without whom this thesis would not have been possible. I have to admit that I was almost unable to cope with her work on Dirichlet-to-Neumann operators, when Kersten asked me to study it shortly after starting the PhD. However, Sonia provided endless support and numerous explanations, in particular related to functional analysis and spectral theory, and eventually, I think, I got it, and we published two articles together.

I thank Roman Kappeler for inviting me to ETH Zürich, where we had many interesting discussions about photonic crystals. He supported me with his deep understanding of photonic crystals and their applications, and encouraged me to work on the numerical aspects of photonic crystal waveguide band structure calculations.

I would also like to thank all the persons that supported me in my professional development before starting the PhD in Berlin. First of all, I would like to thank Matthias Ehrhardt, who supervised my diploma thesis. Apart from his endless support when writing the diploma thesis, I thank him for his encouragement to start a PhD. I thank Georg Vossen for his support during my stay at RWTH Aachen and Fraunhofer ILT. I did not only learn a lot from him about numerical mathematics and its applications in the context of laser welding, but also about the flexibility needed to be a mathematician in an engineering environment. Many thanks go to Martin Grepl for introducing me to the reduced basis method and giving me the possibility to work on this interesting field of research.

The very last words are dedicated to very special persons in my life: my parents, my wife and my children. *Ich möchte mich bei meinen Eltern bedanken, die mich stets unterstützt haben, meinen Weg zu gehen. Sie ließen mir immer alle Freiheiten und gaben mir dennoch unendlich viel Halt, wenn ich diesen brauchte. Und ich möchte mich bei Vanessa bedanken. Ohne sie hätte ich diese Arbeit wahrscheinlich nie begonnen. In jedem Fall wäre ich ohne sie nie damit fertig geworden. Sie gab mir die Ruhe, mich zu konzentrieren, den Freiraum, meine Gedanken zu entwickeln, sowie die Geborgenheit und Liebe, Rückschläge zu verkraften. Außerdem bedanke ich mich bei Edda, dass sie mich immer wieder auf so herrliche Weise von der Arbeit ablenkte, um auf Spielplätzen den Kopf frei zu bekommen. Schließlich möchte ich mich bei Fritz bedanken, dass er exakt so lange warten konnte, zu uns zu stoßen, bis die erste Version dieser Arbeit fertig war, und dafür, dass er mir vom ersten Tag an mit seinem Lachen jegliche Doktorandensorgen genommen hat.*

Abstract

In this thesis, we develop numerical schemes for the accurate and efficient computation of band structures of two-dimensional photonic crystal waveguides, which are periodic nanostructures with a line defect.

The perfectly periodic medium on both sides of the line defect has to be modelled mathematically. For this, we employ Dirichlet-to-Neumann and Robin-to-Robin transparent boundary conditions. These boundary conditions are transparent in the sense that they do not introduce a modelling error, which is in contrast to the well-known supercell method. The numerical realization of these transparent boundary conditions in terms of high-order finite element discretizations addresses the first objective of this work, i. e. to improve the accuracy of photonic crystal waveguide band structure calculations. The realization of Robin-to-Robin transparent boundary conditions is more involved than the realization of Dirichlet-to-Neumann boundary conditions. However, in contrast to Dirichlet-to-Neumann boundary conditions, they do not exhibit any forbidden frequencies for which the boundary conditions are not well-defined or their computation is ill-posed.

Since the eigenvalue problems with Dirichlet-to-Neumann or Robin-to-Robin transparent boundary conditions are nonlinear, efficient numerical schemes for their solution are crucial. We propose an indirect scheme based on Newton's method that is ideally suited for the eigenvalue problems under consideration. Moreover, we develop a path following algorithm, which we apply for the efficient approximation of the eigenpaths of the nonlinear eigenvalue problems, the so-called dispersion curves of the photonic crystal waveguide band structures. This path following algorithm is based on the fact that the dispersion curves are analytic, and hence, a Taylor expansion can be applied. For this, we introduce formulas for the derivatives of the dispersion curves and an adaptive selection of nodes at which a Taylor expansion is computed. With this adaptive selection we can resolve the dispersion curves in full detail while saving computation time.

Our proposed numerical scheme, that includes these two ingredients, i. e. the high-order finite element discretization of the transparent boundary conditions for periodic media and the adaptive path following algorithm, allows for efficiently resolving physical phenomena with high accuracy. For example, we show how to identify mini-stopbands, i. e. avoided crossings of dispersion curves, and we discuss the behaviour of dispersion curves at band edges, which is not possible with standard methods such as the supercell method and an equidistant sampling of dispersion curves.

Zusammenfassung

Diese Dissertation befasst sich mit der Entwicklung von numerischen Verfahren für die akkurate und effiziente Berechnung der Bandstrukturen von zweidimensionalen Photonenkristallwellenleitern. Photonenkristallwellenleiter sind periodische Nanostrukturen mit einem Liniendefekt.

Das perfekt periodische Medium an beiden Seiten des Liniendefekts muss mathematisch modelliert werden. Hierfür werden transparente Dirichlet-zu-Neumann- und Robin-zu-Robin-Randbedingungen verwendet. Diese Randbedingungen sind in dem Sinne transparent, als dass sie, im Gegensatz zu der bekannten Superzellenmethode, keinen Modellierungsfehler verursachen. Die numerische Umsetzung dieser transparenten Randbedingungen in Form von finiten Elementen hoher Ordnung adressiert das erste Ziel der vorliegenden Arbeit, also die Verbesserung der Genauigkeit von Bandstrukturberechnungen für zweidimensionale Photonenkristallwellenleiter. Die Implementation der Robin-zu-Robin-Randbedingungen ist komplizierter als die der Dirichlet-zu-Neumann-Randbedingungen, jedoch haben sie den Vorteil, dass sie für alle Frequenzen wohldefiniert und ihre Berechnung wohlgestellt ist.

Da die Eigenwertprobleme mit Dirichlet-zu-Neumann- oder Robin-zu-Robin-Randbedingungen nichtlinear sind, sind effiziente Methoden für ihre Lösung unabdingbar. Dafür wird ein neuartiges, iteratives Verfahren vorgeschlagen, das auf der Newton-Methode basiert und das ideal auf die zu lösenden Probleme abgestimmt ist. Ferner wird ein Pfadverfolgungsalgorithmus entwickelt, der für die effiziente Approximation der Eigenpfade der nichtlinearen Eigenwertprobleme, den sogenannten Dispersionskurven, angewendet wird. Dieser Pfadverfolgungsalgorithmus basiert auf der Tatsache, dass die Dispersionskurven analytische Funktionen sind und somit eine Taylor-Entwicklung möglich ist. Dazu werden Formeln zur Berechnung der Ableitungen der Dispersionskurven eingeführt und eine adaptive Auswahl der Knotenpunkte vorgeschlagen, an denen eine Taylor-Entwicklung berechnet wird. Durch diese adaptive Auswahl können die Dispersionskurven bei gleichzeitiger Zeitersparnis fein aufgelöst werden.

Das vorgeschlagene, numerische Verfahren für die Berechnung der Bandstrukturen von zweidimensionalen Photonenkristallwellenleitern, welches sowohl die Diskretisierung der transparenten Randbedingungen mit finiten Elementen hoher Ordnung sowie den Pfadverfolgungsalgorithmus enthält, ermöglicht die effiziente Auflösung physikalischer Phänomene mit hoher Genauigkeit. So wird gezeigt, wie mit Hilfe des vorgeschlagenen Verfahrens Ministoppbänder identifiziert werden können. Das sind Bereiche der Bandstruktur, in denen sich zwei Dispersionskurven sehr nahe kommen, ohne sich aber zu schneiden. Ferner kann mit dem vorgeschlagenen Verfahren das Verhalten in einer sehr kleinen Umgebung der Bandkante analysiert werden. Für beide genannten Phänomene gilt, dass sie mit Standardmethoden, wie der Superzellenmethode und einem äquidistanten Abtasten der Dispersionskurven, nicht aufgelöst werden können.

Contents

1 Introduction

Photonic crystals (PhCs) are periodic nanostructures of dielectric material. The periodicity, whose lattice is in the order of the wavelength of visible light, is induced by alternating refractive indices, or in other words, by a periodic dielectric function [JJWM08].

Depending on its frequency light either propagates through PhCs or it is reflected. Intervals of frequencies for which the PhC prohibits the propagation of light, and the light is totally reflected, are called *(photonic) band gaps* or *stop bands*. If the propagation of any polarization and any direction is prohibited, we speak of *complete (photonic) band gaps*, see for example the complete band gap in Figure 1.1. This is why PhCs are sometimes also called *photonic band gap materials*. The *PhC band structure* describes the different behaviour of light propagation and reflection. It shows the *dispersion relation*, which is the relation of the frequency of propagating light in dependence on its direction, which is given in parameterized form in terms of the *quasi-momentum* or *wave vector*. The functions, that describe the dispersion relation, are called *band functions* or, if the quasi-momentum is scalar, we speak of *dispersion curves*.

Due to the definition of allowed and forbidden frequencies of light, PhCs can be regarded as the optical counterpart of crystalline solids, whose periodic potential opens up forbidden energy bands in which electrons cannot propagate through the crystal [Kit04]. See also [Blo62] for more details on band theory for crystalline solids.

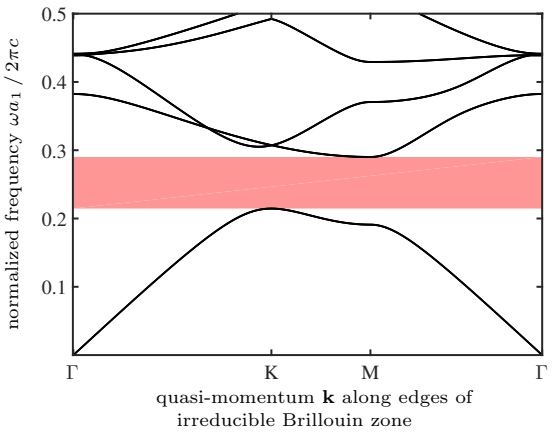


Figure 1.1: Band structure of the 2d PhC related to the 2d PhC waveguide that we will present in Example 2 in Chapter 2. The red area shows the complete band gap of the PhC.

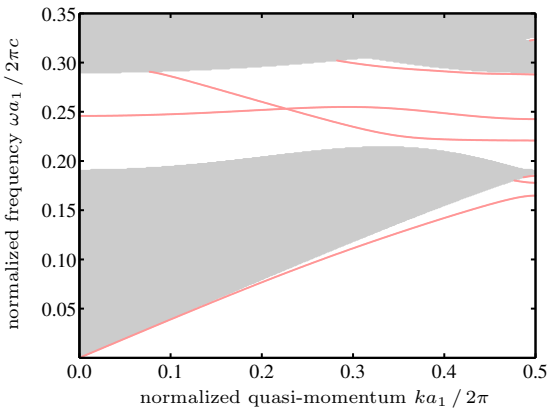


Figure 1.2: Band structure of the 2d PhC waveguide that we will present in Example 2 in Chapter 2. The red lines are dispersion curves that correspond to guided modes, and the grey color shows areas for which propagating PhC modes exist.

In general one has to distinguish between one-dimensional (1d), two-dimensional (2d) and three-dimensional (3d) PhCs, where the number of the dimension stands for the number of axes of periodicity. For example, a stack of dielectric layers is a 1d PhC, whose ability to open up band gaps was already explained in 1887 by Lord Rayleigh [Ray87]. In this work we shall focus on 2d PhCs whose periodicity is usually induced by periodically spaced, straight holes in a dielectric material, or by periodically spaced, straight rods of a dielectric material. In practise, these 2d PhCs have finite extend but it is a common simplification to assume that the device is infinite in the plane that is perpendicular to the holes/rods. These perfectly periodic structures with finite height are called *planar PhCs* or *PhC slabs*. The propagation of light in the plane is determined according to the PhC band structure with its band gaps. Using a stack of layers with different refractive indices, the light in 2d PhC slabs can also be confined in vertical direction,

which is known as *index guiding* [JJWM08]. In Figure 1.3 we present a sketch of such a 2d PhC slab with index guiding in vertical direction. The radiation, or, in other words the loss in vertical direction of PhC slabs is — to the best of our knowledge — still an open question. For homogeneous, open waveguides, however, this question has already been studied in detail [BBDHC09, JH08, JHN12]. Instead of directly studying the properties 2d PhC slabs, it is a usual simplification to consider the corresponding, ideal 2d PhCs [JJWM08]. As we will elaborate in Chapter 2, this leads to a 2d problem whose geometry is sketched in Figure 1.4.

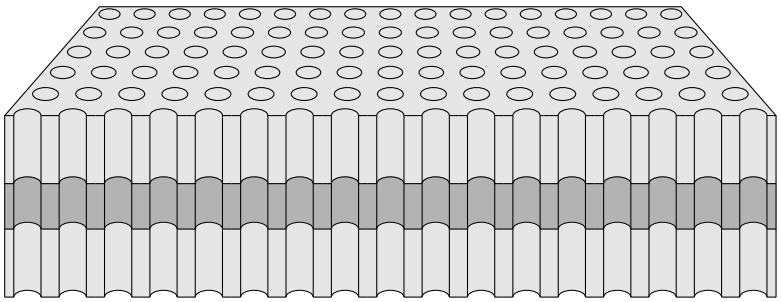


Figure 1.3: Sketch of a 2d PhC slab with index guiding in vertical direction. The refractive indices are chosen such that the light is confined in the centre layer [JJWM08].

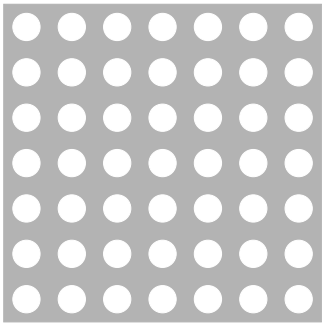


Figure 1.4: Sketch of the 2d approximation of the 2d PhC slab in Figure 1.3.

The mathematical properties of PhC band structures, and in particular 2d PhC band structures, have been studied extensively in the past decades, see for example the review article by Kuchment [Kuc01]. One of the main mathematical questions in the context of PhCs is the existence of band gaps. Even though it is experimentally and numerically evident, that gaps can exist, see for example [FG97, JJWM08] or the computer-assisted proofs in [HPW09], there is only little analytical knowledge for 2d PhCs [Kuc01]. However, there are specific cases for which the existence of gaps could be proved, e. g. high-contrast media [FK96a, FK96b]. If gaps exist the question remains how many gaps can exist. Bethe and Sommerfeld conjectured, that the number in 2d and 3d settings is finite [BS67]. For the 2d case this was shown rather recently in [Vor11].

The numerical computation of PhC band structures is addressed, for example, in the review article by Busch [Bus02]. While finite differences time domain (FDTD) calculations are well-known and established in the engineering community for simulations of finite PhCs, there has been much progress recently in 2d PhC band structure calculations in frequency domain using the *finite element method* (FEM). The proposed methods range from edge FEM [BCG06] over adaptive *hp*-FEM [SK09, GG12] to generalized FEM [BSS11, Bra13]. In this thesis we will follow the ideas in [SK09] and use high-order FEM for our 2d computations.

Due to the existence of band gaps in 2d PhCs light can be guided efficiently in *2d PhC waveguides*. That are PhCs with a line defect, that is usually created by omitting one (PhC W1 waveguide), two (PhC W2 waveguide), or more rows of holes/rods. Inside the PhC band gaps there can exist modes, so called *guided modes* or *trapped modes*, that propagate along the line defect while decaying exponentially in the PhC, i. e. in perpendicular direction to the line defect, see for example the band structure of a 2d PhC W1 waveguide with its dispersion curves corresponding to guided modes in Figure 1.2. For homogeneous line defects, as obtained when omitting one or more rows of holes/rods, e. g. for PhC W1 waveguides, the existence of guided modes was shown in [KO04], while the mathematical justification of this observation in full generality is still under investigation, see for example [AS04] for an approach using Green's functions to describe the band gap structure of 2d PhCs with a line defect. An important feature of 2d PhC waveguides is the possibility to tailor the dispersion of guided modes, and hence, obtaining, for example, slow light modes [Kra08, LWO$^+$08], i. e. guided modes with a small *group velocity* [Bri60]. Slow light modes lead to a simultaneous enhancement of the light intensity and are thus relevant for the construction of devices in nonlinear optics [SJ04]. The group velocity of PhC modes and guided modes in PhC waveguides can be determined with the help of band structure calculations, since the group velocity

is equal to the slope of the dispersion curves, i. e. the derivative of the dispersion curves with respect to the quasi-momentum [KKEJ13].

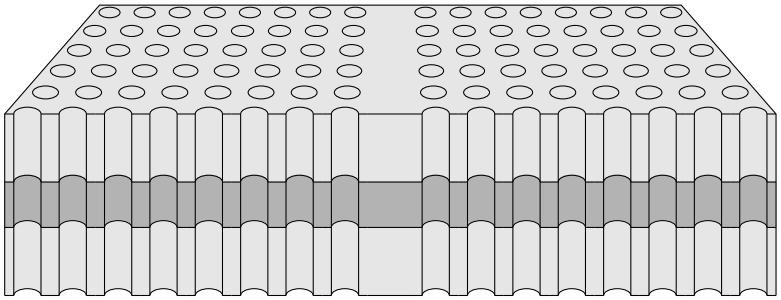


Figure 1.5: Sketch of a 2d PhC W1 slab waveguide with index guiding in vertical direction. The refractive indices are chosen such that the light is confined in the centre layer [JJWM08].

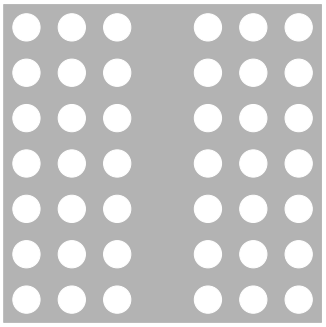


Figure 1.6: Sketch of the 2d approximation of the 2d PhC slab waveguide in Figure 1.5.

Again we note that in practise 2d PhC waveguides have finite extend. Using index guiding in vertical direction, i. e. along the holes/rods as sketched in Figure 1.5, motivates, however, the assumption of vertical invariance. For 2d PhC waveguides with infinite extend in the plane perpendicular to the holes/rods, that we will deal with in this work, for example see the 2d representation of a 2d PhC W1 waveguide in Figure 1.6, a plane wave expansion [Giv99], as used in [BSS11] for the homogeneous exterior domain of 2d PhC waveguides with finitely mainly rows of holes/rods parallel to the line defect, or as in [NS10, NS13] for PhC fibers, is not appropriate since it cannot account for the periodicity of the infinite medium. Moreover, note that homogenization techniques for periodic structures, see for example [BLP78], cannot be applied to the approximation of the periodicity of PhCs, since the wavelength of visible light, which is considered in PhC band structure calculations, is in the order of the lattice of the periodicity, and hence, asymptotic techniques for the approximation of the periodic domain by a homogeneous domain will fail.

Objectives of this work

The frequently used supercell method [Sou05, SK10] is a simple procedure for the approximative computation of guided modes in PhC waveguides. While giving good results for well-confined modes, that are guided modes with a large decay rate in perpendicular direction to the line defect, the supercell method lacks accuracy for modes that are close to the boundaries of the band gaps, the so called *band edges*, since the decay rate for these modes is significantly smaller [Sou05]. Apart from the problem of accuracy, a full band structure calculation is very time-consuming if one aims to resolve all phenomena like *crossings* of dispersion curves, *avoided crossings*, also known as *anti-crossings* [ORB+01, OBS+02], and the behaviour at band edges in full detail. This shows that there is both, a need for

(i) *accuracy*, and
(ii) *efficiency*

in calculating PhC waveguide band structures. This thesis is concerned with these two objectives. The goal of *accuracy* is addressed by introducing *transparent boundary conditions* at the interfaces of line defect and perfectly periodic medium. Boundary conditions are called *transparent* if the solution of the problem with these boundary conditions is identical to the solution of the original problem restricted to the truncated domain. Transparent boundary conditions for periodic media using Dirichlet-to-Neumann (DtN) maps were introduced in [JLF06] for 2d PhCs, see also [FJ09, FCB10, FJL10]. In [Fli13] this approach was rigorously extended to the computation of guided modes in PhC waveguides. Depending on the periodic medium, these DtN maps may not be well-defined at all frequencies and their computation can be ill-posed. Robin-to-Robin (RtR) maps resolve this problem [Fli09]. In this thesis we will present both, DtN transparent boundary conditions and RtR transparent boundary conditions. Note that the

DtN and RtR approaches in [JLF06, Fli09] are different in concept and objective from the DtN maps developed in [YL06, YL07, HL08]. The latter DtN maps are employed for size robust computations in finite periodic structures, while the DtN and RtR maps, that we will deal with in this work, model the infinite periodic medium of PhCs.

The DtN and RtR transparent boundary conditions are an alternative to the frequently used supercell approach. In contrast to the supercell approach they do not introduce a modelling error. In this sense they allow for an *exact* computation of guided modes in PhC waveguides. In this work we will develop the numerical realization of the DtN and RtR approaches, ranging from discretization to the numerical solution of the nonlinear eigenvalue problems.

The second objective of this thesis is — as mentioned above — to improve the efficiency of PhC and PhC waveguide band structure calculations. For this we will develop an adaptive path following algorithm, that can be applied to the linear problems in 2d PhCs and 2d PhC waveguides using the supercell approach as well as to the nonlinear problems of 2d PhC waveguides with DtN or RtR transparent boundary conditions. The proposed method is based on the fact that the dispersion curves in band structures are analytic functions and hence, a Taylor expansion of these functions is possible. We will show how to compute the group velocity as well as any higher derivative of the dispersion curves. Our approach differs significantly from the perturbation theory employed in [SS88, Sip00, HFBW01] for the computation of the group velocity and the *group velocity dispersion*, which is the second dispersion curve derivative, since we develop closed formulas that do not need to be truncated such as the infinite sums in [SS88, Sip00, HFBW01] for the computation of the group velocity dispersion. Moreover, our approach is unique, since it allows for an extension of the formulas to arbitrary orders.

Outline of the thesis

This thesis is organized as follows: In Chapter 2 we elaborate on the mathematical modelling involved in PhC and PhC waveguide band structure calculations, we will review the spectral properties of the associated differential operators, comment on the discretization of the eigenvalue problems using high-order FEM and introduce problem settings of PhCs and PhC waveguides, that we will consider in our numerical examples throughout this thesis. In Chapter 3 we give a very brief review of algorithms for the numerical solution of matrix eigenvalue problem. Apart from reviewing well-known methods, we will also propose a new iterative solver for nonlinear eigenvalue problems that is based on Newton's method. Chapter 4 is dedicated to the computation of the group velocity and any higher derivatives of the dispersion curves of the linear eigenvalue problems related to PhC band structure calculations and PhC waveguide band structure calculations when using the supercell approach. Chapter 5 then deals with the adaptive path following algorithm. First we will generalize the procedure developed in Chapter 4 for the computation of dispersion curve derivatives to general, nonlinear, parameterized matrix eigenvalue problems. Then we propose the adaptive scheme and apply it for an efficient computation of the band structure of a PhC W1 waveguide. In Chapters 6 and 7 we introduce transparent boundary conditions based on DtN and RtR operators, respectively. We will comment on their computation, differentiability and discretization, before employing them to truncate the domain of the eigenvalue problem. We elaborate on the numerical solution of the resulting nonlinear eigenvalue problems using the methods, that we reviewed and proposed in Chapter 3. Finally, we will present extensive numerical results including the application of the adaptive path following algorithm. In Chapter 8 we give concluding remarks and comment on the perspectives of future research.

2 Mathematical modelling of photonic crystal waveguides

In this chapter we will introduce the mathematical formulation of the problems that we shall deal with in this thesis: the computation of modes in 2d PhCs and the computation of guided modes in 2d PhC waveguides. We present the mathematical modelling of these problems and comment on their mathematical properties. Before we start with the introduction of the two problems, we shall elaborate on the description of electromagnetic waves in devices with invariance in one direction, which is assumed to be the case for 2d PhCs and 2d PhC waveguides.

2.1 Electromagnetic waves in two dimensions

The propagation of light in PhCs is described by the macroscopic Maxwell equations without free charges and currents

$$\nabla \times \mathbf{E}(t,\mathbf{x}) = -\tfrac{\partial}{\partial t}\mathbf{B}(t,\mathbf{x}), \tag{2.1a}$$
$$\nabla \times \mathbf{H}(t,\mathbf{x}) = \tfrac{\partial}{\partial t}\mathbf{D}(t,\mathbf{x}), \tag{2.1b}$$
$$\nabla \cdot \mathbf{D}(t,\mathbf{x}) = 0, \tag{2.1c}$$
$$\nabla \cdot \mathbf{B}(t,\mathbf{x}) = 0, \tag{2.1d}$$

where $\mathbf{E} : \mathbb{R}^+ \times \mathbb{R}^3 \to \mathbb{R}^3$ and $\mathbf{H} : \mathbb{R}^+ \times \mathbb{R}^3 \to \mathbb{R}^3$ are the macroscopic electric and magnetic fields, and $\mathbf{D} : \mathbb{R}^+ \times \mathbb{R}^3 \to \mathbb{R}^3$ and $\mathbf{B} : \mathbb{R}^+ \times \mathbb{R}^3 \to \mathbb{R}^3$ are the displacement and magnetic induction fields. The macroscopic Maxwell equations (2.1) are completed by the linear material laws

$$\mathbf{D}(t,\mathbf{x}) = \varepsilon_0 \varepsilon(\mathbf{x}) \mathbf{E}(t,\mathbf{x}), \tag{2.2a}$$
$$\mathbf{B}(t,\mathbf{x}) = \mu_0 \mu(\mathbf{x}) \mathbf{H}(t,\mathbf{x}), \tag{2.2b}$$

with the vacuum and relative dielectric permittivities ε_0 and ε, and the vacuum and relative magnetic permeability μ_0 and μ. These linear material laws are reasonable approximations to the actually nonlinear relations of the fields in case of small amplitudes. See for example the textbook by Jackson [Jac98] for a comprehensive introduction to electromagnetic fields and Maxwell's equations.

In PhCs the material can assumed to be nonmagnetic, i. e. we may set $\mu \equiv 1$. Hence, Eqs. (2.1) and (2.2) reduce to

$$\begin{aligned}
\nabla \times \mathbf{E}(t,\mathbf{x}) &= -\mu_0 \tfrac{\partial}{\partial t}\mathbf{H}(t,\mathbf{x}),\\
\nabla \times \mathbf{H}(t,\mathbf{x}) &= \varepsilon_0 \varepsilon(\mathbf{x}) \tfrac{\partial}{\partial t}\mathbf{E}(t,\mathbf{x}),\\
\nabla \cdot \varepsilon(\mathbf{x})\mathbf{E}(t,\mathbf{x}) &= 0,\\
\nabla \cdot \mathbf{H}(t,\mathbf{x}) &= 0.
\end{aligned}$$

Due to the time-independence of the coefficients our considerations can be reduced to time-harmonic electric and magnetic fields

$$\begin{aligned}
\mathbf{E}(t,\mathbf{x}) &= \mathrm{Re}\Big(\widehat{\mathbf{E}}(\mathbf{x})\mathrm{e}^{-\mathrm{i}\omega t}\Big),\\
\mathbf{H}(t,\mathbf{x}) &= \mathrm{Re}\Big(\widehat{\mathbf{H}}(\mathbf{x})\mathrm{e}^{-\mathrm{i}\omega t}\Big),
\end{aligned}$$

with frequency $\omega \in \mathbb{R}^+$, that satisfy the time-harmonic Maxwell equations

$$\nabla \times \widehat{\mathbf{E}}(\mathbf{x}) = \mathrm{i}\omega\mu_0 \widehat{\mathbf{H}}(\mathbf{x}), \tag{2.3a}$$

$$\nabla \times \widehat{\mathbf{H}}(\mathbf{x}) = -\mathrm{i}\omega\varepsilon_0\varepsilon(\mathbf{x})\widehat{\mathbf{E}}(\mathbf{x}), \tag{2.3b}$$

where the time-harmonic versions of Eqs. (2.1c) and (2.1d), i. e.

$$\nabla \cdot \varepsilon(\mathbf{x})\widehat{\mathbf{E}}(\mathbf{x}) = 0, \tag{2.4a}$$

$$\nabla \cdot \widehat{\mathbf{H}}(\mathbf{x}) = 0, \tag{2.4b}$$

are implicitly satisfied, which can easily be seen by applying the divergence operator to Eqs. (2.3a) and (2.3b), and considering the fact that $\omega > 0$.

Applying the curl operator to (2.3a) and using (2.3b), we obtain

$$\nabla \times \Big(\nabla \times \widehat{\mathbf{E}}(\mathbf{x})\Big) - \frac{\omega^2}{c^2}\varepsilon(\mathbf{x})\widehat{\mathbf{E}}(\mathbf{x}) = 0, \tag{2.5a}$$

where we substituted $\varepsilon_0\mu_0 = c^{-2}$, with the velocity of light c. On the other hand, applying the curl operator to (2.3b) and using (2.3a), we arrive at

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{x})}\nabla \times \widehat{\mathbf{H}}(\mathbf{x})\right) - \frac{\omega^2}{c^2}\widehat{\mathbf{H}}(\mathbf{x}) = 0. \tag{2.5b}$$

As elaborated in Chapter 1, we shall consider 2d PhCs and 2d PhC waveguides as approximations to realistic PhC slabs and PhC slab waveguides. While the latter have finite height, 2d PhCs and 2d PhC waveguides are invariant in the direction of the holes/rods, the x_3-direction, say. In other words, the relative dielectric permittivity ε satisfies $\varepsilon(\mathbf{x}) = \varepsilon(x_1, x_2, 0)$ for all $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$, which motivates that we only look for solutions of the electric field $\widehat{\mathbf{E}}$ and the magnetic field $\widehat{\mathbf{H}}$ that also satisfy this condition, i. e. $\widehat{\mathbf{E}}(\mathbf{x}) = \widehat{\mathbf{E}}(x_1, x_2, 0)$ for all $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ and $\widehat{\mathbf{H}}(\mathbf{x}) = \widehat{\mathbf{H}}(x_1, x_2, 0)$ for all $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$. Then all x_3-derivatives in the differential operators vanish and it is straightforward [Kuc01] to verify that Eqs. (2.5a) and (2.3a) decouple into equations for $(\widehat{E}_3, \widehat{H}_1, \widehat{H}_2)$, the *transverse magnetic (TM) mode*, for which the electric field in x_3-direction satisfies the 2d scalar Helmholtz equation

$$-\Delta\widehat{E}_3(\mathbf{x}) - \frac{\omega^2}{c^2}\varepsilon(\mathbf{x})\widehat{E}_3(\mathbf{x}) = 0, \qquad \mathbf{x} \in \mathbb{R}^2, \tag{2.6}$$

and the magnetic field components $\widehat{H}_1$ and $\widehat{H}_2$ satisfy

$$\widehat{H}_1(\mathbf{x}) = \frac{1}{\mathrm{i}\omega\mu_0}\frac{\partial}{\partial x_2}\widehat{E}_3(\mathbf{x}),$$

$$\widehat{H}_2(\mathbf{x}) = -\frac{1}{\mathrm{i}\omega\mu_0}\frac{\partial}{\partial x_1}\widehat{E}_3(\mathbf{x}).$$

Similarly, Eqs. (2.5b) and (2.3b) decouple into equations for $(\widehat{H}_3, \widehat{E}_1, \widehat{E}_2)$, the *transverse electric (TE) mode*, for which the magnetic field in x_3-direction satisfies the 2d scalar Helmholtz equation

$$-\nabla \cdot \frac{1}{\varepsilon(\mathbf{x})}\nabla\widehat{H}_3(\mathbf{x}) - \frac{\omega^2}{c^2}\widehat{H}_3(\mathbf{x}) = 0, \qquad \mathbf{x} \in \mathbb{R}^2, \tag{2.7}$$

and the electric field components $\widehat{E}_1$ and $\widehat{E}_2$ satisfy

$$\widehat{E}_1(\mathbf{x}) = -\frac{1}{\mathrm{i}\omega\varepsilon_0\varepsilon(\mathbf{x})}\frac{\partial}{\partial x_2}\widehat{H}_3(\mathbf{x}),$$

$$\widehat{E}_2(\mathbf{x}) = \frac{1}{\mathrm{i}\omega\varepsilon_0\varepsilon(\mathbf{x})}\frac{\partial}{\partial x_1}\widehat{H}_3(\mathbf{x}).$$

Remark 2.1. *The curl operator has an infinite-dimensional null-space [Mon03], which yields spurious zero eigenvalues of the eigenvalue problems (2.5) for the time-harmonic electric and magnetic field, respectively, if the divergence equations (2.4) are not explicitly taken into account. For 3d Maxwell equations this issue can be resolved by applying the Helmholtz decomposition [Mon03]. In our case, the infinite-dimensional null-space of the curl operator is implicitly removed by the transformation of the 3d Maxwell eigenvalue problems (2.5) to the 2d scalar Helmholtz eigenvalue problems (2.8), which was motivated by our assumption of a x_3-independent dielectric permittivity and hence, a restriction to x_3-independent solutions.*

In this thesis we shall consider both modes, the TM mode (2.6) and the TE mode (2.7), simultaneously and choose

$$-\nabla\cdot\alpha(\mathbf{x})\nabla U(\mathbf{x})-\omega^2\beta(\mathbf{x})U(\mathbf{x})=0,\qquad \mathbf{x}\in\mathbb{R}^2, \tag{2.8}$$

as our governing equation. In the TM mode, U describes the electric field in x_3-direction and the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are determined through

$$\alpha(\mathbf{x})=1 \qquad \text{and} \qquad \beta(\mathbf{x})=\frac{1}{c^2}\,\varepsilon(\mathbf{x}). \tag{2.9a}$$

On the other hand, in the TE mode, U denotes the magnetic field in x_3-direction and the coefficients are defined by

$$\alpha(\mathbf{x})=\frac{1}{\varepsilon(\mathbf{x})} \qquad \text{and} \qquad \beta(\mathbf{x})=\frac{1}{c^2}. \tag{2.9b}$$

Note that, for simplicity of notation, the velocity of light c is incorporated in the coefficient β.

Finally, we note that (2.8) satisfies important scaling properties. On the one hand, it can easily be verified that a coordinate stretching $\mathbf{x}'=s\mathbf{x}$, with $s\in\mathbb{R}$, results in the same system when rescaling the frequency $\omega'=s^{-1}\omega$. On the other hand, one arrives at the same rescaling of the frequency when choosing the rescaled permittivity $\varepsilon'(\mathbf{x})=s\varepsilon(\mathbf{x})$. These properties illustrate that the problem under consideration does not have a specific length scale.

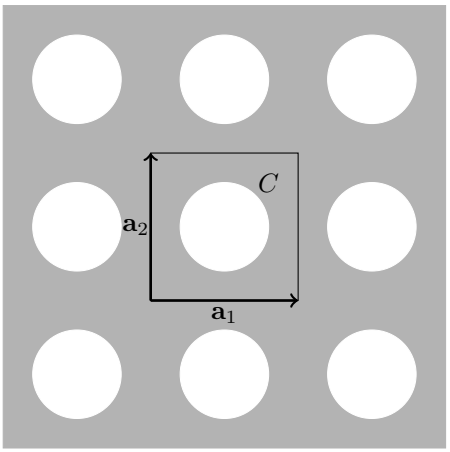
2.2 Modes in two-dimensional photonic crystals

PhCs have a discrete translational symmetry [JJWM08]. More precisely, there exist two linearly independent vectors $\mathbf{a}_1,\mathbf{a}_2\in\mathbb{R}^2$ such that the permittivity ε_{PhC} of 2d PhCs satisfies

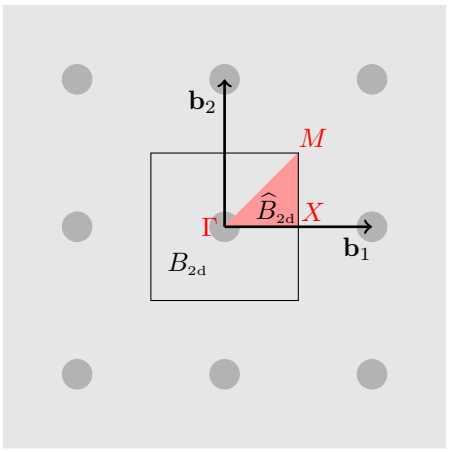
$$\varepsilon_{\text{PhC}}(\mathbf{x}+\mathbf{a}_1+\mathbf{a}_2)=\varepsilon_{\text{PhC}}(\mathbf{x}) \tag{2.10}$$

for all $\mathbf{x}\in\mathbb{R}^2$. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ with smallest possible lengths $a_i=|\mathbf{a}_i|$, $i=1,2$, such that (2.10) is satisfied, are called *lattice vectors* and their lengths are called *lattice constants*. These lattice vectors span a parallelogram — the *unit cell* C of the PhC. The lattice vectors are not defined uniquely by the pattern, or *lattice* of the PhC. Without loss of generality, we can set $\mathbf{a}_1=a_1(1,0)^{\mathrm{T}}$. The lattice vector $\mathbf{a}_2$, however, is not fixed by this choice. For example, it is conventional for the square lattice shown in Figure 2.1a that the lattice vector $\mathbf{a}_2$ is chosen to be $\mathbf{a}_2=a_1(0,1)^{\mathrm{T}}$ with lattice constant $a_2=a_1$. However, it can also chosen to be $\mathbf{a}_2=a_1(1,1)^{\mathrm{T}}$ with length $a_2=\sqrt{2}\,a_1$. On the other hand, for the hexagonal lattice, sometimes also called triangular lattice, shown in Figure 2.2a, two different choices of $\mathbf{a}_2$ are common: while the choice $\mathbf{a}_2=\sqrt{3}\,a_1(0,1)^{\mathrm{T}}$ is orthogonal to $\mathbf{a}_1$ and has length $a_2=\sqrt{3}\,a_1$, we shall prefer $\mathbf{a}_2=\frac{a_1}{2}(1,\sqrt{3})^{\mathrm{T}}$, which has lattice constant $a_2=a_1$. Hence, in case of a square or hexagonal lattice we can choose a unit cell with lattice vectors that have the same length. In this case we may write $a=a_1=a_2$. Even though the scaling properties described above allow for choosing a unitary lattice constant $a=1$, we will explicitly use the lattice constants a, a_1 and a_2 in the sequel of this work.

In the context of PhC we may assume that the permittivity is a piecewise constant, positive function $\varepsilon_{\text{PhC}}:\mathbb{R}^2\to\mathbb{R}^+$, that is bounded from below and above. All theoretical results that we will refer to or develop in this work are also applicable to permittivities that are not piecewise constant. However, all our numerical results are computed for such configurations where ε_{PhC} takes some constant value in the dielectric material and $\varepsilon_0=1$ in air.

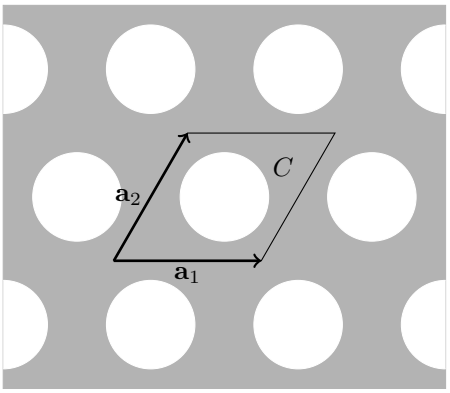


(a) Unit cell C of PhC with square lattice.

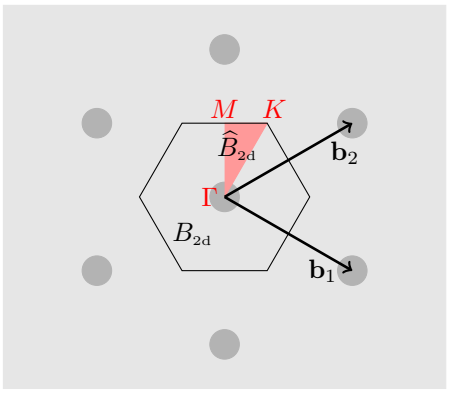


(b) Brillouin zone $B_{2\mathrm{d}}$ of reciprocal lattice.

Figure 2.1: PhC of square lattice (a) with unit cell C, lattice vectors $\mathbf{a}_1 = a(1,0)^{\mathrm{T}}$ and $\mathbf{a}_2 = a(0,1)^{\mathrm{T}}$, and its reciprocal lattice (b) with Brillouin zone $B_{2\mathrm{d}}$, irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$ and reciprocal lattice vectors $\mathbf{b}_1 = \frac{2\pi}{a}(1,0)^{\mathrm{T}}$ and $\mathbf{b}_2 = \frac{2\pi}{a}(0,1)^{\mathrm{T}}$. The irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$ has vertices $\Gamma = (0,0)^{\mathrm{T}}$, $X = \frac{\pi}{a}(1,0)^{\mathrm{T}}$ and $M = \frac{\pi}{a}(1,1)^{\mathrm{T}}$.



(a) Unit cell C of PhC with hexagonal lattice.



(b) Brillouin zone $B_{2\mathrm{d}}$ of reciprocal lattice.

Figure 2.2: PhC of hexagonal lattice (a) with unit cell C, lattice vectors $\mathbf{a}_1 = a(1,0)^{\mathrm{T}}$ and $\mathbf{a}_2 = \frac{a}{2}(1,\sqrt{3})^{\mathrm{T}}$, and its reciprocal lattice (b) with Brillouin zone $B_{2\mathrm{d}}$, irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$ and reciprocal lattice vectors $\mathbf{b}_1 = \frac{2\pi}{a}(1,-1/\sqrt{3})^{\mathrm{T}}$ and $\mathbf{b}_2 = \frac{2\pi}{a}(1,1/\sqrt{3})^{\mathrm{T}}$. The irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$ has vertices $\Gamma = (0,0)^{\mathrm{T}}$, $K = \frac{2\pi}{a}(1/3,1/\sqrt{3})^{\mathrm{T}}$ and $M = \frac{2\pi}{a}(0,1/\sqrt{3})^{\mathrm{T}}$.

Considering the periodicity of the permittivity $\varepsilon_{\mathrm{PhC}}$ and hence, of the coefficients α_{PhC} and β_{PhC}, that can be determined depending on the mode using (2.9), we can apply the Floquet theory [Kuc93] to (2.8). The *Floquet transform* of U reads

$$\widetilde{U}(\mathbf{x},\mathbf{k}) = \sum_{m_1,m_2\in\mathbb{Z}} U(\mathbf{x} - m_1\mathbf{a}_1 - m_2\mathbf{a}_2)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(m_1\mathbf{a}_1+m_2\mathbf{a}_2)},$$

where $\mathbf{k} = (k_1,k_2)^{\mathrm{T}} \in \mathbb{R}^2$ is called *quasi-momentum* or *(Floquet) wave vector*. The Floquet transform can be regarded as analogue of the Fourier transform for periodic media. However, note that the Floquet transform — in contrast to the Fourier transform — still depends on the spatial variable $\mathbf{x}$. Shifting the Floquet transform $\widetilde{U}$ in space by $m_1\mathbf{a}_1 + m_2\mathbf{a}_2$, with $m_1, m_2 \in \mathbb{Z}$, gives the *Floquet condition*

$$\widetilde{U}(\mathbf{x} + m_1\mathbf{a}_1 + m_2\mathbf{a}_2, \mathbf{k}) = \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(m_1\mathbf{a}_1+m_2\mathbf{a}_2)}\widetilde{U}(\mathbf{x},\mathbf{k}), \tag{2.11}$$

which shows that is sufficient to study $\widetilde{U}(\mathbf{x},\mathbf{k})$ only in the unit cell C spanned by the lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. Moreover, we can see easily that the Floquet transform $\widetilde{U}$ is periodic with respect to the quasi-momentum $\mathbf{k}$, i. e.

$$\widetilde{U}(\mathbf{x},\mathbf{k}+m_1\mathbf{b}_1+m_2\mathbf{b}_2)=\widetilde{U}(\mathbf{x},\mathbf{k}),$$

for all $m_1,m_2\in\mathbb{Z}$, where $\mathbf{b}_1,\mathbf{b}_2\in\mathbb{R}^2$ are the *reciprocal lattice vectors* of the *reciprocal lattice* that satisfy

$$\frac{1}{2\pi}\mathbf{a}_i\cdot\mathbf{b}_j\in\mathbb{Z}$$

for all $i,j=1,2$. The reciprocal lattices of the PhCs with square and hexagonal lattice and its reciprocal lattice vectors are shown in Figures 2.1b and 2.2b. For a detailed description of how to construct the reciprocal lattice the reader is referred to, e. g. the books by Kittel [Kit04] or Joannopoulos and coworkers [JJWM08]. The periodicity of the Floquet transform $\widetilde{U}$ with respect to the quasi-momentum $\mathbf{k}$ implies that we can restrict the choice of $\mathbf{k}$ to a subset of $\mathbb{R}^2$, the so called *(first) Brillouin zone* $B_{2\mathrm{d}}$, which is equal to the Wigner-Seitz cell [Kit04] of the reciprocal lattice centered at the origin $\mathbf{k}=0$.

Using the periodicity of the Floquet transform with respect to the quasi-momentum and employing the Floquet condition (2.11) we can deduce that the problem (2.8), which is posed in $\mathbb{R}^2$, can be transformed to a family of problems in the bounded domain C, i. e. for all $\mathbf{k}\in B_{2\mathrm{d}}$ the Floquet transform $\widetilde{U}$ satisfies

$$-\nabla\cdot\alpha_{\mathrm{PhC}}(\mathbf{x})\nabla\widetilde{U}(\mathbf{x},\mathbf{k})-\omega^2\beta_{\mathrm{PhC}}(\mathbf{x})\widetilde{U}(\mathbf{x},\mathbf{k})=0,\qquad \mathbf{x}\in C, \tag{2.12a}$$

with quasi periodic boundary conditions

$$\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{R}}}=\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{a}_1}\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{L}}}, \tag{2.12b}$$

$$\alpha_{\mathrm{PhC}}\,\partial_{\mathbf{n}_{\mathrm{R}}}\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{R}}}=-\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{a}_1}\alpha_{\mathrm{PhC}}\,\partial_{\mathbf{n}_{\mathrm{L}}}\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{L}}}, \tag{2.12c}$$

$$\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{T}}}=\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{a}_2}\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{B}}}, \tag{2.12d}$$

$$\alpha_{\mathrm{PhC}}\,\partial_{\mathbf{n}_{\mathrm{T}}}\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{T}}}=-\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{a}_2}\alpha_{\mathrm{PhC}}\,\partial_{\mathbf{n}_{\mathrm{B}}}\widetilde{U}(\cdot,\mathbf{k})\,|_{\Sigma_{\mathrm{B}}} \tag{2.12e}$$

on the left, right, bottom and top boundaries Σ_{L}, Σ_{R}, Σ_{B} and Σ_{T} of C, where $\partial_{\mathbf{n}_{\mathrm{L}}}\widetilde{U}$, $\partial_{\mathbf{n}_{\mathrm{R}}}\widetilde{U}$, $\partial_{\mathbf{n}_{\mathrm{B}}}\widetilde{U}$ and $\partial_{\mathbf{n}_{\mathrm{T}}}\widetilde{U}$ denote the outward normal derivatives of $\widetilde{U}$ on the boundaries Σ_{L}, Σ_{R}, Σ_{B} and Σ_{T}, respectively, i. e. $\partial_{\mathbf{n}_{\mathrm{L}}}\widetilde{U}=\mathbf{n}_{\mathrm{L}}\cdot\nabla\widetilde{U}$, $\partial_{\mathbf{n}_{\mathrm{R}}}\widetilde{U}=\mathbf{n}_{\mathrm{R}}\cdot\nabla\widetilde{U}$, $\partial_{\mathbf{n}_{\mathrm{B}}}\widetilde{U}=\mathbf{n}_{\mathrm{B}}\cdot\nabla\widetilde{U}$ and $\partial_{\mathbf{n}_{\mathrm{T}}}\widetilde{U}=\mathbf{n}_{\mathrm{T}}\cdot\nabla\widetilde{U}$ with the outward unit normal vectors

$$\mathbf{n}_{\mathrm{L}}=\frac{1}{a_2}\begin{pmatrix}0&-1\\1&0\end{pmatrix}\mathbf{a}_2,\qquad \mathbf{n}_{\mathrm{R}}=-\mathbf{n}_{\mathrm{L}}=\frac{1}{a_2}\begin{pmatrix}0&1\\-1&0\end{pmatrix}\mathbf{a}_2,$$

$$\mathbf{n}_{\mathrm{B}}=\begin{pmatrix}0\\-1\end{pmatrix},\qquad \mathbf{n}_{\mathrm{T}}=-\mathbf{n}_{\mathrm{B}}=\begin{pmatrix}0\\1\end{pmatrix}.$$

Now we introduce some function spaces to rigorously reformulate (2.12a). Let $\mathsf{H}^1(C)$ be the usual space of square integrable functions in C whose weak gradient is also square integrable. Then we define the periodic space

$$\mathsf{H}^1_{\mathrm{p}}(C):=\left\{u\in\mathsf{H}^1(C)\text{ with }u\,|_{\Sigma_{\mathrm{L}}}=u\,|_{\Sigma_{\mathrm{R}}}\text{ and }u\,|_{\Sigma_{\mathrm{B}}}=u\,|_{\Sigma_{\mathrm{T}}}\right\}.$$

Moreover, let $\mathsf{H}^1(\Delta,C)$ be the subspace of $\mathsf{H}^1(C)$ with functions whose Laplacian is square integrable. Then we define

$$\mathsf{H}^1_{\mathrm{p}}(\Delta,C,\alpha):=\left\{u\in\mathsf{H}^1(\Delta,C)\cap\mathsf{H}^1_{\mathrm{p}}(C)\text{ with }\alpha\partial_{\mathbf{n}_{\mathrm{L}}}u\,|_{\Sigma_{\mathrm{L}}}=-\alpha\partial_{\mathbf{n}_{\mathrm{R}}}u\,|_{\Sigma_{\mathrm{R}}}\text{ and }\alpha\partial_{\mathbf{n}_{\mathrm{B}}}u\,|_{\Sigma_{\mathrm{B}}}=-\alpha\partial_{\mathbf{n}_{\mathrm{T}}}u\,|_{\Sigma_{\mathrm{T}}}\right\}.$$

With these definitions and the substitution $\widetilde{U}(\mathbf{x},\mathbf{k})=\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}u(\mathbf{x},\mathbf{k})$, the quasi periodic problem (2.12a) is equivalent to the eigenvalue problem: find $(\omega^2,\mathbf{k})\in\mathbb{R}^+\times B_{2\mathrm{d}}$ such that there exists a non-trivial $u\in\mathsf{H}^1_{\mathrm{p}}(\Delta,C,\alpha_{\mathrm{PhC}})$ that satisfies

$$-(\nabla+\mathrm{i}\mathbf{k})\cdot\alpha_{\mathrm{PhC}}(\nabla+\mathrm{i}\mathbf{k})u-\omega^2\beta_{\mathrm{PhC}}u=0\qquad\text{in }C. \tag{2.13}$$

The eigenvalues ω_n^2, $n \in \mathbb{N}$, of (2.13) in dependence of the quasi-momentum $\mathbf{k} \in B_{2\mathrm{d}}$ define the band structure of the PhCs. The functions $\omega_n^2(\mathbf{k})$ are continuous [Kuc01] and their projection onto the frequency axis defines the spectrum $\widetilde{\sigma}_{\mathrm{PhC}}$ of the operator

$$\widetilde{\mathcal{A}}_{\mathrm{PhC}} := -\frac{1}{\beta_{\mathrm{PhC}}} \nabla \cdot \alpha_{\mathrm{PhC}} \nabla$$

related to the eigenvalue problem (2.12a) with quasi periodic boundary conditions. Let $\sigma_{\mathrm{PhC}}(\mathbf{k})$ denote the spectrum of the operator

$$\mathcal{A}_{\mathrm{PhC}}(\mathbf{k}) := -\frac{1}{\beta_{\mathrm{PhC}}} (\nabla + \mathrm{i}\mathbf{k}) \cdot \alpha_{\mathrm{PhC}} (\nabla + \mathrm{i}\mathbf{k})$$

defined on $\mathsf{H}^1_{\mathrm{p}}(\Delta, C, \alpha_{\mathrm{PhC}})$ and related to the PhC eigenvalue problem (2.13) with periodic boundary conditions. Then the spectrum $\widetilde{\sigma}_{\mathrm{PhC}}$ of $\widetilde{\mathcal{A}}_{\mathrm{PhC}}$ satisfies

$$\widetilde{\sigma}_{\mathrm{PhC}} = \bigcup_{\mathbf{k} \in B_{2\mathrm{d}}} \sigma_{\mathrm{PhC}}(\mathbf{k}).$$

The spectrum $\widetilde{\sigma}_{\mathrm{PhC}}$ can have gaps, that correspond to the band gaps of the PhC, i. e. intervals of frequencies for which light does not propagate in the PhC.

At this point it seems necessary to distinguish between the meanings of the values ω^2 and ω. While the former is the linear eigenvalue of (2.13) and hence, is by definition an element of the spectrum $\sigma_{\mathrm{PhC}}(\mathbf{k})$ of $\mathcal{A}_{\mathrm{PhC}}(\mathbf{k})$, the latter is the frequency and thus, the relevant quantity in sense of physics. Since we introduced a band gap as being a frequency interval for which no light propagates in the PhC, we shall distinguish in this thesis between a gap of the spectrum and a band gap, i. e. if ω^2 is in the gap of the spectrum, ω is in the band gap. In this respect, only the function $\omega_n(\mathbf{k})$ shall be called *band function*, while in literature also $\omega_n^2(\mathbf{k})$ can be denoted by this name.

The band functions $\omega_n(\mathbf{k})$ show several mirror symmetries in the Brillouin zone $B_{2\mathrm{d}}$ allowing for a successive reduction of the Brillouin zone to the so-called *irreducible Brillouin zone* $\widehat{B}_{2\mathrm{d}}$ [JJWM08], see the red triangles in Figures 2.1b and 2.2b for the irreducible Brillouin related to PhCs with square and hexagonal lattice, respectively.

It is widely accepted that for almost all PhC lattices that are of interest, it is sufficient to follow the band functions along the edges of the irreducible Brillouin zone $\widehat{B}$ in order to compute the spectrum and thus, the band gaps of PhCs [JJWM08], see for example the spectrum and the complete band gap of a 2d PhC with hexagonal lattice that we already illustrated in Figure 1.1. However, this cannot be guaranteed in general and in fact, it has be shown [HKSW07] that there exist academic counterexamples.

Finally, let us present the variational formulation of the eigenvalue problem (2.13) of finding modes in 2d PhCs. Multiplying (2.13) with the complex conjugate of some periodic and smooth test function v, integrating over C using integration by parts, where we take the periodicity of u and v into account, and weakening the smoothness requirements on u accordingly, we find that the variational formulation of (2.13) reads: find $(\omega^2, \mathbf{k}) \in \mathbb{R}^+ \times B_{2\mathrm{d}}$ and a non-trivial $u \in \mathsf{H}^1_{\mathrm{p}}(C)$ such that

$$\int_C \alpha(\nabla + \mathrm{i}\mathbf{k})u \cdot (\nabla - \mathrm{i}\mathbf{k})\overline{v} - \omega^2 \beta u \overline{v} \, \mathrm{d}\mathbf{x} = 0 \tag{2.14}$$

for all test functions $v \in \mathsf{H}^1_{\mathrm{p}}(C)$, where we can choose $\mathsf{H}^1_{\mathrm{p}}(C)$ as the space of test functions due to a density argument of the space of periodic and smooth functions in $\mathsf{H}^1_{\mathrm{p}}(C)$. Using the sesquilinear forms

$$\mathfrak{a}_C^{\alpha}(u, v) := \int_C \alpha \nabla u \cdot \nabla \overline{v} \, \mathrm{d}\mathbf{x}, \tag{2.15a}$$

$$\mathfrak{c}_C^{\alpha, i}(u, v) := \int_C \mathrm{i}\alpha \left(u(\partial_i \overline{v}) - (\partial_i u)\overline{v} \right) \, \mathrm{d}\mathbf{x}, \qquad i = 1, 2, \tag{2.15b}$$

$$\mathfrak{m}_C^{\alpha}(u, v) := \int_C \alpha u \overline{v} \, \mathrm{d}\mathbf{x}, \tag{2.15c}$$

$$\mathfrak{m}_C^{\beta}(u, v) := \int_C \beta u \overline{v} \, \mathrm{d}\mathbf{x}, \tag{2.15d}$$

and

$$\mathfrak{b}_C(u,v;\omega,\mathbf{k}) := \mathfrak{a}_C^{\alpha}(u,v) + k_1\mathfrak{c}_C^{\alpha,1}(u,v) + k_2\mathfrak{c}_C^{\alpha,2}(u,v) + |\mathbf{k}|^2\mathfrak{m}_C^{\alpha}(u,v) - \omega^2\mathfrak{m}_C^{\beta}(u,v), \tag{2.15e}$$

the variational formulation (2.14) can be rewritten in the form: find $(\omega^2,\mathbf{k}) \in \mathbb{R}^+ \times B_{2\mathrm{d}}$ and a non-trivial $u \in \mathsf{H}^1_{\mathrm{p}}(C)$ such that

$$\mathfrak{b}_C(u,v;\omega,\mathbf{k}) = 0 \tag{2.16}$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$.

With this understanding of 2d PhCs, their mathematical modelling and related spectral properties, we can now address the problem of finding guided modes in the following section.

2.3 Guided modes in two-dimensional photonic crystal waveguides

2d PhC waveguides can be constructed by introducing a line defects in 2d PhCs. Usually these line defects are introduced by omitting one (PhC W1 waveguide), two (PhC W2 waveguide), or more rows of holes/rods. However, we shall give a more general description of the geometry of 2d PhC waveguides by a piecewise definition of their permittivities. As explained above, the permittivity $\varepsilon_{\mathrm{PhC}} : \mathbb{R}^2 \to \mathbb{R}^+$ of 2d PhCs is bounded from below and above, and satisfies the periodicity condition (2.10). Let the line defect be parallel to and centered at the x_1-axis, and let it have height $a^0_{22} > 0$. Moreover, let $\varepsilon_{\mathrm{defect}} : \mathbb{R}^2 \to \mathbb{R}^+$ denote the permittivity in the line defect that is bounded from below and above, and which is periodic in x_1-direction, i. e. it satisfies $\varepsilon_{\mathrm{defect}}(\mathbf{x}+\mathbf{a}_1) = \varepsilon_{\mathrm{defect}}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^2$, where $\mathbf{a}_1 = a_1(1,0)^{\mathrm{T}}$. Then we choose the piecewise definition

$$\varepsilon_{\mathrm{wg}}(\mathbf{x}) = \begin{cases} \varepsilon^-_{\mathrm{PhC}}(\mathbf{x}), & \text{if } x_2 < -\frac{a^0_{22}}{2}, \\ \varepsilon_{\mathrm{defect}}(\mathbf{x}), & \text{if } -\frac{a^0_{22}}{2} < x_2 < \frac{a^0_{22}}{2}, \\ \varepsilon^+_{\mathrm{PhC}}(\mathbf{x}), & \text{if } x_2 > \frac{a^0_{22}}{2}, \end{cases} \tag{2.17}$$

for the permittivity $\varepsilon_{\mathrm{wg}} : \mathbb{R}^2 \to \mathbb{R}^+$ of a 2d PhC waveguide, where $\varepsilon^{\pm}_{\mathrm{PhC}}$ are two possibly disjoint permittivity functions that each satisfy the periodicity condition (2.10) with possibly different lattice vectors $\mathbf{a}^{\pm}_2$ but matching lattice vectors $\mathbf{a}^{\pm}_1 = a_1(1,0)^{\mathrm{T}}$, see Figure 2.3a. The vector $\mathbf{a}^0_2 = (a^0_{21}, a^0_{22})^{\mathrm{T}}$ can be chosen to match the choices of the unit cells $C^{\pm}_n$, $n \in \mathbb{N}$, on top and bottom of the defect.

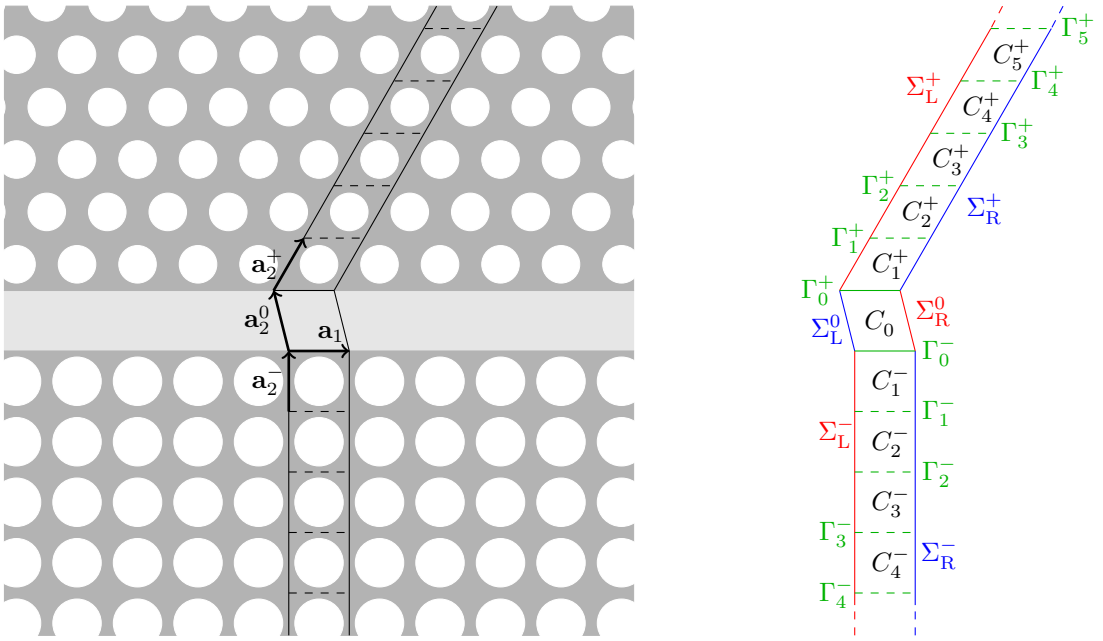
As explained in the previous section, the spectrum of PhCs can exhibit band gaps, i. e. frequency intervals in which there are no eigenvalues of (2.13) for all values of the quasi-momentum. In PhC waveguides there can exist guided modes, which are eigensolutions of the time-harmonic Maxwell's equations and which propagate along the line defect (i. e. along the x_1-axis) while decaying in the directions orthogonal to the line defect (i. e. along the x_2-axis).

As discussed in Section 2.1 the time-harmonic Maxwell's equations decouple in 2d into a TM mode and a TE mode that satisfy a 2d scalar Helmholtz equation. Considering both modes we choose (2.8) as governing equation, i. e.

$$-\nabla\cdot\alpha(\mathbf{x})\nabla U(\mathbf{x}) - \omega^2\beta(\mathbf{x})U(\mathbf{x}) = 0, \qquad \mathbf{x} \in \mathbb{R}^2,$$

where in the TM mode U describes the electric field in x_3-direction and the coefficients α and β are given by $\alpha(\mathbf{x}) = 1$ and $\beta(\mathbf{x}) = c^{-2}\varepsilon_{\mathrm{wg}}(\mathbf{x})$, cf. Eq. (2.9a). On the other hand, in the TE mode, U denotes the magnetic field in x_3-direction and the coefficients are defined by $\alpha(\mathbf{x}) = \varepsilon^{-1}_{\mathrm{wg}}(\mathbf{x})$ and $\beta(\mathbf{x}) = c^{-2}$, cf. Eq. (2.9b). Note that for simplicity of notation, we do not use the subscript "wg" for the coefficients α and β that are defined using the permittivity $\varepsilon_{\mathrm{wg}}$ of the PhC waveguide.

Due to the periodicity of the coefficients α and β in x_1-direction, we can again apply the Floquet theory. However, in contrast to the case of PhCs in the previous section, the periodicity is broken in the direction of $\mathbf{a}_2$ and hence, the Floquet transformation is only one-dimensional. Revisiting all steps in Section 2.2 while neglecting all parts related to the periodicity of the permittivity in the direction of $\mathbf{a}_2$, we find that we can transform the problem in $\mathbb{R}^2$ into a family of problems in the infinite strip



(a) Sketch of a PhC waveguide with homogeneous line defect, semi-infinite PhC of square lattice below the defect, and semi-infinite PhC of hexagonal lattice on top of the defect.

(b) Notation of cells and boundaries of the periodicity strip S.

Figure 2.3: Sketch of a PhC waveguide and its periodicity strip S.

S illustrated in Figure 2.3. Introducing the 1d Brillouin zone $B = [-\frac{\pi}{a_1}, \frac{\pi}{a_1}]$, and the quasi-momentum $k \in B$, the problem reads

$$-\nabla \cdot \alpha(\mathbf{x}) \nabla \widetilde{U}(\mathbf{x}, k) - \omega^2 \beta(\mathbf{x}) \widetilde{U}(\mathbf{x}, k) = 0, \qquad \mathbf{x} \in S, \tag{2.18a}$$

with quasi periodic boundary conditions

$$\widetilde{U}(\cdot, k)|_{\Sigma_\mathrm{R}} = \mathrm{e}^{\mathrm{i}ka_1} \widetilde{U}(\cdot, k)|_{\Sigma_\mathrm{L}}, \tag{2.18b}$$

$$\alpha\, \partial_{\mathbf{n}_\mathrm{R}} \widetilde{U}(\cdot, k)|_{\Sigma_\mathrm{R}} = -\mathrm{e}^{\mathrm{i}ka_1} \alpha\, \partial_{\mathbf{n}_\mathrm{L}} \widetilde{U}(\cdot, k)|_{\Sigma_\mathrm{L}} \tag{2.18c}$$

on the left $\Sigma_\mathrm{L} = \Sigma_\mathrm{L}^+ \cup \Sigma_\mathrm{L}^0 \cup \Sigma_\mathrm{L}^- \subset \partial S$ and right $\Sigma_\mathrm{R} = \Sigma_\mathrm{R}^+ \cup \Sigma_\mathrm{R}^0 \cup \Sigma_\mathrm{R}^- \subset \partial S$ boundaries of S. A guided mode is, by definition, a non trivial solution of (2.18) that satisfies a decay condition for $|x_2| \to \infty$.

Similarly to the spaces $\mathsf{H}^1_\mathrm{p}(C)$ and $\mathsf{H}^1_\mathrm{p}(\Delta, C, \alpha_{\mathrm{PhC}})$ of periodic functions in C, we define the function spaces

$$\mathsf{H}^1_{1\mathrm{p}}(S) := \left\{ u \in \mathsf{H}^1(S) \text{ with } u|_{\Sigma_\mathrm{L}} = u|_{\Sigma_\mathrm{R}} \right\},$$

whose functions implicitly satisfy a decay condition for $|x_2| \to \infty$, and

$$\mathsf{H}^1_{1\mathrm{p}}(\Delta, S, \alpha) := \left\{ u \in \mathsf{H}^1(\Delta, S) \cap \mathsf{H}^1_{1\mathrm{p}}(S) \text{ with } \alpha \partial_{\mathbf{n}_\mathrm{L}} u|_{\Sigma_\mathrm{L}} = -\alpha \partial_{\mathbf{n}_\mathrm{R}} u|_{\Sigma_\mathrm{R}} \right\}$$

of functions in S that are periodic in x_1-direction.

With these definitions and the substitution $\widetilde{U}(\mathbf{x}, k) = \mathrm{e}^{\mathrm{i}kx_1} u(\mathbf{x}, k)$, the eigenvalue problem (2.18) of finding guided modes is equivalent to: find couples $(\omega^2, k) \in \mathbb{R}^+ \times B$ such that there exists a non-trivial $u \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, S, \alpha)$ that satisfies

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha (\nabla + \mathrm{i}k\binom{1}{0}) u - \omega^2 \beta u = 0 \qquad \text{in } S. \tag{2.19}$$

Then we call (ω^2, k) an *eigenvalue couple* of (2.19) with associated eigenmode u. We shall distinguish two different problem formulations:

- the ω-*formulation*, where we fix $k \in B$ which yields a linear eigenvalue problem in ω^2, and
- the k-*formulation*, where we fix $\omega \in \mathbb{R}^+$ and obtain a quadratic eigenvalue problem in k.

However, note that this problem is posed in the unbounded domain S. Therefore, we will later in Chapters 6 and 7 introduce DtN and RtR transparent boundary conditions to truncate the infinite domain S. More precisely, let $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ be the Dirichlet trace spaces of $\mathsf{H}^1_{1\mathrm{p}}(\Delta, C_0, \alpha)$ on $\Gamma_0^\pm$ and let $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ be the corresponding dual spaces. Then we will show in Chapter 6 that under certain assumptions there exist linear DtN maps $\mathcal{D}^\pm(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm), \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm))$ such that the eigenvalue problem (2.19) is equivalent to: find $(\omega^2, k) \in \mathbb{R}^+ \times B$ and a non-trivial $u \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, C_0, \alpha)$ that satisfies

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u - \omega^2\beta u = 0 \qquad \text{in } C_0, \tag{2.20a}$$

$$\pm\alpha\partial_2 u = \mathcal{D}^\pm(\omega, k)\, u \qquad \text{on } \Gamma_0^\pm. \tag{2.20b}$$

On the other hand, in Chapter 7 we will show that the eigenvalue problem (2.19) is equivalent to: find couples $(\omega^2, k) \in \mathbb{R}^+ \times B$ such that there exists a non-trivial $u \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, C_0, \alpha)$ that satisfies

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u - \omega^2\beta u = 0 \qquad \text{in } C_0, \tag{2.21a}$$

$$(\mp\alpha\partial_2 + \mathrm{i}\rho)u = \mathcal{R}^\pm(\omega, k)(\pm\alpha\partial_2 + \mathrm{i}\rho)u \qquad \text{on } \Gamma_0^\pm, \tag{2.21b}$$

where $\mathcal{R}^\pm(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm))$ are linear RtR maps and $\rho \in \mathbb{R} \setminus \{0\}$ is an arbitrary, real, nonzero constant.

In the remainder of this section we will summarize the results of the spectral theory for the eigenvalue problem (2.19) in ω-formulation. To this end, we introduce the operator

$$\mathcal{A}(k) := -\frac{1}{\beta}(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0}),$$

$k \in B$, defined on the function space $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S, \alpha)$ of the whole strip S, that is related to the eigenvalue problem (2.19) with permittivity $\varepsilon_{\mathrm{wg}}$ as defined in Eq. (2.17). Furthermore, we introduce the operators

$$\mathcal{A}^\pm(k) := -\frac{1}{\beta^\pm}(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha^\pm(\nabla + \mathrm{i}k\binom{1}{0}),$$

$k \in B$, defined on $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$, that are related to eigenvalue problems of the form (2.19) posed in the infinite half strips $S^\pm$ with perfectly periodic permittivities $\varepsilon^\pm_{\mathrm{PhC}}$ of the top and bottom PhC respectively. Let $\sigma^\pm_{\mathrm{PhC}}(\mathbf{k})$ denote the spectra of the operators $\mathcal{A}^\pm_{\mathrm{PhC}}(\mathbf{k})$ related to the eigenvalue problem (2.13) when replacing the coefficients α and β by $\alpha^\pm$ and $\beta^\pm$ that correspond to the permittivities $\varepsilon^\pm_{\mathrm{PhC}}$ of the top and bottom PhCs of the waveguide. Then the spectra of the operators $\mathcal{A}^\pm(k)$ are connected to the spectra $\sigma^\pm_{\mathrm{PhC}}(\mathbf{k})$ through

$$\sigma^\pm(k) := \sigma(\mathcal{A}^\pm(k)) = \bigcup_{k_2 \in]-\frac{\pi}{a_2}, \frac{\pi}{a_2}[} \sigma^\pm_{\mathrm{PhC}}(k, k_2).$$

The following results were shown in [Fli13] for the TM mode operator, i. e. for $\alpha \equiv 1$. However, using the same arguments as in [Kuc93, FK97] for operators with perfectly periodic coefficients, in particular for the TE mode operator called *acoustic operator* in [FK97], and applying the Weyl theorem [RS78], we can show

Proposition 2.2. *For all $k \in B$ the operator $\mathcal{A}(k)$ is self-adjoint and positive, i. e. its spectrum satisfies $\sigma(k) \subset \mathbb{R}^+$. Moreover, its essential spectrum $\sigma^{\mathrm{ess}}(k)$, i. e. the spectrum $\sigma(k)$ minus its isolated eigenvalues, satisfies*

$$\sigma^{\mathrm{ess}}(k) = \sigma^+(k) \cup \sigma^-(k)$$

with

$$\sigma^\pm(k) = \mathbb{R}^+ \setminus \bigcup_{n=1}^{N^\pm(k)} G_n^\pm(k),$$

where the gaps $G_n^\pm(k) \subset \mathbb{R}^+$ are open intervals and $N^\pm(k) \in \mathbb{N}_0$ is the number of band gaps.

According to a result [Vor11] on the Bethe-Sommerfeld conjecture [BS67] for periodic Maxwell operators in 2d, we note that the numbers $N^{\pm}(k)$ of band gaps are finite.

We conclude that there exists a number $N(k) \in \mathbb{N}_0$ of gaps $G_n(k) \subset \mathbb{R}^+$, $n = 1, \ldots, N(k)$, such that

$$\sigma^{\mathrm{ess}}(k) = \mathbb{R}^+ \setminus \bigcup_{n=1}^{N(k)} G_n(k),$$

with the set of gaps $\bigcup_{n=1}^{N(k)} G_n(k) = \bigcup_{n=1}^{N^+(k)} G_n^+(k) \cap \bigcup_{n=1}^{N^-(k)} G_n^-(k)$.

Using the theory of self-adjoint operators [RS78], we deduce

Proposition 2.3. *Inside the gaps $G_n(k)$, $n = 1, \ldots, N(k)$, $k \in B$, there exist only isolated eigenvalues of finite multiplicity, which can only accumulate at the boundaries of the gaps $G_n(k)$.*

Let the isolated eigenvalues $\omega_m^2(k) \in \mathbb{R}^+$, $m = 1, \ldots, M(k)$, of $\mathcal{A}(k)$ inside the gaps $G_n(k)$, $n = 1, \ldots, N(k)$, be ordered such that

$$0 < \omega_1^2(k) \leq \ldots \leq \omega_{M(k)}^2(k) \tag{2.22}$$

with $0 \leq M(k) \leq \infty$. Then the functions

$$k \longmapsto \omega_m^2(k)$$

are $\frac{2\pi}{a_1}$-periodic, even and continuous in k [Fli13]. Note that they are not necessarily continuous in B since the number $M(k)$ of eigenvalues $\omega_m^2(k)$ is not constant in B. Using the fact that the self-adjoint operator $\mathcal{A}(k)$ is analytic with respect to the quasi-momentum k, its domain $\mathsf{H}^1_{\mathrm{lp}}(\Delta, S^{\pm}, \alpha)$ is independent of the quasi-momentum k, and the eigenvalues of $\mathcal{A}(k)$ in the band gaps $G_n(k)$ are isolated and have finite multiplicity, we can apply the analytic perturbation theory for self-adjoint, linear operators, see Chapter 7 in [Kat95], and prove a result that is fundamental for the numerical procedures, that we will develop in this thesis.

Theorem 2.4. *Let $M = \max\{M(k) : k \in B\}$. Then there exists a mapping $j(\cdot\,; k) : \{1, \ldots, M\} \to \{1, \ldots, M(k)\}$, for all $k \in B$ such that the functions*

$$k \longmapsto \omega_j(k)$$

are analytic. These functions are called dispersion curves*, or* band functions*. Moreover, the magnitude and phase of the corresponding eigenmodes $u_j(\cdot\,; k)$ can be chosen such that the eigenmodes are also analytic with respect to the quasi-momentum k.*

Note that Theorem 2.4 is valid for all $k \in B$, i. e. the dispersion curves are also analytic at crossings.

It is well known that the first dispersion curve is not analytic in $k = 0$, see for example the band structures in Figures 2.7 and 2.9, that we will present in Section 2.6. Note that this case is explicitly excluded in Theorem 2.4 due to the assumption that the eigenvalues $\omega_j^2(k)$ are positive, see Eq. 2.22.

Finally, we shall state a result related to the eigenfunctions of $\mathcal{A}(k)$ proven in [FK97].

Proposition 2.5. *Let $k \in B$ and let $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$ be an eigenvalue of (2.19). Then the associated eigenfunction $u \in \mathsf{H}^1_{\mathrm{lp}}(\Delta, S, \alpha)$ decays exponentially with $|x_2|$, where the decay rate is proportional to the distance of the eigenvalue ω^2 to the essential spectrum $\sigma^{\mathrm{ess}}(k)$.*

This motivates the notion of the *confinement* of guided modes. We speak of *well-confined* modes, if the decay rate is large, which is in the sense of Proposition 2.5 equivalent to the distance to the essential spectrum. Thus, guided modes close to the *band edge*, i. e. the boundary of essential spectrum and band gap, are not well-confined.

Proposition 2.5 gives the mathematical justification for the supercell method, which we will explain in the following section.

2.4 Model reduction using the supercell approach

Before we will elaborate in Chapters 6 and 7 on the definition and numerical realization of DtN and RtR transparent boundary conditions that are employed for the transformation of the eigenvalue problem (2.19) in the infinite strip S to the eigenvalue problems (2.20) and (2.21) in the defect cell C_0, we will now briefly discuss the frequently used *supercell method*. The supercell method provides access to approximations of guided modes. Based on the observation in Proposition 2.5, that guided modes decay exponentially with $|x_2| \to \infty$, the eigenvalue problem (2.19) is posed in a bounded supercell $S_n \subset S$ instead of the infinite strip S. The computational domain S_n is obtained by simply cutting the infinite strip S after $n \in \mathbb{N}$ periodicity cells of the PhCs on top and bottom, and prescribing periodic boundary conditions at the top and bottom boundaries $\Sigma_{\mathrm{T}} := \Gamma_n^+$ and $\Sigma_{\mathrm{L}} := \Gamma_n^-$. Thus, the problem solved reads: find couples $(\omega^2, k) \in \mathbb{R}^+ \times B$ such that there exists a non-trivial $u \in \mathsf{H}^1_{\mathrm{p}}(\Delta, S_n, \alpha)$ that satisfies

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u - \omega^2 \beta u = 0, \qquad \text{in } S_n. \tag{2.23}$$

Since guided modes decay exponentially as $|x_2| \to \infty$, cf. Proposition 2.5, it can be expected that the modelling error, that is introduced when prescribing periodic boundary conditions after a certain number of holes, is reasonably small. In fact, Soussi [Sou05] showed that the solutions of the supercell method converge exponentially towards the solutions of the exact problem (2.19) if the number n of periodicity cells, that are included in the supercell, tends to infinity. However, as we shall demonstrate in Chapter 6, where we present a comparison of the results of the supercell method with the results of the problem (2.20) with DtN transparent boundary conditions, it is a not possible to control the modelling error of the supercell method for a fixed the number n of periodicity cells, if no a priori knowledge about the confinement of the guided mode is available.

The eigenvalue problem (2.23) of the supercell method is again linear in ω^2 (ω-formulation) and quadratic in k (k-formulation) allowing for standard numerical techniques of PhC band structure calculations to be applied [SK10]. However, note that the eigenvalue problem (2.23) of the supercell method has eigenvalues also inside the essential spectrum $\sigma^{\mathrm{ess}}(k)$ of $\mathcal{A}(k)$. These eigenvalues have to be excluded. To this end, one needs to have access to the essential spectrum, i. e. a full computation of the spectra $\sigma^{\pm}(k) = \sigma(\mathcal{A}^{\pm}(k))$ of the operators related to the PhCs on top and bottom of the guide is needed.

It is important to note that we may not replace the periodic boundary conditions on Σ_{T} and Σ_{B} by, e. g. homogeneous Dirichlet boundary conditions, even though we know that guided modes decay exponentially as $|x_2| \to \infty$. The reason is that homogeneous Dirichlet boundary conditions will produce spurious modes, so called *surface modes*, that are confined at the boundaries Σ_{T} and Σ_{B}, and that may — like guided modes — appear outside the essential spectrum in the band structure calculations and hence, cannot be distinguished from guided modes without taking the mode profile into account.

Note that the number of eigenvalues inside the essential spectrum grows with the number of periodicity cells that are included in the supercell. Using an iterative eigenvalue solver, this implies that one should restrict the eigenvalue computation — as far as possible — to the band gap by a shift and invert strategy, in order not to spoil the performance of the iterative solver.

Similarly to Theorem 2.4, we can state a result on the eigenvalues and corresponding eigenmodes of (2.23). Considering that the differential operator related to (2.23) is self-adjoint and analytic with respect to the quasi-momentum k, and that its domain $\mathsf{H}^1_{\mathrm{p}}(\Delta, S_\alpha)$ is independent of the quasi-momentum, we can again use the theory presented in [Kat95] and show

Theorem 2.6. *For all $k \in B$ the eigenvalues $\omega_j^2(k) \in \mathbb{R}^+$, $j \in \mathbb{N}$, of (2.23) can be ordered such that the* dispersion curves

$$k \longmapsto \omega_j(k)$$

are analytic. Moreover, the magnitude and phase of the corresponding eigenmodes $u_j(\cdot\,; k)$ can be chosen such that the eigenmodes are also analytic with respect to the quasi-momentum k.

Finally, we present the variational formulation of (2.23). Analogously to the procedure applied to derive the variational formulation (2.14) of the eigenvalue problem (2.13) of finding modes in 2d PhCs,

we arrive at the following formulation: find $(\omega^2, k) \in \mathbb{R}^+ \times B$ and a non-trivial $u \in \mathsf{H}^1_{\mathrm{p}}(S_n)$ such that

$$\int_{S_n} \alpha(\nabla + \mathrm{i}k\binom{1}{0})u \cdot (\nabla - \mathrm{i}k\binom{1}{0})\overline{v} - \omega^2 \beta u \overline{v} \,\mathrm{d}\mathbf{x} = 0 \tag{2.24}$$

for all test functions $v \in \mathsf{H}^1_{\mathrm{p}}(S_n)$, or, using the sesquilinear forms

$$\mathfrak{a}^{\alpha}_{S_n}(u,v) := \int_{S_n} \alpha \nabla u \cdot \nabla \overline{v} \,\mathrm{d}\mathbf{x}, \tag{2.25a}$$

$$\mathfrak{c}^{\alpha,1}_{S_n}(u,v) := \int_{S_n} \mathrm{i}\alpha \left(u(\partial_1 \overline{v}) - (\partial_1 u)\overline{v}\right) \,\mathrm{d}\mathbf{x}, \tag{2.25b}$$

$$\mathfrak{m}^{\alpha}_{S_n}(u,v) := \int_{S_n} \alpha u \overline{v} \,\mathrm{d}\mathbf{x}, \tag{2.25c}$$

$$\mathfrak{m}^{\beta}_{S_n}(u,v) := \int_{S_n} \beta u \overline{v} \,\mathrm{d}\mathbf{x}, \tag{2.25d}$$

and

$$\mathfrak{b}_{S_n}(u,v;\omega,k) := \mathfrak{a}^{\alpha}_{S_n}(u,v) + k\mathfrak{c}^{\alpha,1}_{S_n}(u,v) + k^2\mathfrak{m}^{\alpha}_{S_n}(u,v) - \omega^2\mathfrak{m}^{\beta}_{S_n}(u,v), \tag{2.25e}$$

we can write: find $(\omega^2, k) \in \mathbb{R}^+ \times B$ and a non-trivial $u \in \mathsf{H}^1_{\mathrm{p}}(S_n)$ such that

$$\mathfrak{b}_{S_n}(u,v;\omega,k) = 0 \tag{2.26}$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(S_n)$.

2.5 High-order finite element discretization

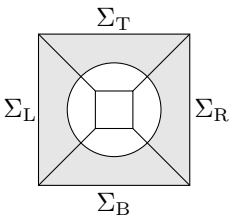
This section is dedicated to the spatial discretization of the variational formulations (2.14) and (2.24) of the eigenvalue problem (2.13) of finding modes in 2d PhCs, and the eigenvalue problem (2.23) in the supercell S_n. As elaborated in the introduction, we employ the *finite element method* (FEM), or shortly, *finite elements* (FE).

The FEM provides discrete subspaces for Sobolev spaces involved in variational formulations. For the variational formulations (2.14) and (2.24) we need to provide FE space of functions that are periodic. To this end, we first elaborate on the FE meshes of the domains C and S_n, respectively.

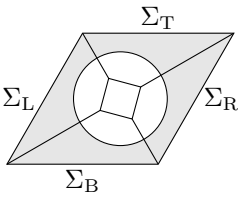
Let the domains C and S_n be partitioned into non-overlapping possibly curved, triangular or quadrilateral subdomains, the *geometrical cells*. Each geometrical cell K is defined as a smooth map F_K of the reference cell $\widehat{K}(K)$, which is for triangular cells the convex hull of the points $(0,0)$, $(1,0)$ and $(0,1)$ and for quadrilaterals the square $[0,1]^2$. The sets of the geometrical cells, the *meshes*, are denoted by $\mathfrak{M}(C)$ and $\mathfrak{M}(S_n)$, see for example the coarse meshes of PhC unit cells with curved, quadrilateral cells in Figure 2.4 and the mesh of a supercell in Figure 2.5. These meshes are assumed to be periodic in direction $\mathbf{a}_1$, i. e. for each edge of a geometrical cell on the left boundary Σ_{L} there is an edge on the right boundary Σ_{R}, which is only shifted by $\mathbf{a}_1$. In particular, this means that the corresponding geometrical cells need to have the same parameterization on the boundaries Σ_{L} and Σ_{R}. Moreover, we assume the mesh $\mathfrak{M}(C)$ of the PhC unit cell to be periodic in direction $\mathbf{a}_2$, i. e. for each edge of a geometrical cell on the bottom boundary Σ_{B} there exists an edge on the top boundary Σ_{T}, which is shifted by $\mathbf{a}_2$. On the other hand, the mesh $\mathfrak{M}(S_n)$ is assumed to be periodic in direction $\mathbf{a}_2^0 + n(\mathbf{a}_2^+ + \mathbf{a}_2^-)$, i. e. for every edge on the bottom boundary $\Sigma_{\mathrm{B}} := \Gamma_n^-$ there is an edge on the top boundary $\Sigma_{\mathrm{T}} := \Gamma_n^+$, that is shifted by $\mathbf{a}_2^0 + n(\mathbf{a}_2^+ + \mathbf{a}_2^-)$.

Based on the meshes $\mathfrak{M}(C)$ and $\mathfrak{M}(S_n)$ we define discrete subspaces of $\mathsf{H}^1_{\mathrm{p}}(C)$ and $\mathsf{H}^1_{\mathrm{p}}(S_n)$ as

$$\mathsf{S}^p_{\mathrm{p}}(\Omega) := \{v \in \mathsf{H}^1_{\mathrm{p}}(\Omega) \cap \mathsf{C}^0(\overline{\Omega}) : v|_K \circ F_K \in \mathsf{P}^p(\widehat{K}(K)) \;\; \forall K \in \mathfrak{M}(\Omega)\},$$



(a) Unit cell of PhC with square lattice.



(b) Unit cell of PhC with hexagonal lattice.

Figure 2.4: Meshes with curved, quadrilateral cells of PhC unit cells C with top, bottom, left and right boundaries $\Sigma_T, \Sigma_B, \Sigma_L, \Sigma_R \subset \partial C$.

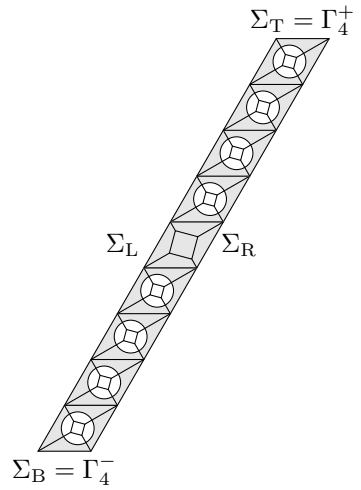


Figure 2.5: Mesh with curved, quadrilateral cells of a supercell S_4 with top, bottom, left and right boundaries $\Sigma_T, \Sigma_B, \Sigma_L, \Sigma_R \subset \partial S_4$ of a PhC W1 waveguide with hexagonal lattice.

where p is a chosen polynomial degree, Ω stands for one of the computational domains C or S_n, and $\mathsf{P}^p(\widehat{K})$ is the space of polynomials with maximal (total) degree p

$$\mathsf{P}^p(\widehat{K}) = \begin{cases} \operatorname{span}\{\widehat{x}_1^\ell \widehat{x}_2^m \mid 0 \le \max\{\ell, m\} \le p\}, & \text{if } \widehat{K} \text{ is a quadrilateral,} \\ \operatorname{span}\{\widehat{x}_1^\ell \widehat{x}_2^m \mid 0 \le \ell + m \le p\}, & \text{if } \widehat{K} \text{ is a triangle,} \\ \operatorname{span}\{\widehat{x}^\ell \quad\; \mid 0 \le \ell \le p\}, & \text{if } \widehat{K} \text{ is an interval.} \end{cases} \tag{2.27}$$

Note that the last case in (2.27) is not needed for the FE spaces introduced above, but it is useful for the FE spaces that we will introduce later in Section 6.1.5 for the discretization of the variational formulations with DtN and RtR transparent boundary conditions.

If the maximal polynomial degree is $p = 1$ the basis functions of $\mathsf{S}_{\mathrm{p}}^p(\Omega)$ are *hat functions*, that take the value one at a single node of the mesh $\mathfrak{M}(\Omega)$ and that vanish at all other nodes. In this case we speak of *linear FEM*. For polynomial degrees larger than one the FEM is said to be of *high-order* [Sch98]. Besides the hat functions, the basis of a high-order FEM consists of functions, the so-called *edge functions*, that can be identified to an edge and that vanish on the closure of all other edges in the mesh. Furthermore, in 2d, high-order FE bases comprise so-called *bubble functions*, that are identified to one cell of the mesh and that vanish in the closure of all other cells.

Let

$$N(C) := \dim \mathsf{S}_{\mathrm{p}}^p(C),$$
$$N(S_n) := \dim \mathsf{S}_{\mathrm{p}}^p(S_n)$$

denote the dimensions of the FE spaces $\mathsf{S}_{\mathrm{p}}^p(C)$ and $\mathsf{S}_{\mathrm{p}}^p(S_n)$. Furthermore, let

$$b_{C,1}, \ldots, b_{C,N(C)} \in \mathsf{S}_{\mathrm{p}}^p(C),$$
$$b_{S_n,1}, \ldots, b_{S_n,N(S_n)} \in \mathsf{S}_{\mathrm{p}}^p(S_n)$$

denote the basis functions of $\mathsf{S}_{\mathrm{p}}^p(C)$ and $\mathsf{S}_{\mathrm{p}}^p(S_n)$, respectively. Then the discrete form of the variational formulation (2.14) of the eigenvalue problem (2.13) of finding modes in 2d PhCs can be written in the form: find $(\omega^2, \mathbf{k}) \in \mathbb{R}^+ \times B_{2\mathrm{d}}$ and a non-trivial $\mathbf{u} \in \mathbb{C}^{N(C)} \setminus \{\mathbf{0}\}$ such that

$$\left(\mathbf{A}_C^\alpha + k_1 \mathbf{C}_C^{\alpha,1} + k_2 \mathbf{C}_C^{\alpha,2} + |\mathbf{k}|^2 \mathbf{M}_C^\alpha - \omega^2 \mathbf{M}_C^\beta\right) \mathbf{u} = \mathbf{0}, \tag{2.28}$$

where the real symmetric matrices $\mathbf{A}_C^\alpha, \mathbf{M}_C^\alpha, \mathbf{M}_C^\beta \in \mathbb{R}^{N(C)\times N(C)}$ and the purely imaginary, Hermitian matrices $\mathbf{C}_C^{\alpha,1}, \mathbf{C}_C^{\alpha,2} \in \mathrm{i}\mathbb{R}^{N(C)\times N(C)}$ have entries

$$\begin{aligned} A_{C,ij}^\alpha &= \mathfrak{a}_C^\alpha(b_{C,j}, b_{C,i}), \\ C_{C,ij}^{\alpha,1} &= \mathfrak{c}_C^{\alpha,1}(b_{C,j}, b_{C,i}), \\ C_{C,ij}^{\alpha,2} &= \mathfrak{c}_C^{\alpha,2}(b_{C,j}, b_{C,i}), \\ M_{C,ij}^\alpha &= \mathfrak{m}_C^\alpha(b_{C,j}, b_{C,i}), \\ M_{C,ij}^\beta &= \mathfrak{m}_C^\beta(b_{C,j}, b_{C,i}), \end{aligned}$$

$i, j = 1, \ldots, N(C)$, with the sesquilinear forms as given in Eq. (2.15). Similarly, the discrete form of the variational formulation (2.24) of the eigenvalue problem (2.23) of finding guided modes in the supercell S_n reads: find $(\omega^2, k) \in \mathbb{R}^+ \times B$ and a non-trivial $\mathbf{u} \in \mathbb{C}^{N(S_n)} \setminus \{\mathbf{0}\}$ such that

$$\left(\mathbf{A}_{S_n}^\alpha + k\mathbf{C}_{S_n}^{\alpha,1} + k^2\mathbf{M}_{S_n}^\alpha - \omega^2\mathbf{M}_{S_n}^\beta\right)\mathbf{u} = \mathbf{0}, \tag{2.29}$$

where the real, symmetric matrices $\mathbf{A}_{S_n}^\alpha, \mathbf{M}_{S_n}^\alpha, \mathbf{M}_{S_n}^\beta \in \mathbb{R}^{N(S_n)\times N(S_n)}$ and the purely imaginary, Hermitian matrix $\mathbf{C}_{S_n}^{\alpha,1} \in \mathrm{i}\mathbb{R}^{N(S_n)\times N(S_n)}$ have entries

$$\begin{aligned} A_{S_n,ij}^\alpha &= \mathfrak{a}_{S_n}^\alpha(b_{S_n,j}, b_{S_n,i}), \\ C_{S_n,ij}^{\alpha,1} &= \mathfrak{c}_{S_n}^{\alpha,1}(b_{S_n,j}, b_{S_n,i}), \\ M_{S_n,ij}^\alpha &= \mathfrak{m}_{S_n}^\alpha(b_{S_n,j}, b_{S_n,i}), \\ M_{S_n,ij}^\beta &= \mathfrak{m}_{S_n}^\beta(b_{S_n,j}, b_{S_n,i}), \end{aligned}$$

$i, j = 1, \ldots, N(S_n)$, with the sesquilinear forms as given in Eq. (2.25).

With the help of the FEM, we are able to compute approximations to the solutions of the eigenvalue problems (2.13) and (2.23) by solving the matrix eigenvalue problems (2.28) and (2.29). There are basically three strategies to improve the accuracy of an existing FE approximation:

- refining the mesh of the computational domain (h-FEM),
- increasing the polynomial degree of the basis functions (p-FEM), or
- a combination of both (hp-FEM).

While h-FEM provides algebraic convergence, p-FEM converges exponentially if the solution is analytic in subdomains that are resolved exactly by the cells of the mesh [Sch98]. This motivates the need of curved cells in our FE mesh in order to exactly resolve the holes/rods of the PhCs. A comprehensive study of the convergence of p-FEM in the context of PhC band structure calculations can be found in [SK09]. Note, that in case of non-smooth material boundaries we can extend the previous and following definitions to hp-adaptive FE spaces.

Concepts — A numerical C++ library for partial differential equations For the implementation of the high-order FEM we employ Concepts, which is a C++ library for the numerical solution of partial differential equations [Con15, FL02]. Originated from a software package for the boundary element method, Concepts has been extended with high-order FEM as well as discontinuous Galerkin methods. Concepts is based on concept-oriented design, i. e. mathematical concepts such as the FE meshes, the FE spaces, the bilinear forms, matrices and vectors are implemented as C++ classes. The object-oriented structure of the library allows the programmer to re-use these concepts in a very flexible way.

Concepts allows for FE meshes with curved cells, such that the circular holes of PhCs can be resolved perfectly, e. g. by the meshes sketched in Figures 2.4 and 2.5. For a detailed description of how curved cells are realized with Concepts, the reader is referred to [Sch08]. There is no upper bound for the polynomial degree of the FE spaces other than the prohibitively increasing condition number of the resulting FE matrices. This enables us to use p-FEM on the coarse meshes sketched above.

Since we assume the permittivity to be piecewise constant, it proves useful to compute separate FE matrices for the dielectric medium and its complement. These matrices can then be exported to Matlab's

binary `mat`-files using the Concepts's class `concepts::MatfileIO`. The matrices are loaded into Matlab with which all further computations, in particular eigenvalue computations, are done. For plotting FE solutions such as eigenvectors, the coefficient vectors, computed with Matlab, are again saved in `mat`-files which are imported into Concepts using `concepts::MatfileIO`. For this purpose the integration rule of the FE space is set to a rule that comprises endpoints, e. g. the trapezoidal rule, and a pointwise evaluation of the FE solution is performed at all quadrature points. This delivers three vectors, the vectors of the x_1- and x_2-components of the quadrature points as well as the vector of the values of the FE solution at the quadrature points. These vectors, together with information on the mesh, is exported to a `mat`-file, from which the data can be loaded into and plotted with Matlab. Both, the pointwise evaluation of the FE solution as well as the export to a `mat`-file is bundled in Concepts's class `graphics::MatlabBinaryGraphics`.

2.6 Examples

In this section we introduce examples of a PhC and a PhC waveguide, that we will use in the following chapters when numerically testing the proposed methods.

Example 1. We consider the TM mode in a PhC of square lattice with lattice constant a, holes of relative radius $\frac{r}{a} = 0.46$ and permittivity $\varepsilon = 1$ that are surrounded by dielectric material of permittivity $\varepsilon = 8$, see Figure 2.6. We shall only consider the Γ-X-interval of the irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$, i. e. we consider $k_1 \in \widehat{B} := [0, \frac{\pi}{a}]$ and $k_2 = 0$. In other words, the eigenvalue problem under consideration is equivalent to the supercell problem (2.23) when replacing the domain S_n by the unit cell C of the PhC described above and sketched in Figure 2.6. For illustration, the band structure of the TM mode along the Γ-X-interval $\widehat{B}$ is presented in Figure 2.7.

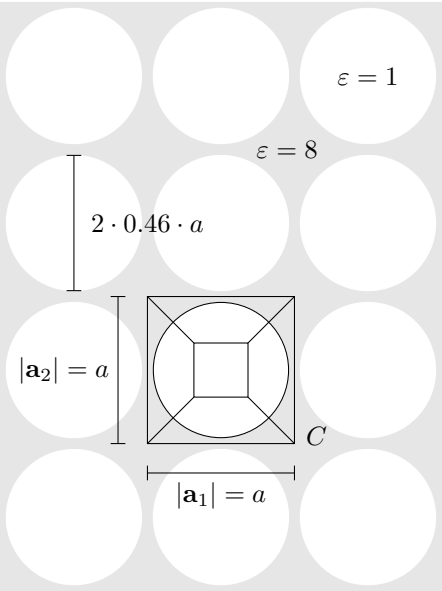


Figure 2.6: Sketch of the PhC with square lattice in Example 1 and its unit cell C with FE mesh of nine quadrilaterals.

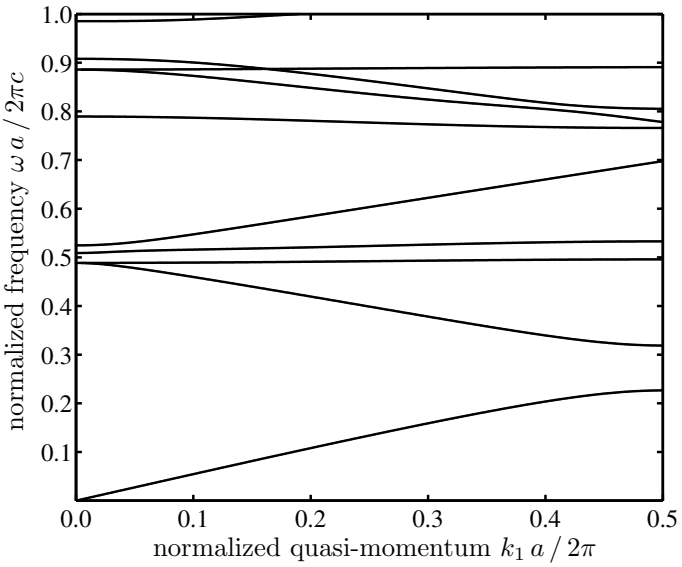


Figure 2.7: Band structure of the TM mode along the Γ-X-interval $\widehat{B} = [0, \frac{\pi}{a}]$ of the irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$ for the PhC of square lattice described in Example 1.

Example 2. We consider the TE mode in a PhC W1 waveguide with hexagonal lattice, i. e. with $\mathbf{a}_1^0 = \mathbf{a}_1^+ = \mathbf{a}_1^- = a_1(1,0)^{\mathrm{T}}$ and $\mathbf{a}_2^0 = \mathbf{a}_2^+ = \mathbf{a}_2^- = \frac{a_1}{2}(1,\sqrt{3})^{\mathrm{T}}$. The dielectric medium of the device has permittivity $\varepsilon = 11.4$ and holes of relative radius $\frac{r}{a_1} = 0.31$ with permittivity $\varepsilon = 1$. Figure 2.9 shows the band structure computed with the help of the supercell method using five periodicity cells on top and

bottom of the guide. The black lines represent the dispersion curves while frequencies with propagating PhC modes are shaded in grey. Note that the complete band gap of the 2d PhC was already shown in Figure 1.1.

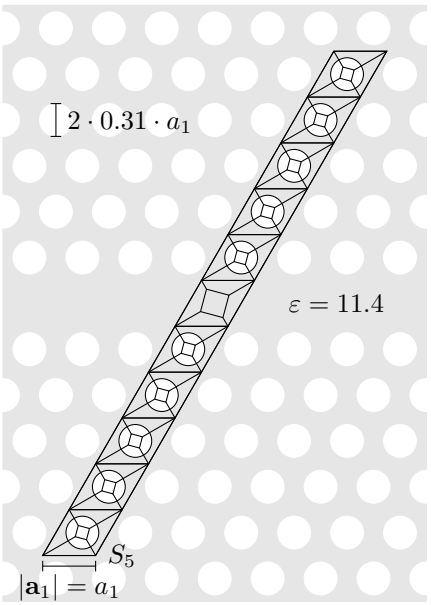


Figure 2.8: Sketch of the PhC W1 waveguide with hexagonal lattice of Example 2 and the supercell S_5 with FE mesh consisting of 95 quadrilaterals, that was used for the computation of the band structure in Figure 2.9.

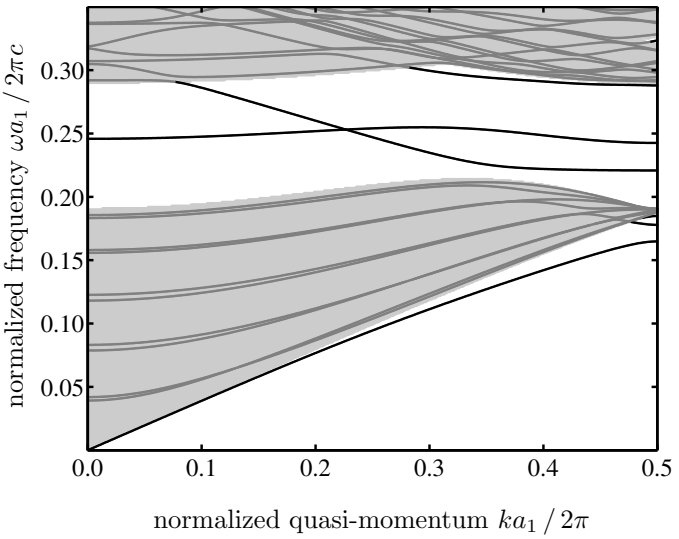


Figure 2.9: Approximation to the band structure of the TE mode in the PhC W1 waveguide of Example 2 using the supercell method with five unit cells on top and bottom of the defect cell C_0. Areas shaded in grey correspond to the essential spectrum $\sigma^{\text{ess}}(k)$, i. e. they show the set of frequencies for which propagating PhC modes exist and the supercell results (grey lines) are spurious. The essential spectrum was computed separately by computing the eigenvalues of a propagation operator, see Definition 6.16. On the other hand, blank areas represent the band gaps, i. e. areas where the eigenvalues of the supercell method (black lines) correspond to guided modes.

3 Numerical solutions of eigenvalue problems

Besides the discretization of the eigenvalue problems related to PhC and PhC waveguide band structure calculations, the solution of the resulting matrix eigenvalue problems is the main numerical task. These matrix eigenvalue problems can either be linear or nonlinear, which rules the choice of the algorithms for their numerical solution. In this chapter we will give a brief review of algorithms for the numerical solution of eigenvalue problems. Furthermore, we shall propose a new iterative procedure to solve nonlinear eigenvalue problems that will prove useful in the context of PhC waveguide band structure calculations with DtN and RtR transparent boundary conditions.

Solution techniques for linear eigenvalue problems are standard [GVL96] and many software packages are available [ABB+99, LSY98]. Similarly, the solution of quadratic eigenvalue problems is well understood [TM01]. Nevertheless, we shall comment on their numerical solution and implementation issues in Section 3.1.

In PhC and PhC waveguide band structure calculations we also have nonlinear eigenvalue problems, e. g. when employing DtN and RtR transparent boundary conditions or when considering dispersive material. Algorithms for the numerical solution of nonlinear eigenvalue problems, in particular of non-polynomial eigenvalue problems, on the other hand, have been a topic of extensive study in the recent decades. While there has been much progress in the development of a wide range of numerical methods, appropriate software packages for nonlinear eigenvalue problems, like those for linear eigenvalue problems, are not yet available [MV04]. In Section 3.2 we will briefly introduce some algorithms, that we will later use in this work, and comment on their requirements and implementation. In Section 3.3 we will finally propose a new iterative solver for nonlinear eigenvalue problems, that is based on Newton's method.

3.1 Algorithms for linear and quadratic eigenvalue problems

The FE discretization (2.29) of the eigenvalue problem (2.23) of finding approximations to guided modes in PhC waveguides by using the supercell approach is a matrix-valued eigenvalue problem that is

- linear in ω^2 when keeping $k \in B$ fixed, and
- quadratic in k when keeping $\omega \in \mathbb{R}^+$ fixed.

Similarly, the FE discretization (2.28) of the eigenvalue problem (2.13) of finding modes in 2d PhCs is a matrix-valued eigenvalue problem that is either linear (in ω^2 when keeping $\mathbf{k} \in B_{2\mathrm{d}}$ fixed) or quadratic (in either component of $\mathbf{k}$ when keeping $\omega \in \mathbb{R}^+$ and the other component of $\mathbf{k}$ fixed).

Let $N \in \mathbb{N}$ denote the number of degrees of freedom. Then the linear eigenvalue problem in ω^2 can be written in short form like

$$(\mathbf{M}_0 - \lambda \mathbf{M}_1)\,\mathbf{u} = \mathbf{0}$$

with eigenvalue $\lambda = \omega^2$, associated eigenvector $\mathbf{u} \in \mathbb{C}^N \setminus \{\mathbf{0}\}$, and Hermitian matrix $\mathbf{M}_0 \in \mathbb{C}^{N\times N}$, and symmetric, positive definite matrix $\mathbf{M}_1 \in \mathbb{R}^{N\times N}$. For this simple sort of generalized, linear, Hermitian, sparse eigenvalue problem we employ the comprehensive software package ARPACK [LSY98, LMSY15]. The FE software Concepts offers an interface to the C++ wrapper of ARPACK [Con15], which allows for a straightforward implementation of the PhC and PhC waveguide band structure calculation, only writing one main program. However, as we elaborated in Section 2.5, we use the possibility of Concepts to export the FE matrices to Matlab's binary data format `mat` and then use Matlab for the post-processing including the solution of linear eigenvalue problems using Matlab's `eigs` function, which is a reimplementation of the ARPACK functions.

The quadratic eigenvalue problem in k, or k_i, $i = 1, 2$, respectively, takes the form

$$\left(\mathbf{M}_0 + \lambda \mathbf{M}_1 + \lambda^2 \mathbf{M}_2\right)\mathbf{u} = \mathbf{0} \tag{3.1}$$

with eigenvalue $\lambda = k$, associated eigenvector $\mathbf{u} \in \mathbb{C}^N \setminus \{\mathbf{0}\}$, symmetric, positive definite matrices $\mathbf{M}_0, \mathbf{M}_2 \in \mathbb{R}^{N\times N}$, and purely imaginary, Hermitian matrix $\mathbf{M}_1 \in \mathrm{i}\mathbb{R}^{N\times N}$. If we want to solve (2.28) for either component of $\mathbf{k} \in B_{2\mathrm{d}}$ while keeping the other component fixed, $\mathbf{M}_0$ is instead a complex-valued, Hermitian matrix with positive definite real part. Quadratic eigenvalue problems can easily be transformed to linear eigenvalue problems of double size [TM01]. Since the three matrices $\mathbf{M}_0$, $\mathbf{M}_1$ and $\mathbf{M}_2$ are Hermitian, we know that the eigenvalues λ are either real or they come in complex conjugate pairs. A suitable linearization that preserves the structure of the spectrum is obtained by substituting $\widetilde{\lambda} = \mathrm{i}\lambda$, introducing

$$\widetilde{\mathbf{u}} = \begin{pmatrix} \widetilde{\lambda}\mathbf{u} \\ \mathbf{u} \end{pmatrix} \in \mathbb{C}^{2N},$$

and solving the linear eigenvalue problem

$$\left(\mathbf{A}_0 + \widetilde{\lambda}\mathbf{A}_1\right)\widetilde{\mathbf{u}} = \mathbf{0}$$

with matrices

$$\mathbf{A}_0 = \begin{pmatrix} -\mathrm{i}\mathbf{M}_1 & \mathbf{M}_0 \\ -\mathbf{M}_0 & \end{pmatrix} \qquad \text{and} \qquad \mathbf{A}_1 = \begin{pmatrix} -\mathbf{M}_2 & \\ & \mathbf{M}_0 \end{pmatrix}.$$

In case of the supercell eigenvalue problem (2.29) the matrices $\mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{2N\times 2N}$ are skew-symmetric and symmetric, respectively, while in the case of the eigenvalue problem (2.28) the matrices $\mathbf{A}_0, \mathbf{A}_1 \in \mathbb{C}^{2N\times 2N}$ are skew-Hermitian and Hermitian, respectively. This linearization is structure preserving in the sense that the skew-Hamiltonian, isotropic, implicitly restarted shift-and-invert Arnoldi algorithm (SHIRA) [MW01] can be applied, that allows for finding the complex conjugate pairs of the eigenvalue λ simultaneously. However, note that we are only interested in the real eigenvalues λ of (3.1), and hence, the linearization [TM01]

$$\left[\begin{pmatrix} \mathbf{M}_1 & \mathbf{M}_0 \\ \mathbf{M}_0 & \end{pmatrix} + \lambda \begin{pmatrix} \mathbf{M}_2 & \\ & -\mathbf{M}_0 \end{pmatrix}\right] \begin{pmatrix} \lambda\mathbf{u} \\ \mathbf{u} \end{pmatrix} = \mathbf{0}$$

proves reasonable when applying ARPACK's algorithms for generalized, Hermitian eigenvalue problems.

3.2 Algorithms for nonlinear eigenvalue problems

Now we consider nonlinear eigenvalue problems as they appear when using DtN or RtR transparent boundary conditions for the exact computation of guided modes in PhC waveguides. To this end, let us consider the nonlinear eigenvalue problem

$$\mathbf{N}(\lambda)\mathbf{u} = \mathbf{0} \tag{3.2}$$

with eigenvalue $\lambda \in \Omega \subseteq \mathbb{C}$ and associated eigenvector $\mathbf{u} = \mathbf{u}(\lambda) \in \mathbb{C}^N \setminus \{\mathbf{0}\}$, where

$$\mathbf{N} : \Omega \to \mathbb{C}^{N\times N} \tag{3.3}$$

is a nonlinear, matrix-valued, holomorphic function. There exists a wide variety of methods to solve (3.2), see for example the comprehensive review of Mehrmann and Voss [MV04]. This variety of methods ranging from projection methods like Arnoldi-type methods or Jacobi-Davidson-type methods, to Newton-type methods and inverse iteration, has been extended in recent years by methods that rely on the concept of invariant pairs that allow for the simultaneous computation of several eigenvalues [Kre09]. This idea was also extended to the continuation of eigenvalues of parameterized, nonlinear eigenvalue problems in [BEK11]. The simultaneous computation of several eigenvalues of (3.2) is also addressed in [Bey12], where an integral method is proposed to solve (3.2) for all its eigenvalues inside a given contour in the complex plane. As an alternative there is the possibility of linearizing the nonlinear matrix function (3.3). For this, Effenberger and Kressner [EK12] proposed an elegant procedure based on Chebyshev interpolation that does not increase the overall size of the system that needs to be solved.

For the numerical solution of nonlinear eigenvalue problems related to waveguides in homogeneous media truncated by DtN transparent boundary conditions Jarlebring and coworkers recently proposed a

tensor infinite Arnoldi method [JMR15]. This method is a computationally advantageous variant of the infinite Arnoldi method [JMM12], i. e. a method that transforms the concept of Arnoldi methods for linear eigenvalue problems to nonlinear eigenvalue problems. It is based on a Taylor expansion of the nonlinear matrix function $\mathbb{N}$ and can be applied to any nonlinear eigenvalue problem with differentiable $\mathbb{N}$. However, problem specific adaptions are crucial for its application. For the waveguide problem these adaptions were presented in [JMR15].

In this work we do not aim to give an extensive study and comparison of methods to solve the nonlinear eigenvalue problems related to PhC waveguide band structure calculations. Instead we will briefly review and apply two methods:

- the method of successive linear problems (MSLP), and
- the linearization based on Chebyshev interpolation.

The former method, which is a widely used variant of inverse iteration [MV04], shall deal as a benchmark for the Newton-type method that we will propose later in Section 3.3. The latter method, on the other hand, is an easy to implement, yet elegant way to simultaneously compute several eigenvalues of (3.2) and will be used later in Chapters 6 and 7, particularly for the k-formulation.

Method of successive linear problems

The MSLP is based on a Taylor expansion of the nonlinear matrix function $\mathbf{N}$, [Ruh73]. Let $\Omega \subset \mathbb{R}$, which is the case for both, the ω-formulation and the k-formulation of PhC band structure calculations. Then, writing (3.3) in the form

$$\mathbf{N}(\lambda + \ell) = \mathbf{N}(\lambda) + \ell \mathbf{N}'(\lambda) + \mathbf{R}(\lambda, \ell)$$

and neglecting the matrix $\mathbf{R}$, whose norm is bounded by

$$\left\|\mathbf{R}(\lambda, \ell)\right\| \leq \frac{\ell^2}{2} \sup_{0<\widehat{\ell}<\ell} \left\|\mathbf{N}''(\lambda + \widehat{\ell})\right\|,$$

we can proceed as described in Algorithm 3.1 to compute an eigenvalue of the nonlinear eigenvalue problem (3.2).

Algorithm 3.1. Method of successive linear problems.

1: Choose start value $\lambda^{(0)} \in \mathbb{R}$.
2: **for** $i = 0, \ldots$ **do**
3: Solve the generalized, linear eigenvalue problem

$$\left(\mathbf{N}(\lambda^{(i)}) + h\mathbf{N}'(\lambda^{(i)})\right) \mathbf{w} = \mathbf{0}$$

for its eigenvalue h with smallest magnitude.
4: **if** $h \approx 0$ **then**
5: **exit**, $\lambda^{(i)}$ is an eigenvalue of (3.2).
6: **end if**
7: Compute new value $\lambda^{(i+1)} = \lambda^{(i)} + h$.
8: **end for**

The MSLP as sketched in Algorithm 3.1 converges quadratically [Ruh73] and its convergence factors were studied in [Jar12].

Chebyshev interpolation

Effenberger and Kressner [EK12] proposed a linearization of nonlinear eigenvalue problems using the Chebyshev interpolation that allows for a simultaneous computation of several eigenvalues that lie on a curve in the complex plane. We consider the nonlinear eigenvalue problem (3.2) with the matrix-valued,

nonlinear function $\mathbf{N} : \Omega \to \mathbb{C}^{N\times N}$, cf. Eq. (3.3). We aim to find a polynomial approximation $\mathbf{N}_d$ of order d to the nonlinear function $\mathbf{N}$, that is valid on a closed curve $I \subset \Omega$. For simplicity, let I be an interval on the real axis. In fact, we will later in Chapters 6 and 7, where we employ the Chebyshev interpolation, only be interested in real eigenvalues, and hence a choice $I_\lambda = [\lambda_\mathrm{a}, \lambda_\mathrm{b}] \subset \mathbb{R}$ is reasonable. On this interval I we project the $d+1$ Chebyshev nodes $\cos(\frac{i+0.5}{d+1}\pi) \in [-1, 1]$, $i = 0, \dots, d$, of the first kind to obtain the $d+1$ projected Chebyshev nodes

$$\lambda_i = \frac{\lambda_\mathrm{b} - \lambda_\mathrm{a}}{2} \cos\left(\frac{i+0.5}{d+1}\,\pi\right) + \frac{\lambda_\mathrm{a} + \lambda_\mathrm{b}}{2} \in I_\lambda, \qquad i = 0, \dots, d.$$

Let $c_j : I_\lambda \to \mathbb{R}$, $\lambda \mapsto c_j(\lambda)$, $j = 0, \dots, d$, denote the first $d+1$ Chebyshev polynomials defined on the interval I_λ, i. e.

$$\begin{aligned} c_0(\lambda) &= 1, \\ c_1(\lambda) &= \lambda, \\ c_{j+2}(\lambda) &= 2\lambda c_{j+1}(\lambda) - c_j(\lambda), \qquad j = 0, \dots, d-2. \end{aligned} \tag{3.4}$$

Then we approximate

$$\mathbf{N}(\lambda) \approx \mathbf{N}_d(\lambda) = \sum_{j=0}^{d} \mathbf{C}_j c_j(\lambda) \tag{3.5}$$

where the $d+1$ matrices $\mathbf{C}_j \in \mathbb{C}^{N\times N}$ are given by the interpolation condition

$$\mathbf{N}(\lambda_i) = \sum_{j=0}^{d} \mathbf{C}_j c_j(\lambda_i) = \sum_{j=0}^{d} \mathbf{C}_j \cos\frac{j(i+0.5)\pi}{d+1}$$

for all $i = 0, \dots, d$, which can be solved efficiently for $\mathbf{C}_j$ using the discrete cosine transformation [ANR74] of the second type, i. e.

$$\begin{aligned} \mathbf{C}_0 &= \frac{1}{d+1} \sum_{i=0}^{d} \mathbf{N}(\lambda_i), \\ \mathbf{C}_j &= \frac{2}{d+1} \sum_{i=0}^{d} \mathbf{N}(\lambda_i) \cos\frac{j(i+0.5)\pi}{d+1}, \qquad j = 1, \dots, d. \end{aligned}$$

Substituting $\mathbf{u}_j(\lambda) := c_j(\lambda)\mathbf{u}$ and searching the kernel of $\mathbf{N}_d(\lambda)$ defined in (3.5), we obtain the polynomial eigenvalue problem

$$\sum_{j=0}^{d} \mathbf{C}_j \mathbf{u}_j(\lambda) = 0,$$

that can be linearized into the general linear eigenvalue problem

$$\begin{pmatrix} \mathbf{0} & \mathbf{I} & & & \\ \mathbf{I} & \mathbf{0} & \mathbf{I} & & \\ & \ddots & \ddots & \ddots & \\ & & \mathbf{I} & \mathbf{0} & \mathbf{I} \\ -\mathbf{C}_0 & \cdots & -\mathbf{C}_{d-3} & \mathbf{C}_d - \mathbf{C}_{d-2} & -\mathbf{C}_{d-1} \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \vdots \\ \mathbf{u}_{d-1} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{I} & & & & \\ & 2\mathbf{I} & & & \\ & & \ddots & & \\ & & & 2\mathbf{I} & \\ & & & & 2\mathbf{C}_d \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \vdots \\ \mathbf{u}_{d-1} \end{pmatrix} \tag{3.6}$$

of dimension $d \cdot N$, where we used the three term recurrence relation (3.4) of the Chebyshev polynomials.

Applying a shift and invert strategy to compute the eigenvalues of (3.6) the matrix of the left hand side needs to be inverted. However, due to the structure of this matrix its inverse can be determined by simply inverting a matrix of size $N \times N$, cf. [EK12]. To explain this, let us assume we want to solve

$$\begin{pmatrix} \mathbf{0} & \mathbf{I} & & & \\ \mathbf{I} & \mathbf{0} & \mathbf{I} & & \\ & \ddots & \ddots & \ddots & \\ & & \mathbf{I} & \mathbf{0} & \mathbf{I} \\ -\mathbf{C}_0 & \cdots & -\mathbf{C}_{d-3} & \mathbf{C}_d - \mathbf{C}_{d-2} & -\mathbf{C}_{d-1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_{d-1} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0 \\ \vdots \\ \mathbf{y}_{d-1} \end{pmatrix}.$$

We can then deduce that

$$\mathbf{x}_1 = \mathbf{y}_0, \qquad \mathbf{x}_{2j+1} = \mathbf{y}_{2j} - \mathbf{x}_{2j-1}, \quad j = 1, \dots, \lfloor d/2 \rfloor - 1,$$

and

$$\mathbf{x}_{2j} = \widetilde{\mathbf{y}}_{2j-1} + (-1)^j \mathbf{x}_0, \quad j = 1, \dots, \lfloor (d-1)/2 \rfloor,$$

with

$$\widetilde{\mathbf{y}}_1 = \mathbf{y}_1, \qquad \widetilde{\mathbf{y}}_{2j+1} = \mathbf{y}_{2j+1} - \widetilde{\mathbf{y}}_{2j-1}, \quad j = 1, \dots, \lfloor (d-1)/2 \rfloor - 1.$$

Finally, $\mathbf{x}_0$ is determined as solution of the N-dimensional linear system

$$\left(\sum_{j=0}^{\lfloor d/2 \rfloor} (-1)^{j+1} \mathbf{C}_{2j} \right) \mathbf{x}_0 = \mathbf{y}_{d-1} + \left(\sum_{j=0}^{\lfloor d/2 \rfloor - 1} \mathbf{C}_{2j+1} \mathbf{x}_{2j+1} \right) + \left(\sum_{j=1}^{\lfloor (d-1)/2 \rfloor} \mathbf{C}_{2j} \widetilde{\mathbf{y}}_{2j-1} \right) - \mathbf{C}_d \mathbf{z},$$

with

$$\mathbf{z} = \begin{cases} \widetilde{\mathbf{y}}_{d-3}, & \text{if } d \text{ is even}, \\ \mathbf{x}_{d-2}, & \text{if } d \text{ is odd}. \end{cases}$$

Thus, it simply remains to invert the matrix $\left(\sum_{j=0}^{\lfloor d/2 \rfloor} (-1)^{j+1} \mathbf{C}_{2j} \right) \in \mathbb{C}^{N \times N}$.

3.3 A new Newton-type method for nonlinear eigenvalue problems

Let us rewrite the nonlinear eigenvalue problem (3.2) in the form

$$\Big(\mathbf{N}_0(\lambda) + \lambda \mathbf{N}_1(\lambda) \Big) \mathbf{u} = \mathbf{0} \tag{3.7}$$

with eigenvalue $\lambda \in \Omega \subseteq \mathbb{C}$ and associated eigenvector $\mathbf{u} = \mathbf{u}(\lambda) \in \mathbb{C}^N \setminus \{\mathbf{0}\}$, $N \in \mathbb{N}$, where $\mathbf{N}_0$ and $\mathbf{N}_1$ are matrix-valued, holomorphic functions

$$\mathbf{N}_i : \Omega \to \mathbb{C}^{N \times N}, \qquad i = 0, 1,$$

and $\mathbf{N}_1$ is regular for all $\lambda \in \Omega$.

For the nonlinear eigenvalue problem (3.7) we propose an iterative solution technique that employs Newton's method. In every iteration we solve a linear eigenvalue problem that is related to the nonlinear problem (3.7) in fixpoint-like fashion. Keeping $\widetilde{\lambda} \in \Omega$ fixed, we introduce the linear eigenvalue problem

$$\Big(\mathbf{N}_0(\widetilde{\lambda}) + \lambda \mathbf{N}_1(\widetilde{\lambda}) \Big) \mathbf{u}(\widetilde{\lambda}) = \mathbf{0}, \tag{3.8a}$$

with eigenvalue $\lambda = \lambda(\widetilde{\lambda}) \in \Omega$ and associated right eigenvector $\mathbf{u}(\widetilde{\lambda}) \in \mathbb{C}^N \setminus \{\mathbf{0}\}$, and

$$\mathbf{v}^{\mathrm{H}}(\widetilde{\lambda}) \Big(\mathbf{N}_0(\widetilde{\lambda}) + \lambda \mathbf{N}_1(\widetilde{\lambda}) \Big) = \mathbf{0}, \tag{3.8b}$$

with associated left eigenvector $\mathbf{v}(\widetilde{\lambda}) \in \mathbb{C}^N \setminus \{\mathbf{0}\}$. If this linear eigenvalue problem admits an eigenvalue λ that is equal to the parameter $\widetilde{\lambda}$, then λ is an eigenvalue of the nonlinear eigenvalue problem (3.7). On the other hand, for all eigenvalues $\widetilde{\lambda}$ of the nonlinear eigenvalue problem (3.7) we can construct a linear eigenvalue problem of the form (3.8) that has an eigenvalue λ which is equal to $\widetilde{\lambda}$. In this sense, the nonlinear eigenvalue problem (3.7) is equivalent to finding the fixpoints of the function

$$\widetilde{\lambda} \longmapsto \lambda(\widetilde{\lambda}).$$

We shall state the following assumption.

Assumption 3.1. *The eigenvalues $\lambda_j(\widetilde{\lambda})$, $1 \leq j \leq N$, of (3.8) can be ordered such that the functions*

$$\widetilde{\lambda} \longmapsto \lambda_j(\widetilde{\lambda})$$

and the corresponding right eigenvectors $\mathbf{u}_j(\widetilde{\lambda}) = \mathbf{u}(\lambda_j(\widetilde{\lambda}))$ are differentiable with respect to $\widetilde{\lambda}$ in Ω.

Remark 3.2. *From the theory presented in [ACL92, ACL93] we know that the eigenvalue λ_j and its associated right and left eigenvectors $\mathbf{u}_j$ and $\mathbf{v}_j$ are analytic in a neighbourhood of $\widetilde{\lambda} \in \Omega$ if $\lambda_j(\widetilde{\lambda})$ is a simple eigenvalue, which is equivalent to*

(i) $\mathbf{N}_0(\widetilde{\lambda}) + \lambda_j(\widetilde{\lambda})\mathbf{N}_1(\widetilde{\lambda})$ *has rank* $n-1$*, and*

(ii) $\mathbf{v}_j^{\mathrm{H}}(\widetilde{\lambda})\mathbf{N}_1(\widetilde{\lambda})\mathbf{u}_j(\widetilde{\lambda}) \neq 0$.

Thus, if conditions (i) and (ii) are satisfied for all $\widetilde{\lambda} \in \Omega$, Assumption 3.1 is fulfilled.

Due to Assumption 3.1 we can differentiate (3.8) with respect to $\widetilde{\lambda}$, which yields

$$\Big(\mathbf{N}_0'(\widetilde{\lambda}) + \lambda_j(\widetilde{\lambda})\mathbf{N}_1'(\widetilde{\lambda}) + \lambda_j'(\widetilde{\lambda})\mathbf{N}_1(\widetilde{\lambda})\Big)\mathbf{u}_j(\widetilde{\lambda}) + \Big(\mathbf{N}_0(\widetilde{\lambda}) + \lambda_j(\widetilde{\lambda})\mathbf{N}_1(\widetilde{\lambda})\Big)\mathrm{d}_{\widetilde{\lambda}}\mathbf{u}_j(\widetilde{\lambda}) = \mathbf{0}. \tag{3.9}$$

Multiplying (3.9) from the left with the conjugate transpose $\mathbf{v}_j^{\mathrm{H}}(\widetilde{\lambda})$ of the left eigenvector $\mathbf{v}_j(\widetilde{\lambda})$ and considering (3.8b), we arrive at

$$\lambda_j'(\widetilde{\lambda}) = -\frac{\mathbf{v}_j^{\mathrm{H}}(\widetilde{\lambda})\left(\mathbf{N}_0'(\widetilde{\lambda}) + \lambda_j(\widetilde{\lambda})\mathbf{N}_1'(\widetilde{\lambda})\right)\mathbf{u}_j(\widetilde{\lambda})}{\mathbf{v}_j^{\mathrm{H}}(\widetilde{\lambda})\mathbf{N}_1(\widetilde{\lambda})\mathbf{u}_j(\widetilde{\lambda})}. \tag{3.10}$$

Remark 3.3. *From the formula (3.10) for the derivative $\lambda_j'(\widetilde{\lambda})$ of the eigenvalue $\lambda_j(\widetilde{\lambda})$ with respect to $\widetilde{\lambda}$ we can see that*

$$\mathbf{v}_j^{\mathrm{H}}(\widetilde{\lambda})\mathbf{N}_1(\widetilde{\lambda})\mathbf{u}_j(\widetilde{\lambda}) \neq 0$$

is a necessary condition for the analyticity of the functions $\widetilde{\lambda} \mapsto \lambda_j(\widetilde{\lambda})$, i. e. a necessary condition for Assumption 3.1 to be satisfied.

Remark 3.4. *If the matrices $\mathbf{N}_0$ and $\mathbf{N}_1$ are Hermitian, the left eigenvector is identical to the right eigenvector, and hence, it is sufficient to solve (3.8) for its right eigenvector.*

The formula (3.10) for the derivative $\lambda_j'(\widetilde{\lambda})$ of the eigenvalue $\lambda_j(\widetilde{\lambda})$ will be used in the iterative scheme that we shall explain in the following.

We introduce the signed distance functions

$$d_j(\widetilde{\lambda}) = \widetilde{\lambda} - \lambda_j(\widetilde{\lambda}), \tag{3.11}$$

$1 \leq j \leq N$, that are — thanks to Assumption 3.1 — continuously differentiable. Hence, we can apply Newton's method to compute the roots of (3.11) for $1 \leq j \leq N$, which are — due our above considerations — eigenvalues of the nonlinear eigenvalue problem (3.7). The global signed distance function

$$d(\widetilde{\lambda}) = \widetilde{\lambda} - \lambda_{j^*}(\widetilde{\lambda}), \tag{3.12}$$

with

$$j^* = j^*(\widetilde{\lambda}) = \arg\min_{1 \leq j \leq N} \left|d_j(\widetilde{\lambda})\right|, \tag{3.13}$$

however, is only continuous and piecewise analytic, but its roots are by definition also eigenvalues of (3.7). The advantage of computing the roots of the global signed distance function d is that we only have to find the roots of a single functions, instead of computing the roots of the N signed distance functions d_j. Note that the outer min-operator in the definition (3.13) is needed in the case that two or more eigenvalues of (3.8) have the same distance to $\widetilde{\lambda}$ in magnitude.

The proposed iterative scheme is then as easy as shown in Algorithm 3.2, where the derivative of the global signed distance function (3.12) is defined as

$$d'(\lambda^{(i)}) := d_{j^*}'(\lambda^{(i)}) = 1 - \lambda_{j^*}'(\lambda^{(i)})$$

with $j^* = j^*(\lambda^{(i)})$ as given in (3.13) and with the derivative λ_{j^*}' of λ_{j^*} with respect to $\lambda^{(i)}$ as presented in Eq. (3.10).

Even though the global signed distance function d is not continuously differentiable, Newton's method as sketched in Algorithm 3.2 converges quadratically [KS86], since the global signed distance function d is piecewise identical to some continuously differentiable signed distance function d_j, $1 \leq j \leq N$.

Algorithm 3.2. Newton's method applied to global signed distance function.

Choose start value $\lambda^{(0)} \in \mathbb{R}$.
for $i = 0, \ldots$ **do**
    Evaluate global signed distance function $d(\lambda^{(i)})$.
    if $d(\lambda^{(i)}) \approx 0$ **then**
        exit, $\lambda^{(i)}$ is an eigenvalue of (3.7).
    end if
    Compute derivative $d'(\lambda^{(i)})$.
    if $d'(\lambda^{(i)}) \approx 0$ **then**
        exit, $\lambda^{(i)}$ is a saddle point, retry with new start value.
    end if
    Compute new value $\lambda^{(i+1)} = \lambda^{(i)} - \frac{d(\lambda^{(i)})}{d'(\lambda^{(i)})}$.
end for

The computational effort of the proposed method is comparable to the effort of the MSLP, see Section 3.2 as in each iteration a linear eigenvalue problem of size $N \times N$ has to be solved for one eigenvalue. Moreover, both methods have the same convergence rate, which can also be seen later in numerical results presented in Chapter 6.

The proposed Newton-type method to solve nonlinear eigenvalue problems of the form (3.7) will later be employed in Chapters 6 and 7 for the nonlinear eigenvalue problems that result from truncating the domain of PhC waveguides using DtN or RtR transparent boundary conditions.

An alternative application of this method in the context of PhC band structure calculations is the eigenvalue problem of finding modes in PhCs with dispersive material, i. e. with frequency-dependent permittivity. When considering the ω-formulation this eigenvalue problem becomes nonlinear and it satisfies the form (3.7) we studied in this section. Note that the k-formulation still remains a quadratic eigenvalue problem. The reader is referred to [ER09, Eng14] for this formulation.

4 Group velocity and higher derivatives of dispersion curves

The group velocity is the first derivative of the dispersion curve $\omega(k)$. Guided modes of small group velocity, also denoted by slow light modes, are of special interest in optics since the reduction of the group velocity simultaneously yields an enhancement of the light intensity [Kra08]. In this chapter, whose results were already published in condensed form in [KS14b], we will derive a closed formula for the group velocity of guided modes in PhC waveguides. This formula can be used for an exact computation of the group velocity for example in waveguide optimizations [LWO+08], replacing the difference quotient to compute the slope of the band functions.

Our approach does not only allow for deriving a closed formula for the group velocity but also for the second derivative, the so-called group velocity dispersion, and for any higher-order derivative of the dispersion curves. In this sense our approach is different from the perturbation theory employed in [SS88, Sip00, HFBW01], where the vector $\mathbf{k} \cdot \mathbf{p}$ approach of electronic band structure theory is transferred to PhC band structure calculations. Our computational procedure has two main advantages:

(i) it is "exact" in the sense that no additional modelling error is introduced in comparison to the perturbation approach in [SS88, Sip00, HFBW01] where an infinite sum for the computation of the group velocity dispersion has to be truncated, and

(ii) it allows for a successive computation of derivatives up to any order with marginal extra computational costs for each additional order.

The latter point motivates an algorithm for efficient band structure calculations that employs these derivatives and that we will present in Chapter 5.

Before we shall derive the formulas for the dispersion curve derivatives in Section 4.2, we will elaborate on the differentiability of the dispersion curves and their corresponding eigenmodes in Section 4.1. A proof of the eigenmode differentiability, that only uses the variational formulation of the eigenvalue problem, is then presented in Section 4.3, and in Section 4.4 we will give concluding remarks.

4.1 Differentiability of dispersion curves and eigenmodes

In this chapter we simultaneously consider the eigenvalue problem (2.13) in the PhC unit cell C, the eigenvalue problem (2.19) in the infinite strip S of a PhC waveguide, and the eigenvalue problem (2.23) in the bounded supercell S_n with $n \in \mathbb{N}$ PhC unit cells $C_i^{\pm}$, $i \leq n$, on top and bottom of the defect cell C_0. All these eigenvalue problems are linear in ω^2. For the sake of simplicity, we shall assume $k_2 = 0$ in the 2d PhC eigenvalue problem (2.13). Then we employ the Floquet transform in x_1-direction, considering the 1d Brillouin zone $B = [-\frac{\pi}{a_1}, \frac{\pi}{a_1}]$, and hence, all mentioned eigenvalue problems are identical except for the domain, which is either C, S or S_n. In the sequel we shall use C for the domain, keeping in mind that it can be interchanged with S and S_n.

In Chapter 6 we will introduce DtN transparent boundary conditions for the interfaces $\Gamma_0^{\pm}$ of C_0 to the semi-infinite periodic strips $S^{\pm}$ on top and bottom of the guide. In that chapter we will also extend the formulas for the dispersion curve derivatives that we will now introduce for the case with periodic boundary conditions.

Let $k \in B$. Directly considering the variational formulation, we search for eigenvalues $\omega_j^2(k) \in \mathbb{R}^+$ and corresponding non-trivial eigenmodes $u_j(k) \equiv u_j(\cdot\,; k) \in \mathsf{H}^1_{\mathrm{p}}(C)$ such that

$$\mathfrak{b}_C(u_j(k), v; \omega_j, k) = 0 \tag{4.1}$$

for all test functions $v \in \mathsf{H}^1_\mathrm{p}(C)$, where the sesquilinear form $\mathfrak{b}_C$ reads

$$\mathfrak{b}_C(u, v; \omega, k) = \mathfrak{a}_C^\alpha(u, v) + k\mathfrak{c}_C^{\alpha,1}(u, v) + k^2\mathfrak{m}_C^\alpha(u, v) - \omega^2\mathfrak{m}_C^\beta(u, v)$$

with

$$\begin{aligned}
\mathfrak{a}_C^\alpha(u, v) &= \int_C \alpha \nabla u \cdot \nabla \overline{v} \, \mathrm{d}\mathbf{x}, \\
\mathfrak{c}_C^{\alpha,1}(u, v) &= \int_C \mathrm{i}\alpha \left(u(\partial_1 \overline{v}) - (\partial_1 u)\overline{v} \right) \, \mathrm{d}\mathbf{x}, \\
\mathfrak{m}_C^\alpha(u, v) &= \int_C \alpha u \overline{v} \, \mathrm{d}\mathbf{x}, \\
\mathfrak{m}_C^\beta(u, v) &= \int_C \beta u \overline{v} \, \mathrm{d}\mathbf{x}.
\end{aligned}$$

Now we give the main result, the proposed formulas rely on.

Theorem 4.1. *The eigenvalues $\omega_j^2(k) \in \mathbb{R}^+$, $j \in \mathbb{N}$, of (4.1) can be ordered such that the dispersion curves*

$$k \longmapsto \omega_j(k)$$

are analytic. In addition, the magnitude and phase of the corresponding eigenmodes $u_j(\cdot\,; k)$ can be chosen such that the eigenmodes are analytic with respect to the quasi-momentum k.

Proof. This theorem is a direct consequence of Theorems 2.4 and 2.6, where the same result is shown for the eigenvalue problems (2.19) and (2.23) in operator formulation using the perturbation theory for linear operators [Kat95]. Since the sesquilinear form in (4.1) is bounded in $\mathsf{H}^1_\mathrm{p}(C)$ and the corresponding linear operator is bounded in $\mathsf{H}^1_\mathrm{p}(\Delta, C, \alpha)$, we conclude that the spectral results of the operator theory directly transfer to the eigenvalue problem (4.1) in variational formulation [Kat95]. □

In addition to the well known eigenvalue analyticity, Theorem 4.1 also guarantees that the corresponding eigenmodes are analytic. Nevertheless, we shall present in Section 4.3 a proof of the differentiability of the eigenmodes only using the variational formulation.

4.2 Dispersion curve derivatives

4.2.1 First derivative of dispersion curves — The group velocity

Thanks to Theorem 4.1 we can take the derivative of Eq. (4.1) with respect to k and obtain

$$\mathfrak{b}_C(\mathrm{d}_k u_j, v; \omega_j, k) = \mathfrak{f}^{(1)}(v) \tag{4.2}$$

for all $v \in \mathsf{H}^1_\mathrm{p}(C)$, with the linear form

$$\mathfrak{f}^{(1)}(v) = \mathfrak{f}^{(1)}(v; k, \omega_j, \omega_j', u_j) = -2k\mathfrak{m}_C^\alpha(u_j, v) - \mathfrak{c}_C^{\alpha,1}(u_j, v) + 2\omega_j\omega_j'\mathfrak{m}_C^\beta(u_j, v) \tag{4.3}$$

and the short notations $\omega_j'(k) := \frac{\partial \omega_j}{\partial k}(k)$ and $\mathrm{d}_k u_j(\cdot\,; k) := \frac{\mathrm{d}u_j}{\mathrm{d}k}(\cdot\,; k) \in \mathsf{H}^1_\mathrm{p}(C)$. Taking $v = u_j$ as test function in Eq. (4.2) yields

$$\mathfrak{f}^{(1)}(u_j; k, \omega_j, \omega_j', u_j) = 0 \tag{4.4}$$

since the sesquilinear form $\mathfrak{b}_C$ is self-adjoint, and hence,

$$\mathfrak{b}_C(\mathrm{d}_k u_j, u_j; \omega_j, k) = \overline{\mathfrak{b}_C(u_j, \mathrm{d}_k u_j; \omega_j, k)} = 0$$

as u_j is an eigenmode of (4.1) at k with associated eigenvalue ω_j^2. We can solve (4.4) for the group velocity ω_j' and obtain

$$\omega_j'(k) = \frac{2k\mathfrak{m}_C^\alpha(u_j, u_j) + \mathfrak{c}_C^{\alpha,1}(u_j, u_j)}{2\omega_j\mathfrak{m}_C^\beta(u_j, u_j)}. \tag{4.5}$$

Note that the group velocity is real-valued since the bilinear forms $\mathfrak{m}_C^\alpha$, $\mathfrak{m}_C^\beta$ and $\mathfrak{c}_C^{\alpha,1}$ are self-adjoint, and thus, e. g.

$$\operatorname{Im}\Big(\mathfrak{c}_C^{\alpha,1}(u_j,u_j)\Big) = \frac{1}{2}\left(\mathfrak{c}_C^{\alpha,1}(u_j,u_j) - \overline{\mathfrak{c}_C^{\alpha,1}(u_j,u_j)}\right) = 0.$$

Considering that $\mathfrak{c}_C^{\alpha,1}(u_j,u_j)$ is real-valued, using integration by parts and the fact that u_j is periodic, we can write

$$\begin{aligned}\mathfrak{c}_C^{\alpha,1}(u_j,u_j) &= \int_C \mathrm{i}\alpha\left(u_j(\partial_1\overline{u}_j) - (\partial_1 u_j)\overline{u}_j\right)\,\mathrm{d}\mathbf{x} = 2\int_C \mathrm{i}\alpha u_j(\partial_1\overline{u}_j)\,\mathrm{d}\mathbf{x} + \int_C \mathrm{i}(\partial_1\alpha)|u_j|^2\,\mathrm{d}\mathbf{x}\\ &= 2\operatorname{Re}\left(\int_C \mathrm{i}\alpha u_j(\partial_1\overline{u}_j)\,\mathrm{d}\mathbf{x}\right) = -2\operatorname{Im}\left(\int_C \alpha u_j(\partial_1\overline{u}_j)\,\mathrm{d}\mathbf{x}\right). \quad (4.6)\end{aligned}$$

Note that the derivative of α, that appears in the above equation, has to be understood in distributional sense and thus, it is well-defined even though we assume only $\alpha \in \mathsf{L}^\infty(\mathbb{R}^2)$. Moreover, it is obvious, that $\int_C(\partial_1\alpha)|u_j|^2\,\mathrm{d}\mathbf{x}$ is well-defined since all other integrals of the equation are well-defined. Thus, we can rewrite the group velocity formula (4.5) in the form

$$\omega_j'(k) = \frac{k\int_C \alpha|u_j|^2\,\mathrm{d}\mathbf{x} - \operatorname{Im}\big(\int_C \alpha\, u_j\,\partial_1\overline{u}_j\,\mathrm{d}\mathbf{x}\big)}{\omega_j\int_C \beta|u_j|^2\,\mathrm{d}\mathbf{x}}.$$

Remark 4.2. *The formula (4.5) for the group velocity contains the eigenmode u_j associated to the eigenvalue $\omega_j^2(k)$. However, the eigenmode is not uniquely defined. If the eigenvalue has multiplicity equal to one, any non-trivial, scalar multiple of an eigenmode is also an eigenmode. However, such a scalar cancels out in (4.5) and the group velocity formula is well-defined. If the eigenvalue has multiplicity larger than one, the situation is more involved. Nevertheless, we claim that the eigenmodes in (4.5) can be chosen as the limit of the eigenmodes associated to the eigenvalues of multiplicity one, that lie on the dispersion curves which intersect at $(\omega_j(k),k)$. For this, it is important that the approximation quality of the eigenmodes is not influenced by the distance to a crossing of dispersion curves. For example, when using the software package ARPACK [LSY98, LMSY15], i. e. an implementation of implicitly restarted Arnoldi iterations, for the numerical solution of the corresponding matrix eigenvalue problem, we can expect that the approximation of eigenvectors of multiple eigenvalues and of eigenvalues that are close to a multiplicity larger than one is of the same quality like the approximation of eigenvectors of simple eigenvalues, since a deflation technique [LS96] is used, [LSY98].*

4.2.2 Higher derivatives of dispersion curves

To simplify the presentation and in accordance to Remark 4.2, we shall assume in the sequel, that the eigenvalue $\omega_j^2(k)$ has multiplicity one and is sufficiently far away from a crossing. In Section 4.2.3 we will discuss what is meant by "sufficiently far away from a crossing" in practise. If the multiplicity of $\omega_j^2(k)$ is larger than one or if the distance of $\omega_j^2(k)$ to a crossing is not sufficient, the reader is referred to Remark 4.5.

In order to extend the procedure to higher order derivatives of the dispersion curves, we have to compute the derivative $\mathrm{d}_k u_j$ of the eigenmode u_j with respect to the quasi-momentum. However, the computation of $\mathrm{d}_k u_j \in \mathsf{H}^1_{\mathrm{p}}(C)$ using Eq. (4.2) is ill-posed since any eigenmode $u_j \in \mathsf{H}^1_{\mathrm{p}}(C)$ solves Eq. (4.2) with zero right hand side and hence, any of these eigenmodes u_j can be added to the solution $\mathrm{d}_k u_j$ of Eq. (4.2) and the equation will still be satisfied. Applying the Fredholm–Riesz–Schauder theory, see for example Section 2.1.4 in [SS11], we can compute a particular solution of Eq. (4.2) by additionally requiring $\mathsf{H}^1(C)$-orthogonality to any of the finitely many [RS78], linearly independent eigenmodes $u_{j,1},\dots,u_{j,m}$. With the above mentioned assumption we look for the particular solution of Eq. (4.2) that is $\mathsf{H}^1(C)$-orthogonal to the single, possibly normalized eigenmode u_j. This orthogonality condition differs from the condition we will introduce in the proof of the differentiability of the eigenmodes, cf. Theorem 4.1, presented in Section 4.3. There we will choose the solution of (4.2) that is $\mathsf{H}^1(C)$-orthogonal to the eigenmode $u_j(\cdot;k_0)$ for some k_0 in the vicinity of k. However, in accordance to Proposition 4.4, that we shall prove later, we can in fact use any extra condition to fix the solution of (4.2), as long as the resulting problem is

well-posed. In this respect, we shall refrain from calling a particular solution of (4.2) the derivative of the eigenmode with respect to the quasi-momentum. Instead, we shall compute an auxiliary function $u_j^{(1)} \in \mathsf{H}^1_{\mathrm{p}}(C)$ and a Lagrange multiplier $\lambda \in \mathbb{C}$ that satisfy

$$\mathfrak{b}_C(u_j^{(1)}, v; \omega_j, k) + \lambda \langle u_j, v \rangle_{\mathsf{H}^1(C)} = \mathfrak{f}^{(1)}(v), \tag{4.7a}$$

$$\langle u_j^{(1)}, u_j \rangle_{\mathsf{H}^1(C)} = 0, \tag{4.7b}$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$, where $\langle \cdot, \cdot \rangle_{\mathsf{H}^1(C)}$ denotes the usual $\mathsf{H}^1(C)$-inner product, i. e. $\langle u, v \rangle_{\mathsf{H}^1(C)} = \int_C \nabla u \cdot \nabla \overline{v} + u\overline{v} \, \mathrm{d}\mathbf{x}$. This auxiliary function satisfies $u_j^{(1)} = \mathrm{d}_k u_j + c u_j$, $c \in \mathbb{C}$, such that $\mathfrak{b}_C(u_j^{(1)}, v; \omega_j, k) = \mathfrak{b}_C(\mathrm{d}_k u_j, v; \omega_j, k)$ for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$.

Remark 4.3. *From the Fredholm alternative, see for example [SS11], and the theory of saddle point problems, see for example [Bra07], and by considering our assumption that the eigenvalue $\omega_j^2(k)$ of (4.1) has multiplicity one, we know that the mixed variational problem (4.7) has a unique solution. The Lagrange multiplier λ of this unique solution is zero, since, testing (4.7a) with $v = u_j$, $\|u_j\|_{\mathsf{H}^1(C)} = 1$, yields $\lambda = \mathfrak{f}^{(1)}(u_j)$, which is identical to zero due to (4.4).*

In order to determine higher derivatives of $k \mapsto \omega_j(k)$ let us introduce the following short notations

$$\omega_j^{(n)}(k) := \frac{\partial^n \omega_j}{\partial k^n}(k) \qquad \text{and} \qquad \mathrm{d}_k^n u_j(\cdot\,; k) := \frac{\mathrm{d}^n u_j}{\mathrm{d}k^n}(\cdot\,; k),$$

$n \in \mathbb{N}_0$. Then taking the n-th derivative of Eq. (4.1) with respect to k yields

$$\mathfrak{b}_C(\mathrm{d}_k^n u_j, v; \omega_j, k) = \mathfrak{f}^{(n)}(v)$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$, where the linear form $\mathfrak{f}^{(n)} = \mathfrak{f}^{(n)}(\cdot\,; k, \omega_j^{(0)}, \dots, \omega_j^{(n)}, u_j^{(0)}, \dots, u_j^{(n-1)})$, that is obtained using binomial and trinomial expansions, reads

$$\begin{aligned}
\mathfrak{f}^{(n)}(v) &= \sum_{p=0}^{n-1} \sum_{q=0}^{n-p} \frac{n!}{p!\, q!\, (n-p-q)!} \omega_j^{(n-p-q)} \omega_j^{(q)} \mathfrak{m}_C^{\beta}(u_j^{(p)}, v) \\
&\quad - \sum_{p=1}^{n} \binom{n}{p} \frac{\partial^p k}{\partial k^p} \, \mathfrak{c}_C^{\alpha,1}(u_j^{(n-p)}, v) - \sum_{p=1}^{n} \binom{n}{p} \frac{\partial^p k^2}{\partial k^p} \, \mathfrak{m}_C^{\alpha}(u_j^{(n-p)}, v) \\
&= \sum_{p=0}^{n-1} \sum_{q=0}^{n-p} \frac{n!}{p!\, q!\, (n-p-q)!} \omega_j^{(n-p-q)} \omega_j^{(q)} \mathfrak{m}_C^{\beta}(u_j^{(p)}, v) \\
&\quad - n \, \mathfrak{c}_C^{\alpha,1}(u_j^{(n-1)}, v) - 2n\, k \, \mathfrak{m}_C^{\alpha}(u_j^{(n-1)}, v) - n(n-1) \, \mathfrak{m}_C^{\alpha}(u_j^{(n-2)}, v),
\end{aligned} \tag{4.8}$$

where we replaced the eigenmode derivatives $\mathrm{d}_k^m u_j(k)$, $1 \le m \le n-1$, by the auxiliary functions $u_j^{(m)}(k)$, and $u_j^{(0)}(k) = u_j(k)$. From this we deduce the n-th derivative of $\omega_j(k)$

$$\begin{aligned}
\omega_j^{(n)}(k) = \frac{1}{2\omega_j \mathfrak{m}_C^{\beta}(u_j, u_j)} \Bigg(& n(n-1) \, \mathfrak{m}_C^{\alpha}(u_j^{(n-2)}, u_j) + 2n\, k \, \mathfrak{m}_C^{\alpha}(u_j^{(n-1)}, u_j) + n \, \mathfrak{c}_C^{\alpha,1}(u_j^{(n-1)}, u_j) \\
& - \sum_{p=1}^{n-1} \sum_{q=0}^{n-p} \frac{n!}{p! q! (n-p-q)!} \omega_j^{(n-p-q)} \omega_j^{(q)} \mathfrak{m}_C^{\beta}(u_j^{(p)}, u_j) \\
& - \sum_{q=1}^{n-1} \binom{n}{q} \omega_j^{(n-q)} \omega_j^{(q)} \mathfrak{m}_C^{\beta}(u_j, u_j) \Bigg).
\end{aligned} \tag{4.9}$$

Analogously to above — using the Lagrange multiplier $\lambda \in \mathbb{C}$ — we can then compute a particular solution $u^{(n)} \in \mathsf{H}^1_{\mathrm{p}}(C)$ of (4.2.2) that satisfies

$$\mathfrak{b}_C(u_j^{(n)}, v; \omega_j, k) + \lambda \langle u_j, v \rangle_{\mathsf{H}^1(C)} = \mathfrak{f}^{(n)}(v) \tag{4.10a}$$

$$\langle u_j^{(n)}, u_j \rangle_{\mathsf{H}^1(C)} = 0 \tag{4.10b}$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$, which is, for $n = 1$, equivalent to Eq. (4.7). Note that the sesquilinear forms on the left hand side of Eq. (4.10) are identical for all $n \in \mathbb{N}$, while the linear forms $\mathfrak{f}^{(n)}$ differ for all orders.

In order to compute $\omega_j^{(n)}$ we have to solve (4.1) for its eigenvalue $\omega_j^2(k)$ and associated eigenmode u_j. Then we successively compute $\omega_j^{(\ell)}$ from (4.9) and solve the linear system (4.10) for $u_j^{(\ell)}$, $\ell = 1, \dots, n-1$. Finally, it remains to compute $\omega_j^{(n)}$ from (4.9). In total we have to solve one eigenvalue problem (4.1), $n-1$ linear systems (4.10), and n algebraic equations (4.9).

Let us discuss the effect of the proposed orthogonality condition in the linear systems (4.7) and (4.10).

Proposition 4.4. *The formula (4.9) for n-th dispersion curve derivative is independent of the orthogonality condition (4.10b).*

Proof. First we note that by construction of (4.10) it is easy to see that for all $n \in \mathbb{N}$

$$u_j^{(n)} = \mathrm{d}_k u_j^{(n-1)} + c^{(n)} u_j$$

with $c^{(n)} \in \mathbb{C}$. Recursively applying this identity yields

$$u_j^{(n)} = \mathrm{d}_k^n u_j + \sum_{i=1}^{n} c^{(i)} \mathrm{d}_k^{n-i} u_j. \tag{4.11}$$

Let us assume that we can properly define and compute $\mathrm{d}_k^n u_j$, $n \in \mathbb{N}$. Replacing the auxiliary functions $u_j^{(n)}$ in (4.9) by $\mathrm{d}_k^n u_j$ we obtain a formula that we denote by $\widetilde{\omega}_j^{(n)}(k)$. Trivially, we have

$$\omega_j^{(1)}(k) = \widetilde{\omega}_j^{(1)}(k).$$

Using Eq. (4.11) for $n = 1$ we find that

$$\begin{aligned}\omega_j^{(2)}(k) = \frac{\mathfrak{m}_C^{\alpha}(u_j,u_j) + 2k\mathfrak{m}_C^{\alpha}(\mathrm{d}_k u_j,u_j) + \mathfrak{c}_C^{\alpha,1}(\mathrm{d}_k u_j,u_j) - 2\omega_j'\omega_j\mathfrak{m}_C^{\beta}(\mathrm{d}_k u_j,u_j) - (\omega_j')^2\mathfrak{m}_C^{\beta}(u_j,u_j)}{\omega_j\mathfrak{m}_C^{\beta}(u_j,u_j)} \cdot \\ + c^{(1)}\frac{2k\mathfrak{m}_C^{\alpha}(u_j,u_j) + \mathfrak{c}_C^{\alpha,1}(u_j,u_j) - 2\omega_j'\omega_j\mathfrak{m}_C^{\beta}(u_j,u_j)}{\omega_j\mathfrak{m}_C^{\beta}(u_j,u_j)}.\end{aligned}$$

Inserting (4.5) shows that the numerator of the second term vanishes and hence, also

$$\omega_j^{(2)}(k) = \widetilde{\omega}_j^{(2)}(k).$$

We proceed by induction. Let $n \in \mathbb{N}$. Assuming that $\omega_j^{(m)}(k) = \widetilde{\omega}_j^{(m)}(k)$ for all $m = 1, \dots, n-1$, we apply (4.11) and obtain

$$\begin{aligned}\omega_j^{(n)}(k) = \widetilde{\omega}_j^{(n)}(k) + \frac{1}{2\omega_j\mathfrak{m}_C^{\beta}(u_j,u_j)}\Bigg(& n(n-1)\sum_{i=1}^{n-2} c^{(i)}\,\mathfrak{m}_C^{\alpha}(\mathrm{d}_k^{n-i}u_j,u_j) \\ & + n\sum_{i=1}^{n-1} c^{(i)}\left(2k\mathfrak{m}_C^{\alpha}(\mathrm{d}_k^{n-i}u_j,u_j) + \mathfrak{c}_C^{\alpha,1}(\mathrm{d}_k^{n-i}u_j,u_j)\right) \\ & - \sum_{p=1}^{n-1}\sum_{q=0}^{n-p}\sum_{i=1}^{p}\frac{n!}{p!q!(n-p-q)!}\omega_j^{(n-p-q)}\omega_j^{(q)}c^{(i)}\mathfrak{m}_C^{\beta}(\mathrm{d}_k^i u_j,u_j)\Bigg).\end{aligned}$$

We sort the right hand side for terms with $c^{(1)}, \dots, c^{(n-1)}$. Using the formula (4.9) for $\omega_j^{(n-i)}$, $1 \le i \le n-1$, which is by assumption identical to $\widetilde{\omega}_j^{(n-i)}$, we find that the $c^{(i)}$-term vanishes and, hence, we can conclude that

$$\omega_j^{(n)}(k) = \widetilde{\omega}_j^{(n)}(k).$$

This shows that the orthogonality condition (4.10b) can in fact be replaced by any other orthogonality condition as long as the resulting system (4.10) is well-posed. □

Finally, let us give a remark on the case of multiple eigenvalues.

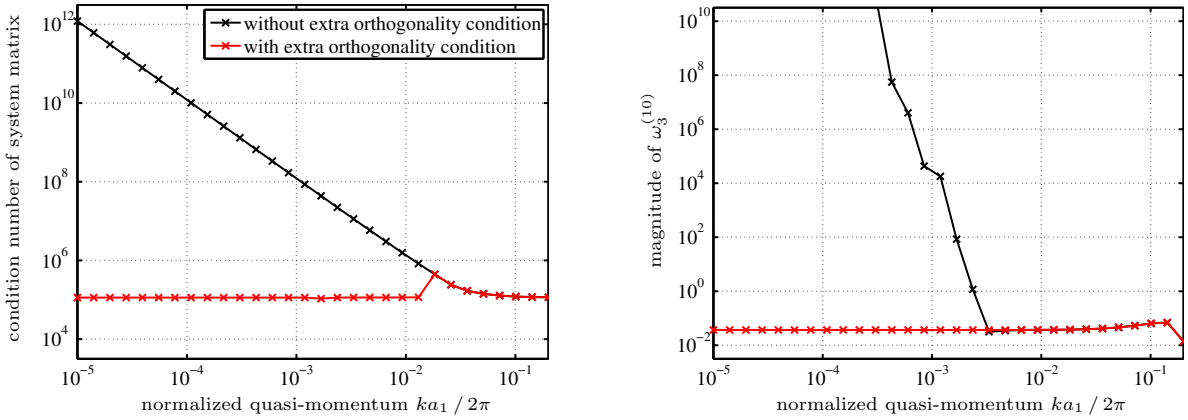
Remark 4.5. *If there are multiple, linearly independent eigenmodes $u_{j,1}, \dots, u_{j,m}$, associated to the eigenvalue $\omega_j^2(k)$ of (4.1), $m-1$ extra orthogonality conditions and Lagrange multipliers $\lambda_2, \dots, \lambda_m \in \mathbb{C}$ have to be added to the linear system (4.10). Note that the eigenmodes $u_{j,1}, \dots, u_{j,m}$ have to be selected as described in Remark 4.2, i. e. as the limit of the eigenmodes corresponding to the eigenvalues of multiplicity one in the vicinity of $\omega_j^2(k)$. The procedure to compute the n-th derivative $\omega_{j,m'}^{(n)}$, $1 \leq m' \leq m$, of $\omega_j(k)$ associated to $u_{j,m'}$ remains the same, only that we have to bear in mind that each eigenmode $u_{j,m'}$ associated to the eigenvalue $\omega_j^2(k)$ yields different quantities $u_{j,m'}^{(n)}$ and $\omega_{j,m'}^{(n)}$.*

4.2.3 Extra orthogonality conditions at simple eigenvalues

The extra orthogonality conditions for the linear system (4.10), mentioned in Remark 4.5 for the case of multiple eigenvalues of (4.1), have to be added to (4.10) also for the case of simple eigenvalues $\omega_j^2(k)$ of (4.1), if there exist other eigenvalues $\omega_{j'}^2(k)$, $j' \neq j$, of (4.1) at k, that are very close to $\omega_j^2(k)$. This is for example always the case in the vicinity of an eigenvalue with multiplicity larger than one. The reason is that the condition number of the matrix related to the linear system (4.10) increases dramatically in such a case.

These extra orthogonality conditions, when posed at simple eigenvalues instead of multiple eigenvalues, imply that Eq. (4.2) and its higher order analogues are not exactly satisfied, since the Lagrange multipliers do not vanish, or in other words, the solution of (4.2), if $\omega_j^2(k)$ is an simple eigenvalue of (4.1), cannot satisfy more than one orthogonality condition.

Therefore, the size of the vicinity, for which extra orthogonality conditions are added to (4.10), has to be chosen very carefully. We observed that reasonably good results can be obtained if orthogonality conditions are added to (4.10) for all eigenvalues $\omega_{j'}^2(k)$ with distance to $\omega_j^2(k)$ smaller than $0.01 \cdot \frac{c^2}{a_1^2}$.



(a) Condition number of system matrix.

(b) Magnitude of tenth dispersion curve derivative.

Figure 4.1: Condition number of the FE system matrix (a) related to the source problem (4.10) and magnitude of the tenth dispersion curve derivative (b) of the third dispersion curve of the band structure of Example 1 when adding an extra orthogonality condition for all eigenvalues closer than $0.01 \cdot \frac{c^2}{a_1^2}$ (red) and when not adding any extra orthogonality conditions (black).

Numerical experiments show that extra orthogonality conditions in the suggested vicinity do not yield a significant error. When comparing the group velocity dispersion, i. e. the second dispersion curve derivative, in the vicinity of a crossing of two dispersion curves, we find that the difference of the computed values for the cases with and without extra orthogonality condition is negligible. Instead the effect of an increasing condition number of the matrix related to the linear system (4.10) is of larger importance. Let us study the behaviour of the condition number numerically. To this end, we consider the band structure of Example 1. The second and third dispersion curves intersect at $k = 0$, see Figure 2.7. For the FE

discretization we choose the coarse mesh with nine quadrilateral cells as sketched in Figure 2.6 and the polynomial degree $p = 5$. In Figure 4.1a we present the condition number of the FE matrix related to the linear system (4.10) for the third dispersion curve in the vicinity of $k = 0$. The black curve shows the condition number if no extra orthogonality condition is taken into account no matter how close another eigenvalue is located. The red curve shows the condition number for the case as suggested above, i. e. we add an orthogonality condition to (4.10) for all eigenvalues $\omega^2(k)$ with distance to $\omega_3^2(k)$ smaller than $0.01 \cdot \frac{c^2}{a_1^2}$. This is the case for all quasi-momenta smaller than $k \approx 0.01\frac{2\pi}{a_1}$, where the eigenvalues on the second dispersion curve have taken into account. We can observe that the condition number of the matrix without extra orthogonality condition increases dramatically in the vicinity of the crossing at $k = 0$. The condition number of the matrix including an extra orthogonality with respect to the eigenmodes on the second dispersion curve, however, only shows a minor increase just before the extra condition is switched on, while it remains almost constant at a relatively low level when we continue to approach the crossing at $k = 0$.

In Figure 4.1b the effect of the increasing condition number on the computation of the dispersion curve derivatives is shown. We present the magnitude of the tenth derivative of the third dispersion curve of the band structure of Example 1. It becomes obvious that the dramatic increase of the condition number, if no extra orthogonality conditions are added to (4.10), also yields a significant increase of the magnitude of the computed value for the dispersion curve derivative. This significant increase, however, is not the correct behaviour as the results for the case including an extra orthogonality condition with respect to the eigenmodes on the second dispersion curve show. This spurious behaviour is even intensified if higher order derivatives are computed. For the tenth derivative as presented in Figure 4.1b it turned out that it is sufficient to add extra orthogonality conditions only for eigenvalues closer than $0.01 \cdot \frac{c^2}{a_1^2}$. If, however, one should be interested in computing even higher orders of the dispersion curve derivatives, one might has to increase the radius within eigenvalues are considered for extra orthogonality conditions.

Extra orthogonality conditions with respect to eigenmodes on close dispersion curves are thus compulsory for the computation of dispersion curve derivatives if other dispersion curves are not sufficiently far away.

4.2.4 Comparison of group velocity formula and difference quotient

Let us now discuss a possible drawback of the group velocity formula (4.5) in the context of a FE discretization. According to the Babuška-Osborn theory on eigenvalue problems [BO91], we expect that — using a FE discretization — the eigenvectors converge with smaller convergence rate than the eigenvalues when increasing the refinement of the discretization. Since approximations to the eigenmodes are needed to compute the group velocity using formula (4.5), we may expect that the convergence of the group velocity formula (4.5) when increasing the refinement of the discretization is of smaller rate than the convergence of the difference quotient of the dispersion curve, that is an approximation to the group velocity which only involves the eigenvalues and no eigenvectors.

In order to analyse this expectation, let us do a convergence study for the setup in Example 1. Figure 4.2 shows the convergence of the error of the group velocity formula (4.5) and the (first order) difference quotient of the first dispersion curve at $k = \frac{\pi}{2a_1}$ when increasing the mesh refinement of a FE discretization of polynomial degree one. The reference solution, on the other hand, is computed with the smallest mesh refinement and with polynomial degree 20. Both, the formula (4.5) for the group velocity as well as the difference quotient converge with the same convergence rate when increasing the refinement of the discretization, which demonstrates that the group velocity formula (4.5) has no disadvantages compared to a difference quotient in a FE discretization.

4.3 Proof of eigenmode differentiability

In Theorem 4.1 we already argued, using perturbation theory for linear operators [Kat95], that not only the eigenvalues $\omega_j^2(k)$ of (4.1) are analytic with respect to the quasi-momentum k, but also the phase and magnitude of the corresponding eigenmodes $u_j(\cdot\,; k)$ can be chosen such that the eigenmodes are analytic

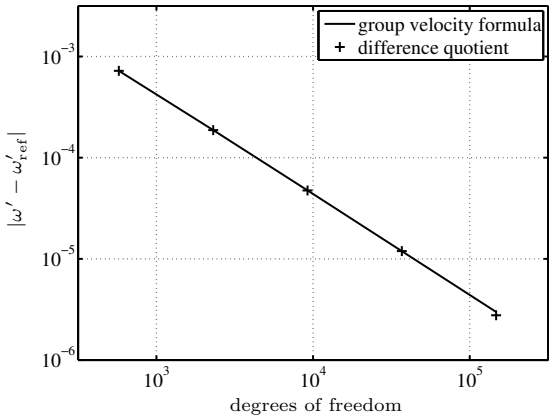


Figure 4.2: Convergence of the error of the group velocity formula (solid line) and the first order difference quotient (markers) when increasing the grid refinement of a FE computation with polynomial degree one. The reference solution is computed with polynomial degree 20.

with respect to the quasi-momentum. Nevertheless, we will now present a proof for the differentiability of the eigenmodes with respect to the quasi-momentum to any order. For this proof we will only argue with the help of the variational formulation of (4.1). However, we will restrict the proof to the case where the eigenvalue $\omega_j^2(k)$ has algebraic multiplicity one in a vicinity of $k = k_0 \in B$.

Then there exists $h_0 > 0$ such that for all $h \in]-h_0, h_0[$ the eigenvalue $\omega_j^2(k_0+h)$ of (4.1) has algebraic multiplicity one, and the eigenmode $u_j(\cdot\,; k_0+h)$ corresponding to the eigenvalue $\omega_j^2(k_0+h)$ is unique up to a complex-valued multiplicative factor. Hence, continuity and differentiability of this eigenmode with respect to k at $k = k_0$ is subject to a complex scaling of $u_j(\cdot\,; k_0+h)$ for all $h \in]-h_0, h_0[$.

Given that the eigenvalue $\omega_j^2(k)$ is analytic in k_0 we prove that the complex scaling can be chosen such that the eigenmode $u_j(\cdot\,; k)$ is differentiable at k_0 to any order.

We start by proving the continuity at k_0.

Lemma 4.6. *Let $u_j(\cdot\,; k_0+h)$, with $h \in]-h_0, h_0[$, be an arbitrary eigenmode of (4.1) at $k = k_0+h$ with $\mathsf{H}^1(C)$-norm one that corresponds to the eigenvalue $\omega_j^2(k_0+h)$. Then, in the limit $h \to 0$, $u_j(\cdot\,; k_0+h)$ is an eigenmode of (4.1) at $k = k_0$ corresponding to the eigenvalue $\omega_j^2(k_0)$ and $u_j(\cdot\,; k_0+h) \to \gamma u_j(\cdot\,; k_0)$ in $\mathsf{H}^1_{\mathrm{p}}(C)$ for some $\gamma \in \mathbb{C}$ with $|\gamma| = 1$.*

Proof. We know that $u_j(k_0) \equiv u_j(\cdot\,; k_0)$ satisfies

$$\mathfrak{b}_C(u_j(k_0), v; \omega_j(k_0), k_0) = 0$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$ and $u_j(k_0+h) \equiv u_j(\cdot\,; k_0+h)$ satisfies

$$\mathfrak{b}_C(u_j(k_0+h), v; \omega_j(k_0+h), k_0+h) = 0$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$. Consequently, the function $e_j(h) \equiv e_j(\cdot\,; h) := u_j(\cdot\,; k_0+h) - u_j(\cdot\,; k_0)$ satisfies

$$\mathfrak{b}_C(e_j(h), v; \omega_j, k_0) = \mathfrak{g}_h(v; k_0, \omega_j, u_j)$$

for all $v \in \mathsf{H}^1_{\mathrm{p}}(C)$ with the linear form

$$\begin{aligned}\mathfrak{g}_h(v; k_0, \omega_j, u_j) = \big(\omega_j^2(k_0+h) - \omega_j^2(k_0)\big)\, \mathfrak{m}_C^{\beta}(u_j(k_0+h), v) \\ - h\,(2k_0+h)\, \mathfrak{m}_C^{\alpha}(u_j(k_0+h), v) - h\, \mathfrak{c}_C^{\alpha,1}(u_j(k_0+h), v). \quad (4.12)\end{aligned}$$

Since ω_j^2 is continuous at $k = k_0$, we know that $\mathfrak{g}_h(v; k_0, \omega_j, u_j) \to 0$ as $h \to 0$. Hence, in the limit $h \to 0$ the function $e_j(\cdot\,; h)$ is either zero or it is an eigenmode of (4.1) at $k = k_0$ with corresponding

eigenvalue $\omega_j^2(k_0)$. Since the algebraic multiplicity of $\omega_j^2(k_0)$ is one, we can conclude that in both cases $\lim_{h\to 0} e_j(\cdot\,;h) = \widetilde{\gamma} u_j(\cdot\,;k_0)$ with some complex number $\widetilde{\gamma} \in \mathbb{C}$. This implies that $\lim_{h\to 0} u_j(\cdot\,;k_0+h) = \gamma u_j(\cdot\,;k_0)$, with $\gamma = \widetilde{\gamma} + 1$. Now we use the fact that the space of all functions in $\mathsf{H}^1_\mathrm{p}(C)$ with magnitude one is closed, and hence, it is — as a closed subspace of a Banach space — also complete. Now taking the norm of the previous equation, we can conclude that $|\gamma| = 1$, which finishes the proof. □

Now let us — in addition to the norm — also fix the phase of the eigenmode. To this end, we define $w := u_j(\cdot\,;k_0)$ as an arbitrary but fixed eigenmode of (4.1) at $k = k_0$ with $\mathsf{H}^1(C)$-norm one. Then we introduce the following problem: find $\widetilde{u}_j(k) \equiv \widetilde{u}_j(\cdot\,;k) \in \mathsf{H}^1_\mathrm{p}(C)$, $k = k_0 + h$, and the Lagrange multiplier $\lambda_{\widetilde{u}_j}(k) \in \mathbb{C}$ that satisfy

$$\mathfrak{b}_C(\widetilde{u}_j(k), v; \omega_j, k) + \lambda_{\widetilde{u}_j}(k)\langle w, v\rangle_{\mathsf{H}^1(C)} = 0, \tag{4.13a}$$

$$\langle \widetilde{u}_j(k), w\rangle_{\mathsf{H}^1(C)} = 1, \tag{4.13b}$$

for all $v \in \mathsf{H}^1_\mathrm{p}(C)$.

Lemma 4.7. *If k is sufficiently close to k_0, the problem (4.13) has a unique solution.*

Proof. First let us prove existence. Let $u_j(\cdot\,;k)$ denote an arbitrary eigenmode of (4.1), that is associated to $\omega_j^2(k)$ and normalized with respect to the $\mathsf{H}^1(C)$-norm, and let $\gamma(h) = \langle u_j(k), w\rangle_{\mathsf{H}^1(C)}$. From Lemma 4.6 we know that $u_j(\cdot\,;k) \to \gamma w$ as $h \to 0$ for some $\gamma \in \mathbb{C}$ with $|\gamma| = 1$. This implies that $\gamma(h) \to \gamma$ as $h \to 0$ and hence, $\gamma(h) \neq 0$ if k is sufficiently close to k_0. Then, $\widetilde{u}_j(\cdot\,;k) = \frac{1}{\gamma(h)} u_j(\cdot\,;k)$ solves (4.13) with the Lagrange multiplier $\lambda_{\widetilde{u}_j}(k) = 0$.

Now let us prove uniqueness. To this end, we assume that — apart from the eigenmode solution $(\widetilde{u}_{j,1}(\cdot\,;k), \lambda_{\widetilde{u}_{j,1}}(k)) = (\widetilde{u}_j(\cdot\,;k), 0) \in \mathsf{H}^1_\mathrm{p}(C) \times \mathbb{C}$ — there exists another solution $(\widetilde{u}_{j,2}(\cdot\,;k), \lambda_{\widetilde{u}_{j,2}}(k)) \in \mathsf{H}^1_\mathrm{p}(C) \times \mathbb{C}$ with $\widetilde{u}_{j,2}(\cdot\,;k) \not\equiv \widetilde{u}_{j,1}(\cdot\,;k)$ or $\lambda_{\widetilde{u}_{j,2}}(k) \neq \lambda_{\widetilde{u}_{j,1}}(k) = 0$.

Since $(\widetilde{u}_{j,2}(\cdot\,;k), \lambda_{\widetilde{u}_{j,2}}(k))$ is a solution of (4.13), we can write

$$\begin{aligned}\mathfrak{b}_C(\widetilde{u}_{j,2}(k), v; \omega_j, k) + \lambda_{\widetilde{u}_{j,2}}(k)\langle w, v\rangle_{\mathsf{H}^1(C)} &= 0,\\ \langle \widetilde{u}_{j,2}(k), w\rangle_{\mathsf{H}^1(C)} &= 1,\end{aligned}$$

for all $v \in \mathsf{H}^1_\mathrm{p}(C)$. Now we test the first equation with $v = \widetilde{u}_j(\cdot\,;k)$. Since this is an eigenmode of (4.1) and it satisfies the constraint condition (4.13b), we obtain $\lambda_{\widetilde{u}_{j,2}}(k) = 0$. However, this implies that $\widetilde{u}_{j,2}(k)$ is an eigenmode of (4.1). But from the constraint condition (4.13b) and the assumption that the two solutions are not identical, it follows that $\widetilde{u}_{j,2}(\cdot\,;k) \not\equiv \gamma \widetilde{u}_{j,1}(\cdot\,;k)$ with some $\gamma \in \mathbb{C}$. Hence, $\widetilde{u}_{j,2}(\cdot\,;k)$ is an eigenmode of (4.1) that is linear independent of the eigenmode $\widetilde{u}_{j,1}(\cdot\,;k)$, which is a contradiction to our assumption that the multiplicity of the eigenvalue problem (4.1) is one at $k \in]k_0 - h_0, k_0 + h_0[$. □

Now we are able to state the following result.

Corollary 4.8. *The unique solution $\widetilde{u}_j(\cdot\,;k)$ of (4.13) is an eigenmode of (4.1) with associated eigenvalue $\omega_j^2(k)$.*

Finally, we can prove continuity of $\widetilde{u}_j(\cdot\,;k)$ at $k = k_0$.

Lemma 4.9. *The eigenmode $\widetilde{u}_j(\cdot\,;k)$ that solves (4.13) is continuous at $k = k_0$.*

Proof. Let $\widetilde{e}_j(h) \equiv \widetilde{e}_j(\cdot\,;h) := \widetilde{u}_j(\cdot\,;k_0+h) - \widetilde{u}_j(\cdot\,;k_0)$, $h \in]-h_0, h_0[$. We introduce the Lagrange multiplier $\lambda_{\widetilde{e}_j}(h) = \lambda_{\widetilde{u}_j}(k_0 + h) - \lambda_{\widetilde{u}_j}(k_0)$ and hence,

$$\mathfrak{b}_C(\widetilde{e}_j(h), v; \omega_j, k_0) + \lambda_{\widetilde{e}_j}(h)\langle w, v\rangle_{\mathsf{H}^1(C)} = \mathfrak{g}_h(v; k_0, \omega_j, \widetilde{u}_j), \tag{4.14a}$$

$$\langle \widetilde{e}_j(h), w\rangle_{\mathsf{H}^1(C)} = \langle \widetilde{u}_j(k_0 + h) - \widetilde{u}_j(k_0), w\rangle_{\mathsf{H}^1(C)}, \tag{4.14b}$$

for all $v \in \mathsf{H}^1_\mathrm{p}(C)$, where the linear form $\mathfrak{g}_h$ is given in (4.12). The term on the right hand side of Eq. (4.14b) vanishes since both functions, $\widetilde{u}_j(\cdot\,;k_0+h)$ and $\widetilde{u}_j(\cdot\,;k_0)$, satisfy Eq. (4.13a). Since $k \mapsto \omega_j(k)$ is continuous at $k = k_0$ we conclude that the right hand side of Eq. (4.14a) tends to zero as $h \to 0$, and hence, — considering that the problem (4.14) is well-posed — we have $\widetilde{e}_j(\cdot\,;h) \to 0$ in $\mathsf{H}^1_\mathrm{p}(C)$ as $h \to 0$, which finishes the proof. □

In order to prove that $\widetilde{u}_j(\cdot\,;k)$ is differentiable at $k=k_0$ we introduce a new mixed variational problem: find $\widetilde{u}'_j(k)\equiv\widetilde{u}'_j(\cdot\,;k)\in\mathsf{H}^1_{\mathrm{p}}(C)$, $k=k_0+h$, and the Lagrange multiplier $\lambda_{\widetilde{u}'_j}(k)\in\mathbb{C}$ that satisfy

$$\mathfrak{b}_C(\widetilde{u}'_j(k),v;\omega_j,k)+\lambda_{\widetilde{u}'_j}(k)\langle w,v\rangle_{\mathsf{H}^1(C)}=\mathfrak{f}^{(1)}(v;k,\omega_j,\omega'_j,\widetilde{u}_j), \tag{4.15a}$$

$$\langle\widetilde{u}'_j(k),w\rangle_{\mathsf{H}^1(C)}=0 \tag{4.15b}$$

for all $v\in\mathsf{H}^1_{\mathrm{p}}(C)$ with the linear form $\mathfrak{f}^{(1)}$ as given in Eq. (4.3). Using the same arguments as in the proof of Lemma 4.7 we can conclude that (4.15) has a unique solution.

Remark 4.10. *Note that in the limit $h\to0$, the problem (4.15) is equivalent to (4.7). In this context, it becomes clear that in the limit $h\to0$ the Lagrange multiplier $\lambda_{\widetilde{u}'_j}(k)$ vanishes, see Remark 4.3. Since the compatibility condition of the mixed variational problem (4.15), i. e. taking the eigenmode $\widetilde{u}_j(\cdot\,;k)$ as test function in (4.15a), yields $\mathfrak{f}^{(1)}(\widetilde{u}_j(k);k,\omega_j,\omega'_j,\widetilde{u}_j)=\lambda_{\widetilde{u}'_j}(k)$, we can conclude that in the limit $h\to0$ the compatibility condition of (4.15) yields the group velocity formula (4.5) with $u_j=\widetilde{u}_j(\cdot\,;k)$.*

Lemma 4.11. *The eigenmode $\widetilde{u}_j(\cdot\,;k)$ that solves (4.13) is Fréchet differentiable with respect to k at $k=k_0$, and $\frac{\mathrm{d}}{\mathrm{d}k}\widetilde{u}_j(\cdot\,;k_0)=\widetilde{u}'_j(\cdot\,;k_0)$.*

Proof. Let $\widetilde{e}'_j(h)\equiv\widetilde{e}'_j(\cdot\,;h):=\frac{1}{h}\big(\widetilde{u}_j(\cdot\,;k_0+h)-\widetilde{u}_j(\cdot\,;k_0)-h\widetilde{u}'_j(\cdot\,;k_0)\big)$, $h\in]-h_0,h_0[$. Introducing the Lagrange multiplier $\lambda_{\widetilde{e}'_j}(h)=\frac{1}{h}\Big(\lambda_{\widetilde{u}_j}(k_0+h)-\lambda_{\widetilde{u}_j}(k_0)-h\lambda_{\widetilde{u}'_j}(k_0)\Big)$, we have

$$\mathfrak{b}_C(\widetilde{e}'_j(h),v;\omega_j,k)+\lambda_{\widetilde{e}'_j}(h)\langle w,v\rangle_{\mathsf{H}^1(C)}=\frac{1}{h}\mathfrak{g}_h(v;k_0,\omega_j,\widetilde{u}_j)-\mathfrak{f}^{(1)}(v;k_0,\omega_j,\omega'_j,\widetilde{u}_j), \tag{4.16a}$$

$$\langle\widetilde{e}'_j(h),w\rangle_{\mathsf{H}^1(C)}=\frac{1}{h}\langle\widetilde{u}_j(k_0+h)-\widetilde{u}_j(k_0),w\rangle_{\mathsf{H}^1(C)}, \tag{4.16b}$$

for all $v\in\mathsf{H}^1_{\mathrm{p}}(C)$. The term on the right hand side of Eq. (4.16b) vanishes since both functions, $\widetilde{u}_j(\cdot\,;k_0+h)$ and $\widetilde{u}_j(\cdot\,;k_0)$, satisfy Eq. (4.13a). Finally, — using the analyticity of the eigenvalue and Lemma 4.9 — we conclude that $\frac{1}{h}\mathfrak{g}_h(\cdot\,;k_0,\omega_j,\widetilde{u}_j)\to\mathfrak{f}^{(1)}(\cdot\,;k_0,\omega_j,\omega'_j,\widetilde{u}_j)$ as $h\to0$, and hence, — considering that the problem (4.16) is well-posed — we have $\widetilde{e}'_j(\cdot\,;h)\to0$ in $\mathsf{H}^1_{\mathrm{p}}(C)$ as $h\to0$, which finishes the proof. □

In order to extend the theory to higher orders we introduce $\widetilde{u}^{(n)}_j(k)\equiv\widetilde{u}^{(n)}_j(\cdot\,;k)\in\mathsf{H}^1_{\mathrm{p}}(C)$, $n\in\mathbb{N}$, $k=k_0+h$, as the unique solution of

$$\mathfrak{b}_C(\widetilde{u}^{(n)}_j(k),v;\omega_j,k)+\lambda_{\widetilde{u}^{(n)}_j}(k)\langle w,v\rangle_{\mathsf{H}^1(C)}=\mathfrak{f}^{(n)}(v;k,\omega_j^{(0)},\dots,\omega_j^{(n)},\widetilde{u}_j^{(0)},\dots,\widetilde{u}_j^{(n-1)}),$$

$$\langle\widetilde{u}^{(n)}_j(k),w\rangle_{\mathsf{H}^1(C)}=0,$$

for all $v\in\mathsf{H}^1_{\mathrm{p}}(C)$ with the Lagrange multiplier $\lambda_{\widetilde{u}^{(n)}_j}\in\mathbb{C}$. Using the same arguments as in the proof of Lemma 4.11 we conclude the following statement.

Lemma 4.12. *The eigenmode $\widetilde{u}_j(\cdot\,;k)$ that solves (4.13) is Fréchet differentiable with respect to k at $k=k_0$ up to any order, and $\frac{\mathrm{d}^n}{\mathrm{d}k^n}\widetilde{u}_j(\cdot\,;k_0)=\widetilde{u}^{(n)}_j(\cdot\,;k_0)$.*

This finishes the proof of the desired result. Since w can be chosen to be any eigenmode of (4.1) at $k=k_0$, we showed that we can choose the magnitude and phase of the eigenmodes of (4.1) associated to eigenvalues that have multiplicity one at $k=k_0$, such that the eigenmodes are continuously differentiable with respect to k at $k=k_0$ to any order.

4.4 Conclusions

In this chapter we derived closed formulas for the group velocity and any higher derivative of dispersion curves of PhC and PhC waveguide band structures.

The formulas rely on the differentiability of the eigenvalues and eigenmodes with respect to the quasi-momentum k, which is a classical result of perturbation theory for linear operators [Kat95] transferred to the variational formulation (4.1). Nevertheless, we provided a proof for the differentiability of the

eigenmodes in Section 4.3. In this proof we only argue with the help of the variational formulation of the eigenvalue problem (4.1). This proof is, however, restricted to the case where the associated eigenvalue as multiplicity one, and hence, it excludes crossings of dispersion curves.

The formulas for higher derivatives of the dispersion curves require the solution of source problems. These source problems vary from order to order only in their right hand side, which implies that the extra cost for computing higher orders is marginal.

In the vicinity of crossings, or whenever two eigenvalues are very close, the condition number of the system matrix related to these source problems becomes prohibitively large. Therefore, we proposed to add extra orthogonality conditions with respect to the eigenmodes of close dispersion curves. These extra orthogonality conditions help to keep the condition number at low levels while not spoiling the numerical results of the dispersion curve derivatives.

The numerical analysis of this effect as well as a mathematical proof for the observed convergence rate of the group velocity formula when enriching the FE discretization is subject to future research.

The results of this chapter will be transferred to a more general, discrete case in Chapter 5, and the formulas for the dispersion curve derivatives will be used in an adaptive Taylor expansion of dispersion curves, that we will also propose in Chapter 5. The generalization of the formulas for the dispersion curve derivatives to the case with DtN transparent boundary conditions for the exact computation of guided modes in PhC waveguides, see Eq. (2.20), will be addressed in Chapter 6.

5 Adaptive path following for parameterized, nonlinear eigenvalue problems

In Chapter 4 we showed that the dispersion curves of PhC and PhC waveguide band structuresare analytic and their derivatives to any order can be computed by closed formulas in variational sense. These facts motivate the approximation of the dispersion curves with a Taylor expansion. We will propose in this chapter an algorithm for an adaptive selection of the quasi-momenta for which Taylor expansions are computed.

Later in Chapters 6 and 7 we will apply this algorithm also to PhC waveguide band structure calculations when using DtN and RtR transparent boundary conditions for truncating the infinite, periodic domain. Note that our algorithm can also be applied to band structure calculations of waveguides with dispersive material like in [ER09, Eng14, SK10].

Before we introduce the adaptive algorithm and show numerical results, we want to generalize the results of Chapter 4 to parameterized, nonlinear eigenvalue problems in discrete form. To this end, we introduce a discrete problem in Section 5.1, before we derive the formulas for the derivatives of its eigenvalues with respect to a parameter up to any order in Section 5.2. Then in Section 5.3 we elaborate on the Taylor expansion and show first numerical results, before we introduce the adaptive algorithm in Section 5.4, whose numerical results, when applied to the dispersion curves of PhC and PhC waveguide band structures, are shown in Section 5.5. The conclusions of this chapter can finally be found in Section 5.6.

The adaptive algorithm was already published in an article together with K. Schmidt [KS14b]. The basic concept and first numerical results can also be found in [KS14a]. The generalization to parameterized, nonlinear eigenvalue problems, however, has not yet been published elsewhere, and the numerical results in Section 5.5 were extended by a convergence study.

5.1 Abstract problem setting

Let $\Omega_\lambda, \Omega_\mu \subseteq \mathbb{C}$ and let a matrix-valued, holomorphic function

$$\mathbf{N} : \Omega_\lambda \times \Omega_\mu \to \mathbb{C}^{N\times N},$$

$N \in \mathbb{N}$, be given. Then we consider the parameterized, nonlinear eigenvalue problem: for any $\mu \in \Omega_\mu$ find eigenvalues $\lambda = \lambda(\mu) \in \Omega_\lambda$ and associated right eigenvectors $\mathbf{u} = \mathbf{u}(\lambda, \mu) \in \mathbb{C}^N \setminus \{\mathbf{0}\}$ and left eigenvectors $\mathbf{v} = \mathbf{v}(\lambda, \mu) \in \mathbb{C}^N \setminus \{\mathbf{0}\}$ such that

$$\mathbf{N}(\lambda, \mu)\mathbf{u} = \mathbf{0} \tag{5.1a}$$

and

$$\mathbf{v}^{\mathrm{H}}\mathbf{N}(\lambda, \mu) = \mathbf{0}. \tag{5.1b}$$

In this chapter we aim to develop an algorithm for the efficient computation of approximations to the eigenvalues $\lambda(\mu)$ of (5.1) for parameters μ inside some curve $I_\mu \subset \Omega_\mu$ in the complex plane. Without loss of generality, we may set $I_\mu = (0, 1)$, since the results for $I_\mu = (0, 1)$ can be transferred to arbitrary real-valued intervals and, with the help of a suitable reparameterization, to arbitrary curves in the complex plane.

In Chapter 4 we argued that the eigenvalues $\omega^2(k)$ of the linear eigenvalue problem (4.1) in variational formulation related to PhC and PhC waveguide band structure calculations can be ordered such that the so-called dispersion curves $k \mapsto \omega_j(k)$ and their corresponding eigenmodes are analytic for all $k \in B$

Let us assume for now that the same is true for the eigenvalue problem (5.1).

Assumption 5.1. *The eigenvalues $\lambda_j(\mu)$, $j \in \mathbb{N}$, of (5.1) can be ordered such that the functions*

$$\mu \longmapsto \lambda_j(\mu), \tag{5.2}$$

the so-called eigenpaths*, are analytic in I_μ, and the corresponding right eigenvectors $\mathbf{u}_j(\mu) = \mathbf{u}(\lambda_j(\mu), \mu)$ are differentiable with respect to μ up to any order.*

Remark 5.2. *Again we note that from the theory presented in [ACL92, ACL93] we know that the eigenvalue λ_j and its associated right and left eigenvectors $\mathbf{u}_j$ and $\mathbf{v}_j$ are analytic in a neighbourhood of $\mu \in I_\mu$ if the eigenvalue $\lambda_j(\mu)$ has multiplicity one, which is equivalent to*

(i) $\mathbf{N}(\lambda_j, \mu)$ has rank $n-1$, and

(ii) $\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j, \mu)\mathbf{u}_j(\mu) \neq 0$.

Thus, if conditions (i) and (ii) are satisfied for all $\mu \in \Omega$, Assumption 5.1 is fulfilled.

In the sequel we shall propose a new adaptive eigenpath following algorithm for the nonlinear eigenvalue problem (5.1). The algorithm relies on Assumption 5.1 that the eigenpaths (5.2) are analytic and hence, a Taylor expansion of these eigenpaths is possible. This Taylor expansion can be understood as eigenvalue continuation technique. The homotopy method, see for example [LG95] for Hermitian problems and [LKK97] for non-Hermitian problems, is a well established technique to follow eigenpaths. The objective of the homotopy method, however, differs from our problem. The homotopy method aims to compute the eigenvalues of a matrix $\mathbf{M}_1$, say, that cannot by computed directly, either because its size is prohibitively large or it does not satisfy certain properties that are beneficial for the numerical computations of its eigenvalues, e. g. self-adjointness or linearity. The idea is to introduce an auxiliary matrix $\mathbf{M}_0$ of the same size like the original matrix $\mathbf{M}_1$, whose eigenvalues are either known or can be computed with significantly smaller effort than the eigenvalues of $\mathbf{M}_1$. Then one introduces the matrix-valued function

$$\mathbf{M}(t) = (1-t)\mathbf{M}_0 + t\mathbf{M}_1, \qquad 0 \leq t \leq 1,$$

and one follows the eigenvalues of $\mathbf{M}(t)$ from the known eigenvalues at $t = 0$ to the desired eigenvalues at $t = 1$. For our problem (5.1), however, we assume that the effort of the eigenvalue computation does not differ significantly for different values of the parameter μ. Hence, we do not aim to depart from some value of $\mu = \mu_0$ where the eigenvalues are known in order to get to some other state $\mu \neq \mu_0$. Instead, we are interested in the eigenpath itself, as for example in our context of PhC and PhC waveguide band structure computations, where the eigenpaths correspond to dispersion curves which form the band structure. Apart from the different objective of the homotopy method, our procedure has a considerable advantage. The Taylor expansion can be computed up to any order, since closed formulas for the eigenpath derivatives up to any order are available, while the homotopy method is, in general, a first order method, that does not take higher derivatives into account. A key feature of our method is to adaptively refine the step size of the path following. This is done by estimating the remainder of the Taylor expansion. For the homotopy method for non-Hermitian problems, Carstensen and coworkers [CGMM11] proposed an adaptive selection of the step size, where they also control the refinement of the FE computation. Such an adaptive FE refinement is is not needed for the application that we consider in this thesis, i. e. band structure calculations for PhCs and PhC waveguides with perfectly circular holes/rods, for example see the sketched geometries in Figures 2.6 and 2.8, since p-FEM on coarse meshes, that perfectly resolve the circular holes/rods, can be expected to converge exponentially, see Section 2.5. However, note that all ideas, that we will present in the following, can also be applied to hp-adaptive FE discretizations, which become crucial, for example, if the hols/rods of the PhC have corners, and hence, the eigenmodes show corner singularities that need to be resolved adaptively. In this case, it proves useful to stick to a FE refinement for the computation of all eigenmode derivatives associated to an eigenvalue, as we shall discuss later in Remark 5.4.

Before we shall present the Taylor expansion and the adaptive algorithm, we will derive the formulas for the eigenpath derivatives in the next section.

5.2 Derivatives of eigenpaths

Due to Assumption 5.1 that the eigenpaths (5.2) are analytic in I_μ and the associated right eigenvectors are differentiable up to any order, we can revisit all steps done in Chapter 4, where we derived formulas for the group velocity and all higher derivatives of the dispersion curves. Formulas for the first and second derivative of eigenpaths were already presented by Lancaster [Lan64]. Formulas for the derivatives of eigenpaths of arbitrary order were developed in [Jan94]. While details for the computation of derivatives especially of higher order are rare in [Jan94], we aim to present the computational procedure in full detail. Moreover, our approach allows the eigenvalues in principle to have multiplicity larger than one, which is explicitly excluded in [Jan94]. If the eigenvalue multiplicity is one, the left and right eigenvectors of (5.1) are uniquely defined up to a scalar constant. If the multiplicity is larger than one, this is not the case as any linear combination of two eigenvectors associated to the same eigenvalue is also an eigenvector. For simplicity and in accordance to the application under consideration, i. e. PhC and PhC waveguide band structure calculations, we assume that the eigenvalue multiplicity is one almost everywhere in I_μ. In situations where this assumption does not hold true we refer to the procedures presented in [AT98, QACT13] for symmetric, linear and quadratic eigenvalue problems, and to the procedure in [AMM07] for general, complex linear eigenvalue problems. If the eigenvalue problem, however, is nonlinear there does not exist — to the best of our knowledge — a procedure of how to uniquely identify eigenvectors of multiple eigenvalues. Recall from Remark 4.2 that we claim that the eigenvectors at crossings of dispersion curves can be determined as the limits of the eigenvectors in the vicinity of the crossing.

5.2.1 First derivative of eigenpaths

We start by differentiating (5.1a) with respect to μ which yields

$$\left(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)+\mathbf{N}_\mu(\lambda_j,\mu)\right)\mathbf{u}_j(\mu)+\mathbf{N}(\lambda_j,\mu)\mathrm{d}_\mu\mathbf{u}_j(\mu)=\mathbf{0}, \tag{5.3}$$

where $\mathbf{N}_\lambda$ and $\mathbf{N}_\mu$ are the partial derivatives of $\mathbf{N}$ with respect to λ and μ, and $\mathrm{d}_\mu\mathbf{u}_j$ is the total derivative of $\mathbf{u}_j$ with respect to μ. Multiplying (5.3) from the left with the left eigenvector $\mathbf{v}_j(\mu)$ and considering (5.1b), we arrive at

$$\lambda_j'(\mu)\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)\mathbf{u}_j(\mu)+\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\mu(\lambda_j,\mu)\mathbf{u}_j(\mu)=0.$$

Thus, we obtain a formula for the derivative $\lambda_j'(\mu)$ of the eigenvalue $\lambda_j(\mu)$ with respect to the parameter μ. It reads

$$\lambda_j'(\mu)=-\frac{\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\mu(\lambda_j,\mu)\mathbf{u}_j(\mu)}{\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)\mathbf{u}_j(\mu)} \tag{5.4}$$

and is the abstract, discrete analogue of the group velocity formula (4.5) for nonlinear eigenvalue problems.

Remark 5.3. *From the formula (5.4) for the derivative $\lambda_j'(\mu)$ of the eigenvalue $\lambda_j(\mu)$ with respect to μ we can see that*

$$\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)\mathbf{u}_j(\mu)\neq 0$$

is a necessary condition for the analyticity of the eigenpath $\mu\mapsto\lambda_j(\mu)$, i. e. a necessary condition for Assumption 5.1 to be satisfied.

The matrix that corresponds to $\mathbf{N}_\lambda$ in the context of PhC and PhC waveguide band structure calculations, in the latter case when using the supercell approach, is the mass matrix, i. e. the matrix that is related to the sesquilinear form

$$\mathfrak{m}_C^\beta(u,v)=\int_C \beta u\overline{v}\,\mathrm{d}\mathbf{x},$$

multiplied with (-2λ), where λ denotes the frequency ω. Since the mass matrix is positive definite, the matrix $\mathbf{N}_\lambda$ is negative definite in this context as long as the frequency ω is strictly positive, which is by definition the case, see Section 2.1, and hence, the matrix $\mathbf{N}_\lambda$ fulfills the condition in Remark 5.3.

5.2.2 Higher derivatives of eigenpaths

In order to extend this procedure to higher orders of the derivatives of $\lambda_j(\mu)$ we have to compute an auxiliary vector $\mathbf{u}_j^{(1)}(\mu)$, that is associated to the total derivative $\mathrm{d}_\mu \mathbf{u}_j(\mu)$ of the eigenvector $\mathbf{u}_j(\mu)$ corresponding to $\lambda_j(\mu)$ in the sense that $\mathbf{u}_j^{(1)}(\mu) = \mathrm{d}_\mu \mathbf{u}_j(\mu) + c\mathbf{u}_j(\mu)$, $c \in \mathbb{C}$, and hence, $\mathbf{N}(\lambda_j,\mu)\mathbf{u}_j^{(1)}(\mu) = \mathbf{N}(\lambda_j,\mu)\mathrm{d}_\mu \mathbf{u}_j(\mu)$.

Since the derivative $\lambda_j'(\mu)$ of the eigenvalue does not depend on the derivative $\mathrm{d}_\mu \mathbf{u}_j(\mu)$ of the eigenvector, we can plug (5.4) into (5.3) and obtain the source problem

$$\mathbf{N}(\lambda_j,\mu)\mathrm{d}_\mu \mathbf{u}_j(\mu) = \mathbf{f}^{(1)}(\mu) \tag{5.5}$$

with the vector of the right hand side

$$\mathbf{f}^{(1)}(\mu) = -\left(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + \mathbf{N}_\mu(\lambda_j,\mu)\right)\mathbf{u}_j(\mu). \tag{5.6}$$

Similarly to Chapter 4 we have to note that (5.5) does not admit a unique solution, since $\mathbf{N}(\lambda_j,\mu)$ is singular. However, additionally imposing orthogonality conditions with respect to some inner product to all linearly independent right eigenvectors of $\mathbf{N}(\lambda_j,\mu)$ corresponding to the eigenvalue $\lambda_j(\mu)$, we can construct a linear system that is well-posed. As mentioned above, let us assume that $\lambda_j(\mu)$ is a simple eigenvalue, i. e. there is only one linearly independent right eigenvector $\mathbf{u}_j(\mu)$ corresponding to $\lambda_j(\mu)$. Then the linear system

$$\begin{pmatrix} \mathbf{N}(\lambda_j,\mu) & \mathbf{Q}\mathbf{u}_j(\mu) \\ \mathbf{u}_j^{\mathrm{H}}(\mu)\mathbf{Q} & \end{pmatrix} \begin{pmatrix} \mathbf{u}_j^{(1)}(\mu) \\ \nu \end{pmatrix} = \begin{pmatrix} \mathbf{f}^{(1)}(\mu) \\ 0 \end{pmatrix}, \tag{5.7}$$

with some symmetric, positive definite matrix $\mathbf{Q} \in \mathbb{R}^{N\times N}$, has a unique solution, where $\nu \in \mathbb{C}$ is a scalar Lagrange multiplier and $\mathbf{u}_j^{(1)}(\mu) \in \mathbb{C}^N$ denotes a particular solution of (5.5), which — in general — cannot be regarded as the derivative of the right eigenvector. We refer to the discussion in Section 4.2.2 why the choice of the orthogonality condition is arbitrary and hence, it is sufficient to compute a particular solution of (5.5) instead of properly defining and computing the derivative of the right eigenvector with respect to the quasi-momentum.

If the eigenvalue $\lambda_j(\mu)$ has multiplicity larger than one, we have to impose orthogonality conditions to an appropriate basis of the eigenspace and carefully select the right eigenvector $\mathbf{u}_j(\mu)$. As mentioned above, we may use the procedure sketched in [AMM07] if the eigenvalue problem is linear, or as presented in [AT98, QACT13] if the problem is symmetric and linear or symmetric and quadratic, respectively. For general, nonlinear problems no such procedure exists, but for the problem under consideration, i. e. PhC and PhC waveguide band structures, we claimed in Remark 4.2 that we can take the limits of the eigenvectors in the vicinity of the crossing of dispersion curves.

Now we repeat these steps, i. e. we differentiate (5.3) with respect to μ, which yields

$$\begin{aligned}\Big(\lambda_j''(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + (\lambda_j'(\mu))^2\mathbf{N}_{\lambda\lambda}(\lambda_j,\mu) + \lambda_j'(\mu)\mathbf{N}_{\mu\lambda}(\lambda_j,\mu) + \mathbf{N}_{\mu\mu}(\lambda_j,\mu)\Big)\mathbf{u}_j(\mu) \\ + 2\Big(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + \mathbf{N}_\mu(\lambda_j,\mu)\Big)\mathrm{d}_\mu \mathbf{u}_j(\mu) + \mathbf{N}(\lambda_j,\mu)\mathrm{d}_\mu^2 \mathbf{u}_j(\mu) = \mathbf{0},\end{aligned} \tag{5.8}$$

where we use the short notations $\mathbf{N}_{\lambda\lambda}$, $\mathbf{N}_{\mu\lambda}$, $\mathbf{N}_{\mu\mu}$ for the second, partial derivatives $\frac{\partial^2}{\partial\lambda^2}\mathbf{N}$, $\frac{\partial^2}{\partial\mu\partial\lambda}\mathbf{N}$, $\frac{\partial^2}{\partial\mu^2}\mathbf{N}$ of $\mathbf{N}$, and $\mathrm{d}_\mu^2\mathbf{u}_j(\mu)$ denotes the second, total derivative of $\mathbf{u}_j(\mu)$ with respect to μ. We multiply (5.8) from the left with the left eigenvector $\mathbf{v}_j(\mu)$ of (5.1b) and obtain a formula for the second derivative $\lambda_j''(\mu)$ of the eigenvalue $\lambda_j(\mu)$ of (5.1) with respect to the parameter μ

$$\begin{aligned}\lambda_j''(\mu) = &-\frac{\mathbf{v}_j^{\mathrm{H}}(\mu)\Big((\lambda_j'(\mu))^2\mathbf{N}_{\lambda\lambda}(\lambda_j,\mu) + \lambda_j'(\mu)\mathbf{N}_{\mu\lambda}(\lambda_j,\mu) + \mathbf{N}_{\mu\mu}(\lambda_j,\mu)\Big)\mathbf{u}_j(\mu)}{\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)\mathbf{u}_j(\mu)} \\ &- \frac{\mathbf{v}_j^{\mathrm{H}}(\mu)\Big(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + \mathbf{N}_\mu(\lambda_j,\mu)\Big)\mathbf{u}_j^{(1)}(\mu)}{\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)\mathbf{u}_j(\mu)},\end{aligned} \tag{5.9}$$

where we could replace $\mathrm{d}_\mu \mathbf{u}_j(\mu)$ by $\mathbf{u}_j^{(1)}(\mu)$ since $\mathbf{u}_j^{(1)}(\mu) = \mathrm{d}_\mu \mathbf{u}_j(\mu) + c\mathbf{u}_j(\mu)$, $c \in \mathbb{C}$, and

$$\mathbf{v}_j^{\mathrm{H}}(\mu)\Big(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + \mathbf{N}_\mu(\lambda_j,\mu)\Big)\mathbf{u}_j(\mu) = 0$$

due to (5.4).

Analogously to (5.7), we can compute a particular solution of (5.8), the auxiliary vector $\mathbf{u}_j^{(2)}(\mu)$, that is associated with the second μ-derivative $\mathrm{d}_\mu^2 \mathbf{u}_j(\mu)$ of the eigenvector $\mathbf{u}_j(\mu)$ corresponding to $\lambda_j(\mu)$. We obtain the linear system

$$\begin{pmatrix} \mathbf{N}(\lambda_j,\mu) & \mathbf{Q}\mathbf{u}_j(\mu) \\ \mathbf{u}_j^{\mathrm{H}}(\mu)\mathbf{Q} & \end{pmatrix} \begin{pmatrix} \mathbf{u}_j^{(2)}(\mu) \\ \nu \end{pmatrix} = \begin{pmatrix} \mathbf{f}^{(2)}(\mu) \\ 0 \end{pmatrix}, \tag{5.10}$$

with the vector of the right hand side

$$\begin{aligned}\mathbf{f}^{(2)}(\mu) = -\Big(\lambda_j''(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + (\lambda_j'(\mu))^2\mathbf{N}_{\lambda\lambda}(\lambda_j,\mu) + \lambda_j'(\mu)\mathbf{N}_{\mu\lambda}(\lambda_j,\mu) + \mathbf{N}_{\mu\mu}(\lambda_j,\mu)\Big)\mathbf{u}_j(\mu) \\ - 2\Big(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + \mathbf{N}_\mu(\lambda_j,\mu)\Big)\mathbf{u}_j^{(1)}(\mu).\end{aligned} \tag{5.11}$$

Remark 5.4. *The matrices of the linear problems (5.7) and (5.10) are identical, and hence, matrix factorizations computed to solve (5.7) can be reused to solve (5.10). This shows, that it is beneficial to stick to a FE refinement for the computation of the eigenmode $\mathbf{u}_j$, the auxiliary vectors $\mathbf{u}_j^{(1)}$ and $\mathbf{u}_j^{(2)}$ associated to the first and second eigenmode derivatives, and, as we shall see in the following, all subsequent auxiliary vectors $\mathbf{u}_j^{(n)}$, $n \geq 3$, associated to higher order eigenmode derivatives.*

This procedure can be repeated successively to find formulas for the derivatives of the eigenvalue $\lambda_j(\mu)$ with respect to the parameter μ up to any order. In the PhC context, we studied in Chapter 4, the matrix $\mathbf{N}$ has polynomial dependences on the parameter $\mu = k$ and eigenvalue $\lambda = \omega$ of order two, which allows for straightforward generalizations of the formulas for the derivatives up to any order. However, in the general, discrete case, we investigate in this chapter, the formulas are not that straightforward as one has to use a multivariant version [CS96] of Faà di Bruno's formula [FdB57]. Taking the n-th derivative of the eigenvalue problem (5.1) with respect to the parameter μ yields

$$\mathrm{d}_\mu^n\Big(\mathbf{N}(\lambda_j,\mu)\mathbf{u}_j(\mu)\Big) = \sum_{m=0}^{n} \binom{n}{m} \mathrm{d}_\mu^{n-m}\mathbf{N}(\lambda_j,\mu)\mathrm{d}_\mu^m \mathbf{u}_j(\mu) = \mathbf{0},$$

i. e.

$$\mathbf{N}(\lambda_j,\mu)\mathrm{d}_\mu^n \mathbf{u}_j(\mu) = \mathbf{f}^{(n)}(\mu) \tag{5.12}$$

where

$$\mathbf{f}^{(n)}(\mu) = -\sum_{m=0}^{n-1} \binom{n}{m} \mathrm{d}_\mu^{n-m}\mathbf{N}(\lambda_j,\mu)\mathbf{u}_j^{(m)}(\mu) \tag{5.13}$$

with the auxiliary vectors $\mathbf{u}_j^{(m)}(\mu)$, $1 \leq m \leq n-1$, associated to the eigenvector derivatives $\mathrm{d}_\mu^m \mathbf{u}_j(\mu)$, and with $\mathbf{u}_j^{(0)}(\mu) = \mathbf{u}_j(\mu)$.

According to the multivariant version of Faà di Bruno's formula presented in [CS96], the total derivative $\mathrm{d}_\mu^n \mathbf{N}(\lambda_j,\mu)$ can be expanded in the form

$$\mathrm{d}_\mu^n \mathbf{N}(\lambda_j,\mu) = \sum_{\substack{m_1,m_2\in\mathbb{N}_0,\\ 1\leq m_1+m_2\leq n}} \frac{\partial^{m_1+m_2}}{\partial\lambda^{m_1}\,\partial\mu^{m_2}}\mathbf{N}(\lambda_j,\mu) \sum_{\mathfrak{N}(n,m_1,m_2)} n! \prod_{i=1}^{n} \frac{\left(\partial_\mu^{\ell_i}\lambda\right)^{q_{1,i}}\left(\partial_\mu^{\ell_i}\mu\right)^{q_{2,i}}}{q_{1,i}!\,q_{2,i}!\,(\ell_i!)^{q_{1,i}+q_{2,i}}}, \tag{5.14}$$

where by convention $0! = 0^0 = 1$, and the set $\mathfrak{N}(n, m_1, m_2)$ reads

$$\begin{aligned}\mathfrak{N}(n, m_1, m_2) = \Bigg\{ &(q_{1,1}, \dots, q_{1,n}, q_{2,1}, \dots, q_{2,n}, \ell_1, \dots, \ell_n) \in \mathbb{N}_0^{3n} \;\Big|\; \exists s \in \{1, \dots, n\} \text{ such that} \\ &q_{1,i} = q_{2,i} = \ell_i = 0 \quad \forall i \in \{1, \dots, n-s\}, \\ &q_{1,i} + q_{2,i} > 0 \quad \forall i \in \{n-s+1, \dots, n\}, \\ &0 < \ell_{n-s+1} < \ldots < \ell_n, \quad \text{and} \\ &\sum_{i=1}^{n} q_{1,i} = m_1, \quad \sum_{i=1}^{n} q_{2,i} = m_2, \quad \sum_{i=1}^{n} (q_{1,i} + q_{2,i})\ell_i = n \Bigg\}.\end{aligned}$$

Due to the term $\left(\partial_\mu^{\ell_i} \mu\right)^{q_{2,i}}$ in (5.14), that takes the form

$$\left(\partial_\mu^{\ell_i} \mu\right)^{q_{2,i}} = \begin{cases} 1, & \text{if } \ell_i = 1 \text{ or if } \ell_i > 1 \text{ and } q_{2,i} = 0, \\ 0, & \text{otherwise,} \end{cases}$$

we can simplify (5.14) to

$$\mathrm{d}_\mu^n \mathbf{N}(\lambda_j, \mu) = \sum_{\substack{m_1, m_2 \in \mathbb{N}_0, \\ 1 \le m_1 + m_2 \le n}} \frac{\partial^{m_1+m_2}}{\partial \lambda^{m_1}\, \partial \mu^{m_2}} \mathbf{N}(\lambda_j, \mu) \sum_{\widetilde{\mathfrak{N}}(n, m_1, m_2)} n! \prod_{i=1}^{n} \frac{\left(\partial_\mu^{\ell_i} \lambda_j\right)^{q_{1,i}}}{q_{1,i}!\; q_{2,i}!\; (\ell_i!)^{q_{1,i}+q_{2,i}}} \tag{5.15}$$

with the set

$$\begin{aligned}\widetilde{\mathfrak{N}}(n, m_1, m_2) = \Bigg\{ &(q_{1,1}, \dots, q_{1,n}, q_{2,1}, \dots, q_{2,n}, \ell_1, \dots, \ell_n) \in \mathbb{N}_0^{3n} \;\Big|\; \exists s \in \{1, \dots, n\} \text{ such that} \\ &q_{1,i} = q_{2,i} = \ell_i = 0 \quad \forall i \in \{1, \dots, n-s\}, \\ &q_{1,i} > 0 \text{ and } q_{2,i} = 0 \quad \forall i \in \{n-s+2, \dots, n\}, \\ &q_{2,n-s+1} = m_2, \\ &0 < \ell_{n-s+1} < \ldots < \ell_n, \text{ where } \ell_{n-s+1} = 1 \text{ if } m_2 > 0, \text{ and} \\ &\sum_{i=1}^{n} q_{1,i} = m_1, \quad \sum_{i=1}^{n} q_{1,i}\ell_i = n - m_2 \Bigg\}.\end{aligned} \tag{5.16}$$

Now we multiply (5.12) with $\mathbf{v}_j^{\mathrm{H}}(\mu)$ from the left and obtain

$$\sum_{m=0}^{n-1} \binom{n}{m} \mathbf{v}_j^{\mathrm{H}}(\mu) \mathrm{d}_\mu^{n-m} \mathbf{N}(\lambda_j, \mu) \mathrm{d}_\mu^m \mathbf{u}_j(\mu) = 0.$$

Using the expansion (5.15), we find that the n-th derivative $\partial_\mu^n \lambda_j(\mu)$ of the eigenvalue $\lambda_j(\mu)$ with respect to μ can be expressed as

$$\begin{aligned}\partial_\mu^n \lambda_j(\mu) = -\Big(\mathbf{v}_j^{\mathrm{H}}(\mu) \mathbf{N}_\lambda(\lambda_j, \mu) \mathbf{u}_j(\mu)\Big)^{-1} \Bigg(&\sum_{m=1}^{n-1} \binom{n}{m} \mathbf{v}_j^{\mathrm{H}}(\mu) \mathrm{d}_\mu^{n-m} \mathbf{N}(\lambda_j, \mu) \mathbf{u}_j^{(m)}(\mu) \\ &+ \sum_{\substack{m_1, m_2 \in \mathbb{N}_0, \\ 1 \le m_1 + m_2 \le n}} \mathbf{v}_j^{\mathrm{H}}(\mu) \frac{\partial^{m_1+m_2}}{\partial \lambda^{m_1}\, \partial \mu^{m_2}} \mathbf{N}(\lambda_j, \mu) \mathbf{u}_j(\mu) \sum_{\widetilde{\widetilde{\mathfrak{N}}}(n, m_1, m_2)} n! \prod_{i=1}^{n} \frac{\left(\partial_\mu^{\ell_i} \lambda_j\right)^{q_{1,i}}}{q_{1,i}!\; q_{2,i}!\; (\ell_i!)^{q_{1,i}+q_{2,i}}} \Bigg)\end{aligned} \tag{5.17}$$

with the set

$$\begin{aligned}\widetilde{\widetilde{\mathfrak{N}}}(n,m_1,m_2) = \Bigg\{ &(q_{1,1},\dots,q_{1,n},q_{2,1},\dots,q_{2,n},\ell_1,\dots,\ell_n) \in \mathbb{N}_0^{3n} \;\Big|\; \exists s \in \{1,\dots,n\} \text{ such that} \\ &q_{1,i} = q_{2,i} = \ell_i = 0 \quad \forall i \in \{1,\dots,n-s\}, \\ &q_{1,i} > 0 \text{ and } q_{2,i} = 0 \quad \forall i \in \{n-s+2,\dots,n\}, \\ &q_{2,n-s+1} = m_2, \\ &0 < \ell_{n-s+1} < \ldots < \ell_n < n, \text{ where } \ell_{n-s+1} = 1 \text{ if } m_2 > 0, \text{ and} \\ &\sum_{i=1}^{n} q_{1,i} = m_1, \quad \sum_{i=1}^{n} q_{1,i}\ell_i = n - m_2 \Bigg\}.\end{aligned}$$

Note that we replaced the derivatives $\mathrm{d}_\mu^m \mathbf{u}_j(\mu)$ of the right eigenvector $\mathbf{u}_j(\mu)$ in (5.17) by the auxiliary vectors $\mathbf{u}_j^{(m)}(\mu)$.

Similarly to the linear systems (5.7) and (5.10), we add an orthogonality condition to the ill-posed problem (5.12), and finally, we can compute the auxiliary vector $\mathbf{u}_j^{(n)}(\mu)$ associated to the n-th derivative $\mathrm{d}_\mu^n \mathbf{u}_j(\mu)$ with respect to μ of the eigenvector $\mathbf{u}_j(\mu)$ corresponding to $\lambda_j(\mu)$ by solving

$$\begin{pmatrix} \mathbf{N}(\lambda_j,\mu) & \mathbf{Q}\mathbf{u}_j(\mu) \\ \mathbf{u}_j^{\mathrm{H}}(\mu)\mathbf{Q} & \end{pmatrix} \begin{pmatrix} \mathbf{u}_j^{(n)}(\mu) \\ \nu \end{pmatrix} = \begin{pmatrix} \mathbf{f}^{(n)}(\mu) \\ 0 \end{pmatrix}, \tag{5.18}$$

where the vector $\mathbf{f}^{(n)}(\mu)$ is given in (5.13). Note that the matrix on the left hand side of (5.18) is identical for all orders $n \in \mathbb{N}$, and hence, matrix factorizations, computed for $n = 1$, can be reused for all orders $n > 1$, cf. Remark 5.4.

The implementation of the multivariant version of Faà di Bruno's formula in (5.15) and (5.17) is very technical. Therefore, we shall introduce a recursive algorithm as an alternative to compute the total derivative $\mathrm{d}_k^n \mathbf{N}$. Note that we can write

$$\begin{aligned}\mathrm{d}_\mu^n \mathbf{N}(\lambda_j,\mu) &= \mathrm{d}_\mu^{n-1}\left(\lambda_j'(\mu)\mathbf{N}_\lambda(\lambda_j,\mu) + \mathbf{N}_\mu(\lambda_j,\mu)\right) \\ &= \mathrm{d}_\mu^{n-1}\mathbf{N}_\mu(\lambda_j,\mu) + \sum_{m=0}^{n-1} \binom{n-1}{m} \partial_\mu^{m+1}\lambda_j(\mu)\, \mathrm{d}_\mu^{n-m-1}\mathbf{N}_\lambda(\lambda_j,\mu)\end{aligned} \tag{5.19}$$

for all $n \in \mathbb{N}$, which motivates the recursive algorithm sketched in Algorithm 5.1.

Apart from this algorithm, Eq. (5.19) also shows that we can write the n-th derivative $\partial_\mu^n \lambda_j(\mu)$ of the eigenvalue $\lambda_j(\mu)$ with respect to μ, that is presented in (5.17) using the multivariant Faà di Bruno formula, in an alternative way, i. e.

$$\begin{aligned}\partial_\mu^n \lambda_j(\mu) = -\Big(\mathbf{v}_j^{\mathrm{H}}(\mu)\mathbf{N}_\lambda(\lambda_j,\mu)\mathbf{u}_j(\mu)\Big)^{-1} \Bigg(&\sum_{m=1}^{n-1} \binom{n}{m} \mathbf{v}_j^{\mathrm{H}}(\mu)\mathrm{d}_\mu^{n-m}\mathbf{N}(\lambda_j,\mu)\mathbf{u}_j^{(m)}(\mu) \\ &+ \mathbf{v}_j^{\mathrm{H}}(\mu)\left(\mathrm{d}_\mu^{n-1}\mathbf{N}_\mu(\lambda_j,\mu) + \sum_{m=0}^{n-2} \binom{n-1}{m} \partial_\mu^{m+1}\lambda_j(\mu)\, \mathrm{d}_\mu^{n-m-1}\mathbf{N}_\lambda(\lambda_j,\mu)\right)\mathbf{u}_j(\mu)\Bigg),\end{aligned} \tag{5.20}$$

where the total derivatives $\mathrm{d}_\mu^m \mathbf{N}$, $\mathrm{d}_\mu^m \mathbf{N}_\lambda$ and $\mathrm{d}_\mu^m \mathbf{N}_\mu$, for $0 \le m \le n-1$, can be evaluated recursively using Algorithm 5.1.

Remark 5.5. *If the matrix $\mathbf{N}$ is Hermitian for all $\mu \in I_\mu$ and $\lambda \in \Omega_\lambda$, the left eigenvector $\mathbf{v}_j$ is identical to the right eigenvector $\mathbf{u}_j$ and hence, we do not need to solve (5.1b) for the left eigenvector.*

5.2.3 Discretization of the formulas for the dispersion curve derivatives

Before we proceed with the Taylor expansion of eigenpaths where we will employ the eigenpath derivatives (5.20), we want to be present briefly the discrete formulas for the dispersion curve derivatives which

Algorithm 5.1. Recursion algorithm for the computation of the total derivative $\mathrm{d}_k^n \mathbf{N}(\lambda_j, \mu)$, $n \in \mathbb{N}_0$.

Requirements: Let the partial derivatives of $\mathbf{N}(\lambda_j, \mu)$ with respect to λ and μ be stored in a triangular array `pN` with entries $\texttt{pN}(m, \ell) = \frac{\partial^{m+\ell}}{\partial \lambda^m \partial \mu^\ell} \mathbf{N}(\lambda_j, \mu)$ for all $m, \ell \in \mathbb{N}_0$ with $m + \ell \leq n$. Furthermore, let the derivatives of $\lambda_j(\mu)$ with respect to μ be given for all orders $m = 0, \ldots, n$.

function RECURSION($n, \texttt{pN}, \partial_\mu^0 \lambda_j, \ldots, \partial_\mu^n \lambda_j$)
  if $n = 0$ **then**
    return $\texttt{pN}(0, 0)$
  else if $n = 1$ **then**
    return $\partial_\mu \lambda_j \cdot \texttt{pN}(1, 0) + \texttt{pN}(0, 1)$
  else
    $\texttt{tN} =$RECURSION($n - 1, \texttt{pN}(0 : n - 1, 1 : n), \partial_\mu^0 \lambda_j, \ldots, \partial_\mu^{n-1} \lambda_j$)
    for $m = 0, \ldots, n - 1$ **do**
      $\texttt{tN} \mathrel{+}= \binom{n-1}{m} \partial_\mu^{m+1} \lambda_j \cdot$ RECURSION($n - 1, \texttt{pN}(1 : n, 0 : n - 1), \partial_\mu^0 \lambda_j, \ldots, \partial_\mu^{n-1} \lambda_j$)
    end for
    return `tN`
  end if
end function

we introduced in Chapter 4 in variational sense. These discrete formulas will later be used in this chapter for numerical applications of the proposed Taylor expansion and adaptive algorithm.

Using the FE discretizations of the 2d PhC eigenvalue problem and the supercell approximation to the eigenvalue problem in 2d PhC waveguides, which we introduced in Eqs. (2.28) and (2.29), we find that the discrete version of the group velocity formula (4.5) reads

$$\omega_j'(k) = \frac{\mathbf{u}_j^{\mathrm{H}}(k) \mathbf{g}^{(1)}(k)}{\mathbf{u}_j^{\mathrm{H}}(k) \mathbf{h}(k)}, \tag{5.21}$$

with

$$\mathbf{g}^{(1)}(k) = \mathbf{g}^{(1)}(k, \mathbf{u}_j(k)) = \left(2k \mathbf{M}_C^\alpha + \mathbf{C}_C^{\alpha,1} \right) \mathbf{u}_j(k)$$

and

$$\mathbf{h}(k) = \mathbf{h}(k, \mathbf{u}_j(k)) = 2\omega_j(k) \mathbf{M}_C^\beta \mathbf{u}_j(k),$$

where $\mathbf{u}_j(k)$ denotes an eigenvector related to the eigenvalue $\omega_j(k)$ of the PhC problem (2.28) with $k_1 = k$ and $k_2 = 0$, or the supercell problem (2.29). Then we compute the auxiliary vector $\mathbf{u}_j^{(1)}(k)$ associated to the first derivative of the eigenvector $\mathbf{u}_j(k)$ with respect to k by solving

$$\begin{pmatrix} \mathbf{A}_C^\alpha + k\mathbf{C}_C^{\alpha,1} + k^2 \mathbf{M}_C^\alpha - \omega^2 \mathbf{M}_C^\beta & (\mathbf{A}_C^1 + \mathbf{M}_C^1) \mathbf{u}_j(k) \\ \mathbf{u}_j^{\mathrm{H}}(k) (\mathbf{A}_C^1 + \mathbf{M}_C^1) & \end{pmatrix} \begin{pmatrix} \mathbf{u}_j^{(1)}(k) \\ \nu \end{pmatrix} = \begin{pmatrix} \omega_j'(k) \mathbf{h}(k) - \mathbf{g}^{(1)}(k) \\ 0 \end{pmatrix},$$

which is the discretization of the mixed variational formulation (4.7). The formula (5.21) can be extended to higher orders without the need of the multivariant version (5.15) of Faà di Bruno's formula or the recursion formula (5.19) and Algorithm 5.1, since the eigenvalue problems (2.28) and (2.29) only show a second-order polynomial dependence on ω and k. In accordance to the formula (4.9) in variational sense, we find

$$\omega_j^{(n)}(k) = \frac{\mathbf{u}_j^{\mathrm{H}}(k) \mathbf{g}^{(n)}(k)}{\mathbf{u}_j^{\mathrm{H}}(k) \mathbf{h}(k)} \tag{5.22}$$

with

$$\begin{aligned}\mathbf{g}^{(n)}(k) &= \mathbf{g}^{(n)}(k, \mathbf{u}_j^{(0)}(k), \ldots, \mathbf{u}_j^{(n-1)}(k)) \\ &= n(n-1)\,\mathbf{M}_C^{\alpha}\mathbf{u}_j^{(n-2)}(k) + 2n\,k\,\mathbf{M}_C^{\alpha}\mathbf{u}_j^{(n-1)}(k) + n\,\mathbf{C}_C^{\alpha,1}\mathbf{u}_j^{(n-1)}(k) \\ &\quad - \sum_{p=1}^{n-1}\sum_{q=0}^{n-p}\frac{n!}{p!q!(n-p-q)!}\omega_j^{(n-p-q)}(k)\omega_j^{(q)}(k)\mathbf{M}_C^{\beta}\mathbf{u}_j^{(p)}(k) \\ &\quad - \sum_{q=1}^{n-1}\binom{n}{q}\omega_j^{(n-q)}(k)\omega_j^{(q)}(k)\mathbf{M}_C^{\beta}(k),\end{aligned} \tag{5.23}$$

where $\mathbf{u}_j^{(0)}(k) = \mathbf{u}_j(k)$. Finally, the auxiliary vector $\mathbf{u}_j^{(n)}(k)$ associated to the n-th derivative of the eigenvector $\mathbf{u}_j(k)$ corresponding to $\omega_j(k)$ is obtained by solving

$$\begin{pmatrix}\mathbf{A}_C^{\alpha} + k\mathbf{C}_C^{\alpha,1} + k^2\mathbf{M}_C^{\alpha} - \omega^2\mathbf{M}_C^{\beta} & (\mathbf{A}_C^1 + \mathbf{M}_C^1)\mathbf{u}_j(k) \\ \mathbf{u}_j^{\mathrm{H}}(k)(\mathbf{A}_C^1 + \mathbf{M}_C^1) & \end{pmatrix}\begin{pmatrix}\mathbf{u}_j^{(n)}(k) \\ \nu\end{pmatrix} = \begin{pmatrix}\omega_j^{(n)}(k)\mathbf{h}(k) - \mathbf{g}^{(n)}(k) \\ 0\end{pmatrix},$$

which is the discrete version of (4.10).

5.3 Taylor expansion of eigenpaths

In this section we explain and demonstrate how to employ the derivatives $\partial_\mu^n\lambda_j(\mu)$, $n \in \mathbb{N}$, of the eigenvalue $\lambda_j(\mu)$ in a Taylor expansion of the eigenpath $\mu \mapsto \lambda_j(\mu)$.

5.3.1 Taylor theorem

Since the eigenpath $\mu \mapsto \lambda_j(\mu)$ is analytic we can apply the Taylor theorem, and hence, for any $\mu_0 \in I_\mu$ and $n \in \mathbb{N}$

$$\lambda_j(\mu) = \sum_{i=0}^{n}\frac{(\mu-\mu_0)^i}{i!}\,\partial_\mu^i\lambda_j(\mu_0) + R_n(\mu), \qquad \mu \in I_\mu, \tag{5.24}$$

with the remainder

$$R_n(\mu) = \frac{1}{n!}\int_{\mu_0}^{\mu}(\mu-\widetilde{\mu})^n\partial_\mu^{n+1}\lambda_j(\widetilde{\mu})\,\mathrm{d}\widetilde{\mu}, \tag{5.25}$$

see for example [Rud64].

The expansion (5.24) can be used to approximate the eigenpath

$$\lambda_j(\mu) \approx \sum_{m=0}^{n}\frac{(\mu-\mu_0)^m}{m!}\,\partial_\mu^m\lambda_j(\mu_0)$$

where the nonlinear eigenvalue problem (5.1) only has to be solved at $\mu = \mu_0$ for λ_j and $\mathbf{u}_j$, and the derivatives $\partial_\mu^n\lambda_j(\mu_0)$ have to be computed according to the procedure described in the previous section. Taylor expansions of analytic functions are known to converge in a vicinity of μ_0 but not necessarily in the whole interval I_μ.

Before we will present numerical results of the Taylor expansion of eigenpaths, we want to derive an estimate of the remainder R_n. Without loss of generality let $\mu > \mu_0$. According to the mean value theorem for integration, there exists $\widehat{\mu} \in [\mu_0, \mu]$ such that the remainder satisfies

$$R_n(\mu) = \frac{1}{n!}\,\partial_\mu^{n+1}\lambda_j(\widehat{\mu})\int_{\mu_0}^{\mu}(\mu-\widetilde{\mu})^n\,\mathrm{d}\widetilde{\mu} = \frac{(\mu-\mu_0)^{n+1}}{(n+1)!}\,\partial_\mu^{n+1}\lambda_j(\widehat{\mu}),$$

which is known as the Lagrange form of the remainder. Clearly,

$$R_n^{\mathrm{UB}}(\mu) = \frac{(\mu-\mu_0)^{n+1}}{(n+1)!}\max_{\widetilde{\mu}\in[\mu_0,\mu]}\partial_\mu^{n+1}\lambda_j(\widetilde{\mu})$$

is an upper bound for $R_n(\mu)$, while

$$R_n^{\mathrm{LB}}(\mu) = \frac{(\mu-\mu_0)^{n+1}}{(n+1)!} \min_{\widetilde{\mu}\in[\mu_0,\mu]} \partial_\mu^{n+1}\lambda_j(\widetilde{\mu})$$

is a lower bound for $R_n(\mu)$. Hence, assuming small variations of $\partial_\mu^{n+1}\lambda_j$ in $[\mu_0,\mu]$, a simple, non-rigorous estimate for the remainder $R_n(\mu)$ is given by

$$R_n^{\mathrm{est}}(\mu) = \frac{(\mu-\mu_0)^{n+1}}{(n+1)!} \partial_\mu^{n+1}\lambda_j(\mu_0). \tag{5.26}$$

5.3.2 Numerical results — Taylor expansion of dispersion curves

For illustration we will now show numerical results for Example 1, i. e. we consider the eigenvalue problem (2.13) of finding TM modes in a 2d PhC with square lattice. Assuming a fixed second component $k_2 = 0$ of the quasi-momentum $\mathbf{k} \in B_{2\mathrm{d}}$, the eigenvalue problem (2.13) of finding modes in 2d PhCs is linear in $\omega^2(k_1)$, where the first component $k_1 \in B = [-\frac{\pi}{a_1}, \frac{\pi}{a_1}]$ of the quasi-momentum $\mathbf{k}$ plays the role of a real-valued parameter. For simplicity of notation let us omit the index "1" in the first component k_1 of the quasi-momentum $\mathbf{k}$. Then we recall from Eq. (2.29) that the discrete form of this eigenvalue problem reads: given $k \in B = [-\frac{\pi}{a_1}, \frac{\pi}{a_1}]$ find eigenvalues $\omega^2(k) \in \mathbb{R}^+$ and associated eigenmodes $\mathbf{u}(k) \in \mathbb{C}^{N(C)}\backslash\{\mathbf{0}\}$ such that

$$\mathbf{N}_C^{\mathrm{TM}}(\omega, k)\mathbf{u} = \mathbf{0}, \tag{5.27}$$

with

$$\mathbf{N}_C^{\mathrm{TM}} : (\omega, k) \longmapsto \mathbf{A}_C^{\alpha=1} + k\mathbf{C}_C^{\alpha=1,1} + k^2\mathbf{M}_C^{\alpha=1} - \omega^2\mathbf{M}_C^{\beta=\varepsilon}. \tag{5.28}$$

Note that choosing the second component $k_2 \neq 0$ of the 2d quasi-momentum $\mathbf{k} \in B_{2\mathrm{d}}$ in the PhC eigenvalue problem (2.13), will only yield the additional matrix $k_2\mathbf{C}_C^{\alpha,2}$ in (5.28).

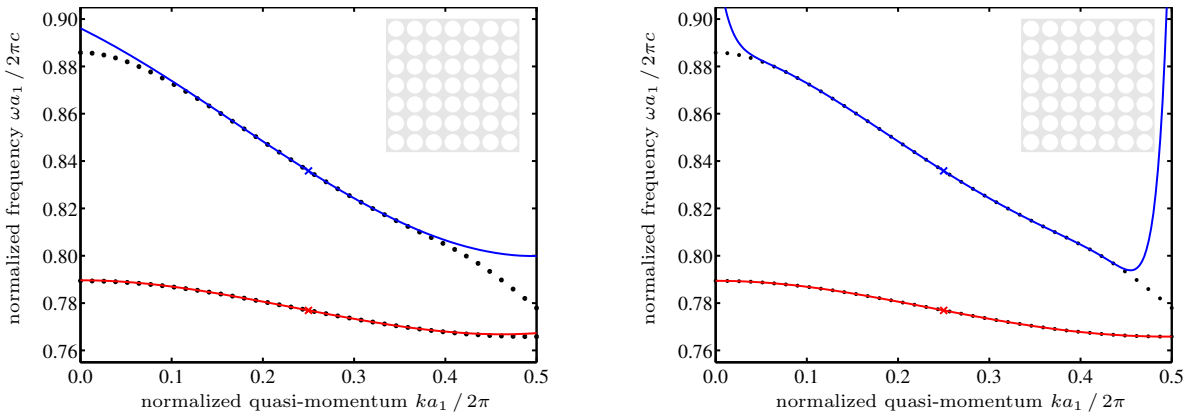
Recall that in the context of PhC band structure calculations the eigenpaths are called dispersion curves. Formulas for its first derivative, i. e. the group velocity, and its higher derivatives in variational formulation were already presented in Chapter 4. In particular, we want to point out that the second-order polynomial dependence of the matrix-valued function $\mathbf{N}_C^{\mathrm{TM}}(\omega, k)$ on the parameter k and the eigenvalue ω leads to a closed formula (5.22) for the n-th derivative $\omega^{(n)}(k)$ of the dispersion curve $\omega(k)$ that does not involve the complicated multivariant version of Faà di Bruno's formula as the general formula (5.17) for the n-th eigenpath derivative does.

We study the TM mode in the Γ-X-interval $\widehat{B} = [0, \frac{\pi}{a_1}]$ of the irreducible Brillouin zone $\widehat{B}_{2\mathrm{d}}$, and compare the dispersion relation $\omega(k)$ at 40 values of k with the results of the Taylor expansion around the centre $k_0 = \frac{\pi}{2a_1}$ of the Γ-X-interval $\widehat{B}$.

In Figure 5.1 we present a comparison of the Taylor expansion of orders $n = 3$ and $n = 20$ with the "exact" sixth and seventh dispersion curve. We can see from Figure 5.1a that already a Taylor expansion of order $n = 3$ provides a good approximation of the sixth dispersion curve (red line). For the presented level of detail, we can only see a difference of the Taylor expansion and the exact curve near $k = \frac{\pi}{a_1}$. The seventh dispersion curve (blue line) is also well approximated in a vicinity of the centre $k_0 = \frac{\pi}{2a_1}$ of the expansion but the error increases towards the borders of $\widehat{B}$, i. e. where $|k - k_0|$ becomes large.

However, increasing the order n of the Taylor expansion does not lead to lower error levels near the end points as can be seen in Figure 5.1b where the Taylor expansion of order $n = 20$ is shown. While the approximation error of the sixth dispersion curve decreases, the approximation error of the seventh dispersion curve becomes even larger near $k = 0$ and $k = \frac{\pi}{a_1}$. This can be explained by analysing the behaviour of the remainder R_n. But before we do so, we present the convergence of the Taylor expansion. In Figure 5.2 the maximum errors over a set of 40 equidistant values of $k \in \widehat{B}$ of the Taylor expansion of the sixth and seventh dispersion curves are plotted with respect to the order n of the Taylor expansion. While we observe exponential convergence of the error of the sixth dispersion curve, the error of the seventh dispersion curve diverges when increasing the order n.

Now let us study the remainder $R_n(k)$ in order to explain the behaviour of the Taylor expansion of the seventh dispersion curve in Figure 5.1 when increasing the order n of the Taylor expansion. In Figure 5.3



(a) Taylor expansion of order $n = 3$.

(b) Taylor expansion of order $n = 20$.

Figure 5.1: Sixth (red) and seventh (blue) dispersion curves of Example 1. Taylor expansion (solid lines) of order $n = 3$ (a) and $n = 20$ (b) around $k_0 = \frac{\pi}{2a_1}$ (crosses) compared to "exact" dispersion curves (dotted lines) evaluated at 40 equidistant values of k.

the estimate $R_n^{\text{est}}(k)$ is presented for the sixth and seventh dispersion curve of the example introduced above. The estimate is evaluated at $k = \frac{\pi}{a_1}$, i. e. where the distance $|k - k_0|$ is maximal. We can see that the estimate $R_n^{\text{est}}(k)$ of the sixth dispersion curve decreases with the order n, which corresponds to the decrease of the actual maximum error presented in Figure 5.2. The estimate $R_n^{\text{est}}(k)$ of the seventh dispersion curve, however, increases with the order n, which explains the increasing error of the Taylor expansion, that was observed in Figure 5.1. In other words, the derivatives $\omega^{(n)}$ increase faster with n than the ratio $\frac{n!}{|k-k_0|^n}$. This means that we have to restrict the computation of the Taylor expansion to a vicinity of k_0 such that $|k - k_0|^n$ is sufficiently small and hence, the ratio $\frac{|k-k_0|^{n+1}}{(n+1)!}$ dominates the estimate $R_n^{\text{est}}(k)$. This fact motivates the definition of an interval in which the estimate of the remainder is bounded by some desired error tolerance. We will define such an interval in the following section where it will be used to develop an adaptive algorithm for the approximation of eigenpaths.

The quality of the non-rigorous estimate $R_n^{\text{est}}(k)$ of the remainder can be seen from a comparison of Figures 5.2 and 5.3. We can see that the maximum error in Figure 5.2 behaves very similar to the estimate of the remainder in Figure 5.3. In fact the effectivity of the estimate, i. e. the ratio of estimate and maximum error, varies between 0.21 ($n = 1$) and 1.31 ($n = 6$) for the sixth dispersion curve, and between 0.02 ($n = 13$) and 2.48 ($n = 17$) for the seventh dispersion curve, and is hence, reasonably close to one.

5.4 An adaptive algorithm for eigenpath following

With the help of the Taylor expansion of eigenpaths and the estimation of its remainder, we can define an adaptive approximation of eigenpaths by a path following algorithm. For the application under consideration in this thesis, i. e. PhC band structure calculations, Spence and Poulton proposed a path following algorithm in [SP05]. They also employ a Taylor expansion of the dispersion curves. However, their algorithm is not adaptive as they use a fixed step size.

5.4.1 Step size control

A key ingredient is the control of the step size. For this we shall use the non-rigorous estimate (5.26) of the remainder (5.25) of the Taylor expansion.

For example, if we want the error of the Taylor expansion of $\lambda_j(\mu)$ around some $\mu_0 \in I_\mu$ to be (roughly) smaller than some error tolerance $\varepsilon_{\text{tol}}^{\text{step}}$ we restrict our expansion to the domain $[\mu_0 - h_{j,n}(\mu_0), \mu_0 +$

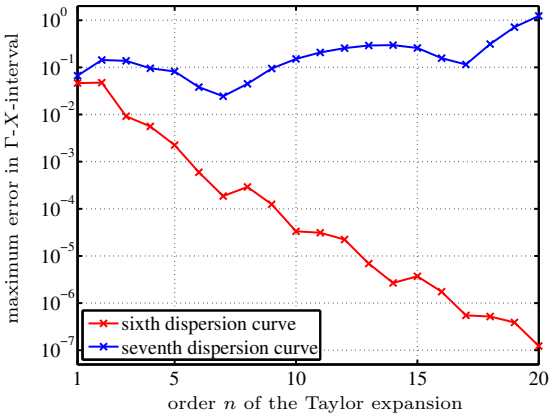


Figure 5.2: Maximum error of the Taylor expansion of the sixth (red) and seventh (blue) dispersion curve of the band structure presented in Figure 5.1 in dependence on the order n of the Taylor expansion. The maximum error is evaluated on an equidistant grid of 40 values of $k \in \widehat{B}$.

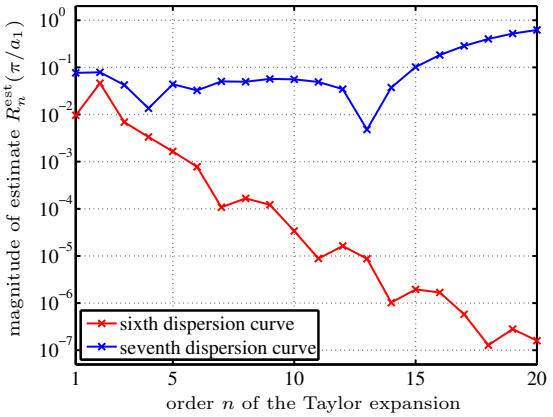


Figure 5.3: Magnitude of the non-rigorous estimate $R_n^{\text{est}}(k)$ of the remainder $R_n(k)$ at $k = \frac{\pi}{a_1}$ of the Taylor expansion of the sixth (red) and seventh (blue) dispersion curve around $k_0 = \frac{\pi}{2a_1}$ of the band structure presented in Figure 5.1 in dependence on the order n of the Taylor expansion.

$h_{j,n}(\mu_0)]$, where the step size $h_{j,n}(\mu_0)$ is obtained from

$$h_{j,n}(\mu_0) = \left(\varepsilon_{\text{tol}}^{\text{step}} \frac{(n+1)!}{|\partial_\mu^{n+1} \lambda_j(\mu_0)|} \right)^{\frac{1}{n+1}}. \tag{5.29}$$

With the help of the step size (5.29), we propose a simple algorithm for the adaptive computation of an approximation to an eigenpath $\lambda_j(\mu)$, $\mu \in I_\mu$, see Algorithm 5.2, i. e. we start by choosing an order n of the Taylor expansion, an error tolerance $\varepsilon_{\text{tol}}^{\text{step}}$ for the step size, and a start start value $\mu^{(0)} \in I_\mu$, e. g. the centre of the interval I_μ, i. e. $\mu^{(0)} = \frac{1}{2}$. Then we compute a set of eigenvalues $\lambda_j(\mu^{(0)})$, $j = 1, \ldots, J$, that, e. g. lie in a desired interval. In the case of PhC waveguide band structure calculations this would be, for example, a band gap. For each eigenvalue $\lambda_j(\mu^{(0)})$ we proceed as follows: We determine the acceptable step size $h_{j,n}(\mu^{(0)})$ according to (5.29). If $h_{j,n}(\mu^{(0)}) \geq \frac{1}{2}$ we approximate $\lambda_j(\mu)$ in I_μ by its Taylor expansion around $\mu^{(0)} = \frac{1}{2}$. Otherwise, we set $\mu^{(-1)} = \mu^{(0)} - h_{j,n}(\mu^{(0)})$ and $\mu^{(+1)} = \mu^{(0)} + h_{j,n}(\mu^{(0)})$, and compute the eigenvalues $\lambda_j(\mu^{(-1)})$ and $\lambda_j(\mu^{(+1)})$ which are closest to their estimation that is obtained by a Taylor expansion of order n around $\mu^{(0)}$ at $\mu^{(-1)}$ and $\mu^{(+1)}$, respectively. Then we compute the acceptable step sizes $h_{j,n}(\mu^{(\pm 1)})$. We continue with this procedure until $\mu^{(-p)} - h_{j,n}(\mu^{(-p)}) \leq 0$ and $\mu^{(+q)} + h_{j,n}(\mu^{(+q)}) \geq 1$ for some $p, q \in \mathbb{N}$.

Note that using an iterative scheme to compute the eigenvalues, it cannot be guaranteed that the eigenvalue, the iterative scheme converges to, is the eigenvalue closest to the start value of the iterative scheme. Thus, lines 12 and 21 of Algorithm 5.2 have to be changed when using an iterative eigenvalue solver. In this case we compute the eigenvalue by choosing the expected location of the eigenvalue as start value of the iterative scheme, having in mind that the result of the iterative solver might not be the closest eigenvalue.

We take the values λ_j and their derivatives $\partial_\mu^i \lambda_j$, $i = 1, \ldots, n$, at $\mu^{(\ell)}$, $\ell = -p, \ldots, q$ and compute an approximation to the dispersion curve using, e. g. an Hermite interpolation [QSS07] or a weighted Taylor expansion where we approximate

$$\lambda_j(\mu) \approx \frac{\mu^{(\ell+1)} - \mu}{\mu^{(\ell+1)} - \mu^{(\ell)}} \sum_{i=0}^{n} \frac{(\mu - \mu^{(\ell)})^i}{i!} \partial_\mu^i \lambda_j(\mu^{(\ell)}) + \frac{\mu - \mu^{(\ell)}}{\mu^{(\ell+1)} - \mu^{(\ell)}} \sum_{i=0}^{n} \frac{(\mu - \mu^{(\ell+1)})^i}{i!} \partial_\mu^i \lambda_j(\mu^{(\ell+1)}), \tag{5.30}$$

if $\mu \in [\mu^{(\ell)}, \mu^{(\ell+1)}]$, $\ell = -p, \ldots, q-1$, and in the intervals $(0, \mu^{(-p)}]$ and $[\mu^{(+q)}, 1)$ we take the Taylor expansion directly. The former approach has the advantage that it delivers a smooth curve but it yields

Algorithm 5.2. Adaptive path following algorithm.

Fix the order $n \in \mathbb{N}$ of the expansion, the error tolerance $\varepsilon_{\text{tol}}^{\text{step}} \ll 1$ and the start value $\mu^{(0)} \in I_\mu$.
Compute a set of eigenvalues $\lambda_j(\mu^{(0)})$, $j = 1, \dots, J$.
for $j = 1, \dots, J$ **do**
    Compute the derivatives $\partial_\mu^i \lambda_j(\mu^{(0)})$, $i = 1, \dots, n+1$.
    Compute the acceptable step size $h_{j,n}(\mu^{(0)})$.
    Set $\mu_j^{(0)} = \mu^{(0)}$.
    Set $p = 0$.
    while $\mu_j^{(-p)} - h_{j,n}(\mu_j^{(-p)}) > \min I_\mu$ **do**
        Set $p = p + 1$.
        Set $\mu_j^{(-p)} = \mu_j^{(-p+1)} - h_{j,n}(\mu_j^{(-p+1)})$.
        Compute the approximative eigenvalue $\widetilde{\lambda}_j(\mu_j^{(-p)})$ at $\mu_j^{(-p)}$ using Taylor expansion of order n around $\mu_j^{(-p+1)}$.
        Compute the eigenvalue $\lambda_j(\mu_j^{(-p)})$ that is closest to $\widetilde{\lambda}_j(\mu_j^{(-p)})$.
        Compute the derivatives $\partial_\mu^i \lambda_j(\mu_j^{(-p)})$, $i = 1, \dots, n+1$.
        Compute the acceptable step size $h_{j,n}(\mu_j^{(-p)})$.
    end while
    Set $q = 0$.
    while $\mu_j^{(q)} + h_{j,n}(\mu_j^{(q)}) < \max I_\mu$ **do**
        Set $q = q + 1$.
        Set $\mu_j^{(q)} = \mu_j^{(q-1)} + h_{j,n}(\mu_j^{(q-1)})$.
        Compute the approximative eigenvalue $\widetilde{\lambda}_j(\mu_j^{(q)})$ at $\mu_j^{(q)}$ using Taylor expansion of order n around $\mu_j^{(q-1)}$.
        Compute the eigenvalue $\lambda_j(\mu_j^{(q)})$ that is closest to $\widetilde{\lambda}_j(\mu_j^{(q)})$.
        Compute the derivatives $\partial_\mu^i \lambda_j(\mu_j^{(q)})$, $i = 1, \dots, n+1$.
        Compute the acceptable step size $h_{j,n}(\mu_j^{(q)})$.
    end while
    Approximate the j-th eigenpath by an Hermite interpolation or a piecewise, weighted Taylor expansion (5.30) of order n using the the eigenvalues λ_j and their derivatives $\partial_\mu^i \lambda_j$, $i = 1, \dots, n$, at $\mu_j^{(\ell)}$, $\ell = -p, \dots, q$.
end for

additional costs for the interpolation. The latter approach, on the other hand, comes with negligible additional costs and its implementation is straightforward.

The computational effort of this algorithm is as follows: In addition to the eigenvalue problem (5.1) at the start value $\mu^{(0)}$, we have to solve for each dispersion curve a total of $p+q$ eigenvalue problems (5.1), $n(p+q+1)$ linear systems (5.18) and $(n+1)(p+q+1)$ algebraic equations (5.17). For each of the $p+q+1$ values of μ we have to compute the acceptable step size using Eq. (5.29), which is a simple scalar equation.

In the following two sections let us introduce additional refinement checks, that will help to improve our approximation.

5.4.2 Backward check

An improvement of the adaptive scheme can be realized by a backward check, i. e. we check if the Taylor expansion around $\mu^{(\ell\pm1)}$ recovers the original value $\lambda_j(\mu^{(\ell)})$ plus/minus some tolerance $\varepsilon_{\text{tol}}^{\text{bwd}}$. If not, it is possible that we mistakenly switched to another eigenpath or the acceptable step size at $\mu^{(\ell\pm1)}$ is much smaller than at $\mu^{(\ell)}$. Then we refine the step size $h_{j,n}(\mu^{(\ell)})$, i. e. we multiply it by a factor σ^{bwd} smaller than one, e. g. $\sigma^{\text{bwd}} = \frac{1}{2}$, and take $\mu^{(\ell\pm1)} = \mu^{(\ell)} \pm \sigma^{\text{bwd}} h_{j,n}(\mu^{(\ell)})$ as subsequent parameter value for our sampling. When carrying out the backward check we also have to solve the eigenvalue problem (5.1)

and compute the derivatives at the boundaries of the interval I_μ, i. e. at $\mu = 0$ and $\mu = 1$, such that the the Taylor expansions around $\mu^{(-p)}$ and $\mu^{(+q)}$ can be validated. The adaptive path following algorithm including backward check is sketched in Algorithm 5.3.

5.4.3 Crossing check

A special emphasis in algorithms for eigenpath following has to be put into the question whether two eigenpaths cross or if they only come very close but avoid a crossing.

In the context of PhC waveguide band structures, such an avoided crossing, or anti-crossing, is called *mini-stopband* [ORB+01]. It is well known, that for symmetric waveguides, e. g. W1 waveguides with square or hexagonal lattice, modes of opposite parity cross while modes of identical parity form a mini-stopband. On the other hand, for waveguides, whose holes/rods on top of the line defect are shifted exactly by $\frac{a_1}{2}$ compared to the holes/rods below the guide, e. g. W2 waveguide with hexagonal lattice, just the opposite holds true: modes of identical parity cross while modes of opposite parity form a mini-stopband. If the waveguide does not satisfy either of these conditions, e. g. if the shift is less than $\frac{a_1}{2}$, all modes avoid to cross and form very narrow mini-stopbands [OBS+02]. Even though this classification allows for an identification of crossings and avoided crossings, we will apply later in Section 5.5 the proposed crossing check also to PhC waveguide band structure calculations, which allows us to identify mini-stopbands without comparing the parities of the modes.

The crossing check works then as described in Algorithm 5.4. After two eigenpaths were approximated with the adaptive scheme described above and it turned out that the approximated eigenpaths cross at some point μ_0, say, we solve the eigenvalue problem (5.1) at μ_0 where we will obtain two close

Algorithm 5.3. Adaptive path following algorithm including backward check.

1: Fix the order $n \in \mathbb{N}$ of the expansion, the error tolerances $\varepsilon_{\text{tol}}^{\text{step}}, \varepsilon_{\text{tol}}^{\text{bwd}} \ll 1$, the refinement factor $\sigma^{\text{bwd}} < 1$ and the start value $\mu^{(0)} \in I_\mu$.
2: Compute a set of eigenvalues $\lambda_j(\mu^{(0)})$, $j = 1, \ldots, J$.
3: **for** $j = 1, \ldots, J$ **do**
4: Compute the derivatives $\partial_\mu^i \lambda_j(\mu^{(0)})$, $i = 1, \ldots, n+1$.
5: Compute the acceptable step size $h_{j,n}(\mu^{(0)})$.
6: Set $\mu_j^{(0)} = \mu^{(0)}$.
7: Set $p = 0$.
8: **while** true **do**
9: Set $p = p + 1$.
10: Set $\mu_j^{(-p)} = \max\left\{\mu_j^{(-p+1)} - h_{j,n}(\mu_j^{(-p+1)}), \min I_\mu\right\}$.
11: Compute the approximative eigenvalue $\widetilde{\lambda}_j(\mu_j^{(-p)})$ at $\mu_j^{(-p)}$ using Taylor expansion of order n around $\mu_j^{(-p+1)}$.
12: Compute the eigenvalue $\lambda_j(\mu_j^{(-p)})$ that is closest to $\widetilde{\lambda}_j(\mu_j^{(-p)})$.
13: Compute the derivatives $\partial_\mu^i \lambda_j(\mu_j^{(-p)})$, $i = 1, \ldots, n+1$.
14: Compute the approximative eigenvalue $\widetilde{\lambda}_j(\mu_j^{(-p+1)})$ at $\mu_j^{(-p+1)}$ using Taylor expansion of order n around $\mu_j^{(-p)}$.
15: **if** $|\widetilde{\lambda}_j(\mu_j^{(-p+1)}) - \lambda_j(\mu_j^{(-p+1)})| > \varepsilon_{\text{tol}}^{\text{bwd}}$ **then**
16: Set $p = p - 1$.
17: Set $h_{j,n}(\mu_j^{(-p)}) = \sigma^{\text{bwd}} h_{j,n}(\mu_j^{(-p)})$.
18: **else if** $\mu_j^{(-p)} = \min I_\mu$ **then**
19: **break**
20: **else**
21: Compute the acceptable step size $h_{j,n}(\mu_j^{(-p)})$.
22: **end if**
23: **end while**

Set $q = 0$.
while true **do**
  Set $q = q + 1$.
  Set $\mu_j^{(q)} = \min\left\{\mu_j^{(q-1)} + h_{j,n}(\mu_j^{(q-1)}), \max I_\mu\right\}$.
  Compute the approximative eigenvalue $\widetilde{\lambda}_j(\mu_j^{(q)})$ at $\mu_j^{(q)}$ using Taylor expansion of order n around $\mu_j^{(q-1)}$.
  Compute the eigenvalue $\lambda_j(\mu_j^{(q)})$ that is closest to $\widetilde{\lambda}_j(\mu_j^{(q)})$.
  Compute the derivatives $\partial_\mu^i \lambda_j(\mu_j^{(q)})$, $i = 1, \ldots, n+1$.
  Compute the approximative eigenvalue $\widetilde{\lambda}_j(\mu_j^{(q-1)})$ at $\mu_j^{(q-1)}$ using Taylor expansion of order n around $\mu_j^{(q)}$.
  if $|\widetilde{\lambda}_j(\mu_j^{(q-1)}) - \lambda_j(\mu_j^{(q-1)})| > \varepsilon_{\text{tol}}^{\text{bwd}}$ **then**
    Set $q = q - 1$.
    Set $h_{j,n}(\mu_j^{(q)}) = \sigma^{\text{bwd}} h_{j,n}(\mu_j^{(q)})$.
  else if $\mu_j^{(q)} = \max I_\mu$ **then**
    break
  else
    Compute the acceptable step size $h_{j,n}(\mu_j^{(q)})$.
  end if
end while
Approximate the j-th eigenpath by an Hermite interpolation or a piecewise, weighted Taylor expansion (5.30) of order n using the the eigenvalues λ_j and their derivatives $\partial_\mu^i \lambda_j$, $i = 1, \ldots, n$, at $\mu_j^{(\ell)}$, $\ell = -p, \ldots, q$.
end for

eigenvalues near the expected crossing. Note that, also if the expected crossing turns out to be an actual crossing, these two eigenvalues are most likely not identical but only very close. Then we compute the first derivatives of these two eigenmodes at μ_0 using the formula (5.4) and compare them with the first derivatives of the approximated eigenpaths (5.30). If the derivatives of the two eigenmodes do not coincide, i. e. the two curves do not cross with the same slope, and each derivative matches well with the derivative of one of the approximated eigenpaths in the sense that the magnitude of the difference does not exceed an error tolerance of $\varepsilon_{\text{tol}}^{\text{xng}}$, we take this as evidence that the two eigenpaths cross. On the other hand, if the two derivatives are very close, i. e. the two eigenpaths have approximately the same slope at μ_0, we also have to compute higher derivatives of the eigenpaths at μ_0 using the formula (5.17) and compare them with the corresponding derivatives of the approximated eigenpaths (5.30). In fact we have to compute and compare at least derivatives of order n, if the derivatives of the two eigenpaths coincide up to order $n-1$. If for all m, with $1 \leq m \leq n$, the derivatives of order m of the two eigenpaths coincide with one of the derivatives of order m of the two approximated eigenpaths, i. e. the magnitude of the difference does not exceed an error tolerance of $\varepsilon_{\text{tol}}^{\text{xng}}$, we shall assume that the two eigenpaths cross. Otherwise, we refine our approximations by additionally applying the adaptive scheme to the eigenpaths around μ_0, taking μ_0 as start value and stopping the scheme if a value of μ is reached for which we already solved the eigenvalue problem (5.1).

If we compare derivatives up to order n we shall denote this test as n-th order crossing check. If $n = 1$, we simply call it crossing check.

Note that the crossing check can also be understood as a validation test for crossings, in particular in the case when the two eigenpaths cross and have the same slope at the crossing. To this end, we also perform the crossing check if the two approximated eigenpaths come very close but do not cross. For those points it is necessary to perform at least a second order crossing check.

Algorithm 5.4. Crossing check of order n.

Requirements: Let two approximated eigenpaths $\widetilde{\lambda}_1 \approx \lambda_1$ and $\widetilde{\lambda}_2 \approx \lambda_j$ with nodes $\mu_j^{(-p_j)}, \dots, \mu_j^{(q_j)}$, $j = 1, 2$, be given. The nodes may either be obtained by Algorithm 5.2 without backward check, or by Algorithm 5.3 including backward check, and the approximative curves may either result from a Hermite interpolation or a weighted Taylor expansion. Let the two approximated eigenpaths intersect at μ_0 with $\mu_j^{(\ell_j)} < \mu_0 < \mu_j^{(\ell_j+1)}$, where $\ell_j \in \{-p_j, \dots, q_j - 1\}$, $j = 1, 2$.

1: Fix the crossing check tolerance $\varepsilon_{\text{tol}}^{\text{xng}} \ll 1$.
2: Compute the two eigenvalues $\lambda_j(\mu_0)$, $j = 1, 2$, at μ_0 that are closest to $\widetilde{\lambda}_1(\mu_0) = \widetilde{\lambda}_2(\mu_0)$.
3: Compute the derivatives $\partial_\mu^i \lambda_j(\mu_0)$, $j = 1, 2$, $i = 1, \dots, n$, of the two eigenpaths at μ_0 up to order n.
4: Set `refine` to false.
5: **for** $i = 1, \dots, n$ **do**
6: **if** $|\partial_\mu^i \lambda_1(\mu_0) - \partial_\mu^i \widetilde{\lambda}_1(\mu_0)| > \varepsilon_{\text{tol}}^{\text{xng}}$ or $|\partial_\mu^i \lambda_2(\mu_0) - \partial_\mu^i \widetilde{\lambda}_2(\mu_0)| > \varepsilon_{\text{tol}}^{\text{xng}}$ **then**
7: Set `refine` to true.
8: **break**
9: **end if**
10: **end for**
11: **if** `refine` **then**
12: Apply Algorithm 5.2 or Algorithm 5.3 with start value μ_0 for the adaptive eigenpath following in the intervals $[\mu_1^{(\ell_1)}, \mu_2^{(\ell_2)}]$ and $[\mu_2^{(\ell_2)}, \mu_1^{(\ell_1)}]$.
13: **end if**

5.5 Adaptive path following of dispersion curves

In this section we want to test the proposed adaptive scheme and show numerical results. We aim to adaptively follow the eigenpaths of PhC and PhC waveguide band structures, the so-called dispersion curves. We start with the TE mode band structure of a PhC W1 waveguide and of a perturbed W1 waveguide, before we will return to the TM mode band structure of Example 1 that we already used in our numerical experiments in Section 5.3.

5.5.1 Band structure of a PhC W1 waveguide

We consider the TE mode band structure of the PhC W1 waveguide with hexagonal lattice introduced in Example 2 and apply the supercell method with a supercell $S_5 \subset \mathbb{R}^2$ of five PhC unit cells on top and bottom of the defect cell to compute approximations to guided modes. In other words, we consider the eigenvalue problem (2.23) whose discrete form, as given in (2.29) reads: given a quasi-momentum $k \in \widehat{B} = [0, \frac{\pi}{a_1}]$ find eigenvalues $\omega^2(k) \in \mathbb{R}^+ \setminus \sigma^{\text{ess}}(k)$ and associated eigenmodes $\mathbf{u}(k) \in \mathbb{C}^{N(S_5)} \setminus \{\mathbf{0}\}$ such that

$$\mathbf{N}_{S_5}^{\text{TE}}(\omega, k)\mathbf{u} = 0, \tag{5.31}$$

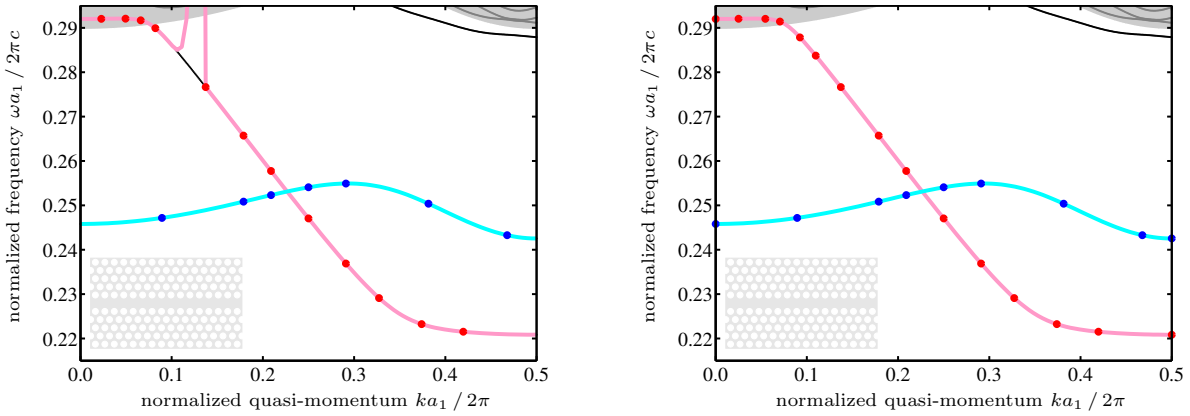
with

$$\mathbf{N}_{S_5}^{\text{TE}} : (\omega, k) \longmapsto \mathbf{A}_{S_5}^{\alpha=1/\varepsilon} + k\mathbf{C}_{S_5}^{\alpha=1/\varepsilon,1} + k^2\mathbf{M}_{S_5}^{\alpha=1/\varepsilon} - \omega^2\mathbf{M}_{S_5}^{\beta=1}.$$

Analogously to (5.27) this eigenvalue problem is linear in $\omega^2(k)$, and formulas for the derivatives of the dispersion curves $k \mapsto \omega(k)$ in variational formulation were already presented in Chapter 4.

Using the adaptive algorithm introduced in Section 5.4 without any additional refinement checks, like the backward check or the crossing check, we aim to compute an approximation to the two dispersion curves in the second band gap, that is approximately located in the frequency interval $[0.2 \cdot \frac{2\pi c}{a_1}, 0.3 \cdot \frac{2\pi c}{a_1}]$, see Figure 2.9 for an illustration of the band structure. We choose a desired error tolerance of $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$ to compute the acceptable step sizes, set the order of the expansion to $n = 10$, and select the start value $k^{(0)} = \frac{\pi}{2a_1}$. When solving (5.31) for its eigenvalues ω^2 at the start value $k^{(0)}$ we omit eigenvalues inside the essential spectrum $\sigma^{\text{ess}}(k^{(0)})$ in order to follow guided modes only. However, note that the procedure can be applied to any eigenpaths of (5.31), not only those corresponding to dispersion curves of guided modes, since all eigenpaths of (5.31) are analytic in B, cf. Theorem 4.1. In fact, we shall continue to

follow the dispersion curves even if they leave the band gap. This will be different later in Chapter 6 where we employ DtN transparent boundary conditions, that are only well-defined in band gaps.



(a) No backward check.

(b) Backward check with $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$.

Figure 5.4: Adaptive Taylor scheme of order $n = 10$ applied to dispersion curves of the hexagonal PhC W1 waveguide. The error tolerance of the step size computation is $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$ and the start value of the iteration is set to $k^{(0)} = \frac{\pi}{2a_1}$.

The results can be seen in Figure 5.4a where the dots indicate the location of the values of k for which the dispersion relation $\omega(k)$ and its derivatives $\omega'(k), \omega^{(2)}(k), \ldots, \omega^{(10)}(k)$ were computed. The lines connecting the dots result from the post-processing, where we chose the weighted Taylor expansion (5.30). Note that the red dispersion curve leaves the band gap at $k \approx 0.1 \cdot \frac{2\pi}{a_1}$ and enters the frequency domain for which propagating PhC modes exist. As elaborated above, we can continue following this curve towards $k = 0$ having in mind, that these supercell eigenmodes are spurious and that this part of the dispersion curve has no physical meaning. Most noticeable is the numerical artifact of the red line between $k \approx 0.08 \cdot \frac{2\pi}{a_1}$ and $k \approx 0.14 \cdot \frac{2\pi}{a_1}$. Recall, that we chose $k^{(0)} = \frac{\pi}{2a_1}$ and hence, we followed the dispersion curve in this part from right to left. For the computation of the acceptable step size the derivatives at $k \approx 0.14 \cdot \frac{2\pi}{a_1}$ are relevant. But obviously, the derivatives at $k \approx 0.08 \cdot \frac{2\pi}{a_1}$ are significantly larger in magnitude than at $k \approx 0.14 \cdot \frac{2\pi}{a_1}$ yielding a smaller step size in the following step (distance to next red dot) and hence, explain the numerical error in the post-processing. This numerical artifact can be eliminated when performing the backward check. The results are presented in Figure 5.4b where we chose an error tolerance of $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ for the backward check.

Clearly, the computational costs of the adaptive scheme including backward check are smaller than the costs of the standard procedure to solve the eigenvalue problem (5.27) for an equidistant sample of quasi-momenta k, if one aims to get the same accuracy as the adaptive scheme. In the post-processing of the adaptive Taylor expansions in Figure 5.4 we chose an equidistant sample of 100 values of the quasi-momentum k for which we computed the weighted Taylor expansion and from which we draw the solid red and blue curves. This shall deal as a reference for the desired accuracy. That means the standard procedure to calculate the band structure is to solve 100 eigenvalue problems (5.27). On the other hand, the adaptive scheme including backward check for the unperturbed waveguide accounts for 25 eigenvalue problems (5.27) and the computation of the frequency derivatives (5.22) of 24 modes, where we solve the eigenvalue problem (5.27) at $k = 0$, $k = \frac{\pi}{a_1}$ and $k = k^{(0)}$ for two eigenvalues in order to save time. Four eigenmodes of these 25 eigenvalue computations are rejected due to the backward check which explains that the number of nodes for which we compute the frequency derivatives (5.22) is smaller than the number of eigenvalue problems (5.27) to be solved. Considering that the computational costs of solving (5.22) for the frequency derivatives is significantly smaller than solving the eigenvalue problem (5.27), we can expect clearly smaller computational costs of the adaptive scheme compared to the standard procedure. Note that the computational advantage of the adaptive scheme especially becomes obvious in the case of a rather simple dispersion curve, e. g. the blue curve in Figure 5.4. In fact, only

a very small number of eigenvalue problems and corresponding source problems have to be computed in order to figure out the dispersion curve's slope correctly.

5.5.2 Band structure with mini-stopband of a perturbed PhC W1 waveguide

Now we want to test the proposed crossing check. Applied to the numerical example above, a W1 waveguide with hexagonal lattice which is symmetric with respect to the line defect, we find that a refinement is not necessary since the approximated slopes and computed group velocities at the projected crossing match well, which is in line with the theory [ORB+01] that says that modes with even parity (red dispersion curve) and modes with odd parity (blue dispersion curve) have to cross and do not form a mini-stopband. Therefore, we shall apply our crossing check to a perturbed configuration. We shift the upper PhC by as little as $10^{-4}a_1$ to the left. This breaks the symmetry and we can expect that the crossing of the two guided modes becomes an avoided crossing.

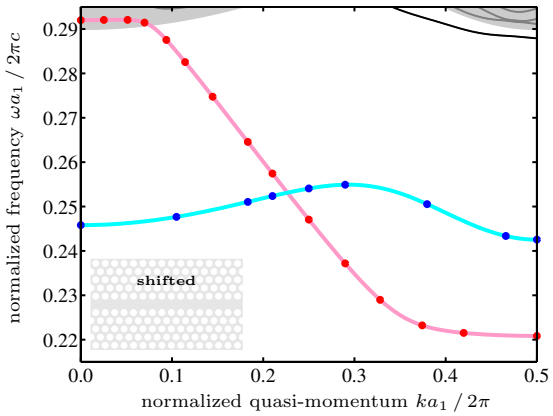
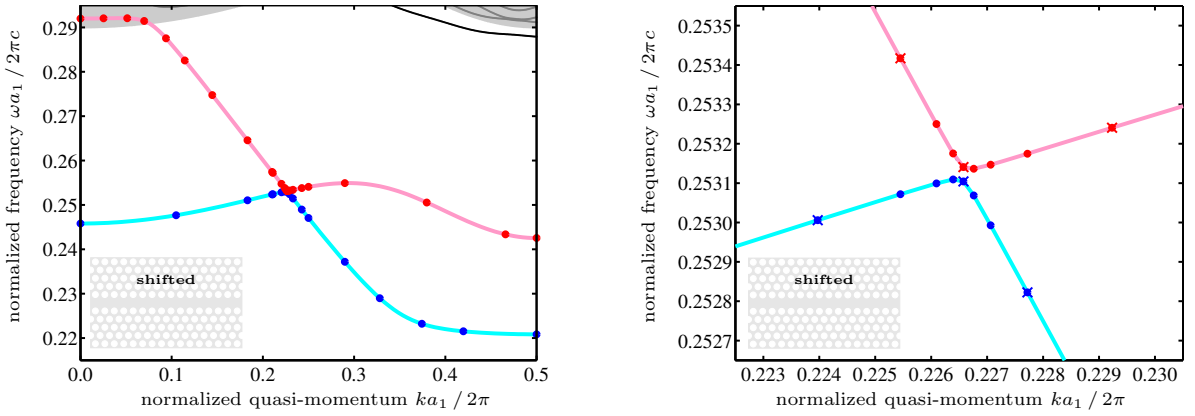


Figure 5.5: Adaptive Taylor scheme of order $n = 10$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ applied to dispersion curves of the perturbed PhC W1 waveguide of hexagonal lattice (upper PhC shifted by $10^{-4}a_1$ to the left). The error tolerance of the step size computation is $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$ and the start value of the iteration is set to $k^{(0)} = \frac{\pi}{2a_1}$.

In Figure 5.5 we show the results of the adaptive Taylor scheme including backward check with tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ applied to the perturbed W1 waveguide of hexagonal lattice. The results are very similar to the results of the unperturbed waveguide in Figure 5.4b. In particular, the approximated dispersion curves cross in Figure 5.5 which we know is not true. Thus, a crossing check as introduced above needs to be performed in order to identify the avoided crossing correctly. The two approximated dispersion curves in Figure 5.5 cross at $k_0 \approx 0.2266 \cdot \frac{2\pi}{a_1}$ and their slopes are approximately $-0.261c$ (red curve) and $0.044c$ (blue curve). But when solving the eigenvalue problem (5.27) at k_0 we find two eigenmodes which have both negative group velocity, and hence — using a tolerance $\varepsilon_{\text{tol}}^{\text{xng}} = 10^{-2}$ for the crossing check — a refinement at k_0 is necessary. The result can be seen in Figure 5.6 where we also show a detailed view of the mini-stopband.

The behaviour of the eigenmodes near the mini-stopband is illustrated in Figure 5.7, where we plotted the six eigenmodes marked with a cross in Figure 5.6b. We observe that the eigenmodes on the upper dispersion curve have even parity for quasi-momenta smaller than the quasi-momentum $k \approx 0.2266 \cdot \frac{2\pi}{a_1}$ of the mini-stopband, see Figure 5.7a, while the eigenmodes on the lower dispersion curve have odd parity, see Figure 5.7d. At the mini-stopband both modes are neither even nor odd, see Figures 5.7b and 5.7e, and when continuing to follow the two dispersion curves to the right, we find that the upper curve has changed its parity to odd, see Figure 5.7c, while the parity of the lower curve becomes even, see Figure 5.7f.

The computational advantage of the proposed adaptive Taylor expansion compared to the standard procedure even increases when trying to identify mini-stopbands. The adaptive scheme including backward



(a)

(b) Detailed view of the mini-stopband.

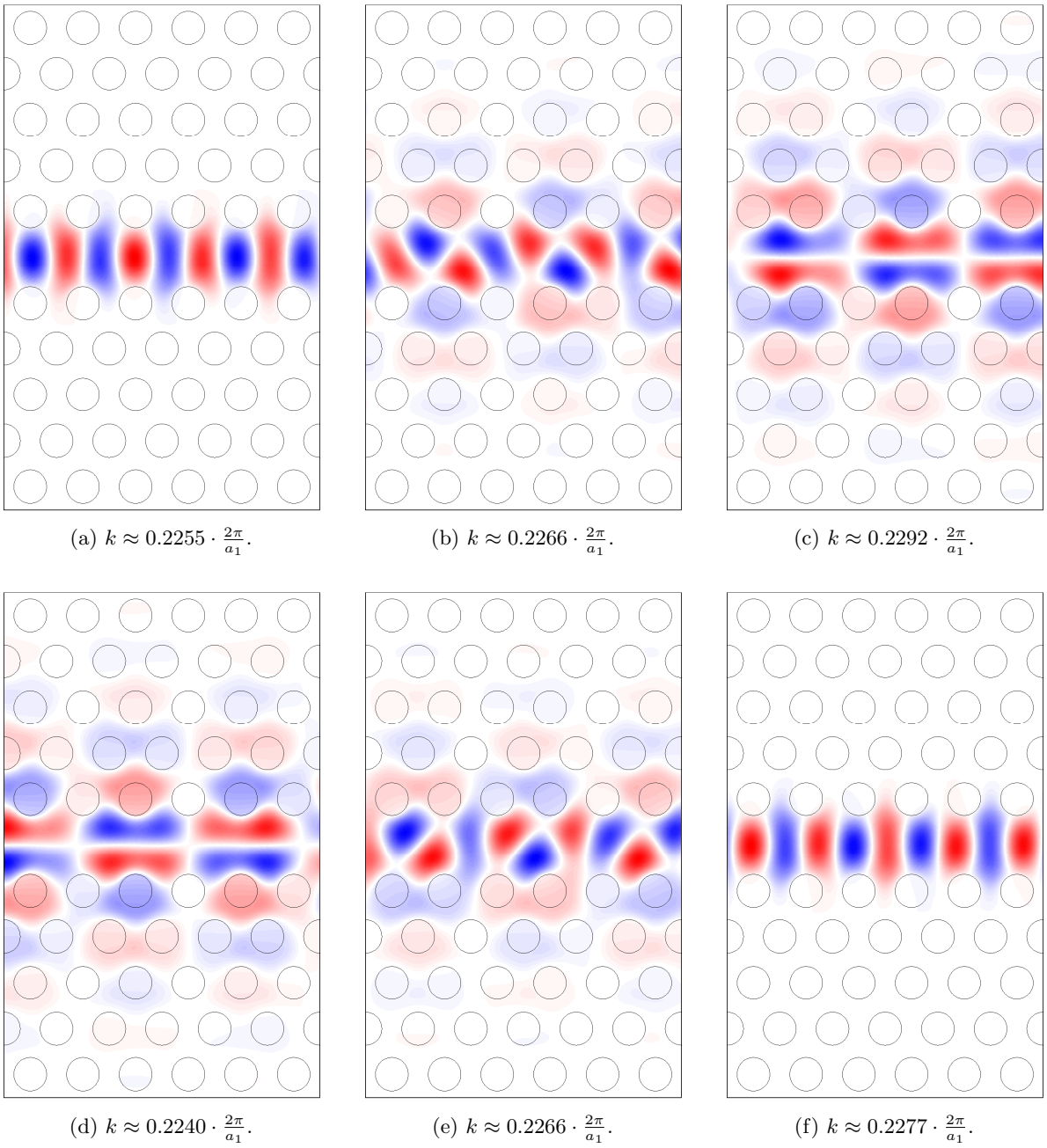
Figure 5.6: Adaptive Taylor scheme of order $n = 10$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ and first order crossing check of tolerance $\varepsilon_{\text{tol}}^{\text{xng}} = 10^{-2}$ applied to dispersion curves of the perturbed PhC W1 waveguide with hexagonal lattice (upper PhC shifted by $10^{-4}a_1$ to the left). The error tolerance of the step size computation is $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$ and the start value of the iteration is set to $k^{(0)} = \frac{\pi}{2a_1}$. For the nodes marked with a cross in (b) the eigenmodes are plotted in Figure 5.7.

check but without crossing check, as presented in Figure 5.5, accounts for 25 eigenvalue problems (5.27) and the computation of the frequency derivatives (5.22) for 24 modes. When additionally performing the crossing check and refining near the avoided crossing, as done in Figure 5.6, we have another 13 eigenvalue problems (5.27) and a total of 26 frequency derivatives (5.22) to solve. We save time by simultaneously refining both dispersion curves together with the same step size and solving (5.22) for two eigenvalues, while computing the frequency derivatives (5.22) of all 26 computed eigenmodes. This makes a total of 38 eigenvalue problems (5.27) and 50 frequency derivatives (5.22), that we have to solve in order to approximate the two dispersion curves of the perturbed waveguide as presented in Figure 5.6. Using the standard procedure to compute the band structure would clearly comprise the solution of more eigenvalue problems (5.27) since a very dense grid of values of the quasi-momentum k is needed in order to resolve the mini-stopband as accurate as in Figure 5.6b.

5.5.3 Dispersion curves intersecting with identical group velocity

Now let us return to Example 1 and the eigenvalue problem (5.27). We want to study the behaviour of our numerical scheme when two dispersion curves intersect at a point but do not cross. This is the case for the second and third dispersion curves at $k = 0$, as can be seen from the band structure in Figure 2.7. Due to symmetry at $k = 0$, we restricted our computations so far to the Γ-X-interval $\widehat{B} = [0, \frac{\pi}{a_1}]$ of the irreducible Brillouin zone $\widehat{B}_{2\text{d}}$. Now let us consider the interval $B = [-\frac{\pi}{a_1}, \frac{\pi}{a_1}]$ having in mind that the band structure is symmetric with respect to the frequency axis at $k = 0$. That means we know in advance that the two dispersion curves, that have a common eigenvalue at $k = 0$, do not cross but touch only. Let us now study if the proposed adaptive scheme can construct this band structure correctly. We choose the start point $k^{(0)} = 0.01 \cdot \frac{2\pi}{a_1}$. Recall that we cannot choose $k^{(0)} = 0$, the centre of the interval B, since the second and third dispersion curves intersect at this point and hence, the multiplicity of the eigenvalue at $k = 0$ is two, which implies that the group velocity formula (5.21), as well as the formula for higher derivatives, Eq. (5.22), is not well-defined without knowledge about the eigenmodes in the vicinity of $k = 0$, as we elaborated in Remark 4.2.

We start by setting the order of the Taylor expansion to $n = 1$. The step size tolerance is chosen to be $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$ and we employ the backward check with tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ but do not use the crossing check. Figure 5.8a shows the numerical result for this configuration. We can see that the second dispersion curve (blue) is computed incorrectly, since from $k = 0$ to the left it follows the third dispersion



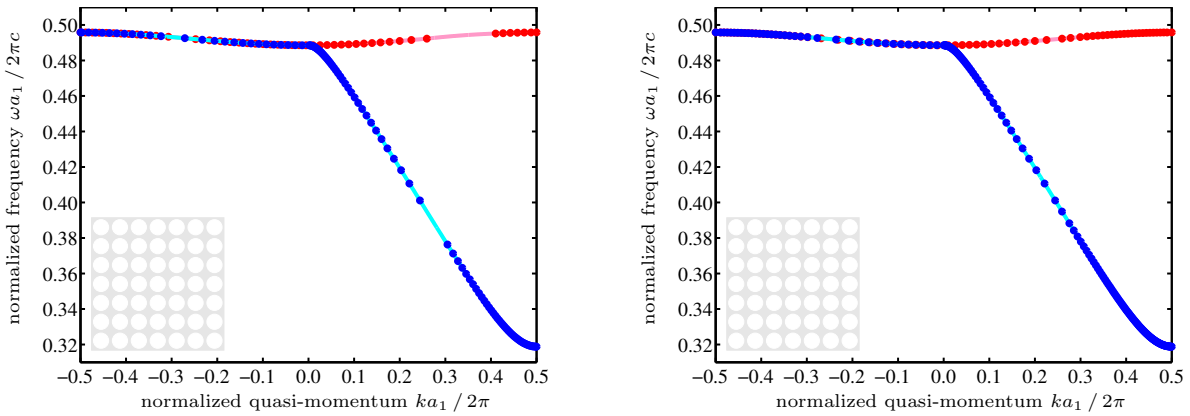
(a) $k \approx 0.2255 \cdot \frac{2\pi}{a_1}$. (b) $k \approx 0.2266 \cdot \frac{2\pi}{a_1}$. (c) $k \approx 0.2292 \cdot \frac{2\pi}{a_1}$.

(d) $k \approx 0.2240 \cdot \frac{2\pi}{a_1}$. (e) $k \approx 0.2266 \cdot \frac{2\pi}{a_1}$. (f) $k \approx 0.2277 \cdot \frac{2\pi}{a_1}$.

Figure 5.7: Real parts of the magnetic field components of the guided modes on the upper (a)–(c) and lower (d)–(e) dispersion curve of the perturbed PhC W1 waveguide with mini-stopband presented in Figure 5.6.

curve (red). When choosing a smaller backward check tolerance, as done in Figure 5.8b, where we set $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-4}$, we do not resolve this problem. In fact, both tolerance parameters, $\varepsilon_{\text{tol}}^{\text{bwd}}$ as well as the step size tolerance $\varepsilon_{\text{tol}}^{\text{step}}$, cannot be chosen small enough since an expansion of first order cannot account for the curvature of the dispersion curve. This explains that an expansion of first order is in general not a good choice no matter how small the tolerance parameters are chosen.

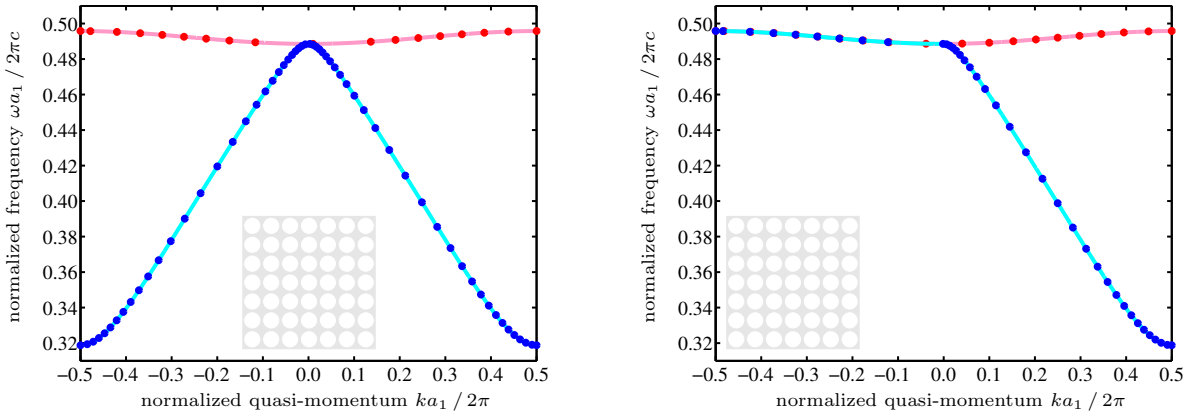
Now we increase the order of the expansion to $n = 2$ and choose again $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$. Figure 5.9a shows that the two dispersion curves are computed correctly. However, using an expansion of order $n = 2$ does not always resolve the problem as Figure 5.9b shows, where we set the start value of the scheme to $k^{(0)} = 0.25 \cdot \frac{2\pi}{a_1}$. It turns out that we were only lucky by previously setting the start value to



(a) Backward check tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$.

(b) Backward check tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-4}$.

Figure 5.8: Adaptive Taylor scheme of order $n = 1$ with backward check applied to the second and third dispersion curves of Example 1. The start value of the scheme is set to $k^{(0)} = 0.01 \cdot \frac{2\pi}{a_1}$.



(a) Start value $k^{(0)} = 0.01 \cdot \frac{2\pi}{a_1}$.

(b) Start value $k^{(0)} = 0.25 \cdot \frac{2\pi}{a_1}$.

Figure 5.9: Adaptive Taylor scheme of order $n = 2$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ applied to the second and third dispersion curves of Example 1.

$k^{(0)} = 0.01 \cdot \frac{2\pi}{a_1}$, where the second derivative is large enough in magnitude to account for the correct slope of the second dispersion curve. In the case presented in Figure 5.9b, however, the tolerance parameters $\varepsilon_{\text{tol}}^{\text{step}}$ and $\varepsilon_{\text{tol}}^{\text{bwd}}$ are chosen too large so that the adaptive scheme does not place a Taylor node close enough to $k = 0$ and hence, the magnitude of the second derivative at the smallest positive node is too small to account for the correct curvature at $k = 0$.

Choosing a smaller backward check tolerance may help to resolve this problem. When selecting a higher order we can also resolve this problem, as shown in Figure 5.10, where we set the order to $n = 3$, keeping the start value at $k^{(0)} = 0.25 \cdot \frac{2\pi}{a_1}$ and leaving the backward check tolerance unchanged.

Alternatively, we can do a second order crossing check in order to resolve this problem even if we keep $n = 2$ and $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$. The two approximated dispersion curves in Figure 5.9b come very close near $k = 0$. In fact, we observe that the two approximated curves cross at $k_0 \approx 0.02 \cdot \frac{2\pi}{a_1}$. At k_0 we solve the eigenvalue problem and compare the first and second derivatives of the dispersion relation with the slopes and curvatures of the approximated curves. It turns out that the second derivatives do not match well with the curvatures of the approximated curves. We refine the approximation as described in the section on the crossing check and find that the left branch of the refined blue curve does not fit to the left branch shown in Figure 5.9b, so that a full computation of the adaptive approximation in the interval $[-\frac{\pi}{a_1}, k_0]$

is necessary. This yields the band structure presented in Figure 5.11, which shows that the adaptive scheme with second order crossing check produces appropriate approximations of two dispersion curves intersecting with the same slope, even if the order is as low as $n = 2$.

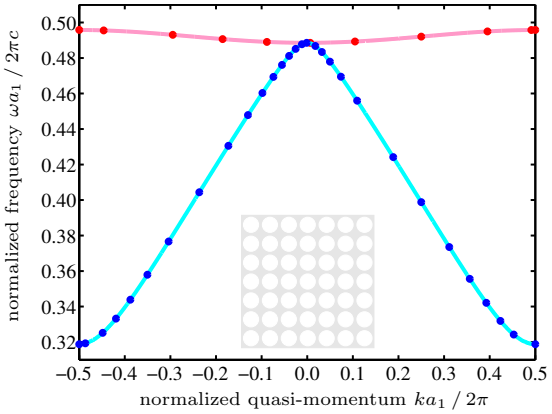


Figure 5.10: Adaptive Taylor scheme of order $n = 3$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ applied to the second and third dispersion curves of Example 1. The start value of the scheme is set to $k^{(0)} = 0.25 \cdot \frac{2\pi}{a_1}$.

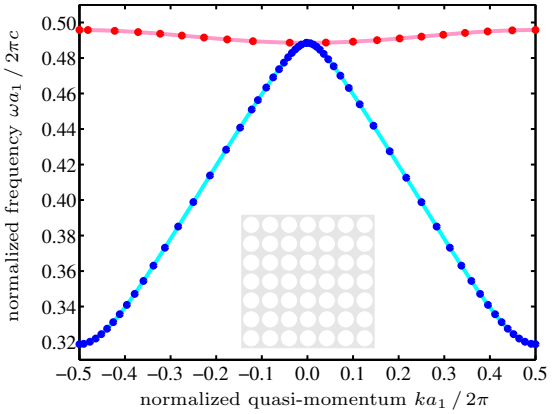


Figure 5.11: Adaptive Taylor scheme of order $n = 2$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ and second order crossing check of tolerance $\varepsilon_{\text{tol}}^{\text{xng}} = 10^{-2}$ applied to the second and third dispersion curves of Example 1. The start value of the scheme is set to $k^{(0)} = 0.25 \cdot \frac{2\pi}{a_1}$.

To summarize, we note that an expansion of order two or larger is needed to correctly identify the behaviour of two curves, that intersect with the same slope. An expansion of order larger than two is preferable in order to decrease the influence of the start value on the approximation. A second order crossing check can be employed to resolve the intersection correctly.

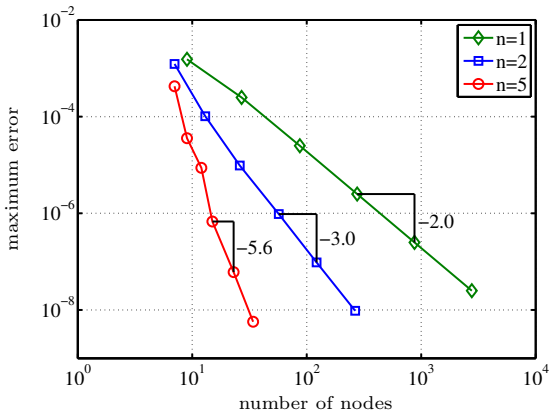
5.5.4 Convergence study

After showing in the previous examples that our proposed adaptive algorithm is applicable to PhC and PhC waveguide band structure calculations and that it can handle various difficulties like mini-stopbands and crossings of identical slope, we now turn to the question what can be said about the error of our algorithm and its convergence.

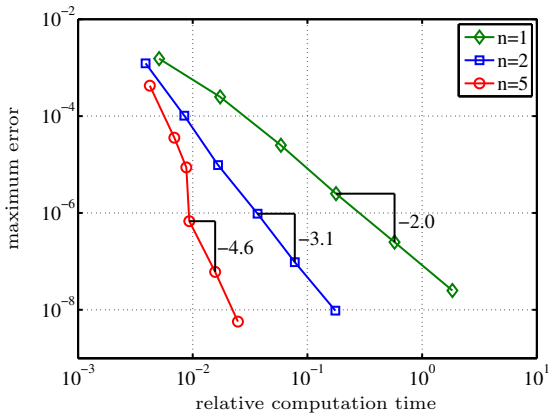
To this end, we study again Example 1. We want to compute the maximum error of our approximation to the fifth dispersion curve in the Γ-X-interval $\widehat{B} = [0, \frac{\pi}{a_1}]$, see the band structure in Figure 2.7. We evaluate the maximum error on an equidistant sample of 10 000 values of the quasi-momentum $k \in \widehat{B}$. We solve the eigenvalue problem (5.27) at each sample point to obtain a reference solution $\omega_{\text{ref}}(k)$ for our approximation.

The fifth dispersion curve of the TM mode band structure of Example 1 shown in Figure 2.7 does not intersect with or comes very close to any other dispersion curve. Therefore, we neither need to apply the crossing check, nor is there a need for additional orthogonality conditions in sense of Section 4.2.3 when computing the derivatives of the eigenmodes.

In Figure 5.12 we show the convergence of the maximum error for three different orders $n = 1, 2, 5$ of the Taylor expansion in our adaptive scheme. The step size and backward check tolerances are chosen to be $\varepsilon_{\text{tol}}^{\text{step}} = \varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-\ell}$, $\ell = 2, \ldots, 7$, i. e. each marker on the three curves for the three different orders corresponds to one of the six different values of the step size and backward check tolerances. Note that the actual maximum error is smaller than the chosen tolerances, which shows the overestimation of the actual error, that was already discussed in Section 5.4.1. In Figure 5.12a the maximum error is compared to the number of nodes in $\widehat{B}$ at which the eigenvalue problem (5.27) is solved and the dispersion curve derivatives up to the respective orders $n = 1, 2, 5$ are computed. Note that for the step size control in fact



(a) Maximum error in comparison to number of nodes.



(b) Maximum error in comparison to computation time.

Figure 5.12: Convergence of the maximum error of adaptive scheme of orders $n = 1, 2, 5$ for different values of the step size and backward check tolerances $\varepsilon_{\text{tol}}^{\text{step}} = \varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-\ell}$, $\ell = 2, \ldots, 7$, when applied to the fifth dispersion curves of the PhC TM mode band structure of Example 1.

the derivatives up to order $n+1$ have to be computed. The results show that the order of the convergence is approximately identical to the order of the scheme plus one. This is the expected convergence rate of Taylor expansions, which can be seen for example from the estimate (5.26) of the remainder (5.25) of the Taylor expansion.

On the other hand, we show in Figure 5.12b the maximum error in relation to the relative computation time, i. e. the ratio of computation time and the time needed to compute the reference solution. While the convergence order for the low order computations, i. e. $n = 1, 2$, is again approximately equal to $n+1$, the convergence order of the maximum error with respect to the computation time for $n = 5$ is smaller than the convergence order of the maximum error with respect to the number of nodes. This shows that the time for computing the dispersion curve derivatives by solving source problems is small but cannot be entirely neglected. Nevertheless, higher order expansions prove beneficial, in particular if we chose small step size and backward check tolerances.

5.6 Conclusions

In this chapter we introduced an adaptive path following algorithm for the eigenvalues of parameterized, nonlinear eigenvalue problems. We derived closed formulas for the derivatives of the eigenpaths and employed these formulas in a Taylor expansion. For the selection of the nodes of the Taylor expansion we proposed an adaptive algorithm based on the estimation of the remainder of the Taylor expansion.

As an example, we employed the proposed scheme to the computation of the dispersion curves of PhC and PhC waveguide band structure calculations, the latter one when using the supercell method. We showed that with the help of an additional refinement technique, the backward check, and the post-processing of the crossing check, we obtain reliable results also in involved cases such as avoided crossings.

With the help of the proposed scheme the computation time of band structure calculations is effectively reduced as our numerical results demonstrate. This reveals that the adaptive path following algorithm is an efficient procedure for PhC and PhC waveguide band structure calculations, and hence, meets the goal of *efficiency* in the context of PhC and PhC waveguide band structure calculations.

In Chapters 6 and 7 we will apply the adaptive algorithm also to problems with DtN and RtR transparent boundary conditions, i. e. to parameterized, nonlinear eigenvalue problems.

6 Dirichlet-to-Neumann transparent boundary conditions

In the introduction we identified two objectives for our work on PhC waveguide band structure calculations. After addressing *efficiency* in the previous chapter, let us now focus on *accuracy*, i. e. the exact computation of guided modes in PhC waveguides.

In Section 2.4 we introduced the supercell approach, which is a simple and frequently used procedure to compute approximations to guided modes in PhC waveguides, i. e. approximations to the solutions of the eigenvalue problem (2.19) in the infinite strip S. The modelling error of the supercell method depends on the confinement of the guided mode, which is the main disadvantage of the supercell method and which motivates the DtN approach presented in [Fli13] for the exact computation of guided modes independent of their confinement. The advantage of not introducing a modelling error comes with the disadvantage of transforming the problem to a nonlinear eigenvalue problem.

In this chapter we will explain the FE discretization and numerical solution of the resulting nonlinear eigenvalue problem in detail. We first published the numerical realization of the DtN approach in [KSF14]. This chapter comprises in addition the extension of the theory in Chapter 4, i. e. the computation of the group velocity and higher derivatives of the dispersion curves, and the procedure in Chapter 5, i. e. the usage of these derivatives in an adaptive Taylor expansion of the dispersion curves, to the case with DtN transparent boundary conditions.

The chapter is organized as follows: in Section 6.1 we introduce the DtN operators and comment on their characterization, differentiability and FE discretization. In Section 6.2 we present a nonlinear eigenvalue problem, that is posed in the defect cell C_0 and which is equivalent to the eigenvalue problem (2.19) in the infinite strip S. We derive formulas for the group velocity and any higher derivative of the dispersion curves, show the FE discretization of the nonlinear eigenvalue problem and comment on its numerical solution, before we present numerical results in Section 6.3, including the results of the path following algorithm. Finally, we give concluding remarks in Section 6.4.

6.1 The Dirichlet-to-Neumann operators

In this section we define the DtN operators, show there characterization using local cell problems and a quadratic operator equation, and prove their differentiability. Finally, we will elaborate on the discretization of the DtN operators and the local cell problems.

6.1.1 Definition of the Dirichlet-to-Neumann operators

As a first step towards the definition of the DtN operators, we introduce Dirichlet problems in the infinite half-strips $S^\pm$: for any $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ find $u^\pm \equiv u^\pm(\cdot;\omega,k,\varphi) \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ such that

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u^\pm - \omega^2\beta u^\pm = 0 \qquad \text{in } S^\pm, \tag{6.1a}$$

$$u^\pm = \varphi \qquad \text{on } \Gamma_0^\pm. \tag{6.1b}$$

Theorem 4.1 in [Fli13] gives the following result.

Theorem 6.1. *Let the unit cells $C_n^\pm$ and $C_{n+1}^\pm$, $n \in \mathbb{N}$, of the infinite half-strips $S^\pm$ be symmetric with respect to their common interface $\Gamma_n^\pm$, i. e. the material functions α and β are axis symmetric in $C_n^\pm \cup C_{n+1}^\pm$ with respect to $\Gamma_n^\pm$. Then the infinite half-strip problems (6.1) are well-posed in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ for all $\omega^2 \notin \sigma^\pm(k)$. If the unit cells of the infinite half-strips $S^\pm$ do not satisfy this symmetry property,*

the problems (6.1) are well-posed in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ *for all* $\omega^2 \notin \sigma^\pm(k)$ *except for a countable set of frequencies* ω^2*.*

The values of ω^2 for which (6.1) is not well-posed correspond to so-called *Dirichlet eigenvalues* of (6.1), i. e. eigenvalues of (6.1) when prescribing homogeneous Dirichlet boundary conditions on $\Gamma_0^\pm$. We shall call these values *global Dirichlet eigenvalues*, in order to distinguish them from Dirichlet eigenvalues of local cell problems we will introduce later.

Remark 6.2. *If the symmetry property in Theorem 6.1 is not fulfilled, the Dirichlet boundary conditions (6.1b) can be replaced by Robin boundary conditions, yielding well-posed problems for any* $\omega^2 \notin \sigma^\pm(k)$ *[Fli09, FJL10]. This approach will be discussed in Chapter 7.*

Remark 6.3. *PhC waveguides with square lattice and circular holes or rods of constant permittivity satisfy the symmetry property mentioned in Theorem 6.1, i. e. their unit cells* $C_n^\pm$ *and* $C_{n+1}^\pm$*,* $n \in \mathbb{N}$*, of the infinite half-strips* $S^\pm$ *are symmetric with respect to their common interface* $\Gamma_n^\pm$*. On the other hand, PhC waveguides with hexagonal lattice, e. g. see the waveguide in Example 2, do not satisfy the symmetry property, and hence, their can exist global Dirichlet eigenmodes in the semi-infinite strips* $S^\pm$ *with hexagonal lattice.*

In the sequel we shall assume that the infinite half-strip problems (6.1) are well-posed. Then, for any $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$, the DtN operators $\mathcal{D}^\pm(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm), \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm))$ are defined as

$$\mathcal{D}^\pm(\omega, k)\varphi = \pm\alpha\, \partial_2 u^\pm(\cdot\,; \omega, k, \varphi)\,|_{\Gamma_0^\pm}\,. \tag{6.2}$$

Proposition 6.4 (Proposition 4.3 in [Fli13])**.** *Let* $k \in B$ *and* $\omega^2 \in \mathbb{R}^+ \setminus \sigma^\pm(k)$*, then the DtN operators* $\mathcal{D}^\pm(\omega, k)$ *are continuous from* $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ *onto* $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ *and their norms are continuous with respect to* $\omega \in \mathbb{R}^+ \setminus \sigma^\pm(k)$*.*

Considering the half-strip problems (6.1) in variational formulation and choosing $u^\pm(\cdot\,; \omega, k, \psi)$, with $\psi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$, as test function, we find that the DtN operators satisfy

$$\begin{aligned}\int_{\Gamma_0^\pm} \mathcal{D}^\pm(\omega, k)\varphi\,\overline{\psi}\,\mathrm{d}s(\mathbf{x}) = -\int_{S^\pm} \alpha(\nabla + \mathrm{i}k\tbinom{1}{0}) u^\pm(\cdot\,; \omega, k, \varphi)\cdot(\nabla - \mathrm{i}k\tbinom{1}{0})\overline{u^\pm}(\cdot\,; \omega, k, \psi)\,\mathrm{d}\mathbf{x} \\ + \omega^2 \int_{S^\pm} \beta u^\pm(\cdot\,; \omega, k, \varphi)\overline{u^\pm}(\cdot\,; \omega, k, \psi)\,\mathrm{d}\mathbf{x}\end{aligned} \tag{6.3}$$

for any $\varphi, \psi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$.

6.1.2 Characterization of the Dirichlet-to-Neumann operators

In Eq. (6.2) the DtN operators are defined via Dirichlet problems (6.1) on an unbounded domain. In this subsection we summarize the results in [JLF06, Fli13] how to compute the DtN operators via local cell problems, i. e. by solving Dirichlet problems on a single periodicity cell, and a stationary Riccati equation.

To this end, we note that the infinite strips $S^\pm$ on top and bottom of the guide can be expressed as union of an infinite number of periodicity cells $C_n^\pm$, $n \in \mathbb{N}$, i. e.

$$S^\pm = \bigcup_{n=1}^{\infty} \left(C_n^\pm \cup \Gamma_n^\pm\right),$$

cf. Figure 2.3b. The top and bottom boundaries of these cells $C_n^\pm$ shall be denoted by $\Gamma_{n-1}^\pm$ and $\Gamma_n^\pm$, i. e.

$$\begin{aligned}\Gamma_0^\pm &= \overline{C_0} \cap \overline{C_1^\pm}, \\ \Gamma_n^\pm &= \overline{C_n^\pm} \cap \overline{C_{n+1}^\pm}, \qquad n \geq 1,\end{aligned}$$

see Figure 2.3b.

We also note that — due to the periodicity and the infinity of the half strips — all cells $C_n^\pm$ can be identified by the first cell $C_1^\pm$ and all boundaries $\Gamma_n^\pm$ can be identified by the first boundary $\Gamma_0^\pm$.

Therefore, let us introduce shift operators $\mathcal{S}_n^\pm \in \boldsymbol{\mathcal{L}}(\mathsf{C}^\infty(\Gamma_0^\pm), \mathsf{C}^\infty(\Gamma_n^\pm))$, $n \in \mathbb{N}$, defined by

$$\mathcal{S}_n^\pm \varphi(\mathbf{x}) = \varphi(\mathbf{x} \mp n\mathbf{a}_2^\pm). \tag{6.4}$$

By a density argument of $\mathsf{C}^\infty(\Gamma_n^\pm)$ in $\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_n^\pm)$ and $\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_n^\pm)$, respectively, we can extend the shift operators $\mathcal{S}_n^\pm$ to functions in $\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_n^\pm)$ and $\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_n^\pm)$. For simplicity of notation we shall write $\mathcal{S}^\pm := \mathcal{S}_1^\pm$. Furthermore, we introduce the inverse $(\mathcal{S}^\pm)^{-1}$ of $\mathcal{S}^\pm$ which is simply given by

$$(\mathcal{S}^\pm)^{-1}\varphi(\mathbf{x}) = \varphi(\mathbf{x} \pm \mathbf{a}_2^\pm). \tag{6.5}$$

These shift operators become important in the FE discretization which we will discuss in Sections 6.1.5 and 6.2.4.

With the help of these operators we can express the trace of the unique solution $u^\pm(\cdot\,;\omega,k,\varphi)$ of the Dirichlet problem (6.1) at the edges $\Gamma_n^\pm$, $n \in \mathbb{N}$, as

$$u^\pm(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_n^\pm} = \mathcal{S}_n^\pm(\mathcal{P}^\pm(\omega,k))^n\varphi,$$

with the *propagation operator* $\mathcal{P}^\pm(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^\pm))$ defined for any $\varphi \in \mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^\pm)$ by

$$\mathcal{P}^\pm(\omega,k)\varphi = (\mathcal{S}^\pm)^{-1}u^\pm(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_1^\pm}\,. \tag{6.6}$$

As shown in [JLF06], the propagation operator $\mathcal{P}^\pm(\omega,k)$ is the unique solution of the so-called *Riccati equation*

$$\mathcal{T}_{10}^\pm(\omega,k)(\mathcal{P}^\pm(\omega,k))^2 + (\mathcal{T}_{00}^\pm(\omega,k) + \mathcal{T}_{11}^\pm(\omega,k))\mathcal{P}^\pm(\omega,k) + \mathcal{T}_{01}^\pm(\omega,k) = 0 \tag{6.7}$$

with spectral radius strictly less than one. Here, the operators $\mathcal{T}_{ij}^\pm(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^\pm), \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm))$, $i,j = 0,1$, are defined by

$$\begin{aligned}
\mathcal{T}_{00}^\pm(\omega,k)\,\varphi &= \quad \mp\alpha\,\partial_2 u_0^\pm(\cdot\,;\omega,k, \quad \varphi))\,|_{\Gamma_0^\pm},\\
\mathcal{T}_{01}^\pm(\omega,k)\,\varphi &= (\mathcal{S}^\pm)^{-1}\big[\pm\alpha\,\partial_2 u_0^\pm(\cdot\,;\omega,k, \quad \varphi))\,|_{\Gamma_1^\pm}\big],\\
\mathcal{T}_{10}^\pm(\omega,k)\,\varphi &= \quad \mp\alpha\,\partial_2 u_1^\pm(\cdot\,;\omega,k,\mathcal{S}^\pm\varphi))\,|_{\Gamma_0^\pm},\\
\mathcal{T}_{11}^\pm(\omega,k)\,\varphi &= (\mathcal{S}^\pm)^{-1}\big[\pm\alpha\,\partial_2 u_1^\pm(\cdot\,;\omega,k,\mathcal{S}^\pm\varphi))\,|_{\Gamma_1^\pm}\big],
\end{aligned}$$

for any $\varphi \in \mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^\pm)$, where $u_i^\pm \equiv u_i^\pm(\cdot;\omega,k,\varphi) \in \mathsf{H}_{1\mathrm{p}}^1(\Delta, C_1^\pm, \alpha)$, $i = 0,1$, solve

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u_i^\pm - \omega^2\beta u_i^\pm = 0 \qquad \text{in } C_1^\pm, \tag{6.8a}$$

with Dirichlet boundary data

$$u_i^\pm = \delta_{ij}\varphi \qquad \text{on } \Gamma_j^\pm. \tag{6.8b}$$

Here and in the following, δ_{ij} denotes the usual Kronecker delta, i. e. $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$. Then the DtN operators $\mathcal{D}^\pm(\omega,k)$ are given by [JLF06]

$$\mathcal{D}^\pm(\omega,k) = -\mathcal{T}_{00}^\pm(\omega,k) - \mathcal{T}_{10}^\pm(\omega,k)\,\mathcal{P}^\pm(\omega,k). \tag{6.9}$$

Remark 6.5. *The Dirichlet cell problems (6.8) are well-posed except for a countable set of frequencies ω, the so-called* local Dirichlet eigenvalues*, and hence, the operators $\mathcal{T}_{ij}$, $i,j = 0,1$, are injective for almost any ω [JLF06] . Moreover, we can show — using the Fredholm theory — that the operators $\mathcal{T}_{00}$, $\mathcal{T}_{11}$ and $\mathcal{T}_{00} + \mathcal{T}_{11}$ are isomorphisms from $\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^\pm)$ onto $\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$. On the other hand, the operators $\mathcal{T}_{01}$ and $\mathcal{T}_{10}$ are compact and hence, they are not bijective [Fli09].*

Remark 6.6. *Replacing the Dirichlet boundary conditions (6.8b) by Robin boundary conditions, the local cell problems (6.8) become well-posed for all frequencies ω [Fli09, FJL10]. The characterization of the DtN operators with the help of local cell problems with Robin boundary conditions will be discussed in Chapter 7.*

6.1.3 Derivatives of the Dirichlet-to-Neumann operators

In this section we show that the DtN operators $\mathcal{D}^\pm(\omega, k)$ are differentiable with respect to the frequency ω and the quasi-momentum k inside the band gaps up to any order. Furthermore, we shall explain how to compute the derivatives of the DtN operators via local cell problems.

The differentiability of the DtN operators with respect to ω and k is an important property that is needed for solving the nonlinear eigenvalue problem with DtN transparent boundary conditions, that we will introduce later in Section 6.2. Moreover, we need the differentiability of the DtN operators up to any order in Section 6.2.3, where we will derive formulas for the derivatives of the dispersion curves with respect to the quasi-momentum k when prescribing DtN transparent boundary conditions on $\Gamma_0^\pm$ in contrast to periodic boundary conditions as done in Chapter 4.

Differentiability of the Dirichlet-to-Neumann operators

Let us again assume that $\omega^2 \notin \sigma^\pm(k)$ and let the half strip problem (6.1) as well as the local cell problems (6.8) be well-posed. Then the DtN operators are defined uniquely and can be computed using Eq. (6.9).

Let $u^\pm(\cdot\,;\omega,k,\varphi)$ be the unique solution in $\mathsf{H}^1_{\mathrm{1p}}(\Delta, S^\pm, \alpha)$ of the Dirichlet problem (6.1). Then we introduce $u^\pm_\omega(\cdot\,;\omega,k,\varphi)$ as the unique solution in $\mathsf{H}^1_{\mathrm{1p}}(\Delta, S^\pm, \alpha)$ of

$$-(\nabla + \mathrm{i}k\tbinom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\tbinom{1}{0})u^\pm_\omega - \omega^2\beta u^\pm_\omega = 2\omega\beta u^\pm \qquad \text{in } S^\pm, \tag{6.10a}$$

$$u^\pm_\omega = 0 \qquad \text{on } \Gamma_0^\pm, \tag{6.10b}$$

and $u^\pm_k(\cdot\,;\omega,k,\varphi)$ as the unique solution in $\mathsf{H}^1_{\mathrm{1p}}(\Delta, S^\pm, \alpha)$ of

$$-(\nabla + \mathrm{i}k\tbinom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\tbinom{1}{0})u^\pm_k - \omega^2\beta u^\pm_k = (2\alpha\,(-k + \mathrm{i}\partial_1) + \mathrm{i}\partial_1\alpha)\,u^\pm \qquad \text{in } S^\pm, \tag{6.11a}$$

$$u^\pm_k = 0 \qquad \text{on } \Gamma_0^\pm. \tag{6.11b}$$

Note that $\partial_1\alpha$ exists almost everywhere in $S^\pm$. The functions $u^\pm_\omega(\cdot\,;\omega,k,\varphi)$ and $u^\pm_k(\cdot\,;\omega,k,\varphi)$ are well-defined for almost any $\omega^2 \notin \sigma^\pm(k)$ thanks to the following proposition.

Proposition 6.7. *Let $\omega^2 \notin \sigma^\pm(k)$ and let the problem (6.1) be well-posed. Then the source problems (6.10) and (6.11) are well-posed.*

Proof. The result directly follows from the fact that by assumption the infinite half-strip problem (6.1) is well-posed for any Dirichlet data $\varphi \in \mathsf{H}^{1/2}_{\mathrm{1p}}(\Gamma_0^\pm)$. Then it is also clear that the problems (6.10) and (6.11), with well-defined source term and homogeneous Dirichlet boundary conditions on $\Gamma_0^\pm$ are well-posed. □

Now we can show the Fréchet-differentiability of $u(\cdot\,;\omega,k,\varphi)$.

Theorem 6.8. *Suppose that $\omega^2 \notin \sigma^\pm(k)$ and that the problem (6.1) is well-posed in a neighbourhood of ω^2. Then for any $\varphi \in \mathsf{H}^{1/2}_{\mathrm{1p}}(\Gamma_0^\pm)$, $u(\cdot\,;\omega,k,\varphi)$ is Fréchet-differentiable with respect to ω and k, and*

$$\frac{\partial u^\pm(\cdot\,;\omega,k,\varphi)}{\partial\omega} = u^\pm_\omega(\cdot\,;\omega,k,\varphi) \qquad \textit{and} \qquad \frac{\partial u^\pm(\cdot\,;\omega,k,\varphi)}{\partial k} = u^\pm_k(\cdot\,;\omega,k,\varphi).$$

Proof. For simplicity of notation, let us write $u^\pm(\omega) = u^\pm(\cdot\,;\omega,k,\varphi)$ and $u^\pm_\omega(\omega) = u^\pm_\omega(\cdot\,;\omega,k,\varphi)$ for any ω. Let $\omega_0^2 \notin \sigma^\pm(k)$ and suppose that the problem (6.1) is well-posed for any $\omega^2 \in ((\omega_0 - h_0)^2, (\omega_0 + h_0)^2)$ with some $h_0 > 0$. It is easy to see that for all $h \in (0, h_0)$

$$e^\pm_\omega(h) := u^\pm(\omega_0 + h) - u^\pm(\omega_0)$$

is a solution in $\mathsf{H}^1_{\mathrm{1p}}(\Delta, S^\pm, \alpha)$ of

$$-(\nabla + \mathrm{i}k\tbinom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\tbinom{1}{0})e^\pm_\omega(h) - \omega_0^2\beta e^\pm_\omega(h) = (2h\omega_0 + h^2)\beta u^\pm(\omega_0 + h) \qquad \text{in } S^\pm,$$

$$e^\pm_\omega(h) = 0 \qquad \text{on } \Gamma_0^\pm.$$

Due to Proposition 6.7, this problem is well-posed and we can deduce $\lim_{h\to 0} e^{\pm}_{\omega}(h) = 0$ in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$, which implies that $u^{\pm}(\omega)$ is continuous at $\omega = \omega_0$.

Now we introduce

$$\tilde{e}^{\pm}_{\omega}(h) := \frac{1}{h}\left(u^{\pm}(\omega_0 + h) - u^{\pm}(\omega_0) - h u^{\pm}_{\omega}(\omega_0)\right).$$

It is straightforward to verify that $\tilde{e}^{\pm}_{\omega}(h)$ is a solution in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$ of

$$\begin{aligned} -(\nabla + \mathrm{i}k(\tfrac{1}{0}))\cdot\alpha(\nabla + \mathrm{i}k(\tfrac{1}{0}))\tilde{e}^{\pm}_{\omega}(h) - \omega_0^2\beta\tilde{e}^{\pm}_{\omega}(h) &= 2\omega_0\beta(u^{\pm}(\omega_0+h) - u^{\pm}(\omega_0)) + h\beta u^{\pm}(\omega_0+h) && \text{in } S^{\pm},\\ \tilde{e}^{\pm}_{\omega}(h) &= 0 && \text{on } \Gamma^{\pm}_0. \end{aligned}$$

Again we can employ Proposition 6.7 and deduce that this problem is well-posed. Finally, using the continuity of $u^{\pm}(\omega)$ at $\omega = \omega_0$ we obtain $\lim_{h\to 0}\tilde{e}^{\pm}_{\omega}(h) = 0$ in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$ and hence, $u^{\pm}(\omega)$ is Fréchet-differentiable with respect to ω at $\omega = \omega_0$ with derivative $u^{\pm}_{\omega}(\cdot\,;\omega_0, k, \varphi)$.

The proof for the derivative with respect to k uses exactly the same ideas. We introduce the short notation $u^{\pm}(k) = u^{\pm}(\cdot\,;\omega, k, \varphi)$ and $u^{\pm}_k(k) = u^{\pm}_k(\cdot\,;\omega, k, \varphi)$ for all $k \in B$. Let $k_0 \in B$ and $\omega^2 \notin \sigma^{\pm}(k_0)$. Suppose that $\omega^2 \notin \sigma^{\pm}(k)$ and that the problem (6.1) is well-posed for any $k \in (k_0, k_0 + h)$ with some $h_0 > 0$. It is easy to see that for all $h \in (0, h_0)$

$$e^{\pm}_k(h) := u^{\pm}(k_0 + h) - u^{\pm}(k_0)$$

is a solution in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$ of

$$\begin{aligned} -(\nabla + \mathrm{i}k_0(\tfrac{1}{0}))\cdot\alpha(\nabla + \mathrm{i}k_0(\tfrac{1}{0}))e^{\pm}_k(h) - \omega^2\beta e^{\pm}_k(h) &\\ = h\left(2\alpha\left(-(k_0+h) + \mathrm{i}\partial_1\right) + \mathrm{i}\partial_1\alpha\right)u^{\pm}(k_0+h) && \text{in } S^{\pm},\\ e^{\pm}_k(h) = 0 && \text{on } \Gamma^{\pm}_0. \end{aligned}$$

This problem is well-posed thanks to Proposition 6.7, and then $\lim_{h\to 0} e^{\pm}_k(h) = 0$ in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$, which implies that $u^{\pm}(\cdot\,;\omega, k, \varphi)$ is continuous at $k = k_0$.

Now we introduce

$$\tilde{e}^{\pm}_k(h) := \frac{1}{h}\left(u^{\pm}(k_0 + h) - u^{\pm}(k_0) - h u^{\pm}_k(k_0)\right),$$

which is a solution in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$ of

$$\begin{aligned} -(\nabla + \mathrm{i}k_0(\tfrac{1}{0}))\cdot\alpha(\nabla + \mathrm{i}k_0(\tfrac{1}{0}))\tilde{e}^{\pm}_k(h) - \omega^2\beta\tilde{e}^{\pm}_k(h) &\\ = \left(2\alpha\left(-k_0 + \mathrm{i}\partial_1\right) + \mathrm{i}\partial_1\alpha\right)\left(u^{\pm}(k_0+h) - u^{\pm}(k_0)\right) &\\ \quad - 2h\alpha u^{\pm}(k_0+h) && \text{in } S^{\pm},\\ \tilde{e}^{\pm}_k(h) = 0 && \text{on } \Gamma^{\pm}_0. \end{aligned}$$

Once more using Proposition 6.7 we can deduce that this problem is well-posed. Finally, employing the continuity of $u^{\pm}(k)$ at $k = k_0$ we obtain $\lim_{h\to 0}\tilde{e}^{\pm}_k(h) = 0$ in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^{\pm}, \alpha)$ and hence, $u^{\pm}(k)$ is Fréchet-differentiable with respect to k at $k = k_0$ with derivative $u^{\pm}_{\omega}(\cdot\,;\omega, k_0, \varphi)$. □

Using the definition (6.2) of the DtN operators $\mathcal{D}^{\pm}(\omega, k)$, we deduce their Fréchet-differentiability with respect to ω and k.

Corollary 6.9. *Suppose that $\omega^2 \notin \sigma^{\pm}(k)$ and that the problem (6.1) is well-posed in a neighbourhood of ω^2. Then the DtN operators $\mathcal{D}^{\pm}(\omega, k)$ are differentiable with respect to the frequency ω and the quasi-momentum k, and for all $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$*

$$\frac{\partial\mathcal{D}^{\pm}}{\partial\omega}(\omega, k)\varphi = \pm\alpha\,\partial_2 u^{\pm}_{\omega}(\cdot\,;\omega, k, \varphi)\big|_{\Gamma^{\pm}_0} \tag{6.12a}$$

and

$$\frac{\partial\mathcal{D}^{\pm}}{\partial k}(\omega, k)\varphi = \pm\alpha\,\partial_2 u^{\pm}_{k}(\cdot\,;\omega, k, \varphi)\big|_{\Gamma^{\pm}_0}\,. \tag{6.12b}$$

Remark 6.10. *Iteratively repeating the same steps as above we can deduce that the DtN operators $\mathcal{D}^\pm(\omega,k)$ are differentiable to any order with respect to the frequency ω and the quasi-momentum k if $\omega^2 \notin \sigma^\pm(k)$ and the problem (6.1) is well-posed in a neighbourhood of ω^2.*

For simplicity of notation, let us write

$$\mathcal{D}^\pm_\omega(\omega,k) = \frac{\partial \mathcal{D}^\pm}{\partial \omega}(\omega,k) \qquad \text{and} \qquad \mathcal{D}^\pm_k(\omega,k) = \frac{\partial \mathcal{D}^\pm}{\partial k}(\omega,k)$$

in the sequel.

Characterization of the derivatives of Dirichlet-to-Neumann operators

For the characterization of the derivatives (6.12) of the DtN operators we employ the same concepts as in Section 6.1.2 for the characterization of the DtN operators. First we will show that the propagation operators $\mathcal{P}^\pm(\omega,k)$ are differentiable with respect to ω and k. Then we note that the same is true for the local DtN operators $\mathcal{T}^\pm_{ij}(\omega,k)$ and present their derivatives of with respect to ω and k. Finally, we show how to compute the derivatives $\mathcal{D}^\pm_\omega(\omega,k)$ and $\mathcal{D}^\pm_k(\omega,k)$ of the DtN operators with the help of these operators.

Analogously to Corollary 6.9, we obtain the following result for the propagation operators.

Corollary 6.11. *Suppose that $\omega^2 \notin \sigma^\pm(k)$ and that the problem (6.1) is well-posed in a neighbourhood of ω^2. Then the propagation operators $\mathcal{P}^\pm(\omega,k)$ are differentiable with respect to the frequency ω and the quasi-momentum k, and for all $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^\pm_0)$*

$$\frac{\partial \mathcal{P}^\pm}{\partial \omega}(\omega,k)\varphi = (\mathcal{S}^\pm)^{-1} u^\pm_\omega(\cdot\,;\omega,k,\varphi)|_{\Gamma^\pm_1} \qquad \text{and} \qquad \frac{\partial \mathcal{P}^\pm}{\partial k}(\omega,k)\varphi = (\mathcal{S}^\pm)^{-1} u^\pm_k(\cdot\,;\omega,k,\varphi)|_{\Gamma^\pm_1}\,.$$

Now we want to characterize first the derivatives of the propagation operators and then the derivatives of the DtN operators via solutions of local cell problems. In the following, we explain the characterization of the derivative with respect to ω, the ideas for the derivatives with respect to k are exactly the same.

To this end, we have to introduce the derivatives of the local DtN operators $\mathcal{T}^\pm_{ij}(\omega,k)$. Let us suppose that the Dirichlet cell problems (6.8) are well-defined and let us introduce for all $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^\pm_0)$ the unique solutions $u^\pm_{\omega,i}(\cdot\,;\omega,k,\varphi)$, $i=0,1$, in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, C^\pm_1, \alpha)$ of the new local cell problems

$$-(\nabla + \mathrm{i}k^{(1)}_0)\cdot\alpha(\nabla + \mathrm{i}k^{(1)}_0)u^\pm_{\omega,i} - \omega^2\beta u^\pm_{\omega,i} = 2\omega\beta u^\pm_i \qquad \text{in } C^\pm_1, \tag{6.13a}$$

$$u^\pm_{\omega,i} = 0 \qquad \text{on } \Gamma^\pm_0 \text{ and } \Gamma^\pm_1, \tag{6.13b}$$

where $u^\pm_i \equiv u^\pm_i(\cdot\,;\omega,k,\varphi)$ are the unique solutions of the local cell problems (6.8). Using exactly the same ideas as above, we can show that the operators $\mathcal{T}^\pm_{ij}(\omega,k)$ are Fréchet differentiable with respect to ω and for all $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^\pm_0)$

$$\begin{aligned}
\frac{\partial \mathcal{T}^\pm_{00}(\omega,k)}{\partial \omega}\varphi &= \mp\alpha\,\partial_2 u^\pm_{\omega,0}(\cdot\,;\omega,k,\quad\ \varphi))|_{\Gamma^\pm_0},\\
\frac{\partial \mathcal{T}^\pm_{01}(\omega,k)}{\partial \omega}\varphi &= (\mathcal{S}^\pm)^{-1}\big[\pm\alpha\,\partial_2 u^\pm_{\omega,0}(\cdot\,;\omega,k,\quad\ \varphi))|_{\Gamma^\pm_1}\big],\\
\frac{\partial \mathcal{T}^\pm_{10}(\omega,k)}{\partial \omega}\varphi &= \mp\alpha\,\partial_2 u^\pm_{\omega,1}(\cdot\,;\omega,k,\mathcal{S}^\pm\varphi))|_{\Gamma^\pm_0},\\
\frac{\partial \mathcal{T}^\pm_{11}(\omega,k)}{\partial \omega}\varphi &= (\mathcal{S}^\pm)^{-1}\big[\pm\alpha\,\partial_2 u^\pm_{\omega,1}(\cdot\,;\omega,k,\mathcal{S}^\pm\varphi))|_{\Gamma^\pm_1}\big].
\end{aligned}$$

Finally, we can uniquely characterize the derivatives of the propagation operators $\mathcal{P}^\pm(\omega,k)$.

Proposition 6.12. *The derivatives of $\mathcal{P}^\pm(\omega,k)$ with respect to ω are the unique solutions in $\mathcal{L}(\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^\pm_0))$ of*

$$\begin{aligned}
&\Big(\mathcal{T}^\pm_{10}(\omega,k)\mathcal{P}^\pm(\omega,k) + \mathcal{T}^\pm_{00}(\omega,k) + \mathcal{T}^\pm_{11}(\omega,k)\Big)\frac{\partial \mathcal{P}^\pm(\omega,k)}{\partial \omega} + \mathcal{T}^\pm_{10}(\omega,k)\frac{\partial \mathcal{P}^\pm(\omega,k)}{\partial \omega}\mathcal{P}^\pm(\omega,k)\\
&\qquad = -\frac{\partial \mathcal{T}^\pm_{10}(\omega,k)}{\partial \omega}\big(\mathcal{P}^\pm(\omega,k)\big)^2 - \left(\frac{\partial \mathcal{T}^\pm_{00}(\omega,k)}{\partial \omega} + \frac{\partial \mathcal{T}^\pm_{11}(\omega,k)}{\partial \omega}\right)\mathcal{P}^\pm(\omega,k) - \frac{\partial \mathcal{T}^\pm_{01}(\omega,k)}{\partial \omega}.
\end{aligned} \tag{6.14}$$

Proof. Differentiating Eq. (6.7) with respect to ω, it is easy to see that the derivatives of $\mathcal{P}^{\pm}(\omega,k)$ with respect to ω are solutions of Eq. (6.14). To deduce uniqueness, it suffices to show that the operator

$$\begin{aligned}\mathcal{T}_{\omega,k}:\ \boldsymbol{\mathcal{L}}(\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})) &\to \boldsymbol{\mathcal{L}}(\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm}))\\ \mathcal{X} &\mapsto \left(\mathcal{T}^{\pm}_{10}(\omega,k)\mathcal{P}^{\pm}(\omega,k)+\mathcal{T}^{\pm}_{00}(\omega,k)+\mathcal{T}^{\pm}_{11}(\omega,k)\right)\mathcal{X}+\mathcal{T}^{\pm}_{10}(\omega,k)\mathcal{X}\mathcal{P}^{\pm}(\omega,k)\end{aligned}$$

is injective. However, injectivity of this operator was already proven in [Coa12], where it occurs in the determination of the DtN operators for time domain problems. Finally, if there exist two solutions $\mathcal{P}_{\omega,1}$ and $\mathcal{P}_{\omega,2}$ of Eq. (6.14) then their difference satisfies $\mathcal{T}_{\omega,k}\left(\mathcal{P}_{\omega,1}-\mathcal{P}_{\omega,2}\right)=0$ and by injectivity of $\mathcal{T}_{\omega,k}$, the two solutions are necessarily the same. □

The techniques for solving the linear operator equation (6.14) on a discrete level will be discussed in Section 6.1.5.

Differentiating Eq. (6.9), we can deduce that the derivatives of the DtN operators $\mathcal{D}^{\pm}(\omega,k)$ with respect to ω read

$$\mathcal{D}^{\pm}_{\omega}(\omega,k)=-\frac{\partial\mathcal{T}^{\pm}_{00}(\omega,k)}{\partial\omega}-\frac{\partial\mathcal{T}^{\pm}_{10}(\omega,k)}{\partial\omega}\mathcal{P}^{\pm}(\omega,k)-\mathcal{T}^{\pm}_{10}(\omega,k)\frac{\partial\mathcal{P}^{\pm}(\omega,k)}{\partial\omega}. \tag{6.15}$$

The derivatives of the propagation operators $\mathcal{P}^{\pm}(\omega,k)$ and of the DtN operators $\mathcal{D}^{\pm}(\omega,k)$ with respect to k are characterized similarly by simply replacing all ω-derivatives in Eqs. (6.14) and (6.15) by k-derivatives. On the other hand, the k-derivatives of the operators $\mathcal{T}^{\pm}_{ij}(\omega,k)$, $i,j=0,1$, are for all $\varphi\in\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ given by

$$\begin{aligned}\frac{\partial\mathcal{T}^{\pm}_{00}(\omega,k)}{\partial k}\varphi &= \mp\alpha\,\partial_2 u^{\pm}_{k,0}(\cdot\,;\omega,k,\ \ \varphi))\big|_{\Gamma_0^{\pm}},\\ \frac{\partial\mathcal{T}^{\pm}_{01}(\omega,k)}{\partial k}\varphi &= (\mathcal{S}^{\pm})^{-1}\big[\pm\alpha\,\partial_2 u^{\pm}_{k,0}(\cdot\,;\omega,k,\ \ \varphi))\big|_{\Gamma_1^{\pm}}\big],\\ \frac{\partial\mathcal{T}^{\pm}_{10}(\omega,k)}{\partial k}\varphi &= \mp\alpha\,\partial_2 u^{\pm}_{k,1}(\cdot\,;\omega,k,\mathcal{S}^{\pm}\varphi))\big|_{\Gamma_0^{\pm}},\\ \frac{\partial\mathcal{T}^{\pm}_{11}(\omega,k)}{\partial k}\varphi &= (\mathcal{S}^{\pm})^{-1}\big[\pm\alpha\,\partial_2 u^{\pm}_{k,1}(\cdot\,;\omega,k,\mathcal{S}^{\pm}\varphi))\big|_{\Gamma_1^{\pm}}\big],\end{aligned}$$

where $u^{\pm}_{k,i}\equiv u^{\pm}_{k,i}(\cdot;\omega,k,\varphi)$ are the unique solutions in $\mathsf{H}^1_{1\mathrm{p}}(\Delta,C_1^{\pm},\alpha)$, $i=0,1$, of

$$-(\nabla+\mathrm{i}k\binom{1}{0})\cdot\alpha(\nabla+\mathrm{i}k\binom{1}{0})u^{\pm}_{k,i}-\omega^2\beta u^{\pm}_{k,i}=(2\alpha\,(-k+\mathrm{i}\partial_1)+\mathrm{i}\partial_1\alpha)\,u_i^{\pm} \qquad \text{in } C_1^{\pm}, \tag{6.16a}$$

$$u^{\pm}_{k,i}=0 \qquad \text{on } \Gamma_0^{\pm} \text{ and } \Gamma_1^{\pm}. \tag{6.16b}$$

Remark 6.13. *In contrast to the Dirichlet cell problems (6.8) to determine the DtN operators, the Dirichlet cell problems (6.13) and (6.16) to compute the ω- and k-derivatives of the DtN operators have homogeneous Dirichlet boundary conditions but a source term that depends on the solutions $u_i^{\pm}$, $i=0,1$, of the original cell problems (6.8).*

Extension to higher order derivatives

As elaborated in Remark 6.10 the DtN operators are differentiable with respect to ω and k up to any order. The same is true for the propagation operators and the local DtN operators. Hence, we can characterize the partial derivatives of the DtN operators with respect to ω and k of any order in a similar fashion like we characterized $\mathcal{D}^{\pm}_{\omega}(\omega,k)$ and $\mathcal{D}^{\pm}_{k}(\omega,k)$.

To this end, let us introduce $u_i^{\pm,(m,n)}(\cdot\,;\omega,k,\varphi)\in\mathsf{H}^1_{1\mathrm{p}}(\Delta,C_1^{\pm},\alpha)$, $m,n\in\mathbb{N}$, $i=0,1$, as the unique solution of

$$-(\nabla+\mathrm{i}k\binom{1}{0})\cdot\alpha(\nabla+\mathrm{i}k\binom{1}{0})u_i^{\pm,(m,n)}-\omega^2\beta u_i^{\pm,(m,n)}=f_i^{\pm,(m,n)} \qquad \text{in } C_1^{\pm}, \tag{6.17a}$$

$$u_i^{\pm,(m,n)}=0 \qquad \text{on } \Gamma_0^{\pm} \text{ and } \Gamma_1^{\pm} \tag{6.17b}$$

with

$$f_i^{\pm,(m,n)} = 2m\omega\beta u_i^{\pm,(m-1,n)} + m(m-1)\beta u_i^{\pm,(m-2,n)} + n\left(2\alpha(-k+\mathrm{i}\partial_1)+\mathrm{i}\partial_1\alpha\right) u_i^{\pm,(m,n-1)} - n(n-1)\alpha\, u_i^{\pm,(m,n-2)}.$$

Note that this notation implies

$$\begin{aligned} u_i^{\pm,(0,0)}(\cdot\,;\omega,k,\varphi) &= u_i^{\pm}(\cdot\,;\omega,k,\varphi),\\ u_i^{\pm,(1,0)}(\cdot\,;\omega,k,\varphi) &= u_{\omega,i}^{\pm}(\cdot\,;\omega,k,\varphi),\\ u_i^{\pm,(0,1)}(\cdot\,;\omega,k,\varphi) &= u_{k,i}^{\pm}(\cdot\,;\omega,k,\varphi). \end{aligned}$$

Introducing the short notation

$$\mathcal{T}_{ij}^{\pm,(m,n)}(\omega,k) := \frac{\partial^{m+n}\mathcal{T}_{i,j}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n},$$

$i,j=0,1$, we have

$$\begin{aligned} \mathcal{T}_{00}^{\pm,(m,n)}(\omega,k)\varphi &= \mp\alpha\,\partial_2 u_0^{\pm,(m,n)}(\cdot\,;\omega,k,\quad\varphi))\big|_{\Gamma_0^{\pm}},\\ \mathcal{T}_{01}^{\pm,(m,n)}(\omega,k)\varphi &= (\mathcal{S}^{\pm})^{-1}\big[\pm\alpha\,\partial_2 u_0^{\pm,(m,n)}(\cdot\,;\omega,k,\quad\varphi))\big|_{\Gamma_1^{\pm}}\big],\\ \mathcal{T}_{10}^{\pm,(m,n)}(\omega,k)\varphi &= \mp\alpha\,\partial_2 u_1^{\pm,(m,n)}(\cdot\,;\omega,k,\mathcal{S}^{\pm}\varphi))\big|_{\Gamma_0^{\pm}},\\ \mathcal{T}_{11}^{\pm,(m,n)}(\omega,k)\varphi &= (\mathcal{S}^{\pm})^{-1}\big[\pm\alpha\,\partial_2 u_1^{\pm,(m,n)}(\cdot\,;\omega,k,\mathcal{S}^{\pm}\varphi))\big|_{\Gamma_1^{\pm}}\big], \end{aligned}$$

for any $m,n\in\mathbb{N}_0$.

With these operators we can characterize the derivative of the propagation operator $\mathcal{P}^{\pm}(\omega,k)$. Similarly to Proposition 6.12, we find that the derivatives

$$\mathcal{P}^{\pm,(m,n)}(\omega,k) := \frac{\partial^{m+n}\mathcal{P}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n}$$

of the propagation operators $\mathcal{P}^{\pm}(\omega,k)$ of order m with respect to ω and order n with respect to k are the unique solutions of

$$\begin{aligned} 0 &= \frac{\partial^{m+n}}{\partial\omega^m\partial\omega^n}\left[\mathcal{T}_{10}^{\pm}\left(\mathcal{P}^{\pm}\right)^2 + \left(\mathcal{T}_{00}^{\pm}+\mathcal{T}_{11}^{\pm}\right)\mathcal{P}^{\pm} + \mathcal{T}_{01}^{\pm}\right]\\ &= \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^3(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathcal{T}_{10}^{\pm,(m_1,n_1)}\mathcal{P}^{\pm,(m_2,n_2)}\mathcal{P}^{\pm,(m_3,n_3)}\\ &\quad + \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^2(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \left(\mathcal{T}_{00}^{\pm,(m_1,n_1)}+\mathcal{T}_{11}^{\pm,(m_1,n_1)}\right)\mathcal{P}^{\pm,(m_2,n_2)}\\ &\quad + \mathcal{T}_{01}^{\pm,(m,n)} \end{aligned} \tag{6.18}$$

with the multinomial coefficient

$$\binom{m}{\mathbf{m}} = \binom{m}{m_1,\dots,m_d} = \frac{m!}{m_1!\cdots m_d!}, \tag{6.19}$$

$d\in\mathbb{N}$, $m\in\mathbb{N}_0$, $\mathbf{m}\in\mathbb{N}^d$, and the sets $\mathfrak{N}^2(m,n)$ and $\mathfrak{N}^3(m,n)$, that are defined by

$$\mathfrak{N}^d(m,n) := \Big\{(\mathbf{m},\mathbf{n})\in\mathbb{N}_0^d\times\mathbb{N}_0^d \;\Big|\; \textstyle\sum_{i=1}^d m_i = m \quad\text{and}\quad \sum_{i=1}^d n_i = n\Big\}. \tag{6.20}$$

Eq. (6.18) can be brought into a similar form like (6.14), i. e.

$$\begin{aligned}\Big(\mathcal{T}_{10}^{\pm}\mathcal{P}^{\pm}+\mathcal{T}_{00}^{\pm}+\mathcal{T}_{11}^{\pm}\Big)\mathcal{P}^{\pm,(m,n)}+\mathcal{T}_{10}^{\pm}\mathcal{P}^{\pm,(m,n)}\mathcal{P}^{\pm}&\\ =-\sum_{(\mathbf{m},\mathbf{n})\in\widetilde{\mathfrak{N}}^3_{\{2,3\}}(m,n)}\binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}}\mathcal{T}_{10}^{\pm,(m_1,n_1)}\mathcal{P}^{\pm,(m_2,n_2)}\mathcal{P}^{\pm,(m_3,m_3)}&\\ -\sum_{(\mathbf{m},\mathbf{n})\in\widetilde{\mathfrak{N}}^2_{\{2\}}(m,n)}\binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}}\left(\mathcal{T}_{00}^{\pm,(m_1,n_1)}+\mathcal{T}_{11}^{\pm,(m_1,n_1)}\right)\mathcal{P}^{\pm,(m_2,n_2)}&\\ -\mathcal{T}_{01}^{\pm,(m,n)},&\end{aligned}\tag{6.21}$$

where the sets $\widetilde{\mathfrak{N}}^2_{\{2\}}(m,n)$ and $\widetilde{\mathfrak{N}}^3_{\{2,3\}}(m,n)$ are defined by

$$\widetilde{\mathfrak{N}}^d_J(m,n):=\Big\{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^d(m,n)\;\Big|\;m_j+n_j\neq m+n\quad\forall j\in J\Big\},\tag{6.22}$$

where $J\subseteq\{1,\dots,d\}$.

Finally, differentiating Eq. (6.9) m times with respect to ω and n times with respect to k, we can deduce that the m-th ω- and n-th k-derivatives

$$\mathcal{D}^{\pm,(m,n)}(\omega,k):=\frac{\partial^{m+n}\mathcal{D}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n}$$

of the DtN operators $\mathcal{D}^{\pm}(\omega,k)$ read

$$\mathcal{D}^{\pm,(m,n)}=-\mathcal{T}_{00}^{\pm,(m,n)}-\sum_{p=0}^{m}\sum_{q=0}^{n}\binom{m}{p}\binom{n}{q}\mathcal{T}_{10}^{\pm,(p,q)}\mathcal{P}^{\pm,(m-p,n-q)}.\tag{6.23}$$

6.1.4 Variational formulation of the local cell problems

In this section we will introduce variational formulations of the local cell problems (6.8) for the computation of $u_i^{\pm}$, $i=0,1$, and of the local cell problems (6.17) for the computation of the ω- and k-derivatives of $u_i^{\pm}$.

To this end, we start by introducing a Dirichlet lift $w_i^{\pm}\equiv w_i^{\pm}(\cdot\,;\varphi)\in\mathsf{H}^1_{\mathrm{lp}}(C_1^{\pm})$ with $w_i^{\pm}|_{\Gamma_j^{\pm}}=\delta_{ij}\varphi$, and the space

$$\mathsf{H}^1_{\mathrm{lp},0}(C_1^{\pm}):=\Big\{u\in\mathsf{H}^1_{\mathrm{lp}}(C_1^{\pm})\text{ with }u|_{\Gamma_0^{\pm}}=u|_{\Gamma_1^{\pm}}=0\Big\}.$$

Then the weak solutions $u_i^{\pm}\equiv u_i^{\pm}(\cdot\,;\omega,k,\varphi)\in\mathsf{H}^1_{\mathrm{lp}}(C_1^{\pm})$ of the Dirichlet cell problems (6.8) can be decomposed into $u_i^{\pm}(\cdot\,;\omega,k,\varphi)=w_i^{\pm}(\cdot\,;\varphi)+u_{i,0}^{\pm}(\cdot\,;\omega,k,\varphi)$, where $u_{i,0}^{\pm}\equiv u_{i,0}^{\pm}(\cdot\,;\omega,k,\varphi)\in\mathsf{H}^1_{\mathrm{lp},0}(C_1^{\pm})$ satisfies

$$\mathfrak{b}_{C_1^{\pm}}(u_{i,0}^{\pm},v;\omega,k)=-\mathfrak{b}_{C_1^{\pm}}(w_i^{\pm},v;\omega,k)\tag{6.24}$$

for all $v\in\mathsf{H}^1_{\mathrm{lp},0}(C_1^{\pm})$, with the sesquilinear form $\mathfrak{b}_{C_1^{\pm}}(\cdot,\cdot\,;\omega,k)$ defined by

$$\mathfrak{b}_{C_1^{\pm}}(u,v;\omega,k):=\mathfrak{a}^{\alpha}_{C_1^{\pm}}(u,v)+k\mathfrak{c}^{\alpha,1}_{C_1^{\pm}}(u,v)+k^2\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u,v)-\omega^2\mathfrak{m}^{\beta}_{C_1^{\pm}}(u,v)\tag{6.25a}$$

with

$$\mathfrak{a}^{\alpha}_{C_1^{\pm}}(u,v):=\int_{C_1^{\pm}}\alpha\nabla u\cdot\nabla\overline{v}\,\mathrm{d}\mathbf{x},\tag{6.25b}$$

$$\mathfrak{c}^{\alpha,1}_{C_1^{\pm}}(u,v):=\int_{C_1^{\pm}}\mathrm{i}\alpha\left(u(\partial_1\overline{v})-(\partial_1u)\overline{v}\right)\,\mathrm{d}\mathbf{x},\tag{6.25c}$$

$$\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u,v):=\int_{C_1^{\pm}}\alpha u\overline{v}\,\mathrm{d}\mathbf{x},\tag{6.25d}$$

$$\mathfrak{m}^{\beta}_{C_1^{\pm}}(u,v):=\int_{C_1^{\pm}}\beta u\overline{v}\,\mathrm{d}\mathbf{x}.\tag{6.25e}$$

The DtN-like operators $\mathcal{T}_{ij}^{\pm}(\omega,k)$, $i,j=0,1$, then satisfy for any $\varphi,\psi\in\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^{\pm})$

$$\begin{aligned}\int_{\Gamma_0^{\pm}}\mathcal{T}_{ij}^{\pm}(\omega,k)\varphi\,\overline{\psi}\,\mathrm{d}s(\mathbf{x}) &= \int_{\Gamma_0^{\pm}}(\mathcal{S}^{\pm})^{-j}\left[\mp(-1)^j\alpha\,\partial_2 u_i^{\pm}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi)\right]\,\overline{\psi}\,\mathrm{d}s(\mathbf{x})\\ &= \int_{\Gamma_j^{\pm}}\mp(-1)^j\alpha\,\partial_2 u_i^{\pm}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi)\,(\mathcal{S}^{\pm})^j\overline{\psi}\,\mathrm{d}s(\mathbf{x})\\ &= \mathfrak{b}_{C_1^{\pm}}(u_i^{\pm}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi);\omega,k),\end{aligned}\tag{6.26}$$

where we used the relations (6.4) and (6.5) and integration by parts.

Note that from Eq. (6.24) it follows that $\mathfrak{b}_{C_1^{\pm}}(u_i^{\pm}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),v;\omega,k)=0$ if $v\in\mathsf{H}^1_{1\mathrm{p},0}(C_1^{\pm})$. The term on the right hand side of Eq. (6.26) has exactly this form only that $w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi)$ has a non-vanishing trace on $\Gamma_j^{\pm}$.

Now we proceed with the variational formulation of (6.17). Considering that

$$\int_{C_1^{\pm}}\partial_1\alpha\,u\,v\,\mathrm{d}\mathbf{x} = -\int_{C_1^{\pm}}\alpha\,(\partial_1 u\,v+u\,\partial_1 v)\,\mathrm{d}\mathbf{x}$$

for all $u,v\in\mathsf{H}^1_{1\mathrm{p},0}(C_1^{\pm})$, it is easy to see that (6.17) is equivalent to: find $u_i^{\pm,(m,n)}\in\mathsf{H}^1_{1\mathrm{p},0}(C_1^{\pm})$ such that

$$\begin{aligned}\mathfrak{b}_{C_1^{\pm}}(u_i^{\pm,(m,n)},v;\omega,k) &= 2m\omega\,\mathfrak{m}^{\beta}_{C_1^{\pm}}(u_i^{\pm,(m-1,n)},v)+m(m-1)\mathfrak{m}^{\beta}_{C_1^{\pm}}(u_i^{\pm,(m-2,n)},v)\\ &\quad-2nk\,\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u_i^{\pm,(m,n-1)},v)-n\,\mathfrak{c}^{\alpha,1}_{C_1^{\pm}}(u_i^{\pm,(m,n-1)},v)-n(n-1)\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u_i^{\pm,(m,n-2)},v)\end{aligned}\tag{6.27}$$

for all $v\in\mathsf{H}^1_{1\mathrm{p},0}(C_1^{\pm})$. Similarly to (6.26) we can deduce that for any $\varphi,\psi\in\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ the derivatives of the local DtN operators $\mathcal{T}_{ij}^{\pm}(\omega,k)$, $i,j=0,1$, of order m with respect to ω and order n with respect to k satisfy

$$\begin{aligned}\int_{\Gamma_0^{\pm}}\mathcal{T}_{ij}^{\pm,(m,n)}(\omega,k)\varphi\,\overline{\psi}\,\mathrm{d}s(\mathbf{x}) &= \mathfrak{b}_{C_1^{\pm}}(u_i^{\pm,(m,n)}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi);\omega,k)\\ &\quad-2m\omega\,\mathfrak{m}^{\beta}_{C_1^{\pm}}(u_i^{\pm,(m-1,n)}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi))\\ &\quad-m(m-1)\,\mathfrak{m}^{\beta}_{C_1^{\pm}}(u_i^{\pm,(m-2,n)}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi))\\ &\quad+2nk\,\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u_i^{\pm,(m,n-1)}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi))\\ &\quad+n\,\mathfrak{c}^{\alpha,1}_{C_1^{\pm}}(u_i^{\pm,(m,n-1)}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi))\\ &\quad+n(n-1)\,\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u_i^{\pm,(m,n-2)}(\cdot\,;\omega,k,(\mathcal{S}^{\pm})^i\varphi),w_j^{\pm}(\cdot\,;(\mathcal{S}^{\pm})^j\psi)).\end{aligned}$$

6.1.5 Discretization

In this section we discuss the discretization of the local cell problems, the solution of the discrete Riccati equation and the computation of the discrete DtN operators and its derivatives. However, before we start with discussing the discretization of the local cell problems we introduce the FE spaces.

High-order finite element spaces

For the discretization of the variational formulations (6.24) of the local cell problems (6.8), we need FE subspaces of $\mathsf{H}^1_{1\mathrm{p}}(C_1^{\pm})$ and its trace spaces $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_i^{\pm})$, $i=0,1$. We shall simultaneously introduce the FE subspace of $\mathsf{H}^1_{1\mathrm{p}}(C_0)$, which we will employ later in Section 6.2, when transforming the eigenvalue problem (2.19) in the infinite strip S to an eigenvalue problem in the defect cell C_0 using the DtN operators $\mathcal{D}^{\pm}$.

To this end, let us first discuss the FE meshes. Similarly to the FE meshes of PhC unit cells and supercells of PhC waveguides, that we discussed in Section 2.5, we assume that the domains C_0 and $C_1^{\pm}$ are partitioned into possibly curved geometrical cells, that are either quadrilaterals or triangles, see for

example the mesh of the domain $C_1^+ \cup C_0 \cup C_1^-$ with curved, quadrilateral cells in Figure 6.1. The meshes $\mathfrak{M}(C_0)$ and $\mathfrak{M}(C_1^\pm)$ are assumed to be periodic in direction $\mathbf{a}_1$, i. e. for each edge of a geometrical cell on the left boundary there exists an edge on the right boundary, which is shifted by $\mathbf{a}_1$. On the other hand, we do not necessarily have to assume that the mesh $\mathfrak{M}(C_0)$ is periodic in direction of $\mathbf{a}_2^0$. However, the meshes $\mathfrak{M}(C_1^\pm)$ have to be again periodic in direction $\mathbf{a}_2^\pm$, i. e. for each edge of a geometrical cell on the boundary $\Gamma_0^\pm$ there exists an edge on the boundary $\Gamma_1^\pm$, that is shifted by $\pm\mathbf{a}_2^\pm$. Moreover, we need that the meshes $\mathfrak{M}(C_0)$ and $\mathfrak{M}(C_0^\pm)$ coincide on their interfaces $\Gamma_0^\pm$, i. e. the set of of edges of $\mathfrak{M}(C_0)$ and $\mathfrak{M}(C_0^\pm)$ are identical and define the geometrical cells of the interface meshes $\mathfrak{M}(\Gamma_0^\pm)$. The geometrical cells K in $\mathfrak{M}(\Gamma_0^\pm)$ can alternatively be defined by affine maps F_K from the reference interval $\widehat{K} = [0, 1]$.

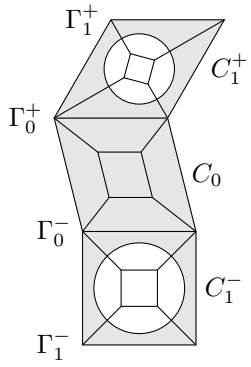


Figure 6.1: Mesh with curved, quadrilateral cells of the defect cell C_0 and the PhC unit cells $C_1^\pm$ on top and bottom, the interfaces $\Gamma_0^\pm$ and the top and bottom boundaries $\Gamma_1^\pm$.

Based on the meshes $\mathfrak{M}(\Omega)$, $\Omega = C_0, C_1^\pm, \Gamma_0^\pm$, we can define discrete subspaces of $\mathsf{H}^1_{1\mathrm{p}}(C_0)$, $\mathsf{H}^1_{1\mathrm{p}}(C_1^\pm)$ and $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ as

$$\mathsf{S}^p_{1\mathrm{p}}(\Omega) := \{v \in \mathsf{H}^1_{1\mathrm{p}}(\Omega) \cap \mathsf{C}^0(\overline{\Omega}) : v|_K \circ F_K \in \mathsf{P}^p(\widehat{K}(K)) \ \ \forall K \in \mathfrak{M}(\Omega)\},$$

where p is the polynomial degree, $\mathsf{P}^p(\widehat{K})$ is the space of polynomials with maximal (total) degree p as defined in (2.27), and Ω is either C_0, $C_1^\pm$ or $\Gamma_0^\pm$. Due to the periodicity of the meshes $\mathfrak{M}(C_1^\pm)$ in direction of $\mathbf{a}_2^\pm$, the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ shifted to $\Gamma_1^\pm$ are enclosed in $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_1^\pm)$ and hence, form a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_1^\pm)$. Thus, we can refrain from defining an independent FE subspace for $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_1^\pm)$.

Let us from now on assume that the FE subspaces of $\mathsf{H}^1_{1\mathrm{p}}(C_0)$, $\mathsf{H}^1_{1\mathrm{p}}(C_1^\pm)$ and $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ have the same maximal (total) polynomial degree p. We will denote the dimensions of these FE spaces by

$$\begin{aligned} N(C_0) &:= \dim \mathsf{S}^p_{1\mathrm{p}}(C_0), \\ N(C_1^\pm) &:= \dim \mathsf{S}^p_{1\mathrm{p}}(C_1^\pm), \\ N(\Gamma_0^\pm) &:= \dim \mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm). \end{aligned}$$

It will prove useful to introduce basis functions $b_{C_0,n}$, $n = 1, \ldots, N(C_0)$, of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ which are ordered such that

- the basis functions with index $n \in \mathfrak{S}(C_0, \Gamma_0^+) := \{1, \ldots, N(\Gamma_0^+)\}$ vanish on Γ_0^-, but their traces on Γ_0^+ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^+)$,
- the basis functions with index $n \in \mathfrak{S}(C_0, \Gamma_0^-) := \{N(\Gamma_0^+) + 1, \ldots, N(\Gamma_0^+) + N(\Gamma_0^-)\}$ vanish on Γ_0^+, but their traces on Γ_0^- build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^-)$, and
- the basis functions with index $n \in \mathfrak{S}(C_0, C_0) := \{N(\Gamma_0^+) + N(\Gamma_0^-) + 1, \ldots, N(C_0)\}$ vanish on $\Gamma_0^\pm$.

With this special ordering we can relate the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ and the traces of the basis functions

of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$, i. e.

$$b_{\Gamma_0^+,n} = \sum_{m=1}^{N(\Gamma_0^+)} Q^+_{C_0,mn} b_{C_0,m}|_{\Gamma_0^+}, \tag{6.28a}$$

$$b_{\Gamma_0^-,n} = \sum_{m=1}^{N(\Gamma_0^-)} Q^-_{C_0,mn} b_{C_0,N(\Gamma_0^+)+m}|_{\Gamma_0^-}, \tag{6.28b}$$

with permutation matrices $\mathbf{Q}^+_{C_0} \in \mathbb{R}^{N(\Gamma_0^+)\times N(\Gamma_0^+)}$ and $\mathbf{Q}^-_{C_0} \in \mathbb{R}^{N(\Gamma_0^-)\times N(\Gamma_0^-)}$. Analogously, we assume that the basis functions $b_{C_1^\pm,n}$, $n = 1,\dots,N(C_1^\pm)$, of $\mathsf{S}^p_{1\mathrm{p}}(C_1^\pm)$ are ordered such that

- the basis functions with index

$$n \in \mathfrak{S}(C_1^\pm,\Gamma_0^\pm) := \{1,\dots,N(\Gamma_0^\pm)\} \tag{6.29a}$$

  vanish on $\Gamma_1^\pm$, but their traces on $\Gamma_0^\pm$ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$,

- the basis functions with index

$$n \in \mathfrak{S}(C_1^\pm,\Gamma_1^\pm) := \{N(\Gamma_0^\pm)+1,\dots,2\,N(\Gamma_0^\pm)\} \tag{6.29b}$$

  vanish on $\Gamma_0^\pm$, but their traces on $\Gamma_1^\pm$ shifted to $\Gamma_0^\pm$, using the shift operator $(\mathcal{S}^\pm)^{-1}$ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ as well, and

- the basis functions with index

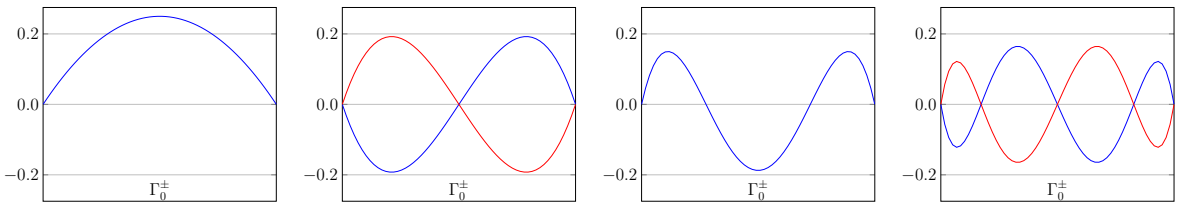
$$n \in \mathfrak{S}(C_1^\pm,C_1^\pm) := \{2\,N(\Gamma_0^\pm)+1,\dots,N(C_1^\pm)\} \tag{6.29c}$$

  vanish on $\Gamma_0^\pm$ and $\Gamma_1^\pm$.

Hence, we obtain analogously to (6.28) a relation of the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ and the traces of the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(C_1^\pm)$, i. e.

$$b_{\Gamma_0^\pm,n} = \sum_{m=1}^{N(\Gamma_0^\pm)} Q^0_{C_1^\pm,mn} b_{C_1^\pm,m}|_{\Gamma_0^\pm} = \sum_{m=1}^{N(\Gamma_0^\pm)} Q^1_{C_1^\pm,mn} b_{C_1^\pm,m+N(\Gamma_0^\pm)}|_{\Gamma_1^\pm} \tag{6.30}$$

with matrices $\mathbf{Q}^i_{C_1^\pm} \in \mathbb{R}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$, $i = 0, 1$.



(a) Edge function of second order. (b) Edge functions of third order. (c) Edge function of fourth order. (d) Edge functions of fifth order.

Figure 6.2: Basis functions of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ according to Karniadakis and Sherwin [KS05] if no h-refinement is applied the edge $\Gamma_0^\pm$, see the coarse mesh in Figure 6.1. Note that due to periodicity, the first order basis function, which is not shown in this figure, is constant with value one. While the basis functions of even order are uniquely defined, the basis functions of odd order are not unique (blue and red curves) and depend on the local and global orientation of the edge.

Using the same sort of shape functions for the spaces $\mathsf{S}^p_{1\mathrm{p}}(C_0)$, $\mathsf{S}^p_{1\mathrm{p}}(C_1^\pm)$ and $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$, the matrices $\mathbf{Q}^\pm_{C_0}$ and $\mathbf{Q}^i_{C_1^\pm}$, $i = 0, 1$, have the structure of permutation matrices with entries ± 1, where there are only

entries -1, if the corresponding edge functions are of odd order and the global and local orientations of the edge, which are responsible for the direction, mismatch. We use a hierarchical family of shape functions proposed by Karniadakis and Sherwin [KS05] based on integrated Legendre polynomials. Their edge functions are shown in Figure 6.2, which illustrates the possibility of mismatching orientations.

Now we ready to discuss the discretization of the local cell problems.

Discretization of the local cell problems

We aim to find approximate solutions $u^{\pm}_{i,h}(\cdot\,;\omega,k,\varphi_h) \in \mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$ to the cell problems (6.24), from which we can construct the discrete local DtN operators $\mathcal{T}_{ij,h}(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0))$, $i,j=0,1$, that can be represented in terms of the local DtN matrices $\mathbf{T}^{\pm}_{ij}(\omega,k) \in \mathbb{C}^{N(\Gamma^{\pm}_0)\times N(\Gamma^{\pm}_0)}$, $i,j=0,1$, with entries

$$T^{\pm}_{ij,mn}(\omega,k) = \int_{\Gamma^{\pm}_0} \mathcal{T}^{\pm}_{ij}(\omega,k) b_{\Gamma^{\pm}_0,n}\, \overline{b_{\Gamma^{\pm}_0,m}}\, \mathrm{d}s(\mathbf{x}). \tag{6.31}$$

With the special ordering of the basis functions and the relation (6.30) between the traces of the basis functions in $\mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$ and $\mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0)$ we can define the discrete Dirichlet lifts

$$w^{\pm}_{0,h}(\cdot\,;b_{\Gamma^{\pm}_0,n}) = \sum_{m=1}^{N(\Gamma^{\pm}_0)} Q^0_{C^{\pm}_1,mn} b_{C^{\pm}_1,m}, \qquad w^{\pm}_{1,h}(\cdot\,;b_{\Gamma^{\pm}_0,n}) = \sum_{m=1}^{N(\Gamma^{\pm}_0)} Q^0_{C^{\pm}_1,mn} b_{C^{\pm}_1,m+N(\Gamma^{\pm}_0)}.$$

Let us introduce the matrix

$$\mathbf{B}_{C^{\pm}_1}(\omega,k) = \mathbf{A}^{\alpha}_{C^{\pm}_1} + k\mathbf{C}^{\alpha,1}_{C^{\pm}_1} + k^2\mathbf{M}^{\alpha}_{C^{\pm}_1} - \omega^2\mathbf{M}^{\beta}_{C^{\pm}_1} \mathbb{C}^{N(C^{\pm}_1)\times N(C^{\pm}_1)} \tag{6.32}$$

where the matrices $\mathbf{A}^{\alpha}_{C^{\pm}_1}, \mathbf{C}^{\alpha,1}_{C^{\pm}_1}, \mathbf{M}^{\alpha}_{C^{\pm}_1}, \mathbf{M}^{\beta}_{C^{\pm}_1} \in \mathbb{R}^{N(C^{\pm}_1)\times N(C^{\pm}_1)}$ have entries

$$\begin{aligned}
A^{\alpha}_{C^{\pm}_1,mn} &= \mathfrak{a}^{\alpha}_{C^{\pm}_1}(b_{C^{\pm}_1,n}, b_{C^{\pm}_1,m}),\\
C^{\alpha,1}_{C^{\pm}_1,mn} &= \mathfrak{c}^{\alpha,1}_{C^{\pm}_1}(b_{C^{\pm}_1,n}, b_{C^{\pm}_1,m}),\\
M^{\alpha}_{C^{\pm}_1,mn} &= \mathfrak{m}^{\alpha}_{C^{\pm}_1}(b_{C^{\pm}_1,n}, b_{C^{\pm}_1,m}),\\
M^{\beta}_{C^{\pm}_1,mn} &= \mathfrak{m}^{\beta}_{C^{\pm}_1}(b_{C^{\pm}_1,n}, b_{C^{\pm}_1,m}),
\end{aligned}$$

$m,n=1,\ldots,N(C^{\pm}_1)$, with the sesquilinear forms as given in Eq. (6.25).

Let $N_0(C^{\pm}_1) := N(C^{\pm}_1) - 2N(\Gamma^{\pm}_0)$. Furthermore, let $\mathbf{e}^{N(\Gamma^{\pm}_0)}_n \in \mathbb{R}^{N(\Gamma^{\pm}_0)}$ be the n-th unit vector of dimension $N(\Gamma^{\pm}_0)$, and let $\mathbf{B}_{C^{\pm}_1}(\Omega_1,\Omega_2;\omega,k)$ with $\Omega_1,\Omega_2 \in \{C^{\pm}_1,\Gamma^{\pm}_0,\Gamma^{\pm}_1\}$ be the block with row indices $\mathfrak{S}(C^{\pm}_1,\Omega_1)$ and column indices $\mathfrak{S}(C^{\pm}_1,\Omega_2)$, cf. Eq. (6.29), of the matrix $\mathbf{B}_{C^{\pm}_1}(\omega,k)$. If the arguments Ω_1 and Ω_2 are replaced by a dot, all rows and columns, respectively, are considered. Then we can write the cell problems for $u^{\pm}_{i,h}$ as linear systems of equations

$$\mathbf{B}_{C^{\pm}_1}(C^{\pm}_1,C^{\pm}_1;\omega,k)\mathbf{u}^{\pm}_{i,0,h}(\omega,k,b_{\Gamma^{\pm}_0,n}) = -\mathbf{B}_{C^{\pm}_1}(C^{\pm}_1,\Gamma^{\pm}_i;\omega,k)\mathbf{Q}^i_{C^{\pm}_1}\mathbf{e}^{N(\Gamma^{\pm}_0)}_n, \quad n=1,\ldots,N(\Gamma^{\pm}_0),$$

where $\mathbf{u}^{\pm}_{i,0,h}(\omega,k,b_{\Gamma^{\pm}_0,n}) \in \mathbb{C}^{N_0(C^{\pm}_1)}$ is the coefficient vector of $u^{\pm}_{i,0,h}(\cdot\,;\omega,k,b_{\Gamma^{\pm}_0,n}) \in \mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1) \cap \mathsf{H}^1_{1\mathrm{p},0}(C^{\pm}_1)$ with respect to the basis functions $b_{C^{\pm}_1,j}$, $j \in \mathfrak{S}(C^{\pm}_1,C^{\pm}_1)$. These discrete local cell problems are well-posed as long as the mesh width h is chosen small enough and the polynomial degree p is large enough [SS11, Thm. 4.2.9], [MS11]. The discrete cell solution $u^{\pm}_{i,h}$ can then be represented by a vector $\mathbf{u}^{\pm}_{i,h}(\omega,k,b_{\Gamma^{\pm}_0,n}) \in \mathbb{C}^{N(C^{\pm}_1)}$, whose entries with indices $\mathfrak{S}(C^{\pm}_1,\Gamma^{\pm}_i)$ are set to the n-th column of $\mathbf{Q}^i_{C^{\pm}_1}$, the entries with indices $\mathfrak{S}(C^{\pm}_1,C^{\pm}_1)$ are set to $\mathbf{u}^{\pm}_{i,0,h}(\omega,k,b_{\Gamma^{\pm}_0,n})$ and the remaining entries are set to zero.

We can collect the coefficient vectors $\mathbf{u}^{\pm}_{i,0,h}(\omega,k,b_{\Gamma^{\pm}_0,n})$ for $n=1,\ldots,N(\Gamma^{\pm}_0)$ in (rectangular) matrices $\mathbf{U}^{\pm}_{i,0,h}(\omega,k) \in \mathbb{C}^{N_0(C^{\pm}_1)\times N(\Gamma^{\pm}_0)}$ that solve

$$\mathbf{B}_{C^{\pm}_1}(C^{\pm}_1,C^{\pm}_1;\omega,k)\mathbf{U}^{\pm}_{i,0,h}(\omega,k) = -\mathbf{B}_{C^{\pm}_1}(C^{\pm}_1,\Gamma^{\pm}_i;\omega,k)\mathbf{Q}^i_{C^{\pm}_1}. \tag{6.33}$$

Similarly to above, we define $\mathbf{U}^{\pm}_{i,h}(\omega,k) \in \mathbb{C}^{N(C_1^{\pm})\times N(\Gamma_0^{\pm})}$, whose rows with indices $\mathfrak{S}(C_1^{\pm},\Gamma_i^{\pm})$ are set to $\mathbf{Q}^i_{C_1^{\pm}}$, the rows with indices $\mathfrak{S}(C_1^{\pm},C_1^{\pm})$ are set to $\mathbf{U}^{\pm}_{i,0,h}(\omega,k)$ and the remaining rows are set to zero.

Inserting the basis functions $b_{\Gamma_0^{\pm},n}$, $n=1,\dots,N(\Gamma_0^{\pm})$, of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ into (6.26) yields

$$\mathbf{T}^{\pm}_{ij}(\omega,k) = (\mathbf{Q}^j_{C_1^{\pm}})^{\mathrm{T}}\mathbf{B}_{C_1^{\pm}}(\Gamma_j^{\pm},\cdot)\mathbf{U}^{\pm}_{i,h}(\omega,k),$$

or

$$\mathbf{T}^{\pm}_{ij}(\omega,k) = (\mathbf{Q}^j_{C_1^{\pm}})^{\mathrm{T}}\mathbf{B}_{C_1^{\pm}}(\Gamma_j^{\pm},C_1^{\pm})\mathbf{U}^{\pm}_{i,0,h}(\omega,k) + (\mathbf{Q}^j_{C_1^{\pm}})^{\mathrm{T}}\mathbf{B}_{C_1^{\pm}}(\Gamma_j^{\pm},\Gamma_i^{\pm})\mathbf{Q}^i_{C_1^{\pm}},$$

which can be rewritten when solving (6.33) for $\mathbf{U}^{\pm}_{i,0,h}$ in the from

$$\mathbf{T}^{\pm}_{ij}(\omega,k) = (\mathbf{Q}^j_{C_1^{\pm}})^{\mathrm{T}}\left(\mathbf{B}_{C_1^{\pm}}(\Gamma_j^{\pm},\Gamma_i^{\pm}) - \mathbf{B}_{C_1^{\pm}}(\Gamma_j^{\pm},C_1^{\pm})\mathbf{B}_{C_1^{\pm}}(C_1^{\pm},C_1^{\pm})^{-1}\mathbf{B}_{C_1^{\pm}}(C_1^{\pm},\Gamma_i^{\pm})\right)\mathbf{Q}^i_{C_1^{\pm}}. \tag{6.34}$$

Here we omitted the (ω,k)-dependence of the matrices $\mathbf{B}_{C_1^{\pm}}(\Omega_1,\Omega_2;\omega,k)$, $\Omega_1,\Omega_2 \in \{C_1^{\pm},\Gamma_0^{\pm},\Gamma_1^{\pm}\}$.

Solution of the discrete Riccati equation

Using the basis functions $b_{\Gamma_0^{\pm},n}$, $n\in\{1,\dots,N(\Gamma_0^{\pm})\}$, of the discrete space $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$, the propagation operators $\mathcal{P}^{\pm}_h(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm}))$ on the discrete spaces are represented by matrices $\mathbf{P}^{\pm}(\omega,k)\in\mathbb{C}^{N(\Gamma_0^{\pm})\times N(\Gamma_0^{\pm})}$ with entries $P^{\pm}_{mn}(\omega,k)\in\mathbb{C}$, $m,n=1,\dots,N(\Gamma_0^{\pm})$, satisfying

$$\mathcal{P}^{\pm}(\omega,k)\, b_{\Gamma_0^{\pm},n} = \sum_{m=1}^{N(\Gamma_0^{\pm})} P^{\pm}_{mn}(\omega,k)\, b_{\Gamma_0^{\pm},m}. \tag{6.35}$$

The Riccati equation (6.7) is fulfilled for any $\varphi \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ the operators are applied to. A discrete Riccati equation results if we apply the operators to a basis of the discrete space $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ and take the duality product with this basis.

Using the matrices $\mathbf{T}^{\pm}_{ij}(\omega,k)$ with entries as given in (6.31) and the propagation matrix $\mathbf{P}^{\pm}(\omega,k)$ as defined in (6.35), we can write the discrete Riccati equation as a linear system of equations

$$\mathbf{T}^{\pm}_{10}(\omega,k)(\mathbf{P}^{\pm}(\omega,k))^2 + (\mathbf{T}^{\pm}_{00}(\omega,k) + \mathbf{T}^{\pm}_{11}(\omega,k))\mathbf{P}^{\pm}(\omega,k) + \mathbf{T}^{\pm}_{01}(\omega,k) = \mathbf{0}. \tag{6.36}$$

Considering that the discretization preserves the periodicity properties of $C_1^{\pm}$ in $\mathbf{a}_2$-direction we deduce that the propagation matrix $\mathbf{P}^{\pm}(\omega,k)$ is the unique matrix satisfying Eq. (6.36) with eigenvalues whose magnitude is strictly less than one.

In [JLF06] Joly and coworkers proposed a modified Newton method to solve the matrix valued problem (6.36) where the spectral constraint is integrated implicitly into the algorithm. This modified Newton method only requires the matrix $\mathbf{T}^{\pm}_{00}(\omega,k)+\mathbf{T}^{\pm}_{11}(\omega,k)$ to be invertible, which is guaranteed by the fact that the corresponding linear operator $\mathcal{T}^{\pm}_{00}(\omega,k)+\mathcal{T}^{\pm}_{11}(\omega,k)$ is an isomorphism, see Remark 6.5, and by the fact that the discrete local cell problems — as already mentioned above — are well-posed as long as the mesh width h is chosen small enough and the polynomial degree p is large enough.

Another method that was sketched in [JLF06] is based on a spectral decomposition of the propagation matrix $\mathbf{P}^{\pm}(\omega,k)$. This spectral decomposition has two main advantages compared to the modified Newton method: first, its computational performance is better, and second, its results have a physical meaning as we will see later in Definition 6.16 and Remark 6.17. Even though it has not been proven that the propagation matrix $\mathbf{P}^{\pm}(\omega,k)$ is diagonalizable — in fact Hohage and Soussi [HS13] showed that the propagation operator $\mathcal{P}^{\pm}(\omega,k)$ of the TM mode is of Jordan type — we will use this spectral method because in practise it seems that the matrix is always diagonalizable. But also if this should not be the case, and the propagation matrix is of Jordan type, we can still use this spectral method in a generalized form by identifying the Jordan blocks and computing the Jordan chains. See [Fli09] for more details. Thus, we seek eigenvalues $\mu^{\pm}(\omega,k)\in\mathbb{C}$ with magnitude strictly less than one and their corresponding eigenvectors $\boldsymbol{\psi}^{\pm}(\omega,k)\in\mathbb{C}^{N(\Gamma_0^{\pm})}$ of the quadratic eigenvalue problem

$$\left[\mathbf{T}^{\pm}_{10}(\omega,k)\left(\mu^{\pm}(\omega,k)\right)^2 + \left(\mathbf{T}^{\pm}_{00}(\omega,k)+\mathbf{T}^{\pm}_{11}(\omega,k)\right)\mu^{\pm}(\omega,k) + \mathbf{T}^{\pm}_{01}(\omega,k)\right]\boldsymbol{\psi}^{\pm}(\omega,k) = 0, \tag{6.37}$$

which can be transformed into the generalized linear eigenvalue problem

$$\begin{pmatrix} -(\mathbf{T}_{00}^{\pm}(\omega,k)+\mathbf{T}_{11}^{\pm}(\omega,k)) & -\mathbf{T}_{01}^{\pm}(\omega,k) \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \boldsymbol{\Psi}^{\pm}(\omega,k) = \mu^{\pm}(\omega,k) \begin{pmatrix} \mathbf{T}_{10}^{\pm}(\omega,k) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \boldsymbol{\Psi}^{\pm}(\omega,k), \tag{6.38}$$

cf. [TM01], with

$$\boldsymbol{\Psi}^{\pm}(\omega,k) = \begin{pmatrix} \mu^{\pm}(\omega,k)\boldsymbol{\psi}^{\pm}(\omega,k) \\ \boldsymbol{\psi}^{\pm}(\omega,k) \end{pmatrix}. \tag{6.39}$$

Now let us come to an important symmetry property of the eigenvalues of the propagation matrix $\mathbf{P}^{\pm}(\omega,k)$. To this end, we first show

Lemma 6.14. *The matrices $\mathbf{T}_{ij}^{\pm}(\omega,k)$, $i,j=1,2$, corresponding to the linear operators $\mathcal{T}_{i,j}^{\pm}$, $i,j=0,1$, are Hermitian, i. e. they satisfy*

$$\mathbf{T}_{ij}^{\pm}(\omega,k)^{\mathrm{T}} = \overline{\mathbf{T}_{ji}^{\pm}(\omega,k)}, \qquad i,j=0,1. \tag{6.40}$$

Proof. The results directly follows from the definition (6.31) of the entries of $\mathbf{T}_{ij}^{\pm}(\omega,k)$, the relation (6.26) and the fact that the bilinear form $\mathfrak{b}_{C_1^{\pm}}$ satisfies $\overline{\mathfrak{b}_{C_1^{\pm}}(u,v;\omega,k)} = \mathfrak{b}_{C_1^{\pm}}(v,u;\omega,k)$. □

Using Lemma 6.14 it is easy to see that the quadratic eigenvalue problem (6.37) satisfies

Proposition 6.15. *If $\mu^{\pm}(\omega,k) \in \mathbb{C}\setminus\{0\}$ is an eigenvalue of (6.37), then $\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-1}$ is also an eigenvalue.*

Proof. Taking the complex conjugate of (6.37) and inserting (6.40) yields

$$\left[\mathbf{T}_{01}^{\pm}(\omega,k)\left(\overline{\mu^{\pm}(\omega,k)}\right)^2 + \left(\mathbf{T}_{00}^{\pm}(\omega,k)+\mathbf{T}_{11}^{\pm}(\omega,k)\right)\overline{\mu^{\pm}(\omega,k)} + \mathbf{T}_{10}^{\pm}(\omega,k)\right]^{\mathrm{T}} \overline{\boldsymbol{\psi}^{\pm}(\omega,k)} = 0.$$

Multiplying with $\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-2}$ and taking the transpose gives

$$\overline{\boldsymbol{\psi}^{\pm}(\omega,k)}^{\mathrm{T}} \left[\mathbf{T}_{10}^{\pm}(\omega,k)\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-2} + \left(\mathbf{T}_{00}^{\pm}(\omega,k)+\mathbf{T}_{11}^{\pm}(\omega,k)\right)\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-1} + \mathbf{T}_{01}^{\pm}(\omega,k)\right] = 0.$$

This implies that there exists a vector $\widetilde{\boldsymbol{\psi}}^{\pm}(\omega,k) \in \mathbb{C}^{N(\Gamma_0^{\pm})}$ such that

$$\left[\mathbf{T}_{10}^{\pm}(\omega,k)\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-2} + \left(\mathbf{T}_{00}^{\pm}(\omega,k)+\mathbf{T}_{11}^{\pm}(\omega,k)\right)\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-1} + \mathbf{T}_{01}^{\pm}(\omega,k)\right] \widetilde{\boldsymbol{\psi}}^{\pm}(\omega,k) = 0,$$

and hence, $\left(\overline{\mu^{\pm}(\omega,k)}\right)^{-1}$ is an eigenvalue of (6.37) with (right) eigenvector $\widetilde{\boldsymbol{\psi}}^{\pm}(\omega,k)$ and left eigenvector $\overline{\boldsymbol{\psi}^{\pm}(\omega,k)}^{\mathrm{T}}$. □

An advantage of the spectral decomposition, that also contributes to its better computational performance compared to the modified Newton method, is that we can directly determine whether ω^2 is inside the discrete approximation of the spectrum $\sigma^{\pm}(k)$.

Definition 6.16. *We call the set of numbers ω^2 for which the quadratic eigenvalue problem (6.37) has eigenvalues with magnitude one the* approximative spectrum $\sigma_h^{\pm}(k)$. *The approximative spectrum in an approximation to the spectrum of the operator $\mathcal{A}^{\pm}(k)$ related to the eigenvalue problem (2.19). Furthermore, define the* approximative essential spectrum $\sigma_h^{\mathrm{ess}}(k) := \sigma_h^{+}(k) \cup \sigma_h^{-}(k)$.

With the help of Proposition 6.15 and Definition 6.16 it is now clear how to compute the spectral decomposition of the propagation matrix $\mathbf{P}^{\pm}(\omega,k)$. We solve the general eigenvalue problem (6.38) for its $2N(\Gamma_0^{\pm})$ eigenvalues $\mu^{\pm}(\omega,k)$. If there exist eigenvalues with magnitude equal to one we stop our computation as we know from Definition 6.16 that this means that ω^2 is in the approximative essential spectrum $\sigma_h^{\mathrm{ess}}(k)$. Otherwise, and in accordance to Proposition 6.15, the $2N(\Gamma_0^{\pm})$ eigenvalues $\mu^{\pm}(\omega,k)$

split into $N(\Gamma_0^\pm)$ eigenvalues with magnitude strictly less than one and $N(\Gamma_0^\pm)$ eigenvalues with magnitude strictly larger than one. While discarding the $N(\Gamma_0^\pm)$ eigenvalues with magnitude strictly larger than one, the $N(\Gamma_0^\pm)$ eigenvalues $\mu^\pm(\omega,k)$ with magnitude strictly less than one and their corresponding eigenvectors $\boldsymbol{\psi}^\pm(\omega,k)$ form the spectral decomposition of the propagation matrix $\mathbf{P}^\pm(\omega,k)$.

Note at this point that we do not introduce a modelling error when we compute the propagation matrix since all $2N(\Gamma_0^\pm)$ eigenvalues of the general eigenvalue problem (6.38) are computed and taken into account. Thus, the only error that we expect is due to the choice of the discretization.

Remark 6.17. *Assuming that $\mathbf{P}^\pm(\omega,k)$ is diagonalizable, the eigenvectors of $\mathbf{P}^\pm(\omega,k)$ form a basis of the traces of the discretized evanescent PhC modes.*

Definition of the discrete Dirichlet-to-Neumann operators

Considering Eq. (6.9) for the characterization of the DtN operators $\mathcal{D}^\pm(\omega,k)$, we can define the discrete DtN operators

$$\mathcal{D}_h^\pm(\omega,k) = -\mathcal{T}_{00,h}^\pm(\omega,k) - \mathcal{T}_{10,h}^\pm(\omega,k)\,\mathcal{P}_h^\pm(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}_{1\mathrm{p}}^p(\Gamma_0^\pm))$$

with the discrete local DtN operators $\mathcal{T}_{ij,h}^\pm(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}_{1\mathrm{p}}^p(\Gamma_0^\pm))$, $i,j=0,1$, and the discrete propagation operators $\mathcal{P}_h^\pm(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}_{1\mathrm{p}}^p(\Gamma_0^\pm))$. Using the matrix representations of these discrete operators, we can compute DtN matrices $\mathbf{D}^\pm(\omega,k) \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$ with entries $D_{mn}^\pm(\omega,k)$, $m,n \in \{1,\dots,N(\Gamma_0^\pm)\}$, that satisfy

$$\mathcal{D}_h^\pm(\omega,k) b_{\Gamma_0^\pm,n} = \sum_{m=1}^{N(\Gamma_0^\pm)} D_{mn}^\pm(\omega,k) b_{\Gamma_0^\pm,m},$$

such that

$$\mathbf{D}^\pm(\omega,k) = -\mathbf{T}_{00}^\pm(\omega,k) - \mathbf{T}_{10}^\pm(\omega,k)\mathbf{P}^\pm(\omega,k), \tag{6.41}$$

cf. Eq. (6.9).

Derivatives of the discrete Dirichlet-to-Neumann operators

Let us define $\mathbf{T}_{ij}^{\pm,(m,n)} \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$, $m,n \in \mathbb{N}_0$, $i,j=0,1$, to be the matrices with entries

$$T_{ij,pq}^{\pm,(m,n)}(\omega,k) = \int_{\Gamma_0^\pm} \mathcal{T}_{ij}^{\pm,(m,n)}(\omega,k)\, b_{\Gamma_0^\pm,q}\, \overline{b}_{\Gamma_0^\pm,p}\, \mathrm{d}s(\mathbf{x}),$$

$p,q = 1,\dots,N(\Gamma_0^\pm)$. Directly taking the m-th ω- and n-th k-derivative of (6.34) in order to get expressions for $\mathbf{T}_{ij}^{\pm,(m,n)}$ is very involved since higher order derivatives of the inverse of $\mathbf{B}_{C_1^\pm}(\omega,k)$ can only be expressed in terms of Faà die Bruno's formula [FdB57]. Thus, we shall explicitly solve (6.33) for $\mathbf{U}_{i,0,h}^\pm(\omega,k) \in \mathbb{C}^{N_0(C_1^\pm)\times N(\Gamma_0^\pm)}$ and construct $\mathbf{U}_{i,h}^\pm(\omega,k) \in \mathbb{C}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$ as described above. Then we recursively solve

$$\begin{aligned}\mathbf{B}_{C_1^\pm}(C_1^\pm,C_1^\pm)\mathbf{U}_{i,0,h}^{\pm,(m',n')} = {}& \mathbf{M}_{C_1^\pm}^\beta(C_1^\pm,\cdot)\left(2m'\omega\mathbf{U}_{i,h}^{\pm,(m'-1,n')} + m'(m'-1)\mathbf{U}_{i,h}^{\pm,(m'-2,n')}\right)\\ &- \mathbf{M}_{C_1^\pm}^\alpha(C_1^\pm,\cdot)\left(2n'k\mathbf{U}_{i,h}^{\pm,(m',n'-1)} + n'(n'-1)\mathbf{U}_{i,h}^{\pm,(m',n'-2)}\right)\\ &- n'\mathbf{C}_{C_1^\pm}^\alpha(C_1^\pm,\cdot)\mathbf{U}_{i,h}^{\pm,(m',n'-1)},\end{aligned}$$

for $\mathbf{U}_{i,0,h}^{\pm,(m',n')}(\omega,k) \in \mathbb{C}^{N_0(C_1^\pm)\times N(\Gamma_0^\pm)}$ for all $m'=0,\dots,m$ and $n'=0,\dots,n$ with $m'+n'>0$, where we define $\mathbf{U}_{i,h}^{\pm,(0,0)}(\omega,k) := \mathbf{U}_{i,h}^\pm(\omega,k)$, and the matrices $\mathbf{U}_{i,h}^{\pm,(m',n')}(\omega,k) \in \mathbb{C}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$, with $m'+n'>0$, are obtained by setting their rows with indices $\mathfrak{S}(C_1^\pm,C_1^\pm)$ to $\mathbf{U}_{i,0,h}^{\pm,(m',n')}(\omega,k)$ and the remaining entries to zero.

Then the matrices $\mathbf{T}_{ij}^{\pm,(m,n)}(\omega,k)$, $m,n\in\mathbb{N}_0$, read

$$\begin{aligned}\mathbf{T}_{ij}^{\pm,(m,n)} &= (\mathbf{Q}^j_{C_1^\pm})^{\mathrm{T}}\mathbf{B}_{C_1^\pm}(\Gamma_j^\pm,\cdot)\mathbf{U}_{i,h}^{\pm,(m,n)}\\ &\quad-(\mathbf{Q}^j_{C_1^\pm})^{\mathrm{T}}\mathbf{M}^\beta_{C_1^\pm}(\Gamma_j^\pm,\cdot)\left(2m\omega\mathbf{U}_{i,h}^{\pm,(m-1,n)}+m(m-1)\mathbf{U}_{i,h}^{\pm,(m-2,n)}\right)\\ &\quad+(\mathbf{Q}^j_{C_1^\pm})^{\mathrm{T}}\mathbf{M}^\alpha_{C_1^\pm}(\Gamma_j^\pm,\cdot)\left(2nk\mathbf{U}_{i,h}^{\pm,(m,n-1)}+n(n-1)\mathbf{U}_{i,h}^{\pm,(m,n-2)}\right)\\ &\quad+n(\mathbf{Q}^j_{C_1^\pm})^{\mathrm{T}}\mathbf{C}^\alpha_{C_1^\pm}(\Gamma_j^\pm,\cdot)\mathbf{U}_{i,h}^{\pm,(m,n-1)}.\end{aligned}$$

The matrices $\mathbf{P}^{\pm,(m,n)}(\omega,k)\in\mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$, i. e. the discrete versions of the derivatives of the propagation operators, can be obtained when transferring the linear operator equation (6.21) into discrete form by replacing all operators with their corresponding matrices. The resulting linear matrix equation is of the form $\mathbf{AP}^{\pm,(m,n)}+\mathbf{BP}^{\pm,(m,n)}\mathbf{C}=\mathbf{D}$, with matrices $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}\in\mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$. It can be transformed into a linear system of equation with $(N(\Gamma_0^\pm))^2$ unknowns, i. e. the entries of $\mathbf{P}^{\pm,(m,n)}(\omega,k)$, cf. [Lan70].

Similarly, we find that the derivatives of the discrete DtN operator read

$$\mathbf{D}^{\pm,(m,n)}(\omega,k)=-\mathbf{T}_{00}^{\pm,(m,n)}(\omega,k)-\sum_{p=0}^{m}\sum_{q=0}^{n}\binom{m}{p}\binom{n}{q}\mathbf{T}_{10}^{\pm,(p,q)}(\omega,k)\mathbf{P}^{\pm,(m-p,n-q)}(\omega,k),\tag{6.42}$$

cf. Eq. (6.23).

6.2 Nonlinear eigenvalue problem with Dirichlet-to-Neumann operators

In the previous section we introduced DtN operators for periodic media, explained their computation and discretization. In this section we now want to show how to employ these operators in order to transform the linear (or quadratic) eigenvalue problem (2.19) on the unbounded domain S to a nonlinear eigenvalue problem posed in the defect cell C_0. We will start with the problem in strong formulation. After introducing a variational formulation, we will elaborate on the discretization of this nonlinear eigenvalue problem and finally, we present numerical solution techniques to solve the nonlinear eigenvalue problem in discretized form.

6.2.1 Main theorem

Now we state the main result of the DtN method.

Theorem 6.18. *[Theorem 4.5 in [Fli13]] Let the problems (6.1) in the semi-infinite strips $S^\pm$ be well-posed. Then the eigenvalue problem (2.19) is equivalent to: find couples $(\omega^2,k)\in\mathbb{R}^+\times B$, with $\omega^2\notin\sigma^{\mathrm{ess}}(k)$, such that there exists a non-trivial $u\in\mathsf{H}^1_{\mathrm{lp}}(\Delta,C_0,\alpha)$ that satisfies*

$$-(\nabla+\mathrm{i}k\tbinom{1}{0})\cdot\alpha(\nabla+\mathrm{i}k\tbinom{1}{0})u-\omega^2\beta u=0 \qquad \text{in } C_0,\tag{6.43a}$$

$$\pm\alpha\,\partial_2 u=\mathcal{D}^\pm(\omega,k)\,u \qquad \text{on } \Gamma_0^\pm.\tag{6.43b}$$

Note that the eigenvalue problem (6.43) — in comparison to problem (2.19) — is posed in the bounded domain C_0 but it is nonlinear with respect to ω and k. Furthermore, note that the characterization of the DtN operators $\mathcal{D}^\pm$ as described in Section 6.1.2 requires the local cell problems (6.8) to be well-posed as well.

6.2.2 Variational formulation

The weak formulation of the nonlinear eigenvalue problem (6.43) is: find couples $(\omega^2,k)\in\mathbb{R}^+\times B$ such that there exists a non-trivial $u\in\mathsf{H}^1_{\mathrm{lp}}(C_0)$ that satisfies

$$\mathfrak{b}_{C_0}(u,v;\omega,k)-\mathfrak{d}(u,v;\omega,k)=0\tag{6.44}$$

for all $v \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$, where the sesquilinear forms $\mathfrak{b}_{C_0}(\cdot,\cdot\,;\omega,k)$ and $\mathfrak{d}(\cdot,\cdot\,;\omega,k)$ are defined as

$$\mathfrak{b}_{C_0}(u,v;\omega,k) := \mathfrak{a}^{\alpha}_{C_0}(u,v) + k\mathfrak{c}^{\alpha,1}_{C_0}(u,v) + k^2\mathfrak{m}^{\alpha}_{C_0}(u,v) - \omega^2\mathfrak{m}^{\beta}_{C_0}(u,v) \tag{6.45a}$$

and

$$\mathfrak{d}(u,v;\omega,k) := \int_{\Gamma_0^+} \mathcal{D}^+(\omega,k)u\,\overline{v}\,\mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-} \mathcal{D}^-(\omega,k)u\,\overline{v}\,\mathrm{d}s(\mathbf{x}), \tag{6.45b}$$

with

$$\mathfrak{a}^{\alpha}_{C_0}(u,v) := \int_{C_0} \alpha\nabla u\cdot\nabla\overline{v}\,\mathrm{d}\mathbf{x}, \tag{6.45c}$$

$$\mathfrak{c}^{\alpha,1}_{C_0}(u,v) := \int_{C_0} \mathrm{i}\alpha\left(u(\partial_1\overline{v}) - (\partial_1 u)\overline{v}\right)\,\mathrm{d}\mathbf{x}, \tag{6.45d}$$

$$\mathfrak{m}^{\alpha}_{C_0}(u,v) := \int_{C_0} \alpha u\overline{v}\,\mathrm{d}\mathbf{x}, \tag{6.45e}$$

$$\mathfrak{m}^{\beta}_{C_0}(u,v) := \int_{C_0} \beta u\overline{v}\,\mathrm{d}\mathbf{x}. \tag{6.45f}$$

Remark 6.19. *Considering Eq. (6.3), it is easy to see that the nonlinear eigenvalue problem (6.44) is equivalent to: find couples $(\omega^2,k)\in\mathbb{R}^+\times B$ and a non-trivial $u\in\mathsf{H}^1_{1\mathrm{p}}(S)$ such that*

$$\int_S \alpha(\nabla + \mathrm{i}k\binom{1}{0})u\cdot(\nabla - \mathrm{i}k\binom{1}{0})\overline{v}\,\mathrm{d}\mathbf{x} - \omega^2\int_S \beta u\overline{v}\,\mathrm{d}\mathbf{x} = 0$$

for all $v\in\mathsf{H}^1_{1\mathrm{p}}(S)$, which is the variational formulation of the eigenvalue problem (2.19), that is linear in ω^2 and quadratic in k, but posed in the infinite strip S.

Before we will discuss the discretization of the nonlinear eigenvalue problem (6.44), we shall prove that the nonlinear eigenvalue problem is symmetric in the Brillouin zone B. This result is needed in Section 6.2.3 when deriving formulas for the derivatives of the dispersion curves. Let us first show some auxiliary results.

Lemma 6.20. *Let $k\in B$ and $\omega^2\in\mathbb{R}^+\setminus\sigma^{\mathrm{ess}}(k)$. Furthermore, let (6.1) be well-posed. Then*

$$\mathfrak{d}(u,v;\omega,k) = \mathfrak{d}(\overline{v},\overline{u};\omega,-k). \tag{6.46}$$

Proof. This is a direct consequence of the definition (6.45b) of the sesquilinear form $\mathfrak{d}$ and the weak formulation (6.3) of the DtN operators $\mathcal{D}^{\pm}$. □

Lemma 6.21. *Let $k\in B$ and $\omega^2\in\mathbb{R}^+\setminus\sigma^{\mathrm{ess}}(k)$. Furthermore, let (6.1) be well-posed. Then*

$$\overline{\mathfrak{d}(u,v;\omega,k)} = \mathfrak{d}(\overline{u},\overline{v};\omega,-k). \tag{6.47}$$

Proof. Using Lemma 6.20 and the fact that

$$\overline{\mathfrak{d}(u,v;\omega,k)} = \mathfrak{d}(v,u;\omega,\overline{k}) = \mathfrak{d}(v,u;\omega,k), \tag{6.48}$$

which follows from the definition (6.45b) of $\mathfrak{d}$, the weak formulation (6.3) of the DtN operators $\mathcal{D}^{\pm}$, and the fact that $\overline{k}=k$ if $k\in B\subset\mathbb{R}$, we can directly conclude Eq. (6.47). □

Lemma 6.22. *For any $(\omega^2,k)\in\mathbb{R}^+\times\mathbb{C}$*

$$\overline{\mathfrak{b}_{C_0}(u,v;\omega,k)} = \mathfrak{b}_{C_0}(\overline{u},\overline{v};\omega,-\overline{k}).$$

Proof. This follows directly from the definition (6.45a) of the sesquilinear form $\mathfrak{b}_{C_0}$. □

Now we are ready to prove

Proposition 6.23. *Let $(\omega^2, k) \in \mathbb{R}^+ \times B$ with $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$ be an eigenvalue couple of the nonlinear eigenvalue problem (6.44) with associated eigenfunction $u \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$. Then $(\omega^2, -k) \in \mathbb{R}^+ \times B$ is an eigenvalue couple of (6.44) with associated eigenfunction $\overline{u} \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$.*

Proof. If $(\omega^2, k) \in \mathbb{R}^+ \times B$ with $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$ is an eigenvalue couple of (6.44) with associated eigenfunction $u \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$, then

$$\overline{\mathfrak{b}_{C_0}(u, v; \omega, k)} - \overline{\mathfrak{d}(u, v; \omega, k)} = \overline{\mathfrak{b}_{C_0}(u, v; \omega, k) - \mathfrak{d}(u, v; \omega, k)} = 0$$

for all $v \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$. Using Lemmas 6.21 and 6.22 as well as the fact that $\overline{k} = k$ if $k \in B \subset \mathbb{R}$, we obtain

$$\mathfrak{b}_{C_0}(\overline{u}, \overline{v}; \omega, -k) - \mathfrak{d}(\overline{u}, \overline{v}; \omega, -k) = 0$$

for all $v \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$, from which the result directly follows. □

6.2.3 Group velocity and higher derivatives of dispersion curves

In this section we transform the procedure to derive formulas for the derivatives of dispersion curves, as presented in Chapter 4 for linear eigenvalue problems, to nonlinear eigenvalue problems with DtN transparent boundary conditions.

We consider the ω-formulation of the variational formulation with DtN transparent boundary conditions as shown in (6.44), i. e. for $k \in B$ we search for eigenvalues $\omega_j^2(k) \in \mathbb{R}^+ \setminus \sigma^{\mathrm{ess}}(k)$ and corresponding non-trivial eigenmodes $u_j(k) \equiv u_j(\cdot\,; k) \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$ such that

$$\mathfrak{b}_{C_0}(u_j(k), v; \omega_j(k), k) - \mathfrak{d}(u_j(k), v; \omega_j(k), k) = 0 \tag{6.49}$$

for all $v \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$, where the sesquilinear forms are defined in (6.45).

In Remark 6.19 we argued that the eigenvalue problem (6.49) with DtN operators is equivalent to the variational formulation of the eigenvalue problem (2.19) in the infinite strip S, for which we showed already in Chapter 4 that the eigenvalues $\omega_j^2(k)$ and their corresponding eigenmodes $u_j(k)$ are analytic with respect to the quasi-momentum k. Hence, analyticity of the eigenvalues and eigenmodes of (6.49) follows directly from this equivalence.

In Section 6.1.3, we showed that the DtN operators $\mathcal{D}^{\pm}$ are differentiable to any order with respect to ω and k inside the band gaps, i. e. $\frac{\partial^{m+n}}{\partial k^m \partial \omega^n}\mathcal{D}^{\pm}(\omega, k)$ are well-defined for any $m, n \in \mathbb{N}$ and can be computed using a set of local cell problems. We shall now use these derivatives for deriving formulas for the group velocity and all higher derivatives of the dispersion relation when prescribing DtN transparent boundary conditions at the top and bottom boundaries Γ_0^+ and Γ_0^-.

Let us start with the first derivative of the dispersion curve, the so-called group velocity. Differentiating (6.49) with respect to k gives

$$\mathfrak{b}_{C_0}(\mathrm{d}_k u_j(k), v; \omega_j(k), k) - \mathfrak{d}(\mathrm{d}_k u_j(k), v; \omega_j(k), k) = \mathfrak{f}^{(1)}_{\mathrm{DtN}}(v) \tag{6.50}$$

for all $v \in \mathsf{H}^1_{\mathrm{1p}}(C_0)$ with the linear form

$$\mathfrak{f}^{(1)}_{\mathrm{DtN}}(v; k, \omega_j, \omega_j', u_j) = \mathfrak{f}^{(1)}(v; k, \omega_j, \omega_j', u_j) + \omega_j' \mathfrak{d}_\omega(u_j, v; \omega, k) + \mathfrak{d}_k(u_j, v; \omega, k),$$

where $\mathfrak{f}^{(1)}$ was already defined in Eq. (4.4) and reads

$$\mathfrak{f}^{(1)}(v; k, \omega_j, \omega_j', u_j) = -2k\mathfrak{m}^{\alpha}_C(u_j, v) - \mathfrak{c}^{\alpha,1}_C(u_j, v) + 2\omega_j\omega_j'\mathfrak{m}^{\beta}_C(u_j, v),$$

and the sesquilinear forms $\mathfrak{d}_\omega$ and $\mathfrak{d}_k$ are defined as

$$\mathfrak{d}_\omega(u, v; \omega, k) = \int_{\Gamma_0^+} \mathcal{D}^+_\omega(\omega, k)\, u\, \overline{v}\, \mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-} \mathcal{D}^-_\omega(\omega, k)\, u\, \overline{v}\, \mathrm{d}s(\mathbf{x}),$$
$$\mathfrak{d}_k(u, v; \omega, k) = \int_{\Gamma_0^+} \mathcal{D}^+_k(\omega, k)\, u\, \overline{v}\, \mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-} \mathcal{D}^-_k(\omega, k)\, u\, \overline{v}\, \mathrm{d}s(\mathbf{x})$$

with the derivatives $\mathcal{D}_\omega^\pm$ and $\mathcal{D}_k^\pm$ of the DtN operators defined in Eq. (6.12). Due to Proposition 6.23 we can test Eq. (6.50) with $v = u$ and find

$$\mathfrak{f}_{\mathrm{DtN}}^{(1)}(u) = 0,$$

which yields the group velocity

$$\omega_j'(k) = \frac{2k\mathfrak{m}_C^\alpha(u_j,u_j) + \mathfrak{c}_C^{\alpha,1}(u_j,u_j) - \mathfrak{d}_k(u_j,u_j;\omega_j,k)}{2\omega_j\mathfrak{m}_{C_0}^\beta(u_j,u_j) + \mathfrak{d}_\omega(u_j,u_j;\omega_j,k)}. \tag{6.51}$$

In comparison to the formula (4.5) of the group velocity for problems with periodic boundary conditions, Eq. (6.51) only has one additional term in the numerator and one additional term in the denominator that are both related to the DtN transparent boundary conditions on $\Gamma_0^\pm$.

In order to extend the procedure to higher order derivatives, recall the short notations for the derivatives of dispersion curves and eigenmodes, that we introduced in Chapter 4, i. e.

$$\omega_j^{(n)}(k) := \frac{\partial^n\omega_j(k)}{\partial k^n} \qquad \text{and} \qquad \mathrm{d}_k^n u_j(\cdot\,;k) := \frac{\mathrm{d}^n u_j(\cdot\,;k)}{\mathrm{d}k^n},$$

$n \in \mathbb{N}_0$. Then the n-th derivative of (6.49) with respect to k reads

$$\frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{b}_{C_0}(u_j(k),v;\omega_j(k),k) - \frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{d}(u_j(k),v;\omega_j(k),k) = 0. \tag{6.52}$$

The first term is equivalent to the n-th derivative of Eq. (4.1) in Section 4.2, i. e.

$$\frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{b}_{C_0}(u_j(k),v;\omega_j(k),k) = \mathfrak{b}_{C_0}(\mathrm{d}_k^n u_j(k),v;\omega_j(k),k) - \mathfrak{f}^{(n)}(v),$$

where the linear form $\mathfrak{f}^{(n)} = \mathfrak{f}^{(n)}(\cdot\,;k,\omega_j^{(0)},\dots,\omega_j^{(n)},u_j^{(0)},\dots,u_j^{(n-1)})$ reads

$$\begin{aligned}\mathfrak{f}^{(n)}(v) = \sum_{p=0}^{n-1}\sum_{q=0}^{n-p}\frac{n!}{p!\,q!\,(n-p-q)!}\omega_j^{(n-p-q)}\omega_j^{(q)}\mathfrak{m}_{C_0}^\beta(u_j^{(p)},v)\\ - n\,\mathfrak{c}_{C_0}^{\alpha,1}(u_j^{(n-1)},v) - 2n\,k\,\mathfrak{m}_{C_0}^\alpha(u_j^{(n-1)},v) - n(n-1)\,\mathfrak{m}_{C_0}^\alpha(u_j^{(n-2)},v),\end{aligned}$$

cf. Eq. (4.8), with the auxiliary functions $u_j^{(m)}(k)$, $1 \le m \le n-1$, associated to the eigenmode derivatives $\mathrm{d}_k^m u_j(k)$ and with $u_j^{(0)}(k) = u_j(k)$. As elaborated in Section 4.2.2 it is sufficient to compute the auxiliary functions $u_j^{(m)}(k)$ instead of properly defining and computing the derivatives $\mathrm{d}_k^m u_j(k)$.

The evaluation of the second term in (6.52), however, is more involved. Recall that we employed a multivariant version [CS96] of Faà di Bruno's formula in Chapter 5 to express the n-th total derivative of a matrix-valued function, cf. Eq. (5.15). As an alternative way to evaluate the n-th total derivative, we proposed the recursive Algorithm 5.1. Let us focus on the latter possibility in this section. Introducing the sesquilinear forms

$$\begin{aligned}\mathfrak{d}^{(n)}(u,v;\omega,k) &= \int_{\Gamma_0^+}\frac{\mathrm{d}^n\mathcal{D}^+(\omega,k)}{\mathrm{d}k^n}\,u\,\overline{v}\,\mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-}\frac{\mathrm{d}^n\mathcal{D}^-(\omega,k)}{\mathrm{d}k^n}\,u\,\overline{v}\,\mathrm{d}s(\mathbf{x}),\\ \mathfrak{d}_\omega^{(n)}(u,v;\omega,k) &= \int_{\Gamma_0^+}\frac{\mathrm{d}^n\mathcal{D}_\omega^+(\omega,k)}{\mathrm{d}k^n}\,u\,\overline{v}\,\mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-}\frac{\mathrm{d}^n\mathcal{D}_\omega^-(\omega,k)}{\mathrm{d}k^n}\,u\,\overline{v}\,\mathrm{d}s(\mathbf{x}),\\ \mathfrak{d}_k^{(n)}(u,v;\omega,k) &= \int_{\Gamma_0^+}\frac{\mathrm{d}^n\mathcal{D}_k^+(\omega,k)}{\mathrm{d}k^n}\,u\,\overline{v}\,\mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-}\frac{\mathrm{d}^n\mathcal{D}_k^-(\omega,k)}{\mathrm{d}k^n}\,u\,\overline{v}\,\mathrm{d}s(\mathbf{x}),\end{aligned}$$

$n \in \mathbb{N}_0$, we expand

$$\frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{d}(u_j(k),v;\omega_j(k),k) = \sum_{m=0}^{n}\binom{n}{m}\mathfrak{d}^{(n)}(\mathrm{d}_k^{n-m}u_j(k),v;\omega_j(k),k),$$

and write

$$\begin{aligned}\mathfrak{d}^{(n)}(\cdot\,,\cdot\,;\omega_j(k),k) &= \frac{\mathrm{d}^{n-1}}{\mathrm{d}k^{n-1}}\Big(\omega_j'\mathfrak{d}_\omega^{(0)}(\cdot\,,\cdot\,;\omega_j(k),k) + \mathfrak{d}_k^{(0)}(\cdot\,,\cdot\,;\omega_j(k),k)\Big)\\ &= \sum_{m=0}^{n-1}\binom{n-1}{m}\omega_j^{(m+1)}\mathfrak{d}_\omega^{(n-m-1)}(\cdot\,,\cdot\,;\omega_j(k),k) + \mathfrak{d}_k^{(n-1)}(\cdot\,,\cdot\,;\omega_j(k),k),\end{aligned} \tag{6.53}$$

which is the analogue of the recursion formula (5.19) in Chapter 5 that motivated Algorithm 5.1. Then we can rewrite (6.54) and find that the n-th derivative $\mathrm{d}_k^n u(k) \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$ of the eigenmode $u_j(k)$ corresponding to $\omega_j(k)$ satisfies

$$\mathfrak{b}_{C_0}(\mathrm{d}_k^n u_j, v; \omega_j, k) - \mathfrak{d}(\mathrm{d}_k^n u_j, v; \omega_j, k) = \mathfrak{f}_{\mathrm{DtN}}^{(n)}(v) \tag{6.54}$$

for all $v \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$, with the linear form

$$\mathfrak{f}_{\mathrm{DtN}}^{(n)}(v) = \mathfrak{f}^{(n)}(v) + \sum_{m=1}^{n} \binom{n}{m} \mathfrak{d}^{(m)}(u_j^{(n-m)}, v; \omega_j, k),$$

where we replaced the derivatives $\mathrm{d}_k^m u_j$, $1 \leq m \leq n-1$, of the eigenmode u_j by the auxiliary functions $u_j^{(m)}$ using the convention $u_j^{(0)} = u_j$. The total derivatives $\mathfrak{d}^{(m)}(\cdot\,,\cdot\,;\omega_j(k),k)$, $1 \leq m \leq n-1$, of $\mathfrak{d}(\cdot\,,\cdot\,;\omega_j(k),k)$ with respect to k can be evaluated recursively using (6.53) as sketched in Algorithm 5.1. Testing Eq. (6.54) with $v = u$ and considering Proposition 6.23 yields

$$\mathfrak{f}_{\mathrm{DtN}}^{(n)}(u) = 0,$$

from which — together with (6.53) — we obtain the n-th derivative of the dispersion relation

$$\begin{aligned}
\omega_j^{(n)}(k) = \Big(2\omega_j \mathfrak{m}_{C_0}^{\beta}(u_j,u_j) + \mathfrak{d}_\omega(u_j,u_j;\omega_j,k)\Big)^{-1} \\
\cdot \Bigg[n(n-1)\mathfrak{m}_{C_0}^{\alpha}(u_j^{(n-2)},u_j) + 2n\,k\,\mathfrak{m}_{C_0}^{\alpha}(u_j^{(n-1)},u_j) + n\,\mathfrak{c}_{C_0}^{\alpha,1}(u_j^{(n-1)},u_j) \\
- \sum_{p=1}^{n-1}\sum_{q=0}^{n-p} \frac{n!}{p!q!(n-p-q)!}\,\omega_j^{(n-p-q)}\omega_j^{(q)}\mathfrak{m}_{C_0}^{\beta}(u_j^{(p)},u_j) \\
- \sum_{q=1}^{n-1} \binom{n}{q} \omega_j^{(n-q)}\omega_j^{(q)}\mathfrak{m}_{C_0}^{\beta}(u_j,u_j) \\
- \sum_{q=1}^{n-1} \binom{n}{q} \mathfrak{d}^{(q)}(u_j^{(n-q)},u_j;\omega_j,k) \\
- \sum_{q=1}^{n-1} \binom{n-1}{q-1} \omega_j^{(q)}\mathfrak{d}_\omega^{(n-q)}(u_j,u_j;\omega_j,k) \\
- \mathfrak{d}_k^{(n-1)}(u_j,u_j;\omega_j,k) \Bigg].
\end{aligned} \tag{6.55}$$

Analogously to the argumentation in Section 4.2, we note that (6.54) is ill-posed. However, by additionally requiring $\mathsf{H}^1(C_0)$-orthogonality to all linearly independent eigenmodes $u_{j,1}(\cdot;k), \ldots, u_{j,m}(\cdot;k) \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$ we can compute a particular solution $u_j^{(n)}(\cdot;k) \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$ of Eq. (6.54). Again — for simplicity — let us assume that there exists only one linearly independent eigenmode $u_j(\cdot;k)$ corresponding to the eigenvalue $\omega_j^2(k)$. Then we seek the auxiliary function $u_j^{(n)}(\cdot;k) \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$ and the Lagrange multiplier $\lambda \in \mathbb{C}$ such that

$$\begin{aligned}
\mathfrak{b}_{C_0}(u_j^{(n)},v;\omega_j,k) - \mathfrak{d}(u_j^{(n)},v;\omega_j,k) + \lambda\langle u_j,v\rangle_{\mathsf{H}^1(C_0)} &= \mathfrak{f}_{\mathrm{DtN}}^{(n)}(v), \\
\langle u_j^{(n)},u_j\rangle_{\mathsf{H}^1(C_0)} &= 0,
\end{aligned}$$

for all $v \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$.

The formula (6.55) is very technical and looks complicated. However, recall that we sketched in Algorithm 5.1 a scheme to compute the total derivatives

$$\begin{aligned}
\mathfrak{d}^{(n)}(\cdot\,,\cdot\,;\omega_j(k),k) &= \frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{d}(\cdot\,,\cdot\,;\omega_j(k),k), \\
\mathfrak{d}_\omega^{(n)}(\cdot\,,\cdot\,;\omega_j(k),k) &= \frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{d}_\omega(\cdot\,,\cdot\,;\omega_j(k),k)
\end{aligned}$$

and

$$\mathfrak{d}_k^{(n)}(\cdot\,,\cdot\,;\omega_j(k),k) = \frac{\mathrm{d}^n}{\mathrm{d}k^n}\mathfrak{d}_k(\cdot\,,\cdot\,;\omega_j(k),k)$$

for all $n \in \mathbb{N}$. Alternatively, the derivatives can be evaluated using a multivariant version of Faà di Bruno's formula for which the reader is referred to Chapter 5. With these derivatives at hand, Eq. (6.55) is only slightly more complicated than the formula (4.9) for the n-th derivative of the dispersion curves in the case with periodic boundary conditions.

6.2.4 Discretization

For the FE discretization of the nonlinear eigenvalue problem in variational formulation (6.44) we need the FE spaces $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ and $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$, that were already introduced in Section 6.1.5. Recall that the basis functions $b_{C_0,n}$, $n = 1,\dots,N(C_0)$, of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ are ordered such that

- the basis functions with index $n \in \mathfrak{S}(C_0,\Gamma_0^+) = \{1,\dots,N(\Gamma_0^+)\}$ vanish on Γ_0^-, but their traces on Γ_0^+ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^+)$,
- the basis functions with index $n \in \mathfrak{S}(C_0,\Gamma_0^-) = \{N(\Gamma_0^+)+1,\dots,N(\Gamma_0^+)+N(\Gamma_0^-)\}$ vanish on Γ_0^+, but their traces on Γ_0^- build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^-)$, and
- the basis functions with index $n \in \mathfrak{S}(C_0,C_0) = \{N(\Gamma_0^+)+N(\Gamma_0^-)+1,\dots,N(C_0)$ vanish on $\Gamma_0^\pm$.

Thus, the traces of the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ on $\Gamma_0^\pm$ and the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ satisfy

$$\begin{aligned} b_{\Gamma_0^+,n} &= \sum_{m=1}^{N(\Gamma_0^+)} Q^+_{C_0,mn} b_{C_0,m}|_{\Gamma_0^+}, \\ b_{\Gamma_0^-,n} &= \sum_{m=1}^{N(\Gamma_0^-)} Q^-_{C_0,mn} b_{C_0,N(\Gamma_0^+)+m}|_{\Gamma_0^-}, \end{aligned}$$

with permutation matrices $\mathbf{Q}^+_{C_0} \in \mathbb{R}^{N(\Gamma_0^+)\times N(\Gamma_0^+)}$ and $\mathbf{Q}^-_{C_0} \in \mathbb{R}^{N(\Gamma_0^-)\times N(\Gamma_0^-)}$, cf. Eq. (6.28).

This relation is important for deriving the discrete form of the variational formulation (6.44). The sesquilinear form $\mathfrak{d}$ is related to the solution in the two semi-infinite strips which is represented by the DtN maps $\mathcal{D}^\pm(\omega,k)$. When inserting the basis functions $b_{\Gamma_0^\pm,n}$ of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ into each of the two integrals in $\mathfrak{d}$ and using the characterization of the DtN operators (6.9) we obtain the matrices as presented in Eq. (6.41), i. e.

$$\mathbf{D}^\pm(\omega,k) = -\mathbf{T}^\pm_{00}(\omega,k) - \mathbf{T}^\pm_{10}(\omega,k)\mathbf{P}^\pm(\omega,k) \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}.$$

However, stating the variational formulation (6.44) in $\mathsf{S}^p_{1\mathrm{p}}(C_0) \subset \mathsf{H}^1_{1\mathrm{p}}(C_0)$ we have to insert in $\mathfrak{d}$ rather the traces of the basis functions $b_{C_0,n}$ on $\Gamma_0^\pm$. We introduce the matrix

$$\mathbf{B}_{C_0}(\omega,k) = \mathbf{A}^\alpha_{C_0} + k\mathbf{C}^{\alpha,1}_{C_0} + k^2\mathbf{M}^\alpha_{C_0} - \omega^2\mathbf{M}^\beta_{C_0} \in \mathbb{C}^{N(C_0)\times N(C_0)} \tag{6.56}$$

where the matrices $\mathbf{A}^\alpha_{C_0}, \mathbf{C}^{\alpha,1}_{C_0}, \mathbf{M}^\alpha_{C_0}, \mathbf{M}^\beta_{C_0} \in \mathbb{R}^{N(C_0)\times N(C_0)}$ have entries

$$A^\alpha_{C_0,mn} = \mathfrak{a}^\alpha_{C_0}(b_{C_0,n}, b_{C_0,m}), \tag{6.57a}$$
$$C^{\alpha,1}_{C_0,mn} = \mathfrak{c}^{\alpha,1}_{C_0}(b_{C_0,n}, b_{C_0,m}), \tag{6.57b}$$
$$M^\alpha_{C_0,mn} = \mathfrak{m}^\alpha_{C_0}(b_{C_0,n}, b_{C_0,m}), \tag{6.57c}$$
$$M^\beta_{C_0,mn} = \mathfrak{m}^\beta_{C_0}(b_{C_0,n}, b_{C_0,m}), \tag{6.57d}$$

$m,n = 1,\dots,N(C_0)$, with the sesquilinear forms as given in Eq. (6.45). Then the discrete form of the nonlinear eigenvalue problem (6.44) reads

$$\left(\mathbf{B}_{C_0}(\omega,k) - \mathbf{D}_{C_0}(\omega,k)\right)\mathbf{u}(\omega,k) = \mathbf{0} \tag{6.58}$$

where $\mathbf{u}(\omega,k) \in \mathbb{C}^{N(C_0)}$ is the coefficient vector of the discrete eigenmode $u_h(\cdot\,;\omega,k) \in \mathsf{S}^p_{1\mathrm{p}}(C_0)$, and $\mathbf{D}(\omega,k) \in \mathbb{C}^{N(C_0)\times N(C_0)}$ is a block matrix of the form

$$\mathbf{D}_{C_0}(\omega,k) = \begin{pmatrix} (\mathbf{Q}^+_{C_0})^{\mathrm{T}}\mathbf{D}^+(\omega,k)\mathbf{Q}^+_{C_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{Q}^-_{C_0})^{\mathrm{T}}\mathbf{D}^-(\omega,k)\mathbf{Q}^-_{C_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

An important issue of the discretization of eigenvalue problems is its stability, i. e. the existence of a minimal dimension of the FE space, such that the standard asymptotic convergence estimates hold for any dimension larger than this threshold. To the best of our knowledge, this issue has not yet been solved for the specific nonlinear eigenvalue problem (6.44). However, numerical evidence shows that the standard asymptotic convergence estimates hold true.

Thus, we can use p-FEM on a coarse grid such as the one sketched in Figure 6.1 for the computation of guided modes in PhC waveguides with smooth material boundaries and can expect exponential convergence.

6.2.5 Numerical solution of the nonlinear eigenvalue problem

Since the DtN operators are differentiable with respect to both, the frequency ω as well as the quasi-momentum k to any order, the nonlinear, matrix-valued function

$$\mathbf{N}_{C_0} : (\omega,k) \longmapsto \mathbf{B}_{C_0}(\omega,k) - \mathbf{D}_{C_0}(\omega,k), \tag{6.59}$$

is holomorphic in ω and k as long as $\omega^2 \notin \sigma_h^{\mathrm{ess}}(k)$ and ω^2 is not a global or local Dirichlet eigenvalue. Hence, we can apply all methods introduced in Section 3.2 to solve the nonlinear eigenvalue problem (6.58) in both formulations, the ω-formulation, where we fix the quasi-momentum and look for frequency eigenvalues, and the k-formulation, where we fix the frequency and search for eigenvalues of the quasi-momentum. In particular, we will employ the method of successive linear problems (MSLP) and the Chebyshev interpolation. While the former is an iterative scheme to compute a single eigenvalue, the latter is a representative of direct methods, that allow for a simultaneous computation of several eigenvalues. The Chebyshev interpolation is a very elegant procedure to solve the nonlinear eigenvalue problem (6.58). However, it comes with the drawback that one needs to be sure that the Chebyshev nodes (in particular the two endpoints) are sufficiently far away from the essential spectrum $\sigma^{\mathrm{ess}}(k)$. This implies that one needs to have a priori knowledge of the spectra $\sigma^{\pm}(k)$ of the operators $\mathcal{A}^{\pm}(k)$ related to the PhCs on top and bottom of the guide. This is similar to the supercell method, where one needs at least a posteriori knowledge of the essential spectrum $\sigma^{\mathrm{ess}}(k)$ to exclude spurious modes.

The Newton-type method, that we proposed in Section 3.3 for eigenvalue problems like (6.58) in ω-formulation, is an alternative to the techniques mentioned above. In contrast to the presentation of the Newton method in Section 3.3, the matrix $\mathbf{N}_{C_0}$ is a function of two parameters, the frequency ω and the quasi-momentum k. Thus, two different versions of this algorithm are possible, i. e. the ω-formulation and the k-formulation. To this end, we shall recall the methodology of Section 3.3, explicitly presenting the algorithms for the problem under consideration.

As a first step towards the Newton-type procedure to solve the nonlinear eigenvalue problem (6.58) we introduce a "simplified" eigenvalue problem with fixed DtN operators. Let $(\omega^2_{\mathcal{D}}, k_{\mathcal{D}}) \in \mathbb{R}^+ \times B$, with $\omega^2_{\mathcal{D}} \notin \sigma_h^{\mathrm{ess}}(k_{\mathcal{D}})$, be arbitrary but fixed. Then the problem: find $\omega^2 = \omega^2(\omega_{\mathcal{D}}, k_{\mathcal{D}}) \in \mathbb{R}^+$ and a non-trivial $\mathbf{u} \in \mathbb{C}^{N(C_0)} \setminus \{\mathbf{0}\}$ such that

$$\left(\mathbf{A}^{\alpha}_{C_0} + k_{\mathcal{D}}\mathbf{C}^{\alpha,1}_{C_0} + k^2_{\mathcal{D}}\mathbf{M}^{\alpha}_{C_0} - \omega^2\mathbf{M}^{\beta}_{C_0} - \mathbf{D}_{C_0}(\omega_{\mathcal{D}}, k_{\mathcal{D}})\right)\mathbf{u} = \mathbf{0} \tag{6.60}$$

is a linear eigenvalue problem, whose solution coincides with the one of (6.58) if $\omega^2 = \omega^2_{\mathcal{D}}$. For this linear eigenvalue problem we state the following important results.

Proposition 6.24. *Let $(\omega^2_{\mathcal{D}}, k_{\mathcal{D}}) \in \mathbb{R}^+ \times B$ with $\omega^2_{\mathcal{D}} \notin \sigma_h^{\mathrm{ess}}(k_{\mathcal{D}})$. Then the eigenvalues $\omega^2_j(\omega_{\mathcal{D}}, k_{\mathcal{D}})$, $1 \le j \le N(C_0)$, of the linear eigenvalue problem (6.60) are real.*

Proof. This result, which is a discrete corollary of Proposition 4.8 in [Fli13], follows directly from the fact that the matrices $\mathbf{M}_{C_0}^{\beta}$ and $\left(\mathbf{A}_{C_0}^{\alpha} + k_{\mathcal{D}}\mathbf{C}_{C_0}^{\alpha,1} + k_{\mathcal{D}}^2\mathbf{M}_{C_0}^{\alpha} - \mathbf{D}_{C_0}(\omega_{\mathcal{D}}, k_{\mathcal{D}})\right)$ are self-adjoint for all $(\omega_{\mathcal{D}}^2, k_{\mathcal{D}}) \in \mathbb{R}^+ \times B$ with $\omega_{\mathcal{D}}^2 \notin \sigma_h^{\text{ess}}(k_{\mathcal{D}})$. □

Since the DtN operators are differentiable with respect to the frequency and the quasi-momentum so is the matrix $\mathbf{D}_{C_0}$. Hence, we can apply the perturbation theory for linear, self-adjoint operators in finite-dimensional spaces, see Chapter 2 in [Kat95], and deduce

Proposition 6.25. *Let $(\omega_{\mathcal{D}}^2, k_{\mathcal{D}}) \in \mathbb{R}^+ \times B$ with $\omega_{\mathcal{D}}^2 \notin \sigma_h^{\text{ess}}(k_{\mathcal{D}})$. Then the eigenvalues $\omega_j^2(\omega_{\mathcal{D}}, k_{\mathcal{D}})$, $1 \leq j \leq N(C_0)$, of the linear eigenvalue problem (6.60) can be ordered such that the functions*

$$(\omega_{\mathcal{D}}, k_{\mathcal{D}}) \longmapsto \omega_j^2(\omega_{\mathcal{D}}, k_{\mathcal{D}})$$

and the corresponding eigenvectors $\mathbf{u}_j(\omega_{\mathcal{D}}, k_{\mathcal{D}})$ are continuously differentiable with respect to $\omega_{\mathcal{D}}$ and $k_{\mathcal{D}}$.

Thanks to Proposition 6.25 we can introduce differentiable *signed distance functions*

$$d_j(\omega_{\mathcal{D}}, k_{\mathcal{D}}) = \omega_{\mathcal{D}}^2 - \omega_j^2(\omega_{\mathcal{D}}, k_{\mathcal{D}}) \in \mathbb{R}. \tag{6.61}$$

Due to the differentiability of the signed distance functions with respect to $\omega_{\mathcal{D}}$ and $k_{\mathcal{D}}$ we can apply Newton's method to compute its roots. As elaborated above, these roots are then also the eigenvalues of the nonlinear eigenvalue problem (6.58).

Let us now introduce the *global signed distance function*

$$d^{\text{g}}(\omega_{\mathcal{D}}, k_{\mathcal{D}}) = d_{j^*}(\omega_{\mathcal{D}}, k_{\mathcal{D}}) \tag{6.62}$$

where

$$j^* = j^*(\omega_{\mathcal{D}}, k_{\mathcal{D}}) = \arg \min_{1 \leq j \leq N(C_0)} |d_j(\omega_{\mathcal{D}}, k_{\mathcal{D}})|.$$

As shown in the numerical results in Section 6.3 this function is not continuous due to sign changes, and hence, not differentiable, however, we shall also see in Section 6.3 that the numerical results of the Newton method applied to the global signed distance function d^{g} are reasonable when using the derivatives of the continuously differentiable signed distance function d_{j^*} with respect to $\omega_{\mathcal{D}}$ or $k_{\mathcal{D}}$, respectively. Applying the Newton method to the differentiable signed distance functions d_j, on the other hand, is not possible, since the ordering of the eigenvalues, such that $\omega_j^2(\omega_{\mathcal{D}}, k_{\mathcal{D}})$ are differentiable is not known in advance. Proposition 6.25 only guarantees that there exists an ordering but it does not say anything about how to find it. For this one can, for example, apply the adaptive path following of dispersion curves as proposed in Chapter 5. However, this implies a huge computational overhead compared to simply applying the Newton method to the global signed distance function.

Algorithm 6.1. Newton's method applied to global signed distance function in ω-formulation.

1: Fix $k_{\mathcal{D}} \in B$ and choose start value $\omega^{(0)} \in \mathbb{R}^+$.
2: **for** $n = 0, \ldots$ **do**
3: **if** $(\omega^{(n)})^2 \in \sigma_h^{\text{ess}}(k_{\mathcal{D}})$ **then**
4: **exit** (and restart with new start value $\omega^{(0)} \in \mathbb{R}^+$)
5: **end if**
6: Solve linear eigenvalue problem (6.60) for ω^2 with $\omega_{\mathcal{D}} = \omega^{(n)}$.
7: Evaluate global signed distance function $d^{\text{g}}(\omega^{(n)}, k_{\mathcal{D}})$.
8: **if** $d^{\text{g}}(\omega^{(n)}, k_{\mathcal{D}}) \approx 0$ **then**
9: **exit**
10: **end if**
11: Compute new value $\omega^{(n+1)} = \omega^{(n)} - \left(\frac{\partial}{\partial \omega_{\mathcal{D}}} d_{j^*}(\omega^{(n)}, k_{\mathcal{D}})\right)^{-1} d^{\text{g}}(\omega^{(n)}, k_{\mathcal{D}})$.
12: **end for**

The iterative scheme in ω-formulation, i. e. keeping $k_{\mathcal{D}} \in B$ fixed and searching for a root $\omega \in \mathbb{R}^+$ of $d^{\text{g}}(\cdot, k_{\mathcal{D}})$, then works as shown in Algorithm 6.1, where the derivative of d_{j^*} with respect to $\omega_{\mathcal{D}}$ can

either be approximated by a difference quotient, or computed with the help of a closed formula, that can be derived using the facts that the DtN operators are differentiable with respect to the frequency, cf. Proposition 6.9, and that the eigenvalues $\omega_j^2(\omega_\mathcal{D}, k_\mathcal{D})$ of the linear eigenvalue problem (6.60) and their corresponding eigenvectors $\mathbf{u}_j(\omega_\mathcal{D}, k_\mathcal{D})$ are continuously differentiable with respect to $\omega_\mathcal{D}$, cf. Proposition 6.25. Then we can proceed as in Section 3.3 and obtain

$$\frac{\partial \omega_j(\omega_\mathcal{D}, k_\mathcal{D})}{\partial \omega_\mathcal{D}} = -\frac{\mathbf{u}_j^{\mathrm{H}} \partial_\omega \mathbf{D}_{C_0}(\omega_\mathcal{D}, k_\mathcal{D}) \mathbf{u}_j}{2\omega_j \mathbf{u}_j^{\mathrm{H}} \mathbf{M}_{C_0}^{\beta} \mathbf{u}_j},$$

cf. Eq. (3.10). Hence, the derivative of the signed distance function d_j with respect to $\omega_\mathcal{D}$ reads

$$\frac{\partial}{\partial \omega_\mathcal{D}} d_j(\omega_\mathcal{D}, k_\mathcal{D}) = 2\omega_\mathcal{D} + \frac{\mathbf{u}_j^{\mathrm{H}} \partial_\omega \mathbf{D}_{C_0}(\omega_\mathcal{D}, k_\mathcal{D}) \mathbf{u}_j}{\mathbf{u}_j^{\mathrm{H}} \mathbf{M}_{C_0}^{\beta} \mathbf{u}_j}.$$

Algorithm 6.2. Newton's method applied to global signed distance function in k-formulation.

1: Fix $\omega_\mathcal{D} \in \mathbb{R}^+$ and choose start value $k^{(0)} \in B$.
2: **for** $n = 0, \dots$ **do**
3: **if** $\omega_\mathcal{D}^2 \in \sigma_h^{\mathrm{ess}}(k^{(n)})$ **then**
4: **exit** (and restart with new start value $k^{(0)} \in B$)
5: **end if**
6: Solve linear eigenvalue problem (6.60) for ω^2 with $k_\mathcal{D} = k^{(n)}$.
7: Evaluate global signed distance function $d^{\mathrm{g}}(\omega_\mathcal{D}, k^{(n)})$.
8: **if** $d^{\mathrm{g}}(\omega_\mathcal{D}, k^{(n)}) \approx 0$ **then**
9: **exit**
10: **end if**
11: Compute new value $k^{(n+1)} = k^{(n)} - \left(\frac{\partial}{\partial k_\mathcal{D}} d_{j^*}(\omega_\mathcal{D}, k^{(n)})\right)^{-1} d^{\mathrm{g}}(\omega_\mathcal{D}, k^{(n)})$.
12: **end for**

The iterative scheme in k-formulation, i. e. keeping $\omega_\mathcal{D} \in \mathbb{R}^+$ fixed and searching for a root $k \in B$ of $d^{\mathrm{g}}(\omega_\mathcal{D}, \cdot)$, works analogously to the ω-formulation and is presented in Algorithm 6.2. Again, the derivative of d_{j^*} with respect to $k_\mathcal{D}$ can either be approximated by a difference quotient, or computed with the help of a closed formula, that can be derived using the facts that the DtN operators are differentiable with respect to the quasi-momentum, cf. Proposition 6.9, and that the eigenvalues $\omega_j^2(\omega_\mathcal{D}, k_\mathcal{D})$ of the linear eigenvalue problem (6.60) as well as their corresponding eigenvectors $\mathbf{u}_j(\omega_\mathcal{D}, k_\mathcal{D})$ are continuously differentiable with respect to $k_\mathcal{D}$, cf. Proposition 6.25. Then we can differentiate (6.60) with respect to $k_\mathcal{D}$ and multiply it from the left with the conjugate transpose of $\mathbf{u}_j$, which yields, similarly to above,

$$\frac{\partial \omega_j(\omega_\mathcal{D}, k_\mathcal{D})}{\partial k_\mathcal{D}} = -\frac{\mathbf{u}_j^{\mathrm{H}} \left(\mathbf{C}_{C_0}^{\alpha,1} + 2k_\mathcal{D} \mathbf{M}_{C_0}^{\alpha} + \partial_k \mathbf{D}_{C_0}(\omega_\mathcal{D}, k_\mathcal{D})\right) \mathbf{u}_j}{2\omega_j \mathbf{u}_j^{\mathrm{H}} \mathbf{M}_{C_0}^{\beta} \mathbf{u}_j},$$

and hence, the derivative of the signed distance function d_j with respect to $k_\mathcal{D}$ reads

$$\frac{\partial}{\partial k_\mathcal{D}} d_j(\omega_\mathcal{D}, k_\mathcal{D}) = \frac{\mathbf{u}_j^{\mathrm{H}} \left(\mathbf{C}_{C_0}^{\alpha,1} + 2k_\mathcal{D} \mathbf{M}_{C_0}^{\alpha} + \partial_k \mathbf{D}_{C_0}(\omega_\mathcal{D}, k_\mathcal{D})\right) \mathbf{u}_j}{\mathbf{u}_j^{\mathrm{H}} \mathbf{M}_{C_0}^{\beta} \mathbf{u}_j}.$$

The selection of the start values $\omega^{(0)}$ and $k^{(0)}$ is of particular importance for the convergence of the Newton method. Suppose that for some $k \in B$ the eigenvalues of the nonlinear eigenvalue problem (6.44) are known. Due to the analyticity of the dispersion curves with respect to ω, cf. Section 6.2.3, it seems reasonable to choose these eigenvalues as start values for the Newton method applied to $d^{\mathrm{g}}(\cdot, k + h)$, i. e. the ω-formulation at the quasi-momentum $k + h$ with some small perturbation h of k. For the k-formulation, however, there generally does not exist such a possibility since the group velocity can be identical to zero, which implies that the inverse of the dispersion curves are in general not analytic in $\mathbb{R}^+$.

We only mention here an approach that is computationally expensive, but can be applied to both, the k-formulation and the ω-formulation. If the results of the supercell method (possibly with a small number of periodicity cells to reduce computation costs) is available, they can deal as start values for the Newton method.

6.3 Numerical results

In this section we present numerical results of the proposed methods to solve the nonlinear eigenvalue problem (6.58) with DtN maps. The numerical example we will discuss in this section is the one presented in Example 2 in Chapter 2, i. e. we study the TE mode band structure of a PhC W1 waveguide with hexagonal lattice, relative permittivity $\varepsilon = 11.4$, and holes of relative radius $\frac{r}{a_1} = 0.31$. Unless otherwise stated, the polynomial degree of the FE computations is set to $p = 5$.

6.3.1 Numerical results of the proposed Newton method

The Newton method, that we proposed in Section 3.3 and specified above in Section 6.2.5 when applied to the nonlinear eigenvalue problem (6.58) with DtN maps, follows the idea of computing the roots of the global signed distance function d^{g} defined in (6.62). To give a first orientation, the magnitude of the global signed distance function d^{g} is plotted in in Figure 6.3. The dark lines indicate to small magnitudes of d^{g} and hence, they show guided modes. The areas left blank correspond to the essential spectrum $\sigma_h^{\mathrm{ess}}(k)$.

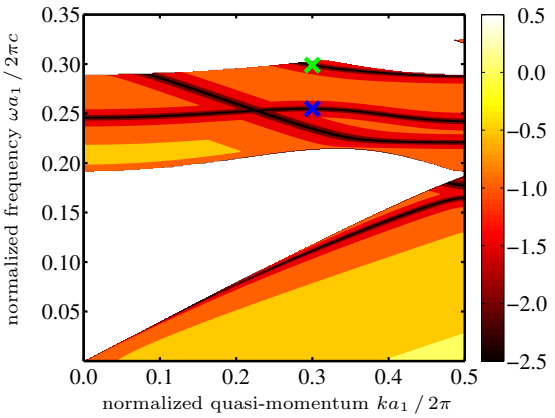
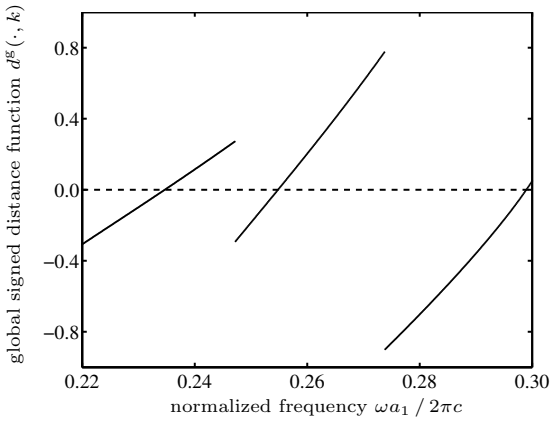


Figure 6.3: Magnitude of global signed distance function d^{g} in logarithmic scale evaluated on a grid of 350×500 (ω, k)-points. The areas left blank correspond to the essential spectrum, i. e. $\omega^2 \in \sigma_h^{\mathrm{ess}}(k)$. The crosses mark the locations of the two guided modes for which we present numerical results in Figures 6.6 and 6.9.

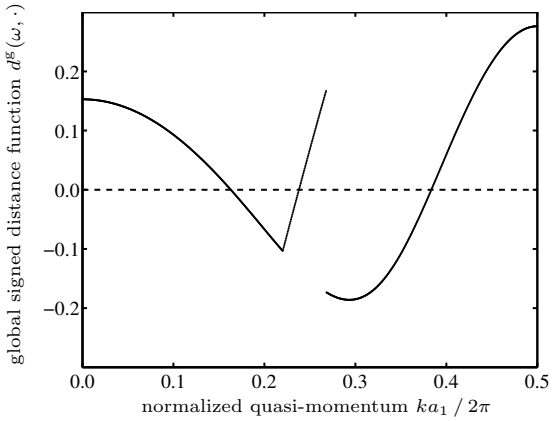
As elaborated above, we know that the global signed distance function d^{g} is discontinuous. This can be seen, for example, in Figure 6.4, where the results of the global signed distance function d^{g} are resolved on a fine scale with respect to the frequency ω for a fixed quasi-momentum $k = 0.3 \cdot \frac{2\pi}{a_1}$, see Figure 6.4a, and with respect to the quasi-momentum k for a fixed frequency $\omega = 0.25 \cdot \frac{2\pi c}{a_1}$, see Figure 6.4b.

Note that for the particular example shown in Figure 6.4, the global signed distance function d^{g} is smooth in a neighbourhood of its roots. However, when two dispersion curves cross, there will not exist a smooth neighbourhood of the global signed distance function, but it will be continuous at this point, since it will tend to zero from both sides. Hence, the application of the Newton method to find the roots of the global signed distance functions, as proposed in Section 6.2.5, is reasonable as long as the start value is sufficiently close to the root.

In the following numerical results we will use the global signed distance function d^{g} in ω-formulation, i. e. we keep the quasi-momentum $k \in B$ fixed and search for roots $\omega \in \mathbb{R}^+$ of $d^{\mathrm{g}}(\cdot, k)$.

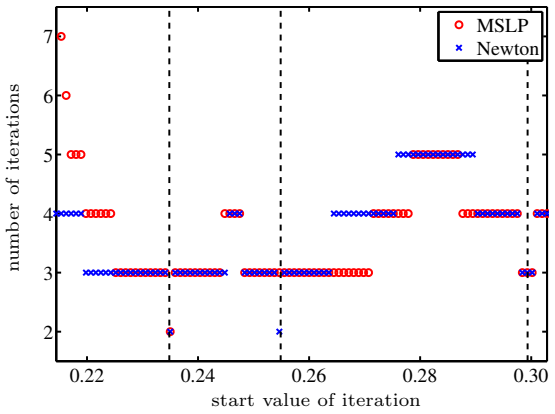


(a) $d^g(\cdot, k)$ with respect to ω for fixed $k = 0.3 \cdot \frac{2\pi}{a_1}$.

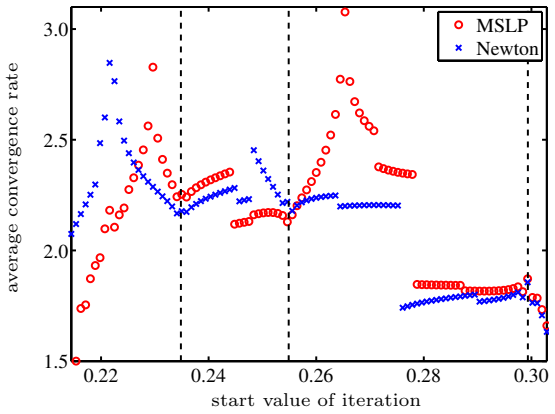


(b) $d^g(\omega, \cdot)$ with respect to k for fixed $\omega = 0.25 \cdot \frac{2\pi c}{a_1}$.

Figure 6.4: Global signed distance function d^g evaluated on an equidistant grid of frequencies ω in the interval $[0.22 \cdot \frac{2\pi c}{a_1}, 0.30 \cdot \frac{2\pi c}{a_1}]$ for a fixed value of $k = 0.3 \cdot \frac{2\pi}{a_1}$ (a), and evaluated on an equidistant grid of quasi-momenta k in the irreducible Brillouin zone $\widehat{B} = [0, \frac{\pi}{a_1}]$ for a fixed value of $\omega = 0.25 \cdot \frac{2\pi c}{a_1}$ (b).



(a) Number of iterations until the step size is below 10^{-13} in magnitude.



(b) Average convergence rate, i. e. the average of all rates per iteration.

Figure 6.5: Number of iterations (a) and average convergence rate (b) of the proposed Newton method (blue crosses) and of the MSLP (red circles) for a sample of 100 start values in the second band gap at $k = 0.3 \cdot \frac{2\pi}{a_1}$. The dashed vertical lines show the locations of the three guided modes in the band gap.

In our first numerical test of the proposed Newton method, we study the convergence of the Newton method in ω-formulation and compare the results with convergence of the frequently used MSLP, which we briefly introduced in Section 3.2. Like the Newton method applied to distance functions, the MSLP is an iterative procedure to solve nonlinear eigenvalue problems, that is supposed to converge quadratically. For a sample of 100 start values of the frequency ω in the second band gap at $k = 0.3 \cdot \frac{2\pi}{a_1}$ we present in Figure 6.5a the number of iterations that are needed until the step size of the Newton method and the MSLP, respectively, are below a threshold of 10^{-13}. The dashed vertical lines show the locations of the three guided modes in the band gap for which the real parts of the magnetic field components are shown in Figure 6.7. Apart from a small number of start values, both, the proposed Newton method and the MSLP need three to five iterations. For most start values the number of iterations needed by the Newton method is identical to the number of iterations of the MSLP. Only for the twelve smallest start values the MSLP constantly needs more iterations than the Newton method and for the start value closest to the lower band edge, the MSLP does not converge at all. This behaviour of the MSLP is linked to the

existence of global Dirichlet eigenvalues and will be explained later in Section 6.3.5. For the same sample of start values, Figure 6.5b shows the average convergence rate, i. e. the average of the numbers $q^{(n)}$ that satisfy

$$\frac{\left|\omega^{(n+1)} - \omega_{\text{ref}}\right|}{\left|\omega^{(n)} - \omega_{\text{ref}}\right|^{q^{(n)}}} = 1$$

for all iterations $n = 0, 1, \ldots, N-1$, where the reference solution ω_{ref} is chosen to be the solution of the respective method after N iterations, i. e. the number that is shown in Figure 6.5a. The rates differ from start value to start value and from method to method but they stay within the interval $[1.5, 3.1]$ around the expected rate of two. However, the average of all rates is approximately the same as the expected rate. The Newton method shows an average rate of 2.0684 and the MSLP's average convergence rate is 2.1004.

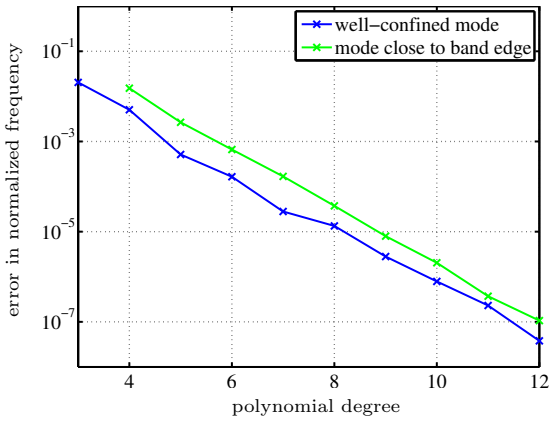


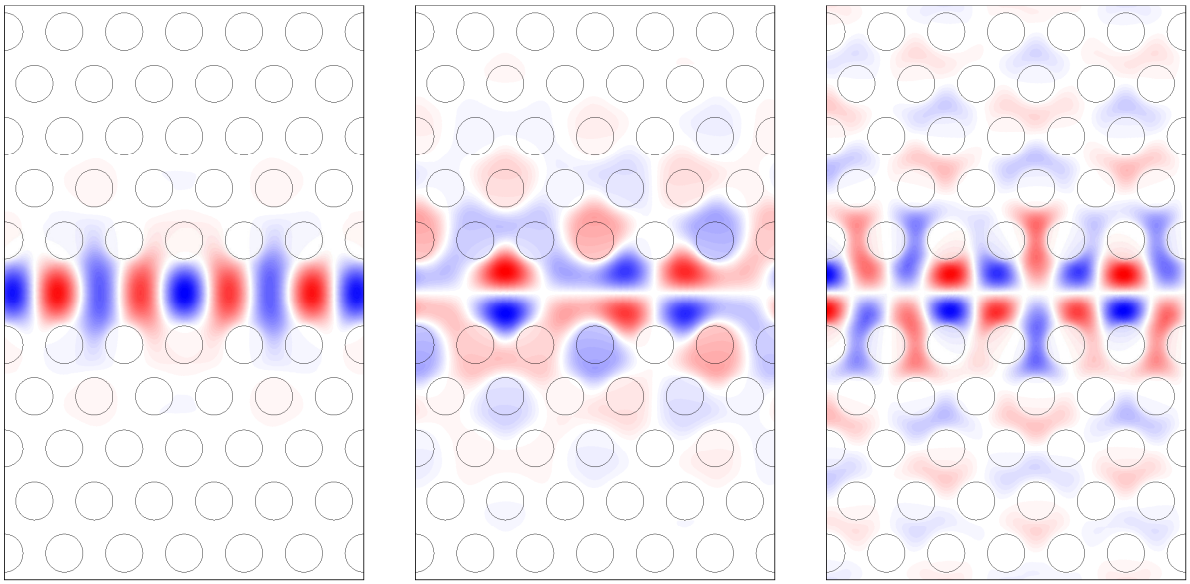
Figure 6.6: Convergence of Newton method with respect to the polynomial degree p when applied to the computation of a well-confined mode (blue) and a mode close to a band edge (green). The reference solution ω_{ref} is computed taking a polynomial degree of $p = 20$.

From Figure 6.5b it seems that for either method the convergence towards the well-confined mode, marked with a blue cross in Figure 6.3, seems of larger rate than the convergence towards the mode close to the band edge, that is marked with a green cross in Figure 6.3. In order to study this seemingly dependence on the confinement of the guided modes in more detail, we show convergence results of the Newton method in ω-formulation with respect to the polynomial degree in Figure 6.6. The reference solution is computed by setting the polynomial degree to $p = 20$ and applying the same iterative scheme. Note that there is no value for the error of the mode close to the band edge for the lowest polynomial degree $p = 3$ since for this degree the mode is inside the approximative essential spectrum and can therefore not be captured. As expected for p-FEM, we can observe exponential convergence with the same convergence rate for both modes independent of their confinement.

This result demonstrates that the DtN transparent boundary conditions, as introduced in this chapter, resolve the problem of the frequently used supercell method, which is known to introduce a modelling error that depends on the confinement of the guided mode and for which we will present numerical results in the following section.

6.3.2 Comparison to the numerical results of the supercell method

The supercell method, that we briefly introduced in Section 2.4, provides approximations to guided modes. The application of the supercell method to the setting of Example 2 was already presented in [SK10]. It was shown that p-FEM converges exponentially when the numerical results are compared to a reference solution that is also obtained with the same supercell. In this section we will to show that



(a) Well-confined, even mode at $\omega \approx 0.235 \cdot \frac{2\pi c}{a_1}$.

(b) Well-confined, odd mode at $\omega \approx 0.255 \cdot \frac{2\pi c}{a_1}$.

(c) Odd mode close to band edge at $\omega \approx 0.299 \cdot \frac{2\pi c}{a_1}$.

Figure 6.7: Real parts of the magnetic field components of the three guided modes in the second band gap at $k = 0.3 \cdot \frac{2\pi}{a_1}$.

this convergence cannot be expected when comparing the results with a reference solution obtained using DtN transparent boundary conditions that do not introduce a modelling error.

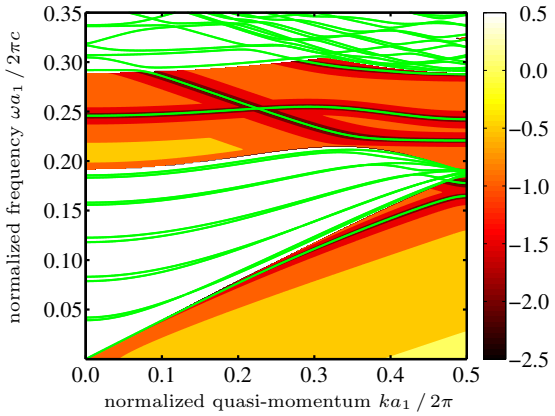
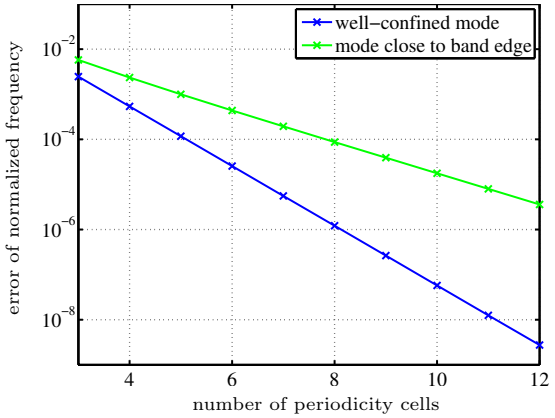


Figure 6.8: Magnitude of global signed distance function d^{g} in logarithmic scale evaluated on a grid of 350×500 (ω, k)-points, and results of the supercell method with five rows of periodicity cells on top and bottom (green lines).

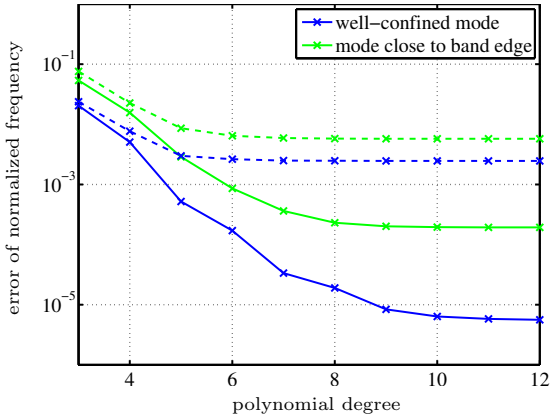
First of all we show a comparison of the supercell band structure with the results of the global signed distance function d^{g}, that were already presented in Figure 6.3. In Figure 6.8 the green lines correspond to the dispersion curves obtained when using the supercell S_5 with five rows of holes on top and bottom of the line defect as shown in Figure 2.8. We can see that inside the band gaps the green lines match well with the dark lines of the global signed distance function d^{g}, that indicate small values of its magnitude $|d^{\mathrm{g}}|$.

The convergence results of the supercell method when applied to the computation of the well-confined mode at $k = 0.3 \cdot \frac{2\pi}{a_1}$ (blue cross in Figure 6.3) and the mode close to the band edge (green cross

in Figure 6.3) are presented in Figure 6.9. On the left we observe an exponential convergence of the results of the supercell method with polynomial degree $p = 7$ towards the roots of the global signed distance function of the nonlinear eigenvalue problem (6.58) with DtN maps and the same polynomial degree $p = 7$, when increasing the number of periodicity cells on top and bottom. However, the rates of convergence differ significantly. The rate of the mode close to the band edge (green) is much smaller than the rate of the well-confined mode (blue), see Figure 6.9a. At this point we have to note that the FE mesh of the supercell method is significantly larger than the mesh of the DtN method sketched in Figure 6.1, see for example the mesh of the supercell with five periodicity cells on top and bottom presented in Figure 2.8 in Chapter 2. Figure 6.9b, on the other hand, where the number of periodicity cells is kept fixed to $n = 3$ and $n = 7$ while the polynomial degree is increased from $p = 3$ to $p = 12$, shows that the error of the supercell method only converges exponentially towards the solution of the roots of the global signed distance function of the nonlinear eigenvalue problem (6.58) with DtN maps and polynomial degree $p = 20$ until a certain error plateau is reached. This error plateau, which is due to the modelling error introduced by the supercell approach, is significantly larger for the mode close the band edge compared to the well-confined mode. These results clearly demonstrate that the supercell method is a good approximation of the exact DtN method for well-confined modes but for modes close to the band edge it produces errors of significantly larger orders.



(a) Fixed polynomial degree $p = 7$.



(b) Fixed number of periodicity cells $n = 3$ (dashed lines) and $n = 7$ (solid lines).

Figure 6.9: Convergence of the well-confined mode (blue) and the mode close to the band edge (green) with respect to the number of the periodicity cells n while keeping the polynomial degree p fixed (a), and with respect to the polynomial degree p while keeping the number of periodicity cells n fixed (b). The reference solution ω_{ref} is computed using the iterative DtN method with a polynomial degree $p = 7$ (a) and $p = 20$ (b).

6.3.3 Numerical results of the direct procedure

Let us now come to the numerical results of the direct Chebyshev interpolation to solve the nonlinear eigenvalue problem (6.58). Recall from Section 6.2.5 that the Chebyshev interpolation requires a priori knowledge of the essential spectrum $\sigma_h^{\text{ess}}(k)$, since the analyticity of the nonlinear matrix function $\mathbf{N}_{C_0}$ in Eq. (6.59) can only be guaranteed in the band gaps, i. e. outside the essential spectrum. Thus, this method is particularly interesting, if we apply it to the k-formulation at frequencies ω that are in a band gap for all $k \in B$, i. e. $\omega^2 \notin \sigma_h^{\text{ess}}(k)$ for all $k \in B$. The convergence of the Chebyshev interpolation is shown in Figure 6.10, where the results of the direct procedure to compute the eigenvalues in the band gap $[0.22 \cdot \frac{2\pi c}{a_1}, 0.28 \cdot \frac{2\pi c}{a_1}]$ using the Chebyshev interpolation is compared to a reference solution computed with Newton's method. We observe an exponential convergence of the mean error of the eigenvalues computed at a sample of 200 frequencies in the band gap. Note that convergence is not monotone. This

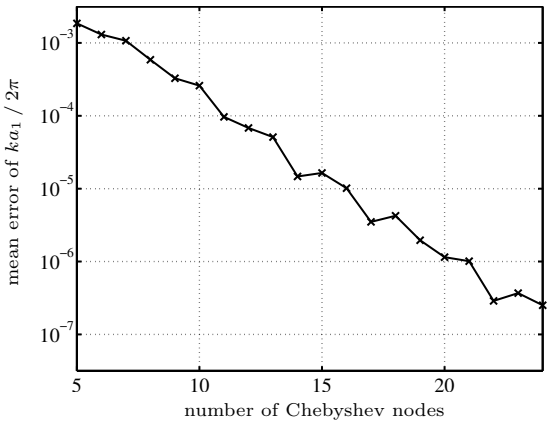


Figure 6.10: Convergence of the mean error of the Chebyshev interpolation in the irreducible Brillouin zone $\widehat{B}$ at a sample of 200 frequencies in the band gap $[0.22 \cdot \frac{2\pi c}{a_1}, 0.28 \cdot \frac{2\pi c}{a_1}]$ with respect to the number of Chebyshev nodes d.

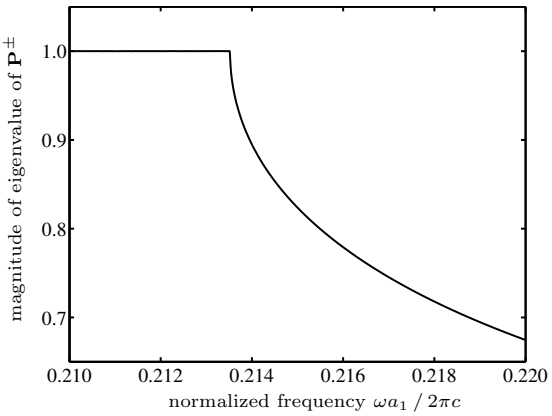
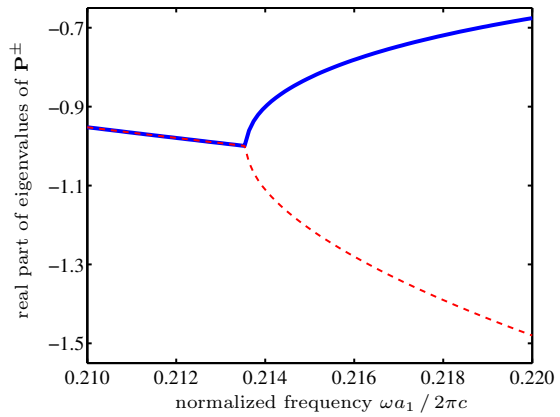
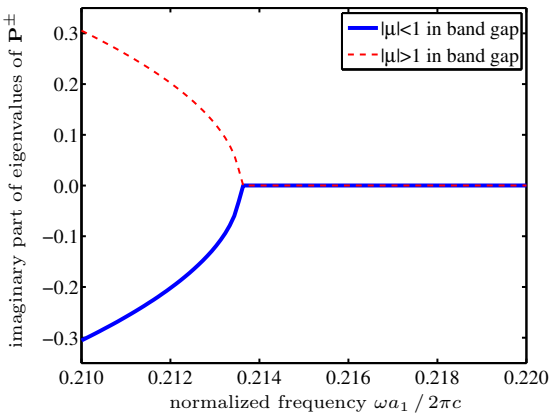


Figure 6.11: Magnitude of the eigenvalue of the propagation matrix $\mathbf{P}^{\pm}$ at $k = 0.3 \cdot \frac{2\pi}{a_1}$ near the band edge at $\omega \approx 0.2135 \cdot \frac{2\pi c}{a_1}$, that has magnitude strictly less than one in the band gap and equal to one in the essential spectrum.

is due to the fact that the Chebyshev nodes are not hierarchical and hence, the error of the Chebyshev interpolation can increase when using more nodes.



(a) Real part.



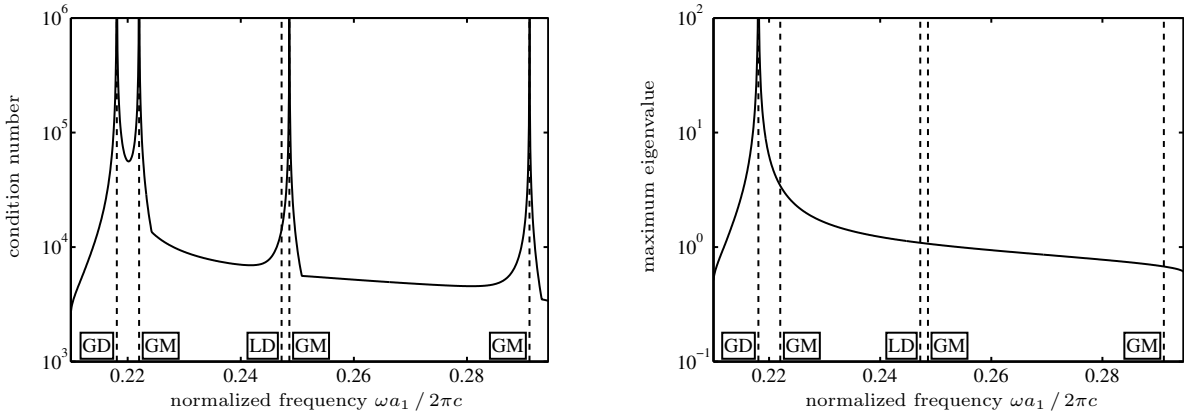
(b) Imaginary part.

Figure 6.12: Real part (a) and imaginary part (b) of the eigenvalues μ of the propagation matrix $\mathbf{P}^{\pm}$ at $k = 0.3 \cdot \frac{2\pi}{a_1}$ near the band edge at $\omega \approx 0.2135 \cdot \frac{2\pi c}{a_1}$, that have magnitude equal to one in the essential spectrum and strictly less than one (blue) and strictly larger than one (red) in the band gap. Hence, the blue curves correspond to the magnitude shown in Figure 6.11.

Analogously, we can observe exponential convergence towards the iterative solution if the direct procedure is applied to the ω-formulation, as done in [KS14a], where we proposed to employ the adaptive path following algorithm introduced in Chapter 5 to compute the essential spectrum $\sigma^{\text{ess}}(k)$ for all $k \in B$. However, we have to take care that the ω-interval is sufficiently far away from the band edge, since the magnitude of the eigenvalue μ of the propagation operator, that changes from $|\mu| = 1$ to $|\mu| < 1$ at the band edge, has a root-like singularity at the band edge as shown in Figures 6.11 and 6.12, and hence, its derivative with respect to ω becomes arbitrarily large near the band edge, which will dominate the derivative of the DtN operator. This drawback of the Chebyshev interpolation is of smaller significance in the k-formulation as long as the chosen frequency interval is not arbitrarily close to the band edge at any $k \in B$, as demonstrated in the convergence analysis in Figure 6.10.

6.3.4 Condition of system and Dirichlet-to-Neumann matrices

Recall that the DtN operators $\mathcal{D}^{\pm}$ are not well-defined, if the Dirichlet problems (6.1) in the semi-infinite strips $S^{\pm}$ are not well-posed, which is the case at so called global Dirichlet eigenvalues of (6.1). Furthermore, recall that the local Dirichlet cell problems (6.8), that we introduced for the characterization of the DtN operators, are ill-posed at so called local Dirichlet eigenvalues. Hence, we expect the nonlinear eigenvalue problem (6.58) to show numerical artifacts at these values.



(a) Condition number of $\mathbf{N}_{C_0}$.

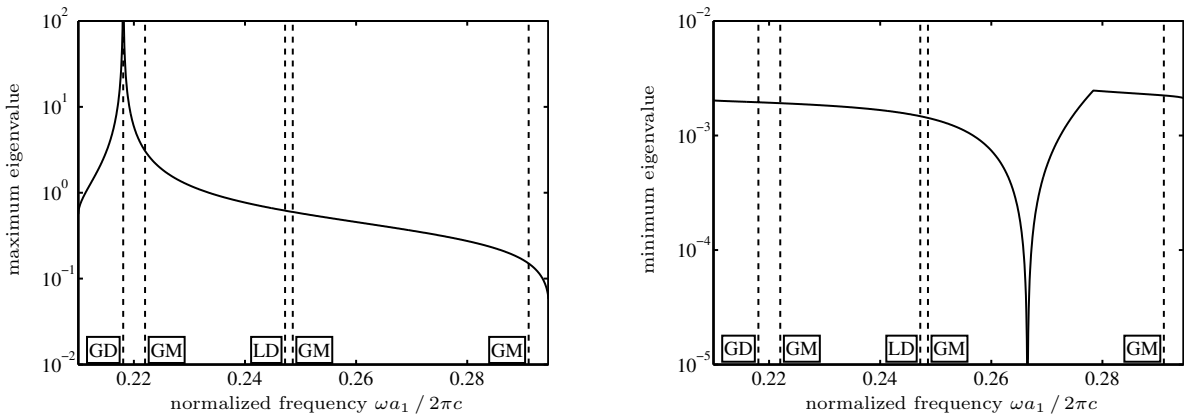
(b) Maximum eigenvalue of $\mathbf{N}_{C_0}$.

Figure 6.13: Condition number (a) and maximum eigenvalue (b) of the system matrix $\mathbf{N}_{C_0}$ in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$. The vertical dashed line show the frequency of the global Dirichlet eigenvalue (GD), the local Dirichlet eigenvalue (LD) and the frequencies of the guided modes (GM).

We start by analysing the condition number and the maximum eigenvalue of the matrix $\mathbf{N}_{C_0}$. Inside the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$, for which the results are presented in Figure 6.13, there are three guided modes, one local Dirichlet eigenvalue and one global Dirichlet eigenvalue. As can be seen from Figure 6.13a the condition number increases at the guided modes (dashed lines labeled "GM"). This is the expected behaviour since by definition the guided modes are eigenvalues of (6.58) and hence, the minimum eigenvalue of $\mathbf{N}_{C_0}$ tends to zero when approaching the guided modes. Apart from these three peaks, the condition number of the system matrix $\mathbf{N}_{C_0}$ also increases in the vicinity of the global Dirichlet eigenvalue (dashed line labeled "GD"), which is due to an increasing maximum eigenvalue of $\mathbf{N}_{C_0}$, see Figure 6.13b. Note that from Figure 6.13a it seems that the local Dirichlet eigenvalue (dashed line labeled "LD") has no influence on the condition number of $\mathbf{N}_{C_0}$. However, we shall study its influence on the condition number in more detail in Figure 6.16.

The location of the local Dirichlet eigenvalues can be determined from a simple linear eigenvalue problem in the cell $C_1^{\pm}$ with homogeneous Dirichlet boundary conditions. In Figure 6.15 the dispersion curves of the local Dirichlet eigenvalue problem are shown in comparison to the values of the global signed distance function d^{g}. The computation of the global Dirichlet eigenvalues, on the other hand, is not as easy as the computation of the local Dirichlet eigenvalues, as the global Dirichlet eigenvalue problem is posed on the infinite half-strips $S^{\pm}$, and hence, the domain needs to be truncated which cannot be done with DtN transparent boundary conditions. In Chapter 7 we will show how to solve the Dirichlet eigenvalue problem in the infinite half-strips $S^{\pm}$ by truncating the domain using RtR operators, see Figure 7.2 for the dispersion curves of global Dirichlet eigenvalues.

The increase of the maximum eigenvalue of the system matrix $\mathbf{N}_{C_0}$ near the global Dirichlet eigenvalue is due to an increase of the maximum eigenvalue of the DtN matrix $\mathbf{D}^{\pm}$ as shown in Figure 6.14a. While the condition number of the DtN matrix also does not seem to be influenced by the existence of a local Dirichlet eigenvalue, its minimum eigenvalue decreases at some point between the second and third guided mode, see Figure 6.14b. At this point the PhC half-strip problem with homogeneous Neumann boundary condition has an eigenvalue — a *global Neumann eigenvalue*. However, this decreasing minimum



(a) Maximum eigenvalue of $\mathbf{D}^{\pm}$.

(b) Minimum eigenvalue of $\mathbf{D}^{\pm}$.

Figure 6.14: Maximum (a) and minimum (b) eigenvalue of the DtN matrix $\mathbf{D}^{\pm}$ in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$. The vertical dashed line show the frequency of the global Dirichlet eigenvalue (GD), the local Dirichlet eigenvalue (LD) and the frequencies of the guided modes (GM).

eigenvalue of the DtN matrix does not influence the condition number of the system matrix due to the matrix $\mathbf{B}_{C_0}$ in the nonlinear eigenvalue problem (6.58).

Now let us study the condition number of the system and DtN matrices in a very small vicinity of the local Dirichlet eigenvalue in more detail. Figure 6.16 shows the condition numbers of the two matrices in dependence on the distance to the local Dirichlet eigenvalue. While we could not observe an effect of the local Dirichlet eigenvalue on the condition number of $\mathbf{N}_{C_0}$ in Figure 6.13a, we can see from Figure 6.16 that condition numbers of both matrices, the system matrix $\mathbf{N}_{C_0}$ and the DtN matrix $\mathbf{D}^{\pm}$, increase significantly near the local Dirichlet eigenvalues. However, note that this significant increase is restricted to a very narrow vicinity of the local Dirichlet eigenvalue. The minimum eigenvalues of the local DtN matrices $\mathbf{T}^{\pm}_{ij}$, $i, j = 0, 1$, decrease in a larger vicinity of the local Dirichlet eigenvalues. However, the generalized eigenvalue problem (6.38), that we have to solve to obtain the propagation matrix $\mathbf{P}^{\pm}$ can be solved using, e. g. Matlab's `eig` function, an implementation of the generalized Schur decomposition, without any numerical artifacts up to a very narrow vicinity of the local Dirichlet eigenvalue.

The effect of global and local Dirichlet eigenvalues will now also be studied in the next two sections, where we will compare the results of the proposed Newton method with the iterative MSLP, and show convergence results of the Newton method and the Chebyshev interpolation in the vicinity of global and local Dirichlet eigenvalues.

6.3.5 Computation of eigenvalues in vicinity of global Dirichlet eigenvalues

As elaborated above, we expect that global Dirichlet eigenvalues influence the performance of our numerical schemes for solving the nonlinear eigenvalue problem (6.58). In order to analyse this influence, let us now compare the proposed Newton-like method with the MSLP. We already showed in Section 6.3.1 that both methods converge with comparable convergence rates.

In Figure 6.17 we present the step sizes of the MSLP and the proposed Newton method in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$ when using different frequencies ω as start value. The vertical, dashed lines show the locations of the guided modes, i. e. the eigenvalues of the nonlinear eigenvalue problem (6.58). Both step size curves have roots at the guided modes and their slopes are negative at these roots which implies that the methods will converge well to the eigenvalues. While the step size of the Newton method does not change its behaviour at the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$, the step size of the MSLP has another root at the global Dirichlet eigenvalue. This can be explained by the fact that not only the maximum eigenvalue of the system matrix $\mathbf{N}_{C_0}$ tends to infinity at the global Dirichlet eigenvalue, see Figure 6.13b, but also the maximum eigenvalue of the derivative $\partial_\omega \mathbf{N}_{C_0}$ of the system matrix. Hence,

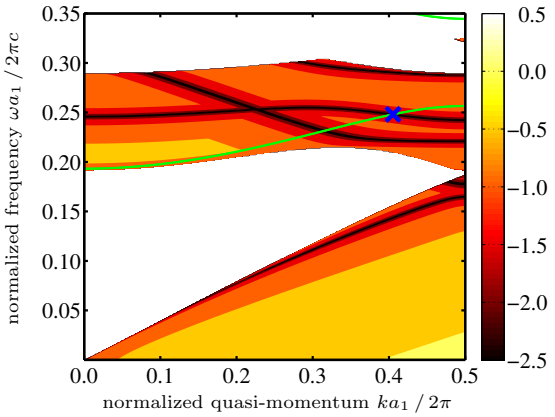


Figure 6.15: Magnitude of global signed distance function d^{g} in logarithmic scale evaluated on a grid of 350×500 (ω, k)-points, and local Dirichlet eigenvalues (green lines). The blue cross indicates the location of the eigenvalue for which convergence results are shown in Figure 6.19.

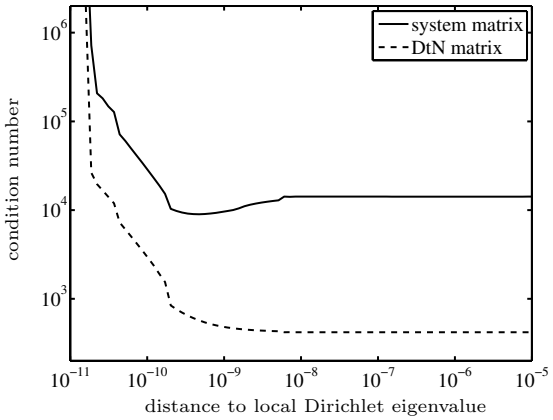


Figure 6.16: Condition number of the system matrix $\mathbf{N}_{C_0}$ (solid line) and the DtN matrix $\mathbf{D}^{\pm}$ (dashed line) in the vicinity of the local Dirichlet eigenvalue in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$.

if $\partial_\omega \mathbf{N}_{C_0}$ has an arbitrarily large eigenvalue, the generalized eigenvalue problem to be solved for the MSLP step size, see Algorithm 3.1, has an eigenvalue zero. However, the slope of the MSLP step size at the global Dirichlet eigenvalue is positive, which means that the MSLP does not converge to the global Dirichlet eigenvalue. But the sign change of the MSLP step size at the global Dirichlet eigenvalue implies that the radius of convergence of the MSLP towards the guided mode at $\omega \approx 0.222 \cdot \frac{2\pi c}{a_1}$ is bounded by the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$, whereas the radius of convergence of the proposed Newton method is not affected by the global Dirichlet eigenvalue.

This demonstrates that our proposed Newton method is preferable to other iterative solvers like the MSLP whose radius of convergence is bounded by infinite eigenvalues of the derivative of the nonlinear matrix.

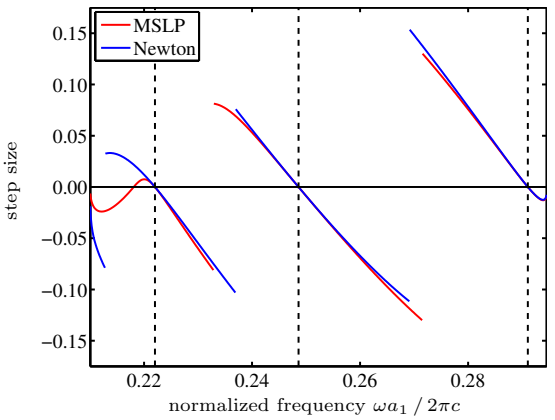


Figure 6.17: Step sizes of the MSLP (red) and the Newton method applied to the global signed distance function (blue) in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$ at different start values of frequency ω. The vertical, dashed lines show the locations of the guided modes.

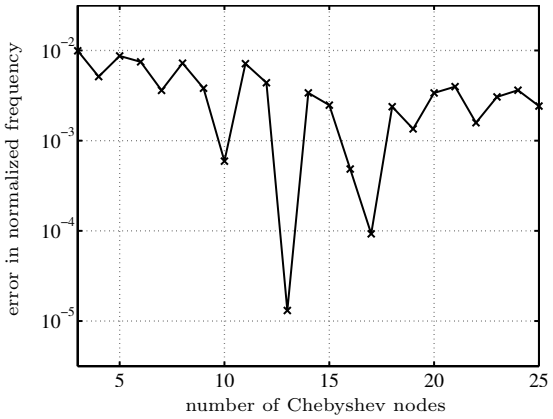


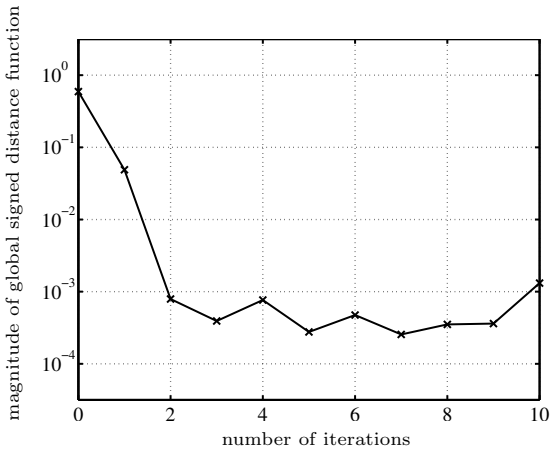
Figure 6.18: Absolute error of the Chebyshev interpolation in ω-formulation when applied to the computation of the guided mode in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$ that is closest to the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$. The ω-interval of the interpolation is chosen to be $[0.215 \cdot \frac{2\pi c}{a_1}, 0.245 \cdot \frac{2\pi c}{a_1}]$.

Now let us apply the linearization based on Chebyshev interpolation to the computation of the guided mode in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$ that is closest to the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$. We employ the Chebyshev interpolation in ω-formulation and choose the interval $[0.215 \cdot \frac{2\pi c}{a_1}, 0.245 \cdot \frac{2\pi c}{a_1}]$, that comprises both, an eigenvalue of the nonlinear eigenvalue problem (6.58) as well as a global Dirichlet eigenvalue. In Figure 6.18 we present the error in the normalized frequency with respect to the number of Chebyshev nodes. Even though the chosen interval is relatively small, especially compared to the irreducible Brillouin zone used for the Chebyshev interpolation in ω-formulation presented in Figure 6.10, the Chebyshev interpolation does not converge. This is due to the fact that the interval contains a global Dirichlet eigenvalue, which implies that for any sufficiently large number of Chebyshev nodes there will be a Chebyshev node that is very close to the global Dirichlet eigenvalue and hence, the nonlinear matrix function $\mathbf{N}_{C_0}$, that is evaluated at all Chebyshev nodes, has a prohibitively large condition number that spoils the eigenvalue computation. This means that the Chebyshev interpolation will always fail to identify eigenvalues that are close to global Dirichlet eigenvalues.

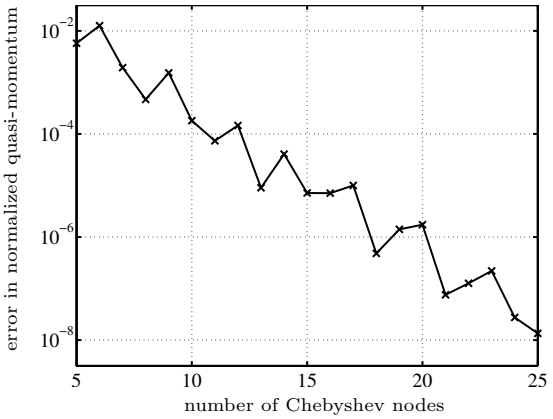
6.3.6 Computation of eigenvalues in vicinity of local Dirichlet eigenvalues

Now let us analyse the behaviour of the Newton method and the Chebyshev interpolation close to a local Dirichlet eigenvalue. Since the condition number of the DtN matrices $\mathbf{D}^{\pm}$ and the system matrix $\mathbf{N}_{C_0}$ only increase in a very narrow vicinity of the local Dirichlet eigenvalue, see Figure 6.16, we can only expect the local Dirichlet eigenvalue to influence the performance of the numerical schemes in this narrow vicinity. To this end, we shall study in the section the convergence of the numerical schemes towards a common eigenvalue of the nonlinear eigenvalue problem (6.58) and the local Dirichlet problem in the cell $C_1^{\pm}$. The blue cross in Figure 6.15 at $(\omega, k) \approx (0.248 \cdot \frac{2\pi c}{a_1}, 0.405 \cdot \frac{2\pi}{a_1})$ marks such a common eigenvalue for which we will now present numerical results.

In Figure 6.19a the magnitude of the global signed distance function is shown in dependence on the number of iterations of the Newton method when using the start value $\omega^{(0)} = 0.263 \cdot \frac{2\pi c}{a_1}$ at $k \approx 0.405 \cdot \frac{2\pi}{a_1}$. We see that the method does not converge to the common eigenvalue $\omega \approx 0.248 \cdot \frac{2\pi c}{a_1}$ of the nonlinear eigenvalue problem (6.58) and the local Dirichlet problem in the cell $C_1^{\pm}$. Instead the magnitude of the global signed distance function remains almost constant after two iterations at a level of 10^{-3}. This is due to the fact that the local Dirichlet problems are ill-posed at the Dirichlet eigenvalues. The closer one comes to such a Dirichlet eigenvalue the larger the condition number of the local DtN matrices $\mathbf{T}_{ij}^{\pm}$, $i, j = 1, 2$, becomes and hence, the more the error of the DtN matrices increases.



(a) ω-formulation of Newton method with start value $\omega^{(0)} = 0.263 \cdot \frac{2\pi c}{a_1}$.



(b) k-formulation of Chebyshev interpolation in irreducible Brillouin zone $\widehat{B} = [0, \frac{\pi}{a_1}]$.

Figure 6.19: Convergence of the Newton method in ω-formulation (a) and the Chebyshev interpolation in k-formulation (b) applied to the computation of the common eigenvalue $(\omega, k) \approx (0.248 \cdot \frac{2\pi c}{a_1}, 0.405 \cdot \frac{2\pi}{a_1})$ of the Dirichlet cell problem and the nonlinear eigenvalue problem with DtN transparent boundary conditions, see blue cross in Figure 6.15.

Now we want to apply the Chebyshev interpolation in k-formulation to the computation of the common eigenvalue $(\omega, k) \approx (0.248 \cdot \frac{2\pi c}{a_1}, 0.405 \cdot \frac{2\pi}{a_1})$ of the nonlinear eigenvalue problem with DtN transparent boundary conditions and the local Dirichlet problem, i. e. we fix the frequency to $\omega \approx 0.248 \cdot \frac{2\pi c}{a_1}$ and chose the irreducible Brillouin zone $\widehat{B} = [0, \frac{\pi}{a_1}]$ as interval for the Chebyshev interpolation. In Figure 6.19b the error in the normalized frequency is shown in comparison to the number of Chebyshev nodes. We can see that error decreases exponentially, where we again have to point out that the convergence is not monotone since the Chebyshev nodes are not hierarchical.

The observed convergence of the Chebyshev interpolation towards a common eigenvalue of the nonlinear eigenvalue problem with DtN transparent boundary conditions and the local Dirichlet problem stands in contrast to the diverging Newton method. The reason for this is that the condition numbers of the DtN matrices $\mathbf{D}^{\pm}$ and the system matrix $\mathbf{N}_{C_0}$ only increase in a very narrow vicinity of the local Dirichlet eigenvalue, see Figure 6.16. While this anyhow effects the Newton method, while approaching this Dirichlet eigenvalue, the Chebyshev interpolation is not effected by the increasing condition number since all Chebyshev nodes, even for larger orders, are sufficiently far away from the Dirichlet eigenvalue.

This situation, however, changes if — in addition to the local Dirichlet eigenvalue — there exists a global Dirichlet eigenvalue inside the interval of the Chebyshev interpolation as our numerical results in Section 6.3.5 showed. We shall analyse this in more detail in the next chapter in Section 7.3.4, where we will apply the Chebyshev interpolation also in ω-formulation which turns out to be problematic in this case.

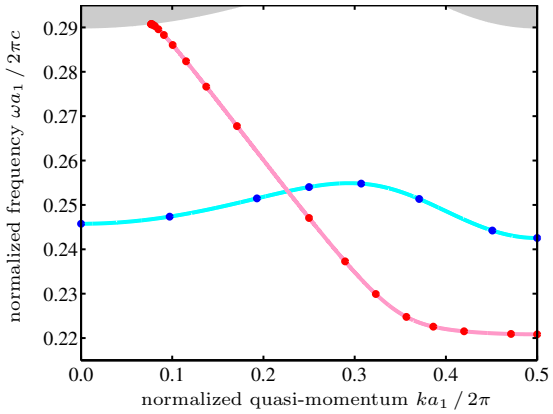
6.3.7 Adaptive path following of dispersion curves

Finally, we want to apply the adaptive path following algorithm based on piecewise Taylor expansions of the dispersion curves, that we introduced in Chapter 5, to the nonlinear eigenvalue problem (6.58) with DtN transparent boundary conditions. For this we employ the formulas for the group velocity and higher derivatives of the dispersion curves, that we derived in Section 6.2.3.

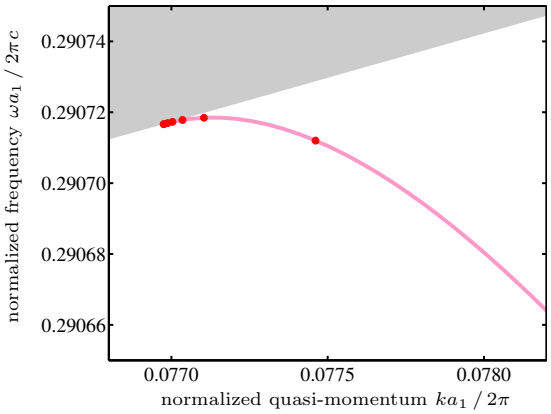
Let $n \in \mathbb{N}$ be the order of the Taylor expansions. As done in Chapter 5, we select a start value $k^{(0)} \in B$ and compute the eigenvalues in a frequency interval $I_\omega \subset \mathbb{R}^+ \setminus \sigma_h^{\text{ess}}(k^{(0)})$ within a band gap. We employ the direct method based on Chebyshev interpolation for the simultaneous computation of all eigenvalues of (6.58) in I_ω. For all eigenvalues, that were found in I_ω, we proceed as presented in Algorithm 5.2 for the case without backward check or as presented in Algorithm 5.3 including backward check, i. e.

(i) we compute the dispersion curve derivatives up to order $n+1$ using Eq. (6.55),
(ii) we evaluate the acceptable step size (5.29) of the Taylor expansion of order n,
(iii) we add the step size to and subtract it from the current node to obtain the next nodes of the quasi-momentum,
(iv) we compute an approximation to the eigenvalue at the next nodes using the Taylor expansion of order n around the current node,
(v) we employ the proposed Newton-like method, or some other iterative scheme, in ω-formulation for the computation of an eigenvalue using the expected location as start value, and then
(vi) we continue to follow the dispersion curve to the left and right, possibly applying additional refinement checks such as the backward check, see Section 5.4.2.

In contrast to the situation in Chapter 5, where we applied the adaptive path following to the supercell approximation, we now have to take the essential spectrum implicitly into account. In Chapter 5 we could simply continue to follow the dispersion curves when they entered the essential spectrum. Now, when using DtN transparent boundary conditions, we cannot continue to follow the dispersion curves across band edges since the DtN operators are not well-defined in the essential spectrum. We resolve this problem as follows: as soon as we find that the expected location of the frequency eigenvalue is outside the band gap, or if an iterative scheme for the computation of the eigenvalue of (6.58), such as the proposed Newton-like method, does not converge in the band gap, we reduce the step size of the path following algorithm, e. g. by the factor $\frac{1}{2}$. If the step size decreases below some threshold $\varepsilon_{\text{tol}}^{\text{edge}}$ during this refinement, we stop to follow the dispersion curve, taking it as granted that the dispersion curve hits the band edge.



(a) Dispersion curves in the second band gap.



(b) Detailed view of the red dispersion curve in the vicinity of the band edge.

Figure 6.20: Adaptive Taylor scheme of order $n = 5$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ applied to the dispersion curves in the second band gap. The error tolerance of the step size computation is $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$, the minimum step size of the band edge refinement is $\varepsilon_{\text{tol}}^{\text{edge}} = 10^{-5}$, and the start value of the iteration is set to $k^{(0)} = \frac{\pi}{2a_1}$.

In Figure 6.20 we present the results of the adaptive Taylor expansion of order $n = 5$ including the additional backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$. Similarly to the results in Chapter 5, the dots indicate the location of the values of k for which the eigenvalues $\omega(k)$ of (6.58) and the dispersion curve derivatives $\omega'(k), \omega^{(2)}(k), \ldots, \omega^{(6)}(k)$ were computed. The lines connecting the dots result from the post-processing, where we again chose the weighted Taylor expansion (5.30).

Note that the red dispersion curve hits the band edge. For this curve the band edge refinement technique, that we described above, was employed with minimum step size $\varepsilon_{\text{tol}}^{\text{edge}} = 10^{-5}$. A detailed view of the dispersion curve at the band edge is shown in Figure 6.20b.

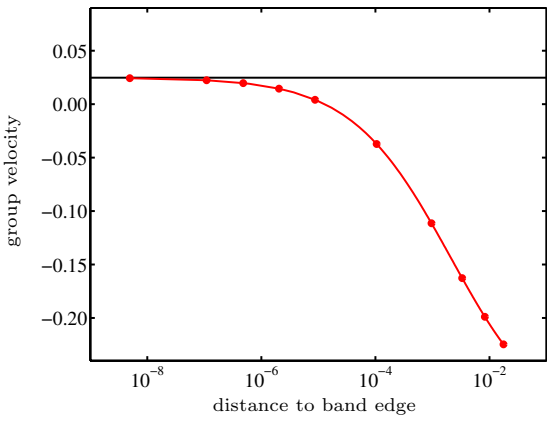


Figure 6.21: Group velocity (red) of the red dispersion curve in Figure 6.20 in dependence on the distance to the band edge. The black line shows the slope of the band edge at the position where the red dispersion curve hits the band edge.

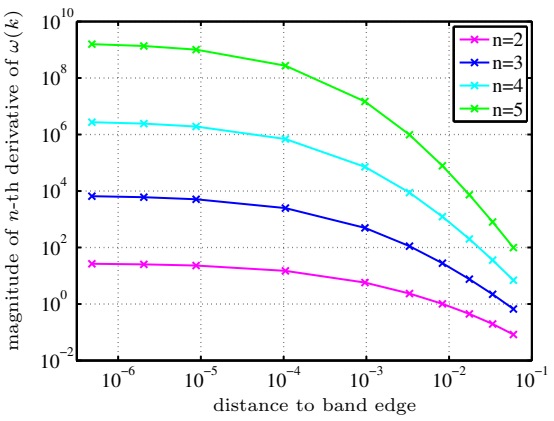
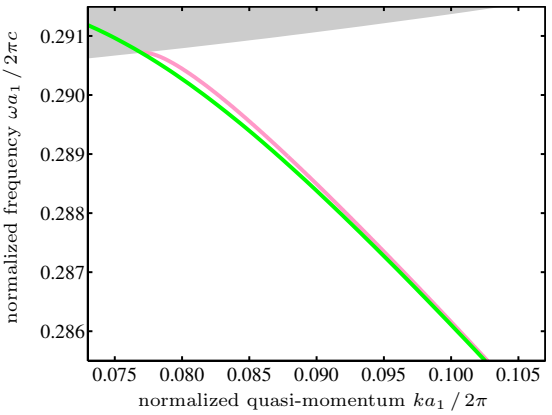


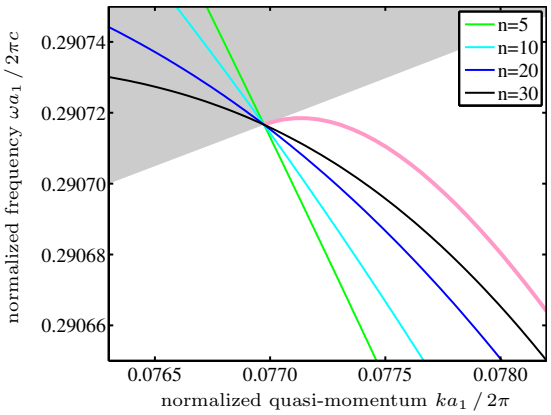
Figure 6.22: Magnitude of the n-th derivatives, $n = 2, \ldots, 5$, of the red dispersion curve in Figure 6.20 in dependence on the distance to the band edge.

Most notable is the tangential behaviour of the dispersion curve at the band edge, i. e. the group velocity of the dispersion curve converges to the slope of the band edge, see Figure 6.21. This surprising behaviour raises the question if the numerical results of the DtN method are reliable or if the results

shown in Figure 6.20b are spurious. First we note that the condition number of the DtN matrix $\mathbf{D}^{\pm}$ does not increase at the band edges, which can bee seen from the maximum and minimum eigenvalues of $\mathbf{D}^{\pm}$ presented in Figure 6.14. However, in Section 6.3.3 we pointed out that the results of the Chebyshev interpolation are not reliable if the Chebyshev nodes are too close to the band edge, which is due to the behaviour of the eigenvalue of the propagation matrix $\mathbf{P}^{\pm}$ that has magnitude strictly less than one in the band gap and magnitude equal to one in the essential spectrum, see Figure 6.11. As argued in Section 6.3.3 this behaviour leads to arbitrarily large derivatives of the propagation operator, and hence, of the DtN operator. In fact, this can also be observed from the numerical results, and it is linked to the increase of the magnitudes of the derivatives of the dispersion curves in the vicinity of the band edge, as presented in Figure 6.22. Nevertheless, as long as the computation of the two close eigenvalues of the propagation matrix $\mathbf{P}^{\pm}$, presented in Figure 6.12, is reliable, which seems to be the case even when using standard procedures such as Matlab's `eig` function, which is — as mentioned already above — an implementation of the generalized Schur decomposition, we can expect that the nonlinear matrix $\mathbf{N}_{C_0}$ of the nonlinear eigenvalue problem (6.58) is also reliable in the vicinity of band edges. Thus, the red dots in Figure 6.20b can expected to be in fact eigenvalues of the nonlinear eigenvalue problem (6.58) and therefore, the tangential behaviour of the dispersion curve at the band edge, as shown in Figure 6.20b, is correct. Finally, let us note that this behaviour cannot be captured by the supercell method due to its prohibitively large modelling error in the vicinity of band edges. This is shown in Figure 6.23, where we compare the results of the adaptive Taylor scheme, that we presented in Figure 6.20, with the result of the supercell method. In Figure 6.23a we can see that the results when using a supercell with $n = 5$ cells on top and bottom of the defect cell C_0 (green) differ in the vicinity of the band edge from the results when using DtN transparent boundary conditions (red). However, at the exact position of the band edge, the difference is again very small. This effect is studied in more detail in Figure 6.23b, where we compare our results of the adaptive Taylor expansion with DtN transparent boundary conditions from Figure 6.20 with supercell results, when using $= 5, 10, 20, 30$ cells on top and bottom of the defect cell C_0. It becomes clear that the position of the dispersion curve entering the essential spectrum is modelled correctly by the supercell method, but the group velocity of the dispersion curve is not. The larger the supercell the smaller is the error in the group velocity, but also for a supercell with $n = 30$ cells on top on bottom of C_0 (black line) the group velocity is far from being identical to the slope of the band edge.



(a) Comparison with the results of the supercell method, when using $n = 5$ unit cells on top and bottom of the defect cell C_0.



(b) Comparison with the results of the supercell method, when using $n = 5, 10, 20, 30$ unit cells on top and bottom of the defect cell C_0.

Figure 6.23: Comparison of the results of the adaptive Taylor scheme with DtN transparent boundary conditions (red), that were presented already in Figure 6.20, with the results of the supercell method (green, cyan, blue, black) in the vicinity of the band edge.

With the adaptive path following of dispersion curves for the problem (6.58) we developed an algorithm that is both, *efficient* and *exact*. Moreover, we can with the help of this algorithm effectively reduce the

influence of local and global Dirichlet eigenvalues. In the previous numerical examples, we showed that the proposed Newton-like method does not converge towards common eigenvalues of (6.58) and the local Dirichlet problems, see Section 6.3.6, and that the convergence towards eigenvalues of (6.58) in the vicinity of global Dirichlet eigenvalues is not possible using the Chebyshev interpolation or limited to small radii of convergence when using the MSLP. If there is a global Dirichlet eigenvalue in the frequency interval I_ω at the selected start value $k^{(0)}$, the Chebyshev interpolation, that we suggested for the simultaneous computation of all eigenvalues of (6.58) in I_ω at $k^{(0)}$, is not applicable. In other words, if the condition number of the matrix $\mathbf{N}_{C_0}(\omega, k^{(0)})$ evaluated at one of the Chebyshev nodes in I_ω exceeds some threshold, e. g. 10^6, we cannot expect the Chebyshev interpolation to deliver reasonable results. In this case we perform a pointwise evaluation of the global signed distance function in I_ω to find appropriate start values for the computation of all eigenvalues of (6.58) in I_ω at $k^{(0)}$ using the Newton method. In the view of the numerical results presented in Figure 6.17 we can expect that the Newton method also converges to eigenvalues of (6.58) even if they are very close to global Dirichlet eigenvalues. If the case of the unlikely event that an eigenvalue of (6.58) at $k^{(0)}$ is simultaneously a global Dirichlet eigenvalue, which implies that convergence towards this of any method is lost, we have to choose a different value for $k^{(0)}$. If during the adaptive algorithm a node turns out to be too close to a local or global Dirichlet eigenvalue, we can simply shift the node by a small amount to restore convergence of the proposed Newton method, or any other iterative scheme.

6.4 Conclusions

In this chapter we showed the high-order FE discretization and numerical solution of the DtN approach, that was presented in [Fli13] for the exact computation of guided modes in PhC waveguides. DtN maps for periodic media are computed by solving local Dirichlet problems and a quadratic eigenvalue problem. Using these DtN maps we transformed the eigenvalue problem for the computation of guided modes, that is posed in an unbounded domain, to a nonlinear eigenvalue problem in a unit cell.

We pointed out that the DtN maps are not well-defined at global Dirichlet eigenvalues and their computation is ill-posed at local Dirichlet eigenvalues, and showed that these Dirichlet eigenvalues lead to ill-conditioned matrices of the nonlinear eigenvalue problem, which implies that — depending on the numerical scheme — convergence towards eigenvalues of the nonlinear eigenvalue problem, that are very close to local or global Dirichlet eigenvalues, can be lost.

We showed that the DtN operators are differentiable with respect to the frequency and the quasi-momentum which is a requirement of nonlinear eigenvalue solvers. Moreover, we discussed the computation of the derivatives of the DtN operators to any order. We also explained the computation of the derivatives of the DtN operators to arbitrary orders.

We applied the iterative Newton method, that we proposed in Chapter 3, to the nonlinear eigenvalue problem and found that it overcomes a problem of other iterative solvers, like the MSLP, that is related to the existence of global Dirichlet eigenvalues. As an alternative to the iterative solvers we applied the Chebyshev interpolation as a direct procedure to solve the nonlinear eigenvalue problem, which proves especially useful in the k-formulation. Numerical examples showed an exponential convergence for p-FEM independent of the confinement of the guided mode, which stands in contrast to the supercell method for which we showed numerically a significant dependence on the confinement of the guided mode.

We extended the theory in Chapter 4 and derived formulas for the group velocity of guided modes and any higher order derivative of the dispersion curves in the case of DtN transparent boundary conditions. We applied these derivatives in an adaptive Taylor expansion of the dispersion curves, which was proposed in Chapter 5. For this we introduced a band edge check that is needed to follow the dispersion curves that leave the band gap. With this adaptive path following of dispersion curves of the problem with DtN transparent boundary conditions we developed an algorithm for PhC waveguide band structure calculations that is both, *efficient* and *exact*. It overcomes the problem of the modelling error introduced by the supercell method, since the DtN transparent boundary conditions model the exterior, periodic domain exactly, and it is time-efficient due to the adaptive selection of nodes for the piecewise Taylor expansion, and only a very little number of nonlinear eigenvalue problems have to be solved. In particular,

we showed that our adaptive scheme is able to resolve the behaviour of the dispersion curves in the vicinity of band edges, which is not possible with the supercell method.

The question that remains is the question of how to overcome the problem of local and global Dirichlet eigenvalues. This is addressed in the following chapter, where we will introduce RtR transparent boundary conditions for periodic media.

7 Robin-to-Robin transparent boundary conditions

RtR transparent boundary conditions for periodic media resolve the problem of global and local Dirichlet eigenvalues, that limit the application of DtN transparent boundary conditions, that we presented in the previous chapter.

We will introduce RtR operators, that are used to truncate the infinite periodic medium of PhC waveguides, and show that the eigenvalue problem (2.19) of finding guided modes in PhC waveguides, that is posed in the infinite strip S, is equivalent to a nonlinear eigenvalue problem with RtR transparent boundary conditions, that is posed in the defect cell C_0. In contrast to the case with DtN transparent boundary conditions as discussed in the previous chapter, this equivalence also holds true at global and local Dirichlet eigenvalues.

The numerical realization of RtR transparent boundary conditions, that we will discuss in this chapter, was first published together with S. Fliss and K. Schmidt [FKS15]. In this work, however, we propose an alternative why for the characterization of the RtR operators. Moreover, we discuss the differentiability of the RtR operators with respect to the frequency and quasi-momentum and elaborate on the computation of the derivatives. This was not published in [FKS15] but it is needed for the numerical solution of the resulting nonlinear eigenvalue problem, in particular when applying the adaptive path following algorithm proposed in Chapter 5.

The outline of this chapter is similar to one of Chapter 6. In Section 7.1 we introduce the RtR operators, discuss their characterization and differentiability, and comment on their FE discretization. In Section 7.2 we present the nonlinear eigenvalue problem, that is equivalent to (2.19), show its FE discretization and comment on its numerical solution, before we present numerical results in Section 7.3 and give concluding remarks in Section 7.4.

7.1 The Robin-to-Robin operators

In this section we define the RtR operators, show there characterization using local cell problems and a quadratic operator equation, and prove their differentiability. Finally, we will elaborate on the discretization of the RtR operators and the local cell problems.

7.1.1 Definition of the Robin-to-Robin operators

The RtR operators are defined through Robin problems in the infinite half-strips $S^{\pm}$. But before we introduce these problems, let us give some introductory remarks on the RtR operators and all other operators that we will introduce later in this section and that map a Robin trace to another Robin trace. We will classify Robin traces in this work by *forward* and *backward*. Note that any Robin trace can be split into a Neumann trace, that has a certain direction, and a Dirichlet trace. We will denote a Robin trace as forward, if its Neumann trace points away from the line defect, and, on the other hand, the Robin trace is called backward, if its Neumann trace points towards the line defect. For example, let $v \in \mathsf{H}^1_{\mathrm{lp}}(\Delta, S, \alpha)$, then $\alpha\,\partial_2 v$ is a forward Robin trace in the infinite half-strip S^+ whereas it is a backward Robin trace in S^-. While the directions of the Neumann traces vary in this work (either forward or backward), the Dirichlet traces are always the same. The Robin traces that we will deal with in this work always take the form $\pm\alpha\,\partial_2 v + \mathrm{i}\rho v$ with some constant $\rho \in \mathbb{R} \setminus \{0\}$.

Now let us come to the Robin problems in the infinite half-strips $S^{\pm}$. For any forward Robin trace

$\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ we seek $u^\pm(\varphi) \equiv u^\pm(\cdot\,;\omega,k,\varphi) \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ such that

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u^\pm(\varphi) - \omega^2\,\beta\,u^\pm(\varphi) = 0 \qquad \text{in } S^\pm, \tag{7.1a}$$

$$(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\varphi) = \varphi \qquad \text{on } \Gamma_0^\pm. \tag{7.1b}$$

The following result, that was proved in [Fli09], is the main advantage of the RtR method compared to the DtN method.

Theorem 7.1. *If $\omega^2 \notin \sigma^\pm(k)$, the half-strip problems (7.1) are well-posed in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$.*

Theorem 7.1 guarantees well-posedness of (7.1) for all frequencies $\omega^2 \notin \sigma^\pm(k)$, i. e. in particular also for global Dirichlet eigenvalues for which the half-strip problem (6.1) with Dirichlet boundary conditions at $\Gamma_0^\pm$ is not well-posed.

Then, for any forward Robin trace $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ on $\Gamma_0^\pm$, the RtR operators $\mathcal{R}^\pm(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm))$ are defined as the backward Robin trace of $u^\pm(\cdot\,;\omega,k,\varphi)$ on $\Gamma_0^\pm$, i. e.

$$\mathcal{R}^\pm(\omega,k)\varphi = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_0^\pm}\,. \tag{7.2}$$

The following result will prove useful for the characterization of the RtR operators.

Lemma 7.2. *Let $\omega^2 \notin \sigma^\pm(k)$. Then the RtR operators $\mathcal{R}^\pm(\omega,k)$ are invertible.*

Proof. Let us introduce the auxiliary half-strip problems: for any backward Robin trace $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ find $\widetilde{u}^\pm(\varphi) \equiv \widetilde{u}^\pm(\cdot\,;\omega,k,\varphi) \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ such that

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})\widetilde{u}^\pm(\varphi) - \omega^2\,\beta\,\widetilde{u}^\pm(\varphi) = 0 \qquad \text{in } S^\pm, \tag{7.3a}$$

$$(\mp\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^\pm(\varphi) = \varphi \qquad \text{on } \Gamma_0^\pm. \tag{7.3b}$$

Like (7.1) these problems are well-posed in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ if $\omega^2 \notin \sigma^\pm(k)$, cf. Theorem 7.1. This implies that

$$u^\pm((\pm\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^\pm(\varphi)) = \widetilde{u}^\pm(\varphi) \tag{7.4a}$$

and

$$\widetilde{u}^\pm((\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\varphi)) = u^\pm(\varphi) \tag{7.4b}$$

for all $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$. Now we introduce

$$\widetilde{\mathcal{R}}^\pm(\omega,k)\varphi = (\pm\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^\pm(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_0^\pm}\,. \tag{7.5}$$

Using (7.5), the definition of the RtR operators (7.2), and the identities (7.4), we conclude that

$$\widetilde{\mathcal{R}}^\pm(\omega,k)\mathcal{R}^\pm(\omega,k)\varphi = (\pm\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^\pm((\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\varphi)\,|_{\Gamma_0^\pm})\,|_{\Gamma_0^\pm} = (\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\varphi)\,|_{\Gamma_0^\pm} = \varphi$$

and

$$\mathcal{R}^\pm(\omega,k)\widetilde{\mathcal{R}}^\pm(\omega,k)\varphi = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm((\pm\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^\pm(\varphi)\,|_{\Gamma_0^\pm})\,|_{\Gamma_0^\pm} = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^\pm(\varphi)\,|_{\Gamma_0^\pm} = \varphi$$

for all $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$, which finishes the proof. □

7.1.2 Characterization of the Robin-to-Robin operators

In this subsection we explain how we can compute the RtR operators using local cell problems and a quadratic operator equation.

First, we note that the infinite strips $S^\pm$ on top and bottom of the guide can be expressed as union of an infinite number of periodicity cells $C_n^\pm$, $n \in \mathbb{N}$, i. e. $S^\pm = \bigcup_{n=1}^\infty (C_n^\pm \cup \Gamma_n^\pm)$, cf. Figure 2.3b. The top and bottom boundaries of these cells $C_n^\pm$ shall be denoted by $\Gamma_{n-1}^\pm$ and $\Gamma_n^\pm$, i. e. $\Gamma_0^\pm = \overline{C_0} \cap \overline{C_1^\pm}$ and

$\Gamma_n^\pm = \overline{C_n^\pm} \cap \overline{C_{n+1}^\pm}$ for $n \geq 1$. We also note that — due to the periodicity and the infinity of the half strips — all cells $C_n^\pm$ can be identified by the first cell $C_1^\pm$ and all boundaries $\Gamma_n^\pm$ can be identified by the first boundary $\Gamma_0^\pm$. This implies that we can identify all functions of $C_n^\pm$ by functions of $C_1^\pm$, and, similarly, all functions of $\Gamma_n^\pm$ by functions of $\Gamma_0^\pm$. Analogously to Chapter 6 we introduce shift operators $\mathcal{S}_n^\pm \in \boldsymbol{\mathcal{L}}(\mathsf{C}^\infty(\Gamma_0^\pm), \mathsf{C}^\infty(\Gamma_n^\pm))$, $n \in \mathbb{N}$, defined by

$$\mathcal{S}_n^\pm \varphi(\mathbf{x}) = \varphi(\mathbf{x} \mp n\mathbf{a}_2^\pm) \tag{7.6}$$

By a density argument of $\mathsf{C}^\infty(\Gamma_n^\pm)$ in $\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_n^\pm)$ and $\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_n^\pm)$, respectively, we can extend the shift operators $\mathcal{S}_n^\pm$ to functions in $\mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_n^\pm)$ and $\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_n^\pm)$. For simplicity of notation we shall write $\mathcal{S}^\pm := \mathcal{S}_1^\pm$. Furthermore, we introduce the inverse $(\mathcal{S}^\pm)^{-1}$ of $\mathcal{S}^\pm$ which is simply given by

$$(\mathcal{S}^\pm)^{-1}\varphi(\mathbf{x}) = \varphi(\mathbf{x} \pm \mathbf{a}_2^\pm). \tag{7.7}$$

We start by introducing two propagation operators.

- The forward-forward propagation operator $\mathcal{P}_{\mathrm{ff}}^\pm(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm))$, defined by

  $$\mathcal{P}_{\mathrm{ff}}^\pm(\omega, k)\varphi = (\mathcal{S}^\pm)^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\cdot\,; \omega, k, \varphi)\,|_{\Gamma_1^\pm},$$

  maps the forward Robin trace $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ on $\Gamma_0^\pm$ to the forward Robin trace of the infinite half-strip solution $u^\pm(\varphi)$ of (7.1) on $\Gamma_1^\pm$ shifted to $\Gamma_0^\pm$. As argued in [FKS15], this operator is compact, injective and its spectral radius is strictly less than one.

- The forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^\pm(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm))$, defined by

  $$\mathcal{P}_{\mathrm{fb}}^\pm(\omega, k)\varphi = (\mathcal{S}^\pm)^{-1}(\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm(\cdot\,; \omega, k, \varphi)\,|_{\Gamma_1^\pm},$$

  maps the forward Robin trace $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ on $\Gamma_0^\pm$ to the backward Robin trace of the infinite half-strip solution $u^\pm(\varphi)$ on $\Gamma_1^\pm$ shifted to $\Gamma_0^\pm$.

Now we define local cell problems: for any forward Robin trace $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ on $\Gamma_0^\pm$ and any backward Robin trace $\psi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_1^\pm)$ on $\Gamma_1^\pm$ find $u_{\mathrm{loc}}^\pm(\varphi, \psi) \equiv u_{\mathrm{loc}}^\pm(\cdot\,; \omega, k, \varphi, \psi) \in \mathsf{H}_{1\mathrm{p}}^1(\Delta, C_1^\pm, \alpha)$ as solution of

$$-(\nabla + \mathrm{i}k_0^{(1)}) \cdot \alpha(\nabla + \mathrm{i}k_0^{(1)})u_{\mathrm{loc}}^\pm(\varphi, \psi) - \omega^2\beta\, u_{\mathrm{loc}}^\pm(\varphi, \psi) = 0 \qquad \text{in } C_1^\pm, \tag{7.8a}$$

$$(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^\pm(\varphi, \psi) = \varphi \qquad \text{on } \Gamma_0^\pm, \tag{7.8b}$$

$$(\mp\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^\pm(\varphi, \psi) = \psi \qquad \text{on } \Gamma_1^\pm. \tag{7.8c}$$

These local cell problems are well-posed for all $(\omega^2, k) \in \mathbb{R}^+ \times B$. The corresponding Dirichlet cell problems (6.8), however, that we used in Chapter 6 to characterize the DtN operators, are only well-posed if we exclude for each $k \in B$ the countable set of local Dirichlet eigenvalues, i. e. eigenvalues of (6.8a) with homogeneous homogeneous Dirichlet boundary conditions at $\Gamma_0^\pm$ and $\Gamma_1^\pm$.

With the solutions of the local cell problems (7.8) we define

- the local forward-backward RtR operator

  $$\mathcal{T}_{\mathrm{fb}}^\pm(\omega, k)\varphi = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^\pm(\varphi, 0)\,|_{\Gamma_0^\pm}, \tag{7.9a}$$

  which maps the forward Robin trace φ on $\Gamma_0^\pm$ to the backward Robin trace of the local cell solution $u_{\mathrm{loc}}^\pm(\varphi, 0)$ on $\Gamma_0^\pm$,

- the local forward-forward RtR operator

  $$\mathcal{T}_{\mathrm{ff}}^\pm(\omega, k)\varphi = (\mathcal{S}^\pm)^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^\pm(\varphi, 0)\,|_{\Gamma_1^\pm}, \tag{7.9b}$$

  which maps the forward Robin trace φ on $\Gamma_0^\pm$ to the forward Robin trace of the local cell solution $u_{\mathrm{loc}}^\pm(\varphi, 0)$ on $\Gamma_1^\pm$ shifted to $\Gamma_0^\pm$,

- the local backward-backward RtR operator

$$\mathcal{T}^{\pm}_{\mathrm{bb}}(\omega,k)\varphi = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(0,\mathcal{S}^{\pm}\varphi)\,|_{\Gamma^{\pm}_0}, \tag{7.9c}$$

  which maps the backward Robin trace φ on $\Gamma^{\pm}_0$ to the backward Robin trace of the local cell solution $u^{\pm}_{\mathrm{loc}}(0,\mathcal{S}^{\pm}\varphi)$ on $\Gamma^{\pm}_0$, and

- the local backward-forward RtR operator

$$\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega,k)\varphi = (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(0,\mathcal{S}^{\pm}\varphi)\,|_{\Gamma^{\pm}_1}, \tag{7.9d}$$

  which maps the backward Robin trace φ on $\Gamma^{\pm}_0$ to the forward Robin trace of the local cell solution $u^{\pm}_{\mathrm{loc}}(0,\mathcal{S}^{\pm}\varphi)$ on $\Gamma^{\pm}_1$ shifted to $\Gamma^{\pm}_0$.

With the help of the local cell solutions and the propagation operators $\mathcal{P}^{\pm}_{\mathrm{ff}}$ and $\mathcal{P}^{\pm}_{\mathrm{fb}}$ we can express the infinite half strip solution $u^{\pm}(\varphi)$ in the cell $C^{\pm}_n$, $n\in\mathbb{N}$, as

$$\begin{aligned} u^{\pm}(\varphi)\,|_{C^{\pm}_n} &= u^{\pm}_{\mathrm{loc}}((\mathcal{P}^{\pm}_{\mathrm{ff}})^{n-1}\varphi, \mathcal{S}^{\pm}\mathcal{P}^{\pm}_{\mathrm{fb}}(\mathcal{P}^{\pm}_{\mathrm{ff}})^{n-1}\varphi) \\ &= u^{\pm}_{\mathrm{loc}}((\mathcal{P}^{\pm}_{\mathrm{ff}})^{n-1}\varphi, 0) + u^{\pm}_{\mathrm{loc}}(0, \mathcal{S}^{\pm}\mathcal{P}^{\pm}_{\mathrm{fb}}(\mathcal{P}^{\pm}_{\mathrm{ff}})^{n-1}\varphi), \end{aligned} \tag{7.10}$$

since the solutions of the local cell problems (7.8) are linear in the data (φ,ψ). Evaluating the forward Robin trace of the infinite half-strip solution $u^{\pm}(\varphi)$ on $\Gamma^{\pm}_1$ using Eq. (7.10), we obtain an equation for the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}$ in terms of the local RtR operators $\mathcal{T}^{\pm}_{\mathrm{ff}}$ and $\mathcal{T}^{\pm}_{\mathrm{bf}}$, and the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}$

$$\begin{aligned} \mathcal{P}^{\pm}_{\mathrm{ff}}(\omega,k)\varphi &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}(\varphi)\,|_{\Gamma^{\pm}_1} \\ &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(\varphi,0)\,|_{\Gamma^{\pm}_1} + (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(0,\mathcal{S}^{\pm}\mathcal{P}^{\pm}_{\mathrm{fb}}\varphi)\,|_{\Gamma^{\pm}_1} \\ &= \mathcal{T}^{\pm}_{\mathrm{ff}}\varphi + \mathcal{T}^{\pm}_{\mathrm{bf}}\mathcal{P}^{\pm}_{\mathrm{fb}}\varphi. \end{aligned} \tag{7.11}$$

On the other hand, identifying the backward Robin trace of the infinite half-strip solution $u^{\pm}(\varphi)$ on $\Gamma^{\pm}_1$ by the backward Robin trace of the infinite half-strip solution $u^{\pm}(\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi)$ on $\Gamma^{\pm}_0$, and evaluating this trace using Eq. (7.10), we obtain an equation for the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}$ in terms of the local RtR operators $\mathcal{T}^{\pm}_{\mathrm{fb}}$ and $\mathcal{T}^{\pm}_{\mathrm{bb}}$, and the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}$

$$\begin{aligned} \mathcal{P}^{\pm}_{\mathrm{fb}}(\omega,k)\varphi &= (\mathcal{S}^{\pm})^{-1}(\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}(\varphi)\,|_{\Gamma^{\pm}_1} \\ &= (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}(\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi)\,|_{\Gamma^{\pm}_0} \\ &= (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi,0)\,|_{\Gamma^{\pm}_0} + (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(0,\mathcal{S}^{\pm}\mathcal{P}^{\pm}_{\mathrm{fb}}\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi)\,|_{\Gamma^{\pm}_0} \\ &= \mathcal{T}^{\pm}_{\mathrm{fb}}\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi + \mathcal{T}^{\pm}_{\mathrm{bb}}\mathcal{P}^{\pm}_{\mathrm{fb}}\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi. \end{aligned} \tag{7.12}$$

Using the local RtR operators $\mathcal{T}^{\pm}_{\mathrm{fb}}(\omega,k)$ and $\mathcal{T}^{\pm}_{\mathrm{bb}}(\omega,k)$, as well as the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega,k)$, we can characterize the RtR operator $\mathcal{R}^{\pm}(\omega,k)$ defined in (7.2), which maps a forward Robin trace on $\Gamma^{\pm}_0$ to a backward Robin trace on $\Gamma^{\pm}_0$, by

$$\mathcal{R}^{\pm}(\omega,k) = \mathcal{T}^{\pm}_{\mathrm{fb}}(\omega,k) + \mathcal{T}^{\pm}_{\mathrm{bb}}(\omega,k)\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega,k). \tag{7.13}$$

Now let us come to the problem of computing the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega,k)$. In [FKS15] we argued that the local backward-forward RtR operator $\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega,k)$ is invertible for all $(\omega^2,k)\in\mathbb{R}^+\times B$, and its inverse can be computed with the help of the auxiliary local cell problem: for any forward Robin traces $\varphi\in\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$ on $\Gamma^{\pm}_0$ and $\psi\in\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_1)$ on $\Gamma^{\pm}_1$ find $\widetilde{u}^{\pm}_{\mathrm{loc}}(\varphi,\psi)\equiv\widetilde{u}^{\pm}_{\mathrm{loc}}(\cdot\,;\omega,k,\varphi,\psi)\in\mathsf{H}^1_{1\mathrm{p}}(\Delta,C^{\pm}_1,\alpha)$ as solution of

$$-(\nabla+\mathrm{i}k^{(1)}_0))\cdot\alpha(\nabla+\mathrm{i}k^{(1)}_0))\widetilde{u}^{\pm}_{\mathrm{loc}}(\varphi,\psi) - \omega^2\beta\,\widetilde{u}^{\pm}_{\mathrm{loc}}(\varphi,\psi) = 0 \qquad \text{in } C^{\pm}_1, \tag{7.14a}$$

$$(\pm\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^{\pm}_{\mathrm{loc}}(\varphi,\psi) = \varphi \qquad \text{on } \Gamma^{\pm}_0, \tag{7.14b}$$

$$(\pm\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^{\pm}_{\mathrm{loc}}(\varphi,\psi) = \psi \qquad \text{on } \Gamma^{\pm}_1. \tag{7.14c}$$

However, this auxiliary local cell problem is not well-posed for all $(\omega^2, k) \in \mathbb{R}^+ \times B$, and hence, the local backward-forward RtR operator $\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)$ is not invertible for all $(\omega^2, k) \in \mathbb{R}^+ \times B$. Thus, we cannot proceed as presented in [FKS15] and express the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$ for all $(\omega^2, k) \in \mathbb{R}^+ \times B$ in terms of the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega, k)$ by simply applying the inverse of $\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)$ to Eq. (7.11), and then using this identity together with Eq. (7.12) to obtain a quadratic operator equation for the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega, k)$.

Instead, we directly insert (7.11) into (7.12) and obtain the quadratic operator equation

$$\mathcal{T}^{\pm}_{\mathrm{bb}}\mathcal{P}^{\pm}_{\mathrm{fb}}\mathcal{T}^{\pm}_{\mathrm{bf}}\mathcal{P}^{\pm}_{\mathrm{fb}} + \mathcal{T}^{\pm}_{\mathrm{bb}}\mathcal{P}^{\pm}_{\mathrm{fb}}\mathcal{T}^{\pm}_{\mathrm{ff}} + \left(\mathcal{T}^{\pm}_{\mathrm{fb}}\mathcal{T}^{\pm}_{\mathrm{bf}} - \mathcal{I}\right)\mathcal{P}^{\pm}_{\mathrm{fb}} + \mathcal{T}^{\pm}_{\mathrm{fb}}\mathcal{T}^{\pm}_{\mathrm{ff}} = 0 \tag{7.15}$$

for the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$, where we omitted — for simplicity of presentation — the (ω, k)-dependence of the operators.

This quadratic operator equation does not uniquely define the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$. A characterization of $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$ using (7.15) is particularly difficult since the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$ — in contrast to the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega, k)$ — does not necessarily have spectral radius strictly less than one. Therefore, it will prove useful, to additionally present the characterization of the RtR operators using the quadratic operator equation for the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega, k)$, which is — as argued above — only valid if (7.14) is well-posed.

Lemma 7.3. *Let the auxiliary local cell problem (7.14) be well-posed. Then local RtR operator $\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)$ is invertible.*

Proof. Let $\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)$ be defined for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$ by

$$\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)\varphi = (\mathcal{S}^{\pm})^{-1}(\mp\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^{\pm}_{\mathrm{loc}}(0, \mathcal{S}^{\pm}\varphi)\,|_{\Gamma^{\pm}_1}\,. \tag{7.16}$$

Note that $\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)$ is well-defined, since by assumption the solution $\widetilde{u}^{\pm}_{\mathrm{loc}}(\cdot\,; \omega, k, 0, \mathcal{S}^{\pm}\varphi)$ of the auxiliary local cell problem (7.14) exists and is unique. Since the usual local cell problem (7.8) is also well-posed, we can deduce that

$$\widetilde{u}^{\pm}_{\mathrm{loc}}(0, (\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(0, \mathcal{S}^{\pm}\varphi)\,|_{\Gamma^{\pm}_1}) = u^{\pm}_{\mathrm{loc}}(0, \mathcal{S}^{\pm}\varphi)$$

for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$. Using the definition (7.9d) of the local backward-forward RtR operator $\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)$, we deduce that

$$(\mathcal{S}^{\pm})^{-1}(\mp\alpha\,\partial_2 + \mathrm{i}\rho)\widetilde{u}^{\pm}_{\mathrm{loc}}(0, \mathcal{S}^{\pm}\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)\varphi)\,|_{\Gamma^{\pm}_1} = \varphi$$

for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$, which implies by definition of $\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)$ that

$$\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)\varphi = \varphi$$

for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$. On the other hand, we can also show that

$$(\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{loc}}(0, \mathcal{S}^{\pm}\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)\varphi)\,|_{\Gamma^{\pm}_1} = \varphi$$

for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$, i. e.

$$\mathcal{T}^{\pm}_{\mathrm{bf}}(\omega, k)\widetilde{\mathcal{T}}^{\pm}_{\mathrm{fb}}(\omega, k)\varphi = \varphi$$

for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$, which finishes the proof. □

Let the auxiliary local cell problem (7.14) be well-posed. Using Lemma 7.3 we can then rewrite (7.11) in the form

$$\mathcal{P}^{\pm}_{\mathrm{fb}}\varphi = \left(\mathcal{T}^{\pm}_{\mathrm{bf}}\right)^{-1}\left(\mathcal{P}^{\pm}_{\mathrm{ff}}\varphi - \mathcal{T}^{\pm}_{\mathrm{ff}}\varphi\right). \tag{7.17}$$

Inserting this equality into Eq. (7.12) yields a quadratic operator equation, the so-called *Riccati equation*,

$$\mathcal{T}^{\pm}_{\mathrm{bb}}\left(\mathcal{T}^{\pm}_{\mathrm{bf}}\right)^{-1}\left(\mathcal{P}^{\pm}_{\mathrm{ff}}\right)^2 + \left(\mathcal{T}^{\pm}_{\mathrm{fb}} - \left(\mathcal{T}^{\pm}_{\mathrm{bf}}\right)^{-1} - \mathcal{T}^{\pm}_{\mathrm{bb}}\left(\mathcal{T}^{\pm}_{\mathrm{bf}}\right)^{-1}\mathcal{T}^{\pm}_{\mathrm{ff}}\right)\mathcal{P}^{\pm}_{\mathrm{ff}} + \left(\mathcal{T}^{\pm}_{\mathrm{bf}}\right)^{-1}\mathcal{T}^{\pm}_{\mathrm{ff}} = 0. \tag{7.18}$$

Proposition 7.4. *Let $\omega^2 \notin \sigma^{\pm}(k)$ and let the auxiliary local cell problem (7.14) be well-posed. Then the forward-forward propagation operator $\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k)$ is the unique solution of the Riccati equation (7.18) with spectral radius strictly less than one.*

Proof. We showed already that $\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k)$ is solution of the Riccati equation (7.18). To show that it is the unique solution, we use the same ideas as in [JLF06, Fli09] and suppose that $\widetilde{\mathcal{P}}_{\mathrm{ff}}^{\pm}$ is also a solution. Let us introduce

$$\widetilde{\mathcal{P}}_{\mathrm{fb}}^{\pm} = \left(\mathcal{T}_{\mathrm{bf}}^{\pm}\right)^{-1}\left(\widetilde{\mathcal{P}}_{\mathrm{ff}}^{\pm} - \mathcal{T}_{\mathrm{ff}}^{\pm}\right)$$

and define for all $\varphi \in \mathsf{H}_{\mathrm{lp}}^{-1/2}(\Gamma_0^{\pm})$

$$v^{\pm}(\varphi)|_{C_n^{\pm}} = u_{\mathrm{loc}}^{\pm}((\widetilde{\mathcal{P}}_{\mathrm{ff}}^{\pm})^{n-1}\varphi, \mathcal{S}^{\pm}\widetilde{\mathcal{P}}_{\mathrm{fb}}^{\pm}(\widetilde{\mathcal{P}}_{\mathrm{ff}}^{\pm})^{n-1}\varphi).$$

We can see easily that $v^{\pm}(\varphi)$ satisfies the boundary condition (7.1b) and is solution of (7.1a) in each cell $C_n^{\pm}$. We can also show the continuity of the forward and the backward traces on each $\Gamma_n^{\pm}$ because $\widetilde{\mathcal{P}}_{\mathrm{ff}}^{\pm}$ is solution of (7.18) and by definition of $\widetilde{\mathcal{P}}_{\mathrm{fb}}^{\pm}$. Finally, $v^{\pm}(\varphi) \in \mathsf{L}^2(S^{\pm})$ because the spectral radius of $\widetilde{\mathcal{P}}_{\mathrm{fb}}^{\pm}$ is strictly less than one. Due to well-posedness of (7.1), $v^{\pm}(\varphi)$ is necessarily equal to $u^{\pm}(\varphi)$ for each φ, and in particular their traces on $\Gamma_1^{\pm}$ coincide. Hence, the operator $\widetilde{\mathcal{P}}_{\mathrm{ff}}^{\pm}$ is identical to $\mathcal{P}_{\mathrm{ff}}^{\pm}$. □

With the help of Proposition 7.4 we can deduce a similar result for the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$.

Corollary 7.5. *Let $\omega^2 \notin \sigma^{\pm}(k)$ and let the auxiliary local cell problem (7.14) be well-posed. Then the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ is the unique solution of the quadratic operator equation (7.15) such that*

$$\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k) = \mathcal{T}_{\mathrm{ff}}^{\pm}(\omega,k) + \mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$$

has spectral radius strictly less than one.

Proof. By construction of the quadratic operator equation (7.15), it is clear that $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ is a solution. On the other hand, uniqueness directly follows from Proposition 7.4 and the fact that the mapping

$$\mathcal{X} \longmapsto \mathcal{T}_{\mathrm{ff}}^{\pm}(\omega,k) + \mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)\mathcal{X},$$

that is needed to compute the forward-forward propagation operator $\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k)$ from the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$, cf. Eq. (7.11), is injective if and only if the auxiliary local cell problem (7.14) is well-posed and hence, the local backward-forward RtR operator $\mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)$ is invertible. □

Thanks to Corollary 7.5 the unique characterization of the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ is now clear as long as the auxiliary local cell problem (7.14) is well-posed. In this case, we can also compute the forward-forward propagation operator $\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k)$, which is uniquely characterized by the Riccati equation (7.18) due to Proposition 7.4, and then employ (7.17) for the unique computation of the forward-backward operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$.

If, on the other hand, the auxiliary local cell problem (7.14) is not well-posed, and hence, the local backward-forward RtR operator $\mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)$ is not invertible, a unique characterization of the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ using the quadratic operator equation (7.15) is still possible. However, we cannot argue like in Corollary 7.5 since the local backward-forward RtR operator $\mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)$ is not injective. Instead we shall use the RtR operator $\mathcal{R}^{\pm}(\omega,k)$ to map between the propagation operators. For this, we insert (7.13) into (7.12), which gives

$$\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k) = \mathcal{R}^{\pm}(\omega,k)\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k). \tag{7.19}$$

Since the RtR operator is invertible, see Lemma 7.2, we can prove the following result.

Proposition 7.6. *Let $\omega^2 \notin \sigma^{\pm}(k)$. Then the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ is the unique solution of the quadratic operator equation (7.15) such that*

$$\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k) = \left(\mathcal{R}^{\pm}(\omega,k)\right)^{-1}\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k) \tag{7.20}$$

has spectral radius strictly less than one.

Proof. The fact that the forward-backward propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega,k)$ is a solution of (7.15) is clear from the construction of (7.15). To show uniqueness, we can proceed exactly like in the proof of Proposition 7.4, additionally taking into account that the forward-forward propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega,k)$ is uniquely defined by Eq. (7.20) since the RtR operator is invertible for all $\omega^2 \notin \sigma^{\pm}(k)$, see Lemma 7.2. □

Now we are able to uniquely characterize the RtR operator $\mathcal{R}^{\pm}(\omega,k)$ using (7.13) no matter if the auxiliary local cell problem (7.14) is well-posed.

However, note that the auxiliary local cell problem (7.14) can be chosen to be well-posed for specific $(\omega^2,k) \in \mathbb{R}^+ \times B$ by carefully selecting the constant $\rho \in \mathbb{R}\setminus\{0\}$. It has to be mentioned, that for the sake of differentiability of the operators, which is a crucial prerequisite of all numerical schemes for the solution of the resulting nonlinear eigenvalue problem, the coefficient ρ has to be differentiable with respect to ω and k. However, allowing ρ to vary smoothly, will yield significantly more complicated formulas for the derivatives of the operators. This explains why it is beneficial to assume that ρ is in fact a constant for all $(\omega^2,k) \in \mathbb{R}^+ \times B$. Nevertheless, it might be possible to select this constant such that the auxiliary local cell problem (7.14) is well-posed for all (ω^2,k) in a subdomain of $\mathbb{R}^+ \times B$, that is of interest in the numerical computation. This will be demonstrated later in the numerical experiments presented in Section 7.3.

It remains to comment on the numerical solution of the quadratic operator equations (7.15) and (7.18). For the numerical solution of the discrete version of the Riccati equation (7.18) we shall propose later in Section 7.1.5 an eigendecomposition, that we already proposed in Chapter 6 for the computation of the Dirichlet propagation operators involved in the characterization of DtN operators. For the numerical solution of the discrete version of the quadratic operator equation (7.15), on the other hand, we will propose in Section 7.1.5 a Newton method similar to the procedure described in [JLF06] for the computation of the Dirichlet propagation operators, that implicitly takes the condition on the spectral radius of $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega,k)$ into account.

Finally, let us come to an important result on the relation of the proposed RtR approach and the DtN approach as presented in Chapter 6. To this end, we recall the Dirichlet problems (6.1) in the infinite half-strips $S^{\pm}$. In order to distinguish between the solutions $u^{\pm}(\varphi)$ of the Robin problems (7.1) in the infinite half-strips $S^{\pm}$ and the solutions of the Dirichlet problems (6.1), we shall denote the latter by $u^{\pm}_{\mathrm{DtN}}(\varphi_{\mathrm{DtN}}) \equiv u^{\pm}_{\mathrm{DtN}}(\cdot\,;\omega,k,\varphi_{\mathrm{DtN}}) \in \mathsf{H}^1_{1\mathrm{p}}(\Delta,S^{\pm},\alpha)$ with some Dirichlet trace $\varphi_{\mathrm{DtN}} \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$. Recall that this Dirichlet problem is only well-posed for $(\omega^2,k) \in \mathbb{R}^+ \times B$ with $\omega^2 \notin \sigma^{\pm}(k)$ except a countable set of frequencies — the global Dirichlet eigenvalues, i. e. eigenvalues of (7.1a) with homogeneous Dirichlet boundary condition at $\Gamma^{\pm}_0$. Furthermore, let $\mathcal{P}^{\pm}_{\mathrm{DtN}}(\omega,k) \in \boldsymbol{\mathcal{L}}(\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0))$ denote the Dirichlet propagation operator (6.6) of the DtN approach, i. e. for $\varphi_{\mathrm{DtN}} \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$ we define $\mathcal{P}^{\pm}_{\mathrm{DtN}}\varphi_{\mathrm{DtN}} = (\mathcal{S}^{\pm})^{-1}u^{\pm}_{\mathrm{DtN}}(\varphi_{\mathrm{DtN}})\,|_{\Gamma^{\pm}_1}$. Then we can show the following result.

Proposition 7.7. *Let $(\omega^2,k) \in \mathbb{R}^+ \times B$ with $\omega^2 \notin \sigma^{\pm}(k)$ and let the Dirichlet problems (6.1) on the infinite half-strips $S^{\pm}$ be well-posed. Then the following holds true: If $\mu^{\pm}_{\mathrm{DtN}} \in \mathbb{C}$ is an eigenvalue of $\mathcal{P}^{\pm}_{\mathrm{DtN}}(\omega,k)$ with associated eigenfunction $\varphi^{\pm}_{\mathrm{DtN}} \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$, i. e. $\mathcal{P}^{\pm}_{\mathrm{DtN}}(\omega,k)\varphi^{\pm}_{\mathrm{DtN}} = \mu^{\pm}_{\mathrm{DtN}}\varphi^{\pm}_{\mathrm{DtN}}$, then $\mu^{\pm}_{\mathrm{DtN}}$ is also an eigenvalue of the forward-forward RtR propagation operator $\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega,k)$ with associated eigenfunction*

$$\varphi^{\pm} = \pm\alpha\,\partial_2 u^{\pm}_{\mathrm{DtN}}(\varphi^{\pm}_{\mathrm{DtN}})\,|_{\Gamma^{\pm}_0} + \mathrm{i}\rho\varphi^{\pm}_{\mathrm{DtN}} \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0).$$

Proof. Let $\mu^{\pm}_{\mathrm{DtN}} \in \mathbb{C}$ be an eigenvalue of $\mathcal{P}^{\pm}_{\mathrm{DtN}}(\omega,k)$ with associated eigenfunction $\varphi^{\pm}_{\mathrm{DtN}} \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$. Then $u^{\pm}_{\mathrm{DtN}}(\varphi^{\pm}_{\mathrm{DtN}})$ solves the Robin problem (7.1) with $\varphi^{\pm} = \varphi^{\pm}_{\mathrm{RtR}} := \pm\alpha\,\partial_2 u^{\pm}_{\mathrm{DtN}}(\varphi^{\pm}_{\mathrm{DtN}})\,|_{\Gamma^{\pm}_0} + \mathrm{i}\rho\varphi^{\pm}_{\mathrm{DtN}}$. But this implies that $u^{\pm}(\varphi^{\pm}_{\mathrm{RtR}}) \equiv u^{\pm}_{\mathrm{DtN}}(\varphi^{\pm}_{\mathrm{DtN}})$ and hence,

$$\begin{aligned}
\mathcal{P}^{\pm}_{\mathrm{ff}}(\omega,k)\varphi^{\pm}_{\mathrm{RtR}} &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}(\varphi^{\pm}_{\mathrm{RtR}})\,|_{\Gamma^{\pm}_1} \\
&= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{DtN}}(\varphi^{\pm}_{\mathrm{DtN}})\,|_{\Gamma^{\pm}_1} \\
&= \mu^{\pm}_{\mathrm{DtN}}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}_{\mathrm{DtN}}(\varphi^{\pm}_{\mathrm{DtN}})\,|_{\Gamma^{\pm}_0} \\
&= \mu^{\pm}_{\mathrm{DtN}}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^{\pm}(\varphi^{\pm}_{\mathrm{RtR}})\,|_{\Gamma^{\pm}_0} \\
&= \mu^{\pm}_{\mathrm{DtN}}\varphi^{\pm}_{\mathrm{RtR}},
\end{aligned}$$

which finishes the proof. □

7.1.3 Derivatives of the Robin-to-Robin operators

In this part of the section on the RtR operators we will show their differentiability with respect to the frequency ω and the quasi-momentum k, and present the characterization of the derivatives. The derivatives are needed in Section 7.2.6 where we introduce numerical techniques to solve a nonlinear eigenvalue problem with RtR operators, that we will introduce later in Section 7.2.

Recall that we assume the coefficient ρ to be constant. This implies, in particular, that ρ does not dependent on ω and k.

Differentiability of the Robin-to-Robin operators

Let $u^\pm(\varphi) \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ be the unique solution of the Robin problem (7.1) with forward Robin trace $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ on $\Gamma_0^\pm$. Then we introduce $u^\pm_\omega(\varphi) \equiv u^\pm_\omega(\cdot\,;\omega,k,\varphi)$ as the unique solution in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ of

$$-(\nabla + \mathrm{i}k^{(1)}_0)\cdot\alpha(\nabla + \mathrm{i}k^{(1)}_0)u^\pm_\omega(\varphi) - \omega^2\beta\, u^\pm_\omega(\varphi) = 2\omega\beta u^\pm(\varphi) \qquad \text{in } S^\pm, \tag{7.21a}$$

$$(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm_\omega(\varphi) = 0 \qquad \text{on } \Gamma_0^\pm, \tag{7.21b}$$

and $u^\pm_k(\varphi) \equiv u^\pm_k(\cdot\,;\omega,k,\varphi)$ as the unique solution in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, S^\pm, \alpha)$ of

$$-(\nabla + \mathrm{i}k^{(1)}_0)\cdot\alpha(\nabla + \mathrm{i}k^{(1)}_0)u^\pm_k(\varphi) - \omega^2\beta\, u^\pm_k(\varphi) = (2\alpha\,(-k + \mathrm{i}\partial_1) + \mathrm{i}\partial_1\alpha)\, u^\pm(\varphi) \qquad \text{in } S^\pm, \tag{7.22a}$$

$$(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm_\omega(\varphi) = 0 \qquad \text{on } \Gamma_0^\pm. \tag{7.22b}$$

Following exactly the same argumentation as in the proofs of Proposition 6.7 and Theorem 6.8 we deduce the two following results.

Proposition 7.8. *Let $\omega^2 \notin \sigma^\pm(k)$. Then the source problems (7.21) and (7.22) are well-posed.*

Theorem 7.9. *Let $\omega^2 \notin \sigma^\pm(k)$. Then for any $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$ the unique solution $u^\pm(\cdot\,;\omega,k,\varphi)$ of the infinite half-strip problem (7.1) is Fréchet-differentiable with respect to ω and k, and*

$$\frac{\partial u^\pm(\cdot\,;\omega,k,\varphi)}{\partial\omega} = u^\pm_\omega(\cdot\,;\omega,k,\varphi) \qquad \textit{and} \qquad \frac{\partial u^\pm(\cdot\,;\omega,k,\varphi)}{\partial k} = u^\pm_k(\cdot\,;\omega,k,\varphi).$$

Using the definition (7.2) of the RtR operators $\mathcal{R}^\pm(\omega,k)$, we deduce their Fréchet-differentiability with respect to ω and k.

Corollary 7.10. *Suppose that $\omega^2 \notin \sigma^\pm(k)$. Then the RtR operators $\mathcal{R}^\pm(\omega,k)$ are differentiable with respect to the frequency ω and the quasi-momentum k, and for all $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^\pm)$*

$$\frac{\partial\mathcal{R}^\pm}{\partial\omega}(\omega,k)\varphi = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm_\omega(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_0^\pm} \tag{7.23a}$$

and

$$\frac{\partial\mathcal{R}^\pm}{\partial k}(\omega,k)\varphi = (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u^\pm_k(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_0^\pm}\,. \tag{7.23b}$$

Remark 7.11. *Iteratively repeating the above procedure, we find that the RtR operators $\mathcal{R}^\pm(\omega,k)$ are differentiable to any order with respect to the frequency ω and the quasi-momentum k if $\omega^2 \notin \sigma^\pm(k)$.*

Characterization of the derivatives of Robin-to-Robin operators

For the characterization of the derivatives (7.23) of the RtR operators we employ the same concepts as in Section 7.1.2 for the characterization of the RtR operators. First we will show that the forward-forward propagation operators $\mathcal{P}^\pm_{\mathrm{ff}}(\omega,k)$ and the forward-backward propagation operators $\mathcal{P}^\pm_{\mathrm{fb}}(\omega,k)$ are differentiable with respect to ω and k. Then we note that the same is true for the local RtR operators

$\mathcal{T}_{ij}^{\pm}(\omega,k)$, $i,j\in\{\mathrm{f},\mathrm{b}\}$, and present their derivatives with respect to ω and k. Finally, we show how to compute the derivatives of the RtR operators with respect to the frequency and the quasi-momentum.

Analogously to the differentiability of the RtR operators in Corollary 7.10 we obtain the differentiability of the forward-forward propagation operators $\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k)$ and the forward-backward propagation operators $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$.

Corollary 7.12. *Suppose that $\omega^2\notin\sigma^{\pm}(k)$. Then the propagation operators $\mathcal{P}_{\mathrm{ff}}^{\pm}(\omega,k)$ and $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ are differentiable with respect to the frequency ω and the quasi-momentum k, and for all $\varphi\in\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^{\pm})$*

$$\begin{aligned}
\frac{\partial\mathcal{P}_{\mathrm{ff}}^{\pm}}{\partial\omega}(\omega,k)\varphi &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2+\mathrm{i}\rho)u_{\omega}^{\pm}(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_1^{\pm}},\\
\frac{\mathcal{P}_{\mathrm{fb}}^{\pm}}{\partial\omega}(\omega,k)\varphi &= (\mp\alpha\,\partial_2+\mathrm{i}\rho)u_{\omega}^{\pm}(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_1^{\pm}},\\
\frac{\partial\mathcal{P}_{\mathrm{ff}}^{\pm}}{\partial k}(\omega,k)\varphi &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2+\mathrm{i}\rho)u_{k}^{\pm}(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_1^{\pm}},\\
\frac{\mathcal{P}_{\mathrm{fb}}^{\pm}}{\partial k}(\omega,k)\varphi &= (\mp\alpha\,\partial_2+\mathrm{i}\rho)u_{k}^{\pm}(\cdot\,;\omega,k,\varphi)\,|_{\Gamma_1^{\pm}}\,.
\end{aligned}$$

Now it remains to characterize the derivatives of the propagation operators and the derivatives of the RtR operators by means of local cell problems.

First, let us introduce the local cell solutions $u_{\mathrm{loc},\omega}^{\pm}(\varphi,\psi)\equiv u_{\mathrm{loc},\omega}(\cdot\,;\omega,k,\varphi,\psi)$ as the unique solution in $\mathsf{H}_{1\mathrm{p}}^{1}(\Delta,C_1^{\pm},\alpha)$ of

$$-(\nabla+\mathrm{i}k\binom{1}{0})\cdot\alpha(\nabla+\mathrm{i}k\binom{1}{0})u_{\mathrm{loc},\omega}^{\pm}(\varphi,\psi)-\omega^2\beta\,u_{\mathrm{loc},\omega}^{\pm}(\varphi,\psi)=2\omega\beta u_{\mathrm{loc}}^{\pm}(\varphi,\psi)\qquad\text{in } C_1^{\pm},\tag{7.24a}$$

$$(\pm\alpha\,\partial_2+\mathrm{i}\rho)u_{\mathrm{loc},\omega}^{\pm}(\varphi,\psi)=0\qquad\text{on }\Gamma_0^{\pm},\tag{7.24b}$$

$$(\mp\alpha\,\partial_2+\mathrm{i}\rho)u_{\mathrm{loc},\omega}^{\pm}(\varphi,\psi)=0\qquad\text{on }\Gamma_1^{\pm},\tag{7.24c}$$

where $u_{\mathrm{loc}}^{\pm}(\varphi,\psi)\equiv u_{\mathrm{loc}}^{\pm}(\cdot\,;\omega,k,\varphi,\psi)$ are the unique solutions of the local cell problems (7.8). Analogously as above, we can show that the local RtR operators $\mathcal{T}_{\mathrm{fb}}^{\pm}(\omega,k)$, $\mathcal{T}_{\mathrm{ff}}^{\pm}(\omega,k)$, $\mathcal{T}_{\mathrm{bb}}^{\pm}(\omega,k)$ and $\mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)$ are Fréchet-differentiable with respect to ω, and for all $\varphi\in\mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^{\pm})$ we have

$$\begin{aligned}
\frac{\partial\mathcal{T}_{\mathrm{fb}}^{\pm}}{\partial\omega}(\omega,k)\,\varphi &= \qquad(\mp\alpha\,\partial_2+\mathrm{i}\rho)u_{\mathrm{loc},\omega}^{\pm}(\varphi,\quad 0)\,|_{\Gamma_0^{\pm}},\\
\frac{\partial\mathcal{T}_{\mathrm{ff}}^{\pm}}{\partial\omega}(\omega,k)\,\varphi &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2+\mathrm{i}\rho)u_{\mathrm{loc},\omega}^{\pm}(\varphi,\quad 0)\,|_{\Gamma_1^{\pm}},\\
\frac{\partial\mathcal{T}_{\mathrm{bb}}^{\pm}}{\partial\omega}(\omega,k)\,\varphi &= \qquad(\mp\alpha\,\partial_2+\mathrm{i}\rho)u_{\mathrm{loc},\omega}^{\pm}(0,\mathcal{S}^{\pm}\varphi)\,|_{\Gamma_0^{\pm}},\\
\frac{\partial\mathcal{T}_{\mathrm{bf}}^{\pm}}{\partial\omega}(\omega,k)\,\varphi &= (\mathcal{S}^{\pm})^{-1}(\pm\alpha\,\partial_2+\mathrm{i}\rho)u_{\mathrm{loc},\omega}^{\pm}(0,\mathcal{S}^{\pm}\varphi)\,|_{\Gamma_1^{\pm}}\,.
\end{aligned}$$

Differentiating the quadratic operator equation (7.15) for the computation of the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ with respect to ω yields the derivatives of the forward-backward propagation operators $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ as solution of

$$\begin{aligned}
&\mathcal{T}_{\mathrm{bb}}^{\pm}\frac{\partial\mathcal{P}_{\mathrm{fb}}^{\pm}}{\partial\omega}\mathcal{P}_{\mathrm{ff}}^{\pm}+\left(\mathcal{R}^{\pm}\mathcal{T}_{\mathrm{bf}}^{\pm}-\mathcal{I}\right)\frac{\partial\mathcal{P}_{\mathrm{fb}}^{\pm}}{\partial\omega}\\
&\qquad=\frac{\partial\mathcal{T}_{\mathrm{bb}}^{\pm}}{\partial\omega}\mathcal{P}_{\mathrm{fb}}^{\pm}\mathcal{T}_{\mathrm{bf}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm}+\mathcal{T}_{\mathrm{bb}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm}\frac{\partial\mathcal{T}_{\mathrm{bf}}^{\pm}}{\partial\omega}\mathcal{P}_{\mathrm{fb}}^{\pm}+\frac{\partial\mathcal{T}_{\mathrm{bb}}^{\pm}}{\partial\omega}\mathcal{P}_{\mathrm{fb}}^{\pm}\mathcal{T}_{\mathrm{ff}}^{\pm}+\mathcal{T}_{\mathrm{bb}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm}\frac{\partial\mathcal{T}_{\mathrm{ff}}^{\pm}}{\partial\omega}\\
&\qquad\quad+\frac{\partial\mathcal{T}_{\mathrm{fb}}^{\pm}}{\partial\omega}\mathcal{T}_{\mathrm{bf}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm}+\mathcal{T}_{\mathrm{fb}}^{\pm}\frac{\partial\mathcal{T}_{\mathrm{bf}}^{\pm}}{\partial\omega}\mathcal{P}_{\mathrm{fb}}^{\pm}+\frac{\partial\mathcal{T}_{\mathrm{fb}}^{\pm}}{\partial\omega}\mathcal{T}_{\mathrm{ff}}^{\pm}+\mathcal{T}_{\mathrm{fb}}^{\pm}\frac{\partial\mathcal{T}_{\mathrm{ff}}^{\pm}}{\partial\omega},
\end{aligned}\tag{7.25}$$

where we omitted — for simplicity of notation — the (ω,k)-dependence of the operators. The form of this linear operator equation is equivalent to the one of (6.14) for the computation of the derivatives of the Dirichlet propagation operators in Chapter 6. The techniques for solving the linear operator equation (7.25) on a discrete level are hence also equivalent to the DtN case presented in Chapter 6 and will be revisited in Section 7.1.5.

It remains to verify that the derivatives $\frac{\partial \mathcal{P}^{\pm}_{\mathrm{fb}}}{\partial \omega}(\omega, k)$ of the forward-backward propagation operators $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$ are uniquely defined by (7.25). This however, remains an open question, since we have not been able to prove injectivity of the mapping

$$\mathcal{X} \longmapsto \mathcal{T}^{\pm}_{\mathrm{bb}} \mathcal{X} \mathcal{P}^{\pm}_{\mathrm{ff}} + \left(\mathcal{R}^{\pm} \mathcal{T}^{\pm}_{\mathrm{bf}} - \mathcal{I}\right) \mathcal{X}. \tag{7.26}$$

However, numerical evidence shows, that injectivity on a discrete level is guaranteed.

Finally, the derivatives of the RtR operators $\mathcal{R}^{\pm}(\omega, k)$ with respect to ω are obtained by differentiating Eq. (7.2), which yields

$$\frac{\partial \mathcal{R}^{\pm}}{\partial \omega} = \frac{\partial \mathcal{T}^{\pm}_{\mathrm{fb}}}{\partial \omega} + \frac{\partial \mathcal{T}^{\pm}_{\mathrm{bb}}}{\partial \omega} \mathcal{P}^{\pm}_{\mathrm{fb}} + \mathcal{T}^{\pm}_{\mathrm{bb}} \frac{\partial \mathcal{P}^{\pm}_{\mathrm{fb}}}{\partial \omega}. \tag{7.27}$$

The derivatives of the propagation operator $\mathcal{P}^{\pm}_{\mathrm{fb}}(\omega, k)$ and the RtR operator $\mathcal{R}^{\pm}(\omega, k)$ with respect to k are characterized similarly by simply replacing all ω-derivatives in Eqs. (7.25) and (7.27) by k-derivatives. On the other hand, the k-derivatives of the local RtR operators $\mathcal{T}^{\pm}_{ij}(\omega, k)$, $i, j \in \{\mathrm{f}, \mathrm{b}\}$, are for any $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$ given by

$$\begin{aligned}
\frac{\partial \mathcal{T}^{\pm}_{\mathrm{fb}}}{\partial k}(\omega, k)\, \varphi &= \phantom{(\mathcal{S}^{\pm})^{-1}} (\mp\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm}_{\mathrm{loc},k}(\varphi, \quad 0)\big|_{\Gamma^{\pm}_0}, \\
\frac{\partial \mathcal{T}^{\pm}_{\mathrm{ff}}}{\partial k}(\omega, k)\, \varphi &= (\mathcal{S}^{\pm})^{-1} (\pm\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm}_{\mathrm{loc},k}(\varphi, \quad 0)\big|_{\Gamma^{\pm}_1}, \\
\frac{\partial \mathcal{T}^{\pm}_{\mathrm{bb}}}{\partial k}(\omega, k)\, \varphi &= \phantom{(\mathcal{S}^{\pm})^{-1}} (\mp\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm}_{\mathrm{loc},k}(0, \mathcal{S}^{\pm}\varphi)\big|_{\Gamma^{\pm}_0}, \\
\frac{\partial \mathcal{T}^{\pm}_{\mathrm{bf}}}{\partial k}(\omega, k)\, \varphi &= (\mathcal{S}^{\pm})^{-1} (\pm\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm}_{\mathrm{loc},k}(0, \mathcal{S}^{\pm}\varphi)\big|_{\Gamma^{\pm}_1},
\end{aligned}$$

where $u^{\pm}_{\mathrm{loc},k}(\varphi, \psi) \equiv u^{\pm}_{\mathrm{loc},k}(\cdot\,; \omega, k, \varphi, \psi)$ with $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_0)$ and $\psi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma^{\pm}_1)$ are the unique solutions in $\mathsf{H}^1_{1\mathrm{p}}(\Delta, C^{\pm}_1, \alpha)$ of

$$\begin{aligned}
-(\nabla + \mathrm{i}k\tbinom{1}{0}) \cdot \alpha (\nabla + \mathrm{i}k\tbinom{1}{0}) u^{\pm}_{\mathrm{loc},k}(\varphi, \psi) - \omega^2 \beta\, u^{\pm}_{\mathrm{loc},k}(\varphi, \psi) & \\
= (2\alpha\,(-k + \mathrm{i}\partial_1) + \mathrm{i}\partial_1 \alpha)\, u^{\pm}_{\mathrm{loc}}(\varphi, \psi) \qquad & \text{in } C^{\pm}_1, && \text{(7.28a)} \\
(\pm\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm}_{\mathrm{loc},k}(\varphi, \psi) = 0 \qquad & \text{on } \Gamma^{\pm}_0, && \text{(7.28b)} \\
(\mp\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm}_{\mathrm{loc},k}(\varphi, \psi) = 0 \qquad & \text{on } \Gamma^{\pm}_1. && \text{(7.28c)}
\end{aligned}$$

Extension to higher order derivatives

In Remark 7.11 we pointed out that the RtR operators are differentiable with respect to ω and k up to any order. We can conclude that the same holds true for the local RtR operators and the propagation operators. Hence, we can characterize the partial derivatives of the RtR operators with respect to ω and k of any order similarly to the first order derivatives.

Let us introduce $u^{\pm,(m,n)}_{\mathrm{loc}}(\varphi, \psi) \equiv u^{\pm,(m,n)}_{\mathrm{loc}}(\cdot\,; \omega, k, \varphi, \psi) \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, C^{\pm}_1, \alpha)$, $m, n \in \mathbb{N}_0$, $m + n \geq 1$, as the unique solution of

$$\begin{aligned}
-(\nabla + \mathrm{i}k\tbinom{1}{0}) \cdot \alpha (\nabla + \mathrm{i}k\tbinom{1}{0}) u^{\pm,(m,n)}_{\mathrm{loc}}(\varphi, \psi) - \omega^2 \beta\, u^{\pm,(m,n)}_{\mathrm{loc}}(\varphi, \psi) = f \qquad & \text{in } C^{\pm}_1, && \text{(7.29a)} \\
(\pm\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm,(m,n)}_{\mathrm{loc}}(\varphi, \psi) = 0 \qquad & \text{on } \Gamma^{\pm}_0, && \text{(7.29b)} \\
(\mp\alpha\, \partial_2 + \mathrm{i}\rho) u^{\pm,(m,n)}_{\mathrm{loc}}(\varphi, \psi) = 0 \qquad & \text{on } \Gamma^{\pm}_1 && \text{(7.29c)}
\end{aligned}$$

with

$$\begin{aligned}
f = 2m\omega\beta u^{\pm,(m-1,n)}_{\mathrm{loc}}(\varphi, \psi) + m(m-1)\beta u^{\pm,(m-2,n)}_{\mathrm{loc}}(\varphi, \psi) & \\
+ n\,(2\alpha(-k + \mathrm{i}\partial_1) + \mathrm{i}\partial_1\alpha)\, u^{\pm,(m,n-1)}_{\mathrm{loc}}(\varphi, \psi) - n(n-1)\alpha\, u^{\pm,(m,n-2)}_{\mathrm{loc}}(\varphi, \psi) &
\end{aligned}$$

and the convention

$$\begin{aligned}
u_{\mathrm{loc}}^{\pm,(0,0)}(\cdot\,;\omega,k,\varphi,\psi) &= u_{\mathrm{loc}}^{\pm}(\cdot\,;\omega,k,\varphi,\psi),\\
u_{\mathrm{loc}}^{\pm,(1,0)}(\cdot\,;\omega,k,\varphi,\psi) &= u_{\mathrm{loc},\omega}^{\pm}(\cdot\,;\omega,k,\varphi,\psi),\\
u_{\mathrm{loc}}^{\pm,(0,1)}(\cdot\,;\omega,k,\varphi,\psi) &= u_{\mathrm{loc},k}^{\pm}(\cdot\,;\omega,k,\varphi,\psi).
\end{aligned}$$

Using the short notation

$$\begin{aligned}
\mathcal{T}_{\mathrm{fb}}^{\pm,(m,n)}(\omega,k) &:= \frac{\partial^{m+n}\mathcal{T}_{\mathrm{fb}}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n},\\
\mathcal{T}_{\mathrm{ff}}^{\pm,(m,n)}(\omega,k) &:= \frac{\partial^{m+n}\mathcal{T}_{\mathrm{ff}}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n},\\
\mathcal{T}_{\mathrm{bb}}^{\pm,(m,n)}(\omega,k) &:= \frac{\partial^{m+n}\mathcal{T}_{\mathrm{bb}}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n},\\
\mathcal{T}_{\mathrm{bf}}^{\pm,(m,n)}(\omega,k) &:= \frac{\partial^{m+n}\mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n},
\end{aligned}$$

the partial derivatives of the local RtR operators read

$$\begin{aligned}
\mathcal{T}_{\mathrm{fb}}^{\pm,(m,n)}(\omega,k)\varphi &= \qquad\quad (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\quad 0)\,|_{\Gamma_0^\pm},\\
\mathcal{T}_{\mathrm{ff}}^{\pm,(m,n)}(\omega,k)\varphi &= (\mathcal{S}^\pm)^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\quad 0)\,|_{\Gamma_1^\pm},\\
\mathcal{T}_{\mathrm{bb}}^{\pm,(m,n)}(\omega,k)\varphi &= \qquad\quad (\mp\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^{\pm,(m,n)}(0,\mathcal{S}^\pm\varphi)\,|_{\Gamma_0^\pm},\\
\mathcal{T}_{\mathrm{bf}}^{\pm,(m,n)}(\omega,k)\varphi &= (\mathcal{S}^\pm)^{-1}(\pm\alpha\,\partial_2 + \mathrm{i}\rho)u_{\mathrm{loc}}^{\pm,(m,n)}(0,\mathcal{S}^\pm\varphi)\,|_{\Gamma_1^\pm}
\end{aligned}$$

for all $m,n\in\mathbb{N}_0$ with $m+n\geq 1$.

With the help of these local RtR operators we can compute any partial derivative

$$\mathcal{P}_{\mathrm{fb}}^{\pm,(m,n)}(\omega,k) := \frac{\partial^{m+n}\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n},$$

$m,n\in\mathbb{N}_0$, $m+n\geq 1$, of the forward-backward propagation operator $\mathcal{P}_{\mathrm{fb}}^{\pm}(\omega,k)$ with respect to ω and k. Taking the m-th derivative with respect to ω and the n-th derivative with respect to k of the quadratic operator equation (7.15) yields

$$\begin{aligned}
0 =& \frac{\partial^{m+n}}{\partial\omega^m\partial\omega^n}\left[\mathcal{T}_{\mathrm{bb}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm}\mathcal{T}_{\mathrm{bf}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm} + \mathcal{T}_{\mathrm{bb}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm}\mathcal{T}_{\mathrm{ff}}^{\pm} + \left(\mathcal{T}_{\mathrm{fb}}^{\pm}\mathcal{T}_{\mathrm{bf}}^{\pm} - \mathcal{I}\right)\mathcal{P}_{\mathrm{fb}}^{\pm} + \mathcal{T}_{\mathrm{fb}}^{\pm}\mathcal{T}_{\mathrm{ff}}^{\pm}\right]\\
=& -\mathcal{P}_{\mathrm{fb}}^{\pm,(m,n)}\\
&+ \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^4(m,n)}\binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}}\mathcal{T}_{\mathrm{bb}}^{\pm,(m_1,n_1)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_2,n_2)}\mathcal{T}_{\mathrm{bf}}^{\pm,(m_3,n_3)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_4,n_4)}\\
&+ \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^3(m,n)}\binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}}\mathcal{T}_{\mathrm{bb}}^{\pm,(m_1,n_1)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_2,n_2)}\mathcal{T}_{\mathrm{ff}}^{\pm,(m_3,n_3)}\\
&+ \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^3(m,n)}\binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}}\mathcal{T}_{\mathrm{fb}}^{\pm,(m_1,n_1)}\mathcal{T}_{\mathrm{bf}}^{\pm,(m_2,n_2)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_3,n_3)}\\
&+ \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^2(m,n)}\binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}}\mathcal{T}_{\mathrm{fb}}^{\pm,(m_1,n_1)}\mathcal{T}_{\mathrm{ff}}^{\pm,(m_2,n_2)},
\end{aligned}$$

with multinomial coefficients as given defined Eq. (6.19). This can be brought into a similar form

like (7.25), i. e.

$$\begin{aligned}
\mathcal{T}_{\mathrm{bb}}^{\pm}\mathcal{P}_{\mathrm{fb}}^{\pm,(m,n)}\mathcal{P}_{\mathrm{ff}}^{\pm} + \left(\mathcal{R}^{\pm}\mathcal{T}_{\mathrm{bf}}^{\pm} - \mathcal{I}\right)\mathcal{P}_{\mathrm{fb}}^{\pm,(m,n)} \\
= - \sum_{(\mathbf{m},\mathbf{n})\in\widetilde{\mathfrak{N}}^4_{\{2,4\}}(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathcal{T}_{\mathrm{bb}}^{\pm,(m_1,n_1)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_2,n_2)}\mathcal{T}_{\mathrm{bf}}^{\pm,(m_3,n_3)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_4,n_4)} \\
- \sum_{(\mathbf{m},\mathbf{n})\in\widetilde{\mathfrak{N}}^3_{\{2\}}(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathcal{T}_{\mathrm{bb}}^{\pm,(m_1,n_1)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_2,n_2)}\mathcal{T}_{\mathrm{ff}}^{\pm,(m_3,n_3)} \\
- \sum_{(\mathbf{m},\mathbf{n})\in\widetilde{\mathfrak{N}}^3_{\{3\}}(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathcal{T}_{\mathrm{fb}}^{\pm,(m_1,n_1)}\mathcal{T}_{\mathrm{bf}}^{\pm,(m_2,n_2)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_3,n_3)} \\
- \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^2(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathcal{T}_{\mathrm{fb}}^{\pm,(m_1,n_1)}\mathcal{T}_{\mathrm{ff}}^{\pm,(m_2,n_2)},
\end{aligned} \tag{7.30}$$

where we omitted — for simplicity of presentation — the (ω, k)-dependence of the operators, and the sets $\mathfrak{N}^2$ and $\widetilde{\mathfrak{N}}^d_J$, $d = 3, 4$, $J = \{2\}, \{3\}, \{2, 4\}$, are defined in Eqs. (6.20) and (6.22), respectively.

Finally, differentiating Eq. (7.13) m times with respect to ω and n times with respect to k, we deduce that the derivatives

$$\mathcal{R}^{\pm,(m,n)}(\omega,k) := \frac{\partial^{m+n}\mathcal{R}^{\pm}(\omega,k)}{\partial\omega^m\partial k^n}$$

of the RtR operators read

$$\mathcal{R}^{\pm,(m,n)} = \mathcal{T}_{\mathrm{fb}}^{\pm,(m,n)} + \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^2(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathcal{T}_{\mathrm{bb}}^{\pm,(m_1,n_1)}\mathcal{P}_{\mathrm{fb}}^{\pm,(m_2,n_2)} \tag{7.31}$$

with the set $\mathfrak{N}^2(m, n)$ defined in (6.20).

7.1.4 Variational formulation of the local cell problems

The derivation of the weak formulation of the local cell problem (7.8) is straightforward. Rewriting the boundary conditions (7.8b) and (7.8c) in the form

$$\mp\alpha\,\partial_2 u_{\mathrm{loc}}^{\pm}(\varphi,\psi) = \mathrm{i}\rho\, u_{\mathrm{loc}}^{\pm}(\varphi,\psi) - \varphi \qquad \text{on } \Gamma_0^{\pm}, \tag{7.32a}$$

$$\pm\alpha\,\partial_2 u_{\mathrm{loc}}^{\pm}(\varphi,\psi) = \mathrm{i}\rho\, u_{\mathrm{loc}}^{\pm}(\varphi,\psi) - \psi \qquad \text{on } \Gamma_1^{\pm}, \tag{7.32b}$$

we can deduce that Eq. (7.8) is equivalent to: for given forward Robin trace φ on $\Gamma_0^{\pm}$ and backward Robin trace ψ on $\Gamma_1^{\pm}$ find $u_{\mathrm{loc}}^{\pm}(\varphi,\psi) \in \mathsf{H}^1_{1\mathrm{p}}(C_1^{\pm})$ such that

$$\mathfrak{b}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm}(\varphi,\psi), v; \omega, k) - \mathrm{i}\rho \sum_{j=0,1} \int_{\Gamma_j^{\pm}} u_{\mathrm{loc}}^{\pm}(\varphi,\psi)\,\overline{v}\,\mathrm{d}s(\mathbf{x}) = -\int_{\Gamma_0^{\pm}} \varphi\,\overline{v}\,\mathrm{d}s(\mathbf{x}) - \int_{\Gamma_1^{\pm}} \psi\,\overline{v}\,\mathrm{d}s(\mathbf{x}) \tag{7.33}$$

for all $v \in \mathsf{H}^1_{1\mathrm{p}}(C_1^{\pm})$, with the sesquilinear form $\mathfrak{b}_{C_1^{\pm}}$ as given in Eq. (6.25a).

Once the local cell solutions $u_{\mathrm{loc}}^{\pm}(\varphi,\psi)$ are known, we can compute the local RtR operators by inserting (7.32) into the definition (7.9) of the local RtR operators which yields

$$\mathcal{T}_{\mathrm{fb}}^{\pm}(\omega,k)\varphi = 2\mathrm{i}\rho\, u_{\mathrm{loc}}^{\pm}(\varphi, 0)|_{\Gamma_0^{\pm}} - \varphi, \tag{7.34a}$$

$$\mathcal{T}_{\mathrm{ff}}^{\pm}(\omega,k)\varphi = 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1} u_{\mathrm{loc}}^{\pm}(\varphi, 0)|_{\Gamma_1^{\pm}}, \tag{7.34b}$$

$$\mathcal{T}_{\mathrm{bb}}^{\pm}(\omega,k)\varphi = 2\mathrm{i}\rho\, u_{\mathrm{loc}}^{\pm}(0, \mathcal{S}^{\pm}\varphi)|_{\Gamma_0^{\pm}}, \tag{7.34c}$$

$$\mathcal{T}_{\mathrm{bf}}^{\pm}(\omega,k)\varphi = 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1} u_{\mathrm{loc}}^{\pm}(0, \mathcal{S}^{\pm}\varphi)|_{\Gamma_1^{\pm}} - \varphi \tag{7.34d}$$

for any $\varphi \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$.

We proceed with the variational formulation of the local cell problems (7.29) for the computation of the ω- and k-derivatives of the local cell solutions $u_{\mathrm{loc}}^{\pm}(\varphi,\psi)$. Similarly to Eq. (6.27) we find that (7.29) is equivalent to: seek $u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\psi)\in \mathsf{H}^1_{1\mathrm{p}}(C_1^{\pm})$ such that

$$\begin{aligned}\mathfrak{b}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\psi),v;\omega,k) - \mathrm{i}\rho\sum_{j=0,1}\int_{\Gamma_j^{\pm}} u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\psi)\,\overline{v}\,\mathrm{d}s(\mathbf{x})\\ = 2m\omega\,\mathfrak{m}^{\beta}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm,(m-1,n)}(\varphi,\psi),v) + m(m-1)\mathfrak{m}^{\beta}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm,(m-2,n)}(\varphi,\psi),v)\\ - 2nk\,\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm,(m,n-1)}(\varphi,\psi),v) - n\,\mathfrak{c}^{\alpha,1}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm,(m,n-1)}(\varphi,\psi),v)\\ - n(n-1)\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u_{\mathrm{loc}}^{\pm,(m,n-2)}(\varphi,\psi),v)\end{aligned}\tag{7.35}$$

for all $v\in \mathsf{H}^1_{1\mathrm{p}}(C_1^{\pm})$ and $m,n\in\mathbb{N}_0$ with $m+n\geq 1$. Then the derivatives of the local RtR operators applied to $\varphi\in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ read

$$\begin{aligned}\mathcal{T}_{\mathrm{fb}}^{\pm,(m,n)}(\omega,k)\varphi &= 2\mathrm{i}\rho\qquad u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\quad 0)\,|_{\Gamma_0^{\pm}},\\ \mathcal{T}_{\mathrm{ff}}^{\pm,(m,n)}(\omega,k)\varphi &= 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1}u_{\mathrm{loc}}^{\pm,(m,n)}(\varphi,\quad 0)\,|_{\Gamma_1^{\pm}},\\ \mathcal{T}_{\mathrm{bb}}^{\pm,(m,n)}(\omega,k)\varphi &= 2\mathrm{i}\rho\qquad u_{\mathrm{loc}}^{\pm,(m,n)}(0,\mathcal{S}^{\pm}\varphi)\,|_{\Gamma_0^{\pm}},\\ \mathcal{T}_{\mathrm{bf}}^{\pm,(m,n)}(\omega,k)\varphi &= 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1}u_{\mathrm{loc}}^{\pm,(m,n)}(0,\mathcal{S}^{\pm}\varphi)\,|_{\Gamma_1^{\pm}}\end{aligned}$$

for any $m,n\in\mathbb{N}_0$ with $m+n\geq 1$.

7.1.5 Discretization

In Section 7.1.4 we introduced a variational formulation for the local cell problems to compute the local RtR operators (7.9). In this section we now want to describe the computation of the discrete RtR maps.

For this, we employ the FE spaces $\mathsf{S}^p_{1\mathrm{p}}(C_1^{\pm})$ and $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$, that were introduced in Section 6.1.5 as subspaces of $\mathsf{H}^1_{1\mathrm{p}}(C_1^{\pm})$ and $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ with polynomial degree p and dimensions $N(C_1^{\pm})$ and $N(\Gamma_0^{\pm})$, respectively. As we shall assume that permittivity $\varepsilon_{\mathrm{wg}}$ and thus, the coefficient functions do not jump on the boundaries $\Gamma_0^{\pm}$ we can expect the Neumann and Robin traces on $\Gamma_0^{\pm}$ to be in $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ and hence, $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ is an appropriate FE subspace of the dual space $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ of the Neumann and Robin traces on $\Gamma_0^{\pm}$. If this additional smoothness condition of the coefficient functions is not satisfied, we shall use the biorthogonal basis proposed by Wohlmuth [Woh01] as subspace of $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$. Finally, recall that we defined the meshes $\mathfrak{M}(C_1^{\pm})$ to be periodic in direction $\mathbf{a}_2^{\pm}$ and thus, the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ shifted to $\Gamma_1^{\pm}$ build a basis of the corresponding FE subspace of $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_1^{\pm})$. The same is valid for the FE subspaces of $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ and $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_1^{\pm})$. Recall, in particular, that this implies that the basis functions $b_{C_1^{\pm},n}$, $n=1,\dots,N(C_1^{\pm})$, of the FE spaces $\mathsf{S}^p_{1\mathrm{p}}(C_1^{\pm})$ can be ordered such that

- the basis functions with index $n\in\mathfrak{S}(C_1^{\pm},\Gamma_0^{\pm})=\{1,\dots,N(\Gamma_0^{\pm})\}$ vanish on $\Gamma_1^{\pm}$, but their traces on $\Gamma_0^{\pm}$ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$,
- the basis functions with index $n\in\mathfrak{S}(C_1^{\pm},\Gamma_1^{\pm})=\{N(\Gamma_0^{\pm})+1,\dots,2\,N(\Gamma_0^{\pm})\}$ vanish on $\Gamma_0^{\pm}$, but their traces on $\Gamma_1^{\pm}$ shifted to $\Gamma_0^{\pm}$, using the shift operator $(\mathcal{S}^{\pm})^{-1}$ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ as well, and
- the basis functions with index $n\in\mathfrak{S}(C_1^{\pm},C_1^{\pm})=\{2\,N(\Gamma_0^{\pm})+1,\dots,N(C_1^{\pm})\}$ vanish on $\Gamma_0^{\pm}$ and $\Gamma_1^{\pm}$.

Thus, the traces of the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(C_1^{\pm})$ on $\Gamma_i^{\pm}$, $i=0,1$, are related to the basis functions $b_{\Gamma_0^{\pm},n}$, $n=1,\dots,N(\Gamma_0^{\pm})$, of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ such that

$$b_{\Gamma_0^{\pm},n}=\sum_{m=1}^{N(\Gamma_0^{\pm})} Q^0_{C_1^{\pm},mn}\,b_{C_1^{\pm},m}|_{\Gamma_0^{\pm}}=\sum_{m=1}^{N(\Gamma_0^{\pm})} Q^1_{C_1^{\pm},mn}\,b_{C_1^{\pm},N(\Gamma_0^{\pm})+m}|_{\Gamma_1^{\pm}},$$

with permutation matrices $\mathbf{Q}^i_{C_1^{\pm}}\in\mathbb{R}^{N(\Gamma_0^{\pm})\times N(\Gamma_0^{\pm})}$, $i=0,1$, cf. Eq. (6.30).

Discretization of the local cell problems

In this subsection we aim to compute the discrete versions of the local RtR operators $\mathcal{T}^{\pm}_{\mathrm{fb}}$, $\mathcal{T}^{\pm}_{\mathrm{ff}}$, $\mathcal{T}^{\pm}_{\mathrm{bb}}$ and $\mathcal{T}^{\pm}_{\mathrm{bf}}$ in order to access the discrete forward-forward propagation operators and discrete RtR operators in the following subsections.

Using the FE spaces $\mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$ and $\mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0)$ we derive a discrete form of the local cell problems (7.8): for given forward Robin trace $\varphi_h \in \mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0)$ on $\Gamma^{\pm}_0$ and backward Robin trace $\psi_h = \mathcal{S}^{\pm}\tilde{\varphi}$, $\tilde{\varphi} \in \mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0)$, on $\Gamma^{\pm}_1$ find $u^{\pm}_{\mathrm{loc,h}}(\varphi_h, \psi_h) \in \mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$ such that

$$\begin{aligned}\mathfrak{b}_{C^{\pm}_1}(u^{\pm}_{\mathrm{loc,h}}(\varphi_h,\psi_h), v_h; \omega, k) - \mathrm{i}\rho \sum_{j=0,1} \int_{\Gamma^{\pm}_j} u^{\pm}_{\mathrm{loc,h}}(\varphi_h,\psi_h)\, \overline{v}_h \,\mathrm{d}s(\mathbf{x}) \\ = -\int_{\Gamma^{\pm}_0} \varphi_h\, \overline{v}_h \,\mathrm{d}s(\mathbf{x}) - \int_{\Gamma^{\pm}_1} \psi_h\, \overline{v}_h \,\mathrm{d}s(\mathbf{x}) \quad (7.36)\end{aligned}$$

for all $v_h \in \mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$, cf. Eq. (7.33). These discrete local cell problems are well-posed as long as the mesh width h is chosen small enough and the polynomial degree p is large enough [SS11, Thm. 4.2.9], [MS11].

The discrete local RtR operators are then defined as

$$\mathcal{T}^{\pm}_{\mathrm{fb},h}(\omega,k)\varphi_h = 2\mathrm{i}\rho \quad u^{\pm}_{\mathrm{loc,h}}(\varphi_h, \quad 0)|_{\Gamma^{\pm}_0} - \varphi_h, \tag{7.37a}$$

$$\mathcal{T}^{\pm}_{\mathrm{ff},h}(\omega,k)\varphi_h = 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1} u^{\pm}_{\mathrm{loc,h}}(\varphi_h, \quad 0)|_{\Gamma^{\pm}_1}, \tag{7.37b}$$

$$\mathcal{T}^{\pm}_{\mathrm{bb},h}(\omega,k)\varphi_h = 2\mathrm{i}\rho \quad u^{\pm}_{\mathrm{loc,h}}(0, \mathcal{S}^{\pm}\varphi_h)|_{\Gamma^{\pm}_0}, \tag{7.37c}$$

$$\mathcal{T}^{\pm}_{\mathrm{bf},h}(\omega,k)\varphi_h = 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1} u^{\pm}_{\mathrm{loc,h}}(0, \mathcal{S}^{\pm}\varphi_h)|_{\Gamma^{\pm}_1} - \varphi_h, \tag{7.37d}$$

cf. Eq. (7.34). Since the discrete local cell problems (7.36) are well-posed, it follows that the discrete local RtR operators (7.37) inherit their properties from the continuous local RtR operators (7.9). In particular, $\mathcal{T}^{\pm}_{\mathrm{bf},h}$ is invertible as long as the discrete auxiliary local cell problem, find $\widetilde{u}^{\pm}_{\mathrm{loc,h}}(0,\psi_h) \in \mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$ such that

$$\begin{aligned}\mathfrak{b}_{C^{\pm}_1}(\widetilde{u}^{\pm}_{\mathrm{loc,h}}(0,\psi_h), v_h; \omega, k) - \mathrm{i}\rho \int_{\Gamma^{\pm}_0} \widetilde{u}^{\pm}_{\mathrm{loc,h}}(0,\psi_h)\, \overline{v}_h \,\mathrm{d}s(\mathbf{x}) + \mathrm{i}\rho \int_{\Gamma^{\pm}_1} \widetilde{u}^{\pm}_{\mathrm{loc,h}}(0,\psi_h)\, \overline{v}_h \,\mathrm{d}s(\mathbf{x}) \\ = \int_{\Gamma^{\pm}_1} \psi_h\, \overline{v}_h \,\mathrm{d}s(\mathbf{x}) \quad (7.38)\end{aligned}$$

for all $v_h \in \mathsf{S}^p_{1\mathrm{p}}(C^{\pm}_1)$, is well-posed.

Now we want to transform the discretized local cell problems (7.36) into linear systems of equations, and represent the discrete local RtR operators (7.37) in terms of matrices. With the help of the basis functions $b_{\Gamma^{\pm}_0,n}$, $n \in \{1,\dots,N(\Gamma^{\pm}_0)\}$, of the discrete space $\mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0)$, we seek matrix representations of the discrete local RtR operators $\mathcal{T}^{\pm}_{ij,h}$, $i,j \in \{\mathrm{f},\mathrm{b}\}$, i. e. we search for matrices $\mathbf{T}^{\pm}_{ij} \in \mathbb{C}^{N(\Gamma^{\pm}_0)\times N(\Gamma^{\pm}_0)}$ with entries $T^{\pm}_{ij,mn}$, $m,n \in \{1,\dots,N(\Gamma^{\pm}_0)\}$ such that

$$\mathcal{T}^{\pm}_{ij,h} b_{\Gamma^{\pm}_0,n} = \sum_{m=1}^{N(\Gamma^{\pm}_0)} T^{\pm}_{ij,mn} b_{\Gamma^{\pm}_0,m} \in \mathsf{S}^p_{1\mathrm{p}}(\Gamma^{\pm}_0), \qquad i,j \in \{\mathrm{f},\mathrm{b}\}. \tag{7.39}$$

We recall the matrix $\mathbf{B}_{C^{\pm}_1} \in \mathbb{C}^{N(C^{\pm}_1)\times N(C^{\pm}_1)}$ from Eq. (6.32), and introduce the matrices $\mathbf{M}_{C^{\pm}_1,\Gamma^{\pm}_i} \in \mathbb{R}^{N(C^{\pm}_1)\times N(C^{\pm}_1)}$, $i = 0,1$, with entries

$$M_{C^{\pm}_1,\Gamma^{\pm}_i,mn} = \int_{\Gamma^{\pm}_i} b_{C^{\pm}_1,n}\, \overline{b}_{C^{\pm}_1,m} \,\mathrm{d}s(\mathbf{x}), \qquad i = 0,1,$$

$m,n \in \{1,\dots,N(C^{\pm}_1)\}$, related to the boundary integrals in Eq. (7.36). Furthermore, we define the matrix

$$\mathbf{S}_{C^{\pm}_1}(\omega,k) := \mathbf{B}_{C^{\pm}_1}(\omega,k) - \mathrm{i}\rho \sum_{j=0,1} \mathbf{M}_{C^{\pm}_1,\Gamma^{\pm}_j},$$

and introduce the notation $\mathbf{M}_{C_1^\pm,\Gamma_i^\pm}(\Gamma_i^\pm) \in \mathbb{R}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$ $i = 0, 1$, for the block matrix of $\mathbf{M}_{C_1^\pm,\Gamma_i^\pm}$ with column indices $\mathfrak{S}(C_1^\pm, \Gamma_i^\pm)$.

Let $\mathbf{e}_n^{N(\Gamma_0^\pm)} \in \mathbb{R}^{N(\Gamma_0^\pm)}$, $n \in \{1, \dots, N(C_1^\pm)\}$ denote the n-th unit vector of dimension $N(C_1^\pm)$. Then the matrix form of the discrete local cell problem (7.33) reads

$$\mathbf{S}_{C_1^\pm}(\omega, k)\, \mathbf{u}^\pm_{\mathrm{loc,h}}(b_{\Gamma_0^\pm,m}, b_{\Gamma_0^\pm,n}) = -\mathbf{M}_{C_1^\pm,\Gamma_0^\pm}(\Gamma_0^\pm)\, \mathbf{Q}^0_{C_1^\pm}\, \mathbf{e}_m^{N(\Gamma_0^\pm)} - \mathbf{M}_{C_1^\pm,\Gamma_1^\pm}(\Gamma_1^\pm)\, \mathbf{Q}^1_{C_1^\pm}\, \mathbf{e}_n^{N(\Gamma_0^\pm)}, \tag{7.40}$$

with $m, n \in \{1, \dots, N(\Gamma_0^\pm)\}$, where $\mathbf{u}^\pm_{\mathrm{loc,h}}(b_{\Gamma_0^\pm,m}, b_{\Gamma_0^\pm,n}) \in \mathbb{C}^{N(C_1^\pm)}$ are the coefficient vectors of the discrete local cell solutions $u_{\mathrm{loc,h}}(\cdot\,; \omega, k, b_{\Gamma_0^\pm,m}, b_{\Gamma_0^\pm,n}) \in \mathsf{S}^p_{1\mathrm{p}}(C_1^\pm)$ with respect to the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(C_1^\pm)$.

Collecting the vectors $\mathbf{u}^\pm_{\mathrm{loc,h}}(b_{\Gamma_0^\pm,m}, 0)$, $m \in \{1, \dots, N(\Gamma_0^\pm)\}$, in matrices $\mathbf{U}^\pm_{\mathrm{loc},h,0} \in \mathbb{C}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$ satisfying

$$\mathbf{S}_{C_1^\pm}(\omega, k)\, \mathbf{U}^\pm_{\mathrm{loc},h,0} = -\mathbf{M}_{C_1^\pm,\Gamma_0^\pm}(\Gamma_0^\pm)\, \mathbf{Q}^0_{C_1^\pm},$$

and the vectors $\mathbf{u}^\pm_{\mathrm{loc,h}}(0, b_{\Gamma_0^\pm,n})$, $n \in \{1, \dots, N(\Gamma_0^\pm)\}$, in matrices $\mathbf{U}^\pm_{\mathrm{loc},h,1} \in \mathbb{C}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$ satisfying

$$\mathbf{S}_{C_1^\pm}(\omega, k)\, \mathbf{U}^\pm_{\mathrm{loc},h,1} = -\mathbf{M}_{C_1^\pm,\Gamma_1^\pm}(\Gamma_1^\pm)\, \mathbf{Q}^1_{C_1^\pm},$$

we can — using Eq. (7.37) and Eq. (7.39) — deduce

$$\begin{aligned}
\mathbf{T}^\pm_{\mathrm{fb}} &= 2\mathrm{i}\rho \mathbf{Q}^0_{C_1^\pm} \mathbf{U}^\pm_{\mathrm{loc},h,0}(\Gamma_0^\pm) - \mathbf{I},\\
\mathbf{T}^\pm_{\mathrm{ff}} &= 2\mathrm{i}\rho \mathbf{Q}^1_{C_1^\pm} \mathbf{U}^\pm_{\mathrm{loc},h,0}(\Gamma_1^\pm),\\
\mathbf{T}^\pm_{\mathrm{bb}} &= 2\mathrm{i}\rho \mathbf{Q}^0_{C_1^\pm} \mathbf{U}^\pm_{\mathrm{loc},h,1}(\Gamma_0^\pm),\\
\mathbf{T}^\pm_{\mathrm{bf}} &= 2\mathrm{i}\rho \mathbf{Q}^1_{C_1^\pm} \mathbf{U}^\pm_{\mathrm{loc},h,1}(\Gamma_1^\pm) - \mathbf{I},
\end{aligned}$$

where $\mathbf{U}^\pm_{\mathrm{loc},h,i}(\Gamma_j^\pm)$, $i, j \in \{0, 1\}$, denotes the block matrix of $\mathbf{U}^\pm_{\mathrm{loc},h,i}$ with row indices $\mathfrak{S}(C_1^\pm, \Gamma_j^\pm)$.

To summarize, it is sufficient to solve the block system

$$\begin{pmatrix} \mathbf{S}_{C_1^\pm}(\omega,k) & \mathbf{0} & \mathbf{0} \\ -2\mathrm{i}\rho\mathbf{Q}^0_{C_1^\pm}\mathbf{I}(\Gamma_0^\pm, C_1^\pm) & \mathbf{I} & \mathbf{0} \\ -2\mathrm{i}\rho\mathbf{Q}^1_{C_1^\pm}\mathbf{I}(\Gamma_1^\pm, C_1^\pm) & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{U}^\pm_{\mathrm{loc},h,0} & \mathbf{U}^\pm_{\mathrm{loc},h,1} \\ \mathbf{T}^\pm_{\mathrm{fb}} & \mathbf{T}^\pm_{\mathrm{bb}} \\ \mathbf{T}^\pm_{\mathrm{ff}} & \mathbf{T}^\pm_{\mathrm{bf}} \end{pmatrix} = \begin{pmatrix} -\mathbf{M}_{C_1^\pm,\Gamma_0^\pm}(\Gamma_0^\pm)\mathbf{Q}^0_{C_1^\pm} & -\mathbf{M}_{C_1^\pm,\Gamma_1^\pm}(\Gamma_1^\pm)\mathbf{Q}^1_{C_1^\pm} \\ -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix}$$

for the matrices $\mathbf{T}^\pm_{\mathrm{fb}}$, $\mathbf{T}^\pm_{\mathrm{ff}}$, $\mathbf{T}^\pm_{\mathrm{bb}}$ and $\mathbf{T}^\pm_{\mathrm{bf}}$, where the rectangular matrices $\mathbf{I}(\Gamma_i^\pm, C_1^\pm) \in \mathbb{R}^{N(\Gamma_0^\pm)\times N(C_1^\pm)}$, $i \in \{0, 1\}$, are the block matrices of the $N(C_1^\pm) \times N(C_1^\pm)$ identity matrix with row indices $\mathfrak{S}(C_1^\pm, \Gamma_i^\pm)$.

Finally, we note that the matrices $\mathbf{T}^\pm_{ij}$, $i, j \in \{\mathrm{f}, \mathrm{b}\}$, map coefficient vectors of FE functions in $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ onto coefficient vectors of other FE functions in $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$, i. e. they map in strong sense. Hence, products of local RtR operators, as they appear in the Riccati equation (7.18) and the formula (7.13) for the computation of the RtR operator, can be realized by simply multiplying the matrices $\mathbf{T}^\pm_{ij}$, $i, j \in \{\mathrm{f}, \mathrm{b}\}$. This is in contrast to the matrices of the local DtN operators in Chapter 6, which are computed in weak sense, and hence, cannot be multiplied directly.

Computation of the discrete propagation operators

In this subsection we face the problem of computing discrete approximations to the forward-forward propagation operators $\mathcal{P}^\pm_{\mathrm{ff}}(\omega, k)$ and the forward-backward propagation operators $\mathcal{P}^\pm_{\mathrm{fb}}(\omega, k)$.

If the discrete auxiliary local cell problem (7.38) is well-posed and hence, the discrete local backward-forward operator $\mathcal{T}^\pm_{\mathrm{bf},h}(\omega, k)$ is invertible, we search for discrete operators $\mathcal{P}^\pm_{\mathrm{ff},h}(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm))$ with spectral radius strictly less than one that satisfy the discrete operator equation

$$\mathcal{T}^\pm_{\mathrm{bb},h}\left(\mathcal{T}^\pm_{\mathrm{bf},h}\right)^{-1}\left(\mathcal{P}^\pm_{\mathrm{ff},h}\right)^2 + \left(\mathcal{T}^\pm_{\mathrm{fb},h} - \left(\mathcal{T}^\pm_{\mathrm{bf},h}\right)^{-1} - \mathcal{T}^\pm_{\mathrm{bb},h}\left(\mathcal{T}^\pm_{\mathrm{bf},h}\right)^{-1}\mathcal{T}^\pm_{\mathrm{ff},h}\right)\mathcal{P}^\pm_{\mathrm{ff},h} + \left(\mathcal{T}^\pm_{\mathrm{bf},h}\right)^{-1}\mathcal{T}^\pm_{\mathrm{ff},h} = 0, \tag{7.41}$$

cf. Eq. (7.18). Similarly, we introduce the discrete forward-backward propagation operator $\mathcal{P}^\pm_{\mathrm{fb},h}(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm))$ by

$$\mathcal{P}^\pm_{\mathrm{fb},h} = \left(\mathcal{T}^\pm_{\mathrm{bf},h}\right)^{-1}\left(\mathcal{P}^\pm_{\mathrm{ff},h} - \mathcal{T}^\pm_{\mathrm{ff},h}\right),$$

cf. Eq. (7.17).

Using the basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ we want to express the discrete forward-forward propagation operator $\mathcal{P}^\pm_{\mathrm{ff},h}$ in matrices $\mathbf{P}^\pm_{\mathrm{ff}} \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$ with coefficients $P^\pm_{\mathrm{ff},mn}$, $m,n \in \{1,\dots,N(\Gamma_0^\pm)\}$, that satisfy

$$\mathcal{P}^\pm_{\mathrm{ff},h} b_{\Gamma_0^\pm,n} = \sum_{m=1}^{N(\Gamma_0^\pm))} P^\pm_{\mathrm{ff},mn} b_{\Gamma_0^\pm,m}. \tag{7.42}$$

Using this definition and the definition of the matrices $\mathbf{T}^\pm_{ij}$, $i,j \in \{\mathrm{f},\mathrm{b}\}$, in (7.39) we can write Eq. (7.41) as a quadratic matrix-valued equation

$$\mathbf{T}^\pm_{\mathrm{bb}} \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \left(\mathbf{P}^\pm_{\mathrm{ff}}\right)^2 + \left(\mathbf{T}^\pm_{\mathrm{fb}} - \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} - \mathbf{T}^\pm_{\mathrm{bb}} \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \mathbf{T}^\pm_{\mathrm{ff}}\right) \mathbf{P}^\pm_{\mathrm{ff}} + \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \mathbf{T}^\pm_{\mathrm{ff}} = \mathbf{0}. \tag{7.43}$$

Considering that the discretization preserves the periodicity properties of $C_1^\pm$ in $\mathbf{a}_2$-direction we deduce that the forward-forward propagation matrix $\mathbf{P}^\pm_{\mathrm{ff}}$ is the unique matrix satisfying Eq. (7.43) with eigenvalues whose magnitudes are strictly less than one.

Similarly, we express the discrete forward-backward propagation operator $\mathcal{P}^\pm_{\mathrm{fb},h}$ in matrices $\mathbf{P}^\pm_{\mathrm{fb}} \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$ with coefficients $P^\pm_{\mathrm{fb},mn}$, $m,n \in \{1,\dots,N(\Gamma_0^\pm)\}$, that satisfy

$$\mathcal{P}^\pm_{\mathrm{fb},h} b_{\Gamma_0^\pm,n} = \sum_{m=1}^{N(\Gamma_0^\pm))} P^\pm_{\mathrm{fb},mn} b_{\Gamma_0^\pm,m}. \tag{7.44}$$

Then we can rewrite Eq. (7.1.5) in terms of matrix equation

$$\mathbf{P}^\pm_{\mathrm{fb}} = \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \left(\mathbf{P}^\pm_{\mathrm{ff}} - \mathbf{T}^\pm_{\mathrm{ff}}\right).$$

Analogously to Chapter 6 for the computation of the discrete Dirichlet propagation operator, we propose a spectral decomposition to compute $\mathbf{P}^\pm_{\mathrm{ff}}$. Even though we cannot guarantee that $\mathbf{P}^\pm_{\mathrm{ff}}$ is diagonalizable the spectral decomposition has proven to be an efficient and reliable approach to compute $\mathbf{P}^\pm_{\mathrm{ff}}$. If, however, the propagation matrix $\mathbf{P}^\pm_{\mathrm{ff}}$ is in fact of Jordan type and hence, cannot be diagonalized, we can still use this spectral method in a generalized form by identifying the Jordan blocks and computing the Jordan chains, as shown in [Fli09] for the DtN method.

Thus, we want to find eigenvalues $\mu^\pm(\omega,k) \in \mathbb{C}$ with magnitude strictly less than one and their corresponding eigenvectors $\boldsymbol{\psi}^\pm(\omega,k) \in \mathbb{C}^{N(\Gamma_0^\pm)}$ of the quadratic eigenvalue problem

$$\left[\mathbf{T}^\pm_{\mathrm{bb}} \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} (\mu^\pm)^2 + \left(\mathbf{T}^\pm_{\mathrm{fb}} - \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} - \mathbf{T}^\pm_{\mathrm{bb}} \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \mathbf{T}^\pm_{\mathrm{ff}}\right) \mu^\pm + \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \mathbf{T}^\pm_{\mathrm{ff}}\right] \boldsymbol{\psi}^\pm = 0, \tag{7.45}$$

which can be transformed into the generalized linear eigenvalue problem

$$\begin{pmatrix} -\left(\mathbf{T}^\pm_{\mathrm{fb}} - \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} - \mathbf{T}^\pm_{\mathrm{bb}} \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \mathbf{T}^\pm_{\mathrm{ff}}\right) & -\left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} \mathbf{T}^\pm_{\mathrm{ff}} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \boldsymbol{\Psi}^\pm = \mu^\pm \begin{pmatrix} \mathbf{T}^\pm_{\mathrm{bb}} \left(\mathbf{T}^\pm_{\mathrm{bf}}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \boldsymbol{\Psi}^\pm, \tag{7.46}$$

cf. [TM01], with $\boldsymbol{\Psi}^\pm = \begin{pmatrix} \mu^\pm \boldsymbol{\psi}^\pm \\ \boldsymbol{\psi}^\pm \end{pmatrix}$.

Now let us present an important result of the spectral decomposition. If $\omega^2 \in \mathbb{R}^+ \setminus \sigma^{\mathrm{ess}}(k)$ is not a global or local Dirichlet eigenvalue, i. e. an eigenvalue of the infinite half-strip problem (7.1a) or the local cell problem (7.8a) with homogeneous Dirichlet boundary conditions, the following result is a direct consequence of Proposition 7.7 and Proposition 6.15 in Chapter 6. If, however, ω^2 is such a global or local Dirichlet eigenvalue we conjecture that the result still holds true.

Conjecture 7.13. *If $\mu^\pm(\omega,k) \in \mathbb{C} \setminus \{0\}$ is an eigenvalue of (7.45), then $\left(\overline{\mu^\pm(\omega,k)}\right)^{-1}$ is also an eigenvalue.*

As a by-product and analogously to the case with DtN operators, the spectral decomposition of the propagation matrix $\mathbf{P}^\pm_{\mathrm{ff}}(\omega,k)$ yields information whether ω^2 is inside the discrete approximation of the essential spectrum $\sigma^{\mathrm{ess}}(k)$.

Definition 7.14. *We call the set of numbers ω^2 for which the quadratic eigenvalue problem (7.45) has eigenvalues with magnitude one* approximative spectrum $\sigma_h^{\pm}(k)$. *Furthermore, we define* $\sigma_h^{\text{ess}}(k) := \sigma_h^{+}(k) \cup \sigma_h^{-}(k)$.

With the help of Conjecture 7.13 and Definition 7.14 it is now clear how to compute the spectral decomposition of the propagation matrix $\mathbf{P}_{\text{ff}}^{\pm}(\omega, k)$. We solve the general eigenvalue problem (7.46) for its $2N(\Gamma_0^{\pm})$ eigenvalues $\mu^{\pm}(\omega, k)$. If there exist eigenvalues with magnitude equal to one we stop our computation as we know from Definition 7.14 that this means that ω^2 is in the approximative essential spectrum $\sigma_h^{\text{ess}}(k)$. Otherwise, and in accordance to Conjecture 7.13, the $2N(\Gamma_0^{\pm})$ eigenvalues $\mu^{\pm}(\omega, k)$ split into $N(\Gamma_0^{\pm})$ eigenvalues with magnitude strictly less than one and $N(\Gamma_0^{\pm})$ eigenvalues with magnitude strictly larger than one. While discarding the $N(\Gamma_0^{\pm})$ eigenvalues with magnitude strictly larger than one, the $N(\Gamma_0^{\pm})$ eigenvalues $\mu^{\pm}(\omega, k)$ with magnitude strictly less than one and their corresponding eigenvectors $\boldsymbol{\psi}^{\pm}(\omega, k)$ form the spectral decomposition of the propagation matrix $\mathbf{P}_{\text{ff}}^{\pm}(\omega, k)$.

Now let us discuss the procedure for the case where the discrete auxiliary local cell problem (7.38) is not well-posed. Then the discrete local backward-forward RtR operator $\mathcal{T}_{\text{bf},h}^{\pm}$ is not invertible and hence, we cannot compute the discrete forward-backward propagation operator $\mathcal{P}_{\text{fb},h}^{\pm}$ by first solving (7.41) for the discrete forward-forward propagation operator $\mathcal{P}_{\text{ff},h}^{\pm}$ with spectral radius strictly less than one, and then using Eq. (7.1.5). Instead we have to solve

$$\mathcal{T}_{\text{bb},h}^{\pm}\mathcal{P}_{\text{fb},h}^{\pm}\mathcal{T}_{\text{bf},h}^{\pm}\mathcal{P}_{\text{fb},h}^{\pm} + \mathcal{T}_{\text{bb},h}^{\pm}\mathcal{P}_{\text{fb},h}^{\pm}\mathcal{T}_{\text{ff},h}^{\pm} + \left(\mathcal{T}_{\text{fb},h}^{\pm}\mathcal{T}_{\text{bf},h}^{\pm} - \mathcal{I}\right)\mathcal{P}_{\text{fb},h}^{\pm} + \mathcal{T}_{\text{fb},h}^{\pm}\mathcal{T}_{\text{ff},h}^{\pm} = 0, \tag{7.47}$$

i. e. the discrete version of the quadratic operator equation (7.15), for the discrete forward-backward propagation operator $\mathcal{P}_{\text{fb},h}^{\pm}(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}_{1\text{p}}^{p}(\Gamma_0^{\pm}))$. Using the matrix notation introduced above, Eq. (7.47) can be rewritten in matrix form

$$\mathbf{F}(\mathbf{P}_{\text{fb}}^{\pm}) := \mathbf{T}_{\text{bb}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm}\mathbf{T}_{\text{bf}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm} + \mathbf{T}_{\text{bb}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm}\mathbf{T}_{\text{ff}}^{\pm} + \left(\mathbf{T}_{\text{fb}}^{\pm}\mathbf{T}_{\text{bf}}^{\pm} - \mathbf{I}\right)\mathbf{P}_{\text{fb}}^{\pm} + \mathbf{T}_{\text{fb}}^{\pm}\mathbf{T}_{\text{ff}}^{\pm} = \mathbf{0}. \tag{7.48}$$

Since $\mathbf{T}_{\text{bf}}^{\pm}$ is singular, we cannot uniquely compute $\mathbf{P}_{\text{fb}}^{\pm}$ from Eq. (7.48) by additionally requiring the matrix

$$\mathbf{P}_{\text{ff}}^{\pm} = \mathbf{T}_{\text{ff}}^{\pm} + \mathbf{T}_{\text{bf}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm}$$

to have spectral radius strictly less than one. Instead, we take Proposition 7.6 into account and assume that the discretization preserves the invertibility of the RtR operator $\mathcal{R}^{\pm}(\omega, k)$, cf. Lemma 7.2. Then we may argue that $\mathbf{P}_{\text{fb}}^{\pm}$ is the unique solution of (7.48) such that

$$\mathbf{P}_{\text{ff}}^{\pm} = \left(\mathbf{T}_{\text{fb}}^{\pm} + \mathbf{T}_{\text{bb}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm}\right)^{-1}\mathbf{P}_{\text{fb}}^{\pm}$$

has spectral radius strictly less than one. Employing this observation, we propose a heuristic numerical scheme for the computation of $\mathbf{P}_{\text{fb}}^{\pm}$, that is motivated by the Newton method presented in [JLF06]. For this, we need the directional derivative of the matrix function $\mathbf{F}$ in (7.48) with respect to $\mathbf{P}_{\text{fb}}^{\pm}$. It reads

$$\mathrm{D}\mathbf{F}(\mathbf{P}_{\text{fb}}^{\pm})\mathbf{H} = \mathbf{T}_{\text{bb}}^{\pm}\mathbf{H}\mathbf{T}_{\text{bf}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm} + \mathbf{T}_{\text{bb}}^{\pm}\mathbf{P}_{\text{fb}}^{\pm}\mathbf{T}_{\text{bf}}^{\pm}\mathbf{H} + \mathbf{T}_{\text{bb}}^{\pm}\mathbf{H}\mathbf{T}_{\text{ff}}^{\pm} + \left(\mathbf{T}_{\text{fb}}^{\pm}\mathbf{T}_{\text{bf}}^{\pm} - \mathbf{I}\right)\mathbf{H}.$$

The Newton method for the computation of $\mathbf{P}_{\text{fb}}^{\pm}$ works as shown in Algorithm 7.1.

In Chapter 6 we argued that the spectral decomposition is preferable compared to the heuristic Newton method for the computation of the discrete Dirichlet propagation operator since its results have a physical meaning. The same is true for the computation of the discrete forward-forward propagation operator as explained in Definition 7.14. Hence, we shall prefer this procedure as long as we can guarantee that the discrete auxiliary local cell problem (7.38) is well-posed. As mentioned already above, the constant $\rho \in \mathbb{R} \setminus \{0\}$ may be chosen such that (7.38) is well-posed for all values of $(\omega^2, k) \in \mathbb{R}^+ \times B$ under consideration, as we shall demonstrate in the numerical results in Section 7.3.

Definition of the discrete Robin-to-Robin operators

Considering Eq. (7.13) for the characterization of the RtR operators $\mathcal{R}^{\pm}(\omega, k)$, we can define the discrete RtR operators

$$\mathcal{R}_h^{\pm}(\omega, k) = \mathcal{T}_{\text{fb},h}^{\pm}(\omega, k) + \mathcal{T}_{\text{bb},h}^{\pm}(\omega, k)\mathcal{P}_{\text{fb},h}^{\pm}(\omega, k) \in \boldsymbol{\mathcal{L}}(\mathsf{S}_{1\text{p}}^{p}(\Gamma_0^{\pm})).$$

Algorithm 7.1. Newton's method for the computation of the discrete forward-backward propagation operator.

1: Choose start matrix $\mathbf{P}^{\pm}_{\mathrm{fb},(0)} \in \mathbb{C}^{N(\Gamma_0^{\pm})\times N(\Gamma_0^{\pm})}$, e. g. $\mathbf{P}^{\pm}_{\mathrm{fb},(0)} = \mathbf{0}$.
2: Choose tolerance $\varepsilon > 0$ for the stopping criterion.
3: **for** $i = 0, 1, \ldots$ **do**
4: Compute $\mathbf{H} \in \mathbb{C}^{N(\Gamma_0^{\pm})\times N(\Gamma_0^{\pm})}$ as solution of

$$\mathrm{D}\mathbf{F}(\mathbf{P}^{\pm}_{\mathrm{fb},(i)})\mathbf{H} = \mathbf{F}(\mathbf{P}^{\pm}_{\mathrm{fb},(i)}).$$

5: **if** $\|\mathbf{H}\| < \varepsilon$ **then**
6: **exit**
7: **end if**
8: Set $\widetilde{\mathbf{P}} = \mathbf{P}^{\pm}_{\mathrm{fb},(i)} - \mathbf{H}$.
9: Compute spectral radius s of $\left(\mathbf{T}^{\pm}_{\mathrm{fb}} + \mathbf{T}^{\pm}_{\mathrm{bb}}\widetilde{\mathbf{P}}\right)^{-1}\widetilde{\mathbf{P}}$.
10: **if** $s \leq 1$ **then**
11: Set $\mathbf{P}^{\pm}_{\mathrm{fb},(i+1)} = \widetilde{\mathbf{P}}$.
12: **else**
13: Set $\mathbf{P}^{\pm}_{\mathrm{fb},(i+1)} = \frac{1}{s}\widetilde{\mathbf{P}}$.
14: **end if**
15: **end for**

Using the matrix representations of the discrete local RtR operators and the discrete propagation operators, we can compute RtR matrices $\mathbf{R}^{\pm}(\omega, k) \in \mathbb{C}^{N(\Gamma_0^{\pm})\times N(\Gamma_0^{\pm})}$ with entries $R^{\pm}_{mn}$, $m, n \in \{1, \ldots, N(\Gamma_0^{\pm})\}$, that satisfy

$$\mathcal{R}^{\pm}_h(\omega, k) b_{\Gamma_0^{\pm},n} = \sum_{m=1}^{N(\Gamma_0^{\pm})} R^{\pm}_{mn} b_{\Gamma_0^{\pm},m},$$

such that

$$\mathbf{R}^{\pm} = \mathbf{T}^{\pm}_{\mathrm{fb}} + \mathbf{T}^{\pm}_{\mathrm{bb}}\mathbf{P}^{\pm}_{\mathrm{fb}},$$

cf. Eq. (7.13).

Definition and computation of the derivatives of the discrete Robin-to-Robin operators

The derivatives of the discrete local RtR operators (7.37) of order $m \in \mathbb{N}_0$ with respect to ω and order $n \in \mathbb{N}_0$ with respect to k, $m + n \geq 1$, are defined by

$$\mathcal{T}^{\pm,(m,n)}_{\mathrm{fb},h}(\omega, k)\varphi_h = 2\mathrm{i}\rho \quad u^{\pm,(m,n)}_{\mathrm{loc,h}}(\varphi_h, \quad 0)\big|_{\Gamma_0^{\pm}}, \tag{7.49a}$$

$$\mathcal{T}^{\pm,(m,n)}_{\mathrm{ff},h}(\omega, k)\varphi_h = 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1} u^{\pm,(m,n)}_{\mathrm{loc,h}}(\varphi_h, \quad 0)\big|_{\Gamma_1^{\pm}}, \tag{7.49b}$$

$$\mathcal{T}^{\pm,(m,n)}_{\mathrm{bb},h}(\omega, k)\varphi_h = 2\mathrm{i}\rho \quad u^{\pm,(m,n)}_{\mathrm{loc,h}}(0, \mathcal{S}^{\pm}\varphi_h)\big|_{\Gamma_0^{\pm}}, \tag{7.49c}$$

$$\mathcal{T}^{\pm,(m,n)}_{\mathrm{bf},h}(\omega, k)\varphi_h = 2\mathrm{i}\rho\,(\mathcal{S}^{\pm})^{-1} u^{\pm,(m,n)}_{\mathrm{loc,h}}(0, \mathcal{S}^{\pm}\varphi_h)\big|_{\Gamma_1^{\pm}}, \tag{7.49d}$$

for any $\varphi \in \mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$, where $u^{\pm,(m,n)}_{\mathrm{loc,h}}(\varphi_h, \psi_h) \in \mathsf{S}^p_{1\mathrm{p}}(C_1^{\pm})$ satisfy

$$\begin{aligned}
&\mathfrak{b}_{C_1^{\pm}}(u^{\pm,(m,n)}_{\mathrm{loc,h}}(\varphi_h, \psi_h), v; \omega, k) - \mathrm{i}\rho \sum_{j=0,1} \int_{\Gamma_j^{\pm}} u^{\pm,(m,n)}_{\mathrm{loc,h}}(\varphi_h, \psi_h)\, \overline{v}\, \mathrm{d}s(\mathbf{x}) \\
&\qquad = 2m\omega\, \mathfrak{m}^{\beta}_{C_1^{\pm}}(u^{\pm,(m-1,n)}_{\mathrm{loc,h}}(\varphi_h, \psi_h), v) + m(m-1)\mathfrak{m}^{\beta}_{C_1^{\pm}}(u^{\pm,(m-2,n)}_{\mathrm{loc,h}}(\varphi_h, \psi_h), v) \\
&\qquad\quad - 2nk\, \mathfrak{m}^{\alpha}_{C_1^{\pm}}(u^{\pm,(m,n-1)}_{\mathrm{loc,h}}(\varphi_h, \psi_h), v) - n\, \mathfrak{c}^{\alpha,1}_{C_1^{\pm}}(u^{\pm,(m,n-1)}_{\mathrm{loc,h}}(\varphi_h, \psi_h), v) \\
&\qquad\quad - n(n-1)\mathfrak{m}^{\alpha}_{C_1^{\pm}}(u^{\pm,(m,n-2)}_{\mathrm{loc,h}}(\varphi_h, \psi_h), v)
\end{aligned} \tag{7.50}$$

for all $v_h \in \mathsf{S}^p_{1\mathrm{p}}(C_1^\pm)$. Then we search for matrices $\mathbf{T}_{ij}^{\pm,(m,n)} \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$ with entries $T_{ij,pq}^{\pm,(m,n)}$, $p,q \in \{1,\dots,N(\Gamma_0^\pm)\}$ such that

$$\mathcal{T}_{ij,h}^{\pm,(m,n)} b_{\Gamma_0^\pm,q} = \sum_{p=1}^{N(\Gamma_0^\pm)} T_{ij,pq}^{\pm,(m,n)} b_{\Gamma_0^\pm,q}, \qquad i,j \in \{\mathrm{f},\mathrm{b}\}.$$

To this end, we solve (7.50) for $u_{\mathrm{loc,h}}^{\pm,(m,n)}(b_{\Gamma_0^\pm,\ell},0)$ and $u_{\mathrm{loc,h}}^{\pm,(m,n)}(0,b_{\Gamma_0^\pm,\ell})$, $1 \le \ell \le N(\Gamma_0^\pm)$, and collect the coefficient vectors with respect to the basis $b_{C_1^\pm,1},\dots,b_{C_1^\pm,N(C_1^\pm)}$ in matrices $\mathbf{U}_{\mathrm{loc},h,0}^{\pm,(m,n)} \in \mathbb{C}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$ and $\mathbf{U}_{\mathrm{loc},h,1}^{\pm,(m,n)} \in \mathbb{C}^{N(C_1^\pm)\times N(\Gamma_0^\pm)}$, respectively, i. e. we solve

$$\begin{aligned}\mathbf{S}_{C_1^\pm}(\omega,k)\,\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m',n')} &= \mathbf{M}_{C_1^\pm}^{\beta}\left(2m'\omega\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m'-1,n')} + m'(m'-1)\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m'-2,n')}\right)\\ &\quad - \mathbf{M}_{C_1^\pm}^{\alpha}\left(2n'k\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m',n'-1)} + n'(n'-1)\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m',n'-2)}\right)\\ &\quad - n'\mathbf{C}_{C_1^\pm}^{\alpha}\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m',n'-1)}\end{aligned}$$

for all $m' = 0,\dots,m$ and $n' = 0,\dots,n$ with $m'+n' \ge 1$, where $\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(0,0)} = \mathbf{U}_{\mathrm{loc},h,i}^{\pm}$, $i = 0,1$. According to (7.49) we have

$$\begin{aligned}\mathbf{T}_{\mathrm{fb}}^{\pm,(m,n)} &= 2\mathrm{i}\rho\mathbf{Q}_{C_1^\pm}^{0}\mathbf{U}_{\mathrm{loc},h,0}^{\pm,(m,n)}(\Gamma_0^\pm),\\ \mathbf{T}_{\mathrm{ff}}^{\pm,(m,n)} &= 2\mathrm{i}\rho\mathbf{Q}_{C_1^\pm}^{1}\mathbf{U}_{\mathrm{loc},h,0}^{\pm,(m,n)}(\Gamma_1^\pm),\\ \mathbf{T}_{\mathrm{bb}}^{\pm,(m,n)} &= 2\mathrm{i}\rho\mathbf{Q}_{C_1^\pm}^{0}\mathbf{U}_{\mathrm{loc},h,1}^{\pm,(m,n)}(\Gamma_0^\pm),\\ \mathbf{T}_{\mathrm{bf}}^{\pm,(m,n)} &= 2\mathrm{i}\rho\mathbf{Q}_{C_1^\pm}^{1}\mathbf{U}_{\mathrm{loc},h,1}^{\pm,(m,n)}(\Gamma_1^\pm),\end{aligned}$$

where $\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m,n)}(\Gamma_j^\pm)$, $i,j \in \{0,1\}$, denote the block matrices of $\mathbf{U}_{\mathrm{loc},h,i}^{\pm,(m,n)}$ with row indices $\mathfrak{S}(C_1^\pm,\Gamma_j^\pm)$.

The matrices $\mathbf{P}_{\mathrm{fb}}^{\pm,(m,n)}(\omega,k) \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$, i. e. the discrete versions of the derivatives of the forward-backward propagation operators, can simply be obtained when transferring the linear operator equation (7.30) into discrete form by replacing all operators with their corresponding matrices. Similarly to the procedure discussed in Chapter 6 for the computation of the derivatives of the Dirichlet propagation matrices, the resulting linear matrix equation can be transformed into a linear system of equation with $(N(\Gamma_0^\pm))^2$ unknowns, i. e. the entries of $\mathbf{P}_{\mathrm{ff}}^{\pm,(m,n)}(\omega,k)$, see [Lan70] for more details.

Finally, we obtain the derivatives of the discrete RtR operators

$$\mathbf{R}^{\pm,(m,n)} = \mathbf{T}_{\mathrm{fb}}^{\pm,(m,n)} + \sum_{(\mathbf{m},\mathbf{n})\in\mathfrak{N}^2(m,n)} \binom{m}{\mathbf{m}}\binom{n}{\mathbf{n}} \mathbf{T}_{\mathrm{bb}}^{\pm,(m_1,n_1)}\mathbf{P}_{\mathrm{fb}}^{\pm,(m_2,n_2)} \tag{7.51}$$

cf. Eq. (7.31), with the set $\mathfrak{N}^2(m,n)$ defined in (6.20).

7.2 Nonlinear eigenvalue problem with Robin-to-Robin operators

In the previous section we introduced RtR operators for periodic media, explained their computation and discretization. In this section we now want to show how to employ these operators in order to transform the linear (or quadratic) eigenvalue problem (2.19) on the unbounded domain S to a nonlinear eigenvalue problem posed in the defect cell C_0. We will start with the problem in strong formulation. After introducing a variational formulation, we will elaborate on the discretization of this nonlinear eigenvalue problem and finally, we present numerical solution techniques to solve the nonlinear eigenvalue problem in discretized form.

7.2.1 Main theorem

We start with the main result of the RtR method [Fli09].

Theorem 7.15. *The eigenvalue problem (2.19) posed in the unbounded domain S is equivalent to: find eigenvalue couples $(\omega^2, k) \in \mathbb{R}^+ \times B$, with $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$, such that there exists a non-trivial $u \in \mathsf{H}^1_{1\mathrm{p}}(\Delta, C_0, \alpha)$ that satisfies*

$$-(\nabla + \mathrm{i}k\binom{1}{0}) \cdot \alpha(\nabla + \mathrm{i}k\binom{1}{0})u - \omega^2 \beta\, u = 0 \qquad \text{in } C_0, \tag{7.52a}$$

$$(\mp\alpha\, \partial_2 + \mathrm{i}\rho)u = \mathcal{R}^{\pm}(\omega, k)(\pm\alpha\, \partial_2 + \mathrm{i}\rho)u \qquad \text{on } \Gamma_0^{\pm}. \tag{7.52b}$$

Note that the problem (7.52) — in comparison to problem (2.19) — is posed in the bounded domain C_0 but it is nonlinear with respect to ω and k due to the highly nonlinear dependence of the RtR operators on ω and k.

7.2.2 Mixed variational formulation

In order to derive a variational formulation of the nonlinear eigenvalue problem (7.52) with RtR operators, we introduce Lagrange multipliers $\lambda^{\pm} \in \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ defined by

$$\lambda^{\pm} = \pm\alpha\, \partial_2 u|_{\Gamma_0^{\pm}}$$

for the Neumann trace on $\Gamma_0^{\pm}$. Using the linearity of the RtR operators, we deduce that the Robin boundary condition (7.52b) can be rewritten in the form

$$(\mathcal{I} + \mathcal{R}^{\pm}(\omega, k)) \left(\pm\alpha\, \partial_2 u|_{\Gamma_0^{\pm}}\right) = \mathrm{i}\rho(\mathcal{I} - \mathcal{R}^{\pm}(\omega, k))u|_{\Gamma_0^{\pm}}, \tag{7.53}$$

where $\mathcal{I}$ denotes the identity operator. Thus, a mixed variational formulation of the nonlinear eigenvalue problem (7.52) reads: find eigenvalue couples $(\omega^2, k) \in \mathbb{R}^+ \times B$ with $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$, and associated eigenmodes $(u, \lambda^+, \lambda^-) \in \mathsf{H}^1_{1\mathrm{p}}(C_0) \times \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^+) \times \mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^-)$ such that

$$\mathfrak{b}_{C_0}(u, v; \omega, k) - \int_{\Gamma_0^+} \lambda^+\, \overline{v}\, \mathrm{d}s(\mathbf{x}) - \int_{\Gamma_0^-} \lambda^-\, \overline{v}\, \mathrm{d}s(\mathbf{x}) = 0, \tag{7.54a}$$

$$\mathrm{i}\rho \int_{\Gamma_0^+} \left(\mathcal{I} - \mathcal{R}^+(\omega, k)\right) u\, \overline{\psi^+}\, \mathrm{d}s(\mathbf{x}) - \int_{\Gamma_0^+} \left(\mathcal{I} + \mathcal{R}^+(\omega, k)\right) \lambda^+\, \overline{\psi^+}\, \mathrm{d}s(\mathbf{x}) = 0, \tag{7.54b}$$

$$\mathrm{i}\rho \int_{\Gamma_0^-} \left(\mathcal{I} - \mathcal{R}^-(\omega, k)\right) u\, \overline{\psi^-}\, \mathrm{d}s(\mathbf{x}) - \int_{\Gamma_0^-} \left(\mathcal{I} + \mathcal{R}^-(\omega, k)\right) \lambda^-\, \overline{\psi^-}\, \mathrm{d}s(\mathbf{x}) = 0, \tag{7.54c}$$

for all $(v, \psi^+, \psi^-) \in \mathsf{H}^1_{1\mathrm{p}}(C_0) \times \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^+) \times \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^-)$, where the sesquilinear form $\mathfrak{b}_{C_0}$ is given in Eq. (6.45a).

7.2.3 Variational formulation with Dirichlet-to-Neumann operators

Now we aim to derive an alternative variational formulation which employs DtN operators and which is — in contrast to the mixed variational formulation (7.54) — symmetric with respect to the trial and test spaces. However, this formulation is not well-posed at all frequencies in the band gap as one has to exclude the global Dirichlet eigenvalues.

Again, we use the fact that the RtR operators $\mathcal{R}^{\pm}(\omega, k)$ are linear and the Robin boundary condition (7.52b) can be rewritten in the form (7.53). Then we present an important result [Fli09] on the operator $(\mathcal{I} + \mathcal{R}^{\pm}(\omega, k))$, which also appears in Eqs. (7.54b)–(7.54c), where it is applied to the Lagrange multipliers $\lambda^{\pm}$.

Proposition 7.16. *Let $k \in B$ and $\omega^2 \in \mathbb{R}^+ \setminus \sigma^{\mathrm{ess}}(k)$. Furthermore, we assume that ω^2 is not a global Dirichlet eigenvalue, i. e. an eigenvalue of the infinite half-strip problem (7.1a) with homogeneous Dirichlet boundary condition on $\Gamma_0^{\pm}$. Then the operator $(\mathcal{I} + \mathcal{R}^{\pm}(\omega, k))$ is invertible.*

Proof. By definition of the RtR operator and the Robin problems in the infinite half-strips, the operator $(\mathcal{I} + \mathcal{R}^{\pm}(\omega, k))$ satisfies

$$(\mathcal{I} + \mathcal{R}^{\pm}(\omega, k))\, \varphi = 2\mathrm{i}\rho\, u^{\pm}(\cdot\,; \omega, k, \varphi)|_{\Gamma_0^{\pm}} \in \mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm}),$$

for any $\varphi \in \mathsf{H}_{1\mathrm{p}}^{-1/2}(\Gamma_0^\pm)$ where $u^\pm(\cdot\,;\omega,k,\varphi)$ is the unique solution of (7.1). Using the same ideas as in Lemma 7.3, we can show that for any $\varphi_{\mathrm{DtN}} \in \mathsf{H}_{1\mathrm{p}}^{1/2}(\Gamma_0^\pm)$ the inverse of $(\mathcal{I} + \mathcal{R}^\pm(\omega,k))$ is defined by

$$\varphi_{\mathrm{DtN}} \longmapsto \frac{(\pm\alpha\,\partial_2 + \mathrm{i}\rho)}{2\mathrm{i}\rho}\, u^\pm_{\mathrm{DtN}}(\varphi_{\mathrm{DtN}})\big|_{\Gamma_0^\pm}, \tag{7.55}$$

where $u^\pm_{\mathrm{DtN}}(\varphi_{\mathrm{DtN}})$ is the unique solution of the infinite half-strip problem (6.1) with Dirichlet boundary condition on $\Gamma_0^\pm$. Note that the solution $u^\pm_{\mathrm{DtN}}(\varphi_{\mathrm{DtN}})$ of (6.1) is in fact unique since we assume that $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$ is not a global Dirichlet eigenvalue. Hence, $(\mathcal{I} + \mathcal{R}^\pm(\omega,k))$ is invertible with inverse (7.55) if $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$ is not a global Dirichlet eigenvalue. □

Using Proposition 7.16 and Eq. (7.53), we can deduce that the Robin boundary condition (7.52b) is equivalent to the Neumann boundary condition

$$\pm\alpha\,\partial_2 u = \mathcal{D}^\pm_{\mathrm{RtR}}(\omega,k)u \qquad \text{on } \Gamma_0^\pm,$$

cf. Eq. (6.43b), with the DtN operator

$$\mathcal{D}^\pm_{\mathrm{RtR}}(\omega,k) = \mathrm{i}\rho\,(\mathcal{I} + \mathcal{R}^\pm(\omega,k))^{-1}\,(\mathcal{I} - \mathcal{R}^\pm(\omega,k)), \tag{7.56}$$

if ω^2 is not a global Dirichlet eigenvalue.

Then the derivation of the corresponding weak formulation of the problem with DtN operators is straightforward: find eigenvalue couples $(\omega^2,k) \in \mathbb{R}^+ \times B$, with $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$, and associated eigenmodes $u \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$ such that

$$\mathfrak{b}_{C_0}(u,v;\omega,k) - \mathfrak{d}_{\mathrm{RtR}}(u,v;\omega,k) = 0 \tag{7.57}$$

for all $v \in \mathsf{H}^1_{1\mathrm{p}}(C_0)$, with the sesquilinear form

$$\mathfrak{d}_{\mathrm{RtR}}(u,v;\omega,k) := \int_{\Gamma_0^+} \mathcal{D}^+_{\mathrm{RtR}}(\omega,k)u\,\overline{v}\,\mathrm{d}s(\mathbf{x}) + \int_{\Gamma_0^-} \mathcal{D}^-_{\mathrm{RtR}}(\omega,k)u\,\overline{v}\,\mathrm{d}s(\mathbf{x}), \tag{7.58}$$

cf. Eq. (6.45b).

Remark 7.17. *Let $k \in B$ and $\omega^2 \in \mathbb{R}^+ \setminus \sigma^{\mathrm{ess}}(k)$. Furthermore, let us assume that ω^2 is not a global Dirichlet eigenvalue, i. e. an eigenvalue of the infinite half-strip problem (7.1a) with homogeneous Dirichlet boundary condition on $\Gamma^\pm$. Then the DtN operator $\mathcal{D}^\pm_{\mathrm{RtR}}(\omega,k)$ is well-defined and can be computed according to Eq. (7.56). If, additionally, ω^2 is not equal to a local Dirichlet eigenvalue, i. e. an eigenvalue of the local cell problem (7.8a), then the DtN operator* (6.9) *denoted by $\mathcal{D}^\pm(\omega,k)$, that is based on the DtN approach as described in Chapter 6 and can be computed using local Dirichlet problems, exists and is well-defined. This implies — according to Theorem 7.15 and its analogue for the DtN approach, Theorem 6.18 — that $\mathcal{D}^\pm_{\mathrm{RtR}}(\omega,k) = \mathcal{D}^\pm(\omega,k)$ for all $(\omega^2,k) \in \mathbb{R}^+ \times B$ with $\omega^2 \notin \sigma^{\mathrm{ess}}(k)$ except for a countable set of frequencies — the global and local Dirichlet eigenvalues.*

7.2.4 Group velocity and higher derivatives of dispersion curves

In Section 6.2.3 we derived formulas for the group velocity and higher derivatives of the dispersion curve for problems with DtN transparent boundary conditions. This means that the formulas derived in Section 6.2.3 can directly be applied to the variational eigenvalue problem (7.57) with DtN transparent boundary conditions based on local Robin problems. For this, we only need to explain how we compute the partial derivatives of the DtN operators $\mathcal{D}^\pm_{\mathrm{RtR}}$ with respect to the frequency ω and the quasi-momentum k from the partial derivatives of the RtR operators $\mathcal{R}^\pm$, whose computation was explained in Section 7.1.3.

For the differentiation of (7.56) we can either apply Faà di Bruno's formula [FdB57] in combination with multinomial expansions, or we apply a recursion algorithm similar to Algorithm 5.1. Since the former approach leads to very complicated formulas, we prefer the second approach even though it does not yield closed formulas. First using binomial expansions, we find that the m-th ω- and n-th k-derivative

$$\mathcal{D}^{\pm,(m,n)}_{\mathrm{RtR}}(\omega,k) := \frac{\partial^{m+n}\mathcal{D}^\pm_{\mathrm{RtR}}(\omega,k)}{\partial\omega^m\partial k^n}$$

of $\mathcal{D}^{\pm}_{\mathrm{RtR}}$ reads

$$\mathcal{D}^{\pm,(m,n)}_{\mathrm{RtR}} = \mathrm{i}\rho \sum_{m'=0}^{m}\sum_{n'=0}^{n} \binom{m}{m'}\binom{n}{n'} \partial_\omega^{m'}\partial_k^{n'}\left((\mathcal{I}+\mathcal{R}^{\pm})^{-1}\right)\, \partial_\omega^{m-m'}\partial_k^{n-n'}(\mathcal{I}-\mathcal{R}^{\pm}). \tag{7.59}$$

While the second term in (7.59) can simply be expressed as

$$\partial_\omega^{m-m'}\partial_k^{n-n'}(\mathcal{I}-\mathcal{R}^{\pm}) = \begin{cases} \mathcal{I}-\mathcal{R}^{\pm}, & \text{if } m'=m \text{ and } n'=n,\\ -\mathcal{R}^{\pm,(m-m',n-n')}, & \text{otherwise,}\end{cases}$$

the evaluation of the first term is more involved. For this, we first note that for all $\ell \in \mathbb{N}$

$$\begin{aligned}\partial_k^n\left((\mathcal{I}+\mathcal{R}^{\pm})^{-\ell}\right) &= \partial_k^{n-1}\left(-\ell(\mathcal{I}+\mathcal{R}^{\pm})^{-\ell-1}\partial_\omega\mathcal{R}^{\pm}\right)\\ &= -\ell\sum_{n'=0}^{n-1}\binom{n-1}{n'}\partial_k^{n'}\left((\mathcal{I}+\mathcal{R}^{\pm})^{-\ell-1}\right)\partial_\omega^{n-n'}\mathcal{R}^{\pm}.\end{aligned} \tag{7.60}$$

Recursively applying (7.60), we find that we can express $\partial_k^n\left((\mathcal{I}+\mathcal{R}^{\pm})^{-1}\right)$ in a sum of products of $(\mathcal{I}+\mathcal{R}^{\pm})^{-1}$ and (higher order) partial derivatives of $\mathcal{R}^{\pm}$ with respect to k. Note that this can be obtained analogously by applying Faà di Bruno's formula to $\partial_k^n\left((\mathcal{I}+\mathcal{R}^{\pm})^{-1}\right)$. Now it remains to take the m-th derivative with respect to ω of this sum. Each summand is a product of several terms of $(\mathcal{I}+\mathcal{R}^{\pm})^{-1}$ and $\partial_k^{n'}\mathcal{R}^{\pm}$, $n' \in \mathbb{N}$. Hence, the m-th derivative with respect to ω of each product can be expressed in terms of a multinomial expansion.

For the mixed variational eigenvalue problem (7.54) with RtR operators, however, we have to extend the procedure, that was developed in Chapter 4 and transferred to problems with DtN transparent boundary conditions in Section 6.2.3, to non-self-adjoint problems. This was already done in discrete sense in Chapter 5, where we do not require the nonlinear matrix function to be Hermitian. Therefore, we shall refrain from revisiting the procedure developed in Chapter 5 and transferring it into variational sense.

However, we would like to comment on the requirements of the procedure in Chapter 5, and whether these are fulfilled in our case. Most important is the analyticity of the dispersion curves and the differentiability of the corresponding eigenmodes to any order. In Section 6.2.3 we argued that the analyticity of the dispersion curves and their corresponding eigenmodes of the variational formulation of the linear eigenvalue problem (2.19) in the infinite strip S, that was discussed in Chapter 4, directly transfers to variational problems of the form (6.44) with DtN transparent boundary conditions due to the equivalence of the variational formulations in the sense of Remark 6.19. Considering Theorem 7.15, the same is true for the dispersion curves and their corresponding eigenmodes of the eigenvalue problem (7.52) with RtR operators and the linear eigenvalue problem (2.19) in the infinite strip S. Assuming that this property is preserved by the mixed variational formulation (7.54), we can analogously to the case of DtN transparent boundary conditions, derive formulas for the group velocity and any higher derivative of the dispersion curves. For this we use the fact that the RtR operators are differentiable to any order with respect to the frequency ω and the quasi-momentum k, see Section 7.1.3.

7.2.5 Discretization

Now let us elaborate on the discretization of the variational formulations introduced above. In addition to the FE space $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ of polynomial degree p and dimension $N(\Gamma_0^{\pm})$, that was already used in Section 7.1.5 as discrete subspace of $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$, we recall $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ as the FE subspace of $\mathsf{H}^1_{1\mathrm{p}}(C_0)$ with total polynomial degree p and dimension $N(C_0)$ as introduced in Section 6.1.5. Furthermore, recall that we assumed in Section 7.1.5 that the material coefficients do not jump on the boundaries $\Gamma_0^{\pm}$ which implies that the Neumann and Robin traces on $\Gamma_0^{\pm}$ are in $\mathsf{H}^{1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$ and hence, we shall take $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^{\pm})$ as discrete subspace of $\mathsf{H}^{-1/2}_{1\mathrm{p}}(\Gamma_0^{\pm})$. Recall from Section 6.1.5 that we ordered the basis functions $b_{C_0,n}$, $n=1,\dots,N(C_0)$, of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ such that

- the basis functions with index $n \in \mathfrak{S}(C_0,\Gamma_0^+) = \{1,\dots,N(\Gamma_0^+)\}$ vanish on Γ_0^-, but their traces on Γ_0^+ build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^+)$,

- the basis functions with index $n \in \mathfrak{S}(C_0, \Gamma_0^-) = \{N(\Gamma_0^+)+1, \dots, N(\Gamma_0^+)+N(\Gamma_0^-)\}$ vanish on Γ_0^+, but their traces on Γ_0^- build a basis of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^-)$, and
- the basis functions with index $n \in \mathfrak{S}(C_0, C_0) = \{N(\Gamma_0^+)+N(\Gamma_0^-)+1, \dots, N(C_0)$ vanish on $\Gamma_0^\pm$.

With this ordering of the basis functions of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$, their traces on $\Gamma_0^\pm$ and the basis functions $b_{\Gamma_0^\pm}$, $n = 1, \dots, N(\Gamma_0^\pm)$, of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ are related through

$$b_{\Gamma_0^+,n} = \sum_{m=1}^{N(\Gamma_0^+)} Q^+_{C_0,mn} b_{C_0,m}|_{\Gamma_0^+},$$
$$b_{\Gamma_0^-,n} = \sum_{m=1}^{N(\Gamma_0^-)} Q^-_{C_0,mn} b_{C_0,N(\Gamma_0^+)+m}|_{\Gamma_0^-},$$

with permutation matrices $\mathbf{Q}^+_{C_0} \in \mathbb{R}^{N(\Gamma_0^+)\times N(\Gamma_0^+)}$ and $\mathbf{Q}^-_{C_0} \in \mathbb{R}^{N(\Gamma_0^-)\times N(\Gamma_0^-)}$, cf. Eq. (6.28).

Analogously to the discretization of the local cell problems in Section 7.1.5, we recall the matrix $\mathbf{B}_{C_0}(\omega,k) \in \mathbb{C}^{N(C_0)\times N(C_0)}$ from Eq. (6.56), and introduce the matrices $\mathbf{M}_{C_0,\Gamma_0^\pm} \in \mathbb{R}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}$ with entries

$$M_{C_0,\Gamma_0^\pm,mn}(k) = \int_{\Gamma_0^\pm} b_{C_0,n}\, \overline{b}_{C_0,m}\, \mathrm{d}s(\mathbf{x}),$$

$m, n \in \mathfrak{S}(C_0, \Gamma_0^\pm)$, related to the boundary integrals in Eq. (7.54).

With these definitions and the special ordering of the basis functions $b_{C_0,n}$ of the space $\mathsf{S}^p_{1\mathrm{p}}(C_0)$ described above, the discretization of the mixed variational formulation (7.54) reads

$$\begin{pmatrix} \mathbf{B}_{C_0} & -\mathbf{I}(C_0,\Gamma_0^+)\mathbf{M}_{C_0,\Gamma_0^+}(\mathbf{Q}^+_{C_0})^{\mathrm{T}} & -\mathbf{I}(C_0,\Gamma_0^-)\mathbf{M}_{C_0,\Gamma_0^-}(\mathbf{Q}^-_{C_0})^{\mathrm{T}} \\ \mathrm{i}\rho\mathbf{I}(\Gamma_0^+,C_0)\mathbf{M}_{C_0,\Gamma_0^+}(\mathbf{Q}^+_{C_0})^{\mathrm{T}}(\mathbf{I}-\mathbf{R}^+) & -\mathbf{M}_{C_0,\Gamma_0^+}(\mathbf{Q}^+_{C_0})^{\mathrm{T}}(\mathbf{I}+\mathbf{R}^+) & \mathbf{0} \\ \mathrm{i}\rho\mathbf{I}(\Gamma_0^-,C_0)\mathbf{M}_{C_0,\Gamma_0^-}(\mathbf{Q}^-_{C_0})^{\mathrm{T}}(\mathbf{I}-\mathbf{R}^-) & \mathbf{0} & -\mathbf{M}_{C_0,\Gamma_0^-}(\mathbf{Q}^-_{C_0})^{\mathrm{T}}(\mathbf{I}+\mathbf{R}^-) \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\lambda}_h^+ \\ \boldsymbol{\lambda}_h^- \end{pmatrix} = \mathbf{0}, \tag{7.61}$$

where $\mathbf{u} = \mathbf{u}(\omega,k) \in \mathbb{C}^{N(C_0)}$ is the coefficient vector of the discrete eigenmode $u_h(\cdot\,;\omega,k) \in \mathsf{S}^p_{1\mathrm{p}}(C_0)$ with respect to the basis functions $b_{C_0,n}$ of $\mathsf{S}^p_{1\mathrm{p}}(C_0)$, and $\boldsymbol{\lambda}_h^\pm = \boldsymbol{\lambda}_h^\pm(\omega,k) \in \mathbb{C}^{N(\Gamma_0^\pm)}$ are the coefficient vectors of the discrete Lagrange multipliers $\lambda_h^\pm(\cdot\,;\omega,k) \in \mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$ with respect to the basis functions $b_{\Gamma_0^\pm,n}$ of $\mathsf{S}^p_{1\mathrm{p}}(\Gamma_0^\pm)$, and the rectangular matrices $\mathbf{I}(\Gamma_0^\pm, C_0) \in \mathbb{R}^{N(\Gamma_0^\pm)\times N(C_0)}$ and $\mathbf{I}(C_0,\Gamma_0^\pm) \in \mathbb{R}^{N(C_0)\times N(\Gamma_0^\pm)}$ are the block matrices of the $N(C_0)\times N(C_0)$ identity matrix with row, respectively column, indices $\mathfrak{S}(C_0,\Gamma_0^\pm)$.

If we choose the variational formulation (7.57) with DtN operators instead of the mixed variational formulation (7.54) we obtain the discrete equation

$$\left(\mathbf{B}_{C_0}(\omega,k) - \mathbf{D}_{C_0,\mathrm{RtR}}(\omega,k)\right)\mathbf{u} = \mathbf{0}, \tag{7.62}$$

where the matrix $\mathbf{D}_{C_0,\mathrm{RtR}}(\omega,k) \in \mathbb{C}^{N(C_0)\times N(C_0)}$ is given by

$$\mathbf{D}_{C_0,\mathrm{RtR}}(\omega,k) = \begin{pmatrix} \mathbf{D}^+_{\mathrm{RtR}}(\omega,k) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^-_{\mathrm{RtR}}(\omega,k) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

with

$$\mathbf{D}^\pm_{\mathrm{RtR}}(\omega,k) = \mathrm{i}\rho\mathbf{M}_{C_0,\Gamma_0^\pm}(\mathbf{Q}^\pm_{C_0})^{\mathrm{T}}\left(\mathbf{I}+\mathbf{R}^\pm(\omega,k)\right)^{-1}\left(\mathbf{I}-\mathbf{R}^\pm(\omega,k)\right)\mathbf{Q}^\pm_{C_0} \in \mathbb{C}^{N(\Gamma_0^\pm)\times N(\Gamma_0^\pm)}. \tag{7.63}$$

On the continuous level we showed that the operator $(\mathcal{I}+\mathcal{R}^\pm(\omega,k))$ is only invertible if ω^2 is not a global Dirichlet eigenvalue, cf. Proposition 7.16. As long as the FE discretization is fine enough, we can equivalently observe that the matrix $(\mathbf{I}+\mathbf{R}^\pm(\omega,k))$ has full rank if and only if ω^2 is not a global Dirichlet eigenvalue. This implies that the DtN matrix (7.63) is only well-defined if ω^2 is not a global Dirichlet eigenvalue.

Again we have to note that — to the best of our knowledge — the stability of the FE discretizations of the two formulations has not yet been studied. However, numerical evidence shows that the standard asymptotic convergence estimates hold true for both formulations and thus, we can again use p-FEM on a coarse grid, cf. Figure 6.1, and can expect exponential convergence.

7.2.6 Numerical solution of the nonlinear eigenvalue problem

The numerical solution of the nonlinear eigenvalue problem (7.62) with DtN operators based on local Robin problems can be obtained analogously to the solution of the nonlinear eigenvalue problem (6.58) with DtN operators based on local Dirichlet problems. Since the RtR operators $\mathcal{R}^{\pm}(\omega, k)$ are differentiable with respect to ω and k, cf. Section 7.1.3, so are the DtN operators $\mathcal{D}^{\pm}_{\mathrm{RtR}}(\omega, k)$, and hence, all methods introduced in Section 3.2, in particular, the MSLP and the Chebyshev interpolation, can be applied to (7.62). But also the iterative Newton method, proposed in Section 3.3 and discussed in Section 6.2.5 for the case with DtN operators can be applied to (7.62).

On the other hand, the discretized nonlinear eigenvalue problem (7.61) with RtR operators can also be solved iteratively and directly. Due to the differentiability of the RtR operators the block matrix in (7.61) is differentiable with respect to ω and k up to any order. Hence, it is possible to apply the direct and indirect methods introduced in Section 3.2 in both formulations, the ω-formulation and the k-formulation.

Applying the Newton-like method, that we proposed in Section 3.3, to the nonlinear eigenvalue problem (7.61) is also possible. However, we cannot proceed analogously to Section 6.2.5, where we applied it to the nonlinear problem (6.58) with DtN transparent boundary conditions. The main ideas of the proposed Newton-type method is to transform the nonlinear eigenvalue problem into fixpoint problem. For this we introduce a linear eigenvalue problem with fixed RtR operators. Let $k_{\mathcal{R}} \in B$ and $\omega^2_{\mathcal{R}} \in \mathbb{R}^+ \setminus \sigma^{\mathrm{ess}}(k_{\mathcal{R}})$ be arbitrary but fixed. Then the problem: find $\omega^2 = \omega^2(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \in \mathbb{R}$ and a non-trivial $\mathbf{u}(\omega) \in \mathbb{C}^{N(C_0)} \setminus \{\mathbf{0}\}$ that satisfies

$$\Big(\widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) - \omega^2 \widetilde{\mathbf{B}}\Big) \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\lambda}_h^+ \\ \boldsymbol{\lambda}_h^- \end{pmatrix} = \mathbf{0}, \tag{7.64}$$

is a linear eigenvalue problem with matrices

$$\widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = \begin{pmatrix} \mathbf{A}^{\alpha}_{C_0} + k_{\mathcal{R}} \mathbf{C}^{\alpha,1}_{C_0} + k^2_{\mathcal{R}} \mathbf{M}^{\alpha}_{C_0} & -\mathbf{I}_{(C_0,\Gamma_0^+)} \mathbf{M}_{C_0,\Gamma_0^+} (\mathbf{Q}^+_{C_0})^{\mathrm{T}} & -\mathbf{I}_{(C_0,\Gamma_0^-)} \mathbf{M}_{C_0,\Gamma_0^-} (\mathbf{Q}^-_{C_0})^{\mathrm{T}} \\ \mathrm{i}\rho \mathbf{I}_{(\Gamma_0^+,C_0)} \mathbf{M}_{C_0,\Gamma_0^+} (\mathbf{Q}^+_{C_0})^{\mathrm{T}} (\mathbf{I} - \mathbf{R}^+) & -\mathbf{M}_{C_0,\Gamma_0^+} (\mathbf{Q}^+_{C_0})^{\mathrm{T}} (\mathbf{I} + \mathbf{R}^+) & \mathbf{0} \\ \mathrm{i}\rho \mathbf{I}_{(\Gamma_0^-,C_0)} \mathbf{M}_{C_0,\Gamma_0^-} (\mathbf{Q}^-_{C_0})^{\mathrm{T}} (\mathbf{I} - \mathbf{R}^-) & \mathbf{0} & -\mathbf{M}_{C_0,\Gamma_0^-} (\mathbf{Q}^-_{C_0})^{\mathrm{T}} (\mathbf{I} + \mathbf{R}^-) \end{pmatrix}$$

and

$$\widetilde{\mathbf{B}} = \begin{pmatrix} \mathbf{M}^{\beta}_{C_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{A}^{\alpha}_{C_0}, \mathbf{C}^{\alpha,1}_{C_0}, \mathbf{M}^{\alpha}_{C_0}, \mathbf{M}^{\beta}_{C_0} \in \mathbb{R}^{N(C_0) \times N(C_0)}$ are defined in (6.57) and $\widetilde{\mathbf{R}}^{\pm} = \mathbf{R}^{\pm}(\omega_{\mathcal{R}}, k_{\mathcal{R}})$. However, this linear eigenvalue problem does not satisfy the requirements given in Section 3.3 as $\widetilde{\mathbf{B}}$ is not regular. Hence, we shall instead substitute $\lambda = -\omega^{-2}$ and solve the linear eigenvalue problem

$$\Big(\widetilde{\mathbf{B}} + \lambda \widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}})\Big) \widetilde{\mathbf{u}} = \mathbf{0} \tag{7.65a}$$

with right eigenvector $\widetilde{\mathbf{u}} \in \mathbb{C}^{N(C_0)+N(\Gamma_0^+)+N(\Gamma_0^-)}$

$$\widetilde{\mathbf{v}}^{\mathrm{H}} \Big(\widetilde{\mathbf{B}} + \lambda \widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}})\Big) = \mathbf{0} \tag{7.65b}$$

with left eigenvector $\widetilde{\mathbf{v}} \in \mathbb{C}^{N(C_0)+N(\Gamma_0^+)+N(\Gamma_0^-)}$. Numerical evidence shows that the matrix $\widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}})$ is regular for all $\omega_{\mathcal{R}} \in \mathbb{R}^+$ and $k_{\mathcal{R}} \in B$ with $\omega^2_{\mathcal{R}} \notin \sigma^{\mathrm{ess}}_h(k_{\mathcal{R}})$, as long as the discretization is fine enough, and hence, we can solve (7.65) for its eigenvalue λ closest to $\lambda_{\mathcal{R}} := -\omega^{-2}_{\mathcal{R}}$ using a shift and invert strategy. If λ is an eigenvalue of (7.65) with $\lambda = -\omega^{-2}_{\mathcal{R}}$, then $(\omega^2_{\mathcal{R}}, k_{\mathcal{R}})$ is an eigenvalue couple of the nonlinear eigenvalue problem (7.61).

In Chapter 6 we also introduced a linear eigenvalue problem with fixed DtN operators, see Eq. (6.60). For that problem we could show that the eigenvalues are real and continuously differentiable with respect to both parameters $\omega_{\mathcal{D}}$ and $k_{\mathcal{D}}$. For the case of RtR transparent boundary conditions, numerical evidence shows that the eigenvalues of (7.65) are real and can be ordered such that the functions

$(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \mapsto \lambda_j(\omega_{\mathcal{R}}, k_{\mathcal{R}})$ are continuously differentiable, but we cannot give a rigorous proof for these results. We can only note that, considering the equivalence of the DtN and RtR approach in the sense of Eq. (7.56) and Remark 7.17, we deduce that the desired properties also hold true for the nonlinear eigenvalue problem (7.54) in mixed variational formulation if $\omega_{\mathcal{R}}^2$ is not a global or local Dirichlet eigenvalue, i. e. an eigenvalue of the infinite half-strip problem (7.1a) with homogeneous Dirichlet boundary conditions, or an eigenvalue of the local cell problem (7.8a) with homogeneous Dirichlet boundary conditions, respectively. Numerical evidence then shows that these properties are inherited by the discrete eigenvalue problem (7.65).

Let us now directly introduce the global signed distance function

$$d_{\mathrm{RtR}}^{\mathrm{g}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = \omega_{\mathcal{R}}^2 - \omega_{j^*}^2(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \tag{7.66}$$

where

$$j^* = j^*(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = \arg \min_{1 \le j \le N(C_0)+N(\Gamma_0^+)+N(\Gamma_0^-)} |\omega_{\mathcal{R}}^2 - \omega_j^2(\omega_{\mathcal{R}}, k_{\mathcal{R}})|$$

and

$$\omega_j^2(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = -\lambda_j^{-1/2}(\omega_{\mathcal{R}}, k_{\mathcal{R}}). \tag{7.67}$$

Similarly to the global signed distance function (6.62) for the case with DtN operators, $d_{\mathrm{RtR}}^{\mathrm{g}}$ can at most be piecewise differentiable.

The Newton method applied to $d_{\mathrm{RtR}}^{\mathrm{g}}$ works similarly to the case with DtN operators as sketched in Algorithm 6.1 for the ω-formulation, i. e. when keeping $k_{\mathcal{R}}$ fixed and searching for a frequency such that $d_{\mathrm{RtR}}^{\mathrm{g}} = 0$, or as in Algorithm 6.2 for the k-formulation, i. e. when keeping $\omega_{\mathcal{R}}$ fixed and searching for a quasi-momentum such that $d_{\mathrm{RtR}}^{\mathrm{g}} = 0$.

However, the computation of the derivative of the global signed distance function is slightly more involved due to the substitution (7.67) and the fact that (7.65) is not Hermitian, which implies that we have to solve (7.65) also for its left eigenvectors $\widetilde{\mathbf{v}}$. Considering that

$$\frac{\partial}{\partial \lambda_{\mathcal{R}}} \widetilde{\mathbf{A}} = \frac{\partial \omega_{\mathcal{R}}}{\partial \lambda_{\mathcal{R}}} \frac{\partial}{\partial \omega_{\mathcal{R}}} \widetilde{\mathbf{A}} = -\frac{1}{2} \omega_{\mathcal{R}}^3 \frac{\partial}{\partial \omega_{\mathcal{R}}} \widetilde{\mathbf{A}},$$

and $\frac{\partial}{\partial \lambda_{\mathcal{R}}} \widetilde{\mathbf{B}} = \mathbf{0}$, we obtain

$$\frac{\partial}{\partial \omega_{\mathcal{R}}} d_{\mathrm{RtR}}^{\mathrm{g}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = 2\omega_{\mathcal{R}} - \omega_{j^*}^2 \frac{\widetilde{\mathbf{v}}_{j^*}^{\mathrm{H}} \widetilde{\mathbf{A}}_{\omega_{\mathcal{R}}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \widetilde{\mathbf{u}}_{j^*}}{\widetilde{\mathbf{v}}_{j^*}^{\mathrm{H}} \widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \widetilde{\mathbf{u}}_{j^*}},$$

where

$$\widetilde{\mathbf{A}}_{\omega_{\mathcal{R}}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathrm{i}\rho \mathbf{I}(\Gamma_0^+, C_0) \mathbf{M}_{C_0, \Gamma_0^+} (\mathbf{Q}_{C_0}^+)^{\mathrm{T}} \widetilde{\mathbf{R}}_{\omega}^+ & -\mathbf{M}_{C_0, \Gamma_0^+} (\mathbf{Q}_{C_0}^+)^{\mathrm{T}} \widetilde{\mathbf{R}}_{\omega}^+ & \mathbf{0} \\ -\mathrm{i}\rho \mathbf{I}(\Gamma_0^-, C_0) \mathbf{M}_{C_0, \Gamma_0^-} (\mathbf{Q}_{C_0}^-)^{\mathrm{T}} \widetilde{\mathbf{R}}_{\omega}^- & \mathbf{0} & -\mathbf{M}_{C_0, \Gamma_0^-} (\mathbf{Q}_{C_0}^-)^{\mathrm{T}} \widetilde{\mathbf{R}}_{\omega}^- \end{pmatrix}.$$

On the other hand, the derivative of $d_{\mathrm{RtR}}^{\mathrm{g}}$ with respect to the parameter $k_{\mathcal{R}}$ reads

$$\frac{\partial}{\partial k_{\mathcal{R}}} d_{\mathrm{RtR}}^{\mathrm{g}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = -\omega_{j^*}^2 \frac{\widetilde{\mathbf{v}}_{j^*}^{\mathrm{H}} \widetilde{\mathbf{A}}_{k_{\mathcal{R}}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \widetilde{\mathbf{u}}_{j^*}}{\widetilde{\mathbf{v}}_{j^*}^{\mathrm{H}} \widetilde{\mathbf{A}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) \widetilde{\mathbf{u}}_{j^*}},$$

where

$$\widetilde{\mathbf{A}}_{k_{\mathcal{R}}}(\omega_{\mathcal{R}}, k_{\mathcal{R}}) = \begin{pmatrix} \mathbf{C}_{C_0}^{\alpha,1} + 2k_{\mathcal{R}} \mathbf{M}_{C_0}^{\alpha} & \mathbf{0} & \mathbf{0} \\ -\mathrm{i}\rho \mathbf{I}(\Gamma_0^+, C_0) \mathbf{M}_{C_0, \Gamma_0^+} (\mathbf{Q}_{C_0}^+)^{\mathrm{T}} \widetilde{\mathbf{R}}_{k}^+ & -\mathbf{M}_{C_0, \Gamma_0^+} (\mathbf{Q}_{C_0}^+)^{\mathrm{T}} \widetilde{\mathbf{R}}_{k}^+ & \mathbf{0} \\ -\mathrm{i}\rho \mathbf{I}(\Gamma_0^-, C_0) \mathbf{M}_{C_0, \Gamma_0^-} (\mathbf{Q}_{C_0}^-)^{\mathrm{T}} \widetilde{\mathbf{R}}_{k}^- & \mathbf{0} & -\mathbf{M}_{C_0, \Gamma_0^-} (\mathbf{Q}_{C_0}^-)^{\mathrm{T}} \widetilde{\mathbf{R}}_{k}^- \end{pmatrix}.$$

Note that the fact that $\widetilde{\mathbf{B}}$ is singular also implies that (7.65) has $2N(\Gamma_0^{\pm})$ zero eigenvalues and, hence, the computation of the eigenvalues of (7.65) with smallest magnitude is not meaningful, i. e. we always should assign a non-zero shift when applying a shift and invert strategy.

7.3 Numerical results

In the numerical examples we want to study the performance of the RtR method in comparison to the DtN method when applied to the computation of eigenvalues that are close to local or global Dirichlet eigenvalues. The DtN operators are not well-defined at global Dirichlet eigenvalues and their computation using local Dirichlet problems is ill-posed at local Dirichlet eigenvalues. In the numerical results of the DtN method presented in Section 6.3 we showed that DtN transparent boundary conditions based on local Dirichlet problems in fact lead to severe convergence problems when solving for the eigenvalues of the nonlinear eigenvalue problem.

The setup that we will discuss in this section is again the one presented in Example 2, i. e. we study the TE mode band structure of a PhC W1 waveguide with hexagonal lattice, relative permittivity $\varepsilon = 11.4$ and holes of relative radius $\frac{r}{a_1} = 0.31$. The polynomial degree of the FE computations set to $p = 5$ in all following numerical experiments.

As mentioned in Section 7.1.2, we aim to choose the constant $\rho \in \mathbb{R}\backslash\{0\}$ such that the auxiliary local cell problem (7.14) is well-posed for all values of (ω^2, k) under consideration. For the setup described above $\rho = 3$ seems to be a reasonable choice as can be seen from Figure 7.1, where we show the eigenvalues of the auxiliary local cell problem (7.14) in comparison to the global signed distance function $d^{\mathrm{g}}_{\mathrm{RtR}}$ representing the band structure of the PhC waveguide. Note that there are no eigenvalues of the auxiliary local cell problem in the second band gap of the PhC waveguide, which is the area we will focus in all following numerical experiments.

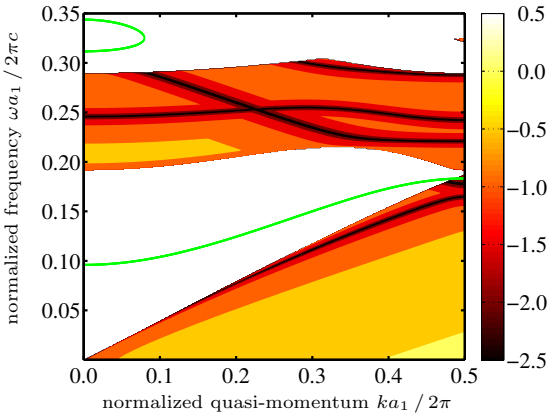


Figure 7.1: Magnitude of global signed distance function $d^{\mathrm{g}}_{\mathrm{RtR}}$ in logarithmic scale evaluated on a grid of 350×500 (ω, k)-points. The areas left blank correspond to the essential spectrum, i. e. $\omega^2 \in \sigma_h^{\mathrm{ess}}(k)$. The green curves represent eigenvalues of the auxiliary local cell problem (7.14).

7.3.1 Computation of global Dirichlet eigenvalues

Before we will analyse the convergence towards eigenvalues of the nonlinear eigenvalue problem (7.61), that are close to global or local Dirichlet eigenvalues, we consider the computation of global Dirichlet eigenvalues, i. e. Dirichlet eigenvalues in semi-infinite periodic strips. By construction, this computation is not possible when using DtN transparent boundary conditions. Using RtR transparent boundary conditions, however, resolves this problem, and global Dirichlet eigenvalues can be computed. Global Dirichlet eigenvalues also have a certain physical meaning. If the periodic medium of a semi-infinite 2d PhC is connected to a perfectly conducting magnetic (TE mode) or electric (TM mode) material, this can be modelled mathematically by homogeneous Dirichlet boundary conditions at the interface. The eigenmodes corresponding to global Dirichlet eigenvalues are confined at the surface of the semi-infinite 2d PhC. In this respect these eigenmodes can be called *surface modes*. However, since the term *surface mode* usually refers to modes that are confined at the interface towards air or vacuum [JJWM08], we

shall denote the eigenmodes associated to global Dirichlet eigenvalues by *Dirichlet surface modes*.

Using RtR operators as introduced in this chapter, the Dirichlet surface mode problem can be reduced to a nonlinear eigenvalue problem on the interface, i. e. to a problem in 1d. However, for the sake of simplicity and in order to reuse the procedures proposed for PhC waveguides, we consider the 2d problem. This means we assume the permittivity of the defect cell C_0 to be identical to the permittivity of the cells C_i^-, $i \in \mathbb{N}$, of the semi-infinite PhC below the line defect, which is now — to be precise — not a line defect anymore. Then we impose homogeneous Dirichlet boundary conditions at Γ_0^+ and RtR transparent boundary conditions at Γ_0^-. On a discrete level this can be realized by replacing the second block of rows in the system matrix of Eq. (7.61) by

$$\left(\mathbf{M}_{C_0,\Gamma_0^+}(\mathbf{Q}_{C_0}^+)^{\mathrm{T}} \quad \mathbf{0}\right) \in \mathbb{R}^{N(\Gamma_0^+)\times(N(C_0)+N(\Gamma_0^+)+N(\Gamma_0^-)}.$$

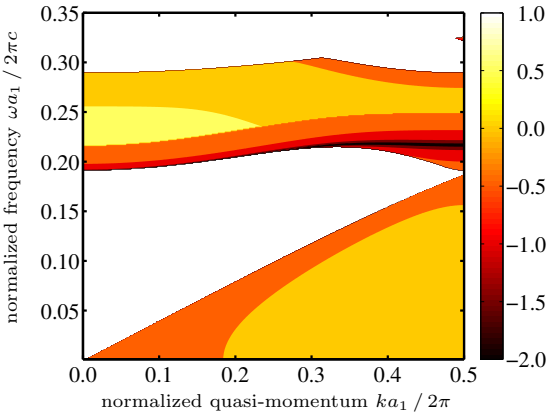


Figure 7.2: Magnitude of global signed distance function $d^{\mathrm{g}}_{\mathrm{RtR}}$ in logarithmic scale of Dirichlet surface mode problem evaluated on a grid of 350×500 (ω, k)-points.

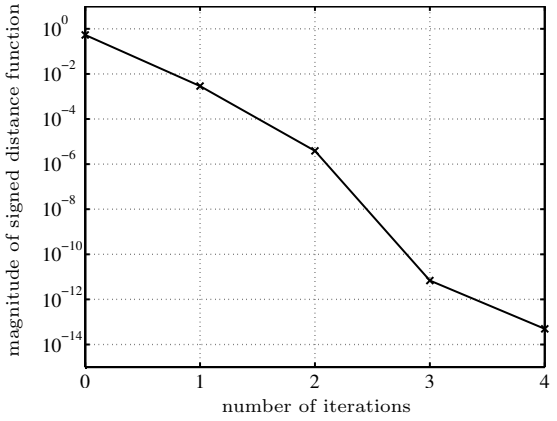


Figure 7.3: Convergence of the signed distance function $|d^{\mathrm{g}}_{\mathrm{RtR}}|$ when applying the Newton method in ω-formulation to the computation of the Dirichlet surface mode at $\omega \approx 0.218\cdot\frac{2\pi c}{a_1}$ and $k = 0.4\cdot\frac{2\pi}{a_1}$. The start value of the iterative schemes is chosen to be $\omega^{(0)} = 0.24\cdot\frac{2\pi c}{a_1}$.

In Figure 7.2 we show the magnitude of the global signed distance function $d^{\mathrm{g}}_{\mathrm{RtR}}$. Recall that the dark lines correspond to small values of $|d^{\mathrm{g}}_{\mathrm{RtR}}|$, and hence represent Dirichlet surface modes.

For the Dirichlet surface mode in the second band gap at $k = 0.4\cdot\frac{2\pi}{a_1}$ we show exemplarily the convergence of the Newton method as proposed in Section 7.2.6. The results are presented in Figure 7.3, where we chose the start value $\omega^{(0)} = 0.24\cdot\frac{2\pi c}{a_1}$. We can see the method converges exponentially towards the eigenvalue at $\omega \approx 0.218\cdot\frac{2\pi c}{a_1}$ with rate slightly larger than quadratic.

In Figure 7.4 the real part of the Dirichlet surface mode at $\omega \approx 0.218\cdot\frac{2\pi c}{a_1}$ and $k = 0.4\cdot\frac{2\pi}{a_1}$ is plotted. It demonstrates well that Dirichlet surface modes are modes that are guided at the surface and decay exponentially in the periodic medium.

7.3.2 Condition of system and Dirichlet-to-Neumann matrices

Similarly to Section 6.3.4, where we studied the condition of the system matrix and the DtN matrix of the problem (6.58) with DtN transparent boundary conditions based on local Dirichlet problems, let us now analyse the condition of the corresponding matrices in the case of RtR transparent boundary conditions and DtN transparent boundary conditions based on local Robin problems, respectively.

Let $\mathbf{N}_{\mathrm{DtN}}$ denote the system matrix of the left hand side of the nonlinear eigenvalue problem (6.58) with DtN maps that are based on local Dirichlet problems. On the other hand, we shall denote the system matrix of the nonlinear eigenvalue problem (7.61) with RtR maps by $\mathbf{N}_{\mathrm{RtR}}$.

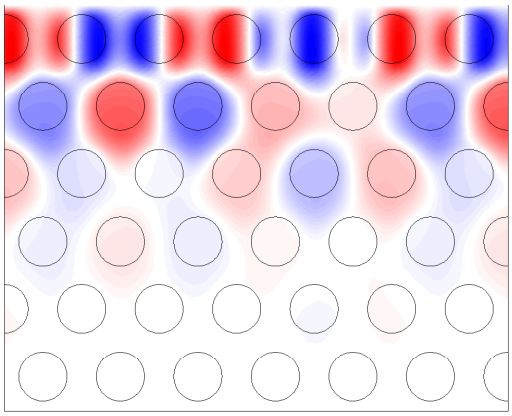
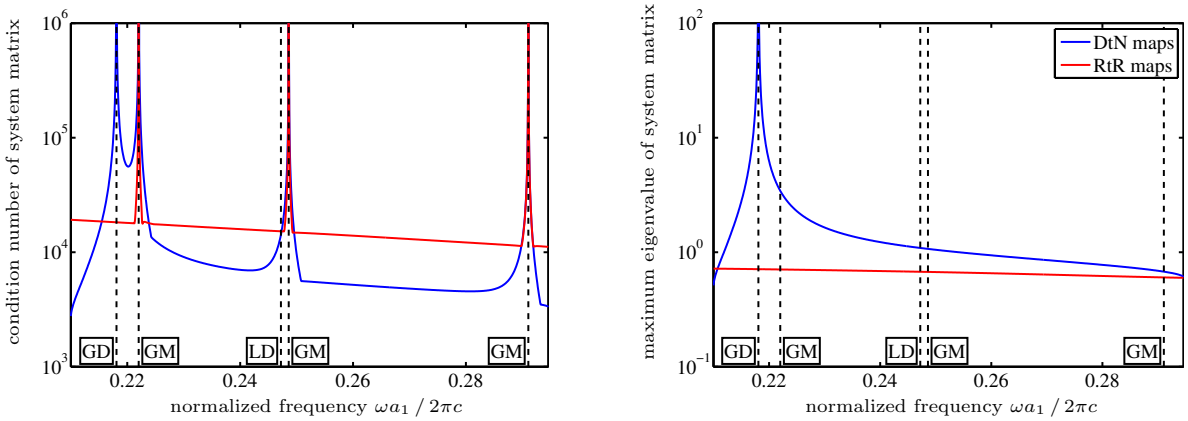


Figure 7.4: Real part of the Dirichlet surface mode at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$ and $k = 0.4 \cdot \frac{2\pi}{a_1}$.



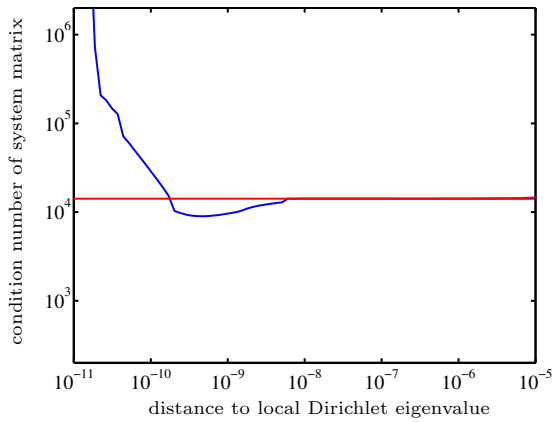
(a) Condition numbers of system matrices.

(b) Maximum eigenvalues of system matrices.

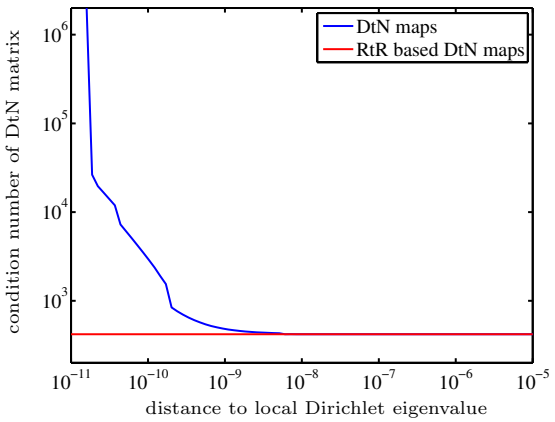
Figure 7.5: Condition number (a) and maximum eigenvalue (b) of the system matrix $\mathbf{N}_{\mathrm{DtN}}$ with DtN maps (blue) and the system matrix $\mathbf{N}_{\mathrm{RtR}}$ with RtR maps (red) in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$. The vertical dashed lines show the frequency of the global Dirichlet eigenvalue (GD), the local Dirichlet eigenvalue (LD) and the frequencies of the guided modes (GM).

In Figure 6.13 we presented the condition number and the maximum eigenvalue of the DtN system matrix $\mathbf{N}_{\mathrm{DtN}}$ in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$. In Figure 7.5 we show the condition number and the maximum eigenvalue of the RtR system matrix $\mathbf{N}_{\mathrm{RtR}}$ in comparison to the results for $\mathbf{N}_{\mathrm{DtN}}$. Again the vertical dashed lines show the location of the three guided modes (labeled "GM"), the global Dirichlet eigenvalue (labeled "GD") and the local Dirichlet eigenvalue (labeled "LD"). Similarly to $\mathbf{N}_{\mathrm{DtN}}$ the condition number of $\mathbf{N}_{\mathrm{RtR}}$ increases at the guided modes, see Figure 7.5a, since the minimum eigenvalue decreases in the vicinity of an eigenvalue of the nonlinear eigenvalue problem (7.61). As we already saw in Chapter 6, the condition number of the system matrix $\mathbf{N}_{\mathrm{DtN}}$ with DtN maps also increases in the vicinity of the global Dirichlet eigenvalue, which is due to an increasing maximum eigenvalue of $\mathbf{N}_{\mathrm{DtN}}$, that is presented in 7.5b. On the other hand, the condition number of the system matrix $\mathbf{N}_{\mathrm{RtR}}$ with RtR maps as well as its maximum eigenvalue do not increase in the vicinity of the global Dirichlet eigenvalue. In fact, the maximum eigenvalue of $\mathbf{N}_{\mathrm{RtR}}$ remains almost constant in the band gap, see Figure 7.5b.

Now let us study the condition number of the system matrix of the problem (7.62) with DtN maps based on local Robin problems and the condition number of the corresponding DtN matrices $\mathbf{D}^{\pm}_{C_0,\mathrm{RtR}}$ in a very small vicinity of the local Dirichlet eigenvalue in more detail. We already saw in Section 6.3.4 that the condition numbers of their counterparts based on local Dirichlet problems increase significantly in a very small vicinity of the local Dirichlet eigenvalue. Figures 7.6a and 7.6b show the condition numbers of



(a) Condition numbers of system matrices.



(b) Condition numbers of DtN matrices.

Figure 7.6: Condition numbers of the system matrices (a) of the problem (6.58) with DtN transparent boundary conditions based on local Dirichlet problems (blue) and of the problem (7.62) with DtN transparent boundary conditions based on local Robin problems (red), and condition numbers of the corresponding DtN matrices (b) in the vicinity of the local Dirichlet eigenvalue in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$.

the two matrices in dependence on the distance to the local Dirichlet eigenvalue and in comparison to the results of the matrices based on local Dirichlet problems, that were already presented in Figure 6.16. As expected, the condition numbers of the matrices related to local Robin problems remain constant in the vicinity of the local Dirichlet eigenvalue, while the condition numbers of the matrices that are based on local Dirichlet problems increase dramatically in a small vicinity of the local Dirichlet eigenvalue. Again we want to point out that the increase of the condition number of the matrices that are based on local Dirichlet problems is limited to a very narrow vicinity of the local Dirichlet eigenvalue.

7.3.3 Computation of eigenvalues in vicinity of global Dirichlet eigenvalues

In Section 6.3.5 we showed that the proposed Newton-like method applied to the nonlinear eigenvalue problem (6.58) with DtN transparent boundary conditions does not show any convergence problems even in the presence of global Dirichlet eigenvalues, see Figure 6.17. However, we could show that the frequently used MSLP suffers from a reduced radius of convergence that is limited by the global Dirichlet eigenvalue.

For the nonlinear eigenvalue problem (7.61) with RtR transparent boundary conditions we do not expect such a behaviour, since we could see in the previous section that the condition number of the system matrix of (7.61) does not increase in the vicinity of the global Dirichlet eigenvalue, see Figure 7.5a.

In Figure 7.7 we present the step sizes of the MSLP and the Newton method in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$. The vertical, dashed lines show the locations of the guided modes, i. e. the eigenvalues of the nonlinear eigenvalue problem (7.61). Similarly to the results for DtN transparent boundary conditions presented in Figure 6.17, both step size curves have roots with negative slope at the guided modes which means that the methods will converge to these eigenvalues. As expected the two curves do not change their behaviour at the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$, which is in contrast to the results with DtN transparent boundary conditions, where the step size curve of the MSLP has another root.

Now let us study the behaviour of the Chebyshev interpolation for the computation of the guided mode in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$ that is closest to the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$. In Figure 6.18 we saw that the Chebyshev interpolation applied to the problem with DtN transparent boundary conditions does not converge. Since the system matrix $\mathbf{N}_{\mathrm{RtR}}$ is not spoilt by the presence of the global Dirichlet eigenvalue, see Figure 7.5, we expect the Chebyshev interpolation applied to the problem (7.61) with RtR transparent boundary conditions to converge. In Figure 7.8 we show

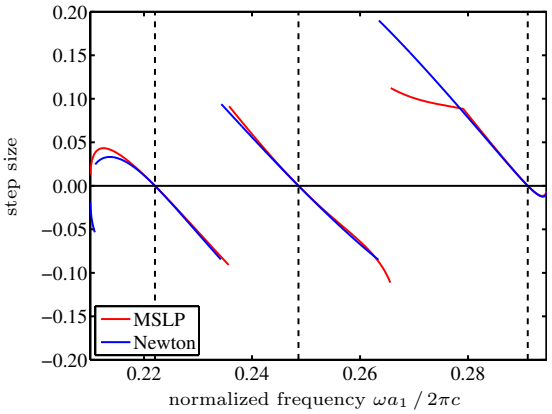


Figure 7.7: Step sizes of the MSLP (red) and the Newton method applied to the global signed distance function (blue) in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$. The vertical, dashed lines show the locations of the guided modes.

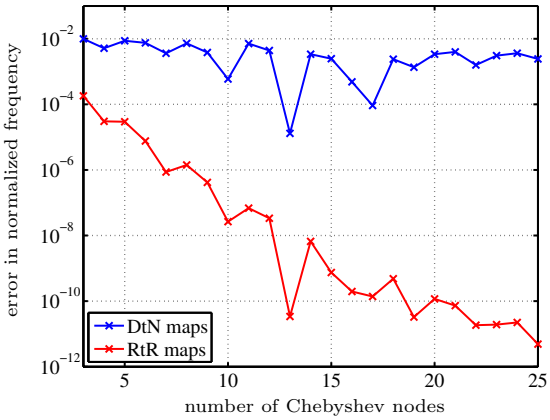


Figure 7.8: Absolute error of the Chebyshev interpolation in ω-formulation when applied to the nonlinear eigenvalue problem with DtN maps (blue) and RtR maps (red) for the computation of the guided mode in the second band gap at $k = 0.4 \cdot \frac{2\pi}{a_1}$ that is closest to the global Dirichlet eigenvalue at $\omega \approx 0.218 \cdot \frac{2\pi c}{a_1}$. The ω-interval of the interpolation is chosen to be $[0.215 \cdot \frac{2\pi c}{a_1}, 0.245 \cdot \frac{2\pi c}{a_1}]$.

the magnitude of the error in the normalized frequency of the Chebyshev interpolation applied to the problem with DtN maps (blue) and with RtR maps (red). In fact, we see that the error of the Chebyshev interpolation applied to (7.61) converges, while the error of the Chebyshev interpolation applied to the problem with DtN transparent boundary conditions does not converge. Again note that also in the case of RtR maps, the error does not decrease monotonically since the Chebyshev nodes are not hierarchical.

The results in Figures 7.7 and 7.8 show that RtR transparent boundary conditions effectively resolve the convergence problems of the numerical schemes applied to the nonlinear eigenvalue problem with DtN transparent boundary conditions in the vicinity of global Dirichlet eigenvalues.

7.3.4 Computation of eigenvalues in vicinity of local Dirichlet eigenvalues

Finally, we want to analyse the behaviour of the Newton method close to a local Dirichlet eigenvalue when applied to the nonlinear eigenvalue problem (7.62) with DtN transparent boundary conditions, that are based on local Robin problems, and when applied to the nonlinear eigenvalue problem (6.58) with DtN transparent boundary conditions, that are based on local Dirichlet problems.

In Section 6.3.6 we already saw that the Newton method applied to the problem (6.58) with DtN transparent boundary conditions, that are based on local Dirichlet problems, does not converge to a common eigenvalue of the nonlinear eigenvalue problem and the local Dirichlet problem, see Figure 6.19a. On the other hand, we also showed in Section 6.3.6, that the Chebyshev interpolation may converge to such a common eigenvalue, even though when applied to the nonlinear eigenvalue problem with DtN transparent boundary conditions based on local Dirichlet problems, see Figure 6.19b.

Since the computation of DtN maps using local Robin problems is — in contrast to the computation of DtN maps using local Dirichlet problems — not ill-posed at local Dirichlet eigenvalues, we expect either method to converge to a common eigenvalue of the nonlinear eigenvalue problem and the local Dirichlet problem.

For orientation we show in Figure 7.9 the magnitude of the global signed distance function $d^{\mathrm{g}}_{\mathrm{RtR}}$. The dark lines indicate small values of $|d^{\mathrm{g}}_{\mathrm{RtR}}|$ and therefore, represent the eigenvalues of the nonlinear eigenvalue problem (7.61). The green lines, on the other hand, show the local Dirichlet eigenvalues.

In Figure 7.10 we present the convergence of the Newton method to the common eigenvalue $\omega \approx$

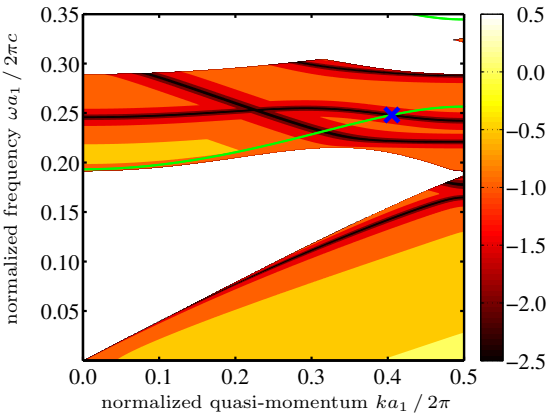


Figure 7.9: Magnitude of global signed distance function $d^{\mathrm{g}}_{\mathrm{RtR}}$ in logarithmic scale evaluated on a grid of 350×500 (ω, k)-points. The green lines represent the local Dirichlet eigenvalues, i. e. eigenvalues of the local cell problem (7.8a) with homogeneous Dirichlet boundary conditions. The blue cross indicates the location of the eigenvalue for which convergence results are shown in Figures 7.10, 7.11 and 7.12.

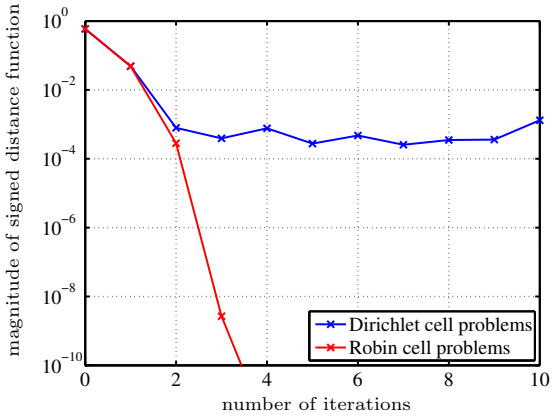


Figure 7.10: Convergence of the signed distance function $|d^{\mathrm{g}}|$ when applying the Newton method with DtN operators based on local Dirichlet problems (blue) and local Robin problems (red) to the computation of the common eigenvalue $(\omega, k) \approx (0.248 \cdot \frac{2\pi c}{a_1}, 0.405 \cdot \frac{2\pi}{a_1})$ of the Dirichlet cell problem and the nonlinear eigenvalue problem (blue cross in Figure 7.9). The start value of the iterative schemes is chosen to be $\omega^{(0)} = 0.263 \cdot \frac{2\pi c}{a_1}$.

$0.248 \cdot \frac{2\pi c}{a_1}$ at $k \approx 0.405 \cdot \frac{2\pi}{a_1}$ of the local Dirichlet problem and the nonlinear eigenvalue problem of the PhC waveguide, whose location in the band structure is marked with a blue cross in Figure 7.9. The red curve represents the error of the Newton method applied to the problem with DtN maps based on local Robin problems. For comparison we show in Figure 7.10 again the error of the Newton method applied to the problem with DtN maps based on local Dirichlet problems (blue curve), that we presented already in Figure 6.19a. While the latter error only converges until it reaches an error level of order 10^{-3}, the error of the Newton method applied to the problem (7.62) with DtN transparent boundary conditions based on Robin problems converges exponentially below 10^{-10}. This is due to the fact that the local Dirichlet problems are ill-posed at Dirichlet eigenvalues, while the local Robin problems are not. The closer one comes to such a Dirichlet eigenvalue the larger the error of the DtN maps become when computing them with local Dirichlet problems.

Now we want to compare the two sorts of DtN transparent boundary conditions when applying the direct method based on a Chebyshev interpolation. Again we aim to compute the common eigenvalue of the local Dirichlet problem and the nonlinear eigenvalue problem, that is marked as a blue cross in Figure 7.9. Analogously to the numerical test presented in Figure 6.19b for the case of DtN transparent boundary conditions based on local Dirichlet problems, we choose the k-formulation, fix the frequency to $\omega \approx 0.248 \cdot \frac{2\pi c}{a_1}$ and set the k-interval to the irreducible Brillouin zone $\widehat{B} = [0, \frac{\pi}{a_1}]$. In Figure 7.11 a comparison of the convergence of the Chebyshev interpolation is shown for the case with DtN transparent boundary conditions based on Dirichlet cell problems (blue) and Robin cell problems (red). We can see that the rate of convergence is in both cases the same, where we again want to point out that the convergence is not monotone since the Chebyshev nodes are not hierarchical.

In Section 6.3.6 we elaborated that the observed convergence for the case of DtN transparent boundary conditions based on local Dirichlet problems is due to the fact that all Chebyshev nodes are sufficiently far away from the local Dirichlet eigenvalue, which is reasonable since the increase of the condition number of the system and DtN matrices close to a local Dirichlet eigenvalue is limited to a very narrow vicinity, see Figure 7.6.

From Figure 7.2 we can see that there does not exist a global Dirichlet eigenvalue in the irreducible Brillouin zone at $\omega \approx 0.248 \cdot \frac{2\pi c}{a_1}$ such that the results of the Chebyshev interpolation in the interval

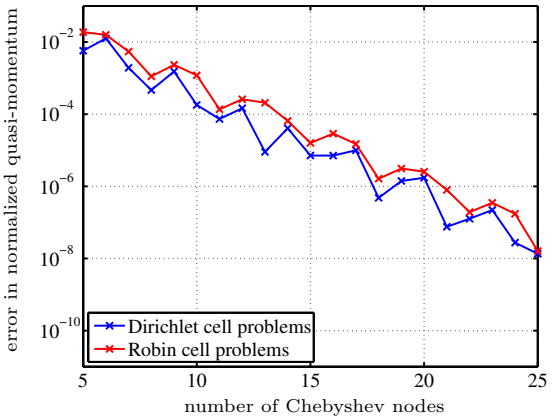


Figure 7.11: Absolute error of the Chebyshev interpolation in k-formulation when applied to the problem with DtN maps based on local Dirichlet problems (blue) and local Robin problems (red) for the computation of the common eigenvalue of the local Dirichlet problem and the nonlinear eigenvalue problem at $(\omega, k) \approx (0.248 \cdot \frac{2\pi c}{a_1}, 0.405 \cdot \frac{2\pi}{a_1})$, that is marked with a blue cross in Figure 7.9. The interval of the interpolation is chosen to be the irreducible Brillouin zone $\widehat{B} = [0, \frac{\pi}{a_1}]$.

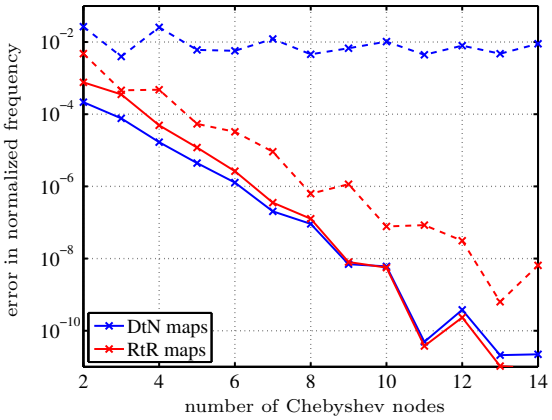


Figure 7.12: Absolute error of the Chebyshev interpolation in ω-formulation when applied to the problem with DtN maps (blue) and RtR maps (red) for the computation of the common eigenvalue of the local Dirichlet problem and the nonlinear eigenvalue problem at $(\omega, k) \approx (0.248 \cdot \frac{2\pi c}{a_1}, 0.405 \cdot \frac{2\pi}{a_1})$, that is marked with a blue cross in Figure 7.9. The interval of the interpolation is chosen to be $[0.235 \cdot \frac{2\pi c}{a_1}, 0.265 \cdot \frac{2\pi c}{a_1}]$ (solid lines) and $[0.215 \cdot \frac{2\pi c}{a_1}, 0.265 \cdot \frac{2\pi c}{a_1}]$ (dashed lines).

$\widehat{B}$ are not spoilt by the presence of global Dirichlet eigenvalues. Figure 7.12 shows the error of the Chebyshev interpolation in ω-formulation when choosing the interval $[0.235 \cdot \frac{2\pi c}{a_1}, 0.265 \cdot \frac{2\pi c}{a_1}]$ (solid lines) and $[0.215 \cdot \frac{2\pi c}{a_1}, 0.265 \cdot \frac{2\pi c}{a_1}]$ (dashed lines). Now the blue curves show the errors of the Chebyshev interpolation when applied to the problem (7.62) with DtN transparent boundary conditions based on local Robin problems, while the red curves represent the corresponding results when applied to the problem (7.61) with RtR transparent boundary conditions. While for the smaller interval both methods, the DtN and the RtR method, converge nicely, we observe that convergence is lost for the larger interval when using DtN transparent boundary conditions. The difference is that the larger interval contains a global Dirichlet eigenvalue, see Figure 7.2.

This puts the numerical results on the Chebyshev interpolation applied to the problem (6.58) with DtN transparent boundary conditions based on local Dirichlet problems, that were presented in Section 6.3.6 and also in Figure 7.11, into perspective. In general, we do not have a priori knowledge about the location of global Dirichlet eigenvalues and hence, cannot argue that the Chebyshev interpolation converges in either case towards a common eigenvalues of the nonlinear eigenvalue problem and the local Dirichlet problem.

Note that this problem also transfers to the computation of guided modes that are not equal to local or global Dirichlet eigenvalues when using the Chebyshev interpolation of the nonlinear problem with DtN transparent boundary conditions in an interval that is not sufficiently far away from global Dirichlet eigenvalues.

7.3.5 Adaptive path following of dispersion curves

Finally, we apply the adaptive path following algorithm based on piecewise Taylor expansions, that was introduced in Chapter 5, to the nonlinear eigenvalue problem (7.61) with RtR transparent boundary conditions. As elaborated already above in Section 7.2.4, it is very cumbersome to compute the derivatives of the DtN operators $\mathcal{D}^{\pm}_{\mathrm{RtR}}$, that are based on local Robin problems, which makes the application of the piecewise Taylor expansion to the nonlinear eigenvalue problem (7.62) very involved and, due to the need

of a recursive procedure to compute the derivatives of $\mathcal{D}^{\pm}_{\mathrm{RtR}}$, also inefficient. Moreover, the eigenvalue problem (7.62) with DtN transparent boundary conditions does not resolve the problem related to global Dirichlet eigenvalues. Therefore, we shall refrain from applying the adaptive path following algorithm to (7.62) and concentrate instead on the nonlinear eigenvalue problem (7.61) with RtR transparent boundary conditions.

In Section 7.2.4 we discussed the analyticity of the dispersion curves of (7.61), which is a key requirement of the adaptive path following algorithm, see Assumption 5.1, and we commented on the computation of the dispersion curve derivatives, that we will need for the piecewise Taylor expansions.

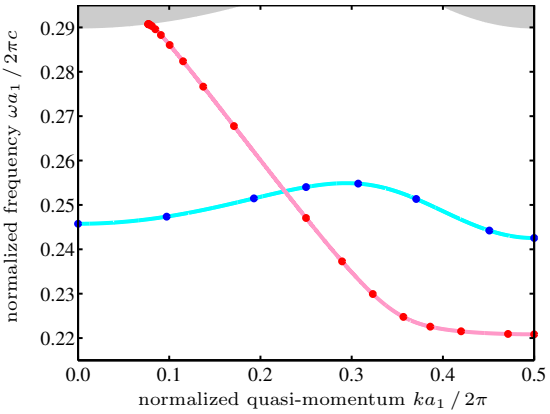
In Section 6.3.7 we applied the adaptive path following to the problem (6.58) with DtN transparent boundary conditions that are based on local Dirichlet problems. In addition to the procedure described Algorithm 5.2 for an adaptive path following without backward check or as presented in Algorithm 5.3 including backward check, we introduced the *band edge check* in Section 6.3.7, that was needed due to the fact, that the DtN operators are not well-defined outside the band gaps and hence, the path following has to stop at band edges. The same is true for the case with RtR transparent boundary conditions. The RtR operators are only well-defined in the band gaps and hence, we need to employ the band edge refinement as described in Section 6.3.7.

Again let us sketch briefly the adaptive path following algorithm: Let $n \in \mathbb{N}$ be the order of the Taylor expansions. We select a start value $k^{(0)} \in \widehat{B}$, e. g. $k^{(0)} = \frac{\pi}{2a_1}$, and compute the eigenvalues in a frequency interval $I_\omega \subset \mathbb{R}^+ \setminus \sigma_h^{\mathrm{ess}}(k^{(0)})$. Then we employ the Chebyshev interpolation for the simultaneous computation of all eigenvalues of (7.61) in I_ω. For all computed eigenvalues at $k^{(0)} \in \widehat{B}$ we proceed as presented in Algorithm 5.2 for the case without backward check or as presented in Algorithm 5.3 including backward check, i. e.

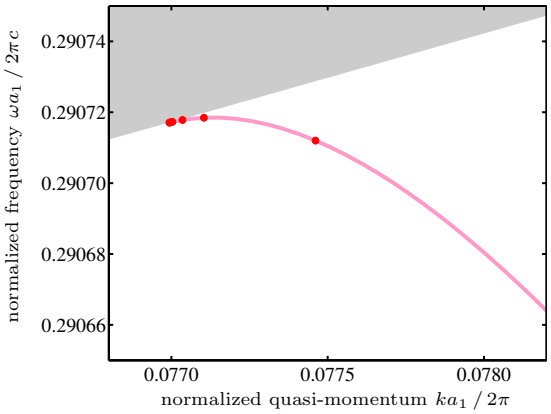
(i) we compute the dispersion curve derivatives up to order $n+1$,
(ii) we evaluate the acceptable step size (5.29) of the Taylor expansion of order n,
(iii) we add the step size to and subtract it from the current node to obtain the next nodes of the quasi-momentum,
(iv) we compute an approximation to the eigenvalue at the next nodes using the Taylor expansion of order n around the current node,
(v) we employ the proposed Newton-like method, or some other iterative scheme, in ω-formulation for the computation of an eigenvalue using the expected location as start value, and then
(vi) we continue to follow the dispersion curve to the left and right, possibly applying additional refinement checks such as the backward check, see Section 5.4.2, until we either reach the boundaries of $\widehat{B}$ or a band edge, which is identified by the band edge refinement proposed in Section 6.3.7.

In Figure 7.13 we present the results of the adaptive Taylor expansion of order $n = 5$ including backward check of tolerance $\varepsilon_{\mathrm{tol}}^{\mathrm{bwd}} = 10^{-2}$. Recall that the dots indicate the location of the values of k for which the eigenvalues $\omega(k)$ of (7.61) and the dispersion curve derivatives $\omega'(k), \omega^{(2)}(k), \ldots, \omega^{(6)}(k)$ were computed. The lines connecting the dots result from the post-processing, where we again chose the weighted Taylor expansion (5.30). Note that the red dispersion curve hits the band edge. For this dispersion curve the band edge refinement technique, that we described in Section 6.3.7, was employed with minimum step size $\varepsilon_{\mathrm{tol}}^{\mathrm{edge}} = 10^{-5}$. The detailed view of this dispersion curve in the vicinity of the band edge, which we present in Figure 7.13b, again shows that the group velocity of the dispersion curve converges to the slope of the band edge as we discussed already in Section 6.3.7.

With the adaptive path following of dispersion curves for the problem (7.61) with RtR transparent boundary conditions, we resolved the problem related to global and local Dirichlet eigenvalues while restoring the efficiency of the computation. In Section 6.3.7 we argued that the adaptive path following of dispersion curves applied to the problem (6.58) with DtN transparent boundary conditions based on local Dirichlet problems effectively reduces the influence of global and local Dirichlet eigenvalues. In fact, the difference of the results for the case with DtN operators, that were presented in Figure 6.20, and for the case with RtR operators as shown above in Figure 7.13 is negligible. Thus, for the adaptive path following of the two dispersion curves in the second band gap it seems that there is no need to switch to the more involved RtR transparent boundary conditions. Moreover, one has to take into account that the computation of the derivatives of the RtR operators $\mathcal{R}^{\pm}$ requires the evaluation of larger sums than



(a) Dispersion curves in the second band gap.



(b) Detailed view of the red dispersion curve in the vicinity of the band edge.

Figure 7.13: Adaptive Taylor scheme of order $n = 5$ with backward check of tolerance $\varepsilon_{\text{tol}}^{\text{bwd}} = 10^{-2}$ applied to the dispersion curves in the second band gap of the nonlinear eigenvalue problem (7.61) with RtR transparent boundary conditions . The error tolerance of the step size computation is $\varepsilon_{\text{tol}}^{\text{step}} = 10^{-4}$, the minimum step size of the band edge refinement is $\varepsilon_{\text{tol}}^{\text{edge}} = 10^{-5}$, and the start value of the iteration is set to $k^{(0)} = \frac{\pi}{2a_1}$.

the computation of the derivatives of the DtN operators $\mathcal{D}^{\pm}$. Depending on the implementation, this may significantly reduce the efficiency compared to the case with DtN transparent boundary conditions. Furthermore, the problem (7.61) with RtR transparent boundary conditions is not Hermitian, which implies that we need to solve (7.61) also for its left eigenvectors in order to evaluate the dispersion curve derivatives. This is not the case for the Hermitian, nonlinear eigenvalue problem (6.58) with DtN transparent boundary conditions, that we only need to solve for its right eigenvectors.

To this end, we propose to resolve the major difficulty of the adaptive path following when applied to the eigenvalue problem (6.58) with DtN transparent boundary conditions, that is related to the existence of global Dirichlet eigenvalues in the frequency interval I_ω at the start value $k^{(0)}$ of the adaptive scheme. Recall that the Chebyshev interpolation applied to (6.58) may yield inaccurate eigenvalues due to the presence of global Dirichlet eigenvalues in the interval I_ω, see Section 7.3.3. Therefore, we suggest to use the Chebyshev interpolation of the nonlinear problem (7.61) with RtR operators to simultaneously identify all eigenvalues in I_ω at the start value $k^{(0)}$, and then proceed with the adaptive path following applied to the eigenvalue problem (6.58) with DtN transparent boundary conditions. With this we keep the efficiency of the adaptive path following as proposed in Section 6.3.7 whilst resolving its major problem, since the adaptive selection of the nodes can be altered slightly such that the convergence of the iterative scheme is not effected by the presence of local or global Dirichlet eigenvalues during the computation.

7.4 Conclusions

In this chapter we showed the high-order FE discretization of RtR operators based on local cell problems with given Robin data. These operators are then employed for the exact computation of guided modes in 2d PhC waveguides and Dirichlet surface modes in semi-infinite 2d PhCs by transforming the problems to nonlinear eigenvalue problems with RtR transparent boundary conditions.

The RtR operators are well-defined and the local Robin problems are well-posed at all frequencies in band gaps and therefore, resolve the problems related to local and global Dirichlet eigenvalues when using DtN transparent boundary conditions based on local Dirichlet problems as introduced in Chapter 6.

As an alternative to the problem with RtR transparent boundary conditions we introduced a problem with DtN transparent boundary conditions whose DtN maps are based on local Robin problems. With this

problem we recover the benefits of problems with DtN transparent boundary conditions, in particular self-adjointness, whilst resolving the problem related to local Dirichlet eigenvalues. However, this problem is by definition not well-defined at global Dirichlet eigenvalues, which leaves the nonlinear eigenvalue problem with RtR transparent boundary condition as the only problem statement that is well-defined for all values in band gaps.

We showed that the RtR operators are differentiable with respect to the frequency and the quasi-momentum which is a requirement of various nonlinear eigenvalue solvers, for example the MSLP and the Chebyshev interpolation. Moreover, we discussed the computation of the derivatives of the RtR operators up to any order. As first order derivatives can be computed fairly straightforward with only little more effort than the computation of the corresponding derivatives of the DtN operators presented in Chapter 6, methods for the solution of the nonlinear eigenvalue problem such as the MSLP and the new Newton-like method proposed in Section 3.3, that require access to the first order derivatives of the RtR operators, can easily be applied.

We also extended the computation of the derivatives of the RtR operators to arbitrary orders, which we employed in an adaptive path following of the dispersion curves. However, in comparison to the case with DtN transparent boundary conditions, the computation of higher order derivatives is more involved as there are considerably larger sums to be computed. Moreover, the fact that the problem with RtR operators is not self-adjoint decreases the efficiency of the path following algorithm since also the left eigenvectors of the nonlinear eigenvalue problems have to be computed at all nodes of the piecewise Taylor expansion. Therefore, we came to the conclusion that the usage of the DtN method in an adaptive path following of dispersion curves of PhC waveguide band structures can be reasonable, since the adaptive selection of the nodes of the piecewise Taylor expansions reduces the problematic effect of local and global Dirichlet eigenvalues, as we already argued in Chapter 6.

What remains to be done in the future is to develop a deeper understanding of the properties of the (local) RtR operators such that the injectivity of the mapping (7.26) can be proved. Similarly, this deeper understanding is needed for a proof of Conjecture 7.13, where we claimed that the eigenvalues of the discrete forward-forward propagation operator come in complex conjugate pairs.

8 Conclusions and outlook

The main objective of this thesis was to develop a numerical scheme for the *accurate* and *efficient* approximation of 2d PhC waveguide band structures. For this we proposed

(i) a high-order FE discretization of DtN and RtR transparent boundary conditions, and
(ii) an adaptive path following of dispersion curves of PhC and PhC waveguide band structures.

Let us now in Section 8.1 point out the main contributions of this thesis related to these two aspects, before we will comment in Section 8.2 on the perspectives of future research in the field of PhC waveguide band structure calculations.

8.1 Contributions of this work

High-order FE discretization of DtN transparent boundary conditions In Chapter 6 we developed a high-order FE discretization of DtN transparent boundary conditions for the periodic medium of 2d PhC waveguides. The presented DtN transparent boundary conditions for waveguides were proposed by S. Fliss [Fli13] and are based on DtN operators that were introduced by P. Joly and co-workers [JLF06] for 2d PhCs with local defects. We explained in Chapter 6 the discretization of these operators by means of high-order FE spaces. Since we assume the holes/rods of the PhCs to be perfectly circular, we can resolve the computational domain with the help of coarse meshes of cells with curved edges. These coarse meshes need no further refinement since p-FEM converges exponentially in this case. The DtN transparent boundary conditions are employed to truncate the unbounded domain of the eigenvalue problem related to PhC waveguide band structure calculations. The resulting eigenvalue problem is, however, nonlinear. In comparison to the frequently used supercell method, that gives approximations to guided modes in PhC waveguides, DtN transparent boundary conditions are *exact* in the sense that they do not introduce an additional modelling error. For the numerical solution of the nonlinear eigenvalue problem, in particular when using the adaptive path following algorithm, differentiability of the DtN operators is crucial. We showed in Chapter 6 that the DtN operators are differentiable to any order with respect to the frequency and quasi-momentum, and explained the computation of their derivatives. The drawback of the DtN transparent boundary conditions, apart form the fact that the eigenvalue problem becomes nonlinear, is that the DtN operators are not well-defined at global Dirichlet eigenvalues and their computation is ill-posed at local Dirichlet eigenvalues.

High-order FE discretization of RtR transparent boundary conditions In Chapter 7 we developed a high-order FE discretization of RtR transparent boundary conditions for the periodic medium of 2d PhC waveguides, that resolve the problems of DtN transparent boundary conditions related to global and local Dirichlet eigenvalues. The approach, that goes back to the PhD thesis of S. Fliss [Fli09], is very similar to DtN transparent boundary conditions. Instead of solving local cell problems with Dirichlet data, we solve local cell problems with Robin data, which resolves the problem related to local Dirichlet eigenvalues. For the discretization of these local cell problems with given Robin data we can reuse the high-order FE spaces, that were already introduced in Chapter 6 for the FE discretization to the DtN transparent boundary conditions. Similarly to the DtN operators, we proved in Chapter 7 that the RtR operators are differentiable to any order with respect to the frequency and quasi-momentum, and we explained the computation of their derivatives, which are needed in the numerical solution of the nonlinear eigenvalue problem with RtR transparent boundary conditions. Similarly to the DtN transparent boundary conditions, the RtR transparent boundary conditions are employed to truncate the unbounded computational domain of the eigenvalue problem related to PhC waveguide band structure

calculations. The resulting eigenvalue problem is also nonlinear but — in contrast to the problem with DtN transparent boundary conditions — it is non-Hermitian.

Newton-like method for iteratively solving nonlinear eigenvalue problems In Chapter 3 we proposed a new iterative solver for nonlinear eigenvalue problems, that is based on Newton's method. This method, which aims to find the roots of a signed distance function, i. e. the difference of the parameter plugged into the nonlinear matrix function and the eigenvalue of an associated parameterized, linear eigenvalue problem, is comparable in convergence and effort with the well-known MSLP. We employed this Newton-like method for the numerical solution of the nonlinear eigenvalue problems with DtN and RtR transparent boundary conditions and compared its results with the ones of the MSLP and of a linearization technique based on Chebyshev interpolation. In particular for the case with DtN transparent boundary conditions our proposed method is preferable compared to the MSLP since its radius of convergence is not spoilt by the presence of global Dirichlet eigenvalues.

Computation of dispersion curve derivatives In Chapter 4 we developed closed formulas for the group velocity and any higher derivative of the dispersion curves of PhC and PhC waveguide band structures. We generalized this procedure to the computation of eigenpath derivatives of general, nonlinear matrix eigenvalue problems in Chapter 5. Subject that the eigenpaths and their associated eigenvectors are differentiable with respect to the parameter, the procedure for the computation of the eigenpath derivatives can be applied to Hermitian as well as non-Hermitian problems. For non-Hermitian problems we only need to compute — in addition to the right eigenvectors — also the left eigenvectors of the nonlinear eigenvalue problem. In this respect, the procedure can be applied to the linear problems in 2d PhCs and 2d PhC waveguides using the supercell approach as well as to the nonlinear problems of 2d PhC waveguides with DtN or RtR transparent boundary conditions.

Adaptive path following for parameterized, nonlinear eigenvalue problems In Chapter 5 we proposed an adaptive path following algorithm for the eigenpaths of parameterized, nonlinear eigenvalue problems. This algorithm is based on a weighted, piecewise Taylor expansion for which we employ the derivatives of the eigenpaths. The selection of the parameter values for which a Taylor expansion is computed is done by estimating the remainder of the Taylor expansion. This procedure yields small step sizes when the eigenpaths change their behaviour on a small scale, and larger step sizes otherwise. The quality of this adaptive approximation is improved by employing additional refinement checks. The backward check ensures that the Taylor expansion around each node gives a good approximation to its adjacent nodes. The crossing check is a post-processing procedure to validate crossings of eigenpaths and to distinguish them from avoided crossings. We employed this algorithm to the adaptive approximation of dispersion curves of PhC and PhC waveguide band structures when using the supercell approach and demonstrated the ability to correctly identify mini-stopbands, i. e. avoided crossings of dispersion curves. We showed that this algorithm effectively reduces the computational costs of band structure calculations. In Chapters 6 and 7 we applied the adaptive scheme to the problems with DtN and RtR transparent boundary conditions, where an additional band edge check is needed, since the DtN and RtR operators are only well-defined in the band gap. Hence, developed an algorithm for *accurate* and *efficient* PhC waveguide band structure calculations. This algorithm is *accurate* in the sense that — in contrast to the case of the supercell method — no additional modelling error is introduced, and it is *efficient* since there are only a little number of nonlinear eigenvalue problems to be solved to approximate the dispersion curves with the help of our adaptive Taylor expansion. In particular, it allows for studying the behaviour of dispersion curves in the vicinity of band edges, which is not possible with the supercell method.

8.2 Outlook

Finally, let us summarize the open questions, that we could not address in this thesis, and mention some perspectives for future research in the field of PhC waveguide band structure calculations.

Numerical analysis of the formulas for the eigenpath derivatives In thesis we did not discuss the numerical analysis of the formulas for the eigenpath derivatives. We only presented numerical results for the convergence of the group velocity formula when increasing the mesh refinement of our FE approximation, see Figure 4.2. The numerical analysis should also address the problem related to the ill-posed source problems in the vicinity of crossings, that need to be solved for the computation of eigenpath derivatives of order two or larger. We resolved this problem by adding additional orthogonality conditions. These extra constraints, however, yield that the source problem is not solved exactly. We showed numerically that this does not spoil the computation of the derivatives but a better understanding of this effect is needed.

Numerical analysis of the DtN and RtR transparent boundary conditions There has not been done any numerical analysis of the DtN and RtR transparent boundary conditions. For example, we argued in Chapters 6 and 7 that numerical evidence shows that the standard asymptotic convergence estimates hold true for the discretized, nonlinear eigenvalue problems with DtN and RtR transparent boundary conditions. However, to the best of our knowledge, a rigorous proof of this observation has not been found.

Properties of RtR operators For the computation of the derivatives of the RtR operators we introduced a mapping, see Eq. (7.26), which we use to characterize the derivatives of the forward-backward propagation operators. This characterization is only unique if (7.26) is injective, which is numerically evident but which has not been shown analytically. Similarly, a deeper understanding of the properties of the RtR and local RtR operators is needed such that we can prove Conjecture 7.13, where we claim that the eigenvalues of the discrete forward-forward propagation operator come in complex conjugate pairs.

Radiation losses in vertical direction of PhC slabs Another interesting topic of future research, that was out of scope in this thesis, is to study radiation losses in vertical direction of PhC slabs. The vertical direction of realistic PhC waveguides cannot be assumed to be invariant. Index guiding as used for 2d PhC waveguides, see for example the 2d PhC W1 slab waveguide sketched in Figure 1.5, reduces the effect of vertical radiation, but it cannot be neglected completely. Instead, a mathematical analysis of the vertical radiation losses in a similar fashion like done in [JH08, JHN12] for homogeneous, open waveguides is needed for PhC slabs.

References

[AMM07] N.P. van der Aa, H.G. ter Morsche, and R.R.M. Mattheij. Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem. *Electron. J. Linear Algebra*, 16(1):300–314, 2007. (Cited on pages 43 and 44.)

[ANR74] N. Ahmed, T. Natarajan, and K.R. Rao. Discrete cosine transform. *IEEE Trans. Comput.*, 100(1):90–93, 1974. (Cited on page 24.)

[AS04] H. Ammari and F. Santosa. Guided waves in a photonic bandgap structure with a line defect. *SIAM J. Appl. Math.*, 64(6):2018–2033, 2004. (Cited on page 2.)

[ABB+99] E. Anderson, Z. Bai, C. Bischof, L.S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D.C. Sorensen. *LAPACK users' guide*. SIAM, Philadelphia, PA, USA, 3rd edition, 1999. (Cited on page 21.)

[ACL92] A.L. Andrew, K.W.E. Chu, and P. Lancaster. Sensitivities of eigenvalues and eigenvectors of problems nonlinear in the eigenparameter. *Appl. Math. Lett.*, 5(3):69–72, 1992. (Cited on pages 26 and 42.)

[ACL93] A.L. Andrew, K.W.E. Chu, and P. Lancaster. Derivatives of eigenvalues and eigenvectors of matrix functions. *SIAM J. Matrix Anal. Appl.*, 14(4):903–926, 1993. (Cited on pages 26 and 42.)

[AT98] A.L. Andrew and R.C.E. Tan. Computation of derivatives of repeated eigenvalues and the corresponding eigenvectors of symmetric matrix pencils. *SIAM J. Matrix Anal. Appl.*, 20(1):78–100, 1998. (Cited on pages 43 and 44.)

[BO91] I. Babuška and J. Osborn. Eigenvalue problems. *Handbook of numerical analysis*, II:641–787, 1991. (Cited on page 35.)

[BLP78] A. Bensoussan, J.L. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*. North-Holland, Amsterdam, The Netherlands, 1978. (Cited on page 3.)

[BS67] H.A. Bethe and A. Sommerfeld. *Elektronentheorie der Metalle*. Heidelberger Taschenbücher. Springer, Berlin/Heidelberg, Germany, 1967. (Cited on pages 2 and 14.)

[Bey12] W.-J. Beyn. An integral method for solving nonlinear eigenvalue problems. *Linear Algebra Appl.*, 436(10):3839–3863, 2012. (Cited on page 22.)

[BEK11] W.-J. Beyn, C. Effenberger, and D. Kressner. Continuation of eigenvalues and invariant pairs for parameterized nonlinear eigenvalue problems. *Numer. Math.*, 119(3):489–516, 2011. (Cited on page 22.)

[Blo62] E.I. Blount. Formalisms of band theory. *Solid State Phys.*, 13:305–373, 1962. (Cited on page 1.)

[BCG06] D. Boffi, M. Conforti, and L. Gastaldi. Modified edge finite elements for photonic crystals. *Numer. Math.*, 105(2):249–266, 2006. (Cited on page 2.)

[BBDHC09] A.-S. Bonnet-BenDhia, G. Dakhia, C. Hazard, and L. Chorfi. Diffraction by a defect in an open waveguide: a mathematical analysis based on a modal radiation condition. *SIAM J. Appl. Math.*, 70(3):677–693, 2009. (Cited on page 2.)

[Bra07] D. Braess. *Finite elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, Cambridge, UK, 3rd edition, 2007. (Cited on page 32.)

[Bra13] H. Brandsmeier. *Standard and generalized hp-finite element discretizations for periodic structures*. PhD thesis, ETH Zürich, Zürich, Switzerland, 2013. (Cited on page 2.)

[BSS11] H. Brandsmeier, K. Schmidt, and C. Schwab. A multiscale hp-FEM for 2D photonic crystal bands. *J. Comput. Phys.*, 230(2):349–374, 2011. (Cited on pages 2 and 3.)

[Bri60] L. Brillouin. *Wave propagation and group velocity*, volume 8 of *Pure and applied physics*. Academic Press, New York, NY, USA, 1960. (Cited on page 2.)

[Bus02] K. Busch. Photonic band structure theory: assessment and perspectives. *C. R. Physique*, 3(1):53–66, 2002. (Cited on page 2.)

[CGMM11] C. Carstensen, J. Gedicke, V. Mehrmann, and A. Miedlar. An adaptive homotopy approach for non-selfadjoint eigenvalue problems. *Numer. Math.*, 119(3):557–583, 2011. (Cited on page 42.)

[Coa12] J. Coatléven. Helmholtz equation in periodic media with a line defect. *J. Comput. Phys.*, 231(4):1675–1704, 2012. (Cited on page 71.)

[Con15] Concepts development team. *Homepage of the numerical C++ library Concepts*. http://www.concepts.math.ethz.ch, 2015. (Cited on pages 18 and 21.)

[CS96] G. Constantine and T. Savits. A multivariate Faà di Bruno formula with applications. *Trans. Amer. Math. Soc.*, 348(2):503–520, 1996. (Cited on pages 45 and 84.)

[EK12] C. Effenberger and D. Kressner. Chebyshev interpolation for nonlinear eigenvalue problems. *BIT*, 52(4):933–951, 2012. (Cited on pages 22, 23, and 24.)

[Eng14] C. Engström. Spectral approximation of quadratic operator polynomials arising in photonic band structure calculations. *Numer. Math.*, 126(3):413–440, 2014. (Cited on pages 27 and 41.)

[ER09] C. Engström and M. Richter. On the spectrum of an operator pencil with applications to wave propagation in periodic and frequency dependent materials. *SIAM J. Appl. Math.*, 70(1):231–247, 2009. (Cited on pages 27 and 41.)

[FdB57] F. Faà di Bruno. Note sur une nouvelle formule de calcul differentiel. *The Quarterly Journal of Pure and Applied Mathematics*, 1:359–360, 1857. (Cited on pages 45, 80, and 125.)

[FG97] A. Figotin and Y.A. Godin. The computation of spectra of some 2d photonic crystals. *J. Comput. Phys.*, 136(2):585–598, 1997. (Cited on page 2.)

[FK97] A. Figotin and A. Klein. Localized classical waves created by defects. *J. Stat. Phys.*, 86:165–177, 1997. (Cited on pages 13 and 14.)

[FK96a] A. Figotin and P. Kuchment. Band-gap structure of spectra of periodic dielectric and acoustic media. I. Scalar model. *SIAM J. Appl. Math.*, 56(1):68–88, 1996. (Cited on page 2.)

[FK96b] A. Figotin and P. Kuchment. Band-gap structure of spectra of periodic dielectric and acoustic media. II. Two-dimensional photonic crystals. *SIAM J. Appl. Math.*, 56(6):1561–1620, 1996. (Cited on page 2.)

[Fli09] S. Fliss. *Etude mathématique et numérique de la propagation des ondes dans des milieux périodiques localement perturbés*. PhD thesis, École doctorale de l'École Polytechnique, Palaiseau, France, 2009. (Cited on pages 3, 4, 66, 67, 78, 106, 110, 120, 123, 124, and 141.)

[Fli13] S. Fliss. A Dirichlet-to-Neumann approach for the exact computation of guided modes in photonic crystal waveguides. *SIAM J. Sci. Comput.*, 35(2):B438–B461, 2013. (Cited on pages 3, 13, 14, 65, 66, 81, 88, 103, and 141.)

[FCB10] S. Fliss, E. Cassan, and D. Bernier. Computation of light refraction at the surface of a photonic crystal using DtN approach. *JOSA B*, 27(7):1492–1503, 2010. (Cited on page 3.)

[FJ09] S. Fliss and P. Joly. Exact boundary conditions for time-harmonic wave propagation in locally perturbed periodic media. *Appl. Numer. Math.*, 59(9):2155–2178, 2009. (Cited on page 3.)

[FJL10] S. Fliss, P. Joly, and J.-R. Li. Exact boundary conditions for wave propagation in periodic media containing a local perturbation. In M. Ehrhardt, editor, *Wave propagation in periodic media*, volume 1, chapter 5, pages 108–134. Bentham Science Publishers, Sharjah, UAE, 2010. (Cited on pages 3, 66, and 67.)

[FKS15] S. Fliss, D. Klindworth, and K. Schmidt. Robin-to-Robin transparent boundary conditions for the computation of guided modes in photonic crystal wave-guides. *BIT*, 55(1):81–115, 2015. (Cited on pages 105, 107, 108, and 109.)

[FL02] P. Frauenfelder and C. Lage. Concepts — an object-oriented software package for partial differential equations. *ESAIM: Math. Model. Numer. Anal.*, 36(5):937–951, 2002. (Cited on page 18.)

[GG12] S. Giani and I.G. Graham. Adaptive finite element methods for computing band gaps in photonic crystals. *Numer. Math.*, 121(1):31–64, 2012. (Cited on page 2.)

[Giv99] D. Givoli. Recent advances in the DtN FE method. *Arch. Comput. Methods Eng.*, 6(2):71–116, 1999. (Cited on page 3.)

[GVL96] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996. (Cited on page 21.)

[HKSW07] J.M. Harrison, P. Kuchment, A. Sobolev, and B. Winn. On occurrence of spectral edges for periodic operators inside the Brillouin zone. *J. Phys. A*, 40(27):7597, 2007. (Cited on page 10.)

[HFBW01] D. Hermann, M. Frank, K. Busch, and P. Wölfe. Photonic band structure computations. *Opt. Express*, 8(3):167–172, 2001. (Cited on pages 4 and 29.)

[HPW09] V. Hoang, M. Plum, and C. Wieners. A computer-assisted proof for photonic band gaps. *Z. Angew. Math. Phys.*, 60(6):1035–1052, 2009. (Cited on page 2.)

[HS13] T. Hohage and S. Soussi. Riesz bases and Jordan form of the translation operator in semi-infinite periodic waveguides. *J. Math. Pures Appl.*, 100(1):113–135, 2013. (Cited on page 78.)

[HL08] Z. Hu and Y.Y. Lu. Improved Dirichlet-to-Neumann map method for modeling extended photonic crystal devices. *Opt. Quantum Electron.*, 40(11-12):921–932, 2008. (Cited on page 4.)

[Jac98] J.D. Jackson. *Classical electrodynamics*. John Wiley & Sons, New York, NY, USA, 3rd edition, 1998. (Cited on page 5.)

[Jan94] M.S. Jankovic. Exact nth derivatives of eigenvalues and eigenvectors. *J. Guid. Control Dynam.*, 17(1):136–144, 1994. (Cited on page 43.)

[Jar12] E. Jarlebring. Convergence factors of Newton methods for nonlinear eigenvalue problems. *Linear Algebra Appl.*, 436(10):3943–3953, 2012. (Cited on page 23.)

[JMR15] E. Jarlebring, G. Mele, and O. Runborg. The waveguide eigenvalue problem and the tensor infinite Arnoldi method. Technical report, arXiv:1503.02096, 2015. (Cited on page 23.)

[JMM12] E. Jarlebring, W. Michiels, and K. Meerbergen. A linear eigenvalue algorithm for the nonlinear eigenvalue problem. *Numer. Math.*, 122(1):169–195, 2012. (Cited on page 23.)

[JH08] C. Jerez-Hanckes. *Modeling elastic and electromagnetic surface waves in piezoelectric transducers and optical waveguides*. PhD thesis, École doctorale de l'École Polytechnique, Palaiseau, France, 2008. (Cited on pages 2 and 143.)

[JHN12] C. Jerez-Hanckes and J.-C. Nédélec. Asymptotics for Helmholtz and Maxwell solutions in 3-d open waveguides. *Commun. Comput. Phys.*, 11:629–646, 2012. (Cited on pages 2 and 143.)

[JJWM08] J.D. Joannopoulos, S.G. Johnson, J.N. Winn, and R.D. Meade. *Photonic crystals: Molding the flow of light*. Princeton University Press, Princeton, NJ, USA, 2nd edition, 2008. (Cited on pages 1, 2, 3, 7, 9, 10, and 130.)

[JLF06] P. Joly, J.-R. Li, and S. Fliss. Exact boundary conditions for periodic waveguides containing a local perturbation. *Commun. Comput. Phys.*, 1(6):945–973, 2006. (Cited on pages 3, 4, 66, 67, 78, 110, 111, 121, and 141.)

[KS05] G. Karniadakis and S.J. Sherwin. *Spectral/hp element methods for computational fluid dynamics*. Numerical mathematics and scientific computation. Oxford University Press, Oxford, UK, 2005. (Cited on pages 76 and 77.)

[KKEJ13] P. Kaspar, R. Kappeler, D. Erni, and H. Jäckel. Average light velocities in periodic media. *J. Opt. Soc. Am. B*, 30(11):2849–2854, 2013. (Cited on page 3.)

[Kat95] T. Katō. *Perturbation theory for linear operators*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin/Heidelberg, Germany, 1995. (Cited on pages 14, 15, 30, 35, 38, and 88.)

[Kit04] C. Kittel. *Introduction to solid state physics*. Wiley, New York, NY, USA, 8th edition, 2004. (Cited on pages 1 and 9.)

[KS14a] D. Klindworth and K. Schmidt. Dirichlet-to-Neumann transparent boundary conditions for photonic crystal wave-guides. *IEEE Trans. Magn.*, 50:217–220, 2014. (Cited on pages 41 and 95.)

[KS14b] D. Klindworth and K. Schmidt. An efficient calculation of photonic crystal band structures using Taylor expansions. *Commun. Comput. Phys.*, 16(5):1355–1388, 2014. (Cited on pages 29 and 41.)

[KSF14] D. Klindworth, K. Schmidt, and S. Fliss. Numerical realization of Dirichlet-to-Neumann transparent boundary conditions for photonic crystal wave-guides. *Comput. Math. Appl.*, 67(4):918–943, 2014. (Cited on page 65.)

[KS86] M. Kojima and S. Shindo. Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. *Oper. Res. Soc. Japan*, 29(4):352–374, 1986. (Cited on page 26.)

[Kra08] T.F. Krauss. Why do we need slow light? *Nat. Photonics*, 2:448–450, 2008. (Cited on pages 2 and 29.)

[Kre09] D. Kressner. A block Newton method for nonlinear eigenvalue problems. *Numer. Math.*, 114:355–372, 2009. (Cited on page 22.)

[Kuc93] P. Kuchment. *Floquet theory for partial differential equations*. Birkhäuser, Basel, Switzerland, 1993. (Cited on pages 8 and 13.)

[Kuc01] P. Kuchment. The mathematics of photonic crystals. In G. Bao, L. Cowsar, and W. Masters, editors, *Mathematical modeling in optical science*, chapter 7, pages 207–272. SIAM, Philadelphia, PA, USA, 2001. (Cited on pages 2, 6, and 10.)

[KO04] P. Kuchment and B.S. Ong. On guided waves in photonic crystal waveguides. In P. Kuchment, editor, *Waves in periodic and random media*, volume 339 of *Contemp. Math.*, pages 105–115. American Math. Society, Providence, RI, USA, 2004. (Cited on page 2.)

[Lan64] P. Lancaster. On eigenvalues of matrices dependent on a parameter. *Numer. Math.*, 6(1):377–387, 1964. (Cited on page 43.)

[Lan70] P. Lancaster. Explicit solutions of linear matrix equations. *SIAM Rev.*, 12(4):544–566, 1970. (Cited on pages 81 and 123.)

[LMSY15] R.B. Lehoucq, K. Maschhoff, D.C. Sorensen, and C. Yang. *Homepage of ARPACK*. http://www.caam.rice.edu/software/ARPACK/, 2015. (Cited on pages 21 and 31.)

[LS96] R.B. Lehoucq and D.C. Sorensen. Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM J. Matrix Anal. Appl.*, 17(4):789–821, 1996. (Cited on page 31.)

[LSY98] R.B. Lehoucq, D.C. Sorensen, and C. Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*, volume 6. SIAM, 1998. (Cited on pages 21 and 31.)

[LWO+08] J. Li, T.P. White, L. O'Faolain, A. Gomez-Iglesias, and T.F. Krauss. Systematic design of flat band slow light in photonic crystal waveguides. *Opt. Express*, 16(9):6227–6232, 2008. (Cited on pages 2 and 29.)

[LG95] S.H. Lui and G.H. Golub. Homotopy method for the numerical solution of the eigenvalue problem of self-adjoint partial differential operators. *Numer. Algorithms*, 10(2):363–378, 1995. (Cited on page 42.)

[LKK97] S.H. Lui, H.B. Keller, and T.W.C. Kwok. Homotopy method for the large, sparse, real nonsymmetric eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 18(2):312–333, 1997. (Cited on page 42.)

[MV04] V. Mehrmann and H. Voss. Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods. *GAMM-Mitt.*, 27(2):121–152, 2004. (Cited on pages 21, 22, and 23.)

[MW01] V. Mehrmann and D. Watkins. Structure-preserving methods for computing eigenpairs of large sparse skew-hamiltonian/hamiltonian pencils. *SIAM J. Sci. Comput.*, 22(6):1905–1925, 2001. (Cited on page 22.)

[MS11] J.M. Melenk and S. Sauter. Wavenumber explicit convergence analysis for Galerkin discretizations of the Helmholtz equation. *SIAM J. Numer. Anal.*, 49(3):1210–1243, 2011. (Cited on pages 77 and 118.)

[Mon03] P. Monk. *Finite element methods for Maxwell's equations*. Oxford University Press, Oxford, UK, 2003. (Cited on page 7.)

[NS10] R.A. Norton and R. Scheichl. Convergence analysis of planewave expansion methods for 2d Schrödinger operators with discontinuous periodic potentials. *SIAM J. Numer. Anal.*, 47(6):4356–4380, 2010. (Cited on page 3.)

[NS13] R.A. Norton and R. Scheichl. Planewave expansion methods for photonic crystal fibres. *Appl. Numer. Math.*, 63(0):88–104, 2013. (Cited on page 3.)

[OBS+02] S. Olivier, H. Benisty, C.J.M. Smith, M. Rattier, C. Weisbuch, and T.F. Krauss. Transmission properties of two-dimensional photonic crystal channel waveguides. *Opt. Quant. Electron.*, 34(1–3):171–181, 2002. (Cited on pages 3 and 54.)

[ORB+01] S. Olivier, M. Rattier, H. Benisty, C. Weisbuch, C.J.M. Smith, R.M. De la Rue, T.F. Krauss, U. Oesterle, and R. Houdre. Mini-stopbands of a one-dimensional system: The channel waveguide in a two-dimensional photonic crystal. *Phys. Rev. B*, 63(11):1133111–1133114, 2001. (Cited on pages 3, 54, and 58.)

[QACT13] J. Qian, A.L. Andrew, D. Chu, and R.C.E. Tan. Computing derivatives of repeated eigenvalues and corresponding eigenvectors of quadratic eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 34(3):1089–1111, 2013. (Cited on pages 43 and 44.)

[QSS07] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical mathematics*, volume 37 of *Texts in applied mathematics*. Springer, Berlin/Heidelberg, Germany, 2nd edition, 2007. (Cited on page 52.)

[Ray87] Lord Rayleigh. On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure. *Philos. Mag. Ser. 5*, 24(147):145–159, 1887. (Cited on page 1.)

[RS78] M. Reed and B. Simon. *Methods of modern mathematical physics*, volume 4: Analysis of operators. Academic Press, New York, NY, USA, 1978. (Cited on pages 13, 14, and 31.)

[Rud64] W. Rudin. *Principles of mathematical analysis*, volume 3. McGraw-Hill, New York, NY, USA, 1964. (Cited on page 49.)

[Ruh73] A. Ruhe. Algorithms for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 10(4):674–689, 1973. (Cited on page 23.)

[SS11] S. Sauter and C. Schwab. *Boundary element methods*. Springer, Berlin/Heidelberg, Germany, 2011. (Cited on pages 31, 32, 77, and 118.)

[Sch08] K. Schmidt. *High-order numerical modeling of highly conductive thin sheets*. PhD thesis, ETH Zürich, Zürich, Switzerland, 2008. (Cited on page 18.)

[SK10] K. Schmidt and R. Kappeler. Efficient computation of photonic crystal waveguide modes with dispersive material. *Opt. Express*, 18(7):7307–7322, 2010. (Cited on pages 3, 15, 41, and 92.)

[SK09] K. Schmidt and P. Kauf. Computation of the band structure of two-dimensional photonic crystals with *hp* finite elements. *Comput. Methods Appl. Mech. Engrg.*, 198:1249–1259, 2009. (Cited on pages 2 and 18.)

[Sch98] C. Schwab. *p- and hp-finite element methods: Theory and applications in solid and fluid mechanics*. Oxford University Press, Oxford, UK, 1998. (Cited on pages 17 and 18.)

[Sip00] J.E. Sipe. Vector $k \cdot p$ approach for photonic band structures. *Phys. Rev. E*, 62(4):5672–5677, 2000. (Cited on pages 4 and 29.)

[SJ04] M. Soljacic and J.D. Joannopoulos. Enhancement of nonlinear effects using photonic crystals. *Nat. Mater.*, 3:211–219, 2004. (Cited on page 2.)

[Sou05] S. Soussi. Convergence of the supercell method for defect modes calculations in photonic crystals. *SIAM J. Numer. Anal.*, 43(3):1175–1201, 2005. (Cited on pages 3 and 15.)

[SP05] A. Spence and C. Poulton. Photonic band structure calculations using nonlinear eigenvalue techniques. *J. Comput. Phys.*, 204(1):65–81, 2005. (Cited on page 51.)

[SS88] C.M. de Sterke and J.E. Sipe. Envelope-function approach for the electrodynamics of nonlinear periodic structures. *Phys. Rev. A*, 38:5149–5165, 1988. (Cited on pages 4 and 29.)

[TM01] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Rev.*, 43(2):235–286, 2001. (Cited on pages 21, 22, 79, and 120.)

[Vor11] M. Vorobets. On the Bethe-Sommerfeld conjecture for certain periodic Maxwell operators. *J. Math. Anal. Appl.*, 377(1):370–383, 2011. (Cited on pages 2 and 14.)

[Woh01] B.I. Wohlmuth. A mortar finite element method using dual spaces for the Lagrange multiplier. *SIAM J. Numer. Anal.*, 38(3):989–1012, 2001. (Cited on page 117.)

[YL06] L. Yuan and Y.Y. Lu. An efficient bidirectional propagation method based on Dirichlet-to-Neumann maps. *IEEE Photon. Technol. Lett.*, 18(18):1967–1969, 2006. (Cited on page 4.)

[YL07] L. Yuan and Y.Y. Lu. A recursive-doubling Dirichlet-to-Neumann-map method for periodic waveguides. *J. Lightw. Technol.*, 25(11):3649–3656, 2007. (Cited on page 4.)