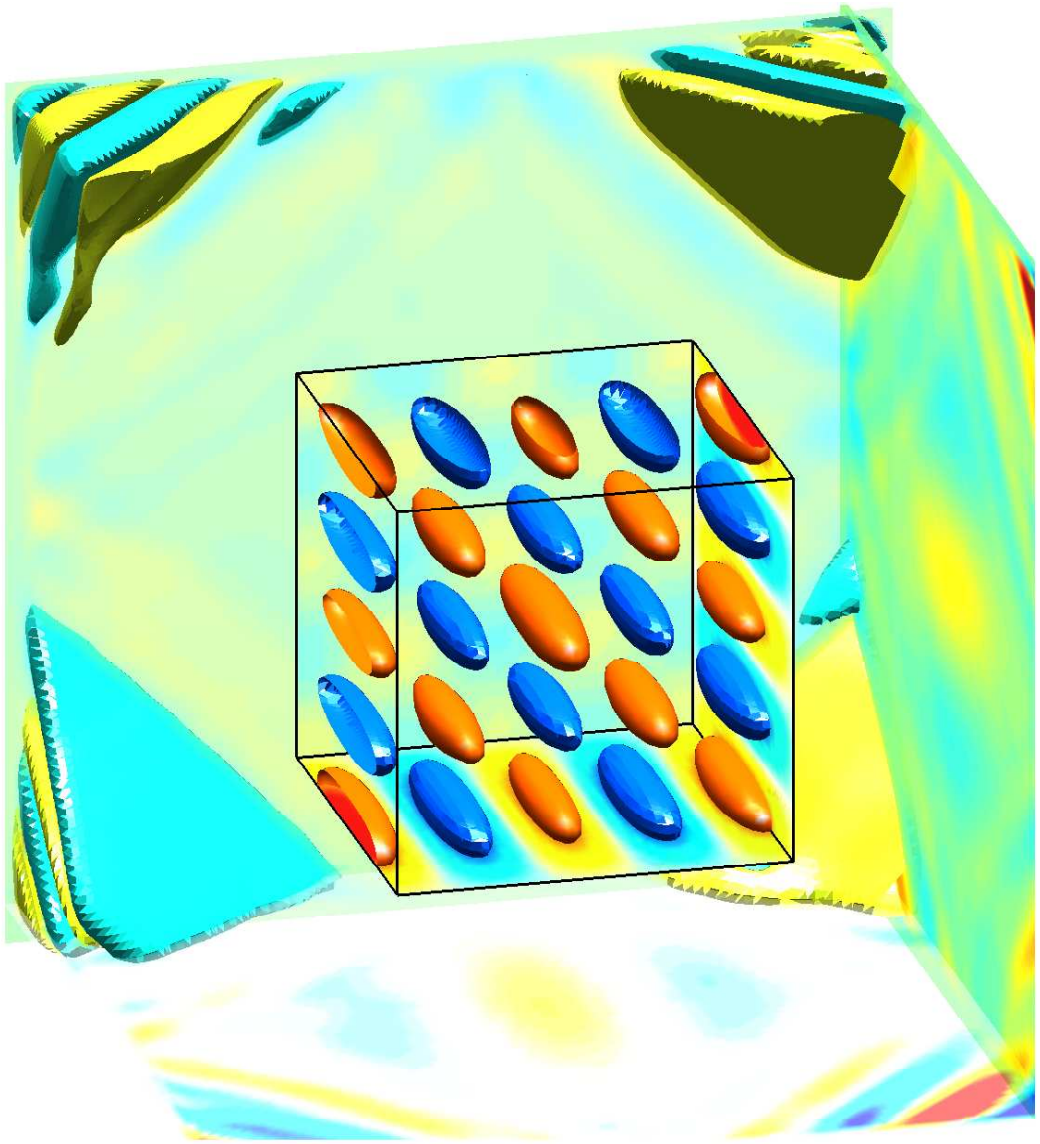
Master thesis - Electrical Engineering

Modeling of wave propagation in open domains

A Krylov subspace approach

Jörn Zimmerling 4047621

Delft University of Technology
CAS–MS–2014–04



6 of June 2014

TU Delft Delft University of Technology

Specialization Mircoelectronics,
Faculty of Electrical Engineering,
Mathematics and Computer Science

Modeling of wave propagation in open domains (May 20, 2014)

Typeset by the author in Nimbus Roman 11 pt using LaTeX.

Preface

In this Master thesis we investigate the use of several Krylov subspace methods for approximating time- and frequency-domain solutions of wave equations in open structures, including dispersive media. The research presented here is in the process of being published at the time of the defense.
This thesis is submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering at Delft University of Technology.

The research described in this thesis was a joint project between two research groups, namely, the Circuits and Systems Research Group and the Optics Research Group. The thesis committee consist in equal parts of people form both groups. The Circuits and Systems Research Group is represented by its head prof.dr.ir. A.-J. van der Veen and dr.ir. R.F. Remis, whereas the Optics Research Group is represented by its head prof.dr. H.P. Urbach and ir. L. Wei.

Corrections
This is the second version of this thesis, where the feedback of the committee has been processed. In this paragraph the changes with respect to the first version are listed. I would like to thank the committee for the feedback on the thesis and pointing out mistakes.

- Figures 4.27 and 4.26 are updated as the medium parameters in the text did not correspond to the ones used in the simulation.
- Minor language and notation corrections

Jörn Zimmerling
Delft, May 2014

Acknowledgments

First and foremost I would like to express my gratitude to my supervisor Rob Remis, for his guidance throughout the past years. Further I am thankful to him for motivating me for science, and his patience during reading and correcting this thesis.
I would like to thank my supervisor for optics related subjects Lei Wei, he offered great support especially in my first month at the institute. I learned a lot from Lei during our technical discussions.
I am sincerely grateful to Vladimir Druskin and Mikhail Zaslavsky from Schlumberger-Doll-Research

for their input and feedback on my work. Both are leading experts in their field and I could benefit greatly from their advice and expertise.
I would like to acknowledge and express my gratitude to the Optics Research group for the valuable discussions and motivational coffee breaks. Without the group the past year would have been less enjoyable.
Finally, I would like to thank my family and friends for their support during the past years.

Modeling of wave propagation in open domains
A Krylov subspace approach

This thesis is submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Jörn Zimmerling B.Sc
born in Pinneberg, Germany

The work presented in this thesis was performed at:

The Circuits and Systems Group
Department Microelectronics & Computer Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

and

The Optics Research Group
Department of Imaging Physics
Faculty of Applied Sciences
Delft University of Technology

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF ELECTRICAL ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **" Modeling of wave propagation in open domains "** by **Jörn Zimmerling** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 6 of June 2014

Chairman: _____
prof.dr.ir. A.-J. van der Veen

Advisor: _____
dr.ir. R.F. Remis

Committee Members: _____
prof.dr. H.P. Urbach

_____
ir. L. Wei

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Simulation and modeling are key driving forces in physics and engineering. Without powerful simulation tools it is difficult if not impossible to design and optimize the complex systems of today and tomorrow. The design of a ship, for example, or a bridge, or nanoantennas, or electronic circuits, or sensors – all require accurate modeling and effective simulation tools.
In many of these complex systems, electromagnetic (or acoustic) wave fields play a crucial role. The propagation of waves needs to be accurately modeled and simulated in a wide variety of applications ranging from photonics to seismic exploration. To this end, numerous modeling and simulation tools have been developed and many of these tools are based on the so-called Finite-Difference Time-Domain method (FDTD method) proposed by Yee [2] already in 1966. Around that time, a computer had no more than a hundred transistors and less than 200 kB of memory. The number of applications that could be effectively solved by FDTD was therefore quite limited. Times have changed, however, and presently essentially everyone owns and uses a computer. The transistor count of a modern day processor exceeds a billion and FDTD implementations on (super)computers can now easily handle wave field problems with millions or even billions of unknowns.
To push the limits of computable configurations beyond the line of hardware scaling, new computational algorithms are needed to reduce memory use and computation time. In this respect, Krylov subspace methods have a proven track record in many different areas of scientific computing [3]. Their use in hyperbolic wave field problems has been limited, however. The main reason is that most wave field problems encountered in practice take place on unbounded domains and up until recently it is was simply unclear how to simulate outward propagating waves in a computationally tractable manner within a Krylov subspace setting. This problem was resolved in [4] by using a so-called global optimal complex-scaling method for domain truncation combined with a stability correction procedure to obtain stable wave field approximations in the time- and frequency-domain. Complex-scaling was already introduced in the 1970s to simulate open quantum systems (see [5]) and can be seen as a special case of the well-known Perfectly Matched Layer (PML) technique proposed by Berenger in 1994 [6]. The latter technique is nowadays widely used in various FDTD implementations and commercial FDTD simulations tools. It utilizes frequency-dependent sponge layers that completely surround the finite domain of interest. The layers depend in a nonlinear manner on frequency and perfectly absorb outgoing waves thereby simulating the extension to infinity. By simply fixing the frequency of the PML the complex-scaling method can be obtained. In this way, a linearized, frequency-independent PML is obtained that works well for frequencies of operation close to the fixed PML frequency. The complex-scaling method presented in [4] is a generalization of the above approach in which the frequency-independent PML optimally absorbs propagating waves for a whole range of frequencies not necessarily close to the fixed PML frequency. In [7], this approach was generalized to optimally absorb evanescent and propagating fields simultaneously. The optimality in both

approaches leads to PMLs containing a small number of grid points. This is desirable, since a PML is only used to simulate the extension to infinity and the field inside such a layer is of no interest.
A consequence of implementing the optimal scaling-method for domain truncation is that the resonances and anti-resonances of the open system are approximated simultaneously. Consequenty, the complex-scaled system cannot be solved in a similar manner as the unscaled system, since otherwise unstable field approximations will result. Fortunately, in [4] it is shown that stable field approximations can be obtained from the complex-scaled system by evaluating a stability-corrected wave function. This function reduces to the unscaled wave field solution if complex-scaling is switched off and produces accurate, symmetry-preserving field approximations.
With complex-scaling and stability-correction into place, it is now possible to apply Krylov reduction to open wave field problems. Fast convergence can be expected, since infinity can be seen as an absorber and Krylov techniques have been shown to exhibit fast convergence for parabolic diffusion problems [8–11]. In [4] and [12] it was shown that so-called polynomial and extended Krylov subspace techniques do indeed exhibit fast convergence and outperform FDTD in the time-domain or conventional linear system solvers in the frequency-domain. Since the Krylov field approximations are basically expansions in approximate open resonant modes, convergence is particularly fast for configurations in which the fields are highly resonant. Examples of such configurations can be found in plasmonics, photonics, nano-antenna design, etc.
In this thesis we apply and validate the above mentioned Krylov solution methods to a variety of configurations and present a generalization of these techniques for media exhibiting general second-order relaxation phenomena. Specifically,

- we show that the resonances found via Krylov subspace reduction coincide with the resonances of a simple one-dimensional open system for which the resonances can be determined analytically;
- we show that the bandstructure of periodic structures can be determined via Krylov reduction techniques;
- we validate the above mentioned Krylov methods by computing time-domain field responses for 1D, 2D, and 3D configurations;
- we determine and visualize the dominant scattering modes of various open structures via Krylov reduction and show that significant order reduction can be achieved, especially for highly resonating fields;
- we introduce resonance expansions in reduced order modeling to obtain small order approximations of frequency-domain transfer functions and time-domain Greens functions after polynomial Krylov subspace reduction;
- we show the potential of rational Krylov subspace reduction for wave problems on open domains;

and, finally,

- for media showing second-order relaxation effects, we present a new symmetry preserving Krylov subspace formulation of Maxwell's equations in 2D and 3D, and
- we use the newly developed Krylov subspace technique for 3D wave fields to efficiently compute the spontaneous decay rate of a quantum emitter located near a golden nanorod. Moreover, we also show that a *single* reduced-order model approximates the spontaneous decay rate over a complete frequency band of interest.

Outline This thesis is organized as follows: In Chapter 2 we discuss the basic equations that govern the behavior of electromagnetic and acoustic wave fields. These equations are discretized on a so-called staggered Yee grid and the spatial derivatives are approximated by two-

point finite-difference formulas. The symmetry properties of the resulting state-space representation are discussed and media with and without dispersion are considered. Subsequently, we briefly address our implementation of the complex-scaling method, the stability-correction procedure, and a short discussion on resonant field expansions is included as well.
Chapter 3 introduces model order reduction via projection methods. Several Krylov subspace methods are reviewed and compared. Algorithms to construct the subspaces are also presented. To illustrate the reduction methods developed in Chapters 2 and 3, several electromagnetic examples are presented in Chapter 4. Polynomial and rational Krylov subspace methods are demonstrated and their performance is compared. Finally, Chapter 5 draws conclusions from the research documented in this thesis and gives an outlook to future work.
All abbreviations used throughout the document are listed in Appendix A. The nomenclature we tried to obey as consistently as possible is given in Appendix B.

Chapter 2

State-Space Representations for Electromagnetic and Acoustic Wave Fields

In this chapter we discuss the basic equations that govern the behavior of electromagnetic and acoustic wave fields. For electromagnetic fields, these equations are obviously given by Maxwell's equations, while for acoustic fields we consider the linearized acoustic wave equations. In Section 2.1 both sets of equations are written in a general matrix operator form that serves as a starting point for our reduced-order modeling approach. Furthermore, in Section 2.2 we discuss the spatial discretization procedure for Maxwell's equations. We do not discuss the discretization of the acoustic wave equations, since this discretization procedure is essentially the same as for Maxwell's equations. The full-order solution of the resulting semi-discrete system is presented in Section 2.3. As suggested by the title of this thesis, we are interested in wave equations on open domains. Therefore absorbing boundary conditions are introduced into the formalism in Section 2.4. This will lead to instabilities, however, and a stability correction is introduced in the same section to obtain a stable solution. Finally, resonance expansions are introduced as an efficient tool to obtain small order representations of open wave problem solutions.

2.1 System formulation

2.1.1 Maxwell Equations

The behavior of the electromagnetic field is governed by Maxwell's equations

$$-\nabla \times \mathbf{H} + \varepsilon_0 \partial_t \mathbf{E} + \mathbf{J}^{\text{ind}} = -\mathbf{J}^{\text{ext}} \tag{2.1}$$

and

$$\nabla \times \mathbf{E} + \mu_0 \partial_t \mathbf{H} + \mathbf{K}^{\text{ind}} = -\mathbf{K}^{\text{ext}}. \tag{2.2}$$

Equation (2.1) is also known as the Ampère-Maxwell law, while Eq. (2.2) is known as Faraday's law. In the above equations, $\mathbf{E}$ is the electric field strength (V/m) and $\mathbf{H}$ is the magnetic field strength (A/m). Furthermore, ε_0 (F/m) and μ_0 (H/m) are the permittivity and permeability of vacuum, respectively, while $\mathbf{J}^{\text{ext}}$ (A/m^2) and $\mathbf{K}^{\text{ext}}$ (V/m^2) are the external (impressed) electric and magnetic current densities. Finally, $\mathbf{J}^{\text{ind}}$ (A/m^2) and $\mathbf{K}^{\text{ind}}$ (V/m^2) take the presence of matter into account and represent electric and magnetic current densities that are induced inside a piece of matter. To properly take into account the various physical processes that take place within a material, it is customary to write the induced electric current density as

$$\mathbf{J}^{\text{ind}} = \mathbf{J}^{\text{cond}} + \partial_t \mathbf{P}, \tag{2.3}$$

where $\mathbf{J}^{\text{cond}}(\text{A/m}^2)$ is the induced conduction current and $\mathbf{P}$ is the polarization vector (C/m^2). The induced magnetic current density is written as

$$\mathbf{K}^{\text{ind}} = \mu_0 \partial_t \mathbf{M}, \tag{2.4}$$

where $\mathbf{M}$ is the magnetization (A/m). For most materials, the induced conduction current and the polarization vector depend only on the electric field strength, while the magnetization only depends on the magnetic field strength. Explicitly, for a general class of materials we have

$$\mathbf{J}^{\text{cond}}(\mathbf{x},t) = \int_{t'=0}^{t} \overline{\overline{\kappa}}_{\text{c}}(\mathbf{x},t')\mathbf{E}(\mathbf{x},t-t')\,\mathrm{d}t', \tag{2.5}$$

$$\mathbf{P}(\mathbf{x},t) = \varepsilon_0 \int_{t'=0}^{t} \overline{\overline{\kappa}}_{\text{e}}(\mathbf{x},t')\mathbf{E}(\mathbf{x},t-t')\,\mathrm{d}t', \tag{2.6}$$

and

$$\mathbf{M}(\mathbf{x},t) = \int_{t'=0}^{t} \overline{\overline{\kappa}}_{\text{m}}(\mathbf{x},t')\mathbf{H}(\mathbf{x},t-t')\,\mathrm{d}t', \tag{2.7}$$

where

- $\overline{\overline{\kappa}}_{\text{c}}$ is the conduction relaxation tensor $(\text{S/m}\cdot\text{s})$,
- $\overline{\overline{\kappa}}_{\text{e}}$ is the dielectric relaxation tensor (s^{-1}), and
- $\overline{\overline{\kappa}}_{\text{m}}$ is the magnetic relaxation tensor (s^{-1}).

Notice that only past values of the electric and magnetic field strength can contribute to the induced current densities at time t, since all materials are causal.

In this thesis, two types of materials will be considered, namely, instantaneously reacting materials and materials with relaxation (noble metals, for example). We first discuss instantaneously reacting media and then consider media with relaxation.

Instantaneously reacting media

For instantaneously reacting materials, the above relaxation tensors become

$$\overline{\overline{\kappa}}_{\text{c}} = \overline{\overline{\sigma}}(\mathbf{x})\delta(t), \quad \overline{\overline{\kappa}}_{\text{e}} = \overline{\overline{\chi}}_{\text{e}}(\mathbf{x})\delta(t), \quad \text{and} \quad \overline{\overline{\kappa}}_{\text{m}} = \overline{\overline{\chi}}_{\text{m}}(\mathbf{x})\delta(t), \tag{2.8}$$

where $\delta(t)$ is the Dirac distribution operative at $t=0$, $\overline{\overline{\sigma}}$ is the conductivity tensor (S/m), and $\overline{\overline{\chi}}_{\text{e}}$ and $\overline{\overline{\chi}}_{\text{m}}$ are the electric and magnetic susceptibility tensors, respectively (both dimensionless). Substitution of the above relaxation functions in the constitutive relations of Eqs. (2.5) – (2.7) gives

$$\mathbf{J}^{\text{cond}}(\mathbf{x},t) = \overline{\overline{\sigma}}(\mathbf{x})\mathbf{E}(\mathbf{x},t), \quad \mathbf{P}(\mathbf{x},t) = \varepsilon_0\overline{\overline{\chi}}_{\text{e}}(\mathbf{x})\mathbf{E}(\mathbf{x},t), \quad \text{and} \quad \mathbf{M}(\mathbf{x},t) = \overline{\overline{\chi}}_{\text{m}}(\mathbf{x})\mathbf{H}(\mathbf{x},t), \tag{2.9}$$

and the induced currents for instantaneously reacting materials are given by

$$\mathbf{J}^{\text{ind}} = \overline{\overline{\sigma}}(\mathbf{x})\mathbf{E}(\mathbf{x},t) + \varepsilon_0\overline{\overline{\chi}}_{\text{e}}(\mathbf{x})\partial_t\mathbf{E}(\mathbf{x},t) \quad \text{and} \quad \mathbf{K}^{\text{ind}} = \mu_0\overline{\overline{\chi}}_{\text{m}}(\mathbf{x})\mathbf{H}(\mathbf{x},t). \tag{2.10}$$

Substituting these expressions in Maxwell's equations (2.1) and (2.2) gives

$$-\nabla\times\mathbf{H} + \overline{\overline{\sigma}}\mathbf{E} + \overline{\overline{\varepsilon}}\partial_t\mathbf{E} = -\mathbf{J}^{\text{ext}} \tag{2.11}$$

and

$$\nabla\times\mathbf{E} + \overline{\overline{\mu}}\partial_t\mathbf{H} = -\mathbf{K}^{\text{ext}}, \tag{2.12}$$

where we have introduced the permittivity and permeability tensor as

$$\bar{\bar{\varepsilon}}(\mathbf{x}) = \varepsilon_0 \left[\bar{\bar{\mathbf{I}}} + \bar{\bar{\chi}}_{\mathrm{e}}(\mathbf{x})\right] \quad \text{and} \quad \bar{\bar{\mu}}(\mathbf{x}) = \mu_0 \left[\bar{\bar{\mathbf{I}}} + \bar{\bar{\chi}}_{\mathrm{m}}(\mathbf{x})\right], \tag{2.13}$$

respectively, and $\bar{\bar{\mathbf{I}}}$ is the diagonal unit tensor.

Written out in full, the Maxwell system of Eqs. (2.11) and (2.12) can be written in matrix form as (see [13])

$$(\mathcal{D} + \mathcal{S} + \mathcal{M}\partial_t)\mathcal{F} = \mathcal{Q}', \tag{2.14}$$

where the spatial derivatives are contained in the spatial differentiation matrix

$$\mathcal{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & \partial_z & -\partial_y \\ 0 & 0 & 0 & -\partial_z & 0 & \partial_x \\ 0 & 0 & 0 & \partial_y & -\partial_x & 0 \\ 0 & -\partial_z & \partial_y & 0 & 0 & 0 \\ \partial_z & 0 & -\partial_x & 0 & 0 & 0 \\ -\partial_y & \partial_x & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.15}$$

Furthermore, the medium parameters are contained in the medium matrices

$$\mathcal{S} = \begin{bmatrix} \bar{\bar{\sigma}} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{M} = \begin{bmatrix} \bar{\bar{\varepsilon}} & 0 \\ 0 & \bar{\bar{\mu}} \end{bmatrix}. \tag{2.16}$$

From energy considerations it follows that the conductivity tensor $\bar{\bar{\sigma}}$ is symmetric and semi-positive definite, while the permittivity and permeability tensors $\bar{\bar{\varepsilon}}$ and $\bar{\bar{\mu}}$ are symmetric and positive definite. Given the definitions of the medium matrices, it is easy to see that $\mathcal{S}$ is symmetric and semi-positive definite as well, while $\mathcal{M}$ is symmetric and positive-definite. Furthermore, for isotropic media we have

$$\bar{\bar{\sigma}} = \sigma\bar{\bar{\mathbf{I}}}, \quad \bar{\bar{\varepsilon}} = \varepsilon\bar{\bar{\mathbf{I}}}, \quad \text{and} \quad \bar{\bar{\mu}} = \mu\bar{\bar{\mathbf{I}}}, \tag{2.17}$$

where σ, ε, and μ are the scalar conductivity, permittivity, and permeability medium parameters. In this thesis, we consider isotropic media only.

Finally, the field and source vector are given by

$$\mathcal{F} = [E_x, E_y, E_z, H_x, H_y, H_z]^T, \tag{2.18}$$

and

$$\mathcal{Q}' = -[J_x^{\mathrm{ext}}, J_y^{\mathrm{ext}}, J_z^{\mathrm{ext}}, K_x^{\mathrm{ext}}, K_y^{\mathrm{ext}}, K_z^{\mathrm{ext}}]^T, \tag{2.19}$$

respectively. For many external sources as encountered in practice, the time dependence of the source can be factored out and we have $\mathcal{Q}' = w(t)\mathcal{Q}$, where $\mathcal{Q}$ is time independent. The scalar function w is called the source wavelet or source signature and vanishes for $t < 0$ if the external source is switched on at $t = 0$.

Media exhibiting relaxation

We consider isotropic media with relaxation tensors given by

$$\bar{\bar{\kappa}}_{\mathrm{c}} = \mathbf{0}, \quad \bar{\bar{\kappa}}_{\mathrm{m}}(\mathbf{x}, t) = \chi_{\mathrm{m}}(\mathbf{x})\delta(t)\bar{\bar{\mathbf{I}}}, \quad \text{and} \quad \bar{\bar{\kappa}}_{\mathrm{e}} = \kappa_{\mathrm{e}}\bar{\bar{\mathbf{I}}}. \tag{2.20}$$

We now have

$$\mathbf{J}^{\mathrm{cond}} = \mathbf{0}, \quad \mathbf{M} = \chi_{\mathrm{m}}\mathbf{H}, \quad \text{and} \quad \mathbf{P}(t) = \varepsilon_0 \int_{t'=0}^{t} \kappa_{\mathrm{e}}(t')\mathbf{E}(t - t')\,\mathrm{d}t', \tag{2.21}$$

where we have suppressed the spatial dependence for notational convenience. For most materials the reaction to the presence of an electromagnetic field is conveniently described by a relaxation function of the form

$$\kappa_e(t) = (\varepsilon_\infty - 1)\delta(t) + \tilde{\chi}_e(t), \tag{2.22}$$

where ε_∞ is the relative permittivity at high frequencies. The polarization vector now becomes

$$\mathbf{P} = \varepsilon_0(\varepsilon_\infty - 1)\mathbf{E} + \tilde{\mathbf{P}} \quad \text{with} \quad \tilde{\mathbf{P}} = \varepsilon_0 \int_{t'=0}^{t} \tilde{\chi}_e(t')\mathbf{E}(t-t')\,\mathrm{d}t' \tag{2.23}$$

and for the induced electric-current density we obtain

$$\mathbf{J}^{\text{ind}} = \mathbf{J}^{\text{cond}} + \partial_t \mathbf{P} = \varepsilon_0(\varepsilon_\infty - 1)\partial_t \mathbf{E} + \partial_t \tilde{\mathbf{P}}. \tag{2.24}$$

Furthermore, the induced magnetic-current density is given by

$$\mathbf{K}^{\text{ind}} = \mu_0 \partial_t \mathbf{M} = \mu_0 \chi_m \partial_t \mathbf{H} \tag{2.25}$$

and substituting these induced-current densities in Maxwell's equations, we arrive at

$$-\nabla \times \mathbf{H} + \partial_t \mathbf{D} = -\mathbf{J}^{\text{ext}}, \tag{2.26}$$
$$\nabla \times \mathbf{E} + \mu \partial_t \mathbf{H} = -\mathbf{K}^{\text{ext}}, \tag{2.27}$$

where $\mu = \mu_0(1+\chi_m)$ and $\mathbf{D} = \varepsilon\mathbf{E} + \tilde{\mathbf{P}}$ with $\varepsilon = \varepsilon_0\varepsilon_\infty$.

Different materials are described by different relaxation functions $\tilde{\chi}_e$. By specifying this function or by specifying the differential equation satisfied by $\tilde{\mathbf{P}}$, we describe how a piece of matter reacts to the presence of an electromagnetic field. Specifically, we consider the following different types of matter:

- *Debye material*: $\tilde{\mathbf{P}}$ satisfies

  $$\tau \partial_t \tilde{\mathbf{P}} + \tilde{\mathbf{P}} = \varepsilon_0(\varepsilon_s - \varepsilon_\infty)\mathbf{E}. \tag{2.28}$$

  Here, τ is the characteristic relaxation time of the material and ε_s is the static relative permittivity. Debye materials describe relaxation of non-interacting dipoles. An example of a Debye material is human tissue.
- *Drude material*: $\tilde{\mathbf{P}}$ satisfies

  $$\partial_t^2 \tilde{\mathbf{P}} + \gamma_p \partial_t \tilde{\mathbf{P}} = \varepsilon_0 \omega_p^2 \mathbf{E}. \tag{2.29}$$

  Here, ω_p is the volume plasma frequency and $\gamma_p = 1/\tau$ is the collision frequency with τ the relaxation time of the material. Examples of Drude materials include noble metals in the infrared region.
- *Lorentz material*: $\tilde{\mathbf{P}}$ satisfies

  $$\partial_t^2 \tilde{\mathbf{P}} + 2\delta \partial_t \tilde{\mathbf{P}} + \omega_0^2 \tilde{\mathbf{P}} = \varepsilon_0(\varepsilon_s - \varepsilon_\infty)\omega_0^2 \mathbf{E}. \tag{2.30}$$

  Here, ω_0 is the resonant plasma frequency of bound electrons and δ is the (mainly radiative) damping constant. Examples of Lorentz materials are noble metals in the visible range.

All three cases can be covered using the generic form

$$\beta_3 \partial_t^2 \tilde{\mathbf{P}} + \beta_2 \partial_t \tilde{\mathbf{P}} + \beta_1 \tilde{\mathbf{P}} = \beta_0 \mathbf{E}, \tag{2.31}$$

where for each material the different medium parameters β_i are listed in Table 2.1.

Table 2.1: Parameters to obtain common dispersion models with the general second-order dispersion model.

Medium	β_0	β_1	β_2	β_3
Normalization	L^2/c_0^2	L^2/c_0^2	L/c_0	1
Lorentz	$\varepsilon_0(\varepsilon_s-\varepsilon_\infty)\omega_0^2$	ω_0^2	2δ	1
Drude	$\varepsilon_0\omega_p^2$	0	γ_p	1
Debye	$\varepsilon_0(\varepsilon_s-\varepsilon_\infty)$	1	τ	0
Conductivity	σ	0	1	0

Introducing the auxiliary variable

$$\mathbf{U} = -\partial_t \tilde{\mathbf{P}} \tag{2.32}$$

we can write the above second-order constitutive relation in first-order form as

$$\beta_3 \partial_t \mathbf{U} + \beta_2 \mathbf{U} - \beta_1 \tilde{\mathbf{P}} + \beta_0 \mathbf{E} = \mathbf{0}. \tag{2.33}$$

Putting everything together, we have arrived at the first-order representation

$$-\nabla \times \mathbf{H} - \mathbf{U} + \varepsilon \partial_t \mathbf{E} = -\mathbf{J}^{\text{ext}} \tag{2.34}$$
$$\mathbf{U} + \partial_t \tilde{\mathbf{P}} = \mathbf{0}, \tag{2.35}$$
$$\beta_2 \mathbf{U} - \beta_1 \tilde{\mathbf{P}} + \beta_0 \mathbf{E} + \beta_3 \partial_t \mathbf{U} = \mathbf{0}, \tag{2.36}$$
$$\nabla \times \mathbf{E} + \mu \partial_t \mathbf{H} = -\mathbf{K}^{\text{ext}}, \tag{2.37}$$

and written out in full, the above Maxwell system can also be written in matrix-operator form as

$$(\mathcal{D} + \mathcal{S} + \mathcal{M}\partial_t)\,\mathcal{F} = \mathcal{Q}', \tag{2.38}$$

where the spatial differentiation matrix is given by

$$\mathcal{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \partial_z & -\partial_y \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\partial_z & 0 & \partial_x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \partial_y & -\partial_x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\partial_z & \partial_y & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \partial_z & 0 & -\partial_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\partial_y & \partial_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.39}$$

and the medium matrices $\mathcal{S}$ and $\mathcal{M}$ are given by

$$\mathcal{S} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \beta_0 & 0 & 0 & -\beta_1 & 0 & 0 & \beta_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_0 & 0 & 0 & -\beta_1 & 0 & 0 & \beta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_0 & 0 & 0 & -\beta_1 & 0 & 0 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.40}$$

and

$$\mathcal{M} = \begin{bmatrix} \varepsilon & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix}, \tag{2.41}$$

respectively. Finally, the field and source vectors are given by

$$\mathcal{F} = [E_x, E_y, E_z, \tilde{P}_x, \tilde{P}_y, \tilde{P}_z, U_x, U_y, U_z, H_x, H_y, H_z]^T \tag{2.42}$$

and

$$Q' = -[J_x^{\text{ext}}, J_y^{\text{ext}}, J_z^{\text{ext}}, 0, 0, 0, 0, 0, 0, K_x^{\text{ext}}, K_y^{\text{ext}}, K_z^{\text{ext}}]^T, \tag{2.43}$$

respectively.

Media exhibiting relaxation – Two-dimensional configurations

We consider H-polarized fields in a two-dimensional configuration that is invariant in the z-direction. Consequently, we may set $\partial_z = 0$ and for H-polarized fields the full set of Maxwell's equations simplifies to

$$\partial_x E_y - \partial_y E_x + \mu \partial_t H_z = -K_z^{\text{ext}}, \tag{2.44}$$
$$-\partial_y H_z - U_x + \varepsilon \partial_t E_x = -J_x^{\text{ext}}, \tag{2.45}$$
$$\partial_x H_z - U_y + \varepsilon \partial_t E_y = -J_y^{\text{ext}}, \tag{2.46}$$
$$U_x + \partial_t \tilde{P}_x = 0, \tag{2.47}$$
$$U_y + \partial_t \tilde{P}_y = 0, \tag{2.48}$$
$$\beta_3 \partial_t U_x + \beta_2 U_x - \beta_1 \tilde{P}_x + \beta_0 E_x = 0, \tag{2.49}$$

and

$$\beta_3 \partial_t U_y + \beta_2 U_y - \beta_1 \tilde{P}_y + \beta_0 E_y = 0. \tag{2.50}$$

This system can also be written in the form

$$(\mathcal{D}+\mathcal{S}+\mathcal{M}\partial_t)\,\mathcal{F}=\mathcal{Q}', \tag{2.51}$$

where the field and source vectors are now given by

$$\mathcal{F}=[H_z,E_x,E_y,\tilde{P}_x,\tilde{P}_y,U_x,U_y]^T \tag{2.52}$$

and

$$\mathcal{Q}'=-[K_z^{\text{ext}},J_x^{\text{ext}},J_y^{\text{ext}},0,0,0,0]^T, \tag{2.53}$$

respectively, while the spatial differentiation matrix is given by

$$\mathcal{D}=\begin{bmatrix} 0 & -\partial_y & \partial_x & 0 & 0 & 0 & 0\\ -\partial_y & 0 & 0 & 0 & 0 & 0 & 0\\ \partial_x & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.54}$$

Finally, the medium matrices are given by

$$\mathcal{S}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & -1\\ 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & \beta_0 & 0 & -\beta_1 & 0 & \beta_2 & 0\\ 0 & 0 & \beta_0 & 0 & -\beta_1 & 0 & \beta_2 \end{bmatrix} \tag{2.55}$$

and

$$\mathcal{M}=\begin{bmatrix} \mu & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & \varepsilon & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \varepsilon & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \beta_3 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \beta_3 \end{bmatrix}. \tag{2.56}$$

2.1.2 Linear Acoustics

To illustrate that other types of wave phenomena can be captured within the same framework, let us consider acoustic wave motion in linear, time invariant, instantaneously reacting, anisotropic, and inhomogeneous fluids. In the low-velocity approximation, the acoustic wave equations are given by the equation of motion

$$\nabla p+\rho\partial_t\mathbf{v}=\mathbf{f} \tag{2.57}$$

and the deformation equation

$$\nabla\cdot\mathbf{v}+\overline{\overline{\kappa}}\partial_t p=q. \tag{2.58}$$

In these equations, p is the acoustic pressure (Pa) and $\mathbf{v}$ is the particle velocity (m/s). Furthermore, $\mathbf{f}$ is the volume source density of volume force (N/m^3) and q is the volume density of injection rate (s^{-1}). Finally, ρ is the volume density of mass (kg/m^3) and $\overline{\overline{\kappa}}$ is the compressibility tensor (Pa^{-1}). Written out in full, we again arrive at a system of the form

$$(\mathcal{D}+\mathcal{S}+\mathcal{M}\partial_t)\,\mathcal{F} = \mathcal{Q}' \tag{2.59}$$

where this time the field and source vectors are given by

$$\mathcal{F} = [p, v_x, v_y, v_z]^T \tag{2.60}$$

and

$$\mathcal{Q}' = [q, f_x, f_y, f_z]^T, \tag{2.61}$$

respectively, while the differentiation matrix is given by

$$\mathcal{D} = \begin{bmatrix} 0 & \partial_x & \partial_y & \partial_z \\ \partial_x & 0 & 0 & 0 \\ \partial_y & 0 & 0 & 0 \\ \partial_z & 0 & 0 & 0 \end{bmatrix}. \tag{2.62}$$

Finally, the medium matrices for linearized acoustics are

$$\mathcal{S} = 0 \quad \text{and} \quad \mathcal{M} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \kappa_{xx} & \kappa_{xy} & \kappa_{xz} \\ 0 & \kappa_{yx} & \kappa_{yy} & \kappa_{yz} \\ 0 & \kappa_{zx} & \kappa_{zy} & \kappa_{zz} \end{bmatrix}. \tag{2.63}$$

From energy considerations it follows that the compressibility tensor $\overline{\overline{\kappa}}$ is symmetric and positive definite. Since $\rho > 0$ it then follows that the medium matrix $\mathcal{M}$ is symmetric and positive definite as well. Finally, we mention that for isotropic media, the compressibility tensor reduces to $\overline{\overline{\kappa}} = \kappa\overline{\overline{\mathbf{I}}}$, where κ is the scalar compressibility.

2.2 Spatial discretization of the Maxwell equations

In this section we discuss the spatial finite-difference discretization of Maxwell's equations for isotropic media. Discretization of the linearized acoustic field equations is not discussed, since these equations can be handled in essentially the same manner as Maxwell's equations.

Spatial discretization is carried out using two-point finite-difference formulas for the spatial derivatives that appear in Maxwell's equations. To this end, we introduce primary and dual grids in each Cartesian direction and subsequently define the finite-difference approximations of the electric and magnetic field strength on certain Cartesian products of these grids. For instance, the primary and dual grids in the y-direction are defined as

$$\Omega_y^{\mathrm{p}} = \{y_q \in \mathbb{R}, q = 0, 1, \ldots, Q+1, y_q > y_{q-1}\}, \tag{2.64}$$

and

$$\Omega_y^{\mathrm{d}} = \{\hat{y}_q \in \mathbb{R}, q = 1, \ldots, Q+1, \hat{y}_{q+1} > \hat{y}_q\}, \tag{2.65}$$

respectively, with corresponding step sizes given by

$$\delta_{y,q} = y_q - y_{q-1} \quad \forall q = 1, \ldots, Q+1 \quad \text{and} \quad \hat{\delta}_{y,q} = \hat{y}_{q+1} - \hat{y}_q \quad \forall q = 1, \ldots, Q. \tag{2.66}$$

Furthermore, in each Cartesian direction the first and last node of the primary grid are always located on the boundary of the computational domain and only so-called staggered grids are considered. These grids are characterized by the interlacing property

$$y_0 < \hat{y}_1 < y_1 < \hat{y}_2 < \cdots < \hat{y}_{Q+1} < y_{Q+1}, \tag{2.67}$$

showing that primary and dual nodes always have neighbouring nodes of the opposite kind. Primary and dual grids in the x- and z-directions are defined similarly with $P+2$ primary nodes in the x-direction and $R+2$ nodes in the z-direction. As a result, we have a total of six one-dimensional spatial grids given by

$$\Omega_x^{\mathrm{p}}, \Omega_x^{\mathrm{d}}, \Omega_y^{\mathrm{p}}, \Omega_y^{\mathrm{d}}, \Omega_z^{\mathrm{p}}, \quad \text{and} \quad \Omega_z^{\mathrm{d}}.$$

Now to truncate our domain of interest, we impose Perfect Electrically Conducting (PEC) boundary conditions at the boundary of our computational domain (absorbing boundary conditions are introduced later). Consequently, we have to set the tangential components of the electric field strength to zero at this boundary. Denoting the finite-difference approximations of the x-, y-, and z-components of the electric and magnetic field strength by $\{e_x, e_y, e_z\}$ and $\{h_x, h_y, h_z\}$, respectively, we therefore define

$$\begin{gathered} e_x \text{ on } \Omega_x^{\mathrm{d}} \times \Omega_y^{\mathrm{p}} \times \Omega_z^{\mathrm{p}} \\ e_y \text{ on } \Omega_x^{\mathrm{p}} \times \Omega_y^{\mathrm{d}} \times \Omega_z^{\mathrm{p}} \\ e_z \text{ on } \Omega_x^{\mathrm{p}} \times \Omega_y^{\mathrm{p}} \times \Omega_z^{\mathrm{d}} \end{gathered}$$

and

$$\begin{gathered} h_x \text{ on } \Omega_x^{\mathrm{p}} \times \Omega_y^{\mathrm{d}} \times \Omega_z^{\mathrm{d}} \\ h_y \text{ on } \Omega_x^{\mathrm{d}} \times \Omega_y^{\mathrm{p}} \times \Omega_z^{\mathrm{d}} \\ h_z \text{ on } \Omega_x^{\mathrm{d}} \times \Omega_y^{\mathrm{d}} \times \Omega_z^{\mathrm{p}} \end{gathered}$$

since the tangential components of the electric field strength are then located at the outer boundary of our computational domain. Consequently, we can properly take the PEC boundary conditions into account by simply setting these components to zero. Note that the normal components of the magnetic field strength at the boundary then automatically vanish as well. The resulting grid is also known as a Yee grid [2].

As mentioned above, we use two-point finite-difference formulas to approximate the partial derivatives in Maxwell's equations. These formulas are implemented using bidiagonal differentiation matrices. For example, y-differencing a field quantity defined on the primary grid in the y-direction is carried out by acting with the finite-difference matrix

$$\mathsf{Y} = \begin{bmatrix} \delta_{y;1}^{-1} & 0 & \cdots & \cdots & \cdots & 0 \\ -\delta_{y;2}^{-1} & \delta_{y;2}^{-1} & 0 & \cdots & \cdots & \vdots \\ 0 & -\delta_{y;3}^{-1} & \delta_{y;3}^{-1} & 0 & \cdots & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & \cdots & \cdots & \cdots & -\delta_{y;Q}^{-1} & \delta_{y;Q}^{-1} \\ 0 & \cdots & \cdots & \cdots & 0 & -\delta_{y;Q+1}^{-1} \end{bmatrix} \tag{2.68}$$

on these field quantities. Observe that this matrix takes the PEC material boundary conditions into account and as a result matrix Y is a nonsquare $(Q+1)$-by-Q matrix that maps from the

primary to the dual grid in the y-direction. Introducing the $(Q+1)$-by-$(Q+1)$ diagonal matrix of primary step sizes as

$$\mathsf{W}_y = \mathrm{diag}(\delta_{y;1}, \delta_{y;2}, ..., \delta_{y;Q+1}) \tag{2.69}$$

and the Q-by-$(Q+1)$ bidiagonal matrix $\mathrm{bidiag}_Q(-1,1)$ with -1 on the diagonal and $+1$ on the first upper diagonal, the above differentiation matrix can also be written as

$$\mathsf{Y} = -\mathsf{W}_y^{-1}\mathrm{bidiag}_Q(-1,1)^T. \tag{2.70}$$

The above difference matrix acts on field quantities defined on the primary grid in the y-direction. In a similar manner, we can define a difference matrix that acts on field quantities defined on the dual grid. Specifically, if we introduce the Q-by-Q diagonal step size matrix

$$\hat{\mathsf{W}}_y = \mathrm{diag}(\hat{\delta}_{y;1}, \hat{\delta}_{y;2}, ..., \hat{\delta}_{y;Q}) \tag{2.71}$$

then the difference matrix

$$\hat{\mathsf{Y}} = \hat{\mathsf{W}}_y^{-1}\mathrm{bidiag}_Q(-1,1) \tag{2.72}$$

computes two-point finite-differences of field quantities defined on the dual grid in the y-direction. This matrix is clearly Q-by-$(Q+1)$ and maps from the dual grid to the primary grid. Moreover, we also have the obvious symmetry relation

$$\hat{\mathsf{Y}}^T\hat{\mathsf{W}}_y = -\mathsf{W}_y\mathsf{Y}. \tag{2.73}$$

Difference matrices in the x- and z-direction are defined in an analogous manner. In particular, introducing the diagonal step size matrices

$$\mathsf{W}_x = \mathrm{diag}(\delta_{x;1}, \delta_{x;2}, ..., \delta_{x;P+1}), \qquad \hat{\mathsf{W}}_x = \mathrm{diag}(\hat{\delta}_{x;1}, \hat{\delta}_{x;2}, ..., \hat{\delta}_{x;P}), \tag{2.74}$$

$$\mathsf{W}_z = \mathrm{diag}(\delta_{z;1}, \delta_{z;2}, ..., \delta_{z;R+1}), \qquad \hat{\mathsf{W}}_z = \mathrm{diag}(\hat{\delta}_{z;1}, \hat{\delta}_{z;2}, ..., \hat{\delta}_{z;R}), \tag{2.75}$$

the difference matrices in the x- and z-direction are given by

$$\mathsf{X} = -\mathsf{W}_x^{-1}\mathrm{bidiag}_P(-1,1)^T, \qquad \hat{\mathsf{X}} = \hat{\mathsf{W}}_x^{-1}\mathrm{bidiag}_P(-1,1), \tag{2.76}$$

$$\mathsf{Z} = -\mathsf{W}_z^{-1}\mathrm{bidiag}_R(-1,1)^T, \qquad \hat{\mathsf{Z}} = \hat{\mathsf{W}}_z^{-1}\mathrm{bidiag}_R(-1,1), \tag{2.77}$$

and we have the symmetry relations

$$\hat{\mathsf{X}}^T\hat{\mathsf{W}}_x = -\mathsf{W}_x\mathsf{X} \qquad \text{and} \qquad \hat{\mathsf{Z}}^T\hat{\mathsf{W}}_z = -\mathsf{W}_z\mathsf{Z}. \tag{2.78}$$

With the introduction of all these differentiation matrices the semi-discrete state-space representation of the Maxwell equations can be formulated. In this thesis we work out three cases, namely, three-dimensional electromagnetic wave propagation in instantaneously reacting media and in media exhibiting relaxation, and two-dimensional electromagnetic wave propagation in media with relaxation. Other formulations can be found in Appendix C.

2.2.1 Electromagnetic state-space representation for instantaneously reacting media in three dimensions

Now that we have described the spatial grid and the corresponding differentiation matrices, we are in a position to discretize Maxwell's equations for isotropic and instantaneously reacting media. Using the Yee grid introduced above, approximating the partial derivatives by two-point finite-difference formulas, and arranging the unknowns in lexicographical order, we arrive at the state-space representation

$$(\mathsf{D}+\mathsf{S}+\mathsf{M}\partial_t)\,\mathsf{f} = \mathsf{q}'. \tag{2.79}$$

The order of this system is denoted by n and it is typically very large for real-world 3D problems (millions or even a billion of unknowns is not uncommon). Furthermore, the spatial differentiation matrix is given by

$$\mathsf{D} = \begin{bmatrix} 0 & \mathsf{D}_\mathrm{h} \\ \mathsf{D}_\mathrm{e} & 0 \end{bmatrix}, \tag{2.80}$$

with

$$\mathsf{D}_\mathrm{h} = \begin{bmatrix} 0 & \hat{\mathsf{Z}} \otimes \mathsf{I}_Q \otimes \mathsf{I}_{P+1} & -\mathsf{I}_R \otimes \hat{\mathsf{Y}} \otimes \mathsf{I}_{P+1} \\ -\hat{\mathsf{Z}} \otimes \mathsf{I}_{Q+1} \otimes \mathsf{I}_P & 0 & \mathsf{I}_R \otimes \mathsf{I}_{Q+1} \otimes \hat{\mathsf{X}} \\ \mathsf{I}_{R+1} \otimes \hat{\mathsf{Y}} \otimes \mathsf{I}_P & -\mathsf{I}_{R+1} \otimes \mathsf{I}_Q \otimes \hat{\mathsf{X}} & 0 \end{bmatrix} \tag{2.81}$$

and

$$\mathsf{D}_\mathrm{e} = \begin{bmatrix} 0 & -\mathsf{Z} \otimes \mathsf{I}_{Q+1} \otimes \mathsf{I}_P & \mathsf{I}_{R+1} \otimes \mathsf{Y} \otimes \mathsf{I}_P \\ \mathsf{Z} \otimes \mathsf{I}_Q \otimes \mathsf{I}_{P+1} & 0 & -\mathsf{I}_{R+1} \otimes \mathsf{I}_Q \otimes \mathsf{X} \\ -\mathsf{I}_R \otimes \mathsf{Y} \otimes \mathsf{I}_{P+1} & \mathsf{I}_R \otimes \mathsf{I}_{Q+1} \otimes \mathsf{X} & 0 \end{bmatrix}, \tag{2.82}$$

and $\otimes$ is the Kronecker (tensor) product. Furthermore, matrix S is given by

$$\mathsf{S} = \begin{bmatrix} \mathsf{M}_\sigma & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.83}$$

where M_σ is a diagonal semi-positive definite matrix with (averaged) conductivity values on its diagonal. The medium matrix M is given by

$$\mathsf{M} = \begin{bmatrix} \mathsf{M}_\varepsilon & 0 \\ 0 & \mathsf{M}_\mu \end{bmatrix} \tag{2.84}$$

and both M_ε and M_μ are diagonal and positive definite medium matrices with averaged permittivity and permeability values on their diagonal. The field vector is of the form

$$\mathsf{f} = [\mathsf{e}_x^T, \mathsf{e}_y^T, \mathsf{e}_z^T, \mathsf{h}_x^T, \mathsf{h}_y^T, \mathsf{h}_z^T]^T, \tag{2.85}$$

where all field quantities are stored in lexicographical order in the corresponding field vectors e_i and h_i, $i = x, y, z$. Finally, the finite-difference approximations of the external sources are stored in the source vector

$$\mathsf{q}' = -[\mathsf{j}_x^{\mathrm{ext};T}, \mathsf{j}_y^{\mathrm{ext};T}, \mathsf{j}_z^{\mathrm{ext};T}, \mathsf{k}_x^{\mathrm{ext};T}, \mathsf{k}_y^{\mathrm{ext};T}, \mathsf{k}_z^{\mathrm{ext};T}]^T. \tag{2.86}$$

Before discussing the various symmetry properties of the state-space representation of Eq. (2.79), we first introduce the system matrix, which characterizes the complete system. In particular, by premultiplying Eq. (2.79) by the inverse of the medium matrix, we obtain

$$(\mathsf{A} + \mathsf{I}\partial_t)\,\mathsf{f} = \mathsf{M}^{-1}\mathsf{q}', \tag{2.87}$$

where the system matrix is introduced as

$$\mathsf{A} = \mathsf{M}^{-1}(\mathsf{D} + \mathsf{S}). \tag{2.88}$$

The solution of the state-space representation can conveniently be written in terms of this matrix as we will show in Section 2.3.

Symmetry relations

To discuss the symmetry properties satisfied by the differentiation matrix, we first introduce the diagonal and positive definite step size matrices

$$\mathsf{W}_{\mathrm{e}} = \begin{bmatrix} \hat{\mathsf{W}}_z \otimes \hat{\mathsf{W}}_y \otimes \mathsf{W}_x & 0 & 0 \\ 0 & \hat{\mathsf{W}}_z \otimes \mathsf{W}_y \otimes \hat{\mathsf{W}}_x & 0 \\ 0 & 0 & \mathsf{W}_z \otimes \hat{\mathsf{W}}_y \otimes \hat{\mathsf{W}}_x \end{bmatrix}, \tag{2.89}$$

and

$$\mathsf{W}_{\mathrm{h}} = \begin{bmatrix} \mathsf{W}_z \otimes \mathsf{W}_y \otimes \hat{\mathsf{W}}_x & 0 & 0 \\ 0 & \mathsf{W}_z \otimes \hat{\mathsf{W}}_y \otimes \mathsf{W}_x & 0 \\ 0 & 0 & \hat{\mathsf{W}}_z \otimes \mathsf{W}_y \otimes \mathsf{W}_x \end{bmatrix}. \tag{2.90}$$

Using the symmetry relations of Eqs. (2.73) and (2.78) it is now easily verified that

$$\mathsf{D}_{\mathrm{h}}^T \mathsf{W}_{\mathrm{e}} = -\mathsf{W}_{\mathrm{h}} \mathsf{D}_{\mathrm{e}} \quad \text{and} \quad \mathsf{D}_{\mathrm{e}}^T \mathsf{W}_{\mathrm{h}} = -\mathsf{W}_{\mathrm{e}} \mathsf{D}_{\mathrm{h}}. \tag{2.91}$$

Furthermore, with

$$\mathsf{W}_+ = \begin{pmatrix} \mathsf{W}_{\mathrm{e}} & 0 \\ 0 & \mathsf{W}_{\mathrm{h}} \end{pmatrix} \quad \text{and} \quad \mathsf{W}_- = \begin{pmatrix} \mathsf{W}_{\mathrm{e}} & 0 \\ 0 & -\mathsf{W}_{\mathrm{h}} \end{pmatrix} \tag{2.92}$$

we also have

$$\mathsf{D}^T \mathsf{W}_+ = -\mathsf{W}_+ \mathsf{D} \quad \text{and} \quad \mathsf{D}^T \mathsf{W}_- = \mathsf{W}_- \mathsf{D}. \tag{2.93}$$

These symmetry relations can be translated into symmetry relations for the system matrix. In particular, for lossless media the system matrix is given by $\mathsf{A} = \mathsf{M}^{-1}\mathsf{D}$ and satisfies the symmetry property

$$\mathsf{A}^T \mathsf{M} \mathsf{W}_+ = -\mathsf{M} \mathsf{W}_+ \mathsf{A}. \tag{2.94}$$

In other words, for instantaneously reacting and lossless media, the system matrix is skew-symmetric with respect ot the energy inner product

$$\langle \mathsf{x}, \mathsf{y} \rangle_{\mathrm{en}} = \mathsf{y}^T \mathsf{M} \mathsf{W}_+ \mathsf{x}. \tag{2.95}$$

This inner product is positive definite and therefore induces a norm which is essentially equal to the stored electromagnetic energy within the system. More precisely, with f any vector of the form as given by Eq. (2.85), the norm $\frac{1}{2}\langle \mathsf{f}, \mathsf{f} \rangle_{\mathrm{en}}$ approximates

$$\mathcal{E}(\mathbb{D}) = \frac{1}{2} \iiint_{\mathbf{x} \in \mathbb{D}} \varepsilon |\mathbf{E}|^2 \, \mathrm{d}V + \frac{1}{2} \iiint_{\mathbf{x} \in \mathbb{D}} \mu |\mathbf{H}|^2 \, \mathrm{d}V, \tag{2.96}$$

which is the electromagnetic energy stored on the domain of interest $\mathbb{D}$.

For lossy media, the system matrix is given by $\mathsf{A} = \mathsf{M}^{-1}(\mathsf{D} + \mathsf{S})$ and the skew-symmetry property of Eq. (2.94) is lost. The system matrix does have another symmetry property, however. To be specific, introducing the free-field Lagrangian bilinear form

$$\langle \mathsf{x}, \mathsf{y} \rangle_{\mathrm{L}} = \mathsf{y}^T \mathsf{M} \mathsf{W}_- \mathsf{x} \tag{2.97}$$

and using the second symmetry relation of Eq. (2.93), we observe that the system matrix is symmetric with respect to this form, that is,

$$\langle \mathsf{A}\mathsf{x}, \mathsf{y} \rangle_{\mathrm{L}} = \langle \mathsf{x}, \mathsf{A}\mathsf{y} \rangle_{\mathrm{L}} \quad \text{for all } \mathsf{x}, \mathsf{y} \in \mathbb{R}^n. \tag{2.98}$$

The reason for calling the above form a free-field Lagrangian bilinear form is that $\frac{1}{2}\langle \mathsf{f}, \mathsf{f} \rangle_{\mathrm{L}}$ approximates

$$\mathcal{L}_{\mathrm{free}}(\mathbb{D}) = \frac{1}{2} \iiint_{\mathbf{x} \in \mathbb{D}} \varepsilon |\mathbf{E}|^2 \, \mathrm{d}V - \frac{1}{2} \iiint_{\mathbf{x} \in \mathbb{D}} \mu |\mathbf{H}|^2 \, \mathrm{d}V, \tag{2.99}$$

which is the free-field Lagrangian. We stress that the bilinear form $\langle \mathsf{x}, \mathsf{y} \rangle_{\mathrm{L}}$ does *not* induce a norm, since it is not positive definite. Finally, we mention that from the symmetry relation of Eq. (2.98) it follows that the discretized field quantities satisfy reciprocity (see [13]).

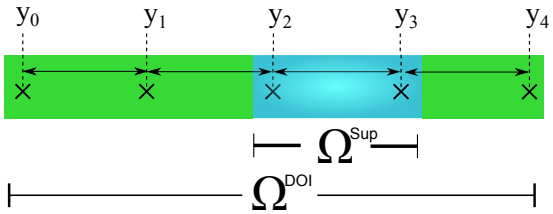
2.2.2 Electromagnetic state-space representation for media exhibiting relaxation in three dimensions

Similar to the instantaneously reacting case, the discretized Maxwell's equations for isotropic media exhibiting relaxation can be obtained. The main difference is the presence of the polarization vectors **P** and **U** in the field equations. These vectors are only active at points where the dispersive material is present. From a storage point of view, it is therefore advantageous to only keep the finite-difference approximations of **P** and **U** at these points in memory. Since the electric field strength approximations are defined over the total computational domain, we need a support matrix to implement the local dispersion relations. To this end, we introduce selection or logical projection matrices, which select the relevant electric field strength components from the total electric field vector. For example, if $\mathsf{I}_y^{\mathrm{sup}}$ is the support matrix of a dispersive material in the y-direction and e_y contains all finite-difference approximations of the y-component of the electric field strength, then the vector $\mathsf{I}_y^{\mathrm{sup}}\mathsf{e}_y$ contains only those y-components of E_y located within the dispersive material. An illustration of how the support matrix is constructed is shown in Figure 2.1.

Using this definition of the support matrices, the constitutive relation of Eq. (2.33) relating the electric and polarization fields to each other can be implemented in a straightforward manner. For example, for the y-component of Eq. (2.33) we have

$$\mathsf{B}_{3,y}\partial_t \mathsf{u}_y + \mathsf{B}_{2,y}\mathsf{u}_y - \mathsf{B}_{1,y}\mathsf{p}_y + \mathsf{B}_{0,y}\mathsf{I}_y^{\mathrm{sup}}\mathsf{e}_y = 0, \tag{2.100}$$

where the matrices $\mathsf{B}_{\{0,1,2,3\},y}$ are diagonal matrices with (averaged) medium values on their diagonal, according to Table 2.1.



$$\mathsf{I}^{\mathrm{sup}} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.101}$$

Figure 2.1: Example of a support matrix given a one-dimensional domain. Grid points y_2 and y_3 are within the dispersive medium and thus $y_2, y_3 \in \Omega^{\mathrm{sup}}$. The rows of $\mathsf{I}^{\mathrm{sup}}$ are the basis vectors of Ω^{sup} expressed in the basis vectors of Ω^{DOI}.

Using the Yee grid introduced earlier, approximating the partial derivatives by two-point finite-difference formulas, and arranging the unknowns in lexicographical order, we now again arrive at the state-space representation

$$(\mathsf{D} + \mathsf{S} + \mathsf{M}\partial_t)\,\mathsf{f} = \mathsf{q}'. \tag{2.102}$$

In this equation, the spatial differentiation matrix is given by

$$\mathsf{D} = \begin{bmatrix} 0 & 0 & 0 & \mathsf{D}_{\mathrm{h}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathsf{D}_{\mathrm{e}} & 0 & 0 & 0 \end{bmatrix}, \tag{2.103}$$

with D_{h} and D_{e} similar to the instantaneously reacting case.

Furthermore, matrix S is given by

$$\mathsf{S} = \begin{bmatrix} 0 & 0 & -\mathsf{I}^{\mathrm{sup};T} & 0 \\ 0 & 0 & \mathsf{I} & 0 \\ \mathsf{B}_0\mathsf{I}^{\mathrm{sup}} & -\mathsf{B}_1 & \mathsf{B}_2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{2.104}$$

where $\mathsf{B}_{\{0,1,2\}}$ are diagonal matrices only defined on the support of the dispersive media. In addition, I^{sup} is the total support matrix and the medium matrix M is given by

$$\mathsf{M} = \begin{bmatrix} \mathsf{M}_\varepsilon & 0 & 0 & 0 \\ 0 & \mathsf{I} & 0 & 0 \\ 0 & 0 & \mathsf{B}_3 & 0 \\ 0 & 0 & 0 & \mathsf{M}_\mu \end{bmatrix}, \tag{2.105}$$

where B_3 is again a dispersion matrix and both M_ε and M_μ are diagonal and positive definite medium matrices with averaged permittivity and permeability values on their diagonal. The field vector is now of the form

$$\mathsf{f} = [\mathsf{e}_x^T, \mathsf{e}_y^T, \mathsf{e}_z^T, \mathsf{p}_x^T, \mathsf{p}_y^T, \mathsf{p}_z^T, \mathsf{u}_x^T, \mathsf{u}_y^T, \mathsf{u}_z^T, \mathsf{h}_x^T, \mathsf{h}_y^T, \mathsf{h}_z^T]^T, \tag{2.106}$$

where all field quantities are stored in lexicographical order in the corresponding field vectors e_i, p_i, u_i, and h_i, $i = x,y,z$. Finally, the finite-difference approximations of the external sources are stored in the source vector

$$\mathsf{q}' = -[\mathsf{j}_x^{\text{ext};T}, \mathsf{j}_y^{\text{ext};T}, \mathsf{j}_z^{\text{ext};T}, 0,0,0,0,0,0, \mathsf{k}_x^{\text{ext};T}, \mathsf{k}_y^{\text{ext};T}, \mathsf{k}_z^{\text{ext};T}]^T. \tag{2.107}$$

Symmetry relations

The system matrix for media exhibiting relaxation given by

$$\mathsf{A} = \mathsf{M}^{-1}(\mathsf{D} + \mathsf{S}) \tag{2.108}$$

is symmetric with respect to the matrix WM with

$$\mathsf{W} = \begin{bmatrix} \mathsf{W}_\text{e} & 0 & 0 & 0 \\ 0 & \mathsf{W}_\text{p} & 0 & 0 \\ 0 & 0 & -\mathsf{W}_\text{u} & 0 \\ 0 & 0 & 0 & -\mathsf{W}_\text{h} \end{bmatrix} \tag{2.109}$$

and W_e and W_h as defined in Eqs. (2.89) and (2.90). Furthermore, W_u and W_p are given by

$$\mathsf{W}_\text{u} = \mathsf{B}_0^{-1}\mathsf{I}^{\text{sup}}\mathsf{W}_\text{e}\mathsf{I}^{\text{sup};T} \quad \text{and} \quad \mathsf{W}_\text{p} = \mathsf{B}_1\mathsf{W}_\text{u} = \mathsf{B}_1\mathsf{B}_0^{-1}\mathsf{I}^{\text{sup}}\mathsf{W}_\text{e}\mathsf{I}^{\text{sup};T}. \tag{2.110}$$

Introducing the bilinear form

$$\langle \mathsf{x}, \mathsf{y} \rangle_{\text{disp}} = \mathsf{y}^T\mathsf{WMx} \tag{2.111}$$

this symmetry relation can also be expressed as

$$\langle \mathsf{Ax}, \mathsf{y} \rangle_{\text{disp}} = \langle \mathsf{x}, \mathsf{Ay} \rangle_{\text{disp}} \quad \text{for all } \mathsf{x}, \mathsf{y} \in \mathbb{R}^n. \tag{2.112}$$

Now if the dispersive material occupies the bounded domain $\mathbb{D}_{\text{disp}} \subset \mathbb{D}$, we have that $\frac{1}{2}\langle \mathsf{f}, \mathsf{f} \rangle_{\text{disp}}$ approximates

$$\mathcal{L}_{\text{free}}(\mathbb{D}) + \frac{1}{2}\iiint_{\mathbf{x}\in\mathbb{D}_{\text{disp}}} \frac{\beta_1}{\beta_0}|\mathbf{P}|^2 \,\mathrm{d}V - \frac{1}{2}\iiint_{\mathbf{x}\in\mathbb{D}_{\text{disp}}} \frac{\beta_3}{\beta_0}|\partial_t\mathbf{P}|^2 \,\mathrm{d}V, \tag{2.113}$$

where $\mathcal{L}_{\text{free}}(\mathbb{D})$ is the free-field Lagrangian given in Eq. (2.99).

2.2.3 Electromagnetic state-space representation for media exhibiting relaxation in two dimensions

For two-dimensional H-polarized problems the analysis is similar to the three-dimensional case. We therefore only state the results. The state-space representation is again given by

$$(\mathsf{D}+\mathsf{S}+\mathsf{M}\partial_t)\,\mathsf{f}=\mathsf{q}'. \tag{2.114}$$

In this equation, the spatial differentiation matrix is given by

$$\mathsf{D}=\begin{bmatrix} 0 & -(\mathsf{Y}\otimes\mathsf{I}_P) & (\mathsf{I}_Q\otimes\mathsf{X}) & 0 & 0 & 0 & 0\\ -(\hat{\mathsf{Y}}\otimes\mathsf{I}_P) & 0 & 0 & 0 & 0 & 0 & 0\\ (\mathsf{I}_Q\otimes\hat{\mathsf{X}}) & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.115}$$

and matrix S is given by

$$\mathsf{S}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & -\mathsf{I}_x^{\mathrm{sup};T} & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & -\mathsf{I}_y^{\mathrm{sup};T}\\ 0 & 0 & 0 & 0 & 0 & \mathsf{I} & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \mathsf{I}\\ 0 & \mathsf{B}_{0,x}\mathsf{I}_x^{\mathrm{sup}} & 0 & -\mathsf{B}_{1,x} & 0 & \mathsf{B}_{2,x} & 0\\ 0 & 0 & \mathsf{B}_{0,y}\mathsf{I}_y^{\mathrm{sup}} & 0 & -\mathsf{B}_{1,y} & 0 & \mathsf{B}_{2,y} \end{bmatrix}, \tag{2.116}$$

where $\mathsf{B}_{\{0,1,2\},\{x,y\}}$ are diagonal matrices with (averaged) medium values on their diagonal. Furthermore, $\mathsf{I}^{\mathrm{sup}}$ is the support matrix defining the support of the medium exhibiting relaxation and the medium matrix M is given by

$$\mathsf{M}=\begin{bmatrix} \mathsf{M}_\mu & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & \mathsf{M}_{\varepsilon;x} & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \mathsf{M}_{\varepsilon;y} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \mathsf{I} & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & \mathsf{I} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \mathsf{B}_{3;x} & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \mathsf{B}_{3;y} \end{bmatrix} \tag{2.117}$$

with $\mathsf{B}_{3,x}$ and $\mathsf{B}_{3,y}$ dispersion matrices and $\mathsf{M}_{\varepsilon;x}$, $\mathsf{M}_{\varepsilon;y}$,and M_μ diagonal and positive definite medium matrices with averaged permittivity and permeability values on their diagonal. The field vector is of the form

$$\mathsf{f}=[\mathsf{h}_z^T,\mathsf{e}_x^T,\mathsf{e}_y^T\,\mathsf{p}_x^T,\mathsf{p}_y^T,\mathsf{u}_x^T,\mathsf{u}_y^T]^T, \tag{2.118}$$

where all field quantities are stored in lexicographical order in the corresponding field vectors e_i, p_i, u_i, and h_i, $i=x,y,z$. Finally, the finite-difference approximations of the external sources are stored in the source vector

$$\mathsf{q}'=-[\mathsf{k}_z^{\mathrm{ext};T},\mathsf{j}_x^{\mathrm{ext};T},\mathsf{j}_y^{\mathrm{ext};T},0,0,0,0]^T. \tag{2.119}$$

Symmetry relations

The system matrix for media exhibiting relaxation given by

$$\mathsf{A} = \mathsf{M}^{-1}(\mathsf{D}+\mathsf{S}) \tag{2.120}$$

is symmetric with respect to the matrix WM with

$$\mathsf{W} = \begin{bmatrix} \mathsf{W}_{\mathrm{h}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathsf{W}_{\mathrm{e}_x} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathsf{W}_{\mathrm{e}_y} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathsf{W}_{\mathrm{p}_x} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathsf{W}_{\mathrm{p}_y} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathsf{W}_{\mathrm{u}_x} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathsf{W}_{\mathrm{u}_y} \end{bmatrix}, \tag{2.121}$$

where

$$\begin{aligned}
\mathsf{W}_{\mathrm{h}} &= (\mathsf{W}_y \otimes \mathsf{W}_x), && (2.122)\\
\mathsf{W}_{\mathrm{e}_x} &= -(\hat{\mathsf{W}}_y \otimes \mathsf{W}_x), && (2.123)\\
\mathsf{W}_{\mathrm{e}_y} &= -(\mathsf{W}_y \otimes \hat{\mathsf{W}}_x), && (2.124)\\
\mathsf{W}_{\mathrm{p}_x} &= -\mathsf{B}_{1,x}\mathsf{B}_{0,x}^{-1}\mathsf{I}_x^{\mathrm{sup}}(\hat{\mathsf{W}}_y \otimes \mathsf{W}_x)\mathsf{I}_x^{\mathrm{sup};T}, && (2.125)\\
\mathsf{W}_{\mathrm{p}_y} &= -\mathsf{B}_{1,y}\mathsf{B}_{0,y}^{-1}\mathsf{I}_y^{\mathrm{sup}}(\mathsf{W}_y \otimes \hat{\mathsf{W}}_x)\mathsf{I}_y^{\mathrm{sup};T}, && (2.126)\\
\mathsf{W}_{\mathrm{u}_x} &= \mathsf{B}_{0,x}^{-1}\mathsf{I}_x^{\mathrm{sup}}(\hat{\mathsf{W}}_y \otimes \mathsf{W}_x)\mathsf{I}_x^{\mathrm{sup};T}, && (2.127)
\end{aligned}$$

and

$$\mathsf{W}_{\mathrm{u}_y} = \mathsf{B}_{0,y}^{-1}\mathsf{I}_y^{\mathrm{sup}}(\mathsf{W}_y \otimes \hat{\mathsf{W}}_x)\mathsf{I}_y^{\mathrm{sup};T}. \tag{2.128}$$

2.3 Full order solution

As we have seen, the state-space representation of Maxwell's equations can be written in terms of the system matrix as

$$(\mathsf{A}+\mathsf{I}\partial_t)\,\mathsf{f} = \mathsf{M}^{-1}\mathsf{q}', \tag{2.129}$$

where the system matrix is given by

$$\mathsf{A} = \mathsf{M}^{-1}(\mathsf{D}+\mathsf{S}). \tag{2.130}$$

From now on we consider source vectors for which the time dependence can be factored out, that is, $\mathsf{q}' = w(t)\mathsf{q}$, with $w(t)$ the source wavelet that vanishes for $t<0$ and q a time-independent vector. With vanishing initial conditions, the field vector f is then given by

$$\mathsf{f}(t) = w(t) * \eta(t)\exp(-\mathsf{A}t)\,\mathsf{M}^{-1}\mathsf{q} \quad \text{for } t>0, \tag{2.131}$$

where the asterisk denotes convolution in time, $\eta(t)$ is the Heaviside unit step function, and $\exp(-\mathsf{A}t)$ is the matrix exponential or evolution operator. In the Laplace domain the solution is given as

$$\hat{\mathsf{f}} = \hat{w}(s)(\mathsf{A}+s\mathsf{I})^{-1}\mathsf{M}^{-1}\mathsf{q}. \tag{2.132}$$

From the above two expressions it can be seen that computing field quantities in the time-domain amounts to evaluating the action of the matrix exponential function on the source vector, while in the frequency-domain the action of the resolvent of A on the source vector is required.

Furthermore, if the source vector q is a finite-difference approximation of the Dirac distribution and $w(t) = \delta(t)$, then

$$\mathsf{g}(t) = \eta(t) \exp(-\mathsf{A}t)\, \mathsf{M}^{-1}\mathsf{q} \tag{2.133}$$

is a numerical approximation of the continuous Green's function.

For real-world applications, direct evaluation of the evolution operator using some decomposition of matrix A is not feasible, since the order of this matrix is simply too large (millions or even a billion of unknowns is not uncommon). Fortunately, only the action of the evolution operator on the source vector (or the action of the resolvent on the source vector in frequency-domain problems) is required to obtain the fields. This allows us to efficiently compute the electromagnetic field quantities using Krylov subspace techniques. These techniques are discussed in the next chapter. Before doing so, however, we first briefly discuss our perfectly matched layer implementation and some of its consequences. We are then in a position to address the problem of computing propagating wave fields in open domains via Krylov subspace projection methods.

2.4 Perfectly matched layers, the complex-scaling method, and stability correction

In a numerical approach, the computational domain is compact and has bounded support, since the memory and computation power of all computational architectures are finite. Therefore a boundary condition needs to be imposed on the boundary of the computational domain.

As already stated in the title of this thesis, one of the goals is modeling of resonances in *open* systems. For this purpose, an optimal PML is inserted around the computational domain, which itself has Dirichlet boundary conditions. This PML simulates the extension of the domain to infinity.

Perfectly matched layers (PMLs) were introduced by Berenger in 1994 [6]. The problem with this technique is that it leads to nonlinear eigenproblems for spatial dimensions larger than one. The first step is to linearize the PML by fixing the PML frequency. The PML technique then essentially reduces to the so-called complex-scaling method, which was introduced in the 1970s to study open quantum systems. However, this scaling method is only effective for frequencies around the fixed PML frequency. This problem was resolved in [4], where an optimal and global complex-scaling method was presented. The method is optimal for propagating waves and this optimality leads to PMLs with a small number of layers. Furthermore, a global complex-scaling method means that the method is effective on an entire interval of frequencies not necessarily close to the fixed PML frequency. The technique was further refined in [7] resulting in an optimal and global complex-scaling method for evanescent and propagating waves. In this thesis, we use these optimal scaling methods to simulate the extension to infinity. In practice, this amounts to implementing complex step sizes within a layer (the PML layer) that completely surrounds the computational domain. The particular values for the step sizes follow from the optimization schemes presented in [4] and [7].

After incorporation of the optimal complex-scaling method, the system matrix is obviously no longer real-valued, since the step sizes within the PML layer are complex. All symmetry relations discussed in the previous sections still hold, however, provided we keep using the transpose operator instead of the Hermitian transpose. In other words, Lagrangian or dispersion symmetry on $\mathbb{R}^n$ turns into complex Lagrangian or dispersion symmetry on $\mathbb{C}^n$.

Stability Correction

The problem with the (optimal) complex-scaling method is that both the open resonances and anti-resonances of the system are approximated simultaneously. The anti-resonances lead to un-

stable field representations in the time-domain and the question now arises how to obtain stable time-domain solutions for the complex-scaled Maxwell system. This question was answered by Druskin and Remis in [4], where it is shown that stable and causal time-domain wavefield solutions can be computed via the stability-corrected wave function

$$\mathsf{f}(t) = w(t) * 2\eta(t)\mathrm{Re}\left[\eta(\mathsf{A})\exp(-\mathsf{A}t)\,\mathsf{M}^{-1}\mathsf{q}\right] \qquad \text{for } t > 0, \tag{2.134}$$

where $\eta(z)$ is the Heaviside unit step function defined as

$$\eta(z) = \begin{cases} 1 & \text{if } \mathrm{Re}(z) > 0, \\ \frac{1}{2} & \text{if } \mathrm{Re}(z) = 0, \\ 0 & \text{if } \mathrm{Re}(z) < 0. \end{cases} \tag{2.135}$$

We note that without complex-scaling, the stability-corrected wave function of Eq. (2.134) reduces to Eq. (2.131). Finally, we remark that the real and imaginary parts of the complex stability-corrected Green's function

$$\mathsf{g}_\mathrm{c}(t) = 2\eta(t)\eta(\mathsf{A})\exp(-\mathsf{A}t)\,\mathsf{M}^{-1}\mathsf{q} \qquad \text{for } t > 0 \tag{2.136}$$

are related to each other via a generalized Hilbert transform.

Resonance expansions

To illustrate which scattering poles contribute to the solution for a given wavelet $w(t)$, let us assume for simplicity that matrix A can be diagonalized, that is, we can write

$$\mathsf{A} = \mathsf{S}\Lambda\mathsf{S}^T\mathsf{W}\mathsf{M}, \tag{2.137}$$

where the eigenvectors s_i of A form the columns of matrix S and the corresponding eigenvalues λ_i are stored on the diagonal of the diagonal matrix Λ. Substitution of the eigendecomposition of Eq. (2.137) in Eq. (2.134) gives

$$\mathsf{f}(t) = w(t) * 2\eta(t)\mathrm{Re}\left[\sum_{i=1}^{n}\langle \mathsf{s}_i, \mathsf{M}^{-1}\mathsf{q}\rangle_\mathrm{disp}\eta(\lambda_i)\exp(-\lambda_i t)\,\mathsf{s}_i\right] \qquad \text{for } t > 0. \tag{2.138}$$

This result shows that only the resonant modes s_i that are excited by the source, that is, modes for which

$$\langle \mathsf{s}_i, \mathsf{M}^{-1}\mathsf{q}\rangle_\mathrm{disp} \neq 0$$

and with resonances in

$$\Lambda^\mathrm{res} := \{z \in \mathbb{C} :\ 0 \leq \mathrm{Re}(z) < \sigma_\mathrm{max}, -\omega_\mathrm{min} < \mathrm{Im}(z) < -\omega_\mathrm{max}\} \tag{2.139}$$

essentially contribute to the solution on a given time interval $[T_\mathrm{min}, \infty)$ with an error of the form $O[\exp(-\sigma_\mathrm{max}T_\mathrm{min})]$ exponentially decaying in time [14]. Therefore, σ_max determines the quality of the early-time approximation, while ω_min and ω_max are determined by the spectrum of the source wavelet $w(t)$.

For later reference we finally mention that the quality factor of a resonance λ_i is defined as

$$Q = \frac{|\mathrm{Im}(\lambda_i)|}{2\mathrm{Re}(\lambda_i)}. \tag{2.140}$$

Chapter 3

Krylov methods for model order reduction

3.1 Introduction

In this chapter model order reduction via projection methods is introduced in order to solve the systems defined in the last chapter. Specifically, we focus on the computation of the action of the evolution operator $\exp(-\mathsf{A}t)$ on the scaled source vector $\mathsf{M}^{-1}\mathsf{q}$. Since matrix A is large, sparse, and symmetric with respect to the Lagrangian or dispersion bilinear form, we pay particular attention to Krylov projection methods.
Section 3.2 introduces reduced-order modeling via general projection methods. It shows how a projection approach leads to a reduced-order solution of the defined problem. After this section the question of how to choose the projection subspace $\mathcal{K}$ is still open. As mentioned above, in this thesis Krylov subspaces are chosen as projection subspaces. Polynomial, extended and rational Krylov subspaces are introduced in Section 3.3. Their convergence behavior, performance, and construction is reviewed and discussed.

3.2 ROM via projection methods

In this section, reduced-order approximations via projection are introduced. Let us start with the state-space representation

$$(\mathsf{A}+\mathsf{I}\partial_t)\mathsf{g} = \delta(t)\mathsf{M}^{-1}\mathsf{q}. \tag{3.1}$$

We are interested in the solution $\mathsf{g}(t)$ for $t > 0$ with vanishing initial conditions.

To construct an approximate solution to the above problem, we consider an m-dimensional subspace $\mathcal{K}_m$ ($m < n$) of $\mathbb{C}^n$ and seek an approximation g_m to the above system that belongs to this subspace.

Let the vectors v_1, v_2, ..., v_m form a W^{s}-orthogonal set of basis vectors, where $\mathsf{W}^{\mathrm{s}} = \mathsf{M}\mathsf{W}_{-}$ for instantaneously reacting media or $\mathsf{W}^{\mathrm{s}} = \mathsf{M}\mathsf{W}$ for dispersive media. The approximate solution is now expanded in terms of these basis vectors as

$$\mathsf{g}_m(t) = \alpha_1(t)\mathsf{v}_1 + \alpha_2(t)\mathsf{v}_2 + \ldots + \alpha_m(t)\mathsf{v}_m = \mathsf{V}_m\mathsf{a}_m(t), \tag{3.2}$$

where V_m has the column partitioning $\mathsf{V}_m = (\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_m)$ and $\mathsf{a}_m(t)$ is the m-by-1 vector of expansion coefficients. To find these coefficients, we first introduce the residual that corresponds to the approximation g_m as

$$\begin{aligned} \mathsf{r}_m &= \delta(t)\mathsf{M}^{-1}\mathsf{q} - \mathsf{A}\mathsf{g}_m - \partial_t\mathsf{g}_m \\ &= \delta(t)\mathsf{M}^{-1}\mathsf{q} - \mathsf{A}\mathsf{V}_m\mathsf{a}_m - \mathsf{V}_m\partial_t\mathsf{a}_m. \end{aligned} \tag{3.3}$$

The coefficients now follow from the requirement that the residual is W^{s}-orthogonal to $\mathcal{K}_m$. In other words, we require that the residual satisfies the pseudo-Galerkin condition

$$\mathsf{r}_m \overset{\mathsf{W}^{\mathrm{s}}}{\perp} \mathcal{K}_m \quad \text{or} \quad \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{r}_m = \mathsf{0} \qquad \text{for } t > 0. \tag{3.4}$$

Substitution of the residual in this condition gives

$$\delta(t) \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{M}^{-1} \mathsf{q} - \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{A} \mathsf{V}_m \mathsf{a}_m - \mathsf{D}_m \partial_t \mathsf{a}_m = 0, \tag{3.5}$$

where we have used $\mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{V}_m = \mathsf{D}_m = \mathrm{diag}(\delta_1, \delta_2, ..., \delta_m) =: \mathsf{D}_m$, since the basis vectors are W^{s}-orthogonal. Setting

$$\mathsf{T}_m = \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{A} \mathsf{V}_m, \tag{3.6}$$

the above equation can also be written as

$$(\mathsf{D}_m^{-1} \mathsf{T}_m + \mathsf{I}_m \partial_t) \mathsf{a}_m = \delta(t) \mathsf{D}_m^{-1} \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{M}^{-1} \mathsf{q} \qquad \text{for } t > 0. \tag{3.7}$$

The solution of the above equation (with vanishing initial conditions) is easily found as

$$\mathsf{a}_m(t) = \eta(t) \exp\left(-\mathsf{D}_m^{-1} \mathsf{T}_m t\right) \mathsf{D}_m^{-1} \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{M}^{-1} \mathsf{q} \qquad \text{for } t > 0 \tag{3.8}$$

and the approximate projected solution follows as

$$\mathsf{g}_m(t) = \eta(t) \mathsf{V}_m \exp\left(-\mathsf{D}_m^{-1} \mathsf{T}_m t\right) \mathsf{D}_m^{-1} \mathsf{V}_m^T \mathsf{W}^{\mathrm{s}} \mathsf{M}^{-1} \mathsf{q} \qquad \text{for } t > 0. \tag{3.9}$$

To evaluate this approximation, the exponent of a small m-by-m matrix $\mathsf{D}_m^{-1} \mathsf{T}_m$ needs to be evaluated instead of the exponent of the large matrix A. Furthermore, since we are interested in modes that are excited by the source vector $\mathsf{M}^{-1} \mathsf{q}$, it makes sense to choose the first basis vector proportional to the source vector. Specifically, if we take

$$\mathsf{v}_1 = n_{\mathrm{s}}^{-1} \mathsf{M}^{-1} \mathsf{q} \tag{3.10}$$

with $n_{\mathrm{s}} = \|\mathsf{v}_1\|_2$ so that v_1 has a unit 2-norm, we have

$$\mathsf{M}^{-1} \mathsf{q} = n_{\mathrm{s}} \mathsf{v}_1 = n_{\mathrm{s}} \mathsf{V}_m \mathsf{e}_1, \tag{3.11}$$

where e_1 is first m-by-1 canonical basis vector. Substitution in our approximate solution then gives

$$\mathsf{g}_m(t) = \eta(t) n_{\mathrm{s}} \mathsf{V}_m \exp\left(-\mathsf{D}_m^{-1} \mathsf{T}_m t\right) \mathsf{e}_1 \qquad \text{for } t > 0. \tag{3.12}$$

Finally, if we now consider a source with a source signature $w(t)$ and incorporate stability-correction, we arrive at the field approximation

$$\mathsf{f}_m(t) = w(t) * 2\eta(t) n_{\mathrm{s}} \mathrm{Re}\left[\mathsf{V}_m \eta(\mathsf{D}_m^{-1} \mathsf{T}_m) \exp\left(-\mathsf{D}_m^{-1} \mathsf{T}_m t\right) \mathsf{e}_1\right] \qquad \text{for } t > 0. \tag{3.13}$$

Below we will show that by exploiting the symmetry property $\mathsf{A}^T \mathsf{W}^{\mathrm{s}} = \mathsf{W}^{\mathrm{s}} \mathsf{A}$, we can construct the basis vectors v_i via a Lanczos algorithm using a three-term recurrence relation if $\mathcal{K}_m$ is a polynomial Krylov space generated by matrix A and the source vector $\mathsf{M}^{-1} \mathsf{q}$. Moreover, matrix T_m is then given by

$$\mathsf{T}_m = \mathsf{D}_m \mathsf{H}_m, \tag{3.14}$$

where H_m is a tridiagonal m-by-m matrix that satisfies

$$\mathsf{H}_m^T \mathsf{D}_m = \mathsf{D}_m \mathsf{H}_m \quad \text{which is precisely the reduced counterpart of } \mathsf{A}^T \mathsf{W}^{\mathrm{s}} = \mathsf{W}^{\mathrm{s}} \mathsf{A}. \tag{3.15}$$

The projected field approximation then becomes

$$\mathsf{f}_m(t) = w(t) * 2\eta(t) n_{\mathrm{s}} \mathrm{Re}\left[\mathsf{V}_m \eta(\mathsf{H}_m) \exp(-\mathsf{H}_m t) \mathsf{e}_1\right] \qquad \text{for } t > 0. \tag{3.16}$$

3.3 Projection subspace

As mentioned in the introduction, Krylov subspaces show good convergence while approximating matrix functions or eigenvalues of matrices. Iterative solvers based on Krylov subspace approaches include GMRES, BiCGstab, and QMR [15]. They are widely used in commercial software and industry. Three types of Krylov subspaces suggested in the literature are now introduced. Then algorithms to construct these subspaces in software are given, together with a short literature review of the described methods.
In this thesis, polynomial, extended, and rational Krylov subspaces are considered, given in this order by

$$\mathcal{K}_m^{\mathrm{PKS}} = \mathrm{span}\{\mathsf{b}, \mathsf{Ab}, \mathsf{A}^2\mathsf{b}, \ldots, \mathsf{A}^{\mathrm{m}-1}\mathsf{b}\}, \tag{3.17}$$

$$\mathcal{K}_{n;m}^{\mathrm{EKS}} = \mathrm{span}\{\mathsf{A}^{-\mathrm{n}+1}\mathsf{b}, \ldots, \mathsf{A}^{-1}\mathsf{b}, \mathsf{b}, \mathsf{Ab}, \mathsf{A}^2\mathsf{b}, \ldots, \mathsf{A}^{\mathrm{m}-1}\mathsf{b}\}, \tag{3.18}$$

$$\mathcal{K}_m^{\mathrm{RKS}} = \mathrm{span}\{(\mathsf{A}+\sigma_1\mathsf{I})^{-1}\mathsf{b}, (\mathsf{A}+\sigma_2\mathsf{I})^{-1}(\mathsf{A}+\sigma_1\mathsf{I})^{-1}\mathsf{b}, \ldots, \prod_{\mathrm{i}=1}^{\mathrm{m}}(\mathsf{A}+\sigma_{\mathrm{i}}\mathsf{I})^{-1}\mathsf{b}\}. \tag{3.19}$$

Polynomial Krylov subspaces basically build polynomials in the system matrix A on the seed vector b. The extended Krylov subspace builds polynomials in A and A^{-1} on the seed vector b and therefore *extends* polynomial Krylov subspaces with negative powers of the system matrix A.

A rational Krylov subspace consists of products of rational matrix functions on the seed vector b. The rational matrix functions thereby represent single (complex) frequency solutions of the system itself. In the following subsections it is shown how a basis for these subspaces can be constructed.

For extended and polynomial Krylov subspaces, recursion relations to construct the basis can be found in the case the system matrix is symmetric with respect to a weighted pseudo-inner product. This leads to Lanczos-type algorithms, whereas Arnoldi algorithms can be used without the presence of such a symmetry. The Arnoldi and Lanczos algorithms for polynomial Krylov subspaces as well as the Arnoldi algorithm for rational Krylov subspaces are presented in the following. The Lanczos algorithm for extended Krylov subspaces can be found in [16,17].

3.3.1 Polynomial Krylov subspaces

Two algorithms are presented for constructing a basis for polynomial Krylov subspaces (PKS), namely, the Arnoldi and Lanczos algorithms. The Arnoldi algorithm has no symmetry constraints on the iteration matrix and utilizes the standard inner product on $\mathbb{C}^n$. All basis vectors need to be kept in memory in the Arnoldi algorithm. By contrast, the Lanczos algorithm exploits the symmetry of the system matrix and uses the Lagrangian or dispersion bilinear form to generate a basis. The basis vectors can be obtained via three-term recurrence relations and consequently at each iteration only the last two Lanczos vectors need to be kept in memory.

Arnoldi decomposition

For Arnoldi type algorithms the system matrix has no symmetry constraint [15]. The Arnoldi algorithm constructs an orthonormal Krylov basis V_m and produces the so-called Arnoldi decomposition

$$\mathsf{AV}_m = \mathsf{V}_m\mathsf{H}_m + \mathsf{p}_m\mathsf{e}_m^T \tag{3.20}$$

of matrix A. In this expression, H_m is an m-by-m upper Hessenberg matrix whose eigenvalues approximate the ones of A. The residual $\mathsf{p}_m\mathsf{e}_m^T$ is a rank one matrix and p_m is orthogonal to the Krylov subspace. Given a starting vector b, the Arnoldi decomposition can be computed

via a modified Gramm Schmidt (MGS) procedure as shown in Algorithm 1. In this algorithm, the previous basis vector is multiplied by the system matrix. After that the projections on all previous basis vectors are subtracted from the vector in a modified Gramm-Schmidt manner.

Algorithm 1 Arnoldi MGS algorithm

$\gamma_1 = ||\mathsf{b}||_2$
$\mathsf{v}_1 = \mathsf{b}/\gamma_1$
for $j = 1, 2, \ldots, m-1$ **do**
  $\mathsf{p}_j = \mathsf{A}\mathsf{v}_j$
  for $i = 1, \ldots, j$ **do**
    $h_{i,j} = \langle \mathsf{p}_j, \mathsf{v}_i \rangle$
    $\mathsf{p}_j = \mathsf{p}_j - h_{i,j}\mathsf{v}_i$
  end for
  $h_{j+1,j} = ||\mathsf{p}_j||_2$
  $\mathsf{v}_{j+1} = \mathsf{p}_i / h_{j+1,j}$
end for

Lanczos decomposition

In the case the system matrix is symmetric with respect to a weighted inner product, a three-term recursion relation between subsequent vectors can be found. This allows fast decomposition with an algorithm complexity linear proportional to the subspace size. The Arnoldi algorithm needs projections on all earlier vectors, which makes its computational cost quadratically dependent on the subspace size.

Since a three-term recursion relation can be found, each new vector in the basis only needs to be orthogonalized with respect to the two previous vectors. Therefore each iteration of the Lanczos algorithm has the same computational costs, and only three Lanczos vectors need to fit the memory of the computational architecture to achieve the decomposition.
As shown in the last section, the Arnoldi algorithm constructs the decomposition

$$\mathsf{A}\mathsf{V}_m = \mathsf{V}_m\mathsf{H}_m + \mathsf{p}_m\mathsf{e}_m^T \tag{3.21}$$

with the residual $\mathsf{p}_m\mathsf{e}_m^T$ a matrix of rank one and p_m orthogonal to V_m.

To derive the Lanczos algorithm, we replace the standard inner product on $\mathbb{C}^n$ by the bilinear form $\langle \mathsf{x}, \mathsf{y} \rangle_{\mathsf{W}^s} = \mathsf{y}^T\mathsf{W}^s\mathsf{x}$ and we normalize the basis vectors such that each vector has unit 2-norm. With $\delta_i = \langle \mathsf{v}_i, \mathsf{v}_i \rangle_{\mathsf{W}^s}$, we then arrive at the decomposition

$$\mathsf{A}\mathsf{V}_m = \mathsf{V}_m\mathsf{H}_m + \mathsf{p}_m\mathsf{e}_m^T, \tag{3.22}$$

with $\mathsf{V}_m^T\mathsf{W}^s\mathsf{V}_m = \mathrm{diag}(\delta_1, \delta_2, ..., \delta_m) = \mathsf{D}_m$ and p_m is W^s-orthogonal to V_m. From the above decomposition, we obtain

$$\mathsf{V}_m^T\mathsf{W}^s\mathsf{A}\mathsf{V}_m = \mathsf{V}_m^T\mathsf{W}^s\mathsf{V}_m\mathsf{H}_m = \mathsf{D}_m\mathsf{H}_m \tag{3.23}$$

and from this equation it follows that $\mathsf{D}_m\mathsf{H}_m$ is symmetric, since A is symmetric with respect to W^s. In other words, we have

$$\mathsf{H}_m^T\mathsf{D}_m = \mathsf{D}_m\mathsf{H}_m \qquad \text{which is the reduced counterpart of } \mathsf{A}^T\mathsf{W}^s = \mathsf{W}^s\mathsf{A}. \tag{3.24}$$

Furthermore, since $\mathsf{D}_m\mathsf{H}_m$ is symmetric, the Hessenberg structure of H_m simplifies to a tridiagonal structure and consequently the basis vectors can be constructed via a three-term recurrence relation. The following Lanczos algorithm implements the above steps by constructing a W^s-orthogonal basis for the Krylov space $\mathcal{K}_m^{\mathrm{PKS}}$.

Algorithm 2 Modified Lanczos algorithm

$\zeta_1 = ||\mathsf{b}||_2$
$\mathsf{v}_1 = \mathsf{b}/\zeta_1$
$\delta_0 = 1$
$\delta_1 = \langle \mathsf{v}_1, \mathsf{v}_1 \rangle_{\mathsf{W}^s}$
for $i = 1, 2, \ldots, m-1$ **do**
  $\alpha_i = \delta_i^{-1} \langle \mathsf{A}\mathsf{v}_i, \mathsf{v}_i \rangle_{\mathsf{W}^s}$
  $\mathsf{p}_i = \mathsf{A}\mathsf{v}_i - \alpha_i \mathsf{v}_i - \zeta_i \delta_i \delta_{i-1}^{-1} \mathsf{v}_{i-1}$
  $\zeta_{i+1} = ||\mathsf{p}_i||_2$
  $\mathsf{v}_{i+1} = \mathsf{p}_i/\zeta_{i+1}$
  $\delta_{i+1} = \langle \mathsf{v}_{i+1}, \mathsf{v}_{i+1} \rangle_{\mathsf{W}^s}$
end for

Observe from the algorithm that the basis vectors are generated via the three-term recurrence relation

$$\zeta_{i+1}\mathsf{v}_{i+1} = \mathsf{A}\mathsf{v}_i - \alpha_i \mathsf{v}_i - \zeta_i \frac{\delta_i}{\delta_{i-1}} \mathsf{v}_{i-1} \tag{3.25}$$

and consequently the ith basis vector (a.k.a. Lanczos vector) is given by

$$\mathsf{v}_i = p_{i-1}(\mathsf{A})\mathsf{b}, \tag{3.26}$$

where $p_{i-1}(\cdot)$ is the so-called Lanczos polynomial. To show the relation between Lanczos polynomials and orthogonal polynomials of one variable, Favard's Theorem is introduced as

Theorem 1. Favard's Theorem: *Polynomials satisfying*

$$p_{n+1}(x) = (a_n x + b_n)p_n - c_n p_{n-1} \quad x \in [-1,1] \quad a_n, c_n > 0 \tag{3.27}$$

are orthogonal with respect to a weighted inner product space.

Examples of such orthogonal polynomials are Legendre and Chebychev polynomials. The solution of the Laplacian operator on a spherical domain is given by Legendre polynomials given in the notation used in Eq. (3.27) as

$$a_n = \frac{2n+1}{n+1}, b_n = 0, c_n = \frac{n}{n+1}, p_0 = 1, p_1 = x \tag{3.28}$$

or Chebyshev polynomials of first kind widely used for interpolation are given by

$$a_n = 2, b_n = 0, c_n = 1, p_0 = 1, p_1 = x. \tag{3.29}$$

Comparing the recursion relation for the Lanczos vectors with Favard's Theorem given in Eq. (3.27), it can be seen that it fits the same formulation. In [18], an extensive treatment on the relation between Lanczos vectors and orthogonal polynomials can found.

3.3.2 Extended Krylov subspace

For extended Krylov subspaces, five term recursion relations can be found as shown in [16, 17]. It was shown that extended Krylov subspaces can approximate matrix functions efficiently, having smaller errors than polynomial subspaces for the same dimension. This paper further discusses the relation between Laurent polynomials and the extended Krylov subspaces.

Approximation of general parameter-dependent matrix equations and of transfer function by projection onto an extended Krylov subspace was investigated in [19]. This research showed the

effectiveness of this approach for realistic applications such as Sylvester equations and shifted systems.

Within the extended Krylov subspace approach for electromagnetic systems, one basically needs to solve Poisson systems while projecting the previous vector on A^{-1}. For wave equations in open domains it was recently shown that extended Krylov subspace approximations can outperform conventional methods especially when low-frequency solutions are required [12].

3.3.3 Rational Krylov subspace

In the rational subspace approach, the system is projected on a sequence of single frequency solutions of the system with b as source. A general rational Krylov subspace is a subspace of the form

$$\mathcal{K}_m^{\mathrm{RKS}} = \mathrm{span}\{(\mathsf{A}+\sigma_1\mathsf{I})^{-1}\mathsf{b}, (\mathsf{A}+\sigma_2\mathsf{I})^{-1}(\mathsf{A}+\sigma_1\mathsf{I})^{-1}\mathsf{b}, \ldots, \prod_{\mathrm{i}=1}^{\mathrm{m}}(\mathsf{A}+\sigma_{\mathrm{i}}\mathsf{I})^{-1}\mathsf{b}\}. \tag{3.30}$$

With properly selected shifts these subspaces can approximate a matrix function within a region of interest with only a small fraction of the dimension needed in polynomial Krylov subspaces. The systems that need to be solved are Helmholtz like systems for different complex frequencies σ_i. Because of a lack of efficient Helmholtz preconditioners for iterative solvers, these systems are computationally expensive, which for certain problems however is outweighted by the fast convergence.

The convergence is strongly dependent on the choice of the initial shifts. Several shift selection techniques have been developed in literature. The work done in [20] presents an algorithm which adapts its shifts and tunes its spectrum towards the spectrum of operator A. It was shown that this approach was superior to equidistant shifts in the case A has a non-uniform spectrum. The problem of selecting optimal shifts for rational Krylov subspaces for the Maxwell diffusion equation was reduced to a third kind Zolotarev problem in [11, 21] and showed improved convergence compared with equidistance shifts. Finally, in [22] it was shown that rational Krylov subspace approaches are attractive for solving the inverse problem. The paper [23] presents an iterative H2 optimal RKS algorithm for approximating dynamic systems.

Using modified Gramm-Schmidt projections for orthogonalization of the basis and given certain shifts denoted by $\sigma = \{\sigma_1, \ldots, \sigma_m\}$, a rational Arnoldi MGS algorithm can be defined as shown in Algorithm 3.

Algorithm 3 Rational Arnoldi MGS algorithm

$\gamma_1 = ||\mathsf{b}||_2$
$\mathsf{v}_0 = \mathsf{b}/\gamma_1$
for $j = 1, 2, \ldots, m$ **do**
  $\mathsf{p}_j = (\mathsf{A}+\sigma_i)^{-1}\mathsf{v}_{j-1}$
  for $i = 1, \ldots, j-1$ **do**
    $\mathsf{p}_j = \mathsf{p}_j - \langle \mathsf{p}_j, \mathsf{v}_i \rangle \mathsf{v}_i$
  end for
  $\mathsf{v}_j = \mathsf{p}_j / ||\mathsf{p}_j||_2$
end for

Within this algorithm the system $\mathsf{p}_j = (\mathsf{A}+\sigma_i)^{-1}\mathsf{v}_{j-1}$ can be solved iteratively or directly using sparse solvers. If only a single shift is used so $\sigma_i = \sigma_{\mathrm{fix}} \forall i = 1, \ldots, m$, an LU (or other) decomposition of the shifted system $(\mathsf{A}+\sigma_{\mathrm{fix}})$ is computationally attractive as the same LU decomposition can be used in every iteration. In this thesis the use of iterative solvers was not explored,

only direct sparse solvers were used.

3.3.4 Convergence Comparison

To obtain insight into how PKS, EKS, and RKS approximate the spectrum of a matrix, we compared these methods from a viewpoint of spectral approximation via power methods. The power method, inverse power method, and shifted inverse power method are reviewed and linked to PKS, EKS, and RKS subspace approximations. All power methods are well described by Saad in [24].

In these power methods, the matrix A, A^{-1}, or $(A-\sigma I)^{-1}$ is raised to a large power and multiplied by a vector in an iterative manner. In the case the matrix has one dominant eigenvalue, and the starting vector has a component in the direction of the associated eigenvector, the method converges towards the dominant eigenpair.

In Algorithm 4 the value α_j converges to the eigenvalue with largest modulus $\lambda_{\max}$, and v_j to its associated eigenvector under the condition that there is only one eigenvalue of largest modulus $\lambda_{\max}$. Furthermore, v_0 needs to have a component in the invariant subspace of $\lambda_{\max}$ for convergence.

Using the inverse of a matrix in the power method, the eigenvalue with the smallest modulus can be found. In the inverse power method as presented in Algorithm 5, the value of α_j converges to $1/\lambda_{\min}$ with $\lambda_{\min}$ the eigenvalues of smallest modulus. Furthermore, v_j converges to the associated eigenvector. For convergence similar constrains as for the power method holds.

The whole spectrum of matrix A can be shifted by a constant value which leads to shifted power methods. By subtracting a value σ from the diagonal of A the whole spectrum shifts by $-\sigma$. The eigenvalue λ_σ of A which is closest to σ becomes the smallest eigenvalue of the shifted system as illustrated in Figure 3.1. Therefore, given specific convergence constraints similar to the power method, the shifted inverse power method as shown in Algorithm 6 converges. The value $\frac{1}{\alpha}$ converges to $\lambda_\sigma - \sigma$ and v_j to the eigenvector of λ_σ.

Algorithm 4 Power method

Chose nonzero start vector v_0
Chose tolerance tol
$j = 0$
while $||Av_j - \alpha_j v_j|| > tol$ **do**
  $j = j+1$
  $\alpha_j =$ Element of Av_{j-1} with max. mod.
  $v_j = \frac{1}{\alpha_j} Av_{j-1}$
end while

Algorithm 5 Inverse power method

Chose nonzero start vector v_0
Chose tolerance tol
$j = 0$
while $||Av_j - \frac{1}{\alpha_j} v_j|| > tol$ **do**
  $j = j+1$
  $\alpha_j =$ Element of $A^{-1}v_{j-1}$ with max. mod.
  $v_j = \frac{1}{\alpha_j} A^{-1} v_{j-1}$
end while

Algorithm 6 Inverse shifted power method

Chose nonzero start vector v_0
Chose tolerance tol
$j = 0$
while $||Av_j - (\frac{1}{\alpha_j} + \sigma)v_j|| > tol$ **do**
  $j = j+1$
  $\alpha_j =$ Element of $(A-\sigma I)^{-1} v_{j-1}$ with max. modulus
  $v_j = \frac{1}{\alpha_j} (A-\sigma I)^{-1} v_{j-1}$
end while

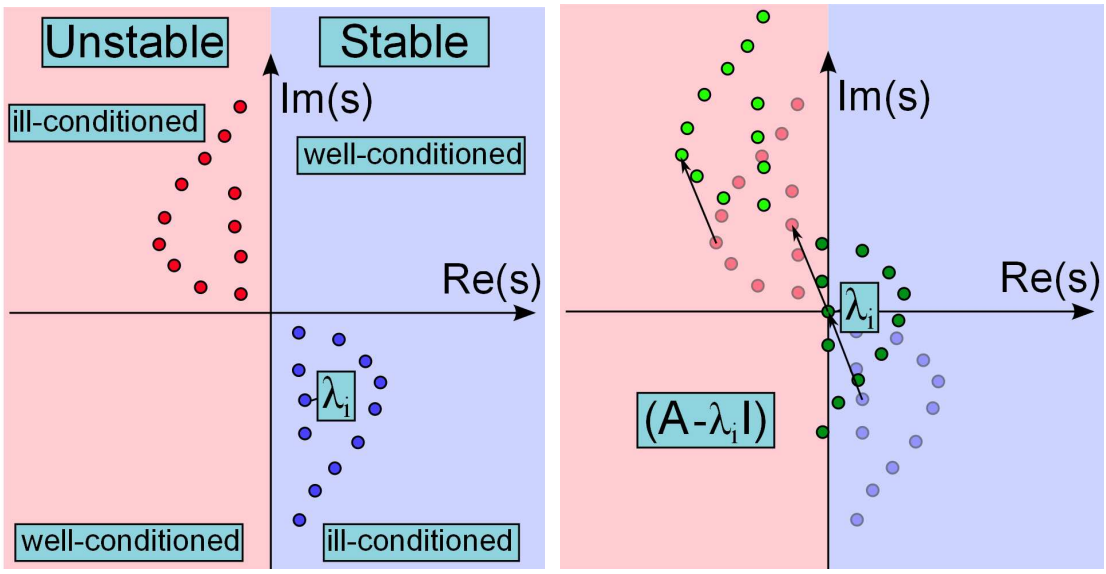
Krylov subspace methods can be seen as a variation of power methods. Where power methods target to approximate one specific eigenpair, Krylov subspace methods target to approximate the whole or at least a substantial part of the spectrum of a matrix. At each iteration the new vector is orthogonalized with respect the whole basis, such that once an eigenvector is within the subspace, the following vectors have no component in the invariant subspace of the associated eigenvalue. The latter is a requirement for convergence towards that specific eigenvalue. Therefore, the method approximates new eigenvectors with every iteration.

By comparing the basic iteration of the rational, extended, and polynomial Krylov subspace with the described power methods, similarities can be found. The ordinary power method has the same basic iteration as the PKS method, while the EKS combines inverse and ordinary power method. Finally, the rational Krylov subspace method has the same core iteration as the inverse shifted power method.

Although this comparison is a large simplification of the convergence of Krylov subspace methods, especially when multiple shifts are used, it leads to basic and rough insight into how the subspace methods converge.

The PKS approach approximates large eigenvalues first. With increasing dimension of the subspace the poles of the reduced-order system converge towards smaller eigenvalues. The polynomial Krylov subspace approach therefore converges from the outside to the inside of the spectrum of matrix A.

For a RKS approach the shifts σ can be selected in such a manner that the projected system converges form a point in the complex plane outwards, or on a certain domain of interest in the region of the selected shifts. This domain of interest can be defined as a rectangle in the complex plane via the bandwidth of the input signal and the time of interest. The former puts a constraint on the imaginary part and the latter on the real part of the poles of relevance. The illustration in Figure 3.1 shows which choice of shifts leads to well- and ill-conditioned systems and also shows the effect of shifting a system. The shifts in the RKS approach can be used to adapt the spectrum of the approximation subspace to the spectrum of the operator. According to the power method analogy, the shifts should be chosen as mirror images of the poles of interest.



(a) Poles of the original system.

(b) Poles of the shifted system.

Figure 3.1: Illustration showing the influence of shifts on the spectrum of the operator A in complex plane.

The EKS approach can be viewed as a rational Krylov subspace approach with one interpolation point at the origin of the extended complex plane and the other at infinity. It therefore converges from the origin outwards and from infinity inwards. Figure 3.2 shows a schematic comparison of the convergence behavior of the three Krylov methods.

In the numerical experiments with RKS presented in the next chapter, direct solvers are to solve the Helmholtz-type systems. However, for larger systems iterative solvers might need to be incorporated into the RKS algorithm, such that a shift selection in the well-conditioned quadrant might be beneficial. Using the conjugated transpose instead of the mirror image of the poles of interest for instance leads to well-posed systems. In fact for poles belonging to a high quality factor resonance, both choices are very similar and both show good convergence.

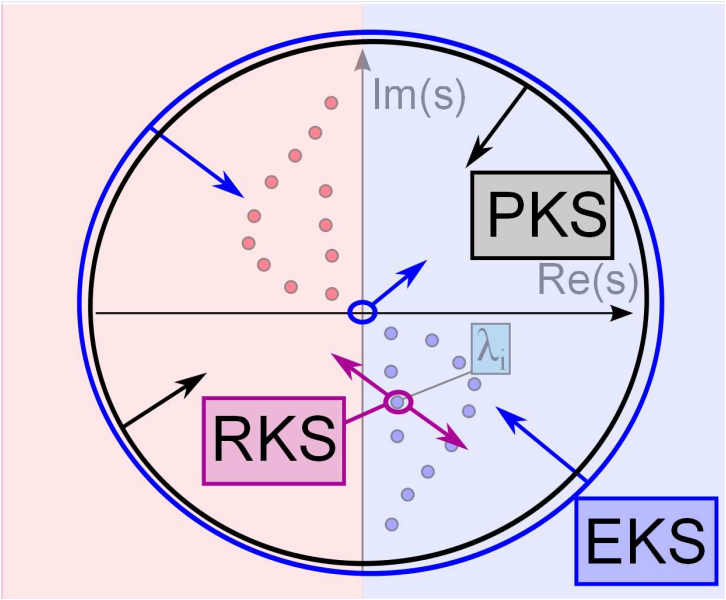


Figure 3.2: Convergence of eigenvalues schematically shown for polynomial (PKS), extended (EKS) and rational (RKS) Krylov subspace projections. The stable poles are shown in the fourth quadrant.

Chapter 4

Results

The numerical examples illustrating the developed methods are presented in this chapter. First model order reduction of instantaneously reacting media is presented in one, two, and three spatial dimensions, given in Section 4.1, Section 4.2 and Section 4.3, respectively. The examples are chosen to illustrate the convergence properties of the various Krylov methods in the time- and frequency-domain, as well as the capability of the methods to deal with very large systems.

Two numerical examples are presented for media exhibiting relaxation, namely, a two-dimensional and a three-dimensional one using a Drude dispersion model. The two-dimensional example shown in Section 4.4 is used to validate the method and shows convergence of the time-domain result to an ADE-FDTD comparison simulation.

The three-dimensional example given in Section 4.5 is used two compare the frequency-domain result of the reduced-order model with the frequency-domain result of an aperiodic Fourier modal method used in [1] to calculate the spontaneous decay rate of a quantum emitter in close proximity to a golden nanorod.

Polynomial Krylov subspaces are used in all numerical experiments. In addition, model order reduction using rational Krylov subspaces is presented for all two-dimensional examples. Finally, the performance of the different algorithms for the chosen examples are compared in Section 4.6.

4.1 Instantaneously reacting materials – One-dimensional configurations

In this section two one-dimensional examples of instantaneously reacting media are given to show that the numerical scattering poles coincide with the analytical poles. Therefore it can be concluded that the numerical method describes the physics correctly. It validates the approach of truncating a domain with an optimal PML to simulate its extension to infinity and find its resonances.

The Lanczos algorithm as presented in Algorithm 2 is used for all examples shown in this section. All domains are truncated with an optimal PML for propagating and evanescent waves. The optimization interval is tuned to the frequency of operation.

4.1.1 Dielectric slab

The first investigated configuration is a one-dimensional electromagnetic wave problem. A dielectric slab with $\varepsilon_r^{\text{slab}} = 4$ and width of λ_{peak} is simulated. The configuration is shown in Figure 4.1. As a source wavelet, the time derivative of a Gaussian pulse is used, with a maximum of the spectrum at λ_{peak}.

The acoustical equivalent of the described configuration is studied in [25]. With this first example it will be shown that the results obtained in [25] can be reproduced with a model order

reduction approach. In the cited work the full equations are solved in the frequency-domain. Therefore the approach presented in this thesis is advantageous in two ways. First, the model order reduction approach allows for smaller models and faster computation. Second, the cited approach can only calculate Fourier domain solutions and no direct solutions in the time-domain can be obtained.

In order to check the convergence of the results calculated with the reduced-order model, an FDTD algorithm with a leapfrog time stepping scheme is used for comparison as described in [26, 27]. The domain is truncated using the optimal PML and the problem size is $n = 1019$. (500 secondary grid points, 499 primary grid points, 4×5 PML layers). The convergence of the Lanczos reduced-order model towards the FDTD comparison response is show in Figure 4.2.

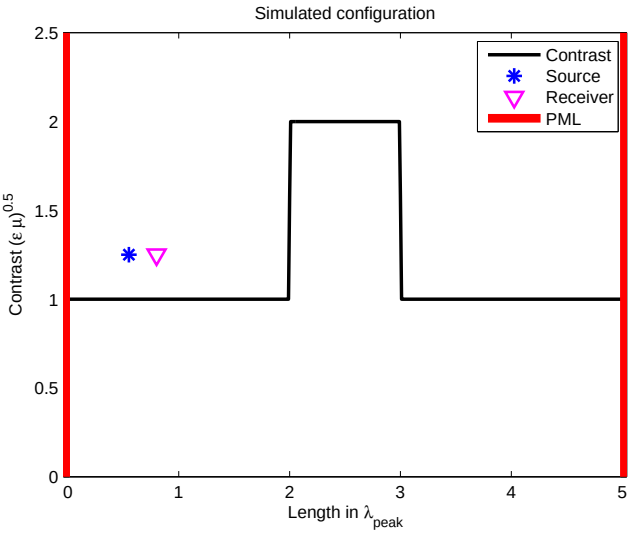


Figure 4.1: Simulated configuration: A one-dimensional slab with source and receiver on one side.

The analytical poles of the system are calculated in [28] and follow from the condition

$$1 = \left(\frac{Y^{\text{vac}} - Y^{\text{slab}}}{Y^{\text{vac}} + Y^{\text{slab}}}\right)^2 \exp\left(-2\sqrt{s\mu^{\text{slab}}(\sigma^{\text{slab}} + s\varepsilon^{\text{slab}})}d^{\text{slab}}\right)$$

with

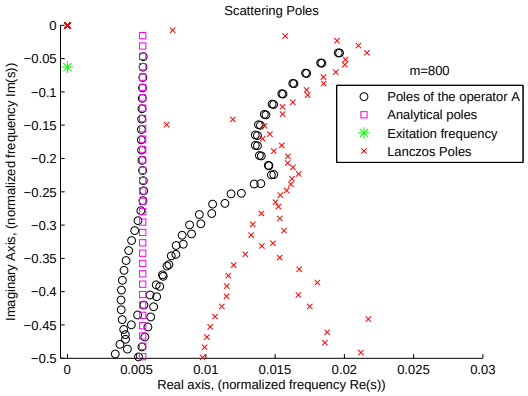
$$Y = \sqrt{\frac{\sigma + s\varepsilon}{s\mu}}.$$

For the given configuration and with normalized parameters for numerical stability, the scattering poles are given by

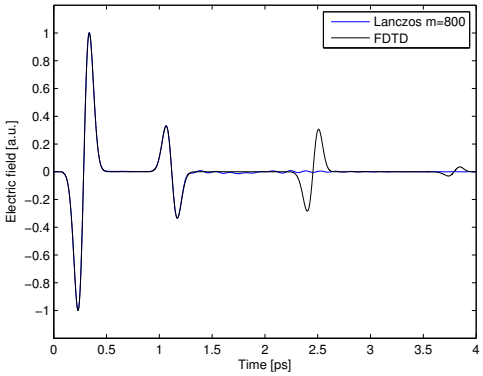
$$s^{\text{sc}} = \frac{1}{4d^{\text{slab}}}\ln(9) - \frac{\mathrm{i}\pi}{2d^{\text{slab}}}n \quad \forall n \in \mathbb{N}^+.$$

Figure 4.2 shows the location of the analytical poles, the poles of the operator A, and the poles of the Lanczos reduced-order model as they converge. It can be seen that the late scattering events that happen inside the slab are modeled correctly, as soon as the Lanczos poles converge to the analytical scattering poles.

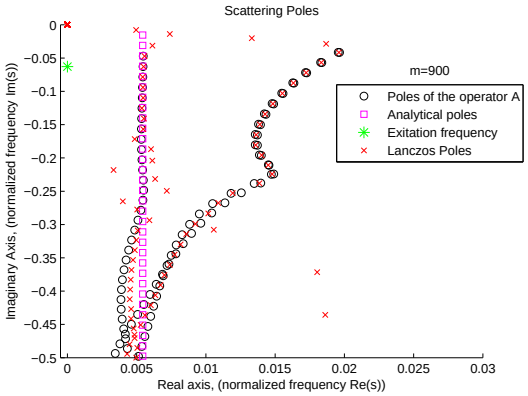
Furthermore, it can be seen that the Lanczos poles approximate the poles of A. Furthermore, for high imaginary values of s, the poles of A do not match the analytical poles, since the spatial discretization and PML are not tuned for those frequencies. The figure also shows that the scattering poles are relatively small in absolute value compared with the other poles of A. Therefore the convergence of the Lanczos algorithm is not optimal, as it usually finds the large eigenvalues first.



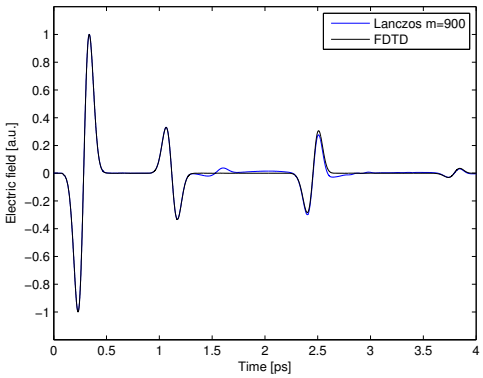
(a) Poles of the system in comparison with the Lanczos approximation for $m = 800$.



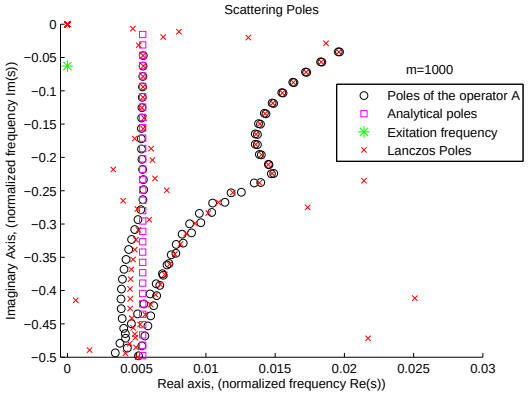
(b) Time-domain result for $m = 800$.



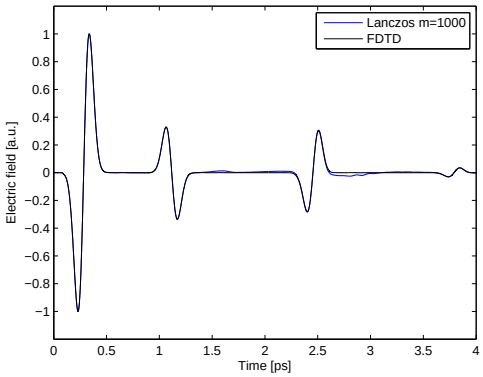
(c) Poles of the system in comparison with the Lanczos approximation for $m = 900$.



(d) Time-domain result for $m = 900$.



(e) Poles of the system in comparison with the Lanczos approximation for $m = 1000$.

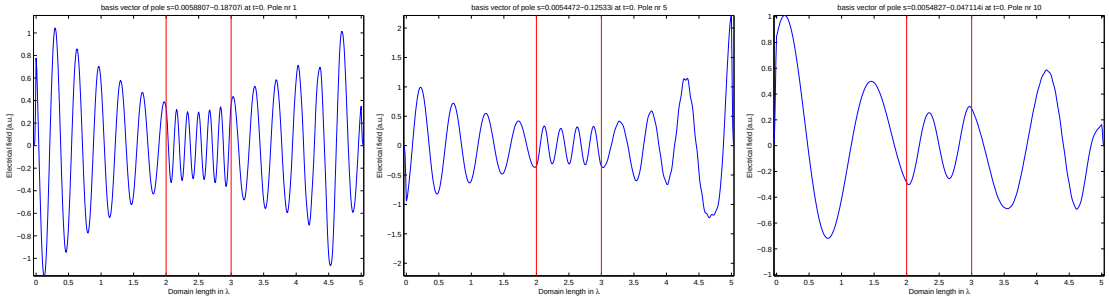


(f) Time-domain result for $m = 1000$.

Figure 4.2: Convergence of the ROM poles and time-domain solution with increasing ROM dimension.

The eigenfunctions shown in Figure 4.3 show good agreement with the eigenfunctions given in [25], where the acoustic equivalent case was studied. In the results shown, even and uneven eigenfunctions can be found, equivalent to the cited work. The amplitude of the waveform increases with increasing distance from the origin. In addition, the eigenfunctions decay with increasing time or once they have reached the PML. When viewed in time it can be seen that standing waves can be found inside the slab and traveling waves outside the slab. The eigen-

functions modeled correctly correspond to $2\pi - 12\pi$ resonances. The corresponding poles have been labeled in Figure 4.4.



(a) 12π resonane of pole #1. (b) 8π resonance of pole #5. (c) 3π resonance of pole #10.

Figure 4.3: Snapshot at $t = 0$ of some of the eigenfunctions that belong to the scatteringpoles.

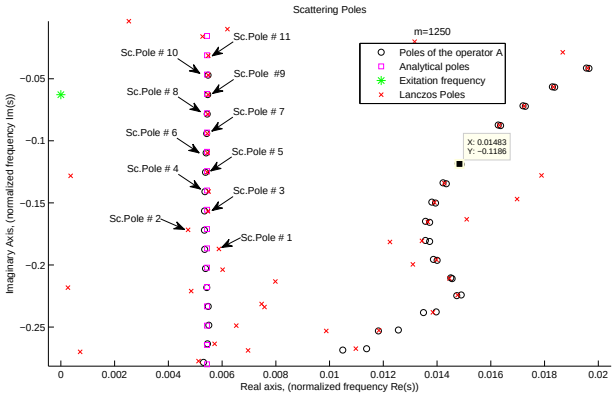


Figure 4.4: Labeled scattering poles to show the eigenvalues of the resonances.

4.1.2 Photonic crystal

The second one-dimensional example is a periodic structure. Structures like this can for instance be found in lasers or chips of integrated circuits. The simulated configuration is illustrated in Figure 4.5. The source is placed in the middle of the configuration. The thickness of the layers is $a_p = 3\ \mu$m and $b_p = 30\ \mu$m and the total finite lattice consists of 41 primitive cells. Finally, we use a step size of $\delta = 1\ \mu$m and nine PML layers resulting in a total number of $n = 2761$ unknowns.

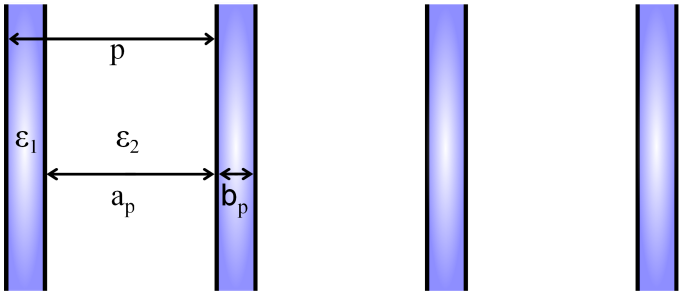


Figure 4.5: Illustration of the simulated 1D periodic structure. The parameters are given by $a_p = 3\ \mu$m, $b_p = 30\ \mu$m and $\varepsilon_1 = 3$, $\varepsilon_2 = 1$.

An infinite periodic structure can also be analyzed using the plane wave expansion method described in Chapter 4 of [29]. With this method the band structure of an infinite periodic structure can be calculated. Now the Lanczos poles, which match the poles of the original operator A, of the finite periodic structure can be compared to the band structure obtained by the plane wave expansion method.

The comparison is shown in Figure 4.6, where the normalized imaginary Laplace frequencies are depicted on the y-axis. On the x-axis the PWE method depicts the wave vector in the fundamental interval (Brillouin zone) $[-\pi/(a_p+b_p), \pi/(a_p+b_p)]$whereas the Krylov subspace approximation plots the normalized real Laplace frequencies.

The figure shows that the Krylov subspace approach is able to capture the band structure of a periodic structure. Moreover, the frequencies associated with the middle of a band have higher radiation losses that the frequencies at the edges of a band. In other words, the quality factor of the resonances in the structure of the frequencies at the sides of a band is higher.

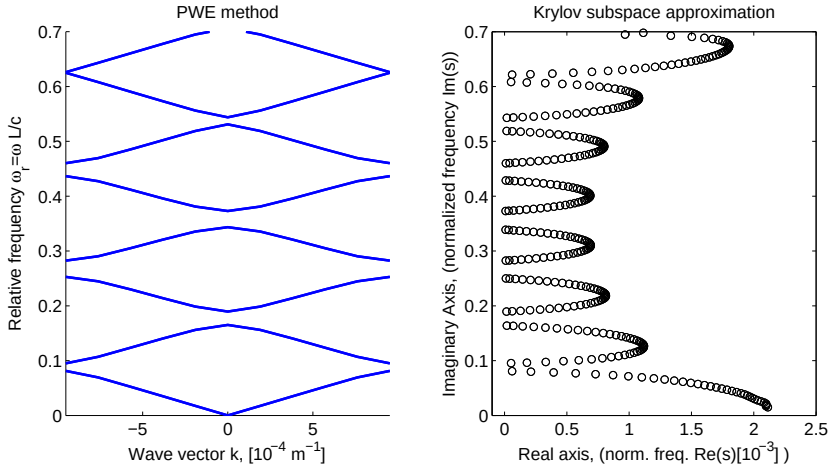


Figure 4.6: Band structure of the photonic crystal, calculated via plane wave expansion, and using the Krylov subspace approach.

4.1.3 Conclusion

In this section it was shown that truncating the domain with an optimal PML is a valid approach to simulate an open system. The numerically calculated resonances match the analytical resonances.

The one-dimensional slab experiment showed that the analytical poles match the poles of the truncated operator A. The one-dimensional periodic structure showed that even multiple scattering events in a periodic structure are modelled correctly by the chosen approach. The resulting scattering poles showed a clear band structure. It can be concluded that the numerical method describes the physics correctly.

4.2 Instantaneously reacting materials – Two-dimensional configurations

In this section two-dimensional systems are analyzed in order to show that the order of a system can be reduced and the correct time-domain behavior can be obtained without loss of accuracy. Furthermore, it will be shown that the field can be expanded in a small number of scattering poles using the scattering expansions of Eq. (2.138). The analyzed configurations are a dielectric box, a photonic crystal waveguide, and a layered Earth model.

4.2.1 Dielectric box

Polynomial Krylov MOR

The first investigated configuration is a two-dimensional electromagnetic wave problem. A dielectric box with a relative permittivity $\varepsilon_r = 4$ and width of 50 μm is simulated. The source and receiver are placed inside the box. The configuration is shown in Figure 4.7. The time derivative of a Gaussian pulse, with a maximum of the spectrum at $\lambda_{peak} = 94$ μm, is used as source wavelet for an H-polarized electromagnetic field.

The domain is truncated using an optimal PML and the problem size is $n = 37000$. The convergence of the Lanczos reduced-order model towards the FDTD comparison response takes 1500 iterations and is shown in Figure 4.8.

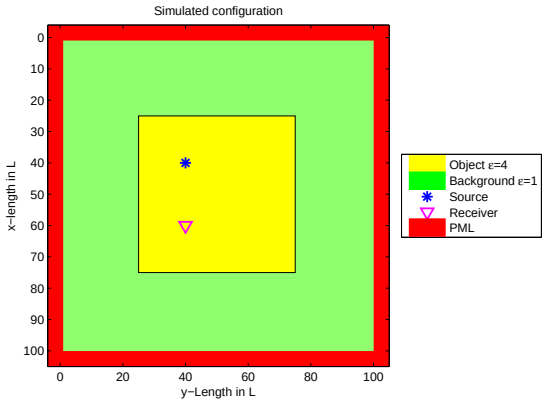
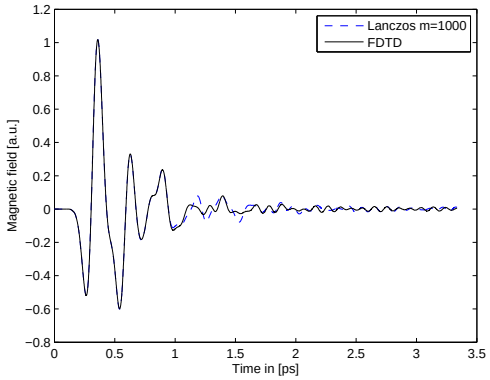
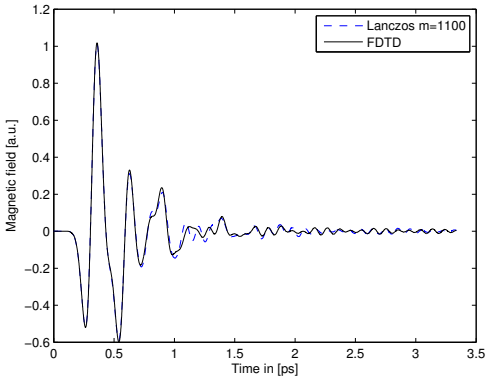


Figure 4.7: Two dimensional dielectric box with source and receiver inside the box.

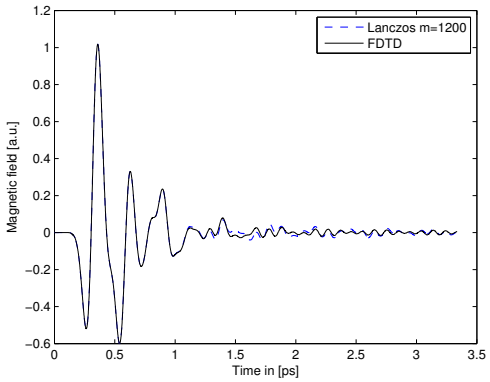
We can now try to expand the field in a reduced set of resonances that are located in a rectangular subdomain of the complex plane determined by the frequency band of the pulse and the length of the time interval of observation. To select the dominant scattering poles, two



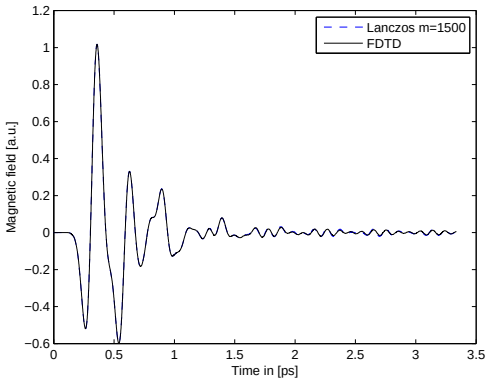
(a) Time-domain result for $m = 1000$.



(b) Time-domain result for $m = 1100$.



(c) Time-domain result for $m = 1200$.



(d) Time-domain result for $m = 1500$.

Figure 4.8: Convergence of the ROM system response with increasing ROM dimension.

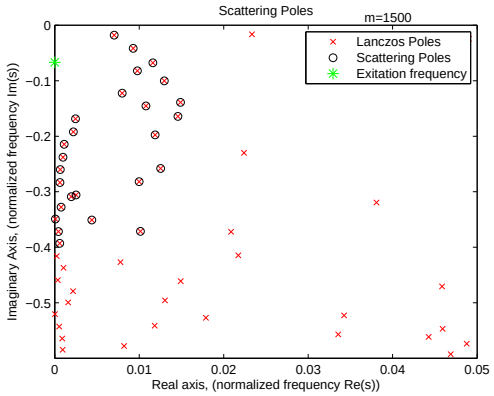
aspects are of importance. First, the time window of interest puts a constraint on the real part of the relevant poles and second, the imaginary part of the poles can be bounded by the frequency spectrum of the input signal.

For the studied problem, 26 scattering poles lie in the rectangular subdomain of interest. The selected poles and corresponding time-domain signal are shown in Figure 4.9. It can be seen that the 26 poles describe the system response well except for times earlier than 0.4 ps. Here the impact of the expected error dropping exponentially in time can be seen, as all late times are modelled correctly.

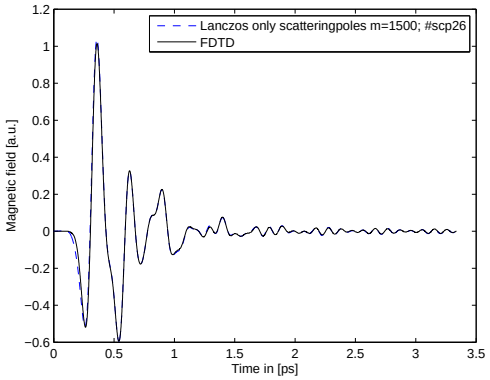
By selecting the thirteen poles with an 1/e decay faster and slower than 1 ps, we can choose whether early or late time events are modeled correctly. It can be seen that changing the constraint on the real part of the poles of interest changes the time interval on which the expansion converges.

The experiment shows that the scattering response can be expanded in a small number of scattering poles. The Lanczos reduced-order model is already a factor 20 smaller than the full order model, but the result shows that dependent on the time interval of observation a model order reduction of up to 1000 is possible.

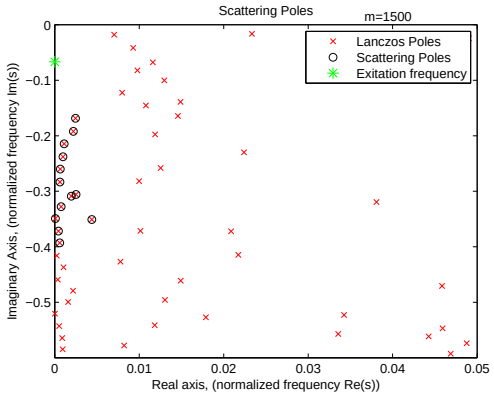
Snapshots at $t = 0$ of two of the resonances f^{res} that belong to the scattering poles are shown in Figure 4.10 (calculated via Eq. (2.138)). Only the fundamental magnetic field is shown. When viewed in time, the eigenfunctions show a standing wave pattern inside the box and traveling waves are present outside the box.



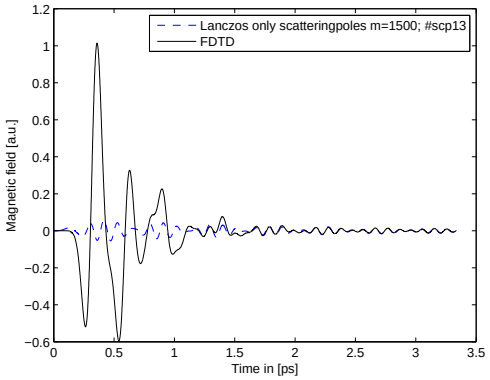
(a) The 26 scattering poles needed to model the whole time-domain solution.



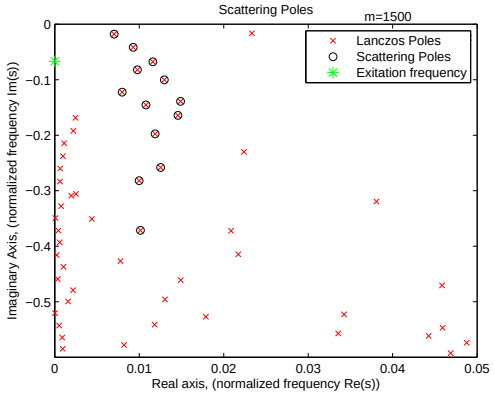
(b) Time solution expanded in 26 scattering poles.



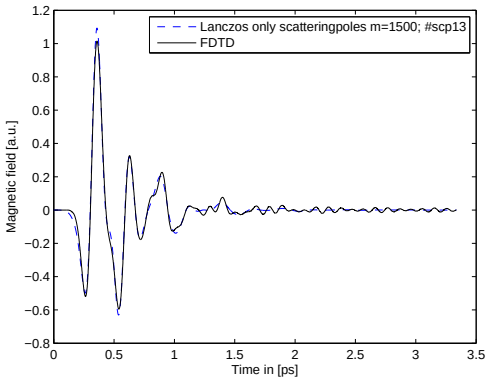
(c) The 13 scattering poles needed to model the late time-domain solution.



(d) Time solution expanded in 13 scattering poles with the smallest real part.

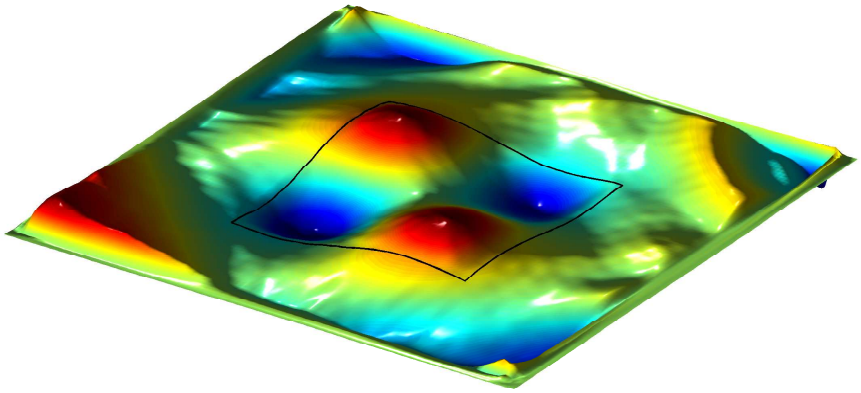


(e) The 13 scattering poles needed to model the early time-domain solution.

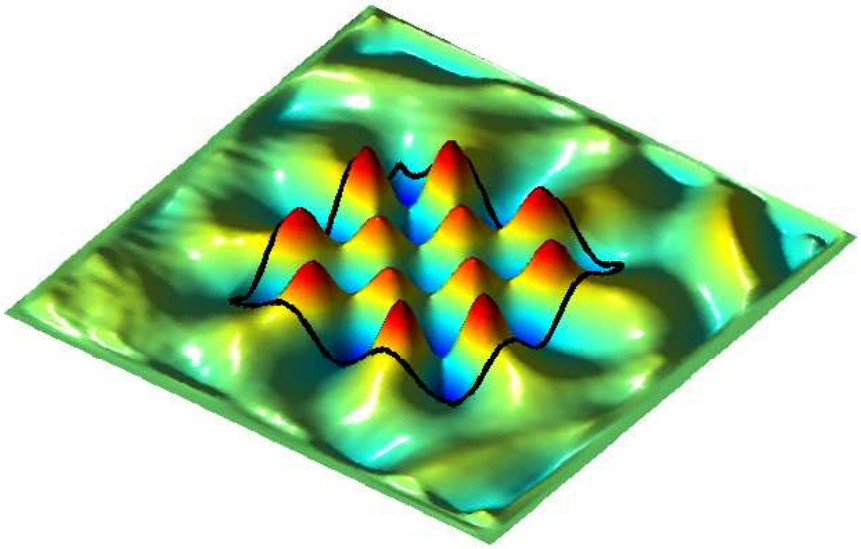


(f) Time solution expanded in 13 scattering poles with the largest real part.

Figure 4.9: Expansion of the system response in the scattering poles of the system. Left: the selected scattering poles to capture the field approximation. Right: resulting field approximation. Early times are captured by selecting poles with a small real part.



(a) $2\pi-$ resonance of the box.



(b) $5\pi-$ resonance of the box.

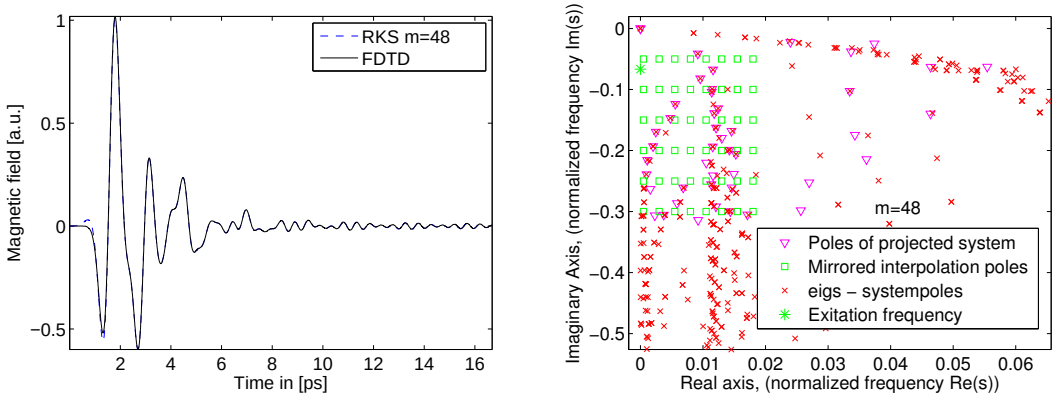
Figure 4.10: Snapshots of two resonance eigenfunctions f^{res} of the system in the time-domain. The black box indicates the boundary of the box.

Rational Krylov MOR

The dielectric box example together with the RKS approach is used to show that the method converges using equidistant shifts. Initial pole selection is done in a manner similar to defining a selection criterion for resonance expansions. The imaginary part of interest is defined by the bandwidth of the used pulse, whereas the real part is set by the time interval of interest.

Subsequently, equidistant shifts are used to fill the defined square. Using these shifts to construct the rational Krylov subspace, an approximate solution can be obtained by projection of the system matrix. In Figure 4.11 the response approximated by the RKS and FDTD are compared. It can be seen that the whole system response is approximated well after $m = 48$ iterations.

Furthermore, the interpolation points of the RKS together with the final system poles and the eigenvalues of the non-reduced system are shown in the right part of the figure. It can be seen from the figure that most poles of the projected system approximate an eigenvalue of the system matrix and the operator is well approximated inside the rectangle defined by the initial shifts, whereas only individual poles are approximated outside the rectangle.



(a) System response of the reduced model.

(b) Poles of the reduced and original system.

Figure 4.11: Rational Krylov subspace approximation of the box problem with equidistant shifts.

4.2.2 Photonic crystal waveguide

In the second two-dimensional example, a photonic crystal waveguide is simulated as depicted in Figure 4.12. The configuration is simulated using E-polarized fields and has $n \approx 0.7$ million unknowns. The depicted cylinders have a relative permittivity of $\varepsilon_r = 11.56$, whereas vacuum is assumed in between the cylinders. As a current source, a Gaussian pulse shifted to a frequency of $f_s = 192.21$ THz was used. This frequency lies in the band gap of the simulated photonic crystal, such that the wave can only propagate in the path of crystal.

Polynomial Krylov MOR

The convergence of the electrical field at the receiver towards an FDTD comparison result is shown in Figure 4.13. After roughly 3000 Lanczos iterations the signal converged to the FDTD result. Again the signal at the receiver can be expanded in resonance eigenfunctions of the system. In the rectangle of interest, only a single pole (!) can be found as shown in Figure 4.14(a). This single pole can describe the time-domain solution at the receiver as shown in Figure 4.14(b). Essentially this means that a single mode is present in the photonic waveguide. This example clearly shows the impact of resonance expansions for certain type of structures.

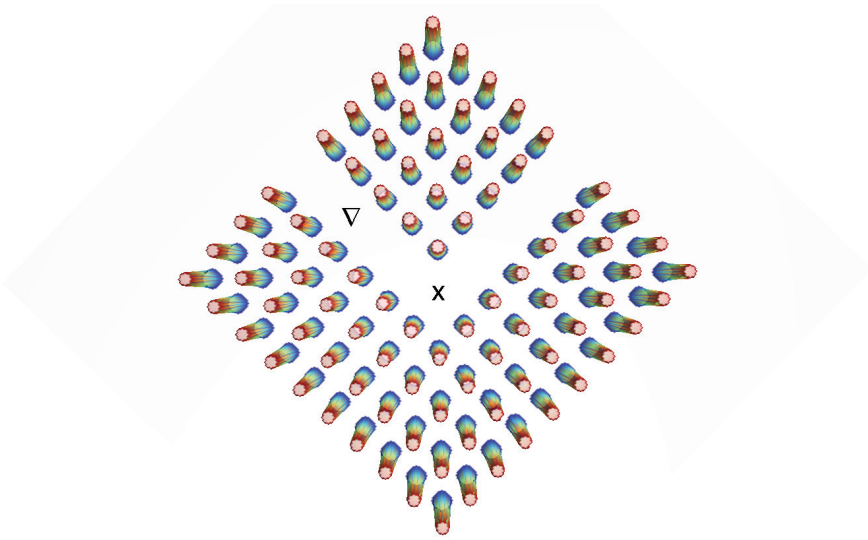
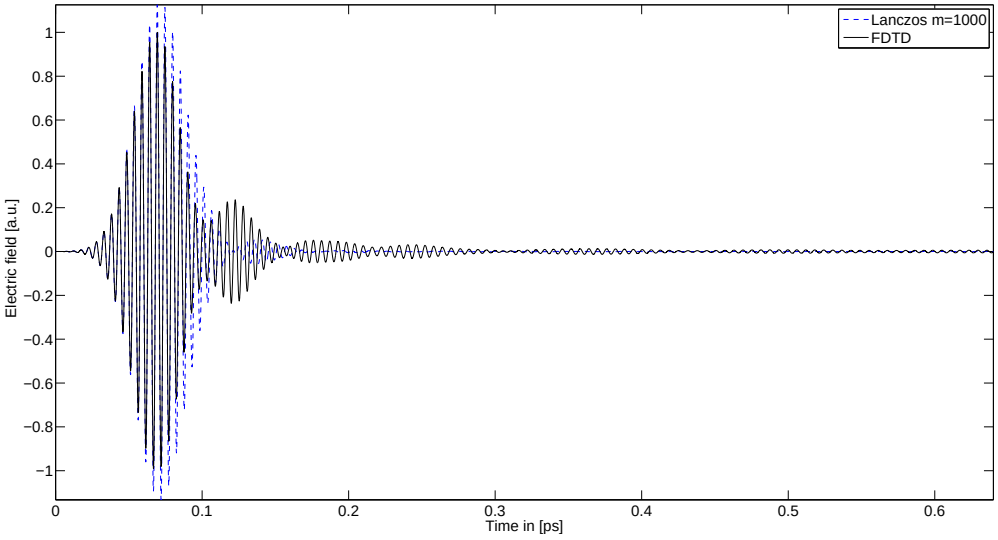
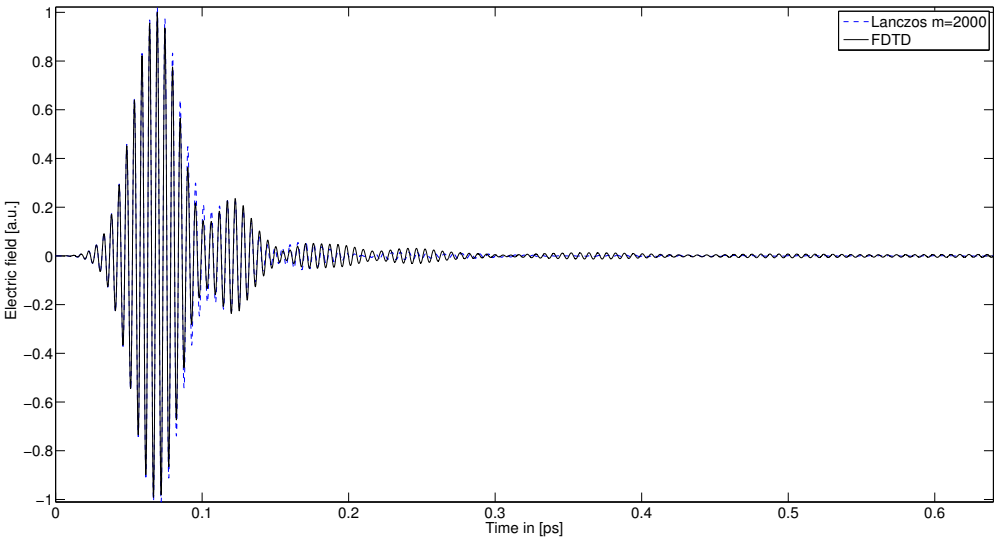


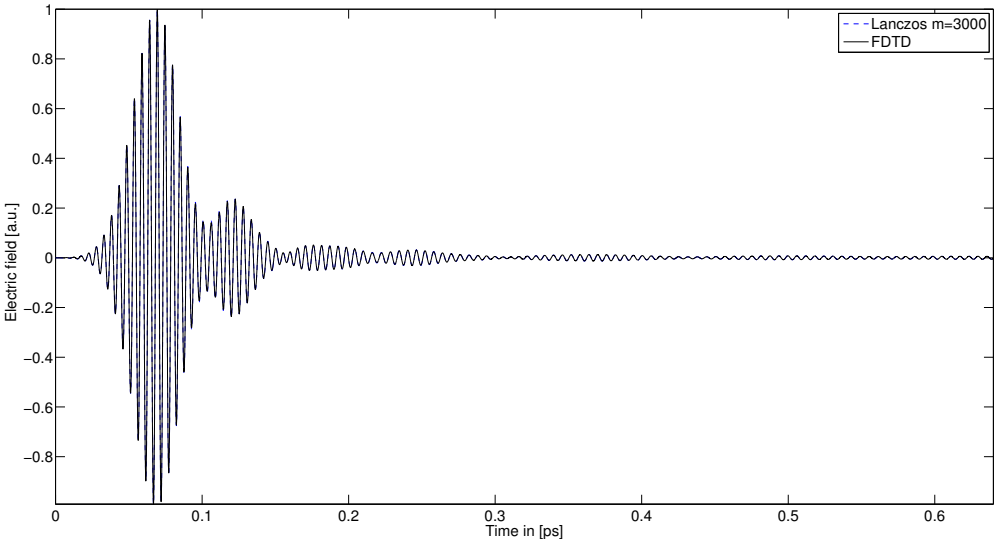
Figure 4.12: Simulated photonic waveguide crystal. The source is marked by an X whereas the receiver is marked as a triangle.



(a) Time-domain result for $m = 1000$.

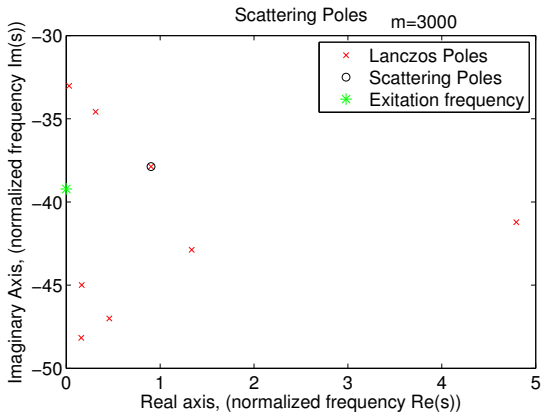


(b) Time-domain result for $m = 2000$.

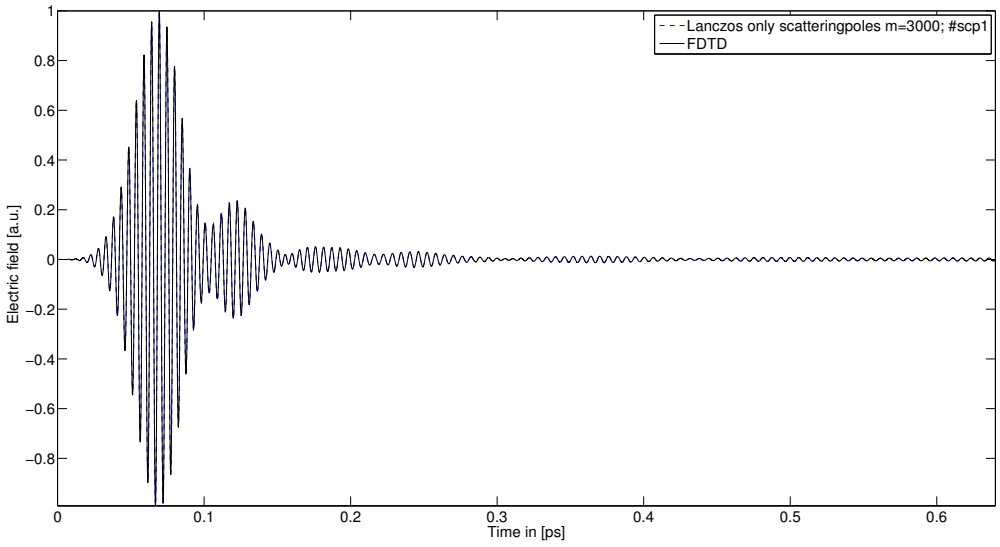


(c) Time-domain result for $m = 3000$.

Figure 4.13: Convergence of the ROM system response with increasing ROM dimension.



(a) Used scattering pole in the expansion.



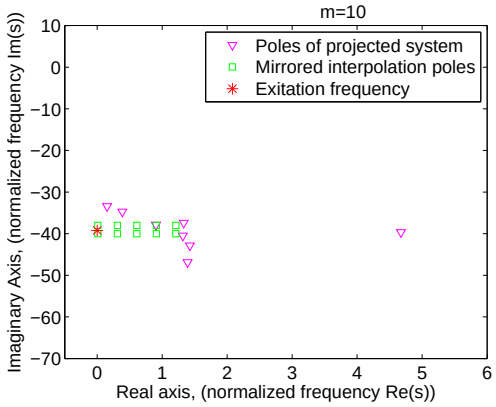
(b) Time-domain result for $m = 3000$ with one scattering pole.

Figure 4.14: Expansion of the system response in the scattering poles of the system.

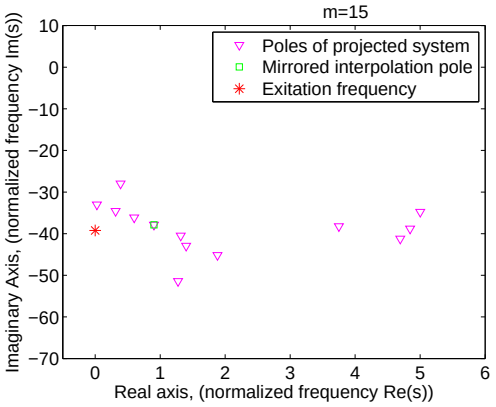
Rational Krylov MOR

In the last section it was shown that the response of the photonic crystal waveguide example could be expanded in a single pole. Therefore, next to the standard method of using equidistant shifts, also single shifts are used in this example. As a single shift, the pole found in the scattering expansions of the last chapter is used. Obviously, this is *a priori* information which is normally not available. However it shows that convergence is influenced by the choice of shifts. The initial shifts and final system poles of both approaches are shown in Figure 4.15, where equidistant shifts are used in Figure 4.15 (a,c), while a single shift is used in Figure 4.15 (b). The resulting time-domain responses are shown in Figure 4.16.

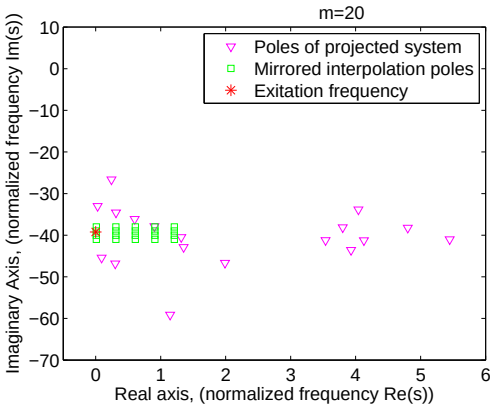
After ten iterations, the early-time response is well approximated independent of the choice of shifts. The RKS using equidistant shifts converged after 20 iterations, where only 15 iterations are needed in the case of using a single shift. In Figure 4.15(b) it can be seen that the interpolation pole (mirror image of σ) coincides with a system pole. This system pole is the same pole which lead to a resonance expansion with only one mode in the last chapter.



(a) Poles of the reduced system and interpolation shifts for m=10.

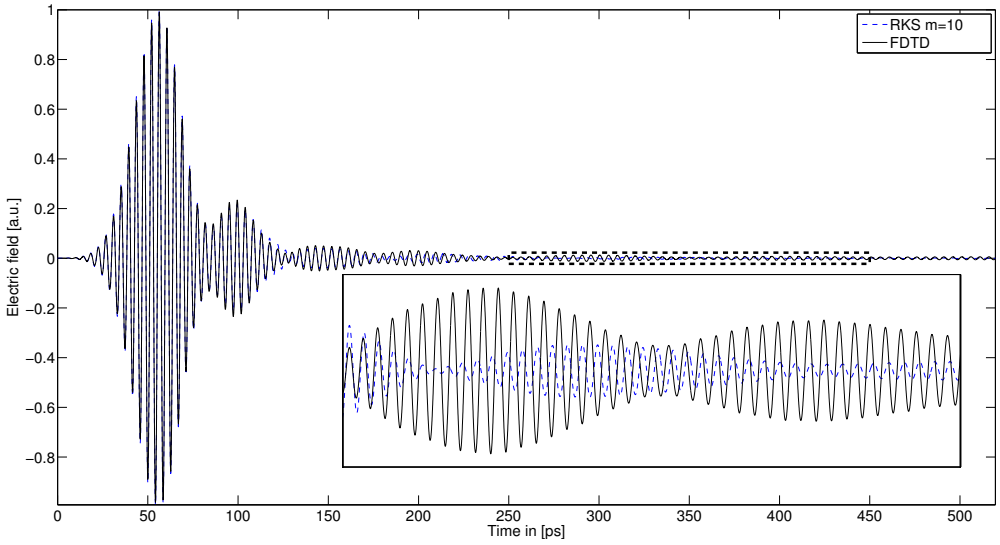


(b) Poles of the reduced system using a single shift and m=15

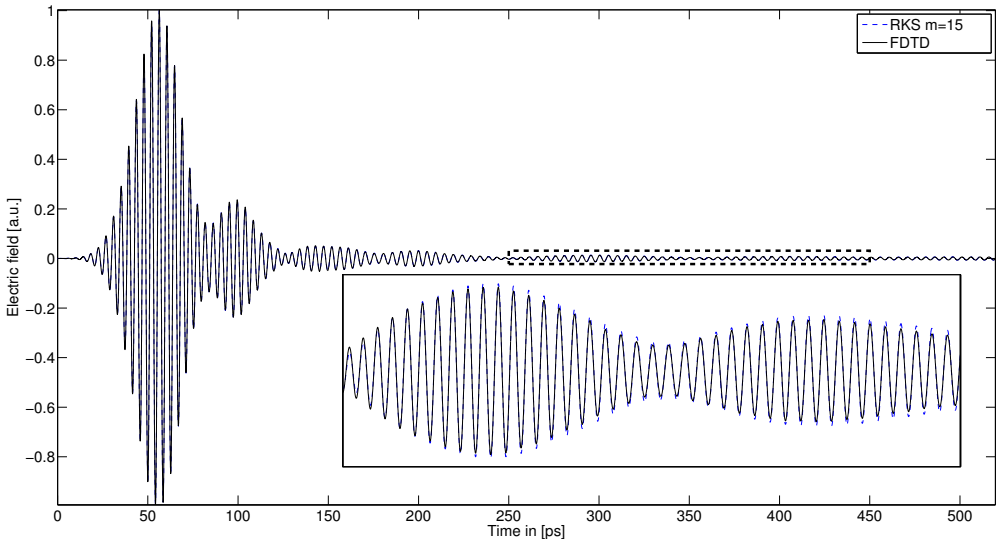


(c) Poles of the reduced system and interpolation shifts for m=20.

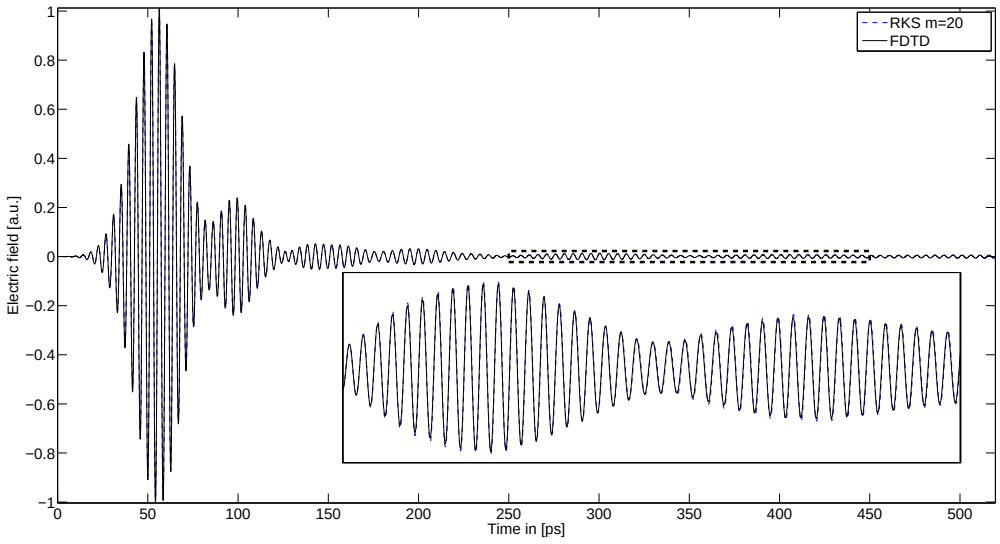
Figure 4.15: Rational Krylov subspace approximation of the spectrum of the crystal problem with equidistant shifts.



(a) System response for a RKS dimension of m=10.



(b) Converged system response for a RKS dimension of m=15 using a single shift.



(c) Converged system response for a RKS dimension of m=20.

Figure 4.16: Rational Krylov subspace approximation of the crystal problem response with equidistant shifts and single shifts.

4.2.3 Layered Earth Example

In this example, a layered Earth structure is analyzed. The simulated configuration is shown in Figure 4.17 and includes a horizontal Earth layer and a vertical drilling hole. For this problem the problem size is $n \approx 545000$. In industry these kind of structures are simulated for drilling hole characterization. Within inversion schemes they are simulated using FDTD codes followed by Prony-type methods to obtain the dominant poles of the transfer function responses between a source and a receiver. Using a Krylov model order reduction approach has the advantage that the transfer function follows from the configuration and algorithm itself.

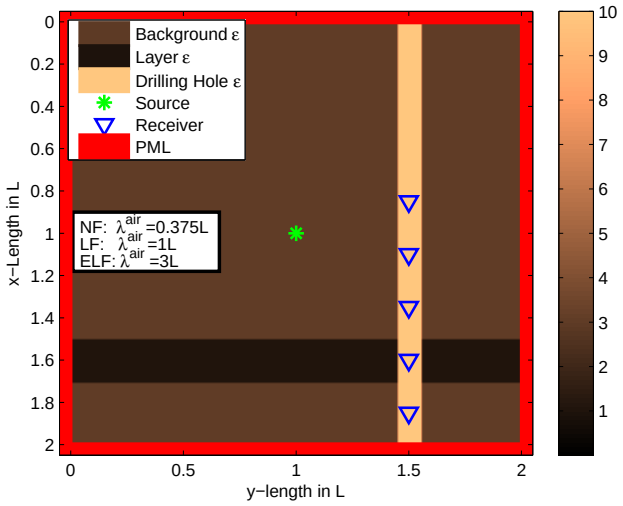


Figure 4.17: Simulated configuration: Earth layer and drilling hole. Top receiver is labeled as receiver 1 and the lowest as receiver 5.

Polynomial Krylov MOR

As a source wavelet, again the time derivative of a Gaussian pulse is used (Ricker wavelet). The wave field along a line from receiver 1, the top receiver, to receiver 5, the lowest receiver, is shown in Figure 4.18. The wave field at every receiver separately is shown in Figure 4.19. The response converged after roughly 3500 iterations of the polynomial Krylov subspace method. It can be seen that the pulse at receiver 4 resonates much stronger than for the other receivers.

The poles in the region of interest are shown in Figure 4.20. The poles close to the imaginary axis are the scattering poles of interest. They can be related to resonances in the two slabs, namely, the Earth layer and the drilling hole.

Now we can expand the solution in the scattering poles labeled by the circles in Figure 4.20. There are 76 poles in the bandwidth of the pulse and have a reasonable small real part. The result for receiver 3 is shown in Figure 4.20, which shows that a good approximation can be found with 76 poles for the chosen configuration. It can be seen that only a very small error arises for early times. For this example this is no problem, since the error occurs for time instances smaller than the physical arrival time, which makes filtering easy.

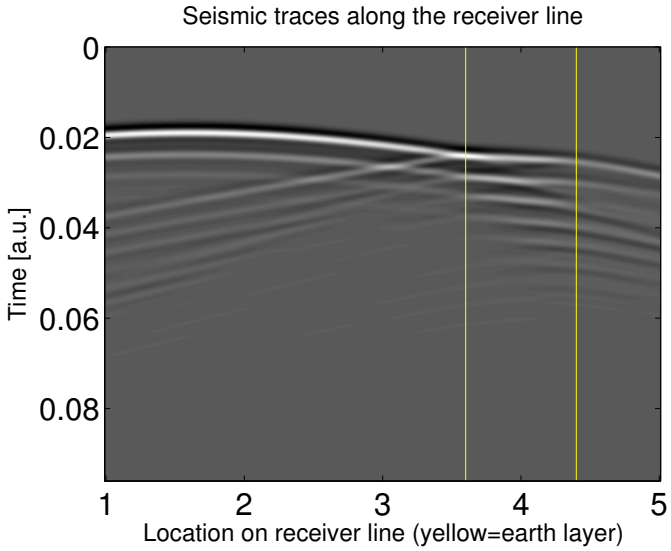
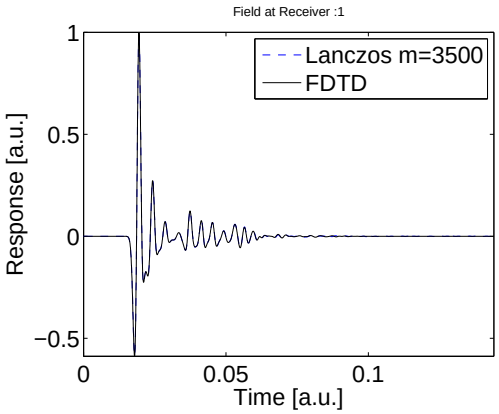


Figure 4.18: Wavefield along a line from receiver 1 to 5 in the drilling hole. The yellow lines mark the horizontal position of the Earth layer.

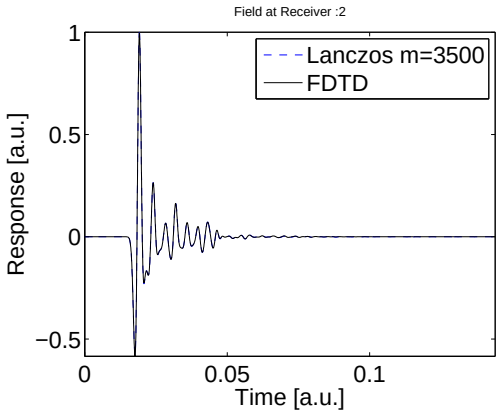
Rational Krylov MOR

In this example, two different choices for the shifts of the rational Krylov subspace were made. The shifts of the rational subspace where chosen either as 120 poles along a line parallel to the imaginary axis, or based on the 76 eigenvalues of the polynomial subspace, which were used to expand the field in Figure 4.20. The RKS method converges for both chosen sets of poles, but fails to converge in the case less poles are used, as shown in Figure 4.21. This shows that adaption of the interpolation points of the RKS to the spectrum of the approximated operator improves convergence.

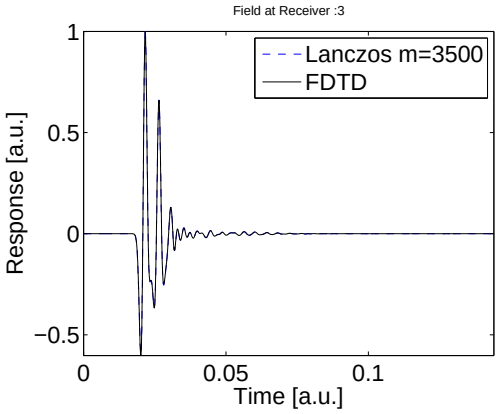
After decomposition, the weight of the poles can be sorted, in order to identify the most contributing modes. These modes are illustrated in Figure 4.22. It was shown that the pole with an eigenfunction as shown in Figure 4.22(a) is dominant at all receivers except for the one next to the Earth layer. The field resonates in the drilling hole. The pole with an eigenfunction shown in Figure 4.22(b) is dominant at the receiver next to the Earth layers. This eigenfunction shows a resonating field at the cross section of the Earth layer and the drilling hole, with outgoing waves to the surrounding medium. In Figure 4.22(c-d) other resonance modes are shown. Typical quality factors are between one and 30.



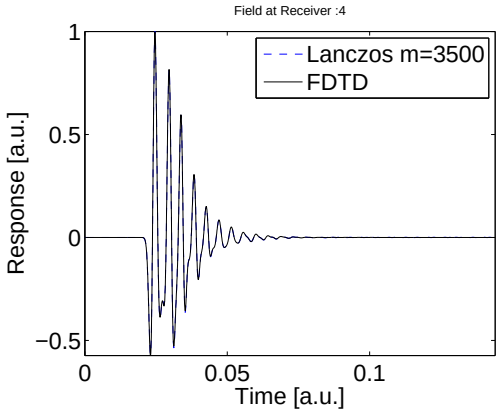
(a) Response at receiver 1 with m=3500.



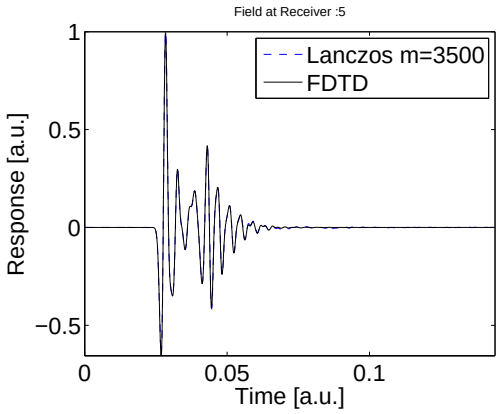
(b) Response at receiver 2 with m=3500.



(c) Response at receiver 3 with m=3500.

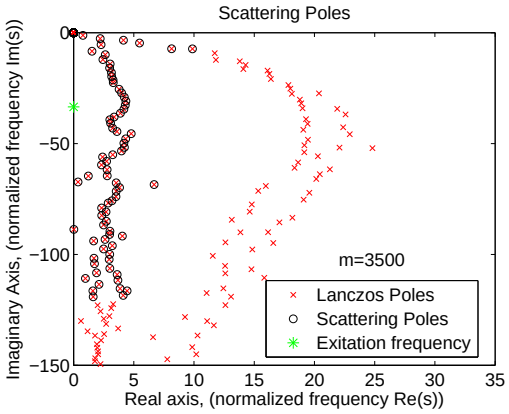


(d) Response at receiver 4 with m=3500.

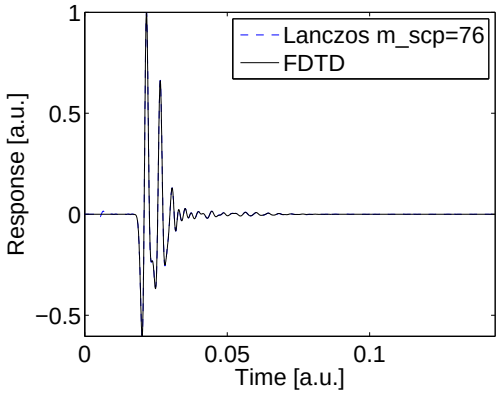


(e) Response at receiver 5 with m=3500.

Figure 4.19: Signal at the five receivers. Receiver 1 is the top receiver, receiver 5 the lower one.

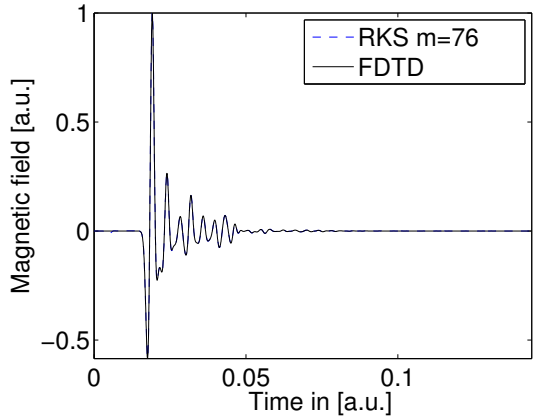


(a) Reduced system and expansion poles.

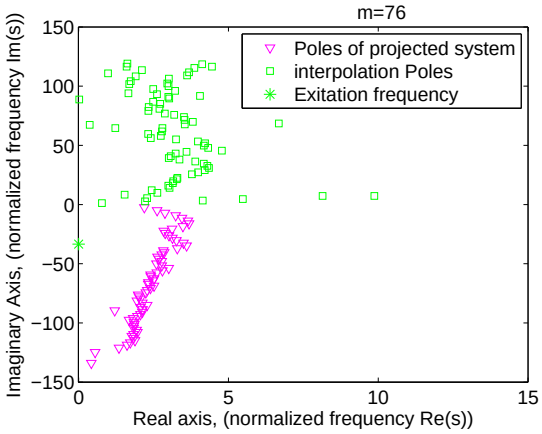


(b) Field at receiver 3 expanded in the scattering poles.

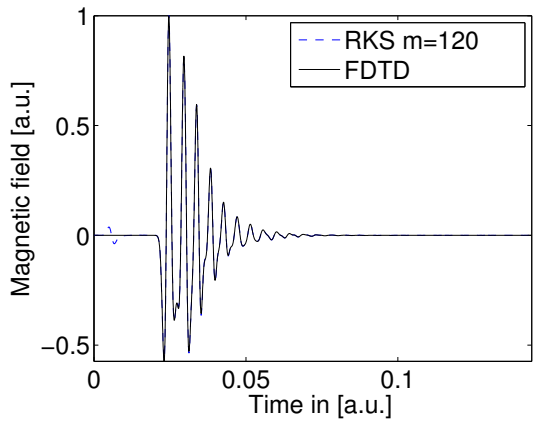
Figure 4.20: Scattering pole expansion of the layered Earth problem.



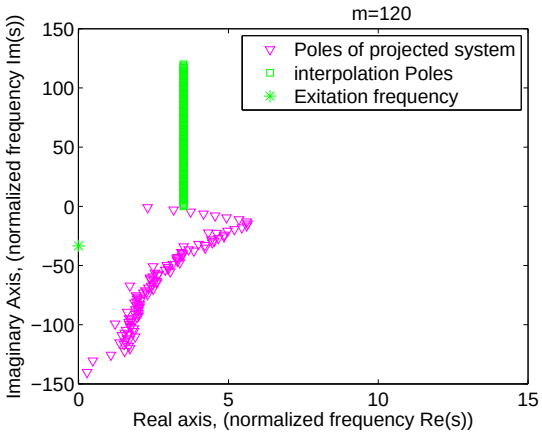
(a) System response for a RKS dimension of m=76 with spectrally adapted interpolation points.



(b) Interpolation poles and system poles for m=76 with spectrally adapted interpolation points.

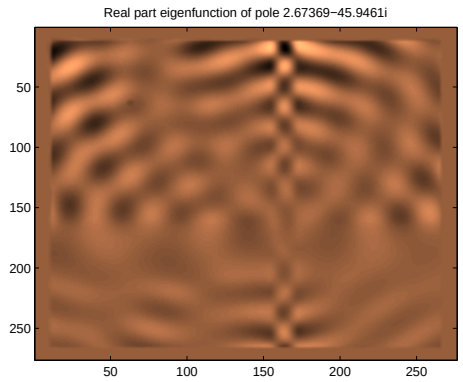


(c) System response for a RKS dimension of m=120 with equidistant interpolation points.

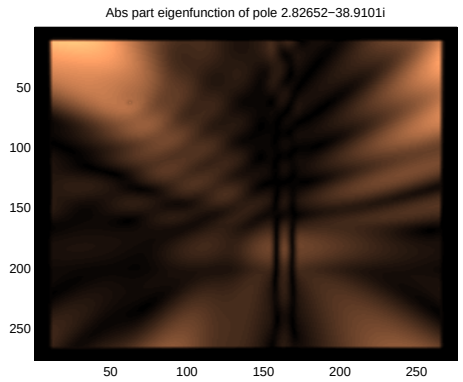


(d) Interpolation poles and system poles for m=120 with equidistant interpolation points.

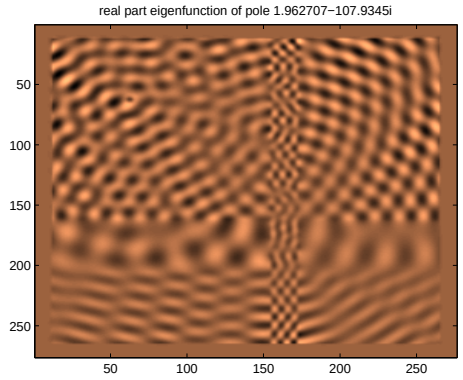
Figure 4.21: Rational Krylov subspace approximation of the layered Earth response using equidistance and spectrally adapted shifts.



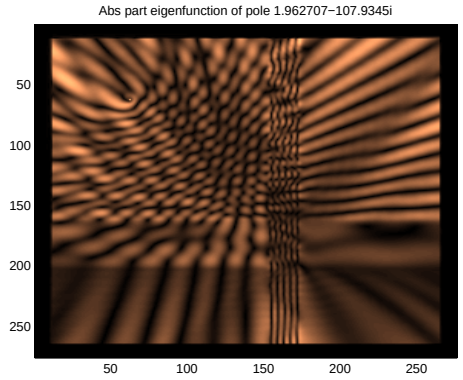
(a) Dominant mode at the receivers 1,2,3,5. Position in top 10 poles:(1,1,5,X,3)



(b) Dominant mode at the receiver 4. Position in top 10 poles:(X,X,X,1,9)



(c) Real part of resonance in drilling hole.



(d) Absolute part of resonance in drilling hole.

Figure 4.22: Eigenfunctions of several poles of the layered Earth problem.

4.2.4 Conclusion

The two-dimensional experiments show that the transient system response converges towards the corresponding FDTD results. The resonance expansions showed that the solution to a scattering problem can be expanded in a small number of scattering poles. In the example of the box with H-polarized fields, 26 poles were sufficient to approximate the time-domain solution.

Using a rational Krylov subspace approach, it was shown that convergence can be reached within very small subspaces compared with the problem size. The small subspace dimension makes it feasible to store all basis vectors in memory, which allows for the computation of modes even for large problems. The results showed that the convergence is strongly dependent on the shift selection, as spectral adaption of the shifts towards the problem enhanced the convergence for the photonic crystal waveguide and layered Earth example.

4.3 Instantaneously reacting materials – A three-dimensional configuration

4.3.1 Dielectric box

In this section a three-dimensional example is analyzed. This case is studied in order to show that the method can handle large systems and significantly reduce their order.

The investigated configuration is a three-dimensional electromagnetic wave problem. A dielectric box with a relative permittivity $\varepsilon_r = 4$ and width and height of 50 μm is simulated. As a source, an x-directed dielectric dipole is used. The time derivative of a Gaussian pulse, with a maximum of the spectrum at $\lambda_{peak} = 94$ μm was used as excitation wavelet. The configuration is shown in Figure 4.23.

The domain is truncated using the optimal PML and the problem size is $n \approx 8.4 \cdot 10^6$. In this section only the approximated resonance fields are presented, as it has already been shown several times throughout the document that the method converges.

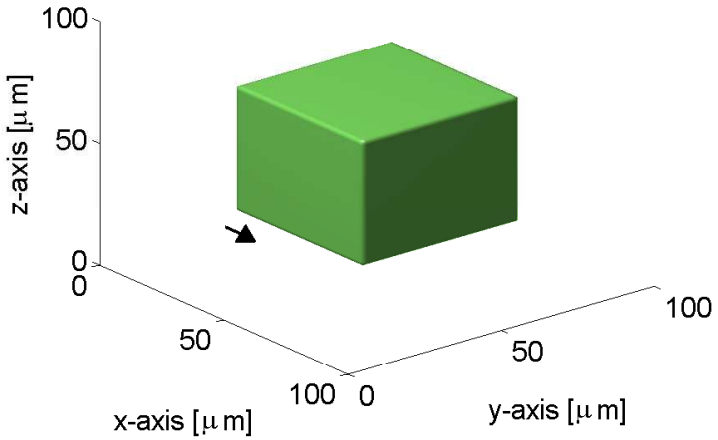
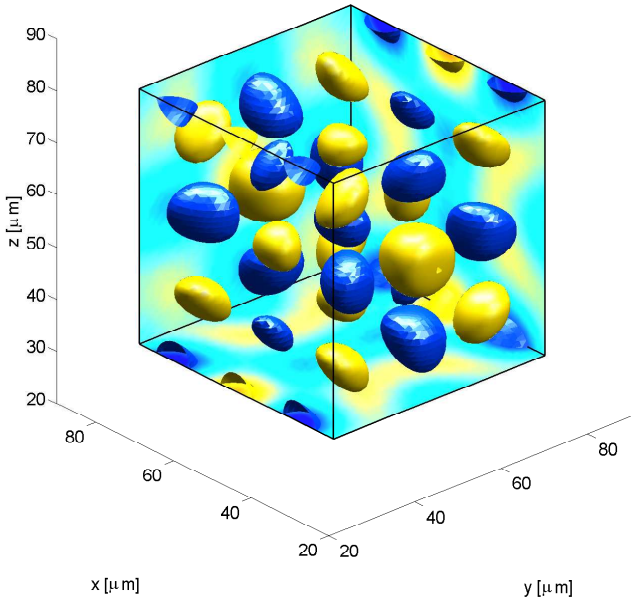


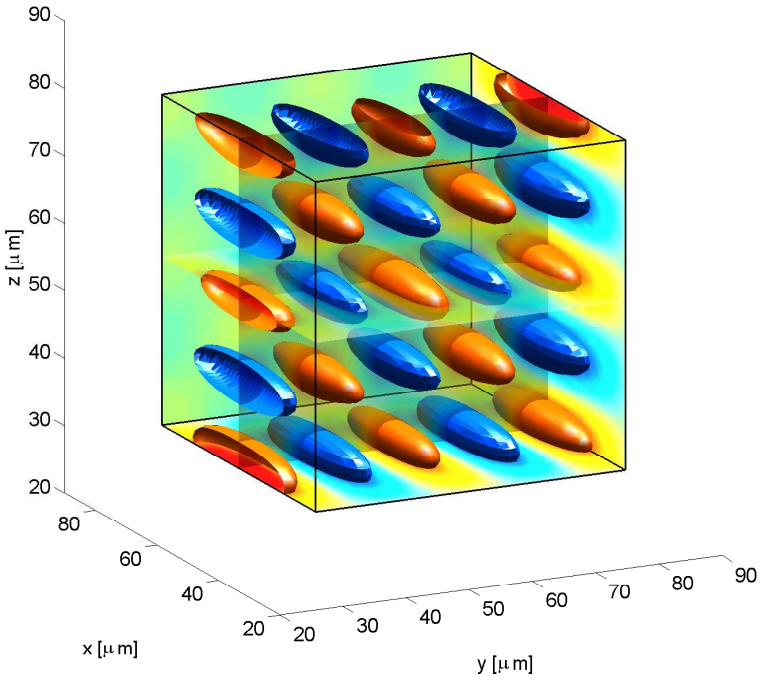
Figure 4.23: Three-dimensional dielectric box with dipole source next to the box.

Two of the approximated resonance modes are given in Figure 4.24 computed using m=10000 Lanczos iterations. The eigenmodes found by the Lanczos approximation are dependent on source location, as it is used as a seed vector in the Lanczos algorithm. Only resonances which are excited and thus contribute to the wavefield are calculated.

The resonance mode given in Figure 4.24 has a resonance wavelength of $2\pi c/\omega_1^{res} = 32.6 + 0.27\text{i}$ μm. This resonance field is the E_x field which is the dominant field quantity inside the box. The surface plots correspond to the 60% maximum of the resonance field. The resonance shows a 4π resonance field in the y- and z-direction, but only a fundamental resonance in the x-direction, which is the direction of the excitation dipole. Finally, the E_x field of the second pole $2\pi c/\omega_2^{res} = 38.3 + 0.47\text{i}$ μm shows a 3π resonance in all spatial directions.



(a) Imaginary part of the scattering resonance for E_x corresponding to the pole $2\pi c/\omega = 38.3 + 0.47\mathrm{i}\ \mu\mathrm{m}$



(b) Imaginary part of the scattering resonance for E_x corresponding to the pole $2\pi c/\omega^{\mathrm{res}} = 32.6 + 0.27\mathrm{i}\ \mu\mathrm{m}$

Figure 4.24: Resonant field in a three-dimensional box.

4.4 Media exhibiting relaxation – a two-dimensional configuration

4.4.1 Dispersive box with Drude relaxation

The simulated configuration is shown in Figure 4.25. A two-dimensional H-polarized configuration was chosen. The dielectric function of the box is given by the Drude model using the model parameters of gold as given in [30] ($\omega_p = 13.6$ PHz, $\gamma_p = 0.1$ PHz, $\varepsilon_\infty = 1$). As a source wavelet a Ricker Wavelet with a maximum in the spectrum at $\lambda_{peak} = 350$ nm is used. At this frequency the relative permitivity is $\varepsilon_r(\lambda_{peak}) \approx -40$. The configuration measures 100 nm$\times$100 nm and the box has dimensions 50 nm$\times$50 nm. A uniform grid with a step size of $\delta = 1$ nm is used. The problem size is $n = 47000$. If we compare this with the corresponding nondispersive case, we observe that the total number of unknowns increased by approximately 10000. As a reference simulation, an ADE-FDTD scheme as described in [31] is used.

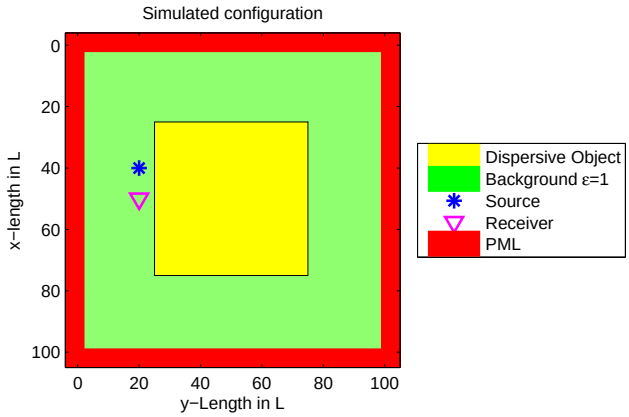


Figure 4.25: Simulated configuration: Square nanorod excited by a line source with L=1 nm.

Polynomial Krylov MOR

The convergence of the Lanczos reduced-order model towards the ADE-FDTD result is given in Figure 4.26(a). It can be seen that the model converged after 6500 Lanczos iterations on the interval of interest. Furthermore, it is shown that the late-time behavior can be expanded in 16 scattering poles with high quality factor.

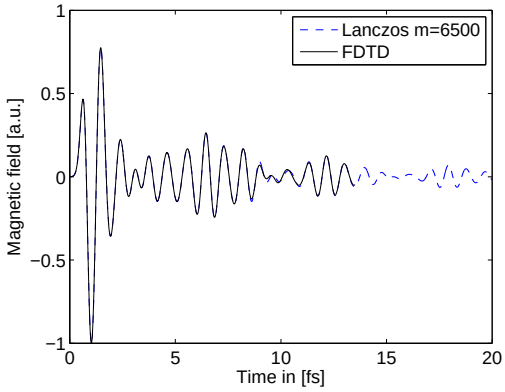
For the PKS approach to converge it needs a subspace dimension which is more than four times bigger than the non-dispersive one. The small increase in the problem size from $n = 37$k to $n = 47$k cannot explain this fourfold increase in dimensionality. The location of the main contributing scattering poles in Figure 4.26(c) shows, however, that they are small in modulus compared with the wide range of the other poles. Most other modes have a relatively large real part, such that they describe damped modes. All contributing poles have a high quality factor and are clustered at the imaginary axis. In the case of the dielectric box, the main contributing poles did not cluster at one point in the spectrum, but were spread out in a region. Therefore, the PKS can approximate distinct poles in the dielectric case with low dimensional subspaces, but it needs a large dimension to obtain distinct modes for the dispersive case. PKS methods converge from the pole of largest modulus towards the pole of smallest modulus. Thus the increase in subspace dimension needed for convergence can be linked to the highly resonating structure, with modes of small modulus clustered at the imaginary axis. In the last sections

it was shown that the convergence of RKS is superior to PKS for systems described by a few resonances. Therefore an RKS approach is used to obtain a faster and low-dimensional field approximation.

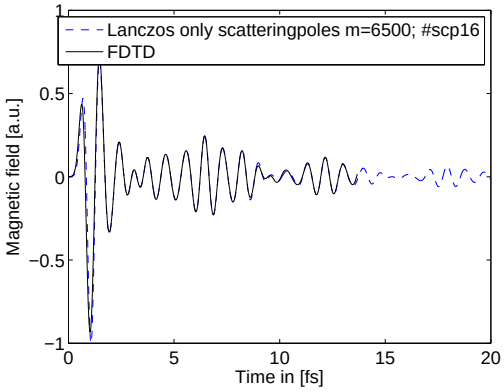
Rational Krylov MOR

The whole system can also be projected onto a rational Krylov subspace. The poles describing resonances with high quality factors are of interest, such that purely imaginary shifts are used. As interpolation points, chosen with no *a priori* information about the system, equidistant pure imaginary shifts between $\sigma_{\min} = 5\mathrm{i}$ and $\sigma_{\max} = 70\mathrm{i}$ were used. Using a subspace size of m=50, convergence of the time-domain result is obtained. The converged response and poles of the system are given in Figure 4.27. In Figure 4.28 the resonance fields of some of the poles are shown and it can be seen that the electric field in the x-direction indeed satisfies the boundary conditions which have to hold at the surface of the box.

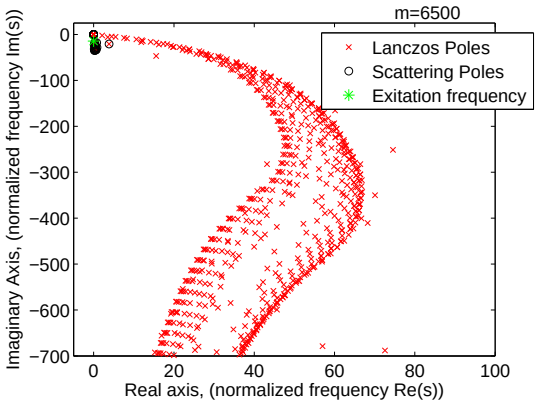
The wave field of the resonances is localized at the surface of the dispersive object, whereas standing waves inside the object were found in the non-dispersive case. For negative index materials, surface plasmons can be excited forming an oscillation between the free-electron gas of the metal and the electric field of the wave. The resonance field shows typical plasmon behavior, as it is localized at the surface and decays exponentially away from the surface of the object. Typically, the quality factors of the resonances range between $Q = 100$ and $Q = 300$.



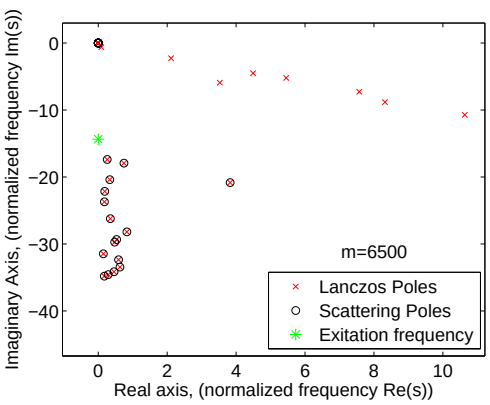
(a) Time-domain response at the receiver of the Lanczos reduced-order model of dimension $m = 6500$ and ADE-FDTD comparison simulation.



(b) Expension of the scattering response in the scattering poles of interest as shown in (d).

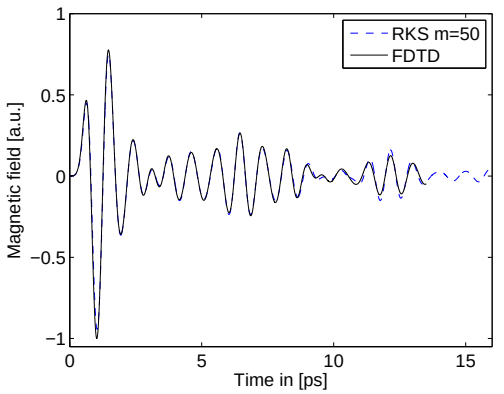


(c) Poles of the reduced-order model.

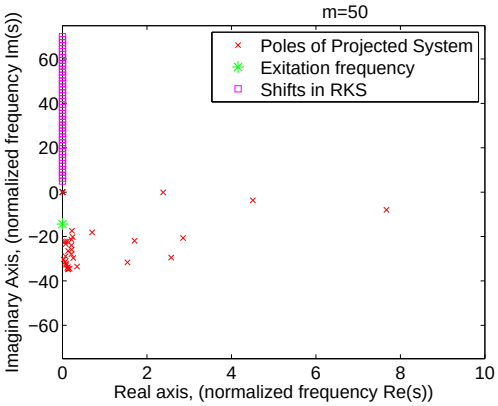


(d) Poles of the reduced-order model with high quality factor in the bandwidth of the pulse.

Figure 4.26: Scattering response using polynomial Krylov subspace reduction.

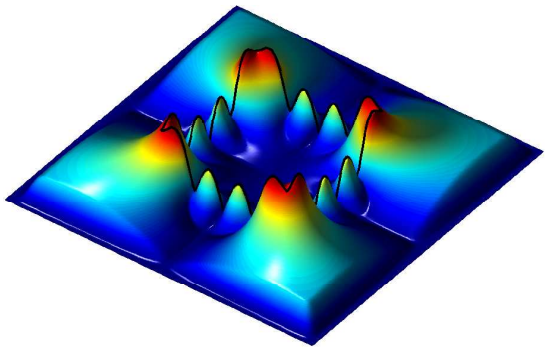


(a) Time-domain response at the receiver of the RKS model of dimension $m = 50$ and ADE-FDTD comparison simulation.

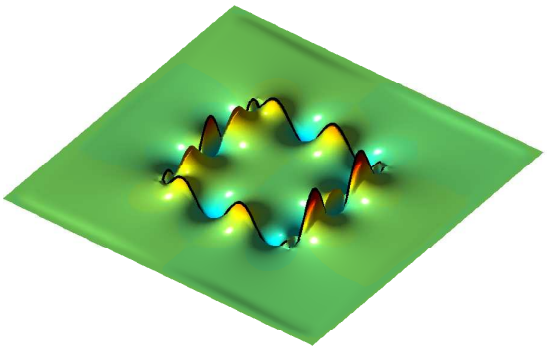


(b) Approximated poles of the RKS reduction and interpolation poles.

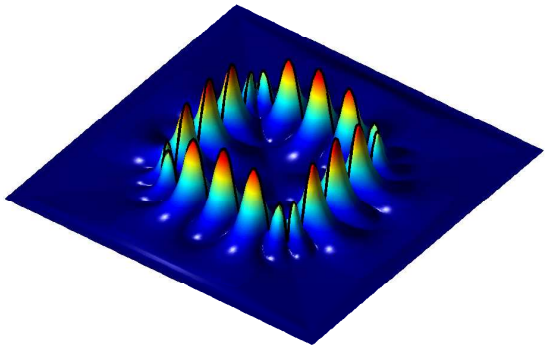
Figure 4.27: Results of the RKS reduction for the two-dimensional dispersive box.



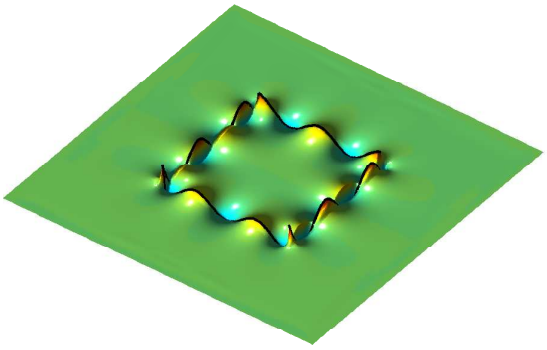
(a) Absolute H-field of pole $\lambda_{res} = 166 + 0.23i$ nm.



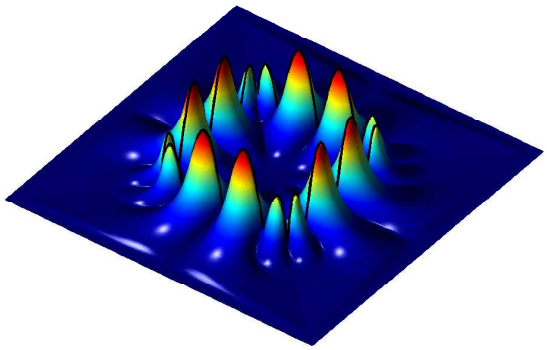
(b) Real E_x-field of pole $\lambda_{res} = 166 + 0.23i$ nm.



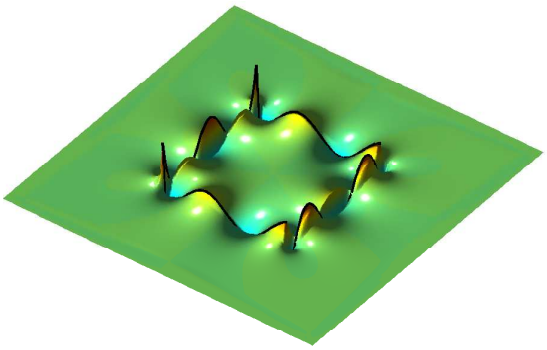
(c) Absolute H-field of pole $\lambda_{res} = 166 + 0.73i$ nm.



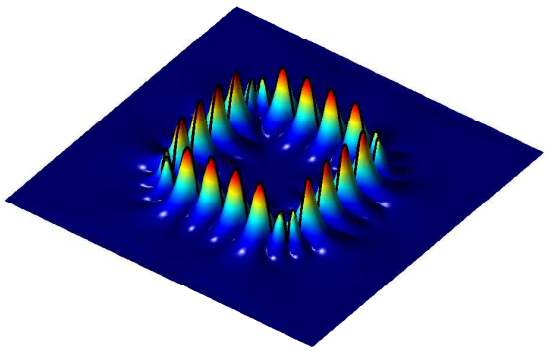
(d) Real E_x-field of pole $\lambda_{res} = 166 + 0.73i$ nm.



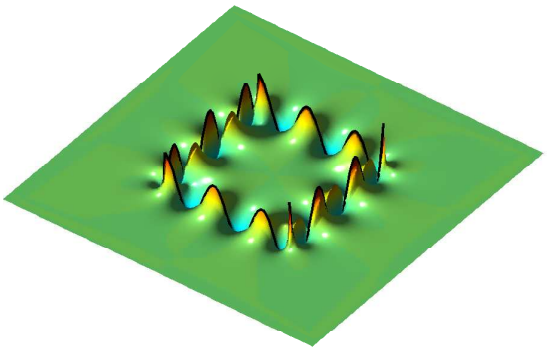
(e) Absolute H-field of pole $\lambda_{res} = 168 + 0.75i$ nm.



(f) Real E_x-field of pole $\lambda_{res} = 168 + 0.75i$ nm.

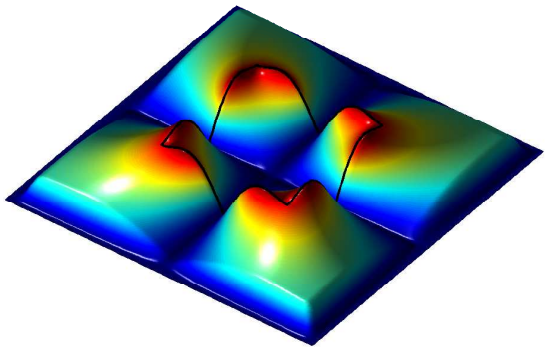


(g) Absolute H-field of pole $\lambda_{res} = 163 + 0.70i$ nm.

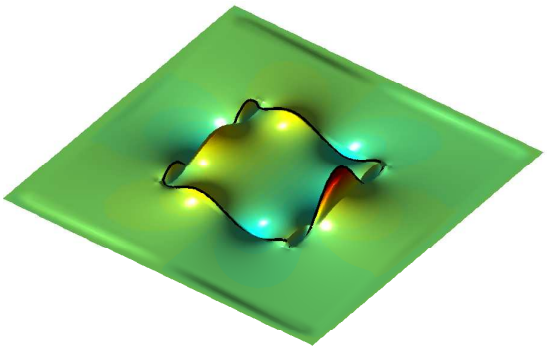


(h) Real E_x-field of pole $\lambda_{res} = 163 + 0.70i$ nm.

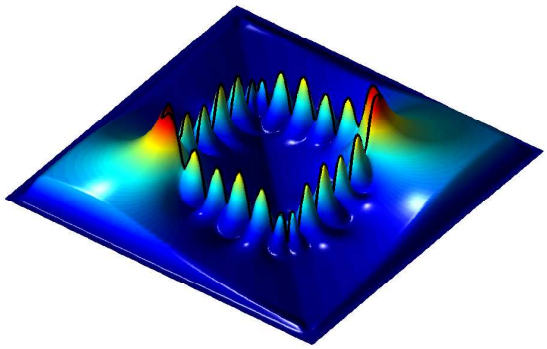
Figure 4.28: Eigenvectors of the system matrix A. Absolute value of the magnetic field is depicted on the left hand side, whereas E_x is depicted to the right.



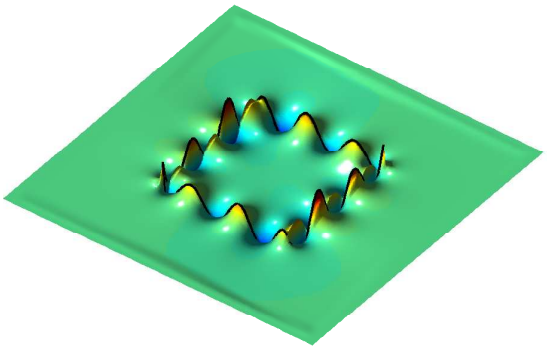
(a) Absolute H-field of pole $\lambda_{res} = 174 + 1.14i$ nm.



(b) Real E_x-field of pole $\lambda_{res} = 174 + 1.14i$ nm.



(c) Absolute H-field of pole $\lambda_{res} = 162 + 0.5i$ nm.



(d) Real E_x-field of pole $\lambda_{res} = 162 + 0.5i$ nm.

Figure 4.29: Continued from Figure 4.28.

4.5 Media exhibiting relaxation – A three-dimensional configuration

4.5.1 Spontaneous decay rate of a dipole near a nanorod

The last example targets to calculate the electromagnetic field of a quantum emitter in close proximity to a plasmonic resonator. This electromagnetic field is needed to calculate the enhancement of the spontaneous decay rate of a quantum emitter by the resonator. This enhancement is known as the Purcell effect.

The spontaneous decay rate of a two-level quantum system located at $\mathbf{x} = \mathbf{x}_0$ can be computed classically according to the formula [30]

$$\Gamma = \frac{2}{\hbar} \mathrm{Im}\left[\overline{\mathbf{p}} \cdot \hat{\mathbf{E}}(\mathbf{x}_0)\right], \tag{4.1}$$

where $\hat{\mathbf{E}}$ is the electric field at the location of the electrical dipole $\hat{\mathbf{J}}^{\mathrm{ext}} = -\mathrm{i}\omega\mathbf{p}\delta(\mathbf{x} - \mathbf{x}_0)$ with dipole moment $\mathbf{p}$. Using our Krylov reduction approach, the spontaneous decay rate can be computed using the formula

$$\overline{\mathbf{p}} \cdot \hat{\mathbf{E}}(\mathbf{x}_0) \approx \mathrm{i}\omega\gamma \mathsf{e}_1^T \left[r(\mathsf{H}_m) + r(\bar{\mathsf{H}}_m)\right] \mathsf{e}_1, \tag{4.2}$$

where the constant γ and the function $r(z)$ are given by

$$\gamma = \langle \mathsf{M}^{-1}\mathsf{q}, \mathsf{M}^{-1}\mathsf{q} \rangle_{\mathrm{disp}} \quad \text{and} \quad r(z) = \frac{\eta(z)}{z - \mathrm{i}\omega}. \tag{4.3}$$

In other words, a *single* model provides decay rate approximations for all frequencies (wavelengths). To illustrate the effectiveness and impact of the approach, we consider a configuration close to the one simulated in [1]. In this paper the authors simulate the spontaneous decay rate of a cylindrical golden nanorod. Equivalent to the cited work, a Drude medium is chosen with medium parameters $\omega_{\mathrm{p}} = 12.6$ PHz, $\gamma_{\mathrm{p}} = 0.141$ PHz and $\varepsilon_\infty = 1$ embedded in a background medium with a relative permittivity $\varepsilon_{\mathrm{r}} = 1.5^2$. After discretization the problem size is about $n = 17 \cdot 10^6$.

In [1], the cylindrical golden nanorod is simulated using an aperiodic Fourier modal based method (a.k.a. rigorous coupled-wave analysis (RCWA)). A version developed to evaluate axial symmetric resonators is used, as described in [32]. By using the axial symmetry of the structure, analytic expressions for the modes in the radial direction can be found, such that the modes need to be calculated on an effectively two-dimensional grid. According to the paper, a single mode calculation takes about 15 minutes on a modern personal computer, with no indication of the problem size.

In the method used in this thesis, all excited modes are approximated at the same time. Moreover, for this example a full three-dimensional simulation was used, whereas the cited work effectively solves two-dimensional systems. Obviously, the axial symmetry constraint of the method in [1] limits the class of objects that can be simulated. Finally, our method is able to transform the solution directly to the time-domain, whereas the aperiodic Fourier model method is a pure frequency-domain method.

Building a Krylov subspace of $m = 3500$ via the developed Lanczos algorithm takes about 1.3 hours. However, once the reduced system is obtained, the Fourier domain solution at the source for 1000 different wavelengths between 0.7 μm and 1.2 μm is obtained in 0.7 seconds by a direct sparse solver. The cited work in [1] needs 15 minutes for a single frequency point. Therefore, our method is superior after evaluating only five frequency points even though our approach solves a full three-dimensional problem as oppose to the two-dimensional ones solved in [1].

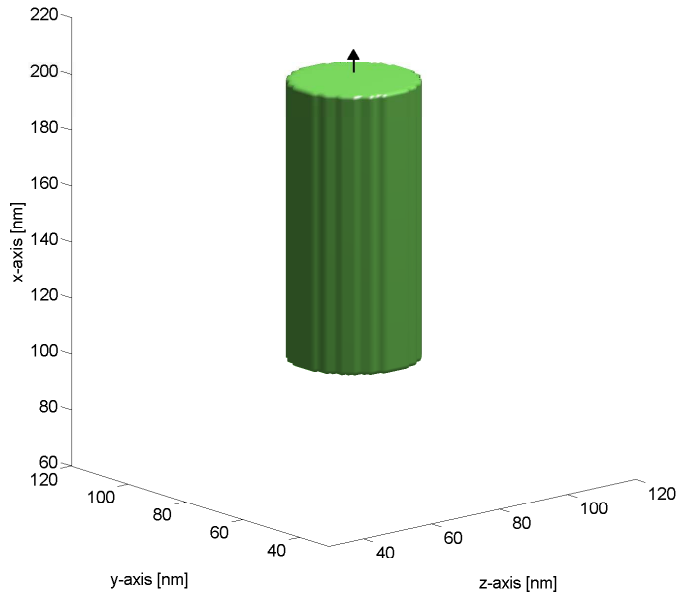


Figure 4.30: Simulated configuration: cylindrical nanorod excited by a dipole molecule.

Within 48 hours, the full order system could neither be solved directly or by the iterative solver GMRES on the used personal computer, because of memory limitations. Five cores of the Server HPC16 could not solve a single frequency point using restarted GMRES. It did not converge to the set tolerance of 10^{-6} after 80 outer and 20 inner iterations, which took 4.3 hours. At that point the residual had a value of $2 \cdot 10^{-4}$. This directly shows the impact of the developed PKS Lanczos algorithm for large system matrices.

In Figure 4.31 the spontaneous decay rate of the simulated cylindrical dipole configuration and the cylinder configuration of [1] are compared. The peak shapes look very similar, and the maximum in the decay rate is reached at the same wavelength.

4.6 Conclusion

4.6.1 Performance comparison

In this section the computation times of the numerical examples and different approaches are compared. The computation times[1] for all two-dimensional examples are provided in Table 4.1. The compared methods are FDTD, RKS and PKS. FDTD is only included in the comparison, to show that speed-ups can be achieved even for small time intervals. Furthermore, it shows that the Lanczos decomposition of a large matrix is faster than FDTD, for the same number of iterations, which is shown by the computation time of the photonic crystal. No such speed-up can be seen for the two-dimensional dielectric box example as the system matrix is too small. In this case the computational cost of the Lanczos algorithm is not dominated by taking matrix vector product computations.

The computation time of the PKS approach is split up into Lanczos decomposition and total computation time. This choice is made as a large fraction of the computation time is spend on solving the eigenvalue problem of the reduced-order system, without making use of

[1] Matlab implementation on an Intel i5-3470 CPU @ 3.2 GHz 8GB 64-bit Windows 7 architecture.

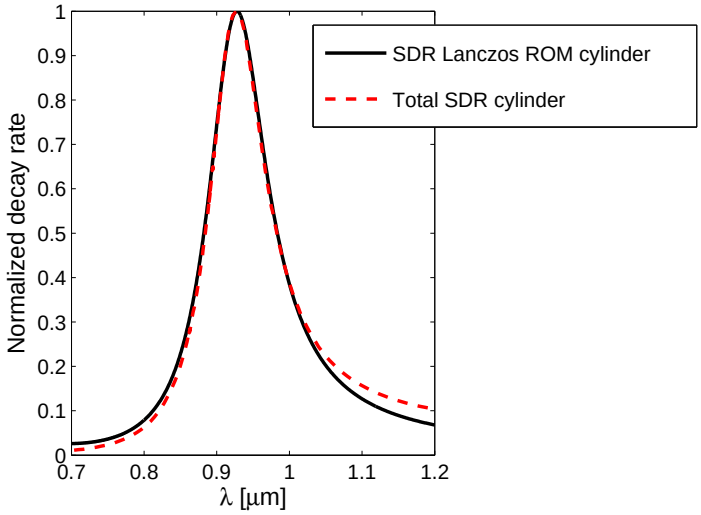


Figure 4.31: Spontaneous decay rate of the nanorod dipole configuration using the Lanczos algorithm and the method described in [1]. The decay rate of the cylinder configuration was calculated using the Lanczos reduced-order model with m=3500.

Table 4.1: Computation time comparison of FDTD, PKS, and RKS.

	Box $n = 37$k		Crystal $n = 695$k		layered Earth $n = 545$k		Drude box $n = 47.4$k	
	Iterations	Time [s]	Iterations	Time [s]	Iterations	Time [s]	Iterations	Time [s]
FDTD	10000	16	16000	3228	3000	245	8200†	36†
PKS Lanczos	1500	6.2	3500	99	3500	77.5	6500	28
PKS total	1500	17.8	3500	217	3500	195	6500	754
RSK	48	7.7	15	68	76	275	50	12

† ADE-FDTD schemes were used to compute the result in a medium exhibiting relaxation

its symmetry and tridiagonal structure. The eigenvalues of a symmetric tridiagonal structure can be efficiently calculated via recursion relations, which can speed up the PKS algorithm. Therefore significant speed up can be achieved by developing a dedicated eigenpair solver for this problem.

The computation times in Table 4.1 show that the RKS approach outperforms FDTD and PKS for the dielectric box and photonic crystal example. No speed-up can be seen for the layered Earth example. However, for all structures the model order is significantly reduced using the RKS approach with respect to the PKS approach. This makes the RKS attractive for inversion schemes, as the transfer function between two points can be obtained in terms of a small number of modes. The Earth layer and drilling hole structure extends into the PML, such that the resonances found in the structure are rather lossy and of low quality factor. A structure with a response dominated by a few scattering poles can be approximated by a lower order model and thus gain from an RKS approach.

The system response of the example exhibiting relaxation is dominated by a few surface modes with high quality factor. All contributing modes can be approximated by an RKS of dimension $m = 50$, which shows superior convergence over PKS. Given the fact that the system response is described by a few resonances with high quality factor, the RKS approach outperforms PKS. The used example has a relatively small problem size, such that the systems inside the RKS can be solved fast using sparse solvers. For this example, the comparison with ADE-FDTD is only included to give a rough indication how the speed of the Lanczos algorithm compares

to the speed of ADE-FDTD. The quality factor of the plasmonic resonance is high, and even 8200 ADE-FDTD steps are not enough to cover the complete response of the pulse as shown in Figure 4.26. The computation time of the reduced-order model could be reduced significantly by using an RKS approach. It outperformed the ADE-FDTD method after 2800 time steps.

The impact of the standard eigenvalue solver which was used in all experiments can clearly be seen in this example. As the PKS approach needs a high dimensional subspace the eigenvalue problem takes about 96% of the computation time. A dedicated eigenvalue solver can yield significant speed up especially for high dimensional subspaces. For large matrices one could also use a PKS approach to reduce the size of the system matrix. After that an RKS approach can be used on the reduced system matrix, such that neither large eigenvalue problems nor very large sparse systems need to be solved.

The two-dimensional dispersive box problem leads to a system matrix of small dimension such that the PKS runtime is not dominated by matrix vector products, which normally makes it superior for large systems. The impact of the developed Lanczos algorithm for dispersive media is of increasing impact for larger system matrices. The three-dimensional dispersive example for instance could only be solved with the PKS approach, solving single frequency systems with a direct sparse solver appeared to be infeasible. The developed Lanczos algorithm allows decomposition of a system matrix with a computation time proportional to $m \cdot n$, rather than $m^2 \cdot n$ in an Arnoldi algorithm. Furthermore, the memory used by the Lanczos algorithm is reduced by a factor $3/m$ compared with the Arnoldi algorithm.

4.6.2 Summary of Results

In this chapter, a novel model order reduction approach for dispersive media was presented. Polynomial and rational Krylov subspace approaches together with resonance expansions were used to model wave propagation. Convergence in the time- and frequency-domain was demonstrated. The time-domain results of the reduced-order models match the ones produced by FDTD. Moreover, the spontaneous decay rate of a quantum emitter close to a three-dimensional object has been calculated, showing convergence in the frequency-domain.

In conclusion, the RKS approach shows a better performance on highly resonating structures with only a few excited modes. PKS shows superior performance, when a large number of modes contribute or the system matrix is too large to efficiently obtain single frequency solutions for RKS projections.

The advantage of Krylov subspace methods over FDTD increases with increasing size of the time interval of interest, as the Krylov method converges on the complete causal time-domain and FDTD only on a bounded time interval. Especially resonating structures with high quality factors need long simulation times, which motivates the choice of Krylov methods for micro cavities, photonic crystals, and for plasmonic configurations.

The RKS convergence is heavily dependent on shift selection. In this chapter, domains of interest were used to adapt the spectrum of the reduced-order model to an interval of interest. However, no optimal shift selection was used. To obtain optimal rational Krylov subspace reduced-order models the process of shift selection should be based on a mathematical formulation of optimality.

For use in inversion schemes, H2 optimal models have desirable properties. The error of the reduced-order transfer function lies in the null space of the approximate Jacobians, which makes H2-optimal models insensitive to approximation errors. Finding constraints for H2 optimal pole selection is therefore a desirable goal for future work.

Chapter 5

Conclusions

5.1 Summary of the Results

In this Master thesis we investigated the use of several Krylov subspace methods for the modeling of time- and frequency-domain wave propagation on open domains in instantaneously reacting and dispersive media. The basic wave equations were discretized on a staggered finite-difference grid and the extension of the domain to infinity was simulated via an optimized complex-scaling method (PML). The resulting state-space representation has a large order which can easily run into the millions. To reduce this order, we exploited the sparsity and symmetry properties of the system matrix in Lanczos-type and Arnoldi algorithms and projected the state-space representation onto several different Krylov subspaces. Approximate field solutions were obtained by exactly solving the projected Krylov system and incorporating stability correction.
In Chapter 2, the discrete symmetry preserving operator formulation of the Maxwell equations was introduced. A novel symmetry preserving formulation for media exhibiting general second-order relaxation was developed. Furthermore, resonance expansions were introduced as an efficient tool to obtain low-order models and transfer functions.
In Chapter 3, model-order reduction via projection onto subspaces was discussed. To obtain the field approximations in the time- and frequency-domain, polynomial and rational Krylov subspaces were used and the efficient construction of their bases was also described briefly. We also presented a short discussion on how the different algorithms approximate the spectrum of the system matrix by making a comparison with the power method and its variants.
In Chapter 4, we illustrated the theory developed in the previous chapters by several numerical experiments. Polynomial and rational Krylov subspace approaches were applied to wave propagation problems involving instantaneously reacting and dispersive media in one-, two- and three-dimensions. For simple one-dimensional systems, it was shown that the resonances obtained via model-order reduction match the analytical resonances of the configuration. Our method was further validated in 1D, 2D, and 3D by comparing the Krylov reduction responses with field responses computed by FDTD. In all cases, our reduced-order models showed good agreement with FDTD results. We also showed that with our reduced-order modeling approach it is possible to determine the bandgap structures of (quasi-)periodic arrays. Furthermore, we showed that it is possible to determine the dominant resonant modes of an open system and we illustrated these modes for a wide range of applications again in 1D, 2D, and 3D. In addition, it was demonstrated that the full time-domain response of a photonic crystal can be determined using a single resonance only. Finally, the newly developed symmetry preserving formulation of the general second-order dispersive Maxwell system allowed for model-order reduction of dispersive media via the Lanczos algorithm. Convergence in the frequency- and time-domain was shown by comparing computed field responses with ADE-FDTD field responses. The im-

pact of this new formulation was illustrated by a full vectorial three-dimensional example with approximately 17 million unknowns.
Our experiments clearly revealed some of the advantages and disadvantages of PKS and RKS. It was shown that RKS outperforms PKS for structures dominated by a few open resonance modes. Next we found that the convergence of RKS strongly depends on the choice of interpolation points. Spectral adaption of the interpolation points to the spectrum of the system matrix was shown to be desirable, which leads to the need of a mathematical formulation of optimal shift selection. Furthermore, it was shown that the semi-discrete nature of our approach makes it superior to conventional methods like the FDTD method, since a Krylov method constructs an interpolatory polynomial or rational function that provides us with field approximations for *all* causal times, whereas FDTD is a time stepping method that needs to compute the field response step by step, starting at the time instant when the source was switched on. Moreover, the transfer function between a given source and a given receiver is explicitly calculated in a Krylov method and in the frequency-domain the Krylov subspace approach outperforms conventional methods as soon as the solution needs to be evaluated at more than one frequency.

At this point we want to put the presented work into perspective. Figure 5.1 gives an overview of Krylov subspace approaches developed over the years for the full-wave Maxwell system and for diffusive electromagnetic fields on *unbounded* domains. In this thesis we used the PKS approach for wave equations with no relaxation and extended it to electromagnetic wave problems involving media exhibiting relaxation. We further started to explore the use of RKS approaches previously used for diffusion equations in case of instantaneously reacting materials and media exhibiting relaxation.

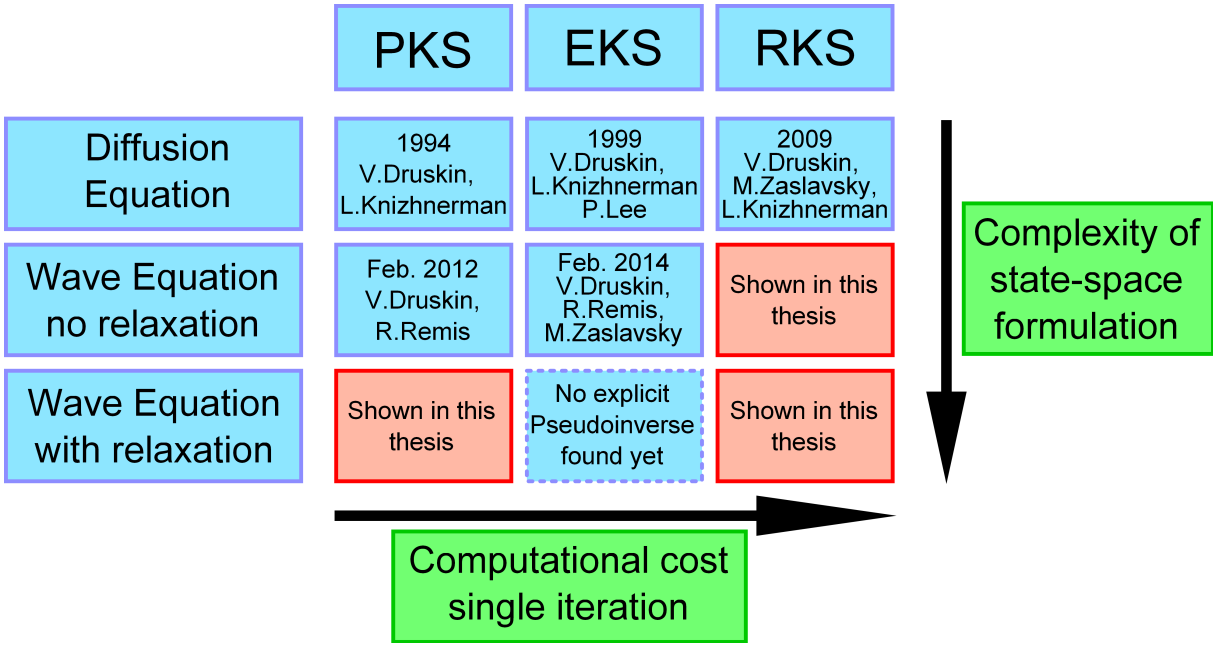


Figure 5.1: Overview of the Krylov subspace approaches discussed in the literature and in this thesis.

5.2 Future work

More research is needed in the field of rational Krylov subspace model-order reduction. In this work we only showed the potential of this approach. A significant reduction in computation times was achieved, especially when the interpolation points (shifts) are adapted to the spectrum

of the system matrix. However, a robust model-order reduction framework requires a mathematical basis for optimal shift selection. For example, H2 optimality constraints lead to rational reduced-order models satisfying the necessary conditions for the time-integrated squared-error to be minimum. Further speed-ups of the RKS method can be achieved by implementing effective preconditioners for Helmholtz-type systems. Another reasonable approach for Maxwell equations involving media exhibiting relaxation is the EKS approach. Recently the impact of the EKS approach was shown in [12]. For plasmonic structures this approach shows potential, as plasmonic structures are usually dominated by a few modes of small modulus clustered close to the imaginary axis.
In the field of optics, the easily measurable far-field is of main importance. Thus near-to-far field transformations can be implemented as documented in Chapter 8.2 of [27] in order to obtain the diffraction orders of a system as output. Moreover, future work should include the incorporation of Floquet quasi-periodic boundary conditions in order to make the method suited for structures like gratings and antenna arrays.
Finally, to illustrate the impact of the model-order reduction framework on full wavefield inversion, the developed framework should be incorporated into Gauss-Newton inversion schemes.

Bibliography

[1] C. Sauvan, J. P. Hugonin, I. S. Maksymov, and P. Lalanne, "Theory of the spontaneous optical emission of nanosize photonic and plasmon resonators," *Phys. Rev. Lett.*, vol. 110, p. 237401, Jun 2013. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.110.237401

[2] K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," *Antennas and Propagation, IEEE Transactions on*, vol. 14, no. 3, pp. 302–307, 1966.

[3] A. Antoulas, *Approximation of Large-scale Dynamical Systems*, ser. Advances in Design and Control. Society for Industrial and Applied Mathematics, 2005. [Online]. Available: http://books.google.nl/books?id=lYfRIJlA1mkC

[4] V. Druskin and R. Remis, "A Krylov stability-corrected coordinate-stretching method to simulate wave propagation in unbounded domains," *SIAM Journal on Scientific Computing*, vol. 35, no. 2, pp. B376–B400, 2013. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/12087356X

[5] N. Moiseyev, "Quantum theory of resonances: calculating energies, widths and cross-sections by complex scaling," *Physics Reports*, vol. 302, no. 5-6, pp. 212 – 293, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0370157398000027

[6] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *J. Comput. Phys.*, vol. 114, no. 2, pp. 185–200, Oct. 1994. [Online]. Available: http://dx.doi.org/10.1006/jcph.1994.1159

[7] V. Druskin, S. Güttel, and L. Knizhnerman, "Near-optimal perfectly matched layers for indefinite Helmholtz problems," *not yet published*, 2013. [Online]. Available: http://eprints.ma.man.ac.uk/2039/01/zol_helm.pdf

[8] V. Druskin and L. Knizhnerman, "Spectral approach to solving three-dimensional Maxwell's diffusion equations in the time and frequency domains," *Radio Science*, vol. 29, no. 4, pp. 937–953, 1994.

[9] V. Druskin, L. Knizhnerman, and P. Lee, "New spectral Lanczos decomposition method for induction modeling in arbitrary 3-D geometry," *GEOPHYSICS*, vol. 64, no. 3, pp. 701–706, 1999. [Online]. Available: http://library.seg.org/doi/abs/10.1190/1.1444579

[10] M. Zaslavsky and V. Druskin, "Solution of time-convolutionary Maxwell's equations using parameter-dependent Krylov subspace reduction," *Journal of Computational Physics*, vol. 229, no. 12, pp. 4831 – 4839, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021999110001294

[11] L. Knizhnerman, V. Druskin, and M. Zaslavsky, "On optimal convergence rate of the rational Krylov subspace reduction for electromagnetic problems in unbounded domains," *SIAM Journal on Numerical Analysis*, vol. 47, no. 2, pp. 953–971, 2009.

[12] V. Druskin, R. Remis, and M. Zaslavsky, "An Extended Krylov Subspace Model-Order Reduction Technique to Simulate Wave Propagation in Unbounded Domains," *ArXiv e-prints*, Feb. 2014.

[13] R. Remis, "Reduced-order modeling of transient elecromagnetic fields," Doctoral thesis, Technische Universiteid Delft, Delft University Press, Mekelweg 4, 2628 CD Delft, Netherlands, october 1998.

[14] S.-H. Tang and M. Zworski, "Resonance expansions of scattered waves," *Communications on Pure and Applied Mathematics*, vol. 53, no. 10, pp. 1305–1334, 2000. [Online]. Available: http://dx.doi.org/10.1002/1097-0312(200010)53:10<1305::AID-CPA4>3.0.CO;2-#

[15] Y. Saad, *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[16] C. Jagels and L. Reichel, "Recursion relations for the extended Krylov subspace method," *Linear Algebra and its Applications*, vol. 434, no. 7, pp. 1716 – 1732, 2011, special Issue: NIU. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0024379510004507

[17] C. Jagels and L. Reichel, "The extended Krylov subspace method and orthogonal Laurent polynomials," *Linear Algebra and its Applications*, vol. 431, no. 3 - 4, pp. 441 – 458, 2009, special Issue in honor of Henk van der Vorst. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002437950900130X

[18] C. G. Green, "Connections between Lanczos iteration and orthogonal polynomials," Master thesis, University of Washington, Pend Oreille Rd NE, Seattle, WA 98105, United States, 2001.

[19] V. Simoncini, "Extended Krylov subspace for parameter dependent systems," *Applied Numerical Mathematics*, vol. 60, no. 5, pp. 550 – 560, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168927410000206

[20] V. Druskin, C. Lieberman, and M. Zaslavsky, "On adaptive choice of shifts in rational Krylov subspace reduction of evolutionary problems," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2485–2496, Aug. 2010. [Online]. Available: http://dx.doi.org/10.1137/090774082

[21] V. Druskin, L. Knizhnerman, and M. Zaslavsky, "Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts," *SIAM J. Sci. Comput.*, vol. 31, no. 5, pp. 3760–3780, Oct. 2009. [Online]. Available: http://dx.doi.org/10.1137/080742403

[22] V. Druskin, V. Simoncini, and M. Zaslavsky, "Solution of the time-domain inverse resistivity problem in the model reduction framework part i. one-dimensional problem with siso data," *SIAM Journal on Scientific Computing*, vol. 35, no. 3, pp. A1621–A1640, 2013. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/110852607

[23] S. Gugercin, A. C. Antoulas, and C. Beattie, "H_2 model reduction for large-scale linear dynamical systems," *SIAM journal on matrix analysis and applications*, vol. 30, no. 2, pp. 609–638, 2008.

[24] Y. Saad, *Numerical methods for eigenvalue problems*. Manchester University Press, 1992.

[25] S. Hein, T. Hohage, and W. Koch, "On resonances in open systems," *Journal of Fluid Mechanics*, vol. 506, pp. 255–284, May 2004.

[26] R. Remis, "The FDTD method," *Lecture notes course: Computational electromagnetics part B*, 2005.

[27] A. Taflove and S. C. Hagness, "Computational electromagnetics: the finite-difference time-domain method," *Artech House,*, 2005.

[28] M. Verweij, P. Van den Berg, and H. Blok, *Electromagnetic Waves, An Introductory Course*, 1st ed. Leegwaterstraat 42, Delft: VSSD, 1999.

[29] I. A. Sukhoivanov and I. V. Guryev, *Photonic Crystals, Physics and Practical Modeling*. Springer Berlin Heidelberg, 2009.

[30] L. Novotny and B. Hecht, *Principles of nano-optics*. Cambridge university press, 2012.

[31] J. Zimmerling, "Surface Plasmon Propagation: Time Domain Simulation of Surface Plasmon Polariton Propagation," *Bachelor Honors Track report*, june 2012.

[32] A. Armaroli, A. Morand, P. Benech, G. Bellanca, and S. Trillo, "Three-dimensional analysis of cylindrical microresonators based on the aperiodic fourier modal method," *J. Opt. Soc. Am. A*, vol. 25, no. 3, pp. 667–675, Mar 2008. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-25-3-667

[33] S. Asvadurov, V. Druskin, M. N. Guddati, and L. Knizhnerman, "On optimal finite-difference approximation of PML," *SIAM Journal on Numerical Analysis*, vol. 41, no. 1, pp. pp. 287–305, 2004. [Online]. Available: http://www.jstor.org/stable/4101158

Appendix A

List of abbreviations

List of abbreviations

et al.	et alii (*lat.* and others)
etc.	et cetera (*lat.* and so forth)
a.k.a	also known as
PWE	Plane Wave Expansion
PKS	Polynomial Krylov Subspace
EKS	Extended Krylov Subspace
RKS	Rational Krylov Subspace
FDTD	Finite-Difference Time-Domain
FVM	Finite Volume Method
FEM	Finite Element Method
ROM	Reduced Order Model
PML	Perfectly Matched Layers
DE	Differential Equation
NtD	Neuman to Dirichlet
ADE	Auxiliary differential equations
PEC	Perfectly Electrically Conducting
ABC	Absorbing Boundary Conditions
EM	Electromagnetic
MOR	Model Order Reduction

Appendix B

Nomenclature

The Nomenclature used throughout the thesis is shown in table B.1.

Table B.1: Nomenclature used in this document.

Quantity	Notation
Scalar	a
Continuous Operator	$\mathcal{D}$
Continuous Field Quantity	$\mathbf{E}$
Kronecker Product	$\otimes$
Vector	e
Matrix	M
Lexicographical Ordering	$\mathsf{e} = vec(\mathsf{E})$
Identity Matrix (P by P)	I_P
Matrix Transpose	A^T
Hermitan Transpose	$\mathsf{A}^\dagger$
Truncated/Stabilized Matrix	$\tilde{\mathsf{A}}$
Laplace Transform	$\hat{F}(s) = \mathcal{L}\{f(t)\}$
Fourier Transform	$\breve{F}(s) = \mathcal{F}\{f(t)\}$
Hilbert Transform	$g(t) = \mathcal{H}\{f(t)\}$
heaviside step function	$\eta(t)$
delta function	$\delta(t)$
Euclidean norm	$\|\|\cdot\|\|$
Complex conjugate	$\bar{\cdot}$
bidiagonal Matrix of size N×N+1	$\text{bidiag}_N(\mathsf{a},\mathsf{b})/\text{bidiag}_N(a_i,b_i)$
tridiagonal Matrix	$\text{tridiag}(\mathsf{a},\mathsf{b},\mathsf{c})/\text{tridiag}(a_i,b_i,c_i)$
Upper Hessenberg matrix with elements h_{ij}	$\text{upperhess}(h_{ij})$
polinomial in $\cdot$ of order i	$p^i(\cdot)$
Rank two tensor	$\bar{\bar{\varepsilon}}$

Appendix C

Electromagnetic system formulation

In this appendix we present the discrete system formulation for one-dimensional problems (Section C.1), two-dimensional problems (Section C.2) and finally three-dimensional problems (Section C.3).

C.1 One-dimensional system formulation

In this section the algorithms used to simulate 1D EM-wave propagation are documented. In the Maxwell equation we choose to set all fields to zero except $\mathcal{E}_z$ and $\mathcal{H}_x$. Therefore the EM-wave propagates along the y-direction. The Maxwell equations in the Laplace domain now read

$$\partial_y \hat{\mathcal{E}}_z + \mu_r s \hat{\mathcal{H}}_x = -\hat{\mathcal{K}}^{\text{ext}}, \tag{C.1}$$
$$\partial_y \hat{\mathcal{H}}_x + \varepsilon_r s \hat{\mathcal{E}}_z + \sigma \hat{\mathcal{E}}_z = -\hat{\mathcal{J}}^{\text{ext}}, \tag{C.2}$$
$$\tag{C.3}$$

or in operator form

$$(\mathcal{D} + \mathcal{S} + \mathcal{M}s)\hat{\mathcal{F}} = \hat{\mathcal{Q}}'. \tag{C.4}$$

With $\mathcal{D}$ the spatial differentiation operator, $\mathcal{M}$ the medium matrices, $\hat{\mathcal{F}}$ containing the field quantities and $\hat{\mathcal{Q}}'$ the vector containing the electromagnetic sources.

C.1.1 Discretization

In order to simulate the system we introduce a finite difference grid. At both ends of the domain an optimal PML is used to simulate the extension of the system to infinity. Therefore we first define two regions in our domain being $\Omega_{y,\text{PML}}$ and $\Omega_{y,\text{DOI}}$. The latter denotes the domain of interest, whereas the first denotes the region that truncates the DOI with a PML. Let k denote the width of the PML and N the size of the domain of interest. We now define the primary grid as

$$\Omega^{\text{p}}_{y,\text{DOI}} = \{y_p, q = 0, 1, \ldots, N, y_q > y_{q-1}\}, \tag{C.5}$$

and the dual grid as

$$\Omega^{\text{d}}_{y,\text{DOI}} = \{\hat{y}_p, q = 1, \ldots, N, \hat{y}_q > \hat{y}_{q-1}\}. \tag{C.6}$$

The grid in the PML is given by

$$\Omega^{1,\text{p}}_{y,\text{PML}} = \{y_p, q = -k, \ldots, -1, y_q > y_{q-1}\}, \tag{C.7}$$

and

$$\Omega^{2,\text{p}}_{y,\text{PML}} = \{y_p, q = N+1, \ldots, N+k, y_q > y_{q-1}\}. \tag{C.8}$$

The dual grid inside the PML is given by

$$\Omega^{1,\mathrm{d}}_{y,\mathrm{PML}} = \{\hat{y}_p, q = -k,\ldots,N, \hat{y}_q > \hat{y}_{q-1}\}, \tag{C.9}$$

and

$$\Omega^{2,\mathrm{d}}_{y,PML} = \{\hat{y}_p, q = N+1,\ldots,N+k, \hat{y}_q > \hat{y}_{q-1}\}. \tag{C.10}$$

The step sizes in the DOI are given by

$$\delta_{y,p} = y_p - y_{p-1} \quad \text{and} \quad \hat{\delta}_{y,p} = \hat{y}_p - \hat{y}_{p-1}. \tag{C.11}$$

We choose a uniform grid with all step sizes to be equal to δ. The optimal step sizes inside the PML are calculated using the approach described in [7]. The step size $\hat{h}_1$ is given by $\frac{\delta}{2}$ plus the complex contribution given by the algorithm presented in [7], in order for the primary and the dual PML to coincide at both ends where the DOI is truncated by the PML.
The step sizes inside the PML are denoted

$$h_q \text{ and } \hat{h}_q \quad \forall q = 1,2,\ldots,k \tag{C.12}$$

and are chosen symmetric on both ends of the domain. The grid is illustrated in Figure C.1. A homogeneous Dirichlet boundary condition is imposed at the electric field approximation at the outer most grid point. This boundary condition is known as PEC boundary condition in EM theory, as a perfect electrically conducting contact does not allow any tangential electric field. The full discrete system can now be written as

$$(\mathsf{D}+\mathsf{S}+\mathsf{M}s)\hat{\mathsf{f}} = \hat{\mathsf{q}}', \tag{C.13}$$

with $\mathsf{f} = [e_z(y_{-(k-1)}),\ldots,e_z(y_{N+k-1}), h_x(\hat{y}_{-k}),\ldots,h_x(\hat{y}_{N+k})]$, where the gridpoint labeled 0 is only included for the electric fieldstrength approximation. The gridpoints $e_z(y_{-k})$ and $e_z(y_{N+k})$ are not included in the field vector as they are boundary points. The discrete spatial differentia-

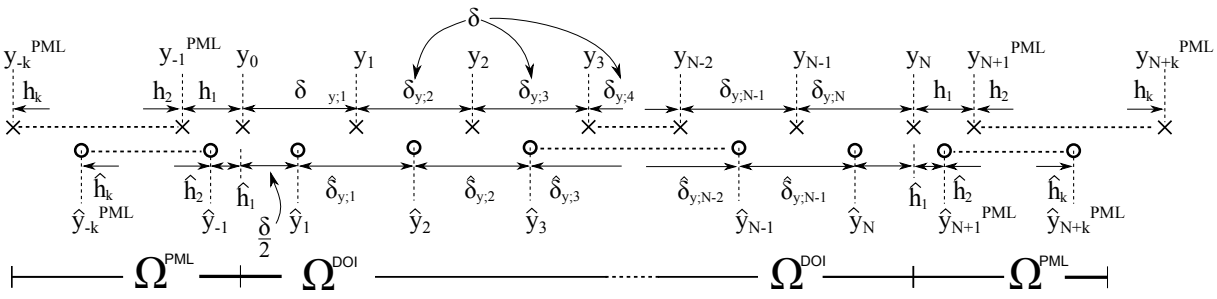


Figure C.1: Illustration of the grid used for the 1D simulation. Note that the step sizes between points $\{\hat{y}_{-1}, \hat{y}_1\}$ and $\{\hat{y}_N, \hat{y}_{N+1}\}$ are given by $\delta/2$ plus the complex distribution from the PML.

tion operator matrix D is given by

$$\mathsf{D} = \begin{pmatrix} 0 & \hat{\mathsf{Y}} \\ \mathsf{Y} & 0 \end{pmatrix}, \tag{C.14}$$

with $\hat{\mathsf{Y}}$ and Y the dual and primary differentiation matrix given by

$$\mathsf{Y} = \begin{bmatrix} h_k^{-1} & & & & & & & & & \\ -h_{k-1}^{-1} & h_{k-1}^{-1} & & & & & & & & \\ & \ddots & \ddots & & & & & & & \\ & & -\delta_{y;1}^{-1} & \delta_{y;1}^{-1} & & & & & & \\ & & & \ddots & \ddots & & & & & \\ & & & & -\delta_{y;N}^{-1} & \delta_{y;N}^{-1} & & & & \\ & & & & & -h_{k-1}^{-1} & h_{k-1}^{-1} & & & \\ & & & & & & \ddots & \ddots & & \\ & & & & & & & -h_{k-1}^{-1} & h_{k-1}^{-1} & \\ & & & & & & & & -h_k^{-1} \end{bmatrix} \tag{C.15}$$

$\hat{\mathsf{Y}}$ is the equivalent upper bidiagonal matrix containing the dual step sizes.

$$\hat{\mathsf{Y}} = \mathrm{diag}(\hat{h}_k, \ldots, \hat{h}_1, \hat{\delta}_1, \ldots, \hat{\delta}_{N-1}, \hat{h}_1, \ldots, \hat{h}_k)^{-1} \cdot \mathrm{bidiag}_{N+2k-1}(-1,1) \tag{C.16}$$

We now further introduce the step size matrixes W_y and $\hat{\mathsf{W}}_y$ as

$$\mathsf{W}_y = \mathrm{diag}(h_k, \ldots, h_1, \delta_1, \ldots, \delta_N, h_1, \ldots, h_k) \tag{C.17}$$

and the dual step size matrix as

$$\hat{\mathsf{W}}_y = \mathrm{diag}(\hat{h}_k, \ldots, \hat{h}_1, \hat{\delta}_1, \ldots, \hat{\delta}_{N-1}, \hat{h}_1, \ldots, \hat{h}_k). \tag{C.18}$$

The differentiation matrices can now comfortably be written as

$$\hat{\mathsf{Y}} = \hat{\mathsf{W}}_y^{-1} \cdot \mathrm{bidiag}_{N+2k-1}(-1,1) \quad \textit{and} \quad \mathsf{Y} = -\mathsf{W}_y^{-1} \cdot \mathrm{bidiag}_{N+2k-1}(-1,1)^T, \tag{C.19}$$

where the function $bidiag_K(a,b)$ makes a upper bidiagonal matrix of size $K \times K+1$ with diagonal elements a and upper diagonal elements b. Further we define the full step size matrix as

$$\mathsf{W} = \begin{pmatrix} \hat{\mathsf{W}}_y & 0 \\ 0 & \mathsf{W}_y \end{pmatrix}. \tag{C.20}$$

The medium matrix M is defined as

$$\mathsf{M} = \mathrm{diag}(\varepsilon(y_{-(k-1)}), \ldots, \varepsilon(y_{N+k-1}), \mu(\hat{y}_{-k}), \ldots, \mu(\hat{y}_{N+k})), \tag{C.21}$$

and

$$\mathsf{S} = \mathrm{diag}(\sigma(y_{-(k-1)}), \ldots, \sigma(y_{N+k-1}), 0, \ldots, 0). \tag{C.22}$$

There are $N+2k+1$ primary grid points and $N+2k$ dual grid points. The outer most primary grid points have PEC boundary conditions. Therefore the matrix M is square and has a length of $2N+4k-1$. At both interfaces of two media the average of the neighbouring medium parameters are chosen.

C.1.2 Symmetry properties

We note that the operator matrix $\mathsf{M}^{-1}\mathsf{D}$ is in general neither symmetric nor anti-symmetric. However we note that

$$\hat{\mathsf{Y}}^T \hat{\mathsf{W}}_y = -\mathsf{W}_y \mathsf{Y} = \mathrm{bidiag}(-1,1)^T, \tag{C.23}$$

which leads to the anti-commutativity relation

$$\mathsf{D}^T\mathsf{W} = -\mathsf{W}\mathsf{D}. \tag{C.24}$$

Using this it can be shown that the opterator A is anti-symmetric with respect to the MW weighted inner product.

$$\begin{aligned}
\langle \mathsf{x}|\mathsf{A}\mathsf{y}\rangle_{\mathsf{MW}} &= \mathsf{y}^\mathsf{T}\mathsf{A}^\mathsf{T}\mathsf{M}\mathsf{W}\mathsf{x} && \text{(C.25)}\\
&= \mathsf{y}^\mathsf{T}\mathsf{D}^\mathsf{T}\mathsf{W}\mathsf{x} && \text{(C.26)}\\
&= -\mathsf{y}^\mathsf{T}\mathsf{W}\mathsf{D}\mathsf{x} && \text{(C.27)}\\
&= -\mathsf{y}^\mathsf{T}\mathsf{W}\mathsf{M}\mathsf{M}^{-1}\mathsf{D}\mathsf{x} && \text{(C.28)}\\
&= -\mathsf{y}^\mathsf{T}\mathsf{M}\mathsf{W}\mathsf{M}^{-1}\mathsf{D}\mathsf{x} && \text{(C.29)}\\
&= -\langle \mathsf{A}\mathsf{x}|\mathsf{y}\rangle_{\mathsf{MW}} && \text{(C.30)}
\end{aligned}$$

Therefore we can use a modified Lanczos algorithm using this anti-symmetry. Since A is anti-symmetric in the weighted inner product we know that H is anti-symmetric in the classical sense ($\mathsf{H}^T = -\mathsf{H}$). Therefore we use the three term recurrence relation $\mathsf{A}\mathsf{v}_i = \beta_{i+1}\mathsf{v}_{i+1} - \beta_i\mathsf{v}_{i-1}$. The term containing v_i drops out for anti-symmetric matrices, as $\alpha_i = -\alpha_i$ is only true for $\alpha_i = 0$.
In the lossy case we have a nonzero matrix S; The solution to Eq. (C.13) in the Laplace domain is now given by

$$\hat{\mathsf{f}} = (\mathsf{M}^{-1}\mathsf{D} + \mathsf{M}^{-1}\mathsf{S} + s\mathsf{I})^{-1}\mathsf{M}^{-1}\hat{\mathsf{q}}'. \tag{C.31}$$

Therefore we see that the system matrix is now given by

$$\mathsf{A} = \mathsf{M}^{-1}\mathsf{D} + \mathsf{M}^{-1}\mathsf{S} \tag{C.32}$$

where the first part is skew-symmetric with respect to the WM weighted inner product and the second part is a diagonal matrix which is obviously symmetric. Therefore the latter part commutes with every diagonal matrix. Now we use the fact that the skew-symmetry of A can be transformed into a symmetry by adding the weight factor δ^- to the inner product with

$$\delta^- = \mathrm{diag}(\underbrace{1,1,\ldots,1}_{\#\ \mathsf{N}+2\cdot\mathsf{k}-1},\underbrace{-1,-1,\ldots,-1}_{\#\ \mathsf{N}+2\cdot\mathsf{k}}) \tag{C.33}$$

Therefore the commutation relation

$$\mathsf{A}^T\delta^-\mathsf{W}\mathsf{M} = \delta^-\mathsf{W}\mathsf{M}\mathsf{A}, \tag{C.34}$$

holds which leads to a symmetry in the $\delta^-\mathsf{WM}$ weighted pseudo inner product.

C.2 Two-dimensional system formulation

In this section the formulation used to simulate 2D EM-wave propagation for H- and E-polarized fields is presented. The detailed derivation of the H-polarized system matrix can be found in [31]. For the E-polarized waves it can be found in [26]. A short derivation of the H-polarized algorithm will be presented, whereas the results for E-polarization are simply given. The first part deals with H-polarized fields.
At last the symmetry matrices used in the Lanczos algorithm are explained.

C.2.1 System Formulation for H-polarized fields

In the Maxwell equation we choose to set all fields to zero except $\hat{\mathcal{E}}_x$, $\hat{\mathcal{E}}_y$ and $\hat{\mathcal{H}}_z$. Therefore the EM-wave propagates in the xy-plane. The Maxwell equations in the Laplace domain equations now read

$$-\partial_y \hat{\mathcal{H}}_z + s\varepsilon\hat{\mathcal{E}}_x + \sigma\hat{\mathcal{E}}_x = -\hat{\mathcal{J}}_x^{\text{ext}}, \tag{C.35}$$
$$\partial_x \hat{\mathcal{H}}_z + s\varepsilon\hat{\mathcal{E}}_y + \sigma\hat{\mathcal{E}}_y = -\hat{\mathcal{J}}_y^{\text{ext}}, \tag{C.36}$$
$$\partial_x \hat{\mathcal{E}}_y - \partial_y \hat{\mathcal{E}}_x + s\mu\hat{\mathcal{H}}_z = -\hat{\mathcal{K}}_z^{\text{ext}}, \tag{C.37}$$

or in operator form

$$(\mathcal{D} + \mathcal{S} + \mathcal{M}s)\hat{\mathcal{F}} = \hat{\mathcal{Q}}' \tag{C.38}$$

With $\mathcal{D}$ the spatial differentiation operator, $\mathcal{M}, \mathcal{S}$ the medium matrices, $\hat{\mathcal{F}}$ containing the field quantities and $\hat{\mathcal{Q}}'$ the vector containing the electromagnetic sources.

Discretization

We introduce primary and secondary grids in x- and y-direction. There are P secondary grid points in the x-direction and Q secondary points in the y-direction. Further there are k PML layers at all sides. Therefore we define $Qk = Q + 2\cdot k$ and $Pk = P + 2\cdot k$
Following this the discrete field approximations, excluding the Dirichlet boundary grid points are given by

$$\hat{\mathsf{H}}_z \quad \text{size} \qquad (Pk)\times(Qk) \tag{C.39}$$
$$\hat{\mathsf{E}}_x \quad \text{size} \qquad (Pk)\times(Qk-1) \tag{C.40}$$
$$\hat{\mathsf{E}}_y \quad \text{size} \qquad (Pk-1)\times(Qk) \tag{C.41}$$

Seen the fact that our discretization has PEC at the outer most grid points $\hat{\mathsf{H}}_z(\hat{x},\hat{y})$ is defined on the dual grid points in the x- and y-direction. Further $\hat{\mathsf{E}}_y(x,\hat{y})$ is defined on the primary x- and dual y-grid. Finally $\hat{\mathsf{E}}_x(\hat{x},y)$ is defined on the dual x- and primary y-grid.
We introduce the discretized spatial differentiation matrices as

$$\mathsf{Y} = \begin{bmatrix} h_k^{-1} & & & & & & & & & & \\ -h_{k-1}^{-1} & h_{k-1}^{-1} & & & & & & & & & \\ & \ddots & \ddots & & & & & & & & \\ & & -\delta_{y;1}^{-1} & \delta_{y;1}^{-1} & & & & & & & \\ & & & \ddots & \ddots & & & & & & \\ & & & & -\delta_{y;Q}^{-1} & \delta_{y;Q}^{-1} & & & & & \\ & & & & & -h_{k-1}^{-1} & h_{k-1}^{-1} & & & & \\ & & & & & & \ddots & \ddots & & & \\ & & & & & & & -h_{k-1}^{-1} & h_{k-1}^{-1} & & \\ & & & & & & & & -h_k^{-1} & & \end{bmatrix} \tag{C.42}$$

and the dual spatial discretization matrix $\hat{\mathsf{Y}}$ is the equivalent upper bidiagonal matrix containing the dual step sizes.

$$\mathsf{W}_y = \text{diag}(\hat{h}_k,\ldots,\hat{h}_1,\hat{\delta}_{y;1},\ldots,\hat{\delta}_{y;Q-1},\hat{h}_1,\ldots,\hat{h}_k) \quad \text{and} \quad \hat{\mathsf{Y}} = \mathsf{W}_y^{-1}\cdot\text{bidiag}_{Q+2k-1}(-1,1). \tag{C.43}$$

Equivalently we can define the matrixes

$$\mathsf{W}_x = \mathrm{diag}(\hat{h}_k, \ldots, \hat{h}_1, \delta_{x;1}, \ldots, \delta_{x;P}, \hat{h}_1, \ldots, \hat{h}_k) \quad \text{and} \quad \mathsf{X} = -\mathsf{W}_x^{-1} \cdot \mathrm{bidiag}_{P+2k-1}(-1,1)^T. \tag{C.44}$$

and

$$\hat{\mathsf{W}}_x = \mathrm{diag}(\hat{h}_k, \ldots, \hat{h}_1, \hat{\delta}_{x;1}, \ldots, \hat{\delta}_{x;P-1}, \hat{h}_1, \ldots, \hat{h}_k) \quad \text{and} \quad \hat{\mathsf{X}} = \hat{\mathsf{W}}_x^{-1} \cdot \mathrm{bidiag}_{P+2k-1}(-1,1), \tag{C.45}$$

in the x-direction. Using these differentiation matrixes the Maxwell equations can be written in their discrete form as

$$-\hat{\mathsf{H}}_z \hat{\mathsf{Y}}^T + \varepsilon s \hat{\mathsf{E}}_x + \sigma \hat{\mathsf{E}}_x = -\mathsf{J}_x, \tag{C.46}$$
$$\hat{\mathsf{X}} \hat{\mathsf{H}}_z + \varepsilon s \hat{\mathsf{E}}_y + \sigma \hat{\mathsf{E}}_y = -\mathsf{J}_y, \tag{C.47}$$
$$\mathsf{X} \hat{\mathsf{E}}_y - \hat{\mathsf{E}}_x \mathsf{Y}^T + \mu s \hat{\mathsf{H}}_z = -\mathsf{K}_z. \tag{C.48}$$

In these equations $\hat{\mathsf{E}}$ is a matrix containing the electric field approximation. In order to be able to define a overall system matrix A in the form

$$(\mathsf{A} + s\mathsf{I})\mathsf{f} = \mathsf{M}^{-1}\mathsf{q}', \tag{C.49}$$

all matrixes containing field quantities need to be transformed into vectors. We use lexicographical ordering of the 2D quantities. Let the operator vec() be the lexicographical mapping operator with

$$\hat{\mathsf{e}}_y = \mathrm{vec}(\hat{\mathsf{E}}_y). \tag{C.50}$$

In Matlab this is obtained by using `h=reshape(H,Pk*Qk,1)`. Further let the Kronecker product of two matrixes A and B be given by $\mathsf{A} \otimes \mathsf{B}$. The identity connecting lexicographical ordering and the Kroneker product used to derive the final system matrix is given by

$$\mathrm{vec}(\mathsf{B}\mathsf{X}\mathsf{A}^T) = (\mathsf{A} \otimes \mathsf{B})\mathrm{vec}(\mathsf{X}). \tag{C.51}$$

$$-(\hat{\mathsf{Y}} \otimes \mathsf{I}_{Pk})\hat{\mathsf{h}}_z + s\mathsf{M}_{\varepsilon;x}\hat{\mathsf{e}}_x + \mathsf{M}_{\sigma;x}\hat{\mathsf{e}}_x = -\mathsf{j}_x^{\mathrm{ext}}, \tag{C.52}$$
$$(\mathsf{I}_{Qk} \otimes \hat{\mathsf{X}})\hat{\mathsf{h}}_z + s\mathsf{M}_{\varepsilon;y}\hat{\mathsf{e}}_y + \mathsf{M}_{\sigma;y}\hat{\mathsf{e}}_y = -\mathsf{j}_y^{\mathrm{ext}}, \tag{C.53}$$
$$(\mathsf{I}_{Qk} \otimes \mathsf{X})\hat{\mathsf{e}}_y - (\mathsf{Y} \otimes \mathsf{I}_{Pk})\hat{\mathsf{e}}_x + s\mathsf{M}_{\mu}\hat{\mathsf{h}}_z = -\mathsf{k}_z^{\mathrm{ext}}, \tag{C.54}$$

with M the medium matrices containing the medium parameters in lexicographical ordering. We now introduce D, the differentiation matrix given by,

$$\mathsf{D} = \begin{bmatrix} 0 & -(\hat{\mathsf{Y}} \otimes \mathsf{I}_{Pk}) & 0 \\ -(\mathsf{Y} \otimes \mathsf{I}_{Pk}) & 0 & (\mathsf{I}_{Qk} \otimes \mathsf{X}) \\ 0 & (\mathsf{I}_{Qk} \otimes \hat{\mathsf{X}}) & 0 \end{bmatrix}. \tag{C.55}$$

The medium matrices are given by,

$$\mathsf{M} = \begin{bmatrix} \mathsf{M}_{\varepsilon;x} & 0 & 0 \\ 0 & \mathsf{M}_{\mu} & 0 \\ 0 & 0 & \mathsf{M}_{\varepsilon;y} \end{bmatrix} \tag{C.56}$$

and

$$\mathsf{S} = \begin{bmatrix} \mathsf{M}_{\sigma;x} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathsf{M}_{\sigma;y} \end{bmatrix}. \tag{C.57}$$

Finally the field and source vectors are given by

$$\hat{\mathsf{f}} = [\hat{\mathsf{e}}_x^T, \hat{\mathsf{h}}_z^T, \hat{\mathsf{e}}_y^T]^T \text{ and } \hat{\mathsf{q}}' = [(-\mathsf{j}_x^{\text{ext}})^T, (-\hat{\mathsf{k}}_z^{\text{ext}})^T, (-\mathsf{j}_y^{\text{ext}})^T]^T, \tag{C.58}$$

respectively. Using these definitions the system can be written as

$$(\mathsf{D} + \mathsf{M}s + \mathsf{S})\hat{\mathsf{f}} = \hat{\mathsf{q}}', \tag{C.59}$$

or as

$$(\mathsf{A} + s\mathsf{I})\hat{\mathsf{f}} = \mathsf{M}^{-1}\hat{\mathsf{q}}', \tag{C.60}$$

where the system matrix has been defined as

$$\mathsf{A} = \mathsf{M}^{-1}\mathsf{D} + \mathsf{M}^{-1}\mathsf{S}. \tag{C.61}$$

Symmetry properties

For the lossless Maxwell equations we define the following anti-commutation relationship,

$$\mathsf{D}^T\mathsf{W} = -\mathsf{W}^T\mathsf{D}, \tag{C.62}$$

with

$$\mathsf{W} = \begin{bmatrix} (\hat{\mathsf{W}}_y \otimes \mathsf{W}_x) & 0 & 0 \\ 0 & (\mathsf{W}_y \otimes \mathsf{W}_x) & 0 \\ 0 & 0 & (\mathsf{W}_y \otimes \hat{\mathsf{W}}_x) \end{bmatrix}. \tag{C.63}$$

For the lossy Maxwell equations we define the following commutation relationship,

$$\mathsf{D}^T\mathsf{W}_- = \mathsf{W}_-^T\mathsf{D}, \tag{C.64}$$

with

$$\mathsf{W}_- = \begin{bmatrix} -(\hat{\mathsf{W}}_y \otimes \mathsf{W}_x) & 0 & 0 \\ 0 & (\mathsf{W}_y \otimes \mathsf{W}_x) & 0 \\ 0 & 0 & -(\mathsf{W}_y \otimes \hat{\mathsf{W}}_x) \end{bmatrix}. \tag{C.65}$$

C.2.2 System Formulation for E-polarized fields

We write the system of Maxwell equations as

$$(\mathcal{D} + \mathcal{M}_1 + \mathcal{M}_2 s)\hat{\mathcal{F}} = \hat{\mathcal{Q}}', \tag{C.66}$$

with

$$\hat{\mathcal{F}} = [\hat{\mathcal{H}}_x, \hat{\mathcal{E}}_z, \hat{\mathcal{H}}_y]^T \tag{C.67}$$

and

$$\hat{\mathcal{Q}}' = [-\hat{\mathcal{K}}_x^{\text{ext}}, -\hat{\mathcal{J}}_z^{\text{ext}}, -\hat{\mathcal{K}}_y^{\text{ext}}]^T \tag{C.68}$$

Discretization

Seen the fact that our discretization has PEC at the outer most grid points $\hat{\mathsf{E}}_z(x,y)$ is defined on the primary grid points in the x- and y-direction. Further $\hat{\mathsf{H}}_x(x,\hat{y})$ is defined on the primary x- and dual y-grid. Last $\hat{\mathsf{H}}_y(\hat{x},y)$ on the dual x- and primary y-grid. We define the field quantities not coinciding with the PEC as

$$\hat{\mathsf{E}}_z \quad \text{size} \qquad (Pk) \times (Qk) \tag{C.69}$$

$$\hat{\mathsf{H}}_x \quad \text{size} \qquad (Pk) \times (Qk+1) \tag{C.70}$$

$$\hat{\mathsf{H}}_y \quad \text{size} \qquad (Pk+1)\times(Qk) \tag{C.71}$$

Using this discretization we can define the following differentation matrix

$$\mathsf{D} = \begin{bmatrix} 0 & (\mathsf{Y}\otimes \mathsf{I}_{Pk}) & 0 \\ (\hat{\mathsf{Y}}\otimes \mathsf{I}_{Pk}) & 0 & -(\mathsf{I}_{Qk}\otimes \hat{\mathsf{X}}) \\ 0 & -(\mathsf{I}_{Qk}\otimes \mathsf{X}) & 0 \end{bmatrix}. \tag{C.72}$$

The medium matrices are given by,

$$\mathsf{M} = \begin{bmatrix} \mathsf{M}_{\mu;x} & 0 & 0 \\ 0 & \mathsf{M}_{\varepsilon} & 0 \\ 0 & 0 & \mathsf{M}_{\mu;y} \end{bmatrix}, \tag{C.73}$$

and

$$\mathsf{S} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathsf{M}_{\sigma} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{C.74}$$

where M_ε, M_σ etc. contain the (averaged) medium parameters on the grid points. The exact definitions can be found in [26]. Finally the field and source vectors are given by

$$\hat{\mathsf{f}} = [\hat{\mathsf{h}}_x^T, \hat{\mathsf{e}}_z^T, \hat{\mathsf{h}}_y^T]^T \text{ and } \hat{\mathsf{q}}' = [(-\mathsf{k}_x^{\text{ext}})^T, (-\hat{\mathsf{j}}_z^{\text{ext}})^T, (-\mathsf{k}_y^{\text{ext}})^T]^T, \tag{C.75}$$

respectively.
Using these definitions the system can be written as

$$(\mathsf{D}+\mathsf{M}s+\mathsf{S})\hat{\mathsf{f}} = \hat{\mathsf{q}}', \tag{C.76}$$

or as

$$(\mathsf{A}+s\mathsf{I})\hat{\mathsf{f}} = \mathsf{M}^{-1}\hat{\mathsf{q}}', \tag{C.77}$$

where the system matrix has been defined as

$$\mathsf{A} = \mathsf{M}^{-1}\mathsf{D}+\mathsf{M}^{-1}\mathsf{S}. \tag{C.78}$$

Symmetry properties

For the lossless Maxwell equations we define the following anti-commutation relationship,

$$\mathsf{D}^T\mathsf{W} = -\mathsf{W}^T\mathsf{D}, \tag{C.79}$$

with

$$\mathsf{W} = \begin{bmatrix} (\mathsf{W}_y\otimes \hat{\mathsf{W}}_x) & 0 & 0 \\ 0 & (\hat{\mathsf{W}}_y\otimes \hat{\mathsf{W}}_x) & 0 \\ 0 & 0 & (\hat{\mathsf{W}}_y\otimes \mathsf{W}_x) \end{bmatrix}. \tag{C.80}$$

For the lossy Maxwell equations we define the following commutation relationship

$$\mathsf{D}^T\mathsf{W}_- = \mathsf{W}_-^T\mathsf{D}, \tag{C.81}$$

with

$$\mathsf{W}_- = \begin{bmatrix} -(\mathsf{W}_y\otimes \hat{\mathsf{W}}_x) & 0 & 0 \\ 0 & (\hat{\mathsf{W}}_y\otimes \hat{\mathsf{W}}_x) & 0 \\ 0 & 0 & -(\hat{\mathsf{W}}_y\otimes \mathsf{W}_x) \end{bmatrix}. \tag{C.82}$$

C.3 Three-dimensional system formulation

For three-dimensional systems our standard formulation as

$$(\mathcal{D}+\mathcal{S}+\mathcal{M}\partial_t)\mathcal{F} = \mathcal{Q}' = w(t)\mathcal{Q}, \tag{C.83}$$

is used again with

$$\mathcal{F} = [\mathcal{E}_x, \mathcal{E}_y, \mathcal{E}_z, \mathcal{H}_x, \mathcal{H}_y, \mathcal{H}_z]^T, \quad \text{and} \quad \mathcal{Q}' = [-\mathcal{J}_{\mathrm{x}}^{\mathrm{ext}}, -\mathcal{J}_{\mathrm{y}}^{\mathrm{ext}}, -\mathcal{J}_{\mathrm{z}}^{\mathrm{ext}}, -\mathcal{K}_{\mathrm{x}}^{\mathrm{ext}}, -\mathcal{K}_{\mathrm{y}}^{\mathrm{ext}}, -\mathcal{K}_{\mathrm{z}}^{\mathrm{ext}}]^T. \tag{C.84}$$

and

$$\mathcal{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & \partial_z & -\partial_y \\ 0 & 0 & 0 & -\partial_z & 0 & \partial_x \\ 0 & 0 & 0 & \partial_y & -\partial_x & 0 \\ 0 & -\partial_z & \partial_y & 0 & 0 & 0 \\ \partial_z & 0 & -\partial_x & 0 & 0 & 0 \\ -\partial_y & \partial_x & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{M} = \begin{pmatrix} \overline{\overline{\varepsilon}} & 0 \\ 0 & \overline{\overline{\mu}} \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} \overline{\overline{\sigma}} & 0 \\ 0 & 0 \end{pmatrix}, \tag{C.85}$$

with $\overline{\overline{\varepsilon}}, \overline{\overline{\mu}}$ and $\overline{\overline{\sigma}}$ the permittivity, permeability and conductivity tensor.

C.3.1 Discretization

We use the Yee discretization of the three dimensional Maxwell equation as given in [2]. Therefore we define primary and dual grids in all spatial coordinates. The letters P, Q, R denote the number of dual step sizes in the x, y, z- direction, respictively. Using this form of discretization we can rewrite Eq. (C.83) in its discrete form as

$$(\mathsf{D}+\mathsf{S}+\mathsf{M}\partial_t)\mathsf{f} = \mathsf{q}' \tag{C.86}$$

with

$$\mathsf{f} = [\mathsf{e}_x^T, \mathsf{e}_y^T, \mathsf{e}_z^T, \mathsf{h}_x^T, \mathsf{h}_y^T, \mathsf{h}_z^T]^T, \quad \text{and} \quad \mathsf{q}' = [-\mathsf{j}_x^{\mathrm{ext};T}, -\mathsf{j}_y^{\mathrm{ext};T}, -\mathsf{j}_z^{\mathrm{ext};T}, -\mathsf{k}_x^{\mathrm{ext};T}, -\mathsf{k}_y^{\mathrm{ext};T}, -\mathsf{k}_z^{\mathrm{ext};T}]^T. \tag{C.87}$$

We assume Yee discretization and introduce the step size matrices as

$$\mathsf{W}_k = diag(\delta_{k;1}, \delta_{k;2}, \dots) \quad \text{and} \quad \hat{\mathsf{W}}_k = diag(\hat{\delta}_{k;1}, \hat{\delta}_{k;2}, \dots) \quad \forall k = x, y, z. \tag{C.88}$$

and the second order differentiation matrices, with Dirichlet boundaries as

$$\mathsf{K} = -\mathsf{W}_k^{-1} \cdot \mathrm{bidiag}(-1, 1)^T \quad \text{and} \quad \hat{\mathsf{K}} = \hat{\mathsf{W}}_k^{-1} \cdot \mathrm{bidiag}(-1, 1) \quad \forall \{\mathsf{K}, k\} = \{\mathsf{X}, x\}, \{\mathsf{Y}, y\}, \{\mathsf{Z}, z\}. \tag{C.89}$$

Now we can define the discrete spatial differentiation operator as

$$\mathsf{D} = \begin{pmatrix} 0 & \mathsf{D}_{\mathrm{h}} \\ \mathsf{D}_{\mathrm{e}} & 0 \end{pmatrix}, \tag{C.90}$$

with

$$\mathsf{D}_{\mathrm{h}} = \begin{bmatrix} 0 & (\hat{\mathsf{Z}} \otimes \mathsf{I}_{Qk} \otimes \mathsf{I}_{Pk+1}) & -(\mathsf{I}_{Rk} \otimes \hat{\mathsf{Y}} \otimes \mathsf{I}_{Pk+1}) \\ (-\hat{\mathsf{Z}} \otimes \mathsf{I}_{Qk+1} \otimes \mathsf{I}_{Pk}) & 0 & (\mathsf{I}_{Rk} \otimes \mathsf{I}_{Qk+1} \otimes \hat{\mathsf{X}}) \\ (\mathsf{I}_{Rk+1} \otimes \hat{\mathsf{Y}} \otimes \mathsf{I}_{Pk}) & -(\mathsf{I}_{Rk+1} \otimes \mathsf{I}_{Qk} \otimes \hat{\mathsf{X}}) & 0 \end{bmatrix}, \tag{C.91}$$

and

$$\mathsf{D}_\mathrm{e} = \begin{bmatrix} 0 & -(\mathsf{Z}\otimes \mathsf{I}_{Qk+1}\otimes \mathsf{I}_{Pk}) & (\mathsf{I}_{Rk+1}\otimes \mathsf{Y}\otimes \mathsf{I}_{Pk}) \\ (\mathsf{Z}\otimes \mathsf{I}_{Qk}\otimes \mathsf{I}_{Pk+1}) & 0 & -(\mathsf{I}_{Rk+1}\otimes \mathsf{I}_{Qk}\otimes \mathsf{X}) \\ -(\mathsf{I}_{Rk}\otimes \mathsf{Y}\otimes \mathsf{I}_{Pk+1}) & (\mathsf{I}_{Rk}\otimes \mathsf{I}_{Qk+1}\otimes \mathsf{X}) & 0 \end{bmatrix}. \tag{C.92}$$

Furthermore we define

$$\mathsf{W}_\mathrm{e} = \begin{bmatrix} \hat{\mathsf{W}}_z\otimes \hat{\mathsf{W}}_y\otimes \mathsf{W}_x & 0 & 0 \\ 0 & \hat{\mathsf{W}}_z\otimes \mathsf{W}_y\otimes \hat{\mathsf{W}}_x & 0 \\ 0 & 0 & \mathsf{W}_z\otimes \hat{\mathsf{W}}_y\otimes \hat{\mathsf{W}}_x \end{bmatrix}. \tag{C.93}$$

and

$$\mathsf{W}_\mathrm{h} = \begin{bmatrix} \mathsf{W}_z\otimes \mathsf{W}_y\otimes \hat{\mathsf{W}}_x & 0 & 0 \\ 0 & \mathsf{W}_z\otimes \hat{\mathsf{W}}_y\otimes \mathsf{W}_x & 0 \\ 0 & 0 & \hat{\mathsf{W}}_z\otimes \mathsf{W}_y\otimes \mathsf{W}_x \end{bmatrix}. \tag{C.94}$$

C.3.2 Symmetry properties

We note the commutation relations

$$-\mathsf{W}_k\mathsf{K} = \hat{\mathsf{K}}^T\hat{\mathsf{W}}_k \quad \forall\{\mathsf{K},k\} = \{\mathsf{X},x\},\{\mathsf{Y},y\},\{\mathsf{Z},z\}, \tag{C.95}$$

such that we can define the symmetry matrix that satisfies the commutation relation with the operator matrix A as

$$\mathsf{A}^T\mathsf{W}^\mathrm{s} = \mathsf{W}^\mathrm{s}\mathsf{A}, \tag{C.96}$$

where the symmetry matrix is given by

$$\mathsf{W}^\mathrm{s} = \mathsf{M}\begin{pmatrix} \mathsf{W}_\mathrm{e} & 0 \\ 0 & -\mathsf{W}_\mathrm{h} \end{pmatrix}. \tag{C.97}$$

Appendix D

Optimal PML formulation

D.1 Derivation of optimal step sizes

To enhance the understanding in what way the PML is considered as optimal, the derivation for propagating waves for a 2D Helmholtz operator will be followed as documented in [33]. The derivations for more general optimal PMLs follow, besides being more complicated, the same principle.
Considering a Helmholtz equation on a two-dimensional half space $\Omega \subset \mathbb{R}^2$ with $\Omega := \{y \in \mathbb{R}, x \in \mathbb{R} | x < 0\}$

$$\Delta \mathrm{u} + \omega^2 \mathrm{u} = 0 \quad \text{on } \Omega. \tag{D.1}$$

The half space $x > 0$ is used to insert a PML in order to absorb waves traveling in positive x-direction, under various angles. After a Fourier transform with respect to y-coordinates the Helmholtz systems is rewritten as

$$u_{xx} - \lambda u = 0 \quad \text{on } \Omega. \tag{D.2}$$

with $\lambda = \kappa_y^2 - \omega^2$. The solution to the upper equation is

$$u = a \exp\left(-\sqrt{\lambda} x\right) \text{ with } \sqrt{\lambda} = \begin{cases} \sqrt{\kappa_y^2 - \omega^2} & \text{if } |\frac{\kappa_y}{\omega}| \geq 1 \text{ describing evanescent waves} \\ i\sqrt{\omega^2 - \kappa_y^2} & \text{if } |\frac{\kappa_y}{\omega}| < 1 \text{ describing propagating waves} \end{cases} \tag{D.3}$$

The solution given in Eq. (D.3) at $x = 0$ directly defines the impedance condition

$$\left.\frac{u}{u_x}\right|_{x=0} = -\frac{1}{\sqrt{\lambda}}, \tag{D.4}$$

mathematically also known as Neuman to Dirichlet (NtD) map, as it converts Neuman to Dirichlet data.
Within this thesis coupled first order differential equations are of interest such that optimal step sizes on a primary and dual grid need to be obtained. Therefore the second order formulation is changed to first order formulation as

$$u_x = \sqrt{\lambda} v \quad \text{and} \quad v_x = \sqrt{\lambda} u, \tag{D.5}$$

such that the NtD seen at the interface map of Eq. (D.4) becomes $\left.\frac{u}{v}\right|_{x=0} = -1$. The step sizes of the PML now need to approximate this NtD over a predefined range of angles. Therefore an expression for the impedance inside the PML, inserted in the half space $x > 0$, is derived. The

complex stretching function χ is introduced such that the first order differential equation inside the PML becomes

$$\chi^{-1}u_x = \sqrt{\lambda}v \quad \text{and} \quad \chi^{-1}v_x = \sqrt{\lambda}u \tag{D.6}$$

with the complex stretching function

$$\chi(x, i\omega) = \alpha(x) + \frac{\beta(x)}{i\omega}. \tag{D.7}$$

Introducing k primary and dual step sizes inside the PML denoted by h_j and $\hat{h}_j$ for $j = 1, 2, \ldots, k$, the stretched differential equation can be written in finite difference form as

$$\frac{u(x_{j+1}) - u(x_j)}{\chi(\hat{x}_j)h_j} = \sqrt{\lambda}v(\hat{x}_j), \quad \frac{v(\hat{x}_{j+1}) - v(\hat{x}_j)}{\chi(x_j)\hat{h}_j} = \sqrt{\lambda}u(x_j), \quad \forall j = 1, \ldots, k. \tag{D.8}$$

Therefore the stretched step sizes inside the PML are given by

$$h_j^c = \chi(\hat{x}_j)h_j \text{ and } \hat{h}_j^c = \chi(x_j)\hat{h}_j \quad \forall j = 1, \ldots, k \tag{D.9}$$

The finite difference NtD is now given by

$$\frac{u(x_1)}{v(\hat{x}_1)} = -f_k^{(\alpha,\beta)} \tag{D.10}$$

where $f_k^{(\alpha,\beta)}$ can be found as continued fraction by subsequent substitution of Equations (D.8) into each other, assuming a Dirichlet boundary condition at the fictive node x_{k+1}, as

$$f_k^{(\alpha,\beta)} = \cfrac{1}{\hat{h}_1^c\sqrt{\lambda} + \cfrac{1}{h_1^c\sqrt{\lambda} + \cfrac{1}{\hat{h}_2^c\sqrt{\lambda} + \cdots + \cfrac{1}{\hat{h}_k^c\sqrt{\lambda} + \cfrac{1}{h_k^c\sqrt{\lambda}}}}}} \tag{D.11}$$

Now a minimization problem arises as the difference between the upper NtD map and the NtD of the half space $x < 0$ needs to be minimalized. For this example propagating waves are considered only. Therefore $\sqrt{\lambda}$ is imaginary and is redefined as

$$\sqrt{\lambda} = i\omega\sqrt{\gamma}, \quad \text{with} \quad \gamma = 1 - (\frac{\kappa_y}{\omega})^2 \ \sqrt{\gamma} \in]0, 1] \tag{D.12}$$

Therefore with $\sqrt{\gamma}$ the angle of the incoming wave is defined. The PML is only optimized over a range incoming wave vectors, such that the optimization interval is defined as $\gamma \in [\gamma_{min}, 1]$. Using this redefinition the stretched step sizes are defined as

$$h_j^c = [\beta(\hat{x}_j) + i\omega\alpha(\hat{x}_j)]h_j\sqrt{\gamma} \text{ and } \hat{h}_j^c = [\beta(x_j) + i\omega\alpha(x_j)]\hat{h}_j \quad \forall j = 1, \ldots, k \tag{D.13}$$

For $f_k^{(\alpha,\beta)}$ to only depend implicitly via γ on frequency, α is set to zero. This is unique for the derivation of PMLs for propagating waves, as pure imaginary step sizes only add damping, but no propagation delay, which damps evanescent waves. Further it can be seen that changing the step sizes and keeping $\beta = 1$ has the same impact as fixing the step sizes and changing β.
The problem is thus reformulated from finding the optimal stretching function β to finding the optimal step sizes that minimize the impedance mismatch at the interface $x = 0$. The NtD map

is no longer a function of α and β such that the notation of $f_k^{(0,1)}$ is changed to f_k.

Now the PML error is given by the difference of the NtD maps at the interface $x = 0$, such that the minimization problem reads

$$\delta_k = \min_{h_j, \hat{h}_j} \max_{\gamma_{min} \leq \gamma \leq 1} |1 - f_k(\gamma)| \tag{D.14}$$

This problem can be rewritten as a third type Zolotarev problem of rational approximation in the complex plane. Therefore a change of variable is used introducing $z = \frac{\gamma}{\gamma_{min}}$ and the rational function $m(z)$ as

$$\sqrt{z} m(z) = f_k(\gamma_{min} z). \tag{D.15}$$

$f_k(\gamma_{min} z)$ is a continued fraction in $\sqrt{\gamma}$ and can therefore be rewritten as rational function of order $[(2k-1)/2k]$ in $\sqrt{\gamma}$. Therefore $m(z)$ can be written as a rational function of order $[(k-1)/k]$ in γ. Finding the optimal rational function that minimizes δ_k as

$$\delta_k = \min_{h_j, \hat{h}_j} \max_{1 \geq z \geq \frac{1}{\gamma_{min}}} \left| 1 - \sqrt{z} m(z) \right|, \tag{D.16}$$

was solved by the Russian mathematician Y.I.Zolotarev in the 19th century in the scope of rational approximation theory. Zolotarevs solution denoted as $m_Z(z)$ is given by

$$m_z(z) = D \frac{\prod_{n=1}^{k-1} (z + c_{2n})}{\prod_{n=1}^{k} (z + c_{2n-1})} \tag{D.17}$$

with

$$c_n = \frac{\mathrm{sn}^2\left(\frac{nK(\sqrt{1-\gamma_{min}})}{2k}; \sqrt{1-\gamma_{min}} \right)}{\mathrm{cn}^2\left(\frac{nK(\sqrt{1-\gamma_{min}})}{2k}; \sqrt{1-\gamma_{min}} \right)} \quad n = 1, 2, \ldots, 2k-1 \tag{D.18}$$

where sn and cn are Jacobi elliptic functions and $K(x)$ denotes the complete elliptic integral. The scalar D is defined via the symmetry condition

$$\max_{[1, 1/\gamma_{min}]} [1 - \sqrt{z} m_Z(z)] = - \max_{[1, 1/\gamma_{min}]} [1 - \sqrt{z} m_Z(z)]. \tag{D.19}$$

Rewriting Zolotarevs function in pole-residue form allows to uniquely retrieve the imaginary step sizes.

In order to obtain PML step sizes that are matched to the grid the discretized NtD map should be used in Eq. (D.4) rather than the continuous NtD map. To optimize the PML for evanescent wave and propagating waves $\sqrt{\lambda}$ is not rewritten in the form of Eq. (D.12).
For one-dimensional problems the optimal PML can be simplified as shown in Appendix D.

D.2 Small angle approximation

In this section the expressions obtained for a 2D optimal PML are simplified for small angles of incidence and for 1D problems. Therefore the limit of

$$\kappa_y \to 0 \tag{D.20}$$

is of interest, such the wave has no spatial frequency in the direction perpendicular to the PML. This approach can also be used when designing 1D PMLs.

We start with our derivation at Eq. (D.12) by noting that $\gamma = 1$, and further that in the limit of $\kappa_y \to 0$ we also obtain $\gamma_{min} = 1$. Further the eliptic modulus in Eq. (D.18) is given by

$$\vartheta = \sqrt{1-\gamma_{min}} = 0. \tag{D.21}$$

The solution to Zolotarevs problem as shown in Eq. (D.17) is given by

$$m_Z(z) = D\frac{\prod_{n=1}^{k-1}(z+c_{2n})}{\prod_{n=1}^{k}(z+c_{2n-1})}, \tag{D.22}$$

with

$$c_n = \frac{\mathrm{sn}^2\left(\frac{nK(\vartheta)}{2K};\vartheta\right)}{\mathrm{cn}^2\left(\frac{nK(\vartheta)}{2K};\vartheta\right)}, \tag{D.23}$$

with sn and cn the Jacobi elliptical functions, $K(\vartheta)$ the complete elliptic integral and k the number of PML layers. Since $\vartheta = 0$ in the small angle approximation we note that

$$K(0) = \frac{\pi}{2}, \quad \mathrm{sn}(x;0) = \sin(x), \quad \text{and} \quad \mathrm{cn}(x;0) = \cos(x). \tag{D.24}$$

Therefore Zolotarevs rational function is now given by

$$m_Z(z) = D\frac{\prod_{n=1}^{k-1}\left(z+\tan^2\left(\frac{n\pi}{2k}\right)\right)}{\prod_{n=1}^{k}\left(z+\tan^2\left(\frac{(2n-1)\pi}{4k}\right)\right)}. \tag{D.25}$$

D.2.1 Determining D

The constant D is uniquely determined by Eq. (D.19). Since the interval $z = C[1, 1/\gamma_{min}]$ becomes a single point we can analytically determine D via

$$1-\sqrt{z}m_Z(z) = -1+\sqrt{z}m_Z(z) \tag{D.26}$$

$$\sqrt{z}m_Z(z) = 1 \tag{D.27}$$

$$D = \frac{\prod_{n=1}^{k}\left(1+\tan^2\left(\frac{(2n-1)\pi}{4k}\right)\right)}{\prod_{n=1}^{k-1}\left(1+\tan^2\left(\frac{n\pi}{2k}\right)\right)} \tag{D.28}$$

We now write $1+\tan^2(x) = \frac{1}{\cos^2(x)}$ and obtain

$$D = \frac{\prod_{n=1}^{k-1}\cos^2\left(\frac{n\pi}{2k}\right)}{\prod_{n=1}^{k}\cos^2\left(\frac{(2n-1)\pi}{4k}\right)}. \tag{D.29}$$

We now use the identity

$$\prod_{n=1}^{k}\cos\left(\frac{(2n-1)\pi}{4k}\right) = \frac{\sqrt{2}}{2^k}, \tag{D.30}$$

which can be obtained from setting a k-th order Butterworth low pass filter transfer functions equal to the cut of frequency. Squaring the identity we can write

$$D = 2^{2k-1}\cdot\prod_{n=1}^{k-1}\cos^2\left(\frac{n\pi}{2k}\right). \tag{D.31}$$

To replace the second identity we first rewrite the cosine product into a sine product. For odd k we can write

$$\prod_{n=1}^{k-1}\cos^2\left(\frac{n\pi}{2k}\right) = \prod_{n=1}^{\frac{k-1}{2}}\cos^2\left(\frac{n\pi}{2k}\right)\cos^2\left(\frac{(k-n)\pi}{2k}\right) \tag{D.32}$$

$$= \prod_{n=1}^{\frac{k-1}{2}}\sin^2\left(\frac{n\pi}{2k}\right)\sin^2\left(\frac{(k-n)\pi}{2k}\right) \tag{D.33}$$

$$= \prod_{n=1}^{k-1}\sin^2\left(\frac{n\pi}{2k}\right) \tag{D.34}$$

$$\tag{D.35}$$

and for even k we can write

$$\prod_{n=1}^{k-1}\cos^2\left(\frac{n\pi}{2k}\right) = \cos^2\left(\frac{\pi}{4}\right)\prod_{n=1}^{\frac{k-2}{2}}\cos^2\left(\frac{n\pi}{2k}\right)\cos^2\left(\frac{(k-n)\pi}{2k}\right) \tag{D.36}$$

$$= \sin^2\left(\frac{\pi}{4}\right)\prod_{n=1}^{\frac{k-2}{2}}\sin^2\left(\frac{n\pi}{2k}\right)\sin^2\left(\frac{(k-n)\pi}{2k}\right) \tag{D.37}$$

$$= \prod_{n=1}^{k-1}\sin^2\left(\frac{n\pi}{2k}\right). \tag{D.38}$$

$$\tag{D.39}$$

Now we use the identity

$$\prod_{n=1}^{k-1}\sin\left(\frac{n\pi}{k}\right) = \frac{k}{2^{k-1}}, \tag{D.40}$$

and rewrite it by noting that the sin function is symmetric with respect to the point $\frac{\pi}{2}$ as

$$\prod_{n=1}^{k-1}\sin\left(\frac{n\pi}{2k}\right) = \frac{\sqrt{k}}{2^{k-1}}. \tag{D.41}$$

Therefore we can square this identity and obtain our final expression for D as

$$\prod_{n=1}^{k-1}\cos^2\left(\frac{n\pi}{2k}\right) = \frac{k}{2^{2k-2}} \tag{D.42}$$

to obtain

$$D = \frac{k\cdot 2^{2k-1}}{2^{2\cdot k-2}} = 2k. \tag{D.43}$$

D.2.2 Calculation of step sizes

To obtain the optimal step sizes the rational function $m_Z(z)$ if written in its pole-residue form

$$m_Z(z) = D\frac{\prod_{n=1}^{k-1}\left(z+\tan^2\left(\frac{n\pi}{2k}\right)\right)}{\prod_{n=1}^{k}\left(z+\tan^2\left(\frac{(2n-1)\pi}{4k}\right)\right)} = D\sum_{i=1}^{k}\frac{w_i}{z-\theta_i}. \tag{D.44}$$

Obviously the poles of the functions coincide such that $\theta_i = -c_{2n-1} = -\tan^2\left(\frac{(2n-1)\pi}{4k}\right)$. The numbers w_i needed to convert the Zolotarev function to its pole residue form is now given by

$$w_i = \frac{(z-\theta_i)}{D}m_Z(z) \tag{D.45}$$

$$= \frac{\prod_{n=1}^{k-1}\left(-\tan^2\left(\frac{(2i-1)\pi}{4k}\right)+\tan^2\left(\frac{n\pi}{2k}\right)\right)}{\prod_{n=1}^{i-1}\left(-\tan^2\left(\frac{(2i-1)\pi}{4k}\right)+\tan^2\left(\frac{(2n-1)\pi}{4k}\right)\right)\prod_{n=i+1}^{k}\left(-\tan^2\left(\frac{(2i-1)\pi}{4k}\right)+\tan^2\left(\frac{(2n-1)\pi}{4k}\right)\right)}, \tag{D.46}$$

where it appears that

$$\sum_{i=1}^{k} w_i = 1. \tag{D.47}$$

Following the derivation in the previous section it can be seen that the optimal DtN map is given by

$$f_k(\gamma) = \sqrt{\gamma}D\sum_{i=1}^{k}\frac{\sqrt{\gamma_{\min}}w_i}{\gamma-\gamma_{\min}\theta_i}. \tag{D.48}$$

This DtN map must be the same as the numerical one of the PML. Therefore the DtN map needs to be computed and brought into pole-residue form. This creates an inverse-eigenvalue problem which can be solved uniquely by a Lanczos algorithm. From the derivation it follows that

$$\hat{h}_1 = \frac{1}{D} = \frac{1}{2k}, \quad \mu_i = -\tan^2\left(\frac{(2i-1)\pi}{4k}\right), \tag{D.49}$$

and

$$s_i = \sqrt{w_i}. \tag{D.50}$$

Further by following the first step of the Laczos algorithm it can be found that

$$a_1 = -\frac{1}{\hat{h}_1 h_1} = \sum_{i=1}^{k} w_i\mu_i = -\sum_{i=1}^{k} w_i\tan^2\left(\frac{(2i-1)\pi}{4k}\right), \tag{D.51}$$

with a the diagonal and b the off diagonal coefficients in the Lanczos coefficient matrix. Following the Lanczos algorithm the step sizes are found as

$$h_1 = \frac{1}{\hat{h}_1}\frac{1}{\sum_{i=1}^{k} w_i\tan^2\left(\frac{(2i-1)\pi}{4k}\right)}, \tag{D.52}$$

which after substitution of w_i can be written as

$$h_1 = \frac{2k}{\frac{4}{3}k^2-\frac{1}{3}} = \frac{6k}{4k^2-1}. \tag{D.53}$$

Following the Lanczos algorithm we now find

$$\hat{h}_2 = \frac{1}{\hat{h}_1\cdot(b_2h_1)^2} \tag{D.54}$$

or

$$\hat{h}_2 = \frac{9\cdot 2k}{(4k^2-1)^2}\frac{1}{\sum_{i=1}^{k}\left(\frac{4}{3}k^2-\frac{1}{3}-\tan^2\left(\frac{(2i-1)\pi}{4k}\right)\right)^2 w_i}. \tag{D.55}$$