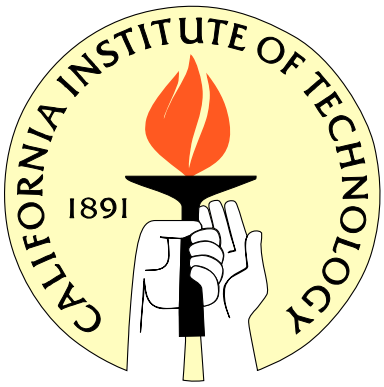
Nonreflecting Boundary Conditions Obtained From Equivalent Sources For Time-Dependent Scattering Problems

Thesis by
David Hoch

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy



California Institute of Technology
Pasadena, California

2008
(Defended May 16, 2008)

עיני תמיד אל-יהוה
כי הוא יוציא מרשת רגלי

Acknowledgments

I would like to thank my advisor, Professor Oscar P. Bruno, for sharing his vast experience, providing careful guidance and unwavering support, and for his continual encouragement. He introduced me to many advanced spectral methods and suggested the topic of this thesis. I enjoyed the many discussions we had, both mathematical and nonmathematical, which I will always keep in good memory.

I wish to express my gratitude to Professor Tom Y. Hou for encouraging me to join the Applied and Computational Mathematics Program at Caltech and for his kind support especially during the first graduate year.

Special thanks are due to Professor Christophe Geuzaine, for providing me with Gmsh, a three-dimensional finite element mesh generator he and his coworkers developed, which can be downloaded from http://geuz.org/gmsh. He also helped me to integrate the mesh generator into my existing code and provided very useful C++ classes.

Tim Elling confirmed and provided certain numerical results in connection with equivalent sources in Appendix B.2 which I am very thankful for.

I would also like to extend my thanks to Professor Thomas Hagstrom for helpful suggestions and references to related recent works.

Our talented and generous systems manager Chad Schmutzer was always willing and ready to solve the countless unpredictable problems I encountered with computers, and I owe him a special thank you for all the time he saved me.

For her thoughtfulness, kindness, and friendliness, I would like to thank our department's administrator, Ms. Sheila Shull. I also wish to thank the other faculty and staff of Caltech with whom I have had the privilege of interacting and especially the rest of my thesis committee, Professors Dan Meiron and Joe Shepherd.

I am grateful to Professor Marcus J. Grote and Professor Christoph Schwab, who supported me in

my decision to pursue a scientific career in the United States. Dr. Schwab taught me the basics of partial differential equations and *hp* finite elements. Dr. Grote introduced me to the world of nonreflecting boundary conditions.

Finally, I would like to express special thanks to my dear wife Betsy, my parents and grandparents, who all constantly encouraged and supported me during difficult times.

Abstract

In many engineering applications, scattering of acoustic or electromagnetic waves from a body of arbitrary shape is considered in an infinite medium. Solving the underlying partial differential equations with a standard numerical method such as finite elements or finite differences requires truncating the unbounded domain of definition into a finite computational region. As a consequence, an appropriate boundary condition must be prescribed at the artificial boundary. Many approaches have been proposed for this fundamental problem in the field of wave scattering. All of them fall into one of three main categories.

The first class of methods is based on mathematical approximations or physical heuristics. These boundary conditions are often local in space and time, therefore easy to implement and run in short computing times. However, these approaches give rise to spurious reflections at the artificial boundary, no matter how refined the discretization is, which travel back into the computational domain and corrupt the solution.

A second group consists of accurate and convergent methods. However, these formulations are usually nonlocal in time and space, thus harder to implement and often more expensive than the computation of the interior scheme itself.

Finally, there are methods which are accurate and fast. These approaches are often local in time, and the nonlocality in space is confined to a closed surface rather than the whole computational domain. The drawback of these approaches lies in the fact that the outer boundary must be taken to be either a sphere, a plane, or a cylinder. For many applications of interest, this may require use of a computational domain much larger than actually needed, which leads to an expensive overall numerical scheme.

This work introduces a new methodology in order to compute the fields at the artificial boundary. The boundary condition is both nonlocal in space and time, but the nonlocal behavior is confined to a finite number of points in time and to a surface in space. Like the second class of methods described

above, the proposed algorithm is accurate and numerically convergent, yet its computational cost is less than the underlying portion of the volumetric calculation. And, unlike the third category, this new approach allows us to choose the artificial boundary to be arbitrarily close to the scatterer. This method is based on a novel concept of "equivalent source" representations which allows a highly accurate and fast evaluation of the boundary condition when used in combination with fast Fourier transforms. We present a variety of numerical results demonstrating the accuracy and effectiveness of the proposed approach.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Scattering theory has played a central role in twentieth-century mathematical physics. Indeed, in fields like radar and sonar technology, earthquake simulation, aeroacoustics, medical applications of computerized tomography, or even quantum chemistry, scattering problems have attracted and challenged scientists for well over a hundred years.

The mathematical models are based on physical conservation laws, and lead to partial differential equations, whose solution may generally be obtained only by means of numerical methods. In many of these scattering problems, the phenomenon of interest is local but embedded in a large surrounding medium. Boundary effects arising from the exterior of that large region are often negligible, which allows modeling it as an infinite, unbounded domain. Sommerfeld [83] proposed a radiation condition at infinity which ensures well-posedness of the problem. In his honor, this condition is nowadays well known as the Sommerfeld radiation condition, which guarantees that the wave is purely outgoing and decaying as it approaches infinity. Standard numerical methods, such as finite differences (FD) and finite elements (FEM), can approximately solve the partial differential equation. However, this usually requires truncation of the unbounded domain and introduction of an artificial boundary condition, because the finite resources of a computer do not allow simulation of a natural phenomenon in a truly infinite domain. If the artificial condition on the truncated boundary does not behave like the actual condition at infinity, spurious reflections will be generated, which will propagate back into the local region of interest and thus pollute the solution.

A typical scattering problem consists of a bounded obstacle Γ, a source term f, and possibly an incident wave u_i. A wave is generated away from the solid object and propagates freely in space, until

it eventually hits the obstacle and is scattered. This reflected wave is called the scattered field u_s—see Figure 1.1.

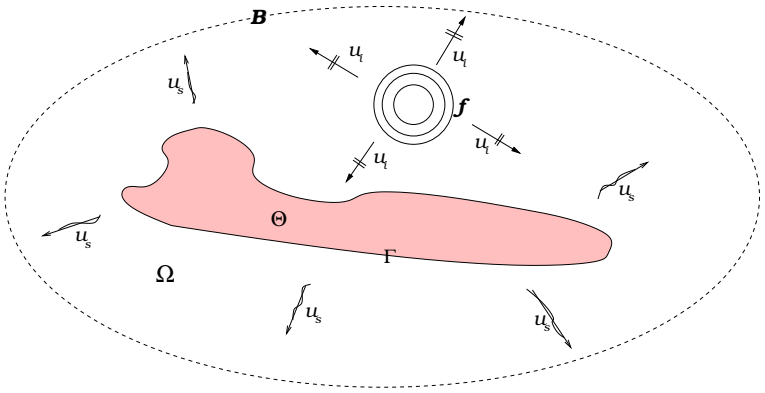


Figure 1.1: A typical time-dependent scattering problem

At any time, the total wave field u is the superposition of u_i and u_s. In all of the examples considered in this thesis the assumption is made that the scatterer is not penetrable. This assumption is in fact immaterial to the methods developed in this contribution: the computational boundary conditions we develop are applicable irrespective of whether the scatterer is penetrable or not, as long as it occupies a finite region in space. An impenetrable object is called *sound-soft* if the total wave field vanishes on its boundary, which leads to the Dirichlet boundary condition $u_s = -u_i$ on Γ. In contrast, an acoustic *sound-hard* obstacle requires the normal velocity of the total field to vanish on its boundary, which implies the Neumann condition $\partial_\nu u_s = -\partial_\nu u_i$ on Γ, where ν is the unit outward normal on Γ. In more general impenetrable models, the so-called impedance boundary condition of the form $\partial_\nu u + i\lambda u = 0$ on Γ is considered, where λ is a positive constant (see [24]).

The incident field is typically known, and the direct scattering problem is to determine u_s from the knowledge of u_i and the partial differential equation governing the wave motion.

1.1 Historical review

The list of computational boundary conditions proposed previously for the time-dependent scattering problem is indeed extensive. A survey and bibliography which includes the research in this field up to 1992 can be found in the monograph by Givoli [37]. Other comprehensive and more recent reviews can be located in Tsynkov [87] and Hagstrom [47, 48, 49]; the latter describe methods which can deliver arbitrary accuracy at acceptable computational cost.

Perhaps the most famous of the existing computational boundary conditions was introduced by Lindman [62], and further expanded by Engquist and Majda [29, 30] in the late seventies. These authors developed an exact boundary condition in terms of a pseudo-differential operator, and obtained an increasingly accurate sequence of local operators by applying Padé approximations on a certain square root function. In 1980, Bayliss and Turkel [8] used a large distance expansion of the solution and also obtained a sequence of boundary conditions. In the mid-eighties, Higdon [52] derived boundary conditions, which are perfectly absorbing at certain angles of incidence. As an alternative, Jiang and Wong [58] used a similar approach and obtained boundary conditions, which are perfectly absorbing for wave packets traveling at a certain group velocity. In 1994, Bérenger [9] introduced the perfectly matched layer (PML) for Maxwell's equations. This technique is based on the construction of an artificial layer surrounding the computational domain which would completely absorb the outgoing wave, i.e., the PML acts as reflectionless interface. Bérenger's original formulation is only weakly well-posed. A clearer understanding of PML as a complex coordinate stretching emerged in [21]. Later formulations became mathematically more clear (see [74, 77]). The PML has emerged as one of the preferred computational boundary conditions, as it provides geometric flexibility and has the potential for generalizations to inhomogeneous or even nonlinear systems. Besides, the implementation is simple, and, although the method is not directly based on an exact formulation and requires a complex parameter selection process, it is yet convergent for many applications.

Methods based on exact or convergent formulations go back to the seventies as well. One of the first nonreflecting boundary conditions was proposed by Smith [82] in 1974. He discovered that reflections from the outer boundary may be completely eliminated by adding together the solutions of the appropriate Dirichlet and Neumann problems. However, this requires running the overall interior scheme many times, making it too expensive and as a consequence useless for realistic applications. Ting and Miksis [86] suggested in the mid-eighties an approach based on Kirchhoff's formula, which was later implemented by Givoli and Cohen in [38]. This method has potential for high accuracy and the boundary can exhibit flexible topologies. He and Weston [51] developed a fully vector version of the scheme for Maxwell's equations. A drawback is the costly retarded potential evaluation. Alpert, Greengard, and Hagstrom [4] developed a nonreflecting boundary condition based on Fourier and Laplace transforms. This technique involves a costly convolution operation. However, they showed in [3] that the convolution kernel can be compressed for planar, spherical, and cylindrical boundaries. Independently, a similar

approach to the compression of boundary kernels was proposed by Lubich and Schädle [63]. Another acceleration method based on fast multipole expansions was proposed by Michielssen et al. [67, 69].

Sofronov [84], and, independently, Grote and Keller [40, 41] developed and implemented an integro-differential approach in three dimensions and demonstrated that high accuracy can be achieved. Grote and Keller extended their ideas to Maxwell's equations [43] and to the elastodynamic equation [44]. The cost to compute the boundary condition is reasonable and smaller than the interior scheme. The drawback of these approaches is that a spherical boundary must be prescribed, making the volumetric portion of the computation in case of an elongated scatterer unnecessarily expensive.

Ryaben'kii and Tsynkov [88] constructed for the time-dependent wave equation an auxiliary function satisfying a forced wave equation in free space which agrees with the solution of the original problem at the artificial boundary. They demonstrated that the auxiliary function can be computed efficiently using Fourier methods exploiting the strong Huyghens principle. Tsynkov later applied this idea to Maxwell's equations [89].

1.2 Overview

In this thesis, we shall study the three-dimensional time-dependent scalar wave equation, which, in case of a compressible fluid, can be derived from the conservation of mass and Newton's second law (see [54]). In this specific case, the wave equation describes a pressure field, and the solution to the equation is the amplitude of the pressure for a given point in space and time.

In Chapter 2, we review the wave equation defined in an unbounded domain. The Sommerfeld condition at infinity is essential for the purely outgoing character of the waves. The Kirchhoff representation plays a crucial role in this thesis and is discussed in Section 2.2. In 2.3, we present the main basis of our proposed nonreflecting computational boundary condition. The details of the algorithm are specified in the subsequent chapters.

We introduce and investigate equivalent sources for the Helmholtz equation in Section 3.1. Instead of dealing with true sources in a volumetric domain, it is possible to compute artificial sources on a certain surface which surrounds the domain of the true sources. We call these artificial distributions "equivalent sources." They represent the actual field with high accuracy at any point which lies outside a small neighborhood of the sources. Section 3.2 deals with the extension of the concept to the time domain. While the treatment of periodic data is straightforward, a new approach needs to be developed to process

nonperiodic data. The concept of continuation Fourier series proves to be of great significance in these regards. The special treatment for Fourier expansion of nonperiodic functions is discussed extensively in Section 3.3. Numerous numerical results are provided, demonstrating the high accuracy of the equivalent source technique in the time-dependent case.

In Chapter 4, we propose a technique based on equivalent sources which computes the data on the artificial boundary efficiently. The basis of our algorithm is comparable to [38, 86], but the use of equivalent sources in our approach accelerates the boundary data evaluation significantly: in the references [38, 86], the dominant work arises from the computational boundary, while we demonstrate that, in our approach, the interior computation is the dominant cost.

Our methodology is exact in the sense that no spurious reflections develop at the artificial boundary and thus, clean convergence is obtained as discretizations are refined appropriately. Methods such as [8, 9, 29, 30, 31, 32, 51, 52, 58, 60, 61, 62, 75, 77], in contrast, suffer from the problem of spurious reflections, which may result in corruption of the numerical solutions.

Although in references [65, 66, 68, 69] it was shown that the time domain multipole method can adequately be used to obtain data on the computational boundary, we have not been able to locate in any of these references information that would allow us to determine the efficiency of the approach in terms of computational time. In this thesis, we demonstrate efficient computing times of our approach, and our method may also prove to be more advantageous than the accelerated multipole technique discussed in [65, 66, 68, 69] in terms of implementation and accuracy.

Chapter 2

Wave equation in unbounded domains

2.1 Model problem

We consider a bounded domain $\Theta \subset \mathbb{R}^3$ with boundary Γ. At an arbitrary point $(\boldsymbol{x}, t) \in \mathbb{R}^3 \backslash \Theta \times \mathbb{R}_+$, the scattered field $u_s(\boldsymbol{x}, t)$ for our model problem is a real or complex valued function solving the following acoustic sound-hard problem

$$\begin{aligned}
\frac{1}{c^2}\frac{\partial^2}{\partial t^2}u_s - \Delta u_s &= f(\boldsymbol{x}, t) \quad \text{in } \mathbb{R}^3 \backslash \Theta \times (0, \infty) && (2.1)\\
u_s(\boldsymbol{x}, 0) &= u_0(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^3 \backslash \Theta && (2.2)\\
\frac{\partial}{\partial t}u_s(\boldsymbol{x}, 0) &= \dot{u}_0(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^3 \backslash \Theta && (2.3)\\
\boldsymbol{\nu} \cdot \nabla u_s &= g(\boldsymbol{x}, t) \quad \text{on } \Gamma \times (0, \infty) && (2.4)\\
\lim_{r \to \infty} r \left(\frac{\partial}{\partial r}u_s + \frac{1}{c}\frac{\partial}{\partial t}u_s \right) &= 0, \qquad r = |\boldsymbol{x}|. && (2.5)
\end{aligned}$$

We assume that the forcing term $f(\boldsymbol{x}, t)$ along with the initial conditions $u_0(\boldsymbol{x})$ and $\dot{u}_0(\boldsymbol{x})$ have a compact support that lies within a domain $\Omega \subset \mathbb{R}^3$, which surrounds the scatterer entirely and thus has Γ as its inner boundary. The scattered field u_s propagates with a finite constant velocity c away from the scatterer in the medium $\mathbb{R}^3 \backslash \Theta$. The geometry is depicted in Figure 1.1. The given function $g(\boldsymbol{x}, t)$ in (2.4) is equal to $-\partial_\nu u_i$ as discussed in Chapter 1. We note that the Neumann condition (2.4) on Γ corresponds to a sound-hard obstacle and it would be replaced by the Dirichlet condition $u_s = -u_i$ in case of a sound-soft scatterer. Finally, the Sommerfeld condition (2.5) at infinity insures that the scattered wave is purely outgoing.

A solution to this problem exists, is unique, and depends continuously on the data (see, e.g., [24] and references therein): the problem (2.1)–(2.5) is well-posed in the sense of Hadamard.

The solution of (2.1)–(2.5) can be expressed by means of an integral representation which is known in the literature as Kirchhoff's formula. This formula can be easily obtained by transforming the given equations into the Fourier space, which results in a Helmholtz problem whose solution can be expressed by means of a frequency-domain integral representation, and then transforming the result back into the time domain. The result of this calculation is summarized in the next section. In Appendix A, the well-known details associated with this result are reviewed: in Appendix A.1 the Helmholtz problem is formulated, Appendix A.2 discusses the integral representation and, Appendix A.3 transforms the solution into the time domain.

2.2 Kirchhoff representation

In the rest of this thesis, we simplify the notation by omitting the subscript s in the scattered field u_s. A well-known Kirchhoff formula for the solution of the problem (2.1)–(2.5) is summarized in the following theorem.

Theorem 2.2.1. *Let $\boldsymbol{r} = \boldsymbol{x} - \tilde{\boldsymbol{x}}$ and $r = |\boldsymbol{x} - \tilde{\boldsymbol{x}}|$. Then, using the definitions in Section 2.1, for any point $\boldsymbol{x} \in \mathbb{R}^3 \backslash \Theta$, we have*

$$u(\boldsymbol{x},t) \;=\; u^v(\boldsymbol{x},t) + u^m(\boldsymbol{x},t) + u^d(\boldsymbol{x},t), \tag{2.6}$$

where

$$u^v(\boldsymbol{x},t) = \frac{1}{4\pi}\int_\Omega \frac{f(\tilde{\boldsymbol{x}}, t-\frac{r}{c})}{r} d\tilde{\boldsymbol{x}}, \tag{2.7}$$

$$u^m(\boldsymbol{x},t) = \frac{1}{4\pi}\int_\Gamma \frac{1}{r}\frac{\partial u}{\partial \nu(\tilde{\boldsymbol{x}})}\left(\tilde{\boldsymbol{x}}, t-\frac{r}{c}\right) ds(\tilde{\boldsymbol{x}}), \tag{2.8}$$

$$u^d(\boldsymbol{x},t) = \frac{1}{4\pi}\int_\Gamma \frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{r^2}\left\{\frac{u(\tilde{\boldsymbol{x}}, t-\frac{r}{c})}{r} + \frac{1}{c}\frac{\partial u}{\partial t}\left(\tilde{\boldsymbol{x}}, t-\frac{r}{c}\right)\right\} ds(\tilde{\boldsymbol{x}}). \tag{2.9}$$

Kirchhoff's formula (2.6) is an integral expression for the solution of (2.1).

We emphasize two crucial properties of this solution:

1) In absence of the forcing f, the field u is given by a surface integration over the scatterer Γ. If f does not vanish, the volumetric term (2.7) must be added. Because of the assumption that f has a compact support in Ω, this integration would be confined to a finite three-dimensional domain.

2) The integrands in the surface integrals (2.8) and (2.9) depend on *retarded* values of the field u, its derivative in time $\partial_t u$, and its normal derivative $\partial_\nu u$ on the scatterer's surface Γ.

2.3 Scattering solver

We recall that the domain of interest Ω is chosen large enough such that the supports of the functions $f(\boldsymbol{x},t), u_0(\boldsymbol{x})$, and $\dot{u}_0(\boldsymbol{x})$ lie within Ω. Finite elements can handle complex geometries of Ω and thus are a suitable choice to resolve the computational domain accurately. However, the truncated problem is only well-posed if a boundary condition is imposed on the outer boundary $\mathcal{B}$. It is not straightforward how to reformulate Sommerfeld's radiation condition (2.5) at infinity to the finite boundary $\mathcal{B}$. As mentioned in the Introduction (Section 1.1), many approaches have been proposed to solve this fundamental problem.

In this section, we present the main basis of the new convergent computational boundary condition we introduce in this thesis. The computational boundary condition is computed from information *inside* the domain. Initially, the scattered field vanishes outside of the compact supported regions of $f(\boldsymbol{x},t), u_0(\boldsymbol{x})$ and $\dot{u}_0(\boldsymbol{x})$. We introduce a closed surface $\mathcal{S}$ which surrounds the union of these domains. The waves propagate with the constant velocity c from $\mathcal{S}$ into the infinite space and arrive at the surface $\mathcal{B}$ no earlier than $t_{\min} \equiv l_{\min}/c$, where $l_{\min}$ is the minimum distance from $\mathcal{S}$ to $\mathcal{B}$. Therefore, for the time $t \in I_0 \equiv [0, t_{min}]$, the required boundary condition on $\mathcal{B}$ is trivial, and the following well-posed scattering problem can be solved at $(\boldsymbol{x},t) \in \Omega \times I_0$ with any appropriate numerical scheme:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}u - \Delta u = f(\boldsymbol{x},t) \quad \text{in } \Omega \times I_0 \tag{2.10}$$

$$u(\boldsymbol{x},0) = u_0(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \tag{2.11}$$

$$\frac{\partial}{\partial t}u(\boldsymbol{x},0) = \dot{u}_0(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \tag{2.12}$$

$$\boldsymbol{\nu} \cdot \nabla u = g_\Gamma(\boldsymbol{x},t), \quad \text{on } \Gamma \times I_0 \tag{2.13}$$

$$\mathcal{L}[u] = 0, \quad \text{on } \mathcal{B} \times I_0, \tag{2.14}$$

where a suitable linear operator in (2.14) needs to be specified which will be discussed in Section 4.1.4.

The solution $u(\boldsymbol{x},t)$ of (2.10)–(2.14) for $(\boldsymbol{x},t) \in \Omega \times I_0$ represents the field as it travels from Γ to the outer boundary $\mathcal{B}$. During that time, the wave passes through $\mathcal{S}$; we accumulate this data on that surface. Once the wave arrives at $\mathcal{B}$, the interior scheme needs to be interrupted, because the field is now nonvanishing on the artificial boundary and therefore boundary data need to be provided. The values on $\mathcal{S}$ can be regarded as true sources, and the Kirchhoff formula (2.6) expresses that the superposition of all the infinitesimal source distributions make up the total scattered field at all points outside of $\mathcal{S}$. We note that by construction, the compact support of f is inside of the closed surface $\mathcal{S}$, therefore, the evaluation of $u(\boldsymbol{x},t)$ outside of the closed surface of $\mathcal{S}$ is restricted to surface integrations:

$$u(\boldsymbol{x},t) \quad = \quad u^m(\boldsymbol{x},t) + u^d(\boldsymbol{x},t), \tag{2.15}$$

where

$$u^m(\boldsymbol{x},t) = \frac{1}{4\pi}\int_{\mathcal{S}} \frac{1}{r}\frac{\partial u}{\partial \nu(\tilde{\boldsymbol{x}})}\left(\tilde{\boldsymbol{x}}, t-\frac{r}{c}\right) ds(\tilde{\boldsymbol{x}}), \tag{2.16}$$

$$u^d(\boldsymbol{x},t) = \frac{1}{4\pi}\int_{\mathcal{S}} \frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{r^2}\left\{\frac{u(\tilde{\boldsymbol{x}}, t-\frac{r}{c})}{r} + \frac{1}{c}\frac{\partial u}{\partial t}\left(\tilde{\boldsymbol{x}}, t-\frac{r}{c}\right)\right\} ds(\tilde{\boldsymbol{x}}). \tag{2.17}$$

The numerical approximations on the r.h.s. of equations (2.10)–(2.13) are known for $t \in I_1 = [t_{\min}, 2\cdot t_{\min}]$, and thus we are able to use the interior solver in Ω to compute the approximated solution for that time interval. The accumulated field on $\mathcal{S}$ for $t \in I_0 \cup I_1$ allows us to compute new data on $\mathcal{B}$ for $t \in I_2 \equiv [2\cdot t_{\min}, 3\cdot t_{\min}]$, which in turn can be used to use the interior algorithm for the time interval I_2. This leads to an iterative process. In the mid-eighties, Ting and Miksis [86] proposed to use (2.15) as an exact nonreflecting boundary condition, which ten years later was numerically implemented by Givoli and Cohen [38]. These authors considered $\mathcal{L}$ to be either the identity- or the Neumann-operator. In [38], it is reported that the overall algorithm exhibits numerical long-time instability: the solution converges up to a certain time to the true solution until an instability develops which manifests by the appearance of rapidly growing oscillations. Givoli and Cohen propose in [38] to remove the instability by the use of a dissipative interior scheme. The disadvantage, though, is obvious: this eliminates the use of all popular and well-understood nondissipative schemes. As we shall demonstrate in Section 4.1.4, the problem of long-time instability for nondissipative numerical schemes can be easily solved by choosing $\mathcal{L}$ to be an appropriate Sommerfeld type operator. The most significant drawback of using (2.6) directly as

the nonreflecting boundary condition is thus the long computing time: the convolution-like operations in (2.16) and (2.17) make the open boundary algorithm more expensive than the interior solver. In this thesis, we propose an approach which significantly accelerates the evaluation of the integrals (2.16)–(2.17) without degrading accuracy and, as a result, the evaluation of the computational boundary conditions is significantly faster than the overall interior PDE algorithm. This is achieved through the use of certain "equivalent sources" that, placed on an appropriate Cartesian mesh, provide useful representations of the field values. The details of the construction are addressed in subsequent chapters. This construction leads to Algorithm 2.3.1, which summarizes the procedure to determine the approximated solution to the scattering problem (2.1)–(2.5) on a finite domain. Defining $I_k = [kt_{\min}, (1+k)t_{\min}]$ for $k = 0, \ldots, N_{\max}$ the overall algorithm reads:

Algorithm 2.3.1. *Scattering solver*

1. *Initially, the scattered field propagates from Γ to $\mathcal{B}$ in the time interval I_0; the wave field thus vanishes on $\mathcal{B}$ during that time. An appropriate interior scheme can be used to solve equations (2.10) to (2.14) in the three-dimensional computational domain Ω. This interior solver needs to be interrupted once the first wave arrives at the outer boundary $\mathcal{B}$.*

2. *The nonreflecting boundary condition solver (step 3) and the interior solver (step 4) are iteratively invoked for $k = 1, \ldots, N_{\max}$.*

3. *The accumulated data on $\mathcal{S} \times I_{k-1} \cup \ldots \cup I_{k-1-m}$ for a suitable integer m in $0 < m < k$ can be used to apply the equivalent source algorithm (EQS) that shall be developed in this thesis to compute the boundary data $g_{\mathcal{B},h}(\boldsymbol{x}, t)$ on $\mathcal{B} \times I_k$.*

4. *All information on the r.h.s. of the system*

$$\begin{aligned}
\frac{1}{c^2}\frac{\partial^2}{\partial t^2}u - \Delta u &= f(\boldsymbol{x}, t) \qquad \text{in } \Omega \times I_k && (2.18)\\
u(\boldsymbol{x}, kt_{\min}) &= u_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega && (2.19)\\
\frac{\partial}{\partial t}u(\boldsymbol{x}, kt_{\min}) &= \dot{u}_k(\boldsymbol{x}), \qquad \boldsymbol{x} \in \Omega && (2.20)\\
\boldsymbol{\nu} \cdot \nabla u &= g_\Gamma(\boldsymbol{x}, t), \quad \text{on } \Gamma \times I_k && (2.21)\\
\mathcal{L}[u] &= g_{\mathcal{B},h}(\boldsymbol{x}, t), \ \text{on } \mathcal{B} \times I_k, && (2.22)
\end{aligned}$$

is known. Thus, an appropriate interior numerical method can be used to obtain the solution in

$\Omega \times I_k$.

Chapter 3

Time domain equivalent sources

In this chapter we consider time-dependent wave fields generated by volumetric source distributions inside a cube, and we formulate a methodology to represent that wave to a high-order accuracy by equivalent sources positioned on any given pair of opposite faces of the cube. A corresponding methodology was introduced in [13, 14] for the Helmholtz equation for the frequency domain. A review of the material introduced previously is presented in Appendix B. In Section 3.1, in turn, we give an extensive discussion about the behavior of the approximated frequency domain field for specific parameter values. Then, our extension of the concept to time-periodic functions is presented in Section 3.2: a time-periodic function can be accurately approximated by a Fourier representation, thus enabling us to use the equivalent source technique for every wave number separately. Section 3.3 is devoted to the more realistic case when the time dependent field is nonperiodic. A Fourier transform cannot be used directly, since this would give rise to the Gibbs phenomenon. A partition of unity method could be applied to such a signal to split the initially nonperiodic wave into wave packets that can be viewed as periodic functions. However, as shown in 3.3.1, this method may affect the computing time of the nonreflecting boundary algorithm negatively. The continuation method introduced in 3.3.2 overcomes the shortcomings of the partition of unity approach. In 3.3.3 the necessity of defining a time buffer in connection with the computational boundary condition is explored. Finally, we close this chapter with numerical experiments in 3.3.4 demonstrating the effectiveness and accuracy of the approach.

3.1 Parameter value identification

In order to identify relevant parameters for the accuracy of the equivalent sources, we give here a brief review of the two-face approach. Let us assume that true sources are embedded in a cube of edge size H and generate a time-harmonic field with wave number k. It has been established in [13, 14], and is also discussed in Appendix B, that there are equivalent sources on two opposite faces D_1 and D_2 of the cube which approximate the initial wave to a high accuracy outside of a certain neighborhood of the source distributions.

In practice, these artificial point sources are constructed as follows: the two selected faces D_1 and D_2 are each discretized using a set of $S \times S$ equidistant nodes. This gives rise to two Cartesian uniform two-dimensional grids $\tau_S^{(1)}$ and $\tau_S^{(2)}$ of mesh size $\Delta_S = H/(S-1)$ which are located at the two faces D_1 and D_2, respectively, so there is a total of $n_S = 2 \times S \times S$ such points. An equivalent monopole source ξ_j and an equivalent dipole source η_j are placed at each node $\boldsymbol{y}_j \in \tau_S^{(1)} \cup \tau_S^{(2)}$. The geometry is depicted in Figure 3.1.

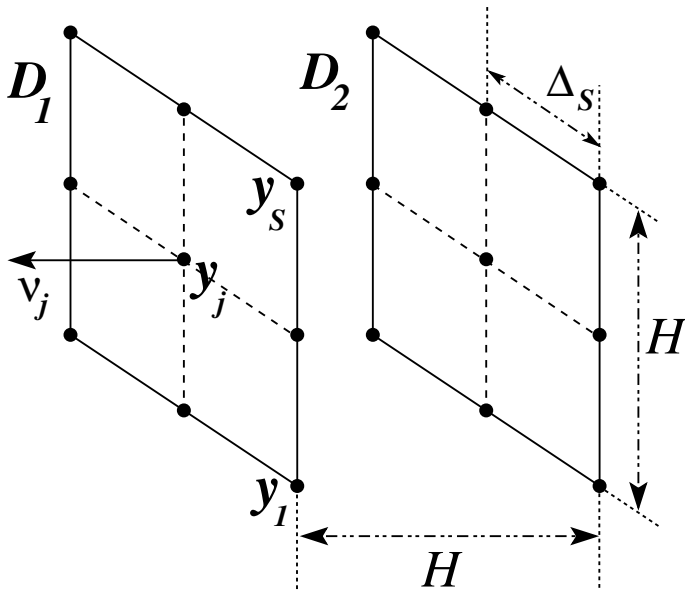


Figure 3.1: The two discretized faces in three dimensions

A second, larger cube also centered about the origin is constructed. We denote the union of its six faces by $\mathcal{S}_C$ and call it the collocation surface. Each face of this collocation cube is discretized into $C \times C$ equidistant nodes, i.e., the mesh size is $\Delta_C = H_C/(C-1)$, where H_C denotes the edge length of the collocation cell. Under the assumption that the field $\hat{u}$ is known at the collocation points $\boldsymbol{x}_l$ with

$l \in \{1, \ldots, n_C\}$, where $n_C = 6 \times C \times C - 12 \times C + 8$ (see Section 3.1.1), the values of $\boldsymbol{\xi}_j = \xi(\boldsymbol{y}_j)$ and $\boldsymbol{\eta}_j = \eta(\boldsymbol{y}_j)$ for $j \in \{1, \ldots, n_S\}$ are obtained by solving the overdetermined system

$$\hat{\boldsymbol{u}} = \begin{bmatrix} \boldsymbol{A}_m & \boldsymbol{A}_d \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{bmatrix}, \tag{3.1}$$

where the vector $\hat{\boldsymbol{u}}$ is defined by

$$\hat{\boldsymbol{u}} = \begin{bmatrix} \hat{u}(\boldsymbol{x}_1) \\ \vdots \\ \hat{u}(\boldsymbol{x}_l) \\ \vdots \\ \hat{u}(\boldsymbol{x}_{n_C}) \end{bmatrix}, \tag{3.2}$$

the monopole matrix is given by

$$\{\boldsymbol{A}_m\}_{l,j} = \frac{e^{ikr_{l,j}}}{4\pi r_{l,j}}, \tag{3.3}$$

and the dipole matrix is

$$\{\boldsymbol{A}_d\}_{l,j} = \frac{e^{ikr_{l,j}}}{4\pi r_{l,j}^2} \left(\frac{1}{r_{l,j}} - ik \right) \boldsymbol{\nu}_j \cdot \boldsymbol{r}_{l,j}. \tag{3.4}$$

In (3.3) and (3.4), we use the notation $\boldsymbol{r}_{l,j} = \boldsymbol{x}_l - \boldsymbol{y}_j, r_{l,j} = |\boldsymbol{r}_{l,j}|$ and $\boldsymbol{\nu}_j$ is the unit normal to the two faces at $\boldsymbol{y}_j \in \tau_S^{(1)} \cup \tau_S^{(2)}$. The overdetermined system (3.1) is solved in the least-square sense by means of a singular value decomposition. The computational cost of this procedure is then an order $\mathcal{O}(n_C \cdot n_S^2)$ operation. Hence, both S and C should be reasonably small, in order to avoid large computing times. However, it should be noted that the singular value decomposition needs to be performed only once. The equivalent sources evaluate then the field at any desired point outside of the collocation cell by the matrix-vector multiplication (3.1). This algorithm is summarized in Appendix B.2.1. In the following subsection, we present the results of a variety of computational tests designed to demonstrate

the dependence of the algorithm's performance on the values of the various associated parameters.

3.1.1 Accuracy as a function of collocation cube size

Le us consider a time-harmonic field with wave number k radiating from a point within a cubic cell c_i of side length H. On two opposite faces D_1 and D_2 of c_i, an appropriate number of locations for the equivalent sources $n_S = 2 \times S \times S$ is selected (compare with Figure 3.1). We note that in [12, 13, 14], it is proposed to place the equivalent sources on the points of the extended planes of D_1 and D_2 which lie within the union of two circular domains concentric with (and containing) the faces of c_i. The radius of these domains is chosen to be equal to (or slightly larger than) half the length of the diagonals of the faces. However, our finding is that in the context of the present work there is no disadvantage in terms of accuracy and computing time if we choose to place equivalent sources directly on the Cartesian grids $\tau_S^{(1)} \cup \tau_S^{(2)}$, and thus this shall be our standard choice in this thesis. We refer to Appendix B.2 for further discussion on this issue. We recall that in order to evaluate the equivalent sources, the field needs to be specified at the $n_C = 6 \times C \times C - 12 \times C + 8$ collocation points of the surface $\mathcal{S}_\mathcal{C}$. The number n_C results as sources are placed on each one of the six faces on the cube: on the first of the three pairs of opposite faces, we place $2 \times C \times C$ Cartesian points; on the second pair of opposite faces, only $2 \times C \times C - 4 \times C$ new positions can be located; and finally, on the last opposite pair, only the $2 \times C \times C - 8 \times C + 8$ interior points of the Cartesian grid can be selected. We assume that the wave values at the collocation points along with the parameters k, H, S, C are known. The purpose of this subsection is to determine the dimensions of a suitable collocation cube which is characterized by the edge length H_C. To this end, we select the specific values $k = 10, S = 7$, and $C = 7$ (which means that $n_S = 98$ and $n_C = 218$). Numerical results suggest it is best to choose n_C at least $2n_S$. Numerical experiments further indicate that under these constraints, any other choice of parameters for k, S, C lead quantitatively to the same conclusions. We consider the five different values $2H, 2.5H, 3H, 4H$, and $5H$ for H_C. Once the collocation cube is known, Algorithm B.2.1 can be used to obtain an approximation of the wave at any point outside of the collocation cube At the fixed point in space $P = [0, 1.25, 1.26]^t$, we compute the numerical error $E_P(H_C)$ to the exact solution in absolute norm. The results are displayed in Figure 3.2 for the four different values $0.00025, 0.0025, 0.025$, and 0.25 for H. We note that the point P lies outside of the collocation cube for all five choices of H_C. Figure 3.2 leads to the following observation: for any fixed H, the equivalent source computation yields a more accurate solution at P as the distance

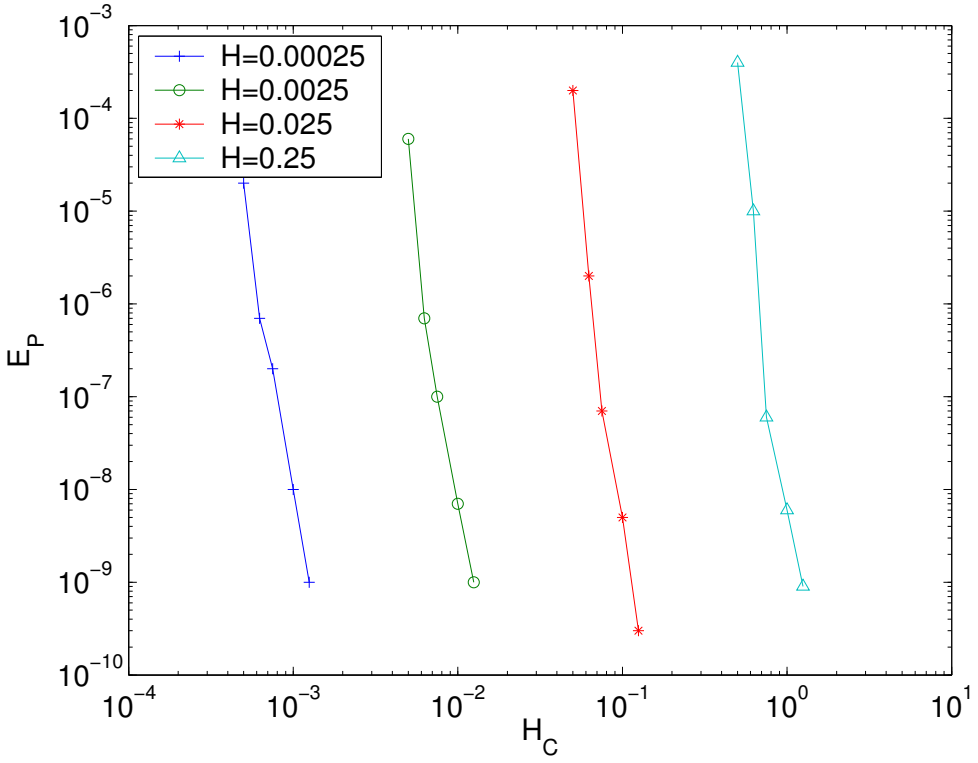


Figure 3.2: The error $E_P(H_C)$ as a function of H_C at $P = [0, 1.25, 1.26]^t$ for different values of H

from the two faces $D_1 \cup D_2$ to the collocation points *increases*. Clearly, in this experiment the increased distance is realized by increasing the size of H_C while keeping H constant, and by doing so, the surface $\mathcal{S}_\mathcal{C}$ comes closer to the point P. It might be thought that the decreased distance from P to the collocation points influences the conclusion "the bigger the collocation cube, the more accurate the equivalent source computation."

To see whether that is indeed the case, we consider the following experiment: for the value $H = 0.25$, we select one of the five parameters for H_C from the first example, and we evaluate the error $E(P)$ at the location $P = [0, l, 0]^t$, where l takes one of the nine values $0.6H, 0.8H, 1.2H, 1.4H, 1.5H, 1.6H, 2H$, or $3H$. Clearly, the first point lies for all five collocation boxes between the surfaces of c_i and $\mathcal{S}_\mathcal{C}$, while the last point is positioned outside the collocation cube. In Figure 3.3, we plot $E(P)$ as a function of P for the smallest and the largest collocation box, i.e., $H_C = 2H$ and $H_C = 5H$, respectively. The quantitative behavior is obvious: starting just outside c_i and moving toward $\mathcal{S}_\mathcal{C}$, the accuracy increases until the collocation surface is reached. Continuing moving in the same direction, the behavior remains unchanged, i.e., the farther away from the collocation box, the more accurate. Looking back at the first

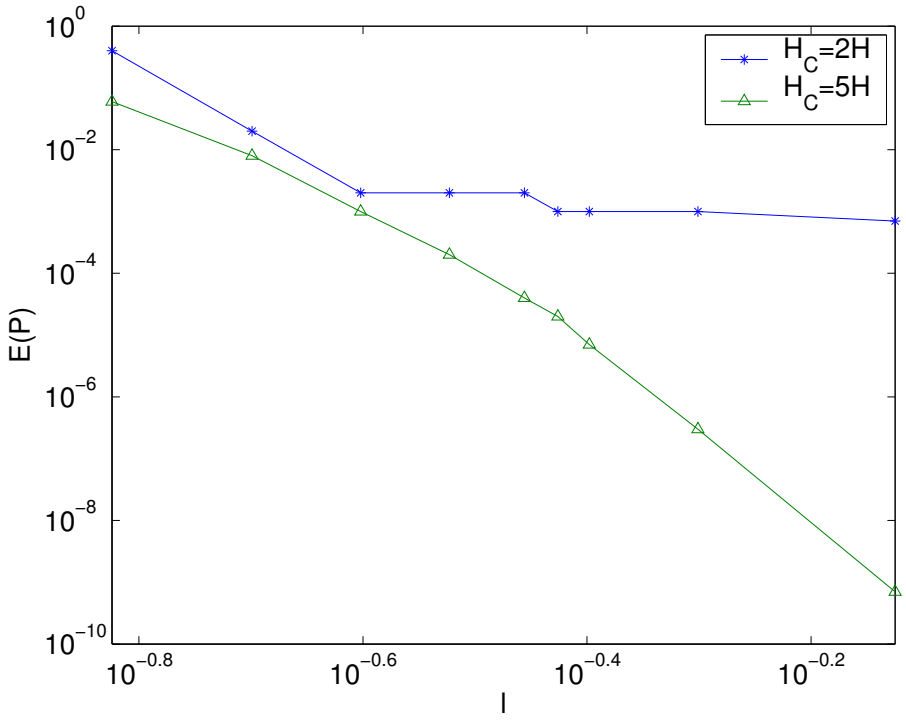


Figure 3.3: The error $E(P)$ as a function of $P = [0, l, 0]^t$ for two different values of H_C

experiment, we see that as the collocation cube increases, the more accurate the solution, despite the fact that the point under consideration is closer. This leads to the conclusion that a larger collocation box yields more accurate results.

As we will explain later, ideally, we want to choose H_C as small as possible, but on the other hand, the scheme should be as accurate as possible. These two trends conflict each other, and the selection $H_C = 3H$ seems to be a reasonable compromise. Thus, $H_C = 3H$ is our standard choice if not stated otherwise.

3.1.2 Point source at the origin

In this subsection, we assume that the point source $\tilde{\boldsymbol{x}}$ of the Green's function (A.12) is located at the origin. The two faces D_1 and D_2 are centered about the origin, and given a certain wave number k and side length H, we seek to determine the number of sources necessary to obtain a prescribed accuracy from Algorithm B.2.1. The wave length is proportional to $1/k$ (see (A.9)), and therefore, Δ_S should be proportional to $1/k$ to adequately resolve the wave in space. Table 3.1 shows the accuracy of the two-face approach for various parameters k, H, n_S, and n_C. The table is meant as an illustration only, to demonstrate how the change of various parameters affects the accuracy of the solution. In the first column, we consider four specific values for the wave number, i.e., $k = 0, 25, 100,$ and 300. The value in parentheses next to the wave number is its inverse, which gives a rough idea what the closest distance between two equivalent sources is expected to be. Studies show that the biggest error to the exact solution in absolute norm outside of the collocation box are found close to the surface $\mathcal{S}_{\mathcal{C}}$, as has been established in the last Section 3.1.1. In the second column, different sizes H of the panel length are considered. The entries in the third, fourth, and fifth columns are all linked to the parameter S. To see that the resolution of the equivalent sources lies in the right range, it is helpful to compare the closest distance between the sources with the inverted values of the wave number. It is thus useful to consider the value Δ_S, which is given in the third column. The entries of the fourth column represent the number of sources S along the panel-side H. The total amount of equivalent sources n_S are tabulated in the fifth column. These entries are helpful for estimating computing times. Similarly, the parameters Δ_C, C and n_C are given in the sixth, seventh, and eighth columns, respectively. Finally, the values of the last column correspond to the absolute error at the point $[0, 3H, 0]^t$. Table 3.1 shows that, as we expect, for a fixed k and H, the accuracy increases as Δ_S and Δ_C are decreased. Interestingly, as we increase H and keep the wave number k fixed, we can use larger Δ_S and still obtain the same order of accuracy. The number of sources, however, will generally increase as can be seen in Table 3.1 and also in Figure 3.5.

Now we pose the following question: given a wave number k and a panel length H, what mesh size Δ_S is at least required to obtain a prescribed accuracy of the field? The answer can be found in Figures 3.4 and 3.5 for wave numbers up to $k = 60$ and panel lengths up to $H = 0.5$. The parameters are chosen in such a way that the error in these figures is at least of the order $\mathcal{O}(10^{-6})$ outside of a sphere about the origin of radius $3H$. The algorithm which determines the parameters picks for any S automatically $S+1$

k, $(\frac{1}{k})$	H	Δ_S	S	n_S	Δ_C	C	n_C	error
0, (∞)	0.025	0.0125	3	18	0.025	4	56	$3 \cdot 10^{-4}$
					0.01875	5	98	$1 \cdot 10^{-5}$
		0.0083333	4	32	0.01875	5	98	$4 \cdot 10^{-6}$
					0.015	6	152	$3 \cdot 10^{-6}$
		0.00625	5	50	0.01875	5	98	$1 \cdot 10^{-6}$
					0.015	6	152	$6 \cdot 10^{-7}$
		0.005	6	72	0.015	6	152	$6 \cdot 10^{-7}$
		0.0041667	7	98	0.0125	7	218	$8 \cdot 10^{-9}$
		0.0035714	8	128	0.0107143	8	296	$3 \cdot 10^{-9}$
	0.075	0.025	4	32	0.05625	5	98	$1 \cdot 10^{-6}$
		0.01875	5	50	0.05625	5	98	$5 \cdot 10^{-7}$
		0.0125	7	98	0.0375	7	218	$4 \cdot 10^{-9}$
		0.0107143	8	128	0.0321429	8	296	$1 \cdot 10^{-9}$
	0.5	0.25	3	18	0.5	4	56	$1 \cdot 10^{-5}$
		0.0833333	7	98	0.25	7	218	$7 \cdot 10^{-10}$
25, (0.04)	0.025	0.0125	3	18	0.025	4	56	$4 \cdot 10^{-4}$
		0.0083333	4	32	0.01875	5	98	$1 \cdot 10^{-5}$
		0.00625	5	50	0.015	6	152	$1 \cdot 10^{-6}$
	0.075	0.01875	5	50	0.05625	5	98	$9 \cdot 10^{-7}$
		0.0107143	8	128	0.0321429	8	296	$6 \cdot 10^{-9}$
	0.5	0.125	5	50	0.375	5	98	$2 \cdot 10^{-2}$
		0.0833333	7	98	0.214286	8	296	$2 \cdot 10^{-4}$
		0.0714286	8	128	0.214286	8	296	$7 \cdot 10^{-5}$
		0.0625	9	162	0.1875	9	386	$8 \cdot 10^{-7}$
100, (0.01)	0.025	0.00625	5	50	0.015	6	152	$1 \cdot 10^{-6}$
	0.075	0.015	6	72	0.045	6	152	$1 \cdot 10^{-4}$
					0.0375	7	218	$2 \cdot 10^{-5}$
		0.0125	7	98	0.0125	8	296	$9 \cdot 10^{-7}$
300, (0.003333)	0.025	0.0041667	7	98	0.0125	7	218	$4 \cdot 10^{-6}$
	0.075	0.00625	13	338	0.0160714	15	1178	$8 \cdot 10^{-7}$

Table 3.1: The field is generated by a point source located at the origin. The table demonstrates the accuracy of the two-face approach for various parameters k, H, n_S, and n_C.

collocation points along the large cube of size $3H$. If the accuracy fails to be achieved, it changes C up to $S + 5$, which means that there might be certain cases where the desired accuracy could be obtained by choosing a lower S and more than $S + 5$ collocation points. In Figure 3.4, we observe that in general, as k is held constant and H gradually increases, the mesh size Δ_S can be chosen larger to obtain the same accuracy of at least $\mathcal{O}(10^{-6})$. But Figure 3.5 also reveals that "the larger H, the better" does not always hold. For $k = 60$ fixed, for example, we need to have only $S = 4$ sources along the panel length $H = 0.025$, while it is required to have $S = 16$ for $H = 0.5$. If we were to put equivalent sources on faces of the prescribed length $H = 0.5$, it would make more sense to partition them into a couple of panels of smaller lengths, to make the underlying singular value decomposition computation less expensive.

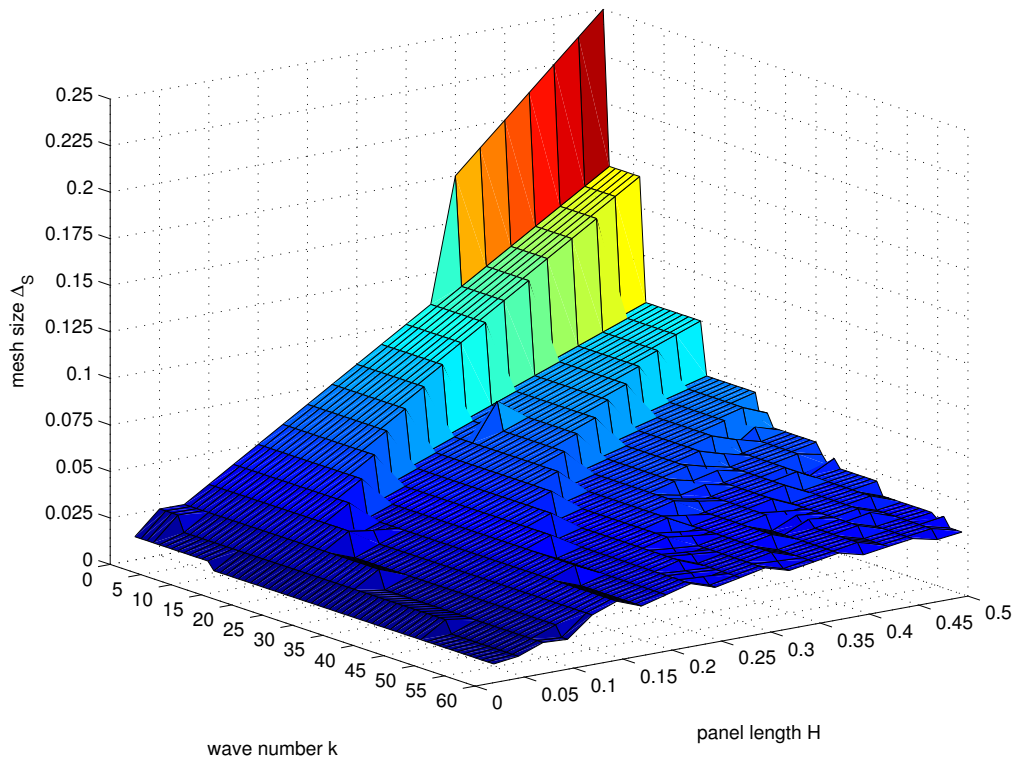


Figure 3.4: The mesh size Δ_S as a function of the wave number k and the panel length H to obtain at least $\mathcal{O}(10^{-6})$ accuracy in the field values

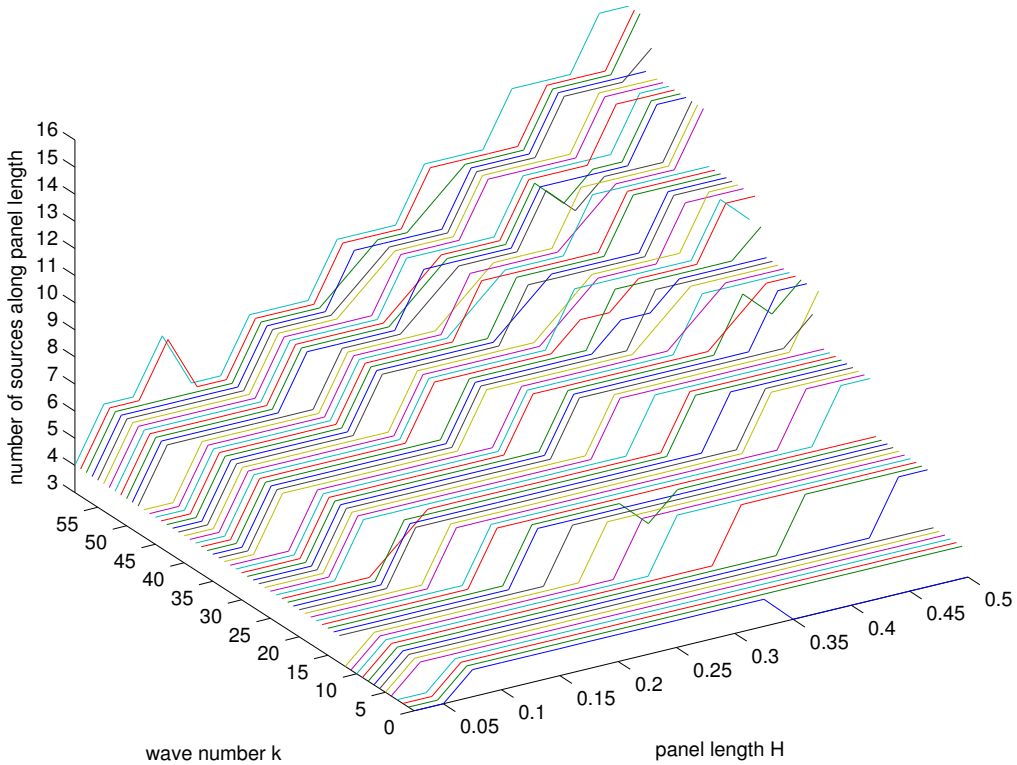


Figure 3.5: The number of equivalent sources S along one panel length as a function of the wave number k and the panel length H to obtain at least $\mathcal{O}(10^{-6})$ accuracy in the field values

3.1.3 Point source at the most challenging location

Numerical experiments indicate that it is most challenging to position the test source $\tilde{\boldsymbol{x}}$ at $[H/2, 0, -H/2]$, or in the middle of one of the other three edges which connect the two faces D_1 and D_2. Table 3.2 displays numerical accuracy for various parameter values. The structure of the table is the same as in Table 3.1, but these two tables are not meant to directly compare with each other. Rather, Table 3.2 is supposed to demonstrate that high accuracy of order $\mathcal{O}(10^{-7})$ for wave numbers up to $k = 100$ and panel lengths H between 0.01 and 0.05 can be achieved by selecting S and C in the range of 10 to 15. Computational results in [14] suggest that the accuracy increases by increasing kH, and the error estimate (B.17) gives a rough order for the values considered there ($kH = 8, 12$, and 16). Here, we are interested in achieving a high accuracy for smaller values of kH. The results in Table 3.2 show that a significantly higher accuracy can be achieved for small values kH than estimate (B.17) indicates. In fact, Table 3.2 demonstrates that at least a $\mathcal{O}(10^{-7})$ accuracy can be achieved for kH in the range of 0.25 to 5. For a given wave number k and panel length H, the error is largely influenced by the choices of Δ_S and Δ_C. The results for $k = 100$ in Table 3.2 suggest that for a fixed wave number, the accuracy indeed increases with higher H: for the values $H = 0.01$ and $H = 0.025$ for example, the error of $\mathcal{O}(10^{-7})$ does not improve as S and C are increased; for $H = 0.05$ however, the numerical approximation exceeds this limit with appropriate values S and C.

The diagrams in Figures 3.6 to 3.9 display the relationship that is needed between k, H, and S so that the numerical solution approximates the true solution to at least $\mathcal{O}(10^{-5})$. Figure 3.10 displays the number of collocation points C used along one edge. The algorithm used here to determine the parameters tests systematically for S fixed sources $C = S$ and more collocation points until the desired accuracy is achieved. If the desired accuracy fails to be reached for that number of sources, the algorithm increases S by one and the search for the lowest value C starts again. This is repeated until the appropriate values S and C for a given k are found. The procedure is performed for all wave numbers of interest.

The rest of Chapter 3 deals with the extension of Algorithm B.2.1 to the time domain: time dependent waves propagate from source distributions located inside of the cubic cell c_i of edge length H into the infinite three-dimensional space. The goal is to find time-dependent equivalent sources on the faces D_1 and D_2 yielding a high-order representation of the field. At any point $\boldsymbol{x} \in \mathcal{S}_C$, we imagine to decompose the wave into wave packets: this can be realized with the partition of unity (see Section 3.3.1

k	H	Δ_S	S	n_S	Δ_C	C	n_C	error
25	0.01	0.00111	10	200	0.00333	10	488	$1 \cdot 10^{-8}$
		0.000909	12	288	0.002727	12	728	$2 \cdot 10^{-7}$
		0.000714	15	450	0.001875	17	1538	$2 \cdot 10^{-7}$
	0.025	0.00277	10	200	0.00833	10	488	$2 \cdot 10^{-8}$
		0.00227	12	288	0.006818	12	728	$5 \cdot 10^{-8}$
	0.05	0.00556	10	200	0.01667	10	488	$2 \cdot 10^{-8}$
		0.004545	12	288	0.013636	12	728	$4 \cdot 10^{-8}$
50	0.01	0.00111	10	200	0.00333	10	488	$1 \cdot 10^{-8}$
		0.000909	12	288	0.002727	12	728	$1 \cdot 10^{-7}$
		0.000714	15	450	0.001875	17	1538	$2 \cdot 10^{-7}$
	0.025	0.00277	10	200	0.00833	10	488	$4 \cdot 10^{-8}$
		0.00227	12	288	0.006818	12	728	$9 \cdot 10^{-8}$
	0.05	0.00455	10	200	0.01667	10	488	$9 \cdot 10^{-8}$
		0.004545	12	288	0.01363	12	728	$2 \cdot 10^{-7}$
75	0.01	0.001111	10	200	0.0033333	10	488	$9 \cdot 10^{-7}$
		0.000909	12	288	0.0027273	12	728	$2 \cdot 10^{-7}$
		0.000714	15	450	0.001875	17	1538	$3 \cdot 10^{-7}$
	0.025	0.002778	10	200	0.0083333	10	488	$8 \cdot 10^{-8}$
		0.002273	12	288	0.0068182	12	728	$1 \cdot 10^{-7}$
		0.001786	15	450	0.0046875	17	1538	$2 \cdot 10^{-7}$
	0.05	0.005556	10	200	0.0166667	10	488	$4 \cdot 10^{-7}$
		0.004545	12	288	0.0136364	12	728	$3 \cdot 10^{-7}$
		0.003571	15	450	0.009375	17	1538	$2 \cdot 10^{-7}$
	0.075	0.008333	10	200	0.025	10	488	$2 \cdot 10^{-6}$
		0.006818	12	288	0.0204545	12	728	$3 \cdot 10^{-7}$
		0.005357	15	450	0.0140625	17	1538	$2 \cdot 10^{-7}$
100	0.01	0.00111	10	200	0.00333	10	488	$1 \cdot 10^{-6}$
		0.000909	12	288	0.002727	12	728	$3 \cdot 10^{-7}$
		0.000714	15	450	0.001875	17	1538	$3 \cdot 10^{-7}$
	0.025	0.002778	10	200	0.008333	10	488	$2 \cdot 10^{-7}$
		0.002272	12	288	0.006818	12	728	$4 \cdot 10^{-7}$
		0.001786	15	450	0.004688	17	1538	$4 \cdot 10^{-7}$
	0.05	0.005556	10	200	0.016667	10	488	$1 \cdot 10^{-6}$
		0.004545	12	288	0.013636	12	728	$4 \cdot 10^{-7}$
		0.003571	15	450	0.009375	17	1538	$2 \cdot 10^{-8}$

Table 3.2: The field is generated by a point source located at $[H/2, 0, -H/2]$. The table displays the accuracy of the two-face approach for various parameters k, H, n_S, and n_C.

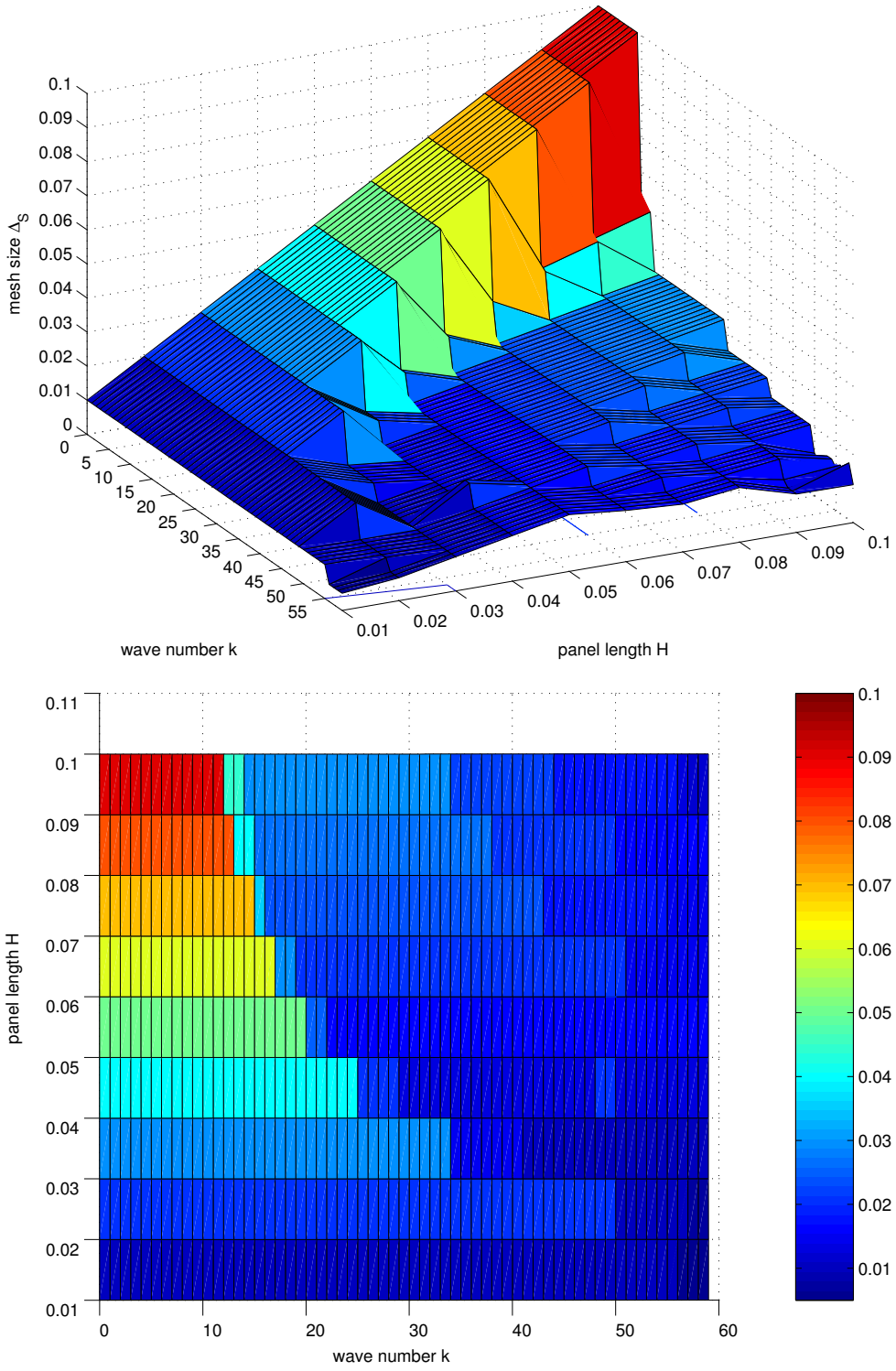


Figure 3.6: Hardest case: The mesh size Δ_S in relation to the wave number k and the panel length H to obtain at least $\mathcal{O}(10^{-5})$ accuracy in the field values

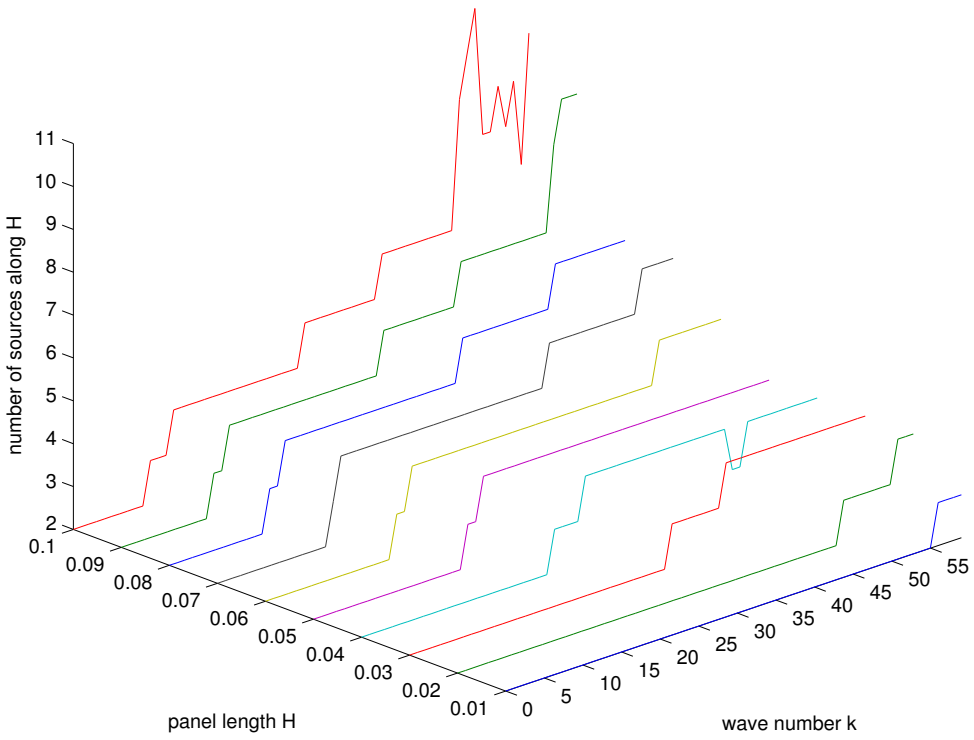


Figure 3.7: Hardest case: The number of equivalent sources S along one panel length in dependence of the wave number k and the panel length H to obtain at least $\mathcal{O}(10^{-5})$ accuracy in the field values

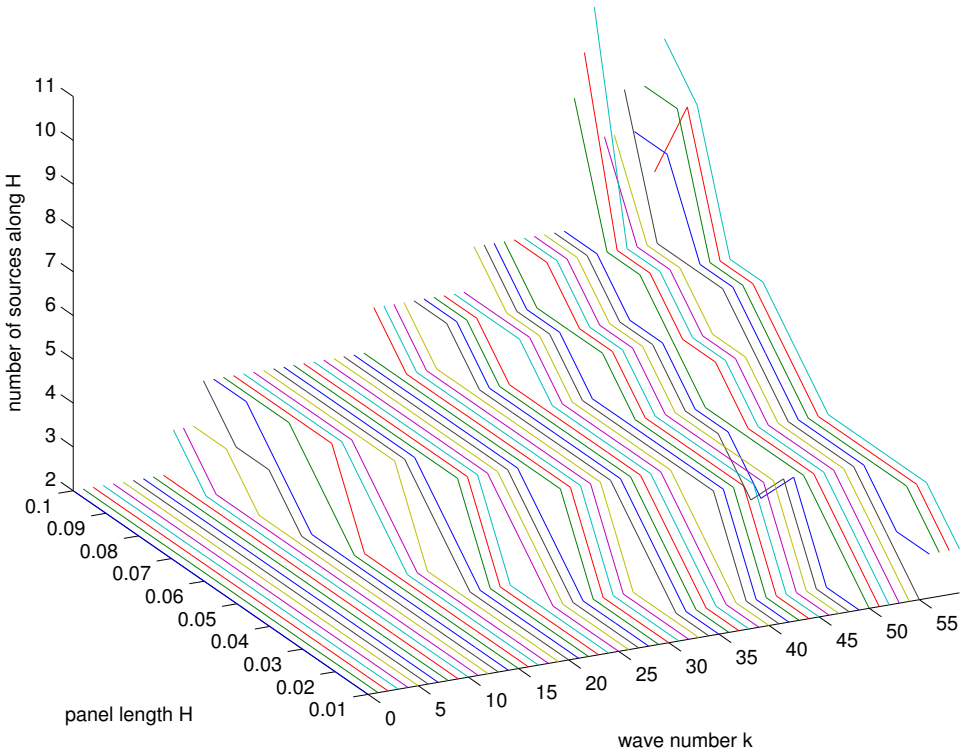


Figure 3.8: Hardest case: The number of equivalent sources S along one panel length in dependence of the wave number k and the panel length H to obtain at least $\mathcal{O}(10^{-5})$ accuracy in the field values

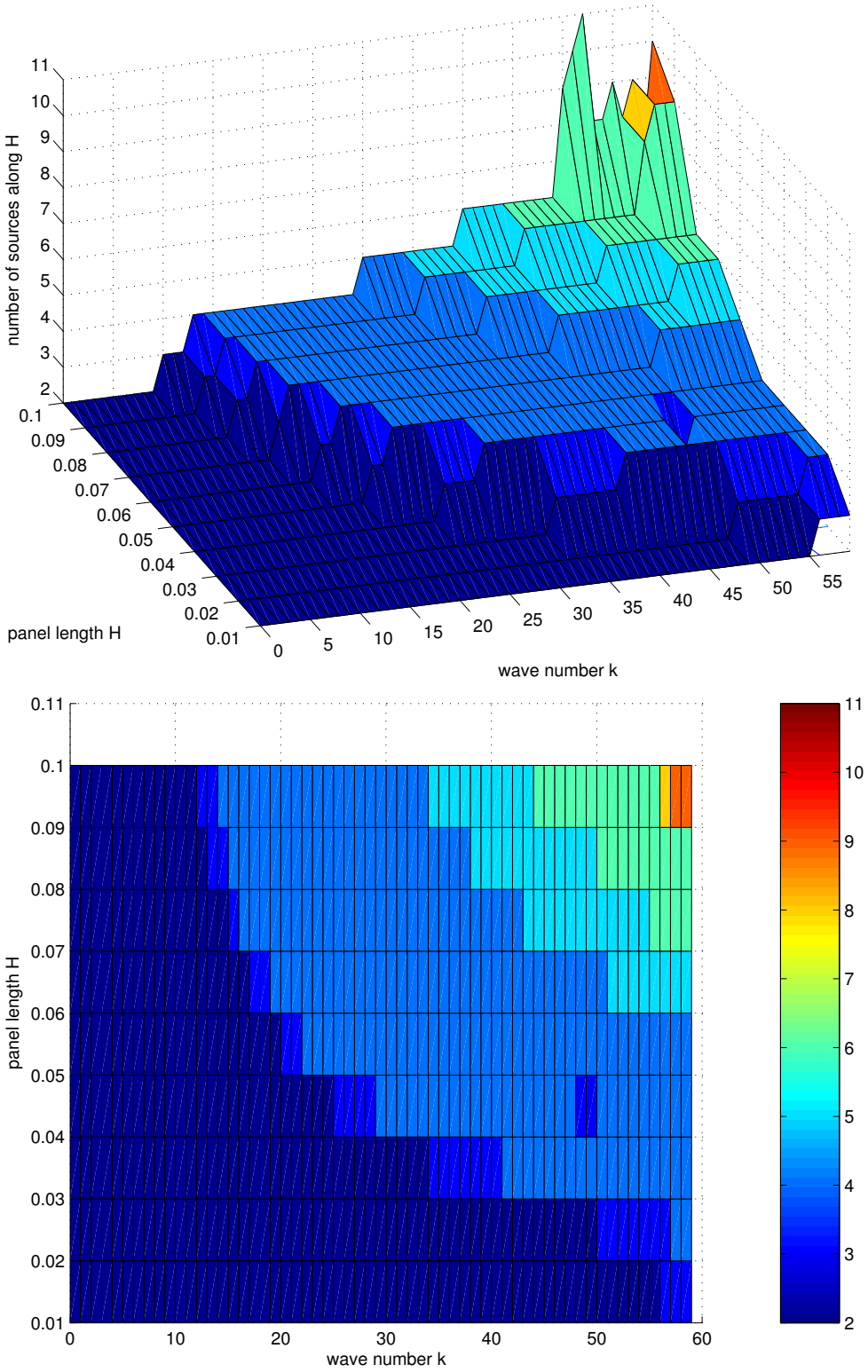


Figure 3.9: Hardest case: Plots 3.7 and 3.8 combined

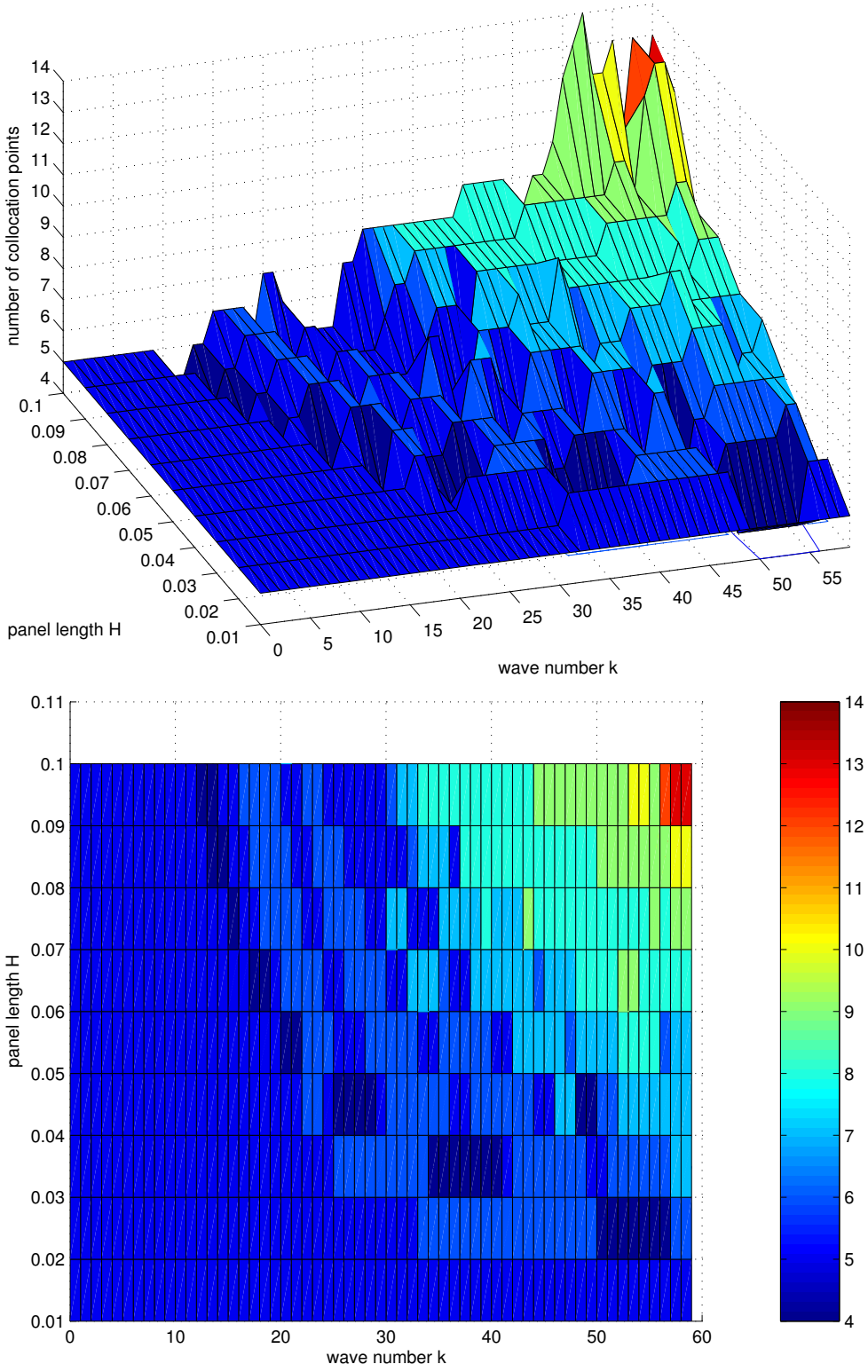


Figure 3.10: Hardest case: The number of collocation points C in dependence of the wave number k and the panel length H to obtain at least $\mathcal{O}(10^{-5})$ accuracy in the field values

for details). We associate the collocation surface with the Kirchhoff surface, and, in view of Kirchhoff's formula, the partition of the solution of the wave equations can be evaluated packetwise at any point outside of $\mathcal{S}_\mathcal{C}$ before adding the packets by superposition principle to the wave function together. Therefore, it is natural to develop and study the concept of the equivalent sources for time-periodic waves first, which we do in the next section. The interest of this discussion is mainly theoretical, however, since, as mentioned in Section 3.3.1, an alternative approach introduced in Section 3.3, based on a certain "continuation method" for Fourier series, can be significantly more efficient in practice.

We close this section with a final remark. Let us assume that a time dependent wave is propagating from a point source into the open three-dimensional space. The point source is located in c_i between the faces D_1 and D_2. In view of equation (2.1), this means that the forcing term takes the form

$$f(\boldsymbol{x},t) \quad = \quad 4\pi\delta(\boldsymbol{x}-\boldsymbol{x}_0)s(t), \tag{3.5}$$

where $\boldsymbol{x}_0$ is the position of the point source and $s(t)$ is an arbitrarily, sufficiently smooth function representing the strength of the source at time t. We recall that the Kirchhoff representation (2.6) solves (2.1). Since the problem is purely outgoing from a point source into the three-dimensional space, the surface integrals (2.8) and (2.9) vanish and the solution (2.6) simplifies to

$$u(\boldsymbol{x},t) = \int_{\mathbb{R}^3} \frac{f(\tilde{\boldsymbol{x}},t-r/c)}{4\pi r} d\tilde{\boldsymbol{x}} = \frac{s(t-r/c)}{r}, \tag{3.6}$$

where r is the distance from $\boldsymbol{x}$ to the location of the source $\boldsymbol{x}_0$. This simple model enables us to evaluate the exact solution to the wave equation very easily without any numerical integration rules. Under the assumption that the field is known on the collocation surface $\mathcal{S}_\mathcal{C}$, our goal is to compute an equivalent source distribution on the two faces D_1 and D_2, which represents the wave outside of $\mathcal{S}_\mathcal{C}$. This is discussed in subsequent sections.

We note that this point source solution can also be used to construct a simple solution to the more complicated case when a scatterer is present: we place a fictitious point source *inside* the scatterer. We know that (3.6) solves the wave equation; thus, if we impose on the scatterer's surface a boundary condition that assumes the form (3.6) for the field values, the exact solution to the problem (2.1)–(2.5) is equation (3.6) and especially, its evaluation does not involve any numerical approximation.

3.2 The time-dependent periodic case

In this section we assume that the continuous wave function $u(\boldsymbol{x},t)$ is T-periodic at any point $\boldsymbol{x}$ outside of c_i, i.e., $u(\boldsymbol{x},t) = u(\boldsymbol{x},t+T)$. Thus, for $\boldsymbol{x} \in \mathbb{R}^3 \backslash c_i$, the field can be expanded by the Fourier series

$$u(\boldsymbol{x},t) \quad = \quad \sum_{n=-\infty}^{\infty} \hat{u}_n(\boldsymbol{x}) e^{\frac{2\pi i}{T} nt}, \tag{3.7}$$

with the Fourier coefficients

$$\hat{u}_n(\boldsymbol{x}) \quad = \quad \frac{1}{T} \int_0^T u(\boldsymbol{x},t) e^{-\frac{2\pi i}{T} nt} dt. \tag{3.8}$$

Substituting (3.7) into the homogeneous wave equation, multiplying by $e^{-2\pi imt/T}$ and integrating over the time domain $[0,T]$ leads to the Helmholtz equations

$$\Delta \hat{u}_n + k_n^2 \hat{u}_n \quad = \quad 0, \quad \text{for } \boldsymbol{x} \in \mathbb{R}^3 \backslash c_i, \tag{3.9}$$

where the wave numbers are defined as

$$k_n \quad = \quad \frac{2\pi}{cT} n. \tag{3.10}$$

In view of the two-face approach, it is clear that monopole equivalent sources $\xi_n^{(l)}$ and dipole equivalent sources $\eta_n^{(l)}$ can be found on the two discs D_l for each frequency index n. The Fourier coefficients $\hat{u}_n$ are thus represented to high-order accuracy by the corresponding *frequency*-dependent equivalent sources. Under the assumption that the Fourier coefficients of the considered waves are rapidly converging to zero as $|n|$ increases, only few frequency modes need to be considered to achieve a very accurate approximation of (3.7). We summarize the procedure in

Algorithm 3.2.1.

1. Transform the given data $u(\boldsymbol{x},t_m)$ at the collocation points $\boldsymbol{x} \in \mathcal{S}_{\mathcal{C}}$ into the Fourier space which gives $\hat{u}_n(\boldsymbol{x})$ for $\{m,n\} \in \{0,\ldots,N-1\}$.

2. Apply Algorithm B.2.1 to the N Fourier coefficient. This results in the equivalent sources $\xi_n^{(l)} \cup \eta_n^{(l)}$ on the panels D_l for $l = 1,2$.

3. Algorithm B.5.1 can now be applied to evaluate $\hat{u}_n(\boldsymbol{x})$ on any Cartesian grid $\tau_F^{(3)}$ outside of $\mathcal{S}_\mathcal{C}$ fast (see Appendix B for details).

4. The inverse Fourier transform in time at $\boldsymbol{x} \in \tau_F^{(3)}$ gives the approximation to $u(\boldsymbol{x}, t_m)$.

3.2.1 Numerical example

As an example, let us assume that the propagation velocity is $c = 1$ and take $s(t)$ (see equation (3.5)) to equal the Gaussian function

$$s(t) \quad = \quad e^{-(t-t_0)^2/\sigma^2}, \tag{3.11}$$

with $\sigma = 0.4$ and $t_0 = 3$. The function's values outside the interval $0 \leq t \leq 6$ are no larger than $\mathcal{O}(10^{-25})$, and thus repeating this function periodically with period 6 gives rise to a discretization of a periodic smooth function of period $T = 6$ up to rounding errors. The function and its discrete Fourier transform with $N = 32$ points are plotted in Figure 3.11. Applying the inverse Fourier transform to the Fourier coefficients $\{\hat{s}_m\}_{m=0}^{N-1}$ gives the approximated values at the N points. The maximum absolute error at these points to the original function $s(t)$ is $3 \cdot 10^{-6}$. To determine how well the N frequencies approximate the function at other points, we can extend the Fourier spectrum by zero padding, i.e., the modified Fourier coefficients $\{\tilde{s}_m\}_{m=0}^{\tilde{N}-1}$ take the form

$$\tilde{s}_m = \begin{cases} \hat{s}_m, & \text{if } m \in \{0, \ldots, N/2\} \ , \\ \hat{s}_{N-\tilde{N}+m}, & \text{if } m \in \{\tilde{N}-1, \ldots, \tilde{N}-N/2+1\} \ , \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

In (3.12) we assume that $\tilde{N} = 2^\alpha N$, where α is a positive integer. Applying an inverse Fourier transform to the zero-padded coefficients samples the approximated function at $\tilde{N}$ equidistant points in the physical time domain. The two lower pictures of Figure 3.11 illustrate this for $\tilde{N} = 2N$. For a fixed N, we compare the zero-padded function with the exact solution at a large number $\tilde{N}$ (= 32,768) of equidistant points and compute the maximum absolute error e_∞. The results along with the maximum errors in the first and second derivatives are reported in Table 3.3. In our numerical code, the indices

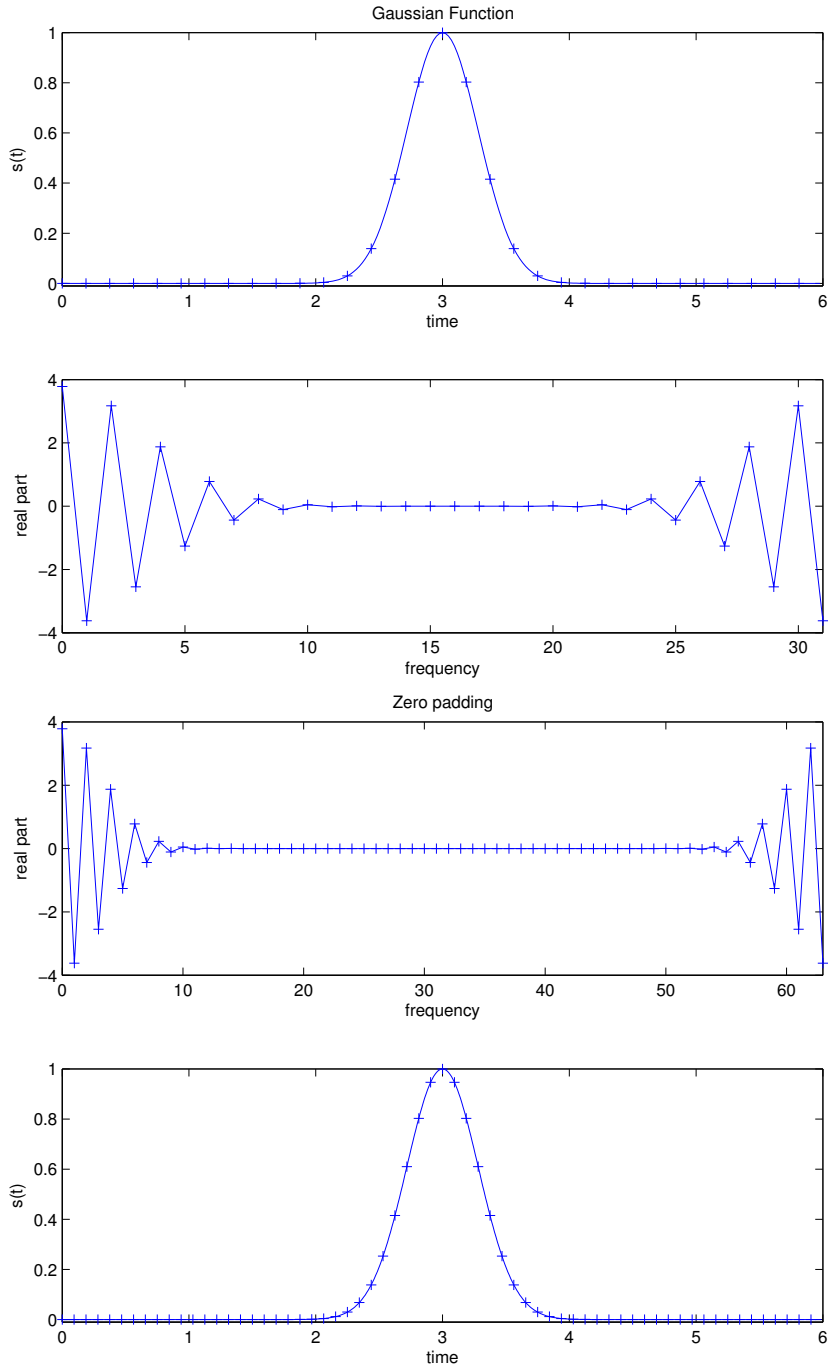


Figure 3.11: The source strength s(t) and zero padding with $N = 32$ and $\tilde{N} = 64$

N	$k_{\max}$	e_∞	$\partial/\partial_t e_\infty$	$\partial^2/\partial_t^2 e_\infty$
16	8.38	0.022	0.183	1.69
32	16.76	$3.50 \cdot 10^{-6}$	$5.88 \cdot 10^{-5}$	$9.94 \cdot 10^{-4}$
64	33.51	$4.44 \cdot 10^{-16}$	$1.81 \cdot 10^{-15}$	$4.22 \cdot 10^{-14}$
128	67.02	$4.44 \cdot 10^{-16}$	$3.96 \cdot 10^{-15}$	$2.00 \cdot 10^{-13}$
256	134.04	$4.44 \cdot 10^{-16}$	$8.66 \cdot 10^{-15}$	$7.32 \cdot 10^{-13}$

Table 3.3: Accuracy of truncating the Fourier series of $s(t)$

larger than $N/2$ correspond to negative frequencies in the Fourier space; therefore, the maximum wave number is $k_{\max} = \pi N/(cT)$. The more points N are used, the larger the resolved frequency spectrum and the better the Fourier expansion approximates the original function. From Table 3.3, it is evident that a number $N = 64$ of modes suffice to approximate the original function along with its first two derivatives almost to machine precision.

In order to demonstrate the properties of approximation of our time-dependent equivalent source approximations, we now select a point source position at $\boldsymbol{x}_0 = [H/2, 0, -H/2]$ and use (3.6) to evaluate the given field at all collocation points $\boldsymbol{x}_C \in \mathcal{S}_\mathcal{C}$. The panel length of the two faces is $H = 0.0625$. The results in Figures 3.8–3.10 suggest to select $S = 5$ and $C = 8$ to approximate the true field with the equivalent sources at least to the order of $\mathcal{O}(10^{-5})$. Algorithm 3.2.1 can now be applied to evaluate the solution at any point outside of the collocation cell. Figure 3.12 displays the field in the time and frequency domains for all $n_C = 296$ collocation points. In Figure 3.13 we plot the exact solution $u(\boldsymbol{x}, t)$ (blue solid) and its numerical approximation $u_h(\boldsymbol{x}, t)$ (green crosses) at the point $\boldsymbol{x} = [0, 0, 0.1876]$. The error in the maximum norm at the point $\boldsymbol{x}$ is defined by

$$e_\infty(\boldsymbol{x}) \equiv \max_{m \in \{0,\dots,N-1\}} |u(\boldsymbol{x}, t_m) - u_h(\boldsymbol{x}, t_m)|. \tag{3.13}$$

Table 3.4 displays errors up to the second time derivative at $\boldsymbol{x} = [0, 0, 0.1876]$ for fixed $H = 0.0625, N = 64, T = 6$ and for different values S and C. The entries of this table confirm that, in accordance with the results displayed in Figures 3.8–3.10, the accuracy of $\mathcal{O}(10^{-5})$ for $S = 5, C = 8$ is achieved.

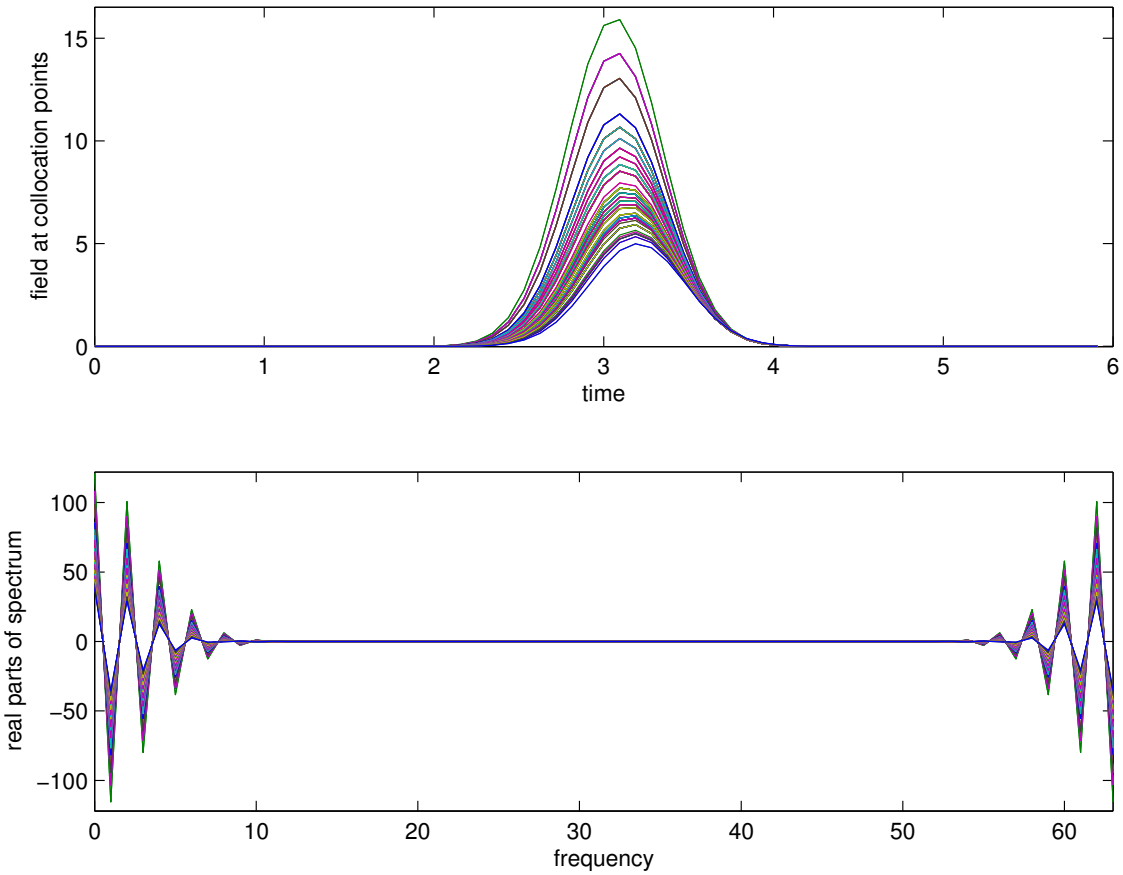


Figure 3.12: The field and its spectrum at all collocation points for $N = 64, C = 8$

$H = 0.0625, N = 64, k_{max} \approx 34, T = 6$						
S	n_S	C	n_C	$e_\infty(\boldsymbol{x})$	$\partial_t e_\infty(\boldsymbol{x})$	$\partial_{tt} e_\infty(\boldsymbol{x})$
3	18	5	98	$2.5 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$	$3.5 \cdot 10^{-2}$
		6	152	$1.7 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$2.4 \cdot 10^{-2}$
		7	218	$1.7 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$2.5 \cdot 10^{-2}$
4	32	5	98	$6.3 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$	$8.9 \cdot 10^{-3}$
		6	152	$1.2 \cdot 10^{-4}$	$3.0 \cdot 10^{-4}$	$1.8 \cdot 10^{-3}$
		7	218	$1.6 \cdot 10^{-4}$	$4.0 \cdot 10^{-4}$	$2.4 \cdot 10^{-3}$
5	50	8	296	$1.3 \cdot 10^{-5}$	$4.3 \cdot 10^{-5}$	$2.5 \cdot 10^{-4}$
		9	386	$5.9 \cdot 10^{-6}$	$2.9 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$
		10	488	$8.0 \cdot 10^{-6}$	$3.3 \cdot 10^{-5}$	$1.9 \cdot 10^{-4}$
6	72	7	218	$3.6 \cdot 10^{-5}$	$8.5 \cdot 10^{-5}$	$4.8 \cdot 10^{-4}$
		8	296	$3.0 \cdot 10^{-6}$	$6.7 \cdot 10^{-6}$	$3.1 \cdot 10^{-5}$
		9	386	$8.0 \cdot 10^{-6}$	$1.8 \cdot 10^{-5}$	$9.8 \cdot 10^{-5}$
7	98	7	218	$2.7 \cdot 10^{-5}$	$7.5 \cdot 10^{-5}$	$3.8 \cdot 10^{-4}$
		8	296	$4.5 \cdot 10^{-6}$	$1.1 \cdot 10^{-5}$	$5.0 \cdot 10^{-5}$
		9	386	$5.4 \cdot 10^{-7}$	$2.0 \cdot 10^{-6}$	$8.3 \cdot 10^{-6}$

Table 3.4: Accuracy of the solution and its first two time derivatives at $\boldsymbol{x} = [0, 0, 0.1876]$

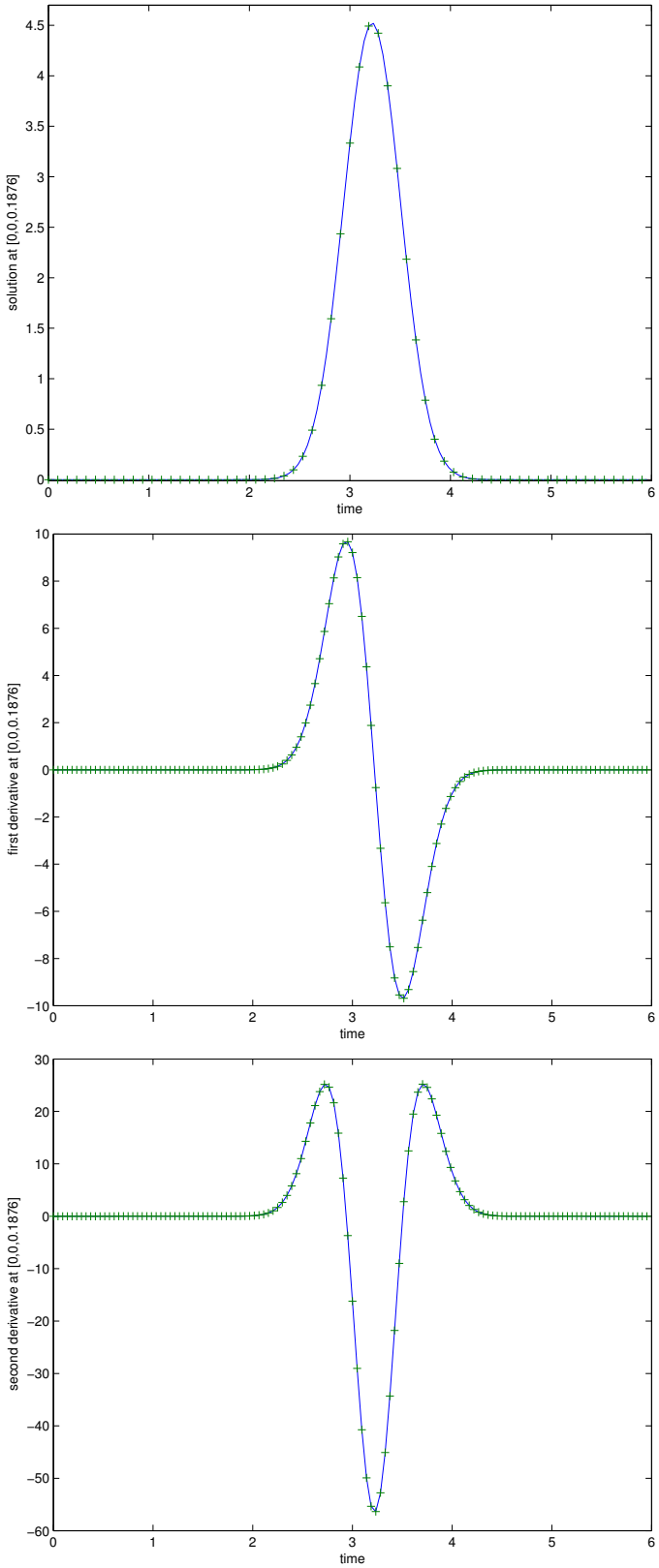


Figure 3.13: Solution (top), its first (center), and second (bottom) time derivative at [0,0,0.1876]. The crosses correspond to the numerical values, the solid line is the exact solution.

3.3 The time-dependent nonperiodic case

The general case does not have periodic sources. To discuss this problem let us consider a wave field that, at time $t = 0$, begins expanding from $\boldsymbol{x}_0$ into the infinite space. On their way, the waves travel through the collocation surface $\mathcal{S}_\mathcal{C}$. For a fixed time $t = T$ we would like to compute equivalent sources which represent the portion of the wave that passed through $\mathcal{S}_\mathcal{C}$. In general, the data on $\mathcal{S}_\mathcal{C}$ is nonperiodic in the interval $t \in [0, T]$, i.e., $u(\boldsymbol{x}_C, 0) \neq u(\boldsymbol{x}_C, T)$. In view of the Kirchhoff representation (2.6) to (2.9), the known portion of the wave on $\mathcal{S}$ in time can be regarded as arising from distribution of sources; certainly the given data are generally not periodic. Yet we cannot directly adapt the technique described in the last section to evaluate the field at a point outside of $\mathcal{S}_\mathcal{C}$, since that method relies on the Fourier expansion (3.7), which in the present context would give rise to the Gibbs phenomenon and, thus, to an extremely poor approximation. Instead, in this section we study a strategy that makes use of partitions of unity and results in wave packets of finite support defined on $\mathcal{S}_\mathcal{C}$. The equivalent source code from the last section can then be applied to the resulting wave packets. But difficulties arise when we use this splitting in connection with nonreflecting open boundary conditions: too many frequency modes may be introduced, which result in high computing costs. An alternative technique that does not suffer from this difficulty shall be discussed in Section 3.3.2.

3.3.1 Partition of unity

The concept of the partition of unity (POU) allows us to split a given wave function $u(\boldsymbol{x}, t)$ at any point $\boldsymbol{x} \in \mathcal{S}_\mathcal{C}$ into K wave packets $\{u_m(\boldsymbol{x}, t)\}_{m=1}^K$, of which the functions associated with the indices $m \in \{2, \ldots, K-1\}$ have finite support in time for $K \geq 3$.

A partition of unity (see reference [12]) may be defined by

$$w_m(t) \quad = \quad \frac{W_m(t)}{\sum_{\ell=1}^K W_\ell(t)}, \quad m = 1, \ldots, K, \tag{3.14}$$

where $W_m(t)$ is given by

$$W_m(t) = \begin{cases} v(t; t_1^{(1)}, t_2^{(1)}), & \text{if } m = 1, \\ v(-t; -t_2^{(K)}, -t_1^{(K)}), & \text{if } m = K, \\ v(-t; -t_2^{(m)}, -t_1^{(m)}) \cdot v(t; t_3^{(m)}, t_4^{(m)}), & \text{if } 1 < m < K \text{ and } K \geq 3, \end{cases} \tag{3.15}$$

with $t_1^{(m)} < t_2^{(m)} < t_3^{(m)} < t_4^{(m)}$ and

$$v(t;t_1,t_2) = \begin{cases} 1, & \text{if } t \le t_1, \\ 0, & \text{if } t \ge t_2, \\ \exp\left(2\frac{t_2-t_1}{t-t_2}\exp(-\frac{t_2-t_1}{t-t_1})\right), & \text{otherwise.} \end{cases} \tag{3.16}$$

Then, we define K wave packets at $\boldsymbol{x} \in \mathcal{S}_{\mathcal{C}}$ by multiplying the total field with the corresponding window function:

$$u_m(\boldsymbol{x},t) \;=\; u(\boldsymbol{x},t)w_m(t). \tag{3.17}$$

In view of Kirchhoff's formula and the linear superposition principle, $u(\boldsymbol{x},t)$ at any point $\boldsymbol{x}$ *outside* of $\mathcal{S}_{\mathcal{C}}$ can be obtained by evaluating each wave packet $u_m(\boldsymbol{x},t)$ separately and then adding all overlapping parts in time together to the total field strength:

$$u(\boldsymbol{x},t) \;=\; \sum_{\ell=1}^{K} u_\ell(\boldsymbol{x},t). \tag{3.18}$$

Computational boundary condition based on POU

In Algorithm 2.3.1 we described a method that acts as a nonreflecting computational boundary condition by making use of the POU along with the equivalent source Algorithm 3.2.1. The alternative approach described in what fallows is not optimal in terms of computing cost, this discussion rather serves as a preliminary step to motivate the equivalent source algorithm as a transparent boundary condition without imposing restrictive requirements on the wave function, such as periodicity. To do this we embed the scatterer into a cube c_i of side length H and assume that all compact supported regions of the initial conditions and the forcing term lie strictly within the collocation box. The cell c_i consists of three pairs of opposite faces, which we denote by $D_1^{(l)} \cup D_2^{(l)}$ for $l \in \{1,2,3\}$. The collocation cube in turn is located in a larger cube with boundary $\mathcal{B}$, and the computational domain Ω for the interior solver in Algorithm 2.3.1 is bounded internally by the surface of the scatterer Γ, and externally by $\mathcal{B}$. The waves,

initially originating from the interior of the collocation cube, eventually expand into the open space. As explained in Section 2.3, the interior algorithm needs to be interrupted once the first wave arrives at $\mathcal{B}$. The collocation surface $\mathcal{S}_\mathcal{C}$ acts as accumulator of the waves which pass through. Without loss of generality, we assume that the first wave arrives at $\mathcal{S}_\mathcal{C}$ at the time $t = 0$ and at $\mathcal{B}$ at the time $t = T$. In practice, we choose $T \leq l_{\min}/c$, where $l_{\min}$ is the minimum distance from $\mathcal{S}_\mathcal{C}$ to $\mathcal{B}$. In general, the data gathered during the time interval $I_0 = [0, T]$ on $\mathcal{S}_\mathcal{C}$ is neither periodic nor has a compact time support, but a wave packet of finite support can be constructed by multiplying the data with a window function w_1, where $t_2^{(1)}$ coincides with T, and $t_1^{(1)}$ is ideally as close to T as possible (see Figures 3.14 and 3.15). The smaller $\Delta t^{(1)} \equiv t_2^{(1)} - t_1^{(1)}$, the more Fourier coefficients are required to resolve w_1 accurately with a Fourier series, but the larger the interval $[0, T - \Delta t^{(1)}]$ in which u_1 coincides with u. Algorithm 3.2.1 can now be applied on the wave packet to compute the field u_1 on $\mathcal{B}$, which arrives there at the time $t = T$ and represents u up to the time $t = 2T - \Delta t^{(1)}$ (see Figure 3.15). We note that the equivalent source algorithm is used for each of the three opposite two-face pairs $D_1^{(l)} \cup D_2^{(l)}, l \in \{1, 2, 3\}$. Each pair's equivalent sources represent the field outside of $\mathcal{S}_\mathcal{C}$ to high order. But an efficient evaluation of the waves is only possible if the plane on which the field needs to be computed is parallel to the two faces. If many points are defined on $\mathcal{B}$, it is advantageous to evaluate the equivalent sources for each one of the three dimensions separately and make use of the fast evaluation technique described in Appendix B.3. Once the boundary data is computed, the interior solver can be used to evaluate $u(\boldsymbol{x}, t)$ in Ω for the time interval $I_1 = [T, 2T - \Delta t^{(1)}]$. On the collocation surface, a subsequent wave packet u_2 is constructed. The parameters of w_2 are selected as follows: $t_1^{(2)} = t_1^{(1)}, t_2^{(2)} = t_2^{(1)}$, and $t_4^{(2)} = 2T - \Delta t^{(1)}$. Even though $t_3^{(2)}$ in principle is free to be chosen, it is most reasonable to set it to equal $t_4^{(2)} - \Delta t^{(1)}$. This way, no additional modes are introduced in the frequency spectrum of w_2. Equivalent sources on $D_1^{(l)} \cup D_2^{(l)}$ are computed with Algorithm 3.2.1 to represent u_2, which evaluate the field data on $\mathcal{B}$. We emphasize an important subtlety: the first time Algorithm 3.2.1 is used, the data of u_1 on $\mathcal{B}$ in the time-interval $[2T - \Delta t^{(1)}, 2T]$ is not considered for the interior computation, because that part corresponds to the transition of the window function w_1 from one to zero. The second time Algorithm 3.2.1 is invoked, we obtain in the same interval $[2T - \Delta t^{(1)}, 2T]$ the corresponding counterpart of the partition of unity, i.e., the superposition $u_1 + u_2$ in $[2T - \Delta t^{(1)}, 2T]$ forms the total scattered field u on $\mathcal{B}$. The domain $[2T, 3T - \Delta t_2^{(2)} - \Delta t^{(1)}]$ does not correspond to a transition part of w_2, and thus, $u_2 = u$ in that interval on the outer boundary. Thus, the data on the boundary $\mathcal{B}$ is known for the interval $[2T - \Delta t^{(1)}, 3T - \Delta t_2^{(2)} - \Delta t^{(1)}]$ (see Figure

3.15). By induction, we can use the same argument: when Algorithm 3.2.1 is invoked the m-th time, the nonreflecting boundary condition is known up to the time $(m+1)T - \sum_{k=1}^{m} \Delta t_2^{(k)}$. The procedure is graphically demonstrated in Figures 3.14 and 3.15. The wave function used in this example is of the form (3.6) with $s(t) = W(t) \cdot \exp(\sin(5.4\pi t - 2.7\pi) - \cos(2\pi t))$, where $W(t)$ is a window function with the parameters $t_1 = 0, t_2 = 0.28, t_3 = 1.575, t_4 = 1.75$.

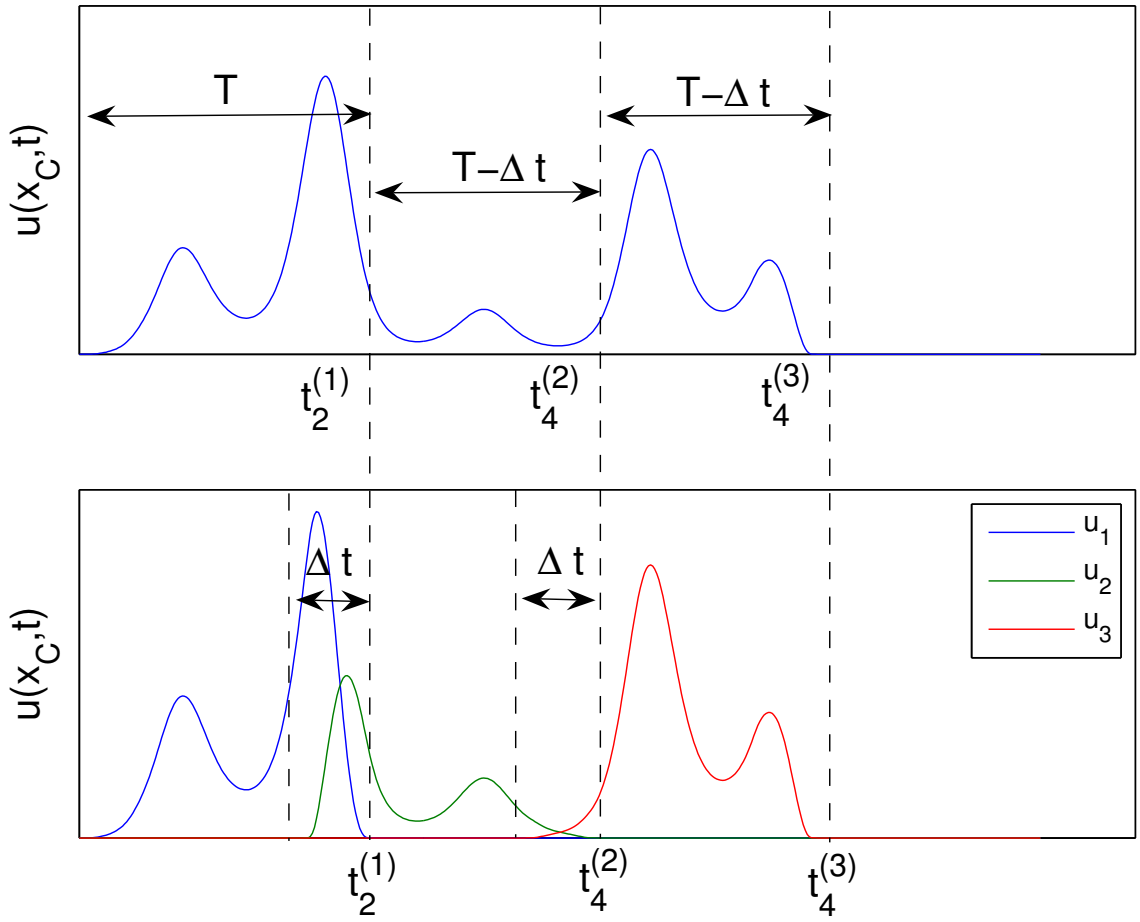


Figure 3.14: Graphical development of the data at the collocation point $\boldsymbol{x}_C$. The interior solver needs to be interrupted at the time $t_2^{(1)}$ (top). The accumulated data at $\boldsymbol{x}_C$ is multiplied by w_1 (bottom). The equivalent source Algorithm 3.2.1 evaluates the wave packet u_1 at $\boldsymbol{x}_\mathcal{B}$, and the interior solver can be applied to compute the field at $\boldsymbol{x}_C$ up to the time $t_4^{(2)}$ (top); compare also with Figure 3.15. The wave packet u_2 is constructed at $\boldsymbol{x}_C$ (bottom), and Algorithm 3.2.1 evaluates the wave u_2 at $\boldsymbol{x}_\mathcal{B}$; using the interior solver again leads to gathered data at $\boldsymbol{x}_C$ up to the time $t_4^{(3)}$ (top), which can be split into u_3 by an appropriate window function w_3 (bottom).

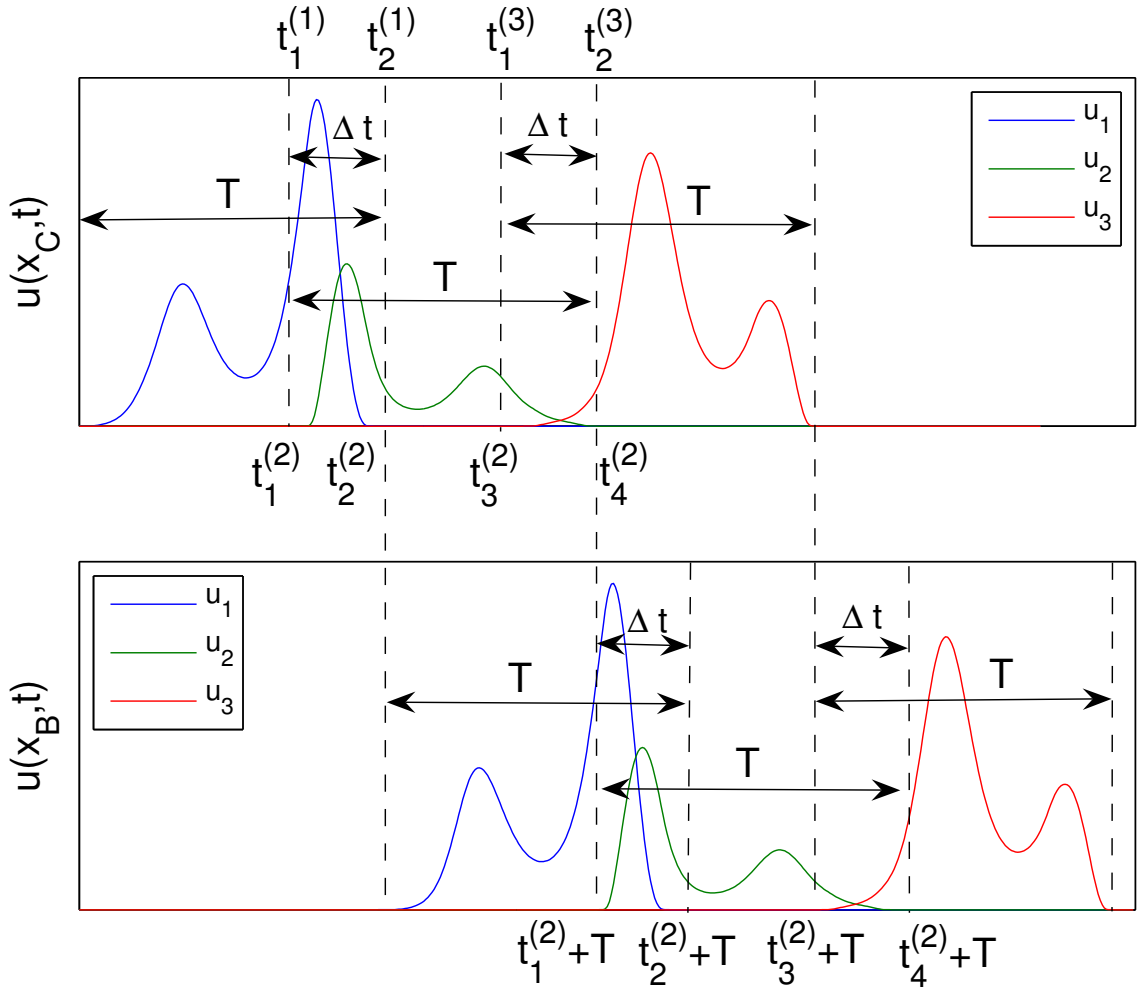


Figure 3.15: Graphical development of evaluating the boundary data at $\boldsymbol{x}_{\mathcal{B}}$: Given the solution from the interior solver at $\boldsymbol{x}_C$ up to the time $t_2^{(1)}$, the wave packet u_1 is constructed (top), which is evaluated at all boundary points $\boldsymbol{x}_{\mathcal{B}}$ with Algorithm 3.2.1 (bottom). Based on the information on $\mathcal{B}$, the interior solver computes u up to time $t_4^{(2)} = t_1^{(2)} + T$. The POU at $\boldsymbol{x}_C$ gives u_2 (top), and the arrival of this wave packet at $\boldsymbol{x}_{\mathcal{B}}$ can be computed with Algorithm 3.2.1 again (bottom). Note that by superposition of u_1 and u_2 at $\boldsymbol{x}_{\mathcal{B}}$, new boundary data for u are known from $t_1^{(2)} + T$ to $t_3^{(2)} + T$: this enables the interior solver to evaluate u up to the time $t_3^{(2)} + T$ in the whole computational domain Ω (top). This leads to an iterative process.

Disadvantage of the POU approach

Even though the POU-based procedure is completely sound and has proved to be useful in our development of computational boundary conditions, we have found that an alternate approach, based on a certain "continuation method" for the resolution of the Gibbs phenomenon, can be significantly more advantageous in terms of computational accuracy and speed. Indeed, the POU we introduced may give rise to additional high-frequency modes which cannot be neglected if the corresponding wave packet is to be resolved to high order. We illustrate this by revisiting the numerical experiment 3.2.1. We adapt exactly the same geometry and parameters as in that example, with one crucial difference: we split the known waves on $\mathcal{S}_\mathcal{C}$ (see Figure 3.12 top) into two packets u_1 and u_2, i.e., $K = 2$ in (3.14)–(3.18). The two transition parameters for w_1 are chosen as $t_1^{(1)} = 3$ fixed and $3 < t_2^{(1)} \leq 6$ variable. Similarly, $t_1^{(2)} = 3$ and $t_2^{(2)} = t_2^{(1)}$. Our purpose in considering different values for $t_2^{(1)}$ is to explore the impact that the choice $\Delta t \equiv \Delta t^{(1)} = \Delta t^{(2)}$ may have on the total number N of frequency modes that are needed to resolve the wave to the same accuracy as in the periodic case: Table 3.3 shows that $N = 64$ modes in the truncated series suffice to approximate the Gaussian function to machine precision, and Table 3.4 confirms that the parameters $S = 5$ and $C = 8$ produce approximations of the field with the time-dependent equivalent sources to at least $\mathcal{O}(10^{-5})$ accuracy for $H = 0.0625$ in the *periodic* case. We use Algorithm 3.2.1 with the same parameters for S and C, and expect to choose $N > 64$ if the accuracy of $\mathcal{O}(10^{-5})$ is not achieved: this would indicate that the window functions introduce additional Fourier coefficients which cannot be neglected for the desired range of convergence. Clearly, the larger Δt, the fewer frequency modes are necessary in the truncated Fourier series. On the other hand, we wish to choose Δt as small as possible. In Table 3.5, we display the results of Algorithm 3.2.1 for the error e_∞ at the point $[0, 0, 3H]^t$, which is the biggest error amongst those arising from the two wave functions u_1 and u_2. As we can see from Table 3.5, if $\Delta t = 3$, the expected accuracy is obtained with the same truncation parameter $N = 64$ as in the periodic case. For $\Delta t \leq 2$, however, $N = 64$ does not suffice to achieve such accuracies: for $\Delta t = 2$ we need $N = 128$; $\Delta t = 1$ requires $N = 256$; and $\Delta t = 0.5$ even demands more than $N = 1024$ to obtain an error of $\mathcal{O}(10^{-5})$ in the field values. If the outer boundary $\mathcal{B}$ is chosen very close to $\mathcal{S}_\mathcal{C}$, these observations highlight a significant issue: if $T = l_{\min}/c$ is small, the interval Δt in which the transition of the window function from one to zero takes place is the smaller, because Δt ought to occupy only a small portion of the domain $[0, T]$. If the field u is not highly oscillatory, the geometry forces us to choose steep window functions which require many more frequency modes N than

$t_1^{(1)}$	$t_2^{(1)}$	N	e_∞
3	6	64	$6.5 \cdot 10^{-5}$
3	5	64	$4.8 \cdot 10^{-4}$
		128	$4.9 \cdot 10^{-5}$
3	4	64	$1.6 \cdot 10^{-2}$
		128	$8.4 \cdot 10^{-4}$
		256	$5.0 \cdot 10^{-5}$
3	3.5	64	$3.5 \cdot 10^{-1}$
		128	$9.3 \cdot 10^{-2}$
		256	$3.7 \cdot 10^{-3}$
		512	$4.7 \cdot 10^{-4}$
		1,024	$4.5 \cdot 10^{-4}$

Table 3.5: The maximum error e_∞ at $\boldsymbol{x} = [0, 0, 3H]^t$ of the two wave packets u_1 and u_2. The equivalent source computation is performed with the parameters $S = 5, C = 8$, and $H = 0.0625$. The number of Fourier modes N are increased if the error fails to be in the order $\mathcal{O}(10^{-5})$.

64 to yield the expected accuracy. In practice, this makes the approach considered above more expensive than it might be. The computational boundary condition based on the POU is thus not as efficient as may be desirable in terms of computing times; therefore we will not explore it further in this thesis.

Instead of exploring the POU further, we propose to apply a Fourier continuation method (see [16, 17]) to the wave functions on $\mathcal{S}_\mathcal{C}$: the time domain $[0, T]$ is extended into a suitable interval $[t_0, t_1]$, in which we define periodic functions that converge to a high accuracy to the initial waves in $[0, T]$, and Algorithm 3.2.1 is then used on the resulting periodic functions. We discuss the continuation method in the next section and modify it appropriately for our needs.

3.3.2 Continuation method

In this section we describe a certain continuation method [16, 17] for the resolution of the Gibbs phenomenon. Let us consider a function $u(t)$ in the interval $[0,T]$, and let $t_n = (n-1)dt$ for $n \in \{1,\dots,N\}$ denote the discrete sample location in $[0,T]$. We assume that the function u is only known at the N discrete points, and the entries of the vector $\boldsymbol{u} = [u_1,\dots,u_N]^t$ are the given function values at these points. We seek a T_e-periodic function $v(t)$ that is defined in $[t_0,t_1] \supset [0,T]$, i.e., $T_e = t_1 - t_0$, such that it matches the function u in its domain of definition $[0,T]$ to a high accuracy. We express $v(t)$ by

$$v(t) \;=\; \sum_{m=1}^{M} \hat{v}_m e^{-\frac{2\pi i}{T_e}(m-1)t}, \tag{3.19}$$

and notice that $v(t)$ can be sampled at the points $t_n \in [0,T]$. In matrix notation, this can be written as

$$\boldsymbol{v} \;=\; \boldsymbol{A}\hat{\boldsymbol{v}}, \tag{3.20}$$

where $\boldsymbol{v} = [v(t_1),\dots,v(t_N)]^t, \hat{\boldsymbol{v}} = [\hat{v}_1,\dots,\hat{v}_M]^t$ and $\boldsymbol{A}_{nm} = e^{-2\pi i(m-1)t_n/T_e}$. We are looking for the continuation coefficients $\hat{\boldsymbol{v}}$ which approximate the initial function values $\boldsymbol{u}$ the best in the L_2-norm, i.e.,

$$\min_{\hat{\boldsymbol{v}}} \|\boldsymbol{u} - \boldsymbol{A}\hat{\boldsymbol{v}}\|_2. \tag{3.21}$$

The solution to (3.21) can be obtained using a singular value decomposition (SVD). As shown in [16, 17], this approach is highly accurate: the solution converges super-algebraically fast to the given function $u(t)$, that is, the error is of the order $1/M^{s-1}$ if $v \in C^s$. If $s = \infty$, the error converges faster than any power of $1/M$. We refer to [16, 17] for further details.

A numerical example is given in Figure 3.16 for the function $u(t) = s(t-d)/d$ with $T = 3$ and $T_e = 6$. The fixed distance d is equal to the absolute value of the vector $[H/2, 0, -H/2]^t$, where $H = 0.0625$. The Gaussian function $s(t)$ is defined in (3.11) and its parameters are chosen as $\sigma = 0.4$ and $t_0 = 3$. Table 3.6 displays the error of $u(t)$ and $v(t)$ in the maximum norm at $31,000$ equidistant points in $[0,3]$ for various modes M and sampling points N. The table demonstrates the high accuracy of the method. Oversampling (i.e., $N > M$) seems to be particularly useful to obtain a better approximation for a fixed

number of frequencies M: for example, the table shows that doubling M and N from 32 to 64 improves the accuracy of the function by a factor of 10^4.

Singular Value Decomposition

It is pointed out in [78] that the SVD may be misused for ill-conditioned matrices, if small singular values are not adequately zeroed out. The condition numbers (ratio of largest to smallest eigenvalues) for the problems discussed in the previous section range from $\mathcal{O}(10^{11})$ to $\mathcal{O}(10^{16})$: the problem is rather ill conditioned. Computing $\hat{\boldsymbol{v}}$ directly by means of the relation $\hat{\boldsymbol{v}} = \boldsymbol{V} \cdot \boldsymbol{S}^{-1} \cdot \overline{\boldsymbol{U}}$ (where $\boldsymbol{U} \in \mathbb{C}^{N \times M}$ and $V \in \mathbb{C}^{M \times M}$ are unitary matrices) may give unsatisfactory results. The matrix $\boldsymbol{S} \in \mathbb{C}^{M \times M}$ is diagonal with the singular values as its entries. To avoid any difficulties arising from use of the singular value decomposition, we set the (i,i)-entry of the diagonal matrix $\boldsymbol{S}^{-1}$ as the inverse of the singular value s_i only if the relation $\frac{1}{\epsilon} \geq \frac{s_{\max}}{s_i}$ is satisfied, where the value of ϵ is in the range of the floating-point precision and $s_{\max}$ denotes the maximum singular value; if this condition does not hold, we set the entry to zero:

$$\{\boldsymbol{S}^{-1}\}_{ii} = \begin{cases} \frac{1}{s_i}, & \text{if } \frac{1}{\epsilon} \geq \frac{s_{\max}}{s_i} \\ 0, & \text{else.} \end{cases}, \tag{3.22}$$

For example, in Matlab tests (where $\epsilon \approx 2.2204 \cdot 10^{-16}$), we indeed notice that the modified definition (3.22) gives rise to a more accurate result for the least-square solution once the condition number of $\boldsymbol{A}$ reaches the range of $\mathcal{O}(10^{16})$. In the case $M = N = 64$, use of a direct SVD results in errors of $e_\infty, \partial/\partial_t e_\infty$ and $\partial^2/\partial_t^2 e_\infty$ to $6.3 \cdot 10^{-9}, 1.4 \cdot 10^{-6}$, and $2.1 \cdot 10^{-4}$, respectively. Errors are smaller by a factor of 10 by making use of (3.22). The differences are even more extreme for larger condition numbers. The solution of the continuation method is plotted in Figure 3.16 for $N = 128$ and $M = 64$.

The behavior of the function in the extended domain can vary significantly for different N and M, as is demonstrated in Figure 3.17. This is due to the fact that the series coefficients, produced by the continuation method, cannot be bounded by the maximum absolute value of the function approximated. This may result in large values in the extended domain, as can be observed in the upper two plots of Figure 3.17. We discuss an alternative approach in Appendix C, which may be more useful in applications where the large continuation coefficients could cause difficulties. That approach is based

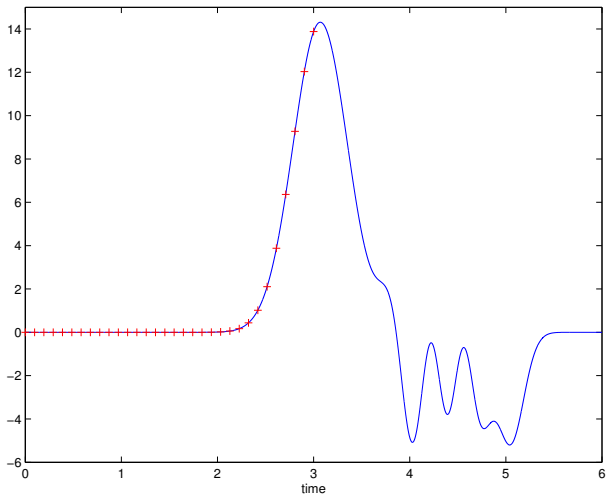


Figure 3.16: The continuation method is applied to the red discrete data in $[0, 3]$ to the extended domain $[0, 6]$, resulting in the blue periodic solution.

on the observation that the Chebyshev coefficients are nicely bounded in terms of the maximum value of the function approximated. The periodicity of the function is achieved by extending the Chebyshev polynomials outside of the standard domain of definition and insuring a compact support by multiplying the basis functions by appropriate window functions (see Appendix C for more details). In our context, the fact that the Fourier coefficients cannot be bounded by the absolute maximum value of the function does not lead to any difficulties, and we adopt this approach to obtain a periodic wave function on $\mathcal{S}_{\mathcal{C}}$.

Finally, we point out that a modified approach, recently introduced in [15] can bypass the large condition number issues. But, in the present context, the described approach is completely satisfactory.

M	N	e_∞	$\partial/\partial_t e_\infty$	$\partial^2/\partial_t^2 e_\infty$
32	32	$4.4 \cdot 10^{-6}$	$4.6 \cdot 10^{-4}$	$3.3 \cdot 10^{-2}$
32	64	$1.0 \cdot 10^{-8}$	$1.8 \cdot 10^{-6}$	$2.1 \cdot 10^{-4}$
32	128	$1.7 \cdot 10^{-9}$	$4.1 \cdot 10^{-7}$	$6.2 \cdot 10^{-5}$
32	256	$1.1 \cdot 10^{-9}$	$3.8 \cdot 10^{-7}$	$5.7 \cdot 10^{-5}$
64	64	$7.3 \cdot 10^{-10}$	$1.6 \cdot 10^{-7}$	$2.5 \cdot 10^{-5}$
64	128	$3.6 \cdot 10^{-13}$	$1.4 \cdot 10^{-10}$	$3.5 \cdot 10^{-8}$
64	256	$4.6 \cdot 10^{-14}$	$1.2 \cdot 10^{-11}$	$3.6 \cdot 10^{-9}$
64	512	$3.0 \cdot 10^{-14}$	$1.3 \cdot 10^{-12}$	$4.7 \cdot 10^{-10}$
128	128	$3.2 \cdot 10^{-9}$	$1.4 \cdot 10^{-6}$	$4.4 \cdot 10^{-4}$
128	256	$3.7 \cdot 10^{-13}$	$2.8 \cdot 10^{-10}$	$1.4 \cdot 10^{-7}$
128	512	$4.1 \cdot 10^{-14}$	$1.7 \cdot 10^{-11}$	$1.4 \cdot 10^{-8}$

Table 3.6: Accuracy of the continuation method

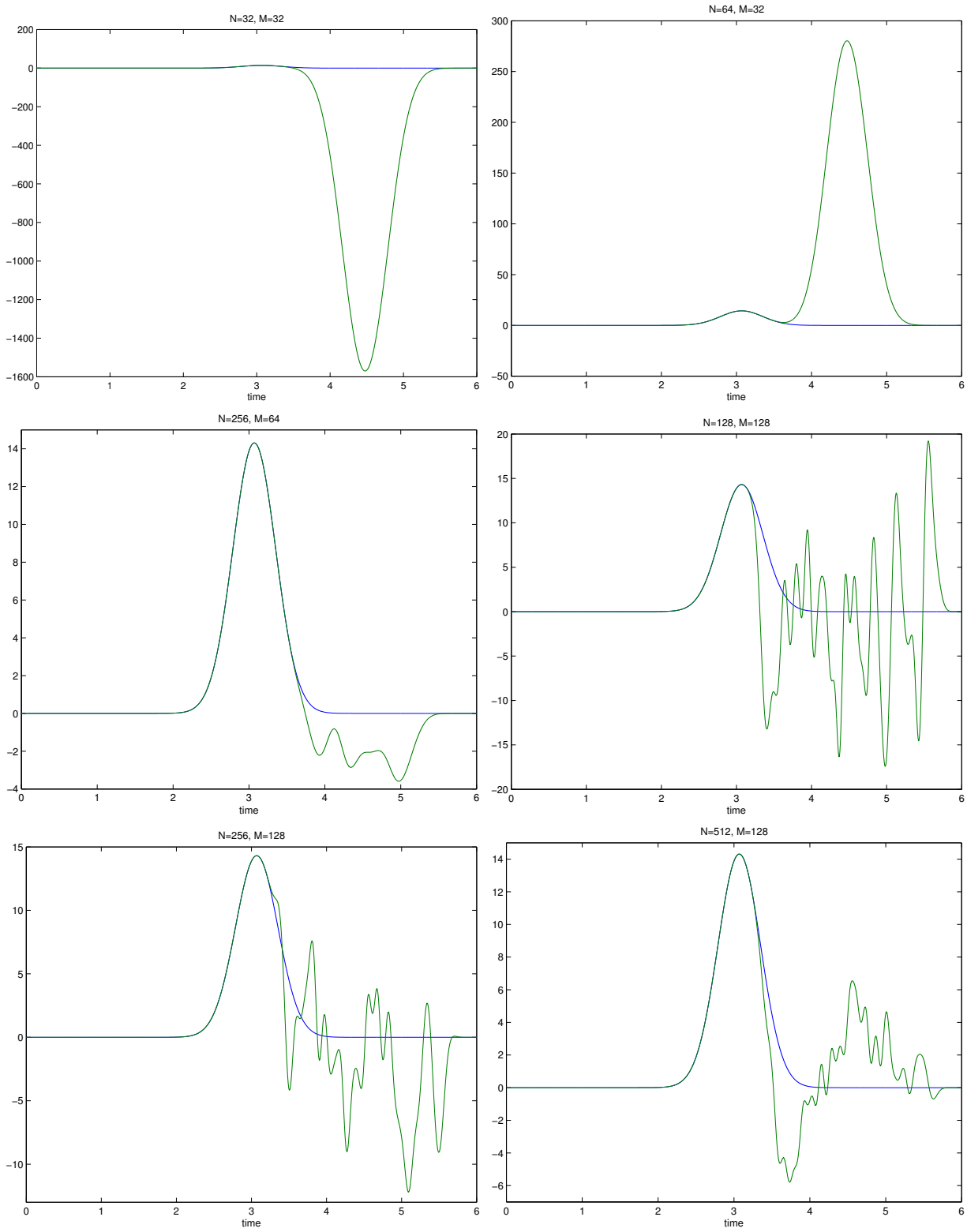


Figure 3.17: The continuation method for different value of N and M. Note the significant differences of the functions in the extension domains. The blue graph corresponds to the exact solution. By construction, the approximated solution (green) matches the blue to a high accuracy in the domain $[0, 3]$ (compare with Table 3.6).

Formulation for real-valued functions

The Fourier continuation method as described by (3.19)–(3.21) has a drawback for purely real functions $u(t)$. As is well known, for a real-valued periodic function the real part of the Fourier coefficients is symmetric and the imaginary part antisymmetric. The least-square solution (3.21) does not insure that these properties are satisfied. As a result, we cannot obtain $v(t)$ by an efficient inverse Fourier transform which takes advantages of the symmetries in the Fourier coefficients by storing only half of the real and imaginary Fourier data. This difficulty can easily be bypassed by reformulating (3.20) into a linear system for half of the real and imaginary coefficients only, which results in solving the $N \times (M-1)$ linear system

$$\boldsymbol{u} \;=\; \hat{v}_0 \cdot \boldsymbol{e}_1 + \boldsymbol{A}_c \cdot \hat{\boldsymbol{v}}^R + \boldsymbol{A}_s \cdot \hat{\boldsymbol{v}}^I \tag{3.23}$$

in the least-square sense for

$$\hat{\boldsymbol{v}} \;=\; \begin{bmatrix} \hat{v}_0 \\ \hat{\boldsymbol{v}}^R \\ \hat{\boldsymbol{v}}^I \end{bmatrix}. \tag{3.24}$$

In (3.23), we use the notation $\boldsymbol{A}_c^{j,k} = 2\cos(\frac{2\pi}{N}jk)$, $\boldsymbol{A}_s^{j,k} = 2\sin(\frac{2\pi}{N}jk)$, $\hat{\boldsymbol{v}}^R = [\hat{v}_1^R, \ldots, \hat{v}_{M/2-1}^R]^t$, $\hat{\boldsymbol{v}}^I = [\hat{v}_1^I, \ldots, \hat{v}_{M/2-1}^I]^t$, and $\boldsymbol{e}_1 = [1, \ldots, 1]^t$. System (3.23) has two significant advantages over (3.20). First, it makes full use of the underlying symmetries. As a result, a purely real linear system which reduces the size by a factor of two in each dimension needs to be solved. Second, an inverse fast Fourier transform (FFT) can now be applied to $\hat{\boldsymbol{v}}$ in order to obtain $\boldsymbol{v}$ accurately.

Table 3.7 demonstrates that the method gives rise to high-order accuracy. An inverse FFT is used to compute the discrete function values in the time domain. Zero padding is performed so that the extended solution is sampled at $2^{16} = 65,536$ points in the extended domain. Notice that the expansions in Table 3.7 are computed with $M-1$ modes, which should be taken into account when comparing the results with Table 3.6.

M	N	e_∞	$\partial/\partial_t e_\infty$	$\partial^2/\partial_t^2 e_\infty$
32	32	$7.1 \cdot 10^{-6}$	$7.4 \cdot 10^{-4}$	$5.4 \cdot 10^{-2}$
32	64	$5.0 \cdot 10^{-8}$	$8.6 \cdot 10^{-6}$	$1.0 \cdot 10^{-3}$
32	128	$9.0 \cdot 10^{-9}$	$2.2 \cdot 10^{-6}$	$3.5 \cdot 10^{-4}$
32	256	$6.0 \cdot 10^{-9}$	$2.0 \cdot 10^{-6}$	$3.1 \cdot 10^{-4}$
64	64	$2.6 \cdot 10^{-10}$	$5.6 \cdot 10^{-8}$	$8.5 \cdot 10^{-6}$
64	128	$7.6 \cdot 10^{-14}$	$3.1 \cdot 10^{-11}$	$7.9 \cdot 10^{-9}$
64	256	$3.2 \cdot 10^{-14}$	$3.9 \cdot 10^{-12}$	$1.2 \cdot 10^{-9}$
64	512	$2.8 \cdot 10^{-14}$	$1.2 \cdot 10^{-11}$	$2.8 \cdot 10^{-9}$
128	128	$1.0 \cdot 10^{-9}$	$4.5 \cdot 10^{-7}$	$1.4 \cdot 10^{-4}$
128	256	$1.9 \cdot 10^{-13}$	$1.3 \cdot 10^{-10}$	$6.6 \cdot 10^{-8}$
128	512	$8.3 \cdot 10^{-14}$	$3.9 \cdot 10^{-11}$	$2.8 \cdot 10^{-8}$

Table 3.7: Accuracy of the modified continuation method

3.3.3 Time buffer

Once the field on $\mathcal{S}_\mathcal{C}$ is periodically extended with the continuation method as described in the last section, this data can be accurately represented by a Fourier series. In principle, we can use Algorithm 3.2.1 and proceed as in the periodic case to evaluate the field outside of $\mathcal{S}_\mathcal{C}$, but caution is needed. To get started, let us revisit the simple geometry we described in Section 3.3.1 where we developed the POU-based approach to act as a nonreflecting computational boundary condition: the scatterer is positioned inside of the cube c_i of side length H. The faces of c_i consist of the three pairs $D_1^{(l)} \cup D_2^{(l)}, l \in \{1,2,3\}$. Further, the collocation box with surface $\mathcal{S}_\mathcal{C}$ is embedded in a larger cube whose surface $\mathcal{B}$ acts as the open boundary, see Figure 3.18 left. As before, we assume that the interior scheme solves for the field in the time interval $I_0 = [0,T]$, and needs to be interrupted at time T, since the first nonvanishing contribution of the waves arrives at $\mathcal{B}$. The goal here is to describe the usage of the continuation method along with Algorithm 3.2.1 to obtain the boundary data on $\mathcal{B}$ for the interval $I_1 = [t_{\min}, T + t_{\min}]$, where $t_{\min} = r_{\min}/c$ and $r_{\min}$ denotes the minimum distance from $\mathcal{S}_\mathcal{C}$ to $\mathcal{B}$. We recall that at any point $\boldsymbol{x} \in \mathcal{B}$, the solution is given by

$$\begin{aligned} u(\boldsymbol{x},t) &= \frac{1}{4\pi}\int_{\mathcal{S}_\mathcal{C}} \frac{1}{r}\frac{\partial u}{\partial \nu(\boldsymbol{y})}(\boldsymbol{y}, t-t_r)\, ds(\boldsymbol{y}) + \\ &\quad \frac{1}{4\pi}\int_{\mathcal{S}_\mathcal{C}} \frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{r^2}\left\{\frac{u(\boldsymbol{y},t-t_r)}{r} + \frac{1}{c}\frac{\partial u}{\partial t}(\boldsymbol{y},t-t_r)\right\} ds(\boldsymbol{y}). \end{aligned} \tag{3.25}$$

where $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{y}, r = |\boldsymbol{r}|, t_r = r/c$ and $\boldsymbol{\nu}(\boldsymbol{y})$ is the outer unit normal of $\mathcal{S}_\mathcal{C}$ at $\boldsymbol{y}$. While we do not use (3.25) directly in our algorithm, the formula remains useful for theoretical purposes. When approximating the given data at $\boldsymbol{y} \in \mathcal{S}_\mathcal{C}$ in the time interval I_0 by means of the continuation method, the question arises as to how to define the extended domain of definition. To this end, we introduce the time buffer $t_{BF} \geq 0$ and denote the extended interval by $[-t_{BF}, T] \cup [T, 2T + t_{BF}]$. The true data of the field is defined in the first interval $[-t_{BF}, T]$, while the artificial part that arises from the Fourier continuation method is located in the second interval $[T, 2T + t_{BF}]$ (see Figure 3.18). We assume here that the past needed information is known on $\mathcal{S}_\mathcal{C}$. Our goal is to determine the conditions on t_{BF}: can the time buffer just be set to zero, or do some restrictions apply? After the continuation method is employed, the data

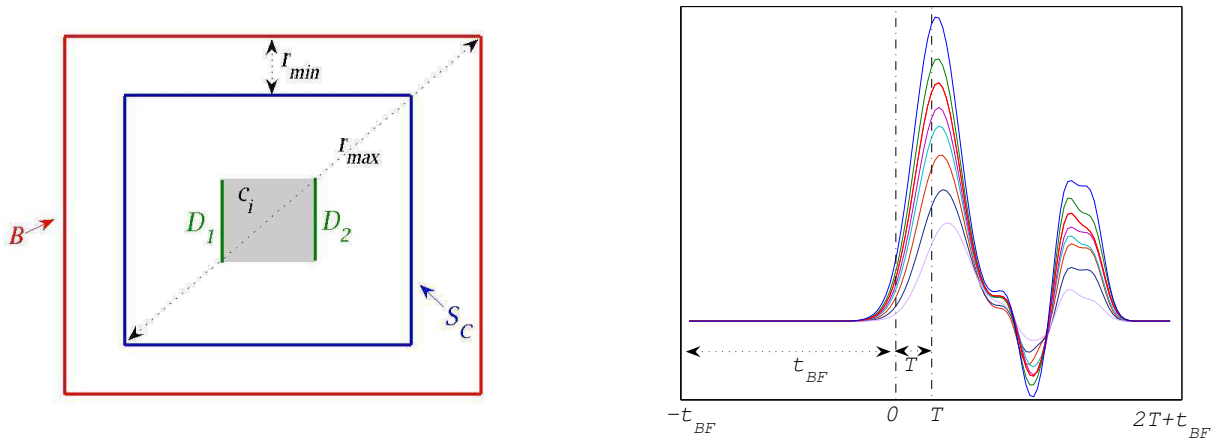


Figure 3.18: Left: Geometry of the problem. Right: Fourier continuation method applied to the source-strengths. The time buffer t_{BF} is necessary to insure that the total induced field does not get corrupted.

on $\mathcal{S}_\mathcal{C}$ is T_e-periodic, i.e., $u(\boldsymbol{y}, t) = u(\boldsymbol{y}, t + T_e)$, where $T_e = 2T + 2t_{BF}$. Focusing on formula (3.25), the integrands can be regarded as a product of $\mathcal{P}u(\boldsymbol{y}, t - t_r)$ with a scalar function (which depends on $\boldsymbol{r}$), where the linear operator $\mathcal{P}$ is the identity, the partial time-derivative, or the projection of the gradient on $\boldsymbol{\nu}$, and the scalar function is continuous for $r > 0$. Careful consideration of (3.25) yields the conclusion that at $(\boldsymbol{x}, t) \in \mathcal{B} \times [-t_{BF}, T] \cup [T, 2T + t_{BF}]$, the solution u is corrupted in certain regions of both time intervals $[-t_{BF}, T]$ and $[T, 2T + t_{BF}]$. Indeed, let us consider the two extreme cases when the infinitesimal wave travels the shortest and longest distance from $\mathcal{S}_\mathcal{C}$ to $\mathcal{B}$, respectively. Denoting by $r_{\min}$ the minimum and by $r_{\max}$ the maximum distance from $\mathcal{S}_\mathcal{C}$ to $\mathcal{B}$, it takes the corresponding infinitesimal waves $t_{\min} = r_{\min}/c$ and $t_{\max} = r_{\max}/c$ to arrive at $\mathcal{B}$, respectively. Focusing first on the wave which travels the minimum distance, we notice that the *periodic* functions $\mathcal{P}u(\boldsymbol{y}, t)$ in (3.25) are *retarded* by

$t_{\min}$; therefore, the data associated with the artificial extension of the continuation functions which is initially in $[T, 2T + t_{BF}]$ on $\mathcal{S}_\mathcal{C}$ can be found in the domain $[-t_{BF}, -t_{BF} + t_{\min}] \cup [t_{\min} + T, 2T + t_{BF}]$. This is graphically depicted in the upper portion of Figure 3.19: the gray shaded areas are associated with this domain. Any other infinitesimal wave that starts from $\mathcal{S}_\mathcal{C}$ arrives at this point on $\mathcal{B}$ later: the corresponding distance traveled is $r > r_{\min}$ and thus the boundaries of the gray shaded regions move to the right for $t_r > t_{\min}$. Under the assumption that t_{BF} is sufficiently large, the gray region from the left will not reach the time $t_{\min} + T$. This means that there is indeed an interval which is not corrupted by the artificial continuation extension function, and, the time $t_{\min} + T$ is the fixed boundary on the right of this domain (see the blue solid line in upper Figure 3.19).

The same argument for the infinitesimal wave which travels the longest distance from $\mathcal{S}_\mathcal{C}$ to $\mathcal{B}$ reveals that the region which is not corrupted from the left is at $-t_{BF} + t_{\max}$ (see middle picture in Figure 3.19) and the gray regions, associated with corruption in the wave function, is the union $[-t_{BF}, -t_{BF} + t_{\max}] \cup [t_{\max} + T, 2T + t_{BF}]$.

Therefore, the time interval for which the wave is not corrupted anywhere in $\mathcal{B}$ is the interval $[-t_{BF} + t_{\max}, t_{\min} + T]$. This domain is depicted graphically by overlapping the two extreme cases described: this results in the lower Figure 3.19. Any other infinitesimal wave from $\mathcal{S}_\mathcal{C}$ has a traveling time t_r which satisfies $t_{\min} < t_r < t_{\max}$, and thus its nonpolluted region will lie always within the interval $\tilde{I}_1 = [-t_{BF} + t_{\max}, t_{\min} + T]$, which is bounded by the two blue solid lines in the bottom of Figure 3.19. We recall that we wish to compute the wave on $\mathcal{B}$ for the interval I_1 and thus $\tilde{I}_1 \supseteq I_1$ must hold. The upper bound of both intervals is $t_{\min} + T$, but the lower bound of I_1 is $t_{\min}$ and of $\tilde{I}_1$ is $t_{\max} - t_{BF}$, respectively. It follows that $t_{\min} \geq t_{\max} - t_{BF}$, which can be rewritten as

$$t_{BF} \geq \frac{r_{\max} - r_{\min}}{c}. \tag{3.26}$$

The condition (3.26) shows that the time buffer t_{BF} is dependent on the geometry of the collocation surface $\mathcal{S}_\mathcal{C}$ and the artificial boundary $\mathcal{B}$ and must be bigger than zero.

We thus have found a preliminary algorithm that solves the scattering problem defined in an unbounded domain. This algorithm successfully incorporates the equivalent source Algorithm 3.2.1 as a crucial part into a computational boundary evaluator. The technique described here proves vastly superior to the POU-based methodology in terms of computational cost. In Algorithm 3.3.1, we summarize the main step of the nonreflecting boundary condition solver given the data on $\mathcal{S}_\mathcal{C}$. The overall scattering

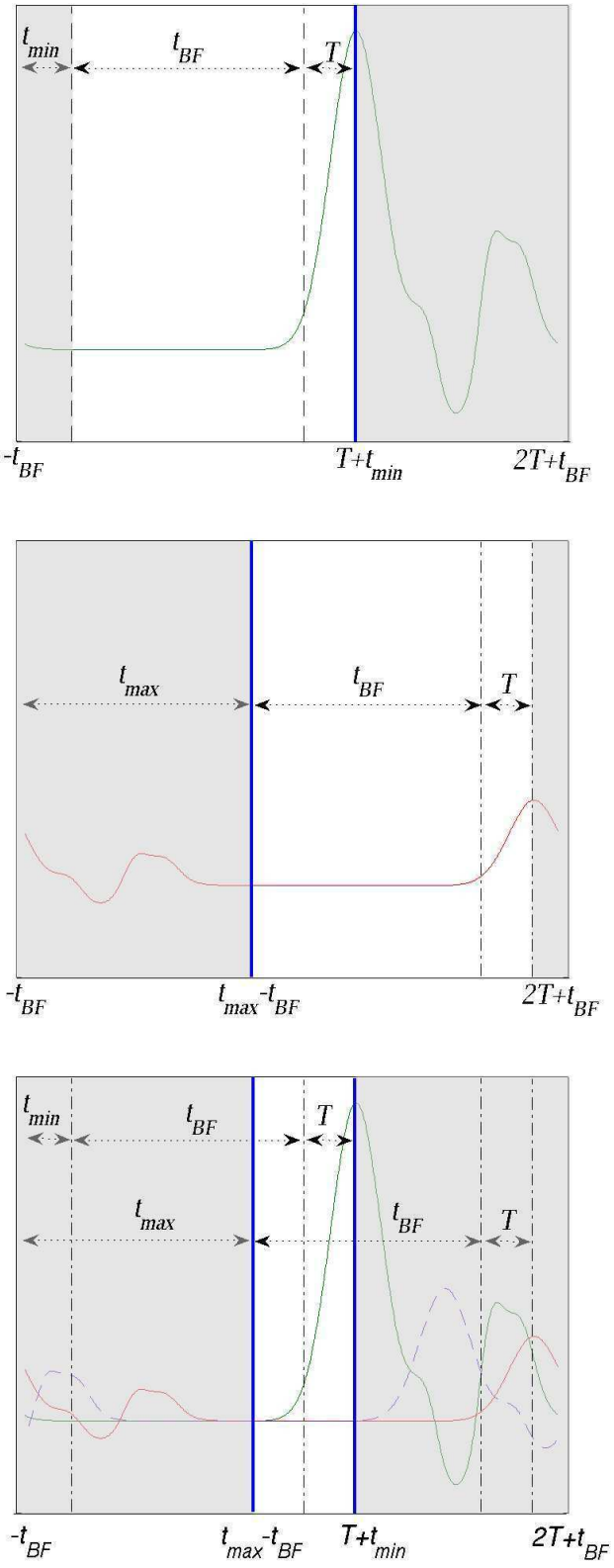


Figure 3.19: Top: The infinitesimal wave which travels the minimum distance $r_{\min}$ from $\mathcal{S}_{\mathcal{C}}$ to $\mathcal{B}$. Middle: The infinitesimal wave that travels the maximum distance $r_{\max}$ from $\mathcal{S}_{\mathcal{C}}$ to $\mathcal{B}$. Bottom: Overlapping the two extreme cases gives the validity of the domain of the field at $\mathcal{B}$.

algorithm is to be coupled with an interior scheme as explained in 2.3.1 and 3.3.1.

Algorithm 3.3.1.

1. Given data on $\mathcal{S}_C$ up to time T, compute the field on $\mathcal{B}$ for times t in the interval $I_1 = [t_{\min}, T+t_{\min}]$. Based on the geometry of $\mathcal{S}_C$ and $\mathcal{B}$, along with the condition (3.26) on the time buffer t_{BF}, we determine the extended domain $[-t_{BF}, 2T+t_{BF}]$, and use the continuation method in this interval to obtain the Fourier continuation functions.

2. At this point we (can and do) apply the periodic case—Algorithm 3.2.1. The solution is valid in the interval $\tilde{I_1} = [-t_{BF} + t_{\max}, t_{\min} + T]$. Notice that if we define the smallest possible time buffer, i.e., $t_{BF} = t_{\max} - t_{\min}$, $\tilde{I_1}$ shrinks to $I_1 = [t_{\min}, T + t_{\min}]$.

In the next section, we demonstrate the performance of Algorithm 3.3.1 on a numerical example. In the following chapter, we then design an improved methodology for the nonreflecting computational boundary condition solver that enables us to take advantage of Algorithm 3.2.1 in an efficient manner, thus leading to our overall proposed nonreflecting boundary condition.

3.3.4 Numerical experiments

We demonstrate the performance of Algorithm 3.3.1 by applying it on the example introduced in Section 3.2.1: we recall that the parameters in the periodic case we considered are $H = 0.0625, S = 5, C = 8, N = 64, c = 1$ to obtain an accuracy of $\mathcal{O}(10^{-5})$. Here, we assume that nonperiodic data on $\mathcal{S}_\mathcal{C}$, shown in the upper portion of Figure 3.20, is given in the interval $[0, 3]$. Below is a plot of the continuation functions defined in the extended time domain $[0, 6]$. Further, we select $t_{BF} \equiv t_{\max} - t_{\min}$ and plot the solution at

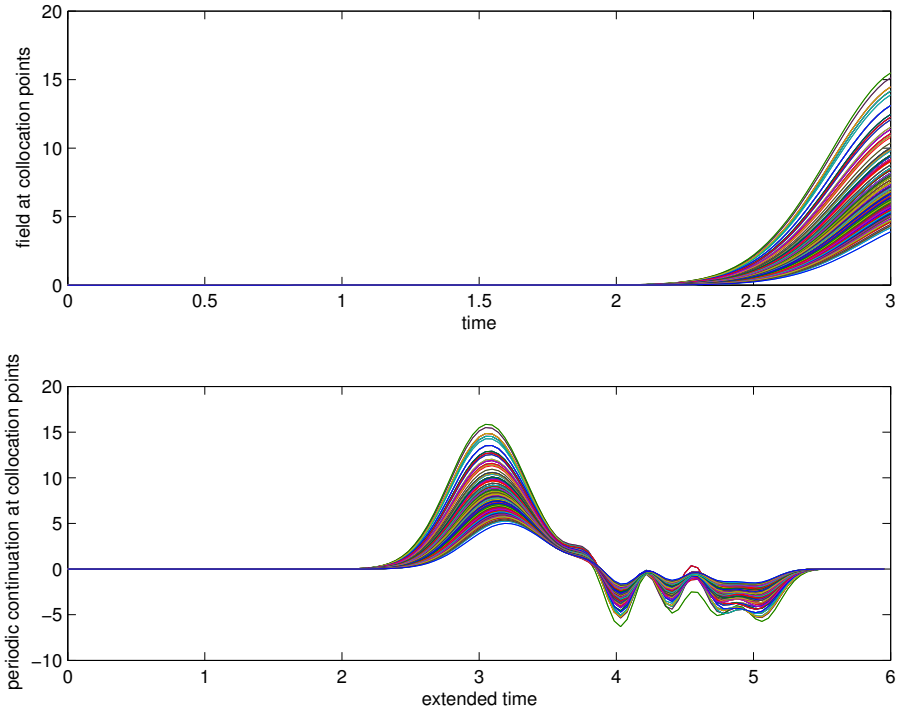


Figure 3.20: The initial data and its periodic extension at the collocation points

the points

$$\boldsymbol{x}_1 = [0, 0, 0.1876]^t, r_{\min} = 0.0939, r_{\max} = 0.2297, t_{BF} = 0.1359$$

$$\boldsymbol{x}_2 = [0, 0, 0.6875]^t, r_{\min} = 0.5938, r_{\max} = 0.7002, t_{BF} = 0.1064$$

$$\boldsymbol{x}_3 = [0, 0, 1.1875]^t, r_{\min} = 1.0938, r_{\max} = 1.1949, t_{BF} = 0.1011$$

$$\boldsymbol{x}_4 = [0, 0, 1.6875]^t, r_{\min} = 1.5938, r_{\max} = 1.6927, t_{BF} = 0.0990$$

$$\boldsymbol{x}_5 = [0, 0, 2.1875]^t, r_{\min} = 2.0938, r_{\max} = 2.1915, t_{BF} = 0.0978$$

$$\boldsymbol{x}_6 = [0, 0, 2.6875]^t, r_{\min} = 2.5938, r_{\max} = 2.6908, t_{BF} = 0.0970$$

$$\boldsymbol{x}_7 = [0, 0, 3.1875]^t, r_{\min} = 3.0938, r_{\max} = 3.1903, t_{BF} = 0.0965$$

$$\boldsymbol{x}_8 = [0, 0, 5.1875]^t, r_{\min} = 5.0938, r_{\max} = 5.1892, t_{BF} = 0.0954$$

which is displayed in Figure 3.21. The green curve corresponds to the numerical solutions obtained with Algorithm 3.3.1. The blue graph represents the exact solution. As we can see, the numerical wave re-enters from the left once it travels out of the time domain on the right. These values do not correspond to the time scale displayed, because the real wave is traveling to infinity and is not periodic in time at a

fixed point in space. The periodicity in time has been artificially created by the continuation method, and the re-entered parts should be translated to the proper time scale to the right. More precisely, re-entered data from the left needs to be shifted by the appropriate multiple of $T_e = 6$ to the right. This issue does not cause a problem for the first six points $\boldsymbol{x}_1$ to $\boldsymbol{x}_6$ if we are only interested in the time interval $[(r_{\max} - r_{BF})/c, r_{\min}/c + T]$. At the point $\boldsymbol{x}_7$, we observe that a part of the well-resolved solution has actually traveled out of the time domain and is found at a misplaced scale to the left. As expected, this effect is even better visible at the farthest point $\boldsymbol{x}_8$, where the entire numerical solution is completely misplaced. Instead of computing the appropriate shift to the correct time scale, this problem may be alternatively solved by choosing a larger value of T_e. Table 3.7 gives the accuracy for various equivalent sources n_S and collocation points n_C at the points $\boldsymbol{x}_1, \boldsymbol{x}_4$, and $\boldsymbol{x}_6$, respectively. The error $e_\infty(\boldsymbol{x})$, defined in (3.13), is computed in the time-interval $[(t_{\min} + t_{BF})/c, t_{\min}/c + T]$.

$t_{BF} = r_{\max} - r_{\min}$							
point	S	n_S	C	n_C	e_∞	$\partial/\partial_t e_\infty$	$\partial^2/\partial_t^2 e_\infty$
$\boldsymbol{x}_1$	5	50	8	296	$1.0 \cdot 10^{-5}$	$2.6 \cdot 10^{-5}$	$3.9 \cdot 10^{-4}$
$\boldsymbol{x}_4$					$2.4 \cdot 10^{-6}$	$1.3 \cdot 10^{-5}$	$7.3 \cdot 10^{-5}$
$\boldsymbol{x}_6$					$1.4 \cdot 10^{-6}$	$8.4 \cdot 10^{-6}$	$4.5 \cdot 10^{-5}$
$\boldsymbol{x}_1$	6	72	8	296	$4.1 \cdot 10^{-6}$	$1.3 \cdot 10^{-5}$	$6.0 \cdot 10^{-5}$
$\boldsymbol{x}_4$					$7.5 \cdot 10^{-7}$	$6.1 \cdot 10^{-6}$	$5.0 \cdot 10^{-5}$
$\boldsymbol{x}_6$					$5.0 \cdot 10^{-7}$	$4.1 \cdot 10^{-6}$	$3.3 \cdot 10^{-5}$
$\boldsymbol{x}_1$	7	98	9	386	$2.5 \cdot 10^{-7}$	$3.6 \cdot 10^{-6}$	$4.0 \cdot 10^{-5}$
$\boldsymbol{x}_4$					$7.0 \cdot 10^{-7}$	$5.9 \cdot 10^{-6}$	$4.7 \cdot 10^{-5}$
$\boldsymbol{x}_6$					$4.7 \cdot 10^{-7}$	$3.9 \cdot 10^{-6}$	$3.1 \cdot 10^{-5}$

Table 3.8: Convergence study of Algorithm 3.3.1

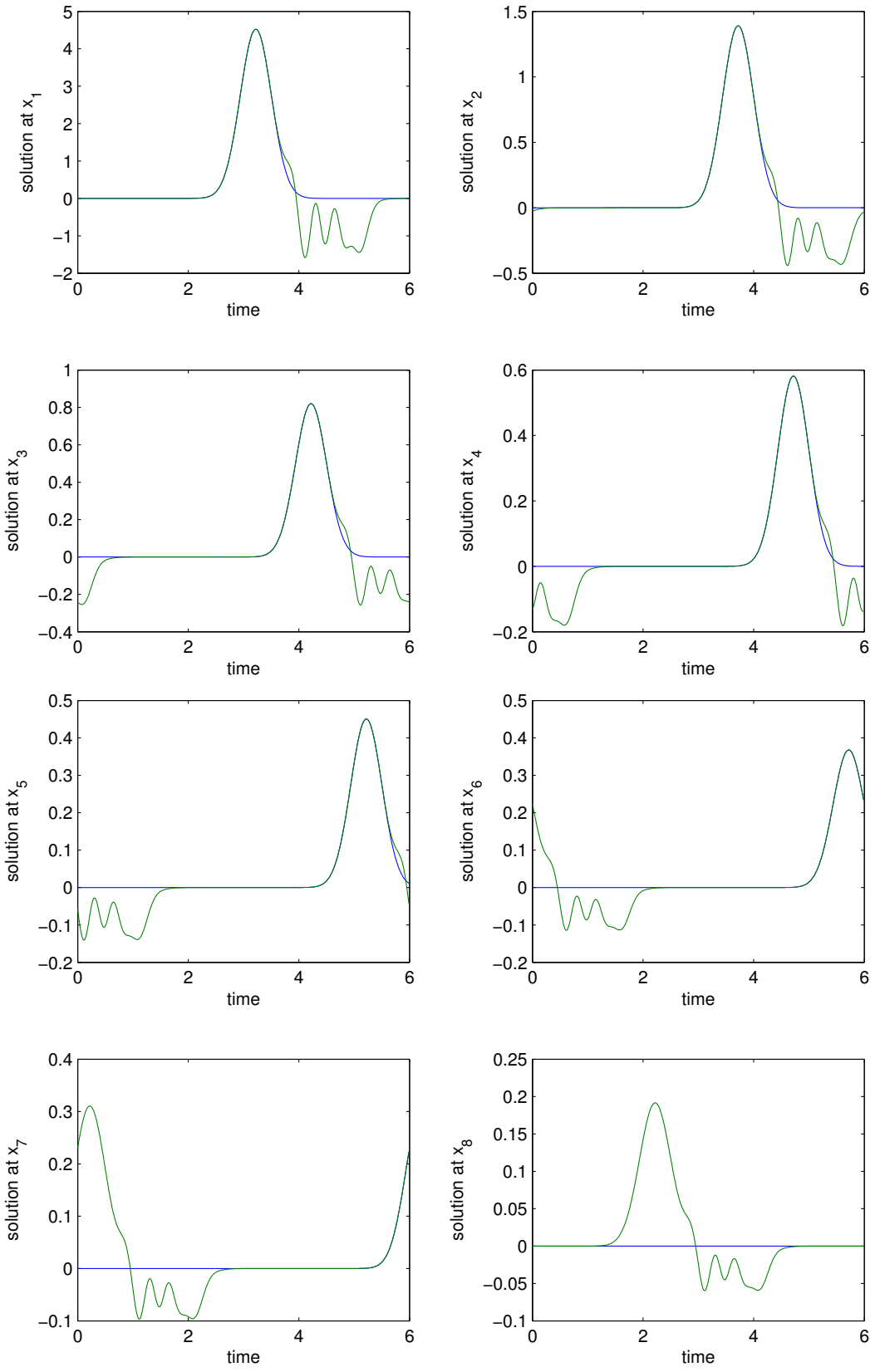


Figure 3.21: The blue curve shows the exact wave function; the green curve is the numerical solution obtained by extending the data periodically with the continuation method at the collocation points, obtaining the corresponding equivalent sources on the two faces D_1 and D_2, and evaluating the field with these sources at the shown points.

Chapter 4

Scattering solver

In this chapter, we introduce our new methodology for evaluation of computational boundary conditions, and, thus, for solution of the scattering problem (2.1)–(2.5) in general unbounded domains. In Algorithm 2.3.1 we formulated the main basis of the boundary condition algorithm without specifying details of evaluation methodologies. Taking advantage of the concepts we developed in the last chapter to formulate a new convergent boundary condition algorithm by detailing aspects left unspecified in Algorithm 2.3.1.

This chapter is organized as follows: in Section 4.1, we present the nonreflecting boundary condition algorithm. First, we introduce an appropriate geometric framework (in 4.1.1) that allows us to identify Algorithm 3.3.1 as a local problem. In Section 4.1.2 we present an efficient computation of the data on the artificial boundary $\mathcal{B}$. In 4.1.3, we discuss an important subtlety: the reconstruction of numerical derivatives from noisy data. This issue is important when coupling our nonreflecting boundary condition with an interior solver to insure the expected convergence rate. (The accuracy of evaluating our nonreflecting boundary condition is not to be lower than that presented in the interior solver (such as FEM), so that the overall convergence rate is generally that given by the interior method). We introduce a novel concept based on Chebyshev interpolation which produces excellent accuracies for numerical differentiation of functions with insignificant computational cost and extremely simple implementations. An aspect concerning adequate choices of the boundary operator $\mathcal{L}$ is discussed in 4.1.4: long-time instability occurs if an inappropriate boundary condition operator is used.

The (standard) interior solver we use to demonstrate the properties of our computational boundary conditions is introduced in Section 4.2. Specifically, we choose a finite element method (FEM) in space, because it can handle complicated geometries and high-order shape functions may be defined

over selected elements as necessary. The details of the FEM formulation are presented in Section 4.2.2. Because of the time dependence, an appropriate time-marching scheme is also a part of the interior solver; well-known explicit and implicit methods are briefly discussed in 4.2.3.

Numerical results in 4.3 confirm the excellent properties of our nonreflecting boundary condition: the overall method exhibits the expected convergence rate as the discretizations are refined, and the computing times which arise in connection with the boundary value evaluation are only a small portion of the time required by the interior scheme. This is a significant advantage over a direct application of Kirchhoff's formula as put forward in [86]. A complexity and storage count of the main operations that are involved in the computational boundary value evaluation is given in 4.4. Finally, we close this thesis with concluding remarks in Section 4.5.

4.1 Nonreflecting boundary condition

This section deals with the design and construction of our nonreflecting boundary condition. First, we introduce in 4.1.1 the geometric framework that is necessary to formulate the boundary data evaluation. Based on that framework, we propose in Section 4.1.2 a new technique that relies on equivalent sources to compute the nonreflecting boundary data. The technique coupled with a numerical interior scheme requires resolution of a crucial detail to insure that the convergence rate of the interior solver is achieved by the overall algorithm: the problem of numerical differentiation of approximated functions without order-of-accuracy deterioration must be addressed. In 4.1.3, we propose a new methodology to address this problem. Another important aspect is the question of long-time stability. We demonstrate that this topic is closely connected with finding a suitable boundary operator. This issue is addressed in Section 4.1.4.

4.1.1 Geometry

We focus on problem (2.1)–(2.5) and recall that we deal with unbounded domains: a scatterer with surface Γ of arbitrary shape occupies the domain Θ. The scattered waves u can travel freely from Γ into the infinite space and are thus functions defined in $\mathbb{R}^3 \backslash \Theta$. We also recall that, per our earlier assumptions, all initial conditions along with the forcing term vanish outside of a certain region that is close to the scatterer. We surround this region with a transparent Kirchhoff surface $\mathcal{S}$: as the waves expand into the infinite domain, they freely travel through $\mathcal{S}$ allowing us to accumulate the wave data. In view of

Kirchhoff's surface representation (2.15), the accumulated information can be used to evaluate the field at any point outside of $\mathcal{S}$. We impose the following two requirements on the construction of $\mathcal{S}$: 1) The surface should be as close to the scatterer as possible and 2) it does not intersect with the regions of compact support for $u_0, \dot{u}_0$, and f. The computational domain Ω surrounds the scatterer Γ and must be defined in such a way that it entirely contains $\mathcal{S}$. Apart from that, Ω can be chosen freely its size will generally depend on the needs of the specific application. Because Ω completely surrounds the scatterer, it is internally bounded by Γ, and externally by the artificial boundary $\mathcal{B}$. For simplicity, in this thesis we assume that $\mathcal{B}$ is the surface of a cuboid, i.e., it consists of the union of the cuboid's six faces $\mathcal{B}_k$: $\mathcal{B} = \cup_{k=1}^{6} \mathcal{B}_k$. However, this is not a requirement for our method to work: in fact, $\mathcal{B}$ can take any shape with insignificant additional computational costs by incorperating relevant techniques from [14]. Next, we introduce an equidistant three-dimensional Cartesian grid τ_H with mesh size H. Without loss of generality, we assume that this grid splits the volume of the cuboid into cubes of edge size H. The union of all the cubes in τ_H that intersect with the scatterer Θ fill a volumetric domain which embeds Γ entirely. We assume that the compactly supported regions of $u_0, \dot{u}_0$, and f also fall into this domain. Thus, the surface of this domain qualifies and is chosen to act as $\mathcal{S}$ (see Figure 4.1). We denote the P faces of τ_H

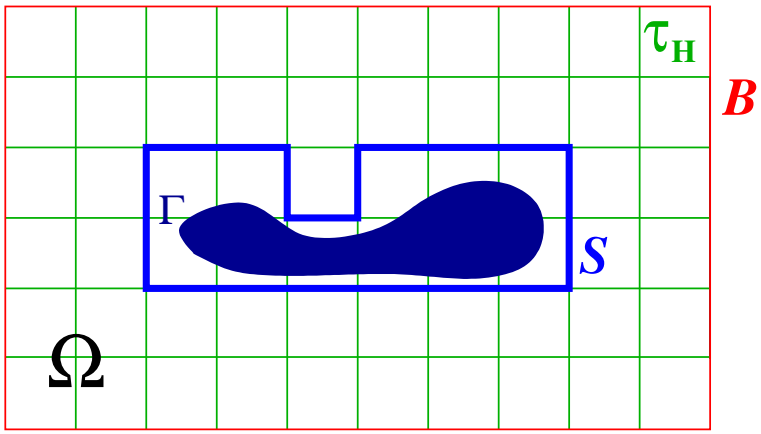


Figure 4.1: Geometry of the global problem: The computational domain Ω with inner boundary Γ and outer boundary $\mathcal{B}$. The Cartesian grid τ_H splits the cuboid into cubes, and the Kirchhoff surface $\mathcal{S}$ embeds the scatterer Γ and is compromised of specially selected faces of τ_H.

which lie on $\mathcal{S}$ by $\mathcal{S}_j$, where $j \in \{1, \ldots, P\}$, i.e., $\mathcal{S} = \cup_{j=1}^{P} \mathcal{S}_j$. To further proceed with the presentation of our computational boundary condition, we need to introduce additional geometrical concepts to each panel S_j, which will result in defining a local problem. In what follows we give the description of the local geometry, and explain in Section 4.1.2 how this relates to the computational boundary condition.

At each pair of opposite edges of each $\mathcal{S}_j$, we place a parallel pair of faces $D^{(l)}_{(j,1)} \cup D^{(l)}_{(j,2)}$ with edge length H perpendicular to $\mathcal{S}_j$. This is illustrated in Figure 4.2. The subscript 1 or 2 distinguishes the disc in a two-face pair and is consistent with the notation of the previous chapter, while $l \in \{1, 2\}$ denotes which one of the two possible two-face pairs is being referred to. Finally, we associate with every $\mathcal{S}_j$ a collocation cube of size $3H$ which is the union of the cube just constructed in Figure 4.2 with its 26 neighboring cubes of the same size H. We denote the surface of the collocation box by $\mathcal{S}_{\mathcal{C}j}$. We now

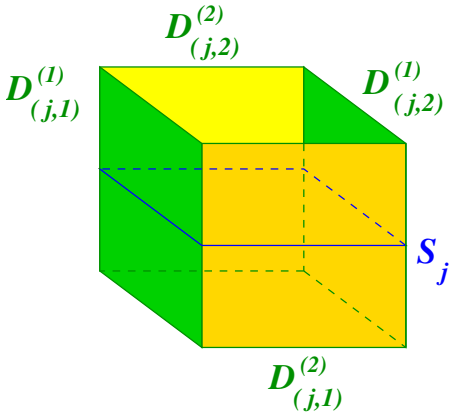


Figure 4.2: Geometry for local panel $\mathcal{S}_j$. The two two-face pairs $D^{(l)}_{(j,1)} \cup D^{(l)}_{(j,2)}$, $l \in \{1, 2\}$, are perpendicular to S_j and are of edge size H.

have all the necessary tools for the design of our nonreflecting boundary condition that acts as boundary solver in step 3 of Algorithm 2.3.1. We give its formulation in the next section.

4.1.2 Evaluation of the computational boundary condition

We recall that in the scattering solver 2.3.1, we left the two main algorithms unspecified, namely, the exact procedure to evaluate the data on $\mathcal{B}$ (step 3) and, the numerical scheme for the PDE system (2.18)–(2.22) that acts as interior solver (step 4). This section deals with step 3, the main topic of this thesis.

We recall that the interior solver cannot proceed, say at $t = T$, when needed data on $\mathcal{B}$ is unspecified. We assume that the interior scheme for that cycle was employed for the time interval $I = [0, T]$. Our goal is to compute the boundary values $(\boldsymbol{x}, t)$ at $\mathcal{B} \times [t_{\min}, T + t_{\min}]$ based on the information we gathered on $\mathcal{S}$ up to the time $t = T$, where $t_{min} \leq T$ denotes the minimum time the wave needs to travel from $\mathcal{S}$ to $\mathcal{B}$.

We point out that in [86], it was proposed to use the Kirchhoff representation (2.15) for this task.

However, the computation is extremely cost-intensive—in fact, significantly more expensive than the volumetric interior calculation. In what follows, we present an approach based on Algorithm 3.3.1 whose computational cost amounts to only a small portion of the three-dimensional interior computational cost.

By our construction of the last section, Kirchhoff's surface $\mathcal{S}$ decomposes into P panels $\{\mathcal{S}_j\}_{j=1}^{P}$. In view of formula (2.15), it is clear that instead of performing the integration over $\mathcal{S}$ directly, we can decompose the computation into P integrals over $\mathcal{S}_j$ and sum the results after evaluating each integral separately. In other words, the data on $\mathcal{S}$ can be considered true sources: an infinitesimal wave expands from the surface into the space. The superposition of these contributions at any point yields the total field. This observation leads to the following idea: instead of using the true sources on $\mathcal{S}_j$ directly, we could substitute them for the equivalent sources. Now if we could reposition the true sources lying on $\mathcal{S}$ to special locations, say on planes on τ_H that are parallel to one of the six faces of $\mathcal{B}$, we could take advantage of a fast evaluation of the field which is based on FFTs, see Appendix B.3.

In what follows, we give the technical details that can be used to implement this idea. On each $\mathcal{S}_j$, we select a two-dimensional mesh $\tau_{\mathcal{S}_j}$ with $S \times S$ nodes and, at each node, we extend the domain of definition in time from $[0,T]$ to $[-t_{BF},T] \cup [T, 2T+t_{BF}]$ by choosing a suitable time buffer t_{BF}. The continuation method can then be applied to obtain a periodic function that approximates the wave in $[0,T]$ to high order, and an accurate representation of the Fourier coefficients can be obtained by means of FFT. For each relevant frequency ω, we can compute the Fourier coefficients on $\boldsymbol{x} \in \mathcal{S}_{C_j}$ by the integral representation

$$\hat{u}_{\mathcal{S}_j}(\boldsymbol{x},\omega) \;=\; \int_{\mathcal{S}_j} \left\{ \hat{u}(\tilde{\boldsymbol{x}}) \frac{\partial G_k(\boldsymbol{x},\tilde{\boldsymbol{x}})}{\partial \nu(\tilde{\boldsymbol{x}})} - \frac{\partial \hat{u}}{\partial \nu}(\tilde{\boldsymbol{x}}) G_k(\boldsymbol{x},\tilde{\boldsymbol{x}}) \right\} ds(\tilde{\boldsymbol{x}}), \tag{4.1}$$

where $G_k(\boldsymbol{x},\tilde{\boldsymbol{x}})$ is the Green's function for the Helmholtz equation (see (A.12)), and the subscript k denotes the wave number (see (A.8)). Note that in a typical application, when the parameters are chosen appropriately, such a local computation has a significantly smaller amount of points on $\mathcal{S}_j$ and $\mathcal{S}_{C_j}$ than on $\mathcal{S}$ and $\mathcal{B}$. As a consequence, the computational effort for these local operations is expected to be much smaller than performing the evaluation globally. Once the local field generated by the distributions of sources on the collocation surface $\mathcal{S}_j$ is known on $\mathcal{S}_{C_j}$, Algorithm B.2.1 can be used to determine the local equivalent sources on the faces S_j and $D^{(l)}_{(j,1)} \cup D^{(l)}_{(j,2)}$ for $l \in \{1,2\}$. We point out that it is not necessary to compute the equivalent sources on $\mathcal{S}_j$; instead, we can just use the true sources. Figure 4.3 graphically depicts such a local problem associated with $\mathcal{S}_j$. These P local problems

are completely independent from each other and are thus a perfect task for parallel computing, should such parallel implementations be pursued. Once the local equivalent sources are known, they need to

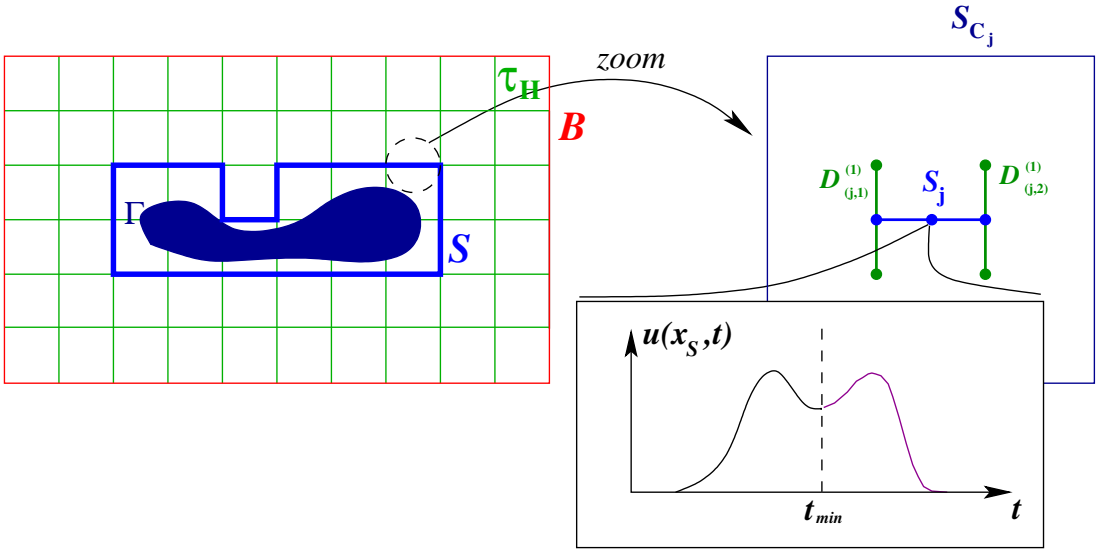


Figure 4.3: Graphical illustration of the local problem: The continuation method is applied to data on $\mathcal{S}_j$ by extending the time-domain from $[0,T]$ to $[-t_{BF},T]\cup[T,2T+t_{BF}]$. After a Fourier transform, the Fourier coefficients are evaluated on $\mathcal{S}_{\mathcal{C}j}$, and finally, the two-face approach can be applied to obtain the local equivalent sources on $D^{(l)}_{(1,j)}\cup D^{(l)}_{(2,j)}$: these sources generate the same local field as the distributions on $\mathcal{S}_j$.

be assembled to global distributions. In the following, we focus our attention on the evaluation of the global field on $\mathcal{B}_k$ for a fixed $k\in\{1,\ldots,6\}$. All local faces parallel to this specific plane are selected, which is either $\mathcal{S}_j$ or one of the two pairs $D^{(1)}_{(j,1)}\cup D^{(1)}_{(j,2)}$, $D^{(2)}_{(j,1)}\cup D^{(2)}_{(j,2)}$. The strength of the sources of the selected planes are added together if they lie at the same node. In this manner, global equivalent sources associated to planes that are parallel to B_k are obtained. They represent to high order the same field as is induced by the true sources on $\mathcal{S}$. On each such plane, the global equivalent sources form a two-dimensional mesh, which, if necessary, can be extended to an equally spaced Cartesian grid by assigning the nodes with no distribution a source of strength zero. Note that all of these planes lie on τ_H—see Figure 4.4 and Figure 4.5. Once the global equivalent sources are defined on Cartesian regular meshes, the techniques described in Appendix B.3 to B.6 can be applied to rapidly compute the data on $\mathcal{B}_k$.

We close this section with a numerical illustration. Let us assume that the scatterer Γ is a sphere with radius $R=0.25$ and $\mathcal{S}$ is the cube of size $2R$ which includes the sphere entirely. We choose three different grids τ_H with the mesh sizes H equal to $2R/5, 2R/7$ and $2R/15$, respectively. The outer

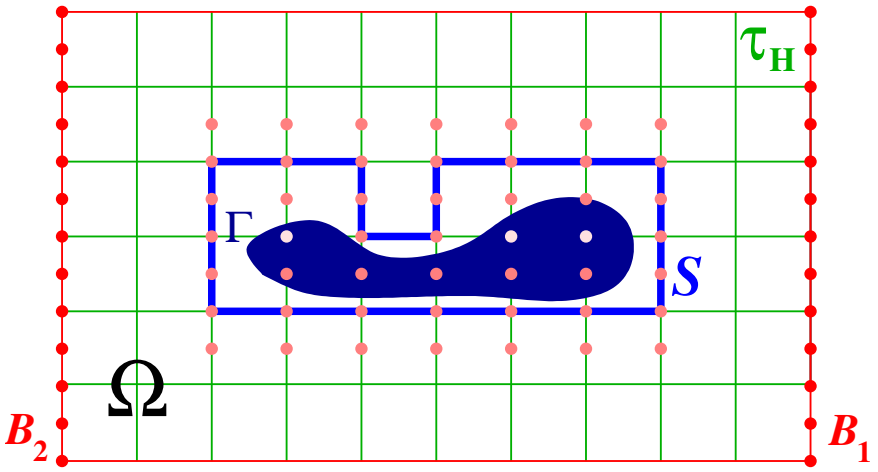


Figure 4.4: Global equivalent sources to compute the field on $\mathcal{B}_1$ and $\mathcal{B}_2$ fast. The locations with a lighter dot correspond to point distributions of zero strength.

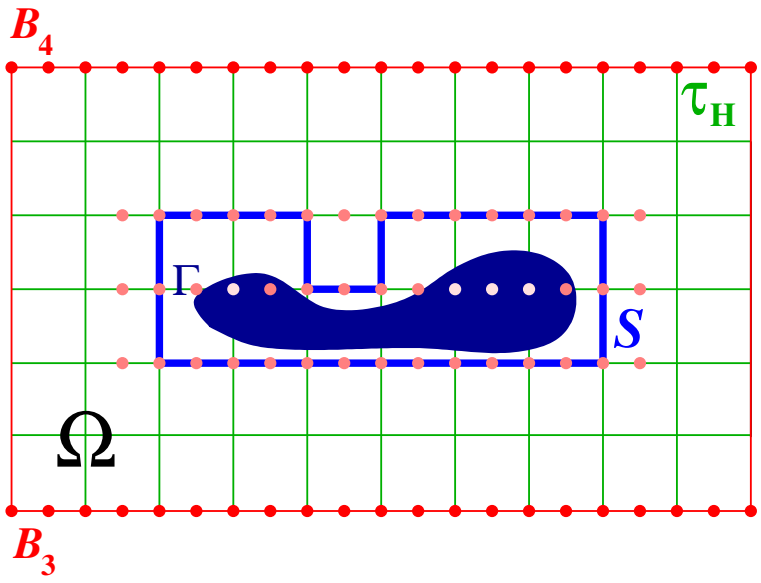


Figure 4.5: Global equivalent sources to compute the field on $\mathcal{B}_3$ and $\mathcal{B}_4$ fast. The locations with a lighter dot correspond to point distributions with zero strength.

boundary $\mathcal{B}$ is the surface of the cube of length $2(R + nH)$ centered about the origin. The integer n equals 2 for the coarsest, 3 for the intermediate, and 6 for the finest mesh τ_H. It follows that the minimum distance from the inner to the outer boundary is $r_{\min} = nH$, i.e., for the three different grids we have $r_{\min} = 0.2, r_{\min} \approx 0.21429$, and $r_{\min} = 0.2$, respectively. The maximum distance from $\mathcal{S}$ to $\mathcal{B}$ is $r_{\max} = \sqrt{3}(2R + r_{\min})$. We define the time buffer to be $t_{BF} = (r_{\max} - r_{\min})/c$, which is according to (3.26) a feasible choice. We assume that the known field at the point $\boldsymbol{x} \in \mathcal{S}$ for $t \in [-t_{BF}, T]$ is of the form (3.6), where $T = 3, r = |\boldsymbol{x}|$ and $s(t)$ is defined in (3.11) with the parameters $\sigma = 0.5, t_0 = 2$. From this information, we compute the field on $\mathcal{B}$ as explained in this section. The wave function on $\mathcal{S}$ is extended from the time interval $[-t_{BF}, T]$ to $[-t_{BF}, 2T + t_{BF}]$ with the continuation method described in Section 3.3.2. The sampling points and modes in time are chosen as $N = M = 32$, and we use zero padding to sample the interval at $N_z = 128$ equidistant points. The propagation velocity of the wave is $c = 1$, which means that the maximum discrete wave number equals $k_{\max} \approx 16.76$. The graphical constructions of Section 3.3.3 still apply, with the appropriate reinterpretation of $r_{\min}$ and $r_{\max}$. The solution u_h on $\mathcal{B}$ is valid for the time interval $I = [t_{\min}, t_{\min} + T]$. The error $e_{\mathcal{B}}$ is defined by

$$e_{\mathcal{B}} = \max_{\boldsymbol{x} \in \mathcal{B}} \max_{t \in I} |u(\boldsymbol{x}, t) - u_h(\boldsymbol{x}, t)|. \tag{4.2}$$

The results are reported in Table 4.1 for different choices of S and C. On $\mathcal{B}$, we also compute the maximum space-time errors $\partial_t e_{\mathcal{B}}$ and $\partial_\nu e_{\mathcal{B}}$, where ν is the outer normal of $\mathcal{B}$. We note that the results for the finest and the coarsest mesh τ_H can be compared directly, because in both cases $r_{\min} = 0.2$ and the constructed surfaces $\mathcal{B}$ coincide in both cases. For the intermediate grid τ_H, however, we have $r_{min} \approx 0.21429$.

4.1.3 High-accuracy differentiation

The problem of evaluating numerical derivatives from approximate data arise in a wide range of areas of numerical analysis and scientific computing, including image processing, solution of integral and differential equations, parameter identification, etc. [27, 20, 35, 46, 45]. The particular problem motivating this work, concerns the evaluation of (4.1) accurately: since the Kirchhoff integral involves both values of the function and its normal derivative, a degradation in the order of the solution accuracy would occur unless such normal derivatives are evaluated with an order of accuracy equal to that implicit in the interior solver formulation. As we demonstrated in 4.3, evaluation of the necessary normal derivatives by

H	S	C	$e_{\mathcal{B}}$	$\partial_t e_{\mathcal{B}}$	$\partial_\nu e_{\mathcal{B}}$
0.1	3	3	$1.2 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$
	5	5	$7.5 \cdot 10^{-6}$	$2.4 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$
	7	7	$3.5 \cdot 10^{-7}$	$1.1 \cdot 10^{-6}$	$3.7 \cdot 10^{-6}$
	9	9	$2.2 \cdot 10^{-8}$	$4.1 \cdot 10^{-7}$	$5.9 \cdot 10^{-7}$
0.07143	3	3	$7.8 \cdot 10^{-4}$	$2.6 \cdot 10^{-3}$	$6.7 \cdot 10^{-3}$
	5	5	$4.3 \cdot 10^{-6}$	$1.7 \cdot 10^{-5}$	$3.9 \cdot 10^{-5}$
	7	7	$2.0 \cdot 10^{-7}$	$2.2 \cdot 10^{-6}$	$2.7 \cdot 10^{-6}$
	9	9	$4.2 \cdot 10^{-8}$	$1.6 \cdot 10^{-6}$	$1.6 \cdot 10^{-6}$
0.03333	3	3	$1.1 \cdot 10^{-3}$	$4.0 \cdot 10^{-3}$	$8.0 \cdot 10^{-3}$
	5	5	$7.3 \cdot 10^{-6}$	$2.7 \cdot 10^{-5}$	$5.2 \cdot 10^{-5}$
	7	7	$1.2 \cdot 10^{-7}$	$5.5 \cdot 10^{-7}$	$9.3 \cdot 10^{-7}$
	9	9	$1.5 \cdot 10^{-8}$	$3.8 \cdot 10^{-7}$	$4.3 \cdot 10^{-7}$

Table 4.1: Accuracy for three different meshes τ_H

means of the technique introduced here allows one to obtain solutions with full order of accuracy—with negligible computational cost.

Previous work on reconstruction of numerical derivatives from scattered noisy data [5, 25, 85, 45, 80, 92, 93, 36, 5] has focused on one of two main approaches: finite differences and regularization. Most of these references are concerned with stability: they seek to eliminate large derivative errors that arise as two function values $f(x_1)$ and $f(x_2)$ with large errors that occur at points x_1 and x_2 that lie very close to each other. In reference [92, 93], for example, an approach (with theoretical error bounds in [93]) is presented which seeks to minimize accuracy degradation by means of a Tikhonov regularization strategy. The methodology presented in this thesis offers a number of advantages over those introduced earlier: 1) It is extremely simple—it only requires simple polynomial interpolations and Chebyshev approximations, both of very low order; 2) The theoretical analysis is straightforward; 3) Unlike the approaches [92, 93], it does not require the solution of a large system of equations; 4) It preserves the accuracy order even for strongly randomized errors; and 5) It is significantly more accurate than previous methods. We emphasize these points through consideration of an example drawn from [93]: for a smooth function selecting (smooth!) errors of the form $0.001 \sin(\pi x_i)$ and inverting a large system of equations, the cited work produced first derivatives of the order of 0.01, while, for this example, the present method produces errors of the order of 0.002. More importantly, our method produced a maximum error of the order of 0.005, even when the error in the function values was of order 0.001 but *random*—so that two neighboring points could have errors of equal magnitude and different sign. Equally appealing results can be obtained from the present methodology for derivatives of arbitrarily high orders. As an additional

reference we mention the contribution [5] where a statistical framework is considered, and bounds on the variances of the errors of the computed derivatives are given; for example, evaluation of the second-order derivatives making up the Laplacian in three dimensions were obtained in that contribution with an error variance of order $h^{6/11}$; corresponding second derivatives of one-dimensional data were given with error variances of order $h^{2/9}$; in this context our method would, in contrast, provide accuracy of high order until the limit imposed by the statistics of the underlying error is reached.

This section is organized as follows: we first introduce our method and establish rigorous error bounds and convergence speeds. In order to make precise the character of the proposed methodology we assume the underlying data contain an error of the order $\mathcal{O}(h^r)$, where h is a measure of the step-size of the possibly non-equi-spaced grid; we then show that, for any order of differentiation, the derivatives obtained by means of the proposed approach contain an error that decreases, likewise, like h^r. We then present a variety of numerical examples, including cases in one and higher dimensions, first and higher derivatives, problems including random errors, and, finally, an example in which the derivatives of a finite element solution of the Poisson equation are computed. In all cases the proposed methodology produces excellent accuracies with insignificant computational cost and extremely simple implementations.

Theory

We consider a function $f(t)$ which is either s times differentiable ($f(t) \in C^s$), infinitely differentiable ($f(t) \in C^\infty$), or analytic ($f(t) \in C^\omega$); without loss of generality we assume f is defined in the interval $[-1, 1]$. (We focus on a one-dimensional problem at first, although, as shown in Section 4.1.3, problems in arbitrary dimensions may be treated by this method.) Let a set of L approximate discrete values of f, $\{\tilde{f}(\theta_\ell)\}_{\ell=1}^L$ be given, in such a way that for a fixed constant C we have

$$|f(\theta_\ell) - \tilde{f}(\theta_\ell)| \leq Ch^r, \quad \ell = 1, \ldots, L, \tag{4.3}$$

where the mesh size h is given by

$$h \equiv \max_{1 \leq \ell \leq L} |\theta_\ell - \theta_{\ell-1}|. \tag{4.4}$$

Roughly speaking, our method is based on using the given data to produce adequate Chebyshev interpolations: in view of the extremely fast convergence of Cheybyshev approximations, Chebyshev expansions

of low-order can be used with high accuracy, thereby avoiding accumulation of sampling errors. In what follows we thus consider low order Chebyshev approximations in the interval $[-1, 1]$; the corresponding Chebyshev points for a Chebyshev expansion of order N will be denoted by

$$t_k = \cos\left(\frac{\pi(k - 1/2)}{N}\right). \tag{4.5}$$

Clearly the values of $\tilde{f}$ are not given on a Chebyshev grid $\{t_k\}$; our strategy thus calls for interpolation by polynomials of adequate degrees to insure preservation of the accuracy order in the function values. Indeed, to guarantee convergence of order r in presence of errors as in equation (4.3), for every point t_k, we find a set of r consecutive neighboring points $\{\theta_j, \ldots, \theta_{j+r}\}$ such that $(\theta_j \leq t_k \leq \theta_{j+r})$, and we obtain an approximation of f at t_k by evaluating at the point t_k the polynomial of order r that interpolates approximate values $\{\tilde{f}(\theta_\ell)\}_{\ell \in \mathcal{N}_k}$; clearly, the values thus obtained approximate the exact function f with an error of the order of h^r:

$$|f(t_k) - \tilde{f}(t_k)| \leq Ch^r, \quad k = 1, \ldots, N. \tag{4.6}$$

Using these values our method then proceeds by evaluating the Chebyshev interpolant

$$\tilde{f}_N(t) \quad = \quad \sum_{i=1}^{N}{}' \tilde{c}_i T_i(t), \tag{4.7}$$

where $T_j(t)$ is the Chebyshev polynomial of degree j, see [11, 78]. (Here we use the usual convention according to which the prime in the summation symbol indicates that the $i = 1$ term in the sum is multiplied by 1/2.)

Algorithm prescription: Using the notations set forth in the previous paragraphs, *an approximate derivative of the function f of an arbitrary order n is obtained via n-fold differentiation of the relation (4.7).* This is the complete prescription of the proposed (exceedingly simple) methodology; an equally simple error analysis, presented in what follows, establishes the excellent properties of the algorithm.

Remark 4.1.1. *We note that the evaluation of the Chebyshev interpolant is not subject to instabilities that arise in regular polynomial interpolation: the Chebyshev coefficients in equation (4.7) can be obtained*

easily by taking advantage of the discrete-orthogonality property

$$\sum_{k=1}^{N} T_i(t_k)T_j(t_k) \quad = \quad k_N\delta_{ij}, \tag{4.8}$$

where the normalization factor k_N in (4.8) is given by

$$k_N = \begin{cases} \frac{N}{2}, & \text{if } i = j \neq 1 \ , \\ N, & \text{if } i = j = 1 \ , \end{cases} \tag{4.9}$$

which gives rise to a stable procedure, embodied in the formula

$$\tilde{c}_j = \frac{2}{N}\sum_{k=1}^{N} \tilde{f}(t_k)T_j(t_k), \tag{4.10}$$

for the evaluation of the Chebyshev interpolant. Similarly, the continuous orthogonality property

$$\int_{-1}^{1} \frac{T_i(t)T_j(t)}{\sqrt{1-t^2}}\mathrm{d}t = k_\pi\delta_{ij} \qquad \text{where} \qquad k_\pi = \begin{cases} \frac{\pi}{2}, & \text{if } i = j \neq 1 \ , \\ \pi, & \text{if } i = j = 1, \end{cases} \tag{4.11}$$

can be used to obtain the alternative formula

$$\tilde{c}_j = \frac{2}{\pi}\int_{-1}^{1} \frac{\tilde{f}_N(t)T_j(t)}{\sqrt{1-t^2}}\mathrm{d}t, \tag{4.12}$$

which, although not part of our numerical implementation, will prove useful in our theoretical analysis.

For notational simplicity, in the subsequent error analysis we assume f is analytic ($s = \omega$); the analysis can be carried in a completely analogous manner, with very similar results, for cases where f is either finitely or infinitely differentiable, but not necessarily an analytic function.

Proceeding to our error analysis, we note first that the error $|f(t) - \tilde{f}_N(t)|$ can be estimated easily. To do this let $f(t)$ be the (rapidly convergent) Chebyshev expansion

$$f(t) = \sum_{j=1}^{\infty}{}' c_j T_j(t) \qquad \text{with} \qquad c_j = \frac{2}{\pi}\int_{-1}^{1} \frac{f(t)T_j(t)}{\sqrt{1-t^2}}\mathrm{d}t, \tag{4.13}$$

and let $f_N(t)$ denote the truncation of this series at order N:

$$f_N(t) \;=\; \sum_{j=1}^{N}{}' c_j T_j(t). \tag{4.14}$$

The triangle inequality then yields

$$|f(t) - \tilde{f}_N(t)| \le |f(t) - f_N(t)| + |f_N(t) - \tilde{f}_N(t)|. \tag{4.15}$$

Under the present hypothesis of analyticity of the function f, the first term on the right-hand side of equation (4.15) is exponentially small: less than, say, $C_1 \exp(-\sigma N)$ for suitable constants C_1 and σ. The second term on the r.h.s. of (4.15), which equals

$$|\sum_{k=1}^{N}{}'(c_k - \tilde{c}_k)T_k(t)|, \tag{4.16}$$

is bounded by $2CNh^r$—since, from equations (4.6), (4.7), (4.12), and (4.13) we have $|c_k - \tilde{c}_k| \le 2Ch^r$. In summary, we have

$$|f(t) - \tilde{f}_N(t)| \;\le\; C_1 e^{-\sigma N} + 2CNh^r. \tag{4.17}$$

It follows that the bound of the error is minimized if

$$N \;\sim\; -\frac{1}{\sigma}\log\left(\frac{2Ch^r}{C_1\sigma}\right), \tag{4.18}$$

indicating that only few terms in the truncated sums are necessary to maximize the Chebyshev approximation. An analogous analysis can be performed if f is only s-times differentiable. This proves the following theorem:

Theorem 4.1.1. *Let $f(t)$ be in C^s and $\tilde{f}_N(t)$ constructed by (4.7). Then, the absolute error can be bounded by (4.17) if $s = \omega$ and by*

$$|f(t) - \tilde{f}_N(t)| \quad \leq \quad \frac{C_2}{N^s} + 2CNh^r \quad \textit{if} \quad s \in \mathbb{N}. \tag{4.19}$$

Further, the optimal N is given by (4.18) in the former case, and by

$$N \sim \left(\frac{sC_2}{2Ch^r} \right)^{\frac{1}{s+1}} \tag{4.20}$$

in the latter case.

The derivative of the error can be estimated very similarly by

$$|f'(t) - \tilde{f}'_N(t)| \leq |f'(t) - f'_N(t)| + |f'_N(t) - \tilde{f}'_N(t)|. \tag{4.21}$$

Looking at the r.h.s of the triangle inequality (4.21), an upper bound for the first term is $C_1 \exp(-\sigma N)$ if $s = \omega$ and C_2/N^{s-1} if $s \in \mathbb{N}$, while the second term can be bounded by $2CN^3h^r$, where we made use of the fact that $|T'_k(t)| \leq k^2$. More generally, the absolute value of the n-th derivative of the Chebyshev polynomial is bounded by $C_n k^{2n}, C_n$ being a suitable constant. This relation can be easily established by the fact that

$$|T_k^{(n)}(t)| \leq T_k^{(n)}(1) \quad \text{and} \quad T_k^{(n)}(1) = \prod_{m=0}^{n-1} \frac{k^2 - m^2}{2m+1}, \tag{4.22}$$

see [1, 11]. The result is summarized in the next theorem:

Theorem 4.1.1. *Under the same assumptions as in Theorem 4.1.1, the error of the n-th order derivative*

of the Chebyshev expansion (4.7) *is bounded by*

$$|f^{(n)}(t) - \tilde{f}_N^{(n)}(t)| \leq C_1 e^{-\sigma N} + 2CC_n N^{(2n+1)} h^r \quad \textit{if} \quad s = \omega, \tag{4.23}$$

$$|f^{(n)}(t) - \tilde{f}_N^{(n)}(t)| \leq C_2 N^{n-s} + 2CC_n N^{(2n+1)} h^r \quad \textit{if} \quad s \in \mathbb{N}; \tag{4.24}$$

clearly in the case that $f \in C^\infty$*, the second estimate in* (4.24) *holds for all* $s \in \mathbb{N}$.

Remark 4.1.1. *In view of the fast convergence to zero of the exponential function and the large powers of* $1/N^{s-n}$ *we take in practice a slightly larger value of* N *(since the constants are not known). As we refine* h*, the truncation parameter* N *virtually does not change. Thus, the error of the derivative is of the order* $\mathcal{O}(h^r)$.

Numerical Examples

Reconstructing the first derivative

We consider the Gaussian function $f(t) = \exp\left(-(t-t_0)^2/\sigma^2\right)$ with the parameters $\sigma = 0.5, t_0 = 0.1$, and choose L discrete points in $[a - h_\theta/2, b + h_\theta/2]$ at the nodes $\theta_\ell = a + (\ell - 3/2)h_\theta$ for $\ell \in \{1, \ldots, L\}$, where $h_\theta = (b-a)/(L-2)$. On the $3L$ points $t_\ell = a + (\ell - 1)h$ with $\ell \in \{1, \ldots, 3L\}$ and $h = (b-a)/(3L-1)$ we define the discrete values $\{\tilde{f}_\ell\}_{\ell=1}^{3L}$ by sampling the corresponding local linear polynomial at t_ℓ, which is obtained from the values $f(\theta_k)$ and $f(\theta_{k+1})$ with $t_\ell \in (\theta_k, \theta_{k+1}]$. Applying the mean value theorem, it is easy to see that

$$|\tilde{f}_\ell - f(t_\ell)| \leq \max_{\xi \in [\theta_k, \theta_{k+1}]} |f''(\xi)| h_\theta^2, \quad t_\ell \in (\theta_k, \theta_{k+1}], \tag{4.25}$$

thus establishing a perturbation $\mathcal{O}(h_\theta^2)$ of $\tilde{f}$ to the analytic function f. From this data we wish to construct an approximated derivative to $f'(t)$, which is also second-order convergent.

First, we remark that a polynomial of any order will generally not achieve the desired result. Indeed, local polynomials of order r can be constructed from the values $\tilde{f}$ corresponding to the nodes $\{t_\ell, \ldots, t_{\ell+r}\}$ which approximate the true function f to second order, but taking the derivative of these

local polynomials yields generally only first order convergence in the L_2- and L_∞-norms, defined by

$$e'_{L_2,p_r} \equiv \sqrt{h_s \sum_{i=1}^{N_s} (p'_r(t_i) - f'(t_i))^2}, \tag{4.26}$$

$$e'_{L_\infty,p_r} \equiv \max_{1\le i\le N_s} |p'_r(t_i) - f'(t_i)|, \tag{4.27}$$

respectively. Here N_s denotes the number of sampling points and h_s is the maximum distance between two neighboring sampling points, and $p'_r(t_i)$ is the corresponding local polynomial of order r differentiated at the point t_i.

We turn our attention to our proposed procedure and seek to define a Chebyshev polynomial in $[a, c]$ for a fixed $c \in (a, b]$. The discrete values $\{\tilde{f}_\ell\}_{\ell=1}^{3L}$ are used along with the quadratic local polynomials to approximate $f(t)$ to second order at the Chebyshev points, and finally, the Chebyshev coefficients are computed by formula (C.9). The derivative of expansion (4.7) can be expressed as

$$p'(t) \equiv \sum_{k=1}^{N} \tilde{c}'_k T_k(t), \tag{4.28}$$

where the $\tilde{c}'_k$ coefficients satisfy the recurrence relation in descending order:

$$\tilde{c}'_N = \tilde{c}'_{N-1} = 0 \tag{4.29}$$

$$\tilde{c}'_{k-1} = \frac{1}{a_{k-1}} \left\{ 2k\tilde{c}_k + \tilde{c}'_{k+1} \right\}, \quad k = N-1, N-2, \ldots, 2, \tag{4.30}$$

with $a_k = 2$ if $k = 1$ and $a_k = 1$ for $k > 1$. Equation (4.28) can be evaluated by Clenshaw's recurrence formula, and the Chebyshev coefficients $\tilde{c}'_k$ are efficiently obtained by a fast cosine transform (see [11] and [78] for more details). In Figure 4.6 the results confirm second-order convergence of (4.28) in both the L_2- and L_∞-norm. Here, the parameters $a = 0, b = 0.26, c = a + (\sqrt{2} - 1)/4 \approx 0.1036$, and $N = 10$ are used. The plots show the superior convergence rate of the Chebyshev polynomial derivative compared to the differentiation of the quadratic polynomial p_2.

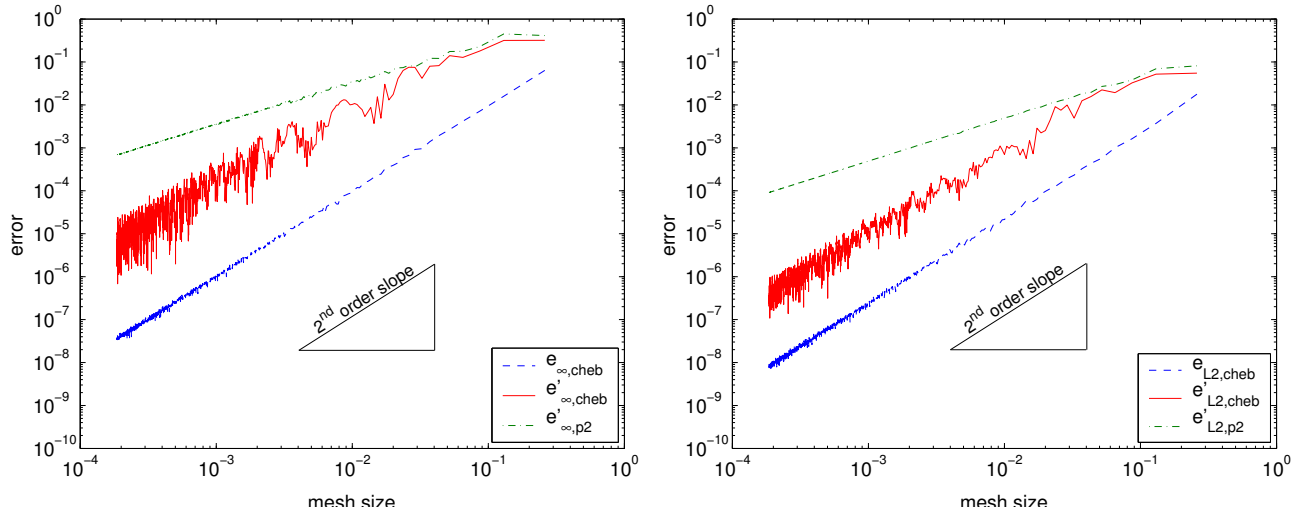


Figure 4.6: Left: second-order convergence of the Chebyshev polynomial (dashed-dotted line) and its first derivative (solid) in the L_∞-norm, as opposed to only first-order convergence of the quadratic polynomial derivative $p_2'(t)$ (dashed). Right: Same error plots as on the left, but in the L_2-norm. To compare the convergence rates, the slopes of the pictured triangles correspond to second-order convergence in both plots.

Reconstruction of higher-order derivatives

Adapting the notation of the first example, let us assume $a = -1, b = 1$. The L values $\{\tilde{f}_\ell\}_{\ell=1}^{L}$ are defined on the equidistant grid in $[a, b]$ with spacing $h = 2/(L-1)$ and are obtained by an r-th order local spline interpolation $s_r(t)$ of the exact function values which are known on the grid points $a - h_\theta\sqrt{0.2}/2 + kh_\theta$, where $k \in \mathbb{Z}, h_\theta = (2+\sqrt{0.2})/(L-1)$ and $r \in [2, 3, 4, 5]$. The suitable number of Chebyshev polynomials N is empirically determined and depends on the order r. The specific values of N used in these examples are reported in the caption of Figure 4.7. The Chebyshev coefficients $\tilde{c}_k$ are determined by (C.9). We note that (4.28)–(4.30) can be extended to obtain the n-th derivative of the Chebyshev polynomial $p(t)$ by

$$p^{(n)}(t) = \sum_{k=1}^{N} \tilde{c}_k^n T_k(t), \tag{4.31}$$

$$\tilde{c}_N^n = \tilde{c}_{N-1}^n = 0 \tag{4.32}$$

$$\tilde{c}_{k-1}^n = \frac{1}{a_{k-1}} \left\{ 2k\tilde{c}_k^{n-1} + \tilde{c}_{k+1}^n \right\}, \quad k = N-1, N-2, \ldots, 2 \tag{4.33}$$

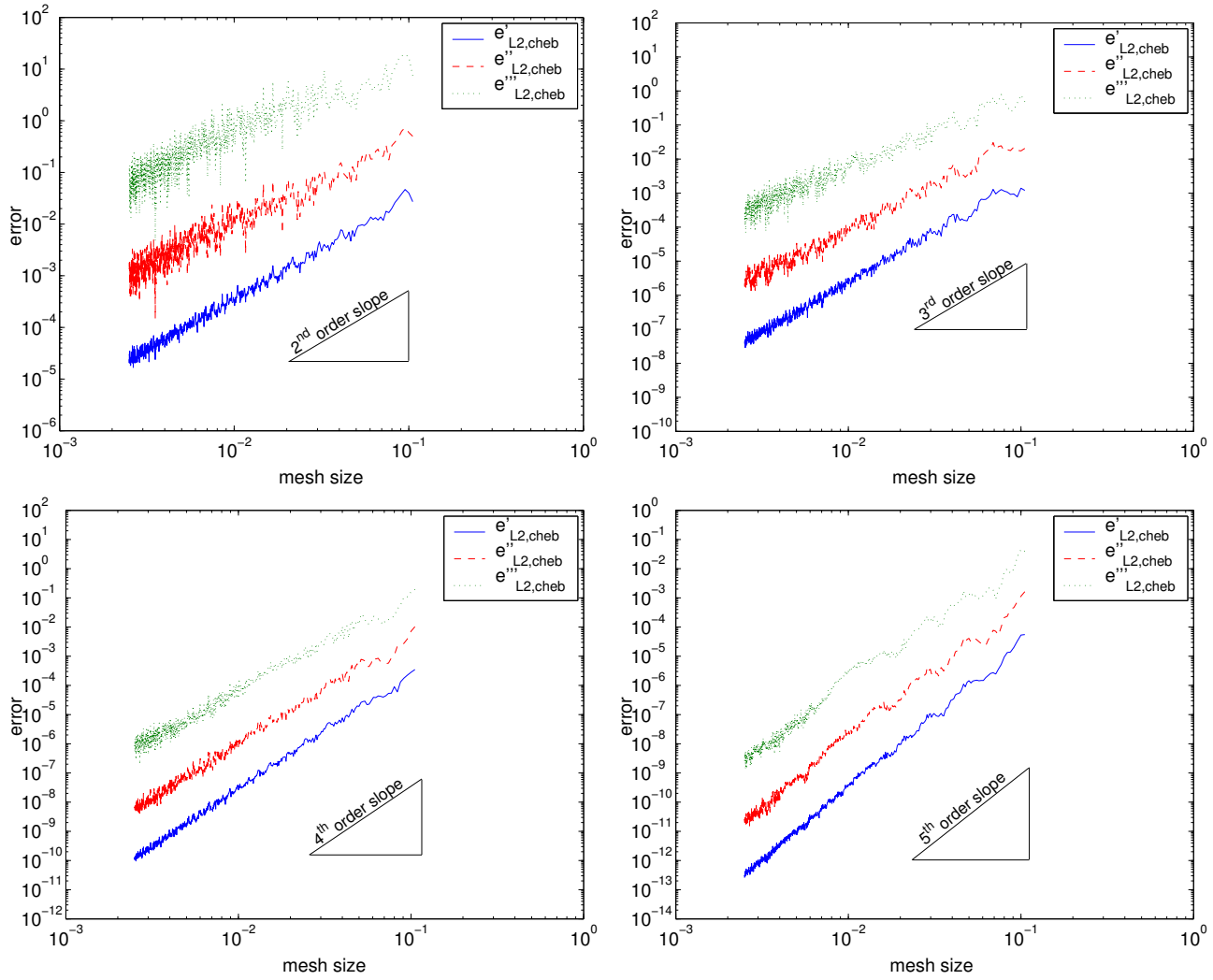


Figure 4.7: Upper left: Construction of the Chebyshev polynomial from local second-order spline functions at $N = 18$ Chebyshev points. The derivatives of the Chebyshev polynomial exhibit all second-order convergence ($n = 1, 2, 3$). Upper right: The values at the $N = 24$ Chebyshev points are obtained from third-order spline polynomials. Therefore, the derivatives of the Chebyshev polynomial are of order three, which is again demonstrated for the first three derivatives. Lower left: $N = 30$ terms are used in the Chebyshev expansion. The values at the Chebyshev points approximate the true function to fourth order. Lower right: Fifth-order convergence of the Chebyshev polynomial derivatives (displayed again for $n = 1, 2, 3$), and $N = 34$ is used. The triangles in all plots exhibit the expected convergence slope as comparison.

(see [11]). In Figure 4.7, the errors in the L_2-norm

$$e_{cheb}^{(n)}(h) \equiv \|f^{(n)}(t) - p^{(n)}(t)\|_{L_2([a,b])} \tag{4.34}$$

are plotted as a function of h_θ for $n \in \{1, 2, 3\}$. The graphs in Figure 4.7 demonstrate that the envelopes of the Chebyshev polynomial derivatives converge with order r, while the derivatives of the spline polynomials would only provide accuracy of order $r - n$. The slopes of the plotted triangles indicate the expected order of convergence.

Comparison to the radial basis function approximation method

We give a numerical comparison of our introduced methodology to the the radial basis function approach discussed in [93]. In that reference, the function $f(x) = \sin(2\pi x)\exp(-x^2)$ is considered and it is assumed that on n uniformly distributed points in the domain $[-2, 2]$ the noisy data is given by $\tilde{f}_i = f_i + 0.001\epsilon \sin(\pi x_i)$ for $i = 1, \ldots, n$. In [93], the case $\epsilon = 1$ is considered. This is a significantly less challenging problem than we are considering since the error varies smoothly from point to point. Wei and Hong [93] use the radial basis function approximation approach and display the root mean square error of the first derivative and n. The root mean square error (RMSE) at N_s arbitrary sampling points $\{x_i\}_{i=1}^{N_s}$ is defined by

$$E(\tilde{f}) \quad = \quad \sqrt{\frac{1}{N_s}\sum_{i=0}^{N_s}(f(x_i) - \tilde{f}(x_i))^2}. \tag{4.35}$$

According to the results presented in [93], the RMSE of the derivative converges up to errors of the order of 0.01 for n large enough. In Figure 4.8, we reproduce the same computation with our method, where $N = 30$ Chebyshev points are used. The RMSE converges to ≈ 0.00223 (much closer to the error in the original function values!) for n large enough, demonstrating a clear superiority of the new approach. We turn our attention to the more challenging problem when ϵ is a uniformly distributed random number in $(0, 1)$. The result is documented on the left plot of Figure 4.8. The RMSE is more oscillatory, yet an excellent convergence of at least 0.005 is obtained. No results for such types of random errors are provided in reference [93].

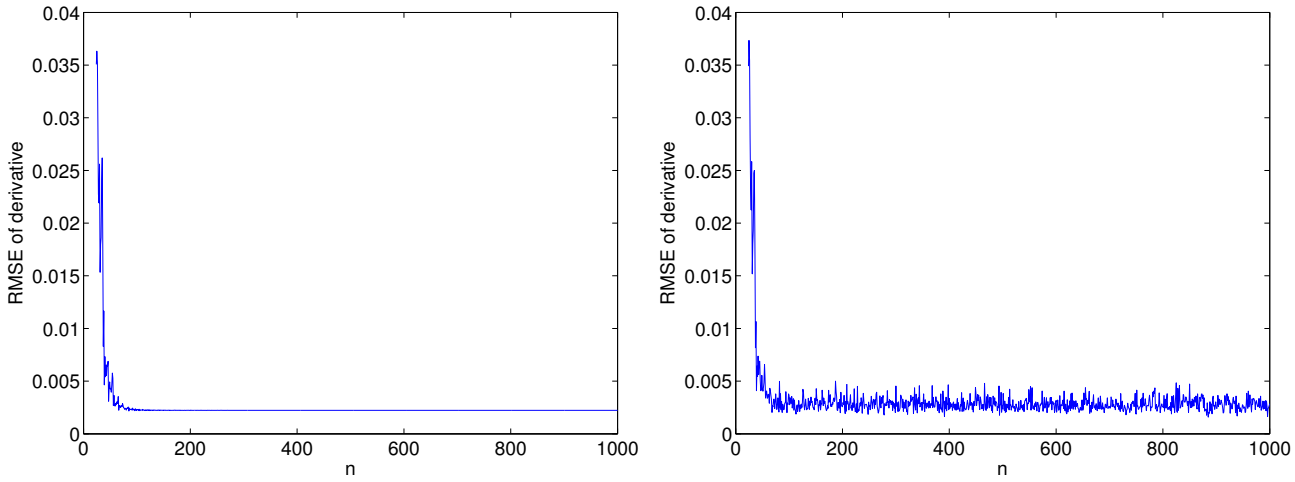


Figure 4.8: Left: Root mean square error versus the number of given sampling points n if $\epsilon = 1$. Right: Same plot as on the left, but this time, ϵ is a uniformly distributed random number in $[0, 1]$.

Reconstructing the gradient from a numerical solution of a partial differential equation

A sphere Γ of radius $R = 0.25$ centered about the origin is located in a cube with the six end-points $[\pm 0.45, \pm 0.45, \pm 0.45]$. The surface of the cube $\mathcal{B}$ and Γ form the outer and inner boundaries of the three-dimensional domain Ω, respectively. We wish to compute the gradient at an arbitrary point in Ω, of the solution of the elliptic partial differential equation

$$\Delta u = f \tag{4.36}$$

$$\boldsymbol{\nu} \cdot \nabla u|_{\Gamma} = g_{\Gamma} \tag{4.37}$$

$$(u + \boldsymbol{\nu} \cdot \nabla u)\,|_{\mathcal{B}} = g_{\mathcal{B}}, \tag{4.38}$$

where $\boldsymbol{\nu}$ is the outer normal of the corresponding surfaces. We solve equations (4.36)–(4.38) with a finite element method. The gradient can be obtained from the solution through the corresponding differentiated local shape functions; as is well known, however, this leads to accuracy loss. Especially, if the gradient data needs to be used as an input for another numerical scheme which is at least of the same order as the finite element method, the result will be generally of an order smaller than the finite elements despite the higher accuracy of the scheme, because the gradient suffers deterioration. For such applications, it is crucial to compute the gradient to the same order of convergence as the solution itself, see [5] for such an example.

The variational formulation of problem (4.36)–(4.38) is:

"Find $u_h \in H_1(\Omega)$, such that

$$(u_h, v)_{\mathcal{B}} + (\nabla u_h, \nabla v)_{\Omega} \quad = \quad (g_{\mathcal{B}}, v)_{\mathcal{B}} + (g_{\Gamma}, v)_{\Gamma} - (f, v)_{\Omega} \tag{4.39}$$

holds for any $v \in H_1(\Omega)$."

We discretize the computational domain Ω by a tetrahedral grid τ with maximum mesh size h and use linear basis function ϕ_i. This leads to the linear system

$$(\boldsymbol{C} + \boldsymbol{K})\, \boldsymbol{u} \quad = \quad \boldsymbol{l}, \tag{4.40}$$

where

$$\boldsymbol{C}_{i,j} = (\phi_i, \phi_j)_{\mathcal{B}}, \quad \boldsymbol{K}_{i,j} = (\phi_i, \phi_j)_{\Omega}, \quad \boldsymbol{l}_i = (g_{\Gamma}, \phi_i)_{\Gamma} + (g_{\mathcal{B}}, \phi_i)_{\mathcal{B}} - (f, \phi_i)_{\Omega}. \tag{4.41}$$

We solve (4.40) with an iterative CG-solver (see [22]) to obtain $\boldsymbol{u}$. For an arbitrary point $\boldsymbol{x}$ in Ω, we select its corresponding element $T_x \in \tau$ and compute the gradient by

$$\nabla_{\phi} u(\boldsymbol{x}) \quad \equiv \quad \sum_{i \in T_x} u_i \nabla \phi_i(\boldsymbol{x}). \tag{4.42}$$

Next, we compute the gradient by the Chebyshev-based approximation algorithm introduced in this section. To do this we use lines parallel to the coordinate axis passing through the point $\boldsymbol{x}$, which either intersect at both ends with $\mathcal{B}$ or intersect at one end with $\mathcal{B}$ and at the other end with Γ, see Figure 4.9 for an example. These three lines form the local coordinate system we use at the point $\boldsymbol{x}$. On each one of its axes, we select the appropriate number of Chebyshev points, determine the corresponding finite element T_c for every Chebyshev point $\boldsymbol{x_c}$ and compute the approximated solution at these points by

$$u_h(\boldsymbol{x_c}) \quad \equiv \quad \sum_{i \in T_c} u_i \phi_i(\boldsymbol{x_c}). \tag{4.43}$$

Once the values at the Chebyshev points are known, we compute the Chebyshev coefficients for every dimension and use (4.28)–(4.30) to obtain the gradient $\nabla_c u$. For our a numerical example we use for the

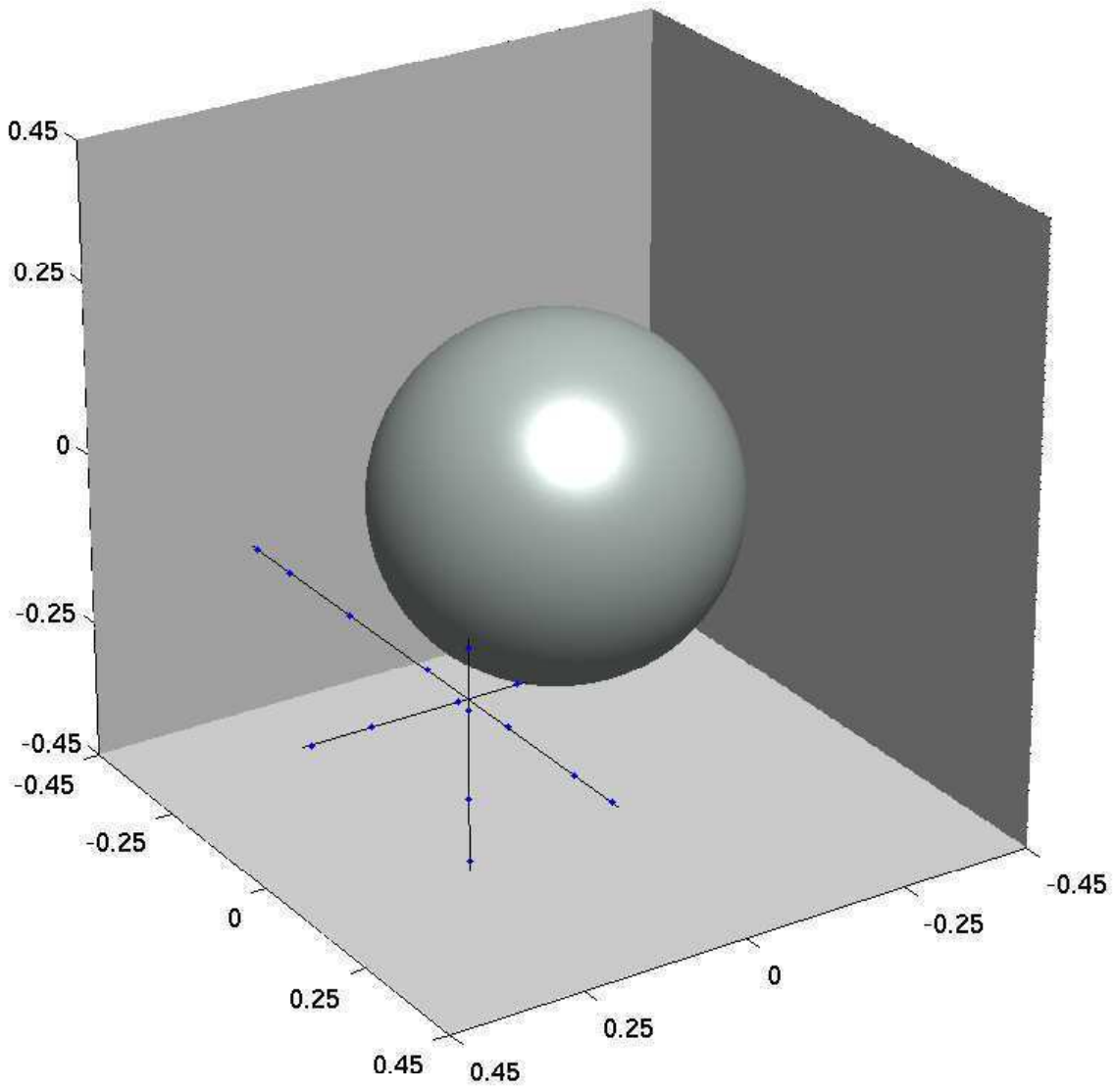


Figure 4.9: The inner boundary Γ and three of the six faces of the outer boundary $\mathcal{B}$ in the background. Inside of Ω, the local coordinate system for the point $[0.2, 0.1, -0.2]$ is constructed. In each dimension, the appropriate number of Chebyshev points is chosen (dots along the corresponding lines).

r.h.s. of equations (4.36)–(4.38) the functions

$$f(\boldsymbol{x}) = e^{|\boldsymbol{x}|^2}\left\{6\cos\left(|\boldsymbol{x}|^2\right)+\sin\left(|\boldsymbol{x}|^2\right)(8|\boldsymbol{x}|^2-6)\right\}, \quad \text{for} \quad \boldsymbol{x}\in\Omega, \tag{4.44}$$

$$g_\Gamma(\boldsymbol{x}) = 2\left(\cos\left(|\boldsymbol{x}|^2\right)-\sin\left(|\boldsymbol{x}|^2\right)\right)e^{|\boldsymbol{x}|^2}\boldsymbol{\nu}\cdot\boldsymbol{x}, \quad \text{for} \quad \boldsymbol{x}\in\Gamma, \tag{4.45}$$

$$g_{\mathcal{B}}(\boldsymbol{x}) = 2\left(\cos\left(|\boldsymbol{x}|^2\right)-\sin\left(|\boldsymbol{x}|^2\right)\right)e^{|\boldsymbol{x}|^2}\boldsymbol{\nu}\cdot\boldsymbol{x}+\cos\left(|\boldsymbol{x}|^2\right)e^{|\boldsymbol{x}|^2}, \quad \text{for} \quad \boldsymbol{x}\in\mathcal{B}, \tag{4.46}$$

so that the exact solution of the system is

$$u(\boldsymbol{x}) = \cos\left(|\boldsymbol{x}|^2\right)e^{|\boldsymbol{x}|^2}, \quad \text{for} \quad \boldsymbol{x}\in\Omega. \tag{4.47}$$

The numerical results in Figure 4.10 display the L_2-error for different mesh sizes of the finite element grid. A clear second-order convergence pattern can be observed for the error of the function and its gradient obtained from the Chebyshev expansion. In contrast, the convergence rate of the gradient evaluated by (4.42) is only accurate to first order.

Conclusion

We have proposed an efficient approach for the evaluation of the derivatives of an approximated function to maximum possible order of convergence. The method is fast and extremely simple, yet powerful. Numerical examples illustrate the higher accuracy and convergence rate over standard methods. This technique can be successfully used to preserve the accuracy of the interior solver in Algorithm 2.3.1 when coupling it with the nonreflecting boundary condition that requires a data input of a derivative (see equation (4.1)) that needs to be computed from the numerical solution of the interior solver. A standard differentiation corrupts the convergence of the overall scheme, because its input data is of lower order. In contrast, the input derivative by our new technique conserves the convergence of the interior scheme, as we demonstrate in Section 4.3.

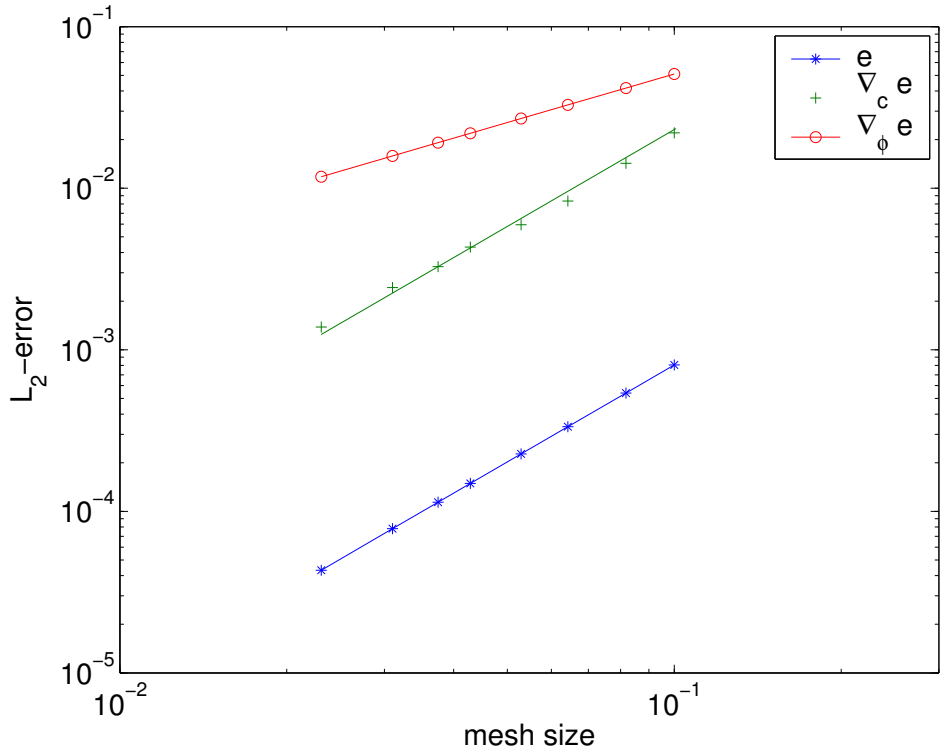


Figure 4.10: Mesh size h versus L_2-error. The improvement in accuracy and convergence rate of the Chebyshev gradient (cross) to the polynomial gradient (circle) is clearly visible: second-order convergence of the finite element solution u_h (star) and the Chebyshev gradient $\nabla_c u_h$, but only first-order convergence of $\nabla_\phi u_h$ to the corresponding exact solutions.

4.1.4 Boundary operator

This section deals with the specification of an appropriate linear operator $\mathcal{L}$ in (2.22). Ting and Miksis proposed in [86] to use Kirchhoff's formula directly as a nonreflecting boundary condition, i.e., these authors chose $\mathcal{L}$ as the identity operator. However, when Givoli and Cohen [38] implemented a numerical scheme that incorporates that boundary condition, they discovered a long-time instability when it is combined with a nondissipative interior finite difference stencil: the numerical solution converges up to a certain time to the correct solution but then a strong instability develops which manifests itself through the appearance of rapidly growing oscillations (see [38]). These authors propose to remove the instability through use of an appropriate dissipative interior scheme. The disadvantage of this approach is that, obviously, it precludes the use of popular and well-understood nondissipative schemes. Therefore, the question arises as to whether there is a more appropriate choice for the operator $\mathcal{L}$ such that the long-time instability does not occur even when a nondissipative interior scheme is used. As it happens, the Sommerfeld-type operator may be the proper choice in (2.22) to avoid the long-time instability. As is known, the Sommerfeld radiation condition at infinity (2.5) eliminates the family of solutions associated with the incoming characteristics and thus insures that the scattering waves which satisfy (2.1)–(2.5) are purely outgoing. It seems intuitively clear that the nonreflecting condition on the finite artificial boundary $\mathcal{B}$ should also have this operator involved. We point out that many proposed transparent boundary conditions are formulated in terms of a Sommerfeld-type operator, and long-time stability has been observed and proven for many of these approaches (see, e.g., [8, 29, 31, 40, 42, 50, 52]). In this section, we present some aspects which demonstrate the impact of the boundary operator on stability.

Example: Impact of different linear boundary operators

We consider the following one-dimensional model problem defined for $(x,t) \in \mathbb{R}_+ \times \mathbb{R}_+$:

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad x > 0, \tag{4.48}$$

$$u(x,0) = \frac{\partial u}{\partial t}(x,0) = 0, \ x > 0, \tag{4.49}$$

$$u(x_\Gamma = 0, t) = g_\Gamma(t). \tag{4.50}$$

The function $g_\Gamma(t)$ in (4.50) is known, and we wish to evaluate the solution numerically in $[0, x_\mathcal{B}]$ for a fixed positive value $x_\mathcal{B} > 0$. This artificial truncation from the semi-infinite domain requires imposing a boundary condition at $x_\mathcal{B}$. The exact local nonreflecting boundary condition at $x_\mathcal{B}$ for problem (4.48)–(4.50) is the well-known Sommerfeld condition

$$\frac{1}{c}\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} \quad = \quad 0, \quad x = x_\mathcal{B}. \tag{4.51}$$

Instead of using (4.51), we wish to use Kirchhoff's formula, which is in the one-dimensional case

$$\mathcal{L}u(x_\mathcal{B}, t) \quad = \quad \mathcal{L}u(x_\mathcal{S}, t - (x_\mathcal{B} - x_\mathcal{S})/c), \quad x = x_\mathcal{B}, \tag{4.52}$$

where we assume that the linear operator $\mathcal{L}$ is either the identity or the Sommerfeld-type operator $\frac{1}{c}\partial_t + \partial_x + \alpha$ for some $\alpha \in \mathbb{R}$ (compare with (4.53) and (4.55), respectively). We note that $\alpha \neq 0$ is needed if a Neumann boundary condition is prescribed at the inner boundary x_Γ: in the case that purely Neumann conditions are imposed on the boundary of the computational domain, the solution is only unique up to a constant; this can be avoided by adding a Dirichlet-type condition on the outer boundary by introducing an $\alpha \neq 0$ value in the corresponding Sommerfeld operator. This simple one-dimensional model problem proves to be very useful to investigate the impact of $\mathcal{L}$ on the stability, since in this case the Kirchhoff formula does not involve integration: if an instability develops in either the presence of a numerical integration rule or when equivalent sources are used, it may be unclear which effect causes the instability. If in turn we observe instability only for certain operators in absence of any integration rule in (4.52), it is clear that the choice of $\mathcal{L}$ lies at the source of the difficulty. In the first case, when $\mathcal{L}$ equals the identity operator, the boundary condition reads

$$u(x_\mathcal{B}, t) \quad = \quad u(x_\mathcal{S}, t - (x_\mathcal{B} - x_\mathcal{S})/c), \quad x = x_\mathcal{B}, \tag{4.53}$$

in the second case, when we deal with the Sommerfeld-type operator with $\alpha = 0$, we have

$$\frac{1}{c}\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} \quad = \quad \frac{1}{c}\frac{\partial u}{\partial y}(x_\mathcal{S}, y)|_{y=t-(x-x_\mathcal{S})/c} - \frac{1}{c}\frac{\partial u}{\partial y}(x_\mathcal{S}, y)|_{y=t-(x-x_\mathcal{S})/c}, \tag{4.54}$$

which simplifies to the Sommerfeld condition (4.51). If $\alpha \neq 0$, the boundary condition is

$$\frac{1}{c}\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} + \alpha u \quad = \quad \alpha u(x_{\mathcal{S}}, t-(x-x_{\mathcal{S}})/c), \quad x = x_{\mathcal{B}}. \tag{4.55}$$

We define an equidistant grid with spacing Δx and Δt and apply central finite differences on (4.55):

$$\frac{u_{j_{\mathcal{B}}}^{n+1} - u_{j_{\mathcal{B}}}^{n-1}}{2\Delta t} + \frac{u_{j_{\mathcal{B}+1}}^{n} - u_{j_{\mathcal{B}-1}}^{n}}{2\Delta x} + \alpha u_{j_{\mathcal{B}}}^{n} = \alpha u_{j_{\mathcal{S}}}^{n-n_{\mathit{diff}}}, \tag{4.56}$$

where $n_{\mathit{diff}} = (j_{\mathcal{B}} - j_{\mathcal{S}})/\lambda$ and the ratio $\lambda \equiv \Delta t/\Delta x$ is chosen such that n_{diff} is an integer. Discretizing equations (4.48)–(4.49) with central finite differences and combining with (4.56), we get the discrete initial conditions ($n = 0$)

$$u_j^0 \quad = \quad 0, \quad u_j^1 = 0, \qquad j \in \{1, \ldots, j_B\}, \tag{4.57}$$

along with the time evolution equations ($n > 1$)

$$u_0^{n+1} \quad = \quad g_\Gamma(t_{n+1}), \tag{4.58}$$

$$u_j^{n+1} \quad = \quad 2\left(1-\lambda^2\right) u_j^n - u_j^{n-1} + \lambda^2 u_{j+1}^n + \lambda^2 u_{j-1}^n, \qquad j \in \{2, \ldots, j_{\mathcal{B}} - 1\}, \tag{4.59}$$

$$u_{j_{\mathcal{B}}}^{n+1} \quad = \quad 2\left(1-\lambda\right) u_{j_{\mathcal{B}}}^n + \frac{\lambda-1}{\lambda+1} u_{j_{\mathcal{B}}}^{n-1} + \frac{2\lambda^2}{\lambda+1} u_{j_{\mathcal{B}}-1}^n + \alpha \frac{2\Delta t \lambda}{\lambda+1}\left(u_{j_{\mathcal{S}}}^{n-n_{\mathit{diff}}} - u_{j_{\mathcal{B}}}^n\right). \tag{4.60}$$

When Kirchhoff's formula (4.53) is used directly as the computational boundary condition, the corresponding discrete system is identical to (4.57)–(4.59), but (4.60) must be replaced by

$$u_{j_{\mathcal{B}}}^{n+1} \quad = \quad u_{j_{\mathcal{S}}}^{n+1-n_{\mathit{diff}}}. \tag{4.61}$$

We choose $\lambda \equiv 1/2, j_{\mathcal{B}} = 21, g_\Gamma = \sin t$, and implement (4.57)–(4.59) along with the open boundary condition (4.61). We observe that an instability develops at $t \approx 8$ if $n_{\mathit{diff}} \neq 1$. In the special case when $n_{\mathit{diff}} \equiv 1$, stability is obtained, as has been discovered and proved in [38]. The numerical solution is plotted in Figure 4.11 for the two cases $n_{\mathit{diff}} \in \{1, 2\}$. Next, we replace (4.61) by (4.60) and find that the numerical system is stable for a suitable choice of the parameter α. Table 4.2 displays the restrictions on α to guarantee stability for $n_{\mathit{diff}} = 1, 2, 3$, and 4. The solution is computed for a large number of time steps (over $5,000,000$ steps), so that we can assume long-time stability if α lies in the corresponding

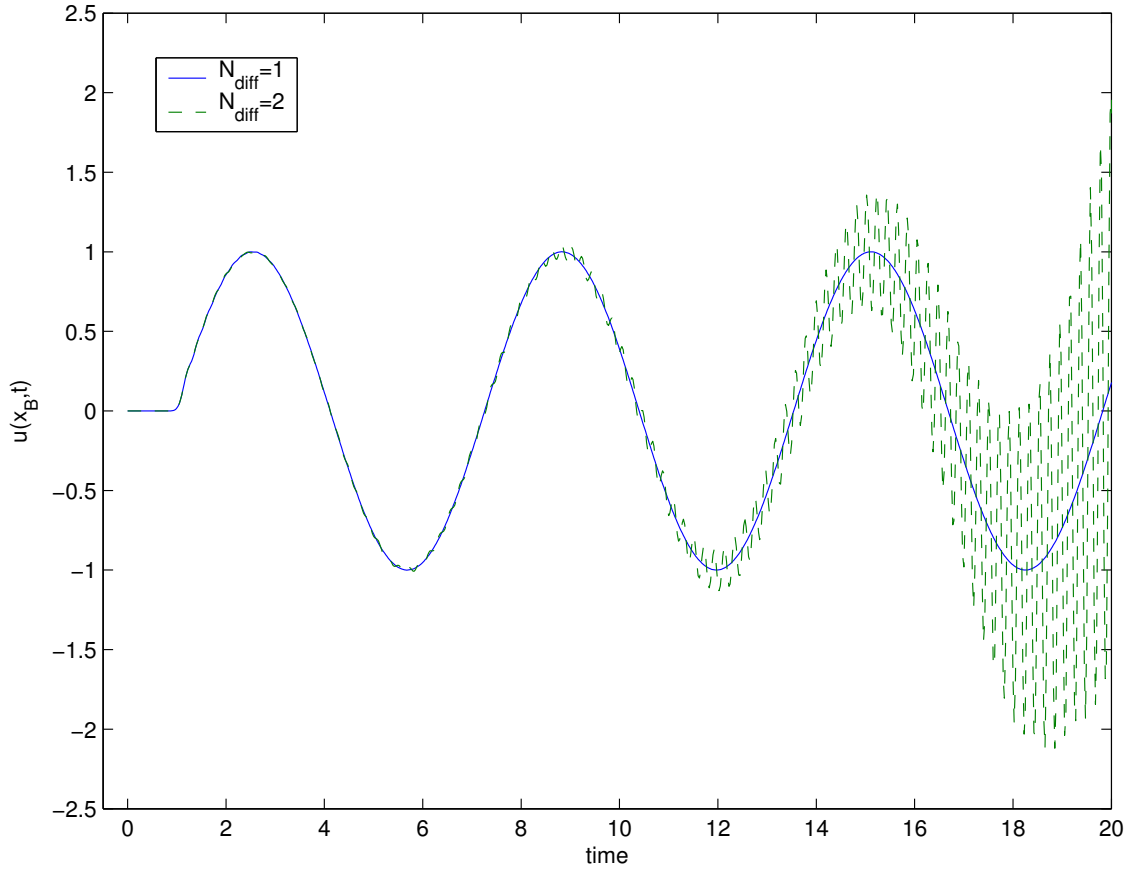


Figure 4.11: Solution $u(x_{\mathcal{B}}, t)$ obtained with boundary condition (4.61). For $n_{diff} = 1$ (blue), the solution is stable, while instability occurs for $n_{diff} = 2$ (green).

range of Table 4.2.

n_{diff}	Stability if $\alpha \in$
1	$[-49, 316]$
2	$[-24, 203]$
3	$[-16, 96]$
4	$[-12, 48]$

Table 4.2: As the distance $x_{\mathcal{B}} - x_{\mathcal{S}}$ increases, the domain of definition for α decreases to insure stability.

Total energy of the system

A proof of stability of a numerical method can often be given on the basis of considering the total energy contained in the solution. The boundedness of the energy of a certain scheme at any time under the assumption that no energy enters the computational domain implies stability. We thus consider the total energy $E(t) = 1/2(\int_{x_\Gamma}^{x_{\mathcal{B}}} (\partial_t u)^2 + 1/c^2 (\nabla u)^2 dx)$ of the system when there is no external forcing and no wave enters the domain $[x_\Gamma, x_{\mathcal{B}}]$. The solution $u(x,t)$, which results solely from the initial conditions $u_0(t), \dot{u}_0(t)$ that are compactly supported in $[x_\Gamma, x_{\mathcal{B}}]$, can freely travel through the boundaries $x_\Gamma = 0$ and $x_{\mathcal{B}} = 1$ and eventually leave the computational domain if nonreflecting boundary conditions are imposed at x_Γ and $x_{\mathcal{B}}$. Taking the derivative of $E(t)$, making use of the homogeneous wave equation and integration by parts shows that the change in the total energy is given by

$$E'(t) \quad = \quad c^2 \frac{\partial u}{\partial t} \frac{\partial u}{\partial x}\Big|_{x_\Gamma}^{x_{\mathcal{B}}}. \tag{4.62}$$

In analogy to (4.55), the modified Sommerfeld boundary condition at x_Γ reads

$$\frac{1}{c}\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x} + \tilde{\alpha} u \quad = \quad \tilde{\alpha} u(x_{\tilde{\mathcal{S}}}, t + (x - x_{\tilde{\mathcal{S}}})/c), \quad x = x_\Gamma, \tag{4.63}$$

and the discretized model problem ($c = 1$) is

$$u_j^{n+1} \quad = \quad 2\left(1 - \lambda^2\right) u_j^n - u_j^{n-1} + \lambda^2 u_{j+1}^n + \lambda^2 u_{j-1}^n, \qquad j \in \{2, \ldots, j_{\mathcal{B}} - 1\}, \tag{4.64}$$

$$u_{j_{\mathcal{B}}}^{n+1} \quad = \quad 2\left(1 - \lambda\right) u_{j_{\mathcal{B}}}^n + \frac{\lambda - 1}{\lambda + 1} u_{j_{\mathcal{B}}}^{n-1} + \frac{2\lambda^2}{\lambda + 1} u_{j_{\mathcal{B}}-1}^n + \alpha \frac{2\Delta t \lambda}{\lambda + 1} \left(u_{j_{\mathcal{S}}}^{n - n_{diff}} - u_{j_{\mathcal{B}}}^n\right), \tag{4.65}$$

$$u_{j_\Gamma}^{n+1} \quad = \quad 2\left(1 - \lambda\right) u_{j_\Gamma}^n + \frac{\lambda - 1}{\lambda + 1} u_{j_\Gamma}^{n-1} + \frac{2\lambda^2}{\lambda + 1} u_{j_\Gamma+1}^n + \tilde{\alpha} \frac{2\Delta t \lambda}{\lambda + 1} \left(u_{j_{\tilde{\mathcal{S}}}}^{n - \tilde{n}_{diff}} - u_{j_\Gamma}^n\right), \tag{4.66}$$

with the two given initial conditions u_j^0 and u_j^1 for $j \in \{1, \ldots, j_{\mathcal{B}}\}$. Dizcretizing (4.62) with central differences, the numerical energy thus satisfies

$$\frac{E^{n+1} - E^{n-1}}{2\Delta t} = \frac{u_{j_{\mathcal{B}}}^{n+1} - u_{j_{\mathcal{B}}}^{n-1}}{2\Delta t} \frac{u_{j_{\mathcal{B}}+1}^{n} - u_{j_{\mathcal{B}}-1}^{n}}{2\Delta x} - \frac{u_{j_\Gamma}^{n+1} - u_{j_\Gamma}^{n-1}}{2\Delta t} \frac{u_{j_\Gamma+1}^{n} - u_{j_\Gamma-1}^{n}}{2\Delta x}. \tag{4.67}$$

Substituting the discretized boundary conditions (4.65) and (4.66) into (4.67) and multiplying by $2\Delta t$ gives

$$\begin{aligned} E^{n+1} - E^{n-1} = & -\frac{(u_{j_{\mathcal{B}}}^{n+1} - u_{j_{\mathcal{B}}}^{n-1})^2 + (u_{j_\Gamma}^{n+1} - u_{j_\Gamma}^{n-1})^2}{2\Delta t} + \\ & \alpha \left(u_{j_{\mathcal{B}}}^{n+1} - u_{j_{\mathcal{B}}}^{n-1}\right) \left(u_{j_s}^{n-n_{diff}} - u_{j_{\mathcal{B}}}^{n}\right) + \tilde{\alpha} \left(u_{j_\Gamma}^{n+1} - u_{j_\Gamma}^{n-1}\right) \left(u_{j_{\tilde{s}}}^{n-\tilde{n}_{diff}} - u_{j_\Gamma}^{n}\right). \end{aligned} \tag{4.68}$$

If we choose $\alpha = \tilde{\alpha} \equiv 0$, i.e., we consider the Sommerfeld condition, we easily see that

$$E^{n+1} - E^{n-1} \quad \leq \quad 0, \tag{4.69}$$

and summing from $n = 1$ to $n = N-1$ leads to $E^N + E^{N-1} \leq E^0 + E^1$, which implies numerical stability. Thus, the condition (4.69) is sufficient for stability.

If $\alpha \neq 0, \tilde{\alpha} \neq 0$, $\left(u_{j_s}^{n-n_{diff}} - u_{j_{\mathcal{B}}}^{n}\right) \neq 0$, and $\left(u_{j_{\tilde{s}}}^{n-\tilde{n}_{diff}} - u_{j_\Gamma}^{n}\right) \neq 0$, the following conditions on α and $\tilde{\alpha}$ suffice for stability:

$$\alpha \leq \frac{u_{j_{\mathcal{B}}}^{n+1} - u_{j_{\mathcal{B}}}^{n-1}}{2\Delta t \left(u_{j_s}^{n-n_{diff}} - u_{j_{\mathcal{B}}}^{n}\right)} \tag{4.70}$$

and

$$\tilde{\alpha} \leq \frac{u_{j_\Gamma}^{n+1} - u_{j_\Gamma}^{n-1}}{2\Delta t \left(u_{j_{\tilde{s}}}^{n-\tilde{n}_{diff}} - u_{j_\Gamma}^{n}\right)}. \tag{4.71}$$

Quantitatively, we expect that the differences in $\left(u_{j_s}^{n-n_{diff}} - u_{j_{\mathcal{B}}}^{n}\right)$ and $\left(u_{j_{\tilde{s}}}^{n-\tilde{n}_{diff}} - u_{j_\Gamma}^{n}\right)$ increase as n_{diff} increases, which leads to more restrictive choices for α and $\tilde{\alpha}$ for larger n_{diff}. Table 4.2 confirms this behavior. If a purely Neumann condition is prescribed at the inner boundary condition, Lemma 4.1.1 shows that $\alpha > 0$ is required to insure uniqueness.

Uniqueness

In the following lemma, we consider a three-dimensional domain Ω, which is internally bounded by the surface of the scatterer Γ and externally by $\mathcal{B}$. We show that if purely Neumann conditions are prescribed on Γ, the α in the boundary operator is required to be positive to guarantee a unique solution. The proof also translates directly to the one-dimensional case.

Lemma 4.1.1. *Let $\alpha(\boldsymbol{x}) > 0$. Then the PDE*

$$\frac{\partial^2}{\partial t^2} u = \Delta u + f \tag{4.72}$$

$$\frac{\partial}{\partial \nu} u|_\Gamma = g_\Gamma(t), \quad \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \nu} + \alpha(\boldsymbol{x}) \right) u|_{\mathcal{B}} = g_{\mathcal{B}}(t) \tag{4.73}$$

$$u(\boldsymbol{x}, 0) = u_0, \qquad \frac{\partial}{\partial t} u(\boldsymbol{x}, 0) = \dot{u}_0, \tag{4.74}$$

defined for $(\boldsymbol{x}, t) \in \Omega \times [0, \infty)$, admits a unique solution.

Proof. Let us assume that there are two different solutions u_1 and u_2 that solve (4.72)–(4.74). Then, the difference function $w \equiv u_1 - u_2$ solves the homogeneous system of (4.72)–(4.74), and the time derivative of its total energy is

$$\frac{d}{dt} E_\Omega(t) = (\partial_{tt} w, \partial_t w)_\Omega + (\nabla \partial_t w, \nabla w)_\Omega. \tag{4.75}$$

The first term on the r.h.s in (4.75) equals to $(\Delta w, \partial_t w)_\Omega$, and integration by parts reveals that $(\partial_{tt} w, \partial_t w)_\Omega = (\partial_\nu w, \partial_t w)_{\partial\Omega} - (\nabla w, \nabla \partial_t w,)_\Omega$. Therefore,

$$\frac{d}{dt} E_\Omega(t) = (\partial_\nu w, \partial_t w)_{\partial\Omega}\,, \tag{4.76}$$

and using the fact that the Neumann condition on Γ vanishes, we even have $\frac{d}{dt} E_\Omega(t) = (\partial_\nu w, \partial_t w)_{\mathcal{B}}$. We replace $\partial_\nu u|_{\mathcal{B}}$ by $-(\partial_t + \alpha) u|_{\mathcal{B}}$, integrate in time from zero to $T > 0$ and obtain

$$E_\Omega(T) = E_\Omega(0) - \int_0^T (\partial_t w, \partial_t w)_{\mathcal{B}}\, dt - \frac{1}{2} \left\{ (\alpha w, w)_{\mathcal{B}}[T] - (\alpha w, w)_{\mathcal{B}}[0] \right\}. \tag{4.77}$$

Clearly, the first and the last term on the r.h.s vanish because of the homogeneous initial condition, and

we are left with

$$E_\Omega(T) \;=\; -\int_0^T (\partial_t w, \partial_t w)_\mathcal{B}\; dt - \frac{1}{2}(\alpha w, w)_\mathcal{B}[T]. \tag{4.78}$$

Obviously, $E(T) \geq 0$ by definition, but $E(T) \leq 0$ by (4.78). Therefore the total energy of w has to vanish for all times, and (4.78) leads to the conditions $\partial_t w|_\mathcal{B} = 0$ and $w|_\mathcal{B} = 0$. If $\alpha = 0$, only the first condition holds, which would yield $w|_\mathcal{B} = c(\boldsymbol{x}_\mathcal{B})$ only. But for $\alpha(\boldsymbol{x}_B) > 0, w|_\mathcal{B} = 0$, and we can apply the analytic continuation principle to find $w(\boldsymbol{x}, t) = 0$ for any t and $\boldsymbol{x} \in \Omega$. Thus, $u_1 = u_2$, which proves uniqueness.

□

Remark 4.1.1.

1. If $\alpha(\boldsymbol{x}, t)$ depends on $\boldsymbol{x}$ *and* t, a sufficient condition for uniqueness is

$$-\alpha(\boldsymbol{x}, t)u(\boldsymbol{x}, t) \leq \partial_t u(\boldsymbol{x}, t) \quad \text{for } \boldsymbol{x} \in \mathcal{B} \text{ and } t \in \mathbb{R}_+. \tag{4.79}$$

2. Combining (4.76) with (4.73), we see that a sufficient condition for stability is

$$g_\mathcal{B}(t) - \alpha(\boldsymbol{x}, t)u(\boldsymbol{x}, t) \leq \partial_t u(\boldsymbol{x}, t), \quad \boldsymbol{x} \in \mathcal{B}. \tag{4.80}$$

3. Many computational boundary conditions proposed previously, which don't exhibit long-time instability, involve the operator $\mathcal{L} = \frac{1}{c}\partial_t + \partial_r + \frac{1}{r}$ when $\mathcal{B}$ is a sphere. We conjecture that for a convex computational domain Ω, when $\boldsymbol{\nu} \cdot \boldsymbol{x} \geq 0$, the operator

$$\mathcal{L}_\alpha \;\equiv\; \frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial \nu} + \frac{\boldsymbol{\nu} \cdot \boldsymbol{x}}{|\boldsymbol{x}|^2} \tag{4.81}$$

gives rise to stability. If a Dirichlet boundary condition is imposed on Γ, we conjecture that the operator

$$\mathcal{L}_S \;\equiv\; \frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial \nu} \tag{4.82}$$

implies stability.

4.2 Interior solver

4.2.1 Variational formulation

We denote the Sobolev space by

$$H^1(\Omega) = \left\{ u \in L^2(\Omega) \mid \frac{\partial u}{\partial x_i} \in L^2(\Omega) \right\}, \quad 1 \leq i \leq 3. \tag{4.83}$$

For $u(t) \in H^1(\Omega)$ and $v \in H^1(\Omega)$, we define $(u,v)_\Omega \equiv \int_\Omega u(\boldsymbol{x},t)\overline{v(\boldsymbol{x})}d\boldsymbol{x}$, $(u,v)_\mathcal{B} \equiv \int_\mathcal{B} u(\boldsymbol{x},t)\overline{v(\boldsymbol{x})}dS_x$ and $(u,v)_\Gamma \equiv \int_\Gamma u(\boldsymbol{x},t)\overline{v(\boldsymbol{x})}dS_x$. Next, we multiply (2.18) by a test function $\overline{v}$, integrate by parts and use (2.19)–(2.22) to obtain the following variational formulation:

Find $u(t) \in H^1(\Omega)$, such that for all $v \in H^1(\Omega)$, we have

$$\left(\frac{\partial^2 u}{\partial t^2}, v\right)_\Omega + \left(\mathcal{L}u - \frac{\partial u}{\partial \nu}, v\right)_\mathcal{B} + (\nabla u, \nabla v)_\Omega = (f,v)_\Omega + (g_\Gamma, v)_\Gamma + (g_{\mathcal{B},h}, v)_\mathcal{B} \tag{4.84}$$

$$(u(0), v)_\Omega = (u_0, v)_\Omega\,, \quad \big(u'(0), v\big)_\Omega = (\dot{u}_0, v)_\Omega\,. \tag{4.85}$$

4.2.2 Finite element formulation

The computational domain is discretized by a mesh τ, which approximates Ω with a finite number of elements K. The finite element space

$$S^{p,1}(\Omega,\tau) = \big\{ u \in H^1(\Omega) \mid u|_K \in \mathcal{P}_p \text{ for } K \in \tau \big\} \tag{4.86}$$

contains all functions in $H^1(\Omega)$ whose restrictions to $K \in \tau$ are polynomials of order p. Replacing the Sobolev space by (4.86), the unknown wave field and the test function can be approximated by

$$u_h(\boldsymbol{x},t) = \sum_{k=1}^{G} u_k(t)\phi_k(\boldsymbol{x}), \qquad v_h(\boldsymbol{x}) = \sum_{k=1}^{G} v_k \phi_k(\boldsymbol{x}) \tag{4.87}$$

where $\{\phi_k(\boldsymbol{x})\}_{k=1}^G$ span the discretized Sobolev space (4.86). The finite element discretization reads:

Find $u_h(t) \in S^{1,p}(\Omega,\tau)$ such that for all $v_h \in S^{1,p}(\Omega,\tau)$

$$\left(\frac{\partial^2 u_h}{\partial t^2}, v_h\right)_\Omega + \left(\mathcal{L}u_h - \frac{\partial u_h}{\partial \nu}, v_h\right)_\mathcal{B} + (\nabla u_h, \nabla v_h)_\Omega = (f, v_h)_\Omega + (g, v_h)_\Gamma + (g_{\mathcal{B},h}, v_h)_\mathcal{B} \tag{4.88}$$

$$(u_h(0), v_h)_\Omega = (u_0, v_h)_\Omega, \qquad (u_h'(0), v_h)_\Omega = (\dot{u}_0, v_h)_\Omega. \tag{4.89}$$

Equations (4.88) and (4.89) are equivalent to the linear second-order system

$$\boldsymbol{M}\ddot{\boldsymbol{u}} + \boldsymbol{C}\dot{\boldsymbol{u}} + (\boldsymbol{K} + \tilde{\boldsymbol{M}})\boldsymbol{u} \;=\; \boldsymbol{l}(t) \tag{4.90}$$

for $\boldsymbol{u}(t) \equiv [u_1(t), \ldots, u_G(t)]^t \in \mathbb{R}^{G\times 1}$ with the initial conditions

$$\boldsymbol{M}\boldsymbol{u}(0) = \boldsymbol{l}_1, \qquad \boldsymbol{M}\dot{\boldsymbol{u}}(0) = \boldsymbol{l}_2. \tag{4.91}$$

The matrices $\boldsymbol{M}$ and $\boldsymbol{K}$ in (4.90) are known as the mass and stiffness matrix, respectively. Their entries are

$$\boldsymbol{M}_{k,l} \;=\; (\phi_k, \phi_l)_\Omega \tag{4.92}$$

$$\boldsymbol{K}_{k,l} \;=\; (\nabla\phi_k, \nabla\phi_l)_\Omega. \tag{4.93}$$

We remark that both matrices are sparse due to the local definitions of the basis functions. Further, the mass matrix $\boldsymbol{M}$ is symmetric positive definite, while the stiffness matrix $\boldsymbol{K}$ is generally only positive semi-definite. However, if Dirichlet boundary condition are prescribed on a part of Γ or $\mathcal{B}$, Poincaré's theorem states that the matrix $\boldsymbol{K}$ in this particular case is even positive definite.

The matrix $\boldsymbol{C}$ in (4.90) takes the form

$$\boldsymbol{C}_{k,l} \;=\; (\phi_k, \phi_l)_\mathcal{B}. \tag{4.94}$$

This matrix is symmetric positive semi-definite and acts as a damping term.

Finally, the matrix $\tilde{\boldsymbol{M}}$ in (4.90) is only nonvanishing if $\mathcal{L}$ is given by (4.81), in which case its entries can be expressed by

$$\tilde{\boldsymbol{M}}_{k,l} \;=\; \left(\frac{\boldsymbol{\nu}\cdot\boldsymbol{x}}{|\boldsymbol{x}|^2}\phi_k, \phi_l\right)_\mathcal{B}. \tag{4.95}$$

That matrix is symmetric positive semi-definite if $\boldsymbol{\nu} \cdot \boldsymbol{x} \geq 0$, which is the case for convex boundaries $\mathcal{B}$. Contrary to the matrices (4.92), (4.93), (4.94) and (4.95), the load vector

$$l_k(t) \quad = \quad (f, \phi_k)_\Omega + (g, \phi_k)_\Gamma + (g_{\mathcal{B},h}, \phi_k)_\mathcal{B} \tag{4.96}$$

is time dependent and thus needs to be updated at every time-step. The linear semi-discrete system (4.90)–(4.91) needs to be solved with an appropriate time-marching scheme, which is discussed in the next section.

4.2.3 Time-marching schemes

There are many suitable time-marching schemes, implicit as well as explicit ones, that can be applied successfully to (4.90)–(4.91). The explicit methods are easy to implement and the solution is obtained with a direct solver; however, restrictions on the time step apply to insure stability. In many cases, only a relatively small time step can be chosen to satisfy the CFL-condition. Implicit methods can have a much larger time step and in many cases no restrictions apply to guarantee stability. The discrete equations need to be obtained through the solution of a linear system, making the computational cost for one time step larger than in an explicit scheme. Therefore, it is important to find out for the specific system to be solved what method works best. Here, we give an example of an explicit and an implicit solver. As a representative for the explicit classes, we consider the leap-frog scheme, which has been very popular in connection with the wave equation. Then, we present an implicit method which is particularly useful with our nonreflecting boundary condition.

Explicit method

We select the leap-frog method to solve (4.90)–(4.91) in time. The time-step Δt needs to be chosen sufficiently small compared to the minimum mesh size $h_{\min}$ of the finite element grid τ to satisfy the CFL-condition of numerical stability, i.e.,

$$\lambda \equiv \frac{\Delta t}{h_{\min}} \leq C_\lambda \tag{4.97}$$

for an appropriate constant $C_\lambda < 1$. The solution vector $\boldsymbol{u}^k$ corresponds to the time $t^k = k\Delta t$. Then the basic step to advance the numerical solution in time for $k = 1, 2, \ldots$ is

$$\boldsymbol{A}\boldsymbol{u}^{k+1} = \boldsymbol{l}(t^k) + \boldsymbol{B}_1\boldsymbol{u}^k + \boldsymbol{B}_2\boldsymbol{u}^{k-1}, \tag{4.98}$$

where

$$\boldsymbol{A} = \left(\frac{1}{\Delta t^2}\boldsymbol{M} + \frac{1}{2\Delta t}\boldsymbol{C}\right), \tag{4.99}$$

and

$$\boldsymbol{B}_1 = \left(\frac{2}{\Delta t^2}\boldsymbol{M} - (\boldsymbol{K} + \tilde{\boldsymbol{M}})\right), \quad \boldsymbol{B}_2 = \left(\frac{1}{2\Delta t}\boldsymbol{C} - \frac{1}{\Delta t^2}\boldsymbol{M}\right). \tag{4.100}$$

The matrix $\boldsymbol{A}$ in (4.99) is symmetric positive definite, and therefore invertible. Also, it is sparse due to the locality of the basis functions $\{\phi_k\}_{k=1}^G$.

Implicit method

To avoid the restriction (4.97), we consider the implicit scheme

$$\boldsymbol{M}\delta_t^2\boldsymbol{u}_k + \boldsymbol{K}\boldsymbol{u}_{k,\theta} + \tilde{\boldsymbol{M}}\boldsymbol{u}_{k,\theta} + \boldsymbol{C}\delta_t\boldsymbol{u}_k = \boldsymbol{l}_{k,\theta}, \tag{4.101}$$

with the definitions

$$\delta_t^2\boldsymbol{u}_k = (\boldsymbol{u}_{k+1} - 2\boldsymbol{u}_k + \boldsymbol{u}_{k-1})/(\Delta t)^2, \tag{4.102}$$

$$\delta_t\boldsymbol{u}_k = (\boldsymbol{u}_{k+1} - \boldsymbol{u}_{k-1})/(2\Delta t), \tag{4.103}$$

$$\boldsymbol{u}_{k,\theta} = \theta\boldsymbol{u}_{k+1} + (1 - 2\theta)\boldsymbol{u}_k + \theta\boldsymbol{u}_{k-1}. \tag{4.104}$$

As noted in [28], this second-order correct approximation in Δt is unconditionally stable for $\theta \geq 1/4$, and the choice $\theta = 1/4$ minimizes the time truncation error over this class of methods. With this specific choice, equation (4.101) can be rewritten as

$$\boldsymbol{A}_1\boldsymbol{u}_{k+1} = \boldsymbol{B}_1\boldsymbol{u}_k + \boldsymbol{B}_2\boldsymbol{u}_{k-1} + \left(\frac{1}{4}\boldsymbol{l}(t_{k+1}) + \frac{1}{2}\boldsymbol{l}(t_k) + \frac{1}{4}\boldsymbol{l}(t_{k-1})\right). \tag{4.105}$$

The sparse matrix

$$\boldsymbol{A}_1 \;=\; \boldsymbol{A} + \frac{1}{4}\left(\boldsymbol{K} + \tilde{\boldsymbol{M}}\right) \tag{4.106}$$

is positive definite. We note that our nonreflecting boundary condition is particularly suitable for the scheme in (4.105). Indeed, the $l(t_{k+1})$ in (4.105) can be easily obtained, since the boundary values are precomputed for a certain time-interval. Therefore, equation (4.105) can be solved directly.

4.3 Numerical examples

In this section, we provide numerical results for the overall scattering solver, Algorithm 2.3.1. In Section 4.3.1, we assume that the scatterer is a sphere and the field is generated by a point source on the z-axis. This makes the problem axis-symmetric, and we can reduce the computation in the three-dimensional domain to two dimensions. This model is particularly useful to study the performance of the *fully three-dimensional* artificial boundary condition at minimum computational cost. The evaluation of the nonreflecting boundary condition based on the equivalent sources is not reduced in dimensionality. The results show that the developed solver satisfies the expected accuracy and exhibits long-time stability. In Section 4.3.2, we present computations for elongated obstacles, for which the nonreflecting boundary evaluator may be especially advantageous over other approaches: many artificial boundary conditions require a large domain Ω either to obtain a prescribed accuracy, or the boundary must be of a special shape, which often makes the definition of the computational domain Ω unnecessary large. In our method, in contrast, the artificial boundary can be in principle as close to the scatterer as desired without compromising accuracy. In Section 4.3.3 we present fully three-dimensional computations showing that the computational times required by the nonreflecting boundary condition algorithm is only a small portion of the overall computation.

4.3.1 Spherical obstacle

In this section we provide some numerical examples of Algorithm 2.3.1 that make use of the explicit time-marching scheme (4.98) to discretize (2.18)–(2.22). The obstacle Γ is the sphere of radius r_s centered about the origin, and the artificial boundary $\mathcal{B}$ is the cylinder with radius r_c and height $2r_c$ along the z-axis from $[-r_c, r_c]$. The source strength $f(\boldsymbol{x}, t)$ is defined in (3.5) and (3.11). The exact solution to the

problem is given by (3.6) as long as the point source lies strictly within the scatterer. In this case, the wave is purely outgoing and the numerical solution approximates the given exact solution provided these functions match on Γ. Assuming that the field and its time derivative are zero initially, the wave enters the computational domain Ω through the inner boundary Γ, since there is no forcing outside the sphere. The Kirchoff's surface $\mathcal{S}$ is chosen as the cube with edge length $2r_s$ which includes the sphere entirely, see Figure 4.12. We note that if $\boldsymbol{x}_0$ in (3.5) is positioned on the z-axis, the problem is axis-symmetric and the three-dimensional finite element computation in Ω can be reduced to two dimensions in the cylindrical coordinates r and z, i.e.,

$$r = \sqrt{x^2 + y^2}, \qquad z = z; \tag{4.107}$$

as mentioned earlier, results of a fully three-dimensional computations are presented in Section 4.3.3. A two-dimensional finite element mesh in cylindrical coordinates (r, z) of this form is depicted in Figure 4.12. For any angle $\theta = \arctan \frac{y}{x}$ the field $u(r, z, t)$ has the same value for fixed r, z, and t. In this particular case, the entries of equations (4.92) to (4.96) take the form

$$\boldsymbol{M}_{k,l} = \int_{\Omega_{r,z}} \phi_k(r,z)\phi_l(r,z) r dr dz \tag{4.108}$$

$$\boldsymbol{K}_{k,l} = \int_{\Omega_{r,z}} \left\{ \frac{\partial \phi_k}{\partial r}(r,z)\frac{\partial \phi_l}{\partial r}(r,z) + \frac{\partial \phi_k}{\partial z}(r,z)\frac{\partial \phi_l}{\partial z}(r,z) \right\} r dr dz \tag{4.109}$$

$$\boldsymbol{C}_{k,l} = \int_{\mathcal{B}_{r,z}} \phi_k(r,z)\phi_l(r,z) r dS_{r,z} \tag{4.110}$$

$$\tilde{\boldsymbol{M}}_{k,l} = \int_{\mathcal{B}_{r,z}} \phi_k(r,z)\phi_l(r,z)\frac{r r_c}{(r^2+z^2)} dS_{r,z} \tag{4.111}$$

$$\begin{aligned} \boldsymbol{l}_k(t) = & \int_{\Omega_{r,z}} f(r,z,t)\phi_k(r,z) r dr dz + \int_{\Gamma_{r,z}} g_\Gamma(r,z,t)\phi_k(r,z) r dS_{r,z} \\ & + \int_{\mathcal{B}_{r,z}} g_{\mathcal{B},h}(r,z,t)\phi_k(r,z) r dS_{r,z}. \end{aligned} \tag{4.112}$$

The reduction of the interior computation to a two-dimensional domain helps us to keep the overall number of unknowns low, so that a direct sparse solver works well to invert the system matrix $\boldsymbol{A}$. In the

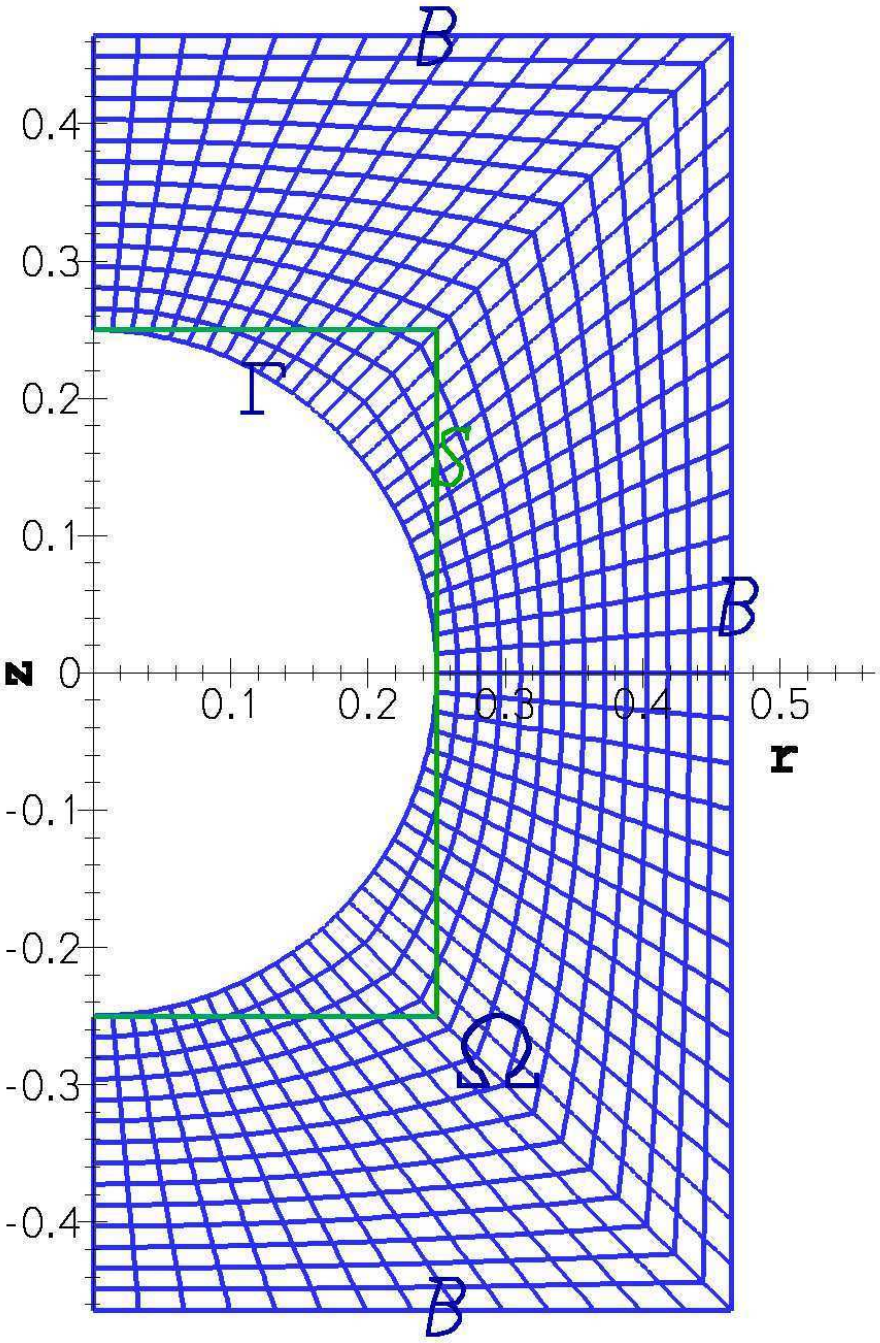


Figure 4.12: Blue: Finite element mesh in cylindrical coordinates (r, z) which discretizes the computational domain Ω. Green: Intersection of the finite element grid with the Kirchhoff's surface $\mathcal{S}$

computations of this section, we specifically use the SuperLU solver, which can be freely downloaded from http://crd.lbl.gov/∼xiaoye/SuperLU.

For the shape functions associated with (4.86), we select the serendipity space (see [81]). We define a family of finite element meshes $\{m_k\}_{k=0}^{8}$ which differ in their maximum mesh sizes h, where m_0 is the coarsest and m_8 the finest grid. Clearly, h needs to be small enough to resolve the minimum wave length. As a "rule of thumb," $kh \equiv 2\pi h/\lambda$ should be no larger than 0.6 which means that every relevant wave length is represented on the finite element mesh by at least 10 points (see [54]); note, however, that owing to pollution effects, the number of points per wavelength needs to be increased as the total acoustic size of the scatterer increases. The mesh size $h_{\min}$ denotes the shortest edge on a finite element grid and $\Delta t = C_\lambda h_{\min}$ is the time step in the leap-frog scheme with the constant value $C_\lambda = 0.1$. We define the instantaneous errors $e_G(t)$ and $e_2(t)$ as

$$e_G(t) \equiv \|\nabla u(.,t) - \nabla u_h(.,t)\|_{l_2(\Omega)} \tag{4.113}$$

$$e_2(t) \equiv \|u(.,t) - u_h(.,t)\|_{l_2(\Omega)}, \tag{4.114}$$

and denote by $e_{G,T}$ and $e_{2,T}$ their maximum value over the time interval $[0,T]$:

$$e_{G,T} \equiv \max_{t\in[0,T]} e_G(t) \tag{4.115}$$

$$e_{2,T} \equiv \max_{t\in[0,T]} e_2(t). \tag{4.116}$$

Demonstration of long-time instability if Neumann boundary operator is used

The numerical results presented in this section are computed using the parameters $c = 1, r_s = 0.25, H = 2{\cdot}r_s/5, r_c = r_s{+}2H, \boldsymbol{x}_0 = [0,0,0]^t, \sigma = 0.5, t_0 = 2.0, S = 5, C = 5, l_{\min} = r_c{-}r_s, l_{\max} = \sqrt{3}(r_s{+}r_c), l_{BF} = l_{\max} - l_{\min} + 0.1, N_t = 32, N_z = 128$ if not stated otherwise. We turn our attention to the choice of the operator $\mathcal{L}$ in (2.22). The blue curves on the upper left of Figure 4.13 display the numerical values of the wave field at the point $\boldsymbol{x}_3 \equiv [0, 0.4, 0]^t$ when the Neumann operator $\mathcal{L}_\nu = \partial_\nu$ is used in (2.22). In the lower left picture, the function $s(t) = \sin(2t)$ is used instead. The green curves are the true solutions. On the right figures the corresponding errors $u_h(\boldsymbol{x}_3, t) - u_{ex}(\boldsymbol{x}_3, t)$ are depicted for $t \in [0, 42]$. As in the examples in reference [38], the solution is accurate at first, but after a certain time an instability develops; compare also with Section 4.1.4.

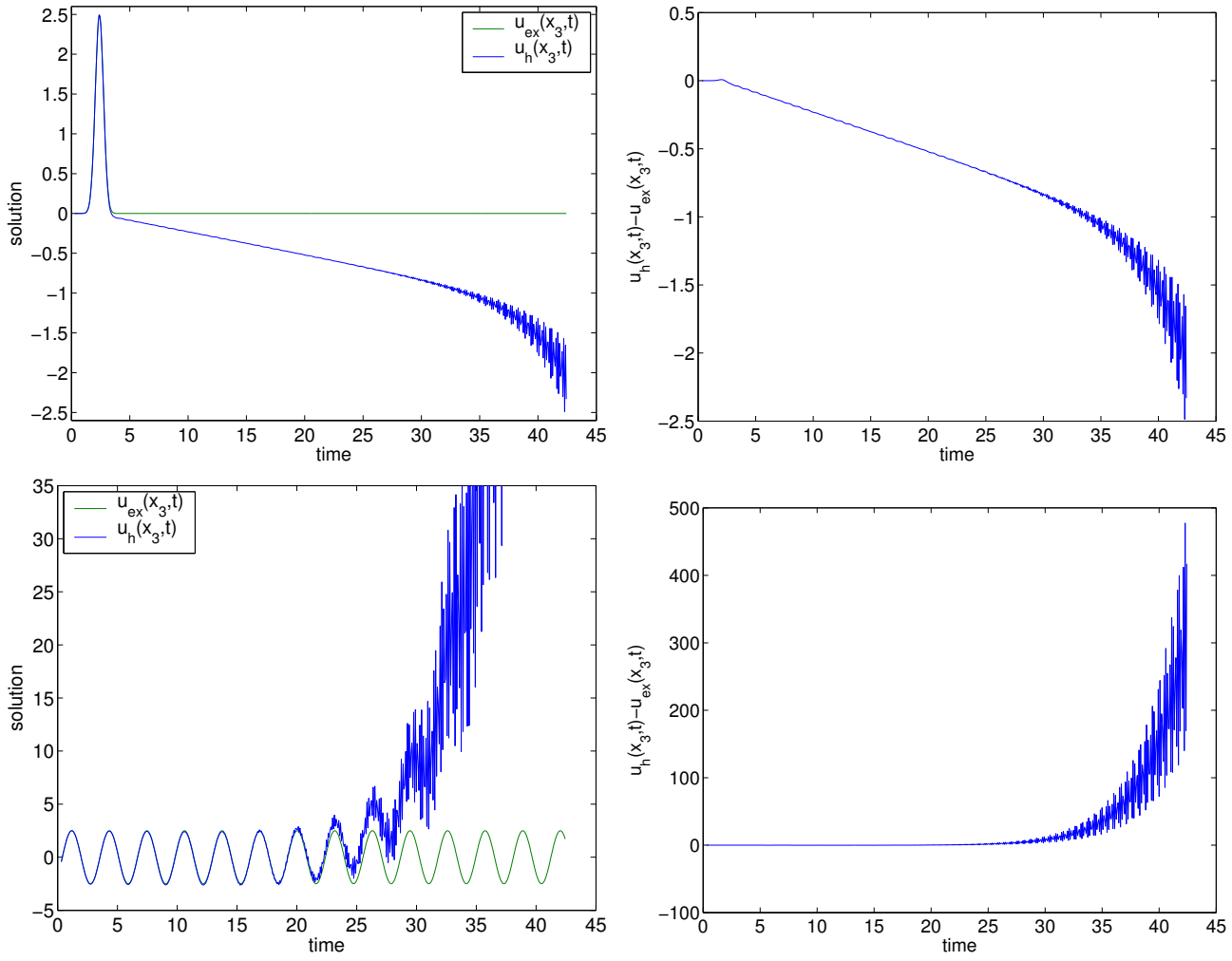


Figure 4.13: Choosing the Neumann operator $\mathcal{L}_\nu$ in (2.22) leads to long-time instability for an interior nondissipative stencil (blue). The green curve is the exact solution. These examples are computed on mesh m_0, and the solution is plotted at $\boldsymbol{x}_3 = [0, 0.4, 0]^t$ (left). On the right, the error of the two curves is displayed. The procedure as explained in 4.1.3 is used to compute the gradient on the Kirchhoff surface.

Long-time stability for Sommerfeld-type operator

The problem of long-time instability is resolved when we replace the operator $\mathcal{L}_\nu$ by (4.81) or by (4.82): the blue curves in Figures 4.14 and 4.16 correspond to computations using the operator $\mathcal{L}_S$ in (2.22) and demonstrate that the numerical solution remains stable. On the upper left, the solution is displayed at $\boldsymbol{x}_3$, while on the upper right its gradient is projected to $\boldsymbol{n} = [1, 1, 1]^t$. The lower pictures show that the errors $u_h(\boldsymbol{x}_3, t) - u_{ex}(\boldsymbol{x}_3, t)$ and $\boldsymbol{n} \cdot \nabla(u_h(\boldsymbol{x}_3, t) - u_{ex}(\boldsymbol{x}_3, t))$ remain bounded and are small. While the gradient seems to behave properly (Figure 4.14, lower right), the numerical solution does not approach zero with high-order accuracy once the wave is supposed to have completely left Ω (Figure 4.14, lower left). We also observe that the error of the solution in Figure 4.16 (lower left) is not symmetric about the time axis as it should be. Rather, it is shifted by a constant. Again, its gradient doesn't exhibit this difficulty. This discrepancy is due to the prescription of purely Neumann data on Γ. Any constant solves the associated homogeneous problem, and therefore uniqueness is not guaranteed. The operator $\mathcal{L}_\alpha$ defined in (4.81) insures that zero is the only constant which solves its homogeneous problem, and therefore uniqueness of the numerical solution is guaranteed even with purely Neumann data on Γ. This is shown for the two different examples in Figures 4.15 and 4.17, respectively. Figure 4.18 demonstrates that the numerical solution remains stable, even after many more time steps.

Convergence

Next, we investigate the accuracy in time of the overall numerical scheme for the two stable boundary operators defined in (4.81) and (4.82). The results in Figure 4.19 are obtained using the Sommerfeld operator $\mathcal{L}_S$ to evaluate $e_2(t)$ and $e_G(t)$ on m_3. We study the importance of reconstructing the gradient of the solution on the Kirchhoff's surface $\mathcal{S}$ to highest possible order: on the top of Figure 4.19, standard linear interpolation of the form $\nabla u_h(\boldsymbol{x}_\mathcal{S}, t_n) = \sum_k \nabla\phi_k(\boldsymbol{x}_\mathcal{S}) u_k(t_n)$ is used, which leads to loss in accuracy. While the gradient of the solution behaves properly, the field itself is not as accurate as it should be. Using the Chebyshev interpolation technique as explained in Section 4.1.3 leads to a significant improvement (middle of Figure 4.19). These specific examples are computed with $N_{cheb,r} = 7$ and $N_{cheb,z} = 14$ in the r and z directions, respectively. Finally, for comparison purposes, at the bottom of Figure 4.19, the computation is performed by imposing the exact values on $\mathcal{B}$ instead of using the computational boundary condition approach. As can be seen in Figure 4.20, using the operator $\mathcal{L}_\alpha$ defined in (4.81) brings significant improvement. The upper picture reveals that the behavior and accuracy is still not

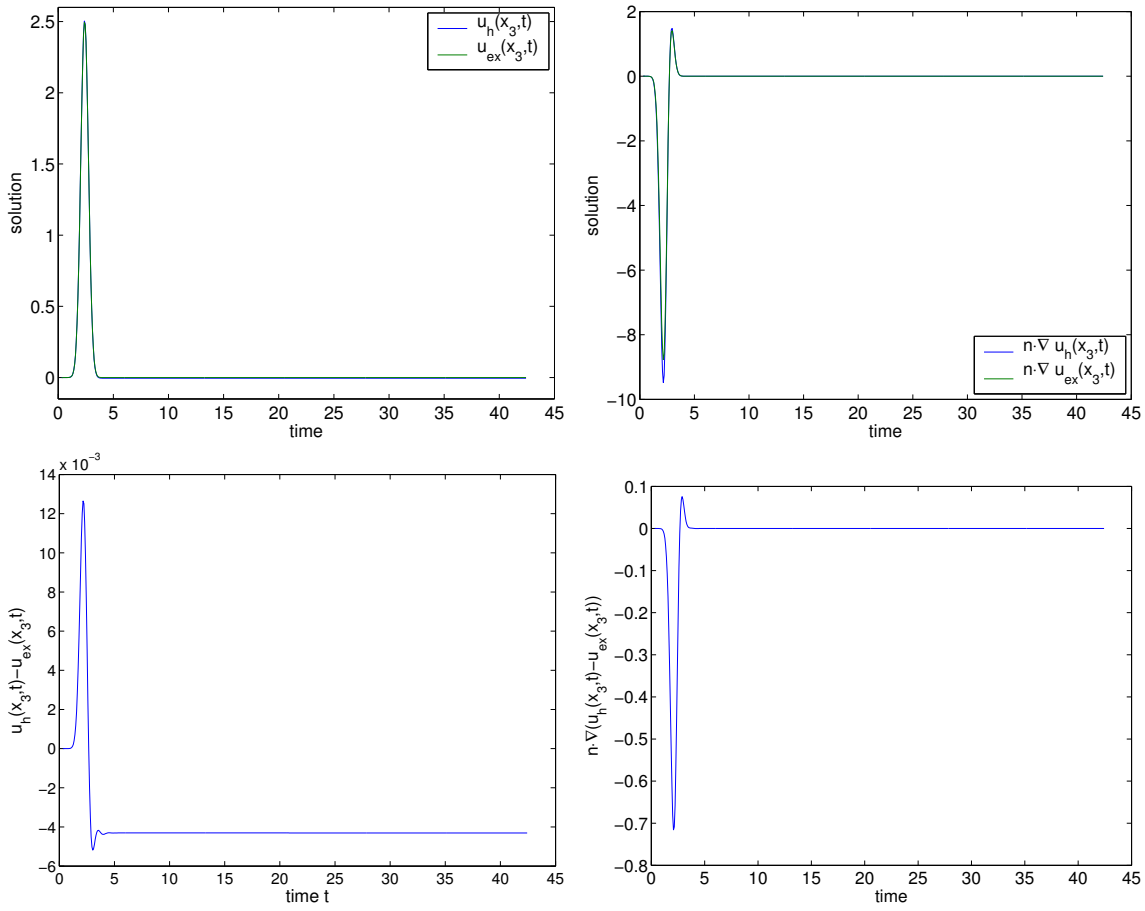


Figure 4.14: Top: The numerical solution computed on mesh m_0 with the Sommerfeld operator $\mathcal{L}_S$. The plots show the solution at $\boldsymbol{x}_3 = [0, 0.4, 0]^t$, and its gradient projected to $\boldsymbol{n} = [1, 1, 1]^t$. The technique developed in 4.1.3 is used to compute the gradient on $\mathcal{S}$. Bottom: The timely difference of the numerical and exact wave field at $\boldsymbol{x}_3$.

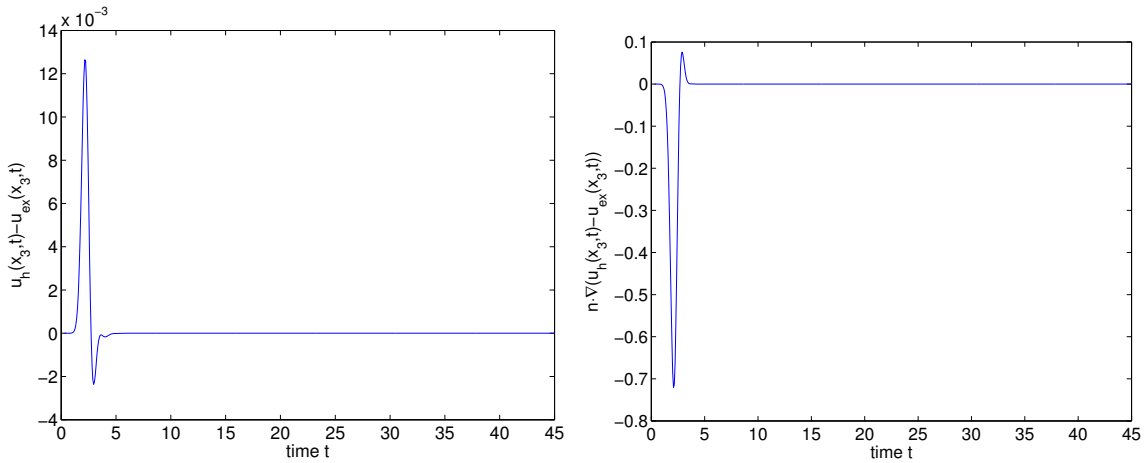


Figure 4.15: Same computation as in Figure 4.14, but $\mathcal{L}_\alpha$ is used in place of the Sommerfeld operator.

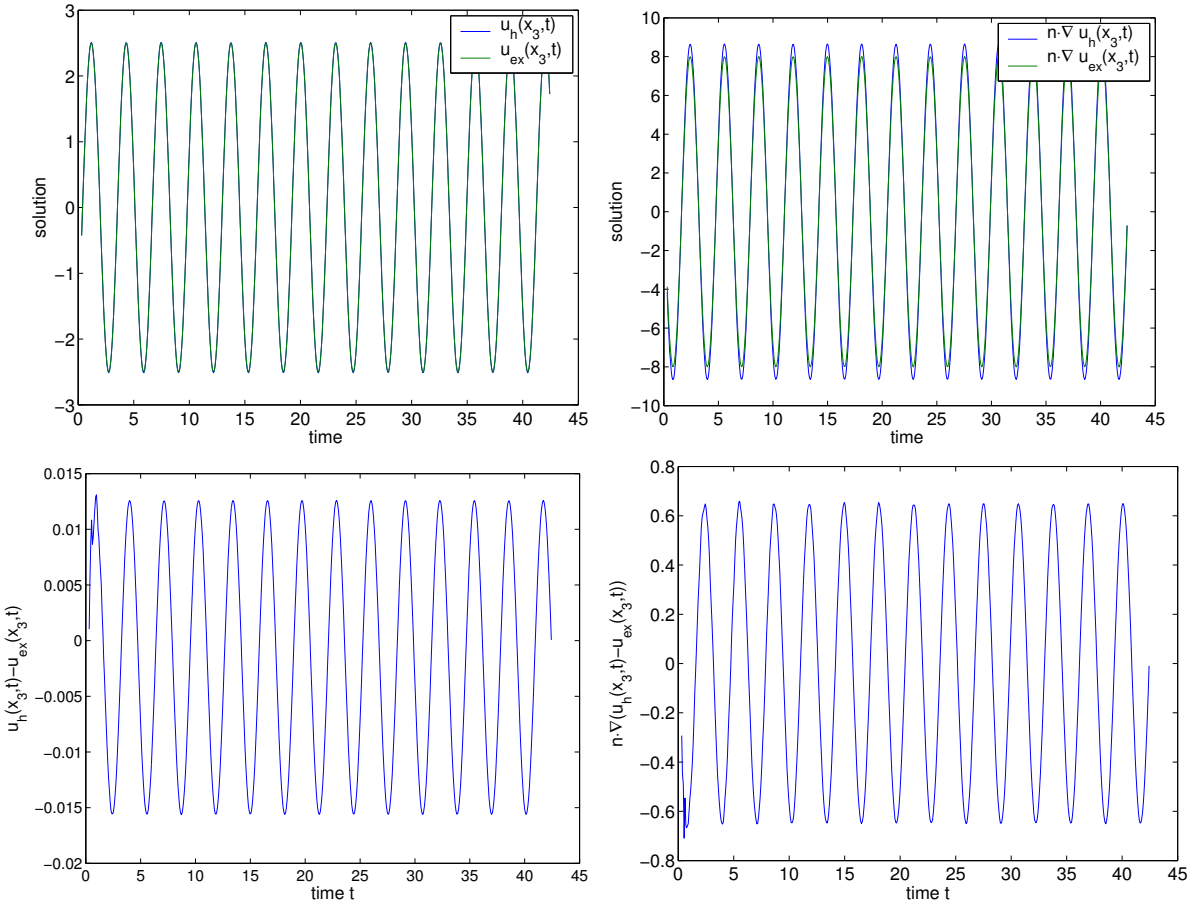


Figure 4.16: Top: The numerical solution computed on mesh m_0 with the Sommerfeld operator $\mathcal{L}_S$. The plots show the solution at $\boldsymbol{x}_3 = [0, 0.4, 0]^t$, and its gradient projected to $\boldsymbol{n} = [1, 1, 1]^t$. The Chebyshev series is used to compute the gradient on $\mathcal{S}$. Bottom: The timely difference of the numerical and exact wave field at $\boldsymbol{x}_3$. Note that the amplitude of the error on the left oscillates between $[-0.0155, 0.0125]$.

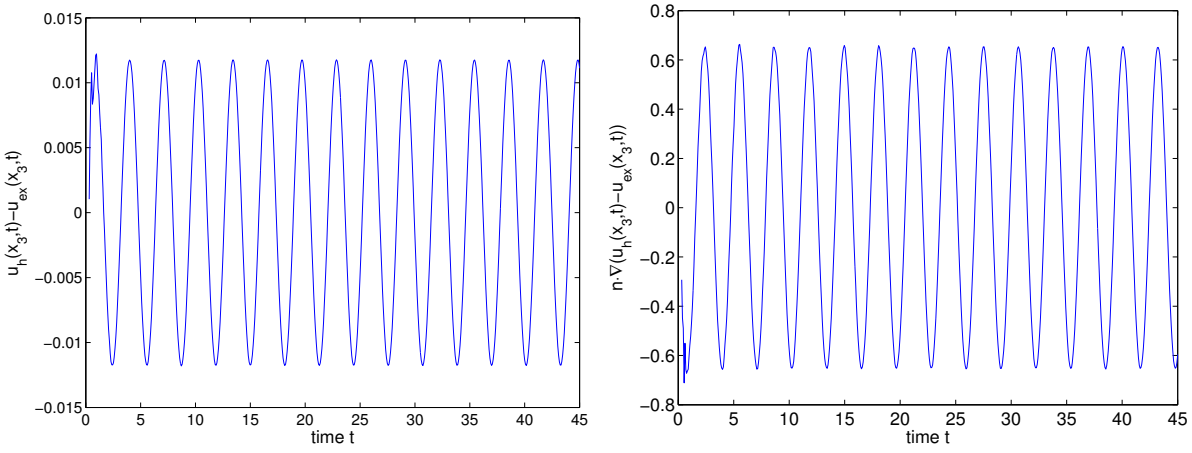


Figure 4.17: Same computation as in Figure 4.14, but $\mathcal{L}_\alpha$ is used in place of the Sommerfeld operator. This time, the error is in both cases symmetric with respect to the time axis.

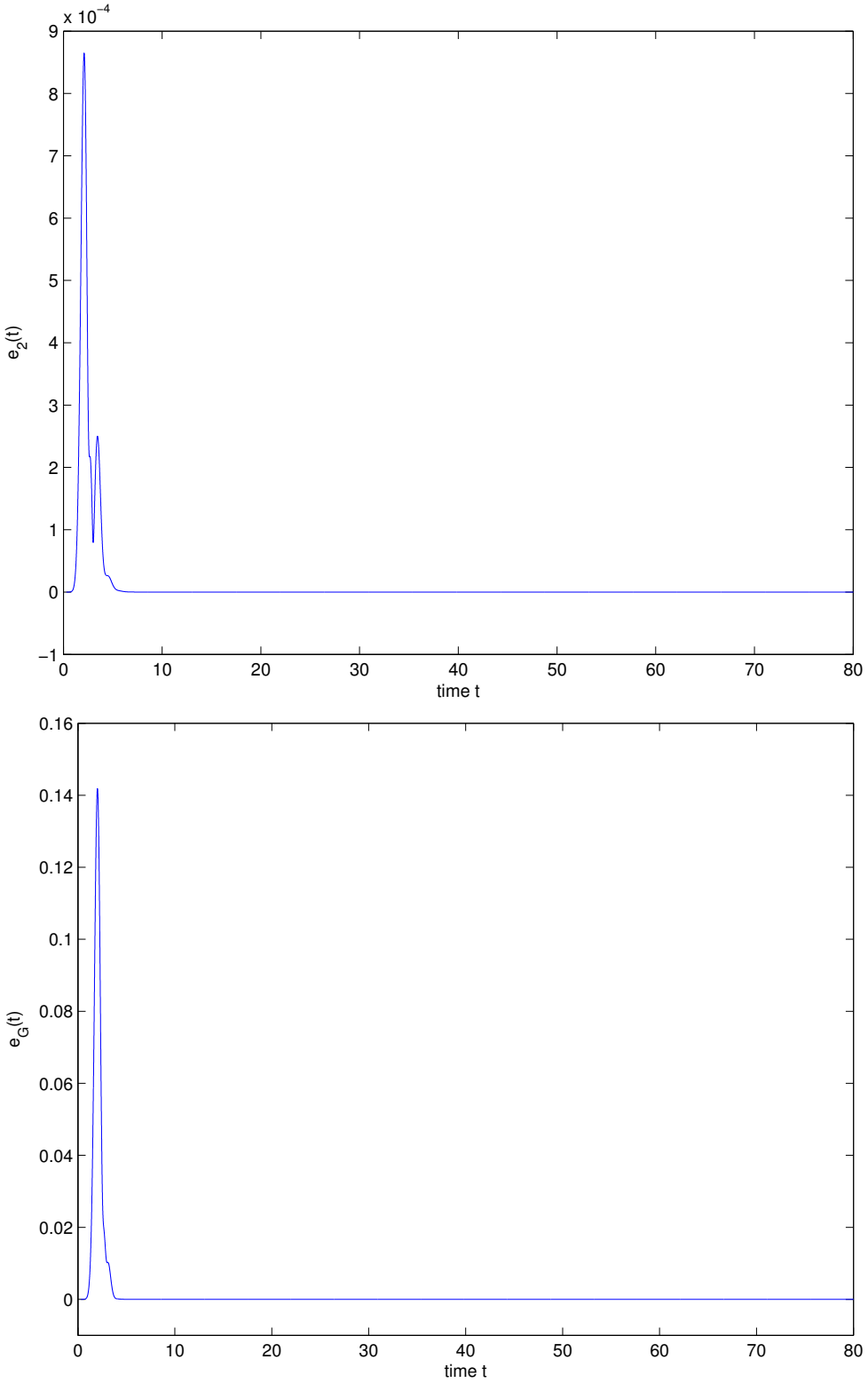


Figure 4.18: Demonstration of long time stability: the errors $e_2(t)$ and $e_G(t)$ are computed up to time $t = 80$ on m_0. This corresponds roughly to 45,860 time steps.

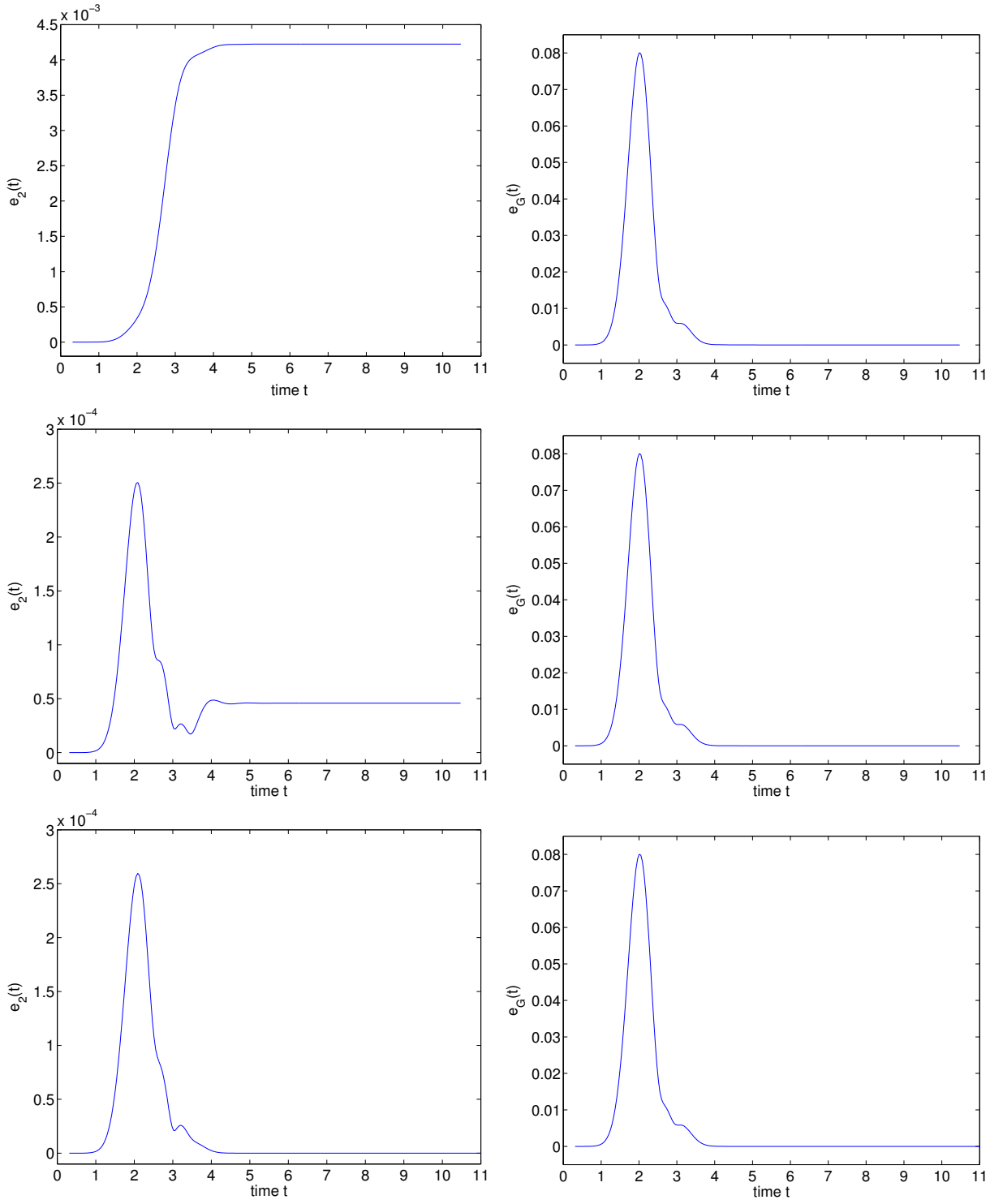


Figure 4.19: The errors $e_2(t)$ (left) and $e_G(t)$ (right) are displayed when using $\mathcal{L}_S$ as the boundary operator. The evaluations for these specific results are performed on m_3. The plots differ in how the gradient on $\mathcal{S}$ is computed. Top: linear interpolation is used to get the gradient. Center: The interpolation method developed in 4.1.3 computes the gradient. Bottom: $\mathcal{L}_S$ is applied on the exact values to obtain $g_{\mathcal{B},h}$. While $e_G(t)$ is identical in all three cases (right), significant differences are observed in $e_2(t)$ (left).

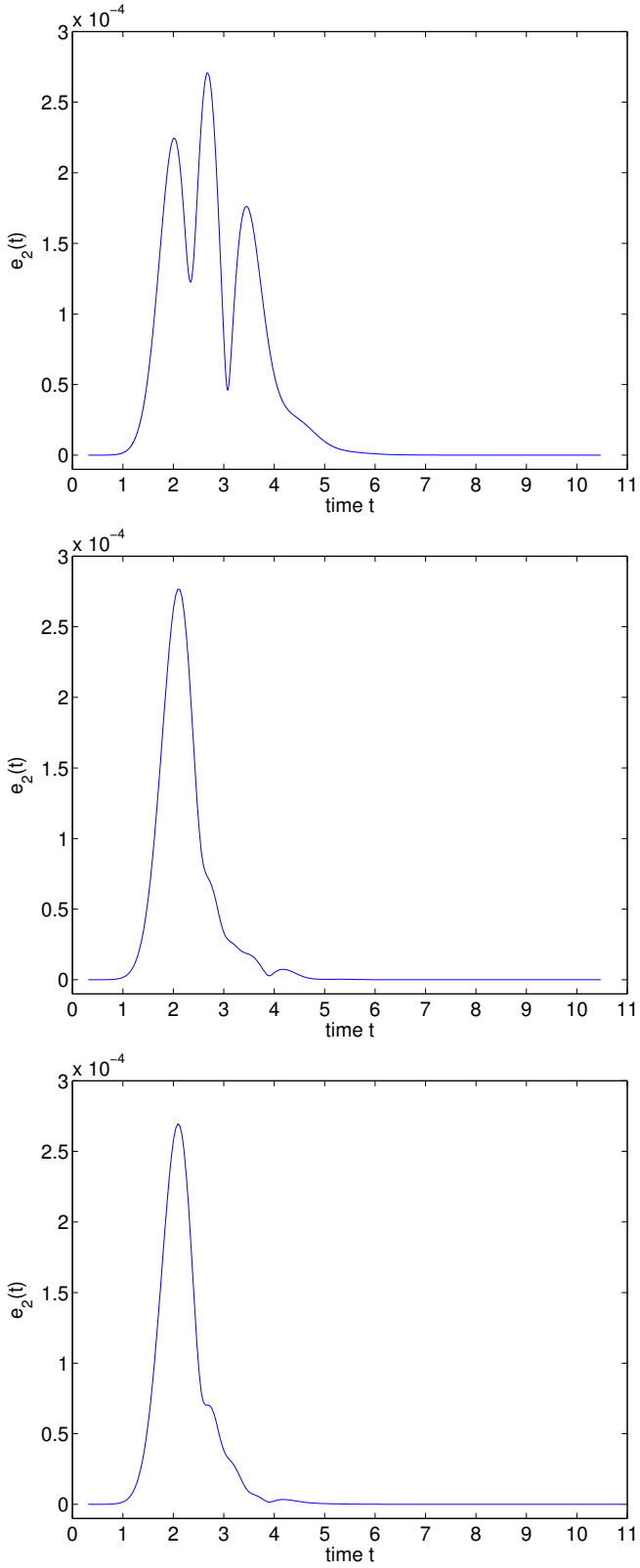


Figure 4.20: The error $e_2(t)$ is displayed when using $\mathcal{L}_\alpha$ as the boundary operator. The evaluations for these specific results are performed on m_3. The plots differ in how the gradient on $\mathcal{S}$ is computed. Top: linear interpolation is used to get the gradient. Center: The interpolation method from Section 4.1.3 computes the gradient. Bottom: $\mathcal{L}_\alpha$ is applied on the exact values to obtain $g_{\mathcal{B},h}$.

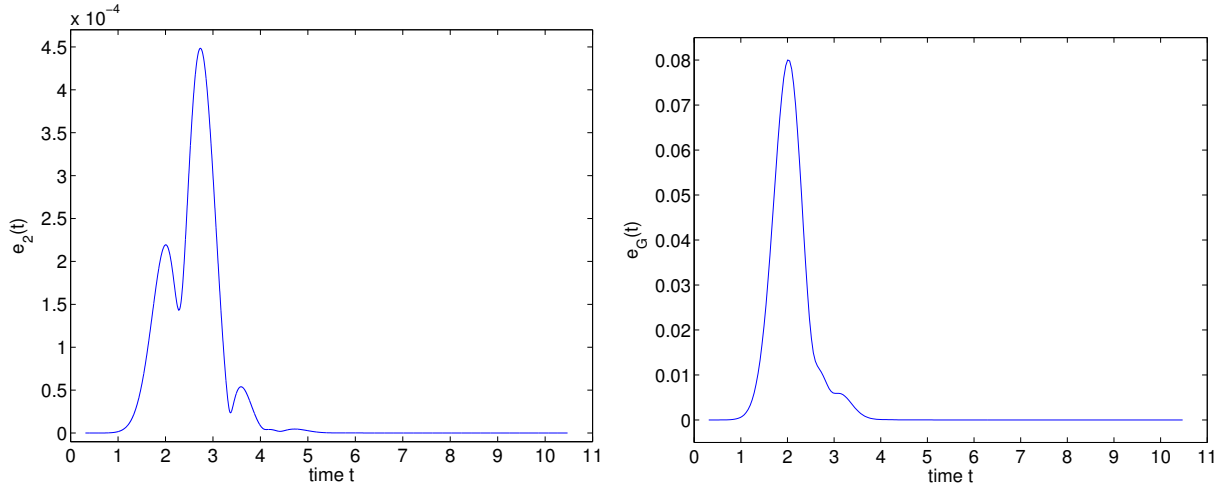


Figure 4.21: The operator $\mathcal{L}_M$ is used to obtain $e_2(t)$ and $e_G(t)$ on the mesh m_3. The inaccuracy in $e_2(t)$ is due to linear interpolation to evaluate the gradient on $\mathcal{S}$. If the technique from Section 4.1.3 is used to evaluate the gradient, the plot is identical to the picture in the center of Figure 4.20.

the same as when the exact data on $\mathcal{B}$ in (2.22) is used—compare with the lower plot. The middle graph demonstrates that the problem is removed when evaluating the gradient on $\mathcal{S}$ with the special Chebyshev expansion technique. Finally, Figure (4.21) demonstrates that also other choices for $\alpha(\boldsymbol{x}) > 0$ can lead to a unique, numerical stable solution. Here, we define

$$\mathcal{L}_M \equiv \frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial \nu} + \frac{1}{|\boldsymbol{x}|}. \tag{4.117}$$

Clearly, if the outer boundary is a sphere, $\boldsymbol{\nu} \cdot \boldsymbol{x}$ equals $|\boldsymbol{x}|$ and the operator in (4.117) is identical to (4.81). This is not the case if $\mathcal{B}$ is the surface of a cuboid, but since $1/|\boldsymbol{x}| > 0$, uniqueness is guaranteed by Lemma 4.1.1, and the numerical results in Figure 4.21 also confirm stability. As in Figure 4.20, the artificial middle arch of the plot in Figure 4.21 on the left, which causes the numerical solution to be less accurate, vanishes when the standard linear interpolation is replaced by the method developed in 4.1.3. The plot in the latter case is identical to the picture in the center of Figure 4.20.

In Figure 4.22, we show the results for $e_{2,T}$ and $e_{G,T}$ on the meshes $\{m_k\}_{k=0}^{K}$ where $K = 8$ if the polynomial degree of the shape functions is $p = 1$ and $K = 4$ if $p = 2$. The maximum value of the error is reached on all grids around the time $t \approx 2$. On a logarithmic scale, we plot the mesh size on the abscissa versus the errors on the ordinate. For linear shape functions $p = 1$, we expect first-order convergence for $e_{G,T}$ and second-order convergence for $e_{2,T}$, which is confirmed in Figure 4.22. We

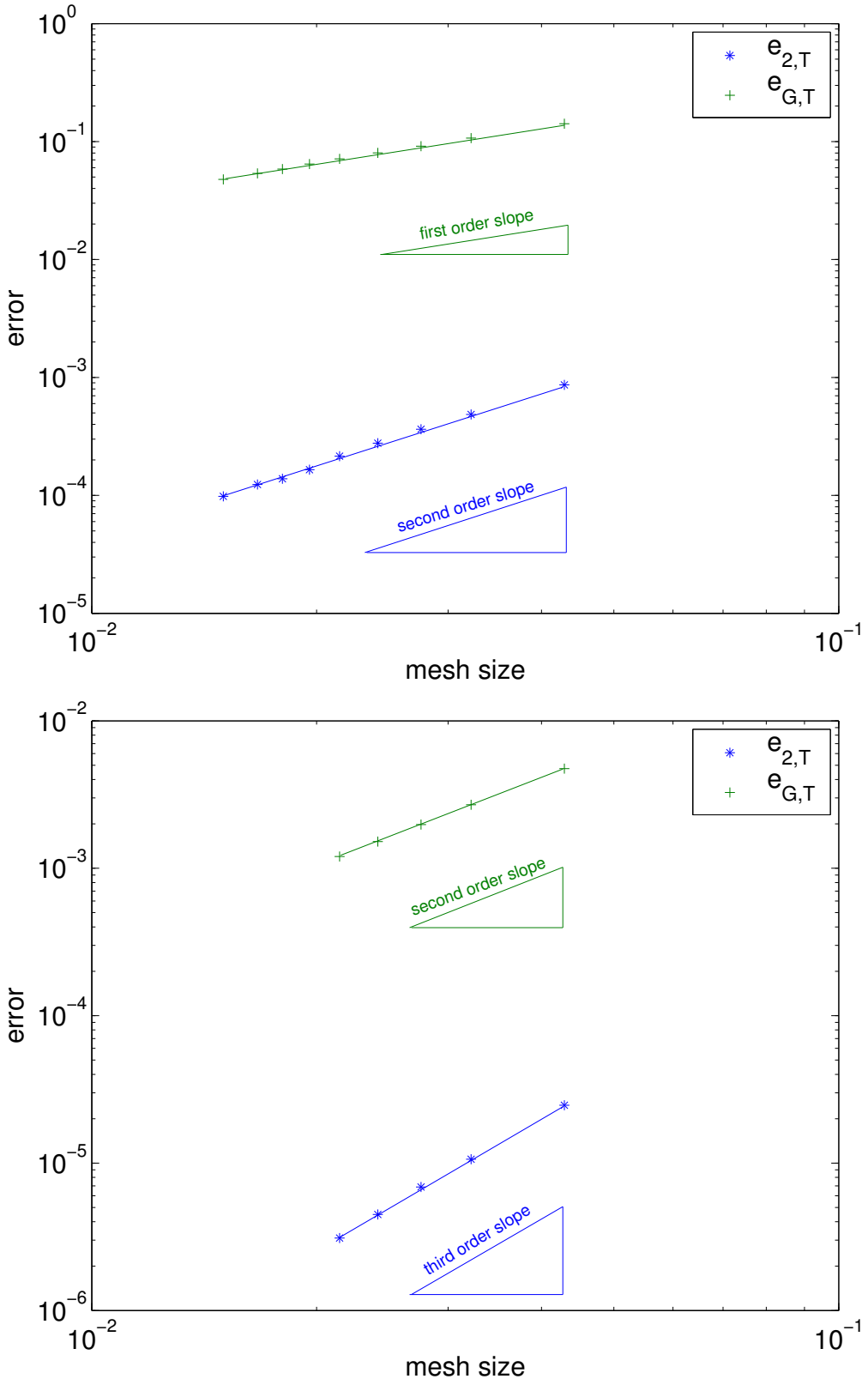


Figure 4.22: Convergence analysis for $p = 1$ (top) and $p = 2$ (bottom). The errors are plotted in the energy norm and L_2-norm, respectively. In all computations the maximum errors in time are around $t \approx 2$

perform similar computations for quadratic shape functions $p = 2$. The results in Figure 4.22 show that $e_{G,T}$ is now second-order and $e_{2,T}$ third-order convergent, as expected. Figure 4.22 thus confirms the high accuracy of the nonreflecting boundary condition: as the discretizations of the FEM are refined, the solution converges according to the order of the interior scheme to the exact solution.

	h	$gdof(p=1)$	$gdof(p=2)$
m_0	0.0429329	450	1,295
m_1	0.0321997	741	2,153
m_2	0.0275997	1,035	3,021
m_3	0.0241498	1,309	3,833
m_4	0.0214665	1,615	4,741
m_5	0.0195652	1,953	5,745
m_6	0.018	2,323	6,845
m_7	0.0166667	2,725	8,041
m_8	0.015	3,388	10,015

Table 4.3: The table displays the mesh sizes for $\{m\}_{k=0}^{8}$ along with the global degree of freedom for $p = 1$ and $p = 2$.

Contour plots

Figures 4.23 and 4.24 show the contour plots of the total field and the scattered field, respectively. In these cases, the point source is located at $\boldsymbol{x}_0 = [0, 0, 0.6]^t$, and the numerical solutions are computed on the FEM mesh m_4. Finally, Figure 4.25 shows the scattered field at selected points in space for the time interval $[0, 10.5]$.

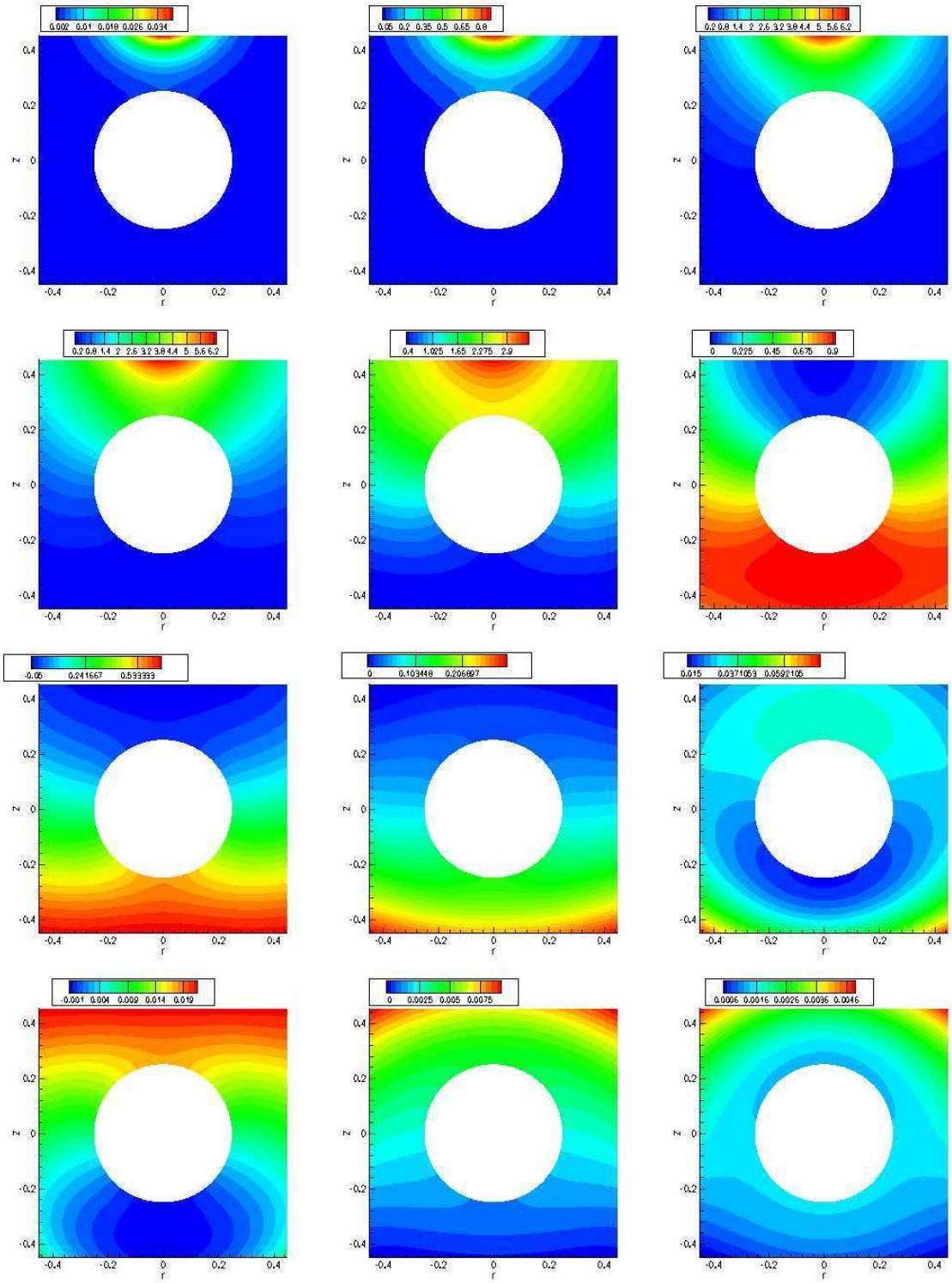


Figure 4.23: Contour plots of the total field for the times $t \approx 1.47, 1.76, \ldots, 4.88$. The point source is located at $\boldsymbol{x}_0 = [0, 0, 0.6]^t$.

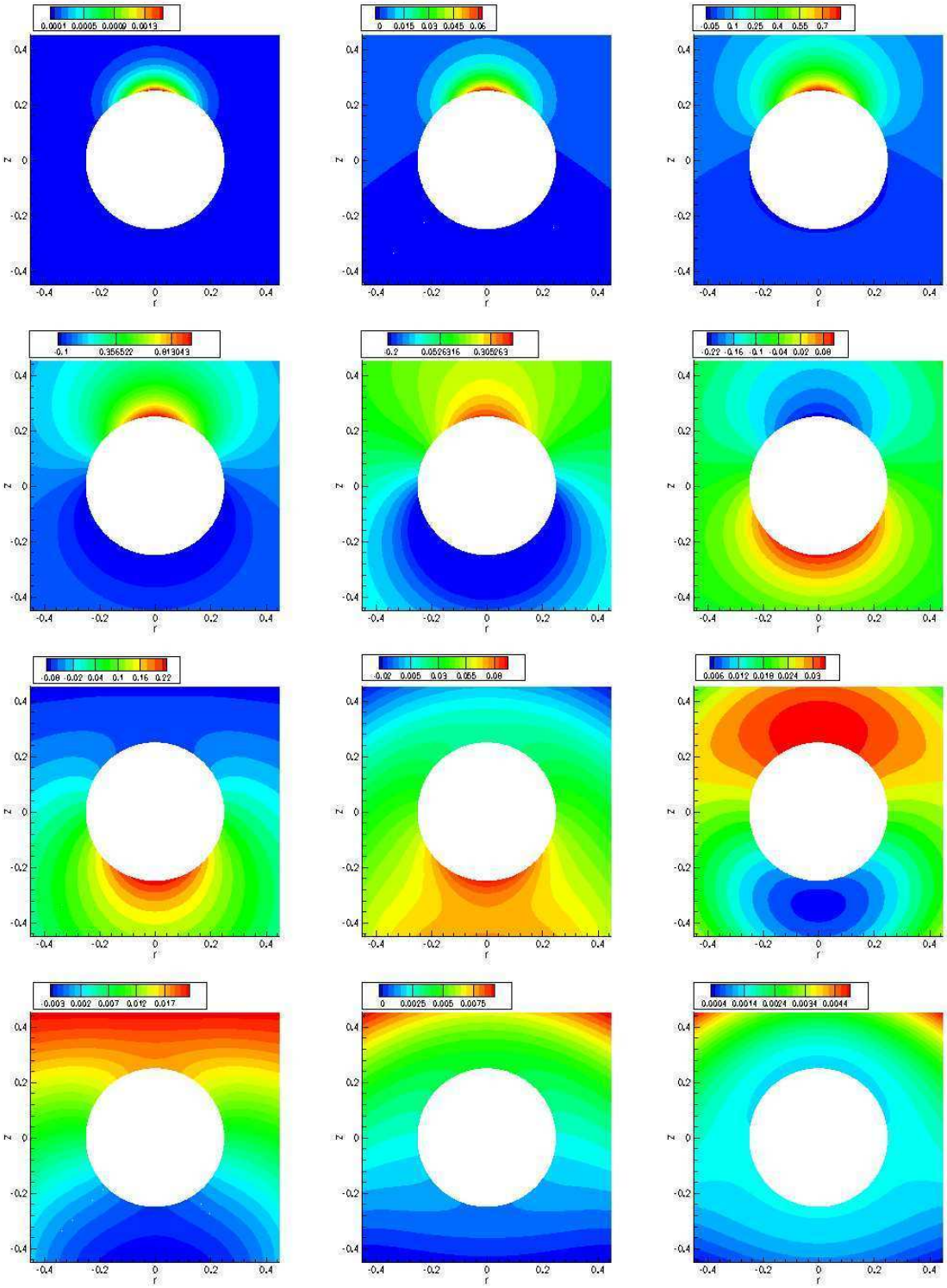


Figure 4.24: Contour plots of the scattered field for the selected times $t \approx 0.71,\ 1.23, 1.69, 1.98,$ $2.26, 2.82, 3.10, 3.39, 3.67, 3.95, 4.23, 4.38$. The point source is located at $\boldsymbol{x}_0 = [0, 0, 0.6]^t$ and the computation was performed on m_4.

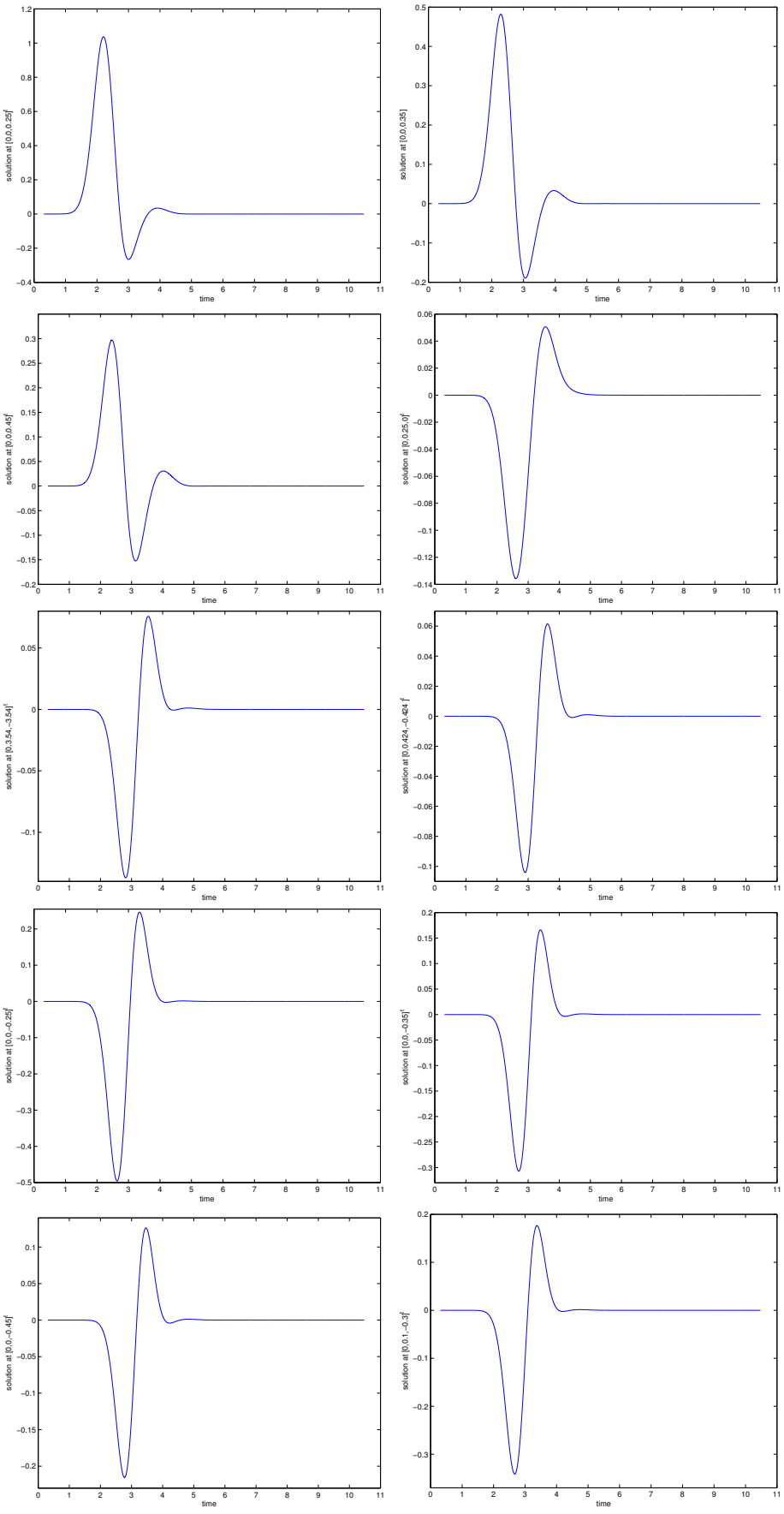


Figure 4.25: Scattered field at selected points in space. The point source acts from $\boldsymbol{x}_0 = [0, 0, 0.6]^t$.

Higher wave numbers

As an illustration of a more challenging problem, we consider $s(t) = \sin(kt)$ in (3.5) and the point source acts from $\boldsymbol{x}_0 = [0, 0, 0.23]^t$ with $k = 14$ or $k = 40$. In the first case, the computation is performed on the meshes $\{m_k\}_{k=0}^{2}$, in the more challenging case, on $\{m_k\}_{k=8}^{10}$, see Table 4.4 for the degree of freedom of the two finest meshes. Plots of the solutions and their corresponding $e_G(t)$ errors are displayed in Figure

	$gdof(p=1)$
m_9	$5,355$
m_{10}	$8,865$

Table 4.4: The degree of freedom of meshes m_9 and m_{10}

4.26. The convergence results are summarized in Table 4.5. The $e_{G,T}$ error is expected to be first-order convergent. Thus, the last column should be one. As we can see, this is achieved for $k = 14$, while the convergence is worse than first order for $k = 40$, and more grid points per wave length are necessary to get the clean first-order convergence rate.

k	h	kh	$e_{G,T}$	$\frac{\log(e^k_{G,T}/e^{k+1}_{G,T})}{\log(h_k/h_{k+1})}$
	0.0429329	0.6	1.50596	
14	0.0321997	0.45	1.13011	0.9980
	0.0275997	0.39	0.942462	1.1779
	0.015	0.6	3.31743	
40	0.0118421	0.47	2.69183	0.884
	0.00918367	0.37	2.14951	0.885

Table 4.5: Comparison between $k = 40, d/\lambda \approx 3.18$ and $k = 14, d/\lambda \approx 1.11$

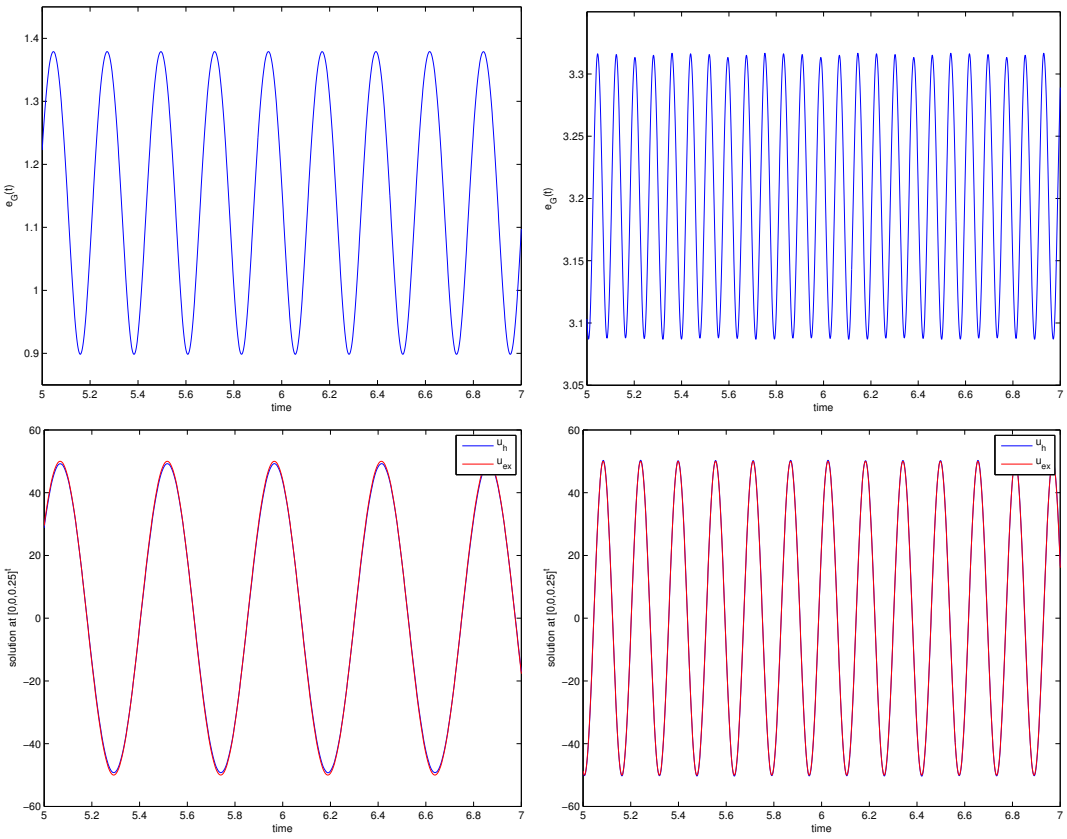


Figure 4.26: In both cases, the resolution is $kh \approx 0.6$ and the point source acts from $s = 0.23$. Left: $k = 14, h_0 \approx 0.043, d/\lambda \approx 1.11$. Right: $k = 40, h_8 = 0.015, d/\lambda \approx 3.18$

4.3.2 Elongated obstacle

In this section, we demonstrate that the computational nonreflecting boundary condition developed in this thesis has a significant advantage over other approaches. For the present example the scatterer is an ellipse with the major axis of length 0.3, the minor axis of 0.02. We surround it by the cuboid of dimensions $0.02 \times 0.02 \times 0.3$, which acts as the Kirchhoff's surface $\mathcal{S}$. As the outer boundary $\mathcal{B}$, we select the cuboid of length $0.1 \times 0.1 \times 0.46$ and position it in such a way that the minimum distance to $\mathcal{S}$ is 0.04 in the x- and z- direction, and 0.08 in the y- direction. Further, we choose $H = 0.02$, which gives 510 panels on $\mathcal{S}$. The maximum wave number is $k_{\max} \approx 16$, and from Figures 3.8 to 3.10 we deduce the parameters $S = 3, C = 3$ suffice to approximate the waves with the equivalent sources everywhere with an error of at least $\mathcal{O}(10^{-5})$. Further, we consider $N_t = 32$ frequency modes in total and we use zero-padding with $N_z = 64$ points when the wave is transformed from the frequency domain back into the physical domain on $\mathcal{B}$. The same Gaussian pulse as in the last section is used, located at $[0, 0, 0.06]^t$. The FEM grid in Figures 4.27 and 4.28 has $1,159$ degrees of freedom, with $h_{\min} \approx 0.00216$ and $h_{\max} \approx 0.01167$. The contour plots in Figure 4.27 and 4.28 represent the total and scattered field, respectively. The plots correspond in both cases to $t \approx 0.639, 3.228, 3.659,$ and 3.875. Using a sphere of radius 0.23 as the artificial boundary $\mathcal{B}$ instead requires roughly ten times more elements in two dimensions and hundred times more elements in three dimensions, and correspondingly higher computational times.

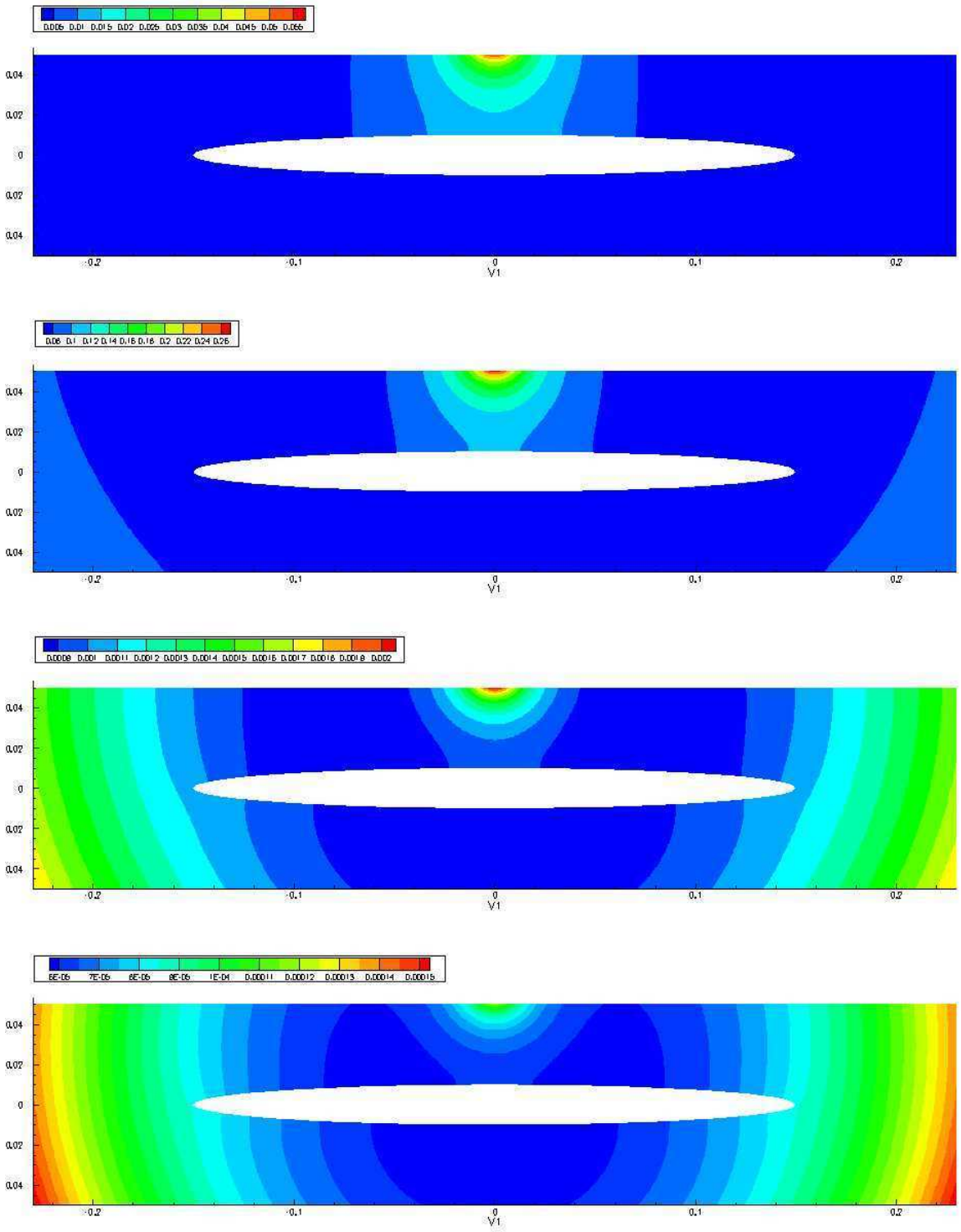


Figure 4.27: Contour plots of the scattered field for the selected times. The point source is located at $\boldsymbol{x}_0 = [0, 0, 0.06]^t$.

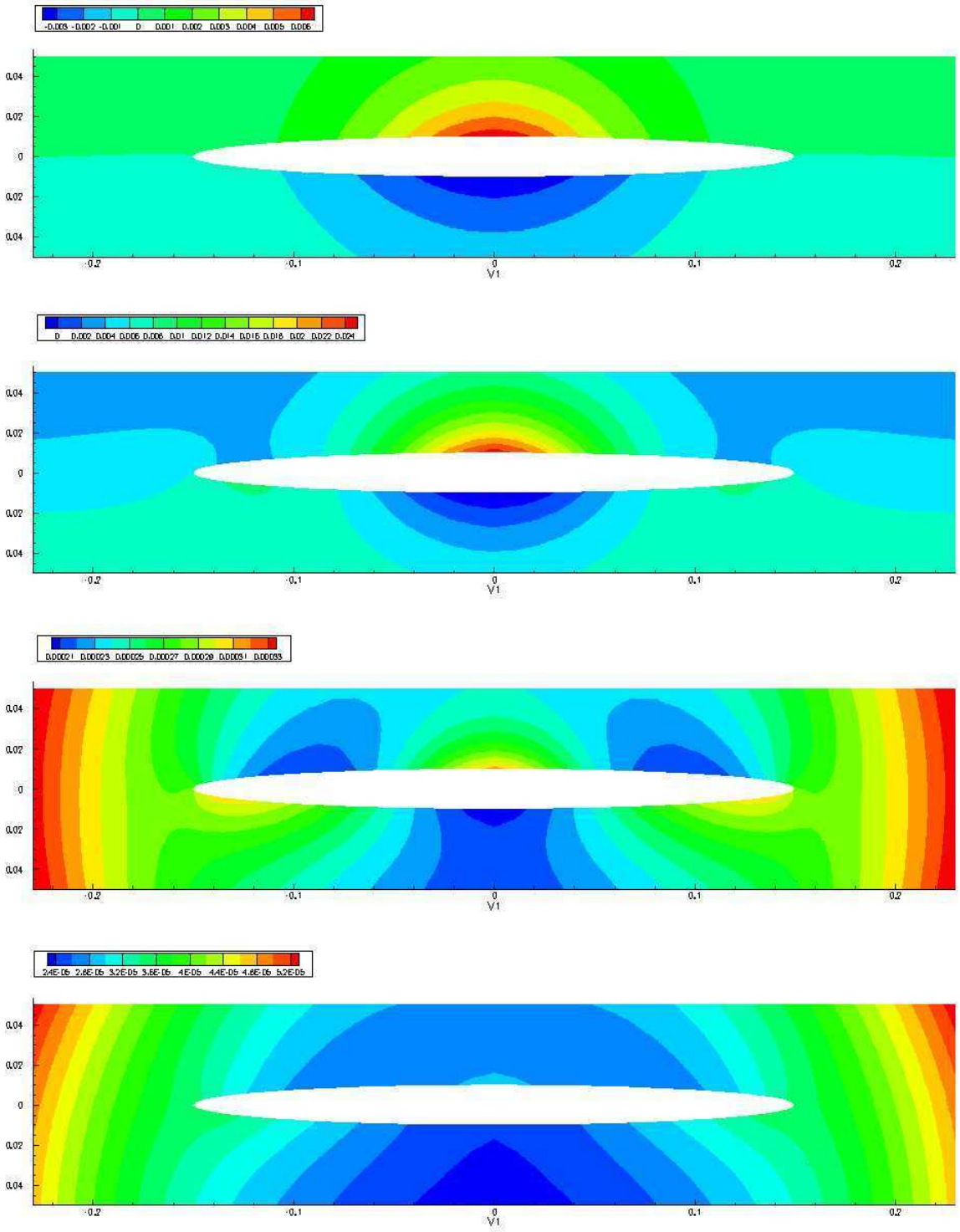


Figure 4.28: Contour plots of the scattered field for the selected times. The point source is located at $\boldsymbol{x}_0 = [0, 0, 0.06]^t$.

4.3.3 Fully three-dimensional example

In this subsection, we present numerical results for the fully three-dimensional version of the sphere in a cube geometry. The spherical scatterer of radius 0.25 is embedded in a cube of length 0.9. Figure 4.29 depicts the scatterer with three faces of the outer boundary in the background. For these computations,

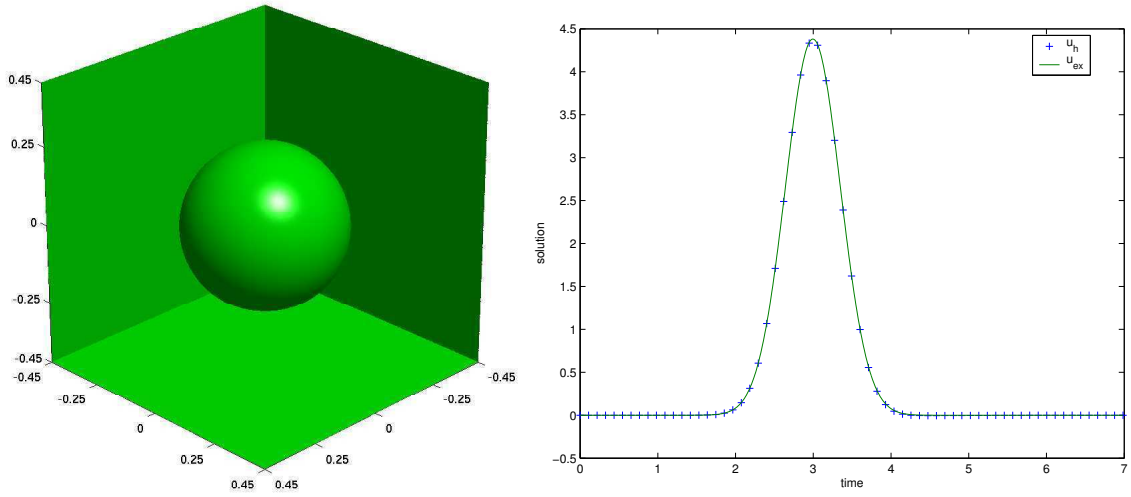


Figure 4.29: Left: Sphere in a Cube. Right: Exact and numerical solution computed on the coarsest grid at $\boldsymbol{x} = [0.25, 0.25, -0.25]^t$

the implicit scheme (4.101) is used along with a CG-solver to invert the matrix $\boldsymbol{A}_1$. Typically, the CG-solver requires up to 50 iterations for every time step. The usage of (4.101) removes the CFL constraint and allows us to choose $\Delta t \approx h_{\min}$. The cubic Kirchhoff surface of length 0.5 consists of 150 panels with $H = 0.1$. The scattered field matches with the outgoing wave generated at the point source $\boldsymbol{x}_0 = [0.15, 0.1, -0.11]^t$. In these computations, the parameters $\sigma = 0.5$ and $t_0 = 2.0$ are used. The numerical solution is computed in the time interval $[0, 7]$. The scattered wave has fully left the computational domain at time $t = 7$. Since the global time step Δt changes with the (maximum) mesh size h, the number of time steps n_t for the FEM computation increases as the mesh is refined. The computational domain is discretized by tetrahedra.

The main purpose of this section is to present comparisons of the computing times of the interior scheme to those arising from the boundary data on $\mathcal{B}$. The first three computations in Table 4.6 are performed on a single Intel(R) Pentium(R) D 3.4 GHz processor with 1 GB RAM. For every cell problem, the parameters $S = 3, C = 3, N_t = 8, N_z = 128$ are used. The "global degrees of freedom" (gdof) equal

the number of nodes of the FEM grid because we consider here a purely Neumann problem with linear shape functions. The exact and numerical solution on the coarsest grid at the point $[0.25, 0.25, -0.25]^t$ is depicted in Figure 4.29 on the left. As is demonstrated in Table 4.6, the portion of the computing time for the nonreflecting boundary condition based on equivalent sources (EQS) is much smaller than the time required for the FEM evaluation. We note that the elapsed time for the FEM algorithm is even larger for an explicit method due to the CFL restriction. The sixth column in the table shows that the evaluation of the open boundary data is in all cases smaller than 10% of the overall computation. The last two computations in Figure 4.6 are performed on a Pentium III (Cascades) CPU with 3.1 GB

gdof	h	n_t	FEM [sec]	EQS [sec]	EQS/(FEM+EQS)
58,256	0.043	640	590	60	9.2 %
151,440	0.031	880	2,029	90	4.3 %
264,672	0.026	1,082	4,220	130	3 %
365,120	0.023	1,202	70,150	1,270	1.8 %
423,696	0.022	1,280	86,000	1,360	1.6 %

Table 4.6: The table displays the computing times of the FEM and EQS algorithm for various mesh sizes h.

RAM. For each cell problem, the parameters $S = 5, C = 5, N_t = 8, N_z = 128$ are used to increase the accuracy on the boundary $\mathcal{B}$. Even so, the EQS algorithm needs significantly less time than the FEM.

Figure 4.31 displays the evolution of the scattered wave which matches a field that is generated by a point source located at $[0, 0, 0.2]$ and expands in the open three-dimensional space. The plots correspond to the selected times $t = 1.3, 1.4, 1.5, 1.6$. The three-dimensional graphs show the contour of the wave on the four surfaces $\{\boldsymbol{x} \in \mathbb{R}^3 : x = 0, 0 < y, z < 0.45\}, \{\boldsymbol{x} \in \mathbb{R}^3 : z = 0, 0 < x, y < 0.45\}, \{\boldsymbol{x} \in \mathbb{R}^3 : y = 0, 0 < x, z < 0.45\}$, and $\{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| = 0.25, x > 0, y > 0, z > 0\}$. Note that the wave perfectly exits the domain without any visible spurious reflections—a fact that is born in a more precise manner by the extremely small quantitative values of the error observed.

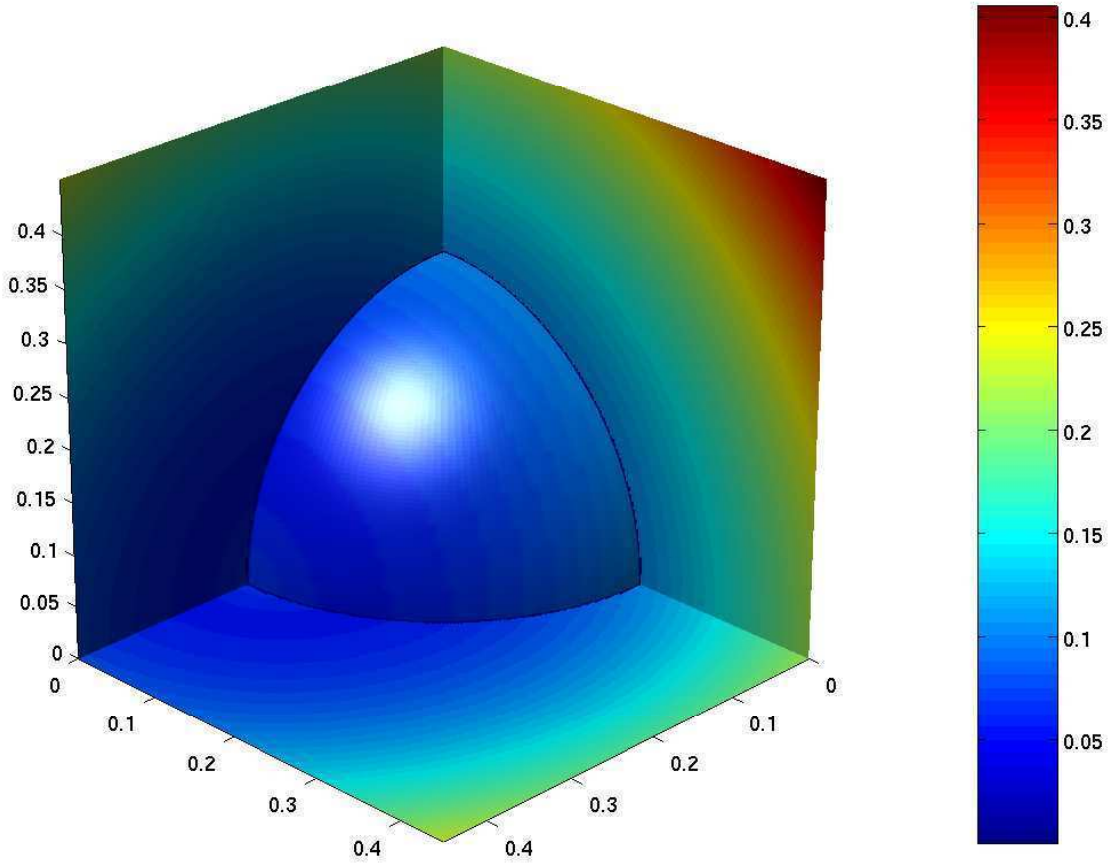


Figure 4.30: Contour plot of the scattered wave with color bar at time $t = 1.6$

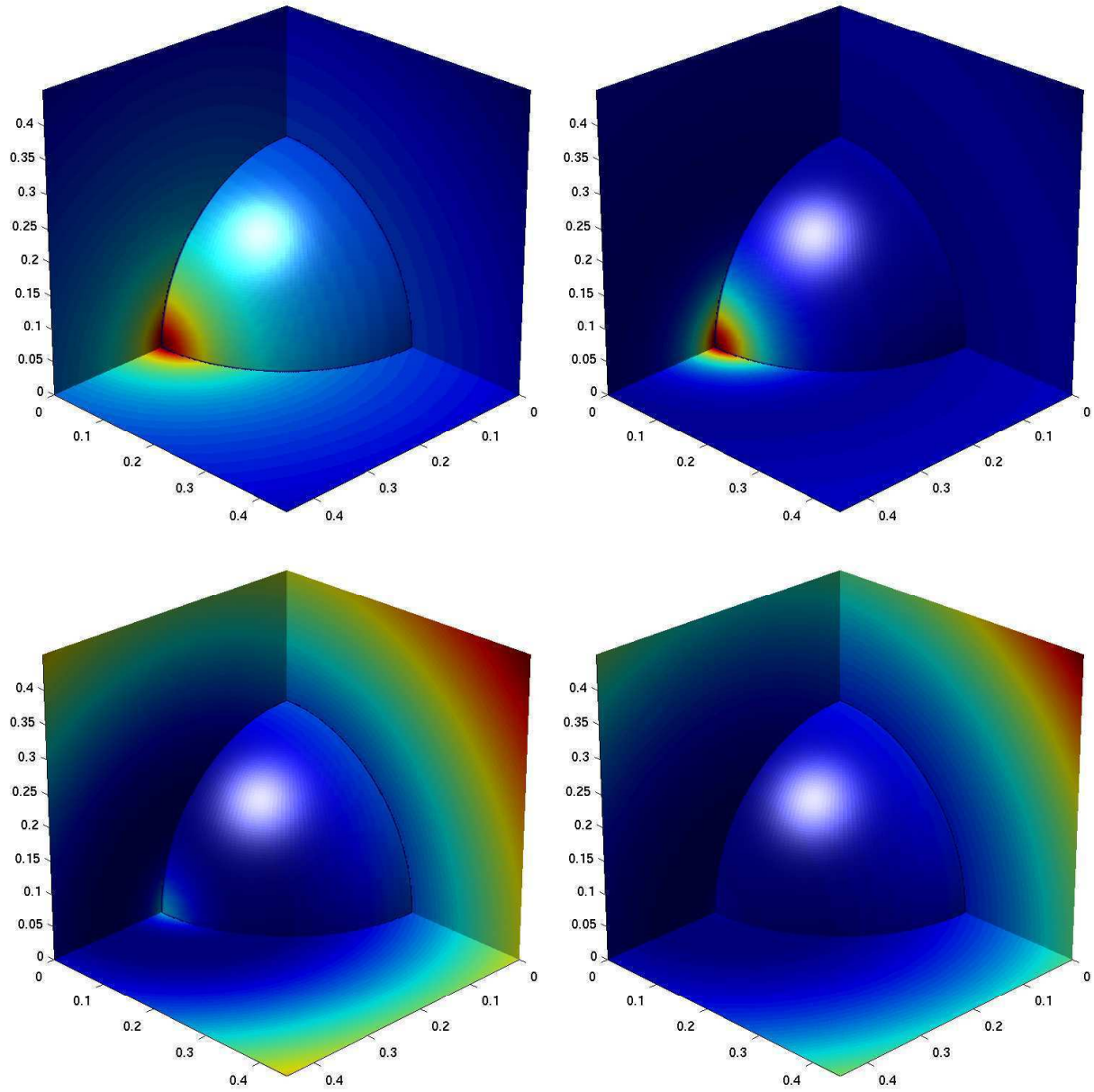


Figure 4.31: Contour plots of the scattered field for different times

4.4 Complexity and storage

The computational cost and storage required for the computational boundary evaluations depends on the specific problem under consideration. Here, our goal is to roughly estimate these costs for typical geometries. We assume that $\mathcal{B}$ is close to the surface $\mathcal{S}$, and that the Kirchhoff's surface consists of n_P panels. To each such panel we associate n_S source locations and n_C collocation points. N_t discrete frequencies are necessary to represent the wave accurately (by means of the continuation method, see Section 3.3.2). Further, we recall that the global equivalent sources are located at n_l layers that are parallel to one of the three faces of the outer boundary, and n_t is the number of time iterations needed for the FEM before it needs to be interrupted to run the computational boundary algorithm. For simplicity, we assume both $\mathcal{B}$ and $\mathcal{S}$ are cubes. The major operations to compute one cycle of boundary data are:

1. Given $u_{\mathcal{S}}(t_j)$ and $\partial_\nu u_{\mathcal{S}}(t_j)$, compute $\hat{u}_{\mathcal{S}}(\omega_j)$ and $\partial_\nu \hat{u}_{\mathcal{S}}(\omega_j)$, where $j \in \{0, \ldots, N_t - 1\}$. At $n_S n_P$ points, a FFT of the size $N_t/2$ has to be performed twice. Thus, this requires operations of the order $\mathcal{O}(2 \cdot n_S n_P \cdot N_t/2 \log_2(N_t/2))$.

2. Evaluate $\hat{u}_C(\omega_j)$ at all collocation points. For $N_t/2$ frequencies, two summations over a product of a Green's function (or its normal derivative) with a true source are involved. Each sum has n_S terms, and must be evaluated at n_C points for n_P panels. The operation count is therefore $\mathcal{O}(2 \cdot N_t/2 \cdot n_S n_C n_P)$.

3. For every cell problem, calculate the local equivalent sources on each of the three two-faces pairs. This is essentially three times a matrix-vector multiplication of the size $2n_C \cdot 4n_S$ for $N_t/2$ frequencies at n_P panels, i.e., the work is $\mathcal{O}(3 \cdot 8 n_S n_C \cdot N_t/2 \cdot n_P)$.

4. Apply the appropriate fast algorithm to obtain the boundary data $\hat{g}_{\mathcal{B}}(\omega_j)$ from the global equivalent sources. For every dimension and frequency, this involves $4n_l$ two-dimensional FFTs of the size $4N$, where N is the number of unknowns on one of the two-dimensional faces $\{\mathcal{B}_k\}_{k=1}^6$, one inverse FFT of the same size and $2n_l N$ multiplications, which is in summary an operation of the order $\mathcal{O}(3 \cdot N_t/2 \cdot 4N\{2n_l + (4n_l + 1)\log_2 4N\})$.

5. An inverse FFT in time for $N_t/2$ relevant frequencies at $6N$ points yields $g_{\mathcal{B}}(t_j)$. This requires $\mathcal{O}(6N \cdot N_t/2 \log_2 N_t/2)$ operations.

The work $\Sigma_{\mathcal{B}}$ and storage $\Xi_{\mathcal{B}}$ requirements for these five steps are summarized in Table 4.7, under the assumption that $n_C \approx 2n_S$ and $n_l \approx \sqrt{n_P/6}$. Thus, the dominant term in the operation count is

step	work $\Sigma_{\mathcal{B}}$	storage $\Xi_{\mathcal{B}}$
1	$\mathcal{O}(n_S n_P \cdot N_t \log_2(N_t/2))$	$\mathcal{O}(2N_t n_S n_P)$
2	$\mathcal{O}(2N_t n_P n_S^2)$	$\mathcal{O}(N_t n_C n_P)$
3	$\mathcal{O}(24N_t n_P n_S^2)$	$\mathcal{O}(n_P \cdot 3 \cdot 2n_S)$
4	$\mathcal{O}(6/\sqrt{6}N_t N\{2\sqrt{n_P} + (4\sqrt{n_P} + \sqrt{6})\log_2(4N)\})$	$\mathcal{O}(6NN_t)$
5	$\mathcal{O}(3N \cdot N_t \log_2(N_t/2))$	$\mathcal{O}(6NN_t)$

Table 4.7: Main work and storage contributions to compute the data on $\mathcal{B}$

$\mathcal{O}(N_t\sqrt{n_P}N\log_2(4N))$. We further note that due to the high accuracy of the equivalent sources, the panel length needs to be altered only very slowly as we refine the FEM mesh, making the work close to an order $\Sigma_{\mathcal{B}} = \mathcal{O}(N_t N \log_2(4N))$ operation. Using the Kirchhoff's formula directly as proposed in [86] costs $\Sigma_K = \mathcal{O}(n_t N^2)$, which is significantly higher than our algorithm for large N. Specifically, the ratio of the number of operations is

$$\frac{\Sigma_K}{\Sigma_{\mathcal{B}}} = \mathcal{O}\left(\frac{N}{\log_2(4N)}\right). \tag{4.118}$$

The total number of operations of a FEM in the computational domain Ω is of the order $\Sigma_\Omega = \mathcal{O}(n_t N^{3/2})$ when linear shape functions are used. In (4.118), we use the fact that typically n_t is of the order of N_t. Thus, the required work for the total scattering Algorithm 2.3.1 is dominated by Σ_Ω for sufficiently large N, while the proposed procedure in [86] is dominated by Σ_K.

The storage of our proposed boundary condition, in turn, is dominated by $\Xi_{\mathcal{B}} = \mathcal{O}(N_t N)$. For an explicit time-marching scheme, the CFL-condition (4.97) must hold, which implies that $n_t = \mathcal{O}(\sqrt{N})$ and therefore, $\Xi_{\mathcal{B}} = \mathcal{O}(N^{3/2})$, which is of the same order as the storage Ξ_K in [86] and the interior FEM scheme with linear shape functions.

The authors in [86] point out that their proposed scheme is typically more efficient than the absorbing boundary conditions developed in [8, 29, 32, 62, 60, 61, 75], which introduce an error proportional to $\mathcal{O}(1/R)^{k+1}$, where R is the size of the computational domain and k is an integer determining the order of approximation. These conditions are local in nature and thus have negligible computing times, but to maintain the accuracy of the FEM requires choosing a computational domain much larger than suffices through the use of Kirchhoff's formula, where one can select the boundary $\mathcal{B}$ as close to the scatterer as

desired. The authors in [86] show that in this sense, their proposed scheme is usually more efficient in terms of work and storage, even though the computation of the boundary condition is more expensive than the interior scheme. More recent approaches [3, 40, 84] which are exact nonreflecting boundary conditions in nature may suffer in certain circumstances from similar disadvantages if a special geometry on $\mathcal{B}$, such as a sphere, is required: in case of elongated scatterers, a vast computational domain Ω needs to be selected, making the overall algorithm inefficient, even though the boundary data are computed fast and no error is introduced outside of the interior discretization order. Our proposed scattering algorithm is thus more advantageous than many other approaches: 1) it improves significantly the computing time of the boundary data compared to reference [86]; 2) it does not introduce any spurious reflection at the artificial boundary as methods proposed in [8, 29, 32, 62, 60, 61, 75]; and 3) it is very flexible in designing the computational domain, contrary to approaches as introduced in [3, 40, 84].

4.5 Conclusion

In this thesis, we have introduced a new, exact nonreflecting boundary condition for the scalar wave equation. The approach is based on equivalent sources, which allows accelerating the computation of the artificial boundary condition considered in [86]: While in that reference, the dominant work arises from the computational boundary, we have demonstrated that in our case, the interior scheme is more expensive than the boundary condition algorithm. In fact, the computing times for our proposed nonreflecting boundary condition is only a small portion compared to an FEM in the three-dimensional computational domain. Our method is asymptotically exact in nature, so that, in practice, no spurious reflections develop at the artificial boundary. Methods such as [8, 9, 29, 30, 31, 32, 51, 52, 58, 60, 61, 62, 75, 77], in contrast, may suffer from this problem, which can corrupt the numerical solution significantly. Finally, our proposed approach allows us to truncate the computational domain from the unbounded space arbitrarily close to the scatterer. Many exact nonreflecting boundary conditions [3, 4, 40, 41, 43, 44, 63, 84] do not allow this.

Even though we have focused in this thesis on the wave equation, the developed techniques directly translate also to Maxwell's equations and to elastodynamic waves under similar conditions as in this work.

Finally, we believe that the scattering Algorithm 2.3.1 may be further accelerated and made more efficient by using a more suitable interior solver. The recently developed ADI algorithm [15] may prove

to be an excellent scheme coupled with our nonreflecting boundary condition, as, without requiring large matrix inversions, it gives rise to unconditional stability; it is thus significantly faster than FEM implementations, and discretizations of very high order are easily achieved.

Appendix A

Review: the wave equation

A.1 Helmholtz problem

We assume that $u_s(\boldsymbol{x},t)$ and $f(\boldsymbol{x},t)$ can be expressed by the the Fourier integral representation

$$u_s(\boldsymbol{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{u}_s(\boldsymbol{x},\omega)e^{-i\omega t}d\omega \tag{A.1}$$

$$f(\boldsymbol{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\boldsymbol{x},\omega)e^{-i\omega t}d\omega, \tag{A.2}$$

with the inverse transformations

$$\hat{u}_s(\boldsymbol{x},\omega) = \int_{-\infty}^{\infty}u_s(\boldsymbol{x},t)e^{i\omega t}dt \tag{A.3}$$

$$\hat{f}(\boldsymbol{x},\omega) = \int_{-\infty}^{\infty}f(\boldsymbol{x},t)e^{i\omega t}dt. \tag{A.4}$$

For (A.3) to hold, the solution must be causal, i.e., it cannot depend on the unbounded past. Then, substituting (A.1) and (A.2) into (2.1), (2.4), and (2.5) gives rise to the time harmonic problem

$$\Delta\hat{u}_s + k^2\hat{u}_s = -\hat{f}(\boldsymbol{x},\omega) \quad \text{in } \mathbb{R}^3\times\mathbb{R} \tag{A.5}$$

$$\partial_\nu\hat{u}_s = \hat{g}(\boldsymbol{x},\omega) \quad \text{on } \Gamma\times\mathbb{R} \tag{A.6}$$

$$\lim_{r\to\infty} r\left(\partial_r\hat{u}_s - ik\hat{u}_s\right) = 0, \quad r=|\boldsymbol{x}|. \tag{A.7}$$

The time harmonic problem is, of course, intimately associated with its time-dependent counterpart. The elliptic partial differential equation (A.5) is known as the inhomogeneous Helmholtz equation with wave number given by the *dispersion relation*

$$k = \frac{\omega}{c}, \tag{A.8}$$

see, e.g., [91]. As is known, the dispersion relation results from substituting a plane wave $e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}$ into the homogeneous part of the wave equation (2.1), where $k = |\boldsymbol{k}|$. The wave length λ, in turn, is given by

$$\lambda \equiv \frac{2\pi}{k}, \tag{A.9}$$

and, similarly, the period T is obtained by

$$T \equiv \frac{2\pi}{\omega}. \tag{A.10}$$

Substituting (A.9) and (A.10) into (A.8) yields

$$\lambda = cT. \tag{A.11}$$

A.2 Integral representation

A.2.1 Green's function

In the following discussion, it is assumed that the speed of sound c is positive. We note that for a fixed $\tilde{\boldsymbol{x}}$, the functions

$$G_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \;=\; \frac{1}{4\pi}\frac{e^{ik|\boldsymbol{x}-\tilde{\boldsymbol{x}}|}}{|\boldsymbol{x}-\tilde{\boldsymbol{x}}|} \tag{A.12}$$

and

$$\tilde{G}_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \;=\; \frac{1}{4\pi}\frac{e^{-ik|\boldsymbol{x}-\tilde{\boldsymbol{x}}|}}{|\boldsymbol{x}-\tilde{\boldsymbol{x}}|} \tag{A.13}$$

both satisfy the homogeneous Helmholtz equation if $\boldsymbol{x} \neq \tilde{\boldsymbol{x}}$. But Sommerfeld's radiation condition (A.7) accepts only (A.12) and eliminates (A.13). The Green's function (A.12) is therefore known to be the radial outgoing fundamental solution of the homogeneous Helmholtz equation in $\mathbb{R}^3\backslash\{\tilde{\boldsymbol{x}}\}$. It can be shown that the solution to the exterior Helmholtz problem behaves at infinity like $\exp(ikr)/r \cdot \alpha(\boldsymbol{r}/r)$, where α is a function of the vector $\boldsymbol{r}/r$, i.e., a function defined on the unit sphere and thus independent of the radial direction r, see [76].

A.2.2 Representation theorem

Theorem A.2.1. *Let Θ be a domain of class C^2, an assumption that we make for simplicity [1], Γ its boundary and ν the unit normal vector on Γ which points outside of Θ. Let $\hat{u}_s(\boldsymbol{x})$ be a sufficiently smooth function (i.e., $\hat{u}_s \in C^2(\mathbb{R}^3\backslash\bar{\Theta}) \cap C(\mathbb{R}^3\backslash\Theta)$) and its normal derivative is defined in the sense that the limit $\partial_\nu \hat{u}_s(\boldsymbol{x}) = \lim_{h\to+0} \nu(\boldsymbol{x}) \cdot \nabla \hat{u}_s(\boldsymbol{x} - h\nu(\boldsymbol{x}))$ exists uniformly on $\boldsymbol{x} \in \Gamma$. Further, assume that $\hat{u}_s$ satisfies the inhomogeneous Helmholtz equation (A.5), and that Ω is a domain that contains the compact support of the inhomogeneity $\hat{f}$ and has Γ as its inner boundary (see Figure A.1). Then, the integral representation*

$$\hat{u}_s(\boldsymbol{x}, \omega) = \hat{u}^v(\boldsymbol{x}, \omega) + \hat{u}_\Gamma(\boldsymbol{x}, \omega), \tag{A.14}$$

with

$$\hat{u}^v(\boldsymbol{x}, \omega) = \int_\Omega \hat{f}(\tilde{\boldsymbol{x}}, \omega) G_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}) d\tilde{\boldsymbol{x}}, \tag{A.15}$$

$$\hat{u}_\Gamma(\boldsymbol{x}, \omega) = \int_\Gamma \left\{ \hat{u}(\tilde{\boldsymbol{x}}) \frac{\partial G_k(\boldsymbol{x}, \tilde{\boldsymbol{x}})}{\partial \nu(\tilde{\boldsymbol{x}})} - \frac{\partial \hat{u}}{\partial \nu}(\tilde{\boldsymbol{x}}) G_k(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \right\} ds(\tilde{\boldsymbol{x}}), \tag{A.16}$$

holds for $\boldsymbol{x} \in \mathbb{R}^3\backslash\bar{\Theta}$.

Proof. The proof can be found in [24], for example. Since the theorem plays a major role in this thesis,

[1] This regularity assumption is a sufficient condition and can be weakened, see [59]. A domain $\Theta \in \mathbb{R}^3$ is said to be of class C^k, if for each point z of the boundary $\partial\Theta$ there exists a neighborhood V_z of z such that the intersection $V_z \cap \bar{\Theta}$ can be mapped bijectively onto the half ball $\{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| < 1, x_3 \geq 0\}$, this mapping and its inverse are k-times continuously differentiable, and the intersection $V_z \cap \partial\Theta$ is mapped onto the disk $\{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| < 1, x_3 = 0\}$.

a proof is provided here for completeness. We suppose that $u, v \in C^2(\bar{\Omega})$ and define

$$(u,v)_\Omega \equiv \int_\Omega u(\tilde{\boldsymbol{x}})\overline{v(\tilde{\boldsymbol{x}})}d\tilde{\boldsymbol{x}} \tag{A.17}$$

$$(u,v)_{\partial\Omega} \equiv \int_{\partial\Omega} u(\tilde{\boldsymbol{x}})\overline{v(\tilde{\boldsymbol{x}})}ds(\tilde{\boldsymbol{x}}). \tag{A.18}$$

By Green's first theorem, we have

$$(\Delta u, v)_\Omega = (\boldsymbol{n}\cdot\nabla u, v)_{\partial\Omega} - (\nabla u, \nabla v)_\Omega \tag{A.19}$$

$$(\Delta v, u)_\Omega = (\boldsymbol{n}\cdot\nabla v, u)_{\partial\Omega} - (\nabla v, \nabla u)_\Omega\,. \tag{A.20}$$

Subtracting (A.20) from (A.19) gives

$$(\Delta u, v)_\Omega - (\Delta v, u)_\Omega = (\boldsymbol{n}\cdot\nabla u, v)_{\partial\Omega} - (\boldsymbol{n}\cdot\nabla v, u)_{\partial\Omega}\,. \tag{A.21}$$

Let $\boldsymbol{x}$ be an arbitrary point in Ω, and let us define $\Omega_\rho = \{\tilde{\boldsymbol{x}} \in \Omega : |\boldsymbol{x}-\tilde{\boldsymbol{x}}| > \rho\}$ and $B_\rho = \Omega\backslash\Omega_\rho$ for any sufficiently small $\rho > 0$, so that the sphere $S(\boldsymbol{x};\rho) = \{\tilde{\boldsymbol{x}} \in \Omega : |\boldsymbol{x}-\tilde{\boldsymbol{x}}| = \rho\}$ is fully contained in Ω without touching its inner boundary Γ or outer boundary $\mathcal{B}$ (see Figure A.1). Further, for a sufficiently large $R > 0$, we define $\Omega_{R,\rho} = \{\tilde{\boldsymbol{x}} \in \mathbb{R}^3\backslash(\Theta \cup B_\rho) : |\tilde{\boldsymbol{x}}| < R\}$, which has Γ as its inner and the sphere $S_R = \{\tilde{\boldsymbol{x}} \in \mathbb{R}^3 : |\tilde{\boldsymbol{x}}| = R\}$ as its outer boundary.

Next, we substitute $v(\tilde{\boldsymbol{x}}) = G(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ and $u(\tilde{\boldsymbol{x}}) = \hat{u}(\tilde{\boldsymbol{x}})$, for $\tilde{\boldsymbol{x}} \in \Omega_{R,\rho}$, into (A.21). We make use of (A.5), the fact that $\hat{f}$ is vanishing outside Ω, $\Delta G(\boldsymbol{x},.) = -k^2 G(\boldsymbol{x},.)$ and $\boldsymbol{\nu} = -\boldsymbol{n}$ on Γ to obtain

$$\begin{aligned}(\hat{u}\ ,\ \boldsymbol{n}\cdot\nabla G(\boldsymbol{x},.))_{S_\rho} - (\boldsymbol{n}\cdot\nabla\hat{u}, G(\boldsymbol{x},.))_{S_\rho} = \left(\hat{f}, G(\boldsymbol{x},.)\right)_{\Omega_\rho} + (\boldsymbol{\nu}\cdot\nabla G(\boldsymbol{x},.), \hat{u})_\Gamma - \\ (\boldsymbol{\nu}\cdot\nabla\hat{u}, G(\boldsymbol{x},.))_\Gamma - (\boldsymbol{n}\cdot\nabla G(\boldsymbol{x},.), \hat{u})_{S_R} + (\boldsymbol{n}\cdot\nabla\hat{u}, G(\boldsymbol{x},.))_{S_R}\,.\end{aligned} \tag{A.22}$$

With the substitution $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \rho(\boldsymbol{y}-\boldsymbol{x})$, we can express the terms on the left-hand side of (A.22) as

$$(\boldsymbol{n}\cdot\nabla\hat{u}, G(\boldsymbol{x},.))_{S_\rho} = \frac{\rho^2 e^{i\rho k}}{4\pi}\int_{S(\boldsymbol{x},1)} \boldsymbol{n}\cdot\nabla_y\hat{u}(\boldsymbol{x}+\rho(\boldsymbol{y}-\boldsymbol{x}))ds(\boldsymbol{y}), \tag{A.23}$$

$$(\hat{u}, \boldsymbol{n}\cdot\nabla G(\boldsymbol{x},.))_{S_\rho} = \frac{(1-ik\rho)e^{ik\rho}}{4\pi}\int_{S(\boldsymbol{x},1)} \hat{u}(\boldsymbol{x}+\rho(\boldsymbol{y}-\boldsymbol{x}))ds(\boldsymbol{y}). \tag{A.24}$$

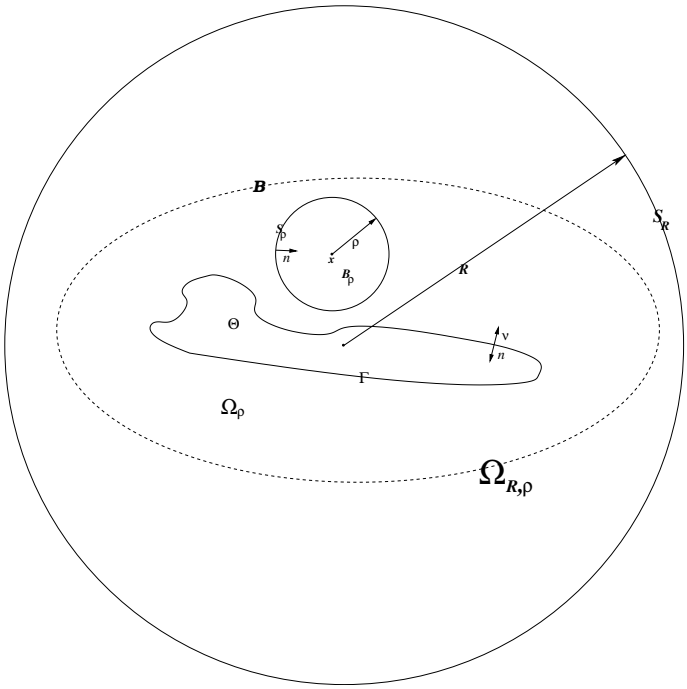


Figure A.1: Visualization of the domains

In the limit $\rho \to 0$, the quantity in equation (A.23) tends to zero, while that in equation (A.24) tends to $\hat{u}(\boldsymbol{x})$.

We turn our attention to the last two terms in (A.22). Adding and subtracting the expression $ik(u, G(\boldsymbol{x}, .))_{S_R}$, we can write them as

$$\left(\left(\frac{\partial}{\partial R} - ik\right)\hat{u}, G(\boldsymbol{x}, .)\right)_{S_R} - \left(\left(\frac{\partial}{\partial R} - ik\right)G(\boldsymbol{x}, .), \hat{u}\right)_{S_R}. \tag{A.25}$$

The absolute value of the first term in (A.25) can be estimated by the fact that the radiation condition $R(\partial_R - ik)\hat{u} \to 0$ as R tends to infinity holds and $|G(\boldsymbol{x}, \tilde{\boldsymbol{x}})| < \frac{C_1}{R}$ for some constants C_1 as $R \to \infty$. This yields

$$\left|\left(\left(\frac{\partial}{\partial R} - ik\right)\hat{u}, G(\boldsymbol{x}, .)\right)_{S_R}\right| \to 0, \quad \text{as R} \to \infty. \tag{A.26}$$

The absolute value of the second term in (A.25) can be estimated by the Schwarz inequality

$$\left|\left(\left(\frac{\partial}{\partial R}-ik\right)G(\boldsymbol{x},.),\hat{u}\right)_{S_R}\right| \;\leq\; \|\hat{u}\|_{S_R}\cdot\left\|\left(\frac{\partial}{\partial R}-ik\right)G(\tilde{\boldsymbol{x}},.)\right\|_{S_R}. \tag{A.27}$$

By direct computation, we see that $\left(\frac{\partial}{\partial R}-ik\right)G(\tilde{\boldsymbol{x}},.)\leq\frac{\tilde{C}}{R^2}$ for some constant $\tilde{C}$, and therefore the inequality $\left\|\left(\frac{\partial}{\partial R}-ik\right)G(\tilde{\boldsymbol{x}},.)\right\|_{S_R}\leq\frac{\tilde{C}}{R}$ holds. We consider (A.19) in the domain $\Omega_R=\{\tilde{\boldsymbol{x}}\in\mathbb{R}^3\backslash\Theta:|\tilde{\boldsymbol{x}}|<R\}$ for $v=\hat{u}$ and $u=\hat{u}$, which results in

$$(\boldsymbol{n}\cdot\nabla\hat{u},\hat{u})_{S_R} \;=\; -(\boldsymbol{n}\cdot\nabla\hat{u},\hat{u})_{\Gamma}-k^2(\hat{u},\hat{u})_{\Omega_R}+(\nabla\hat{u},\nabla\hat{u})_{\Omega_R}. \tag{A.28}$$

Taking the imaginary part of (A.28), it follows that

$$Im\,(\boldsymbol{n}\cdot\nabla\hat{u},\hat{u})_{S_R} \;=\; -Im\,(\boldsymbol{n}\cdot\nabla\hat{u},\hat{u})_{\Gamma}. \tag{A.29}$$

We also note that

$$\left\|(\frac{\partial}{\partial R}-ik)\hat{u}\right\|_{S_R}=\left\|\frac{\partial\hat{u}}{\partial R}\right\|_{S_R}+k^2\|\hat{u}\|_{S_R}+2k\;\mathrm{Im}(\hat{u},\frac{\partial\hat{u}}{\partial R})_{S_R} \tag{A.30}$$

tends to zero as $R\to\infty$ because of the Sommerfeld radiation condition (A.7). Combining (A.29) and (A.30) gives

$$\lim_{R\to\infty}\left\{\left\|\frac{\partial\hat{u}}{\partial R}\right\|_{S_R}+k^2\left\|\hat{u}\right\|_{S_R}\right\} \;=\; 2kIm\,(\boldsymbol{n}\cdot\nabla\hat{u},\hat{u})_{\Gamma}. \tag{A.31}$$

Both terms on the left-hand side of (A.31) are nonnegative. Hence, they must be individually bounded by a constant as $R\to\infty$ since their sum tends to a finite limit, which is the right-hand side of (A.31). We find therefore that (A.27) goes to zero in the limit $R\to\infty$. Passing to the limits $\rho\to 0$ and $R\to\infty$ in (A.22) completes the proof. □

A.3 Proof of the Kirchhoff representation

In the following, we denote $|\boldsymbol{x} - \tilde{\boldsymbol{x}}|$ by r. We establish in this section the well-known Kirchhoff representation by transforming the frequency domain expression (A.14) into an integral representation in the time domain. Kirchhoff's formula is, of course, well known, and the details of its derivation are provided here for the sake of completeness. Applying the inverse Fourier transform to (A.14), we obtain

$$u(\boldsymbol{x},t) \quad = \quad u^v(\boldsymbol{x},t) + u^m(\boldsymbol{x},t) + u^d(\boldsymbol{x},t),$$

where the first term on the r.h.s. can be simplified to

$$u^v(\boldsymbol{x},t) = \frac{1}{2\pi}\int_\Omega \int_{-\infty}^{\infty} \hat{f}(\tilde{\boldsymbol{x}},\omega) G_k(\boldsymbol{x},\tilde{\boldsymbol{x}}) e^{-i\omega t} d\omega d\tilde{\boldsymbol{x}} =$$

$$\frac{1}{2\pi}\int_\Omega \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tilde{\boldsymbol{x}},\tilde{t}) \frac{e^{ikr}}{4\pi r} e^{i\omega(\tilde{t}-t)} d\tilde{t} d\omega d\tilde{\boldsymbol{x}} =$$

$$\frac{1}{2\pi}\int_\Omega \int_{-\infty}^{\infty} \frac{f(\tilde{\boldsymbol{x}},\tilde{t})}{4\pi r} \underbrace{\int_{-\infty}^{\infty} e^{i\omega(\frac{r}{c}-t+\tilde{t})} d\omega}_{2\pi\delta(\frac{r}{c}-t+\tilde{t})} d\tilde{t} d\tilde{\boldsymbol{x}} =$$

$$\int_\Omega \int_{-\infty}^{\infty} \frac{f(\tilde{\boldsymbol{x}},\tilde{t})}{4\pi r} \delta(\frac{r}{c} - t + \tilde{t}) d\tilde{t} d\tilde{\boldsymbol{x}} =$$

$$\int_\Omega \frac{f(\tilde{\boldsymbol{x}}, t - \frac{r}{c})}{4\pi r} d\tilde{\boldsymbol{x}}.$$

A similar computation gives

$$u^m(\boldsymbol{x},t) = \frac{1}{2\pi}\int_\Gamma \int_{-\infty}^{\infty} \frac{\partial u}{\partial \nu(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}},\omega) G_k(\boldsymbol{x},\tilde{\boldsymbol{x}}) e^{-i\omega t} d\omega d\tilde{\boldsymbol{x}} = \int_\Gamma \frac{1}{4\pi r} \frac{\partial u}{\partial \nu(\tilde{\boldsymbol{x}})} \left(\tilde{\boldsymbol{x}}, t - \frac{r}{c}\right) ds(\tilde{\boldsymbol{x}}).$$

Finally, we note that

$$\frac{\partial G(\boldsymbol{x},\tilde{\boldsymbol{x}})}{\partial \nu(\tilde{\boldsymbol{x}})} = \boldsymbol{\nu}_{\tilde{\boldsymbol{x}}} \cdot \nabla_{\tilde{\boldsymbol{x}}} G(\boldsymbol{x},\tilde{\boldsymbol{x}}) = -\frac{\boldsymbol{\nu}_{\tilde{\boldsymbol{x}}} \cdot (\boldsymbol{x}-\tilde{\boldsymbol{x}})}{r}(ik - \frac{1}{r})G(\boldsymbol{x},\tilde{\boldsymbol{x}}) =$$

$$\frac{\boldsymbol{\nu} \cdot (\boldsymbol{x}-\tilde{\boldsymbol{x}})}{4\pi r^2}(\frac{1}{r} - ik)e^{i\omega(\frac{r}{c}-t+\tilde{t})} = \frac{\boldsymbol{\nu} \cdot (\boldsymbol{x}-\tilde{\boldsymbol{x}})}{4\pi r^3}e^{i\omega(\frac{r}{c}-t+\tilde{t})} + \frac{\boldsymbol{\nu} \cdot (\boldsymbol{x}-\tilde{\boldsymbol{x}})}{4\pi r^2}\frac{1}{c}\frac{\partial}{\partial t}e^{i\omega(\frac{r}{c}-t+\tilde{t})}.$$

Defining $\boldsymbol{r} = \boldsymbol{x} - \tilde{\boldsymbol{x}}$, we get

$$u^d(\boldsymbol{x},t) = \frac{1}{2\pi}\int_\Gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial G(\tilde{\boldsymbol{x}},\boldsymbol{x})}{\partial \nu(\tilde{\boldsymbol{x}})} u(\tilde{\boldsymbol{x}},\tilde{t}) e^{i\omega(\tilde{t}-t)} d\omega d\tilde{t} d\tilde{\boldsymbol{x}} =$$

$$\frac{1}{2\pi}\int_\Gamma \int_{-\infty}^{\infty} \left\{ u(\tilde{\boldsymbol{x}},\tilde{t}) \left(\frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{4\pi r^3}\int_{-\infty}^{\infty} e^{i\omega(\frac{r}{c}-t+\tilde{t})} d\omega + \frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{4\pi r^2}\frac{1}{c}\frac{\partial}{\partial t}\int_{-\infty}^{\infty} e^{i\omega(\frac{r}{c}-t+\tilde{t})} d\omega \right) \right\} d\tilde{t} ds(\tilde{\boldsymbol{x}}) =$$

$$\int_\Gamma \int_{-\infty}^{\infty} \left\{ u(\tilde{\boldsymbol{x}},\tilde{t}) \left(\frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{4\pi r^3}\delta(\frac{r}{c} - t + \tilde{t}) + \frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{4\pi r^2}\frac{1}{c}\frac{\partial}{\partial t}\delta(\frac{r}{c} - t + \tilde{t}) \right) \right\} d\tilde{t} ds(\tilde{\boldsymbol{x}}) =$$

$$\int_\Gamma \frac{\boldsymbol{\nu}\cdot\boldsymbol{r}}{4\pi r^2} \left\{ \frac{u(\tilde{\boldsymbol{x}}, t-\frac{r}{c})}{r} + \frac{1}{c}\frac{\partial u}{\partial t}\left(\tilde{\boldsymbol{x}}, t - \frac{r}{c}\right) \right\} ds.$$

A.4 Expansion in spherical harmonics

Spherical harmonics play an important role in the analysis of the Helmholtz equation; an expression of the general solution of the Helmholtz equation in terms of spherical harmonics is provided in this section. We assume that outside of a sphere centered at the origin with radius r_Γ, the homogeneous Helmholtz equation

$$\Delta \hat{u}(\boldsymbol{x}) + k^2 \hat{u}(\boldsymbol{x}) \quad = \quad 0 \quad \text{for } r = |\boldsymbol{x}| \geq r_\Gamma \tag{A.32}$$

holds. Introducing the spherical coordinates

$$\boldsymbol{x} \quad = \quad \begin{pmatrix} r\sin\theta\cos\phi \\ r\sin\theta\sin\phi \\ r\cos\theta \end{pmatrix}, \tag{A.33}$$

and denoting the unit sphere by $\Gamma = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| = 1\}$, the spherical harmonics can be expressed by

$$Y_{nm}(\theta, \phi) \quad = \quad \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_n^{|m|}(\cos\theta) e^{im\phi}, \tag{A.34}$$

where the $P_n^{|m|}$ are the associated Legendre functions. The spherical harmonics form an orthonormal basis of $L^2(\Gamma)$, see, e.g., [76]. We write (A.32) in spherical coordinates

$$\left(\frac{\partial^2}{\partial^2 r} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_S\right)\hat{u} + k^2\hat{u} \quad = \quad 0, \tag{A.35}$$

and we seek solutions of the form

$$\hat{u}(\boldsymbol{x}) \quad = \quad \sum_{n=0}^{\infty}\sum_{m=-n}^{n} u_{mn}(r)\, Y_{mn}(\theta, \phi). \tag{A.36}$$

The expansion (A.36) is substituted into (A.35), and, noticing that the Y_{nm} satisfy the eigenvalue problem

$$\Delta_S Y_{mn} \quad = \quad -n(n+1)\, Y_{mn}, \tag{A.37}$$

for the Laplace-Beltrami operator

$$\Delta_S \quad = \quad \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \tag{A.38}$$

we multiply equation (A.35) by $\overline{Y_{\tilde{m}\tilde{n}}}$ and integrate over the unit sphere to get

$$\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + k^2 - \frac{n(n+1)}{r^2}\right) u_{mn} \quad = \quad 0, \quad r \geq r_\Gamma. \tag{A.39}$$

Dividing this equation by k^2 and making the change of variable $\tilde{r} = kr$ along with $u_{mn}(r) = \tilde{u}(\tilde{r})$ gives rise to the spherical Bessel equation. The radial part of the expansion is therefore given by

$$u_{mn}(r) \quad = \quad c_{mn} h_n^{(1)}(rk) + d_{mn} h_n^{(2)}(rk), \tag{A.40}$$

where c_{mn} and d_{mn} are constants, and where $h_n^{(1,2)}$ are the spherical Hankel functions of the first and second kind,

$$h_n^{(1)}(r) = (-r)^n \left(\frac{1}{r}\frac{\partial}{\partial r}\right)^n \left(\frac{e^{ir}}{r}\right) = (-i)^n \frac{e^{ir}}{r} \sum_{m=0}^{n} (i)^m \beta_m^n \left(\frac{1}{r}\right)^m, \tag{A.41}$$

$$h_n^{(2)}(r) = \overline{h_n^{(1)}(r)}, \quad \beta_m^n = \frac{(m+n)!}{m!(n-m)!2^m}, \tag{A.42}$$

see [76]. Alternatively, the spherical Hankel functions can be defined as the linear combination of the two linearly independent classes of spherical Bessel functions j_n and spherical Neumann functions y_n:

$$h_n^{(1)}(r) = -y_n(r) + ij_n(r) \tag{A.43}$$

$$h_n^{(2)}(r) = -y_n(r) - ij_n(r). \tag{A.44}$$

Note that the definitions of the spherical Hankel functions vary in the literature. [2] In the rest of this section, we investigate the impact of the outer operator on the uniqueness of the solution. This topic is related to Section 4.1.4. As we shall see, a Sommerfeld-type operator is crucial to obtain a uniqueness proof.

Unbounded domain: Sommerfeld's radiation condition at infinity

Only the $h_n^{(1)}$ family consists of outgoing waves and satisfies the radiation condition (A.7) at infinity. More precisely, the expression $r(\partial_r - ik)h_n^{(1)}(rk)$ behaves like $(-i)^n[-exp(ikr)/(rk)]\beta_0^n$ as $r \to \infty$ and $r(\partial_r - ik)h_n^{(2)}(rk) \sim -2(i)^{n+1}exp(-ikr)\beta_0^n$ as $r \to \infty$. It follows that all constants d_{mn} must vanish to satisfy Sommerfeld's boundary condition at infinity. Suppose we know the Dirichlet values of $\hat{u}$ on the sphere with radius r_Γ centered around the origin. The solution in the exterior of the sphere can be written as the expansion

$$\hat{u}(\boldsymbol{x}) = \sum_{n=0}^{\infty} \frac{h_n^{(1)}(rk)}{h_n^{(1)}(r_\Gamma k)} \sum_{m=-n}^{n} c_{mn} Y_{mn}(\theta, \phi), \tag{A.45}$$

$$c_{mn} = (\hat{u}, Y_{mn})_\Gamma \,. \tag{A.46}$$

[2] For example, in [24], the spherical Hankel function of the first kind is defined as $h_n^{(1)}(r) = (-i)^n \frac{e^{ir}}{ir} \sum_{m=0}^{n} (i)^m \beta_m^n \left(\frac{1}{r}\right)^m$, or equivalently as $h_n^{(1)} = j_n + iy_n$.

We note that formula (A.45) holds for any $r_\Gamma > 0$. It can be shown that if the series converges in the mean square sense on the sphere $|\boldsymbol{x}| = r_\Gamma$, then it necessarily converges absolutely and uniformly on compact subsets $|\boldsymbol{x}| > r_\Gamma$. Away from the origin, it solves the homogeneous Helmholtz equation.
We point out that $|h_n^{(1)}(rk)|^2 = \left(\alpha_0^n + \alpha_1^n/(rk)^2 + \ldots + \alpha_n^n/(rk)^{2n}\right)/(rk)^2$ with suitable positive constants $\alpha_0^n, \ldots, \alpha_n^n$ (see [76] for further details) is nonvanishing, and thus k^2 is never an eigenvalue to the operator $-\Delta$ in the annulus, which makes (A.45) the unique radial part of the Helmholtz problem with a Dirichlet condition on r_Γ and Sommerfeld data at infinity. This result is a fundamental difference to a purely Dirichlet or Neumann problem, as we demonstrate in the following subsections: we consider the example of a domain with spherical finite inner and outer boundary, and we seek boundary conditions on the outer sphere that uniquely determine the field in the bounded domain.

Finite domain: Dirichlet or Neumann boundary condition on outer boundary

We consider (A.32) defined in the annulus $r_\Gamma < r < r_\mathcal{B}$ along with the Dirichlet boundary conditions $u|_\Gamma = g_\Gamma$ at the inner sphere and $u|_\mathcal{B} = g_\mathcal{B}$ at the outer sphere with the given Dirichlet data g_Γ and $g_\mathcal{B}$. This leads to

$$u_{mn}^\mathcal{B} = c_{mn} h_n^{(1)}(r_\mathcal{B} k) + d_{mn} h_n^{(2)}(r_\mathcal{B} k), \tag{A.47}$$

$$u_{mn}^\Gamma = c_{mn} h_n^{(1)}(r_\Gamma k) + d_{mn} h_n^{(2)}(r_\Gamma k), \tag{A.48}$$

where $u_{mn}^\Gamma = (g_\Gamma, Y_{mn})_\Gamma$ and $u_{mn}^\mathcal{B} = (g_\mathcal{B}, Y_{mn})_\mathcal{B}$. The constants are given by

$$c_{mn} = \frac{g_{mn}^\Gamma h_n^{(2)}(r_\mathcal{B} k) - g_{mn}^\mathcal{B} h_n^{(2)}(r_\Gamma k)}{h_n^{(1)}(r_\Gamma k) h_n^{(2)}(r_\mathcal{B} k) - h_n^{(1)}(r_\mathcal{B} k) h_n^{(2)}(r_\Gamma k)} \tag{A.49}$$

$$d_{mn} = \frac{g_{mn}^\Gamma h_n^{(1)}(r_\mathcal{B} k) - g_{mn}^\mathcal{B} h_n^{(1)}(r_\Gamma k)}{h_n^{(1)}(r_\mathcal{B} k) h_n^{(2)}(r_\Gamma k) - h_n^{(1)}(r_\Gamma k) h_n^{(2)}(r_\mathcal{B} k)}. \tag{A.50}$$

Equations (A.49) and(A.50) are not valid if their denominator vanishes. This occurs if

$$j_n(r_\mathcal{B} k) y_n(r_\Gamma k) = y_n(r_\mathcal{B} k) j_n(r_\Gamma k). \tag{A.51}$$

Equality (A.51) holds if and only if k^2 is a Dirichlet eigenvalue of the negative Laplacian in the spherical shell. Thus, we see that if we impose Dirichlet boundary conditions on the inner and outer sphere, uniqueness of solutions does not hold for certain wave numbers k. A similar situation arises for Neumann boundary conditions.

Finite domain: Sommerfeld-type boundary condition on the outer boundary

We emphasize that the lack of uniqueness discussed in the previous section does not occur if a Sommerfeld-type operator is imposed on the outer sphere of radius $r_{\mathcal{B}}$ of a spherical shell. In order to demonstrate this fact, let us assume that we are given Dirichlet data at r_Γ and the condition $(\partial_r u - iku)|_{\mathcal{B}} = g_{\mathcal{B}}$. The two equations which determine the constants c_{mn} and d_{mn} are

$$g^\Gamma_{mn} = c_{mn} h^{(1)}_n(r_\Gamma k) + d_{mn} h^{(2)}_n(r_\Gamma k) \tag{A.52}$$

and

$$g^{\mathcal{B}}_{mn} = c_{mn} k \left[\frac{d}{dr} h^{(1)}_n(r)|_{r_{\mathcal{B}} k} - i h^{(1)}_n(r_{\mathcal{B}} k) \right] + d_{mn} k \left[\frac{d}{dr} h^{(2)}_n(r)|_{r_{\mathcal{B}} k} - i h^{(2)}_n(r_{\mathcal{B}} k) \right]. \tag{A.53}$$

The formal solution of the linear system (A.52)–(A.53) is

$$c_{mn} = \frac{g^{\mathcal{B}}_{mn} h^{(2)}_n(r_\Gamma k) - k g^\Gamma_{mn} \left[\frac{d}{dr} h^{(2)}_n(r)|_{r_{\mathcal{B}} k} - i h^{(2)}_n(r_{\mathcal{B}} k) \right]}{D_n} \tag{A.54}$$

$$d_{mn} = \frac{k g^\Gamma_{mn} \left[\frac{d}{dr} h^{(1)}_n(r)|_{r_{\mathcal{B}} k} - i h^{(1)}_n(r_{\mathcal{B}} k) \right] - g^{\mathcal{B}}_{mn} h^{(1)}_n(r_\Gamma k)}{D_n}, \tag{A.55}$$

where

$$D_n = k \left\{ h^{(2)}_n(r_\Gamma k) \left[\frac{d}{dr} h^{(1)}_n(r)|_{r_{\mathcal{B}} k} - i h^{(1)}_n(r_{\mathcal{B}} k) \right] - h^{(1)}_n(r_\Gamma k) \left[\frac{d}{dr} h^{(2)}_n(r)|_{r_{\mathcal{B}} k} - i h^{(2)}_n(r_{\mathcal{B}} k) \right] \right\}$$

simplifies to

$$\begin{aligned} D_n &= 2k\left\{j_n(r_\Gamma k)y_n(r_\mathcal{B} k) - j_n(r_\mathcal{B} k)y_n(r_\Gamma k)\right\} + \\ &\quad 2ki\left\{j_n(r_\Gamma k)y_n'(r_\mathcal{B} k) - y_n(r_\Gamma k)j_n'(r_\mathcal{B} k)\right\}. \end{aligned} \tag{A.56}$$

Therefore,

$$\begin{aligned} |D_n|^2 &= 4k^2\left[j_n(r_\Gamma k)y_n(r_\mathcal{B} k) - j_n(r_\mathcal{B} k)y_n(r_\Gamma k)\right]^2 + \\ &\quad 4k^2\left[j_n(r_\Gamma k)y_n'(r_\mathcal{B} k) - j_n'(r_\mathcal{B} k)y_n(r_\Gamma k)\right]^2, \end{aligned} \tag{A.57}$$

and thus $|D_n|^2$ vanishes if and only if

$$\begin{bmatrix} y_n'(r_\mathcal{B} k) & -j_n'(r_\mathcal{B} k) \\ y_n(r_\mathcal{B} k) & -j_n(r_\mathcal{B} k) \end{bmatrix} \begin{bmatrix} j_n(r_\Gamma k) \\ y_n(r_\Gamma k) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.58}$$

The determinant of the matrix in (A.58) is equal to $y_n(r_\mathcal{B} k)j_n'(r_\mathcal{B} k) - y_n'(r_\mathcal{B} k)j_n(r_\mathcal{B} k)$. This is exactly the negative Wronskian of $j_n(r)$ and $y_n(r)$. Because these two functions are linearly independent, the Wronskian is never zero (in fact, the Wronskian equals $-\frac{1}{(r_\mathcal{B} k)^2}$, see, e.g., [24]). It follows that the matrix in (A.58) is invertible and the solution to the corresponding system is $j_n(r_\Gamma k) = 0$ and $y_n(r_\Gamma k) = 0$. As these two functions never have common zeros (see [1, 76]), this condition can't occur and we conclude that the constants c_{mn} and d_{mn} are uniquely determined by (A.54) and (A.55).

Appendix B

Equivalent sources

In this appendix we describe a method to represent solutions of the Helmholtz equation by means of planar distributions of sources. We consider an outgoing wave $\hat{u}$ that is generated by sources contained in the ball of radius r_Γ centered at the origin: the solution we consider is given by a certain integral defined on a bounded domain (contained within the sphere of radius r_Γ) that is a surface, a volume, or both. Outside of this ball, equation (A.32) holds and the field can thus be represented as an expansion (A.45) for $|\boldsymbol{x}| > r_\Gamma$. Now, we pose the following question: Is it possible to distribute some artificial sources on a disc within the ball, such that these new introduced sources generate a field which coincides to a high accuracy with the initial wave? This question has been affirmatively answered in [13, 14]. In this appendix, we discuss the details and formulate efficient algorithms which are relevant for our purpose of constructing the data on the artificial outer boundary.

B.1 Equivalent source distribution on a disc

Theorem B.1.1. *Let $\epsilon > 0$, $b > a > 0$, $B(\boldsymbol{r}_0; b) = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x} - \boldsymbol{r}_0| < b\}$ and*

$$\hat{u}(\boldsymbol{x}) \;=\; \sum_{n=0}^{\infty} h_n^{(1)}(kr) \sum_{m=-n}^{n} c_{nm} Y_{nm}(\theta, \phi) \tag{B.1}$$

with some bounded constants c_{nm}. Then, there is a combination of single- and double-layer potentials $\hat{u}_D$ with densities supported on the disc $D = \{\boldsymbol{x} \in \mathbb{R}^3 : y = 0, x^2 + z^2 < a^2\}$, which differs from $\hat{u}(\boldsymbol{x})$ in less than ϵ for all $\boldsymbol{x} \in \mathbb{R}^3 \backslash B(\boldsymbol{r}_0; b)$.

Proof. We note that (B.1) satisfies the homogeneous Helmholtz equation (A.32) away from the origin along with the radiation condition at infinity. We denote the origin by $\boldsymbol{r}_0$ and define two new coordinate systems with origins at the points

$$\boldsymbol{r}_1(\delta) = \boldsymbol{r}_0 - \delta \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{r}_2(\delta) = \boldsymbol{r}_0 + \delta \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \tag{B.2}$$

for $\delta > 0$ (see Figure B.1). Next, we define balls B^i_δ around the centers $\boldsymbol{r}_i$, $i = 1$ or 2, with radii *larger*

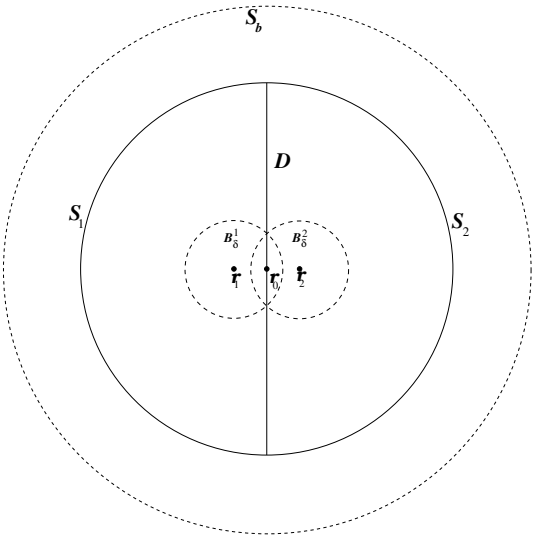


Figure B.1: The geometry of Theorem B.1.1

than δ, i.e., both points $\boldsymbol{r}_0$ and $\boldsymbol{r}_i$ lie in B^i_δ. Now the Dirichlet problems

$$\Delta \hat{u}^i + k^2 \hat{u}^i = 0, \qquad \boldsymbol{x} \in \mathbb{R}^3 \backslash B^i_\delta \tag{B.3}$$

$$\hat{u}^i = \hat{u}|_{\partial B^i_\delta}, \quad \boldsymbol{x} \in \partial B^i_\delta \tag{B.4}$$

$$r^i \left(\partial_{r^i} - ik\right) \hat{u}^i = 0, \qquad \text{as } r^i \to \infty, \tag{B.5}$$

admit the solution

$$\hat{u}^i = \sum_{n=0}^{\infty} h_n^{(1)}(kr^i) \sum_{m=-n}^{n} c^i_{nm} Y_{nm}(\theta^i, \phi^i), \tag{B.6}$$

where the c^i_{nm} are uniquely determined by (A.46) and (B.4). With these constants, this formula defines a solution of the homogeneous Helmholtz equation for $\boldsymbol{x} \in \mathbb{R}^3 \backslash \{\boldsymbol{r}_i\}$. By construction, $\hat{u}$ and $\hat{u}^i$ solve the Helmholtz equation outside B^i_δ with the same boundary conditions on ∂B^i_δ and at infinity. By uniqueness, it follows that $\hat{u} = \hat{u}^i$ outside of B^i_δ. In particular, for any point outside the ball $B(\boldsymbol{r}_0; b)$, the identity $\hat{u}^1 = \hat{u}^2 = \hat{u}$ holds. Therefore, we can write

$$\hat{u} \;=\; \hat{u}^1 + \hat{u}^2 - \hat{u}, \quad \boldsymbol{x} \in \mathbb{R}^3 \backslash B(\boldsymbol{r}_0; b). \tag{B.7}$$

Since all three functions $\hat{u}, \hat{u}^i$ satisfy the homogeneous Helmholtz equation away from their origins, we can make use of the representation theorem in Section A.2.2 to get

$$\hat{u} \;=\; \int_{S_1 \cup S_2} \hat{u}(\tilde{\boldsymbol{x}}) \frac{\partial G}{\partial \nu_i(\tilde{\boldsymbol{x}})}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) - \frac{\partial \hat{u}}{\partial \nu_i}(\tilde{\boldsymbol{x}}) G(\boldsymbol{x}, \tilde{\boldsymbol{x}}) ds(\tilde{\boldsymbol{x}}), \tag{B.8}$$

$$\hat{u}^i \;=\; \int_{S_i \cup D} \hat{u}^i(\tilde{\boldsymbol{x}} - \boldsymbol{r}_i(\delta)) \frac{\partial G}{\partial \nu_i(\tilde{\boldsymbol{x}})}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) - \frac{\partial \hat{u}^i}{\partial \nu_i}(\tilde{\boldsymbol{x}} - \boldsymbol{r}_i(\delta)) G(\boldsymbol{x}, \tilde{\boldsymbol{x}}) ds(\tilde{\boldsymbol{x}}). \tag{B.9}$$

Comparing these results with (B.7) and noticing that $\boldsymbol{\nu}_1 = -\boldsymbol{\nu}_2$ on the disc D, it follows that

$$\hat{u} = \hat{u}_S + \hat{u}_D, \tag{B.10}$$

where

$$\begin{aligned} \hat{u}_S \;=\; & \sum_{i=1}^{2} \int_{S_i} (\hat{u}^i(\tilde{\boldsymbol{x}} - \boldsymbol{r}_i(\delta)) - \hat{u}(\tilde{\boldsymbol{x}})) \frac{\partial G}{\partial \nu_i(\tilde{\boldsymbol{x}})}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) - \\ & \left(\frac{\partial \hat{u}^i}{\partial \nu_i(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}} - \boldsymbol{r}_i(\delta)) - \frac{\partial \hat{u}^i}{\partial \nu(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}}) \right) G(\boldsymbol{x}, \tilde{\boldsymbol{x}}) ds(\tilde{\boldsymbol{x}}), \end{aligned} \tag{B.11}$$

$$\begin{aligned} \hat{u}_D \;=\; & \int_D (\hat{u}^1(\tilde{\boldsymbol{x}} - \boldsymbol{r}_1(\delta)) - \hat{u}^2(\tilde{\boldsymbol{x}} - \boldsymbol{r}_2(\delta))) \frac{\partial G}{\partial \nu_1}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) ds(\tilde{\boldsymbol{x}}) - \\ & \int_D \left(\frac{\partial \hat{u}^1}{\partial \nu_1(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}} - \boldsymbol{r}_1(\delta)) - \frac{\partial \hat{u}^1}{\partial \nu_1(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}} - \boldsymbol{r}_2(\delta)) \right) G(\boldsymbol{x}, \tilde{\boldsymbol{x}}) ds(\tilde{\boldsymbol{x}}). \end{aligned} \tag{B.12}$$

The Green's function and its derivative are bounded in (B.11), because the distance from any point on S to the point $\boldsymbol{x}$ is nonvanishing. In view of the uniform continuity of the functions $\hat{u}$ and $\hat{u}^i$ *away from*

their origins (see Section A.4), for any positive numbers $\tilde{\epsilon}_1$ and $\tilde{\epsilon}_2$ there exists a positive constant $\tilde{\delta}_0$ such that, $|\hat{u}^i(\tilde{\boldsymbol{x}} - \boldsymbol{r}_i(\delta)) - \hat{u}(\tilde{\boldsymbol{x}})| < \tilde{\epsilon_1}$ and $|\frac{\partial \hat{u}^i}{\partial \nu(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}} - \boldsymbol{r}_i(\delta)) - \frac{\partial \hat{u}}{\partial \nu(\tilde{\boldsymbol{x}})}(\tilde{\boldsymbol{x}})| < \tilde{\epsilon}_2$ for all $\delta < \tilde{\delta}_0$. As δ approaches zero, the integrals over S_i vanish, i.e., for any $\epsilon > 0$, there is a δ_0, such that for $\delta < \delta_0$, $|\hat{u}_S| < \epsilon$. We conclude that $|\hat{u} - \hat{u}_D| = |\hat{u}_S| < \epsilon$ for $\delta < \delta_0$. □

B.2 Two-face approach

We assume that the distribution of all known sources are contained in a cubic cell c_i of side H. Let us introduce the spheres $S_{1,2} = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x} \mp \frac{H}{2}[0,1,0]| = \frac{\sqrt{3}}{2}H\}, S_i^{1,2} = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x} \mp \frac{H}{2}[0,1,0]| = \frac{H}{2}\}$, and $S_b = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| = \frac{1+\sqrt{3}}{2}H\}$ (see Figure B.2.1). Theorem B.1.1 implies the following theorem (see [12, 13, 14] for more details):

Theorem B.2.1. *Outside S_b, the field $\hat{u}_{c_i}(\boldsymbol{x})$ induced by sources contained in a cubic cell c_i can be approximated with a prescribed accuracy by the sum of a single- and a double-layer potential with densities distributed over any pair of parallel faces of c_i.*

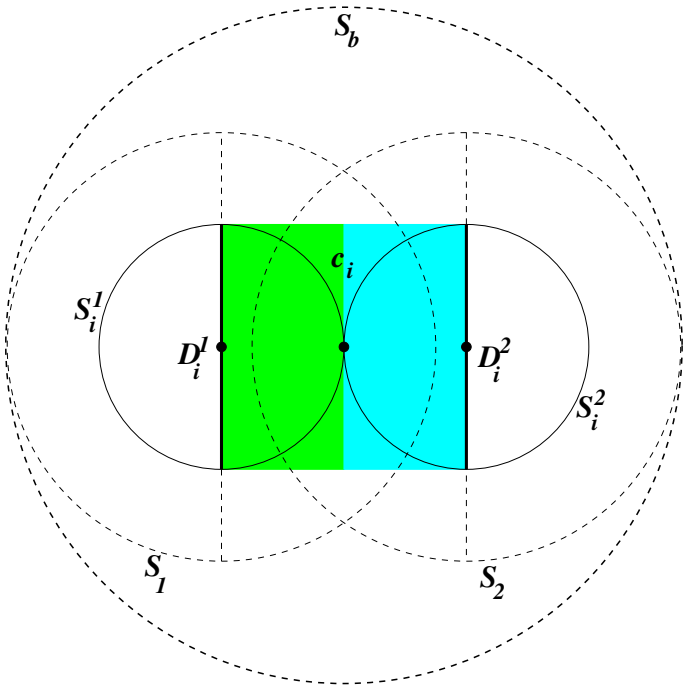


Figure B.2: The geometry of Theorem B.2.1

We select one of the three pairs of the cube's parallel faces and denote it by $D_1 \cup D_2$. Theorem B.2.1

states that there exist sources $\xi(\tilde{\boldsymbol{x}})$ and $\eta(\tilde{\boldsymbol{x}})$ located at $\tilde{\boldsymbol{x}} \in D_1 \cup D_2$, such that

$$\mathcal{P}[\xi,\eta](\boldsymbol{x}) \equiv \int_{D_1\cup D_2} \left(\frac{\partial G}{\partial \nu(\tilde{\boldsymbol{x}})}(\boldsymbol{x},\tilde{\boldsymbol{x}})\xi(\tilde{\boldsymbol{x}}) - G(\boldsymbol{x},\tilde{\boldsymbol{x}})\eta(\tilde{\boldsymbol{x}}) \right) ds(\tilde{\boldsymbol{x}}) \tag{B.13}$$

approximates $\hat{u}$ at any $\boldsymbol{x}$ outside of S_b in absolute value in less than an arbitrary small $\epsilon > 0$, i.e.,

$$|\mathcal{P}[\xi,\eta](\boldsymbol{x}) - \hat{u}(\boldsymbol{x})| < \epsilon. \tag{B.14}$$

In practice, the integrals in (B.13) must be discretized, leading to the representation of the form

$$\mathcal{I}[\xi,\eta](\boldsymbol{x}) \equiv \sum_{n=1}^{n_S} \left(\frac{\partial G}{\partial \nu(\boldsymbol{y}_n)}(\boldsymbol{x},\boldsymbol{y}_n)\ \tilde{\xi}(\boldsymbol{y}_n) - G(\boldsymbol{x},\boldsymbol{y}_n)\tilde{\eta}(\boldsymbol{y}_n) \right), \tag{B.15}$$

where $\tilde{\xi}$ and $\tilde{\eta}$ are the equivalent sources sampled at the integration points and multiplied by the corresponding weights. We note that when studying the truncation error of the Green's function, i.e.,

$$e_N \equiv \left| k \sum_{n=N+1}^{\infty} \sum_{m=-n}^{n} h_n^{(1)}(k|\boldsymbol{x}|) Y_n^m \left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|} \right) j_n(k|\tilde{\boldsymbol{x}}|) \bar{Y}_n^m \left(\frac{\tilde{\boldsymbol{x}}}{|\tilde{\boldsymbol{x}}|} \right) \right|, \tag{B.16}$$

the estimate

$$e_N \leq 3^{-N/2} \tag{B.17}$$

can be obtained under the assumption that $N \approx 2kH$, see [14].

We define the vectors $\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\eta}}$ with the entries $\tilde{\boldsymbol{\xi}}_i = \xi(\boldsymbol{y}_i)$ and $\tilde{\boldsymbol{\eta}}_i = \eta(\boldsymbol{y}_i)$, and consider the exact and approximated field values at n_C distinct points $\boldsymbol{x}_l$ for $l \in \{1,\ldots,n_C\}$. Denoting $\boldsymbol{r}_{l,j} = \boldsymbol{x}_l - \boldsymbol{y}_j$ and $r_{l,j} = |\boldsymbol{r}_{l,j}|$, equation (B.15) can be written in the matrix form

$$\tilde{\boldsymbol{\psi}} = \begin{bmatrix} \boldsymbol{A}_m & \boldsymbol{A}_d \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\xi}} \\ \tilde{\boldsymbol{\eta}} \end{bmatrix}, \tag{B.18}$$

where

$$\tilde{\boldsymbol{\psi}} = \begin{bmatrix} \mathcal{I}[\xi,\eta](\boldsymbol{x}_1) \\ \vdots \\ \mathcal{I}[\xi,\eta](\boldsymbol{x}_l) \\ \vdots \\ \mathcal{I}[\xi,\eta](\boldsymbol{x}_{n_C}) \end{bmatrix}, \tag{B.19}$$

$$\{\boldsymbol{A}_m\}_{l,j} = \frac{e^{ikr_{l,j}}}{4\pi r_{l,j}}, \tag{B.20}$$

and

$$\{\boldsymbol{A}_d\}_{l,j} = \frac{e^{ikr_{l,j}}}{4\pi r_{l,j}^2}\left(\frac{1}{r_{l,j}} - ik\right)\boldsymbol{\nu}_j \cdot \boldsymbol{r}_{l,j}. \tag{B.21}$$

Assuming that the field $\hat{u}$ is known at the n_C distinct collocation points $\boldsymbol{x}_l$, the goal is to find the equivalent source distribution on the two faces. This motivates us to look for monopole sources $\boldsymbol{\xi} \in \mathbb{R}^{n_s \times 1}$ and dipole sources $\boldsymbol{\eta} \in \mathbb{R}^{n_s \times 1}$ such that $\|\hat{\boldsymbol{u}} - (\boldsymbol{A}_m\boldsymbol{\xi} + \boldsymbol{A}_d\boldsymbol{\eta})\|_2$ is minimized, i.e., we seek to solve

$$\min_{\boldsymbol{\xi},\boldsymbol{\eta}} \|\hat{\boldsymbol{u}} - (\boldsymbol{A}_m\boldsymbol{\xi} + \boldsymbol{A}_d\boldsymbol{\eta})\|_2 . \tag{B.22}$$

The solution of the overdetermined system (B.22) can be obtained by the pseudo-inverse of the matrix $[\boldsymbol{A}_m, \boldsymbol{A}_d]$, which involves a singular value decomposition that is generally an order $\mathcal{O}(n_C \cdot n_S^2)$ operation. Once the equivalent sources $\boldsymbol{\xi}, \boldsymbol{\eta}$ are computed, the field can be evaluated to a high accuracy at any point outside of the collocation surface. In practice, we have to determine the position of the collocation points. In [13], it is suggested to embed the cube with the two panels of size H into the center of a three-times-larger cube of size $3H$ and choose points on this surface as collocation points. This leads to the following algorithm:

Algorithm B.2.1.

1. Given two opposite faces of a cube with length H, choose an appropriate equivalent source grid $\tau_S^{(l)}$

on the panels with nodes $\boldsymbol{y}_i$ and mesh size Δ_S—see Figure 3.1. Select equally spaced collocation points $\boldsymbol{x}_l$ on the six faces of a three-times-larger cube that embeds the cell with the equivalent sources. We assume that the field values $\hat{\boldsymbol{u}}_l = \hat{u}(\boldsymbol{x}_l)$ are known at all collocation points.

2. Build the system matrices (B.20) and (B.21).

3. Solve the least-square problem (B.22).

4. The field values can be evaluated at any point outside the collocation surface by the matrix-vector multiplication (B.18).

Effect of various disc sizes on accuracy and efficiency

We seek to substitute the true sources contained in a cubic cell c_i by equivalent sources on the faces of c_i in such a way that the field produced by the true and equivalent sources coincide to within a prescribed numerical accuracy. In [12, 13, 14], it is proposed to place the equivalent sources in a set Π_i. The set Π_i consists of points which lie within the union of two circular domains concentric with (and containing) two opposite faces of c_i. The radius of these domains is chosen to be equal to (or slightly larger than) the length of half of the diagonals of the faces.

We demonstrate that in the context of this thesis it is advantageous in terms of accuracy and computing time to place equivalent sources directly on the Cartesian grids $\tau_S^{(1)} \cup \tau_S^{(2)}$ of the faces $D_1 \cup D_2$ of c_i with edge side H. We set the radius of Π_i to H, in which case it is slightly larger than half the diagonal of the faces. Two separate computations are performed with Algorithm B.2.1 to obtain the equivalent sources on Π_i and $D_1 \cup D_2$, respectively, and the field is then evaluated on a slightly larger cubic surface than the collocation surface. Specifically, the side length of the cube is $3H + 2 \cdot 10^{-4}$. In the Figures B.4 and B.5 we plot the maximum absolute error on that surface versus the number of locations of the equivalent sources n_S for the wave numbers $k = 0.25, k = 1.25, k = 4$, and $k = 12$, respectively. The side length of c_i is chosen as $H = 2$ for this specific computation. We observe the same quantitative behavior in all cases: as n_S increases, the equivalent sources on $D_1 \cup D_2$ produce up to a certain critical value $n_S^{crit,1} \approx 1,000$ more accurate values than the distributions on Π_i. But as n_S approaches this value, the convergence of the distributions on $D_1 \cup D_2$ slows down, while the representation induced by the sources from Π_i continues rapidly converging to the exact solution until n_S reaches a second critical value $n_S^{crit,2} \approx 4,000$. At this point, there is a significant difference between the errors of the two

curves. We note that as kH increases, the numerical solutions for a certain n_S are less accurate. In the Figures 3.6–3.10, we consider a maximal value of $kH = 6$. From Figures B.4–B.5 we conclude that if the prescribed accuracy of the equivalent sources is $\mathcal{O}(10^{-8})$ or less for $kH < 6$, the distribution on $D_1 \cup D_2$ is advantageous over the distribution on Π_i. In the context of this thesis, the errors produced by the finite element and finite difference schemes are larger than this order for the mesh sizes we consider. Thus, we deal with the range $n_S < n_S^{crit,1}$, and choose to distribute the equivalent source locations on $D_1 \cup D_2$ rather than on Π_i.

Finally, we emphasize that the accuracy of the equivalent sources representation is not limited: we can obtain more and more accurate solutions all the way to machine precision by increasing the sizes of the discs in the set Π_i appropriately. We demonstrate this in the example when $H = 2$ and $k = 2$. The radius of the various $\Pi_i(\alpha)$ is defined by $(1 + \alpha)H/2$, where α is a constant. In Figure B.3, we see that higher accuracy is achieved as α is increased appropriately. It should be noted, as we already stated before, that increasing the collocation box also results in higher accuracy.

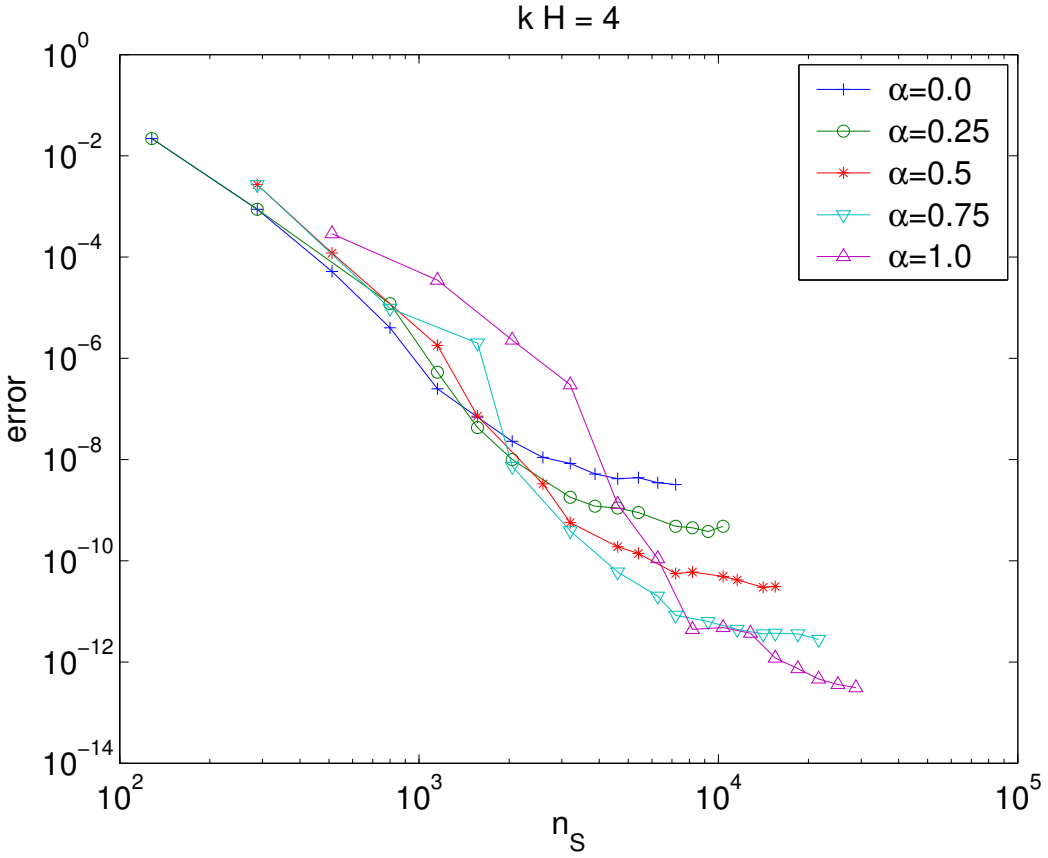


Figure B.3: Convergence study as α increases

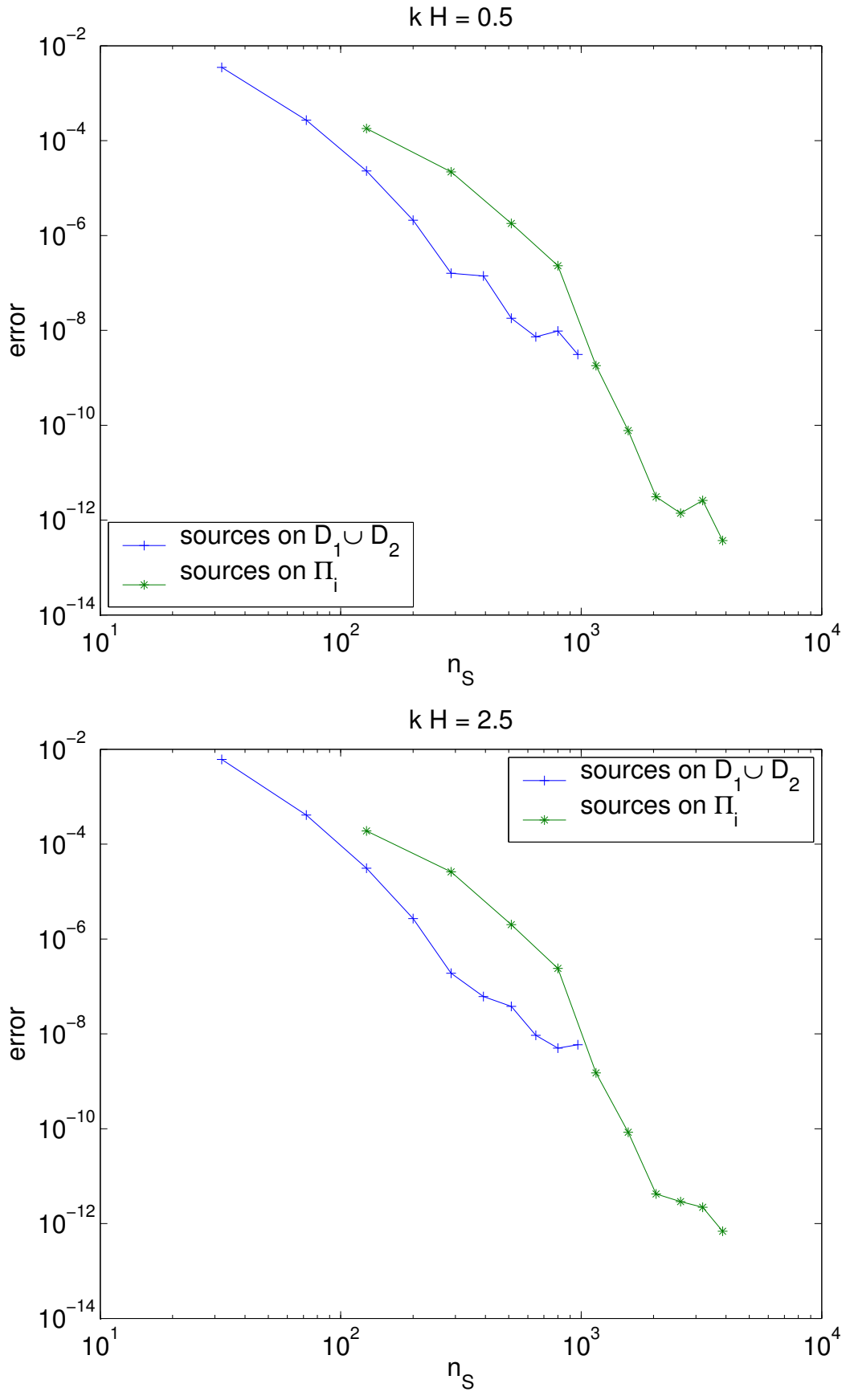


Figure B.4: Error versus the number of locations of the equivalent sources n_S for $kH = 0.5$ and $kH = 2.5$

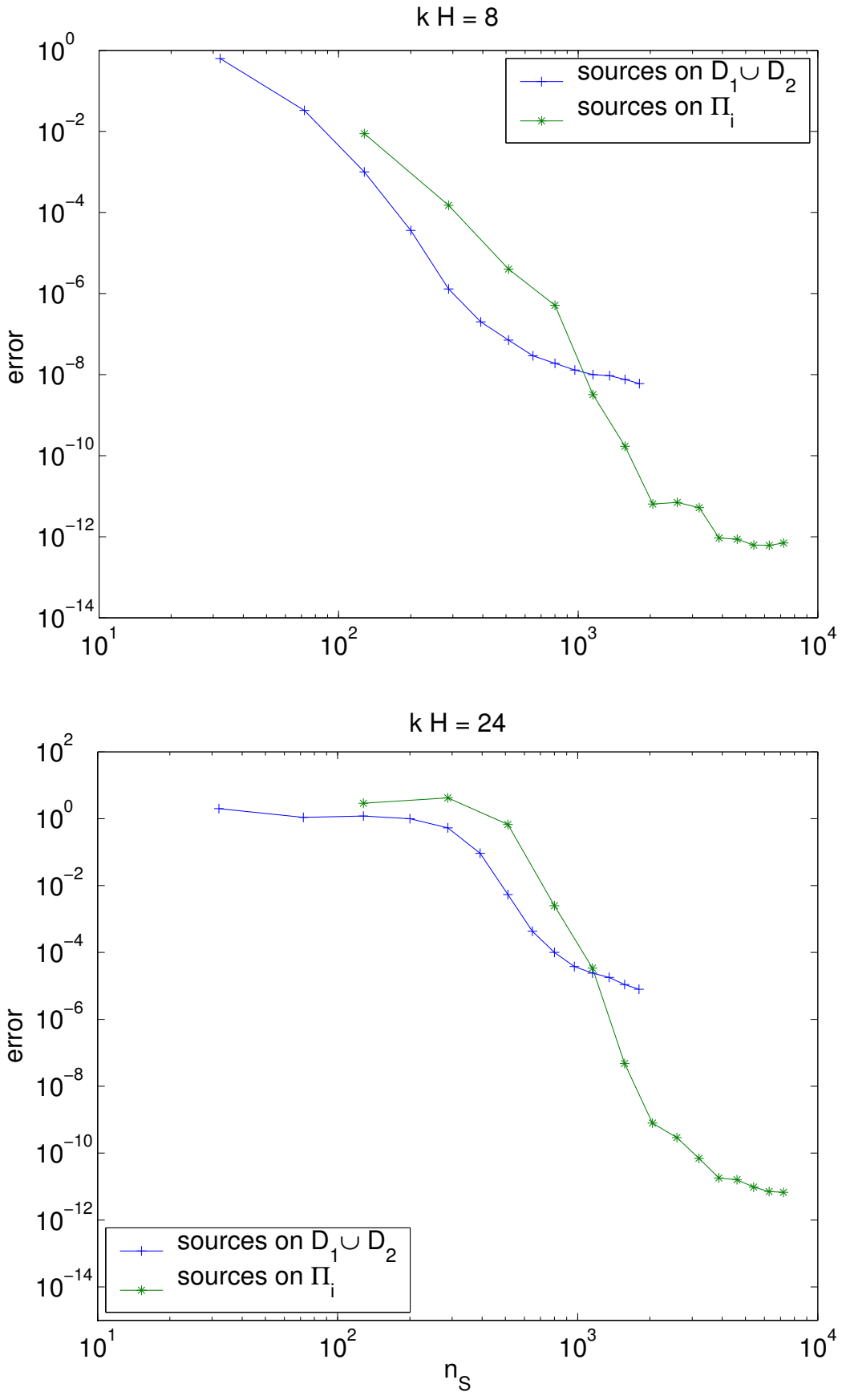


Figure B.5: Error versus the number of locations of the equivalent sources n_S for $kH = 8$ and $kH = 24$

B.3 Fast sampling in space

For special geometries, the convolution (B.18) can be evaluated rapidly by means of a fast Fourier transform (FFT). We explain this on the basis of the two-face approach where the equivalent sources are distributed on the regions $\tau_S^{(1)} \cup \tau_S^{(2)}$—see Figure 3.1. The positions of the equivalent sources on the discs D_l are at $\boldsymbol{y}_j^{(l)}$, where $l \in \{1,2\}$. Further, the two-dimensional multi-index j is defined in the range of $(j_1, j_2) \in \{1,\ldots,S\} \times \{1,\ldots,S\}$. We assume that the plane $\mathcal{B}$ is parallel to the discs $D_1 \cup D_2$, and a two-dimensional equidistant grid $\tau_S^{(3)}$ is constructed on it with nodes $\boldsymbol{x}_k = \boldsymbol{y}_k^{(l)} + \boldsymbol{d}^{(l)}$, where $\boldsymbol{d}^{(l)}$ is the distance vectors from D_l to $\mathcal{B}$, see Figure B.6. We denote the monopole and dipole source distributions

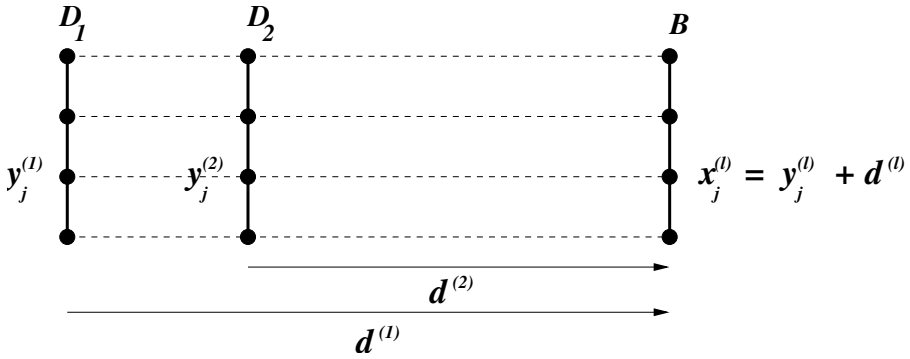


Figure B.6: Geometry to sample on $\mathcal{B}$ with a FFT

on D_l by $\xi_j^{(l)}$ and $\eta_j^{(l)}$, where $j \in \tau_S^{(l)}$ and $l \in \{1,2\}$, and recall that the field at the points $\boldsymbol{x}_k$ can be computed by

$$\psi_k \;=\; \sum_{l=1}^{2} \psi_k^{(l)}, \quad \text{for } k \in \tau_S^{(3)}, \tag{B.23}$$

where

$$\psi_k^{(l)} \;\equiv\; \sum_{j \in \tau_S^{(l)}} \{ m_{k-j}^{(l)} \xi_j^{(l)} + d_{k-j}^{(l)} \eta_j^{(l)} \}, \tag{B.24}$$

and

$$m_{k-j}^{(l)} \equiv G(\boldsymbol{y}_j^{(l)}, \boldsymbol{x}_k), \tag{B.25}$$

$$d_{k-j}^{(l)} \equiv \frac{\partial G}{\partial \nu(\boldsymbol{y}_j^{(l)})}(\boldsymbol{y}_j^{(l)}, \boldsymbol{x}_k). \tag{B.26}$$

We note that the right-hand side of equations (B.25) and (B.26), with $l \in \{1,2\}$ and $\{j,k\} \in \tau_S^{(l)}$, are functions of $\boldsymbol{y}_j^{(l)} - \boldsymbol{x}_k = \boldsymbol{y}_j^{(l)} - \boldsymbol{y}_k^{(l)} - \boldsymbol{d}^{(l)} = \boldsymbol{y}_{j-k}^{(l)} - \boldsymbol{d}^{(l)}$, and therefore, the right-hand side of equation (B.24) can be regarded as a two-dimensional convolution defined on the equi-spaced mesh $\tau_S^{(l)}$. The following lemma indicates how (B.24) can be computed fast with a two-dimensional FFT.

Lemma B.3.1. *Given the nonperiodic discrete values ζ_j and s_l, with $j \in \{0,\ldots,N-1\}$ and $l \in \{-(N-1),\ldots,0,\ldots,N-1\}$, the convolution*

$$v_k \equiv \sum_{j=0}^{N-1} s_{k-j}\zeta_j, \quad k \in \{0,\ldots,N-1\} \tag{B.27}$$

can be evaluated by Fourier transforms as

$$\tilde{v}_k = \mathcal{F}^{-1}\{\mathcal{F}\{\tilde{s}\} \cdot \mathcal{F}\{\tilde{\zeta}\}\}_k, \quad k \in \{0,\ldots,\tilde{N}-1\}, \tag{B.28}$$

where $\tilde{N} \equiv 2N-1$, $\mathcal{F}\{\tilde{v}\}_k \equiv \sum_{m=0}^{\tilde{N}-1} \tilde{v}_m e^{2\pi i k m/\tilde{N}}$, for $j \in \{0,\ldots,\tilde{N}-1\}$ and

$$\tilde{\zeta}_j \equiv \begin{cases} \zeta_j, & \text{if } j \in \{0,\ldots,N-1\} \\ 0, & \text{else} \end{cases} \tag{B.29}$$

$$\tilde{s}_j \equiv \begin{cases} s_j, & \text{if } j \in \{-(N-1),\ldots,0,\ldots,N-1\} \\ s_{j-\tilde{N}}, & \text{if } j \in \{N,\ldots,\tilde{N}-1\} \\ s_{j+\tilde{N}}, & \text{if } j \in \{-(\tilde{N}-1),\ldots,-N\} \end{cases} \tag{B.30}$$

$$\tilde{v}_k \equiv \begin{cases} v_k, & \text{if } k \in \{0, \ldots, N-1\} \\ \sum_{j=0}^{\tilde{N}-1} \tilde{s}_{k-j}\tilde{\zeta}_j, & else \end{cases} \tag{B.31}$$

for $k \in \{N, \ldots, \tilde{N}-1\}$.

Proof. Using the definition of the Fourier transform and (B.27), we have

$$\mathcal{F}\{v\}_k = \sum_{m=0}^{N-1} v_m e^{\frac{2\pi i}{N}km} = \sum_{m=0}^{N-1}\sum_{j=0}^{N-1} s_{m-j}\zeta_j e^{\frac{2\pi i}{N}km}.$$

Exchanging the sums and substituting for $m = n + j$ gives

$$\mathcal{F}\{v\}_k \quad = \quad \sum_{j=0}^{N-1}\sum_{n=-j}^{N-1-j} s_n\zeta_j e^{\frac{2\pi i}{N}k(n+j)}. \tag{B.32}$$

For a fixed j in (B.32), we split the inner sum into two parts:

$$\sum_{n=-j}^{N-1-j} s_n\zeta_j e^{\frac{2\pi i}{N}k(n+j)} \quad = \quad \sum_{n=-j}^{-1} \zeta_j s_n e^{\frac{2\pi i}{N}k(n+j)} + \sum_{n=0}^{N-1-j} \zeta_j s_n e^{\frac{2\pi i}{N}k(n+j)}. \tag{B.33}$$

We set $m = N + n$, and introduce this into the first sum of (B.33):

$$\sum_{n=-j}^{-1} \zeta_j s_n e^{\frac{2\pi i}{N}k(n+j)} = \sum_{m=N-j}^{N-1} \zeta_j s_{m-N} e^{\frac{2\pi i}{N}k(m-N+j)} = \sum_{m=N-j}^{N-1} \zeta_j s_{m-N} e^{\frac{2\pi i}{N}(m+j)}.$$

Therefore, equation (B.33) can be rewritten as

$$\sum_{n=-j}^{N-1-j} s_n \zeta_j e^{\frac{2\pi i}{N}k(n+j)} = \sum_{n=0}^{N-1-j} \zeta_j s_n e^{\frac{2\pi i}{N}k(n+j)} + \sum_{n=N-j}^{N-1} \zeta_j s_{n-N} e^{\frac{2\pi i}{N}(n+j)}. \tag{B.34}$$

If the values of s_n are $N-$periodic, i.e., $s_n = s_{n-N}$ for $n \in \{0, \ldots, N-1\}$ which is depicted in Figure B.7, equation (B.34) simplifies to $\sum_{n=0}^{N-1} \zeta_j s_n e^{\frac{2\pi i}{N}k(n+j)}$ and we have

$$\mathcal{F}\{v\}_k = \sum_{j=0}^{N-1}\sum_{n=0}^{N-1} \zeta_j s_n e^{\frac{2\pi i}{N}k(n+j)} = \sum_{j=0}^{N-1} \zeta_j e^{\frac{2\pi i}{N}kj} \cdot \sum_{n=0}^{N-1} s_n e^{\frac{2\pi i}{N}kn} = \mathcal{F}\{\zeta\}_k \cdot \mathcal{F}\{s\}_k. \tag{B.35}$$

For nonperiodic values, i.e., $s_n \neq s_{n-N}$, we define the extended values $\tilde{s}_j$ and $\tilde{\zeta}_j$ for $j \in \{0, \ldots, \tilde{N} =$

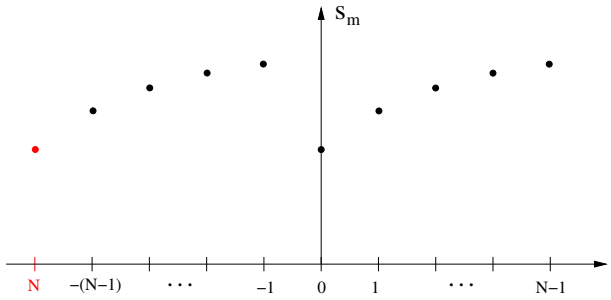


Figure B.7: $N-$periodic discrete values s_n

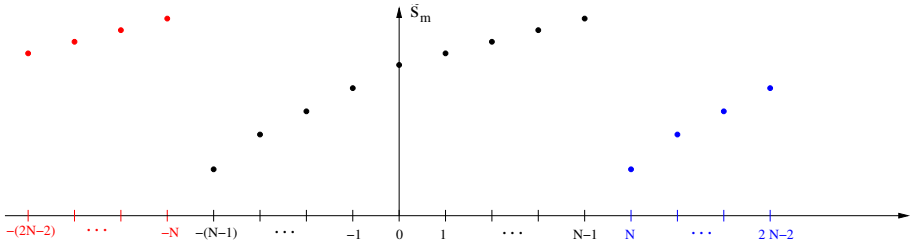


Figure B.8: Extended nonperiodic values s_n to $\tilde{N}$-periodic values $\tilde{s}_n$

$2N-1\}$ as in (B.30) and (B.29), respectively (compare also with Figure B.8). By construction, we have

$$v_k = \sum_{j=0}^{\tilde{N}-1} \tilde{s}_{k-j}\tilde{\zeta}_j, \quad k \in \{0,\dots,N-1\}. \tag{B.36}$$

Thus, the definition $\tilde{v}_k \equiv \sum_{j=0}^{\tilde{N}-1} \tilde{s}_{k-j}\tilde{\zeta}_j$ is equal to v_k for the indices $k \in \{0,\dots,N-1\}$. Since $\tilde{s}_k$ is $\tilde{N}$-periodic, equations (B.32)–(B.35) apply for the $\tilde{N}$ discrete values. We conclude that the Fourier transform $\mathcal{F}\{\tilde{v}\}_k$ can be computed by a multiplication of $\mathcal{F}\{\tilde{s}\}_k$ with $\mathcal{F}\{\tilde{\zeta}\}_k$, and the lemma follows by performing the inverse Fourier transform on $\mathcal{F}\{\tilde{v}\}_k$ and (B.36).

□

Remark B.3.1. *Lemma B.3.1 is also valid for multi-indices, and thus it can be applied to equation (B.24), with s_k substituted for m_k or d_k, and ζ_j replaced by ξ_j or η_j.*

B.4 Evaluation of the field on finer meshes than $\tau_S^{(3)}$

We consider the geometry in Figure B.6. In the last section we described a fast and accurate algorithm for the evaluation of the field on the mesh $\tau_S^{(3)}$ of the plane $\mathcal{B}$. Here, our goal is to evaluate the field on a finer mesh $\tau_F^{(3)}$. Clearly, we could obtain the unknown values from a local interpolation, but this may be below our expected accuracy if the spacing of $\tau_S^{(3)}$ is too large. Instead, we can extend the ideas of the last section to finer grids by zero-padding. In the following, we give the details of this approach.

We define an equidistant Cartesian grid $\tau_F^{(l)}$ of mesh size Δ_F and length $\tilde{H}$ with $F \times F$ nodes in such a way that all points of the coarser mesh $\tau_S^{(l)}$ coincide with nodes belonging to $\tau_F^{(l)}$. We replace the three meshes $\tau_S^{(l)}$ on D_1, D_2, and $\mathcal{B}$ with the finer grids $\tau_F^{(l)}$ and assign to all nodes on D_l which do not coincide with the initial coarser mesh an equivalent source of strength zero. This procedure is also known as zero-padding and it results in defining a discrete source distribution $\xi_F^{(l)}(\boldsymbol{y}_j) \cup \eta_F^{(l)}(\boldsymbol{y}_j)$ on the finer grids $j \in \tau_F^{(l)}$ of the two faces D_l. Thus, we can extend the monopole distribution on $\tau_S^{(l)}$ by

$$\xi_F^{(l)}(\boldsymbol{y}_j) = \begin{cases} \xi_j^{(l)}, & \text{if } (j_1, j_2) \in \tau_S^{(l)} \\ 0, & \text{else,} \end{cases} \tag{B.37}$$

and the dipole distribution by

$$\eta_F^{(l)}(\boldsymbol{y}_j) = \begin{cases} \eta_j^{(l)}, & \text{if } (j_1, j_2) \in \tau_S^{(l)} \\ 0, & \text{else.} \end{cases} \tag{B.38}$$

It is obvious that the field $\psi_F(\boldsymbol{x}_k)$ on $\mathcal{B}$ defined by

$$\psi_F(\boldsymbol{x}_k) \;=\; \sum_{l=1}^{2} \sum_{j \in \tau_F^{(l)}} \{\tilde{m}_{k-j}^{(l)} \xi_F^{(l)}(\boldsymbol{y}_j) + \tilde{d}_{k-j}^{(l)} \eta_F^{(l)}(\boldsymbol{y}_j)\}, \quad k \in \tau_F^{(3)}, \tag{B.39}$$

is identical to (B.23) for a coarse grid point $\boldsymbol{x}_k$ with $k \in \tau_S^{(3)}$. The values $\tilde{m}_{k-j}^{(l)}$ and $\tilde{d}_{k-j}^{(l)}$ are the natural extensions of (B.25) and (B.26) to the finer grids $\{k, j\} \in \tau_F^{(l)}$. Using (B.37) and (B.38), we can simplify (B.39) to

$$\psi_F(\boldsymbol{x}_k) \;=\; \sum_{l=1}^{2} \sum_{j \in \tau_S^{(l)}} \{m_{k-j}^{(l)} \xi_j^{(l)} + d_{k-j}^{(l)} \eta_j^{(l)}\}, \quad k \in \tau_F^{(3)}. \tag{B.40}$$

Clearly, for computational purposes we prefer (B.39) over (B.40), as (B.39) can be evaluated with Lemma B.3.1 fast by two-dimensional FFTs on the meshes $\tau_F^{(l)}$, while it is less obvious how to achieve this with (B.40). Nevertheless, equation (B.40) reveals that (B.39) indeed evaluates the field values correctly, once the equivalent sources $\xi_j^{(l)}$ and $\eta_j^{(l)}$ are known. In a more general case, there are n_l layers where equivalent sources are defined. Equation (B.39) includes $2n_l$ matrix-vector multiplications of size F^4, and the cost to compute ψ_F directly is therefore an $\mathcal{O}(2n_l F^4)$ process. In contrast, computing the convolution in (B.39) with Lemma B.3.1 involves $4n_l$ two-dimensional FFTs of the size $(2F-1)^2$, one inverse FFT of the same size and $2n_l(2F-1)^2$ multiplications, which is in summary an order $\mathcal{O}\left((2F-1)^2\{2n_l + (4n_l+1)\log_2(2F-1)^2\}\right)$ operation.

B.5 Implementation details of formula (B.39)

Lemma B.3.1 can be applied to (B.39) as follows: we identify $\tilde{s}_{k_1,k_2}$ with $\tilde{m}^{(l)}_{k_1,k_2}$ and $\tilde{d}^{(l)}_{k_1,k_2}$ for $\{k_1,k_2\} \in \{-(2F-1),\dots,0,\dots,2F-1\}$. These values are functions of $\boldsymbol{x}_{j_1,j_2} - \boldsymbol{y}^{(l)}_{k_1,k_2}$ for $j \in \tau^{(3)}_F$ and $k \in \tau^{(l)}_F, l \in \{1,2\}$ and can be built in the following way:

$$\tilde{s}^{(l)}_{k_1,k_2} = \begin{cases} f(\boldsymbol{x}_{k_1,k_2} - \boldsymbol{y}^{(l)}_{0,0}), & \text{if } \{k_1,k_2\} \in \tau^{(l)}_F = \{0,\dots,F-1\} \\ f(\boldsymbol{x}_{0,0} - \boldsymbol{y}^{(l)}_{-k_1,-k_2}), & \text{if } \{k_1,k_2\} \in \{-1,\dots,-(F-1)\} \\ f(\boldsymbol{x}_{k_1,0} - \boldsymbol{y}^{(l)}_{0,-k_2}), & \text{if } \{k_1,-k_2\} \in \{1,\dots,F-1\} \\ f(\boldsymbol{x}_{0,k_2} - \boldsymbol{y}^{(l)}_{-k_1,0}), & \text{if } \{-k_1,k_2\} \in \{1,\dots,F-1\}\ , \end{cases} \tag{B.41}$$

where f is the Green's function or its normal derivative:

$$f(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} G(\boldsymbol{x},\boldsymbol{y}), & \text{if } s_k = m_k \\ \frac{\partial G}{\partial \nu(\boldsymbol{y})}(\boldsymbol{x},\boldsymbol{y}), & \text{if } s_k = d_k. \end{cases} \tag{B.42}$$

In practice, we prefer not to deal with negative indices, which can be achieved by shifting these indices by $+(2F-1)$. This results in $2F-1$ periodic values $\tilde{s}_k$ defined for $\{k_1,k_2\} \in \{0,\dots,2F-2\}$.

We summarize the fast evaluation on the grid $\tau^{(3)}_F$ in

Algorithm B.5.1.

1. Given the equivalent sources $\xi^{(l)}_j$ and $\eta^{(l)}_j$ on the meshes $\tau^{(l)}_S$, extend their definition on $\tau^{(l)}_F$ by zero-padding (see equations (B.37) and (B.38)).
2. Build $\tilde{m}^{(l)}_{k_1,k_2}$ and $\tilde{d}^{(l)}_{k_1,k_2}$ according to (B.41).
3. Evaluate the two-dimensional FFTs of $\xi^{(l)}_F, \eta^{(l)}_F, \tilde{m}^{(l)}_{k_1,k_2}$, and $\tilde{d}^{(l)}_{k_1,k_2}$.
4. Multiply the corresponding expressions term by term in the Fourier space.
5. Apply the inverse FFT (B.28) to get the field values on the grid points of $\tau^{(3)}_F$.

Figures B.9 and B.10 display the real and imaginary parts of the functions $\tilde{m}_k^{(1)}, \tilde{d}_k^{(1)}$ and $\tilde{m}_k^{(2)}, \tilde{d}_k^{(2)}$, respectively. The centers of the two faces D_1 and D_2 are located at $[0, -H/2, 0]$ and $[0, H/2, 0]$. The parameters for this example are $F = 5, H = 0.0625, k = 10$, and $d^{(2)} = 4.5 \cdot H$. Figure B.11 gives the contour plots of $\tilde{m}_k^{(1)}$ and $\tilde{d}_k^{(1)}$ for the refined meshes with $F = 9$ and $F = 17$ nodes per side length.

Table B.1 shows the computing times in seconds on a Pentium III (Cascades) CPU to evaluate the field on $\tau_F^{(3)}$ directly by the convolution (CONV) versus the CPU-time required to compute the same values with the fast Algorithm B.5.1, denoted by CFFT. Note that the elapsed times for the CFFT algorithm shown in the table include the build-up time of formulas (B.41) and (B.42), as well as all necessary FFTs, which means that the portion of the purely sampling algorithm in CFFT is even smaller.

S	C	F	CONV	CFFT
11	15	41	1	0
		81	2	1
		161	8	1
		321	34	3
		641	135	14
13	17	49	1	0
		97	4	0
		193	16	0
		385	67	5
		769	270	28
15	19	57	2	0
		113	8	0
		225	31	1
		449	123	6
		897	491	41

Table B.1: CPU-times of CONV and CFFT in sec

B.6 Evaluation on a large surface $\mathcal{B}$

Let us assume the field values need to be computed on a plane $\mathcal{B}$ which is significantly larger than the discs D_l. In such a case, the meshes of the two-faces $\tau_F^{(1)}$ and $\tau_F^{(2)}$ can be appropriately increased to the large grid size on $\mathcal{B}$ and equivalent sources of strength zero assigned to the new nodes. However, this may be not efficient with respect to storage and computing time. Instead, it might be advantageous to split the original large sampling mesh on $\mathcal{B}$ into meshes $\tau_F^{(3,j)}$ for $j \in \{1, \ldots, J\}$ that are of the same size as $\tau_F^{(3)}$ and which may or may not overlap—see Figure B.12 for an example. On each of these meshes

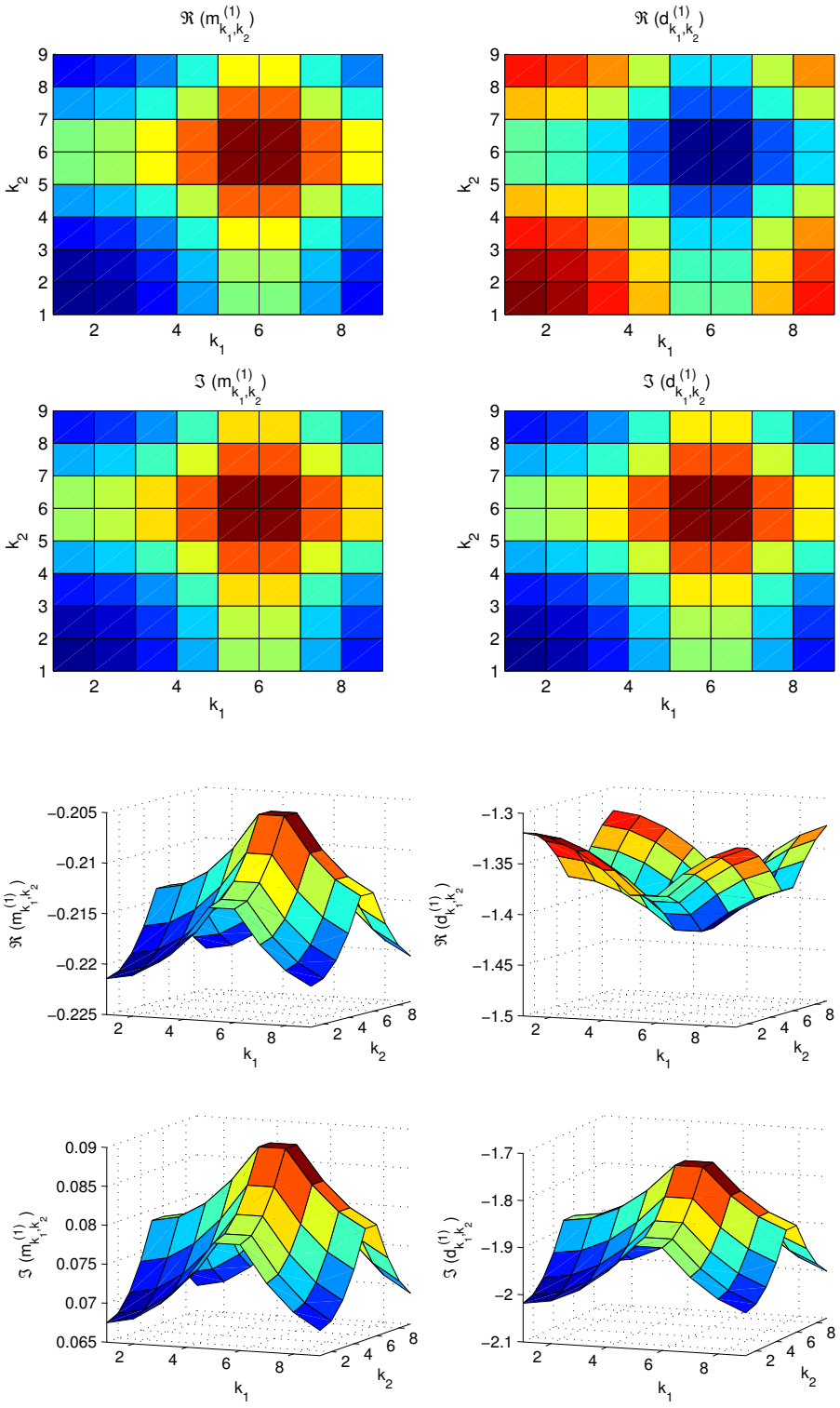


Figure B.9: Real and imaginary parts of $\tilde{m}_k^{(1)}$ and $\tilde{d}_k^{(1)}$ for $F = 5, H = 0.0625, k = 10$

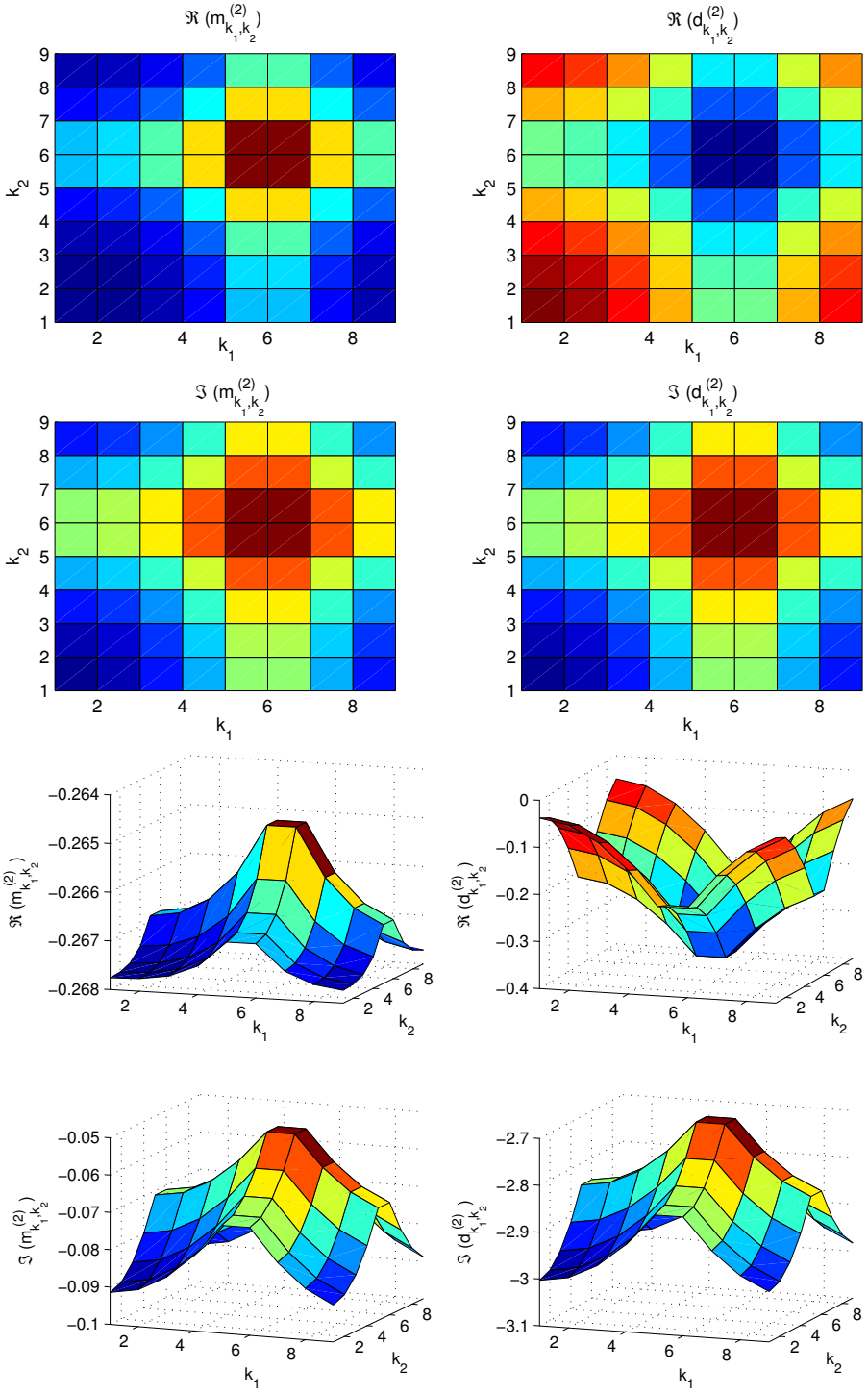


Figure B.10: Real and imaginary parts of $\tilde{m}_k^{(2)}$ and $\tilde{d}_k^{(2)}$ for $F = 5, H = 0.0625, k = 10$

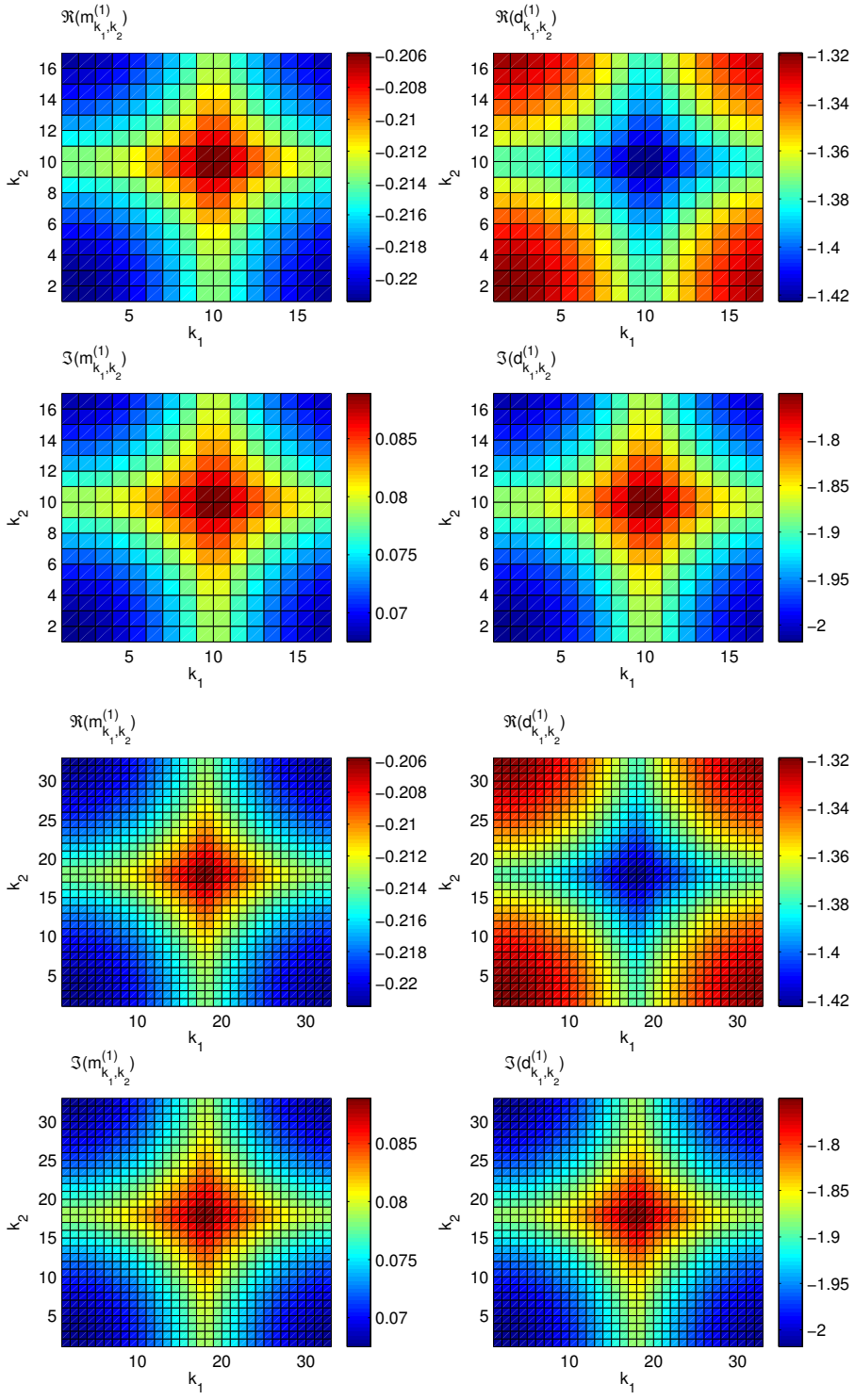


Figure B.11: Real and imaginary parts of $\tilde{m}_k^{(1)}$ and $\tilde{d}_k^{(1)}$ for $F = 9$ and $F = 17$, respectively

the fast evaluation algorithm B.5.1 can be applied. This limits the size of the Fourier transforms to $(2F-1)\times(2F-1)$ and can save a significant amount of storage if the sampling on $\mathcal{B}$ needs to be performed over a large area on an extremely fine mesh. Figure B.12 illustrates splitting the evaluation on $\mathcal{B}$ into $J=9$ smaller computations. In the Figures B.13 and B.14, we plot the real and imaginary parts of $\tilde{m}_k^{(1)}$ and $\tilde{d}_k^{(1)}$ for the meshes $\tau_F^{(3,7)}$ and $\tau_F^{(3,8)}$, respectively.

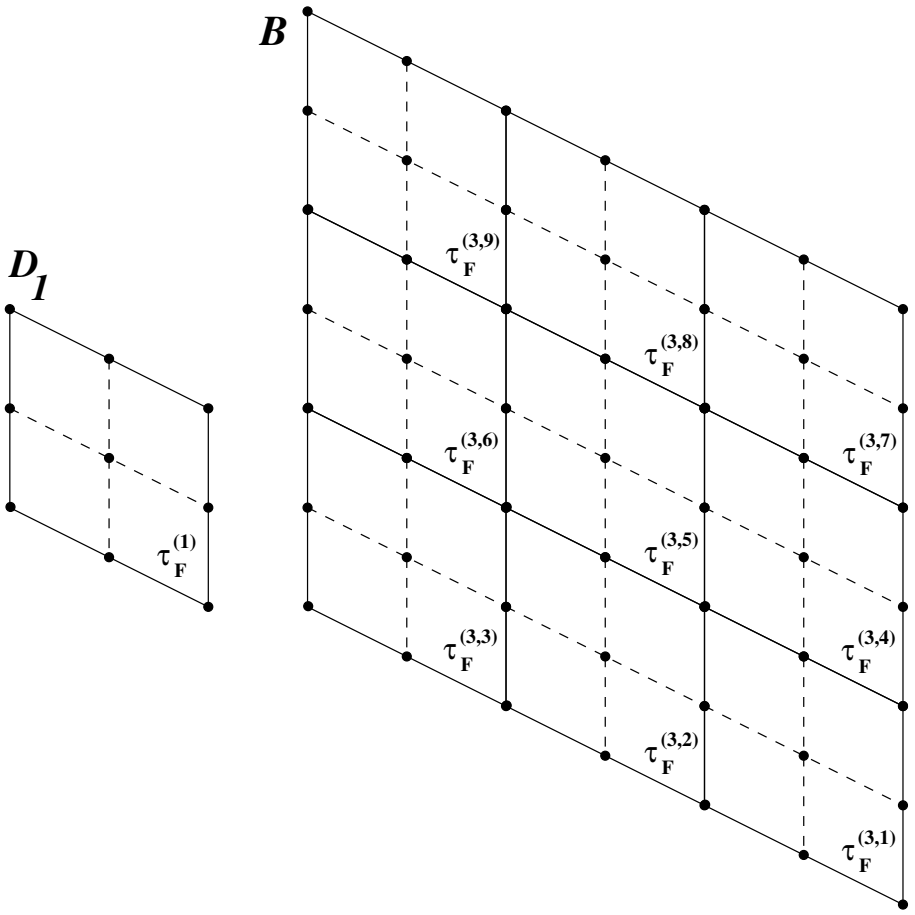


Figure B.12: Splitting the field evaluation on $\mathcal{B}$ into 9 smaller FFT computations on the meshes $\tau_F^{(3,j)}$ for $j=1,\ldots,9$

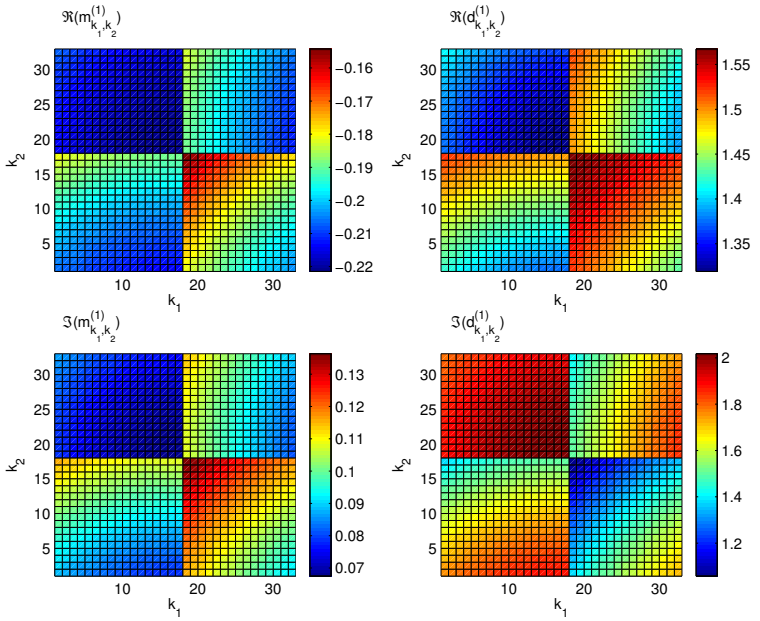


Figure B.13: Real and imaginary parts of $\tilde{m}_k^{(1)}$ and $\tilde{d}_k^{(1)}$ for $F = 17, H = 0.0625$ on $\tau_F^{(3,7)}$

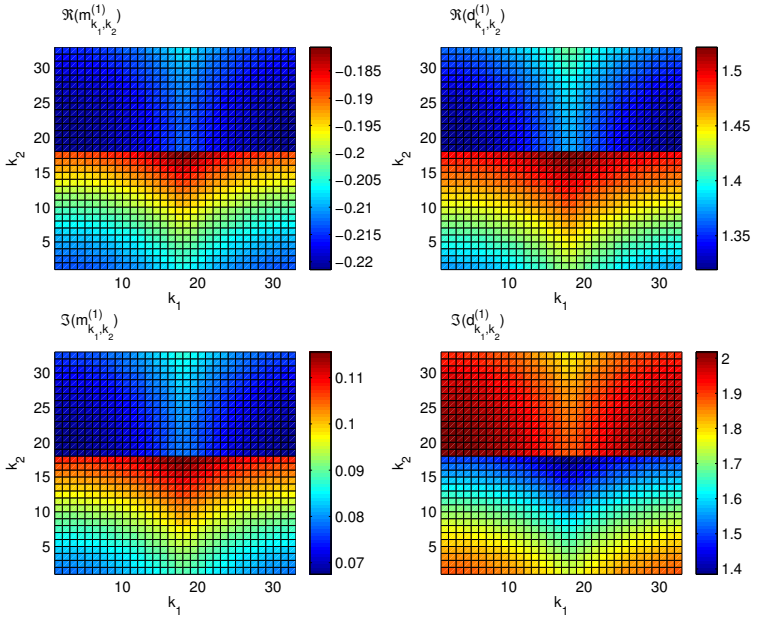


Figure B.14: Real and imaginary parts of $\tilde{m}_k^{(1)}$ and $\tilde{d}_k^{(1)}$ for $F = 17, H = 0.0625$ on $\tau_F^{(3,8)}$

Appendix C

Periodic extension based on Chebyshev approximation

Unlike the Fourier transformed variables of a function, the series coefficients produced by the continuation method cannot be bounded by the maximum absolute value of the function approximated, see [17]. As a result, the continuation coefficients can be quite large, which may lead to large function values in the extended domain, as demonstrated in Figure 3.17. While this does not necessarily seem to be a significant disadvantage for our application, it might be an asset to bound the continuation function in the extended domain by a small constant. As known, the Chebyshev coefficients are nicely bounded in terms of the maximum value of the function approximated. This observation leads us to explore the Chebyshev polynomials $T_n(t)$, with the goal to extend the domain of definition for the Chebyshev series into a larger periodic domain.

We recall some important properties of the polynomials $T_n(t)$. For $t \in [-1, 1]$, the Chebyshev polynomial of degree n is defined (see, e.g. [11] or [39]) by

$$T_n(t) = \cos(n\ \Theta(t)), \tag{C.1}$$

$$\Theta(t) = \arccos(t). \tag{C.2}$$

It is easy to see that $T_0(t) = 1$ and $T_1(t) = t$. The trigonometric identity $\cos(n\ x) = 2\cos(x)\cos((n -$

1) $x) - \cos((n-2)\ x)$ immediately gives the recurrence relation

$$T_n(t) \quad = \quad 2tT_{n-1}(t) - T_{n-2}(t),\ n \geq 2. \tag{C.3}$$

Formula (C.3) along with the initial conditions reveal that T_n is indeed a polynomial of order n. The $T_n(t)$ has n zeros in the interval $[-1, 1]$ which are located at the points

$$t \quad = \quad \cos\left(\frac{\pi(k+\frac{1}{2})}{n}\right), \quad k = 0, 1, \ldots, n-1, \tag{C.4}$$

also known as the Chebyshev points. Also, in the same interval, the Chebyshev polynomial exhibits $n+1$ extrema, which take the values 1 if maximum and -1 if minimum, located at $\cos(\pi k/n)$ for $k = 0, 1, \ldots, n$. Since all the n zeros of $T_n(t)$ are in $[-1, 1]$, we expect the Chebyshev polynomials to grow rapidly to infinity once outside of that interval if $n > 1$. This means that the Chebyshev series, initially defined in $[-1, 1]$,

$$v(t) \quad = \quad \sum_{k=0}^{N-1} c_k T_k(t) - \frac{1}{2}c_0, \tag{C.5}$$

would exhibit a rapid growth to large values if the domain of definition of the Chebyshev polynomials were to be extended outside $[-1, 1]$. Besides, such an extension would not necessarily lead to a periodic function. Both difficulties can be overcome by defining for each mode an appropriate window function $w_k(t) = w(t; -a_k, -1, 1, a_k)$ which is one in $[-1, 1]$ and smoothly decays to zero in $[1, a_k]$ and $[-a_k, -1]$ for an appropriate value $a_k > 1$. A possible definition for the window is the C^∞ function $w(t; t_0, t_1, t_2, t_3) = v(-t; -t_1, -t_0) \cdot v(t; t_2, t_3)$, where

$$v(t; t_0, t_1) = \begin{cases} 1, & \text{if } t \leq t_0\ , \\ 0, & \text{if } t \geq t_1\ , \\ \exp\left(2\frac{t_1-t_0}{t-t_1}\exp(-\frac{t_1-t_0}{t-t_0})\right), & \text{else,} \end{cases} \tag{C.6}$$

see [16]. The expansion

$$\tilde{v}(\tilde{t}) \quad = \quad \sum_{k=0}^{N-1} c_k T_k(\tilde{t}) w_k(\tilde{t}) - \frac{1}{2} c_0 w_0(\tilde{t}), \tag{C.7}$$

is the natural extension of (C.5) to the domain $\tilde{t} \in [-\tilde{T}_e/2, \tilde{T}_e/2]$, where $\tilde{T}_e$ is the period of the extended domain. By definition, we have $\tilde{v}(t) = v(t)$ for $t \in [-1, 1]$. The original function does not need to be confined to $[-1, 1]$. In fact, the transformation

$$x(t) \quad = \quad (t_b - t_a)/2 \cdot t + (t_a + t_b)/2 \tag{C.8}$$

maps the reference interval into any domain $[t_a, t_b]$. We illustrate the accuracy of the method on two concrete examples.

We consider the function in Figure C.1 on the left defined in the interval $[2.8, 3.0]$. Our goal is to construct an expansion (C.7) which is periodic in the interval $[2.6, 3.2]$, i.e., $\tilde{T}_e = 6$, and approximates the given green function in Figure C.1 to high order. We set $N = 35$ and compute the Chebyshev points

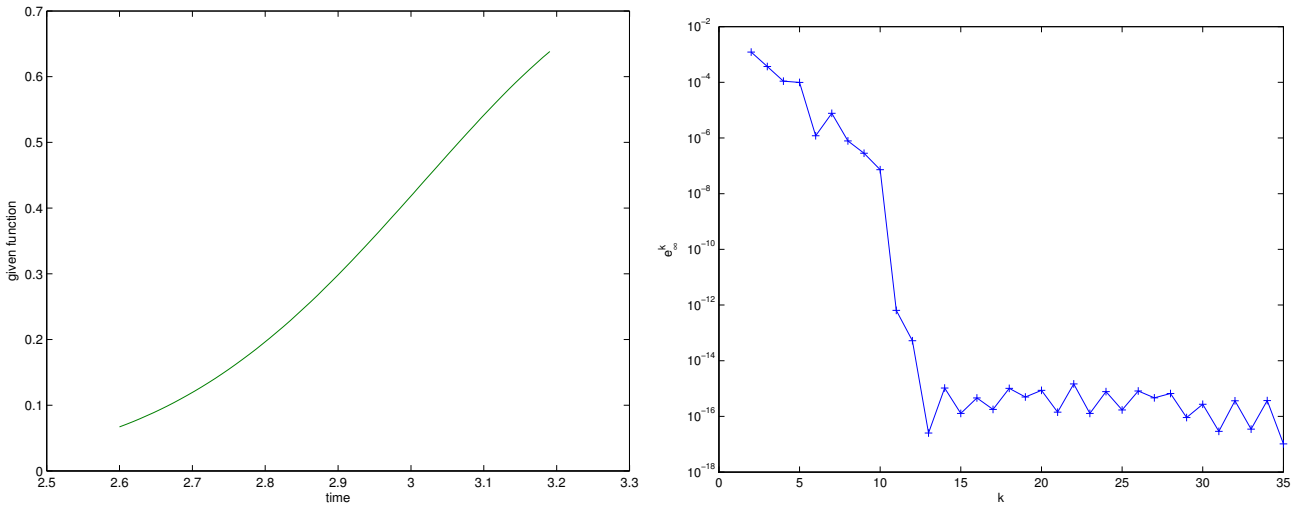


Figure C.1: Left: The function we wish to approximate to high accuracy in $[2.8, 3.0]$.
Right: The error e_∞^k is a measure if a_k has been chosen appropriately.

in the reference interval. Under the assumption that the function values $v(t)$ at these points are known,

we evaluate the Chebyshev coefficients

$$c_l = \frac{2}{N}\sum_{k=0}^{N-1} v\left(\cos\left(\frac{\pi(k+\frac{1}{2})}{N}\right)\right)\cos\left(\frac{\pi l(k+\frac{1}{2})}{N}\right), \quad l=0,\ldots,N-1. \tag{C.9}$$

The subtlety lies in selecting an appropriate a_k. Here, we choose $w_0 = 1$ and

$$a_k = \begin{cases} 3, & \text{if } k \in \{1,\ldots,15\}\ , \\ 1.3, & \text{if } k \in \{16,\ldots,35\}\ . \end{cases} \tag{C.10}$$

On the right of Figure C.1, we compare the discrete values $c_k T_k(t_l) w_k(t_l)$ at M equidistant points in $[-3,3]$ with its zero padded function. More precisely, we apply a FFT to the values of $c_k T_k(t_l) w_k(t_l)$ at $M/2$ equidistant points in $[-3,3]$, and, after zero padding to M frequency coefficients, transform the data back into the physical domain. We plot for every Chebyshev mode k the maximum absolute error

$$e_\infty^k \equiv \max_{l\in\{0,\ldots,M-1\}} |c_k T_k(t_l) w_k(t_l) - \mathcal{F}_M^{-1}\{\mathcal{F}_{M/2}\{c_k T_k w_k\}\}(t_l)| \tag{C.11}$$

on a logarithmic scale. A too large error in (C.11) for some k may indicate that the a_k is not chosen

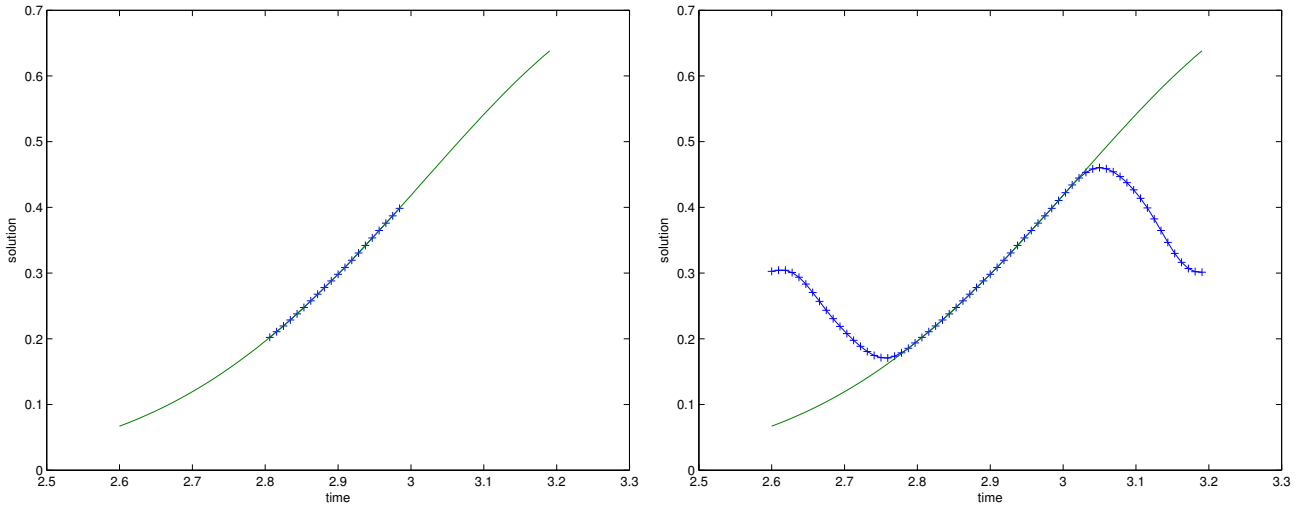


Figure C.2: Left: Continuation function (blue crosses) in the initial domain $[2.8, 3.0]$. Right: Continuation function (blue crosses) in the extended domain $[2.6, 3.2]$

appropriately. Indeed, a slow-to-zero decaying window function $w_k(t)$ can cause the term $c_k T_k w_k$ to rise outside the reference interval rapidly to extremely large values before falling to zero. On the other hand, a steep decay to zero of w_k can introduce too high frequencies in $c_k T_k w_k$. In both cases, the term $c_k T_k w_k$ would be hard to resolve properly. The result on the right side of Figure C.1 suggests that the choice in (C.10) is acceptable. The function (C.7) is plotted in Figure C.2 in both the original and extended domain. The values outside of $[2.8, 3.0]$ are nicely bounded as expected, and, inside the interval, the error to the original function is small: $e_\infty \approx 8.0 \cdot 10^{-15}, \partial_t e_\infty \approx 4.4 \cdot 10^{-11}, \partial_{tt} e_\infty \approx 4.0 \cdot 10^{-7}$.

In the second example, we want to find the Chebyshev continuation of the function plotted in Figure C.3 on the left. It should match with the original function in the domain $[2.8, 3.0]$, but this time we choose $\tilde{T}_e = 4$, i.e., the extended periodic function is supposed to be defined in $[2.7, 3.1]$. We select $w_0 = 1$ and

$$a_k = \begin{cases} 2, & \text{if } k \in \{1, \ldots, 27\} \ , \\ 1.3, & \text{if } k \in \{28, \ldots, 35\} \ . \end{cases} \tag{C.12}$$

The plot in Figure C.3 on the right suggests that this is a suitable choice. Figure C.4 again confirms that the periodic function is bounded by a small constant outside the original interval. Finally, the errors inside of $[2.8, 3.0]$ are $e_\infty \approx 1.5 \cdot 10^{-15}, \partial_t e_\infty \approx 2.6 \cdot 10^{-11}, \partial_{tt} e_\infty \approx 1.2 \cdot 10^{-7}$ demonstrating again the high accuracy of this approach.

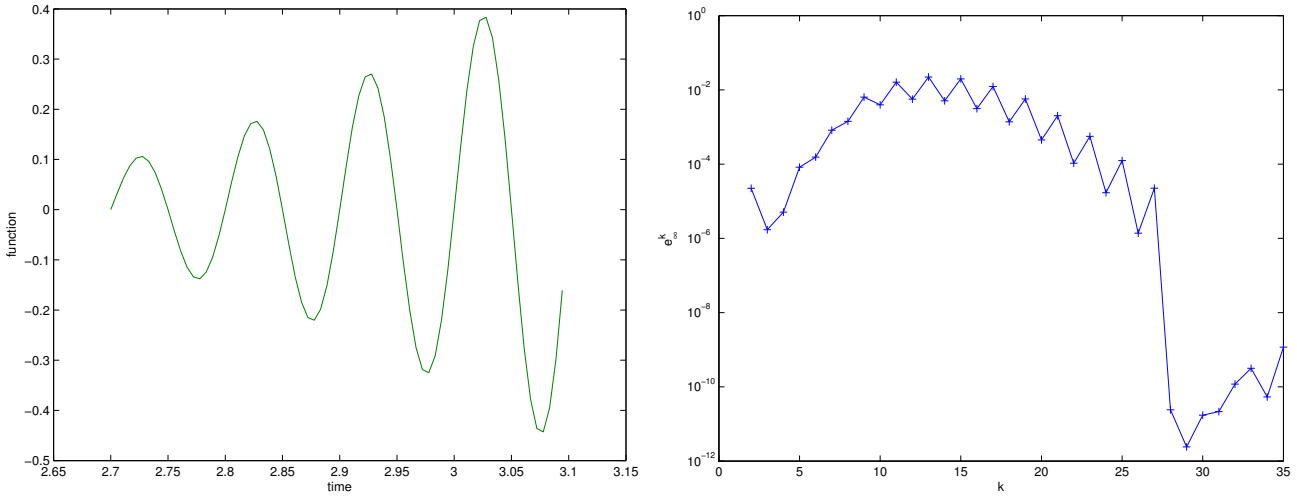


Figure C.3: Left: The function we wish to approximate. Right: The error e_∞^k

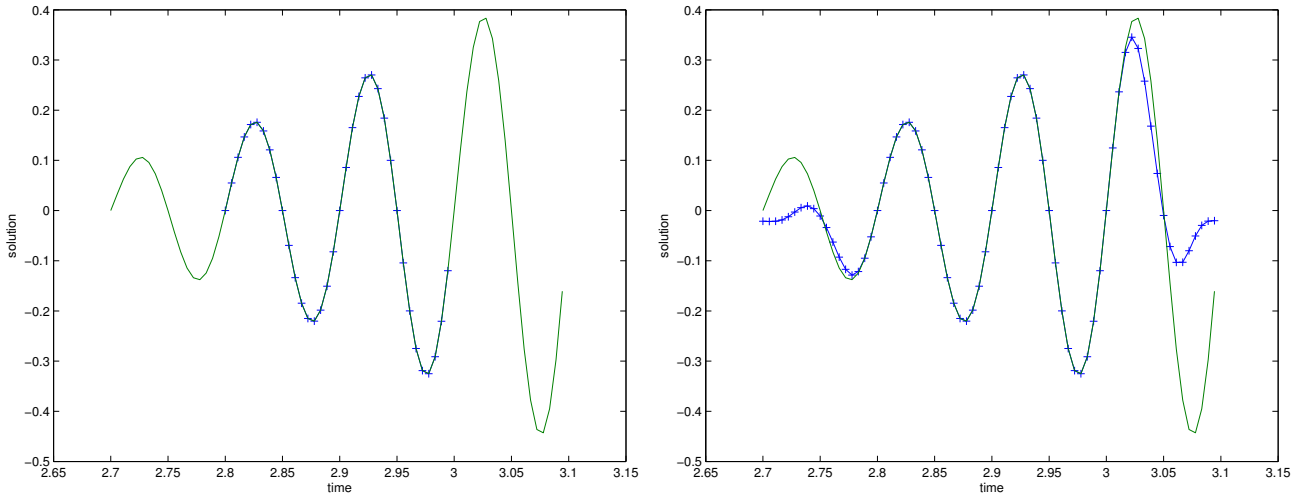


Figure C.4: Left: Continuation function (blue crosses) in $[2.8, 3.0]$. Right: Continuation function (blue crosses) in $[2.7, 3.1]$

Bibliography

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, U.S. Govt. Print. Off., Washington, District of Columbia, 1964

[2] J. D. Achenbach, *Wave Propagation in Elastic Solids*, Elsevier, New York, 1973

[3] B. Alpert, L. Greengard, and T. Hagstrom, *Rapid Evaluation of Nonreflecting Boundary Kernels for Time-Domain Wave Propagation*, *SIAM J. Numer. Anal.*, **37**, 1138–1164 (2000)

[4] B. Alpert, L. Greengard, and T. Hagstrom, *Nonreflecting Boundary Conditions for the Time-Dependent Wave Equation*, *J. Comput. Phys.*, **180**, 270–296 (2002)

[5] R. S. Anderssen and M. Hegland, *For numerical differentiation, dimensionality can be a blessing!*, *Math. Comp.*, **68**, 1121–1141 (1999)

[6] G. A. Baker, *Error Estimates for Finite Element Methods for the Second Order Hyperbolic Equation*, *SIAM J. Numer. Anal.*, **13**, 564–576 (1976)

[7] I. Babuska and B. Szabo, *Finite Element Analysis*, John Wiley & Sons, Inc., New York, 1991

[8] A. Bayliss and E. Turkel, *Radiation Boundary Condition for Wave-Like Equations*, *Commun. Pure Appl. Math.*, **33**, 707–725 (1980)

[9] J. P. Bérenger, *A perfectly matched layer for the absorption of electromagnetic waves*, *J. Comput. Phys.*, **114**, 185–200 (1994)

[10] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Pergamon Press, New York, 1980

[11] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, Dover Publications, Mineola, New York, 2001

[12] O. P. Bruno and L. A. Kunyansky, *Surface scattering in three dimensions: an accelerated high-order solver*, *Proc. R. Soc. Lond.*, **457**, 2921–2934 (2001)

[13] O. P. Bruno and L. A. Kunyansky, *A Fast, High-Order Algorithm for the Solution of Surface Scattering Problems I*, *J. Comp. Phys.*, **169**, 80–110 (2001)

[14] O. P. Bruno and L. A. Kunyansky, *A Fast, High-Order Algorithm for the Solution of Surface Scattering Problems II*, 2001

[15] O. P. Bruno and M. Lyon, *High-order unconditionally-stable FC ADI algorithms for general domains*, 2008

[16] O. P. Bruno and M. M. Pohlman, *A Continuation Method for the Resolution of the Gibbs Phenomenon*, 2004

[17] O. P. Bruno, Y. Han, and M. M. Pohlman, *Accurate, high-order representation of complex three-dimnensional surfaces via Fourier-Continuation analysis*, *J. Comp. Phys.*, **227**, 1094-1125 (2007)

[18] A. Björck, *Numerical Methods for Least Squares Problems*, Siam, Philadelphia, 1996

[19] E. O. Brigham, *The Fast Fourier Transform and its Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1988

[20] J. Cheng, Y. C. Hon, and Y. B. Wang, *A numerical method for the discontinuous solutions of Abel integral equations*, Inverse Problems and Spectral Theory (volume 348 of Contemp. Math., pp. 233–243), Am. Math. Soc., Providence, Rhode Island, 2004

[21] W. Chew and W. Weedon, *A 3-D perfectly matched medium from modified Maxwell's equations with stretched coordinates*, *Microwave Optical Technol. Letter*, **7**, 599–604 (1994)

[22] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, *J. Res. Nat. Bur. Stand.*, **49**, 409–436 (1952)

[23] G. C. Cohen, *Higher-Order Numerical Methods for Transient Wave Equations*, Springer, Berlin, New York, 2002

[24] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, Berlin, New York, 1992

[25] J. Cullum, *Numerical differentiation and regularization*, *SIAM J. Numer. Anal.*, **8**, 254–265 (1971)

[26] J. L. Davis, *Mathematics of wave propagation*, Princeton University Press, Princeton, 2000

[27] S. R. Deans, *The Radon Transform and Some of its Applications*, Wiley, New York, 1983

[28] T. Dupont, *L^2-Estimates for Galerkin Methods for Second-Order Hyperbolic Equations*, *SIAM J. Numer. Anal.*, **10**, 880-889 (1973)

[29] B. Engquist and A. Majda, *Absorbing Boundary Conditions for the Numerical Simulation of Waves*, *Math. Comp.*, **31**, 629–651 (1977)

[30] B. Engquist and A. Majda, *Radiation Boundary Conditions for Acoustic and Elastic Wave Calculation*, *Comm. Pure Appl. Math.*, **32**, 313–357 (1979)

[31] B. Engquist and L. Halpern, *Far Field Boundary-Conditions for Computation over Long-Time*, *Appl. Numer. Math.*, **4**, 21–45 (1988)

[32] K. Feng, *Finite Element Method and Natural Boundary Reduction*, Proceedings of the International Congress of Mathematicians, Warsaw, Poland, 1983

[33] P. Filippi et al., *Acoustic: basic physics, theory, and methods*, Academic Press, San Diego, 1999

[34] J. Fritz, *Plane Waves and Spherical Means applied to Partial Differential Equations*, Interscience Publishers, New York, 1955

[35] R. Gorenflo and S. Vessella, *Abel Integral Equations*, Springer-Verlag, Berlin, New York, 1991

[36] C. W. Groetsch, *Differentiation of approximately specified functions*, *Am. Math. Mon.*, **98**, 847–850 (1991)

[37] D. Givoli, *Numerical Methods for Problems in Infinite Domains*, *Studies in Applied Mechanics*, **33**, Elsevier (1992)

[38] D. Givoli and D. Cohen, *Nonreflecting Boundary Conditions Based on Kirchhoff-Type Formulae*, *J. Comp. Phys.*, **117**, 102–112 (1995)

[39] D. Gottlieb and S.A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 1977

[40] M. Grote and J. Keller, *Exact Nonreflecting Boundary Conditions For The Time Dependent Wave Equation*, *SIAM J. Appl. Math.*, **55**, 280–297 (1995)

[41] M. Grote and J. Keller, *Nonreflecting Boundary Conditions for Time-Dependent Scattering*, *J. Comput. Phys.*, **127**, 52–81 (1996)

[42] W. Bangerth, M. Grote, and K. Hohenegger, *Finite Element Method for Time-Dependent Scattering: Nonreflecting Boundary Condition, Adaptivity, and Energy Decay*, *Comput. Meth. Appl. Mech. Engin.*, **193**, 2453–1482 (2003)

[43] M. Grote and J. Keller, *Nonreflecting Boundary Conditions for Maxwell's equations*, *J. Comput. Phys.*, **139**, 327–342 (1998)

[44] M. Grote, *Nonreflecting Boundary Conditions for Elastodynamic Scattering*, Research Report No. 99-19, Seminar für Angewandte Mathematik, Eidgenössische Technische Hochschule Zürich, 1999

[45] M. Hanke and O. Scherzer, *Inverse problem light: numerical differentiation*, *Am. Math. Mon.*, **108**, 512–521 (2001)

[46] M. Hanke and O. Scherzer, *Error analysis of an equation error method for the identification of the diffusion coefficient in a quasi-linear parabolic differential equation*, *SIAM J. Appl. Math.*, **59**, 1012–1027 (1999)

[47] T. Hagstrom, *Radiation Boundary Condition for the Numerical Simulation of Waves*, *Acta Numerica*, **8**, 47–106 (1999)

[48] T. Hagstrom, *New Results on Absorbing Layers and Radiation Boundary Conditions*, Lecture Notes in Computational Science and Engineering (volume 31, pp.1-42), Springer-Verlag, New York, 2003

[49] T. Hagstrom and S. Lau, *Radiation Boundary Condition for Maxwell's Equation: A Review of Accurate Time-Domain Formulations*, *J. Comput. Math.*, **25**, 305–336 (2007)

[50] T. Hagstrom, *Private communication*, 2005

[51] S. He and V. Weston, *Wave-Splitting and Absorbing Boundary Conditions for Maxwell's Equations on a Curved Surface*, *Math. and Comput. in Simulation*, **50**, 435–455 (1999)

[52] R. L. Higdon, *Numerical Absorbing Boundary Conditions for the Wave Equation*, *Mathematics of Computation*, **49**, 65–90 (1987)

[53] D. Hoch, *A High Order Finite Element Method for Time Dependent Problems in Complex Geometry*, (diploma thesis, Advisor: M. J. Grote, Eidgenössische Technische Hochschule Zürich), 2001

[54] F. Ihlenburg, *Finite Element Analysis of Acoustic Scattering*, Springer, Berlin, New York, 1998

[55] S. Ivar, *Green's Functions and Boundary Value Problems*, Wiley, New York, 1979

[56] J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, 1975

[57] D. S. Jones *The Theory of Electrodynamics*, New York, Macmillan, 1964

[58] H. Jiang and Y.S. Wong, *Absorbing boundary conditions for second-order hyperbolic equations*, *J. Comput. Phys.*, **88**, 205–231 (1990)

[59] O. D. Kellogg, *Foundations of Potential Theory*, Springer, New York, 1929

[60] G. Kriegsmann, *Radiation Condition for Wave Guide Problems*, *SIAM J. Sci. Stat. Comput.,* **3**, 318-326 (1982)

[61] G. Kriegsmann and C. Morawetz, *Solving the Helmholtz Equation for Exterior Problems with Variable Index of Refraction: I*, *SIAM J. Sci. Stat. Comput.,* **1**, 371–385 (1980)

[62] E. L. Lindman, *Free-Spaced Boundary Conditions for the Time Dependent Wave Equation*, *J. Comput. Phys.*, **18**, 66–78 (1975)

[63] C. Lubich and A. Schädle, *Fast convolution for nonreflecting boundary conditions*, *SIAM J. Sci. Comput.*, **24**, 161–182 (2002)

[64] A. J. Majda and A. L. Bertozzi, *Vorticity and Incompressible Flow*, Cambridge University Press, New York, 2002

[65] D. Jiao, M. Lu, E. Michielssen, and J. Jin, *A Fast Time-Domain Finite Element-Boundary Integral Method for Electromagnetic Analysis*, *IEEE Trans. Antennas Propagat.*, **49**, 1453–1461 (2001)

[66] D. Jiao, A. Ergin, B. Shanker, E. Michielssen, and J. Jin, *A Fast Higher-Order Time-Domain Finite Element-Boundary Integral Method for 3-D Electromagnetic Scattering Analysis*, *IEEE Trans. Antennas Propagat.*, **50**, 1192–1202 (2002)

[67] B. Shanker, A. Ergin, M. Lu, and E. Michielssen, *Fast Analysis of Transient Electromagnetic Scattering Phenomena using the Multilevel Plane Wave Time Domain Algorithm*, *IEEE Trans. Antennas Propagat.*, **51**, 628–641 (2003)

[68] M. Lu, M. Lv, and A. Ergin, *Multilevel plane wave time domain-based global boundary kernels for two-dimensional finite difference time domain simulations*, *Radio Science*, **39**, 1–16 (2004)

[69] A. Yilmaz, Z. Lou, E. Michielssen, and J. Jin, *A Single-Boundary Implicit and FFT-Accelerated Time-Domain Finite Element-Boundary Integral Solver*, *IEEE Trans. Antennas Propagat.*, **55**, 1382–1397 (2007)

[70] P. D. Lax, C. S. Morawetz, and R. S. Phillips, *Exponential Decay of Solutions of the Wave Equation in the Exterior of a Star-Shaped Obstacle*, *Comm. Pure Appl. Math., Vol.*, **XVI**, 477–486 (1963)

[71] C. S. Morawetz, *Exponential Decay of Solutions of the Wave Equation*, *Comm. Pure Appl. Math., Vol.*, **XIX**, 439–444 (1966)

[72] C. S. Morawetz, *Notes on Time Decay and Scattering for some Hyperbolic Problems*, Society for Industrial and Appl. Math., Philadelphia, 1975

[73] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968

[74] F. Collino and P. Monk, *The perfectly matched layer in curvilinear coordinates*, *SIAM J. Sci. Comp.*, **19**, 2061–2090 (1998)

[75] G. Mur, *Absorbing Boundary Conditions for the Finite-Difference Approximation of the Time-Domain Electromagnetic-Field Equations*, *IEEE Trans. Electromagnetic Compat.*, **EMC-23**, 377–382 (1981)

[76] J. C. Nedelec, *Acoustic and Electromagnetic Equations*, Springer-Verlag, Berlin, New York, 2001

[77] P. Petropoulos, *Reflectionless sponge layers as absorbing boundary conditions for the numerical solution of Maxwell's equations in rectangular, cylindrical and spherical coordinates*, *SIAM J. Appl. Math.*, **60**, 1037–1058 (2000)

[78] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C,* Cambridge University Press, 1992

[79] A. G. Ramm, *Wave Scattering by Small Bodies of Arbitrary Shapes*, World Scientific, New Jersey, 2005

[80] A. G. Ramm and A. B. Smirnova, *On Stable Numerical Differentiation*, *Math. Comput.*, **70**, 1131–1153 (2001)

[81] C. Schwab, *p- and hp- Finite Element Methods*, Oxford University Press, 1998

[82] W. D. Smith, *A Nonreflecting Plane Boundary for Wave Propagation Problems*, *J. Comp. Phys.*, **15**, 492–503 (1974)

[83] A. Sommerfeld, *Lecture on Theoretical Physics*, Academic Press, 1964

[84] I. L. Sofronov, *Artificial Boundary Conditions of Absolute Transparency for two- and three-dimensional external time-dependent scattering problems*, *Eur. J. Appl. Math.*, **9**, 561–588 (1998)

[85] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-posed Problems*, (Translated from Russian, Preface by translation editor Fritz John Wiley, New York), Scripta Series in Mathematics, V. H. Winston & Sons, Washington, District of Columbia, 1977

[86] L. Ting and M. J. Miksis, *Exact boundary conditions for scattering problems*, *J. Acoust. Soc. Am.*, **80**, 1825 (1986)

[87] S. V. Tsynkov, *Numerical Solutions of Problems on Unbounded Domains. A Review*, *Appl. Num. Math.*, **27**, 465 (1998)

[88] V. S. Ryaben'kii, S. V. Tsynkov, and V. I. Turchaninov, *Global Discrete Artificial Boundary Conditions for Time-Dependent Wave Propagation*, *J. Comput. Phys.,* **174**, 712–758 (2001)

[89] S. V. Tsynkov, *On the Application of Lacunae-Based Methods to Maxwell's equations*, *J. Comput. Phys.,* **199**, 126–149 (2004)

[90] L. Tourrette and L. Halpern, *Absorbing Boundaries and Layers, Domain Decomposition Methods*, Nova Science Publishers, New York, 2001

[91] G. B. Whitham, *Linear And Nonlinear Waves*, J. Wiley & Sons, New York, 1974

[92] Y. B. Wang, X. Z Jia, and J. Cheng, *A numerical differentiation method and its application to reconstruction of discontinuity*, *Inverse Problems*, **18**, 1461–1476 (2002)

[93] T. Wei and Y. C. Hon, *Numerical differentiation by radial basis functions approximation*, *Adv. Comput. Math.*, **27**, 247–272 (2004)