COMPUTATIONAL METHODS FOR WAVE PROPAGATION PROBLEMS IN UNBOUNDED DOMAINS

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Vrushali Bokil

May 2003

COMPUTATIONAL METHODS FOR WAVE PROPAGATION PROBLEMS IN UNBOUNDED DOMAINS

Vrushali Bokil

APPROVED:

Dr. Roland Glowinski, Chairman

Dr. Edward J. Dean

Dr. Tsorng-Whay Pan

Dr. Michael Buksas,
Los Alamos National Laboratory

Dr. Kurt J. Marfurt,
Department of Geosciences

Dr. John L. Bear
Dean, College of Natural Sciences
and Mathematics

ACKNOWLEDGMENTS

I feel very fortunate to have worked with an advisor who is scientifically a leading figure and at the same time an exemplary teacher. I would like to express my utmost gratitude and heartfelt appreciation to my advisor Dr. Roland Glowinski, for his encouragement, enthusiasm, support, and guidance, his confidence in my abilities, and his friendship, which have proved invaluable to me over the last five years.

I have had the privilege of studying at the T7 group of Los Alamos National Laboratory during the last four summers under the guidance of Dr. Mac Hyman and Dr. Michael Buksas. I have greatly benefited from their guidance and scientific expertise. The contagious enthusiasm and encouragement of Dr. Mac Hyman and the strong and steady support of Dr. Michael Buksas have helped me tremendously. Many thanks to both of them for their time, efforts and their concern.

I have received many valuable observations, suggestions and comments from Dr. K. Lipnikov, Dr. T. W. Pan and Dr. J. Toivanen. I am greatly indebted to them for their time and effort. I thank my committee members Dr. M. Buksas, Dr. E. Dean, Dr. K. Marfurt, and Dr. T. W. Pan for taking time out from their busy schedule to be a part of my dissertation defense. A big thank you to Dr. M. Gehrke for her concern in my career, and to Dr. R. Bagby, Dr. B. Keyfitz, Dr. I. Swanson and Dr. P. Teller for their confidence in me.

I would like to thank the staff and members of the Department of Mathematics at the University of Houston for their wonderful help. My research has been made possible by funds from the Los Alamos Computer Science Institute (LACSI) and by DOE.

Finally, to my friends and family, my sisters, and my parents, who are a constant source of encouragement, I have much to thank. Without their unconditional love and support I would not have reached this far.

COMPUTATIONAL METHODS FOR WAVE PROPAGATION PROBLEMS IN UNBOUNDED DOMAINS.

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Vrushali Bokil

May 2003

ABSTRACT

In this dissertation I have proposed a novel fictitious domain method based on a distributed Lagrange multiplier for the solution of the time-dependent problem of scattering by an obstacle. I have introduced the fictitious domain method for the case of the two-dimensional scalar wave equation, as well as for the two-dimensional transverse magnetic (TM) mode of Maxwell's equations, along with a Dirichlet condition on the boundary of the obstacle in each case.

In the case of the two-dimensional scalar wave equation, I have presented a symmetrized operator splitting scheme to decouple the operator that propagates the wave and the operator that enforces the Dirichlet condition on the boundary of the obstacle. I have studied different discretizations for the different subproblems involved in the operator splitting scheme. These include conforming finite elements as well as mixed finite element formulations utilizing the lowest order Nédélec edge elements on rectangular grids. I have presented an analysis of the fictitious domain approach and the symmetrized operator splitting scheme for a one-dimensional wave problem. Comparisons are performed with other relevant numerical schemes, such as the finite difference scheme, that show the advantages of the formulation proposed in this thesis.

I have constructed a mixed finite element formulation for the two-dimensional TM mode of the uniaxial perfectly matched layer (PML) for Maxwell's equations. Energy estimates that demonstrate the well-posedness of the model are presented. I have employed a mixed discretization which utilizes the lowest order Raviart-Thomas elements and bilinear nodal finite elements on rectangular grids. I have performed a plane wave analysis to study the errors that arise due to dispersion, anisotropy, the numerical discretization as well as the termination of the PML by a perfect conductor condition. Finally, I have incorporated the fictitious domain approach into the mixed finite element model for the PML.

Numerical results that validate the effectiveness of the different models are presented in this dissertation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

An important aspect of problems in several physical areas such as electromagnetics, acoustics, elasticity, and seismology, is the ability to accurately simulate wave phenomena in bounded or unbounded media. The simulation of such physical systems involves the numerical solution of partial differential equations (PDE's) that describe the underlying physics. We can identify three physical categories for classifying equations that model wave propagation; the acoustics equations which model mechanical waves in fluids, the elastic systems, which model mechanical waves in solids, and the Maxwell equations, that model the propagation of electromagnetic waves. For most of these applications closed form solutions of the underlying PDE's either do not exist or are intractable. For example, solutions to the time-dependent Maxwell's equations in general form are known only for a few special cases. The difficulty lies in the imposition of boundary conditions. Thus, in many cases numerical approximation is the most convenient way to solve these PDE's. On the other hand, numerical simulation can play an important role in the design and understanding of complex systems.

In this thesis, we will study numerical models for problems involving the acoustics equations and Maxwell's equations in two dimensions. In particular, we are interested in the time dependent problem of scattering by an obstacle, and a related problem of con-

structing numerical models for wave propagation on unbounded domains.

Among alternative mathematical formulations of Maxwell's equations or the wave equation, integral equation (IE) formulations and PDE's lead to two very different computational approaches. In the following sections, we briefly review both these approaches. As PDE formulations and consequent time-domain numerical techniques will be the focus in this thesis, we will emphasize these methods and provide a selected bibliography of developments in this area as they pertain to this thesis.

1.1 Integral Equation Based Methods

With computational methods derived from integral equations, a three-dimensional boundary value problem reduces to a two-dimensional problem over the boundary of the domain of interest, as in the boundary element methods, or in the method of moments [72]. IE solvers involve a smaller number of unknowns than differential equation solvers, because only the induced sources, which exist in a space of smaller dimension, are unknowns. On the other hand, in a differential equation, the field is the unknown quantity. However, integral equation solvers result in dense matrices. Thus, even with a significant reduction in the number of unknowns, a full system matrix (impedance matrix) has to be solved. Integral techniques can be used only wherever analytical Green's functions are available. Thus, the applicability of integral equation based techniques is much more limited as compared to differential equation based methods. One of the most important advantages of IE methods is the treatment of open region problems; the infinite extent of the structure is already accounted for in the computation of the Green's functions, and there is no need for an artificial mesh truncation.

Current research based on special properties of the impedance matrix leading to *fast multipole methods* [32, 39, 117, 118] is presenting attractive alternatives to differential equation based methods.

1.2 Differential Equation Based Methods

In the PDE based approach, there are two types of problems that can be associated with models of wave propagation; problems in the time-domain and problems in the frequency domain. In the first case the solution is obtained via time integration. In the second case, a periodic dependence in time of the solution is imposed *a priori*. A wider range of applications can be analyzed with the time-domain approach. For example, time-domain approaches are better suited for the accurate and efficient simulations of electromagnetic waves through three-dimensional structures over a large range of frequencies. Frequency domain approaches can model harmonic sources, while pulse sources are more easily handled by time-domain methods. Among the many different approaches available for the numerical approximation of the problems, finite difference, finite volume, and finite element methods are some of the most popular schemes used today for the spatial/temporal discretization of the PDE's in question. A review of modeling techniques for electromagnetics can be found in [80, 108].

1.2.1 Finite Difference Methods

Finite difference methods have a long history going back to Leonhard Euler (1707 - 1783) who began to study the calculus of finite differences. Finite difference methods were the first numerical methods to be used to solve the wave equation. The first finite difference method, that was used for the wave equation, was based on centered second-order approximations of the second-order derivatives in time as well of the Laplace operator, leading to a fully explicit scheme [1]. Such an approximation has good stability and dispersion features needed for the wave equation, and is also a nondissipative scheme. The *Yee scheme* [134], which was named the *finite difference time-domain method* (FDTD) by Taflove [123], was among the first techniques used for the solution of the time-domain Maxwell's equations. The Yee algorithm is a fully explicit scheme and combines central differences on a stag-

gered grid in space with a second-order staggered leapfrog temporal method. The electric field $\mathbf{E}$ is evaluated about a unit cubic cell at the centers of edges, where as the magnetic field $\mathbf{H}$ is evaluated at the centers of the faces of the unit cubic cell. It is also a nondissipative scheme. The use of a regular Cartesian grid, well suited for wave propagation, with an explicit scheme in time also makes the finite difference method very efficient from the computational point of view. For a comprehensive bibliography, including various extensions of the Yee scheme see [80, 124].

The disadvantage of the finite difference scheme is that there is difficulty in extending it to more general domains. One technique used in such a case is to employ a *stair-cased* approximation to the irregular boundary. Thus, in solving time-dependent problems of scattering by an obstacle, the finite difference method creates numerical diffraction when the boundary of the obstacle does not fit the mesh grid. The stair-cased approximation to the boundary of the obstacle degrades the numerical solution unless a very fine resolution of grid points is used. Some issues of the staircase problem related to the FDTD method and possible solutions to this problem are studied in [28, 49, 78, 94, 95, 121, 132]. Finite element methods help alleviate this problem, by allowing the mesh to exactly follow the boundary of the scattering object.

As the Yee scheme is an explicit scheme a stability condition, called the Courant-Friedrich-Lewy (CFL) condition, has to be satisfied, which determines a relation between the time step and the spatial mesh step size. Thus, small mesh step sizes will lead to small time steps. Implicit schemes may alleviate this requirement on the time step, though such schemes require the solution of linear systems at each time step. Nédélec *et. al.* considered an implicit scheme for the time integration of Maxwell's equations with a finite element method in [4].

1.2.2 Finite Element Methods

The finite element method (FEM) was first described by Courant [43] in 1942. Usually, finite element methods are *variational techniques*, which optimize an expression that is known to be stationary about the true solution. Generally, FEM's solve for the unknown field quantities by minimizing an energy functional. A related approach is the class of Galerkin methods that are based on *weak* formulations of the PDE's. The use of Galerkin methods to derive FEM's leads to conservative and stable algorithms for most classes of problems in mathematical physics. One of the first applications of finite elements to electromagnetic problems was by Arlett *et. al.* [7]. However, most applications of finite elements to electromagnetics were carried out in the frequency domain. Research in the area of time-domain finite elements has recently gathered momentum. Finite element methods in the time-domain offer some important advantages over the standard finite difference method. The use of *unstructured grids* offers high versatility in the modeling of complex geometries. Field and flux continuity conditions at material interfaces can be handled by the variational approach in a natural way [86]. However, the numerical implementation of finite element methods is usually more difficult than that for finite difference methods. Also, computational efficiency is decreased by the unstructured nature of the data.

The finite element modeling of electromagnetic fields can be done by using nodal elements and/or the edge or face elements which are essentially due to Raviart-Thomas [116] in two-dimensions and Nédélec in three-dimensions [109]. Finite element methods for Maxwell's equations can be based on the first order Maxwell curl equations, or the second order curl-curl (wave) formulations of either the electric field or the magnetic field. Of course, this is also true of other numerical methods like the finite difference method. Other options are vector potential based methods or frequency domain methods based on the Helmholtz equations. The divergence conditions are usually assumed to be implicit in the curl equations, and hence are not incorporated into the numerical models; a practice that is believed to lead to non-physical solutions, called *spurious solutions* [81]. In [8], the authors

have used a constrained wave equation system (the second order curl-curl system) along with a Lagrange multiplier associated with the divergence condition to obtain a mixed finite element method. A FEM involving simultaneous approximation of two or more physical variables is called a mixed finite element method. Very often, a mixed formulation results in saddle point problems. In this case there are stability conditions [25, 11] that impose limitations on discrete finite element spaces. The edge (face) elements proposed by Nédélec are of the mixed type.

Functions of the edge type were first used in [93], though they achieved popularity only after the fundamental theoretical paper by Nédélec [109]. An initial application of these elements was done by Bossavit and Verité [22]. The lowest order edge elements are related to Whitney forms [21] and were independently discovered by Cendes [29]. Edge elements are believed to alleviate many of the problems associated with the nodal elements, such as representing fields in media with discontinuous medium properties. However, spurious solutions have been associated to edge, face and nodal elements [81]. For a discussion and comparison of edge elements and nodal elements see [23, 106, 107]. Mixed formulations in the time-domain using edge elements have achieved much popularity over the past decade. Analyses of different mixed methods and other work is done in this area by P. Monk [38, 89, 99, 100, 101, 102, 103, 104] as well as in [4, 87, 116] and references cited in these papers.

Edge elements of the mixed type and the Raviart-Thomas elements have also been used for the transient wave equation written in the mixed *velocity-stress* form as a system of first order PDE's [35, 47, 60]. The use of mass-lumping techniques in FEM's to create higher order numerical methods for the solution of the transient wave equation has recently been studied by [13, 34, 35, 37]. We use a mixed method involving edge elements in two-dimensions, in the discretization of the wave equation written in the velocity-stress form, in Chapter 3. We also use edge elements due to Raviart and Thomas [116] for the two-dimensional TM mode of Maxwell's equation's in Chapter's 5 and 6.

A major weakness of the FEM is that, unlike the moment methods, the infinite extent of a structure is not accounted for in the variational problem; thus it is relatively difficult to model open region problems. This difficulty is also present in the finite difference methods. Absorbing boundaries, which will be discussed later on in this chapter, have to be used to overcome this deficiency. There are numerous books and many review papers on FEM. For a selected bibliography see [111].

1.2.3 Fictitious Domain Methods

An alternative way of solving scattering problems is to use a *fictitious domain method*. A fictitious domain method is a technique in which the solution to a given problem is obtained by extending the given data to a larger but simpler shaped domain, containing the original domain, and solving the corresponding equations in this larger *fictitious domain*. Fictitious domain methods can be traced back to the 1960's to Saulév [120]. The fictitious domain could be a rectangle or a circle, for example. The advantage of this method is that the problem in the fictitious domain can be discretized on a uniform mesh, independent of the obstacle boundary, thus avoiding the time consuming construction of a boundary fitted mesh as in the FEM. However, there are some classes of fictitious domain methods that use boundary fitted meshes to improve accuracy [92].

A fictitious domain method is also known as a *domain imbedding method* [20] or a *fictitious component method* [55]. A related technique is the *capacitance matrix method* [50, 53]. One class of fictitious domain methods involves using *distributed/boundary Lagrange multipliers* to enforce the boundary conditions on the boundaries of the original smaller domain. This is known as the *functional analytic approach* which leads to *saddle point problems* and has been considered by Y. Kuznetsov [85, 64, 76] and jointly by R. Glowinski, T. W. Pan and J. Périaux [63, 66] among others. This is the class of problems that we will consider in this thesis.

There are other classes of fictitious domain methods. We will mention some of these

briefly. One class of methods uses an *optimal control* approach [9, 74]. In this approach, the system is solved in the fictitious domain, with a distributed/boundary control introduced on the right hand side of the system equations. The control forces the solution to satisfy the required boundary conditions, at least approximately. This approach resembles the Lagrange multiplier technique. In some cases the solution of the optimal control problem is obtained by solving the optimality conditions. The cost function here consists of two parts, one part penalizes the boundary condition and the other part penalizes the control inside the original domain/boundary [74]. This leads to an optimization problem.

Fictitious domain methods are often used to construct a preconditioner for iterative methods, such as *Krylov subspace methods*. One such approach is called an *algebraic* fictitious domain method [55, 92].

Fictitious domain methods were originally developed to handle problems with complex geometries in the stationary case [10, 66]. The application of fictitious domain methods to time dependent problems is relatively new and has recently been studied by R. Glowinski, P. Joly among others [27, 41, 57, 67].

1.2.4 Operator Splitting Methods

The idea behind operator splitting methods (fractional step methods) is to reduce the solution of a complicated problem into a series of subproblems, on smaller time intervals, which are generally simpler to solve. The different subproblems can then be solved using numerical techniques that are best suited for them. Thus a mix of different discretizations can be used for the different subproblems. These subproblems are connected to each other by using the solution of one subproblem as initial conditions to the succeeding one.

Operator splitting methods were introduced in the 1950's by Peaceman and Rachford [110]. Fractional step methods were also investigated by Yanenko [133], Marchuk [90, 91] and others. Second-order accurate splitting schemes were investigated by G. Strang [122], Gottlieb [71] and others [54]. The augmented Lagrangian method and operator splitting

methods have been extensively applied by R. Glowinski *et al.*, to the solution of problems from non-Newtonian fluid mechanics, non-linear elasticity and petroleum engineering seismic explorations [48, 65, 69].

1.3 Absorbing Boundary Conditions and Perfectly Matched Layers

The effective modeling of waves on unbounded domains by numerical methods such as the finite difference method or the finite element method is dependent on the particular absorbing boundary condition used to truncate the computational domain. An early example of an unbounded problem was in limited area weather forecasting where there are no natural boundaries and the boundary data is not known in advance. Another example, that will be considered in this thesis, is scattering by an obstacle, which is encountered in acoustics, electromagnetics as well as elastodynamics in the time-dependent as well as the time-harmonic case. An ideal truncation scheme must ensure that the outgoing waves do not reflect backwards into the computational domain, from the mesh termination surface, and corrupt the solution. Over the years many different solutions have been proposed to simulate wave propagation in unbounded domains. These solution methods can typically be classified into two categories.

The first category comprises of the non reflecting, radiating or absorbing boundary conditions (ABC's). The ABC's can be first, second or higher order boundary conditions. They are applied at the mesh termination surface to truncate the computational volume, as required by any PDE solution. One of the first radiation boundary conditions was in the area of meteorological applications by Orlianski. Some of the seminal works in this area are due to Engquist and Majda [52] who gave a mathematical treatment of ABC's for hyperbolic problems based on pseudo-differential operators. Mur applied this technique to Maxwell's equations [105]. Other important works are [17, 30, 77, 83]. A review of work

done in the area of ABC's can be seen in [127] in the special issue on ABC's [128]. Some of the common problems with ABC's are to do with accuracy control, conformality, ease of parallelization and implementation difficulties with higher order ABC's.

An alternative to ABC's are absorbing layer models [14, 115]. In 1994, J.P. Berenger created a technique called the *perfectly matched layer* (PML) method for the reflectionless absorption of electromagnetics waves [15, 16]. The PML is an absorbing layer that is placed around the computational domain of interest in order to attenuate outgoing radiation. Berenger showed that, the PML allowed perfect transmission of electromagnetic waves across the interface of the computational domain, regardless of the frequency, polarization or angle of incidence of the waves, and the waves are attenuated exponentially with depth in the layer. The absorbing capabilities of the PML are roughly about 3 orders of magnitude better than most ABC's. In this thesis we will construct a PML model for Maxwell's equations in a finite element setting. Since the paper of Berenger a large volume of research on PML's has been carried out, see for example [6, 31, 59, 96, 97, 112, 125, 129].

1.4 Outline of Thesis

In this thesis we propose and analyze a fictitious domain method for the time-dependent problem of scattering by an obstacle. We do this for the scalar wave equation in two dimensions as well as for the TM mode of Maxwell's equations in two-dimensions. We also construct a PML model for the two-dimensional TM mode and implement it in a mixed finite element setting. The outline of the thesis is as follows.

In Chapter 2 we consider operator splitting schemes for a second order problem. In Section 2.3 we construct a two-step operator splitting method for a second-order problem written as a system of first order equations. Each of the subproblems utilizes a Crank-Nicholson update. We analyze the two-step scheme to determine its temporal accuracy. In Section 2.4 we consider a symmetrized version of the two-step operator splitting scheme,

using the symmetrization idea due to Strang [122], in order to obtain a more accurate (in time) method. Numerical validations are made in each case to support the theoretical analyses.

In Chapter 3, we propose a fictitious domain method, based on a distributed Lagrange multiplier, for the solution of the two-dimensional scalar wave equation with a Dirichlet boundary condition; namely a time-dependent problem of scattering by an obstacle. In Section 3.4 we present a fully conforming method for the numerical solution of this problem. In Section 3.5 we propose a symmetrized operator splitting scheme for the solution of the wave problem. The idea of the splitting scheme is to decouple the operator that propagates the wave from the operator that enforces the Dirichlet boundary condition. In Section 3.7 the operator splitting scheme is used in two settings; a purely conforming approach, and an approach based on the velocity-stress formulation of the wave equation. We use the lowest order Nédélec edge elements on rectangles in two-dimensions to approximate the gradient of the solution, and nodal bilinear finite elements to approximate the solution and its time derivative. In Sections 3.8 and 3.9 we present numerical examples to demonstrate the effectiveness of the new fictitious domain method.

In Chapter 4 we perform a 1D plane wave analysis of the fictitious domain method and the operator splitting scheme to obtain expressions for the dispersion relation and the reflection coefficients in each case. A comparison of these methods is done with the FDTD method as well as with another fictitious domain method based on a boundary Lagrange multiplier [27, 41] in Sections 4.4 and 4.5.

In Chapter 5, Sections 5.2-5.5, we construct a *uniaxial anisotropic* perfectly matched layer model for the two-dimensional TM mode of Maxwell's equation using a mixed Galerkin finite element formulation. In Section 5.6 we prove some energy estimates for our PML model. The discrete model, presented in Section 5.7, uses the lowest order Raviart-Thomas finite elements on rectangular grids [116] for the discretization of the magnetic field and nodal bilinear finite elements for the electric field. In Section 5.8 we perform a

dispersion analysis to study the errors present in the discrete PML model. The analysis is based on similar approaches used in [86, 103]. In Section 5.9 we present calculations of the reflection coefficient and analyze the errors present in the discrete PML model which are caused by the discretization scheme, and the termination of the PML layer by means of a *perfect conductor condition* (PEC). Numerical results are presented in Section 5.10 to demonstrate the effectiveness of the new model.

In Chapter 6 we incorporate the fictitious domain method, introduced in Chapter 3, into the two-dimensional TM mode of Maxwell's equations. Using the uniaxial formulation of the perfectly matched layer, presented in Chapter 5, we consider a time-dependent scattering problem by employing the fictitious domain method of Chapter 3. We also consider a first order absorbing boundary condition for Maxwell's equations, called the *Silver-Müller* condition as a basis for comparison with the PML model.

The main results of Chapter 3 will be published in [19] and Chapter 5 has been submitted for a journal publication [18].

Chapter 2

Operator Splitting Methods

2.1 Introduction

Let us consider the initial value problem

$$\begin{cases} \dfrac{d\phi}{dt} + A(\phi, t) = 0, \\ \phi(0) = \phi_0, \end{cases} \tag{2.1}$$

where A is an operator (possibly nonlinear) from a Hilbert space H into itself and $\phi_0 \in H$. Suppose now that operator A has the decomposition

$$A = A_1 + A_2, \tag{2.2}$$

with the *suboperators* A_1 and A_2 being individually simpler operators than A. It is then quite natural to integrate the initial value problem (2.1) by numerical methods taking advantage of the decomposition property (2.2). This can be achieved by *operator splitting* schemes which are the focus of this chapter.

In this chapter we will discuss two different operator splitting schemes, a two-step scheme, and its *symmetrized* version due to G. Strang, for the time integration of (2.1). The temporal accuracy of these two schemes will be discussed for two cases; in the case that suboperators A_1 and A_2 commute, and the case that they do not commute.

Using the theory of the splitting schemes for first order initial value problems, we will apply these schemes to the time integration of initial value problems involving second order ordinary differential equations

$$\begin{cases} \dfrac{d^2\phi}{dt^2} + A\phi = 0, \\ \phi(0) = \phi_0,\ \phi_t(0) = \phi_1, \end{cases} \tag{2.3}$$

where, in (2.3), $\phi(t) \in \mathbb{R}^d$, $\forall t > 0$, $\phi_0, \phi_1 \in \mathbb{R}^d$, and A is a $d \times d$ real, symmetric, positive definite matrix, independent of t.

The presentation in this chapter is based on Chapters 2 and 6 of the book [62] by R. Glowinski. An outline of the chapter is as follows. In Section 2.2 we study a second order hyperbolic problem. In Section 2.3 we construct a two-step operator splitting method for the second-order problem (2.3) written as a system of first order equations. Each of the subproblems uses a Crank-Nicolson update. We analyze the two-step scheme to determine its temporal accuracy. The theoretical result is demonstrated numerically by different examples. In Section 2.4 we consider a symmetrized version of the two-step operator splitting scheme, using the symmetrization idea due to Strang [122], in order to obtain a more accurate (in time) method. Numerical validations are made in each case to support the theoretical analyses.

2.2 A Second Order Problem

We will first rewrite the initial value problem (2.3) in first order form, and study the properties of the resulting system. Let us consider the system (2.3) of second order ordinary differential equations

$$\begin{cases} \dfrac{d^2\phi}{dt^2} + A\phi = 0, \\ \phi(0) = \phi_0,\ \phi_t(0) = \phi_1. \end{cases} \tag{2.4}$$

where, in (2.4), $\phi(t) \in \mathbb{R}^d$, $\forall t > 0$, $\phi_0, \phi_1 \in \mathbb{R}^d$, and A is a $d \times d$ real, symmetric, positive definite matrix, independent of t. We rewrite system (2.3) in first order form by defining a

new variable

$$u = \frac{d\phi}{dt}, \tag{2.5}$$

with which system (2.4) becomes

$$\begin{cases} \dfrac{du}{dt} + A\phi = 0, \\ \dfrac{d\phi}{dt} - u = 0, \\ \phi(0) = \phi_0,\ u(0) = u_0 = \phi_1. \end{cases} \tag{2.6}$$

Let us define the vector

$$\chi = \begin{bmatrix} u & \phi \end{bmatrix}^T. \tag{2.7}$$

Vector χ satisfies system

$$\begin{cases} \dfrac{d\chi}{dt} + \mathcal{A}(\chi, t) = 0, \\ \chi(0) = \chi_0, \end{cases} \tag{2.8}$$

where

$$\chi_0 = \begin{bmatrix} u_0 & \phi_0 \end{bmatrix}^T, \tag{2.9}$$

and

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ -I & 0 \end{bmatrix}. \tag{2.10}$$

Above, I is the $d \times d$ identity matrix. The solution of system (2.8) is given by

$$\chi(t) = \mathrm{e}^{-\mathcal{A}t}\chi_0,\ \forall t \geq 0. \tag{2.11}$$

Since the $d \times d$ matrix A is symmetric and positive definite, it can be diagonalized. Let

$$\Upsilon = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_d),\ \forall k = 1, \ldots, d, \gamma_k \in \mathbb{R},\ \gamma_k > 0, \tag{2.12}$$

be the diagonal matrix of eigenvalues γ_k of matrix A. Let the set $\{w_k\}_{k=1}^d$ be an orthonormal vector basis of $\mathbb{R}^d$ consisting of eigenvectors of A. Here, w_k is the eigenvector corresponding to the eigenvalue γ_k, i.e.,

$$Aw_k = \gamma_k w_k,\ \forall k = 1, 2, \ldots, d. \tag{2.13}$$

Let P be a $d \times d$ matrix such that

$$P = (w_1, w_2, \ldots, w_d)^T, \ \forall k = 1, 2, \ldots, d, \ w_k \in \mathbb{R}^d. \tag{2.14}$$

Then, we have

$$A = P \Upsilon P^T, \text{ with } PP^T = P^T P = I. \tag{2.15}$$

To find the eigenvalues of matrix $\mathcal{A}$ we need to solve for the variable ξ in the system of equations given by

$$\mathcal{A}\mathcal{W} = \xi \mathcal{W}, \tag{2.16}$$

with $\mathcal{W} = [x_1^T, x_2^T]^T, \ \forall i = 1, 2, x_i \in \mathbb{R}^d$. This is equivalent to the system

$$\begin{cases} Ax_2 = \xi x_1, \\ \\ x_1 = -\xi x_2. \end{cases} \tag{2.17}$$

Eliminating x_1 we obtain the matrix equation

$$Ax_2 + \xi^2 x_2 = 0. \tag{2.18}$$

Thus, choosing $x_2 = w_k$ for $k = 1, 2, \ldots, d$, we obtain

$$\gamma_k w_k + \xi^2 w_k = 0. \tag{2.19}$$

Since $w_k \neq 0$, this implies that

$$\xi^2 = -\gamma_k \implies \xi = \pm \mathrm{i}\sqrt{\gamma_k}. \tag{2.20}$$

Thus, the eigenvalues of $\mathcal{A}$ are given by

$$\xi_k = \begin{cases} \mathrm{i}\sqrt{\gamma_k}, & \text{for } k = 1, 2, \ldots, d, \\ -\mathrm{i}\sqrt{\gamma_{k-d}}, & \text{for } k = d+1, d+2, \ldots, 2d. \end{cases} \tag{2.21}$$

and, the corresponding *right* eigenvectors are

$$\mathcal{W}_k = \begin{cases} \dfrac{1}{\sqrt{\gamma_k + 1}}[-\mathrm{i}\sqrt{\gamma_k} w_k^T, w_k^T]^T, & \forall \ k = 1, 2, \ldots, d, \\ \dfrac{1}{\sqrt{\gamma_{k-d} + 1}}[\mathrm{i}\sqrt{\gamma_{k-d}} w_{k-d}^T, w_{k-d}^T]^T, & \forall \ k = d+1, d+2, \ldots, 2d. \end{cases} \tag{2.22}$$

Let Λ be the $2d \times 2d$ diagonal matrix of eigenvalues of $\mathcal{A}$, i.e.,

$$\Lambda = \mathrm{diag}(\xi_1, \xi_2, \ldots, \xi_{2d}). \tag{2.23}$$

Let $\mathcal{P}$ be a $2d \times 2d$ matrix such that

$$\mathcal{P} = (\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{2d})^T, \ \forall k = 1, 2, \ldots, 2d, \ \mathcal{W}_k \in \mathbb{R}^{2d}. \tag{2.24}$$

The columns of the matrix $\mathcal{P}$ are the right eigenvectors of matrix $\mathcal{A}$. Then, the columns of $\mathcal{P}^{-T}$ form the *left* eigenvectors of matrix $\mathcal{A}$ and we have the decomposition [70]

$$\mathcal{P}^{-1}\mathcal{A}\mathcal{P} = \Lambda \implies \mathcal{A} = \mathcal{P}\Lambda\mathcal{P}^{-1}. \tag{2.25}$$

Then, the solution (2.11) can be written as

$$\chi(t) = \mathcal{P}\mathrm{e}^{-\Lambda t}\mathcal{P}^{-1}\chi_0. \tag{2.26}$$

Thus, projecting over the eigenvectors of matrix $\mathcal{A}$ we obtain

$$\chi_k(t) = \mathrm{e}^{-\xi_k t}\chi_{0k}, \ \forall k = 1, 2, \ldots, 2d. \tag{2.27}$$

2.3 A Two-Step Operator Splitting Scheme

2.3.1 First Order Problems

Consider the following initial value problem

$$\begin{cases} \dfrac{d\phi}{dt} + A_1\phi + A_2\phi = 0, \\ \phi(0) = \phi_0, \end{cases} \tag{2.28}$$

where, in (2.28), $\phi(t) \in \mathbb{R}^d$, $\forall t > 0$, $\phi_0 \in \mathbb{R}^d$, and $\forall i = 1, 2, A_i$ is a $d \times d$ real matrix, independent of t. The solution of problem (2.28) is given by

$$\phi(t) = \mathrm{e}^{-(A_1+A_2)t}\phi_0, \ \forall t \geq 0. \tag{2.29}$$

We consider a time discretization step $\Delta t > 0$ and denote $(n+\alpha)\Delta t$ by $t^{n+\alpha}$, with $n \in \mathbb{N}$. Let $\phi^{n+\alpha} \approx \phi(t^{n+\alpha})$. It follows from (2.29) that

$$\phi(t^{n+1}) = \mathrm{e}^{-(A_1+A_2)\Delta t}\phi(t^n),\ \forall n \geq 0. \tag{2.30}$$

Suppose now that matrices A_1 and A_2 commute, i.e., $A_1A_2 = A_2A_1$. Then from (2.30) we have

$$\phi(t^{n+1}) = \mathrm{e}^{-A_1\Delta t}\mathrm{e}^{-A_2\Delta t}\phi(t^n),\ \forall n \geq 0. \tag{2.31}$$

From the relation (2.31), ϕ^{n+1} can be obtained *exactly* from ϕ^n via the solution of

$$\begin{cases} \text{(i)} & \dfrac{dw}{dt} + A_1 w = 0, \text{ on } (t^n, t^{n+1}), \\ \text{(ii)} & w(t^n) = \phi(t^n), \end{cases} \tag{2.32}$$

$$\phi^{n+1/2} = w(t^{n+1}), \tag{2.33}$$

and

$$\begin{cases} \text{(i)} & \dfrac{dw}{dt} + A_2 w = 0, \text{ on } (t^n, t^{n+1}), \\ \text{(ii)} & w(t^n) = \phi^{n+1/2}, \end{cases} \tag{2.34}$$

$$\phi^{n+1} = w(t^{n+1}). \tag{2.35}$$

In (2.33) and (2.34, ii), $\phi^{n+1/2}$ denotes a *predicted* value of ϕ at $t = t^{n+1}$. We can view (2.32) as a *predicting* step and (2.34) as a *correcting* step. Thus, starting from $\phi^0 = \phi_0$, for $n \geq 0$, we can obtain ϕ^{n+1} from ϕ^n via (2.32)-(2.35). This scheme is of the *operator splitting* type and is *exact* if A_1 and A_2 commute. Scheme (2.32)-(2.35) is discussed in [62, 90, 91, 133]. We will refer to the scheme (2.32)-(2.35) as the *two-step* operator splitting scheme.

2.3.2 Second Order Problems and Commuting Suboperators

In this section we construct an operator splitting scheme based on (2.32)-(2.35) for the second order problem (2.3) written in first order form (2.6).

Let α, β be real numbers such that $0 \leq \alpha, \beta \leq 1$ and $\alpha + \beta = 1$. We define the suboperators A_1 and A_2 as

$$A_1 = \alpha A, \ A_2 = \beta A. \tag{2.36}$$

Let

$$\phi^0 = \phi_0, u^0 = \phi_1, \tag{2.37}$$

then, for $n \geq 0$, we obtain ϕ^{n+1} from ϕ^n, and u^{n+1} from u^n via

$$\begin{cases} \text{(i)} & \dfrac{u^{n+1/2} - u^n}{\Delta t} + A_1 \left(\dfrac{\phi^{n+1/2} + \phi^n}{2} \right) = 0, \\ \text{(ii)} & \dfrac{\phi^{n+1/2} - \phi^n}{\Delta t} - \alpha \left(\dfrac{u^{n+1/2} + u^n}{2} \right) = 0, \end{cases} \tag{2.38}$$

to obtain $\phi^{n+1/2}$ and $u^{n+1/2}$, then solve

$$\begin{cases} \text{(i)} & \dfrac{u^{n+1} - u^{n+1/2}}{\Delta t} + A_2 \left(\dfrac{\phi^{n+1} + \phi^{n+1/2}}{2} \right) = 0, \\ \text{(ii)} & \dfrac{\phi^{n+1} - \phi^{n+1/2}}{\Delta t} - \beta \left(\dfrac{u^{n+1} + u^{n+1/2}}{2} \right) = 0. \end{cases} \tag{2.39}$$

Let

$$\chi^n = \begin{bmatrix} u^n & \phi^n \end{bmatrix}^T. \tag{2.40}$$

We can rewrite (2.38) and (2.39) using χ^n. For $n = 0$

$$\chi^0 = \begin{bmatrix} u^0 & \phi^0 \end{bmatrix}^T, \tag{2.41}$$

then, for $n \geq 0$, we obtain χ^{n+1} from χ^n via

$$\text{(i)} \begin{bmatrix} I & \dfrac{\Delta t}{2}\alpha A \\ -\dfrac{\Delta t}{2}\alpha I & I \end{bmatrix} \chi^{n+1/2} = \begin{bmatrix} I & \dfrac{-\Delta t}{2}\alpha A \\ \dfrac{\Delta t}{2}\alpha I & I \end{bmatrix} \chi^n, \tag{2.42}$$

to obtain $\chi^{n+1/2}$, then solve

$$\text{(ii)} \begin{bmatrix} I & \dfrac{\Delta t}{2}\beta A \\ -\dfrac{\Delta t}{2}\beta I & I \end{bmatrix} \chi^{n+1} = \begin{bmatrix} I & \dfrac{-\Delta t}{2}\beta A \\ \dfrac{\Delta t}{2}\beta I & I \end{bmatrix} \chi^{n+1/2}. \tag{2.43}$$

Let $\mathcal{I}$ be the $2d \times 2d$ identity matrix. Then, we can rewrite (2.42), (2.43) as

$$\begin{cases} \text{(i)} \quad (\mathcal{I} + \dfrac{\Delta t}{2}\alpha\mathcal{A})\chi^{n+1/2} = (\mathcal{I} - \dfrac{\Delta t}{2}\alpha\mathcal{A})\chi^{n}, \\ \text{(ii)} \quad (\mathcal{I} + \dfrac{\Delta t}{2}\beta\mathcal{A})\chi^{n+1} = (\mathcal{I} - \dfrac{\Delta t}{2}\beta\mathcal{A})\chi^{n+1/2}. \end{cases} \tag{2.44}$$

Eliminating $\chi^{n+1/2}$ from (2.44) we get

$$\chi^{n+1} = (\mathcal{I} + \frac{\Delta t}{2}\beta\mathcal{A})^{-1}(\mathcal{I} - \frac{\Delta t}{2}\beta\mathcal{A})(\mathcal{I} + \frac{\Delta t}{2}\alpha\mathcal{A})^{-1}(\mathcal{I} - \frac{\Delta t}{2}\alpha\mathcal{A})\chi^{n}. \tag{2.45}$$

The discrete analogues of (2.11) and (2.26) are

$$\chi^{n} = (\mathcal{I} + \frac{\Delta t}{2}\beta\mathcal{A})^{-n}(\mathcal{I} - \frac{\Delta t}{2}\beta\mathcal{A})^{n}(\mathcal{I} + \frac{\Delta t}{2}\alpha\mathcal{A})^{-n}(\mathcal{I} - \frac{\Delta t}{2}\alpha\mathcal{A})^{n}\chi_0, \tag{2.46}$$

and,

$$\chi_i^{n} = \left(\frac{1 - \dfrac{\Delta t}{2}\beta\xi_i}{1 + \dfrac{\Delta t}{2}\beta\xi_i}\right)^{n} \left(\frac{1 - \dfrac{\Delta t}{2}\alpha\xi_i}{1 + \dfrac{\Delta t}{2}\alpha\xi_i}\right)^{n} \chi_{0i}, \ \forall\, i = 1, 2, \ldots 2d. \tag{2.47}$$

In order to study the *accuracy* of the two-step scheme we introduce the following *rational function*

$$\mathcal{R}(\zeta) = \left(\frac{1 - \beta\dfrac{\zeta}{2}}{1 + \beta\dfrac{\zeta}{2}}\right)\left(\frac{1 - \alpha\dfrac{\zeta}{2}}{1 + \alpha\dfrac{\zeta}{2}}\right). \tag{2.48}$$

Expanding $\mathcal{R}(\zeta)$ as a Taylor series in the neighborhood of $\zeta = 0$ we get

$$\mathcal{R}(\zeta) = 1 - (\alpha + \beta)\zeta + (\alpha + \beta)^2\frac{\zeta^2}{2} - (\alpha + \beta)(\alpha^2 + \alpha\beta + \beta^2)\frac{\zeta^3}{4} + \zeta^4\mathcal{O}(1). \tag{2.49}$$

On the other hand we have

$$\mathrm{e}^{-\zeta} = 1 - \zeta + \frac{\zeta^2}{2} - \frac{\zeta^3}{6} + \zeta^4\mathcal{O}(1). \tag{2.50}$$

Comparing (2.49) and (2.50) we observe that the two-step scheme, for the time discretization given in (2.38) and (2.39), is *second order accurate* (local truncation error is $\mathcal{O}(\zeta^3)$) for any pair $\{\alpha, \beta\}$ satisfying $\alpha + \beta = 1$, $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$.

Remark 1 *In* (2.38)-(2.39)*, we have used a Crank Nicolson scheme, which is responsible for the second order accuracy of the two-step scheme. Using a lower accuracy scheme for the two different substeps will result in a first order accurate scheme.*

2.3.3 An Example with Commuting Suboperators

We demonstrate the second order accuracy of the two-step operator splitting scheme for the second order ODE (2.3) by a simple numerical example in which the suboperators A_1 and A_2 commute.

Let us take the operator A in (2.3) to be the identity operator, i.e.,

$$A\phi = \phi, \ \forall \phi \in C^2(\mathbb{R}^+; \mathbb{R}). \tag{2.51}$$

The initial value problem to be considered here is

$$\begin{cases} \dfrac{d^2\phi}{dt^2} + \phi = 0, \\ \\ \phi(0) = 1/2, \ \phi_t(0) = 0. \end{cases} \tag{2.52}$$

The exact solution of the initial value problem (2.52) is

$$\phi_E(t) = \frac{1}{2}\cos t, \ \forall t \geq 0. \tag{2.53}$$

We will apply the two-step splitting scheme (2.38)-(2.39), with the operator decomposition (2.36), to the time integration of problem (2.52) on the time interval $[0, 1]$, to obtain the computed solution ϕ_C. Table 2.1 presents the comparison of the relative errors between the exact solution (2.53) and the computed solution ϕ_C, over the time interval $[0, 1]$. The relative error is defined to be

$$RE = \frac{\|\phi_E - \phi_C\|_2}{\|\phi_E\|_2}, \tag{2.54}$$

where, $\|\psi\|_2$ is the Euclidean norm of ψ given by

$$\|\psi\|_2 = \left(\sum_{k=1}^{N} |\psi(k\Delta t)|^2 \right)^{1/2}, \tag{2.55}$$

with $N = 1/\Delta t$. Table 2.1 presents the relative errors for $\alpha = 0, 0.25, 0.5$. Since the results for α and $1 - \alpha$ are identical, given the nature of the two-step splitting, we have excluded the results for the case $\alpha = 0.75, 1.0$.

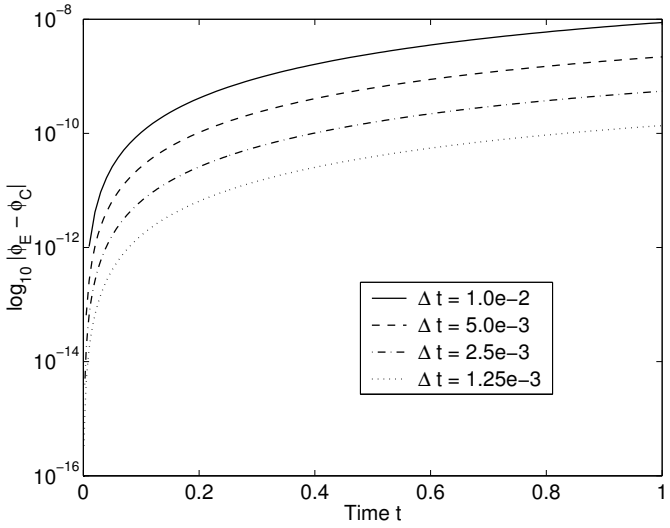

Figure 2.1: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of Δt ($\alpha = 0.5$), in the case that the suboperators commute using the two-step operator splitting scheme.

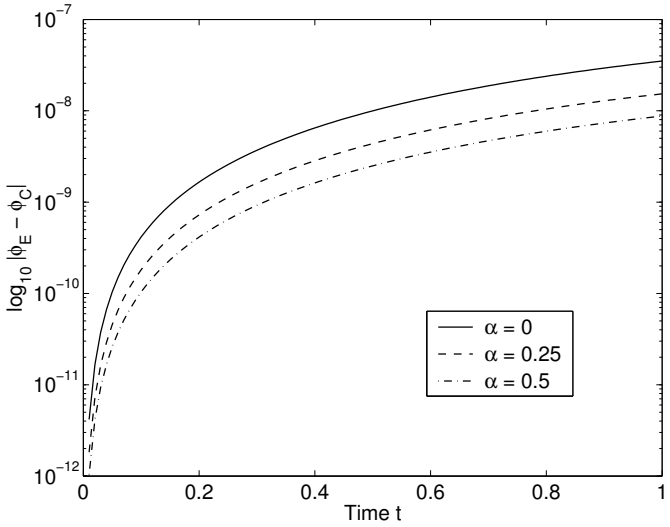

Figure 2.2: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of α ($\Delta t = 0.01$), in the case that the suboperators commute using the two-step operator splitting scheme

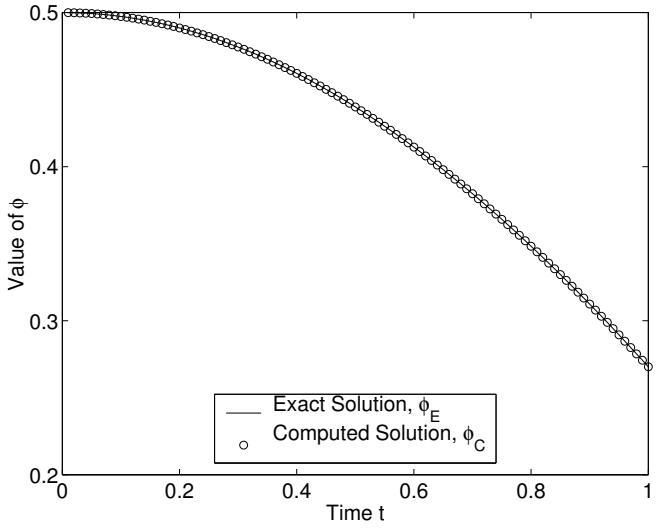

Figure 2.3: Comparison of the exact and computed solution $(\Delta t = 0.01)$, for the case that the suboperators commute, using the two-step operator splitting scheme.

	α		
Δt	0.0	0.25	0.5
1.0e-2	$3.93e-8$	$1.72e-8$	$9.81e-9$
5.0e-3	$9.75e-9$	$4.26e-9$	$2.44e-9$
2.5e-3	$2.43e-9$	$1.06e-9$	$6.07e-10$
1.25e-3	$6.06e-10$	$2.65e-10$	$1.52e-10$

Table 2.1: Comparison of the relative errors between the exact solution and the computed solution using the two-step operator splitting scheme for the case of commuting suboperators.

Figure 2.1 plots $\log_{10}|\phi_E - \phi_C|$, as a function of time, for different values of Δt, with a

fixed $\alpha = 0.5$. The second order temporal behavior of the scheme can be clearly seen in this plot. Figure 2.2 plots $\log_{10} |\phi_E - \phi_C|$, as a function of time, for different values of α, with a fixed $\Delta t = 0.01$. It can be seen that the plot for $\alpha = 0.5$ has the smallest errors. Figure 2.3 compares the exact solution with the computed solution for $\Delta t = 0.01$.

Remark 2 *In Table 2.1 we note that, for a given value of α, the ratio of successive entries, for $\Delta t = \tau$, and $\Delta t = \tau/2$, is approximately 4.0. This demonstrates the second order accuracy of the two-step scheme for this case in which the suboperators commute.*

2.3.4 The Case of Noncommuting Suboperators.

As mentioned in Remark 1, the two-step scheme is first-order accurate only, if low accuracy discretizations are used for the two different substeps of the scheme. First order accuracy also results if the operators A_1 and A_2 do not commute. In Section (2.3.1) we discussed the case in which the suboperators A_1 and A_2 do commute. Let us suppose now that these suboperators do not commute. We then have

$$\begin{aligned} \mathrm{e}^{-(A_1+A_2)\Delta t} &= I - (A_1 + A_2)\Delta t + \frac{1}{2}(A_1 + A_2)^2\Delta t^2 + \mathcal{O}(\Delta t^3) \\ &= I - (A_1 + A_2)\Delta t + \frac{1}{2}(A_1^2 + A_1A_2 + A_2A_1 + A_2^2)\Delta t^2 + \mathcal{O}(\Delta t^3). \end{aligned} \tag{2.56}$$

We have the expansions

$$\mathrm{e}^{-A_1\Delta t} = I - A_1\Delta t + \frac{1}{2}A_1^2\Delta t^2 + \mathcal{O}(\Delta t^3), \tag{2.57}$$

$$\mathrm{e}^{-A_2\Delta t} = I - A_2\Delta t + \frac{1}{2}A_2^2\Delta t^2 + \mathcal{O}(\Delta t^3). \tag{2.58}$$

The expansions (2.57) and (2.58) yield

$$\mathrm{e}^{-A_2\Delta t}\mathrm{e}^{-A_1\Delta t} = I - (A_1 + A_2)\Delta t + \frac{1}{2}(A_1^2 + 2A_2A_1 + A_2^2)\Delta t^2 + \mathcal{O}(\Delta t^3). \tag{2.59}$$

Comparing (2.56) with (2.59), we obtain

$$\mathrm{e}^{-A_2\Delta t}\mathrm{e}^{-A_1\Delta t} - \mathrm{e}^{-(A_1+A_2)\Delta t} = \frac{1}{2}(A_2A_1 - A_1A_2)\Delta t^2 + \mathcal{O}(\Delta t^3). \tag{2.60}$$

Thus, as can be seen from (2.60), in the case that the operators A_1 and A_2 *do not commute*, the two-step operator splitting scheme (2.32)-(2.34) is *first order accurate*.

2.3.5 An Example with Noncommuting Suboperators

In this section we consider a problem with an operator decomposition in which the suboperators do not commute. In this case, we demonstrate with a numerical example that the two-step operator splitting scheme is first order accurate in time.

Let us consider the initial value problem

$$\begin{cases} \dfrac{d^2\phi}{dt^2} + \phi - 1 = 0, \\ \\ \phi(0) = 1/2,\ \phi_t(0) = 0, \end{cases} \tag{2.61}$$

$\phi(t) \in \mathbb{R},\ \forall t \geq 0$. The exact solution of the initial value problem (2.61) is

$$\phi_E(t) = 1 - \frac{1}{2}\cos t,\ \forall t \geq 0. \tag{2.62}$$

The operator A, in this case, is

$$A\psi = \psi - 1,\ \forall \psi \in C^2(\mathbb{R}^+; \mathbb{R}). \tag{2.63}$$

We will apply the two-step splitting scheme (2.42)-(2.43) to the time integration of problem (2.61) with the suboperators being defined as

$$A_1\psi = \alpha\psi - \alpha', A_2\psi = \beta\psi - \beta', \tag{2.64}$$

where, as before $\alpha + \beta = 1$. Also, $\alpha' + \beta' = 1$, as well. Thus, results will be presented for different values of α'. We note in this case that the operators do not commute for all values of α and α'. Indeed

$$A_1A_2\psi = \alpha\beta\psi - \alpha\beta' - \alpha', \tag{2.65}$$

whereas,

$$A_2A_1\psi = \alpha\beta\psi - \alpha'\beta - \beta'. \tag{2.66}$$

Thus, the suboperators A_1 and A_2 commute if

$$\alpha\beta' + \alpha' = \alpha'\beta + \beta' \implies \alpha + \alpha' = 1. \tag{2.67}$$

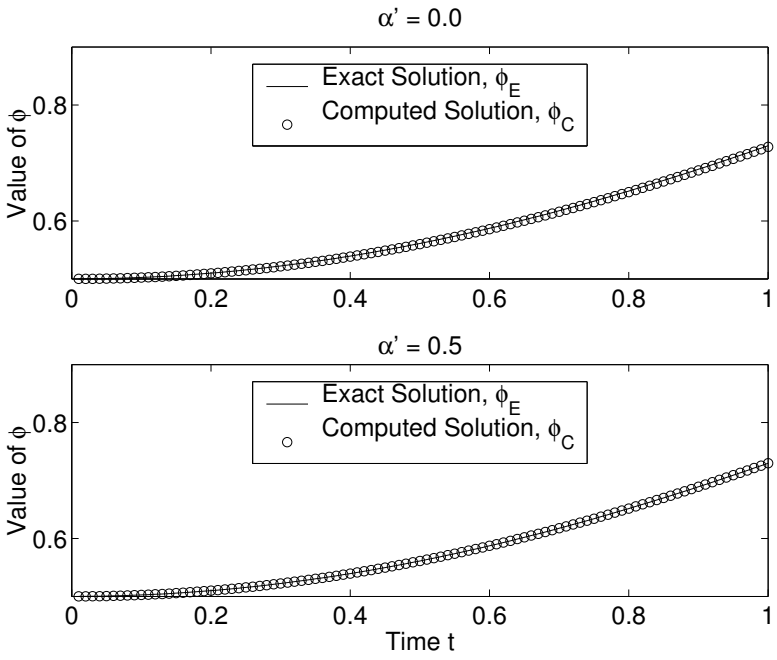

Figure 2.4: Comparison of the exact and computed solutions $(\Delta t = 0.01)$, using the two-step operator splitting scheme for the case of noncommuting suboperators.

	α'		
Δt	0.0	0.25	0.5
1.0e-2	$2.25e-3$	$1.12e-3$	$7.14e-9$
5.0e-3	$1.12e-3$	$5.61e-4$	$1.78e-9$
2.5e-3	$5.60e-4$	$2.80e-4$	$4.43e-10$
1.25e-3	$2.80e-4$	$1.40e-4$	$1.12e-10$

Table 2.2: Comparison of the relative errors between the exact solution and the computed solution using the two-step operator splitting scheme for the case of noncommuting suboperators.

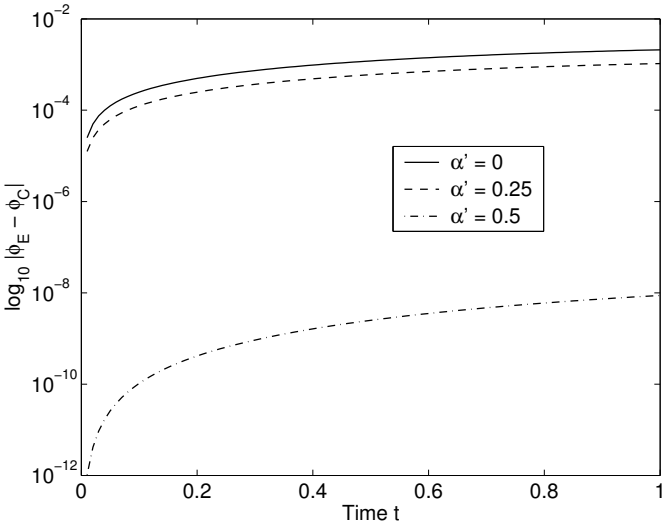

Figure 2.5: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of α' ($\Delta t = 0.01$), for the two-step operator splitting scheme in the case of noncommuting suboperators.

Remark 3 *In this example the operator A, and the suboperators A_1 and A_2 are not linear, but affine. The analysis performed in Section 2.3.2, demonstrating the second order temporal accuracy of the two-step operator splitting scheme, pertains to linear operators. Nevertheless, we will apply our operator splitting schemes to examples in which the operators involved are affine or even nonlinear, in order to get an idea of the numerical accuracy in time of the different examples considered.*

Table 2.2 presents the comparison of the relative errors between the exact solution (2.62) and the computed solution over the time interval $[0, 1]$. We have chosen $\alpha = \beta = 0.5$, and $\alpha' = 0, 0.25, 0.5$. As before, we have excluded the results for the case $\alpha' = 0.75, 1.0$. Figure 2.4 compares the exact solution with the solution computed for $\Delta t = 0.01$ in two cases. The top comparison is for $\alpha' = 0.0$, for which the two-step scheme does not commute, and the bottom comparison is for $\alpha' = 0.5$, for which the two-step scheme does

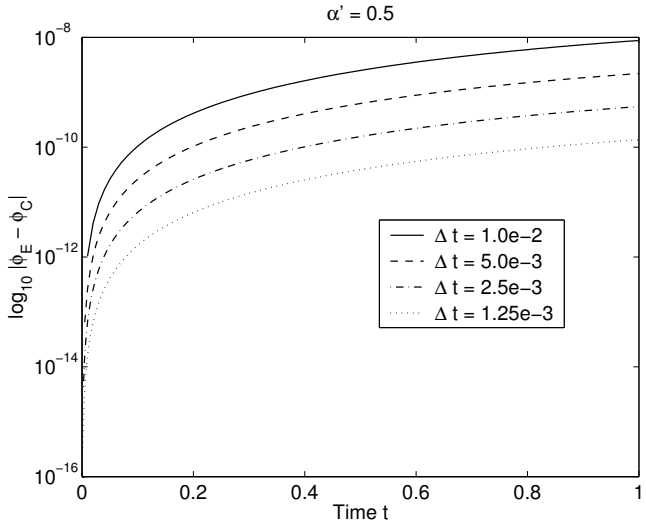

Figure 2.6: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of Δt, for the two-step operator splitting scheme for the case of noncommuting suboperators.

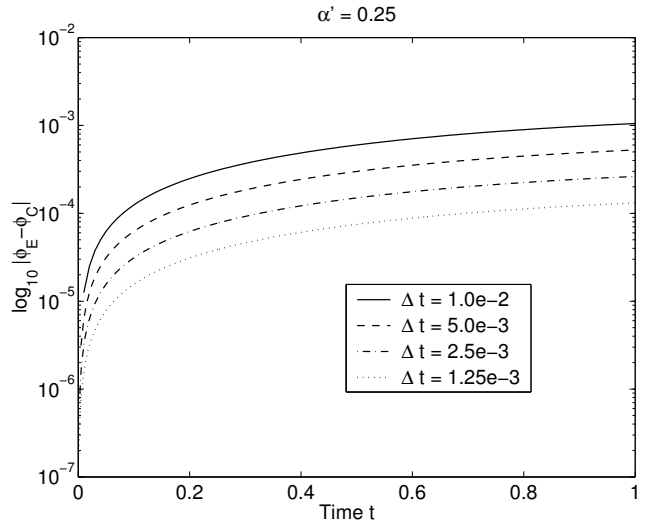

Figure 2.7: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of Δt, for the two-step operator splitting scheme for the case of noncommuting suboperators.

commute. The agreement with the exact solution is better for the case $\alpha' = 0.5$. Figure 2.5 plots $\log_{10} |\phi_E - \phi_C|$, as a function of time, for different values of α', with a fixed $\Delta t = 0.01$. It can be seen that the plot for $\alpha' = 0.5$ has the smallest errors. Figure 2.6 plots $\log_{10} |\phi_E - \phi_C|$, as a function of time, for different values of Δt, with a fixed $\alpha' = 0.5$. The second order temporal behavior of the scheme can be clearly seen in this plot. Figure 2.7 plots $\log_{10} |\phi_E - \phi_C|$ for different values of Δt, with a fixed $\alpha' = 0.25$. The first order temporal behavior of the scheme is evident from the logarithmic plot.

Remark 4 *In Table 2.2 the ratio of successive entries for $\Delta t = \tau$ and $\Delta t = \tau/2$ is approximately 2.0, in the case that $\alpha' = 0$ or $\alpha' = 0.25$. When $\alpha = 0.5$, $\alpha + \alpha' = 1.0$ and the suboperators A_1 and A_2 commute. Thus, in this case, the ratio of successive entries is approximately 4.0. This demonstrates the second order accuracy of the two-step scheme for the case ($\alpha' = 0.5$), in which the suboperators commute, and first order accuracy in the case ($\alpha' = 0, 0.25$) that the suboperators do not commute.*

Remark 5 *Let us suppose that in* (2.1) *we have*

$$A(\phi, t) = B(\phi) - f(t), \tag{2.68}$$

with

$$B = B_1 + B_2. \tag{2.69}$$

In order to apply the two-step scheme (2.32)-(2.35) *or the symmetrized scheme* (2.76)-(2.81)*, which will be presented in Section 2.4.1, to the solution of the initial value problem* (2.1)*, the forcing term f needs to be decomposed as well. To do so, we can employ any* reasonable *decomposition*

$$f = f_1 + f_2. \tag{2.70}$$

A simple decomposition is given by,

$$f_1 = \frac{1}{2} f, \ f_2 = \frac{1}{2} f. \tag{2.71}$$

2.4 Construction of a Second Order Accurate Splitting Scheme Using Symmetrization

2.4.1 First Order Problems

As seen in the previous sections, the two-step operator splitting scheme is first order accurate in the case that the suboperators do not commute. It is possible to construct a variant of the two-step scheme that remains second-order accurate, even in non-commutative cases. The construction of such a scheme was first given by G.Strang [122], and is known as a *symmetrized splitting scheme*.

Let us decompose the operator A in (2.1) as

$$A = \frac{1}{2}A_1 + A_2 + \frac{1}{2}A_1. \tag{2.72}$$

Then

$$\begin{aligned} \mathrm{e}^{-A_1\Delta t/2}\mathrm{e}^{-A_2\Delta t}\mathrm{e}^{-A_1\Delta t/2} &= I - (A_1 + A_2)\Delta t \\ &\quad + \frac{1}{2}(A_1^2 + A_2A_1 + A_1A_2 + A_2^2)\Delta t^2 + \mathcal{O}(\Delta t^3). \end{aligned} \tag{2.73}$$

Comparing (2.73) and (2.56) we have

$$\mathrm{e}^{-A_1\Delta t/2}\mathrm{e}^{-A_2\Delta t}\mathrm{e}^{-A_1\Delta t/2} - \mathrm{e}^{-(A_1+A_2)\Delta t} = \mathcal{O}(\Delta t^3). \tag{2.74}$$

We note that the right hand side of (2.73) vanishes in the case that the operators A_1 and A_2 commute.

We construct a *symmetrized* splitting scheme based on the decomposition (2.72) for the integration of the initial value problem (2.1). Let

$$\phi^0 = \phi_0, \tag{2.75}$$

then, for $n \geq 0$*, we obtain* ϕ^{n+1} *from* ϕ^n *via*

$$\begin{cases} \text{(i)} & \dfrac{dw}{dt} + A_1 w = 0, \text{ on } (t^n, t^{n+1/2}), \\ \text{(ii)} & w(t^n) = \phi(t^n), \end{cases} \tag{2.76}$$

$$\text{to obtain } \phi^{n+1/2} = w(t^{n+1/2}). \tag{2.77}$$

$$\begin{cases} \text{(i)} & \dfrac{dw}{dt} + A_2 w = 0, \text{ on } (t^n, t^{n+1}), \\ \text{(ii)} & w(t^n) = \phi^{n+1/2}, \end{cases} \tag{2.78}$$

$$\text{to obtain } \tilde{\phi}^{n+1} = w(t^{n+1}). \tag{2.79}$$

$$\begin{cases} \text{(i)} & \dfrac{dw}{dt} + A_1 w = 0, \text{ on } (t^{n+1/2}, t^{n+1}), \\ \text{(ii)} & w(t^{n+1/2}) = \tilde{\phi}(t^{n+1}), \end{cases} \tag{2.80}$$

$$\phi^{n+1} = w(t^{n+1}). \tag{2.81}$$

Remark 6 *The splitting scheme* (2.76)-(2.81) *is exact if the operators A_1 and A_2 commute. In this case we have $\phi^n = \phi(t^n), \forall n \geq 0$. In the case that the operators A_1 and A_2 do not commute, the scheme is second order accurate.*

2.4.2 A Symmetrized Splitting Scheme for the Second Order Problem

We construct a symmetrized scheme for the second order problem written in first order form (2.6), based on the scheme ((2.76)-(2.81)). As before, let α, β be real numbers such that $\alpha + \beta = 1$, $0 \leq \alpha, \beta \leq 1$. Let us define the suboperators as

$$A_1 = \alpha A, \ A_2 = \beta A. \tag{2.82}$$

Let

$$\phi^0 = \phi_0, u^0 = \phi_1, \tag{2.83}$$

then, for $n \geq 0$, we obtain ϕ^{n+1} from ϕ^n and u^{n+1} from u^n via

$$\begin{cases} \text{(i)} & \dfrac{u^{n+1/2} - u^n}{\Delta t/2} + A_1 \left(\dfrac{\phi^{n+1/2} + \phi^n}{2} \right) = 0, \\ \text{(ii)} & \dfrac{\phi^{n+1/2} - \phi^n}{\Delta t/2} - \alpha \left(\dfrac{u^{n+1/2} + u^n}{2} \right) = 0, \end{cases} \tag{2.84}$$

to obtain $\phi^{n+1/2}, u^{n+1/2}$.

$$\begin{cases} \text{(i)} & \dfrac{\tilde{u}^{n+1} - u^n}{\Delta t} + A_2 \left(\dfrac{\tilde{\phi}^{n+1} + \phi^{n+1/2}}{2} \right) = 0, \\ \text{(ii)} & \dfrac{\tilde{\phi}^{n+1} - \phi^n}{\Delta t} - \alpha \left(\dfrac{\tilde{u}^{n+1} + u^{n+1/2}}{2} \right) = 0, \end{cases} \tag{2.85}$$

to obtain $\tilde{\phi}^{n+1/2}, \tilde{u}^{n+1/2}$.

$$\begin{cases} \text{(i)} & \dfrac{u^{n+1}-\tilde{u}^{n+1}}{\Delta t/2} + A_1\left(\dfrac{\phi^{n+1}+\tilde{\phi}^{n+1}}{2}\right) = 0, \\ \text{(ii)} & \dfrac{\phi^{n+1}-\tilde{\phi}^{n+1}}{\Delta t/2} - \alpha\left(\dfrac{u^{n+1}+\tilde{u}^{n+1}}{2}\right) = 0. \end{cases} \tag{2.86}$$

We can rewrite (2.84)-(2.86) using χ^n as follows. For $n = 0$

$$\chi^0 = \begin{bmatrix} u_0 & \phi_0 \end{bmatrix}^T, \tag{2.87}$$

then, for $n \geq 0$, we obtain χ^{n+1} from χ^n via

$$\text{(i)} \quad \begin{bmatrix} I & \dfrac{\Delta t}{4}\alpha A \\ -\dfrac{\Delta t}{4}\alpha I & I \end{bmatrix} \chi^{n+1/2} = \begin{bmatrix} I & \dfrac{-\Delta t}{4}\alpha A \\ \dfrac{\Delta t}{4}\alpha I & I \end{bmatrix} \chi^n, \tag{2.88}$$

to obtain $\chi^{n+1/2}$, then solve

$$\text{(ii)} \quad \begin{bmatrix} I & \dfrac{\Delta t}{2}\beta A \\ -\dfrac{\Delta t}{2}\beta I & I \end{bmatrix} \tilde{\chi}^{n+1} = \begin{bmatrix} I & \dfrac{-\Delta t}{2}\beta A \\ \dfrac{\Delta t}{2}\beta I & I \end{bmatrix} \chi^{n+1/2}, \tag{2.89}$$

to obtain $\tilde{\chi}^{n+1}$, then solve

$$\text{(iii)} \quad \begin{bmatrix} I & \dfrac{\Delta t}{4}\alpha A \\ -\dfrac{\Delta t}{4}\alpha I & I \end{bmatrix} \chi^{n+1} = \begin{bmatrix} I & \dfrac{-\Delta t}{4}\alpha A \\ \dfrac{\Delta t}{4}\alpha I & I \end{bmatrix} \tilde{\chi}^{n+1}. \tag{2.90}$$

We can rewrite (2.88)-(2.90) as,

$$\begin{cases} \text{(i)} & (\mathcal{I} + \dfrac{\Delta t}{4}\alpha\mathcal{A})\chi^{n+1/2} = (\mathcal{I} - \dfrac{\Delta t}{4}\alpha\mathcal{A})\chi^n, \\ \text{(ii)} & (\mathcal{I} + \dfrac{\Delta t}{2}\beta\mathcal{A})\tilde{\chi}^{n+1} = (\mathcal{I} - \dfrac{\Delta t}{2}\beta\mathcal{A})\chi^{n+1/2}, \\ \text{(iii)} & (\mathcal{I} + \dfrac{\Delta t}{4}\alpha\mathcal{A})\chi^{n+1} = (\mathcal{I} - \dfrac{\Delta t}{4}\alpha\mathcal{A})\tilde{\chi}^{n+1}. \end{cases} \tag{2.91}$$

Let us define

$$\mathcal{A}_\alpha = (\mathcal{I} + \frac{\Delta t}{4}\alpha\mathcal{A})^{-1}(\mathcal{I} - \frac{\Delta t}{4}\alpha\mathcal{A}),\ \mathcal{A}_\beta = (\mathcal{I} + \frac{\Delta t}{2}\beta\mathcal{A})^{-1}(\mathcal{I} - \frac{\Delta t}{2}\beta\mathcal{A}). \tag{2.92}$$

Eliminating the intermediate variables $\chi^{n+1/2}$ and $\tilde{\chi}^{n+1}$, we get

$$\chi^{n+1} = \mathcal{A}_\alpha \mathcal{A}_\beta \mathcal{A}_\alpha \chi^n. \tag{2.93}$$

The discrete analogues of (2.11) and (2.26) are given by

$$\chi^n = \mathcal{A}_\alpha^n \mathcal{A}_\beta^n \mathcal{A}_\alpha^n \chi_0, \tag{2.94}$$

$$\chi_i^n = \left(\frac{1 - \frac{\Delta t}{2}\beta\xi_i}{1 + \frac{\Delta t}{2}\beta\xi_i} \right)^n \left(\frac{1 - \frac{\Delta t}{4}\alpha\xi_i}{1 + \frac{\Delta t}{4}\alpha\xi_i} \right)^{2n} \chi_{0i}. \tag{2.95}$$

To study the accuracy of the symmetrized scheme we introduce the rational function

$$\mathcal{R}(\zeta) = \left(\frac{1 - \beta\frac{\zeta}{2}}{1 + \beta\frac{\zeta}{2}} \right) \left(\frac{1 - \alpha\frac{\zeta}{4}}{1 + \alpha\frac{\zeta}{4}} \right)^2. \tag{2.96}$$

Expanding $\mathcal{R}(\zeta)$ as a Taylor series in the neighborhood of $\zeta = 0$ we get

$$\mathcal{R}(\zeta) = 1 - (\alpha+\beta)\zeta + (\alpha+\beta)^2\frac{\zeta^2}{2} - ((\alpha+\beta)(\alpha^2+\alpha\beta+\beta^2) - \frac{\alpha^3}{4})\frac{\zeta^3}{4} + \zeta^4\mathcal{O}(1). \tag{2.97}$$

Comparing (2.97) and (2.50) we observe that the symmetrized scheme for the time discretization given in (2.84)-(2.86) is *second order accurate* for any pair $\{\alpha, \beta\}$ satisfying $\alpha + \beta = 1, 0 \leq \alpha, \beta \leq 1$.

2.4.3 An Example with Commuting Suboperators

We apply the symmetrized splitting scheme (2.88)-(2.90) to the time integration of problem (2.52). Figure 2.8 plots the exact solution (2.53) with the solution computed using the symmetrized scheme. Here the decomposition of the operator A is the same as in Section 2.3.3.

In Table 2.3 the ratio of successive entries for $\Delta t = \tau$ and $\Delta t = \tau/2$ is approximately 4.0. This demonstrates the second order accuracy of the symmetrized scheme. Figure 2.9 plots $\log_{10}|\phi_E - \phi_C|$, as a function of time, for different values of α, with a fixed

$\Delta t = 0.01$. It can be seen that the plot for $\alpha = 0.75$ has the smallest errors. Figure 2.10 plots $\log_{10} |\phi_E - \phi_C|$, as a function of time, for different values of Δt, with a fixed $\alpha = 0.5$. The second order temporal behavior of the scheme can be clearly seen in this plot.

	α				
Δt	0.0	0.25	0.5	0.75	1.0
$1.0e-2$	$3.93e-8$	$1.67e-8$	$6.13e-9$	$4.75e-9$	$9.81e-9$
$5.0e-3$	$9.75e-9$	$4.15e-9$	$1.52e-9$	$1.18e-9$	$2.44e-9$
$2.5e-3$	$2.43e-9$	$1.03e-9$	$3.79e-10$	$2.94e-10$	$6.07e-10$
$1.25e-3$	$6.06e-10$	$2.58e-10$	$9.47e-11$	$7.33e-11$	$1.52e-10$

Table 2.3: Comparison of the relative errors between the exact solution and the computed solution using the symmetrized operator splitting scheme for the case of commuting suboperators.

2.4.4 An Example with Noncommuting Suboperators

We demonstrate the second order accuracy of the symmetrized operator splitting scheme for the second order ODE (2.61). We will apply the symmetrized scheme (2.84)-(2.86) to the time integration of problem (2.61). The decomposition of the operator A is the same as in Section 2.3.5.

Figure 2.14 compares the exact solution and the solution computed for $\Delta t = 0.01$ over the time interval $[0, 1]$ for two cases, $\alpha' = 0.0$ (top), and $\alpha' = 0.5$ (bottom). As opposed to the case of the two-step scheme, the case $\alpha' = 0.0$ for the symmetrized scheme also demonstrates second order temporal behavior. Table 2.4 presents the comparison of the relative errors between the exact solution (2.62) and the computed solution obtained by applying the symmetrized scheme as mentioned above. From Table 2.4, we observe that

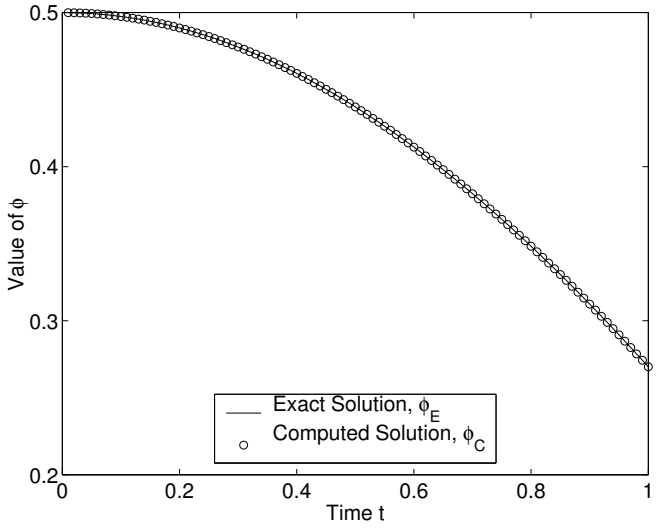

Figure 2.8: Comparison of the exact and computed solutions ($\Delta t = 0.01$), using the symmetrized operator splitting scheme for the case of commuting suboperators.

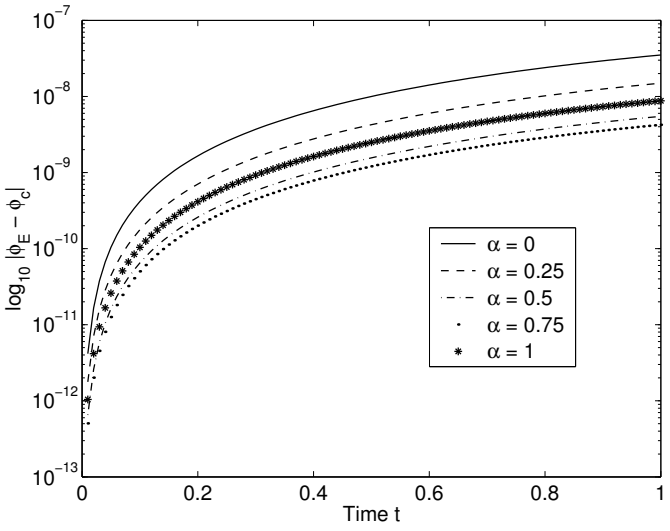

Figure 2.9: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of $\alpha(\Delta t = 0.01)$, for the symmetrized operator splitting scheme in the case of commuting suboperators.

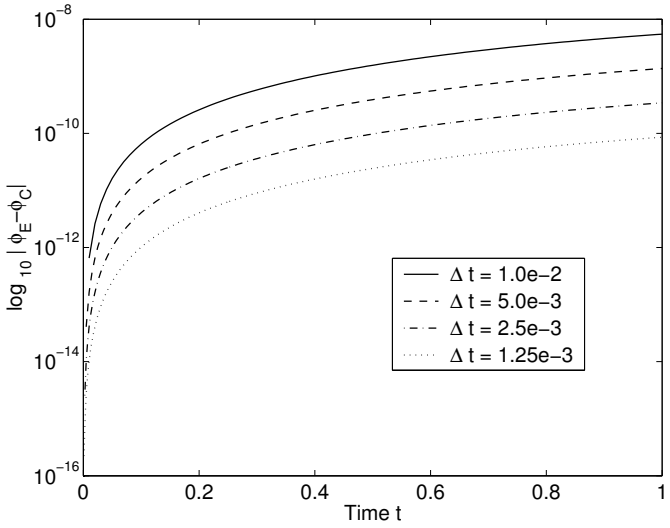

Figure 2.10: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of Δt ($\alpha = 0.5$), for the symmetrized operator splitting scheme for the case of commuting suboperators.

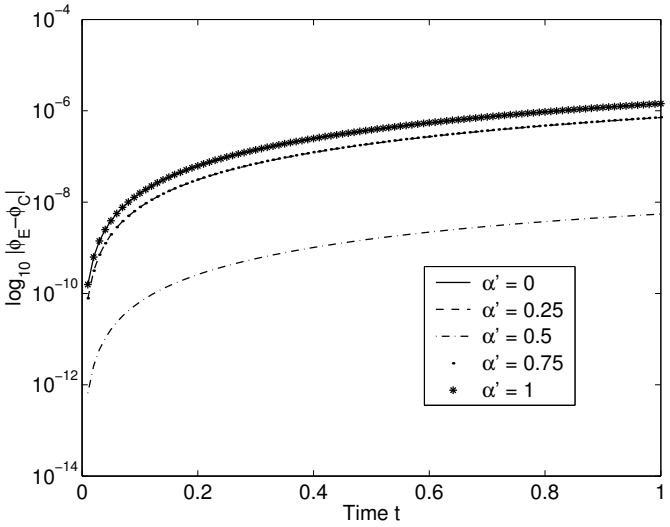

Figure 2.11: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of α' ($\Delta t = 0.01$), for the symmetrized operator splitting scheme in the case of noncommuting suboperators.

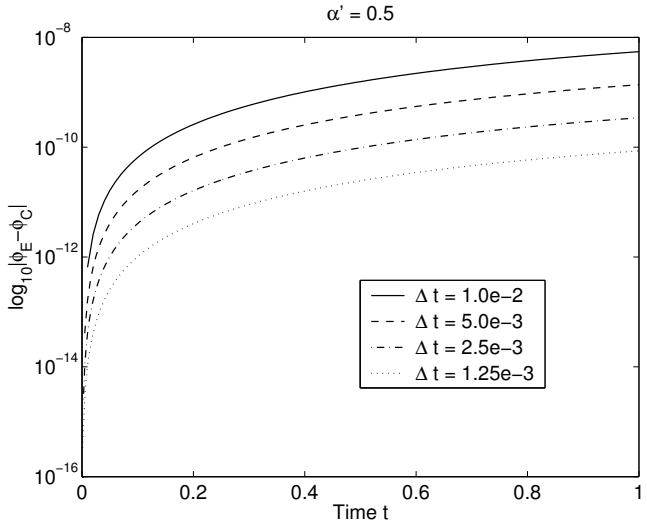

Figure 2.12: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of Δt, for the symmetrized operator splitting scheme in the case of noncommuting suboperators.

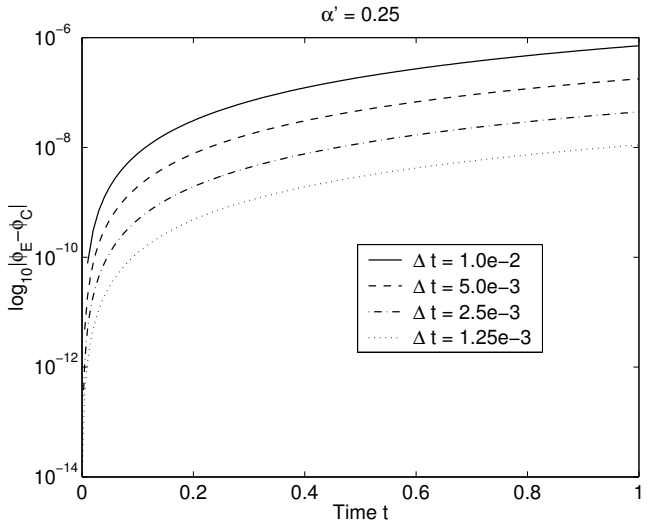

Figure 2.13: A logarithmic plot of the error $|\phi_E - \phi_C|$ as a function of time, for different values of Δt, for the symmetrized operator splitting scheme for the case of noncommuting suboperators.

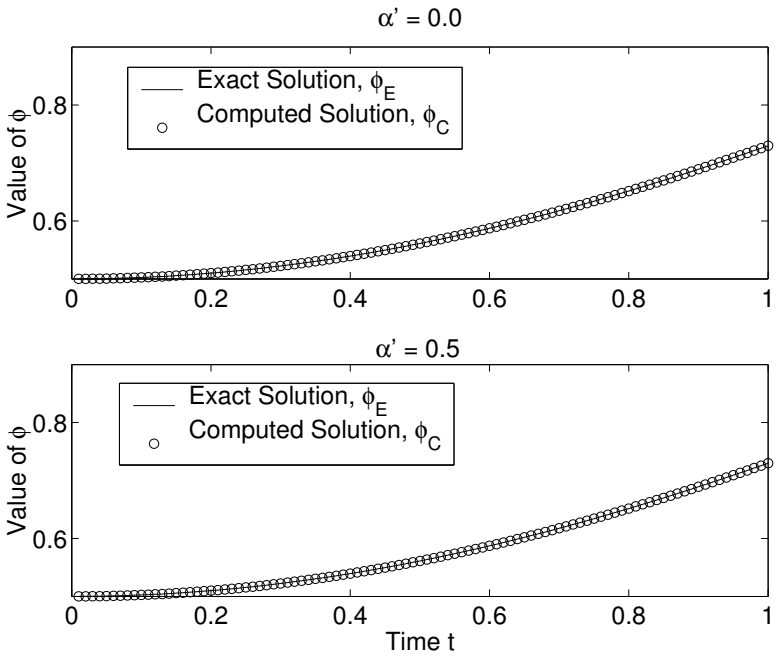

Figure 2.14: Comparison of the exact and computed ($\Delta t = 0.01$), solutions using the symmetrized operator splitting scheme in the case of noncommuting suboperators.

	α'				
Δt	0.0	0.25	0.5	0.75	1.0
$1.0e-2$	$1.13e-6$	$5.64e-7$	$4.46e-9$	$5.73e-7$	$1.14e-6$
$5.0e-3$	$2.82e-7$	$1.40e-7$	$1.11e-9$	$1.43e-7$	$2.84e-7$
$2.5e-3$	$7.03e-8$	$3.50e-8$	$2.77e-10$	$3.55e-8$	$7.08e-8$
$1.25e-3$	$1.75e-8$	$8.74e-9$	$6.92e-11$	$8.88e-9$	$1.77e-8$

Table 2.4: Comparison of the relative errors between the exact solution and the computed solution using the symmetrized operator splitting scheme in the case of noncommuting suboperators.

the symmetrized scheme remains second order accurate in time, for all values of the parameter α', even though the suboperators A_1 and A_2 do not commute.

Figure 2.11 plots $\log_{10}|\phi_E - \phi_C|$, as a function of time, for different values of α', with a fixed $\Delta t = 0.01$. It can be seen that the plot for $\alpha' = 0.5$ has the smallest errors. Figure 2.12 plots $\log_{10}|\phi_E - \phi_C|$, as a function of time, for different values of Δt, with a fixed $\alpha' = 0.5$. The second order temporal behavior of the scheme can be clearly seen in this plot. Figure 2.13 plots $\log_{10}|\phi_E - \phi_C|$ for different values of Δt, with a fixed $\alpha' = 0.25$. As opposed to the two-step scheme which stays first order accurate for this case, the second order temporal behavior of the symmetrized scheme can be clearly seen in this plot.

2.4.5 A Problem with a Nonlinearity

We include a third example in which a nonlinearity has been added. Let us consider the initial value problem

$$\begin{cases} \dfrac{d^2\phi}{dt^2} + \phi + \phi^3 - 1 = 0, \\ \\ \phi(0) = 1/2,\ \phi_t(0) = 0. \end{cases} \tag{2.98}$$

$\phi(t) \in \mathbb{R},\ \forall t \geq 0$. The nonlinear operator A is defined by

$$A\psi = \psi + \psi^3 - 1,\ \forall \psi \in C^2(\mathbb{R}^+; \mathbb{R}) \tag{2.99}$$

We will demonstrate the second order accuracy of the symmetrized splitting scheme (2.84)-(2.86) to the time integration of problem (2.98) on the interval $[0, 1]$, with the suboperators being defined as

$$A_1\psi = \psi^3, A_2\psi = \psi - 1. \tag{2.100}$$

In this case as well, the suboperators A_1 and A_2 do not commute. For comparison, we calculate a *reference solution* ϕ_R of (2.98) by applying to it an explicit scheme that is second order accurate in time. For $t \in [0, 1]$, the reference solution is given by $\phi_R^{n+1} \approx \phi_R((n+1)\Delta t)$ calculated via

$$\begin{cases} \dfrac{\phi_R^{n+1} - 2\phi_R^n + \phi_R^{n-1}}{(\Delta t)^2} + \phi_R^n + (\phi_R^n)^3 - 1 = 0, \\ \phi_R^0 = 1/2, \dfrac{\phi_R^1 - \phi_R^{-1}}{\Delta t} = 0. \end{cases} \tag{2.101}$$

Table 2.5 presents the comparison of the relative errors, between the reference solution, computed via (2.101) with $\Delta t = 3.90625e-5$, and the computed solution over the time interval $[0, 1]$, for $\alpha = 0, 0.25, 0.5, 0.75, 1.0$. As can be seen from the table, the symmetrized splitting scheme is second order accurate in time.

Figure 2.15 compares the reference solution calculated using $\Delta t = 3.90625e - 5$ with the computed solution using $\Delta t = 0.01$. Figure 2.16 plots $\log_{10} |\phi_R - \phi_C|$, as a function of time, for different values of α, with a fixed $\Delta t = 0.01$. Figure 2.17 plots $\log_{10} |\phi_R - \phi_C|$, as a function of time, for different values of Δt, with a fixed $\alpha = 0.5$. Again, the second order temporal behavior of the scheme can be clearly seen in this plot.

	α				
Δt	0.0	0.25	0.5	0.75	1.0
$1.0e-2$	$2.10e-6$	$1.19e-6$	$3.67e-7$	$3.92e-7$	$1.05e-6$
$5.0e-3$	$5.23e-7$	$2.96e-7$	$9.10e-8$	$9.78e-8$	$2.62e-7$
$2.5e-3$	$1.30e-7$	$7.38e-8$	$2.26e-8$	$2.45e-8$	$6.55e-8$
$1.25e-3$	$3.25e-8$	$1.84e-8$	$5.58e-9$	$6.18e-9$	$1.64e-8$

Table 2.5: Comparison of the relative errors between the reference solution and the computed solution using the symmetrized operator splitting scheme for the nonlinear problem.

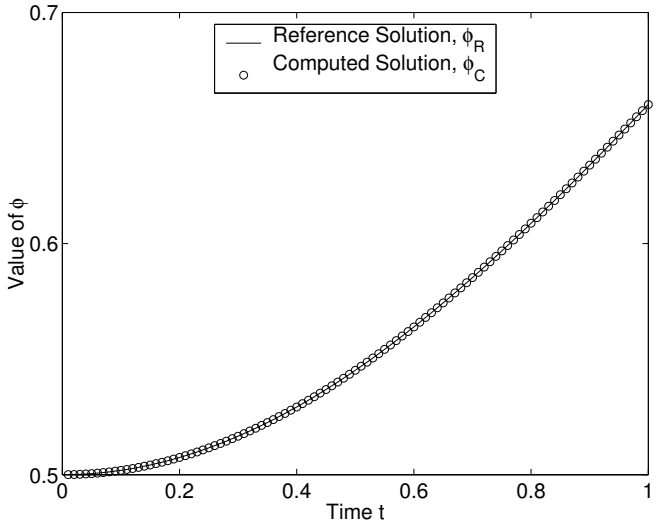

Figure 2.15: Comparison of the reference and computed solutions ($\Delta t = 0.01$), using the symmetrized operator splitting scheme for the nonlinear problem.

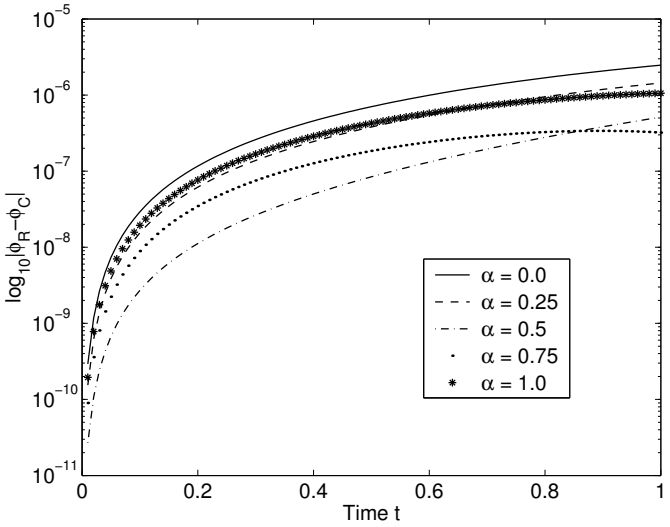

Figure 2.16: A logarithmic plot of the error $|\phi_R - \phi_C|$ as a function of time, for different values of α ($\Delta t = 0.01$), using the symmetrized operator splitting scheme applied to the nonlinear problem.

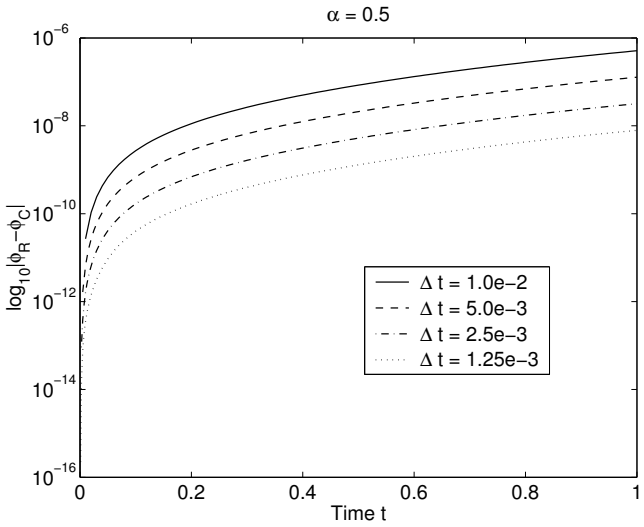

Figure 2.17: A logarithmic plot of the error $|\phi_R - \phi_C|$ as a function of time, for different values of Δt ($\alpha = 0.5$), using the symmetrized operator splitting scheme applied to the nonlinear problem.

Chapter 3

Fictitious Domains, Operator Splitting, and Mixed Finite Elements for Wave Problems

3.1 Introduction

In this chapter we introduce three novel methods for the numerical solution of a wave scattering problem. The first method is a fictitious domain approach that uses distributed Lagrange multipliers. A similar approach has been used in the case of incompressible viscous flow around moving rigid bodies [63, 68]. A fictitious domain approach utilizing boundary Lagrange multipliers for the time dependent scattering problem was introduced in [27, 41, 57]. We introduce two methods which combine a symmetrized operator splitting scheme for time discretization along with a fictitious domain method involving distributed multipliers. Based on the symmetrized scheme encountered in Chapter 2, the symmetrized operator splitting schemes, introduced in this chapter, decouple the propagation of the wave, and the enforcement of the Dirichlet condition on the boundary of the obstacle. The first method is a fully conforming scheme, which uses continuous finite ele-

ments. The second method uses mixed finite elements in space-time for the substeps which propagate the wave.

A brief outline of this chapter is as follows. In Section 3.2 we formulate the time dependent scattering problem. In Section 3.3 we present a new fictitious domain method utilizing distributed multipliers for the solution of our wave problem. We also present results related to the conservation of energy associated with this formulation. In Section 3.4 we present a fully conforming finite element method for the numerical solution of the wave problem. We perform a stability analysis, and we study the energy conservation properties of this numerical model. In Section 3.5 we describe a symmetrized operator splitting scheme for the numerical solution of the fictitious domain model presented in Section 3.3. In Sections 3.6 and 3.7 we present a mixed finite element scheme for the numerical solution of the wave problem, and then combine it with the operator splitting scheme presented in Section 3.5. In Section 3.8 we present some numerical examples to validate the proposed methods.

3.2 Formulation of a Model Wave Problem

We consider the scalar wave equation with constant coefficients. We are interested in studying the scattering of a wave by an obstacle $\omega \subset \mathbb{R}^d$ with $d = 2$ (or $d = 3$). Let c denote the speed of propagation. The scalar evolution problem can be set up as

$$\begin{cases} \text{(i)} & \dfrac{1}{c^2}\dfrac{\partial^2 \Phi}{\partial t^2} - \Delta \Phi = 0, \text{ in } \Omega \setminus \bar{\omega}, \\ \text{(ii)} & \Phi = G, \text{ on } \partial\omega, \\ \text{(iii)} & \dfrac{1}{c}\dfrac{\partial \Phi}{\partial t} + \dfrac{\partial \Phi}{\partial \mathbf{n}} = 0, \text{ on } \Gamma = \partial\Omega, \\ \text{(iv)} & \Phi(0) = \Phi_0, \dfrac{\partial \Phi}{\partial t}(0) = \Phi_1. \end{cases} \tag{3.1}$$

As shown in Figure 3.1, ω is the obstacle which is embedded in the larger bounded domain

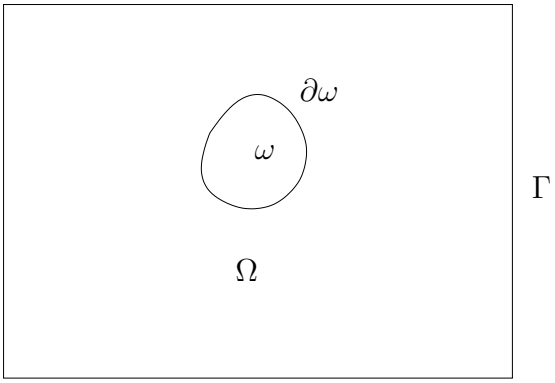

Figure 3.1: The obstacle ω embedded inside the larger domain Ω.

Ω of $\mathbb{R}^2$. We would like to study problem (3.1) in the case of an unbounded exterior domain. One of the ways of simulating the scattering problem in an unbounded domain is to impose an absorbing boundary condition on the boundary of the truncated domain Ω. We impose a first order absorbing boundary condition (3.1, iii) on the (artificial) boundary Γ. In Chapter 5 we will study more sophisticated absorbing boundary conditions like *perfectly matched layers* for wave propagation problems.

3.3 A Fictitious Domain Formulation for the Wave Problem

One of the techniques used to solve time dependent problems of scattering by an obstacle is the finite difference method, which uses a rectangular grid and an explicit scheme in time. This method is computationally very efficient, however the staircase approximation to the obstacle is inaccurate, and it leads to excessive numerical diffraction when the obstacle boundary does not fit the mesh, as seen in Figure 3.2. In this figure the scattering obstacle is a disk, and is approximated by the darkened nodal points.

An alternative and more accurate approximation is given by the fictitious domain

method (FDM) based on Lagrange multipliers. This technique, which was developed to handle problems with complex geometries in the stationary case [10, 66], has recently been applied to time dependent problems [41, 57, 67].

The idea behind the fictitious domain method is to extend the solution Φ inside the obstacle ω, and solve the wave equation (3.1) in the entire domain Ω, which has a simple shape like a square or rectangle [27, 41, 57, 67]. The Dirichlet condition on $\partial\omega$ is enforced via the introduction of a Lagrange multiplier. In [27, 57] a boundary multiplier fictitious domain method is introduced for the wave equation, and for Maxwell's equations. In this chapter, we present a distributed multiplier fictitious domain method for the wave problem (3.1).

Let g be an H^1 - extension of G on ω; using a distributed Lagrange multiplier approach, problem (3.1) is equivalent to the variational one

Find $\{\phi(t), \lambda(t)\} \in H^1(\Omega) \times L^2(\omega)$ such that:

$$\begin{cases} \text{(i)} & \dfrac{1}{c^2}\displaystyle\int_\Omega \frac{\partial^2\phi}{\partial t^2} w \,\mathrm{dx} + \int_\Omega \nabla\phi\cdot\nabla w \,\mathrm{dx} + \frac{1}{c}\int_\Gamma \frac{\partial\phi}{\partial t} w \,\mathrm{d\Gamma} \\ & \quad + \displaystyle\int_\omega \lambda w \,\mathrm{d}\omega = 0, \quad \forall w \in H^1(\Omega), \\ \text{(ii)} & \displaystyle\int_\omega (\phi - g)\mu \,\mathrm{d}\omega = 0, \quad \forall \mu \in L^2(\omega), \\ \text{(iii)} & \phi(0) = \phi_0,\ \dfrac{\partial\phi}{\partial t}(0) = \phi_1, \end{cases} \tag{3.2}$$

in the sense that

$$\phi = \begin{cases} \Phi & \text{on} \quad \Omega \setminus \bar{\omega}, \\ G & \text{on} \quad \partial\omega. \end{cases} \tag{3.3}$$

The function ϕ_0 is chosen to be a H^1 - extension of Φ_0, and ϕ_1 to be at least an L^2-extension of Φ_1. Thus, we have

$$\phi(0) = \phi_0 = \begin{cases} \Phi_0, & \text{on} \quad \Omega \setminus \bar{\omega}, \\ 0, & \text{on} \quad \omega. \end{cases} \quad , \quad \frac{\partial\phi}{\partial t}(0) = \phi_1 = \begin{cases} \Phi_1, & \text{on} \quad \Omega \setminus \bar{\omega}, \\ 0, & \text{on} \quad \omega. \end{cases} \tag{3.4}$$

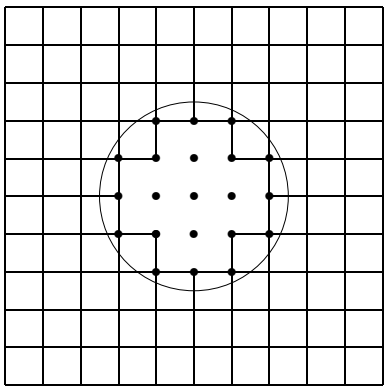

Figure 3.2: A staircase approximation to a scattering disk. The disk is approximated by the highlighted nodal points.

Remark 7 *We note that the first order absorbing boundary condition, (3.1,* iii*), is incorporated into the variational formulation (3.2,* i*), and hence does not have to be explicitly imposed in the functional space chosen for the solution ϕ.*

3.3.1 Conservation of Energy

In this section we derive an *energy identity* from the variational formulation (3.2). The energy identity presented below guarantees the well-posedness of the problem, and the stability of the solution.

Theorem 1 *Let us assume that $g = 0$ in (3.2,* ii*). Then, system* (3.2) *verifies the following energy identity*

$$\frac{d}{dt}\mathcal{E} = -\frac{1}{c}\left\|\frac{\partial\phi}{\partial t}\right\|^2_{L^2(\Gamma)}, \tag{3.5}$$

where the energy $\mathcal{E}$ is defined as

$$\mathcal{E} = \frac{1}{2}\left\{\frac{1}{c^2}\left\|\frac{\partial\phi}{\partial t}\right\|^2_{L^2(\Omega)} + \|\nabla\phi\|^2_{L^2(\Omega)}\right\}, \tag{3.6}$$

with

$$\|\cdot\|_{L^2(\Gamma)} = \left(\int_\Gamma |\cdot|^2 \,\mathrm{d}\Gamma\right)^{1/2}, \text{ and } \|\cdot\|_{L^2(\Omega)} = \left(\int_\Omega |\cdot|^2 \,\mathrm{dx}\right)^{1/2}. \tag{3.7}$$

Thus, (3.5) *implies that the energy does not grow over time, i.e.,*

$$\mathcal{E}(t) \leq \mathcal{E}(0),\ \forall t > 0. \tag{3.8}$$

Proof 1 : *Let us take* $w = \dfrac{\partial \phi}{\partial t}$ *in (3.2,* i*). We obtain*

$$\frac{1}{c^2}\int_\Omega \frac{\partial^2\phi}{\partial t^2}\frac{\partial\phi}{\partial t}\,\mathrm{dx} + \int_\Omega \nabla\phi\cdot\nabla\frac{\partial\phi}{\partial t}\,\mathrm{dx} + \frac{1}{c}\int_\Gamma \left|\frac{\partial\phi}{\partial t}\right|^2 \,\mathrm{d}\Gamma + \int_\omega \lambda\frac{\partial\phi}{\partial t}\,\mathrm{d}\omega = 0. \tag{3.9}$$

This gives us

$$\frac{1}{2}\frac{d}{dt}\left\{\frac{1}{c^2}\left\|\frac{\partial\phi}{\partial t}\right\|^2_{L^2(\Omega)} + \|\nabla\phi\|^2_{L^2(\Omega)}\right\} + \frac{1}{c}\int_\Gamma \left|\frac{\partial\phi}{\partial t}\right|^2 \,\mathrm{d}\Gamma + \int_\omega \lambda\frac{\partial\phi}{\partial t}\,\mathrm{d}\omega = 0. \tag{3.10}$$

From (3.2, ii*), since* $g = 0$*, differentiating with respect to time we have*

$$\int_\omega \frac{\partial\phi}{\partial t}\mu\,\mathrm{d}\omega = 0,\ \forall \mu \in L^2(\omega). \tag{3.11}$$

Taking $\mu = \lambda$ *in* (3.11) *we get*

$$\int_\omega \frac{\partial\phi}{\partial t}\lambda\,\mathrm{d}\omega = 0. \tag{3.12}$$

Substituting (3.12) *in* (3.10)*, and using the definition of the energy* (3.6) *we have*

$$\frac{d}{dt}\mathcal{E} = -\frac{1}{c}\left\|\frac{\partial\phi}{\partial t}\right\|^2_{L^2(\Gamma)}. \tag{3.13}$$

Equation (3.13) implies that there is no dissipation of the waves in the domain Ω. This is the principle of *conservation of energy* for the wave equation.

3.4 A Fully Conforming Method for the Numerical Solution of the Wave Problem

3.4.1 Time discretization

We will use a centered finite difference scheme for the time discretization of the wave problem. On the interval $[0, T]$, let $\Delta t = T/N$ be the time step, where $N \in \mathbb{N}$. Define

$\Phi^k \approx \Phi(k\Delta t)$ and denote $t^k = k\delta t$.

For $n = 0, 1, \ldots, N-1$, on the interval (t^n, t^{n+1}), given ϕ^n, ϕ^{n-1} we will solve the problem

Find $(\phi^{n+1}, \lambda^{n+1}) \in H^1(\Omega) \times L^2(\omega)$ such that:

$$\begin{cases} \text{(i)} & \dfrac{1}{c^2}\displaystyle\int_\Omega \frac{\phi^{n+1} - 2\phi^n + \phi^{n-1}}{\Delta t^2} w \, \mathrm{dx} + \int_\Omega \nabla\phi^n \cdot \nabla w \, \mathrm{dx} + \int_\omega \lambda^{n+1} w \, \mathrm{d}\omega \\ & + \dfrac{1}{c}\displaystyle\int_\Gamma \frac{\phi^{n+1} - \phi^{n-1}}{2\Delta t} w \, \mathrm{d}\Gamma = 0 \ , \ \forall w \in H^1(\Omega), \\ \text{(ii)} & \displaystyle\int_\omega (\phi^{n+1} - g^{n+1})\mu \, \mathrm{d}\omega = 0 \ , \ \forall \mu \in L^2(\omega), \\ \text{(iii)} & \phi^0 = \phi_0, \ \phi^1 - \phi^{-1} = 2\Delta t \phi_1. \end{cases} \tag{3.14}$$

3.4.2 Finite Element Approximation of the Wave Problem

We divide Ω into elementary rectangles, and consider $\mathcal{T}_h$ to be a regular mesh with elements $\{K\}$ of edge length h. We define the finite dimensional space

$$\mathbf{V}_h = \{v_h \mid v_h \in C^0(\bar{\Omega}), v_h|_K \in Q_1, \forall \ K \in \mathcal{T}_h\}, \tag{3.15}$$

which approximates $H^1(\Omega)$. In (3.15), the space Q_1 is defined as

$$Q_1 = P_{11}, \tag{3.16}$$

where, for $k_1, k_2 \in \mathbb{N} \cup \{0\}$

$$P_{k_1 k_2} = \{p(x_1, x_2) | p(x_1, x_2) = \sum_{0 \le i \le k_1} \sum_{0 \le j \le k_2} a_{ij} x_1^i x_2^j, a_{ij} \in \mathbb{R}\}. \tag{3.17}$$

Thus, P_{11} is the space of continuous bilinear functions, and $\mathbf{V}_h$ is the space of continuous piecewise bilinear functions. As $\phi \in H^1(\Omega)$ (which is its natural space), we will choose the space $\mathbf{V}_h$ for the finite element approximation ϕ_h of ϕ. We will use quadrature rules for the calculation of the integrals

$$\int_\Omega v_h \, w_h \, \mathrm{dx} = \sum_{K \in \mathcal{T}_h} \int_K v_h \, w_h \, \mathrm{dx} \ , \forall \ v_h, w_h \in \mathbf{V}_h, \tag{3.18}$$

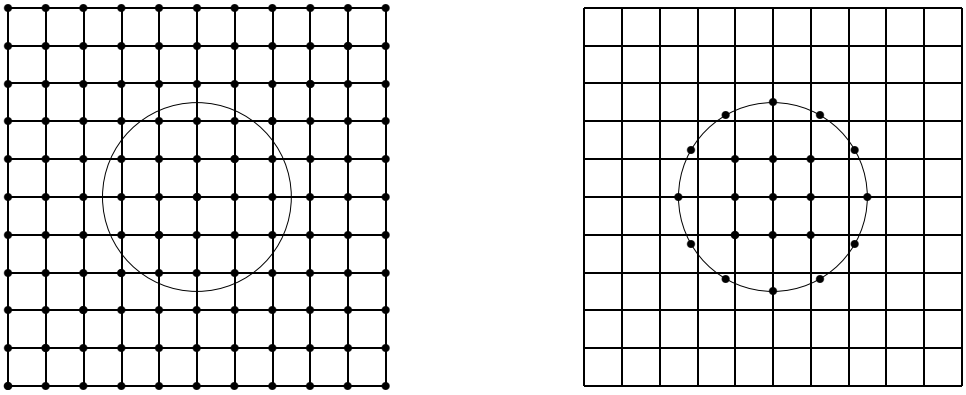

Figure 3.3: The degrees of freedom for the solution ϕ (left), and the degrees of freedom, $\Sigma_h^{\bar{\omega}}$, for the Lagrange multiplier λ (right) in the fictitious domain method, in the case of a scattering disk. The mesh ratio, i.e., the ratio of the step size chosen on the obstacle to the mesh step size, is about 1.3.

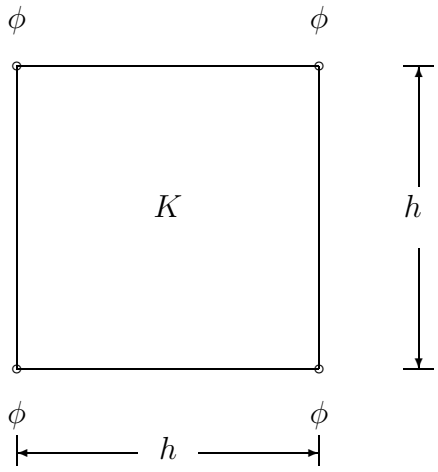

Figure 3.4: A sample domain element K. The degrees of freedom for ϕ are at the vertices of the square.

due to which we obtain a diagonal mass matrix and thus an explicit scheme in time. The use of quadrature formulas to obtain diagonal mass matrices is referred to as *mass-lumping*. Similarly, we use quadrature rules to calculate the boundary integral

$$\int_{\Gamma} v_h \, w_h \, \mathrm{d}\Gamma. \tag{3.19}$$

Let the set of mesh points on $\bar{\Omega}$ be defined as

$$\Sigma_h = \{P \mid P \in \bar{\Omega}, P \text{ is a vertex of } \mathcal{T}_h\}. \tag{3.20}$$

Next, we define the set

$$\Sigma_h^{\bar{\omega}} = \{P | P \in \bar{\omega},\ d(P, \partial\omega) \geq h\} \cup \text{Discrete set of points belonging to } \partial\omega. \tag{3.21}$$

The points on $\partial\omega$ are typically chosen so that their distance is of the order of h. Using the sets defined above, we now define the set $\mathbf{\Lambda}_h$ of the Lagrange multipliers by

$$\mathbf{\Lambda}_h = \{\mu_h \mid \mu_h = \sum_{P \in \Sigma_h^{\bar{\omega}}} \mu_P \chi_P,\ \mu_P \in \mathbb{R}\}, \tag{3.22}$$

with χ_P the characteristic function of the elementary square of center P and edge length h; we clearly have $\mu_h(P) = \mu_P$. We approximate the integrals involving the distributed multiplier by

$$\int_{\omega} \mu_h \, v_h \, \mathrm{dx} \approx h^2 \sum_{P \in \Sigma_h^{\bar{\omega}}} \mu_h(P) \, v_h(P),\ \forall v_h \in \mathbf{V}_h,\ \forall \mu_h \in \mathbf{\Lambda}_h. \tag{3.23}$$

Figure 3.3 illustrates the degrees of freedom for the solution ϕ (left), and a choice for the set $\Sigma_h^{\bar{\omega}}$ in the case of a scattering disk. The ratio of the distance between points on the circle, denoted by $h_{\partial\omega}$, to the mesh step size, h, is about 1.3. We will call this ratio as the *mesh ratio*. In numerical experiments, good results are observed when the mesh ratio is approximately 1.5 or greater [61]. Figure 3.4 represents a sample square domain element in the discretized mesh. The edge length is h, and the degrees of freedom for the solution ϕ are at the vertices of the square.

Using the above definitions, a fully discretized scheme for the wave problem is given by the following fully discrete variational formulation

• **Scheme** FDDM:

Find $(\phi_h^{n+1}, \lambda_h^{n+1}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\left\{\begin{array}{ll} \text{(i)} & \dfrac{1}{c^2}\displaystyle\int_\Omega \frac{\phi_h^{n+1}-2\phi_h^n+\phi_h^{n-1}}{\Delta t^2} w_h \,\mathrm{dx} + \int_\Omega \nabla\phi_h^n \cdot \nabla w_h \,\mathrm{dx} \\ & + \dfrac{1}{c}\displaystyle\int_\Gamma \frac{\phi_h^{n+1}-\phi_h^{n-1}}{2\Delta t} w_h \,\mathrm{d\Gamma} + \int_\omega \lambda_h^{n+1} w_h \,\mathrm{d}\omega = 0 \,, \forall\, w_h \in \mathbf{V}_h, \\ \text{(ii)} & \displaystyle\int_\omega (\phi_h^{n+1} - g^{n+1})\mu_h \,\mathrm{d}\omega = 0 \,, \forall\, \mu_h \in \mathbf{\Lambda}_h, \\ \text{(iii)} & \phi_h^0 = \phi_0,\ \phi_h^1 - \phi_h^{-1} = 2\Delta t \phi_1. \end{array}\right. \tag{3.24}$$

3.4.3 Iterative Solution of the Discrete Problem

For the solution of the system (3.24), at each time step we have to solve a system of linear equations of the form

$$\left\{\begin{array}{lcl} D_h\phi^{n+1} + B_h^T\lambda^{n+1} &=& a, \\ B_h\phi^{n+1} &=& b, \end{array}\right. \tag{3.25}$$

where $D_h \in \mathbb{R}^{N\times N}$ is symmetric positive definite, and $B_h \in \mathbb{R}^{M\times N} (M << N)$. We use the *Schur Complement* of the system (3.25)

$$(B_h D_h^{-1} B_h^T)\lambda^{n+1} = B_h D_h^{-1} a - b, \tag{3.26}$$

to solve for λ^{n+1}. We do this by using a conjugate gradient algorithm in the form given by Glowinski and LeTallec [65] which we present below.

Algorithm 1 *:* **An Uzawa-Type Conjugate Gradient Algorithm**

- 0 $\hat{\lambda}_0$ *is given.*

  0.1) Solve $D_h\hat{\phi}_0 = b - B_h^T\hat{\lambda}_0$.

  0.2) Compute $\hat{g}_0 = c - B_h\hat{\phi}_0$.

  0.3) Set $\hat{w}_0 = \hat{g}_0$.

- 1 *For* $k = 0, 1, 2, \ldots$ *until convergence:*

1.1) Solve $D_h\hat{z}_k = B_h^T\hat{w}_k$.

1.2) $\rho_k = \dfrac{\|\hat{g}_k\|^2}{(B_h\hat{z}_k, \hat{w}_k)}$.

1.3) $\hat{\lambda}_{k+1} = \hat{\lambda}_k - \rho_k\hat{w}_k$.

1.4) $\hat{\phi}_{k+1} = \hat{\phi}_k + \rho_k\hat{z}_k$.

1.5) $\hat{g}_{k+1} = \hat{g}_k + \rho_k B_h\hat{z}_k$.

1.6) If $\dfrac{\|\hat{g}_{k+1}\|^2}{\|\hat{g}_0\|^2} \leq \epsilon$, *(Test of Convergence)*

take $\lambda^{n+1} = \hat{\lambda}_{k+1}$, $\phi^{n+1} = \hat{\phi}_{k+1}$, *Stop.*

If not, proceed to step 1.7.

1.7) $\gamma_k = \dfrac{\|\hat{g}_{k+1}\|^2}{\|\hat{g}_k\|^2}$.

1.8) $\hat{w}_{k+1} = \hat{g}_{k+1} + \gamma_k\hat{w}_k$.

$k = k + 1$.

3.4.4 Stability Analysis and Conservation of Energy

Analogous to the continuous case, we derive a *discrete* energy identity based on the discrete variational formulation (3.24). Using the discrete energy identity we show that the fictitious domain method is stable, with the Courant - Friedrichs - Lewy (CFL) condition being the same as in the case of the problem without an obstacle. We will assume that $g = 0$ in (3.24, ii). We define the bilinear forms

$$\begin{cases} a(\phi_h, \psi_h) &= \displaystyle\int_\Omega \nabla\phi_h \cdot \nabla\psi_h \, \mathrm{dx} \, , \forall \, (\phi_h, \psi_h) \in \mathbf{V}_h \times \mathbf{V}_h, \\ b(\phi_h, \mu_h) &= \displaystyle\int_\omega \phi_h\mu_h \, \mathrm{d}\omega \, , \forall \, (\phi_h, \mu_h) \in \mathbf{V}_h \times \mathbf{\Lambda}_h. \end{cases} \tag{3.27}$$

Next, we define the operator $\mathcal{A}_h : \mathbf{V}_h \to \mathbf{V}_h'$ by

$$(\mathcal{A}_h\phi_h, \psi_h)_{L^2(\Omega)} = a(\phi_h, \psi_h). \tag{3.28}$$

Theorem 2 *If the Courant - Friedrichs - Lewy (CFL) condition*

$$c\Delta t \leq \frac{h}{\sqrt{2}}, \tag{3.29}$$

is satisfied (in 2D), then the operator

$$\mathcal{S}_h = I - \frac{c^2\Delta t^2}{4}\mathcal{A}_h, \tag{3.30}$$

defines a positive quadratic form, the expression

$$\mathcal{E}_h^{n+1/2} = \frac{1}{2}\left\{\frac{1}{c^2}\left(\frac{\phi_h^{n+1}-\phi_h^n}{\Delta t}, \mathcal{S}_h\frac{\phi_h^{n+1}-\phi_h^n}{\Delta t}\right)_{L^2(\Omega)} + \left\|\nabla\left(\frac{\phi_h^{n+1}+\phi_h^n}{2}\right)\right\|^2_{L^2(\Omega)}\right\} \tag{3.31}$$

defines a discrete energy, and system (3.24) *verifies the energy identity*

$$\mathcal{E}_h^{n+1/2} = \mathcal{E}_h^{n-1/2} - \Delta t\left\|\frac{\phi_h^{n+1}-\phi_h^{n-1}}{2\Delta t}\right\|^2_{L^2(\Gamma)}, \ \forall n \in \mathbb{N}, n \geq 0. \tag{3.32}$$

Thus, (3.32) *implies that the discrete energy does not grow over time, i.e.,*

$$\mathcal{E}_h^{n+1/2} \leq \mathcal{E}_h^{n-1/2}, \ \forall n \geq 0. \tag{3.33}$$

Proof 2 : *Using the definition of the operator $\mathcal{A}_h$ we can rewrite the discrete energy as*

$$\mathcal{E}_h^{n+1/2} = \frac{1}{2}\left\{\frac{1}{c^2}\left\|\frac{\phi_h^{n+1}-\phi_h^n}{\Delta t}\right\|^2_{L^2(\Omega)} + \int_\Omega \nabla\phi_h^{n+1}\cdot\nabla\phi_h^n \,\mathrm{dx}\right\}. \tag{3.34}$$

From (3.24, i*), taking $w_h = \dfrac{\phi_h^{n+1}-\phi_h^{n-1}}{2\Delta t}$, we obtain*

$$\begin{aligned}
&\frac{1}{c^2\Delta t}\int_\Omega\left\{\frac{\phi_h^{n+1}-\phi_h^n}{\Delta t} - \frac{\phi_h^n-\phi_h^{n-1}}{\Delta t}\right\}\left\{\frac{\phi_h^{n+1}-\phi_h^{n-1}}{2\Delta t}\right\}\,\mathrm{dx}\\
&+\frac{1}{2\Delta t}\int_\omega \nabla\phi_h^n(\nabla\phi_h^{n+1}-\nabla\phi_h^{n-1})\,\mathrm{dx} + \frac{1}{2\Delta t}\int_\omega \lambda_h^{n+1}(\phi_h^{n+1}-\phi_h^{n-1})\,\mathrm{d}\omega\\
&= -\frac{1}{c}\left\|\frac{\phi_h^{n+1}-\phi_h^{n-1}}{2\Delta t}\right\|^2_{L^2(\Gamma)}.
\end{aligned} \tag{3.35}$$

Using (3.34) *we can rewrite the above as*

$$\frac{\mathcal{E}_h^{n+1/2}-\mathcal{E}_h^{n-1/2}}{\Delta t} + \frac{1}{2\Delta t}\int_\omega \lambda_h^{n+1}(\phi_h^{n+1}-\phi_h^{n-1})\,\mathrm{d}\omega = -\frac{1}{c}\left\|\frac{\phi_h^{n+1}-\phi_h^{n-1}}{2\Delta t}\right\|^2_{L^2(\Gamma)}. \tag{3.36}$$

Next, from (3.24, ii*) by taking* $\mu_h = \lambda_h^{n+1}$*, and* $\mu_h = \lambda_h^{n+3}$ *we have, respectively,*

$$\int_\omega \lambda_h^{n+1}\phi_h^{n+1}\,\mathrm{dx} = 0, \text{ and } \int_\omega \lambda_h^{n+1}\phi_h^{n-1}\,\mathrm{dx} = 0. \tag{3.37}$$

Substituting in (3.37) *in* (3.36) *we have*

$$\frac{\mathcal{E}_h^{n+1/2} - \mathcal{E}_h^{n-1/2}}{\Delta t} = -\frac{1}{c}\left\|\frac{\phi_h^{n+1} - \phi_h^{n-1}}{2\Delta t}\right\|^2_{L^2(\Gamma)}, \tag{3.38}$$

which gives the energy identity (3.32)*, and implies that the energy does not grow with time.*

It remains to show that the operator $\mathcal{S}_h$ *defines a positive quadratic form under the CFL condition* (3.29)*. In 2 dimensions we have [91],*

$$\sup_{w_h \in \mathbf{V}_h} \frac{h^2(\mathcal{A}_h w_h, w_h)_{L^2(\Omega)}}{4(w_h, w_h)_{L^2(\Omega)}} < 2, \tag{3.39}$$

which implies that

$$\frac{h^2(\mathcal{A}_h w_h, w_h)_{L^2(\Omega)}}{4(w_h, w_h)_{L^2(\Omega)}} < 2, \tag{3.40}$$

and hence

$$(w_h, w_h)_{L^2(\Omega)} > \frac{h^2}{2}\frac{(\mathcal{A}_h w_h, w_h)_{L^2(\Omega)}}{4}. \tag{3.41}$$

Using the CFL condition (3.29)*, we have*

$$(w_h, w_h)_{L^2(\Omega)} > \frac{c^2\Delta t^2}{4}(\mathcal{A}_h w_h, w_h)_{L^2(\Omega)}. \tag{3.42}$$

Simplifying the above, we have

$$(w_h, \mathcal{S}_h w_h)_{L^2(\Omega)} = (w_h, (I - \frac{c^2\Delta t^2}{4}\mathcal{A}_h)w_h)_{L^2(\Omega)} > 0, \ \forall w_h \in \mathbf{V}_h. \tag{3.43}$$

Equation (3.43) *implies that the operator* $\mathcal{S}_h$ *is a positive quadratic form. Thus, the CFL condition assures the stability of the scheme* (3.24)*.*

3.5 An Operator Splitting Scheme

In this section, we describe a symmetrized operator splitting scheme for the numerical solution of the wave problem (3.1). The idea behind operator splitting, in the case of the

scattering problem, is to decouple the operator that propagates the wave, and the operator that enforces the Dirichlet condition on the boundary of the obstacle ω. On a time interval of length Δt, we can construct a two-step splitting scheme by separating the solution of (3.1) into two steps. In one step of time length Δt we will propagate the wave, i.e., solve the wave equation in the whole domain Ω, and in the second step of length Δt we will enforce the Dirichlet condition on $\partial\omega$.

As demonstrated in Chapter 2, the two-step schemes are usually first order accurate in time, whereas the Strang symmetrized operator splitting scheme is second order accurate, even if the suboperators involved do not commute. Thus, in order to obtain second order accurate schemes in time we will construct a Strang symmetrization of the two-step scheme. We can do so in two ways. We perform one of the two steps mentioned above on two intervals of length $\Delta t/2$ separated by the other step performed on an interval of length Δt. This gives rise to two different symmetrized operator splitting schemes.

1. Symmetrized scheme 1:

   - Propagate the wave on an interval of length $\Delta t/2$,
   - Enforce the Dirichlet condition on $\partial\omega$ on an interval of length Δt,
   - Propagate the wave on an interval of length $\Delta t/2$.

2. Symmetrized scheme 2:

   - Enforce the Dirichlet condition on $\partial\omega$ on an interval of length $\Delta t/2$,
   - Propagate the wave on an interval of length Δt,
   - Enforce the Dirichlet condition on $\partial\omega$ on an interval of length $\Delta t/2$.

In the rest of this chapter we will demonstrate the numerical implementation of scheme 2. This operator splitting scheme is based on the formulation (3.24). Let us define the velocity u to be the time derivative

$$u = 2\frac{\partial\phi}{\partial t}. \tag{3.44}$$

We will rewrite the wave equation as a system of first order PDE's by the use of the variable u. This will allow us to construct operator splitting schemes similar to those introduced in Chapter 2. Let us define χ_ω to be the characteristic function of the domain ω.

On the interval (t^n, t^{n+1}), given (ϕ^n, u^n), we solve three subproblems to obtain $(\phi^{n+1/2}, u^{n+1/2})$, $(\tilde{\phi}^{n+1}, \tilde{u}^{n+1})$, and (ϕ^{n+1}, u^{n+1}), in that order.

• **Operator Splitting Scheme** OFDDM$_m$:

For $n = 0, 1, 2, \ldots, N-1$, solve:

- SUBPROBLEM $(1)_m$: Find $(\phi^{n+1/2}, u^{n+1/2}, \lambda^{n+1/2})$ solution of:

$$\begin{cases} \dfrac{\partial \phi}{\partial t} - \dfrac{u}{2} = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1/2}), \\ \dfrac{1}{c^2}\dfrac{\partial u}{\partial t} + \lambda \chi_\omega = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1/2}), \\ \phi \chi_\omega = g, \quad \text{in} \quad \Omega \times (t^n, t^{n+1/2}), \\ \phi(t^n) = \phi^n, \ u(t^n) = u^n. \end{cases} \tag{3.45}$$

- SUBPROBLEM $(2)_m$: Find $(\tilde{\phi}^{n+1}, \tilde{u}^{n+1})$ solution of:

$$\begin{cases} \dfrac{\partial \phi}{\partial t} - \dfrac{u}{2} = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1}), \\ \dfrac{1}{c^2}\dfrac{\partial u}{\partial t} - \Delta \phi = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1}), \\ u + c\dfrac{\partial \phi}{\partial \mathbf{n}} = 0, \quad \text{in} \quad \Gamma \times (t^n, t^{n+1}), \\ \phi(t^n) = \phi^{n+1/2}, \ u(t^n) = u^{n+1/2}. \end{cases} \tag{3.46}$$

- SUBPROBLEM $(3)_m$: Find $(\phi^{n+1}, u^{n+1}, \lambda^{n+1})$ solution of:

$$\begin{cases} \dfrac{\partial \phi}{\partial t} - \dfrac{u}{2} = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \dfrac{1}{c^2}\dfrac{\partial u}{\partial t} + \lambda \chi_\omega = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \phi \chi_\omega = g, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \phi(t^{n+1/2}) = \tilde{\phi}^{n+1}, \ u(t^{n+1/2}) = \tilde{u}^{n+1}. \end{cases} \tag{3.47}$$

We will rewrite the scheme OFDDM$_m$, by eliminating the variable u from the equations. Thus, from a first order system of PDE's we move back to the second order wave equation.

• **Operator Splitting Scheme** OFDDM$_s$:

For $n = 0, 1, 2, \ldots, N-1$, given (ϕ^n, u^n) solve:

- SUBPROBLEM $(1)_s$: Find $(\phi^{n+1/2}, \lambda^{n+1/2})$ solution of:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\partial^2\phi}{\partial t^2} + \lambda\chi_\omega = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1/2}), \\ \phi\chi_\omega = g, \quad \text{in} \quad \Omega \times (t^n, t^{n+1/2}), \\ \phi(t^n) = \phi^n,\ \phi_t(t^n) = \dfrac{1}{2}u^n. \end{cases} \tag{3.48}$$

Calculate $u^{n+1/2} = 2\dfrac{\partial\phi}{\partial t}|^{n+1/2}$ as follows: Find $(\hat{\phi}^{n+1}, \hat{\lambda}^{n+1})$ solution of:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\partial^2\phi}{\partial t^2} + \lambda\chi_\omega = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \phi\chi_\omega = g, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \phi(t^n) = \phi^n,\ \phi(t^{n+1/2}) = \phi^{n+1/2}. \end{cases} \tag{3.49}$$

$$\text{Set} \quad u^{n+1/2} = 2\frac{\hat{\phi}^{n+1} - \phi^n}{\Delta t}. \tag{3.50}$$

- SUBPROBLEM $(2)_s$: Find $\tilde{\phi}^{n+1}$ solution of:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\partial^2\phi}{\partial t^2} - \Delta\phi = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1}), \\ \dfrac{2}{c}\dfrac{\partial\phi}{\partial t} + \dfrac{\partial\phi}{\partial \mathbf{n}} = 0, \quad \text{on} \quad \Gamma \times (t^n, t^{n+1}), \\ \phi(t^n) = \phi^{n+1/2},\ \phi_t(t^n) = \dfrac{1}{2}u^{n+1/2}. \end{cases} \tag{3.51}$$

Calculate $\tilde{u}^{n+1} = 2\dfrac{\partial \phi}{\partial t}|^{n+1}$ as follows: Find $\hat{\phi}^{n+2}$ solution of:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\partial^2 \phi}{\partial t^2} - \Delta \phi = 0, \quad \text{in} \quad \Omega \times (t^{n+1}, t^{n+2}), \\ \dfrac{2}{c}\dfrac{\partial \phi}{\partial t} + \dfrac{\partial \phi}{\partial \mathbf{n}} = 0, \quad \text{on} \quad \Gamma \times (t^{n+1}, t^{n+2}), \\ \phi(t^n) = \phi^{n+1/2},\ \phi(t^{n+1}) = \tilde{\phi}^{n+1}. \end{cases} \tag{3.52}$$

$$\text{Set} \quad \tilde{u}^{n+1} = \frac{\hat{\phi}^{n+2} - \phi^{n+1/2}}{\Delta t}. \tag{3.53}$$

- SUBPROBLEM $(3)_s$: Find $(\phi^{n+1}, \lambda^{n+1})$ solution of:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\partial^2 \phi}{\partial t^2} + \lambda \chi_\omega = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \phi \chi_\omega = g, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \phi(t^{n+1/2}) = \tilde{\phi}^{n+1},\ \phi_t(t^{n+1/2}) = \dfrac{1}{2}\tilde{u}^{n+1}. \end{cases} \tag{3.54}$$

Calculate $u^{n+1} = 2\dfrac{\partial \phi}{\partial t}|^{n+1}$ as follows: Find $(\hat{\phi}^{n+3/2}, \hat{\lambda}^{n+3/2})$ solution of:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\partial^2 \phi}{\partial t^2} + \lambda \chi_\omega = 0, \quad \text{in} \quad \Omega \times (t^{n+1}, t^{n+3/2}), \\ \phi \chi_\omega = g, \quad \text{in} \quad \Omega \times (t^{n+1}, t^{n+3/2}), \\ \phi(t^{n+1/2}) = \tilde{\phi}^{n+1},\ \phi(t^{n+1}) = \phi^{n+1}. \end{cases} \tag{3.55}$$

$$\text{Set} \quad u^{n+1} = 2\frac{\hat{\phi}^{n+3/2} - \tilde{\phi}^{n+1}}{\Delta t}. \tag{3.56}$$

We note that, in each subproblem of the operator splitting schemes, we have to perform an additional step over an interval of length, either Δt or $\Delta t/2$, in order to calculate the time derivative u.

We will use a centered finite difference scheme for the time discretization, and the finite element spaces described in Section 3.4 for the space discretization of the subproblems

in scheme OFDDM$_s$. With these space/time discretizations we construct a fully discrete operator splitting scheme for the solution of the wave problem.

On the interval (t^n, t^{n+1}), given (ϕ_h^n, u_h^n) we solve three subproblems to obtain $(\phi_h^{n+1/2}, u_h^{n+1/2}, \lambda_h^{n+1/2})$, $(\tilde{\phi}_h^{n+1}, \tilde{u}_h^{n+1})$, $(\phi_h^{n+1}, u_h^{n+1}, \lambda_h^{n+1})$, in that order, as below.

• **Operator Splitting Scheme** OFDDM:

For $n = 0, 1, 2, \ldots, N-1$, solve:

- SUBPROBLEM $(1)_h$: Find $(\phi_h^{n+1/2}, \lambda_h^{n+1/2}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\phi_h^{n+1/2} - 2\phi_h^n + \bar{\phi}_h^{n-1/2}}{(\Delta t/2)^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\omega \lambda_h^{n+1/2}\, w_h \mathrm{d}\omega \\ = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\phi_h^{n+1/2} - g^{n+1/2})\mu_h \,\mathrm{d}\omega = 0\ , \forall\ \mu_h \in \mathbf{\Lambda}_h, \\ \phi_h^{n+1/2} - \bar{\phi}_h^{n-1/2} = \dfrac{\Delta t}{2}\, u_h^n. \end{cases} \tag{3.57}$$

Calculate $u_h^{n+1/2} = 2\dfrac{\partial \phi}{\partial t}|^{n+1/2}$ as follows:

Find $(\hat{\phi}_h^{n+1}, \hat{\lambda}_h^{n+1}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\hat{\phi}_h^{n+1} - 2\phi_h^{n+1/2} + \phi_h^n}{(\Delta t/2)^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\omega \hat{\lambda}_h^{n+1}\, w_h \mathrm{d}\omega \\ = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\hat{\phi}_h^{n+1} - g^{n+1})\mu_h \,\mathrm{d}\omega = 0\ , \forall\ \mu_h \in \mathbf{\Lambda}_h. \end{cases} \tag{3.58}$$

$$\text{Set} \quad u_h^{n+1/2} = 2\frac{\hat{\phi}_h^{n+1} - \phi_h^n}{\Delta t}. \tag{3.59}$$

- SUBPROBLEM $(2)_h$: Find $\tilde{\phi}_h^{n+1} \in \mathbf{V}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\tilde{\phi}_h^{n+1} - 2\phi_h^{n+1/2} + \bar{\phi}_h^{n-1}}{\Delta t^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\Omega \nabla\phi_h^{n+1/2}\cdot\nabla w_h \,\mathrm{dx} \\ +\dfrac{2}{c}\displaystyle\int_\Gamma \frac{\tilde{\phi}_h^{n+1} - \bar{\phi}_h^{n-1}}{2\Delta t}\, w_h \,\mathrm{d}\Gamma = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \tilde{\phi}_h^{n+1} - \bar{\phi}_h^{n-1} = \Delta t\, u_h^{n+1/2}. \end{cases} \tag{3.60}$$

Calculate $\tilde{u}_h^{n+1} = 2\dfrac{\partial\phi}{\partial t}|^{n+1}$ as follows:

Find $\hat{\phi}_h^{n+2} \in \mathbf{V}_h$ such that

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\hat{\phi}_h^{n+2} - 2\tilde{\phi}_h^{n+1} + \phi_h^{n+1/2}}{\Delta t^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\Omega \nabla\tilde{\phi}_h^{n+1}\cdot\nabla w_h \,\mathrm{dx} \\ +\dfrac{2}{c}\displaystyle\int_\Gamma \frac{\tilde{\phi}_h^{n+2} - \phi_h^{n+1/2}}{2\Delta t}\, w_h \,\mathrm{d}\Gamma = 0, \quad \forall w_h \in \mathbf{V}_h. \end{cases} \tag{3.61}$$

$$\text{Set} \quad \tilde{u}_h^{n+1} = \frac{\hat{\phi}_h^{n+2} - \phi_h^{n+1/2}}{\Delta t}. \tag{3.62}$$

- SUBPROBLEM 3_h: Find $(\phi_h^{n+1}, \lambda_h^{n+1}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\phi_h^{n+1} - 2\tilde{\phi}_h^{n+1} + \bar{\phi}_h^{n-1}}{(\Delta t/2)^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\omega \lambda_h^{n+1}\, w_h \mathrm{d}\omega \\ = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\phi_h^{n+1} - g^{n+1})\mu_h \,\mathrm{d}\omega = 0\ , \forall\ \mu_h \in \mathbf{\Lambda}_h, \\ \phi_h^{n+1} - \bar{\phi}_h^{n-1} = \dfrac{\Delta t}{2}\tilde{u}_h^{n+1}. \end{cases} \tag{3.63}$$

Calculate $u_h^{n+1} = 2\dfrac{\partial\phi}{\partial t}|^{n+1}$ as follows:

Find $(\hat{\phi}_h^{n+3/2}, \hat{\lambda}_h^{n+3/2}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\hat{\phi}_h^{n+3/2} - 2\phi_h^{n+1} + \tilde{\phi}_h^{n+1}}{(\Delta t/2)^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\omega \hat{\lambda}_h^{n+3/2}\, w_h \mathrm{d}\omega \\ = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\hat{\phi}_h^{n+3/2} - g^{n+3/2})\mu_h \,\mathrm{d}\omega = 0\ , \forall\ \mu_h \in \mathbf{\Lambda}_h, \end{cases} \tag{3.64}$$

$$\text{Set} \quad u_h^{n+1} = 2\frac{\hat{\phi}_h^{n+3/2} - \tilde{\phi}_h^{n+1}}{\Delta t}. \tag{3.65}$$

We have used the finite element space $\mathbf{V}_h$ and a fully explicit scheme to solve subproblems $(1)_h$ and $(3)_h$. This approach is mostly interesting to evaluate the accuracy of the splitting method. The idea of subproblems $(1)_h$ and $(3)_h$ is to approximate $L^2(\Omega)$ by the finite element space $\mathbf{V}_h$ as defined in (3.15).

3.6 A Formulation of the 2D Scalar Wave Equation as a First Order System

In the operator splitting scheme OFDDM, introduced in the last section, we notice that the enforcement of the Dirichlet condition on the boundary of ω, and the propagation of the wave are decoupled. Hence, subproblem $(2)_m$ can be formulated using, for example, the *velocity - stress* formulation of the wave equation. A mixed finite element discretization can then be employed for this formulation.

In the velocity - stress formulation, the wave equation is written as a system of first order PDE's involving the velocity u, and the gradient $p = \nabla\phi$ of the solution ϕ. Mixed formulations are useful, for example, in cases where one requires knowledge of the gradient. This formulation is also useful in constructing perfectly matched layer boundary conditions, as will be demonstrated in Chapter 5, for Maxwell's equations.

In the second operator splitting scheme that we will introduce, as mentioned in the introduction to this chapter, we will employ mixed finite elements in the substeps that propagate the wave. To this end we first present a discussion of the velocity - stress formulation of the wave equation.

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain. Wave propagation is modeled by the 2D scalar wave

equation.

$$\left\{ \begin{array}{l} \text{Find} \quad \phi : (0,T) \rightarrow H_0^1(\Omega) \quad \text{such that :} \\ \dfrac{1}{c^2}\dfrac{\partial^2 \phi}{\partial t^2} - \Delta \phi = f, \quad f \in C^0(0,T;L^2(\Omega)), \end{array} \right. \tag{3.66}$$

with the initial conditions

$$\phi(0) = \phi_0 \in H^1(\Omega) \quad ; \quad \frac{\partial \phi}{\partial t}(0) = \phi_1 \in L^2(\Omega). \tag{3.67}$$

For the purposes of the presentation of a mixed formulation we will use Dirichlet boundary conditions on the artificial boundary Γ of the domain Ω. Thus $\phi \in H_0^1(\Omega)$. In the next section, we will go back to absorbing boundary conditions on Γ. Define

$$\left\{ \begin{array}{rcl} \mathbf{p} & = & \nabla \phi, \\ u & = & \dfrac{\partial \phi}{\partial t}. \end{array} \right. \tag{3.68}$$

Using the definitions above in (3.66) we obtain the first order velocity-stress formulation of the wave equation:

$$\left\{ \begin{array}{rcrcl} \dfrac{1}{c^2}\dfrac{\partial u}{\partial t} & - & \nabla \cdot \mathbf{p} & = & f, \quad \text{in} \quad \Omega \times [0,T], \\ \dfrac{\partial \mathbf{p}}{\partial t} & - & \nabla u & = & 0, \quad \text{in} \quad \Omega \times [0,T], \end{array} \right. \tag{3.69}$$

with the initial conditions

$$\mathbf{p}(0) = \mathbf{p}_0 = \nabla \phi_0 \quad ; \quad u(0) = u_0 = \phi_1 \tag{3.70}$$

System (3.69) leads to the *saddle point* problem

$$\left\{ \begin{array}{l} \text{Find} \quad (\mathbf{p}, u) : (0,T) \rightarrow V \times Q \equiv [L^2(\Omega)]^2 \times H_0^1(\Omega) \quad \text{such that :} \\ \quad \dfrac{d}{dt}\displaystyle\int_\Omega \mathbf{p} \cdot \mathbf{q} \, \mathrm{dx} \quad - \quad \int_\Omega \mathbf{q} \cdot \nabla u \, \mathrm{dx} \quad = \quad 0 \ , \ \forall \mathbf{q} \in V, \\ \quad \dfrac{1}{c^2}\dfrac{d}{dt}\displaystyle\int_\Omega u \, w \, \mathrm{dx} \quad + \quad \int_\Omega \mathbf{p} \cdot \nabla w \, \mathrm{dx} \quad = \quad \int_\Omega f w \, \mathrm{dx} \ , \ \forall w \in Q. \end{array} \right. \tag{3.71}$$

We define the bilinear forms

$$\left\{\begin{array}{lcll} a(\mathbf{p},\mathbf{q}) & = & \displaystyle\int_\Omega \mathbf{p}\cdot\mathbf{q}\,\mathrm{dx}\,, & \forall\,(\mathbf{p},\mathbf{q})\in V\times V,\\ c(u,w) & = & \displaystyle\int_\Omega u\,w\,\mathrm{dx}\,, & \forall\,(u,w)\in Q\times Q,\\ b(w,\mathbf{q}) & = & \displaystyle-\int_\Omega \nabla w\cdot\mathbf{q}\,\mathrm{dx}\,, & \forall\,(w,\mathbf{q})\in Q\times V,\\ (f,w) & = & \displaystyle\int_\Omega f\,w\,\mathrm{dx}\,, & \forall\,w\in Q. \end{array}\right. \tag{3.72}$$

Using the bilinear forms defined in (3.72) in (3.71), we can rewrite the saddle point problem (3.71) as:

$$\left\{\begin{array}{l} \text{Find}\quad (\mathbf{p},u):(0,T)\to V\times Q\equiv [L^2(\Omega)]^2\times H_0^1(\Omega)\quad \text{such that :}\\ \qquad\begin{array}{lcll} \dfrac{d}{dt}a(\mathbf{p},\mathbf{q}) & + & b(u,\mathbf{q}) & = \quad 0\,,\ \forall\,\mathbf{q}\in V,\\ \dfrac{1}{c^2}\dfrac{d}{dt}c(u,w) & - & b(w,\mathbf{q}) & = \quad (f,w)\,,\ \forall\,w\in Q. \end{array} \end{array}\right. \tag{3.73}$$

The bilinear form $a(\cdot,\cdot)$ is continuous on $V\times V$, and thus defines a linear continuous operator $\mathcal{A}: V\to V'$ by

$$\langle \mathcal{A}\mathbf{p},\mathbf{q}\rangle_{V'\times V} = a(\mathbf{p},\mathbf{q}). \tag{3.74}$$

Similarly, the bilinear form $b(\cdot,\cdot)$ is continuous on $Q\times V$, and defines a continuous linear operator $\mathcal{B}: Q\to V'$ such that

$$\langle \mathcal{B}w,\mathbf{q}\rangle_{V'\times Q} = b(w,\mathbf{q}) \tag{3.75}$$

For saddle point problems, in the stationary case, the spaces involved have to satisfy certain compatibility conditions, in order for the problem to be well posed [56]. These conditions require the bilinear form a to be coercive on V, or V-*elliptic*, where as the bilinear form b must satisfy an *inf-suf* condition, also called the LBB condition, as seen below. Let

$$V_0 = \text{Ker}\ \mathcal{B}^T = \{\mathbf{q}\in V | b(w,\mathbf{q}) = 0, \forall w\in Q\}. \tag{3.76}$$

$$\begin{cases} \text{(i)} & \text{Coercivity of the bilinear form } a \text{ on Ker } \mathcal{B}^T \implies \\ & \exists\, \alpha > 0 \; s.t \; \forall\, \mathbf{p} \in V_0,\; a(\mathbf{p},\mathbf{p}) \geq \alpha \|\mathbf{p}\|_V^2. \\ \text{(ii)} & \text{The continuous LBB condition} \implies \\ & \exists\, \beta > 0 \; s.t. \; \forall\, w \in Q, \exists\, \mathbf{q} \in V \; s.t. \; b(w,\mathbf{q}) \geq \beta \|w\|_Q \|\mathbf{q}\|_V. \end{cases} \tag{3.77}$$

Since, $V = [L^2(\Omega)]^2$, the coercivity of $a(\cdot,\cdot)$ is immediate. To prove the LBB condition, we employ the Poincaré-Friedrichs inequality. Given $w \in Q = H_0^1(\Omega)$, choose $\mathbf{q} = -\nabla w \in [L^2(\Omega)]^2$. Then

$$\frac{b(w,\mathbf{q})}{\|\mathbf{q}\|} = \frac{-(\mathbf{q},\nabla w)}{\|\mathbf{q}\|} = \frac{(\nabla w,\nabla w)}{\|\nabla w\|} = \|\nabla w\| \geq \frac{1}{\beta'}\|w\|. \tag{3.78}$$

In (3.78), β' comes from Poincaré's inequality, and thus depends only on Ω. The constant β in the LBB condition is then equal to $1/\beta'$, and the saddle point problem is stable [24].

3.7 Combining an Operator Splitting Scheme with a Mixed Finite Element Method

In this section we propose a symmetrized operator splitting method which uses mixed finite elements in space-time for the substeps that propagate the wave. Thus, we will incorporate the scheme presented in Section 3.6 into the operator splitting scheme, OFDDM, presented in Section 3.5.

We will solve the wave equation in the entire domain Ω in one substep, and in the other two substeps we enforce the Dirichlet condition on $\partial\omega$ using a distributed Lagrange multiplier as is done in the scheme OFDDM. Thus, the second operator splitting scheme differs from OFDDM in the formulation and implementation of subproblem $(2)_h$. We use a mixed method in space-time for the approximation of the wave equation. See [35] for a similar mixed method. We define

$$\mathbf{p} = \nabla\phi, \quad u = 2\frac{\partial\phi}{\partial t}. \tag{3.79}$$

We can rewrite subproblem $(2)_m$ (3.46), in mixed form, by using the definition of $\mathbf{p}$ as

• SUBPROBLEM $(2)_m$: Find $\mathbf{p}^{n+1/2}$ satisfying:

$$\begin{cases} \text{(i)} \quad \dfrac{\partial \mathbf{p}}{\partial t} - \dfrac{1}{2}\nabla u = 0, \text{ in } \Omega \times (t^n, t^{n+1/2}), \\ \text{(ii)} \quad \mathbf{p}(t^n) = \nabla \phi^{n+1/2}, u(t^n) = 2\dfrac{\partial \phi}{\partial t}|^{n+1/2}. \end{cases} \tag{3.80}$$

and then

Find $(\tilde{\phi}^{n+1}, \tilde{u}^{n+1})$ satisfying:

$$\begin{cases} \text{(i)} \quad \dfrac{1}{c^2}\dfrac{\partial u}{\partial t} - \nabla \cdot \mathbf{p} = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \text{(ii)} \quad \dfrac{\partial \phi}{\partial t} - \dfrac{u}{2} = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \text{(iii)} \quad \dfrac{1}{c}\, u + \mathbf{p} \cdot \mathbf{n} = 0, \quad \text{on} \quad \Gamma \times (t^{n+1/2}, t^{n+1}), \\ \text{(iv)} \quad \mathbf{p}(t^{n+1/2}) = p^{n+1/2}, u(t^{n+1/2}) = 2\dfrac{\partial \phi}{\partial t}|^{n+1/2}. \end{cases} \tag{3.81}$$

Given $\phi^{n+1/2}$ and $\dfrac{\partial \phi}{\partial t}|^{n+1/2}$, we cannot solve for $\mathbf{p}$ using the system above. Thus we differentiate the $\mathbf{p}$ equation in time to obtain a *wave* equation in $\mathbf{p}$. Let us define the time derivative of u as

$$a = \frac{\partial u}{\partial t} \tag{3.82}$$

Differentiating equation (3.80, i), we rewrite subproblem $(2)_m$ as:

• SUBPROBLEM $(2)_m$: Find $\mathbf{p}^{n+1/2}$ satisfying:

$$\begin{cases} \text{(i)} \quad \dfrac{\partial^2 \mathbf{p}}{\partial t^2} - \dfrac{1}{2}\nabla a = 0, \quad \text{in} \quad \Omega \times (t^n, t^{n+1/2}), \\ \text{(ii)} \quad \mathbf{p}(t^n) = \nabla \phi^{n+1/2}, a(t^n) = \dfrac{\partial u}{\partial t}|^{n+1/2}, \\ \text{(iii)} \quad \dfrac{\partial \mathbf{p}}{\partial t}(t^n) = \dfrac{1}{2}\nabla u^{n+1/2}. \end{cases} \tag{3.83}$$

and then

Find $(\tilde{\phi}^{n+1}, \tilde{u}^{n+1})$ satisfying:

$$\begin{cases} \text{(ii)} & \dfrac{1}{c^2}\dfrac{\partial u}{\partial t} - \nabla \cdot \mathbf{p} = 0, \text{ in } \Omega \times (t^{n+1/2}, t^{n+1}), \\ \text{(iii)} & \dfrac{\partial \phi}{\partial t} - \dfrac{u}{2} = 0, \quad \text{in} \quad \Omega \times (t^{n+1/2}, t^{n+1}), \\ \text{(iv)} & \dfrac{1}{c}\, u + \mathbf{p} \cdot \mathbf{n} = 0, \text{ on } \Gamma \times (t^{n+1/2}, t^{n+1}), \\ \text{(v)} & \mathbf{p}(t^{n+1/2}) = p^{n+1/2}, u(t^{n+1/2}) = 2\dfrac{\partial \phi}{\partial t}|^{n+1/2}, \end{cases} \tag{3.84}$$

The variational formulation for (3.83)-(3.84) is given as:

- SUBPROBLEM $(2)_m$: On $\Omega \times (t^n, t^{n+1})$ solve for $(\mathbf{p}^{n+1/2}, \tilde{\phi}^{n+1}, \tilde{u}^{n+1})$ satisfying:

$$\begin{cases} \dfrac{d^2}{dt^2}\displaystyle\int_\Omega \mathbf{p} \cdot \mathbf{q}\, \mathrm{dx} - \frac{1}{2}\int_\Omega \nabla a \cdot \mathbf{q} = 0\ , \forall\ \mathbf{q} \in [L^2(\Omega)]^2, \\ \dfrac{1}{c^2}\dfrac{d}{dt}\displaystyle\int_\Omega u\, w\, \mathrm{dx} + \int_\Omega \nabla w \cdot \mathbf{p} + \frac{1}{c}\int_\Gamma u\, w\, \mathrm{d\Gamma} = 0\ , \forall\ w \in H^1(\Omega), \\ \dfrac{\partial \phi}{\partial t} - \dfrac{1}{2}u = 0, \\ \mathbf{p}(t^n) = \nabla \phi^{n+1/2},\ \dfrac{\partial \mathbf{p}}{\partial t}(t^n) = \dfrac{1}{2}\nabla u^{n+1/2},\ u(t^n) = 2\dfrac{\partial \phi}{\partial t}|^{n+1/2}, \\ a(t^n) = \dfrac{\partial u}{\partial t}|^{n+1/2}. \end{cases} \tag{3.85}$$

We note that the absorbing boundary condition, $\dfrac{u}{c} + \mathbf{p} \cdot \mathbf{n} = 0$, is incorporated into the variational formulation, and hence does not have to be explicitly added to the finite element spaces.

For the approximation of $\mathbf{p}$, a *discrete* inf-sup condition has to be satisfied in order for the approximation to remain well posed. Since $\phi_h \in Q_1$ on any $K \in \mathcal{T}_h$, we must choose a finite element space $\mathbf{P}_h$, such that the reference space for this approximation is

$$\nabla Q_1 = P_{01} \times P_{10}. \tag{3.86}$$

Thus, the approximation space $\mathbf{P}_h$, for the approximation of $\mathbf{p}$, is chosen to be

$$\mathbf{P}_h = \{\mathbf{q} \mid \forall\ K \in \mathcal{T}_h,\ \mathbf{q}|_K \in P_{01} \times P_{10}\}. \tag{3.87}$$

This space of linear edge elements for the gradient is the lowest order Nédélec space in two-dimensions. The degrees of freedom for the approximations ϕ_h, u_h and of $\mathbf{p}_h$ are staggered in both time and space as shown in Figure 3.5.

As before, we will use a centered finite difference scheme for the time discretization of the wave problem. On the interval $[0,T]$, let $\Delta t = T/N$ be the time step, where $N \in \mathbb{N}$. Define $t^j = j\Delta t$, and $\phi^j \approx \phi(t^j)$, where $j = k$ or $j = k+1/2$, for $k \in \mathbb{Z}$. We now describe the second operator splitting scheme using mixed elements for the solution of (3.1) initialized by

$$u_h^0 = 2\phi_1, \quad \text{and} \quad \phi_h^0 = \phi_0, \tag{3.88}$$

where ϕ_0 and ϕ_1 are defined in (3.2). Define a substep $\tau = \Delta t/2$.

• **Operator Splitting Scheme** MOFDDM:

For $n = 0, 1, 2, \ldots, N-1$, on the interval (t^n, t^{n+1}), given $(\phi_h^n,\ u_h^n)$, solve the following three subproblems:

SUBPROBLEM $(1)_h$: Find $(\phi_h^{n+1/2}, \lambda_h^{n+1/2}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\phi_h^{n+1/2} - 2\phi_h^n + \bar{\phi}_h^{n-1/2}}{(\Delta t/2)^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\omega \lambda_h^{n+1/2}\, w_h \,\mathrm{d}\omega = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\phi_h^{n+1/2} - g^{n+1/2})\mu_h = 0\ , \forall\ \mu_h \in \mathbf{\Lambda}_h, \\ \phi_h^{n+1/2} - \bar{\phi}_h^{n-1/2} = \tau\ u_h^n. \end{cases}$$

Calculate $u_h^{n+1/2} = 2\dfrac{\partial \phi_h}{\partial t}|^{n+1/2}$ and $a_h^{n+1/2} = \dfrac{\partial u_h}{\partial t}|^{n+1/2}$ as follows:

Find $(\hat{\phi}_h^{n+1}, \hat{\lambda}_h^{n+1}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\hat{\phi}_h^{n+1} - 2\phi_h^{n+1/2} + \phi_h^n}{(\Delta t/2)^2}\, w_h \,\mathrm{d}\mathbf{x} + \int_\omega \hat{\lambda}_h^{n+1}\, w_h \,\mathrm{d}\omega = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\hat{\phi}_h^{n+1} - g^{n+1})\mu_h = 0\ , \forall\ \mu_h \in \mathbf{\Lambda}_h. \end{cases} \tag{3.89}$$

$$\begin{aligned} \text{Set} \quad u_h^{n+1/2} &= 2\frac{\hat{\phi}_h^{n+1} - \phi^n}{\Delta t}, \\ a_h^{n+1/2} &= 2\frac{\hat{\phi}_h^{n+1} - 2\phi_h^{n+1/2} + \phi_h^n}{(\Delta t/2)^2}. \end{aligned} \tag{3.90}$$

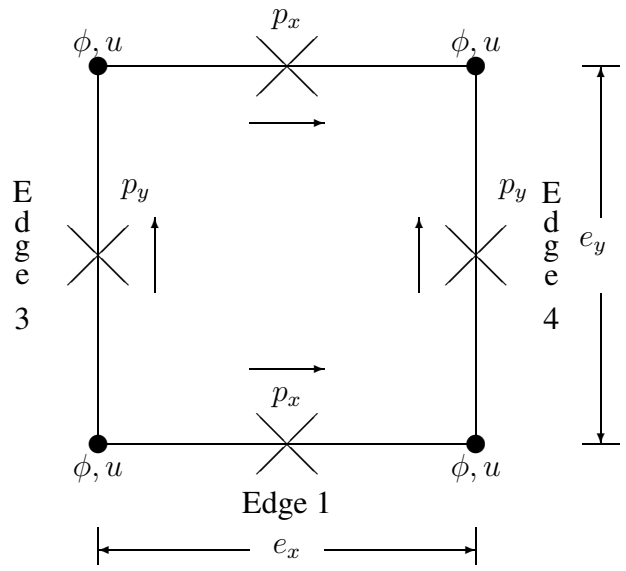

Figure 3.5: A sample domain element K. The degrees of freedom, for the solution ϕ and the velocity u, and the gradient $\mathbf{p} = (p_x, p_y)$, are staggered in space. ϕ, u are bilinear continuous functions with degrees of freedom at the nodes of the square. The degrees of freedom for p_x and p_y are at the midpoints of edges parallel to the x-axis, and y-axis, respectively.

SUBPROBLEM $(2)_h$: Find $(\mathbf{p}_h^{n+1/2}, \tilde{u}_h^{n+1}, \tilde{\phi}_h^{n+1}) \in \mathbf{P}_h \times \mathbf{V}_h \times \mathbf{V}_h$, such that:

$$\begin{cases} \dfrac{1}{\tau^2}\displaystyle\int_\Omega \{\mathbf{p}_h^{n+1/2} - 2\nabla\phi_h^{n+1/2} + \bar{\mathbf{p}}_h^{n-1/2}\}\cdot\mathbf{q}_h \,\mathrm{dx} - \frac{1}{2}\int_\Omega \nabla a_h^{n+1/2}\cdot\mathbf{q}_h \,\mathrm{dx} \\ \quad = 0\ ,\forall\ \mathbf{q}_h \in \mathbf{P}_h, \\ \dfrac{1}{c^2}\displaystyle\int_\Omega \frac{\tilde{u}_h^{n+1} - u_h^{n+1/2}}{\tau}\, w_h \,\mathrm{dx} + \frac{1}{c}\int_\Gamma \frac{\tilde{u}_h^{n+1} + u_h^{n+1/2}}{2}\, w_h \,\mathrm{d\Gamma} \\ \quad + \displaystyle\int_\Omega \nabla w \cdot \mathbf{p}_h^{n+1/2} \,\mathrm{dx} = 0,\ \forall\ w_h \in \mathbf{V}_h, \\ \dfrac{\tilde{\phi}_h^{n+1} - \phi_h^{n+1/2}}{2\tau} = \dfrac{1}{2}\left\{\dfrac{u_h^{n+1/2} + \tilde{u}_h^{n+1}}{2}\right\}, \\ \dfrac{1}{2\tau}\displaystyle\int_\Omega \{\mathbf{p}_h^{n+1/2} - \bar{\mathbf{p}}_h^{n-1/2}\}\cdot\mathbf{q}_h \,\mathrm{dx} = \frac{1}{2}\int_\Omega \nabla u_h^{n+1/2}\cdot\mathbf{q}_h \,\mathrm{dx},\ \forall\ \mathbf{q}_h \in \mathbf{P}_h. \end{cases} \tag{3.91}$$

SUBPROBLEM $(3)_h$: Find $(\phi_h^{n+1}, \lambda_h^{n+1}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\phi_h^{n+1} - 2\tilde{\phi}_h^{n+1} + \bar{\phi}_h^{n-1}}{\tau^2}\, w_h \,\mathrm{d\mathbf{x}} + \int_\omega \lambda^{n+1}\, w_h \,\mathrm{d}\omega = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\phi_h^{n+1} - g^{n+1})\mu_h = 0\ ,\forall\ \mu_h \in \mathbf{\Lambda}_h, \\ \phi_h^{n+1} - \bar{\phi}_h^{n-1} = \tau\ \tilde{u}_h^{n+1}. \end{cases} \tag{3.92}$$

Calculate $u_h^{n+1} = 2\dfrac{\partial\phi_h}{\partial t}|^{n+1}$ as follows:
Find $(\hat{\phi}_h^{n+3/2}, \hat{\lambda}_h^{n+3/2}) \in \mathbf{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_\Omega \frac{\hat{\phi}_h^{n+3/2} - 2\phi_h^{n+1} + \tilde{\phi}_h^{n+1}}{\tau^2}\, w_h \,\mathrm{d\mathbf{x}} + \int_\omega \hat{\lambda}_h^{n+3/2}\, w_h \,\mathrm{d}\omega = 0, \quad \forall w_h \in \mathbf{V}_h, \\ \displaystyle\int_\omega (\hat{\phi}_h^{n+3/2} - g^{n+3/2})\mu_h = 0\ ,\forall\ \mu_h \in \mathbf{\Lambda}_h. \end{cases} \tag{3.93}$$

$$\text{Set} \quad u_h^{n+1} = 2\frac{\hat{\phi}_h^{n+3/2} - \tilde{\phi}_h^{n+1}}{\Delta t}. \tag{3.94}$$

The idea of subproblems $(1)_h$ and $(3)_h$ is to approximate $L^2(\Omega)$ by the finite element space $\mathbf{V}_h$ as defined in (3.15). In these subproblems we use the Uzawa-conjugate gradient algorithm (1) [65], to solve the system of linear equations that arise.

3.8 Scattering by a Disk

3.8.1 Problem Description

We consider the scattering of the harmonic planar waves $e^{-\mathrm{i}(\rho t - \mathbf{k}\cdot\mathbf{x})}$ by a perfectly reflecting disk whose radius is 0.25 meter. The frequency f, is 0.6 GHz, and the wavelength L, is 0.5 meter. The wavenumber is denoted by $\mathbf{k} = (k_x, k_y)$. The angular frequency is $\rho = 2\pi f$. The wave illuminates ω from the left and propagates horizontally. The geometry of the problem is illustrated in Figure 3.6. We have used a rectangular mesh consisting of 113×113 nodes, with the mesh step size $h = 0.5/16$ meter. The time step is $\Delta t = 2\pi/(25\rho)$. For this test problem the exact solution is known when Γ is located at infinity.

3.8.2 Exact Solution

We present here the exact solution for the scattering problem with a circular scatterer. Let the circle be centered at the origin, with radius r_0. The analytic solution for the Dirichlet problem is given by,

$$\phi(r,\theta) = -\sum_{n\in\mathbb{Z}} \mathrm{i}^{|n|} J_{|n|}(\rho r_0) \mathrm{e}^{\mathrm{i}n\theta} \frac{H_n^{(1)}(\rho r)}{H_n^{(1)}(\rho r_0)}, \quad \forall\, r \in [r_0, \infty), \quad \forall\, \theta \in [0, 2\pi], \tag{3.95}$$

where $H_n^{(1)}$ is the *Hankel* function of the first kind and J_n is the *Bessel* function of order n.

3.8.3 Numerical Results

In this section we present plots of the real part of the exact solution (3.95), and the real part of the solution computed using the operator splitting scheme MOFDDM of Section 3.7, for the scattering problem described in Section 3.8.1. We will also present tables of errors of solutions computed using all the three different schemes introduced in this chapter, with respect to the exact solution, as well as with certain reference solutions.

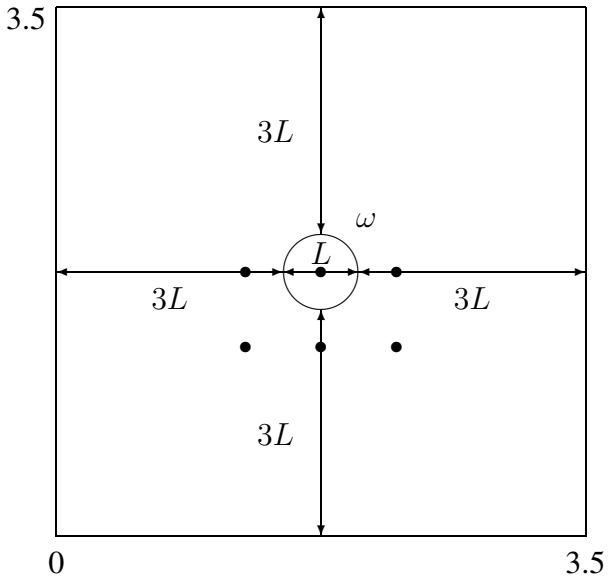

Figure 3.6: Domain Ω with a circular obstacle. The disk is one wavelength in diameter. The distance between the disk and the boundary of the domain is 3 wavelengths ($3L$). The darkened points are $L/2$ units away from the boundary of the disk in the x, and/or y direction.

In Figures 3.7 and 3.8 a top view, and a contour plot, respectively, of the exact solution and the solution computed using the operator splitting scheme MOFDDM is shown. We have used 16 points per wavelength to compute our solution, with a time step of $\Delta t = 6.6667e-11$, such that the CFL condition is

$$\frac{c\Delta t}{h} = 0.64 < \frac{1}{\sqrt{2}}. \tag{3.96}$$

The computed solution is time integrated for 175 time steps, i.e., until $t = 7L/c = 7/f$. Figures 3.9 and 3.10 are contour plots of the exact solution, and solutions to the operator splitting scheme MOFDDM, with refined mesh step sizes using 32 and 64 nodes per wavelength, respectively. In both these plots the time step is also refined such that the CFL

condition is given by (3.96). In Figures 3.11 and 3.12, a full view of the exact solution, and the computed solution for a discretization with 16 nodes per wavelength, respectively, is presented. The figures show a remarkable agreement, considering that the mesh is not locally modified to fit $\partial\omega$, as some other fictitious domain methods do.

In Figure 3.13 we compare a slice of the exact and computed solutions, which is taken in a direction perpendicular to the propagation of the wave, and containing the center of the disk, i.e., through the line $x = 1.75$. We show the comparison for three different discretizations, 16, 32, and 64 nodes per wavelength, after 7 periodic cycles. As the mesh size is refined the agreement of the computed solution with the exact gets better. We note that the value at points which lie along the diameter of the circle is 1, since the boundary condition is imposed on the entire obstacle. Figure 3.15 presents the error, with respect to the exact solution, at points on the line $x = 1.75$. The error seems to be decreasing as $\mathcal{O}(h)$.

In Figure 3.14 we compare a slice of the exact and computed solutions, which is taken in a direction parallel to the propagation of the wave, and containing the center of the disk, i.e., through the line $y = 1.75$. We show the comparison for three different discretizations, 16, 32, and 64 nodes per wavelength after 7 periodic cycles. Again, as the mesh size is refined, the agreement of the computed solution with the exact solution improves. Figure 3.16 presents the error, with respect to the exact solution, at points on the line $y = 1.75$. As before, the error seems to be decreasing as $\mathcal{O}(h)$.

As a last comparison, we compare the evolution in time of the computed and exact solutions at six different points in the mesh for 300 time steps, i.e., until $t = 12L/c$. Figure 3.17 presents the time evolution at the points (1.25,1.25), (1.75, 1.25), and (2.25, 1.25). Figure 3.18 presents the time evolution at the points (1.25,1.75), (1.75, 1.75), and (2.25, 1.75). These points are highlighted in Figure 3.6. Each point is half a wavelength from the boundary of the disk in the x, and/or y direction. The time evolution plots are computed for a discretization with 16 nodes per wavelength.

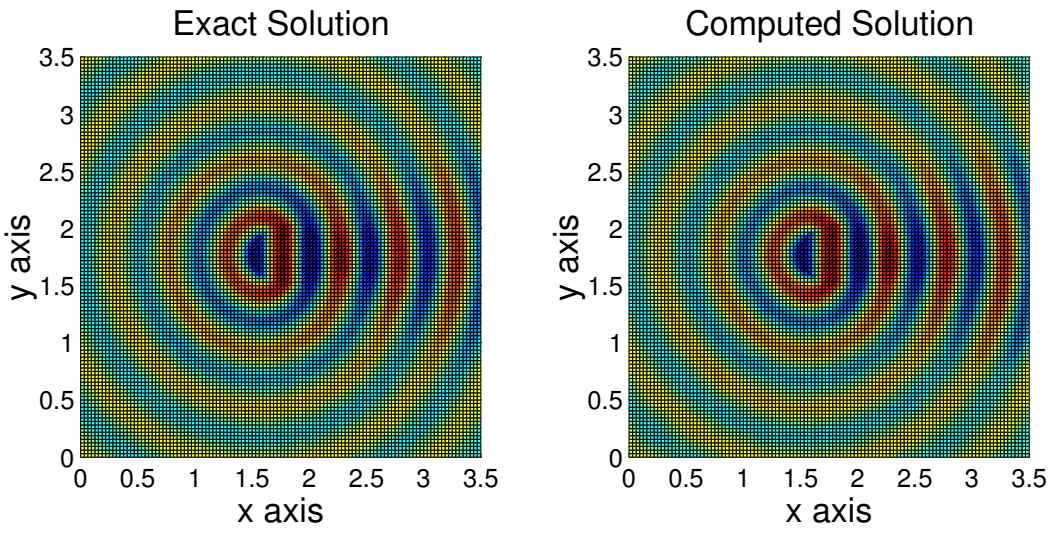

Figure 3.7: Top view of the real parts of the exact and computed solutions for $h = L/16$, and $\Delta t = L/(25c)$.

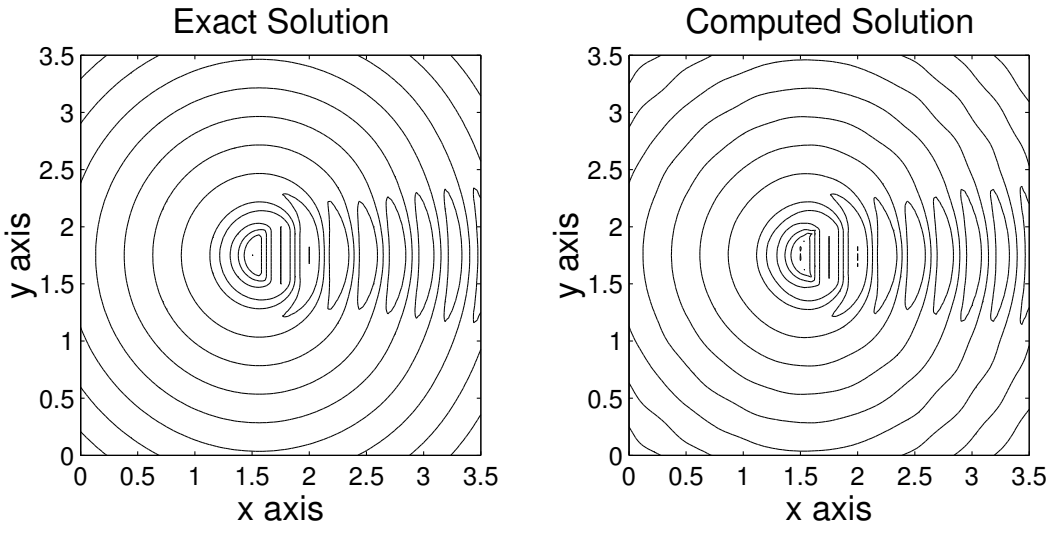

Figure 3.8: Contours of the real parts of the exact and computed solutions for $h = L/16$, and $\Delta t = L/(25c)$.

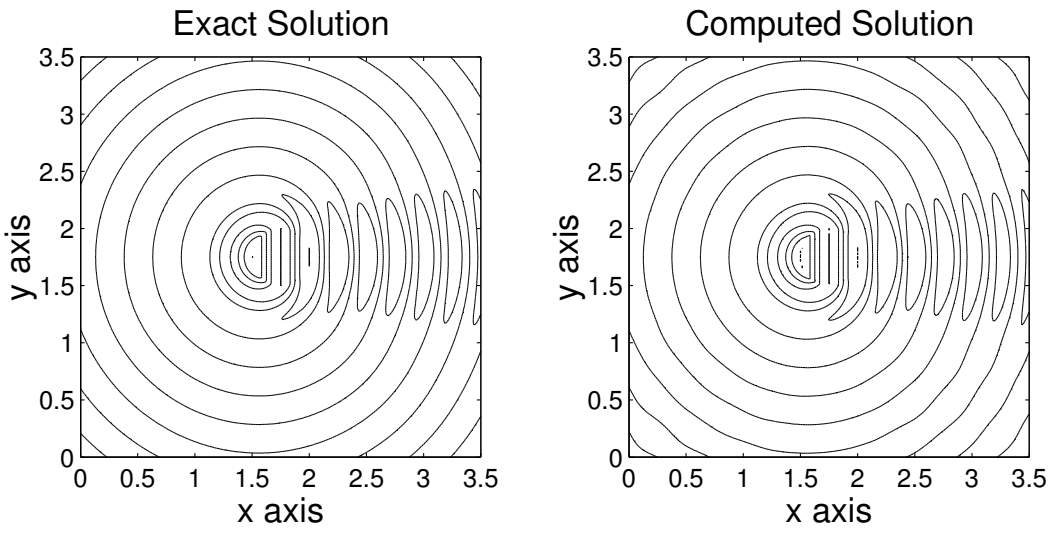

Figure 3.9: Contours of the real parts of the exact and computed solutions for $h = L/32$, and $\Delta t = L/(50c)$.

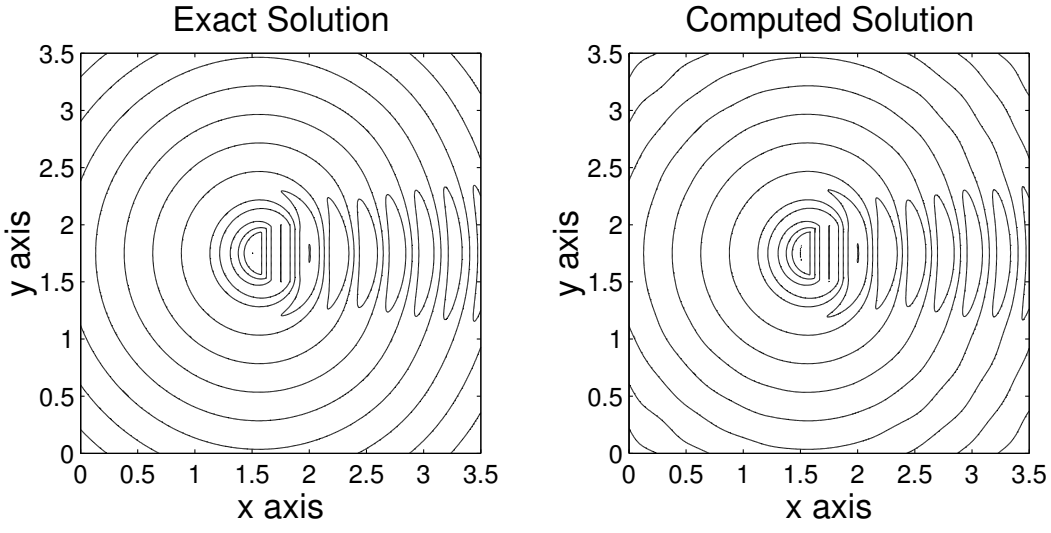

Figure 3.10: Contours of the real parts of the exact and computed solutions for $h = L/64$, and $\Delta t = L/(100c)$.

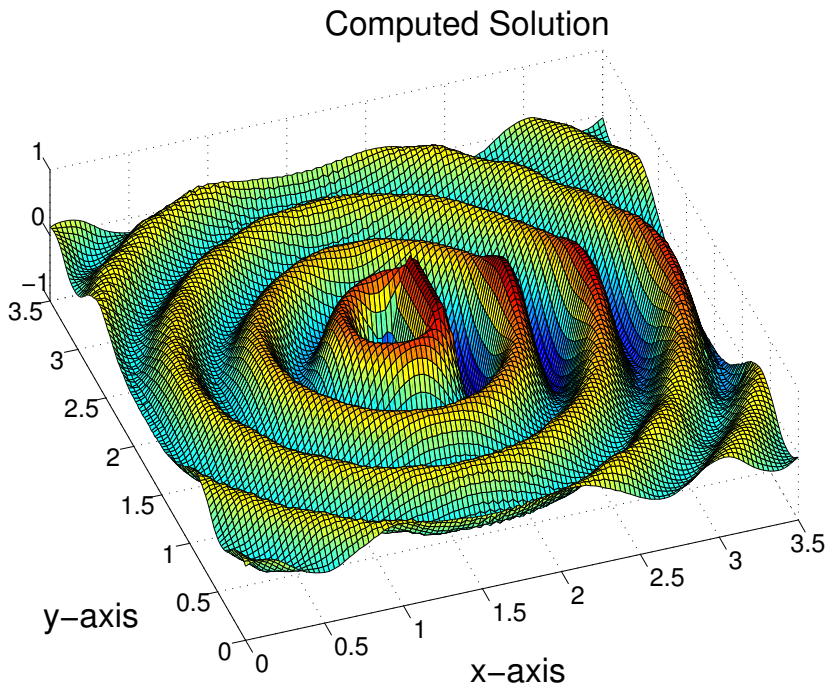

Figure 3.11: Real part of the computed solution for $h = L/16$, and $\Delta t = L/(25c)$.

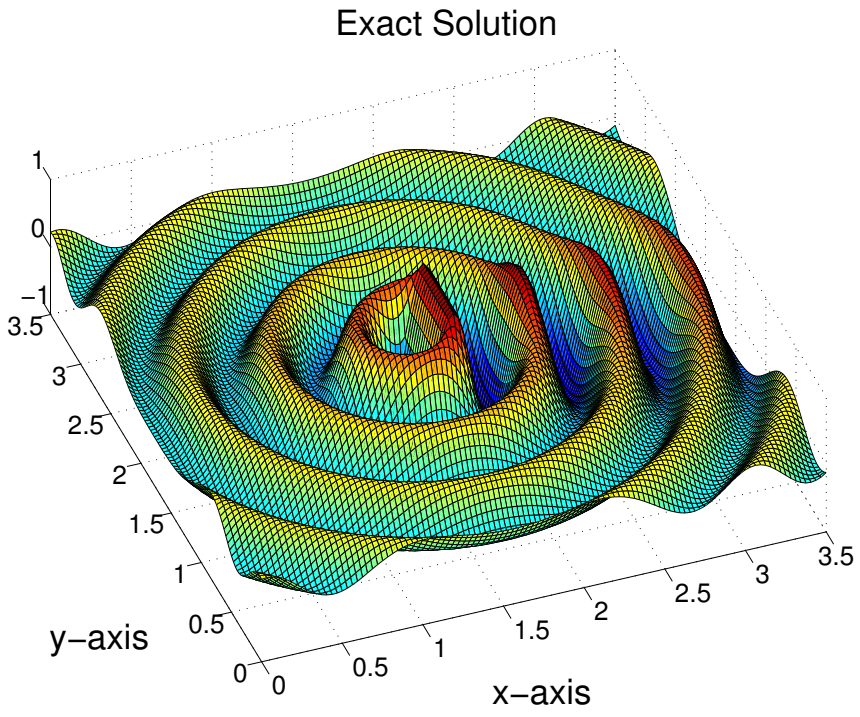

Figure 3.12: Real part of the exact solution.

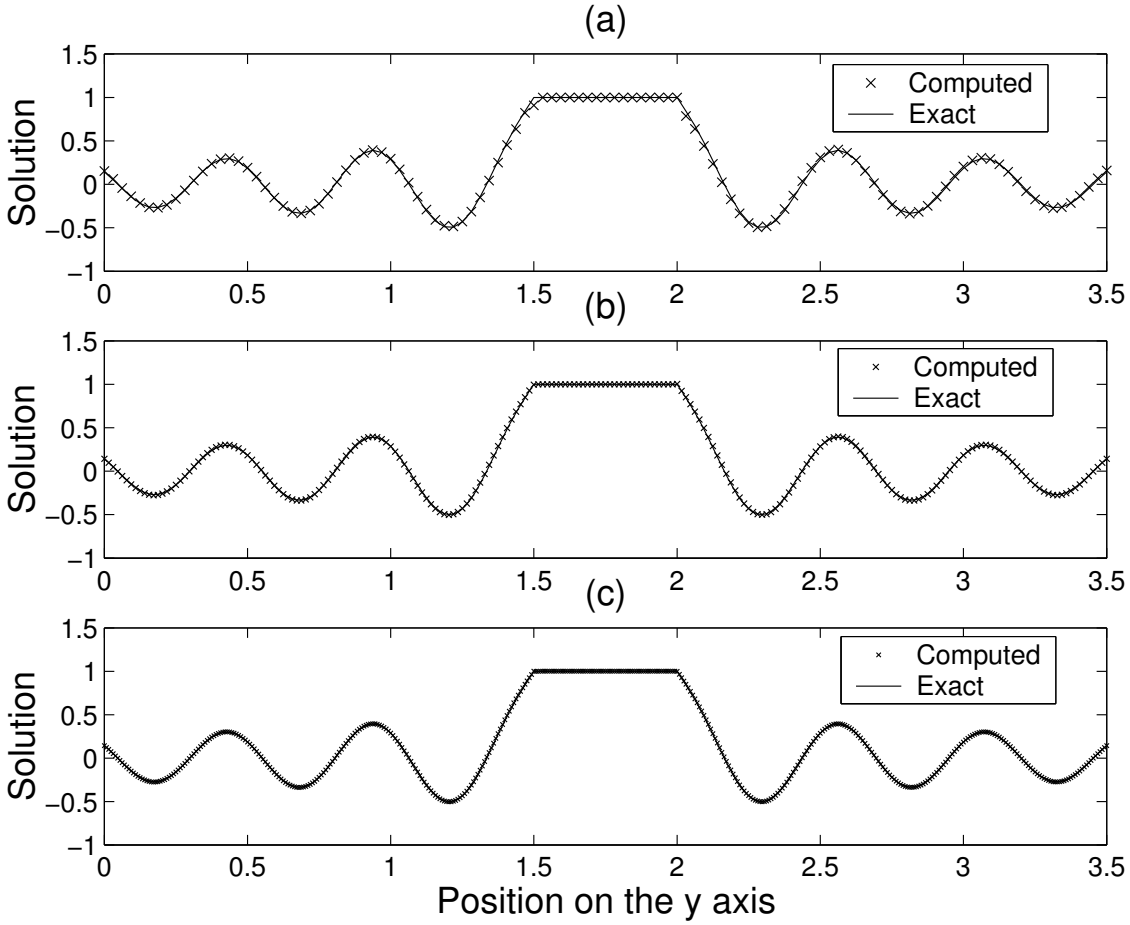

Figure 3.13: Comparison of the real parts of the exact (—) and computed (x) solutions, on the half-line containing the center of ω and perpendicular to the incidence direction for (a) $h = L/16$, (b) $h = L/32$, and (c) $h = L/64$.

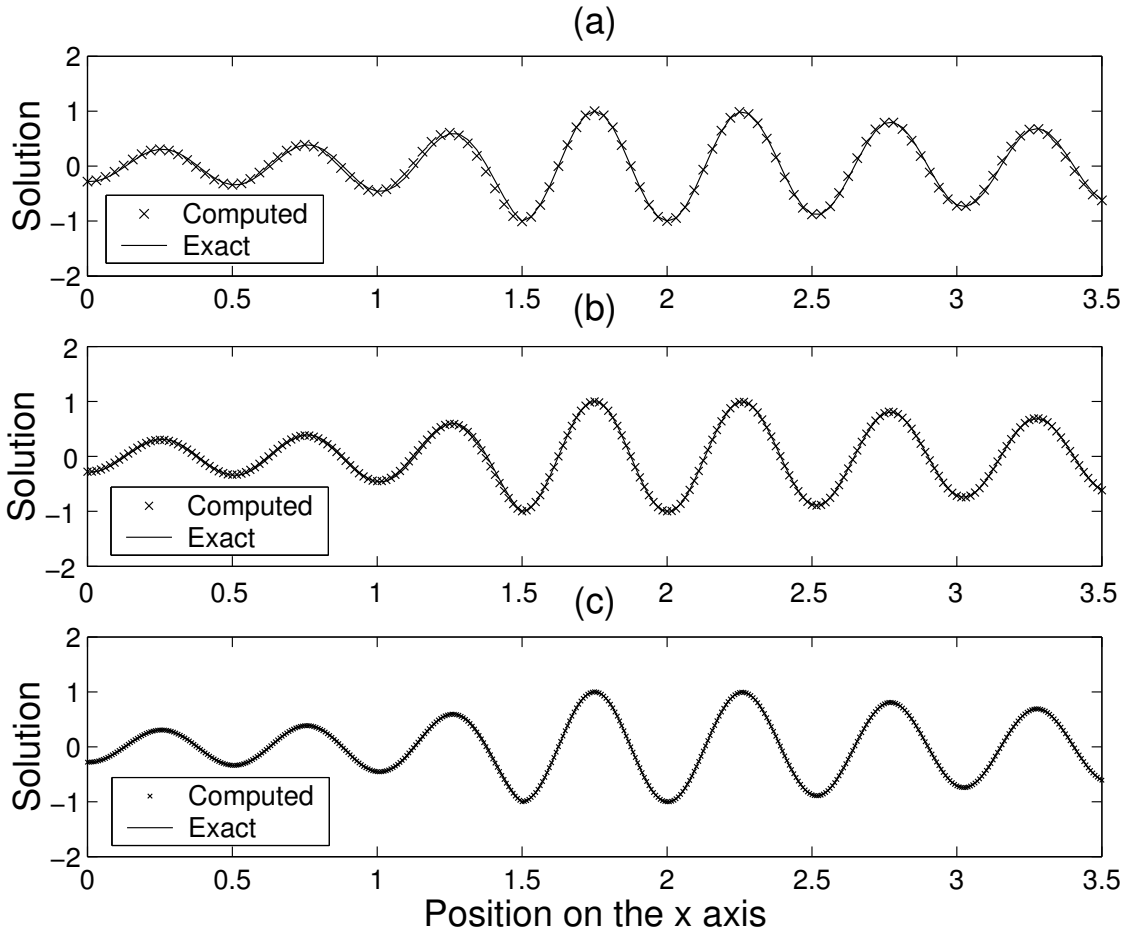

Figure 3.14: Comparison of the real parts of the exact (—) and computed (x) solutions, on the half-line containing the center of ω and parallel to the incidence direction for (a) $h = L/16$, (b) $h = L/32$, and (c) $h = L/64$.

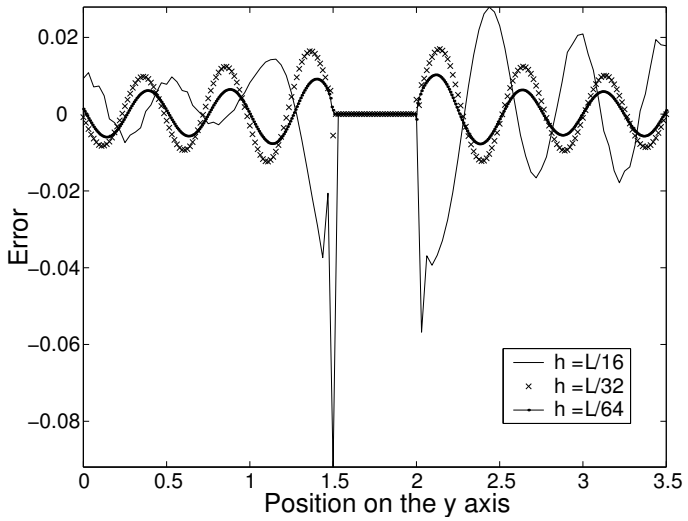

Figure 3.15: Error in the real parts of the computed and exact solutions, on the half-line containing the center of ω and perpendicular to the incidence direction for different values of h.

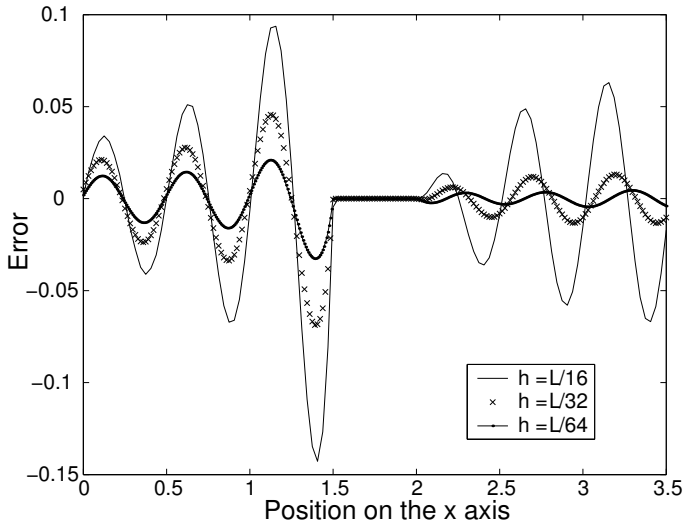

Figure 3.16: Error in the real parts of the computed and exact solutions, on the half-line containing the center of ω and parallel to the incidence direction for different values of h.

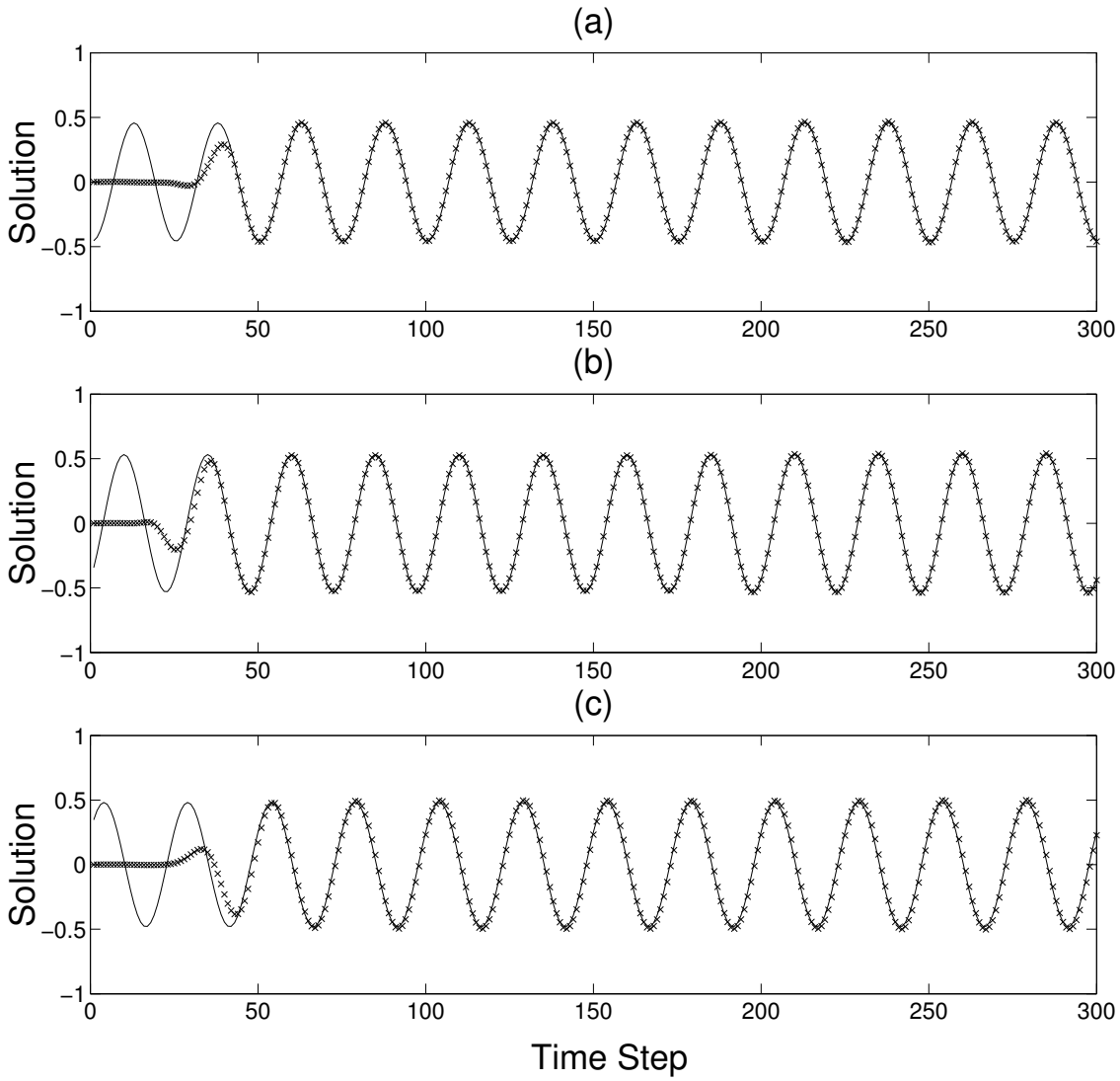

Figure 3.17: Time evolution for 300 time steps of the exact (-), and computed (x) solutions at the points (a)$(1.25, 1.25)$, (b)$(1.75, 1.25)$, and (c) $(2.25, 1.25)$

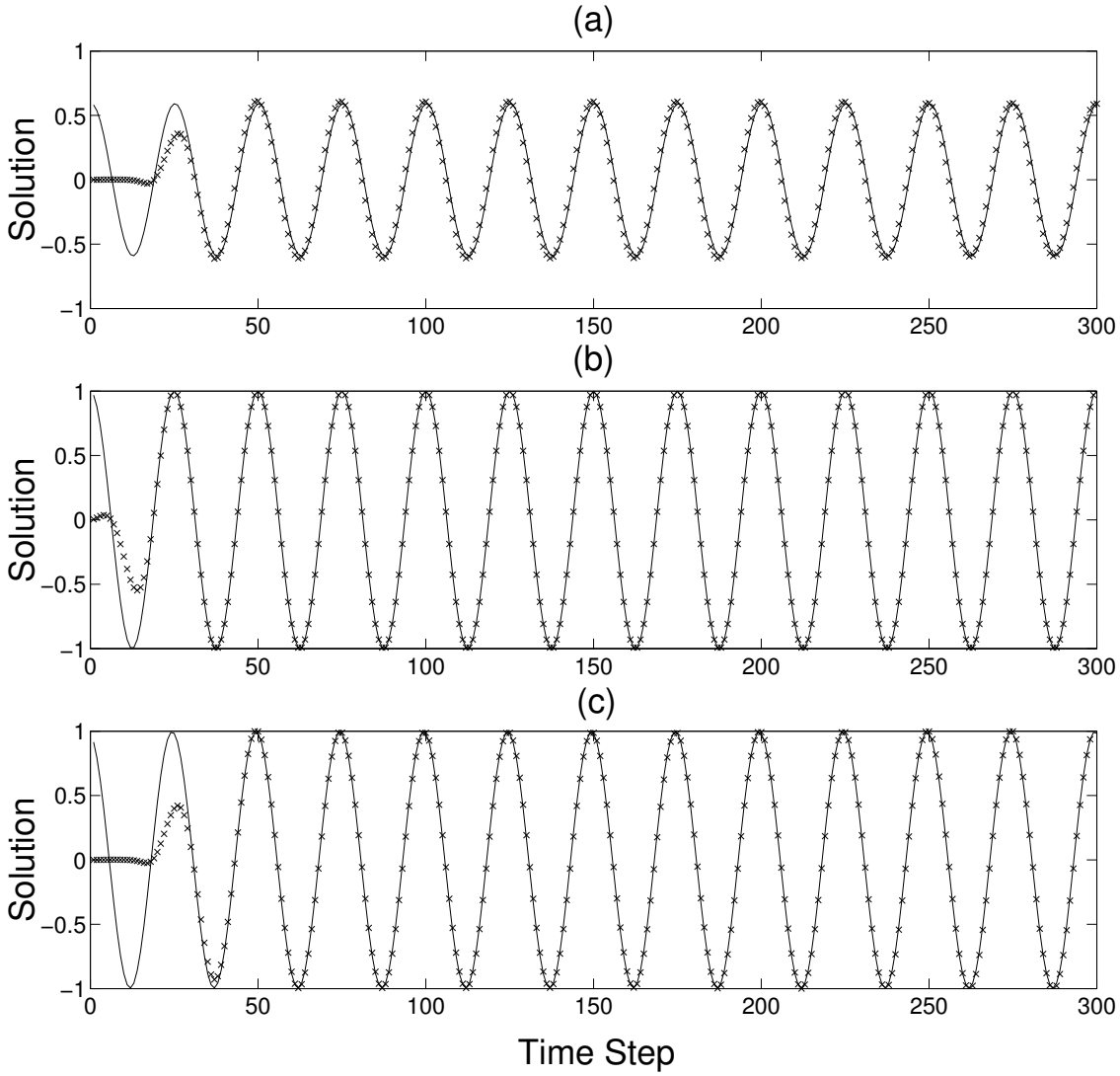

Figure 3.18: Time evolution for 300 time steps of the exact (-), and computed (x) solutions at the points (a)$(1.25, 1.75)$, (b)$(1.75, 1.75)$, and (c) $(2.25, 1.75)$

We now present some error computations for the three new schemes presented in this chapter, namely FDDM, OFDDM, and the scheme MOFDDM. For each scheme, the computations are performed for a mesh ratio $h_{\partial\omega}/h = 2.0$, between the mesh step size, and the step size on the boundary of the disk. The tolerance ϵ for the Uzawa algorithm is taken to be $\epsilon = 10^{-9}$. N is the total number of nodes for each computation. For all the cases presented below, the number of iterations needed for convergence in the Uzawa algorithm, were between 11 and 16, for all values of h and Δt.

In Table 3.1, we present (relative) errors for the schemes, calculated with respect to a reference solution. The reference solution for each scheme is computed via the same scheme, with $h = L/16$, and $\Delta t = L/(800c)$. Thus, the reference solution is refined with respect to Δt, but not with respect to h. The idea here is to determine the temporal order of accuracy of each scheme. Thus, in Table 3.1

$$\text{Relative Error } = \frac{\|\phi_C - \phi_R\|_{L^2(\Omega)}}{\|\phi_R\|_{L^2(\Omega)}}, \tag{3.97}$$

where ϕ_C denotes the computed solution, and ϕ_R denotes the reference solution. The total number of nodes, for this case is $N = 113\times113$. Table 3.1 shows that the ratio of successive errors, for all the three schemes, is approximately 4.00, which suggests the second order temporal accuracy of each scheme.

h	Δt	FDDM		OFDDM		MOFDDM	
		Error	Ratio	Error	Ratio	Error	Ratio
$L/16$	$L/(25c)$	3.262e-2		5.501e-2		4.857e-2	
$L/16$	$L/(50c)$	8.332e-3	3.92	1.349e-2	4.08	1.463e-2	3.32
$L/16$	$L/(100c)$	2.132e-3	3.91	3.319e-3	4.07	3.618e-3	4.04
$L/16$	$L/(200c)$	5.127e-4	4.16	7.891e-4	4.21	8.616e-4	4.20

Table 3.1: Error of the solutions computed with respect to a reference solution refined with respect to Δt but not with respect to h.

In Table 3.2, we present errors for the three schemes, again calculated with respect to a reference solution. In this case, the reference solution for each scheme is calculated via the same scheme, with $h = L/128$, i.e., 128 nodes per wavelength, and $\Delta t = L/(200c)$. Thus, the reference solution is now refined with respect to both h, and Δt. The error in this case is calculated using (3.97). Since the ratios of successive errors for all schemes drop below 4.00, but remain above 2.00, this suggests that the spatial order of accuracy for all the schemes is between $\mathcal{O}(h)$, and $\mathcal{O}(h^2)$.

N	h	Δt	FDDM		OFDDM		MOFDDM	
			Error	Ratio	Error	Ratio	Error	Ratio
113^2	$L/16$	$L/(25c)$	6.480e-2		6.352e-2		1.210e-1	
225^2	$L/32$	$L/(50c)$	2.771e-2	2.34	1.688e-2	3.76	2.748e-2	4.40
449^2	$L/64$	$L/(100c)$	1.033e-2	2.68	6.394e-3	2.64	6.938e-3	3.96

Table 3.2: Error of the solutions computed with respect to a reference solution refined in h, and Δt.

Finally, Table 3.3 shows the error for the three schemes with respect to the exact solution presented in Section 3.8.2. The error in each case is the $L^2(\Omega)$ norm of the difference of the computed solution and the exact solution, divided by the $L^2(\Omega)$ norm of the exact solution.

$$\text{Relative Error } = \frac{\|\phi_C - \phi_E\|_{L^2(\Omega)}}{\|\phi_E\|_{L^2(\Omega)}}, \tag{3.98}$$

where, ϕ_E denotes the exact solution. We claim that the first order absorbing boundary condition on the artificial boundary Γ of the domain Ω dominates the error. Hence, it is difficult to get an idea of the spatial or temporal accuracy of the solution from Table 3.3. In Chapter 6 we will consider perfectly matched layers instead of the first order absorbing boundary condition fo Maxwell's equations. These absorbing layers, as will be shown in Chapter 6, provide much better absorbing capabilities.

N	h	Δt	FDDM		OFDDM		MOFDDM	
			Error	Ratio	Error	Ratio	Error	Ratio
113^2	$L/16$	$L/(25c)$	8.484e-2		7.382e-2		1.251e-1	
225^2	$L/32$	$L/(50c)$	5.507e-2	1.54	4.335e-2	1.70	4.666-2	2.68
449^2	$L/64$	$L/(100c)$	4.377e-2	1.26	3.960e-2	1.09	3.837e-2	1.22
897^2	$L/128$	$L/(200c)$	3.919e-2	1.12	3.787e-2	1.05	3.740e-2	1.03

Table 3.3: Error of the solutions computed with respect to the exact solution.

3.9 Scattering by Multiple Disks

3.9.1 Problem Description

We next consider the scattering of the harmonic planar waves $e^{-\mathrm{i}(\rho t - \mathbf{k}\cdot\mathbf{x})}$ by nine perfectly reflecting disks whose radius is 0.25 meter. The frequency is 0.6 GHz, and the wavelength is 0.5 meter. The wave illuminates ω from the left and propagates horizontally. We have used a rectangular mesh consisting of 321×321 nodes, with the mesh step size $h = 0.5/32$ meter. The time step is $\Delta t = 2\pi/(50\rho)$. For this test problem the exact solution is not known. We compare our results obtained using the scheme MOFDDM, with a reference solution that is obtained by solving a time harmonic problem in which the mesh is locally fitted to the boundary of the obstacles [75].

The details of the computational domain are shown in Figure 3.19. Each disk is one wavelength in diameter. The distance between two neighboring disks is one wavelength in the x, and/or y direction. We have kept the (artificial) boundary Γ at least two and a half wavelengths from each disk.

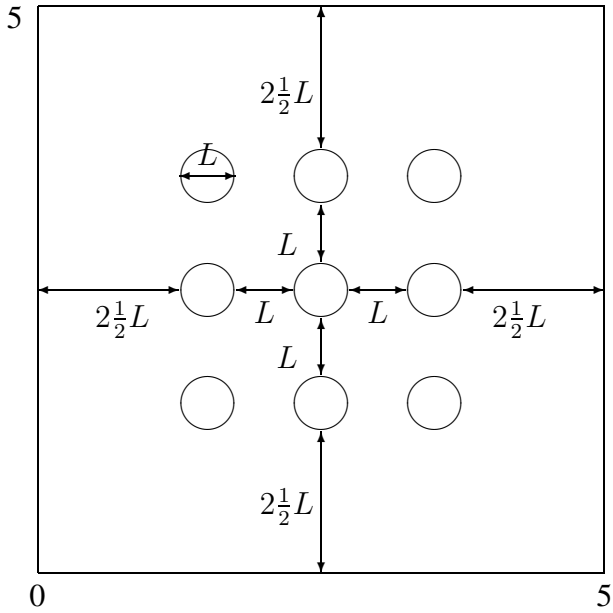

Figure 3.19: Domain Ω with nine disks. Each disk is one wavelength in diameter. The distance between two consecutive disks is one wavelength, and the distance between the outer disks and the boundary of the domain is $2\frac{1}{2}$ wavelengths ($L =$ wavelength).

3.9.2 Numerical Results

We present contour plots of our computed solution, and we compare these plots with the reference solution mentioned in Section 3.9.1. Contour plots are presented in Figures 3.20, 3.21, 3.22, and 3.23, which show the solution after 500, 1000, 1500, and 2000 time steps, respectively. The convergence of the solution to the time harmonic solution is slow, because of the presence of multiple obstacles (a nonconvex obstacle). In Figure 3.24 the contour plot for the time harmonic reference solution is presented.

In Figures 3.25, and 3.26 we compare the evolution in time of the reference, and computed solutions at the points (2.5,1.5), and (2, 1.5). We calculate the time evolution for the time harmonic solution $U(x, y)$ by multiplying it with $\mathrm{e}^{-\mathrm{i}\rho t}$, and considering the real part of this product, i.e., $\mathrm{Re}\left(U(x, y)\mathrm{e}^{-\mathrm{i}\rho t}\right)$. As expected the solutions at the point (2.5,1.5), which is the center of a disk, coincide for all time steps. The agreement of the solutions at the point (2,1.5), which is outside the disk, and at a distance of $L/2$ from the boundary of the disk with center at (2.5,1.5), is not as good, but gets better with time.

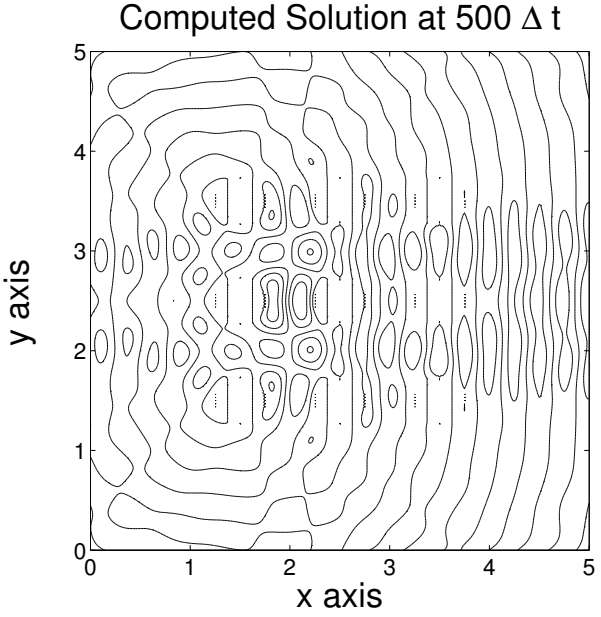

Figure 3.20: Contour plot of the computed solution at $t = 10L/c$.

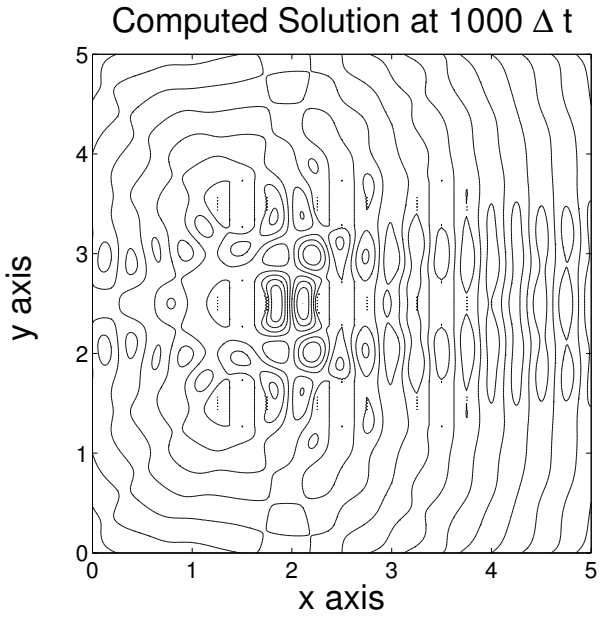

Figure 3.21: Contour plot of the computed solution at $t = 20L/c$.

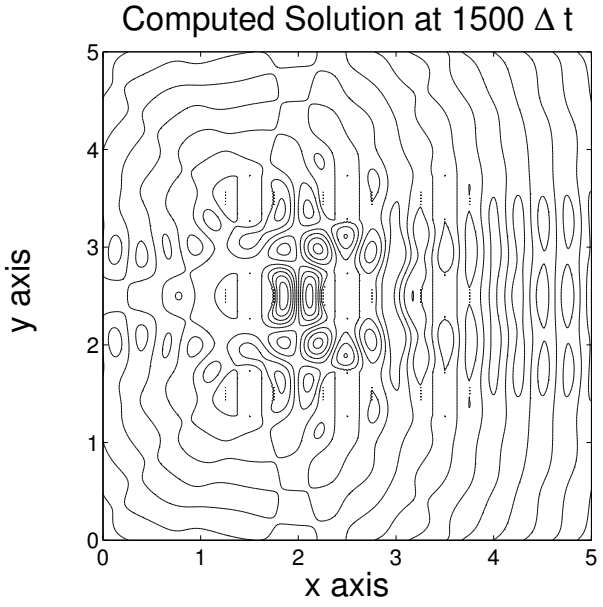

Figure 3.22: Contour plot of the computed solution at $t = 30L/c$.

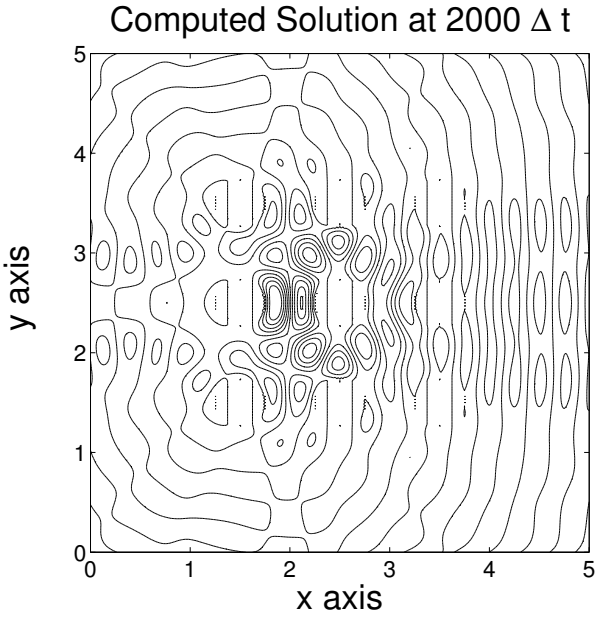

Figure 3.23: Contour plot of the computed solution at $t = 40L/c$.

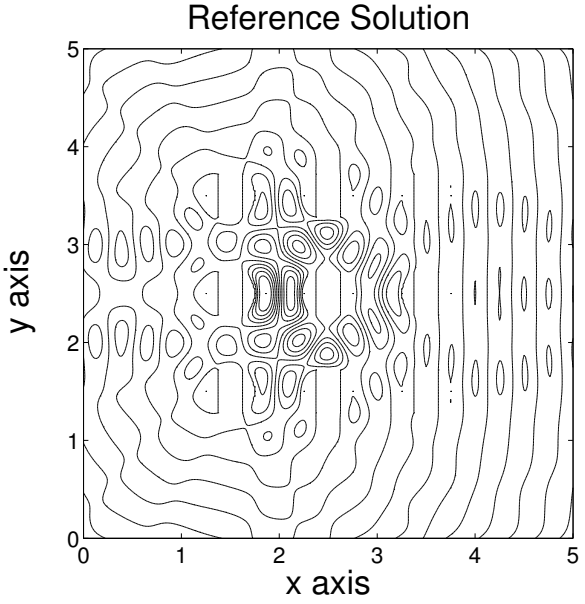

Figure 3.24: Contour plot of the time harmonic solution

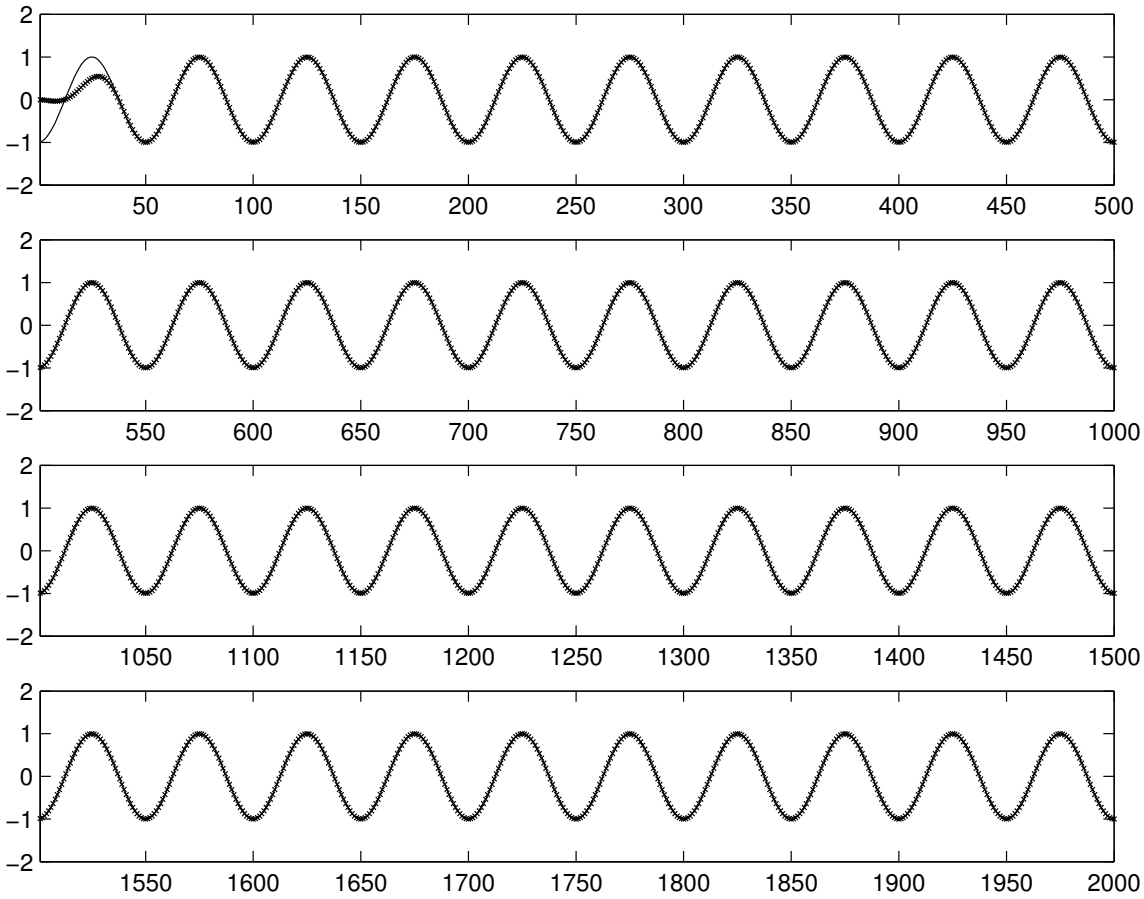

Figure 3.25: Time evolution of the solution at the point $(2.5, 1.5)$. This point is the center of the second circle in the lower layer of circles. (-) denotes the reference solution, and (x) denotes our computed solution

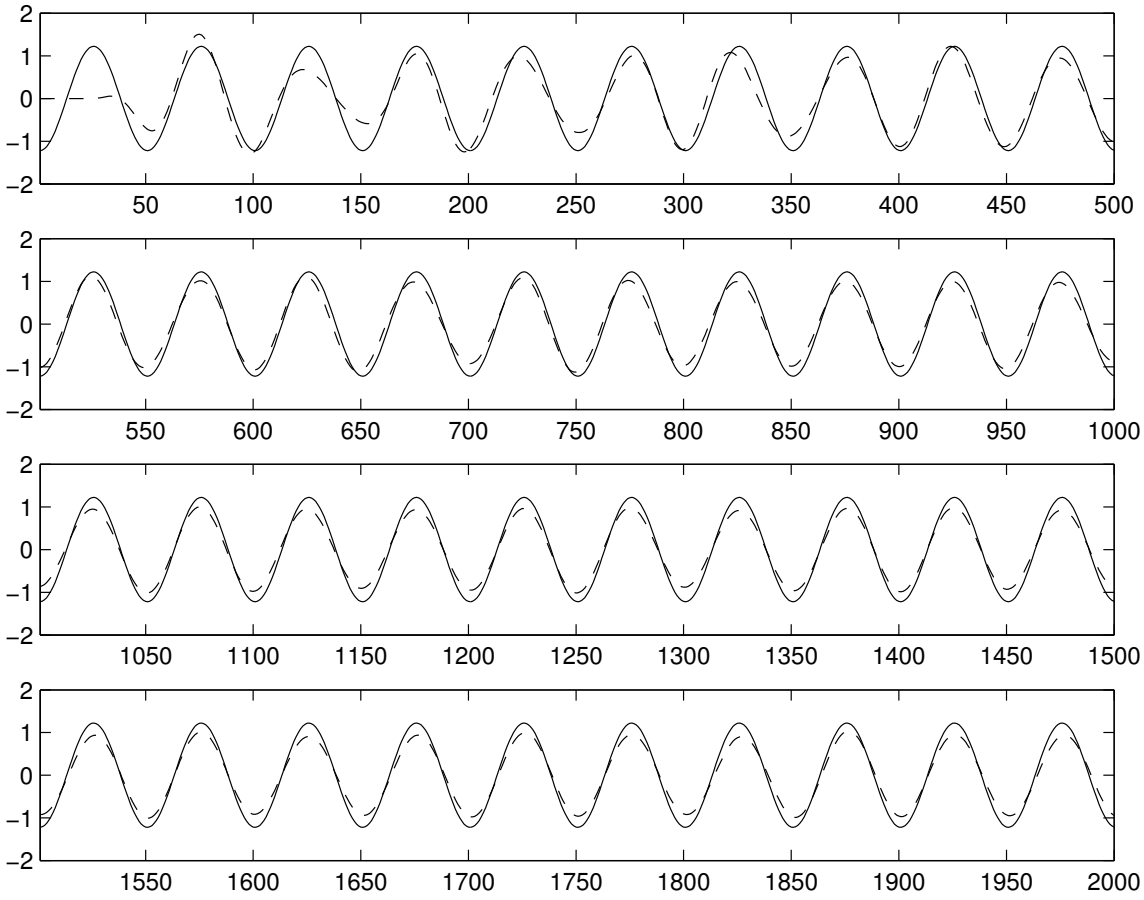

Figure 3.26: Time evolution of the solution at the point $(2, 1.5)$. (-) denotes the reference solution, and (- -) denotes our computed solution

In Figures 3.27, and 3.28 we compare slices of the computed and reference solutions. In Figure 3.27 we compare a slice of the two solutions, which is perpendicular to the direction of propagation of the wave. In Figure 3.28 we compare a slice of the two solutions, which is parallel to the direction of propagation of the wave. In each case the comparison is shown at 500, 1000, 1500 and 2000 time steps. The figures demonstrate the convergence of our solution to the time harmonic reference solution.

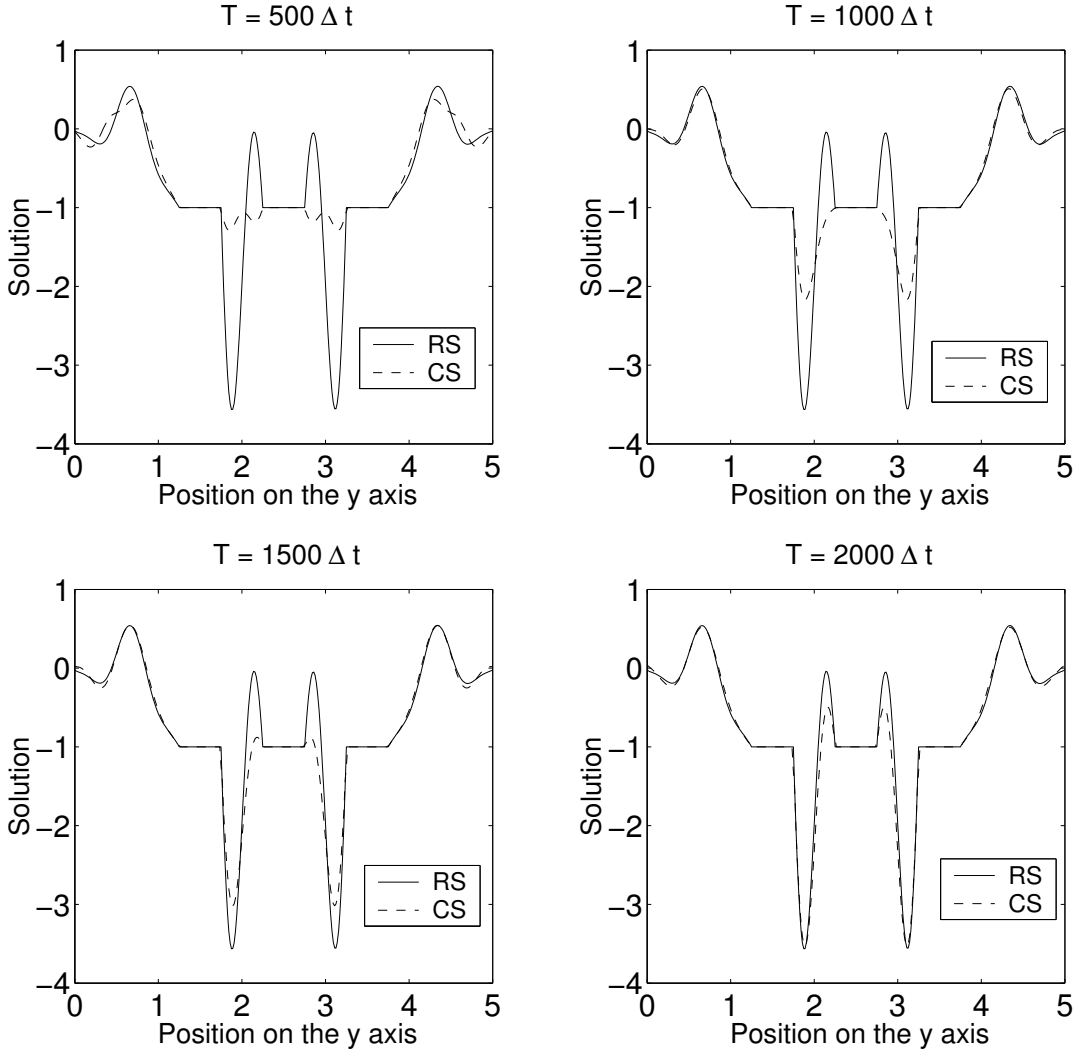

Figure 3.27: Comparison of the reference solution (RS), and the computed solution (CS). In each case a slice of the solution is taken at $x = 2.5$.

As can be seen from the presented plots, in order to get a good comparison between our solution and the reference solution, we have to time-integrate our scheme over a large time interval. In general, it is difficult to obtain the time-harmonic solution in such a case when the obstacle is nonconvex.

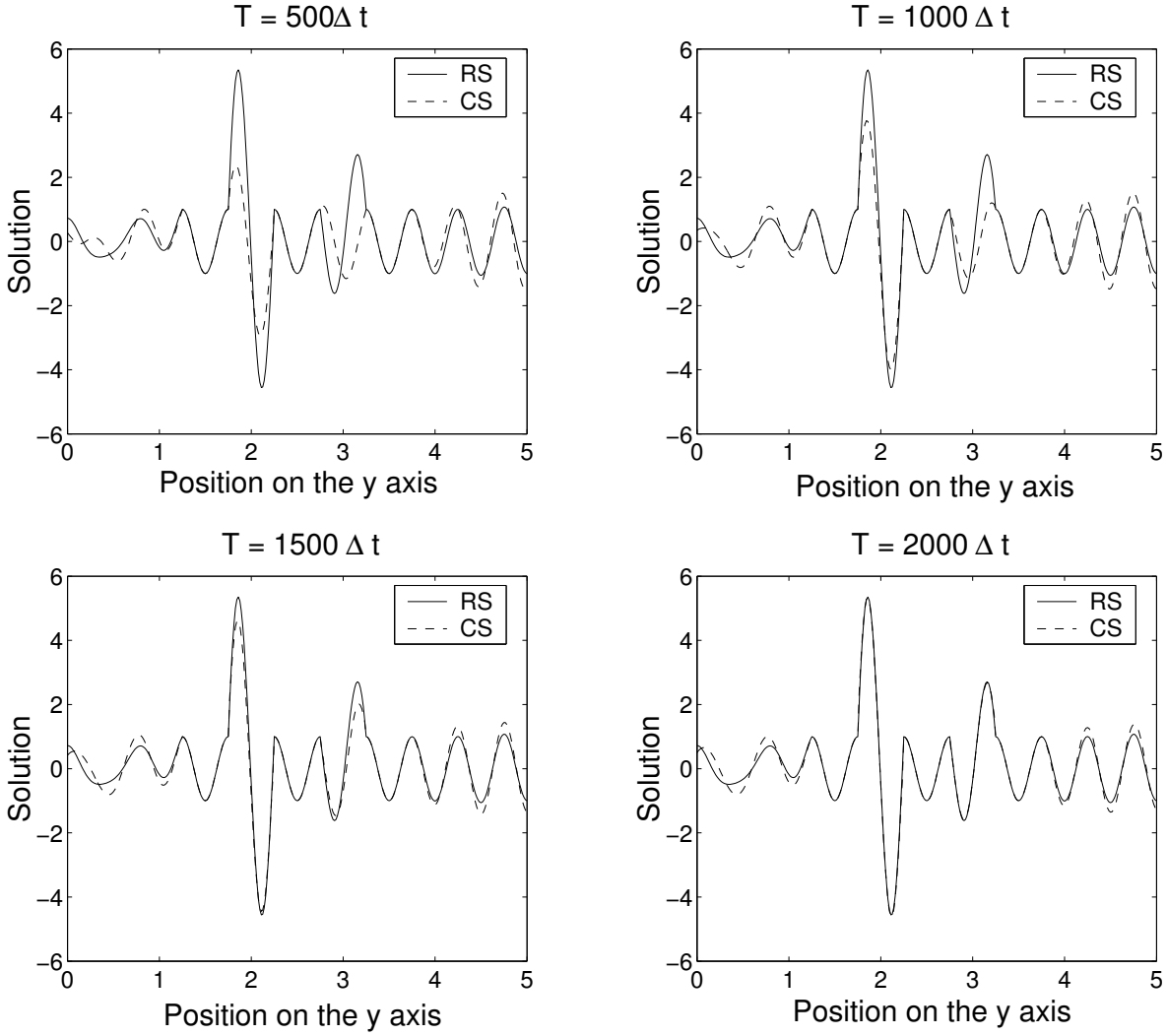

Figure 3.28: Comparison of the reference solution (RS), and the computed solution (CS). In each case a slice of the solution is taken at $y = 2.5$.

Chapter 4

Analysis of the Fictitious Domain Method for a 1D Wave Problem

4.1 Introduction

In Chapter 3 we presented a fictitious domain method and two symmetrized operator splitting schemes for the solution of the wave scattering problem (3.1). In this chapter we will analyze the fictitious domain method, FDDM, and the operator splitting scheme OFDDM, presented in Chapter 3, for a 1D wave propagation problem. We consider the 1D wave equation with a Dirichlet boundary condition. The problem is to find Φ such that

$$\begin{cases} \dfrac{1}{c^2}\dfrac{\partial^2\Phi}{\partial t^2} - \dfrac{\partial^2\Phi}{\partial x^2} = 0,\ x < x_r, \\ \Phi(x = x_r) = 0, \\ \Phi(t = 0) = \Phi_0 \in H^1(\mathbb{R}),\ \ \Phi_0(x = x_r) = 0, \\ \dfrac{d\Phi}{dt}(t = 0) = \Phi_1 \in L^2(\mathbb{R}), \end{cases} \tag{4.1}$$

where $0 \leq x_r < 1$. We will compare the methods FDDM and OFDDM with a finite difference method which we will denote as FDM, and another fictitious domain method, which employs a boundary Lagrange multiplier, introduced in [27, 41, 57]. We will denote

this last method by FDBM.

The outline of this chapter is as follows. In Section 4.2 we describe the fictitious domain formulation FDDM, for the 1D problem (4.1). In Section 4.2.1 we describe the spatial discretization that will be used, and in Section 4.2.2 we will discuss the mass lumping techniques that will be employed. In Section 4.2.3 we discuss the time discretization and present the fully discrete problem. In Sections 4.2.4 we will perform a dispersion analysis for the 1D wave problem to obtain the dispersion relation that applies to the finite difference method and both the fictitious domain methods. In Section 4.3 we present the 1D version of the operator splitting scheme OFDDM, that was introduced in Chapter 3. In Section 4.3.1 we calculate the dispersion relation for the operator splitting scheme. Finally in Section 4.4 we perform a plane wave analysis to obtain the reflection coefficient related to the Dirichlet condition for all the four schemes. In Section 4.5 we present comparisons of all four schemes on the basis of their reflection coefficients.

4.2 A Fictitious Domain Method: FDDM

The fictitious domain formulation with a distributed multiplier, FDDM, proposed in Chapter 3, in the case of the 1D problem (4.1) can be written as:

$$
\left\{
\begin{array}{l}
\text{Find} \quad (\phi, \lambda) \text{ such that :} \\
\\
\phi \in C^1([0,T], H^1(\mathbb{R})), \text{ and } \dfrac{d\phi}{dt} \in C^0([0,T], L^2(\mathbb{R})), \\
\\
\lambda \in L^2(0,T, L^2(x_r, \infty)), \\
\\
\dfrac{1}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \,\mathrm{dx} + \int_{\mathbb{R}} \frac{\partial \phi}{\partial x}\frac{\partial \psi}{\partial x} \,\mathrm{dx} + \int_{x_r}^{\infty} \lambda\psi \,\mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\
\\
\displaystyle\int_{x_r}^{\infty} \mu\phi \,\mathrm{dx} = 0, \ \forall \mu \in L^2(x_r, \infty), \\
\\
\phi(t=0) = \phi_0, \text{ and } \dfrac{d\phi}{dt}(t=0) = \phi_1,
\end{array}
\right. \tag{4.2}
$$

where, the relation between Φ and ϕ is discussed in Chapter 3. As shown in Figure 4.1, the domain ω for the 1D case is the interval (x_r, ∞). Thus, in the fictitious domain formulation, the problem is extended to the entire real line $\mathbb{R}$, and the Dirichlet boundary condition, $\phi(x_r) = 0$, is imposed via the introduction of a distributed Lagrange multiplier λ defined over the domain, $\bar{\omega} = [x_r, \infty)$.

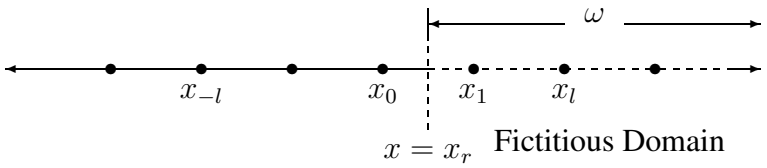

Figure 4.1: The fictitious domain.

4.2.1 Space Discretization

In this section we set up the discrete spatial problem. Let $h > 0,\ h \in \mathbb{R}$ be a parameter. We divide the real line $\mathbb{R}$ into segments $I_l = [lh, (l+1)h],\ l \in \mathbb{Z}$. We will denote $x_l = lh$. We have

$$\mathbb{R} = \cup_{l \in \mathbb{Z}} I_l. \tag{4.3}$$

For the solution ϕ, the 1D finite element space 1D is

$$\mathbf{V}_h = \{\phi_h \in H^1(\mathbb{R}), \forall\, l \in \mathbb{Z}, \phi|_{I_l} \in P_1\}, \tag{4.4}$$

where P_1 is the space of linear continuous functions. Thus, for $\phi_h \in \mathbf{V}_h$ we have

$$\phi_h(x) = \sum_{l \in \mathbb{Z}} \phi_{h,l} w_l(x), \quad \text{with } \ w_l(x) = w(\frac{x}{h} - l), \tag{4.5}$$

where the function $w(x)$ is defined as

$$w(x) = \begin{cases} 1 + x, & \text{if} \quad -1 \leq x \leq 0, \\ 1 - x, & \text{if} \quad 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

We will take $x_r = rh$, with $0 \leq r < 1$. Thus, x_r does not coincide with a nodal point unless $r = 0$. Suppose that the initial functions ϕ_h^0, and ϕ_h^1 are given in the space $\mathbf{V}_h$, and that ϕ_h^0 satisfies the boundary condition $\phi_h^0(x = x_r) = 0$.

We now choose a space for the distributed Lagrange multiplier λ, analogous to the 2D case. Define the sets $\Sigma_h = \{x_l \mid x_l = lh;\ l \in \mathbb{Z}\}$, and

$$\begin{aligned} \Sigma_h^{\bar{\omega}} &= \{x_l \mid x_l \in \Sigma_h \cap [x_r, \infty), \text{dist}(x_l, x_r) \geq h\} \cup \{x_r\}, \\ &= \{x_l \mid l \in \mathbb{N} \text{ and } l \geq 2\} \cup \{x_r\}. \end{aligned} \tag{4.7}$$

Using these sets, the space $\mathbf{\Lambda}_h$ of the Lagrange multipliers is defined as

$$\mathbf{\Lambda}_h = \{\mu_h \mid \mu_h = \sum_{x_l \in \Sigma_h^{\bar{\omega}}} \mu_{h,l} \chi_{(x_l - 1, x_l + 1)},\ \mu_{h,l} \in \mathbb{R}\}, \tag{4.8}$$

where, $\chi_{(a,b)}$ is the characteristic function of the interval (a, b), i.e.,

$$\chi(x) = \begin{cases} 1, & \text{if} \quad a < x < b, \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

We approximate the integrals over $\omega = (x_r, \infty)$ as follows.

$$\int_{x_r}^{\infty} v_h \mu_h \mathrm{dx} \approx h \sum_{x_l \in \Sigma_h^{\bar{\omega}}} v_h(x_l) \mu_h(x_l),\ \forall v_h \in \mathbf{V}_h,\ \forall \mu_h \in \mathbf{\Lambda}_h. \tag{4.10}$$

Thus, the space discrete problem can be written as:

$$\begin{cases} \text{Find} \quad (\phi_h, \lambda_h) \text{ such that :} \\ \phi_h \in C^1([0,T], \mathbf{V}_h), \text{ and } \dfrac{d\phi_h}{dt} \in C^0([0,T], \mathbf{V}_h), \\ \lambda_h \in L^2(0, T, \mathbf{\Lambda}_h), \\ \dfrac{1}{c^2} \displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi_h}{\partial t^2} \psi_h \, \mathrm{dx} + \int_{\mathbb{R}} \frac{\partial \phi_h}{\partial x} \frac{\partial \psi_h}{\partial x} \, \mathrm{dx} + \int_{x_r}^{\infty} \lambda_h \psi_h \mathrm{dx} = 0,\ \forall \psi_h \in \mathbf{V}_h, \\ \displaystyle\int_{x_r}^{\infty} \mu_h \phi_h \mathrm{dx} = 0,\ \forall \mu_h \in \mathbf{\Lambda}_h, \\ \phi_h(t=0) = \phi_{0,h}, \text{ and } \dfrac{d\phi_h}{dt}(t=0) = \phi_{1,h}, \end{cases} \tag{4.11}$$

where $\phi_{0,h}$, and $\phi_{1,h}$, are finite element approximations of the functions ϕ_0, and ϕ_1, respectively.

4.2.2 Mass Lumping Techniques

We will calculate the mass matrix for the discrete problem using the trapezoidal rule in order to obtain a scheme that is explicit in time. The use of quadrature rules to calculate (diagonal) mass matrices is known as *mass lumping*. The trapezoidal rule can be stated as follows. For a continuous function g on the interval $[a, b]$, we have

$$\int_a^b g(x)\ \mathrm{dx} \approx |b-a|\ \left(\frac{g(a)+g(b)}{2}\right). \tag{4.12}$$

Stated briefly, the integral of the function g over the interval $[a, b]$ is its average at the points $x = a$, and $x = b$, multiplied by the length of the interval. Using the trapezoidal rule, the entries of the mass matrix for the discrete scheme (4.11) are,

$$\int_{\mathbb{R}} w_l(x)w_k(x) = \begin{cases} h, & \text{if} \quad l = k, \\ 0, & \text{otherwise.} \end{cases} \tag{4.13}$$

Thus, scheme (4.11) can be written as,

$$\begin{cases} \text{Find} \quad (\phi_{h,l}, \lambda_{h,k}) \text{ such that :} \\ \\ \phi_{h,l} \in C^1([0,T]),\ \forall\ l \in \mathbb{Z},\ \text{and } \lambda_{h,k} \in L^2(0,T),\ k = r,\ \text{or } k \in \mathbb{N}, k \geq 2, \\ \\ \text{(i)}\ \dfrac{1}{c^2}\dfrac{d^2\phi_{h,l}}{dt^2} - \dfrac{\phi_{h,l+1} - 2\phi_{h,l} + \phi_{h,l-1}}{h^2} + \lambda_{h,r}\left((1-r)\delta_{l,0} + r\delta_{l,1}\right) \\ \\ + \displaystyle\sum_{k\geq 2, k\in\mathbb{N}} \lambda_{h,k}\delta_{l,k} = 0,\ \forall\ l\ \in\ \mathbb{Z}, \\ \\ \text{(ii)}\ (1-r)\phi_{h,0} + r\phi_{h,1} = 0, \\ \\ \text{(iii)}\ \phi_{h,l} = 0,\ \text{for } l \in \mathbb{N},\ l \geq 2, \\ \\ \text{(iv)}\ \phi_{h,l}(t=0) = \phi_{0,h,l},\ \text{and } \dfrac{d\phi_{h,l}}{dt}(t=0) = \phi_{1,h,l},\ \forall\ l\ \in\ \mathbb{Z}. \end{cases} \tag{4.14}$$

4.2.3 Time Discretization

The time interval $[0, T]$ is divided into sub intervals $[t^n, t^{n+1}]$ of length Δt, where $t^n = n\Delta t$, for $n \geq 0,\ n \in \mathbb{N}$. The function ϕ_h can be expressed as

$$\phi_h(x, t^n) = \sum_{l \in \mathbb{Z}} \phi_{h,l}^n w_l(x),\ \phi_{h,l}^n \in \mathbb{R}. \tag{4.15}$$

We approximate the time derivative $\dfrac{d^2\phi_{h,l}}{dt^2}$ by a second order centered finite difference method. The fully discrete scheme can then be written as

$$\begin{cases} \text{Find} \quad (\phi_{h,l}^{n+1}, \lambda_{h,k}^{n+1}) \text{ such that } \forall\ l\ \in\ \mathbb{Z}, k = r \text{ or } k\ \in\ \mathbb{N}, k \geq 2: \\ \\ \phi_{h,l}^{n+1} \in \mathbb{R} \text{ and } \lambda_{h,k}^{n+1} \in \mathbb{R}, \\ \\ \text{(i)}\ \dfrac{\phi_{h,l}^{n+1} - 2\phi_{h,l}^n + \phi_{h,l}^{n-1}}{c^2\Delta t^2} - \dfrac{\phi_{h,l+1}^n - 2\phi_{h,l}^n + \phi_{h,l-1}^n}{h^2} \\ \quad + \lambda_{h,r}^{n+1}\left((1-r)\delta_{l,0} + r\delta_{l,1}\right) + \displaystyle\sum_{k \geq 2,\ k \in \mathbb{N}} \lambda_{h,k}^{n+1}\delta_{l,k} = 0,\ \forall\ l\ \in\ \mathbb{Z}, \\ \\ \text{(ii)}\ (1-r)\phi_{h,0}^{n+1} + r\phi_{h,1}^{n+1} = 0, \\ \\ \text{(iii)}\ \phi_{h,l}^{n+1} = 0, \text{ for } l \in \mathbb{N},\ l \geq 2, \\ \\ \text{(iv)}\ \phi_{h,l}^0 = \phi_{0,h,l}, \text{ and } \dfrac{\phi_{h,l}^1 - \phi_{h,l}^{-1}}{2\Delta t} = \phi_{1,h,l},\ \forall\ l\ \in\ \mathbb{Z}. \end{cases} \tag{4.16}$$

To obtain an expression for the distributed Lagrange multiplier, we write the equation for $\phi_{h,l}^{n+1}$ in (4.16, i) for different nodes separately.

- For $l < 0$,
$$\frac{\phi_{h,l}^{n+1} - 2\phi_{h,l}^n + \phi_{h,l}^{n-1}}{c^2\Delta t^2} - \frac{\phi_{h,l+1}^n - 2\phi_{h,l}^n + \phi_{h,l-1}^n}{h^2} = 0. \tag{4.17}$$

- For $l = 0$
$$\frac{\phi_{h,0}^{n+1} - 2\phi_{h,0}^n + \phi_{h,0}^{n-1}}{c^2\Delta t^2} - \frac{\phi_{h,1}^n - 2\phi_{h,0}^n + \phi_{h,-1}^n}{h^2} + \lambda_{h,r}^{n+1}(1-r) = 0. \tag{4.18}$$

- For $l = 1$

$$\frac{\phi_{h,1}^{n+1} - 2\phi_{h,1}^{n} + \phi_{h,1}^{n-1}}{c^2\Delta t^2} - \frac{\phi_{h,2}^{n} - 2\phi_{h,1}^{n} + \phi_{h,0}^{n}}{h^2} + \lambda_{h,r}^{n+1} r = 0. \tag{4.19}$$

- For $l \geq 2$

$$\frac{\phi_{h,l}^{n+1} - 2\phi_{h,l}^{n} + \phi_{h,l}^{n-1}}{c^2\Delta t^2} - \frac{\phi_{h,l+1}^{n} - 2\phi_{h,l}^{n} + \phi_{h,l-1}^{n}}{h^2} + \lambda_{h,l}^{n+1} = 0. \tag{4.20}$$

Multiplying the equation for $l = 0$ by $1 - r$ and the equation for $l = 1$ by r, adding and using the constraint $(1-r)\phi_{h,0}^{k} + r\phi_{h,1}^{k} = 0$, for $k = n-1, n, n+1$, and $\phi_{h,l}^{n+1} = 0,$ for $l \geq 2$, we obtain

$$\begin{aligned} &\text{(i)} \quad \lambda_{h,r}^{n+1} = \frac{(1-3r)\phi_{h,1}^{n} - (2-3r)\phi_{h,0}^{n} + (1-r)\phi_{h,-1}^{n}}{(1-2r+2r^2)h^2}, \\ &\text{(ii)} \quad \lambda_{h,2}^{n+1} = \frac{\phi_{h,1}^{n}}{h^2}, \\ &\text{(iii)} \quad \lambda_{h,l}^{n+1} = 0, \ \forall\, l \geq 3. \end{aligned} \tag{4.21}$$

Remark 8 *If $r = 0$, scheme* (4.16) *reduces to the finite difference scheme with centered differencing in space and time for $l \leq 0$. The constraint equations reduce to*

$$\begin{aligned} \phi_{h,0}^{n+1} &= 0 \\ \phi_{h,l}^{n+1} &= 0, \textit{ for } l \geq 2. \end{aligned} \tag{4.22}$$

Also, from (4.21, i*), $\lambda_{h,r}^{n+1}$ is given by*

$$\lambda_{h,r}^{n+1} = \frac{\phi_{h,1}^{n} - 2\phi_{h,0}^{n} + \phi_{h,-1}^{n}}{h^2}. \tag{4.23}$$

Substituting the above in (4.18) *the equation for $\phi_{h,0}^{n+1}$ becomes*

$$\frac{\phi_{h,0}^{n+1} - 2\phi_{h,0}^{n} + \phi_{h,0}^{n-1}}{c^2\Delta t^2} - \frac{\phi_{h,1}^{n} - 2\phi_{h,0}^{n} + \phi_{h,-1}^{n}}{h^2} + \frac{\phi_{h,1}^{n} - 2\phi_{h,0}^{n} + \phi_{h,-1}^{n}}{h^2} = 0, \tag{4.24}$$

which implies that,

$$\phi_{h,0}^{n+1} - 2\phi_{h,0}^{n} + \phi_{h,0}^{n-1} = 0. \tag{4.25}$$

This is equivalent to $\phi_{h,0}^{n+1} = 0$, given that $\phi_{h,0}^{0} = 0$ and $\phi_{h,0}^{1} = 0$.

4.2.4 Dispersion Analysis

The *dispersion relation* is an equation that relates the angular frequency ρ, the wave number k, and the speed of propagation c. To obtain the dispersion relation for the fictitious domain method, we perform a plane wave analysis of the scheme in the absence of the distributed Lagrange multiplier. As a result, the dispersion relation for this scheme is the same as that for the finite difference scheme with centered differences in space and time. In other words, if we assume the propagation of a plane wave

$$\phi_{h,l}^{n} = \mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}khl}, \tag{4.26}$$

in the scheme

$$\frac{\phi_{h,l}^{n+1} - 2\phi_{h,l}^{n} + \phi_{h,l}^{n-1}}{c^2\Delta t^2} - \frac{\phi_{h,l+1}^{n} - 2\phi_{h,l}^{n} + \phi_{h,l-1}^{n}}{h^2} = 0, l \in \mathbb{Z}, l \leq -1, \tag{4.27}$$

we obtain

$$\begin{aligned}\frac{1}{c^2\Delta t^2}\mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}khl}\left(\mathrm{e}^{-\mathrm{i}\rho\Delta t} - 2 + \mathrm{e}^{\mathrm{i}\rho\Delta t}\right) \\ - \frac{1}{h^2}\mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}khl}\left(\mathrm{e}^{-\mathrm{i}kh} - 2 + \mathrm{e}^{\mathrm{i}kh}\right) = 0.\end{aligned} \tag{4.28}$$

On simplification, we have

$$\frac{1}{c^2\Delta t^2}4\sin^2\left(\frac{\rho\Delta t}{2}\right) - \frac{1}{h^2}4\sin^2\left(\frac{kh}{2}\right) = 0. \tag{4.29}$$

Thus, we obtain the dispersion relation

$$\sin\left(\frac{\rho\Delta t}{2}\right) = \frac{c\Delta t}{h}\sin\left(\frac{kh}{2}\right). \tag{4.30}$$

Solving for k in the above, we have

$$k = \frac{2}{h}\sin^{-1}\left(\frac{h}{c\Delta t}\sin\left(\frac{\rho\Delta t}{2}\right)\right). \tag{4.31}$$

Solving for ρ in (4.30), we have

$$\rho = \frac{2}{\Delta t}\sin^{-1}\left(\frac{c\Delta t}{h}\sin\left(\frac{kh}{2}\right)\right). \tag{4.32}$$

Remark 9 *In the dispersion relation* (4.31)*, if we choose*

$$\eta = \frac{c\Delta t}{h} = 1, \tag{4.33}$$

then, the dispersion relation simplifies to

$$k = \frac{\rho}{c}, \tag{4.34}$$

which is the dispersion relation for the continuous 1D wave equation. The value $\eta = 1$ *is a magic number for which the solution of the finite difference scheme* (4.27)*, is the exact solution to the 1D wave equation [124].*

4.3 An Operator Splitting Scheme

In this section we consider the operator splitting scheme OFDDM, introduced in Chapter 3 for the 1D wave problem (4.1). As before, we define the velocity u to be the time derivative

$$u = 2\frac{\partial \phi}{\partial t}. \tag{4.35}$$

On the interval (t^n, t^{n+1}), given (ϕ^n, u^n), we will solve three subproblems.

•: **Operator Splitting Scheme** OFDDM$_s$:

For $n = 0, 1, 2, \ldots, N-1$, given (ϕ^n, u^n) solve:

- SUBPROBLEM $(1)_m$: On $\mathbb{R} \times (t^n, t^{n+1/2})$, find $(\phi^{n+1/2}, \lambda^{n+1/2})$ via

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \, \mathrm{dx} + \int_{x_r}^{\infty} \lambda\psi \, \mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\ \displaystyle\int_{x_r}^{\infty} \mu\phi \, \mathrm{dx} = 0, \ \forall \mu \in L^2(x_r, \infty), \\ \phi(t^n) = \phi^n, \text{ and } \phi_t(t^n) = \dfrac{1}{2}u^n. \end{cases} \tag{4.36}$$

  $\phi^{n+1/2} = \phi(t^{n+1/2}), \ \lambda^{n+1/2} = \lambda(t^{n+1/2})$.

  Calculate $u^{n+1/2} = 2\dfrac{\partial \phi}{\partial t}|^{n+1/2}$ as follows:

On $\mathbb{R} \times (t^{n+1/2}, t^{n+1})$, find $(\hat{\phi}^{n+1}, \hat{\lambda}^{n+1})$ via

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \, \mathrm{dx} + \int_{x_r}^{\infty} \lambda\psi \, \mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\ \displaystyle\int_{x_r}^{\infty} \mu\phi \, \mathrm{dx} = 0, \ \forall \mu \in L^2(x_r, \infty), \\ \phi(t^{n+1/2}) = \phi^{n+1/2}, \text{ and } \phi(t^n) = \phi^n. \end{cases} \tag{4.37}$$

$$\hat{\phi}^{n+1} = \phi(t^{n+1}), \ \hat{\lambda}^{n+1} = \lambda(t^{n+1}), \text{ and } u^{n+1/2} = 2\frac{\hat{\phi}^{n+1} - \phi^n}{\Delta t}. \tag{4.38}$$

- SUBPROBLEM $(2)_m$: On $\mathbb{R} \times (t^n, t^{n+1})$, find $\tilde{\phi}^{n+1}$ via

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \, \mathrm{dx} + \int_{\mathbb{R}} \frac{\partial \phi}{\partial x}\frac{\partial \psi}{\partial x} \, \mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\ \phi(t^n) = \phi^{n+1/2}, \text{ and } \phi_t(t^n) = \dfrac{1}{2}u^{n+1/2}. \end{cases} \tag{4.39}$$

$\tilde{\phi}^{n+1} = \phi(t^{n+1})$.

Calculate $\tilde{u}^{n+1} = 2\dfrac{\partial \phi}{\partial t}|^{n+1}$ as follows:

On $\mathbb{R} \times (t^{n+1}, t^{n+2})$, find $\hat{\phi}^{n+2}$ satisfying:

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \, \mathrm{dx} + \int_{\mathbb{R}} \frac{\partial \phi}{\partial x}\frac{\partial \psi}{\partial x} \, \mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\ \phi(t^n) = \phi^{n+1/2}, \text{ and } \phi(t^{n+1}) = \tilde{\phi}^{n+1}. \end{cases} \tag{4.40}$$

$$\hat{\phi}^{n+2} = \phi(t^{n+2}), \text{ and } \ \tilde{u}^{n+1} = \frac{\hat{\phi}^{n+2} - \phi^{n+1/2}}{\Delta t}. \tag{4.41}$$

- SUBPROBLEM $(3)_m$: On $\mathbb{R} \times (t^{n+1/2}, t^{n+1})$, find $(\phi^{n+1}, \lambda^{n+1})$ via

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \, \mathrm{dx} + \int_{x_r}^{\infty} \lambda\psi \mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\ \displaystyle\int_{x_r}^{\infty} \mu\phi \mathrm{dx} = 0, \ \forall \mu \in L^2(x_r, \infty), \\ \phi(t^{n+1/2}) = \tilde{\phi}^{n+1}, \text{ and } \phi_t(t^{n+1/2}) = \dfrac{1}{2}\tilde{u}^{n+1}. \end{cases} \tag{4.42}$$

$\phi^{n+1} = \phi(t^{n+1}), \ \lambda^{n+1} = \lambda(t^{n+1})$

Calculate $u^{n+1} = 2\dfrac{\partial \phi}{\partial t}|^{n+1}$ as follows:

On $\mathbb{R} \times (t^{n+1}, t^{n+3/2})$, find $(\hat{\phi}^{n+3/2}, \hat{\lambda}^{n+3/2})$ via

$$\begin{cases} \dfrac{2}{c^2}\displaystyle\int_{\mathbb{R}} \frac{\partial^2 \phi}{\partial t^2}\psi \, \mathrm{dx} + \int_{x_r}^{\infty} \lambda\psi \mathrm{dx} = 0, \ \forall \psi \in H^1(\mathbb{R}), \\ \displaystyle\int_{x_r}^{\infty} \mu\phi \mathrm{dx} = 0, \ \forall \mu \in L^2(x_r, \infty), \\ \phi(t^{n+1/2}) = \tilde{\phi}^{n+1}, \text{ and } \phi(t^{n+1}) = \phi^{n+1}. \end{cases} \tag{4.43}$$

$$\hat{\phi}^{n+3/2} = \phi(t^{n+3/2}), \ \hat{\lambda}^{n+3/2} = \lambda(t^{n+3/2}), \text{ and } u^{n+1} = 2\frac{\hat{\phi}^{n+3/2} - \tilde{\phi}^{n+1}}{\Delta t}. \tag{4.44}$$

Using the space and time discretizations proposed in Sections 4.2.1, 4.2.2, and 4.2.3, we obtain the fully discrete operator splitting scheme:

• **Operator Splitting Scheme** OFDDM:

- SUBPROBLEM $(1)_h$: Find $(\phi_h^{n+1/2}, \lambda_h^{n+1/2})$ satisfying:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\phi_{h,l}^{n+1/2} - 2\phi_{h,l}^{n} + \bar{\phi}_{h,l}^{n-1/2}}{(\Delta t/2)^2} + \lambda_{h,r}^{n+1/2}\left((1-r)\delta_{l,0} + r\delta_{l,1}\right) \\ \quad + \displaystyle\sum_{k\geq 2, \ k\in\mathbb{N}} \lambda_{h,\,k}^{n+1/2}\delta_{l,k} = 0, \ \forall \, l \in \ \mathbb{Z}, \\ (1-r)\phi_{h,0}^{n+1/2} + r\phi_{h,1}^{n+1/2} = 0, \\ \phi_{h,l}^{n+1/2} = 0, \text{ for } l \in \mathbb{N}, \ l \geq 2. \\ \phi_{h,l}^{n+1/2} - \bar{\phi}_{h,l}^{n-1/2} = \dfrac{\Delta t}{2} u_{h,l}^{n}. \end{cases} \tag{4.45}$$

Calculate $u_h^{n+1/2} = 2\dfrac{\partial \phi_h}{\partial t}|^{n+1/2}$ as follows: Find $(\hat{\phi}_h^{n+1}, \hat{\lambda}_h^{n+1})$ satisfying:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\hat{\phi}_{h,l}^{n+1} - 2\phi_{h,l}^{n+1/2} + \phi_{h,l}^{n}}{(\Delta t/2)^2} + \hat{\lambda}_{h,r}^{n+1}\left((1-r)\delta_{l,0} + r\delta_{l,1}\right) \\ \quad + \displaystyle\sum_{k\geq 2, \ k\in\mathbb{N}} \hat{\lambda}_{h,\,k}^{n+1}\delta_{l,k} = 0, \ \forall \, l \ \in \ \mathbb{Z} \\ (1-r)\hat{\phi}_{h,0}^{n+1} + r\hat{\phi}_{h,1}^{n+1} = 0, \\ \hat{\phi}_{h,l}^{n+1} = 0, \text{ for } l \in \mathbb{N}, \ l \geq 2. \end{cases} \tag{4.46}$$

$$\text{Set } u_{h,l}^{n+1/2} = 2\frac{\hat{\phi}_{h,l}^{n+1} - \phi_{h,l}^{n}}{\Delta t},\ \forall\, l\ \in\ \mathbb{Z}. \tag{4.47}$$

- SUBPROBLEM $(2)_h$: Find $\tilde{\phi}_h^{n+1}$ satisfying:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\tilde{\phi}_{h,l}^{n+1} - 2\phi_{h,l}^{n+1/2} + \bar{\phi}_{h,l}^{n-1}}{\Delta t^2} - \dfrac{\phi_{h,l+1}^{n+1/2} - 2\phi_{h,l}^{n+1/2} + \phi_{h,l-1}^{n+1/2}}{h^2} = 0, \\ \dfrac{\tilde{\phi}_{h,l}^{n+1} - \bar{\phi}_{h,l}^{n-1}}{\Delta t} = u_{h,l}^{n+1/2}. \end{cases} \tag{4.48}$$

  Calculate $\tilde{u}_h^{n+1} = 2\dfrac{\partial \phi_h}{\partial t}|^{n+1}$ as follows: Find $\hat{\phi}_h^{n+2}$ satisfying:

$$\frac{2}{c^2}\frac{\hat{\phi}_{h,l}^{n+2} - 2\tilde{\phi}_{h,l}^{n+1} + \phi_{h,l}^{n+1/2}}{\Delta t^2} - \frac{\tilde{\phi}_{h,l+1}^{n+1} - 2\tilde{\phi}_{h,l}^{n+1} + \tilde{\phi}_{h,l-1}^{n+1}}{h^2} = 0, \tag{4.49}$$

$$\text{Set } \tilde{u}_{h,l}^{n+1} = \frac{\hat{\phi}_{h,l}^{n+2} - \phi_{h,l}^{n+1/2}}{\Delta t},\ \forall l\ \in\ \mathbb{Z}. \tag{4.50}$$

- SUBPROBLEM $(3)_h$: Find $(\phi_h^{n+1}, \lambda_h^{n+1})$ satisfying:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\phi_{h,l}^{n+1} - 2\tilde{\phi}_{h,l}^{n+1} + \bar{\phi}_{h,l}^{n-1}}{(\Delta t/2)^2} + \lambda_{h,r}^{n+1}\left((1-r)\delta_{l,0} + r\delta_{l,1}\right), \\ \quad + \displaystyle\sum_{k\geq 2,\ k\in\mathbb{N}} \lambda_{h,\,k}^{n+1}\delta_{l,\,k} = 0,\ \forall\, l\ \in\ \mathbb{Z}, \\ (1-r)\phi_{h,0}^{n+1} + r\phi_{h,1}^{n+1} = 0, \\ \phi_{h,l}^{n+1} = 0,\ \text{for } l \in \mathbb{N},\ l \geq 2, \\ \dfrac{\phi_{h,l}^{n+1} - \bar{\phi}_{h,l}^{n-1}}{\Delta t} = \dfrac{1}{2}\tilde{u}_{h,l}^{n+1}. \end{cases} \tag{4.51}$$

  Calculate $u_h^{n+1} = 2\dfrac{\partial \phi_h}{\partial t}|^{n+1}$ as follows: Find $(\hat{\phi}_h^{n+3/2}, \hat{\lambda}_h^{n+3/2})$ satisfying:

$$\begin{cases} \dfrac{2}{c^2}\dfrac{\hat{\phi}_{h,l}^{n+3/2} - 2\phi_{h,l}^{n+1} + \tilde{\phi}_{h,l}^{n+1}}{\Delta t^2} + \hat{\lambda}_{h,r}^{n+3/2}\left((1-r)\delta_{l,0} + r\delta_{l,1}\right), \\ \quad + \displaystyle\sum_{k\geq 2,\ k\in\mathbb{N}} \hat{\lambda}_{h,\,k}^{n+3/2}\delta_{l,\,k} = 0,\ \forall\, l\ \in\ \mathbb{Z}, \\ (1-r)\hat{\phi}_{h,0}^{n+3/2} + r\hat{\phi}_{h,1}^{n+3/2} = 0, \\ \hat{\phi}_{h,l}^{n+3/2} = 0,\ \text{for } l \in \mathbb{N},\ l \geq 2, \end{cases} \tag{4.52}$$

$$\text{Set } u_{h,l}^{n+1} = 2\frac{\hat{\phi}_{h,l}^{n+3/2} - \tilde{\phi}_{h,l}^{n+1}}{\Delta t},\ \forall\, l\ \in\ \mathbb{Z}. \tag{4.53}$$

We perform an analysis of the operator splitting scheme OFDDM, in the special case when

$$\eta = \frac{c\Delta t}{h} = 1. \tag{4.54}$$

In this case, eliminating the intermediate steps, i.e., subproblem $(2)_h$, eliminating terms in u_h^k and λ_h^k, we obtain an equation for ϕ_h^{n+1} in terms of ϕ_h^n and ϕ_h^{n-1}. This process involves some tedious, though mechanical calculations, and we have used the software MATHEMATICA to this end. Here we present the final equations for $\phi_{h,-1}^{n+1}$, $\phi_{h,0}^{n+1}$, and $\phi_{h,1}^{n+1}$. Let us define the operator S_h as

$$S_h\phi_{h,l}^k = \phi_{h,l+1}^k - 2\phi_{h,l}^k + \phi_{h,l-1}^k,\ \forall\, k\ \in\ \mathbb{N}\cup\{0\},\ l\ \in\ \mathbb{Z}. \tag{4.55}$$

We then define the operator $\mathcal{B}_h$ to be,

$$\mathcal{B}_h\phi_{h,l}^k = 32\phi_{h,l}^k + 16S_h\phi_{h,l}^k + S_h^2\phi_{h,l}^k - 16\phi_{h,l}^{k-1},\ \forall\, k\ \in\ \mathbb{N}\cup\{0\},\ l\ \in\ \mathbb{Z}. \tag{4.56}$$

The equations for $\phi_{h,l}^{n+1}$ at the nodes $l = -1, 0, 1$, are

$$\begin{aligned}
&\text{(i)} \quad \phi_{h,-1}^{n+1} = \frac{1}{16}\mathcal{B}_h\phi_{h,-1}^n,\\
&\text{(ii)} \quad \phi_{h,0}^{n+1} = \frac{r^2}{16\left((1-r)^2+r^2\right)}\mathcal{B}_h\phi_{h,0}^n - \frac{r(1-r)}{16\left((1-r)^2+r^2\right)}\mathcal{B}_h\phi_{h,1}^n,\\
&\text{(iii)} \quad \phi_{h,1}^{n+1} = \frac{(1-r)^2}{16\left((1-r)^2+r^2\right)}\mathcal{B}_h\phi_{h,1}^n - \frac{r(1-r)}{16\left((1-r)^2+r^2\right)}\mathcal{B}_h\phi_{h,0}^n.
\end{aligned} \tag{4.57}$$

The equation (4.57, i), remains true for $l\ \leq\ -1$. The corresponding equation for the fictitious domain method FDDM, as given in (4.17) is,

$$\phi_{h,-1}^{n+1} = \frac{1}{16}\bar{\mathcal{B}}_h\phi_{h,-1}^n, \tag{4.58}$$

with

$$\bar{\mathcal{B}}_h\phi_{h,l}^k = 32\phi_{h,l}^k + 16S_h\phi_{h,l}^k - 16\phi_{h,l}^{k-1},\ \forall\, k\ \in\ \mathbb{N}\cup\{0\},\ l\ \in\ \mathbb{Z}. \tag{4.59}$$

Thus, the operator splitting scheme, introduces the extra term $\frac{1}{16}S_h^2\phi_{h,-1}^n$ in the right hand side of the equation for $\phi_{h,-1}^{n+1}$ (4.57, i), as can be seen from the definition of $\mathcal{B}_h$ in (4.56).

4.3.1 The Dispersion Relation for the Operator Splitting Scheme

As before, the dispersion relation is calculated in the absence of the distributed Lagrange multiplier. Again, we assume the propagation of a plane wave

$$\phi_{h,\,l}^{n} = \mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}khl},\ \forall\, k\ \in\ \mathbb{N}\cup\{0\},\ l\ \in\ \mathbb{Z}, \tag{4.60}$$

in the equation (4.57, i). For $l \leq -1$, from (4.55) and (4.56), we have

$$\begin{aligned}\phi_{h,l}^{n+1} &= \frac{1}{16}\mathcal{B}_h\phi_{h,l}^{n},\\ &= \frac{1}{16}\{32\phi_{h,l}^{n} + 16\phi_{h,l+1}^{n} - 32\phi_{h,l}^{n} + 16\phi_{h,l-1}^{n} + \phi_{h,l+2}^{n} - 2\phi_{h,l+1}^{n} + \phi_{h,l}^{n},\\ &- 2\phi_{h,l+1}^{n} + 4\phi_{h,l}^{n} - 2\phi_{h,l-1}^{n} + \phi_{h,l}^{n} - 2\phi_{h,l-1}^{n} + \phi_{h,l-2}^{n} - 16\phi_{h,l}^{n-1}\}.\end{aligned} \tag{4.61}$$

Collecting like terms together, we get

$$16\phi_{h,l}^{n+1} = \phi_{h,l+2}^{n} + 12\phi_{h,l+1}^{n} + 6\phi_{h,l}^{n} + 12\phi_{h,l-1}^{n} + \phi_{h,l-2}^{n} - 16\phi_{h,l}^{n-1}. \tag{4.62}$$

Substituting (4.60) in (4.62), and subtracting $32\phi_{h,l}^{n}$ from both sides of the equation we have

$$16\phi_{h,l}^{n}\left(\mathrm{e}^{-\mathrm{i}\rho\Delta t} - 2 + \mathrm{e}^{\mathrm{i}\rho\Delta t}\right) = \phi_{h,l+2}^{n} + 12\phi_{h,l+1}^{n} - 26\phi_{h,l}^{n} + 12\phi_{h,l-1}^{n} + \phi_{h,l-2}^{n}. \tag{4.63}$$

Dividing both sides by $\mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}khl}$, simplifying terms and using the identity

$$\sin^2\theta = 4\sin^2(\theta/2) - 4\sin^4(\theta/2), \tag{4.64}$$

we get

$$\begin{aligned}-64\sin^2\left(\frac{\rho\Delta t}{2}\right) &= \mathrm{e}^{-2\mathrm{i}kh} + 12\mathrm{e}^{-\mathrm{i}kh} - 26 + 12\mathrm{e}^{\mathrm{i}kh} + \mathrm{e}^{2\mathrm{i}kh},\\ &= -4\sin^2(kh) - 48\sin^2\left(\frac{kh}{2}\right),\\ &= -64\sin^2\left(\frac{kh}{2}\right) + 16\sin^4\left(\frac{kh}{2}\right).\end{aligned} \tag{4.65}$$

Thus, we get the dispersion relation for the operator splitting scheme OFDDM, to be

$$\sin^2\left(\frac{\rho\Delta t}{2}\right) = \sin^2\left(\frac{kh}{2}\right)\left(1 - \frac{1}{4}\sin^2\left(\frac{kh}{2}\right)\right). \tag{4.66}$$

Comparing, with the dispersion relation for the non split scheme (4.30), we see that there is an additional term $(1/4)\sin^4(\frac{kh}{2})$, that is introduced by the operator splitting. Solving for ρ, in (4.66) we get

$$\rho = \frac{2c}{h}\sin^{-1}\left(\sin(\frac{kh}{2})\sqrt{\left(1-\frac{1}{4}\sin^2(\frac{kh}{2})\right)}\right) \tag{4.67}$$

Expanding ρ as a series in h, we have

$$\rho = kc - \frac{ch^2k^3}{32} - \frac{ch^4k^5}{2048} + \mathcal{O}(h^6). \tag{4.68}$$

If we compare (4.68) with (4.34), we see that

$$[\rho]_{\text{OFDDM}} = [\rho]_{\text{Exact}} + \mathcal{O}(h^2), \tag{4.69}$$

under the condition $\eta = 1$, as given in (4.33).

4.4 A Plane Wave Analysis

In this section, we perform a plane wave analysis of different schemes for the numerical solution of the 1D wave problem (4.1). We will calculate the *reflection coefficient* in each case, and compare the different schemes on this basis. The four different schemes to be considered here are

- The finite difference method FDM,
- The fictitious domain method with a distributed multiplier FDDM,
- The operator splitting scheme OFDDM.
- The fictitious domain method with a boundary multiplier FDBM,

The 1D wave equation (4.1) satisfies a solution of the form,

$$\phi(x,t) = \mathrm{e}^{-\mathrm{i}\rho t}\left(\mathrm{e}^{-\mathrm{i}k(x-x_r)} + R\mathrm{e}^{\mathrm{i}k(x-x_r)}\right). \tag{4.70}$$

The Dirichlet condition $\phi(x = x_r) = 0$ implies that the reflection coefficient R_{cont} is given by

$$R_{\text{cont}} = -1. \tag{4.71}$$

4.4.1 A Finite Difference Method: FDM

In the finite difference scheme, the Dirichlet boundary condition is moved to the nodal point $x_0 = 0$. We consider the finite difference scheme with centered differences in space and time. This scheme can be written as

$$\begin{cases} \text{Find} \quad \phi_{h,l}^{n+1} \text{ such that } \forall\, l \leq 0,\ n \ \in\ \mathbb{N} \cup \{0\} : \\ \\ \phi_{h,l}^{n+1} \in \mathbb{R}, \\ \\ \text{(i)}\ \dfrac{\phi_{h,l}^{n+1} - 2\phi_{h,l}^{n} + \phi_{h,l}^{n-1}}{c^2 \Delta t^2} - \dfrac{\phi_{h,l+1}^{n} - 2\phi_{h,l}^{n} + \phi_{h,l-1}^{n}}{h^2} = 0, \\ \\ \text{(ii)}\ \phi_{h,0}^{n+1} = 0, \\ \\ \text{(iii)}\ \phi_{h,l}^{0} = \phi_{0,\,h,l},\ \text{and}\ \dfrac{\phi_{h,l}^{1} - \phi_{h,l}^{-1}}{2\Delta t} = \phi_{1,\,h,l}, \\ \\ \text{(iv)}\ \phi_{0,h,0} = 0,\ \phi_{1,h,0} = 0. \end{cases} \tag{4.72}$$

In this case we look for a solution in the form

$$\phi_{h,\,l}^{n} = \mathrm{e}^{-\mathrm{i}\rho n \Delta t} \left(\mathrm{e}^{-\mathrm{i}kh(l-r)} + R_{\text{FDM}} \mathrm{e}^{\mathrm{i}kh(l-r)} \right), \text{ if } \quad l \leq 0. \tag{4.73}$$

This gives the superposition of the incident and the reflected wave in the domain $l \leq 0$. Using the condition $\phi_{h,0}^{n} = 0$ in (4.73), we have

$$R_{\text{FDM}} = -\mathrm{e}^{2\mathrm{i}hkr}. \tag{4.74}$$

As a series in h, we have

$$R_{\text{FDM}} = -1 - 2\mathrm{i}hkr + \mathcal{O}(h^2). \tag{4.75}$$

Remark 10 *We note that, if* $r = 0$ *in* (4.74)*, then we have* $R_{\mathrm{FDM}} = -1 = R_{\mathrm{cont}}$*. Otherwise, the numerically reflected wave in the* FDM *scheme is identical to the reflected wave in the continuous case, with a phase error of* $2\mathrm{i}hkr$*.*

4.4.2 A Fictitious Domain Method with a Distributed Multiplier: FDDM

In this case we look for a solution of the form

$$\phi^n_{h,\,l} = \begin{cases} \mathrm{e}^{-\mathrm{i}\rho n\Delta t}\left(\mathrm{e}^{-\mathrm{i}kh(l-r)} + R_{\mathrm{FDDM}}\mathrm{e}^{\mathrm{i}kh(l-r)}\right), & \text{if} \quad l \leq 0, \\ T_{\mathrm{FDDM}}\mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}kh(l-r)}, & \text{if} \quad l = 1, \\ 0, & \text{if} \quad l \geq 2. \end{cases} \tag{4.76}$$

where T_{FDDM} is the transmission coefficient and,

$$\lambda^{n+1}_{h,k} = \begin{cases} \lambda_r\mathrm{e}^{-\mathrm{i}\rho n\Delta t}, & \text{if} \quad k = r, \\ \lambda_2\mathrm{e}^{-\mathrm{i}\rho n\Delta t}, & \text{if} \quad k = 2, \\ 0, & \text{if} \quad k \geq 3. \end{cases} \tag{4.77}$$

Substituting the expressions for $\phi_{h,-1}$, $\phi^n_{h,0}$, and $\phi^n_{h,1}$ from (4.76) in (4.18), and using the dispersion relation (4.30), we get

$$R_{\mathrm{FDDM}}\mathrm{e}^{\mathrm{i}kh(1-r)} - T_{\mathrm{FDDM}}\mathrm{e}^{-\mathrm{i}kh(1-r)} + (1-r)h^2\lambda_r = -\mathrm{e}^{-\mathrm{i}kh(1-r)}. \tag{4.78}$$

Next, substituting the expressions for $\phi_{h,0}$, $\phi^n_{h,1}$, and $\phi^n_{h,2}$ from (4.76) in (4.19), we get

$$-R_{\mathrm{FDDM}}\mathrm{e}^{-\mathrm{i}khr} + T_{\mathrm{FDDM}}\mathrm{e}^{\mathrm{i}khr}(1+\mathrm{e}^{-2\mathrm{i}kh}) + rh^2\lambda_r = \mathrm{e}^{\mathrm{i}khr}. \tag{4.79}$$

Lastly, substituting for $\phi_{h,\,0}$ and $\phi_{h,\,1}$ in the constraint equation (4.16, ii), we have

$$R_{\mathrm{FDDM}}(1-r)\mathrm{e}^{-\mathrm{i}khr} + T_{\mathrm{FDDM}}r\mathrm{e}^{-\mathrm{i}kh(1-r)} = -(1-r)\mathrm{e}^{\mathrm{i}khr}. \tag{4.80}$$

Thus, we have to solve a system of three equations which, written in matrix form are

$$\begin{bmatrix} \xi^{1-r} & -\xi^{-(1-r)} & (1-r) \\ -\xi^{-r} & \xi^r(1+\xi^{-2}) & r \\ (1-r)\xi^{-r} & r\xi^{-(1-r)} & 0 \end{bmatrix} \begin{bmatrix} R \\ T \\ h^2\lambda_r \end{bmatrix} = \begin{bmatrix} -\xi^{-(1-r)} \\ \xi^r \\ -(1-r)\,\xi^r \end{bmatrix}, \tag{4.81}$$

where, in the above

$$\xi = \mathrm{e}^{\mathrm{i}kh}. \tag{4.82}$$

The determinant of system (4.81) is

$$\det\ = -\frac{2r(1-r)}{\xi} - r^2 - (1-r)^2(1+\frac{1}{\xi^2}). \tag{4.83}$$

If the determinant of system (4.81) is nonzero, then the unique solution of (4.81) is given to be

$$\begin{cases} R = \dfrac{-\xi^{2r}\left((\xi(1-r)+r)^2 + (1-r)^2\right)}{(1-r+\xi r)^2 + \xi^2(1-r)^2}, \\ T = \dfrac{\xi r(1-r)(1-\xi)^2}{(1-r+\xi r)^2 + \xi^2(1-r)^2}, \\ \lambda_r = \dfrac{\xi^{r-1}(\xi^2-1)\left((1-r)(1+\xi^2)+\xi r\right)}{h^2\left((1-r+\xi r)^2 + \xi^2(1-r)^2\right)}, \end{cases} \tag{4.84}$$

From (4.21, ii) we have

$$\begin{aligned} \lambda_2 &= \frac{1}{h^2} T \xi^{r-1} \\ &= \frac{\xi^r r(1-r)(1-\xi)^2}{((1-r+\xi r)^2 + \xi^2(1-r)^2)h^2}. \end{aligned} \tag{4.85}$$

As a series in h, we have

$$\begin{cases} R = -1 - \dfrac{2\mathrm{i}kr(1-r)(2-r)h}{2-2r+r^2} + \mathcal{O}(h^2), \\ T = \dfrac{2\mathrm{i}k(r-1)rh}{2-2r+2r^2} + \mathcal{O}(h^2), \\ \lambda_r = \dfrac{2\mathrm{i}k(2-r)}{(2-2r+r^2)h} - \dfrac{2k^2(2-r)^2(1-r)r}{(2-2r+r^2)^2} + \mathcal{O}(h), \\ \lambda_2 = \dfrac{2\mathrm{i}k(r-1)r}{(2-2r+r^2)h} - \dfrac{2k^2(r-1)r^2(2-3r+r^2)}{(2-2r+r^2)^2} + \mathcal{O}(h). \end{cases} \tag{4.86}$$

Remark 11 *If* $r = 0$*, then from* (4.84) *we have*

$$\begin{cases} R = -1, \\ \\ T = 0, \\ \\ \lambda_r = \dfrac{2\mathrm{i}\sin(kh)}{h^2}, \\ \\ \lambda_k = 0, \ \forall\ k \geq 2. \end{cases} \tag{4.87}$$

4.4.3 An Operator Splitting Scheme: OFDDM

As in the previous section, we look for a solution of the form

$$\phi^n_{h,\,l} = \begin{cases} \mathrm{e}^{-\mathrm{i}\rho n\Delta t}\left(\mathrm{e}^{-\mathrm{i}kh(l-r)} + R_{\mathrm{OFDDM}}\mathrm{e}^{\mathrm{i}kh(l-r)}\right), & \text{if} \quad l \leq 0, \\ T_{\mathrm{OFDDM}}\mathrm{e}^{-\mathrm{i}\rho n\Delta t}\mathrm{e}^{-\mathrm{i}kh(l-r)}, & \text{if} \quad l = 1, \\ 0, & \text{if} \quad l \geq 2. \end{cases} \tag{4.88}$$

We substitute the expressions for $\phi_{h,-1}$, $\phi^n_{h,0}$, and $\phi^n_{h,1}$ from (4.88) in (4.57, ii, iii). We also employ the first three terms of the series (4.68) for ρ in the dispersion relation (4.67). Using the software MATHEMATICA to solve the resulting equations for the reflection coefficient R_{OFDDM} and the transmission coefficient T_{OFDDM}, we write a series expansion in h for both these terms. We have,

$$\begin{cases} R_{\mathrm{OFDDM}} = -1 - \dfrac{2\mathrm{i}kr(27 - 42r + 14r^2)h}{26 - 27r + 14r^2} + \mathcal{O}(h^2), \\ \\ T_{\mathrm{OFDDM}} = \dfrac{2\mathrm{i}k(r-1)(15r-1)h}{26 - 27r + 14r^2} + \mathcal{O}(h^2), \end{cases} \tag{4.89}$$

Remark 12 *If* $r = 0$ *we have* $R_{\mathrm{OFDDM}} = -1 = R_{\mathrm{cont}}$*. However, as opposed to the case of the scheme* FDDM*,* $T \neq 0$ *in this case. Thus, the solution is nonzero at the node* $l = 1$*, when* $r = 0$*. We also have* $T = 0$ *when* $r = 1$*.*

4.4.4 A Fictitious Domain Method with a Boundary Multiplier: FDBM

For the last method, we consider a fictitious domain method which utilizes a boundary multiplier. This method was analyzed in [41, 57]. We present here the relevant results. In this method, the Dirichlet boundary condition is imposed at the point x_r as follows.

$$\begin{cases} \text{Find} \quad (\phi_{h,l}^{n+1}, \lambda_{h,r}^{n+1}) \text{ such that } \forall\, l \in \mathbb{Z}, n \in \mathbb{N} \cup \{0\} : \\ \\ \phi_{h,l}^{n+1} \in \mathbb{R}, \text{ and } \lambda_{h,r}^{n+1} \in \mathbb{R}, \\ \\ \text{(i) } \dfrac{\phi_{h,l}^{n+1} - 2\phi_{h,l}^{n} + \phi_{h,l}^{n-1}}{c^2\Delta t^2} - \dfrac{\phi_{h,l+1}^{n} - 2\phi_{h,l}^{n} + \phi_{h,l-1}^{n}}{h^2} \\ \quad + \lambda_{h,r}^{n+1} \left(\dfrac{(1-r)\delta_{l,0} + r\delta_{l,1}}{h} \right) = 0, \\ \\ \text{(ii) } (1-r)\phi_{h,0}^{n+1} + r\phi_{h,1}^{n+1} = 0, \\ \\ \text{(iii) } \phi_{h,l}^{0} = \phi_{0,h,l}, \text{ and } \dfrac{\phi_{h,l}^{1} - \phi_{h,l}^{-1}}{2\Delta t} = \phi_{1,h,l}. \end{cases} \tag{4.90}$$

The dispersion relation for this case is also given by (4.30). In the plane wave analysis of FDBM, we look for a solution of the form

$$\phi_{h,\,l}^{n} = \begin{cases} \mathrm{e}^{-\mathrm{i}\rho n\Delta t} \left(\mathrm{e}^{-\mathrm{i}kh(l-r)} + R_{\mathrm{FDBM}} \mathrm{e}^{\mathrm{i}kh(l-r)} \right), & \text{if} \quad l \leq 0, \\ T_{\mathrm{FDBM}} \mathrm{e}^{-\mathrm{i}\rho n\Delta t} \mathrm{e}^{-\mathrm{i}kh(l-r)}, & \text{if} \quad l \geq 1, \end{cases} \tag{4.91}$$

and

$$\lambda_{h,r}^{n+1} = \lambda_r \mathrm{e}^{-\mathrm{i}\rho n\Delta t}. \tag{4.92}$$

We can solve for the reflection coefficient R_{FDBM}, and the transmission coefficient T_{FDBM} in a similar manner as before. From [57], we have

$$\begin{cases} R_{\mathrm{FDBM}} = \dfrac{-\xi^{2r-1} \left(\xi(1-r) + r \right)^2}{\xi + 2r(1-r)(1-\xi)}, \\ T_{\mathrm{FDBM}} = \dfrac{r(1-r)(1-\xi)^2}{\xi + 2r(1-r)(1-\xi)}, \\ \lambda_r = \dfrac{\xi^{r-1}(1-\xi^2) \left((1-r)\xi + r \right)}{h \left(\xi + 2r(1-r)(1-\xi) \right)}. \end{cases} \tag{4.93}$$

As a series in h, we have

$$\begin{cases} R_{\mathrm{FDBM}} = -1 - \dfrac{2\mathrm{i}kr(1-r)h}{2-2r+r^2} + \mathcal{O}(h^2), \\ T_{\mathrm{FDBM}} = -2\mathrm{i}k(1-r)rh + \mathcal{O}(h^2), \\ \lambda_r = 2\mathrm{i}k - 4k^2(1-r)rh + \mathcal{O}(h^2). \end{cases} \tag{4.94}$$

Remark 13 *If $r = 0$, then from* (4.93) *we have*

$$\begin{cases} R = -1 = R_{\mathrm{cont}}, \\ T = 0, \\ \lambda_r = \dfrac{2\mathrm{i}\sin(kh)}{h}. \end{cases} \tag{4.95}$$

Thus, as in the case of the scheme FDDM*, we have $R = -1$ and $T = 0$.*

4.5 Comparison of Schemes

In this section we compare the four different schemes encountered in this chapter, namely FDM, FDDM, OFDDM, and FDBM, on the basis of their reflection coefficients. Figure 4.2 plots the error in the amplitude of the reflected wave,

$$|R| - |R_{\mathrm{cont}}| = |R| - 1, \tag{4.96}$$

against the number of nodes per wavelength L/h, where L denotes the wavelength. Here R is the reflection coefficient for any one of the four schemes. We plot this error for four different values of $r = 0.25, 0.5, 0.75$, and 1.0 (even though $r < 1$, we still consider this case). From Figure (4.2) we note that the error is the largest in the case of the FDBM scheme, whereas this error is zero ($\approx 10^{-16}$) for the other three schemes, for $r = 0.25, 0.5, 0.75$. For $r = 1.0$ the error is zero for all the four schemes.

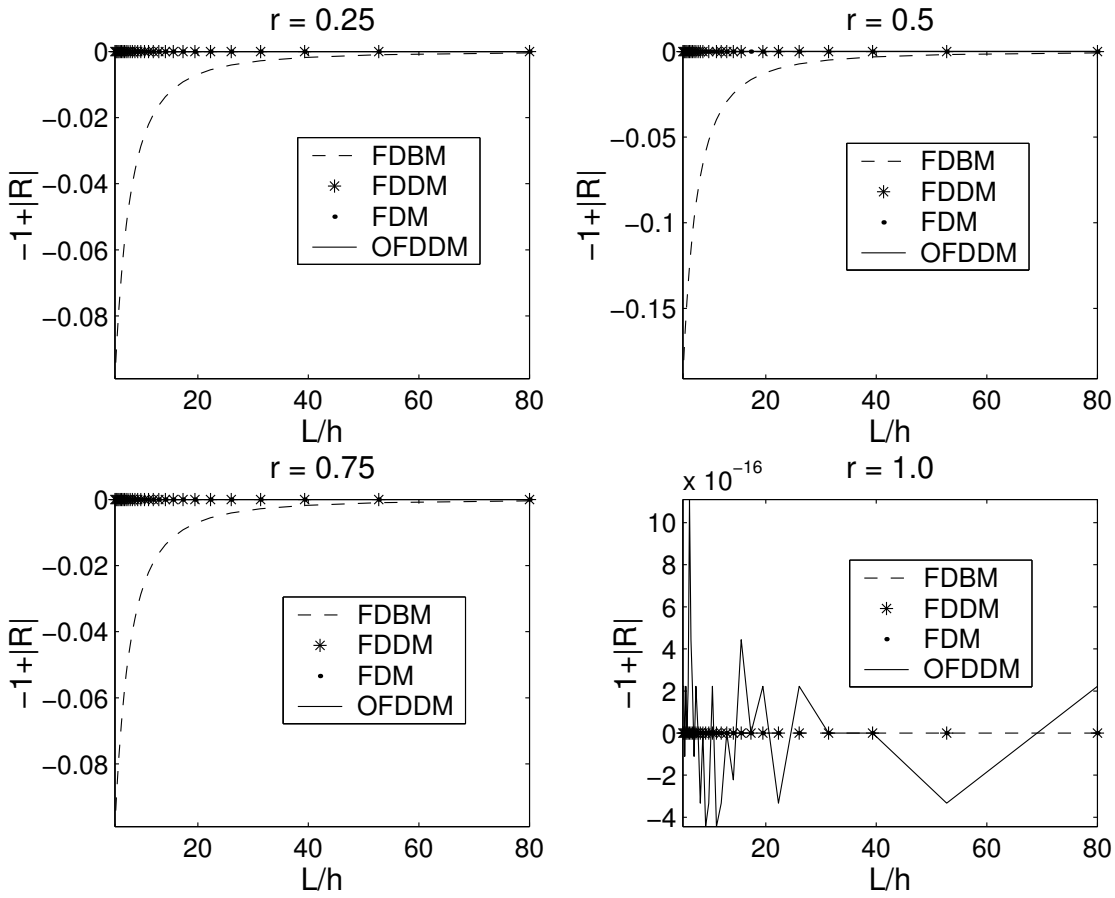

Figure 4.2: Error in the amplitude of the reflected wave versus the number of nodes per wavelength, for different values of r.

In Figure 4.3 we plot the error in the amplitude (4.96) against r, for $5, 10, 20$ and 40 nodes per wavelength. Again, as in Figure 4.2, the error is the largest in the case of the FDBM scheme. This implies that in the scheme FDBM, energy is propagated inside the domain ω, whereas this is not the case for the other three schemes.

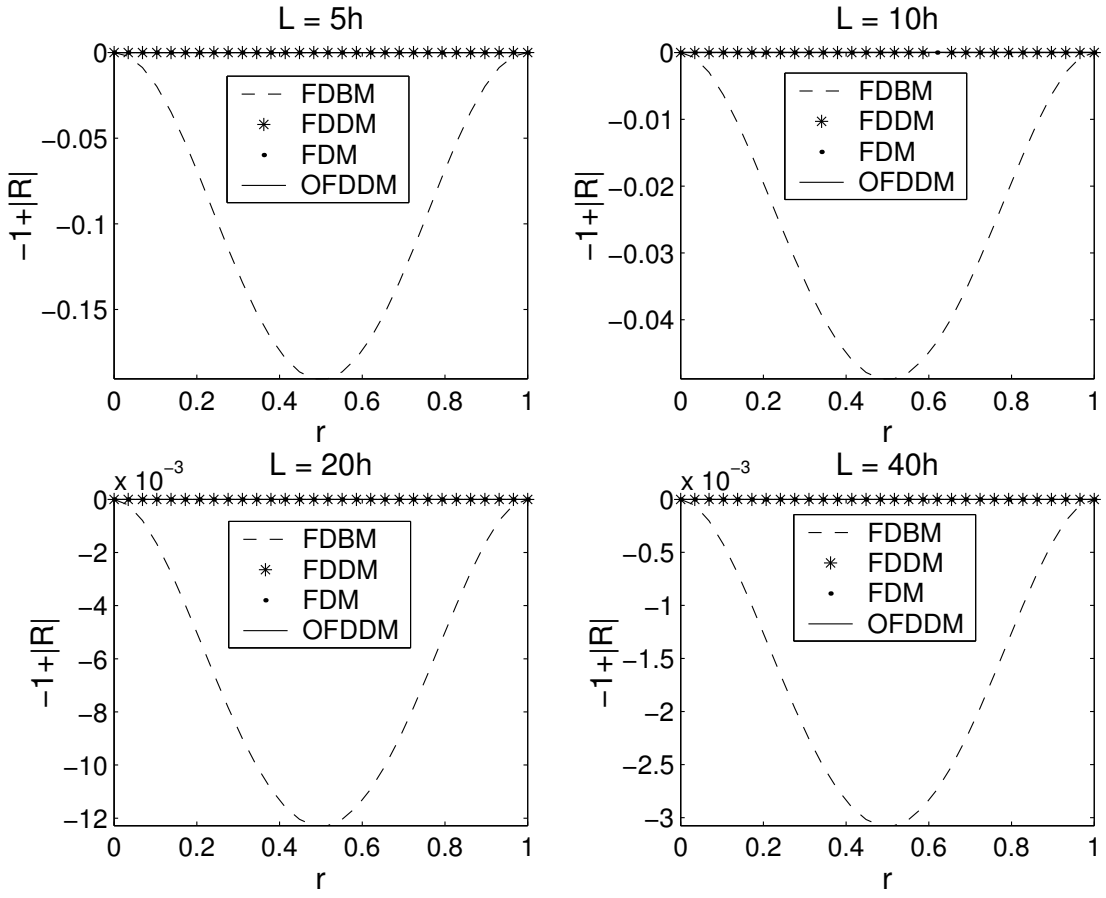

Figure 4.3: Error in the amplitude of the reflected wave versus r, for different number of nodes per wavelength.

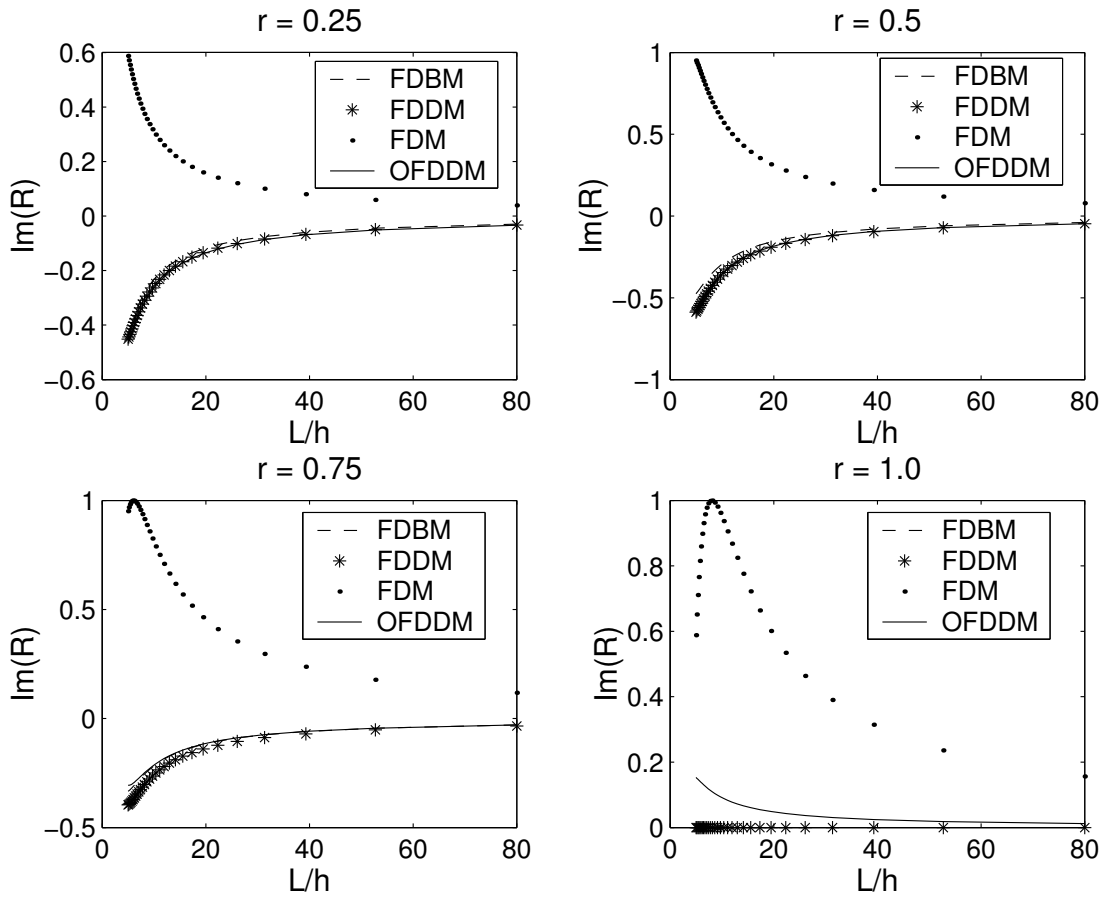

Figure 4.4: The phase error in the reflected wave versus the number of nodes per wavelength, for different values of r.

In Figure 4.4 we plot the phase error in the reflected wave $\mathrm{Im}(R)$, i.e., the imaginary part of the reflection coefficient R, against number of nodes per wavelength, for different values of r. In this case, we see that the phase error is the largest for the finite difference scheme FDM. The phase errors for the other three schemes are comparable.

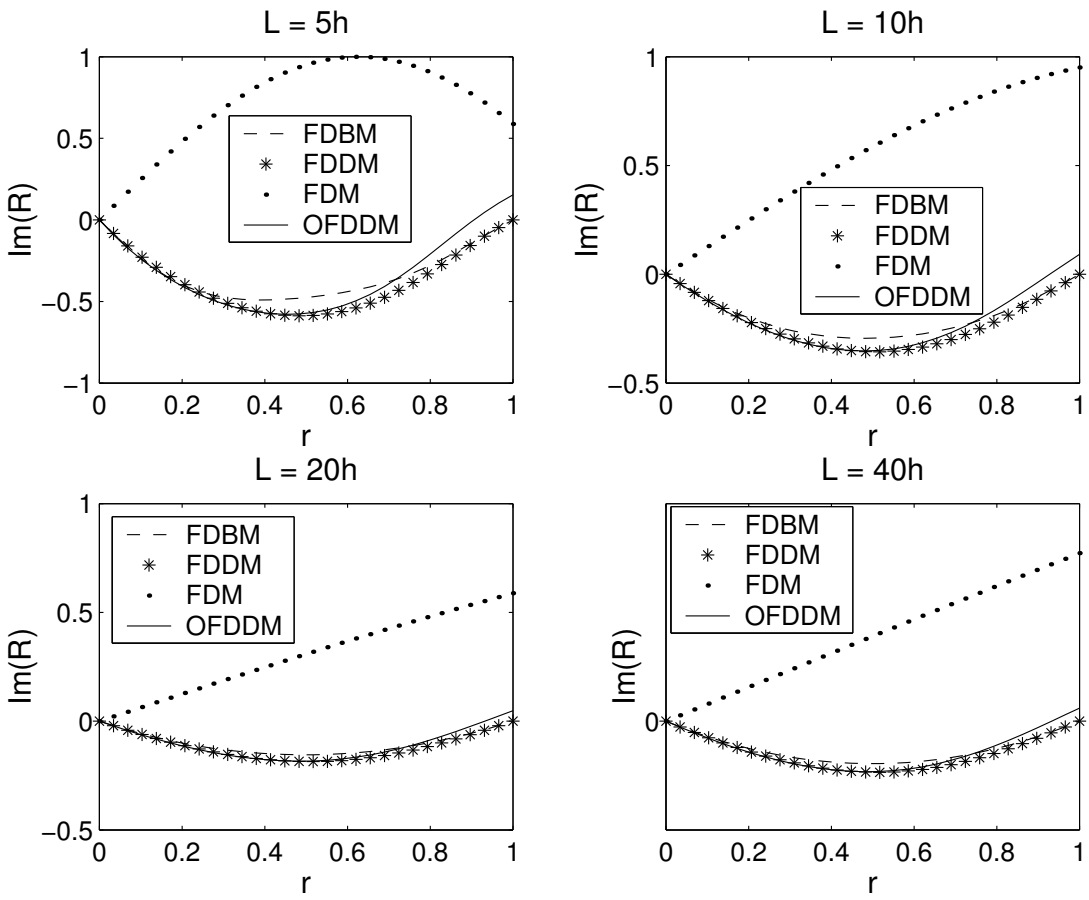

Figure 4.5: The phase error in the reflected wave versus r, for different number of nodes per wavelength.

In Figure 4.5 we plot the phase error in the reflected wave $\mathrm{Im}(R)$, i.e., the imaginary part of the reflection coefficient R, against r. This is done for $5, 10, 20$ and 40 nodes per wavelength. Again, we see that the phase error is the largest for the finite difference scheme FDM. The phase errors for the other three schemes are comparable; however, the operator splitting scheme OFDDM suffers from a phase shift, due to which $R_{\mathrm{OFDDM}} \neq 0$ at $r = 1$. As the number of nodes per wavelength is increased, R_{OFDDM} seems to converge to 0.

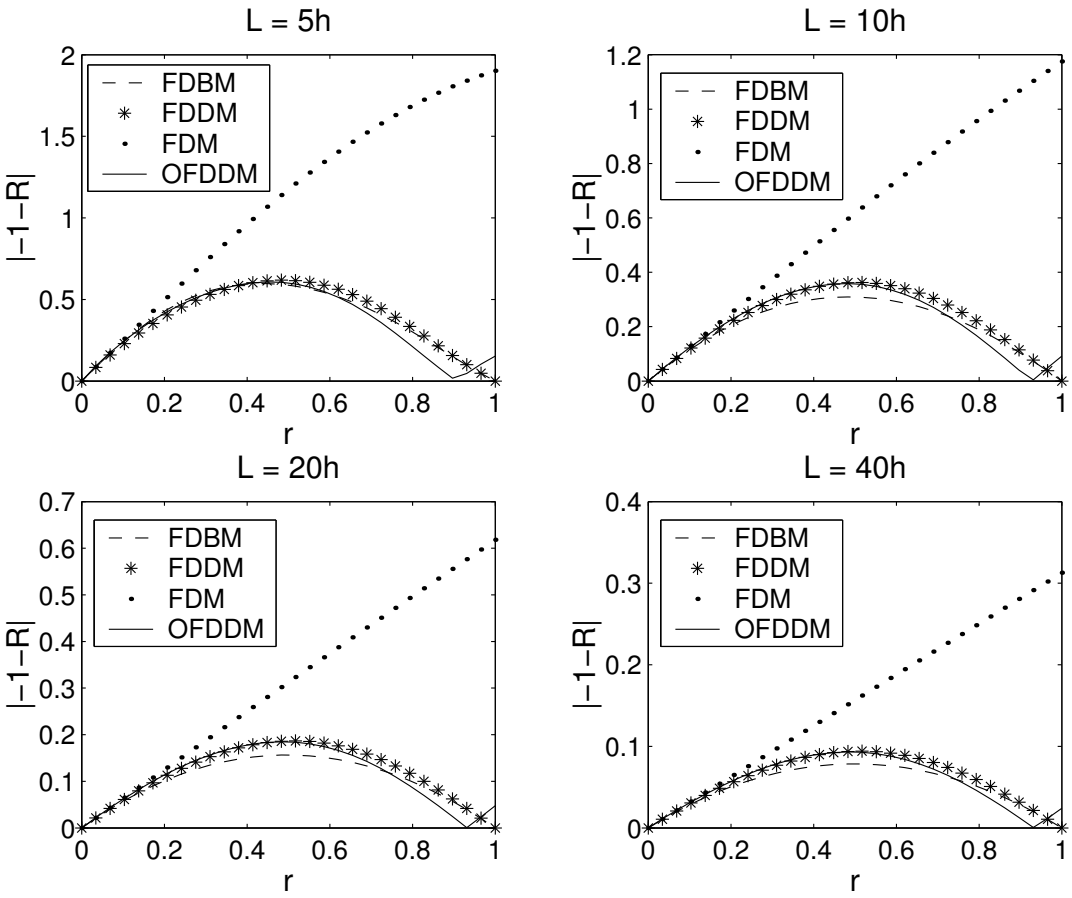

Figure 4.6: Total error in the reflection coefficient versus number of nodes per wavelength, for different values of r.

In Figure 4.6 we plot the total error in the reflection coefficient

$$\text{Total Error } = |R_{\text{cont}} - R| = | - 1 - R|, \tag{4.97}$$

against r for $5, 10, 20$ and 40 nodes per wavelength. The total error is the largest for the finite difference scheme FDM, but decreases as $r \to 0$. The error for all schemes becomes comparable as we increase the number of nodes per wavelength. Again, we can observe a phase shift in the operator splitting scheme OFDDM.

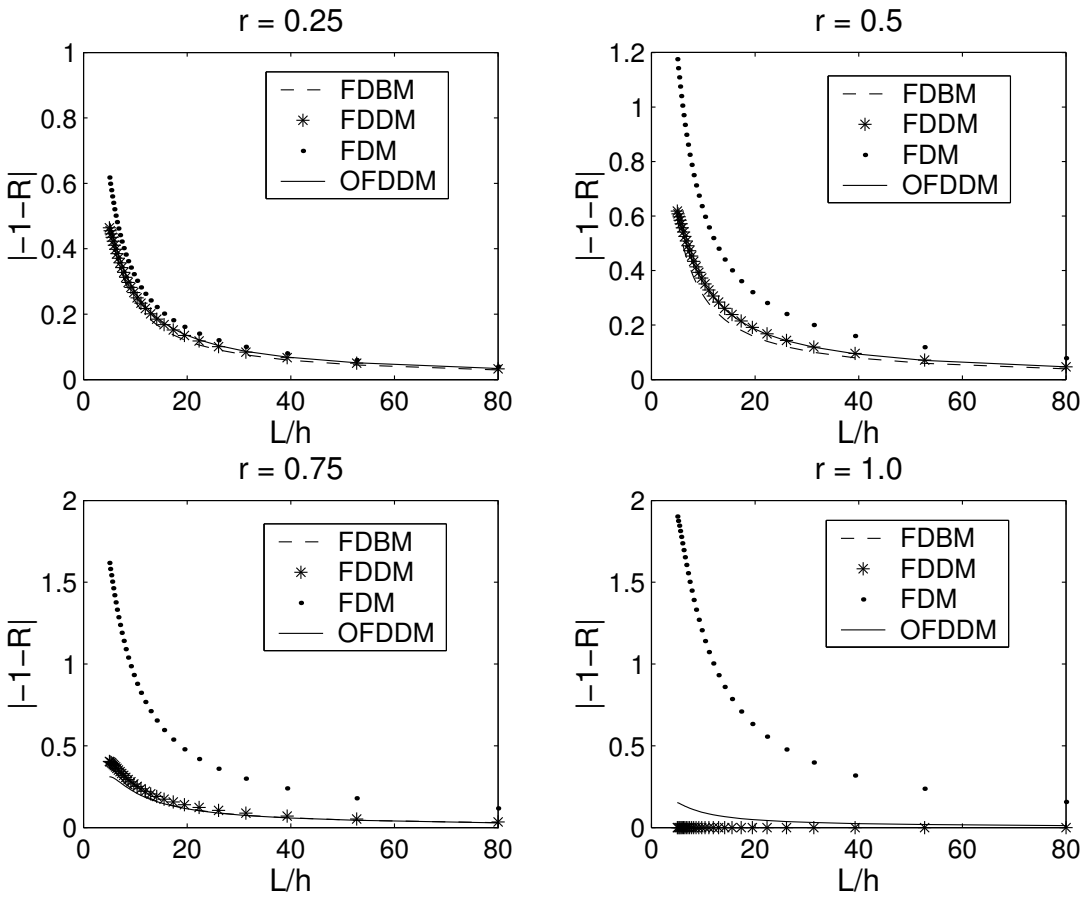

Figure 4.7: Total error in the reflection coefficient versus r, for different number of nodes per wavelength.

In Figure 4.7 we plot the total error in the reflection coefficient (4.97) against the number of nodes per wavelength, for different values of r. Again, the total error is largest for the finite difference scheme FDM, but decreases as $r \to 0$.

To summarize, all the methods presented here are first order with respect to h. However, all the fictitious domain methods improve the accuracy of the reflection coefficient. The fictitious domain method with a distributed multiplier does not propagate energy inside the obstacle as does the method with a boundary multiplier. The operator splitting produces a phase error. The dispersion relation for the operator splitting scheme has an additional error term, as we have shown in the case that $\eta = c\Delta t/h = 1$.

Chapter 5

A 2D Mixed Finite Element Formulation of the Uniaxial Perfectly Matched Layer

5.1 Introduction

The effective modeling of waves on unbounded domains by numerical methods, such as the finite difference method or the finite element method, is dependent on the particular absorbing boundary condition used to truncate the computational domain. In 1994, J. P. Berenger created the *perfectly matched layer (PML)* technique for the reflectionless absorption of electromagnetic waves in the time domain [15]. The PML is an absorbing layer that is placed around the computational domain of interest in order to attenuate outgoing radiation. Berenger showed that the PML allowed perfect transmission of electromagnetic waves across the interface of the computational domain, regardless of the frequency, polarization or angle of incidence of the waves, and the waves are attenuated exponentially with depth in the layer. The discretization of Maxwell's equations introduces errors which cause the PML to be less than perfectly matched. Also, the finite depth of the layer allows the transmitted part of the wave to return to the computational domain. Even so, it has been found that the PML medium can result in reflection errors as minute as -80 dB to -100 dB

[15, 16, 33, 58].

The PML has been constructed in a variety of ways, originally as a split field model obtained by performing a nonphysical splitting of Maxwell's equations [15], then via a complex change of variables [33, 114], and as an anisotropic uniaxial medium [58, 98, 119], among others. All these different approaches have been shown to lead to mathematically equivalent absorbing models [12, 98, 135]. Since its original inception in 1994, PML's have also extended their applicability in areas other than computational electromagnetics, such as acoustics, elasticity etc. [3, 5, 73, 79, 82].

In this chapter we propose a mixed finite element method (FEM), based on the anisotropic uniaxial formulation of the PML (UPML) by Sacks *et al.,* [119], to simulate wave propagation on unbounded domains. A mixed FEM has also been used in [38] which is based on the Zhao-Cangellaris's model for the PML [135]. The underlying partial differential equations in the Zhao-Cangellaris's PML model are second order in time, whereas the proposed uniaxial model has a system of first order PDE's.

The proposed scheme is a finite element counterpart of the 2D finite difference time domain method (FDTD) method, that is popular in computational electromagnetics [134]. An advantage of the FEM is that it can model arbitrary complex geometrical structures effectively. On rectangles, we use continuous piecewise bilinear finite elements to discretize the electric field and the Raviart-Thomas elements [116] to discretize the magnetic field. The degrees of freedom are staggered as in the FDTD scheme. We do not employ mass lumping techniques, i.e., use quadrature rules to obtain diagonal mass matrices, since it is not possible to simulate some anisotropic materials in this case. In general, mass lumping procedures are harder to construct in the case of higher order finite element methods, and this is especially true in the case of Maxwell's equations [36].

In Section 5.2, 5.3 we describe the UPML model and its implementation. In Section 5.4 we derive the 2D TM mode of the UPML model. Next, we describe a mixed finite element formulation for the UPML in Section 5.5. We state some energy decay results

that imply the well posedness of the PML model in Section 5.6. Section 5.7 describes the space and time discretization to be carried out. We perform a dispersion analysis and stability analysis in Section 5.8, and a reflection coefficient analysis in Section 5.9. Finally, we present numerical examples in Section 5.10 that demonstrate the effectiveness of the discrete PML model.

We will denote the angular frequency by ω in this chapter, instead of ρ, as we have been doing in earlier chapters. This is mainly for clarity as we will be dealing with equations in the frequency domain; ω being the standard notation in such cases. Since we are not considering the scattering problem in this chapter, there should be no confusion with the *obstacle* ω encountered in the previous chapters.

5.2 An Anisotropic Perfectly Matched Layer Absorbing Medium

We begin with a form of Maxwell's equations which is suitable for general media, which permit both electric and magnetic currents but do not contain unbalanced electric charges

$$\left\{\begin{array}{rcll} \dfrac{\partial \mathbf{B}}{\partial t} & = & -\nabla\times\mathbf{E} - \mathbf{J_M}, & \text{(Maxwell-Faraday's Law)}, \\ \dfrac{\partial \mathbf{D}}{\partial t} & = & \nabla\times\mathbf{H} - \mathbf{J_E}, & \text{(Maxwell-Ampere's Law)}, \\ \nabla\cdot\mathbf{B} & = & 0, & \text{(Gauss's Law for the magnetic field)}, \\ \nabla\cdot\mathbf{D} & = & 0, & \text{(Gauss's Law for the electric field)}. \end{array}\right. \tag{5.1}$$

Constitutive relations which relate the electric and magnetic fluxes $(\mathbf{D}, \mathbf{B})$ and the electric and magnetic currents $(\mathbf{J_E}, \mathbf{J_M})$ to the electric and magnetic fields $(\mathbf{E}, \mathbf{H})$ are added to these equations to make the system fully determined and to describe the response of a material to the electromagnetic fields. In free space, these constitutive relations are $\mathbf{D} = \epsilon_0\mathbf{E}$ and

$\mathbf{B} = \mu_0\mathbf{H}$, and $\mathbf{J_E} = \mathbf{J_M} = 0$, where ϵ_0 and μ_0 are the permittivity and the permeability of free space. In general, there are different possible forms for these constitutive relationships. In a frequency domain formulation of Maxwell's equations, these can be converted to linear relationships between the dependent and independent quantities with frequency dependent coefficient parameters.

We will derive a PML model in the frequency domain and then obtain a PML model in the time domain by taking the inverse Fourier transforms of the frequency domain equations. To this end, we consider the time-harmonic form of Maxwell's equations (5.1) (with time dependence $e^{\mathrm{i}\rho t}$ given by

$$\begin{cases} \mathrm{i}\omega\hat{\mathbf{B}} &= -\nabla\times\hat{\mathbf{E}} - \hat{\mathbf{J}}_M, \\ \mathrm{i}\omega\hat{\mathbf{D}} &= \nabla\times\hat{\mathbf{H}} - \hat{\mathbf{J}}_E, \\ \nabla\cdot\hat{\mathbf{B}} &= 0, \\ \nabla\cdot\hat{\mathbf{D}} &= 0, \end{cases} \tag{5.2}$$

where for every field vector $\mathbf{V}$, $\hat{\mathbf{V}}$ denotes its Fourier transform and we have the constitutive laws

$$\begin{cases} \hat{\mathbf{B}} &= [\mu]\hat{\mathbf{H}}, \\ \hat{\mathbf{D}} &= [\epsilon]\hat{\mathbf{E}}, \\ \hat{\mathbf{J}}_M &= [\sigma_M]\hat{\mathbf{H}}, \\ \hat{\mathbf{J}}_E &= [\sigma_E]\hat{\mathbf{E}}. \end{cases} \tag{5.3}$$

Here, the square brackets indicate a tensor quantity.

Note that when the density of electric and magnetic charge carriers in the medium is uniform throughout space, then $\nabla\cdot\hat{\mathbf{J}}_E = 0$ and $\nabla\cdot\hat{\mathbf{J}}_M = 0$.

We define new tensors

$$\begin{cases} [\bar{\mu}] &= [\mu] + \dfrac{[\sigma_M]}{\mathrm{i}\omega}, \\ [\bar{\epsilon}] &= [\epsilon] + \dfrac{[\sigma_E]}{\mathrm{i}\omega}. \end{cases} \tag{5.4}$$

Using the definitions (5.4) we define two new constitutive laws that are equivalent to (5.3), given by

$$\begin{cases} \hat{\mathbf{B}}_{\text{new}} &= [\bar{\mu}]\hat{\mathbf{H}}, \\ \hat{\mathbf{D}}_{\text{new}} &= [\bar{\epsilon}]\hat{\mathbf{E}}. \end{cases} \tag{5.5}$$

Using (5.5) in (5.2) Maxwell's equations in time-harmonic form become

$$\begin{cases} \mathrm{i}\omega\hat{\mathbf{B}}_{\text{new}} &= -\nabla\times\hat{\mathbf{E}}, \\ \mathrm{i}\omega\hat{\mathbf{D}}_{\text{new}} &= \nabla\times\hat{\mathbf{H}}, \\ \nabla\cdot\hat{\mathbf{B}}_{\text{new}} &= 0, \\ \nabla\cdot\hat{\mathbf{D}}_{\text{new}} &= 0. \end{cases} \tag{5.6}$$

The split-field PML introduced by Berenger [15] is a hypothetical medium based on a mathematical model. In [98] Mittra and Pekel showed that Berenger's PML was equivalent to Maxwell's equations with a diagonally anisotropic tensor appearing in the constitutive relations for $\mathbf{D}$ and $\mathbf{B}$. For a single interface, the anisotropic medium is *uniaxial* and is composed of both the electric and magnetic permittivity tensors. This uniaxial formulation performs as well as the original split-field PML while avoiding the nonphysical field splitting. As will be shown below, by properly defining a general constitutive tensor $[S]$, we can use the UPML in the interior working volume as well as the absorbing layer. This tensor provides a lossless isotropic medium in the primary computation zone, and individual UPML absorbers adjacent to the outer lattice boundary planes for mitigation of spurious wave reflections. The fields excited within the UPML are also plane wave in nature and satisfy Maxwell's curl equations.

The derivation of the PML properties for the tensor constitutive laws is also done directly by Sacks *et al.,* in [119] and by Gedney in [58]. We follow the derivation by Sacks *et al.,* here. We begin by considering planar electromagnetic waves in free space incident upon a PML half space. Starting with the impedance matching assumption, i.e., the

impedance of the layer must match that of free space: $\epsilon_0^{-1}\mu_0 = [\bar{\epsilon}]^{-1}[\bar{\mu}]$ we have

$$\frac{[\bar{\epsilon}]}{\epsilon_0} = \frac{[\bar{\mu}]}{\mu_0} = [S] = \text{diag}\{a_1, a_1, a_3\}. \tag{5.7}$$

Hence, the constitutive parameters inside the PML layer are

$$[\bar{\epsilon}] = \epsilon_0[S], \text{ and } [\bar{\mu}] = \mu_0[S], \tag{5.8}$$

where $[S]$ is a diagonal tensor.

By considering plane wave solutions of the form

$$\mathbf{V}(\mathbf{x}, t) = \hat{\mathbf{V}}(\mathbf{x})\, \mathrm{e}^{\mathrm{i}(\omega t - \mathbf{k}\cdot\mathbf{x})}, \tag{5.9}$$

for all field vectors $\mathbf{V}$, to the time-harmonic Maxwell's equations with the diagonally anisotropic tensor, where $\mathbf{k} = (k_x, k_y, k_z)$ is the wave vector of the planar electromagnetic wave and $\mathbf{x} = (x, y, z)$, the dispersion relation for waves in the PML are found to be

$$\frac{k_x^2}{a_2 a_3} + \frac{k_y^2}{a_1 a_3} + \frac{k_z^2}{a_1 a_2} = k_0^2 \equiv \omega^2 \mu_0 \epsilon_0 \equiv \frac{\omega^2}{c^2}. \tag{5.10}$$

where, c is the speed of light in free space.

Without loss of generality, we consider a PML layer which fills the positive x half-space and plane waves with wave vectors in the xy- plane ($k_z = 0$). Let θ_i be the angle of incidence of the plane wave measured from the normal to the surface $x = 0$. The standard phase and magnitude matching arguments at the interface yield a generalization of Snell's law

$$\sqrt{a_1 a_3}\sin\theta_t = \sin\theta_i, \tag{5.11}$$

where θ_t is the angle of the transmitted plane wave. By matching the magnitudes of the electric and magnetic fields at the interface, $x = 0$, we have the following values of the

reflection coefficients for the TE and the TM modes:

$$\begin{cases} R^{TE} &= \dfrac{\cos\theta_i - \sqrt{\dfrac{a_3}{a_2}}\cos\theta_t}{\cos\theta_i + \sqrt{\dfrac{a_3}{a_2}}\cos\theta_t}, \\ R^{TM} &= \dfrac{\sqrt{\dfrac{a_3}{a_2}}\cos\theta_t - \cos\theta_i}{\cos\theta_i + \sqrt{\dfrac{a_3}{a_2}}\cos\theta_t}. \end{cases} \tag{5.12}$$

From (5.12) we can see, that by choosing $a_3 = a_2 = a$ and $\sqrt{a_1 a_3} = 1$, the interface is completely reflectionless for any frequency and angle of incidence and polarization. Using (5.5) and (5.7), the constitutive laws for the perfectly matched layer are

$$\begin{cases} \hat{\mathbf{B}}_{\text{new}} &= \mu_0[S]\hat{\mathbf{H}}, \\ \hat{\mathbf{D}}_{\text{new}} &= \epsilon_0[S]\hat{\mathbf{E}}, \end{cases} \tag{5.13}$$

where the tensor $[S]$ is

$$[S] = \begin{bmatrix} a^{-1} & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}. \tag{5.14}$$

The perfectly matched layer is therefore characterized by the single complex number a. Taking it to be the constant $a = \gamma - \mathrm{i}\beta$, and substituting into the dispersion relation (5.10), we get the expression

$$\hat{\mathbf{E}}(x, y, z) = \hat{\mathbf{E}}_0\, \mathrm{e}^{-k_0\beta\cos\theta_t x}\, e^{-\mathrm{i}k_0(\gamma\cos\theta_t x + \sin\theta_t y)} e^{\mathrm{i}\omega t}, \tag{5.15}$$

for the electric field inside of the PML. Hence, we can see that γ determines the wavelength of the wave in the PML, and for $\beta > 0$, the wave is attenuated according to the distance of travel in the x direction.

5.3 Implementation of the Uniaxial PML

To apply the perfectly matched layer to electromagnetic computations, the half infinite layer is replaced with a layer of finite depth and backed with a more conventional boundary condition, such as a perfect electric conductor (PEC). This truncation of the layer will lead to reflections generated at the PEC surface, which can propagate back through the layer to re-enter the computational region. In this case, the reflection coefficient R, is now a function of the angle of incidence θ, the depth of the PML δ, as well as the parameter a in (5.14). Thus, this parameter a for the PML is chosen in order for the attenuation of waves in the PML to be sufficient so that the waves striking the PEC surface are negligible in magnitude. Perfectly matched layers are then placed near each edge (face in 3D) of the computational domain where a non-reflecting condition is desired. This leads to overlapping PML regions in the corners of the domain. As shown in [119], the correct form of the tensor which appears in the constitutive laws for these regions is the product

$$[S] = [S]_x[S]_y[S]_z, \tag{5.16}$$

where component $[S]_\alpha$ in the product in (5.16) is responsible for attenuation in the α direction, for $\alpha = x, y, z$, see Figure 5.1. All three of the component tensors in (5.16) are diagonal and have the forms

$$[S]_x = \begin{bmatrix} s_x^{-1} & 0 & 0 \\ 0 & s_x & 0 \\ 0 & 0 & s_x \end{bmatrix}; \quad [S]_y = \begin{bmatrix} s_y & 0 & 0 \\ 0 & s_y^{-1} & 0 \\ 0 & 0 & s_y \end{bmatrix}; \quad [S]_z = \begin{bmatrix} s_z & 0 & 0 \\ 0 & s_z & 0 \\ 0 & 0 & s_z^{-1} \end{bmatrix}. \tag{5.17}$$

In the above s_x, s_y, s_z are analogous to the complex valued parameter a encountered in Section 5.2, in the analysis of the single PML layer. Here, s_α governs the attenuation of the electromagnetic waves in the α direction for $\alpha = x, y, z$. When designing PML's for implementation, it is important to choose the parameters s_α so that the resulting frequency domain equations can be easily converted back into the time domain. The simplest of these

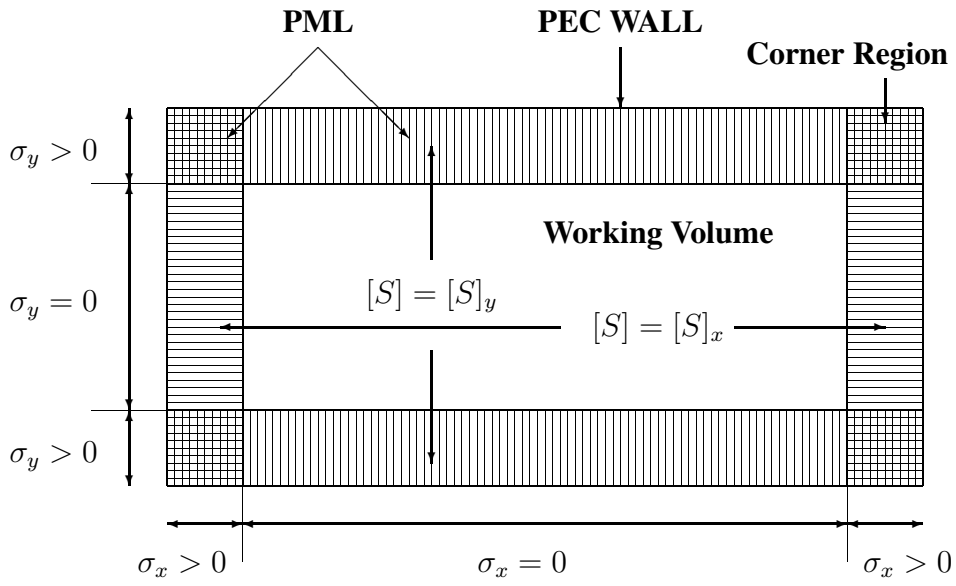

Figure 5.1: PML layers surrounding the domain of interest. In the corner regions of the PML, both σ_x and σ_y are positive and the tensor $[S]$ is the product $[S]_x[S]_y$. In the remaining regions only one of σ_x (left and right PML's) or σ_y (top and bottom PML's) are nonzero and positive. The tensor $[S]$, is thus either $[S]_x$ or $[S]_y$, respectively. The PML is truncated by a perfect electric conductor (PEC).

which we employ here [58] is

$$s_\alpha = 1 + \frac{\sigma_\alpha}{\mathrm{i}\omega\epsilon_0}, \quad \text{where} \quad \sigma_\alpha \geq 0 \quad \alpha = x, y, z. \tag{5.18}$$

The PML interface represents a discontinuity in the conductivities σ_α. To reduce the numerical reflections caused by these discontinuous conductivities, the σ_α are chosen to be functions of the variable α (for e.g., σ_x is taken to be a function of x in the $[S]_x$ component of the PML tensor). Choosing these functions so that $\sigma_\alpha = 0$, i.e., $s_\alpha = 1$ at the interface makes the PML a continuous extension of the medium being matched and reduces numerical reflections at the interface. Increasing the value of σ_α with depth in the layer, allows for greater overall attenuation while keeping down the numerical reflections. Gedney [58] suggests a conductivity profile

$$\sigma_\alpha(\alpha) = \frac{\sigma_{\max}|\alpha - \alpha_0|^m}{\delta^m}; \quad \alpha = x, y, z, \tag{5.19}$$

where δ is the depth of the layer, $\alpha = \alpha_0$ is the interface between the PML and the computational domain, and m is the order of the polynomial variation. Gedney remarks that values of m between 3 and 4 are believed to be optimal. For the conductivity profile (5.19), the PML parameters can be determined for given values of m, δ, and the desired reflection coefficient at normal incidence R_0, as

$$\sigma_{\max} \approx \frac{(m+1)\ln(1/R_0)}{2Z\delta}, \tag{5.20}$$

Z being the characteristic wave impedance of the PML. Empirical testing suggests that, for a broad range of problems, an optimal value of $\sigma_{\max}$ is given by

$$\sigma_{\text{opt}} \approx \frac{m+1}{150\pi h_\alpha \sqrt{\epsilon_r}}, \tag{5.21}$$

where h_α is the space increment in the α direction and ϵ_r is the relative permittivity of the material being modeled. In the case of free space $\epsilon_r = 1$.

5.4 The 2D TM Mode of the Uniaxial PML

From the time-harmonic Maxwell's curl equations in the UPML (5.6) and (5.13), Ampere's and Faraday's laws can be written in the most general form as

$$\begin{cases} \mathrm{i}\omega\mu_0[S]\hat{\mathbf{H}} &= -\nabla\times\hat{\mathbf{E}} \quad ; \quad (\text{Maxwell-Faraday's Law}), \\ \mathrm{i}\omega\epsilon_0[S]\hat{\mathbf{E}} &= \nabla\times\hat{\mathbf{H}} \quad ; \quad (\text{Maxwell-Ampere's Law}). \end{cases} \tag{5.22}$$

In (5.22), $[S]$ is the diagonal tensor defined via (5.16), (5.17)-(5.20). In the presence of this diagonal tensor, a plane wave is purely transmitted into the uniaxial medium. The tensor $[S]$ is no longer uniaxial by strict definition, but rather is anisotropic. However, the anisotropic PML is still referenced as uniaxial, since it is uniaxial in the non overlapping PML regions.

To obtain the 2D model of the UPML, we assume no variation in the z direction (i.e., $\dfrac{\partial}{\partial z} = 0$). In the 2D TM mode the electromagnetic field has three components, E_z, H_x, and H_y. In this case, we have $\sigma_z = 0$ and $s_z = 1$ in the UPML, and the time-harmonic Maxwell's equations (5.22), in the uniaxial medium can be written in scalar form as

$$\begin{cases} \mathrm{i}\omega\mu_0\dfrac{s_y}{s_x}\hat{H}_x &= -\dfrac{\partial\hat{E}_z}{\partial y}, \\ \mathrm{i}\omega\mu_0\dfrac{s_x}{s_y}\hat{H}_y &= -\dfrac{\partial\hat{E}_z}{\partial x}, \\ \mathrm{i}\omega\epsilon_0 s_x s_y\hat{E}_z &= \dfrac{\partial\hat{H}_y}{\partial x} - \dfrac{\partial\hat{H}_x}{\partial y}. \end{cases} \tag{5.23}$$

To avoid a computationally intensive implementation, we do not insert the expressions for s_x, s_y and s_z, obtained via (5.18), into (5.22), and transform to the time domain. Instead, we define suitable constitutive relationships that facilitate the decoupling of the frequency dependent terms [124]. To this end, we introduce the fields

$$\begin{cases} \hat{B}_x &= \mu_0 s_x^{-1}\hat{H}_x, \\ \hat{B}_y &= \mu_0 s_y^{-1}\hat{H}_y, \\ \hat{D}_z &= \mu_0 s_y\hat{E}_z. \end{cases} \tag{5.24}$$

Substituting the definitions (5.24) in (5.23), using the defining relations for s_x and s_y from (5.18), and then transforming into the time domain by using the inverse Fourier transform, yields an equivalent system of time-domain differential equations, which is the 2D TM mode of the uniaxial PML:

$$\left\{\begin{array}{rcl} \dfrac{\partial B_x}{\partial t} &=& -\dfrac{\sigma_y}{\epsilon_0}B_x - \dfrac{\partial E_z}{\partial y}, \\ \dfrac{\partial H_x}{\partial t} &=& \dfrac{1}{\mu_0}\dfrac{\partial B_x}{\partial t} + \dfrac{\sigma_x}{\epsilon_0\mu_0}B_x, \\ \dfrac{\partial B_y}{\partial t} &=& -\dfrac{\sigma_x}{\epsilon_0}B_y + \dfrac{\partial E_z}{\partial x}, \\ \dfrac{\partial H_y}{\partial t} &=& \dfrac{1}{\mu_0}\dfrac{\partial B_y}{\partial t} + \dfrac{\sigma_y}{\epsilon_0\mu_0}B_y, \\ \dfrac{\partial D_z}{\partial t} &=& -\dfrac{\sigma_x}{\epsilon_0}D_z + \dfrac{\partial H_y}{\partial x} - \dfrac{\partial H_x}{\partial y}, \\ \dfrac{\partial E_z}{\partial t} &=& -\dfrac{1}{\epsilon_0}\sigma_y E_z + \dfrac{1}{\epsilon_0}\dfrac{\partial D_z}{\partial t}. \end{array}\right. \tag{5.25}$$

Thus, the PML model consists in solving system (5.25) for the six variables, B_x, B_y, H_x, H_y, D_z, E_z.

5.5 A Mixed Finite Element Formulation for the UPML

Let $\mathbb{D}$ be an open bounded domain of $\mathbb{R}^2$. We surround $\mathbb{D}$ on all sides by PML layers to obtain the domain Ω. Let $\mathbf{H} = (H_x, H_y)$, $\mathbf{B} = (B_x, B_y)$, $E = E_z$ and $D = D$. We rewrite system (5.25) as

$$\left\{\begin{array}{rcl} \dfrac{\partial \mathbf{B}}{\partial t} &=& -\dfrac{1}{\epsilon_0}\Sigma_2\mathbf{B} - \overrightarrow{\operatorname{curl}}\, E, \\ \dfrac{\partial \mathbf{H}}{\partial t} &=& \dfrac{1}{\mu_0}\dfrac{\partial \mathbf{B}}{\partial t} + \dfrac{1}{\epsilon_0\mu_0}\Sigma_1\mathbf{B}, \\ \dfrac{\partial D}{\partial t} &=& -\dfrac{1}{\epsilon_0}\sigma_x D + \operatorname{curl}\mathbf{H}, \\ \dfrac{\partial E}{\partial t} &=& -\dfrac{1}{\epsilon_0}\sigma_y E + \dfrac{1}{\epsilon_0}\dfrac{\partial D}{\partial t}. \end{array}\right. \tag{5.26}$$

Here

$$\Sigma_1 = \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{pmatrix}; \quad \Sigma_2 = \begin{pmatrix} \sigma_y & 0 \\ 0 & \sigma_x \end{pmatrix}. \tag{5.27}$$

In the above, the operator denoted by $\overrightarrow{\mathrm{curl}}$, is a linear differential operator, which is defined as

$$\overrightarrow{\mathrm{curl}}\, v = (\frac{\partial v}{\partial y}, -\frac{\partial v}{\partial x}) \quad \forall\, v \in \mathcal{D}'(\Omega). \tag{5.28}$$

Similarly, the linear differential operator denoted by curl is defined as

$$\mathrm{curl}\, \mathbf{v} = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \quad \forall\, \mathbf{v} = (v_x, v_y) \in \mathcal{D}'(\Omega)^2. \tag{5.29}$$

Here, $\mathcal{D}'(\Omega)$ is the space of distributions on Ω. The operator curl appears as the (formal) transpose of the operator $\overrightarrow{\mathrm{curl}}$ [45], i.e.,

$$\langle \mathrm{curl}\, \mathbf{v}, \phi \rangle = \langle \mathbf{v}, \overrightarrow{\mathrm{curl}}\, \phi \rangle, \forall \mathbf{v} \in \mathcal{D}'(\Omega)^2, \phi \in \mathcal{D}'(\Omega). \tag{5.30}$$

We will solve system (5.26) in Ω, along with PEC conditions on $\partial\Omega$ to terminate the PML, namely,

$$\mathbf{n} \times \mathbf{E} = 0 \quad \text{on} \quad \partial\Omega,$$

where $\mathbf{n}$ is the outward unit normal to $\partial\Omega$. In the case of the 2D TM mode, the PEC condition translates to

$$E = E_z = 0, \quad \text{on} \quad \partial\Omega. \tag{5.31}$$

We also have the initial conditions

$$E(x,0) = E_0, \quad D(x,0) = E_0, \quad \mathbf{H}(x,0) = \mathbf{H}_0, \quad \mathbf{B}(x,0) = \mathbf{H}_0, \quad \text{for} \quad x \in \Omega. \tag{5.32}$$

We consider the following variational formulation of system (5.26) which is suitable for discretization by finite elements.

Find $(E(\cdot,t), D(\cdot,t), \mathbf{H}(\cdot,t), \mathbf{B}(\cdot,t)) \in H_0^1(\Omega) \times H_0^1(\Omega) \times [L^2(\Omega)]^2 \times [L^2(\Omega)]^2$ such that for all $\mathbf{\Psi} \in [L^2(\Omega)]^2$, for all $\phi \in H_0^1(\Omega)$,

$$\left\{ \begin{array}{rcl} \dfrac{d}{dt}\displaystyle\int_\Omega \mathbf{B}\cdot\mathbf{\Psi}\, d\mathbf{x} &=& -\dfrac{1}{\epsilon_0}\displaystyle\int_\Omega \Sigma_2\mathbf{B}\cdot\mathbf{\Psi}\, d\mathbf{x} - \displaystyle\int_\Omega \overrightarrow{\operatorname{curl}}\, E\cdot\mathbf{\Psi}\, d\mathbf{x}, \\ \dfrac{d}{dt}\displaystyle\int_\Omega \mathbf{H}\cdot\mathbf{\Psi}\, d\mathbf{x} &=& \dfrac{1}{\mu_0}\dfrac{d}{dt}\displaystyle\int_\Omega \mathbf{B}\cdot\mathbf{\Psi}\, d\mathbf{x} + \dfrac{1}{\epsilon_0\mu_0}\displaystyle\int_\Omega \Sigma_1\mathbf{B}\cdot\mathbf{\Psi}\, d\mathbf{x}, \\ \dfrac{d}{dt}\displaystyle\int_\Omega D\cdot\phi\, d\mathbf{x} &=& -\dfrac{1}{\epsilon_0}\displaystyle\int_\Omega \sigma_x D\cdot\phi\, d\mathbf{x} + \displaystyle\int_\Omega \overrightarrow{\operatorname{curl}}\,\phi\cdot\mathbf{H}\, d\mathbf{x}, \\ \dfrac{d}{dt}\displaystyle\int_\Omega E\cdot\phi\, d\mathbf{x} &=& -\dfrac{1}{\epsilon_0}\displaystyle\int_\Omega \sigma_y E\cdot\phi\, d\mathbf{x} + \dfrac{1}{\epsilon_0}\dfrac{d}{dt}\displaystyle\int_\Omega D\cdot\phi\, d\mathbf{x}. \end{array} \right. \tag{5.33}$$

We assume that the fields $(E, D, \mathbf{H}, \mathbf{B})$ are sufficiently differentiable in time. We note that, for $E \in L^2(\Omega)$, $\overrightarrow{\operatorname{curl}}\, E = (\dfrac{\partial E}{\partial y}, -\dfrac{\partial E}{\partial x}) \in [L^2(\Omega)]^2$, implies that both the partial derivatives of E must be in $L^2(\Omega)$. Hence we must have $E \in H^1(\Omega)$.

5.6 Energy Estimates for the UPML

Maxwell's equations form a symmetric hyperbolic system. The solution to such a system is strongly well-posed [84]. It is natural then to consider the well-posedness of the different PML models. One of the first papers to touch this subject was by Abarbanel and Gottlieb [2], who showed that Berenger's split-field PML was a weakly well-posed system; thus instabilities could appear in numerical implementations of this model. In [113], the authors demonstrate that the Zhao- Cangellaris's model for the PML is strongly well-posed. Bécache and Joly [12] show this well-posedness explicitly by presenting energy decay results for the 2D TE mode of this model.

We derive energy decay results for the 2D TM mode of the UPML in two cases, these being, σ a positive constant and $\sigma \in L^\infty(\Omega)$. We have derived estimates for the UPML model under the same conditions as done in [12] for the Zhao-Cangellaris model. The Zhao-Cangellaris's model established the equivalence between the Chew-Weedon's PML model [33] based on coordinate stretching, and the anisotropic model by Sacks *et al.,* As

a consequence, the energy decay results for the Zhao-Cangellaris's PML and the UPML appear to be similar. This is to be understood in the sense, that the definitions of the energies involved are identical in the second, third and fourth estimate and almost identical in the first estimate. We present the energy decay results here for the sake of completeness as well as for comparison.

To simplify the analysis, we assume that $\epsilon_0 = \mu_0 = 1$ in the rest of this Section. We also assume that we start with zero initial conditions in (5.32). Let $(\cdot, \cdot)$ denote the $L^2(\Omega)$ inner product.

Energy Estimate 1 *Let us assume that we have a PML in the region $x > 0$. In this case, $\sigma_x = \sigma$ and $\sigma_y = 0$. For a positive constant value of σ, the energy $\mathcal{E}_1^x$ of the PML system defined as*

$$\mathcal{E}_1^x = \frac{1}{2}\left(\left\| \frac{\partial H_x}{\partial t} \right\|^2_{L^2(\Omega)} + \left\| \frac{\partial H_y}{\partial t} \right\|^2_{L^2(\Omega)} + \|\sigma B_y\|^2_{L^2(\Omega)} + \left\| \left(\frac{\partial}{\partial t} + \sigma \right) E \right\|^2_{L^2(\Omega)} \right), \quad (5.34)$$

is a decreasing function of time. It satisfies the identity

$$\frac{d}{dt}\mathcal{E}_1^x = -2\sigma \left\| \frac{\partial H_y}{\partial t} \right\|^2_{L^2(\Omega)} \leq 0. \quad (5.35)$$

Proof 3 : *From (5.25), the UPML formulation for just one absorbing layer parallel to the y axis is given by*

$$\begin{cases} \text{(i)} & \dfrac{\partial B_x}{\partial t} & = & -\dfrac{\partial E}{\partial y}, \\ \text{(ii)} & \dfrac{\partial H_x}{\partial t} & = & \left(\dfrac{\partial}{\partial t} + \sigma \right) B_x, \\ \text{(iii)} & \left(\dfrac{\partial}{\partial t} + \sigma \right) B_y & = & \dfrac{\partial E}{\partial x}, \\ \text{(iv)} & \dfrac{\partial H_y}{\partial t} & = & \dfrac{\partial B_y}{\partial t}, \\ \text{(v)} & \left(\dfrac{\partial}{\partial t} + \sigma \right) D & = & \dfrac{\partial H_y}{\partial x} - \dfrac{\partial H_x}{\partial y}, \\ \text{(vi)} & \dfrac{\partial E}{\partial t} & = & \dfrac{\partial D}{\partial t}. \end{cases} \quad (5.36)$$

Applying the operator $\left(\frac{\partial}{\partial t}+\sigma\right)$ *to (5.36,* i*), the operator* $\frac{\partial}{\partial t}$ *to (5.36,* ii*) and combining the two results, we get*

$$\frac{\partial^2 H_x}{\partial t^2} = -\frac{\partial}{\partial y}\left(\frac{\partial}{\partial t}+\sigma\right)E. \tag{5.37}$$

We note that, since σ *is a constant, the operator* $\left(\frac{\partial}{\partial t}+\sigma\right)$ *commutes with the operators,* $\frac{\partial}{\partial y}$ *and* $\frac{\partial}{\partial x}$*. Taking the inner product of both sides of (5.37) with* $\frac{\partial H_x}{\partial t}$ *we have*

$$\begin{aligned}\left(\frac{\partial^2 H_x}{\partial t^2}, \frac{\partial H_x}{\partial t}\right) &= -\left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial t}+\sigma\right)E, \frac{\partial H_x}{\partial t}\right), \\ \implies \frac{1}{2}\frac{d}{dt}\left\|\frac{\partial H_x}{\partial y}\right\|^2_{L^2(\Omega)} &= -\left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial t}+\sigma\right)E, \frac{\partial H_x}{\partial t}\right).\end{aligned} \tag{5.38}$$

Next, apply the operator $\left(\frac{\partial}{\partial t}+\sigma\right)$ *to (5.36,* iii*) to get*

$$\left(\frac{\partial}{\partial t}+\sigma\right)^2 B_y = \left(\frac{\partial}{\partial t}+\sigma\right)\frac{\partial}{\partial x}E. \tag{5.39}$$

Taking inner products on both sides of (5.39) with $\frac{\partial B_y}{\partial t}$ *we get*

$$\begin{aligned}\left(\left(\frac{\partial}{\partial t}+\sigma\right)^2 B_y, \frac{\partial B_y}{\partial t}\right) &= \left(\left(\frac{\partial}{\partial t}+\sigma\right)\frac{\partial E}{\partial x}, \frac{\partial B_y}{\partial t}\right), \\ \implies \left(\frac{\partial^2 B_y}{\partial t^2}, \frac{\partial B_y}{\partial t}\right) + 2\sigma\left(\frac{\partial B_y}{\partial t}, \frac{\partial B_y}{\partial t}\right) & \\ +\sigma^2\left(B_y, \frac{\partial B_y}{\partial t}\right) &= \left(\left(\frac{\partial}{\partial t}+\sigma\right)\frac{\partial E}{\partial x}, \frac{\partial B_y}{\partial t}\right).\end{aligned} \tag{5.40}$$

From (5.36, iv*), using* $\frac{\partial B_y}{\partial t} = \frac{\partial H_y}{\partial t}$ *in (5.40), we have*

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\left\|\frac{\partial H_y}{\partial t}\right\|^2_{L^2(\Omega)} + 2\sigma\left(\frac{\partial H_y}{\partial t}, \frac{\partial H_y}{\partial t}\right) + \frac{1}{2}\frac{d}{dt}\left\|\sigma B_y\right\|^2_{L^2(\Omega)} & \\ = \left(\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}+\sigma\right)E, \frac{\partial H_y}{\partial t}\right).\end{aligned} \tag{5.41}$$

Finally, to eliminate D*, we apply the operator* $\frac{\partial}{\partial t}$ *to (5.36,* v*) and the operator* $\left(\frac{\partial}{\partial t}+\sigma\right)$ *to (5.36,* vi*), combining the results to get*

$$\left(\frac{\partial}{\partial t}+\sigma\right)\frac{\partial E}{\partial t} = \left(\frac{\partial}{\partial t}\frac{\partial H_y}{\partial x} - \frac{\partial}{\partial t}\frac{\partial H_x}{\partial y}\right). \tag{5.42}$$

Taking the inner products of both sides of (5.42) with $\left(\frac{\partial}{\partial t}+\sigma\right)E$*, we have*

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\left\|\left(\frac{\partial}{\partial t}+\sigma\right)E\right\|^2_{L^2(\Omega)} &= \left(\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}+\sigma\right)E,\left(\frac{\partial}{\partial t}+\sigma\right)E\right)\\ &= \left(\frac{\partial}{\partial t}\frac{\partial H_y}{\partial x},\left(\frac{\partial}{\partial t}+\sigma\right)E\right)-\left(\frac{\partial}{\partial t}\frac{\partial H_x}{\partial y},\left(\frac{\partial}{\partial t}+\sigma\right)E\right).\end{aligned} \tag{5.43}$$

Integrating by parts in the right hand side of (5.43), we have

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\left\|\left(\frac{\partial}{\partial t}+\sigma\right)E\right\|^2_{L^2(\Omega)} &= -\left(\frac{\partial H_y}{\partial t},\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}+\sigma\right)E\right)\\ &\quad+\left(\frac{\partial H_x}{\partial t},\frac{\partial}{\partial y}\left(\frac{\partial}{\partial t}+\sigma\right)E\right).\end{aligned} \tag{5.44}$$

Adding (5.38), (5.41) and (5.44) we have

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\left(\left\|\frac{\partial H_x}{\partial t}\right\|^2_{L^2(\Omega)}+\left\|\frac{\partial H_y}{\partial t}\right\|^2_{L^2(\Omega)}+\|\sigma B_y\|^2_{L^2(\Omega)}+\left\|\left(\frac{\partial}{\partial t}+\sigma\right)E\right\|^2_{L^2(\Omega)}\right)\\ +2\sigma\left\|\frac{\partial H_y}{\partial t}\right\|^2_{L^2(\Omega)}=0.\end{aligned} \tag{5.45}$$

Using definition (5.34) of $\mathcal{E}_1^x$ *in (5.45) and since* $\sigma>0$*, we finally obtain*

$$\frac{d}{dt}\mathcal{E}_1^x(t)=-2\sigma\left\|\frac{\partial H_y}{\partial t}\right\|^2_{L^2(\Omega)}\le 0,$$

which proves (5.35). This implies that $\mathcal{E}_1^x$ *is a decreasing function of time.*

Energy Estimate 2 *Consider a corner domain of the PML where both* σ_x *and* σ_y *are positive and constants. Let* $\mathbf{H}=(H_x,H_y)$*. The solution of the UPML TM mode satisfies the energy inequality*

$$\mathcal{E}_2^C(t)\le\mathcal{E}_2^C(s);\quad \textit{for all}\quad t\ge s, \tag{5.46}$$

where $\mathcal{E}_2^C$ *is the second order energy defined as*

$$\mathcal{E}_2^C(t)=\frac{1}{2}\left(\left\|\frac{\partial^2\mathbf{H}}{\partial t^2}\right\|^2_{L^2(\Omega)}+\left\|\Sigma_2\frac{\partial\mathbf{H}}{\partial t}\right\|^2_{L^2(\Omega)}+\left\|\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right\|^2_{L^2(\Omega)}\right). \tag{5.47}$$

Proof 4 : *The UPML model for this case is*

$$\begin{cases} \text{(i)} & \dfrac{\partial \mathbf{B}}{\partial t} + \Sigma_2 \mathbf{B} &=& -\overrightarrow{\operatorname{curl}}\, E, \\ \text{(ii)} & \dfrac{\partial \mathbf{B}}{\partial t} + \Sigma_1 \mathbf{B} &=& \dfrac{\partial \mathbf{H}}{\partial t}, \\ \text{(iii)} & \left(\dfrac{\partial}{\partial t} + \sigma_x\right) D &=& \operatorname{curl} \mathbf{H}, \\ \text{(iv)} & \left(\dfrac{\partial}{\partial t} + \sigma_y\right) E &=& \dfrac{\partial D}{\partial t}. \end{cases} \tag{5.48}$$

We eliminate $\mathbf{B}$ *from (5.48,* i*) and (5.48,* ii*) by applying the operator* $(\dfrac{\partial}{\partial t} + \Sigma_2)$ *to (5.48,* ii*) and the operator* $(\dfrac{\partial}{\partial t} + \Sigma_1)$ *to (5.48,* i*), and combining the results to get*

$$\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t} + \Sigma_2\right)\mathbf{H} = -\left(\frac{\partial}{\partial t} + \Sigma_1\right)\overrightarrow{\operatorname{curl}}\, E. \tag{5.49}$$

Applying the operator $(\dfrac{\partial}{\partial t} + \Sigma_2)$ *to both sides in the equation above, we get*

$$\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t} + \Sigma_2\right)^2 \mathbf{H} = -\left(\frac{\partial}{\partial t} + \Sigma_2\right)\left(\frac{\partial}{\partial t} + \Sigma_1\right)\overrightarrow{\operatorname{curl}}\, E. \tag{5.50}$$

Taking the inner product of both sides with $\dfrac{\partial^2 \mathbf{H}}{\partial t^2}$ *we have*

$$\left(\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t} + \Sigma_2\right)^2 \mathbf{H}, \frac{\partial^2 \mathbf{H}}{\partial t^2}\right) = -\left(\left(\frac{\partial}{\partial t} + \Sigma_2\right)\left(\frac{\partial}{\partial t} + \Sigma_1\right)\overrightarrow{\operatorname{curl}}\, E, \frac{\partial^2 \mathbf{H}}{\partial t^2}\right). \tag{5.51}$$

$$\begin{aligned} \Longrightarrow\ \frac{1}{2}\frac{d}{dt}\left(\left\|\frac{\partial^2 \mathbf{H}}{\partial t^2}\right\|^2_{L^2(\Omega)} + \left\|\Sigma_2 \frac{\partial \mathbf{H}}{\partial t}\right\|^2_{L^2(\Omega)}\right) + 2\Sigma_2 \left\|\frac{\partial^2 \mathbf{H}}{\partial t^2}\right\|^2_{L^2(\Omega)} \\ = -\left(\left(\frac{\partial}{\partial t} + \Sigma_2\right)\left(\frac{\partial}{\partial t} + \Sigma_1\right)\overrightarrow{\operatorname{curl}}\, E, \frac{\partial^2 \mathbf{H}}{\partial t^2}\right) \end{aligned} \tag{5.52}$$

Next, to eliminate D *we apply the operator* $\dfrac{\partial}{\partial t}$ *to (5.48,* iii*), and the operator* $(\dfrac{\partial}{\partial t} + \sigma_x)$ *to (5.48,* iv*), and combine the results to get*

$$\left(\frac{\partial}{\partial t} + \sigma_x\right)\left(\frac{\partial}{\partial t} + \sigma_y\right) E = \frac{\partial}{\partial t} \operatorname{curl} \mathbf{H}. \tag{5.53}$$

Applying $\dfrac{\partial}{\partial t}$ *to both sides of the equation above, and taking inner products with* $\left(\dfrac{\partial}{\partial t}+\sigma_x\right)\left(\dfrac{\partial}{\partial t}+\sigma_y\right)E$*, we get*

$$\left(\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E,\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right) = \left(\frac{\partial^2}{\partial t^2}\operatorname{curl}\mathbf{H},\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right). \quad (5.54)$$

$$\Longrightarrow \frac{1}{2}\frac{d}{dt}\left\|\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right\|^2_{L^2(\Omega)} = \left(\frac{\partial^2}{\partial t^2}\operatorname{curl}\mathbf{H},\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right). \quad (5.55)$$

Integrating the right hand side by parts and making use of the fact that the operators $\left(\dfrac{\partial}{\partial t}+\sigma_x\right)$ *and* $\left(\dfrac{\partial}{\partial t}+\sigma_y\right)$ *commute, we have*

$$\Longrightarrow \frac{1}{2}\frac{d}{dt}\left\|\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right\|^2_{L^2(\Omega)} = \left(\frac{\partial^2}{\partial t^2}\mathbf{H},\left(\frac{\partial}{\partial t}+\Sigma_2\right)\left(\frac{\partial}{\partial t}+\Sigma_1\right)\overrightarrow{\operatorname{curl}}\,E\right). \quad (5.56)$$

Adding (5.52) *and* (5.56) *we have*

$$\frac{1}{2}\frac{d}{dt}\left(\left\|\frac{\partial^2\mathbf{H}}{\partial t^2}\right\|^2_{L^2(\Omega)}+\left\|\Sigma_2\frac{\partial\mathbf{H}}{\partial t}\right\|^2_{L^2(\Omega)}+\left\|\left(\frac{\partial}{\partial t}+\sigma_x\right)\left(\frac{\partial}{\partial t}+\sigma_y\right)E\right\|^2_{L^2(\Omega)}\right) = -2\Sigma_2\left\|\frac{\partial^2\mathbf{H}}{\partial t^2}\right\|^2_{L^2(\Omega)}<0. \quad (5.57)$$

Using the definition of the energy $\mathcal{E}_2^C$ (5.47)*, and since* $\sigma_x>0$ *and* $\sigma_y>0$*, we have*

$$\frac{d}{dt}\mathcal{E}_2^C(t) = -2\Sigma_2\left\|\frac{\partial^2\mathbf{H}}{\partial t^2}\right\|^2_{L^2(\Omega)}<0, \quad (5.58)$$

which implies that $\mathcal{E}_2^C$ *is a decreasing function of time, and thus* (5.46) *is proved.*

Energy Estimate 3 *Again, assume a PML in the region* $x>0$*. For* $\sigma\in L^\infty(\Omega)$*, the zero order energy* $\mathcal{E}_0^x$ *of the UPML defined by*

$$\mathcal{E}_0^x(t) = \frac{1}{2}\left(\|H_x\|^2_{L^2(\Omega)}+\|H_y\|^2_{L^2(\Omega)}+\|E\|^2_{L^2(\Omega)}+\|H_x-B_x\|^2_{L^2(\Omega)}\right), \quad (5.59)$$

satisfies the energy estimate

$$\mathcal{E}_0^x(t) \leq \mathcal{E}_0^x(0) + 2\left\|\sigma\right\|_\infty \int_0^t \mathcal{E}_0^x(s)ds. \tag{5.60}$$

Proof 5 : *The UPML formulation for just one absorbing layer parallel to the y axis is given by* (5.36) *as seen in Energy estimate 1. Taking the inner product of both sides of (5.36,* i*) with H_x, both sides of (5.36,* iii*) with H_y, both sides of (5.36,* v*) with E, adding the three resulting equations and integrating the right hand side by parts (IBP) we get*

$$\begin{aligned}\left(\frac{\partial B_x}{\partial t}, H_x\right) + \left((\frac{\partial}{\partial t} + \sigma)B_y, H_y\right) + \left((\frac{\partial}{\partial t} + \sigma)D, E\right) = \\ -\left(\frac{\partial E}{\partial y}, H_x\right) + \left(\frac{\partial E}{\partial x}, H_y\right) + \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}, E\right) = 0 \quad (\textit{by IBP}).\end{aligned} \tag{5.61}$$

Assuming that we start from zero initial conditions, from (5.36, ii, iv, vi*), we have*

$$\begin{cases} \text{(i)} & H_x(\cdot,t) = B_x(\cdot,t) + \sigma\tilde{B}_x(\cdot,t), \\ \text{(ii)} & H_y(\cdot,t) = B_y(\cdot,t), \\ \text{(iii)} & E(\cdot,t) = D(\cdot,t), \end{cases} \tag{5.62}$$

where, in the above

$$\tilde{B}_x(\cdot,t) = \int_0^t B_x(\cdot,s)\ \mathrm{ds}. \tag{5.63}$$

From (5.62, ii*), we have*

$$\begin{aligned}\left((\frac{\partial}{\partial t} + \sigma)B_y, H_y\right) &= \left(\frac{\partial H_y}{\partial t}, H_y\right) + (\sigma H_y, H_y) \\ &= \frac{1}{2}\frac{d}{dt}\left\|H_y\right\|^2_{L^2(\Omega)} + (\sigma H_y, H_y)\,.\end{aligned} \tag{5.64}$$

Next, from (5.62, iii*), we have*

$$\begin{aligned}\left((\frac{\partial}{\partial t} + \sigma)D, E\right) &= \left(\frac{\partial E}{\partial t}, E\right) + (\sigma E, E) \\ &= \frac{1}{2}\frac{d}{dt}\left\|E\right\|^2_{L^2(\Omega)} + (\sigma E, E)\,.\end{aligned} \tag{5.65}$$

Finally, from (5.36, ii*), we have*

$$\begin{aligned}\left(\frac{\partial B_x}{\partial t}, H_x\right) &= \left(\frac{\partial H_x}{\partial t} - \sigma B_x, H_x\right) \\ &= \left(\frac{\partial H_x}{\partial t}, H_x\right) - (\sigma B_x, H_x)\,.\end{aligned} \tag{5.66}$$

Using (5.62, i*), we obtain*

$$\begin{aligned}\left(\frac{\partial B_x}{\partial t}, H_x\right) &= \frac{1}{2}\frac{d}{dt}\left\|H_x\right\|^2_{L^2(\Omega)} - \left(\sigma(H_x - \sigma\tilde{B}_x), H_x\right) \\ &= \frac{1}{2}\frac{d}{dt}\left\|H_x\right\|^2_{L^2(\Omega)} - (\sigma H_x, H_x) + \left(\sigma^2\tilde{B}_x, H_x\right).\end{aligned} \tag{5.67}$$

The last term in (5.67) *can be rewritten using (5.62,* i*), (5.36,* ii*) and rearranging terms as*

$$\begin{aligned}\left(\sigma^2\tilde{B}_x, H_x\right) &= (\sigma(H_x - B_x), H_x) \\ &= (\sigma(H_x - B_x), H_x - B_x) + \left(\frac{\partial(H_x - B_x)}{\partial t}, H_x - B_x\right).\end{aligned} \tag{5.68}$$

Using (5.68) *in* (5.67) *we get*

$$\begin{aligned}\left(\frac{\partial B_x}{\partial t}, H_x\right) = \frac{1}{2}\frac{d}{dt}&\left(\left\|H_x\right\|^2_{L^2(\Omega)} + \left\|H_x - B_x\right\|^2_{L^2(\Omega)}\right) \\ &- (\sigma H_x, H_x) + (\sigma(H_x - B_x), H_x - B_x).\end{aligned} \tag{5.69}$$

Substituting (5.64)*,* (5.65) *and* (5.69) *in* (5.61) *we have*

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}&\left(\left\|H_x\right\|^2_{L^2(\Omega)} + \left\|H_y\right\|^2_{L^2(\Omega)} + \left\|E\right\|^2_{L^2(\Omega)} + \left\|H_x - B_x\right\|^2_{L^2(\Omega)}\right) \\ &= (\sigma H_x, H_x) - (\sigma H_y, H_y) - (\sigma E, E) - (\sigma(H_x - B_x), H_x - B_x).\end{aligned} \tag{5.70}$$

The right hand side of (5.70) *can be bounded as*

$$\begin{aligned}&(\sigma H_x, H_x) - (\sigma H_y, H_y) - (\sigma E, E) - (\sigma(H_x - B_x), H_x - B_x) \\ &\leq 2\left\|\sigma\right\|_\infty \{\frac{1}{2}\left(\left\|H_x\right\|^2_{L^2(\Omega)} + \left\|H_y\right\|^2_{L^2(\Omega)} + \left\|E\right\|^2_{L^2(\Omega)} + \left\|H_x - B_x\right\|^2_{L^2(\Omega)}\right)\} \\ &= 2\left\|\sigma\right\|_\infty \mathcal{E}_0^x,\end{aligned} \tag{5.71}$$

where, $\mathcal{E}_0^x$ *is defined in* (5.59)*. Thus, from* (5.59)*,* (5.70) *and* (5.71) *we have*

$$\begin{aligned}&\frac{d}{dt}\mathcal{E}_0^x(t) \leq 2\left\|\sigma\right\|_\infty \mathcal{E}_0^x(t) \\ &\implies \mathcal{E}_0^x(t) \leq \mathcal{E}_0^x(0) + 2\left\|\sigma\right\|_\infty \int_0^t \mathcal{E}_0^x(s)\ \mathrm{ds}.\end{aligned} \tag{5.72}$$

Energy Estimate 4 *Assume that in a corner domain of the PML, both σ_x and σ_y are in $L^\infty(\Omega)$. Let $\mathbf{B} = (B_x, B_y)$. If the product $\sigma_x\sigma_y$ remains positive everywhere in the domain of interest, then the zero order energy of the UPML given by*

$$\mathcal{E}_0^C(t) = \frac{1}{2}\left(\|\mathbf{H}-\mathbf{B}\|^2_{L^2(\Omega)} + \|E\|^2_{L^2(\Omega)} + \|\mathbf{H}\|^2_{L^2(\Omega)} + \left(\sigma_x\sigma_y\tilde{E},\tilde{E}\right)_{L^2(\Omega)}\right), \tag{5.73}$$

where

$$\tilde{E}(t) = \int_0^t E(\cdot,s)ds, \tag{5.74}$$

satisfies the energy estimate

$$\mathcal{E}_0^C(t) \leq \mathcal{E}_0^C(0) + 3\left(\|\sigma_x\|_\infty + \|\sigma_y\|_\infty\right)\int_0^t \mathcal{E}_0^C(s)ds. \tag{5.75}$$

Proof 6 : *The UPML equations in this case are given by* (5.48)*. From (5.48,* i, iii*), we get after integrating by parts,*

$$\left((\frac{\partial}{\partial t}+\Sigma_2)\mathbf{B},\mathbf{H}\right) + \left((\frac{\partial}{\partial t}+\sigma_x)D,E\right) = -\left(\overrightarrow{\text{curl}}\,E,\mathbf{H}\right) + (\text{curl}\,\mathbf{H},E) = 0. \tag{5.76}$$

Consider the term

$$\left((\frac{\partial}{\partial t}+\sigma_x)D,E\right) = \left(\frac{\partial D}{\partial t},E\right) + (\sigma_x D,E)\,. \tag{5.77}$$

From (5.48, iv*), assuming zero initial conditions we have*

$$D(\cdot,t) = E(\cdot,t) + \sigma_y\tilde{E}(\cdot,t), \tag{5.78}$$

where, $\tilde{E}(t)$ is defined in (5.74)*. Thus, from* (5.77) *and* (5.78) *we have*

$$\begin{aligned}\left((\frac{\partial}{\partial t}+\sigma_x)D,E\right) &= \left(\frac{\partial E}{\partial t},E\right) + (\sigma_x E,E) + (\sigma_y E,E) + \left(\sigma_x\sigma_y\tilde{E},E\right)\\ &= \frac{1}{2}\frac{d}{dt}\|E\|^2_{L^2(\Omega)} + \frac{1}{2}\frac{d}{dt}\left(\sigma_x\sigma_y\tilde{E},\tilde{E}\right) + ((\sigma_x+\sigma_y)E,E)\,.\end{aligned} \tag{5.79}$$

From, (5.48, ii*), we have*

$$\mathbf{H}(\cdot,t) = \mathbf{B}(\cdot,t) + \Sigma_1\tilde{\mathbf{B}}(\cdot,t), \tag{5.80}$$

where

$$\tilde{\mathbf{B}}(\cdot,t)=\int_0^t \mathbf{B}(\cdot,s)\,\mathrm{ds} \tag{5.81}$$

From, (5.80) *and* (5.81) *we get*

$$\begin{aligned}\left((\frac{\partial}{\partial t}+\Sigma_2)\mathbf{B},\mathbf{H}\right)&=\left(\frac{\partial \mathbf{H}}{\partial t},\mathbf{H}\right)+((\sigma_y-\sigma_x)B_x,H_x)+((\sigma_x-\sigma_y)B_y,H_y)\\&=\frac{1}{2}\frac{d}{dt}\left\|\mathbf{H}\right\|^2_{L^2(\Omega)}+(\Sigma_2\mathbf{B},\mathbf{H})-(\Sigma_1\mathbf{B},\mathbf{H})\,.\end{aligned} \tag{5.82}$$

Again, from (5.80) *and* (5.81) *we get*

$$\begin{aligned}(\Sigma_1\mathbf{B},\mathbf{H})&=(\sigma_xB_x,H_x)+(\sigma_yB_y,H_y)\\&=\left(\sigma_x(H_x-\sigma_x\tilde{B}_x),H_x\right)+\left(\sigma_y(H_y-\sigma_y\tilde{B}_y),H_y\right)\\&=(\sigma_xH_x,H_x)+(\sigma_yH_y,H_y)-\left(\sigma_x^2\tilde{B}_x,H_x\right)-\left(\sigma_y^2\tilde{B}_y,H_y\right)\\&=(\Sigma_1\mathbf{H},\mathbf{H})-\left(\sigma_x^2\tilde{B}_x,H_x\right)-\left(\sigma_y^2\tilde{B}_y,H_y\right).\end{aligned} \tag{5.83}$$

Using (5.83) *in* (5.82) *we get*

$$\begin{aligned}\left((\frac{\partial}{\partial t}+\Sigma_2)\mathbf{B},\mathbf{H}\right)&=\frac{1}{2}\frac{d}{dt}\left\|\mathbf{H}\right\|^2_{L^2(\Omega)}+(\Sigma_2\mathbf{B},\mathbf{H})-(\Sigma_1\mathbf{H},\mathbf{H})\\&\quad+\left(\sigma_x^2\tilde{B}_x,H_x\right)+\left(\sigma_y^2\tilde{B}_y,H_y\right)\\&=\frac{1}{2}\frac{d}{dt}\left\|\mathbf{H}\right\|^2_{L^2(\Omega)}+(\Sigma_2\mathbf{B},\mathbf{H})-(\Sigma_1\mathbf{H},\mathbf{H})+\left(\Sigma_1^2\tilde{\mathbf{B}},\mathbf{H}\right).\end{aligned} \tag{5.84}$$

We can simplify the last term in (5.84) *as follows*

$$\begin{aligned}\left(\Sigma_1^2\tilde{\mathbf{B}},\mathbf{H}\right)&=(\Sigma_1(\mathbf{H}-\mathbf{B}),\mathbf{H}-\mathbf{B})+(\Sigma_1\mathbf{H},\mathbf{B})\\&\quad+\left(\frac{\partial(\mathbf{H}-\mathbf{B})}{\partial t},\mathbf{H}-\mathbf{B}\right)-(\Sigma_1\mathbf{B},\mathbf{H})\\&=\frac{1}{2}\frac{d}{dt}\left\|\mathbf{H}-\mathbf{B}\right\|^2_{L^2(\Omega)}+(\Sigma_1(\mathbf{H}-\mathbf{B}),\mathbf{H}-\mathbf{B})\,.\end{aligned} \tag{5.85}$$

Also

$$(\Sigma_2\mathbf{B},\mathbf{H})=(\Sigma_2(\mathbf{B}-\mathbf{H}),\mathbf{H})+(\Sigma_2\mathbf{H},\mathbf{H})\,. \tag{5.86}$$

From (5.84)*,* (5.85)*, and* (5.86) *we have*

$$\begin{aligned}\left((\frac{\partial}{\partial t}+\Sigma_2)\mathbf{B},\mathbf{H}\right)&=\frac{1}{2}\frac{d}{dt}\left(\left\|\mathbf{H}\right\|^2_{L^2(\Omega)}+\left\|\mathbf{H}-\mathbf{B}\right\|^2_{L^2(\Omega)}\right)+(\Sigma_2(\mathbf{B}-\mathbf{H}),\mathbf{H})\\&\quad+((\Sigma_2-\Sigma_1)\mathbf{H},\mathbf{H})+(\Sigma_1(\mathbf{H}-\mathbf{B}),\mathbf{H}-\mathbf{B})\end{aligned} \tag{5.87}$$

From (5.79) *and* (5.87) *we have*

$$\left((\frac{\partial}{\partial t}+\sigma_x)D,E\right)+\left((\frac{\partial}{\partial t}+\Sigma_2)\mathbf{B},\mathbf{H}\right)=(\Sigma_1(\mathbf{H}-\mathbf{B}),\mathbf{H}-\mathbf{B})$$
$$+\frac{1}{2}\frac{d}{dt}\left(\|E\|^2_{L^2(\Omega)}+\|\mathbf{H}\|^2_{L^2(\Omega)}+\|\mathbf{H}-\mathbf{B}\|^2_{L^2(\Omega)}+\left(\sigma_x\sigma_y\tilde{E},\tilde{E}\right)\right)$$
$$+((\sigma_x+\sigma_y)E,E)+(\Sigma_2(\mathbf{B}-\mathbf{H}),\mathbf{H})+((\Sigma_2-\Sigma_1)\mathbf{H},\mathbf{H})=0. \quad (5.88)$$

Using the definition of the energy $\mathcal{E}_0^C$ (5.73) *we have*

$$\frac{d}{dt}\mathcal{E}_0^C=((\Sigma_1-\Sigma_2)\mathbf{H},\mathbf{H})-(\Sigma_1(\mathbf{H}-\mathbf{B}),\mathbf{H}-\mathbf{B})-((\sigma_x+\sigma_y)E,E)$$
$$+(\Sigma_2(\mathbf{H}-\mathbf{B}),\mathbf{H})\,. \quad (5.89)$$

We can bound the terms on the right hand hand of (5.89) *as follows. We have*

$$(\Sigma_2(\mathbf{H}-\mathbf{B}),\mathbf{H})+((\Sigma_1-\Sigma_2)\mathbf{H},\mathbf{H})$$
$$\leq\ \{\|\sigma_x\|_\infty+\|\sigma_y\|_\infty\}\left[\frac{1}{2}\left(\|\mathbf{H}\|^2_{L^2(\Omega)}+\|\mathbf{H}-\mathbf{B}\|^2_{L^2(\Omega)}\right)+\|\mathbf{H}\|^2_{L^2(\Omega)}\right]. \quad (5.90)$$

Substituting (5.90) *in* (5.89) *we obtain*

$$\frac{d}{dt}\mathcal{E}_0^C\leq\ \{\|\sigma_x\|_\infty+\|\sigma_y\|_\infty\}\left[\|E\|^2_{L^2(\Omega)}+\frac{3}{2}\|\mathbf{H}\|^2_{L^2(\Omega)}+\frac{3}{2}\|\mathbf{H}-\mathbf{B}\|^2_{L^2(\Omega)}\right]$$
$$\leq 3\{\|\sigma_x\|_\infty+\|\sigma_y\|_\infty\}\mathcal{E}_0^C. \quad (5.91)$$

Integrating, we thus have the energy estimate

$$\mathcal{E}_0^C(t)\leq\mathcal{E}_0^C(0)+3\left(\|\sigma_x\|_\infty+\|\sigma_y\|_\infty\right)\int_0^t\mathcal{E}_0^C(s)ds, \quad (5.92)$$

which is (5.75).

5.7 The Discrete Mixed Finite Element Scheme

5.7.1 Space Discretization

Let Ω now be a union of rectangles such that we can consider a regular mesh $(\mathcal{T}_h)$ with square elements (K) of edge $h>0$ as in Figure 5.2.

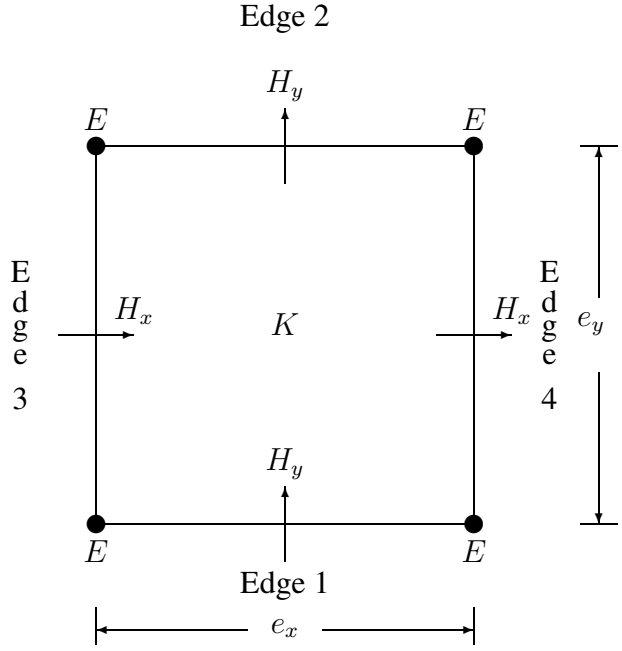

Figure 5.2: A sample domain element K. The degrees of freedom for the electric and magnetic field are staggered in space. E is a bilinear function with degrees of freedom at the nodes of the square. The degrees of freedom for H_x and H_y are the midpoints of edges parallel to the x-axis and y-axis, respectively.

We consider the following approximation space for $\mathbf{H}$ and $\mathbf{B}$:

$$\mathcal{V}_h = \{\mathbf{\Psi}_h \in H(\mathrm{div}, \Omega) |\ \forall K \in \mathcal{T}_h, \mathbf{\Psi}_h|_K \in RT_{[0]}\}, \tag{5.93}$$

where, $RT_{[0]} = P_{10} \times P_{01}$, is the lowest order Raviart Thomas space [116] and for $k_1, k_2 \in \mathbb{N} \cup \{0\}$,

$$P_{k_1 k_2} = \{p(x_1, x_2) | p(x_1, x_2) = \sum_{0 \le i \le k_1} \sum_{0 \le j \le k_2} a_{ij} x_1^i x_2^j\}.$$

The basis functions for H_x and H_y have unity value along one e_y or e_x edge, respectively, and zero over all other edges (see Figure 5.2).

The approximation space for E and D is chosen to be

$$\mathcal{U}_h = \{\phi_h \in H_0^1(\Omega) | \quad \forall K \in \mathcal{T}_h, \phi_h|_K \in Q_1\}, \tag{5.94}$$

where, the space $Q_1 = P_{11}$. The basis functions for E have unity value at one node and are zero at all other nodes. Figure 5.2 shows the locations for the degrees of freedom for both approximation spaces.

Based on the approximation spaces described above, we define the space discrete scheme as:

Find $(E_h(\cdot, t), D_h(\cdot, t), \mathbf{H}_h(\cdot, t), \mathbf{B}_h(\cdot, t)) \in \mathcal{U}_h \times \mathcal{U}_h \times \mathcal{V}_h \times \mathcal{V}_h$ such that for all $\mathbf{\Psi}_h \in \mathcal{V}_h$, for all $\phi_h \in \mathcal{U}_h$,

$$\left\{ \begin{array}{rcl} \frac{d}{dt}\int_\Omega \mathbf{B}_h \cdot \mathbf{\Psi}_h d\mathbf{x} & = & -\frac{1}{\epsilon_0}\int_\Omega \Sigma_1 \mathbf{B}_h \cdot \mathbf{\Psi}_h d\mathbf{x} \ - \ \int_\Omega \overrightarrow{\mathrm{curl}}\, E_h \cdot \mathbf{\Psi}_h d\mathbf{x}, \\ \frac{d}{dt}\int_\Omega \mathbf{H}_h \cdot \mathbf{\Psi}_h d\mathbf{x} & = & \frac{1}{\mu_0}\frac{d}{dt}\int_\Omega \mathbf{B}_h \cdot \mathbf{\Psi}_h d\mathbf{x} \ + \ \frac{1}{\epsilon_0 \mu_0}\int_\Omega \Sigma_2 \mathbf{B}_h \cdot \mathbf{\Psi}_h d\mathbf{x}, \\ \frac{d}{dt}\int_\Omega D_h \cdot \phi_h \, d\mathbf{x} & = & -\frac{1}{\epsilon_0}\int_\Omega \sigma_x D_h \cdot \phi_h \, d\mathbf{x} \ + \ \int_\Omega \overrightarrow{\mathrm{curl}}\, \phi_h \cdot \mathbf{H}_h \, d\mathbf{x}, \\ \frac{d}{dt}\int_\Omega E_h \cdot \phi_h \, d\mathbf{x} & = & -\frac{1}{\epsilon_0}\int_\Omega \sigma_y E_h \cdot \phi_h \, d\mathbf{x} \ + \ \frac{1}{\epsilon_0}\frac{d}{dt}\int_\Omega D_h \cdot \phi_h \, d\mathbf{x}. \end{array} \right. \tag{5.95}$$

5.7.2 Time Discretization

For the time discretization we use a leapfrog scheme, i.e., a centered second order accurate finite difference scheme. For the zero order terms, we use a semi-implicit approximation,

[124] as described below. Let $(\cdot,\cdot)$ denote the L^2 norm in Ω. Define

$$\Delta_t V^n = \frac{V^{n+1/2} - V^{n-1/2}}{\Delta t}, \tag{5.96}$$

and

$$\underline{V}^n = \frac{V^{n+1/2} + V^{n-1/2}}{2}. \tag{5.97}$$

Using the definitions above, we can describe the fully discrete scheme in space and time as

Find $(E_h^{n+1}, D_h^{n+1}, \mathbf{H}_h^{n+\frac{1}{2}}, \mathbf{B}_h^{n+\frac{1}{2}}) \in \mathcal{U}_h \times \mathcal{U}_h \times \mathcal{V}_h \times \mathcal{V}_h$ such that for all $\mathbf{\Psi}_h \in \mathcal{V}_h$, for all $\phi_h \in \mathcal{U}_h$,

$$\left\{\begin{array}{llcll}
\text{(i)} & (\Delta_t \mathbf{B}_h^n, \mathbf{\Psi_h}) & = & -\dfrac{1}{\epsilon_0}(\Sigma_2 \underline{\mathbf{B}_h^n}, \mathbf{\Psi}_h) & - \quad (\overrightarrow{\text{curl}}\, E_h^n\ ,\ \mathbf{\Psi}_h), \\
\text{(ii)} & (\Delta_t \mathbf{H}_h^n, \mathbf{\Psi}_h) & = & \dfrac{1}{\mu_0}(\Delta_t \mathbf{B}_h^n, \mathbf{\Psi}_h) & + \quad \dfrac{1}{\epsilon_0\mu_0}(\Sigma_1 \underline{\mathbf{B}_h^n}, \mathbf{\Psi}_h), \\
\text{(iii)} & (\Delta_t D_h^{n+\frac{1}{2}}, \phi_h) & = & -\dfrac{1}{\epsilon_0}(\sigma_x \underline{D_h}^{n+\frac{1}{2}}, \phi_h) & + \quad (\overrightarrow{\text{curl}}\, \phi_h, \mathbf{H}_h^{n+\frac{1}{2}}), \\
\text{(iv)} & (\Delta_t E_h^{n+\frac{1}{2}}, \phi_h) & = & -\dfrac{1}{\epsilon_0}(\sigma_y \underline{E_h}^{n+\frac{1}{2}}, \phi_h) & + \quad \dfrac{1}{\epsilon_0}(\Delta_t D_h^{n+\frac{1}{2}},\ \phi_h).
\end{array}\right. \tag{5.98}$$

We remark here that the UPML- FEM model (5.98), must satisfy the stability condition (CFL)

$$c\Delta t \leq \frac{1}{\sqrt{3\left(\dfrac{1}{h_x^2} + \dfrac{1}{h_y^2}\right)}}, \tag{5.99}$$

where h_x and h_y are the step sizes along the x and y axes, respectively, and c is the speed of light in free space. If we choose a uniform mesh, i.e., $h_x = h_y = h$, then the corresponding stability condition is

$$\eta = \frac{c\Delta t}{h} \leq \frac{1}{\sqrt{6}}. \tag{5.100}$$

Equation (5.100) will be justified in Section 5.8, where we perform a dispersion analysis for the mixed finite element scheme (5.98). Solving system (5.98) involves the solution of linear systems associated with (5.98, ii), and (5.98, iv), at each time step. We solve these linear systems by a preconditioned conjugate gradient method with the diagonal of the corresponding mass matrices as a preconditioner.

5.8 Dispersion Analysis

The numerical approximation of time-dependent wave problems introduces errors which involve *dissipation*, *dispersion*, and *anisotropy*. The attenuation of the amplitude of the plane wave is referred to as dissipation. The mixed finite element scheme presented in Section 5.7 is a non-dissipative scheme for Maxwell's equations in free space [88]. However, the PML model, in which lower order terms are added, is a dissipative model. As seen in (5.15) waves are attenuated in the PML according to the distance of travel in a given direction.

The numerical model produces waves that propagate at incorrect speeds. The dependence of the velocity of propagation of the numerical sinusoidal waves on frequency is termed as dispersion, which is the focus of this section. In addition, this velocity is also dependent upon direction. This is refered to as numerical anisotropy. All these errors are cumulative in nature which implies that after time integration over long intervals the solution can become polluted and may completely deviate from the correct solution [88].

A dispersive equation is one that admits plane wave solutions of the form $\mathrm{e}^{\mathrm{i}(\omega t - \mathbf{k}\cdot\mathbf{x})}$, but with the property that the speed of propagation of these waves is not independent of $\mathbf{k}$ [126]. Whether a PDE is dispersive or non dispersive, any discrete model of the PDE will be dispersive. For such time-harmonic waves, numerical dispersion results in the creation of a *phase error* in the solution. This is due to the incorrect modeling of the sinusoidal behavior of the propagating wave, as the piecewise polynomial approximation of a finite element method does not exactly match a sine or cosine function [86]. A dispersion analysis, or a plane wave analysis, of a discrete model will help describe the propagation of waves in the numerical method away from the boundaries, In addition, this analysis will also give information on the expected accuracy of the methods. Such an analysis, thus, is very important in understanding the behavior of the numerical model. A study of the dispersion analysis of different mixed methods for the time domain Maxwell's equations is done in

[101]. We now study the properties of the discrete PML model in terms of a numerical dispersion analysis. This analysis, which is performed on regular square elements, does not really model the phase error for arbitrary unstructured meshes; however, it does help to give a relative comparison with other methods [86]. Recall that $\mathbf{k}$ denotes the wave vector for the continuous case. For simplicity, we again assume that $\epsilon_0 = \mu_0 = 1$, hence $c = 1$. We look for solutions to the continuous system (5.33), of the form

$$V(x, y, t) = V_0 \, \mathrm{e}^{\mathrm{i}\omega t - \mathbf{k}\cdot\mathbf{x}}, \tag{5.101}$$

where V is any one of E, D, H_x, H_y, B_x, B_y. Substituting (5.101) in (5.33) shows that ω and $\mathbf{k}$ are related by the dispersion relation

$$\omega^2 = \left(\frac{k_x}{s_x}\right)^2 + \left(\frac{k_y}{s_y}\right)^2 \tag{5.102}$$

Inside the computational domain, where $s_x = s_y = 1$, the dispersion relation is given by

$$\omega^2 = k_x^2 + k_y^2 = |\mathbf{k}|^2 \implies \omega = |\mathbf{k}|. \tag{5.103}$$

Other solutions are $\omega = 0$, or $\omega = -|\mathbf{k}|$. There are two types of velocities that are important here [126]. The *phase velocity* is defined as

$$c = \frac{\omega}{|\mathbf{k}|}, \tag{5.104}$$

which in this case is 1, as we have assumed $\epsilon_0 = \mu_0 = 1$. The *group velocity* is defined to be

$$\mathbf{C}(\mathbf{k}) = \nabla_{\mathbf{k}}\omega(\mathbf{k}) = \frac{\mathbf{k}}{|\mathbf{k}|}. \tag{5.105}$$

It is a well known fact, that the propagation of energy under dispersive partial differential equations is governed by the group velocity [126, 26, 130]. Asymptotically, the energy associated with the wave vector $\mathbf{k}$ moves at the *group speed* $|\mathbf{C}|$, which in the present case is $|\mathbf{C}| = 1$. Thus, regardless of the wave number $\mathbf{k}$, all plane waves move with the same group speed $|\mathbf{C}| = 1$.

We now perform a similar analysis for the semi-discrete system (5.95), where we consider exact integration in time. Let us assume an infinite PML in the region $x > 0$. Thus, $\sigma_y = 0$ and let $\sigma_x = \sigma$. We will look for solutions of the form

$$V(x, y, t) = \hat{V}(x, y)\, \mathrm{e}^{\mathrm{i}\omega t}, \tag{5.106}$$

to the semi-discrete system (5.95). Substituting (5.106) in (5.95), we obtain the time harmonic system

$$\left\{\begin{array}{rcl} \mathrm{i}\omega(\hat{B}_x, \psi_x) & = & -\left(\dfrac{\partial \hat{E}}{\partial y}, \psi_x\right), \\ \mathrm{i}\omega(\hat{H}_x, \psi_x) & = & \mathrm{i}\omega\left(\left(1 + \dfrac{\sigma}{\mathrm{i}\omega}\right)\hat{B}_x, \psi_x\right), \\ \mathrm{i}\omega\left(\left(1 + \dfrac{\sigma}{\mathrm{i}\omega}\right)\hat{B}_y, \psi_y\right) & = & \left(\dfrac{\partial \hat{E}}{\partial x}, \psi_y\right), \\ \mathrm{i}\omega(\hat{H}_y, \psi_y) & = & \mathrm{i}\omega(\hat{B}_y, \psi_y), \\ \mathrm{i}\omega\left(\left(1 + \dfrac{\sigma}{\mathrm{i}\omega}\right)\hat{D}, \phi\right) & = & \left(\hat{H}_x, \dfrac{\partial \phi}{\partial y}\right) - \left(\hat{H}_y, \dfrac{\partial \phi}{\partial x}\right), \\ \mathrm{i}\omega(\hat{E}, \phi) & = & \mathrm{i}\omega(\hat{D}, \phi). \end{array}\right. \tag{5.107}$$

We assume that σ is a piecewise constant function of x with jumps at $x = lh, l = 0, 1, 2, \ldots$, where $h = h_x = h_y$ is the mesh step size. Let

$$\sigma_l = \begin{cases} \text{Value of } \sigma \text{ on } (lh, (l+1)h), & \text{if } l \geq 0, \\ 0, & \text{if } l < 0. \end{cases} \tag{5.108}$$

Using the definition (5.18), we have

$$s_{x,l} = s_l = 1 + \frac{\sigma_l}{\mathrm{i}\omega}. \tag{5.109}$$

Since $\sigma_y = 0$, we have $s_y = 0$. As the PML is in the half space $x > 0$, $x = 0$ is the interface

between the PML and the interior computation region. Let us define

$$\begin{cases} M_x u_{l,m} &= 4u_{l,m} + u_{l-1,m} + u_{l+1,m}, \\ M_y u_{l,m} &= 4u_{l,m} + u_{l,m-1} + u_{l,m+1}, \\ \tilde{S}_x u_{l,m} &= M_y u_{l-1/2,m} - M_y u_{l+1/2,m}, \\ \tilde{S}_y u_{l,m} &= M_x u_{l,m-1/2} - M_x u_{l,m+1/2}, \\ M_z u_{l,m} &= M_x M_y u_{l,m}. \end{cases} \tag{5.110}$$

Consider an interior super element as shown in Figure 5.3. Using the definitions (5.110) in (5.107), we obtain the following system of equations that corresponds to the space discrete finite element scheme (5.95):

$$\begin{cases} M_x \hat{B}_{l,m+1/2} &= \dfrac{\mathrm{i}}{\omega h} M_x(\hat{E}_{l,m+1} - \hat{E}_{l,m}), \\ M_x \hat{H}_{l,m+1/2} &= \left(\dfrac{s_l + s_{l-1}}{2}\right) M_x \hat{B}_{l,m+1/2}, \\ s_l M_y \hat{B}_{l+1/2,m} &= \dfrac{-\mathrm{i}}{\omega h} M_y(\hat{E}_{l+1,m} - \hat{E}_{l,m}), \\ M_y \hat{H}_{l+1/2,m} &= M_y \hat{B}_{l+1/2,m}, \\ \left(\dfrac{s_l + s_{l-1}}{2}\right) M_z \hat{D}_{l,m} &= \dfrac{-6\mathrm{i}}{\omega h}(\tilde{S}_y \hat{H}_{l,m} - \tilde{S}_x \hat{H}_{l,m}), \\ M_z \hat{E}_{l,m} &= M_z \hat{D}_{l,m}. \end{cases} \tag{5.111}$$

Combining the equations in (5.111), we obtain an equation in E by eliminating the other variables, which is

$$\begin{aligned} -\frac{\omega^2 h^2}{6}\left(\frac{s_l + s_{l-1}}{2}\right) M_z \hat{E}_{l,m} = \left(\frac{s_l + s_{l-1}}{2}\right)(M_x \hat{E}_{l,m+1} - 2M_x \hat{E}_{l,m} + M_x \hat{E}_{l,m-1}) \\ + \frac{1}{s_l}(M_y \hat{E}_{l+1,m} - M_y \hat{E}_{l,m}) - \frac{1}{s_{l-1}}(M_y \hat{E}_{l,m} - M_y \hat{E}_{l-1,m}). \end{aligned} \tag{5.112}$$

Let us now look for solutions to (5.112) of the form

$$\hat{E}_{l,m} = \hat{E}_l \, \mathrm{e}^{-\mathrm{i}k_y m h}. \tag{5.113}$$

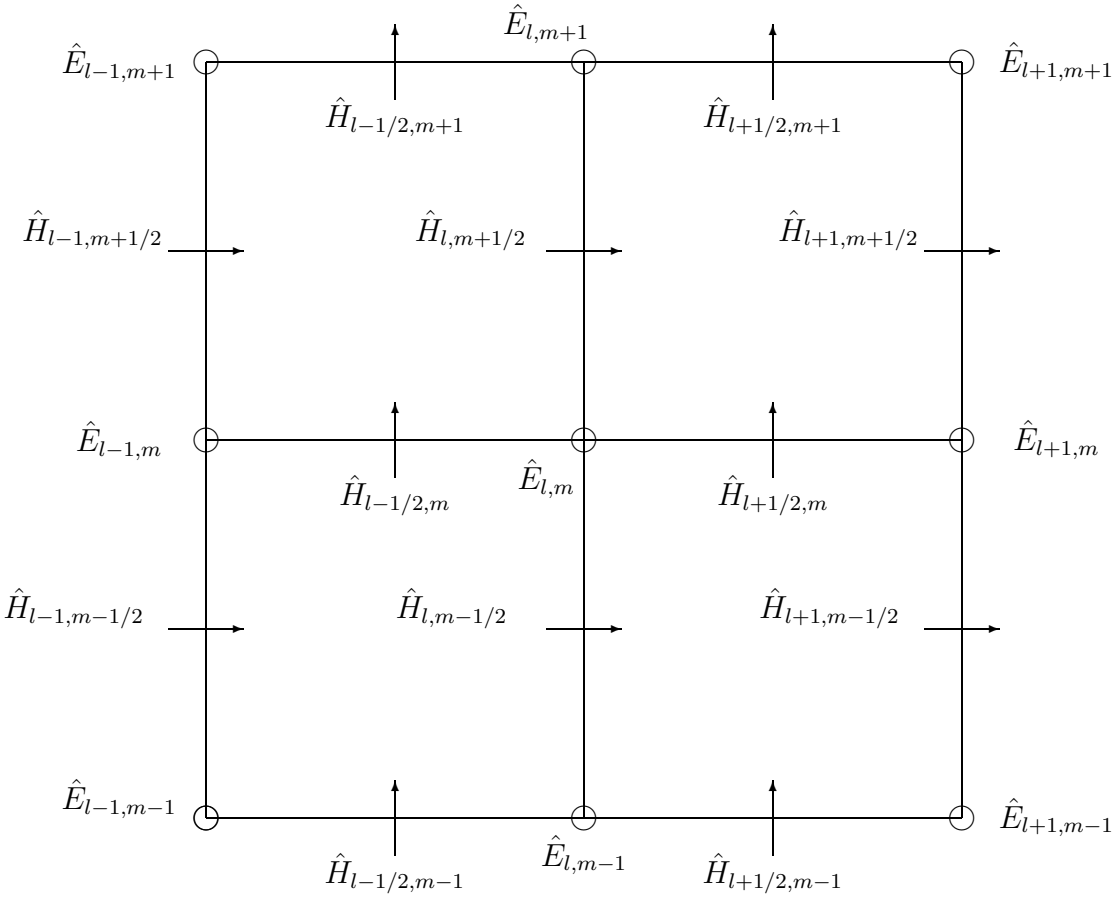

Figure 5.3: Dependency diagram for an interior super element. A degree of freedom $\hat{E}_{l,m}$, away from the boundary of the domain Ω, depends on 8 other electric degrees of freedom and 12 magnetic degrees of freedom.

Substituting (5.113) in (5.112), and after performing some algebra, we obtain

$$-\frac{\zeta\omega^2h^2}{6}\left(\frac{s_l+s_{l-1}}{2}\right)(4\hat{E}_l+\hat{E}_{l-1}+\hat{E}_{l+1})=\frac{1}{s_l}(\hat{E}_{l+1}-\hat{E}_l)-\frac{1}{s_{l-1}}(\hat{E}_l-\hat{E}_{l-1}), \quad (5.114)$$

where the term ζ is defined as

$$\zeta=1-\frac{12}{\omega^2h^2}\left(\frac{\sin^2(k_yh/2)}{1+2\cos^2(k_yh/2)}\right). \quad (5.115)$$

Let k_x and k_x^{pml} be the x components of the wave vector in free space and the PML, respectively. Assuming $\hat{E}_l=\mathrm{e}^{-\mathrm{i}k_xhl}$ in free space, substituting in (5.114), with $\sigma=0$, we obtain the dispersion relation in free space to be

$$\frac{\omega^2h^2}{12}=\frac{\sin^2(k_xh/2)}{1+2\cos^2(k_xh/2)}+\frac{\sin^2(k_yh/2)}{1+2\cos^2(k_yh/2)}. \quad (5.116)$$

The dispersion relation for the FDTD scheme [124] is

$$\frac{\omega^2h^2}{4}=\sin^2(k_xh/2)+\sin^2(k_yh/2). \quad (5.117)$$

Thus, depending on the magnitude and direction of $\mathbf{k}$, the numerically computed wave has an erroneous phase [101]. As a result a plane wave of the form (5.101) will generally move in an incorrect direction at an incorrect speed. However, we can observe that if k_xh and k_yh are small, then from (5.116) we have

$$\begin{aligned}\omega^+(k_x,k_y,h)&=\frac{2\sqrt{3}}{h}\sqrt{\frac{(k_xh)^2+(k_yh)^2}{12}}\\ \implies \omega^+(k_x,k_y,h)&=\sqrt{k_x^2+k_y^2}=|\mathbf{k}|,\end{aligned} \quad (5.118)$$

where ω^+ denotes the positive solution. Similarly from (5.117) we obtain that $\omega^+=|\mathbf{k}|$. This implies that the effects of dispersion can be reduced to any desired level if we choose a fine enough mesh. To derive the (angular) frequency ω for the fully discrete scheme (5.98), we observe that discretization in time corresponds to replacing ωh in (5.116) by $2\frac{h}{\Delta t}\sin\left(\frac{\omega\Delta t}{2}\right)$. Let us denote the frequency for the fully discrete scheme (5.98) by $\omega_{\Delta t}$. In the working volume, where $\sigma=0$, we get the dispersion relation to be

$$\left(\frac{h}{\sqrt{3}\Delta t}\right)^2\sin^2\left(\frac{\omega_{\Delta t}\Delta t}{2}\right)=\frac{\sin^2(k_xh/2)}{1+2\cos^2(k_xh/2)}+\frac{\sin^2(k_yh/2)}{1+2\cos^2(k_yh/2)}. \quad (5.119)$$

Let $\eta = c\Delta t/h$ ($c = 1$). Solving for $\omega^+_{\Delta t}$ (positive solution) in (5.119) we obtain

$$\omega^+_{\Delta t} = \frac{2}{\Delta t}\sin^{-1}\left(\frac{\eta h\omega^+(k_x, k_y, h)}{2}\right), \tag{5.120}$$

where ω^+ is the (positive) solution to (5.116), i.e.,

$$\omega^+ = \frac{2\sqrt{3}}{h}\sqrt{\frac{\sin^2(k_x h/2)}{1 + 2\cos^2(k_x h/2)} + \frac{\sin^2(k_y h/2)}{1 + 2\cos^2(k_y h/2)}}. \tag{5.121}$$

We note that η must be chosen such that the frequency $\omega^+_{\Delta t}$ is real. This in turn implies that the argument of $\sin^{-1}$ in (5.120) is bounded by 1. We thus need

$$\begin{aligned} &\eta\left(\max_{(k_x h, k_y h)\in[0,\pi]\times[0,\pi]} |h\omega^+(k_x, k_y, h)|\right) \leq 2 \\ &\implies \eta \leq \frac{2}{\sqrt{24}} = \frac{1}{\sqrt{6}} \approx 0.4082, \end{aligned} \tag{5.122}$$

which is the assertion (5.100). A detailed explanation of the error in the group velocity for the case of the free space dispersion relation is given in [101].

In the case of the FDTD scheme a similar analysis yields the stability result

$$\eta \leq \frac{1}{\sqrt{2}} \implies \frac{c\Delta t}{h} \leq \frac{1}{\sqrt{2}} \approx 0.7071, \tag{5.123}$$

where c, which is the speed of propagation, has been chosen to be 1. We obtained a similar result in Chapter 3, Section 3.4.4, Theorem 2. Thus, the finite difference scheme can be viewed as a mass-lumped finite element scheme on a rectangular mesh.

Figure 5.4 compares the dispersion in the proposed FEM scheme and the FDTD scheme for free space, for four different wave propagation angles, $\theta = 0, 15, 30, 45$ degrees, with $\eta = 0.4$. The x axis denotes the number of grid points per wavelength, L/h, where L is the free space wavelength, and the y axis is the numerical phase velocity normalized to the speed of light, v_p/c ($c = 1$). We note that, in both the schemes, the phase velocity is the lowest at 45 degrees, implying that the dispersion is the least along the diagonal of the mesh elements; a fact that will be evident in other results to be presented. In Figure 5.5 similar results are presented for $\eta = 0.01$. We note that the dispersion present in the FEM scheme decreases as the value of η is decreased from 0.4 to 0.01, while the reverse is true in the case of the FDTD scheme.

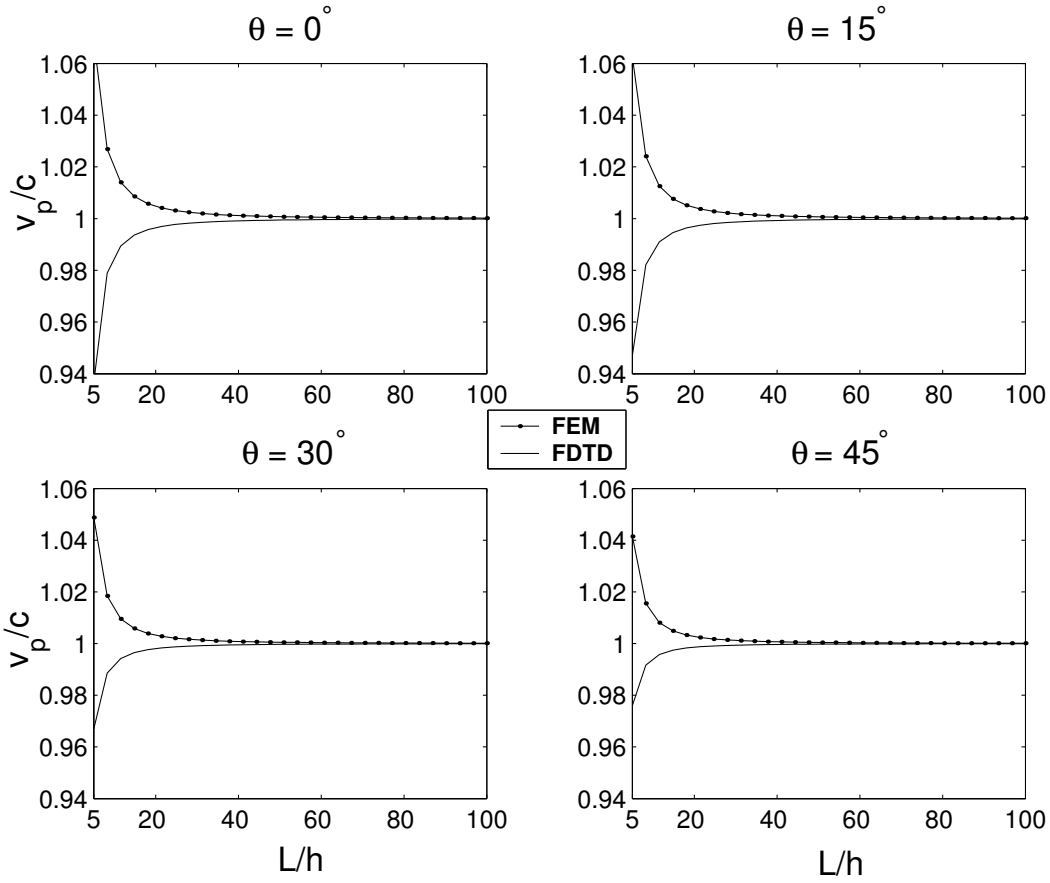

Figure 5.4: Comparison of the dispersion present in the FEM and the FDTD scheme for selected angles of wave propagation. The y axis represents a normalized phase velocity. We see faster than speed of light propagation in the FEM versus slower than the speed of light propagation in the FDTD method. As the number of grid points per wavelength is increased, the phase velocity approaches the speed of light in either case. Here $\eta = 0.4$ for both cases. We notice more dispersion in the FEM as opposed to the FDTD case.

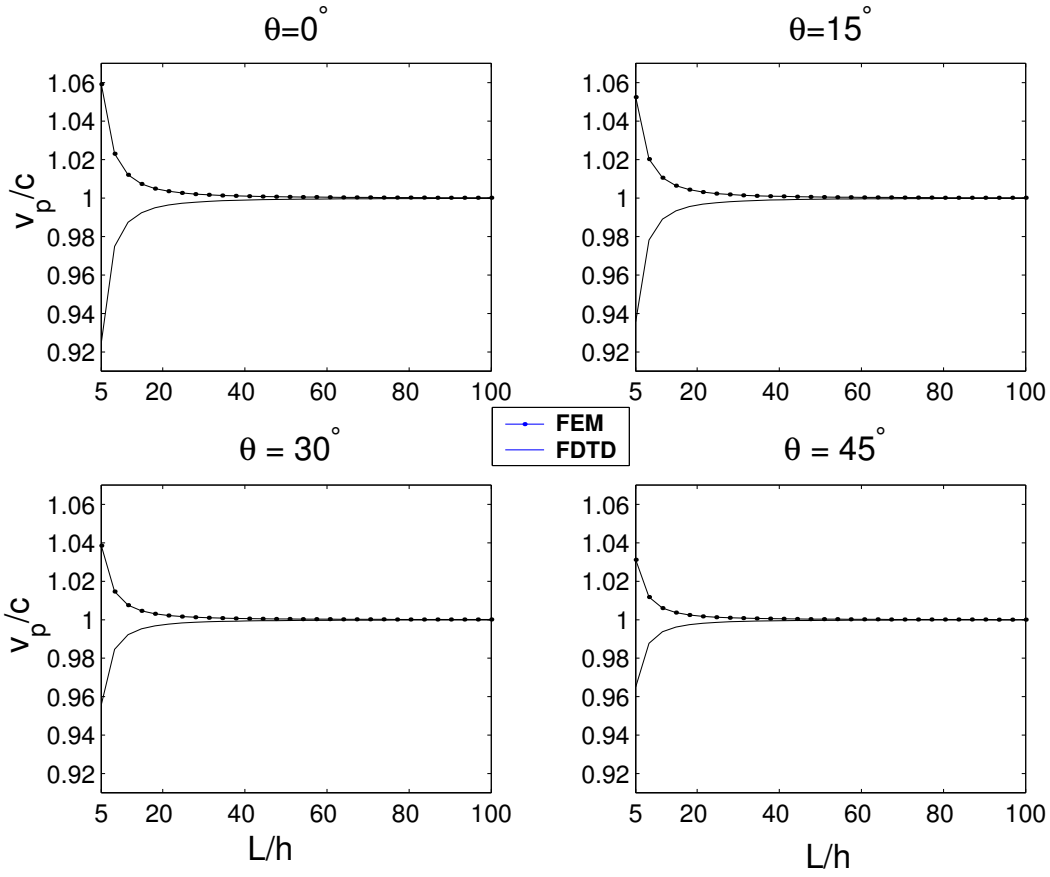

Figure 5.5: Comparison of the dispersion present in the FEM and the FDTD scheme for selected angles of wave propagation. The y axis represents a normalized phase velocity. We see faster than speed of light propagation in the FEM versus slower than the speed of light propagation in the FDTD method. As the number of grid points per wavelength is increased, the phase velocity approaches the speed of light in either case. Here $\eta = 0.01$ for both cases. The dispersion in the FDTD is slightly more than the FEM .

The major effects of dispersion are seen in the case of 10 or less nodes per wavelength. In all cases as L/h becomes large the convergence of v_p to $c\ (=1)$ is clearly seen.

The phase error that results from the dispersion, expressed in degrees per wavelength, is defined as

$$\delta_p = 360^\circ \left| \frac{\tilde{k} - k}{k} \right|, \tag{5.124}$$

where $\tilde{k}$ is the numerical wave number with which the plane wave propagates in the numerical grid. The wave number for the continuous case is k. Figure 5.6 is a polar graph of the phase error for selected values of $L/h = 2\pi/kh$ as a function of θ (dashed lines), for the FEM and FDTD, for $\eta = 0.4$ (top) and $\eta = 0.01$. Due to the symmetry, the same dispersion error is obtained at the angles θ, $\theta + n\ 90^\circ$, and $m\ 90^\circ - \theta$, where m and n are integers. From Figure 5.6, we see that, for both the schemes, the smallest error occurs when the plane wave traverses the elements diagonally ($\theta = \pm\ 45^\circ, \pm\ 135^\circ$). The largest error occurs when the wave is propagating along an axis of the mesh ($\theta = 0^\circ, \pm\ 90^\circ, 180^\circ$) [131]. Similar observations were made earlier in the plots of v_p/c versus L/h. We note that the phase error in the FEM reduces as the value of η is decreased, whereas, in the FDTD scheme, the phase error increases as the value of η is decreased from 0.4 to 0.01. In Figures 5.7 and 5.8 this change in the phase error for the two schemes is more evident. In Figure 5.7, the phase error is plotted for $\eta = 0.4$ (top) and $\eta = 0.2$ (bottom) and in Figure 5.8 for $\eta = 0.1$ (top) and $\eta = 0.01$ (bottom). Again, dispersion effects are more evident for the case of 10 nodes per wavelength. As L/h increases the phase error converges to zero, for all angles of propagation. Thus the effects of dispersion can be reduced to any desired degree by considering a fine enough mesh.

Figure 5.9 is a log-log graph of the phase error δ_p, versus the number of nodes per wavelength L/h, for selected values of the angle θ, for the FDTD scheme and the FEM scheme presented here. We observe that, for both schemes, as L/h is increased, the error becomes smaller. For large values of L/h the graphs are linear, indicating the error to be

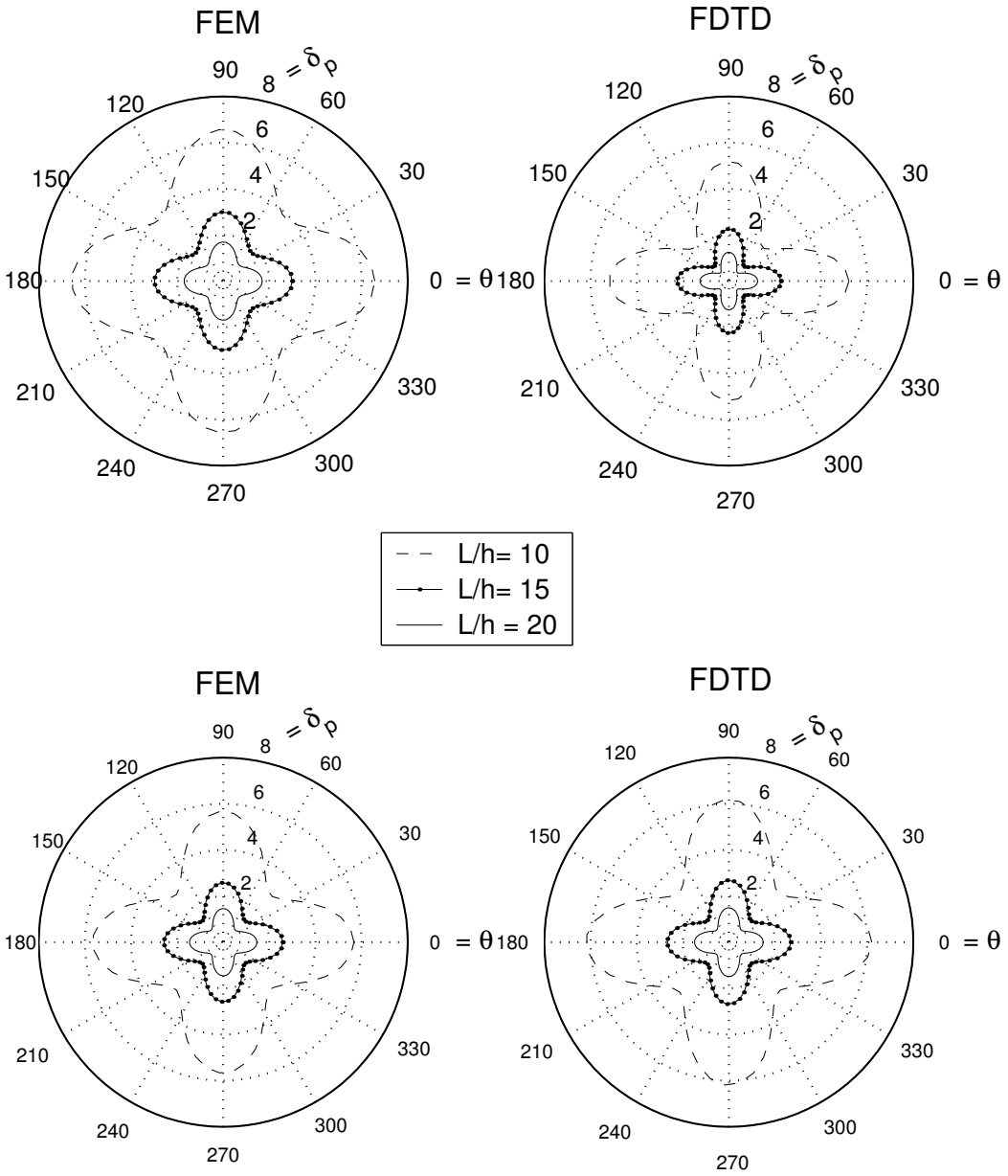

Figure 5.6: Polar graph of the phase error in degrees per wavelength for selected values of L/h, with $\eta = 0.4$ (top) and $\eta = 0.01$ (bottom).

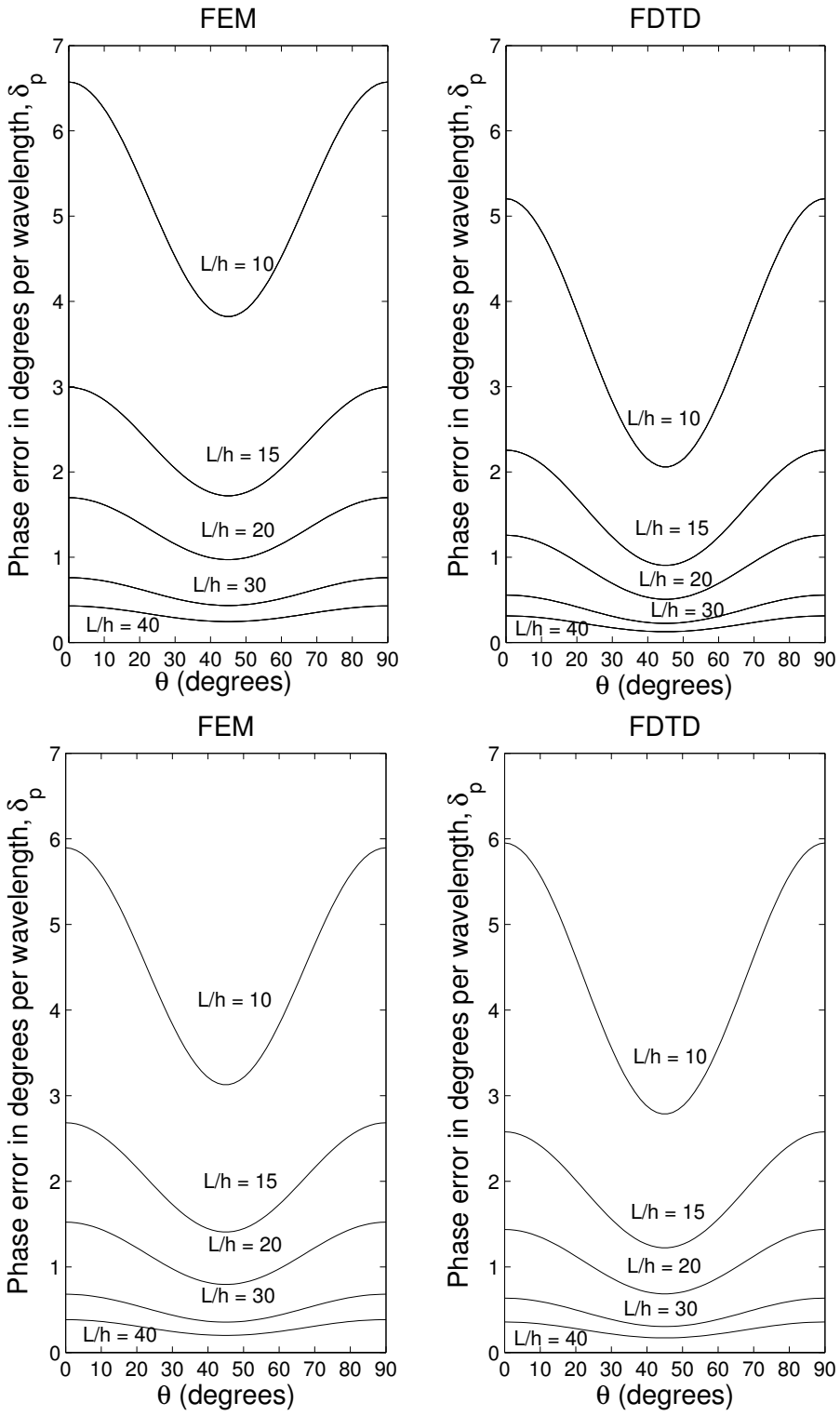

Figure 5.7: Phase error in degrees per wavelength as a function of the angle θ for selected values of L/h, with $\eta = 0.4$ (top) and $\eta = 0.2$ (bottom).

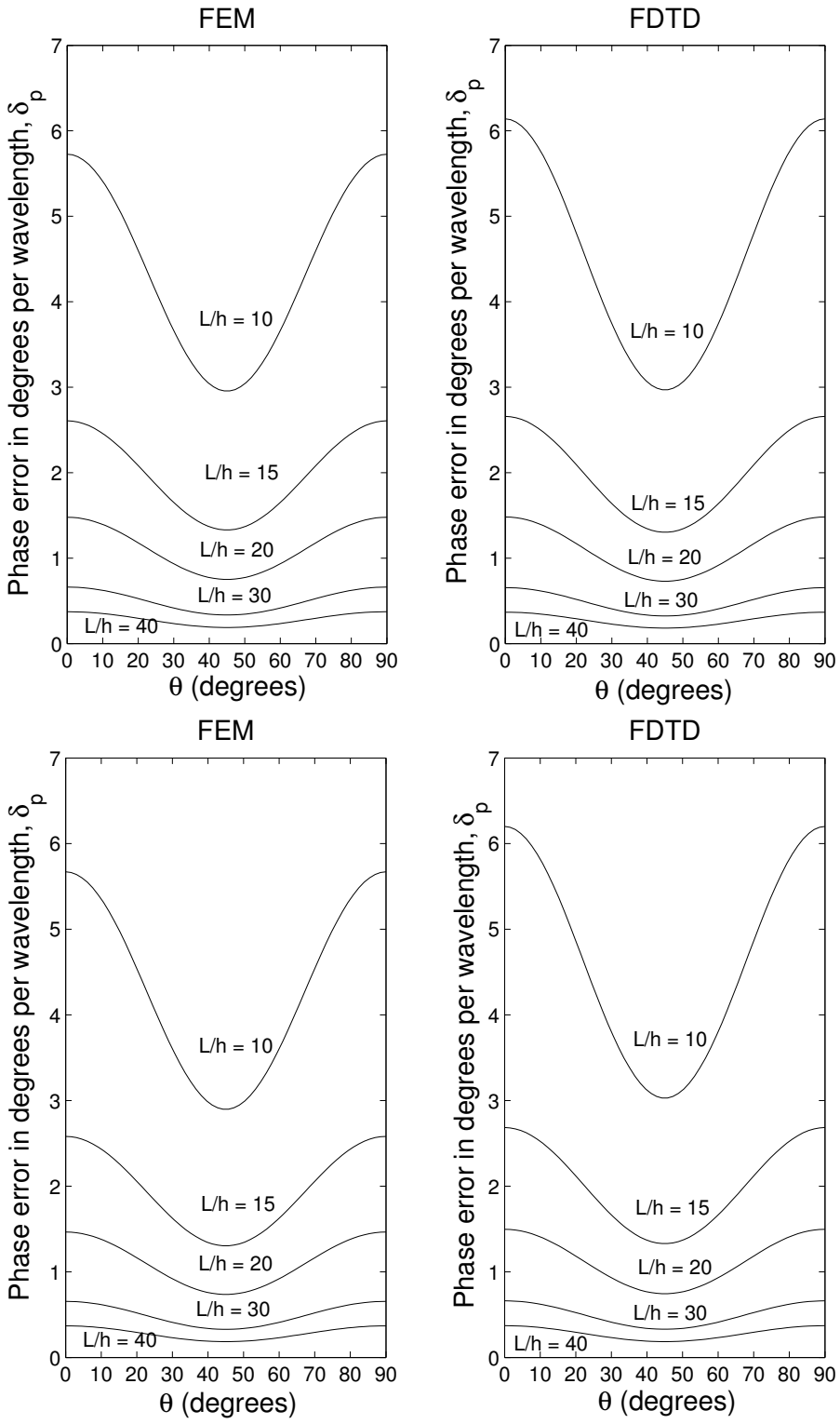

Figure 5.8: Phase error in degrees per wavelength as a function of the angle θ for selected values of L/h, with $\eta = 0.1$ (top) and $\eta = 0.01$ (bottom).

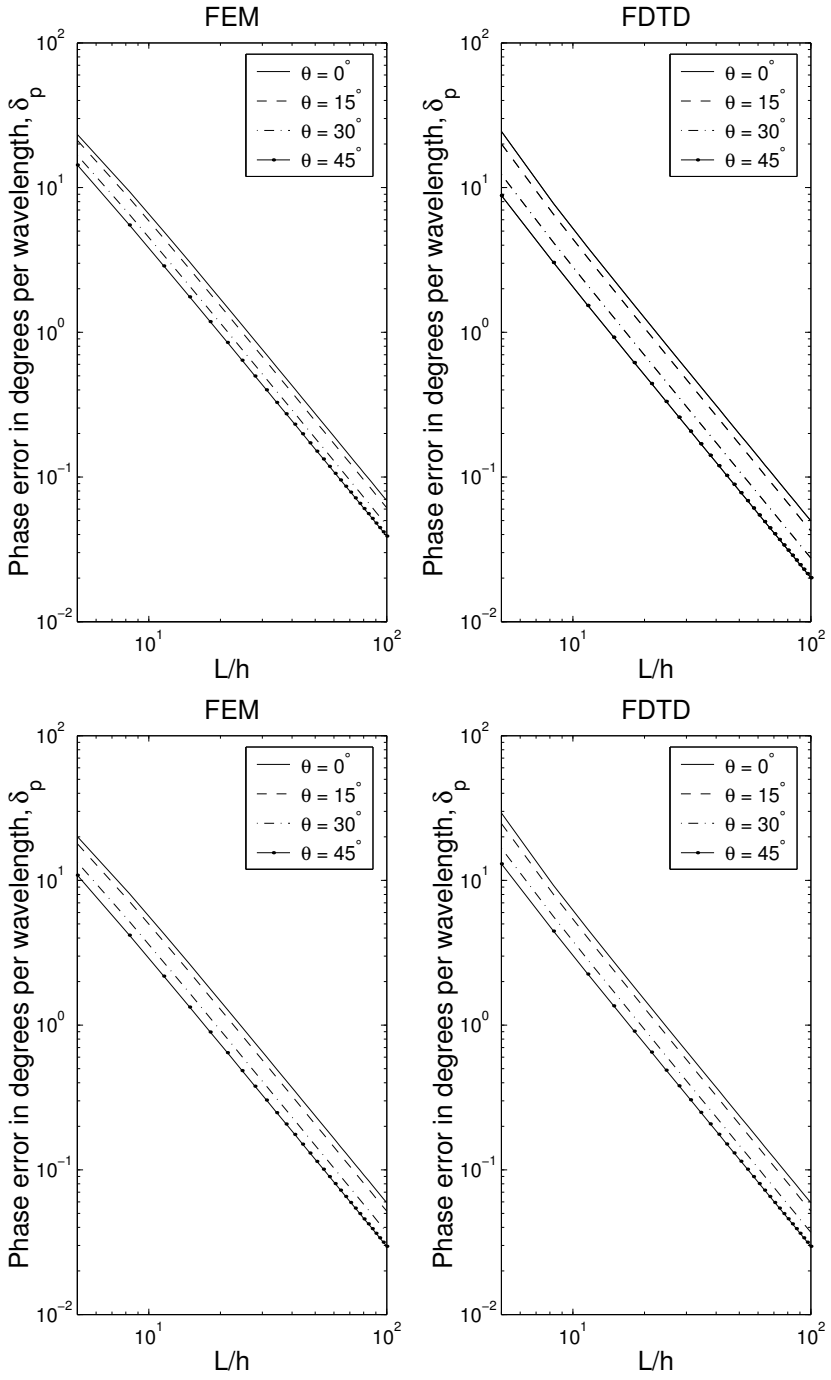

Figure 5.9: Log-log plot of the phase error in degrees per wavelength as a function of L/h for selected values of the angle of incidence θ with $\eta = 0.4$ (top) and $\eta = 0.01$ (bottom).

proportional to the square of the inverse of the number of nodes per wavelength, i.e., to $(h/L)^2$. This implies the second order convergence, with respect to (h/L), of $\tilde{k}$ to k. We also note that the phase error is lower for the FDTD scheme than for the FEM scheme when $\eta = 0.4$, but as η is decreased to 0.01, the effects of dispersion increase for the FDTD scheme.

We present here another important feature of numerical dispersion, which is the anisotropy of the dispersion with respect to the angle of incidence of the propagating wave. In ordinary wave propagation, energy propagates perpendicular to the wave front. When there is anisotropy dispersion, the angle will not be perpendicular. The anisotropy, ϑ, is defined as

$$\vartheta = \max_{0\leq\theta\leq\pi/2}\delta_p - \min_{0\leq\theta\leq\pi/2}\delta_p. \tag{5.125}$$

In Tables 5.1 and 5.2 we present the maximum and minimum values, over all angles of propagation, of the phase error δ_p, for 10, 15, 20, 30, and 40 nodes per wavelength. We also present the anisotropy, in each case, for the FEM and FDTD schemes, for $\eta = 0.4$ and $\eta = 0.01$, respectively. In Table 5.1, with $\eta = 0.4$, we see that the maximum and the minimum values of δ_p are larger in the case of the FEM scheme; however, the anisotropy is less in the case of the FEM.

L/h	FEM			FDTD		
	Max	Min	ϑ	Max	Min	ϑ
10	6.5733	3.8228	2.7505	5.2043	2.0599	3.1444
15	2.9930	1.7204	1.2726	2.2548	0.9044	1.3503
20	1.6981	0.9720	0.7261	1.2573	0.5066	0.7507
30	0.7594	0.4334	0.3260	0.5539	0.2245	0.3309
40	0.4281	0.2441	0.1841	0.3117	0.1261	0.1856

Table 5.1: Anisotropy for $\eta = 0.4$, for selected values of L/h.

In Table 5.2, with $\eta = 0.01$, we see that the maximum and minimum values of the FDTD scheme are now larger than the respective values in the FEM scheme. The anisotropy, however, is still smaller for the case of the FEM.

L/h	FEM			FDTD		
	Max	Min	ϑ	Max	Min	ϑ
10	5.6705	2.8991	2.7714	6.2007	3.0302	3.1705
15	2.5812	1.3043	1.2769	2.6850	1.3298	1.3552
20	1.4643	0.7368	0.7275	1.4970	0.7447	0.7522
30	0.6548	0.3285	0.3263	0.6612	0.3299	0.3312
40	0.3691	0.1850	0.1841	0.3711	0.1854	0.1857

Table 5.2: Anisotropy for $\eta = 0.01$, for selected values of L/h.

The dispersion relation in the PML is obtained in a similar fashion. Let us consider σ to be a constant. Assuming $\hat{E}_l = \mathrm{e}^{-\mathrm{i}k_x^{\mathrm{pml}}hl}$ in the PML, substituting in (5.114), we obtain the dispersion relation in the PML to be

$$\frac{\omega^2 h^2}{12} = \left(\frac{1}{s^2}\right) \frac{\sin^2(k_x^{\mathrm{pml}}h/2)}{1 + 2\cos^2(k_x^{\mathrm{pml}}h/2)} + \frac{\sin^2(k_y h/2)}{1 + 2\cos^2(k_y h/2)}. \tag{5.126}$$

To derive the frequency ω for the fully discrete scheme, we again replace ωh by $2\dfrac{h}{\Delta t}\sin\left(\dfrac{\omega \Delta t}{2}\right)$ to get

$$\frac{h^2}{3(\Delta t)^2}\sin^2\left(\frac{\omega_{\Delta t}\Delta t}{2}\right) = \left(\frac{1}{s^2}\right) \frac{\sin^2(k_x^{\mathrm{pml}}h/2)}{1 + 2\cos^2(k_x^{\mathrm{pml}}h/2)} + \frac{\sin^2(k_y h/2)}{1 + 2\cos^2(k_y h/2)}. \tag{5.127}$$

5.9 Calculation of the Reflection Coefficient

In this section, we study the properties of the discrete PML model by performing a plane wave analysis to calculate the reflection coefficient. As in Section 5.8, for the 2D case, let the numerical wave vector be defined by $\mathbf{k} = (k_x, k_y)$. In the discrete setting the PML

model is no longer perfectly matched, since the discretization introduces some error which manifests itself as spurious reflections. There is also error that is introduced due to the termination of the PML. In this section, we calculate the errors introduced in the discrete model by calculating the reflection coefficient of an infinite PML (to study the errors caused by the discretization) as well as the reflection coefficient of a finite PML (to study the errors introduced by terminating the PML). The calculation of the reflection coefficient follows [42], where the authors have calculated the reflection coefficient for the TE mode of the Zhao-Cangellaris's PML model using the FDTD scheme [38].

For simplicity we again assume in this section that $\epsilon_0 = \mu_0 = 1$. Let us also assume an infinite PML in the region $x > 0$. Thus, $\sigma_y = 0$ and let $\sigma_x = \sigma$. To calculate the reflection coefficient for the infinite PML, we look for solutions to (5.114) of the form:

$$\hat{E}_l = \begin{cases} \mathrm{e}^{-\mathrm{i}k_x hl} + R\,\mathrm{e}^{\mathrm{i}k_x hl}, & \text{for } l < 0, \\ T\,\mathrm{e}^{-k_x^{\mathrm{pml}} hl}, & \text{for } l > 0. \end{cases} \tag{5.128}$$

The reflection coefficient is R, and T is the transmission coefficient. Consider the equations associated to the node at the interface $l = 0$, and one node each on either side of the interface at $l = 1$, and $l = -1$. From (5.114) we have

$$\begin{cases} -\dfrac{\zeta\omega^2 h^2}{6}(4\hat{E}_{-1} + \hat{E}_{-2} + \hat{E}_0) = (\hat{E}_0 - \hat{E}_{-1}) - (\hat{E}_{-1} - \hat{E}_{-2}), \\ -\dfrac{\zeta\omega^2 h^2}{6}\left(\dfrac{s_0}{2}\right)(4\hat{E}_0 + \hat{E}_{-1} + \hat{E}_1) = \dfrac{1}{s_0}(\hat{E}_1 - \hat{E}_0) - (\hat{E}_0 - \hat{E}_{-1}), \\ -\dfrac{\zeta\omega^2 h^2}{6}\left(\dfrac{s_1 + s_0}{2}\right)(4\hat{E}_1 + \hat{E}_0 + \hat{E}_2) = \dfrac{1}{s_1}(\hat{E}_2 - \hat{E}_1) - \dfrac{1}{s_0}(\hat{E}_1 - \hat{E}_0), \end{cases} \tag{5.129}$$

where ζ is defined in (5.115), and s_l is defined in (5.109). Substituting for $\hat{E}_l$ from (5.128) in (5.129), we obtain three equations in the unknowns $\hat{E}_0$, R and T. Solving these resulting equations for R, we can show that the reflection coefficient has the Taylor series expansion

$$R = -\frac{1}{16\omega^2}(\omega^2 - k_y^2)\sigma(\sigma + 2\mathrm{i}\omega)h^2 + \frac{1}{48\omega^3}\sigma^2(\sigma + 2\mathrm{i}\omega)(\omega^2 - k_y^2)^{3/2}h^3 + O(h^4), \tag{5.130}$$

which we have calculated using the software MATHEMATICA. The formula (5.130) implies that the reflection coefficient is proportional to h^2. Thus, the discrete PML model is consistent with the continuous model.

Next, we study the effects of terminating the PML by a PEC. This amounts to setting $E = 0$ at the boundary $x = \delta = Mh$ of the PML, i.e. $E_M = 0$. To obtain the reflection coefficient for this case, we need to write down the equation (5.114) for all the nodes in the PML as well as for the node at the interface of the working volume-PML, E_0 and one node in the working volume, which is h distance away from the interface, i.e., E_{-1}. Assuming that we know the value of E_{-2} we obtain a system of equations to solve:

$$A\mathbf{E} = -(\omega^2 h^2 \zeta + 6) E_{-2} \tilde{e_1}. \tag{5.131}$$

In the above $\mathbf{E} = [E_{-1}, E_0, E_1, \cdots, E_{M-1}]^T$ and $\tilde{e_1} = [1, 0, 0, \cdots]^T$ and the matrix of coefficients obtained from (5.114) is

$$A = \begin{bmatrix} b_{-1} & c_0 & 0 & \dots & \dots \\ a_{-1} & b_0 & c_1 & 0 & \dots \\ 0 & a_0 & b_1 & c_2 & \dots \\ & & \ddots & \ddots & \ddots \\ & \dots & 0 & a_{M-2} & b_{M-1} \end{bmatrix}. \tag{5.132}$$

In the above

$$\begin{aligned} a_l &= \left(\omega^2 h^2 \left(\frac{s_{l+1} + s_l}{2} + \frac{6}{s_l}\right)\right), \\ b_l &= \left(4\omega^2 h^2 \left(\frac{s_{l-1} + s_l}{2}\right) - 6\left(\frac{1}{s_l} + \frac{1}{s_{l-1}}\right)\right), \\ c_l &= \left(\omega^2 h^2 \left(\frac{s_{l-1} + s_{l-2}}{2} + \frac{6}{s_{l-1}}\right)\right). \end{aligned} \tag{5.133}$$

We can solve the system (5.132) for the value of R by using (5.128) for $l = -1$ and $l = -2$. The absolute value of the reflection coefficient is calculated to be

$$|R| = \left| \frac{1 + (\omega^2 h^2 \zeta + 6)\kappa \ \mathrm{e}^{\mathrm{i}k_x h}}{1 + (\omega^2 h^2 \zeta + 6)\kappa \, \mathrm{e}^{-\mathrm{i}k_x h}} \right|. \tag{5.134}$$

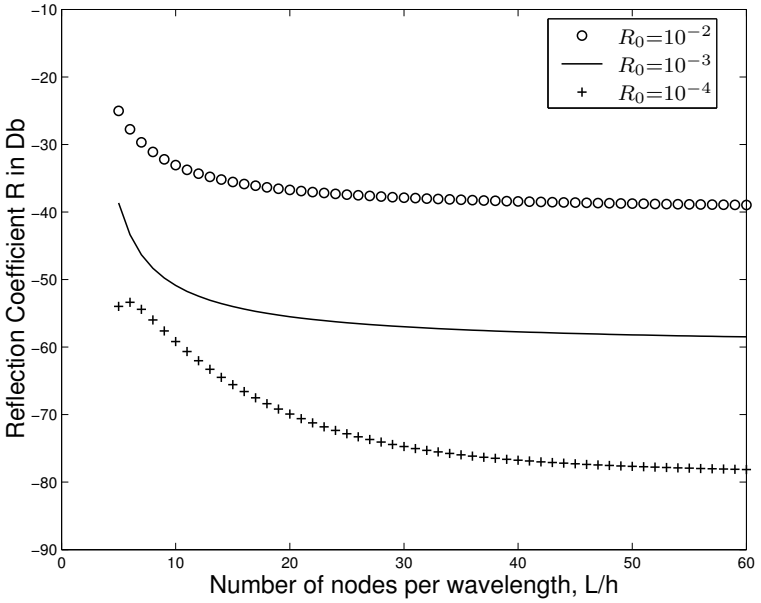

Figure 5.10: Numerical reflection coefficient at normal incidence. We note that, as we increase the number of nodes per wavelength, the numerical reflection coefficient approaches R_0.

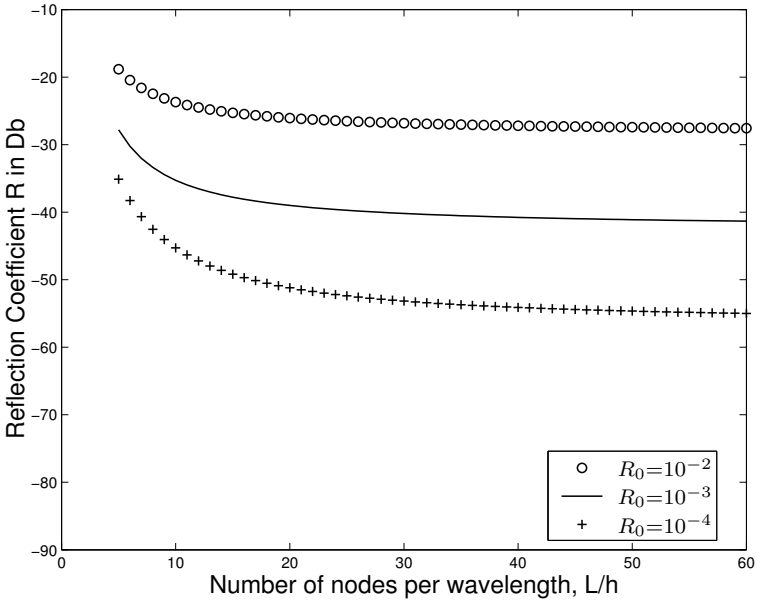

Figure 5.11: Numerical reflection coefficient for $\theta = \pi/4$. We observe more reflection in this case than in the case $\theta = 0$.

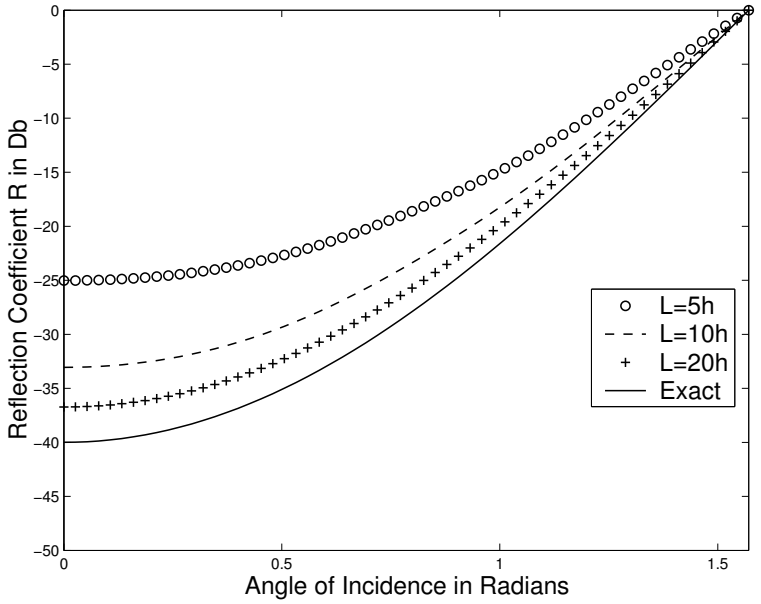

Figure 5.12: Numerical reflection coefficient for $R_0 = 10^{-2}$. As L/h is increased, the numerical reflection coefficient converges to $R_0^{\cos\theta}$.

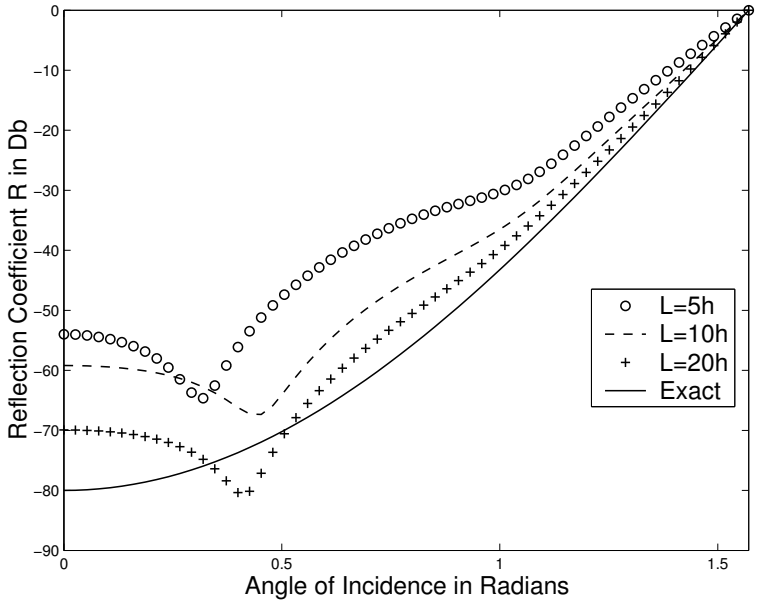

Figure 5.13: Numerical reflection coefficient for $R_0 = 10^{-4}$. As L/h is increased R converges to $R_0^{\cos\theta}$.

where κ is the first diagonal entry in A^{-1}.

Figures 5.10 and 5.11 plot the reflection coefficient in decibels, Db (i.e., $20\log_{10} R$), versus the number of nodes per wavelength L/h, for different values of R_0, where recall from Section 5.3 that R_0 is the reflection coefficient at normal incidence. Figures 5.12 and 5.13 plot the reflection coefficient in Db versus the angle of incidence θ.

We note that the numerical reflection coefficient converges to the reflection coefficient of the continuous model, which is $R_0^{\cos\theta}$, as we increase the value of N. We also note that as θ approaches the value π, the numerical reflection coefficient approaches the value 1. This is a well known behaviour of PML models, i.e., waves that are propagating transversely to the interface between the domain of interest and a single PML, are not absorbed by the PML. However, these waves get absorbed into the corner regions where two PML's overlap. The plots were obtained by considering PML's that are one wavelength thick, i.e., the number of nodes per wavelength is the number of nodes in the PML. The polynomial grading (5.19) was chosen for σ with $m = 2$ and $\sigma_{\max}$ as in (5.20) with $Z = 1$. The PML is in the region $x > 0$ as mentioned before, thus $x_0 = 0$ in (5.19) for $\alpha = x$.

5.10 Absorption of a Pulse on the Boundaries of a Computational Domain

The numerical experiment described in this section evaluates the performance of the UPML when a pulse strikes the boundaries of a computational domain. We measure the amount of reflection that an outward propagating pulse produces as it moves from free space to a boundary surrounded by absorbing PML's as in Figure 5.1.

We choose our domain Ω to be the square $[0, 12] \times [0, 12]$, with a source located at the center $(6, 6)$ of the square [36]. The domain is surrounded by absorbing layers on all four boundaries. We discretize the problem with a rectangular grid composed of 90×90 square elements of step size $h = 2/15$ and the time step is $\Delta t = 0.04/c$, which satisfies the CFL

condition (5.99). The source is taken to be the function

$$f(x, y, t) = f_1(x, y) f_2(t),$$

where

$$f_2(t) = \begin{cases} -2\pi^2 f_0^2 (t - t_0)\, \mathrm{e}^{-\pi^2 f_0^2 (t-t_0)^2}, & \text{if } t \leq 2t_0, \\ 0, & \text{if } t \geq 2t_0. \end{cases} \tag{5.135}$$

In the above $f_0 = \dfrac{c}{20h}$, is the central frequency and $t_0 = 1/f_0$. The function $f_1(x, y)$ is defined as

$$f_1(x, y) = \mathrm{e}^{-7\sqrt{(x-6)^2+(y-6)^2}}. \tag{5.136}$$

We obtain a reference solution by using a similar finite element scheme for the TM mode of Maxwell's equations on a larger domain Ω_R containing 360×360 square elements, and the same mesh step size and time step. Ω_R is terminated using PEC conditions on its boundary. We have used the polynomial grading (5.19) for σ with the optimal value of $\sigma_{\max}$ as given in (5.21) with $m = 3.5$.

The L^2 norm of the error due to numerical reflections, which arise due to the finite PML terminated by PEC conditions, is obtained by subtracting at each time step the field E at any grid point inside Ω, from the field E at the corresponding point in Ω_R, taking the square of this difference and summing such differences over all grid points in Ω. We do the above for three PML's containing 4, 8 and 16 cells. A comparison is presented with respect to the split field PML (SF) of Berenger, using the same test problem. The reference solution for the split-field case is chosen similarly. Figure 5.14 shows the L^2 error between the two reference solutions (Reference Error) for the split-field and the finite element scheme, and the L^2 error of the two schemes for 4, 8 and 16 cell PML's. From Figure 5.14, we can see that the reference error (discretization error) dominates for about 250 time steps. After this, as the wave exits the computational domain, the reflection error due to the PEC backed PML takes over. We have used 20 nodes/wavelength (i.e., $L/h = 20$) in our calculations. As can be seen for a 16 cell PML the reflection error is lower than the reference error. The

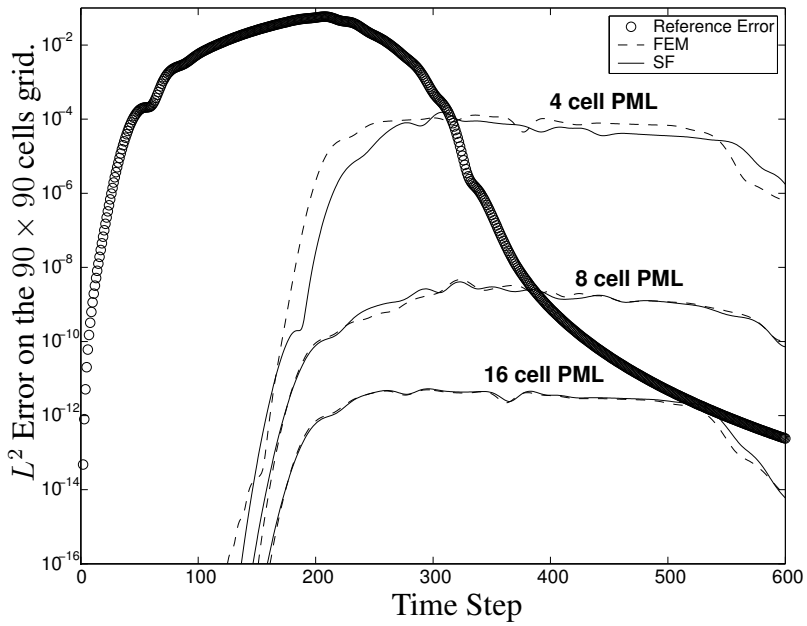

Figure 5.14: Comparison of the L^2 error for the UPML with a mixed finite element scheme and the split field PML with the FDTD scheme on a 90×90 cells grid.

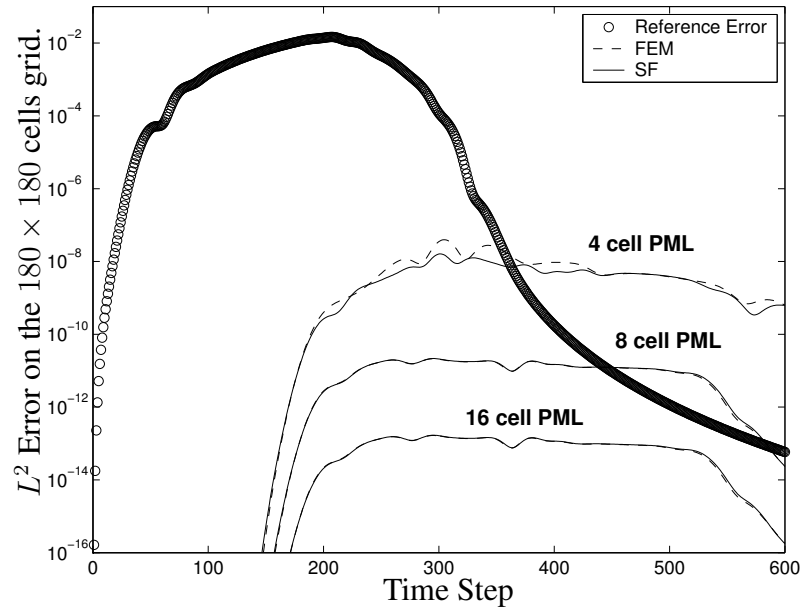

Figure 5.15: Comparison of the L^2 error for the UPML with the mixed finite element scheme and the split field PML with the FDTD scheme for a 180×180 cells grid.

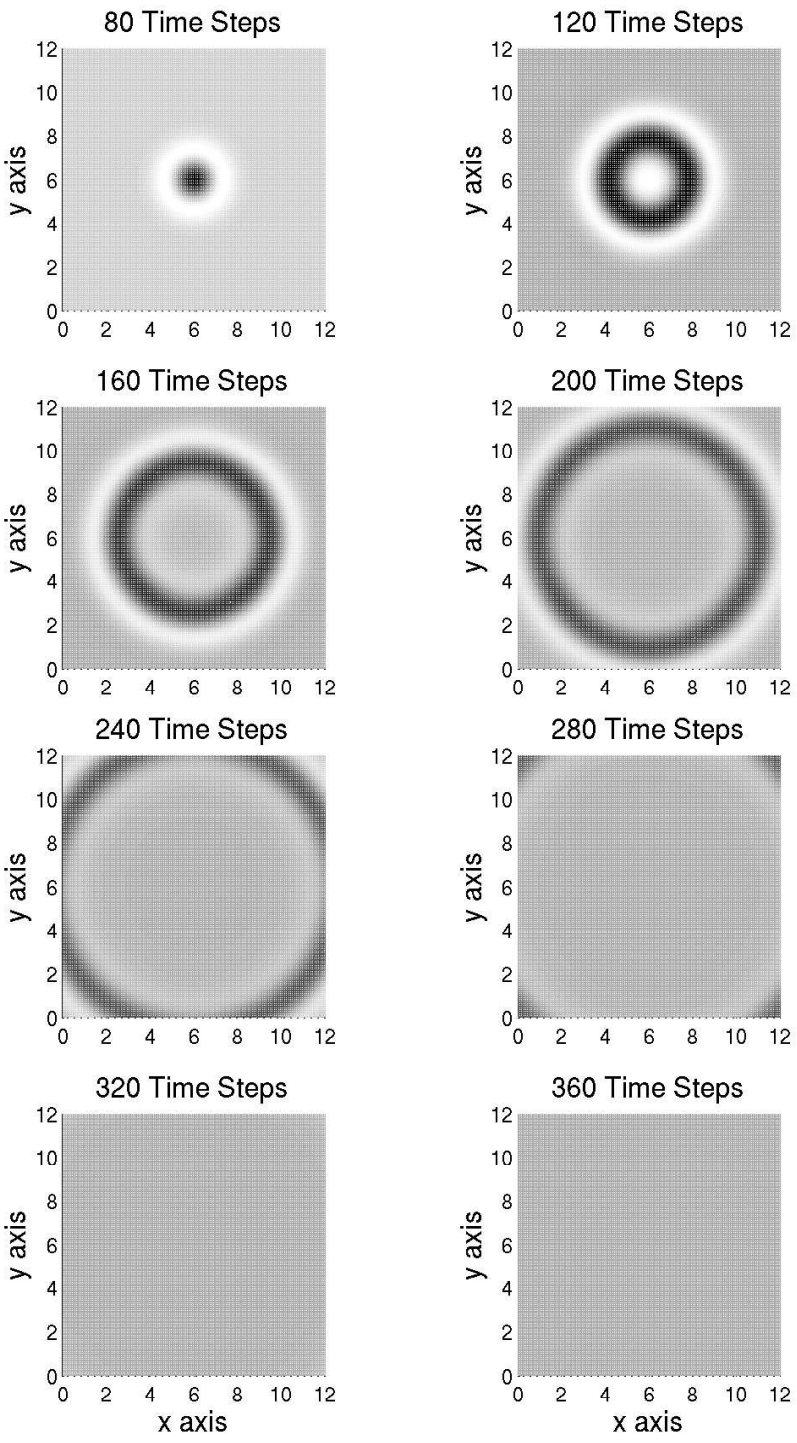

Figure 5.16: Propagation of the wave front for different time steps.

best results were obtained when m in (5.21) was chosen to be between 3 and 4. Figure 5.15 shows the L^2 error between the two reference solutions (reference error) for the split-field and the finite element scheme, and the L^2 error of the two schemes for 4, 8 and 16 cell PML's, for a refined discretization. In this case $h = 1/15$ and $\Delta t = 0.02/c$. From Figure 5.15 we can see that a four cell PML provides a good absorbing layer.

In Figure 5.16 we see the propagation of a pulse on a 180×180 cells domain backed by an eight cell PML. The wave front completely disappears from the domain, as seen in the subplot corresponding to 320 time steps. All subplots are plotted at the same magnitude. At lower magnitudes we can see the numerical reflections that enter the computational domain.

Chapter 6

A Fictitious Domain Formulation for the 2D TM Mode of Maxwell's Equations

6.1 Introduction

Electromagnetic phenomena play an important role in modern technology in different areas such as advanced mobile information systems, the design, development, integration, and testing of antennas, communication signal processing and many more. Applications involve the propagation and scattering of transient electromagnetic signals such as in aircraft radar signature analysis or the nondestructive testing of concrete structures. The study of such applications requires the ability to predict different kinds of electromagnetic effects. Some of the important effects include the radar scattering attributes i.e., radar cross section (RCS) of different complex objects such as airplanes and missiles, the propagation of pulses through dispersive media such as soil or concrete to detect pollutants or hidden targets, interaction of electromagnetic waves with biological media, the interaction of antenna elements with aircrafts and ships, and many more [108].

The complete set of laws for time-varying electromagnetic phenomena can be derived from physical concepts such as electric charges and current density, some universal laws,

such as the conservation of electric charge, Faraday's and Ampere's laws, and constituent laws which are characteristic for a given medium [51]. These laws of electromagnetism are represented by *Maxwell's equations* and are central to predictions such as those described in the paragraph above. Thus, the study of computational techniques for solving wave scattering problems which involve large complex bodies, and the analysis of wave propagation through inhomogeneous media is being widely carried out by many researchers. A problem of interest is specified by stating appropriate boundary conditions that describe the physical situation in question. For example, one specifies the positions of conductors on an integrated circuit, the shape of the case of a cellular telephone, the configuration of an antenna etc. as boundary conditions.

There are many different techniques available for solving the time-dependent problem of scattering by an obstacle. Many such techniques, their advantages and disadvantages, have been outlined in the introductory chapter. In Chapter 3 we introduced a fictitious domain method, based on a distributed Lagrange multiplier, for the solution of the two-dimensional scalar wave equation with a Dirichlet condition on the boundary of an obstacle. In this chapter we will discuss how a similar formulation can be applied to the case of the two-dimensional TM mode of Maxwell's equations. The fictitious domain method has been recently applied to the case of the time dependent Maxwell's equations [40, 41, 44], and to the time harmonic case. In all these cases, a boundary Lagrange multiplier has been used to enforce the Dirichlet condition on the boundary of the obstacle. In Chapter 4 we compared the distributed multiplier to the boundary multiplier in the one-dimensional case and observed some advantages of the distributed multiplier formulation.

Using the uniaxial formulation of the perfectly matched layer, presented in Chapter 5, we consider a time-dependent scattering problem by employing the fictitious domain method of Chapter 3. We will also consider a first order absorbing boundary condition for Maxwell's equations, namely, the *Silver-Müller* absorbing boundary condition in order to provide a comparison with the performance of the PML as well as to gain an understanding

of the errors present in the fictitious domain method.

An outline of the chapter is as follows. In Section 6.2 we present Maxwell's equations in free space and discuss their relation to the wave equation. In Section 6.3 we consider appropriate boundary conditions that model the time dependent problem of scattering by an obstacle, in the case of the wave equation as well as for Maxwell's equations. In Section 6.4 we present two different ways of modeling the scattering problem and consider absorbing boundary conditions for these models in section 6.5. We consider two-dimensional versions of these models in Section 6.6. In Section 6.7 we present a fictitious domain method for the two-dimensional TM mode of Maxwell's equations with a first order absorbing boundary condition. In Section 6.8 we replace the first order absorbing boundary condition by a perfectly matched layer. Numerical experiments are presented to validate our methods in Section 6.9.

6.2 Maxwell's Equations and the Wave Equation

The analysis of electromagnetic problems requires the numerical solution of the *linear* time-domain Maxwell equations. As seen in Chapter 5, these are vector differential equations that describe the evolution in time and space of the electric field $\mathbf{E}$, and the magnetic field $\mathbf{H}$, and can be stated in the most general form as :

$$\left\{ \begin{array}{lrcl} \text{(i)} & \dfrac{\partial \mathbf{B}}{\partial t} & = & -\nabla\times\mathbf{E} \;-\; \mathbf{J_M}, \quad \text{(Maxwell-Faraday's Law)}, \\ \text{(ii)} & \dfrac{\partial \mathbf{D}}{\partial t} & = & \nabla\times\mathbf{H} \;-\; \mathbf{J_E}, \quad \text{(Maxwell-Ampere's Law)}. \end{array} \right. \tag{6.1}$$

The fields $\mathbf{D}$ and $\mathbf{B}$ are called the electric displacement (flux, induction), and the magnetic induction (flux), respectively. $\mathbf{J_E}$ and $\mathbf{J_M}$ are impressed electric and magnetic current densities, respectively. Gauss's laws state conditions on the divergence of $\mathbf{D}$ and $\mathbf{B}$, i.e.,

$$\left\{ \begin{array}{lrcl} \text{(i)} & \nabla\cdot\mathbf{B} & = & 0, \quad \text{(Gauss's Law for the magnetic field)}, \\ \text{(ii)} & \nabla\cdot\mathbf{D} & = & \rho, \quad \text{(Gauss's Law for the electric field)}. \end{array} \right. \tag{6.2}$$

In the above, ρ is the charge density. From (6.1, ii) and (6.2, ii), we obtain the equation for the conservation of charge

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J_E} = 0. \tag{6.3}$$

In general, the system (6.1) is closed with constitutive relations

$$\begin{cases} \mathbf{B} &= \mathcal{G}(\mathbf{E}, \mathbf{H}), \\ \mathbf{D} &= \mathcal{F}(\mathbf{E}, \mathbf{H}). \end{cases} \tag{6.4}$$

In the case of linear (field-independent), isotropic (direction-independent) and non dispersive (frequency independent) materials, the constitutive relations (6.4) become

$$\begin{cases} \mathbf{B} &= \mu\, \mathbf{H}, \\ \mathbf{D} &= \epsilon\, \mathbf{E}, \end{cases} \tag{6.5}$$

where, μ and ϵ are called the (magnetic) permeability and the (electric) permittivity, respectively, and are positive constants characteristic of the medium being considered. Such a medium is called a *perfect medium* [46]. For free space, these quantities have the values

$$\begin{cases} \mu_0 &= 4\pi 10^{-7}\ H/m \ \text{ (Henry's per meter)}, \\ \epsilon_0 &= \dfrac{1}{36\pi 10^{9}}\ F/m \ \text{ (Farad's per meter)}. \end{cases} \tag{6.6}$$

Using the constitutive relations (6.5), Maxwell's equations are given by the system

$$\begin{cases} \mu \dfrac{\partial \mathbf{H}}{\partial t} &= -\nabla \times \mathbf{E} \ - \ \mathbf{J_M}, \ \ \text{(Maxwell-Faraday's Law)}, \\ \epsilon \dfrac{\partial \mathbf{E}}{\partial t} &= \nabla \times \mathbf{H} \ - \ \mathbf{J_E}, \ \ \text{(Maxwell-Ampere's Law)}. \end{cases} \tag{6.7}$$

In addition, if we allow the possibility of electric and magnetic losses that can dissipate electromagnetic fields in materials via conversion to heat energy, we can define an equivalent magnetic current to account for the magnetic loss, i.e.,

$$\mathbf{J_M} = \sigma_M \mathbf{H}, \tag{6.8}$$

and an equivalent electric current to account for the electric loss, i.e.,

$$\mathbf{J_E} = \sigma_E \mathbf{E}. \tag{6.9}$$

In (6.8), σ_M is called an equivalent magnetic resistivity and in (6.9), σ_E is called the electric conductivity. From (6.7), (6.8) and (6.9), we can rewrite Maxwell's equations along with the Gauss divergence laws as

$$\begin{cases} \text{(i)} & \mu\dfrac{\partial \mathbf{H}}{\partial t} &=& -\nabla\times\mathbf{E} &-& \sigma_M\mathbf{H}, & \text{(Maxwell-Faraday's Law)}, \\ \text{(ii)} & \epsilon\dfrac{\partial \mathbf{E}}{\partial t} &=& \nabla\times\mathbf{H} &-& \sigma_E\mathbf{E}, & \text{(Maxwell-Ampere's Law)}, \\ \text{(iii)} & \nabla\cdot\mu\mathbf{H} &=& 0, & & & \text{(Gauss's Law for } \mathbf{H}), \\ \text{(iv)} & \nabla\cdot\epsilon\mathbf{E} &=& \rho, & & & \text{(Gauss's Law for } \mathbf{E}). \end{cases} \tag{6.10}$$

We note that the two divergence equations (6.10, iii) and (6.10, iv) are redundant. This is the case as equation (6.10, iv) can be considered to be the definition of the charge density ρ, whereas equation (6.10, iii) can be deduced from (6.10, i) if we assume that the condition

$$\nabla\cdot\mu\mathbf{H}|_{t=t_0} = 0, \tag{6.11}$$

is satisfied for some $t_0 \geq 0$. If we consider wave propagation in an isotropic homogeneous medium such as vacuum, and the source of waves is far enough so that $\rho = 0$, $\sigma_E = 0$ and also $\sigma_M = 0$, then the equations (6.10) reduce to

$$\begin{cases} \text{(i)} & \mu\dfrac{\partial \mathbf{H}}{\partial t} &=& -\nabla\times\mathbf{E}, \\ \text{(ii)} & \epsilon\dfrac{\partial \mathbf{E}}{\partial t} &=& \nabla\times\mathbf{H}. \end{cases} \tag{6.12}$$

with both the electric and magnetic fields being divergence free. In this case, the equations (6.12) can be decoupled into separate equations for $\mathbf{E}$ and $\mathbf{H}$, given by

$$\begin{cases} \dfrac{\partial^2\mathbf{H}}{\partial t^2} &+& \dfrac{1}{\epsilon\mu}\nabla\times\nabla\times\mathbf{H} &=& 0, \\ \dfrac{\partial^2\mathbf{E}}{\partial t^2} &+& \dfrac{1}{\epsilon\mu}\nabla\times\nabla\times\mathbf{E} &=& 0. \end{cases} \tag{6.13}$$

Taking into account the relation

$$\nabla\times\nabla\times\mathbf{V} = \nabla(\nabla\cdot\mathbf{V}) - \Delta\mathbf{V}, \tag{6.14}$$

as well as the divergence laws (6.10, iii) and (6.10, iv) (with $\rho = 0$), we have

$$\begin{cases} \dfrac{\partial^2 \mathbf{H}}{\partial t^2} - c^2 \Delta \mathbf{H} = 0, \\ \dfrac{\partial^2 \mathbf{E}}{\partial t^2} - c^2 \Delta \mathbf{E} = 0. \end{cases} \tag{6.15}$$

Thus, both the electric and magnetic field vectors satisfy the vector wave equation with velocity

$$c = \frac{1}{\sqrt{\epsilon \mu}}. \tag{6.16}$$

6.3 Boundary Conditions

In this thesis, we are interested in boundary conditions that are related to the time-dependent problem of scattering by an obstacle. Let $d = 2$ or 3.

- For the wave equation, as we have seen before, the appropriate boundary condition is the Dirichlet condition

$$\Phi = G(\mathbf{x}, t), \text{ on } \partial\omega \times (0, T), \ \omega \subset \mathbb{R}^d, \tag{6.17}$$

  for $T > 0$, where Φ satisfies the scalar wave equation

$$\frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} - \Delta \Phi = 0, \text{ in } \mathbb{R}^d / \bar{\omega} \times (0, T). \tag{6.18}$$

  When $G = 0$, we obtain a perfectly reflecting boundary condition.

- For Maxwell's equations, the appropriate condition is called a perfect conductor condition (PEC). A perfect conductor is a medium in which the electric field is zero. It is not possible to have either charge or current in the interior of such a medium. However, there can exist densities of charge ρ_Σ and of current $\mathbf{J}_\Sigma$ on the surface Σ of the perfect conductor. The boundary conditions on the surface of this medium are called perfect conductor conditions, and can be stated as

$$\mathbf{E} \times \mathbf{n} = 0, \text{ on } \partial\omega \times (0, T), \ \omega \subset \mathbb{R}^d, \tag{6.19}$$

where, $\mathbf{n}$ is the outward unit normal to $\partial\omega$. This is a reflecting boundary condition, and implies that the tangential component of the electric field $\mathbf{E}$ is zero on $\partial\omega$.

6.4 The Scattering Problem

As noted before, the divergence equations are redundant and we concentrate on the curl equations. The problem that is of interest in this thesis is the scattering by an obstacle ω with the boundary $\partial\omega$. We are interested in calculating the field in the exterior of ω. There are two different setup's that model this problem:

- Modeled by the wave equation:

$$\begin{cases} \dfrac{1}{c^2}\dfrac{\partial^2\Phi}{\partial t^2} - \Delta\Phi = 0, \text{ in } \mathbb{R}^d/\,\bar{\omega} \times (0,T), \\ \Phi = G(x,t), \quad \text{in } \partial\omega \times (0,T), \\ \Phi(\mathbf{x}, t=0) = \Phi_0(\mathbf{x}), \text{ and } \Phi_t(\mathbf{x}, t=0) = \Phi_1(\mathbf{x}), \quad \text{in } \mathbb{R}^d/\bar{\omega}. \end{cases} \tag{6.20}$$

  This is the problem that we have considered in Chapters 3 and 4.

- Modeled by Maxwell's equations in free space:

$$\begin{cases} \mu_0\dfrac{\partial \mathbf{H}}{\partial t} + \nabla\times\mathbf{E} = 0, \text{ in } \mathbb{R}^d/\bar{\omega} \times (0,T), \\ \epsilon_0\dfrac{\partial \mathbf{E}}{\partial t} - \nabla\times\mathbf{H} = 0, \text{ in } \mathbb{R}^d/\bar{\omega} \times (0,T), \\ \mathbf{E}\times\mathbf{n} = 0, \text{ in } \partial\omega \times (0,T), \\ \mathbf{E}(\mathbf{x}, t=0) = \mathbf{E}_0(\mathbf{x}), \text{ and } \mathbf{H}(\mathbf{x}, t=0) = \mathbf{H}_0(\mathbf{x}), \text{ in } \mathbb{R}^d/\bar{\omega}. \end{cases} \tag{6.21}$$

  We will consider this scattering problem in the remainder of this chapter.

6.5 Absorbing Boundary Conditions

Since numerical simulations of wave propagation have to be performed on a finite numerical domain, we have to truncate the domain $\mathbb{R}^d/\bar{\omega}$ at some artificial boundary Γ. Absorbing

boundary conditions (or layers) must be imposed on Γ to obtain a well posed problem and to simulate the outgoing nature of the waves. In Chapter 3, we considered a first order absorbing boundary condition for the two-dimensional scalar wave equation which is given by

$$\frac{1}{c}\frac{\partial \Phi}{\partial t} + \frac{\partial \Phi}{\partial \mathbf{n}} = 0, \quad \text{on } \Gamma \times (0, T). \tag{6.22}$$

The condition (6.22) is known as an approximate *Sommerfeld* radiation condition.

In Chapter 5 we constructed an absorbing layer called a perfectly matched layer for the two-dimensional Maxwell's equations. In addition to the PML model we will also consider a first order absorbing boundary condition for Maxwell's equations known as the *Silver-Müller* condition, which can be imposed in two different ways as

$$\mathbf{H} \times \mathbf{n} = \sqrt{\frac{\epsilon_0}{\mu_0}}\mathbf{n} \times (\mathbf{E} \times \mathbf{n}), \quad \text{on } \Gamma \times (0, T), \tag{6.23}$$

or as

$$\mathbf{E} \times \mathbf{n} = \sqrt{\frac{\mu_0}{\epsilon_0}}\mathbf{n} \times (\mathbf{H} \times \mathbf{n}), \quad \text{on } \Gamma \times (0, T). \tag{6.24}$$

In the equation (6.23), the term $\mathbf{n} \times (\mathbf{E} \times \mathbf{n})$ is the tangential electric field and $\mathbf{n}$ is the unit outward normal [104]. Similarly, in (6.24) the term $\mathbf{n} \times (\mathbf{H} \times \mathbf{n})$ is the tangential magnetic field.

Let Ω be a bounded domain in $\mathbb{R}^d$, with boundary Γ, enclosing the obstacle ω. The Silver-Müller boundary conditions on Γ model the electromagnetic interactions between the domain Ω and the exterior. They approximate the boundary Γ by its tangent plane. The outgoing electromagnetic plane waves which propagate normally to the boundary Γ of the domain Ω can leave freely without being reflected at the boundary. They are absorbed at the boundary.

The Silver-Müller condition on $\Gamma \times (0, T)$ is equivalent to the Sommerfeld radiation field condition for the Cartesian field components. Applying the Silver-Müller conditions at a finite distance from the scatterer results in an approximate absorbing boundary condition which is exact for outgoing spherical waves [104].

The scattering problem posed in the domain Ω can be stated as:

$$\begin{cases} \text{(i)} & \mu_0 \dfrac{\partial \mathbf{H}}{\partial t} + \nabla \times \mathbf{E} = 0, \quad \text{in } \Omega/\bar{\omega} \times (0,T), \\ \text{(ii)} & \epsilon_0 \dfrac{\partial \mathbf{E}}{\partial t} - \nabla \times \mathbf{H} = 0, \quad \text{in } \Omega/\bar{\omega} \times (0,T), \\ \text{(iii)} & \mathbf{E} \times \mathbf{n} = 0, \quad \text{in } \partial\omega \times (0,T), \\ \text{(iv)} & \mathbf{H} \times \mathbf{n} = \sqrt{\dfrac{\epsilon_0}{\mu_0}} \mathbf{n} \times (\mathbf{E} \times \mathbf{n}), \quad \text{on } \Gamma \times (0,T), \\ \text{(v)} & \mathbf{E}(\mathbf{x}, t=0) = \mathbf{E}_0(\mathbf{x}), \text{ and } \mathbf{H}(\mathbf{x}, t=0) = \mathbf{H}_0(\mathbf{x}), \quad \text{in } \Omega/\bar{\omega}. \end{cases} \tag{6.25}$$

6.6 The TM Mode for Maxwell's Equations in Two Dimensions

As done in Chapter 5, let us assume that neither the electromagnetic field excitation nor the modeled geometry has any variation in the z-direction. That is, we assume that all partial derivatives of the fields with respect to z equal zero, and the structure being modeled extends to infinity in the z direction with no change in the shape or position of its transverse cross section. In this case the six curl equations can be decoupled into two sets of equations each involving three electromagnetic field vectors. Let $\mathbf{E} = (E_x, E_y, E_z)$ and $\mathbf{H} = (H_x, H_y, H_z)$ be the components of the electric and magnetic field vectors, respectively, in a Cartesian coordinate system. In the TE *mode* the electromagnetic field has three components H_z, E_x and E_y. In the TM *mode* the electromagnetic field has the three components E_z, H_x and H_y.

We will consider the TM mode of Maxwell's equations in 2D. Let Ω now be a bounded domain of $\mathbb{R}^2$. Let $\mathbf{H} = (H_x, H_y)$ and let $E = E_z$. Let $\mathbf{n} = (n_x, n_y)$ be the unit normal vector, and let us define the unit vector $\mathbf{t}$ pointing in the tangential direction $\mathbf{t} = (n_y, -n_x)$.

Then the system (6.25) becomes:

$$\begin{cases} \text{(i)} & \mu_0 \dfrac{\partial \mathbf{H}}{\partial t} + \overrightarrow{\text{curl}}\, E = 0, \quad \text{in } \Omega/\bar{\omega} \times (0,T), \\ \text{(ii)} & \epsilon_0 \dfrac{\partial E}{\partial t} - \text{curl}\, \mathbf{H} = 0, \quad \text{in } \Omega/\bar{\omega} \times (0,T), \\ \text{(iii)} & E = 0, \quad \text{in } \partial\omega \times (0,T), \\ \text{(iv)} & \mathbf{H} \cdot \mathbf{t} = \sqrt{\dfrac{\epsilon_0}{\mu_0}} E, \quad \text{on } \Gamma \times (0,T), \\ \text{(v)} & E(\mathbf{x}, t=0) = E_0(\mathbf{x}), \text{ and } \mathbf{H}(\mathbf{x}, t=0) = \mathbf{H}_0(\mathbf{x}), \text{ in } \Omega/\bar{\omega} \end{cases} \tag{6.26}$$

We note that (6.26, iv) is the Silver-Müller condition (6.23) for the TM mode. The cross product $\mathbf{H} \times \mathbf{n}$ can be written as $\mathbf{H} \cdot \mathbf{t}\hat{z}$, where $\hat{z}$ is a unit vector in the z direction. The operators $\overrightarrow{\text{curl}}$ and curl are linear differential operators that were defined in Chapter 5, Section 5.5 in (5.28), (5.29) and (5.30).

Remark 14 *The* TE *and* TM *modes are decoupled since they do not contain any common field vector components. These two modes are completely independent for structures that are composed of isotropic materials or anisotropic materials in which the off diagonal components in the constitutive tensors are absent. [124].*

6.7 A Fictitious Domain Method

We employ the fictitious domain method introduced in Chapter 3 to enforce the Dirichlet boundary condition (6.26, iii) on the boundary $\partial\omega$ of the obstacle ω. The Silver-Müller boundary condition is naturally incorporated into the weak formulation that we construct by integrating the equations (6.26, i, ii) over the domain Ω and by using Green's formula, or equivalently integration by parts, in (6.26, ii). Thus, the Silver-Müller boundary condition does not have to be enforced in the functional spaces. From (6.26) we obtain the problem :

Find $\{\tilde{E}(\cdot,t), \tilde{\mathbf{H}}(\cdot,t), \lambda(\cdot,t)\} \in H^1(\Omega) \times [L^2(\Omega)]^2 \times L^2(\omega)$ such that:

$$\begin{cases} \text{(i)} & \mu_0 \dfrac{d}{dt}\displaystyle\int_\Omega \tilde{\mathbf{H}} \cdot \boldsymbol{\Psi}\, \mathrm{dx} + \int_\Omega \overrightarrow{\mathrm{curl}}\, \tilde{E} \cdot \boldsymbol{\Psi}\, \mathrm{dx} = 0, \quad \forall\, \boldsymbol{\Psi} \in [L^2(\Omega)]^2, \\ \text{(ii)} & \epsilon_0 \dfrac{d}{dt}\displaystyle\int_\Omega \tilde{E}\phi\, \mathrm{dx} - \int_\Omega \tilde{\mathbf{H}} \cdot \overrightarrow{\mathrm{curl}}\, \phi\, \mathrm{dx} + \sqrt{\dfrac{\epsilon_0}{\mu_0}} \displaystyle\int_\Gamma \tilde{E}\phi\, \mathrm{d}\Gamma, \\ & + \displaystyle\int_\omega \lambda\phi\, \mathrm{d}\omega = 0, \quad \forall\, \phi \in H^1(\Omega), \\ \text{(iii)} & \displaystyle\int_\omega \tilde{E}\tau\, \mathrm{d}\omega = 0, \quad \forall\, \tau \in L^2(\omega), \\ \text{(iv)} & \tilde{E}(\mathbf{x}, t=0) = \tilde{E}_0(\mathbf{x}), \text{ and } \tilde{\mathbf{H}}(\mathbf{x}, t=0) = \tilde{\mathbf{H}}_0(\mathbf{x}), \text{ in } \Omega, \end{cases} \tag{6.27}$$

in the sense that

$$\tilde{E} = \begin{cases} E & \text{on} \quad \Omega \setminus \bar{\omega}, \\ 0 & \text{on} \quad \partial\omega. \end{cases} ; \; \tilde{\mathbf{H}} = \begin{cases} \mathbf{H} & \text{on} \quad \Omega \setminus \bar{\omega}, \\ 0 & \text{on} \quad \partial\omega. \end{cases} \tag{6.28}$$

The function $\tilde{E}_0$ is chosen to be a H^1 - extension of E_0, and $\tilde{\mathbf{H}}_0$ to be at least an L^2-extension of $\mathbf{H}_0$. Thus, we have

$$\tilde{E}(\mathbf{x}, t=0) = \begin{cases} E_0(\mathbf{x}) & \text{on} \quad \Omega \setminus \bar{\omega}, \\ 0 & \text{on} \quad \omega. \end{cases}, \quad \tilde{\mathbf{H}}(\mathbf{x}, t=0) = \begin{cases} \mathbf{H}_0(\mathbf{x}) & \text{on} \quad \Omega \setminus \bar{\omega}, \\ 0 & \text{on} \quad \omega. \end{cases} \tag{6.29}$$

In succeeding sections we will, however, drop the $\tilde{}$ symbol on the fields E and $\mathbf{H}$. Thus, the system (6.27) will read:

Find $\{E(\cdot,t), \mathbf{H}(\cdot,t), \lambda(\cdot,t)\} \in H^1(\Omega) \times [L^2(\Omega)]^2 \times L^2(\omega)$ such that:

$$\begin{cases} \text{(i)} & \mu_0 \dfrac{d}{dt}\displaystyle\int_\Omega \mathbf{H} \cdot \boldsymbol{\Psi}\, \mathrm{dx} + \int_\Omega \overrightarrow{\mathrm{curl}}\, E \cdot \boldsymbol{\Psi}\, \mathrm{dx} = 0, \quad \forall\, \boldsymbol{\Psi} \in [L^2(\Omega)]^2, \\ \text{(ii)} & \epsilon_0 \dfrac{d}{dt}\displaystyle\int_\Omega E\phi\, \mathrm{dx} - \int_\Omega \mathbf{H} \cdot \overrightarrow{\mathrm{curl}}\, \phi\, \mathrm{dx} + \sqrt{\dfrac{\epsilon_0}{\mu_0}} \displaystyle\int_\Gamma E\phi\, \mathrm{d}\Gamma, \\ & + \displaystyle\int_\omega \lambda\phi\, \mathrm{d}\omega = 0, \quad \forall\, \phi \in H^1(\Omega), \\ \text{(iii)} & \displaystyle\int_\omega E\tau\, \mathrm{d}\omega = 0, \quad \forall\, \tau \in L^2(\omega), \\ \text{(iv)} & E(\mathbf{x}, t=0) = E_0(\mathbf{x}), \text{ and } \mathbf{H}(\mathbf{x}, t=0) = \mathbf{H}_0(\mathbf{x}) \quad \text{in } \Omega. \end{cases} \tag{6.30}$$

6.7.1 Conservation of Energy

In this section we derive an energy identity from the variational formulation (6.30). The energy identity presented below guarantees the well-posedness of the problem, and the stability of the solution.

Theorem 3 *The system* (6.30) *verifies the following energy identity*

$$\frac{d}{dt}\mathcal{E} = -\sqrt{\frac{\epsilon_0}{\mu_0}}\,\|E\|^2_{L^2(\Gamma)}\,, \tag{6.31}$$

where the energy $\mathcal{E}$ is defined as

$$\mathcal{E} = \frac{1}{2}\left\{\epsilon_0\,\|E\|^2_{L^2(\Omega)} + \mu_0\,\|\mathbf{H}\|^2_{L^2(\Omega)}\right\}, \tag{6.32}$$

with

$$\|\mu\|_{L^2(\Gamma)} = \left(\int_\Gamma |\mu|^2\,\mathrm{d}\Gamma\right)^{1/2}. \tag{6.33}$$

Thus, (6.31) *implies that the energy does not grow over time, i.e.,*

$$\mathcal{E}(t) \leq \mathcal{E}(0),\ \forall t > 0. \tag{6.34}$$

Proof 7 : *Let us take $\phi = E$ in (6.30,* ii*). We obtain*

$$\epsilon_0\frac{d}{dt}\int_\Omega |E|^2\,\mathrm{dx} - \int_\Omega \mathbf{H}\cdot\overrightarrow{\mathrm{curl}}\,E\,\mathrm{dx} + \sqrt{\frac{\epsilon_0}{\mu_0}}\int_\Gamma |E|^2\,\mathrm{d}\Gamma + \int_\omega \lambda E\,\mathrm{d}\omega = 0. \tag{6.35}$$

Next, we take $\mathbf{\Psi} = \mathbf{H}$ in (6.30, i*). With this choice we get*

$$\mu_0\frac{d}{dt}\int_\Omega |\mathbf{H}|^2\,\mathrm{dx} + \int_\Omega \overrightarrow{\mathrm{curl}}\,E\cdot\mathbf{H}\,\mathrm{dx} = 0. \tag{6.36}$$

Adding equations (6.35) *and* (6.36) *we have*

$$\epsilon_0\frac{d}{dt}\int_\Omega |E|^2 + \mu_0\frac{d}{dt}\int_\Omega |\mathbf{H}|^2\,\mathrm{dx} + \sqrt{\frac{\epsilon_0}{\mu_0}}\int_\Gamma |E|^2\,\mathrm{d}\Gamma + \int_\omega \lambda E\,\mathrm{d}\omega = 0, \tag{6.37}$$

which can be rewritten as

$$\frac{1}{2}\frac{d}{dt}\left(\epsilon_0\,\|E\|^2_{L^2(\Omega)} + \mu_0\,\|\mathbf{H}\|^2_{L^2(\Omega)}\right) + \sqrt{\frac{\epsilon_0}{\mu_0}}\int_\Gamma |E|^2\,\mathrm{d}\Gamma + \int_\omega \lambda E\,\mathrm{d}\omega = 0. \tag{6.38}$$

From (6.30, iii*) taking $\tau = \lambda$ in we obtain*

$$\int_{\omega} E\lambda \, \mathrm{d}\omega = 0. \tag{6.39}$$

Substituting (6.39) *in* (6.38)*, and using the definition of the energy* (6.32) *we have*

$$\frac{d}{dt}\mathcal{E} = -\sqrt{\frac{\epsilon_0}{\mu_0}} \, \|E\|^2_{L^2(\Gamma)} \, . \tag{6.40}$$

Equation (6.40) implies that there is no dissipation of the waves in the domain Ω. As seen before this is the principle of *conservation of energy* for the variational formulation (6.30) for Maxwell's equation.

6.7.2 The Discrete Model

For the space discretization we will use the same discrete spaces as described in Chapter 5, Section 5.7.1. The degrees of freedom for E and $\mathbf{H}$ are shown in Figure 5.2. Thus, let Ω be a union of rectangles, and consider a regular mesh $(\mathcal{T}_h)$ with square elements (K) of edge $h > 0$ as in Figure 5.2. On a reference element we choose the Q_1 space of bilinear finite elements to approximate the electric field E and the lowest order Raviart-Thomas space, $RT_{[0]}$ for the discretization of the magnetic field $\mathbf{H}$. Based on these approximation spaces, the space discrete scheme is defined as:

Find $\{E_h(\cdot,t), \mathbf{H}_h(\cdot,t), \lambda_h(\cdot,t)\} \in \mathcal{U}_h \times \mathcal{V}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} \text{(i)} & \mu_0 \dfrac{d}{dt} \displaystyle\int_{\Omega} \mathbf{H}_h \cdot \mathbf{\Psi}_h \, \mathrm{dx} + \int_{\Omega} \overrightarrow{\mathrm{curl}} \, E_h \cdot \mathbf{\Psi}_h \, \mathrm{dx} = 0, \quad \forall \, \mathbf{\Psi}_h \in \mathcal{V}_h, \\ \text{(ii)} & \epsilon_0 \dfrac{d}{dt} \displaystyle\int_{\Omega} E_h \phi_h \, \mathrm{dx} - \int_{\Omega} \mathbf{H}_h \cdot \overrightarrow{\mathrm{curl}} \, \phi_h \, \mathrm{dx} + \sqrt{\frac{\epsilon_0}{\mu_0}} \int_{\Gamma} E_h \phi_h \, \mathrm{d\Gamma}, \\ & + \displaystyle\int_{\omega} \lambda_h \phi_h \, \mathrm{d}\omega = 0, \quad \forall \, \phi_h \in \mathcal{U}_h, \\ \text{(iii)} & \displaystyle\int_{\omega} E_h \tau_h \, \mathrm{d}\omega = 0, \quad \forall \, \tau_h \in \mathbf{\Lambda_h}, \\ \text{(iv)} & E_h(\mathbf{x}, t=0) = E_{0,h}(\mathbf{x}), \text{ and } \mathbf{H}_h(\mathbf{x}, t=0) = \mathbf{H}_{0,h}(\mathbf{x}) \quad \text{in } \Omega. \end{cases} \tag{6.41}$$

In (6.41, iii), the space $\mathbf{\Lambda}_h$ is the space of distributed Lagrange multipliers that was defined in Chapter 3, Section 3.4.2 in (3.22).

For the time discretization, as done in Chapter 5, we will use a staggered leapfrog scheme in which the electric and magnetic field components are obtained at time steps that are 1/2 units apart. For example, we compute the magnetic field at the time step $n+1/2$ and the electric field at the time step $n+1$. In Chapter 5 we have performed all numerical results without resorting to the use of quadrature formulas to obtain an explicit system in time. Here, we use mass lumping to obtain diagonal mass matrices and hence an explicit scheme in time. In this case the scheme, in the absence of the distributed Lagrange multiplier, reduces to the FDTD scheme whose numerical properties (dispersion, anisotropy) were discussed in Chapter 5, Section 5.8. Using similar notations as done in Chapter 5, Section 5.7.2, the fully discrete scheme can be presented as:

Find $\{E_h^{n+1}, \mathbf{H}_h^{n+1/2}, \lambda_h^{n+1}\} \in \mathcal{U}_h \times \mathcal{V}_h \times \mathbf{\Lambda}_h$ such that:

$$
\begin{cases}
\text{(i)} & \mu_0(\Delta_t \mathbf{H}_h^n, \mathbf{\Psi}_h) + (\overrightarrow{\text{curl}}\, E_h^n, \mathbf{\Psi}_h) = 0, \quad \forall\, \mathbf{\Psi}_h \in \mathcal{V}_h, \\
\text{(ii)} & \epsilon_0(\Delta_t E_h^{n+1/2}, \phi_h) - (\mathbf{H}_h^{n+1/2}, \overrightarrow{\text{curl}}\, \phi_h) + \sqrt{\dfrac{\epsilon_0}{\mu_0}}(\underline{E_h}^{n+1/2}, \phi_h)_\Gamma, \\
& +(\lambda_h^{n+1}, \phi_h)_\omega = 0, \quad \forall\, \phi_h \in \mathcal{U}_h, \\
\text{(iii)} & (E_h^{n+1}, \tau_h)_\omega = 0, \quad \forall\, \tau_h \in \mathbf{\Lambda}_h, \\
\text{(iv)} & E_h^0(\mathbf{x}) = E_{0,h}(\mathbf{x}), \text{ and } \mathbf{H}_h^{-1/2}(\mathbf{x}) = 2\mathbf{H}_{0,h}(\mathbf{x}) - \mathbf{H}_h^{1/2}(\mathbf{x}) \quad \text{in } \Omega.
\end{cases}
\tag{6.42}
$$

In the above the notation $(\cdot, \cdot)$ stands for the L^2 inner product of either scalar or vector fields. The notation $(\cdot, \cdot)_X$ refers to the L^2 inner product over the domain X.

Let E^{yee} be the solution to the FDTD (Yee) scheme. Then the update equations for the

scheme (6.42) can be presented as follows. For an interior node (l, m) we have

$$\begin{cases} \text{(i)} \quad H_x|_{l,m+1/2}^{n+1/2} = H_x|_{l,m+1/2}^{n-1/2} - \dfrac{\Delta t}{\mu_o h}(E|_{l,m+1}^{n} - E|_{l,m}^{n}) \\ \text{(ii)} \quad H_y|_{l+1/2,m}^{n+1/2} = H_y|_{l+1/2,m}^{n-1/2} + \dfrac{\Delta t}{\mu_o h}(E^n|_{l+1,m} - E^n|_{l,m}) \\ \text{(iii)} \quad E^{\text{yee}}|_{l,m}^{n+1} = E|_{l,m}^{n} + \dfrac{\Delta t}{\epsilon_0 h}(H_y|_{l+1/2,m}^{n+1/2} - H_y|_{l-1/2,m}^{n+1/2}) \\ \qquad\qquad\qquad\qquad -\dfrac{\Delta t}{\epsilon_0 h}(H_x|_{l,m+1/2}^{n+1/2} - H_x|_{l,m-1/2}^{n+1/2}) \end{cases} \tag{6.43}$$

For a node on the boundary Γ, the boundary integral

$$\int_\Gamma \underline{E_h}^{n+1/2} \phi_h \mathrm{d}\Gamma, \tag{6.44}$$

will contribute terms to both the right hand side and the left hand side of equation (6.43, iii), as this term involves E_h^{n+1}, which is unknown, as well as E_h^n, which is known. In this case (6.43, iii) has to be modified as

$$\text{(iii)} \quad E^{\text{yee}}|_{l,m}^{n+1} = \frac{\gamma^-}{\gamma^+} E|_{l,m}^{n} + \frac{\Delta t}{\epsilon_0 h \gamma^+} S\mathbf{H}|_{l,m}^{n+1/2}. \tag{6.45}$$

In the above

$$\gamma^- = \left(\frac{1}{\beta} - \frac{\kappa c \Delta t}{\alpha h}\right);\ \gamma^- = \left(\frac{1}{\beta} + \frac{\kappa c \Delta t}{\alpha h}\right), \tag{6.46}$$

where, for an interior node $\beta = 1, \alpha = 1, \kappa = 0$, for a boundary node but not a corner node $\beta = 2, \alpha = 2, \kappa = 1$, and for a boundary corner node $\beta = 4, \alpha = 4, \kappa = 1$. Also, S is the stiffness matrix associated with the integral

$$\int_\Omega \mathbf{H}^{n+1/2} \overrightarrow{\text{curl}}\, \phi_h \, \mathrm{dx},\ \forall \phi_h \in \mathcal{U}_h. \tag{6.47}$$

The solution E_h^{n+1} to the scheme (6.42) is obtained from the solution E_h^{yee}, to the FDTD scheme (including boundary terms), by adjusting for the Dirichlet condition on the obstacle via the Lagrange multiplier λ_h^{n+1}. Thus, we will solve a system of the form

Find $(E_h^{n+1}, \lambda_h^{n+1}) \in \mathcal{U}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} D_h E_h^{n+1} + B_h^T \lambda_h^{n+1} = E_h^{\text{yee}}|^{n+1}, \\ \\ B_h E_h^{n+1} = 0, \end{cases} \tag{6.48}$$

where the operator B_h and its transpose B_h^T are associated to the integration formula (3.23) in Chapter 3, Section 3.4.2. D_h is the lumped mass matrix associated to the integral $\int_\Omega E_h\phi_h \, \mathrm{dx}, \ \forall \phi_h \in \mathcal{U}_h$. We will use the Uzawa algorithm 1, described in Chapter 3, to solve this system.

6.8 Implementing a Fictitious Domain Method in the Uniaxial PML

In order to use the uniaxial PML model, instead of the first order Silver-Müller boundary condition, we have to modify the discrete model (5.98) to account for the Dirichlet condition on the boundary of the obstacle. We rewrite this discrete model here for completeness.

Find $(E_h^{\mathrm{yee}}|^{n+1}, D_h^{n+1}, \mathbf{H}_h^{n+\frac{1}{2}}, \mathbf{B}_h^{n+\frac{1}{2}}) \in \mathcal{U}_h \times \mathcal{U}_h \times \mathcal{V}_h \times \mathcal{V}_h$ such that for all $\mathbf{\Psi}_h \in \mathcal{V}_h$, for all $\phi_h \in \mathcal{U}_h$,

$$\left\{\begin{array}{llcll} \text{(i)} & (\Delta_t \mathbf{B}_h^n, \mathbf{\Psi_h}) & = & -\frac{1}{\epsilon_0}(\Sigma_2 \underline{\mathbf{B}_h^n}, \mathbf{\Psi}_h) & - \ (\overrightarrow{\mathrm{curl}}\, E_h^n \ , \ \mathbf{\Psi}_h), \\ \text{(ii)} & (\Delta_t \mathbf{H}_h^n, \mathbf{\Psi}_h) & = & \frac{1}{\mu_0}(\Delta_t \mathbf{B}_h^n, \mathbf{\Psi}_h) & + \ \frac{1}{\epsilon_0\mu_0}(\Sigma_1 \underline{\mathbf{B}_h^n}, \mathbf{\Psi}_h), \\ \text{(iii)} & (\Delta_t D_h^{n+\frac{1}{2}}, \phi_h) & = & -\frac{1}{\epsilon_0}(\sigma_x \underline{D_h}^{n+\frac{1}{2}}, \phi_h) & + \ (\overrightarrow{\mathrm{curl}}\, \phi_h, \mathbf{H}_h^{n+\frac{1}{2}}), \\ \text{(iv)} & (\Delta_t E_h^{\mathrm{yee}}|^{n+\frac{1}{2}}, \phi_h) & = & -\frac{1}{\epsilon_0}(\sigma_y \underline{E_h^{\mathrm{yee}}}|^{n+\frac{1}{2}}, \phi_h) & + \ \frac{1}{\epsilon_0}(\Delta_t D_h^{n+\frac{1}{2}}, \ \phi_h). \end{array}\right. \tag{6.49}$$

In the above

$$\Delta_t E_h^{\mathrm{yee}}|^{n+1/2} = \frac{E_h^{\mathrm{yee}}|^{n+1} - E_h^n}{\Delta t}, \tag{6.50}$$

and

$$\underline{E_h^{\mathrm{yee}}}|^{n+1/2} = \frac{E_h^{\mathrm{yee}}|^{n+1} + E_h^n}{2}. \tag{6.51}$$

Once we obtain the solution to the system (6.49), the solution to the scattering problem is obtained by solving a linear system similar to (6.48). Thus, the problem is:

To find $(E_h^{n+1}, \lambda_h^{n+1}) \in \mathcal{U}_h \times \mathbf{\Lambda}_h$ such that:

$$\begin{cases} D_h E_h^{n+1} + B_h^T \lambda_h^{n+1} = E_h^{\text{yee}}|^{n+1}, \\ \\ B_h E_h^{n+1} = 0. \end{cases} \tag{6.52}$$

6.9 Scattering by a Disk

We repeat the numerical experiment performed in Chapter 3 using the fictitious domain formulation for the two-dimensional TM mode of Maxwell's equations. We do this for the case of the Silver-Müller absorbing boundary condition considered in Section 6.7, as well as for the uniaxial PML model considered in Section 6.8. The problem description is the same as in Chapter 3, Section 3.8.1. We repeat relevant details here for the sake of completeness.

We consider the scattering of the harmonic planar waves $e^{-\mathrm{i}(\rho t - \mathbf{k}\cdot\mathbf{x})}$ by a perfectly reflecting disk whose radius is 0.25 meter. The frequency, f, is 0.6 GHz, and the wavelength, L, is 0.5 meter. The angular frequency is $\rho = 2\pi f$. The wave illuminates ω from the left and propagates horizontally. As before, we have used a rectangular mesh consisting of 113×113 nodes, with the mesh step size $h = 0.5/16$ meter. The time step is $\Delta t = 2\pi/(25\rho)$. We have also considered mesh refinements in order to estimate the accuracy of our solution.

In Figure 6.1 we plot the number of degrees of freedom (DOF) of the Lagrange multiplier on the boundary of the disk, $\partial\omega$, as a function of the mesh ratio, $h_{\partial\omega}/h$, for different discretizations, where $h_{\partial\omega}$ is defined as the step size on the boundary of the disk. As can be seen from Figure 6.1, for fine meshes, as opposed to coarse meshes, bigger changes in the DOF on the boundary of the disk result from a small change in the mesh ratio. The Uzawa algorithm converges in a finite number of iterations for values of $h_{\partial\omega}/h$ between 1.5 and 3. However, for certain values between 1.5 and 1.8 the behavior of the Uzawa algorithm is unstable with respect to number of iterations. Thus we consider values for $h_{\partial\omega}/h$ between

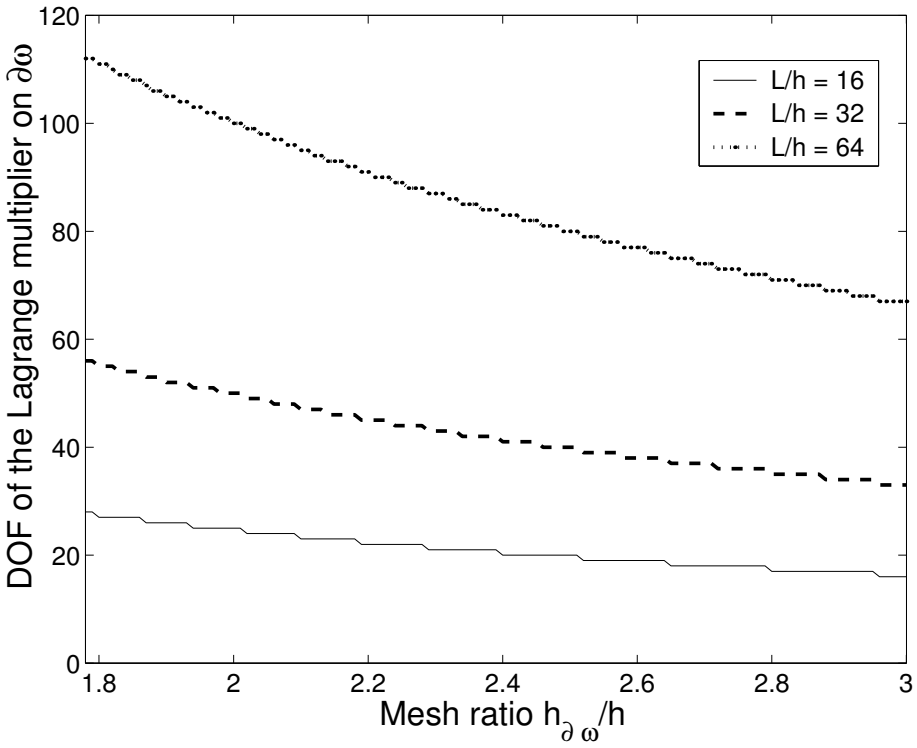

Figure 6.1: The number of degrees of freedom (DOF) of the Lagrange multiplier on the boundary of the disk versus the mesh ratio $h_{\partial\omega}/h$.

1.8 and 3.

Figure 6.2 is a contour plot of the exact solution. In Figure 6.3 we present contour plots of the solutions computed using the fictitious domain method with the Silver-Müller boundary condition (left) and a PML of thickness $L/4$ (right).

In Figures 6.3, 6.4 and 6.5, we plot the error (pointwise difference) between each computed solution and the exact solution for discretizations with 16, 32 and 64 nodes per wavelength, respectively. It is clearly observed that, as the mesh is refined, the error in the fictitious domain method with the Silver-Müller condition is dominated by reflections from the artificial boundary. In the case of the PML model the error in the discretization of the Lagrange multiplier dominates the total error.

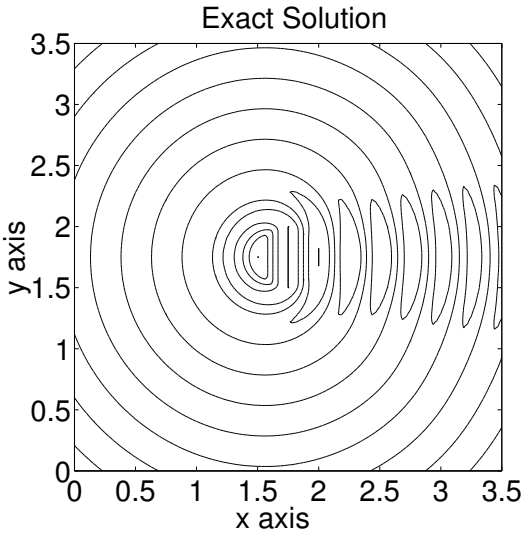

Figure 6.2: Contour plot of the exact solution

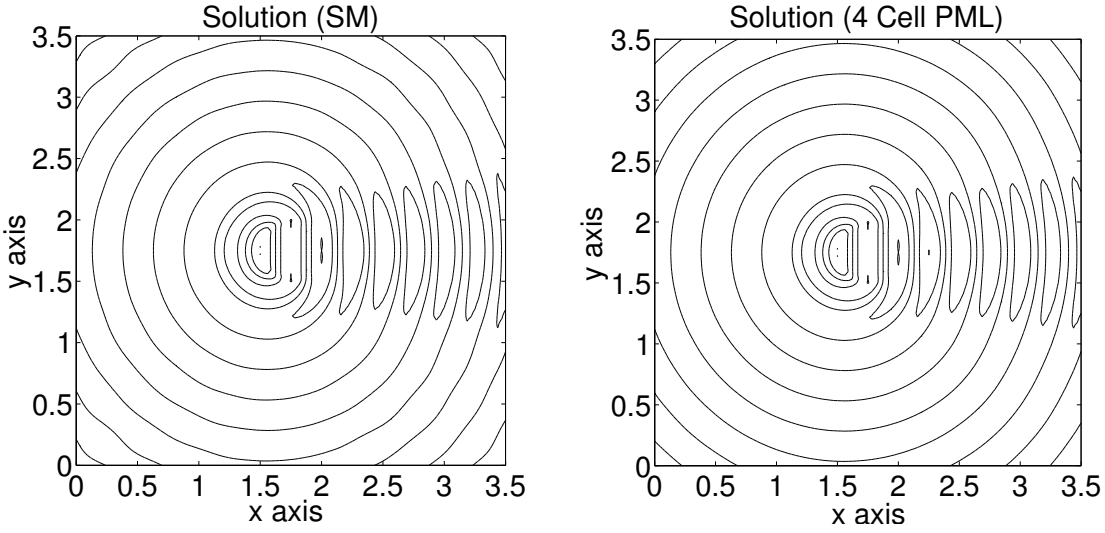

Figure 6.3: Contour plots of the fictitious domain solution with the Silver-Müller boundary condition (left) and the 4 cell PML (right) of thickness $L/4$, for a discretization with 16 nodes per wavelength.

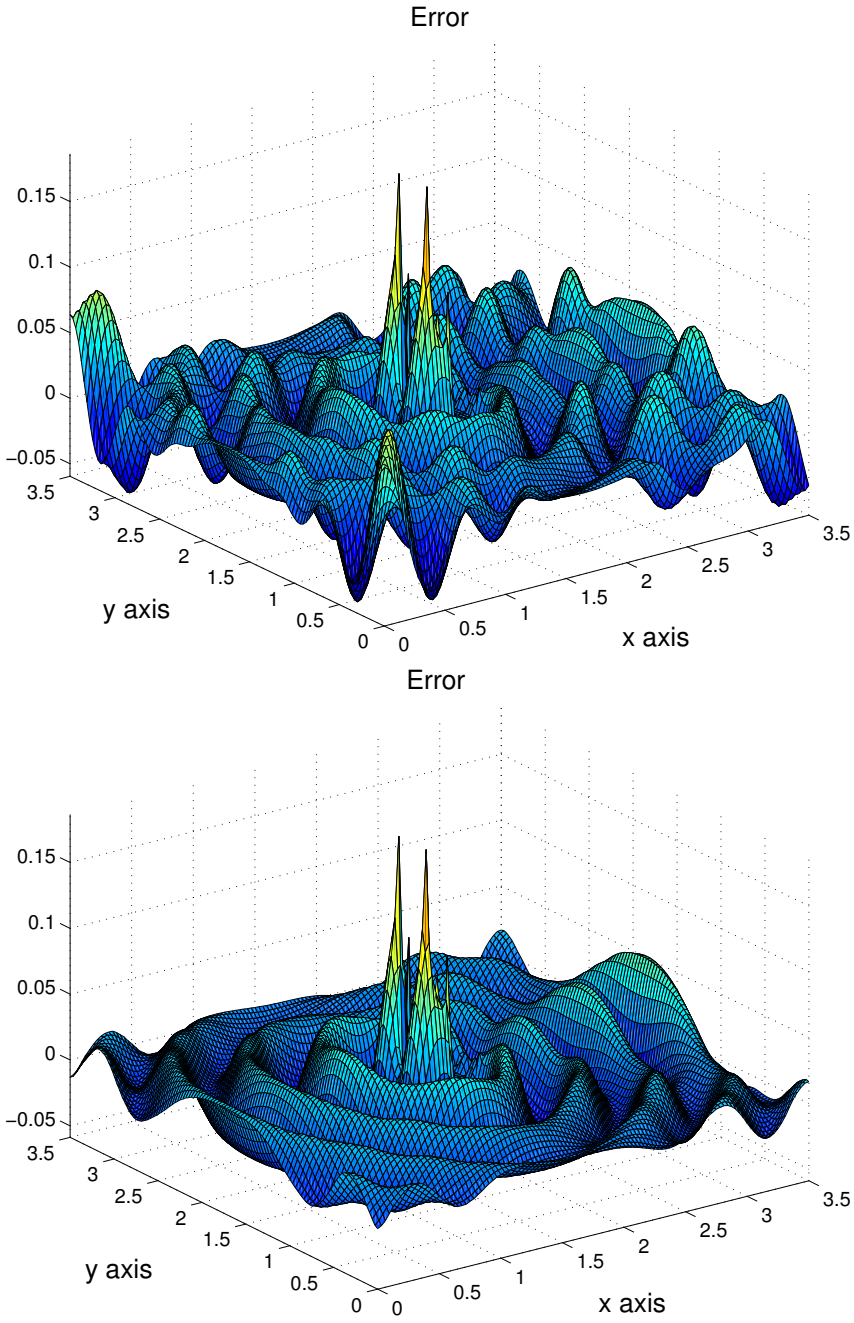

Figure 6.4: Plot of the error between the exact solution and the fictitious domain method with the Silver-Müller boundary condition (top) and the 4 cell PML (bottom) for a discretization with 16 nodes per wavelength.

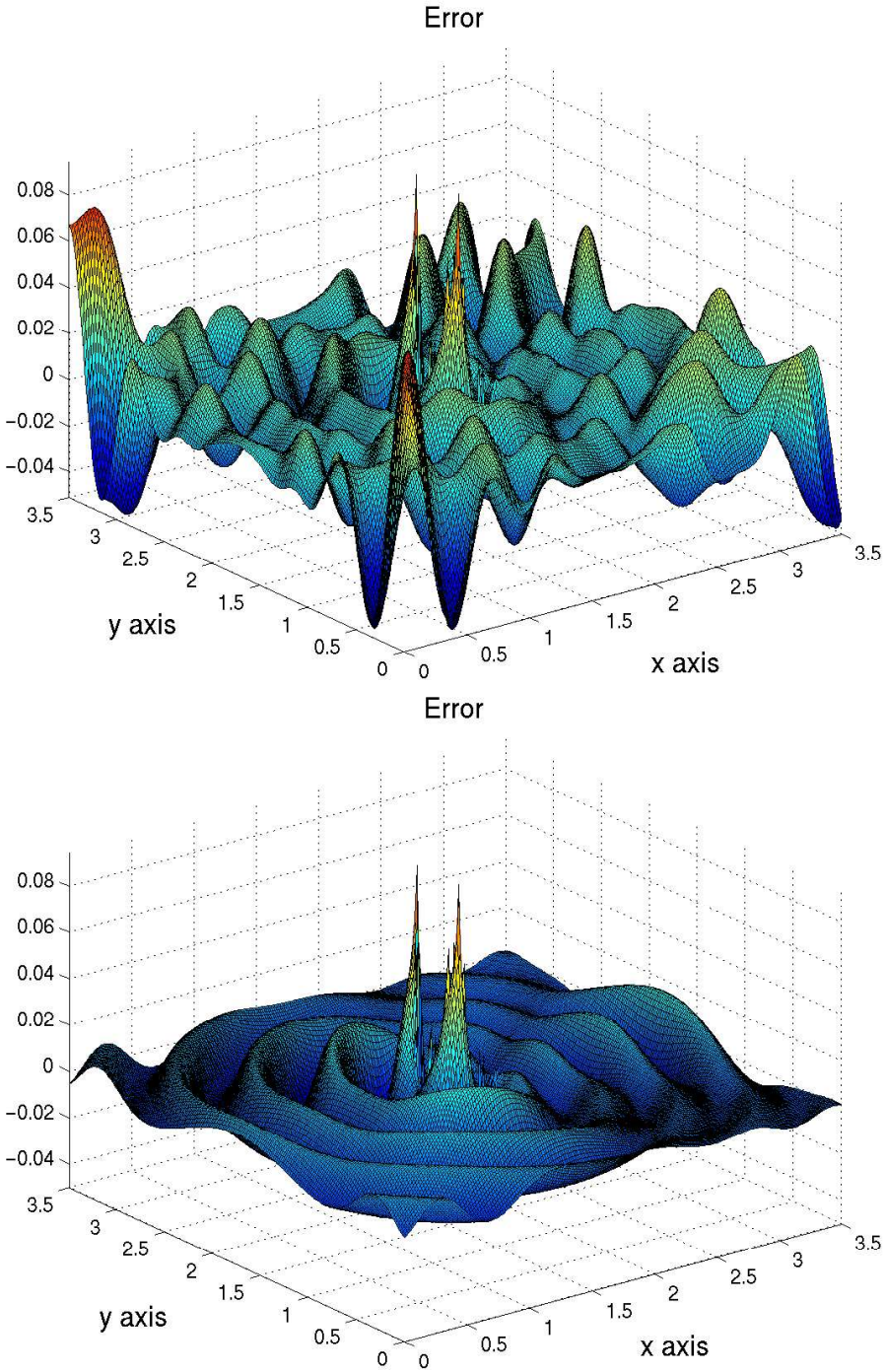

Figure 6.5: Plot of the error between the exact solution and the fictitious domain method with the Silver-Müller boundary condition (top) and the 4 cell PML (bottom) for a discretization with 32 nodes per wavelength.

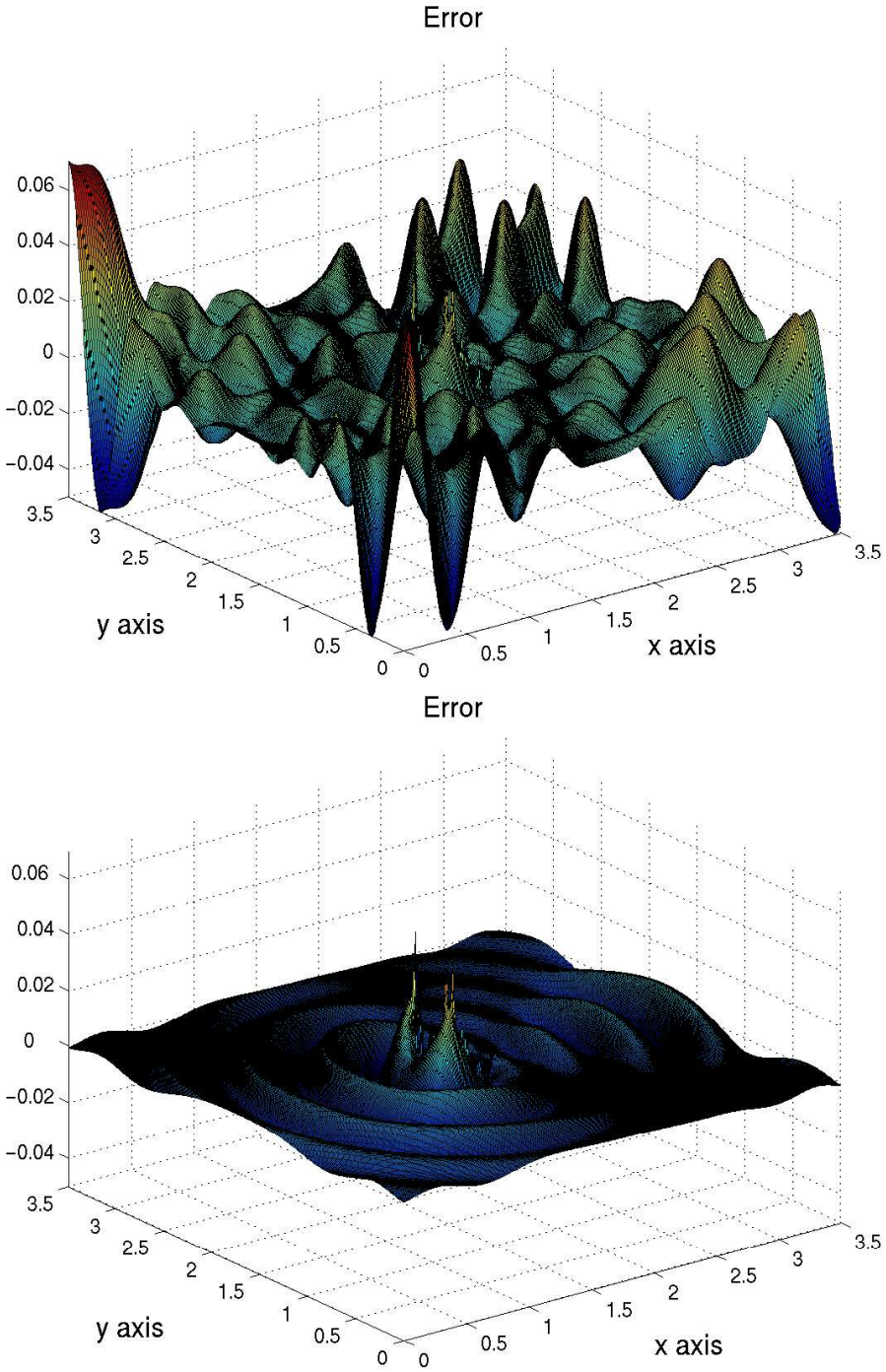

Figure 6.6: Plot of the error between the exact solution and the fictitious domain method with the Silver-Müller boundary condition (top) and the 4 cell PML (bottom) for a discretization with 64 nodes per wavelength.

We define the relative error (RE) between the exact solution and a computed solution as

$$\text{RE} = \frac{\|E_{\text{exact}} - E_C\|_{L^2(\Omega)}}{\|E_{\text{exact}}\|_{L^2(\Omega)}}, \tag{6.53}$$

where E_{exact} stands for the exact solution, and E_C denotes a computed solution. In Figure 6.7 we plot the relative error for the fictitious domain method with a PML of thickness $L/4$, against the mesh ratio $h_{\partial\omega}/h$, for different discretizations. From Figure 6.7 we can see that the relative error can vary by a factor of 2 for different values of the mesh ratio. In Figure 6.8 we plot ratios of relative errors between successive mesh refinements obtained from Figure 6.7. The solid line represents the ratio of the relative error for a discretization with 16 nodes per wavelength, and the relative error of a discretization with 32 nodes per wavelength. Similarly, we plot the ratio of relative errors for discretizations with 32 and 64 nodes per wavelength (dashed line). Again, we observe that the ratios can vary by almost a factor of 3.

In Figures 6.9 and 6.10, we plot the maximum and the minimum iterations counts, respectively, that are required for the convergence of the Uzawa algorithm, as a function of the mesh ratio. These results are for the fictitious domain method with a PML of thickness $L/4$. The number of iterations for the three discretizations is seen to be bounded by 20.

In Table 6.1 we present relative errors and maximum and minimum iteration counts for the fictitious domain method with the Silver-Müller boundary condition. As observed in Figures 6.4, 6.5, and 6.6, reflections from the artificial boundary dominate the error. Thus, we do not expect to see much improvement in the error as the mesh is refined.

In Table 6.2 we present relative errors and maximum and minimum iteration counts for the fictitious domain method with PML's of thickness, $L/4$, $L/2$ and L. The relative error in all three cases is almost the same. Thus, we do not obtain any benefit by increasing the thickness of the PML as the dominating error is due to the discretization of the Lagrange multiplier.

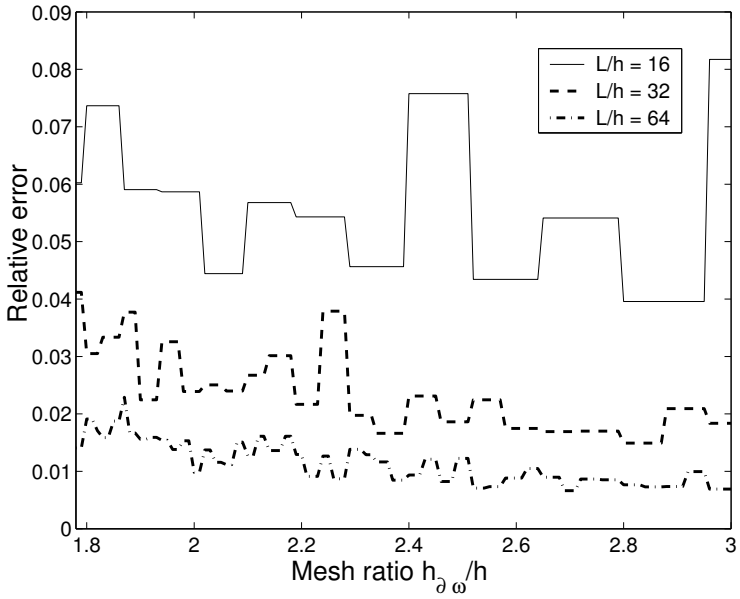

Figure 6.7: Plot of the relative error versus the mesh ratio $h_{\partial\omega}/h$ for three different discretizations.

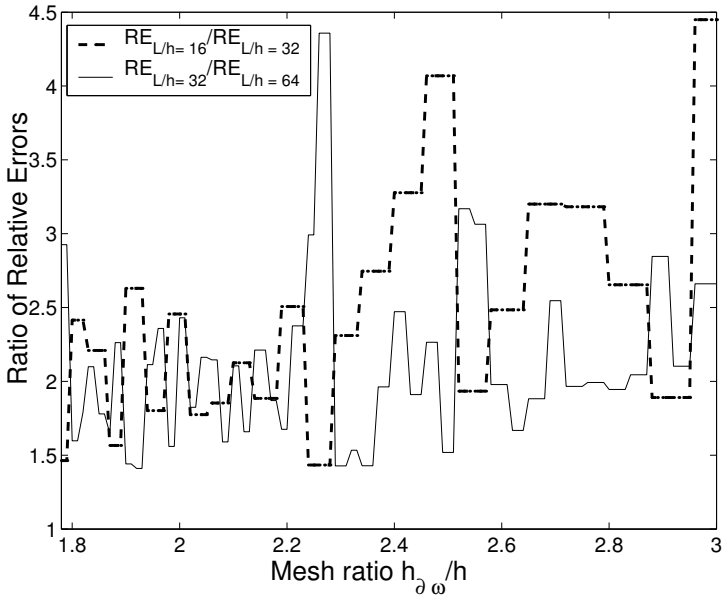

Figure 6.8: Plot of ratios of successive relative errors versus the mesh ratio $h_{\partial\omega}/h$.

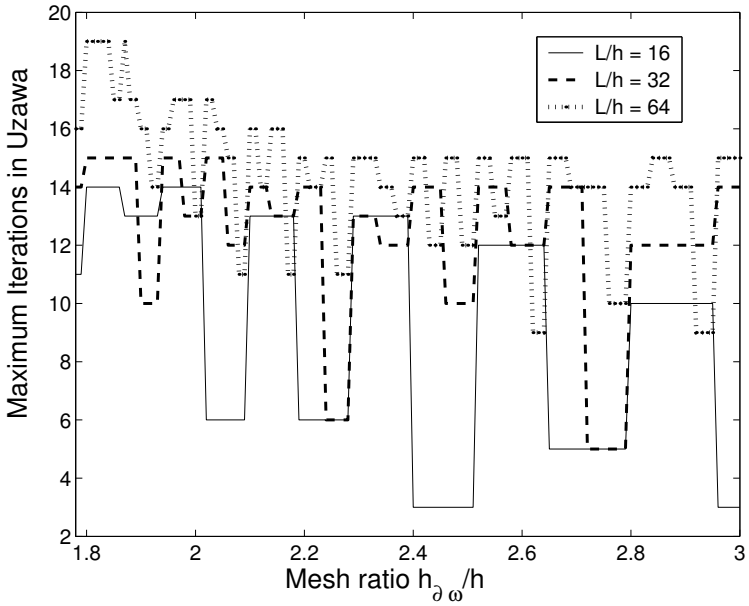

Figure 6.9: The maximum number of iterations required for the Uzawa algorithm versus the mesh ratio $h_{\partial\omega}/h$ for three different discretizations.

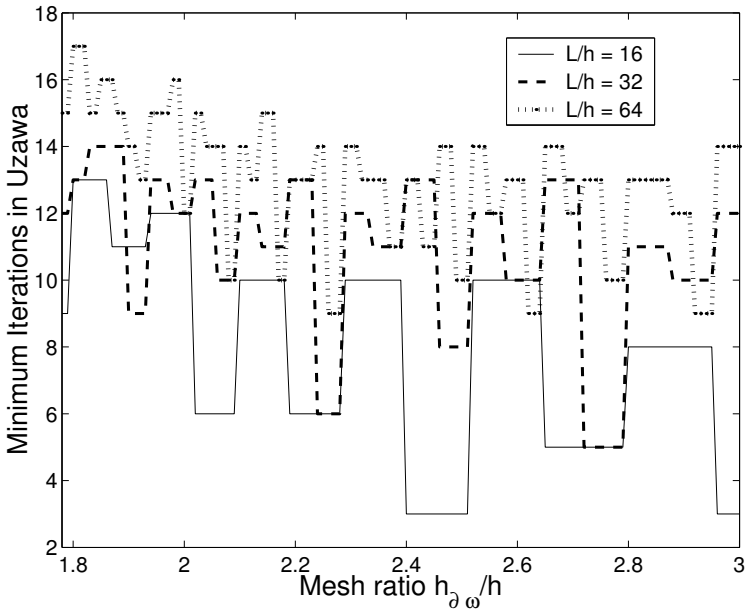

Figure 6.10: The minimum number of iterations required for the Uzawa algorithm versus the mesh ratio $h_{\partial\omega}/h$ for three different discretizations.

h	Δt	SM		
		RE	Max Iter	Min Iter
$L/16$	$L/(25c)$	7.026e-2	14	12
$L/32$	$L/(50c)$	4.519e-2	13	12
$L/64$	$L/(100c)$	3.854e-2	13	12

Table 6.1: Table of relative errors of the fictitious domain solution, with the Silver-Müller (SM) boundary condition, computed with respect to the exact solution for different discretizations.

PML	h	Δt	RE	Max Iter	Min Iter
$L/4$	$L/16$	$L/(25c)$	5.867e-2	14	12
	$L/32$	$L/(50c)$	2.389e-2	13	12
	$L/64$	$L/(100c)$	9.830e-3	13	12
$L/2$	$L/16$	$L/(25c)$	5.815e-2	14	12
	$L/32$	$L/(50c)$	2.389e-2	13	12
	$L/64$	$L/(100c)$	9.830e-3	13	12
L	$L/16$	$L/(25c)$	5.815e-2	14	12
	$L/32$	$L/(50c)$	2.389e-2	13	12
	$L/64$	$L/(100c)$	9.829e-3	13	12

Table 6.2: Table of relative errors of the fictitious domain solutions, for PML's of varying thickness, computed with respect to the exact solution.

This can also be observed in the Figures 6.4, 6.5, and 6.6. A quarter wavelength thick PML is sufficient to obtain significant improvements over the first order boundary condition. As seen in Table 6.2, the relative error for a discretization with 64 nodes per wavelength is

4 times smaller than the error for the same discretization with the Silver-Müller boundary condition.

Finally, in Table 6.3, we present relative errors and maximum and minimum iteration counts for the fictitious domain method with a PML of thickness $L/4$, for different values of the mesh ratio $h_{\partial\omega}/h$, and for different discretizations.

h	Δt	$h_{\partial\omega}/h$	PML ($L/4$)		
			RE	Max Iter	Min Iter
$L/16$	$L/(25c)$	2.0	5.867e-2	14	12
		2.4	7.577e-2	3	3
		2.7	5.412e-2	5	5
$L/32$	$L/(50c)$	2.0	2.389e-2	13	12
		2.4	2.312e-2	14	13
		2.7	1.691e-2	14	13
$L/64$	$L/(100c)$	2.0	9.830e-3	13	12
		2.4	9.355e-3	15	13
		2.7	6.641e-3	14	12
$L/128$	$L/(200c)$	2.0	6.972e-3	16	15
		2.4	4.957e-3	16	14
		2.7	4.679e-3	13	11

Table 6.3: Table of relative errors of the fictitious domain solutions for a PML of thickness $L/4$, computed with respect to the exact solution, for different values of the mesh ratio $h_{\partial\omega}/h$ and different discretizations.

For a given discretization, the relative errors are comparable for different values of the mesh ratio. We observe that the ratios between successive relative errors, for a fixed value of the mesh ratio, is approximately 2. Thus, the spatial accuracy of the fictitious domain method,

based on these results, seems to be about first order.

In Table 6.4 we compare the fictitious domain approach to a staircase approach using the finite difference time domain method. As can be seen from the table, the fictitious domain method provides a significant improvement over the staircase approximation. This is also evident from Figures 6.11 and 6.12, which compare the errors for both methods for 16 and 64 nodes per wavelength.

N	L/h	Staircase	Fictitious Domain
113^2	16	1.959e-1	5.867e-2
225^2	32	9.997e-2	2.389e-2
449^2	64	4.871e-2	9.830e-3
897^2	128	2.619e-2	6.972e-3

Table 6.4: Table of relative errors for the fictitious domain solution for a PML of thickness $L/4$, and relative errors for a staricase approximation for different nodes per wavelength.

So far we have presented results in which the frequency f, and hence the wavelength L, in the domain was fixed. As the frequency is increased, i.e., the wavelength is decreased, the effects of dispersion start to degrade the solution. The error in the solution is no longer dominated by the error in the discretization of the Lagrange multiplier. The error at higher frequencies is dominated by large phase errors, which accumulate over time and can significantly affect the solution. To study the errors that arise at high frequencies, we have calculated the relative errors in the case when $f = 1.2, 2.4$ and 4.8 GHz. The relative error is a combination of the error in the amplitude of the solution as well as the error in the phase. To see the dominance of the phase error, we also calculate a relative amplitude error and a phase error for each frequency. In these calculations the size of the domain is fixed. We use a fixed step size $h = 0.5/128$, which uses 128 nodes per wavelength at the lowest

frequency of 0.6 GHz, on the square domain $[0, 3.5] \times [0, 3.5]$. Thus, we have 897×897 nodes with $N = 897$. The time step is chosen so that the stability condition as before is $\eta = 0.64$. The results are all computed after 1400 iterations. All results are computed using a 4 cell PML and the fictitious domain method. We do not observe any significant reduction in the errors by increasing the thickness of the PML layer.

In Figure 6.13 we plot a top view of the solutions with $f = 0.6$ GHz (top) and $f = 1.2$ GHz (bottom). In Figure 6.14 we plot a top view of the solutions with $f = 2.4$ GHz (top) and $f = 4.8$ GHz (bottom). The computed solutions qualitatively compare well with the exact solution.

Let R_{exact} and I_{exact} denote the real and imaginary parts of the exact solution. Similarly, let R_C and I_C denote the real and imaginary parts of our computed solution. We will define the phase error (PE) as

$$\text{PE}(x) = \tan^{-1}\left(\frac{I_{\text{exact}}(x)}{R_{\text{exact}}(x)}\right) - \tan^{-1}\left(\frac{I_C(x)}{R_C(x)}\right). \tag{6.54}$$

We will calculate the phase error in degrees per node as

$$\text{Phase error } = \frac{360}{2\pi N}\left(\sum_{k=1}^{N^2} |\text{PE}(x_k)|^2\right)^{1/2} \text{degrees/node}, \tag{6.55}$$

where x_k is a node of the finite element triangulation $\mathcal{T}_h$. We will also define the amplitude error (AE) as

$$\text{AE}(x) = \sqrt{(I_{\text{exact}}(x))^2 + (R_{\text{exact}}(x))^2} - \sqrt{(I_C(x))^2 + (R_C(x))^2}. \tag{6.56}$$

We will calculate a relative amplitude error (RAE) for each frequency which will be the ratio of the L^2 norm of the amplitude error to the L^2 norm of the amplitude of the exact solution in the computational domain.

In Figure 6.15 we plot linear grayscale images of the phase error (top) and the amplitude error (bottom) over the square domain.

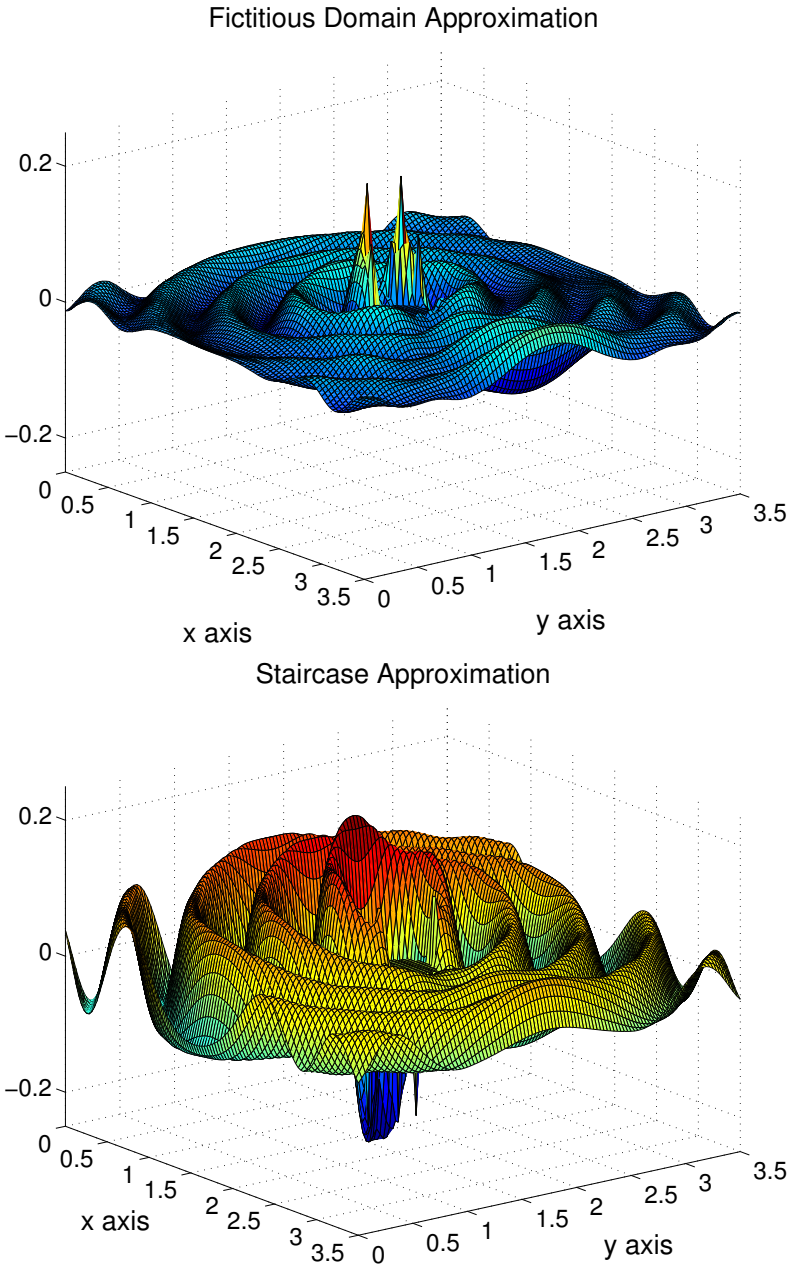

Figure 6.11: Plot of the error between the exact solution and the fictitious domain method with a 4 cell PML (top), and with a staircase approximation (bottom), for a discretization with 16 nodes per wavelength

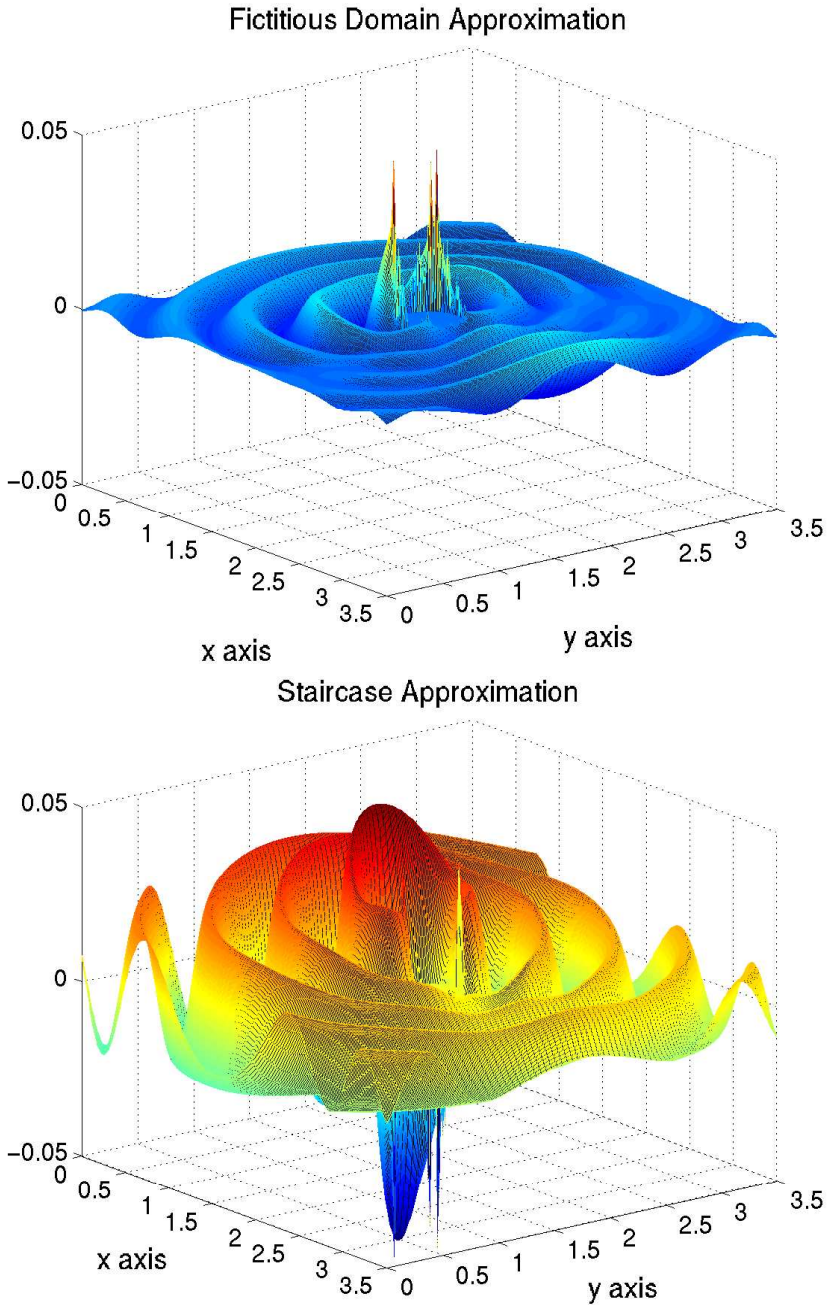

Figure 6.12: Plot of the error between the exact solution and the fictitious domain method with a 4 cell PML (top), and with a staircase approximation (bottom), for a discretization with 64 nodes per wavelength

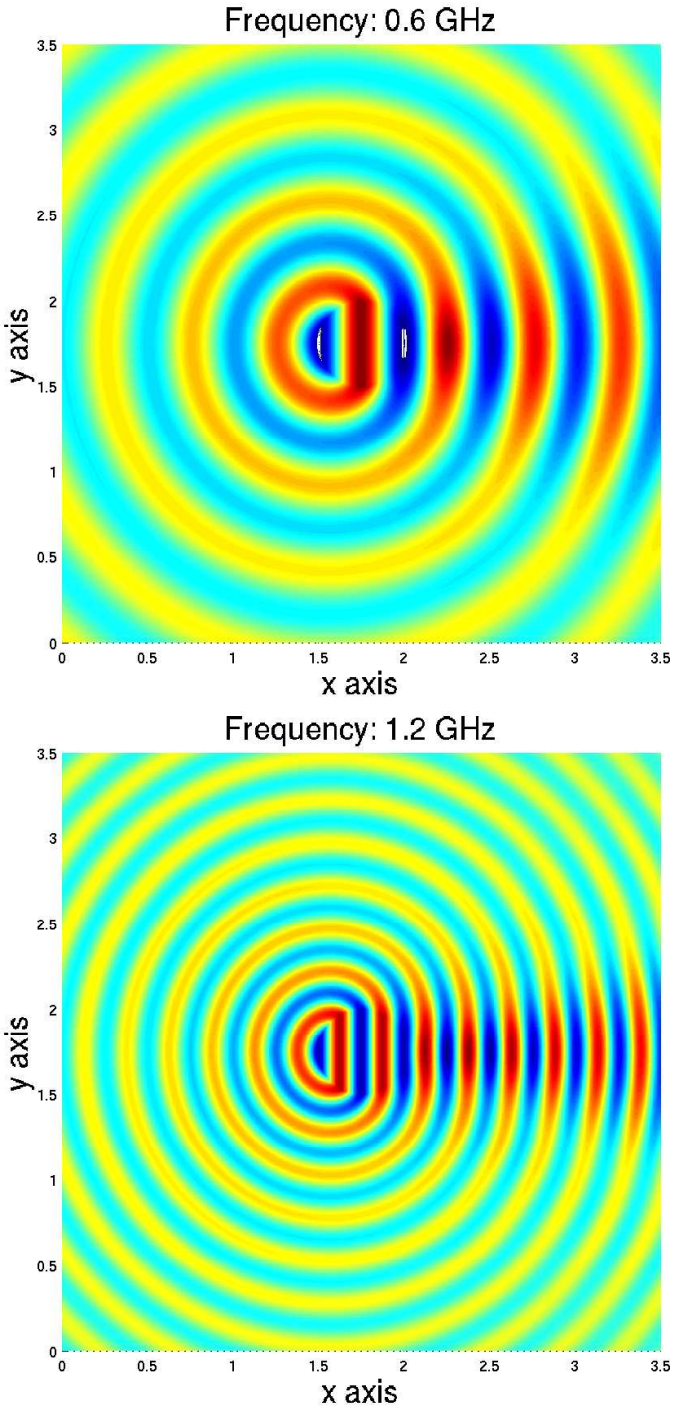

Figure 6.13: A top view of the computed solution (real part) for a harmonic planar wave with frequency $f = 0.6$ GHz and $L = 0.5$m (top), and for a harmonic planar wave with frequency $f = 1.2$ GHz and $L = 0.25$m (bottom).

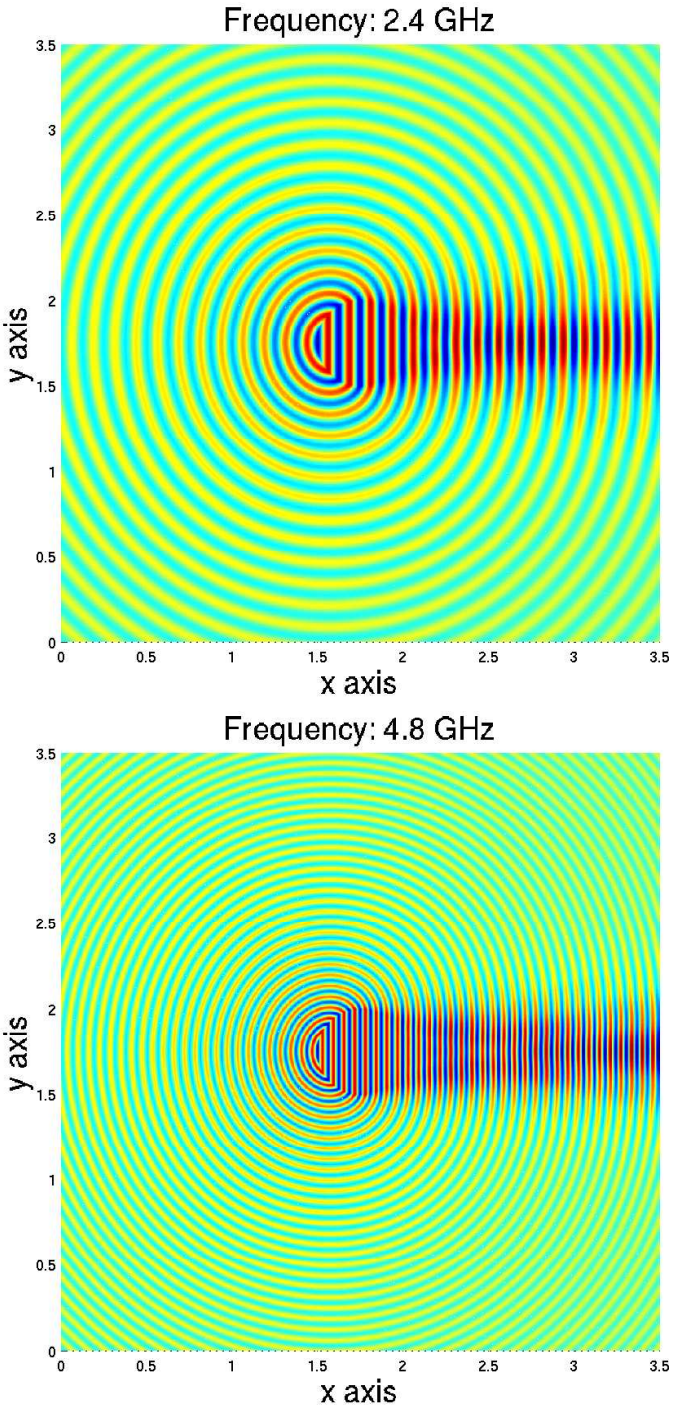

Figure 6.14: A top view of the computed solution (real part) for a harmonic planar wave with frequency $f = 2.4$ GHz and $L = 0.125$m (top), and for a harmonic planar wave with frequency $f = 4.8$ GHz and $L = 0.0625$m (bottom).

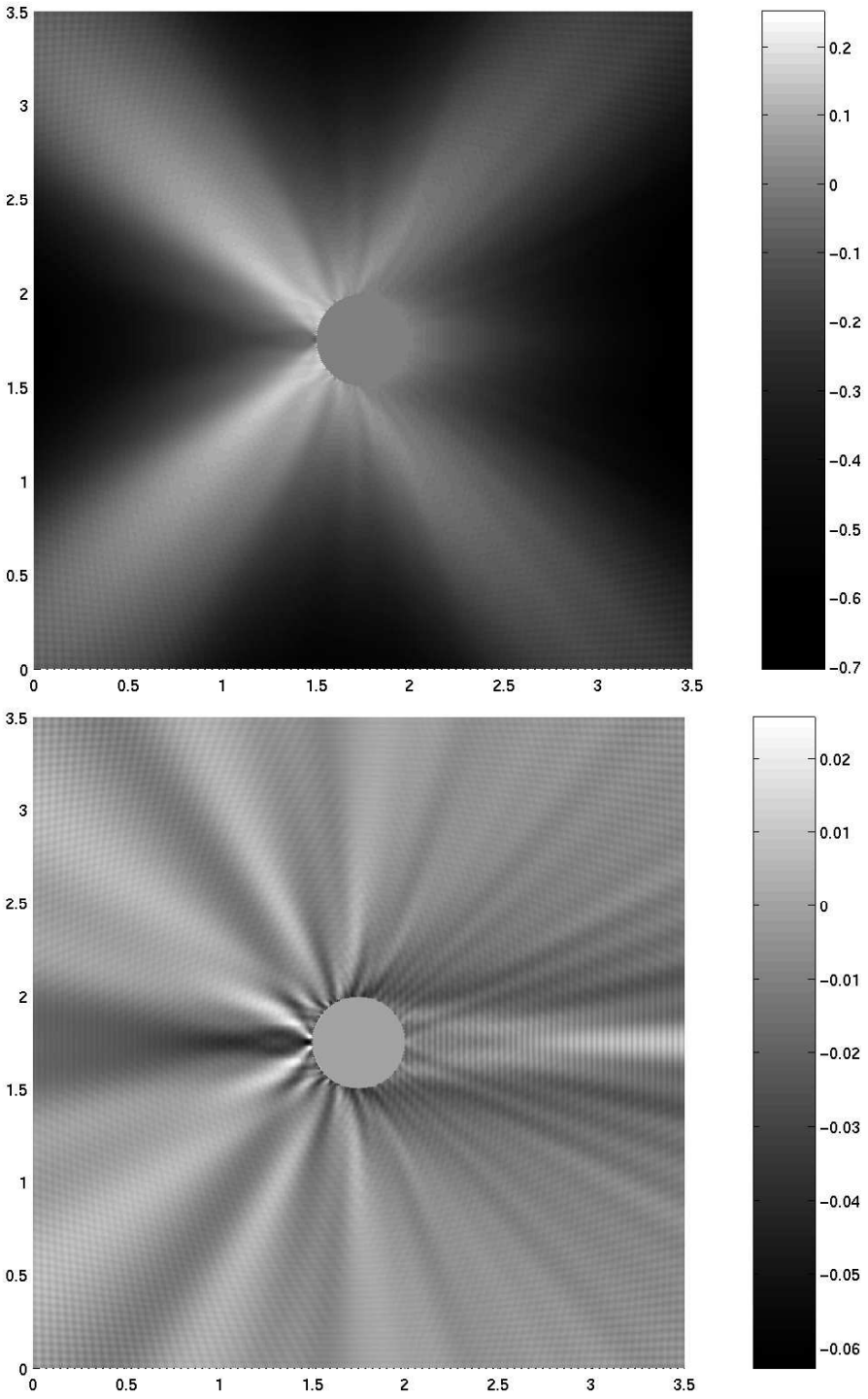

Figure 6.15: A linear gray scale image of the phase error in radians (top) and the amplitude error (bottom) over the square domain $[0, 3.5] \times [0, 3.5]$.

In Figure 6.15 we can see that the phase error is the smallest along the grid diagonals and it is the largest along the axis of the mesh.

In Table 6.5 we present the (total) relative errors for the real and imaginary parts of the solution. As expected the relative errors increase as the frequency is increased. We also compare the relative amplitude error RAE, calculated using the amplitude error AE defined in (6.56), and the phase error in degrees per node, defined in (6.55), for each case. The relative amplitude error is significantly better than the total relative errors. As can be seen from the table the phase error increases significantly at higher frequencies. For $f = 0.6$ GHz, the phase error is 0.37 degrees per node. This error increases to 15.69 degrees per node when $f = 4.8$ GHz.

f (GHz)	L (m)	L/h	RE (Real)	RE (Imag)	RAE	Phase
0.6	0.5	128	6.97e-3	7.77e-3	3.31e-3	0.37
1.2	0.25	64	1.57e-2	1.52e-2	5.72e-3	0.73
2.4	0.125	32	4.97e-2	4.72e-2	1.11e-2	2.29
4.8	0.0625	16	2.89e-1	2.93e-1	2.57e-2	15.69

Table 6.5: Table of errors for the fictitious domain solution for a PML of thickness $L/4$, at different frequencies. The relative error for the real and imaginary parts of the solution is given. RAE is a relative amplitude error and the phase error in degrees per node in each case is provided.

Chapter 7

Conclusion

In this dissertation we have proposed a novel fictitious domain method for the time dependent problem of scattering by an obstacle. The fictitious domain method is based on a distributed Lagrange multiplier which is used to enforce the Dirichlet condition on the boundary of the scatterer. We have implemented the fictitious domain approach for the two-dimensional scalar wave equation, and the two-dimensional TM mode of Maxwell's equations with Dirichlet conditions on the boundary of the obstacle in each case.

We have proposed and analyzed an operator splitting scheme, and its symmetrized version, for a second order problem, written as a system of first order equations. The temporal accuracy of the schemes are analyzed and verified numerically by some simple test cases.

For the case of the two-dimensional scalar wave equation we have presented a novel symmetrized operator splitting scheme, that decouples the operator responsible for the wave propagation, and the operator that enforces the Dirichlet condition on the boundary of the scatterer. Once the operators are decoupled, numerical schemes that are best suited for each subproblem are used to discretize and solve the different subproblems. We use a mixed finite element scheme, based on the velocity-stress formulation of the wave equation, for the propagation of the wave, and a conforming finite element approach, utilizing the new fictitious domain method, for the enforcement of the Dirichlet condition on

the boundary of the scatterer. The use of the distributed Lagrange multiplier involves the solution of a saddle point problem at each time step. We solve this saddle point problem using a conjugate gradient algorithm due to R. Glowinski and P. LeTallec [65]. One of the main advantages of the Lagrange multiplier approach is that the stability condition (CFL) for the numerical scheme is the same as the condition in the absence of the scatterer [61]. However, for saddle point problems, certain inf-sup conditions have to be satisfied between the different finite element spaces employed for the approximation of the solution and the Lagrange multiplier [56].

In the mixed finite element scheme, the degrees of freedom for the solution and its time derivative, and the degrees of freedom for the gradient are staggered in space as well as in time. The finite element space that is used to approximate the gradient is the lowest order Nédélec space of linear edge elements in two-dimensions. The staggering of the spatial and temporal components of the degrees of freedom resembles the FDTD approach. In a numerical experiment of scattering by a disk we see a remarkable agreement between the plots of the exact solution of the problem, and the problem computed using the fictitious domain approach considering that the mesh is not modified locally to fit the boundary of the scatterer, as some other fictitious domain methods do. We have demonstrated the second order accuracy of these schemes with respect to time.

We have compared our results for a problem with 9 scattering obstacles to the result computed by [75], since for this problem we do not have an exact solution. In [75], the authors have considered a time-harmonic approach in which the mesh is locally fitted to the boundary of the obstacles. Again we see good agreement in the quality of both solutions.

As the scattering problem is an unbounded exterior problem, absorbing boundary conditions have been utilized to simulate the outgoing nature of propagating waves. We have used a first order absorbing boundary condition, the Sommerfeld condition for the wave equation and the Silver-Müller condition for Maxwell's equations. In both cases we have observed that the error due to the first order absorbing boundary condition dominates the

total error in numerical simulations. Thus, mesh refinement does not not help in improving the L^2 error between the computed and exact solutions.

To obtain more accurate absorbing boundary conditions, we have considered an absorbing layer based on Berenger's perfectly matched layer technique. We have presented a mixed finite element scheme for the numerical solution of the 2D TM mode of the uniaxial PML model for Maxwell's equations. In the proposed FEM, the underlying system of PDE's is first order in time as compared to the finite element implementation of the Zhao-Cangellaris's PML model [135], which has second order in time PDE's. On rectangles, the spatial discretization uses bilinear finite elements for the electric field, and the lowest order Raviart-Thomas divergence conforming elements for the magnetic field. Our method is a generalization of the FDTD scheme which can be interpreted as a mass-lumped FEM. In general finite element methods can model arbitrary complex geometrical structures effectively, whereas it is difficult to treat complex domains with the FDTD scheme.

We have proved energy estimates for the presented PML model under certain assumptions. The estimates and their proofs are analogous to those proved in [12] for the Zhao-Cangellaris's PML. We perform a dispersion analysis to compare the effects of dispersion in the FEM and the FDTD scheme. The analysis also shows that our numerical method is consistent with the continuous model. We have also analyzed the anisotropy present in the two models. From the analysis it is evident that the effects of dispersion can be reduced to any desired degree by considering a fine enough mesh. In the case of a finite absorbing layer, we have demonstrated the effects of terminating the PML for our method. As in the case of other discrete PML models, the reflection coefficient depends on the angle of incidence, and decreases as the length of the layer is increased. We have performed numerical experiments to compare our model with the split-field PML model of Berenger. Based on the energy estimates, the dispersion and reflection coefficient analysis of the mixed FEM, and our numerical results, we conclude that the proposed scheme has absorbing properties that are comparable to those of other PML models, including the Zhao-Cangellaris's PML

model, and Berenger's split field PML. The proposed mixed FEM can be extended to 3D by using a combination of Nédelec's elements and Nédelec-Raviart-Thomas elements for the discretization of the electric and magnetic fields, respectively.

We have incorporated the fictitious domain approach in the PML model for Maxwell's equations. We see a significant improvement in the L^2 error between the exact solution and the computed solution, as compared to the case of the Silver-Müller absorbing boundary condition, for the time-dependent problem of scattering by a disk. As the results show, a PML which is a quarter of a wavelength thick is sufficient to obtain good absorbing properties. The dominant error in this case is due to the discretization of the Lagrange multiplier. Thus, increasing the PML thickness does not improve the relative error between the exact and computed solutions. The fictitious domain method appears to have first order spatial accuracy based on our results.

For a one-dimensional wave problem, on the basis of a reflection coefficient analysis, we compare our new fictitious domain method and the symmetrized operator splitting scheme, with the FDTD scheme as well as another fictitious domain method based on a boundary Lagrange multiplier. We show the superiority of the fictitious domain methods as compared to the FDTD method. We also note that our fictitious domain method has certain desirable properties in common with the FDTD scheme, namely the lack of propagation of energy to the interior of the fictitious domain. The operator splitting scheme suffers from a phase shift that results in additional error in the reflection coefficient.

More analysis is needed to understand the behavior of the operator splitting methods. The extension of the fictitious domain method to the full Maxwell's equations is not straight forward. As our method is based on the use of the space of Q_1 bilinear finite elements on rectangles in two-dimensions, this method cannot be directly extended to the Nédélec edge elements in three dimensions. The analogous approach would be to consider nodal Lagrange elements on cubic meshes in three-dimensions. Future endeavors will be aimed at clarifying these ideas and performing detailed analysis of the methods considered here.

Bibliography

[1] S. Abarbanel and D. Gottlieb. A note on the leap-frog scheme in two and three dimensions. *J. Comput. Phys.*, 21:351–355, 1976.

[2] S. Abarbanel and D. Gottlieb. A mathematical analysis of the PML method. *J. Comput. Phys.*, 134:357–363, 1997.

[3] S. Abarbanel, D. Gottlieb, and J. S. Hesthaven. Well-posed perfectly matched layers for advective acoustics. *J. Comput. Phys.*, 154:266–283, 1999.

[4] J. C. Adam, A. G. Serveniere, J.-C. Nédélec, and P.-A. Raviart. Study of an implicit scheme for integrating Maxwell's equations. *Comput. Methods Appl. Mech. Engrg.*, 22(3):327–346, 1980.

[5] A. Ahland, D. Schulz, and E. Voges. Accurate mesh truncation for Schrödinger equations by a perfectly matched layer absorber: Application to the calculation of optical spectra. *Phys. Rev. B*, 60(8):109–112, August 1999.

[6] W. Andrew, C. Balanis, and P. Tirkas. A comparison of the Berenger perfectly matched layer and the Lindman higher-order ABC's for the FDTD method. *IEEE Microwave Guided Wave Lett.*, 5(6):192–194, June 1995.

[7] P. L. Arlett, A. K. Bahrani, and O. C. Zienkiewicz. Application of finite elements to the solution of Helmholtz's equation. *Proc. IEEE*, 115:1762–1766, 1968.

[8] F. Assous, P. Degond, E. Heintze, P.-A. Raviart, and J. Seger. On a finite element method for solving the three-dimensional Maxwell equations. *J. Comput. Phys.*, 109:222–237, 1993.

[9] C. Atamian, G. V. Dinh, R. Glowinski, J. He, and J. Périaux. On some imbedding methods applied to fluid dynamics and electromagnetics. *Comput. Methods Appl. Mech. Engrg.*, 91:1271–1299, 1991.

[10] C. Atamian and P. Joly. Une analyse de la méthode des domaines fictifs pour le probleme de Hemholtz extérieur. Technical Report 1378, INRIA, 1991.

[11] I. Babuska. The finite element method with Lagrangian multipliers. *Numer. Math.*, 20:179–192, 1973.

[12] E. Bécache and P. Joly. On the analysis of Berenger's perfectly matched layers for Maxwell's equations. *Math. Model. Numer. Anal.*, 36(1):87–119, 2002.

[13] E. Bécache, P. Joly, and C. Tsogka. An analysis of new mixed finite elements for the approximation of wave propagation problems. *SIAM J. Numer. Anal.*, 37(4):1053–1084, 2000.

[14] J. P. Berenger. *Actes du Colloque CEM, (Tregastel, France)*, 1983.

[15] J. P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 114(2):185–200, October 1994.

[16] J. P. Berenger. Three dimensional perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.*, 127:363–379, 1996.

[17] J. Blaschak and G. A. Kriegsmann. A comparative study of absorbing boundary conditions. *J. Comput. Phys.*, 77:109–139, July 1988.

[18] V. Bokil and M. Buksas. A 2D mixed finite element formulation of the uniaxial perfectly matched layer. *J. Comput. Phys.*, submitted.

[19] V. Bokil and R. Glowinski. A fictitious domain method with operator splitting for wave problems in mixed form. In G. C. Cohen, E. Heikkola, P. Joly, and P. Neittaanmäki, editors, *Proceedings of the Sixth International Conference on Mathematical and Numerical Aspects of Wave Propagation*. Springer-Verlag, June 2003, to appear.

[20] C. Börgers and O. B. Widlund. On finite element domain imbedding methods. *SIAM J. Numer. Anal.*, 27:963–978, 1990.

[21] A. Bossavit. Mixed finite elements and the complex of Whitney forms. In J. Whiteman, editor, *The Mathematics of Finite Elements and Applications*, Volume V1, pages 137–144. Academic Press, 1988.

[22] A. Bossavit and I. Mayergoyz. Edge elements for scattering problems. *IEEE Trans. Magn.*, 25:2816–2821, July 1989.

[23] A. Bossavit and J. C. Verité. The TRIFOU code: Solving the 3-D eddy currents problem by using H as the state variable. In *Digest of Summaries of the COMPUMAG Conference*, pages 286–290, Genoa, Italy, 1983.

[24] D. Braess. *Finite Elements : Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 2nd edition, April 2001.

[25] F. Brezzi. On the existence and uniqueness of saddle point problems arising from Lagrange multipliers. *RAIRO Modél. Math. Anal. Numer.*, 8-R2:129–151, 1974.

[26] L. Brillouin. *Wave Propagation and Group Velocity*. Pure Appl. Phys. Academic Press, New York and London, 1960.

[27] M. O. Bristeau, V. Girault, R. Glowinski, T. W. Pan, J. Périaux, and Y. Xiang. On a fictitious domain method for flow and wave problems. In R. Glowinski, editor, *Domain Decompostition Methods in Sciences and Engineering*, pages 361–386, 1997.

[28] A. C. Cangellaris and D. B. Wright. Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in FDTD simulations of electromagnetic phenomenon. *IEEE Trans. Antennas Propagat.*, 39:1518–1525, October 1991.

[29] Z. Cendes and M. Barton. New vector finite elements for three-dimensional magnetic computation. *J. Appl. Phys.*, 61:3919–3921, 1987.

[30] C. Cerjan, D. Kosloff, R. Kosloff, and M. Reshef. A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophys.*, 50:705–708, 1985.

[31] Z. Chen, J. Xu, and J. Ma. FDTD validations of a nonlinear PML scheme. *IEEE Microwave Guided Wave Lett.*, 9(3):93–95, March 1999.

[32] W. C. Chew, J.-M. Jin, C.-C. Lu, E. Michielssen, and J. M. Song. Fast solution methods in electromagnetics. *IEEE Trans. Antennas Propagat.*, 45(3):533–543, March 1997.

[33] W. C. Chew and W. H. Weedon. A 3D perfectly matched medium from modified Maxwells equations with stretched coordinates. *Microwave Opt. Technol. Lett.*, 7(13):599–604, September 1994.

[34] G. Cohen. *Higher-Order Numerical Methods for Transient Wave Equations*. Springer-Verlag, 2002.

[35] G. Cohen and S. Fauqueux. Mixed finite elements with mass-lumping for the transient wave equation. *J. Comput. Acoust.*, 8(1):171–188, 2000.

[36] G. Cohen, P. Joly, J. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.*, 38(6):2047–2078, 2001.

[37] G. Cohen and P. Monk. Gauss point mass lumping schemes for Maxwell's equations. *Numer. Methods Partial Differential Equations*, 14:63–88, 1998.

[38] G. Cohen and P. Monk. Mur-Nédélec finite element schemes for Maxwell's equations. *Computer Methods in Applied Mechanics and Engineering*, 169:197–217, 1999.

[39] R. Coifman, V. Rokhlin, and S. Wandzura. The fast multipole method for the wave equation: A pedestrian prescription. *IEEE Trans. Antennas Propagat.*, 35:7–12, 1993.

[40] F. Collino, S. Garcés, and P. Joly. A fictitious domain method for conformal modeling of the perfect electric conductors in the FDTD method. *IEEE Trans. Antennas Propagat.*, 46(10):1519–1526, October 1998.

[41] F. Collino, P. Joly, and F. Millot. Fictitious domain method for unsteady problems: Application to electromagnetic scattering. *J. Comput. Phys.*, 138:907–938, 1997.

[42] F. Collino and P. Monk. Optimizing the perfectly matched layer. *Comput. Methods Appl. Mech. Engrg.*, 164:157–171, 1998.

[43] R. L. Courant. Variational methods for the solution of problems of equilibrium and vibration. *Bull. Amer. Math. Soc.*, 5:1–23, 1943.

[44] W. Dahmen, T. Klint, and K. Urban. On fictitious domain formulations for Maxwell's equations. *Found. Comput. Math.*, 2003, to appear.

[45] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*, Volume 3. Springer-Verlag, Berlin Heidelberg, 1990.

[46] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*, Volume 1. Springer-Verlag, Berlin Heidelberg, 1990.

[47] E. Dean and R. Glowinski. Domain decompositions of wave problems using a mixed finite element method. In P. E. Bjorstad, M. S. Espedal, and D. Keyes, editors, *Ninth International Conference on Domain Decomposition Methods*, pages 326–333, 1998.

[48] E. Dean, R. Glowinski, and C. Li. Applications of operator splitting methods to the numerical solution of nonlinear problems in continuum mechanics and physics. In J. Goldstein, S. Rosencrans, and G. Sod, editors, *Mathematics Applied to Science*, pages 13–64. Academic Press, Boston, 1988.

[49] H. H. Dridi, J. S. Hesthaven, and A. Ditkowski. Staircase-free finite difference time-domain formulation for general materials in complex geometries. *IEEE Trans. Antennas Propagat.*, 49(5):749–756, 2001.

[50] M. Dryja. A capacitance matrix method for Dirichlet problem on polygonal region. *Numer. Math.*, 39:51–64, 1982.

[51] G. Duvaut and J.-L. Lions. *Inequalities in Mechanics and Physics*. Springer-Verlag, 1980.

[52] B. Engquist and A. Majda. Absorbing boundary conditions for numerical simulation of waves. *Math. Comp.*, 31(139):629–651, 1977.

[53] O. G. Ernst. A finite element capacitance matrix method for exterior Hemholtz problems. *Numer. Math.*, 75:175–204, 1996.

[54] G. Fairweather and A. Mitchell. A high accuracy alternating direction method for the wave equation. *J. Inst. Math. Appl.*, 1:309–316, 1965.

[55] S. A. Finogenov and Y. A. Kuznetsov. Two-stage fictitious components method for solving the Dirichlet boundary value problem. *Soviet J. Numer. Anal. Math. Modeling*, 3:301–323, 1988.

[56] M. Fortin and F. Brezzi. *Mixed and Hybrid Finite Element Methods*, Volume 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, July 1991.

[57] S. Garces. *Une méthode de domaines fictifs pour la modélisation des structures rayonnantes tridimensionnelles*. PhD thesis, Ecole Nationale Supérieure de Aéronautique et de Espace, 1997.

[58] S. D. Gedney. An anisotropic perfectly matched layer absorbing media for the truncation of FDTD lattices. *IEEE Trans. Antennas Propagat.*, 44(12):1630–1639, December 1996.

[59] S. D. Gedney and A. Roden. Applying Berenger's perfectly matched layer (PML) boundary condition to non-orthogonal FDTD analyses of planar microwave circuits. In *1995 URSI Radio Science Meeting Digest, Newport Beach, CA*, pages 18–23, June 1995.

[60] T. Geveci. On the application of mixed finite elements to the wave equation. *RAIRO. Modél. Math. Anal. Numer.*, 22(2):243–250, 1988.

[61] V. Girault and R. Glowinski. Error analysis of a fictitious domain method applied to a Dirichlet problem. *Japan J. Indust. Appl. Math.*, 12(3):487–514, October 1995.

[62] R. Glowinski. Numerical Methods for Fluids (Part 3). In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, Volume IX. North-Holland, Amsterdam, 2003, to appear.

[63] R. Glowinski, T. Hesla, D. D. Joseph, T. W. Pan, and J. Périaux. Distributed Lagrange multiplier methods for particulate flows. In M.-O. Bristeau, G. Etgen,

F. Fitzgibbon, J.-L. Lions, J. Périaux, and M. F. Wheeler, editors, *Computational Science for the 21st Century*, pages 270–279, 1997.

[64] R. Glowinski and Y. Kuznetsov. On the solution of the Dirichlet problem for linear elliptic operators by a distributed Lagrange multiplier method. *C. R. Acad. Sci. Paris Sér. I Math.*, 327:693–698, 1998.

[65] R. Glowinski and P. LeTallec. *Augmented Lagrangian and Operator Splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia, 1989.

[66] R. Glowinski, T. W. Pan, and J. Périaux. A fictitious domain method for Dirichlet problem and applications. *Comput. Methods Appl. Mech. Eng.*, 111:283–303, 1994.

[67] R. Glowinski, T. W. Pan, and J. Périaux. A fictitious domain method for external incompressible viscous flow modeled by Navier-Stokes equations. *Comp. Math. Appl Mech. Eng.*, 112:113–148, 1994.

[68] R. Glowinski, T. W. Pan, and J. Périaux. Distributed Lagrange multiplier methods for incompressible viscous flow around moving rigid bodies. *Comput. Methods Appl. Mech. Engrg.*, 151:181–194, 1998.

[69] R. Glowinski and Q. Tran. Constrained optimization in reflection tomography. *East-West J. Numer. Anal.*, 1:213–234, 1993.

[70] G. Golub and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press Ltd., London, third edition, 1996.

[71] D. Gottlieb. Strang-type difference schemes for multidimensional problems. *SIAM J. Numer. Anal.*, 9:650–661, 1972.

[72] R. F. Harrington. *Field Computation by Moment methods*. MacMillan, 1968.

[73] M. E. Hayder, F. Q. Hu, and M. Y. Hussaini. Towards perfectly absorbing conditions for Euler equations. In *AIAA 13th CFD Conference, Paper 97-2075*, 1997.

[74] J. He. *Méthodes de domaines fictifs en méchanique des fluides: Applications aux écoulements potentiels instationnaires autour d'obstacles mobiles*. PhD thesis, Université Paris VI, 1994.

[75] E. Heikkola, Y. A. Kuznetsov, P. Neittaanmäki, and J. Toivanen. Fictitious domain methods for the numerical solution of two-dimensional scattering problems. *J. Comput. Phys.*, 145:89–109, 1998.

[76] E. Heikkola, T. Rossi, P. Tarvainen, and Y. Kuznetsov. Efficient preconditioners based on fictitious domains for elliptic FE-problems with Lagrange multipliers. In H. G. Bock, F. Brezzi, R. Glowinski, G. Kanschat, Y. Kuznetsov, J. Périaux, and R. Rannacher, editors, *ENUMATH 97*, pages 646–661. World Scientific, 1998.

[77] R. L. Higdon. Numerical absorbing boundary conditions for the wave equation. *Math. Comp.*, 49(179):65–90, July 1987.

[78] R. Holland. Pitfalls of staircase meshing. *IEEE Trans. Electromagn. Compat.*, 35(4):434–439, November 1993.

[79] F. Q. Hu. On absorbing boundary conditions for linearized Euler equations by a perfectly matched layer. *J. Comput. Phys.*, 129:201–219, 1996.

[80] T. H. Hubing. Survey of numerical electromagnetic modeling techniques. Technical Report TR91-1-001.3, University of Missouri-Rolla, September 1991.

[81] B. N. Jiang, J. Wu, and L. A. Povinelli. The origin of spurious solutions in computational electromagnetics. *J. Comput. Phys.*, 125:104–123, 1996.

[82] D. Johnson, C. Furse, and A. Tripp. Application and optimization of the perfectly matched layer boundary condition for geophysical simulations. *Microwave Opt. Technol. Lett.*, 25(4):254–255, May 2000.

[83] R. Kosloff and D. Kosloff. Absorbing boundaries for wave propagation problems. *J. Comput. Phys.*, 63:363–376, 1986.

[84] H.-O. Kreiss and J. Lorenz. *Initial Boundary Value Problems and the Navier-Stokes Equations*, Volume 136 of *Pure Appl. Math.* Academic Press, Boston, USA, San Diego, CA, 1989.

[85] Y. A. Kuznetsov. Iterative analysis of finite element problems with Lagrange multipliers. In M.-O. Bristeau, G. Etgen, F. Fitzgibbon, J.-L. Lions, J. Périaux, and M. F. Wheeler, editors, *Computational Science for the 21st Century*, pages 170–178, 1997.

[86] J.-F. Lee, R. Lee, and A. Cangellaris. Time-domain finite element methods. *IEEE Trans. Antennas Propagat.*, 45(3):430–442, 1997.

[87] R. L. Lee and N. K. Madsen. A mixed finite element formulation for Maxwell's equations in the time-domain. *J. Comput. Phys.*, 88:284–304, 1990.

[88] Y. Liu. Fourier analysis of numerical algorithms for the Maxwell's equations. *J. Comput. Phys.*, 124:396–416, 1996.

[89] C. G. Makridakis and P. Monk. Time-discrete finite element schemes for Maxwell's equations. *Math. Model. Numer. Anal.*, 29(2):171–197, 1995.

[90] G. I. Marchuk. *Methods of Numerical Mathematics*. Springer-Verlag, New York, NY, 1975.

[91] G. I. Marchuk. Splitting and Alternating Direction Methods. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, pages 197–462. North-Holland, 1990.

[92] G. I. Marchuk, Y. A. Kuznetsov, and A. M. Matsokin. Fictitious domain and domain decomposition methods. *Sov. J. Numer. Anal. Math. Modelling*, 1:3–35, 1986.

[93] J. McMahon. Lower bounds for the electrostatic capacity of a cube. *Proc. Royal Irish Acad.*, 55(Sect. A):133–167, 1953.

[94] P. Mezzanotte and L. Roselli. Reply to comment on "a simple way to model curved metal boundaries in FDTD algorithm avoiding staircase approximation". *IEEE Microwave Guided Wave Lett.*, 6(4):184, April 1996.

[95] P. Mezzanotte, L. Roselli, and R. Sorrentino. A simple way to model curved metal boundaries in FDTD algorithm avoiding staircase approximation. *IEEE Microwave Guided Wave Lett.*, 5(8):267–269, August 1995.

[96] R. Mittra and U. Pekel. A finite element method frequency domain application of the perfectly matched layer (PML) concept. *Microwave Opt. Technol. Lett.*, 9(3):117–122, June 1995.

[97] R. Mittra and U. Pekel. Mesh truncation in the finite element frequency-domain method with a perfectly matched layer (PML) applied in conjunction with analytic and numerical absorbing boundary conditions. *Microwave Opt. Technol. Lett.*, 9(4):244–249, July 1995.

[98] R. Mittra and U. Pekel. A new look at the perfectly matched layer (PML) concept for the reflectionless absorption of electromagnetic waves. *IEEE Microwave Guided Wave Lett.*, 5(3):84–86, March 1995.

[99] P. Monk. A mixed method for approximating Maxwell's equations. *SIAM J. Numer. Anal.*, 28(6):1610–1634, December 1991.

[100] P. Monk. Analysis of a finite element method for Maxwell's equations. *SIAM J. Numer. Anal.*, 29(3):714–729, 1992.

[101] P. Monk. A comparison of three mixed methods for the time-dependent Maxwell's equations. *SIAM J. Sci. Stat. Comput.*, 13(5):1097–1122, September 1992.

[102] P. Monk. An analysis of Nédélec's method for the spatial discretization of Maxwell's equations. *J. Comput. Appl. Math.*, 47:101–121, 1993.

[103] P. Monk. A dispersion analysis of finite element methods for Maxwell's equations. *SIAM J. Sci. Comput.*, 15(4):916–937, 1994.

[104] P. Monk and K. Parrott. Analysis of finite element time domain methods in electromagnetic scattering. Technical Report 96/25, Oxford University Computing Laboratory, Numerical Analysis Group, Oxford, England OX1 3QD, December 1996.

[105] G. Mur. Absorbing boundary conditions for the finite difference approximation of the time domain electromagnetic field equations. *IEEE Trans. Electromagn. Compat.*, 23(4):377–382, 1981.

[106] G. Mur. Edge elements, their advantages and their disadvantages. *IEEE Trans. Magn.*, 30(5):3552–3557, September 1994.

[107] G. Mur. The fallacy of edge elements. *IEEE Trans. Magn.*, 34(5):3244–3247, September 1998.

[108] A. Nachman. Minireview: A brief perspective on computational electromagnetics. *J. Comput. Phys.*, 126:237–239, 1996.

[109] J. C. Nédélec. A new family of mixed finite elements in $\mathbb{R}^3$. *Numer. Math.*, 50:57–81, 1986.

[110] D. Peaceman and H. Rachford. The numerical solution of parabolic and elliptic differential equations. *J. Soc. Indust. Appl. Math.*, 3:28–41, 1955.

[111] G. Pelosi, R. Coccioli, and S. Selleri. *Quick finite elements for electromagnetic waves*, chapter 7. Artech House, June 1998.

[112] P. G. Petropoulos. Reflectionless sponge layers as absorbing boundary conditions for the numerical solution of Maxwell equations in rectangular, cylindrical and spherical coordinates. *SIAM J. Appl. Math.*, 60(3):1037–1058, 2000.

[113] P. G. Petropoulos, L. Zhao, and A. C. Cangellaris. A reflectionless sponge layer absorbing boundary condition for the solution of Maxwell's equations with higher order staggered finite difference schemes. *J. Comput. Phys.*, 139:184–208, 1998.

[114] C. M. Rappaport. Perfectly matched absorbing boundary conditions based on anisotropic lossy mapping of space. *IEEE Microwave Guided Wave Lett.*, 5(3):90–92, March 1995.

[115] C. M. Rappaport and L. Bahrmasel. An absorbing boundary condition based on an anechoic absorber for EM scattering computation. *J. Electromagn. Waves Appl.*, 6(12):1621–1234, 1992.

[116] P. A. Raviart and J. M. Thomas. A mixed finite element method for 2nd order elliptic problems. In I. Galligani and E. Magenes, editors, *Proc. of Math. Aspects on the Finite Element Method*, Volume 606 of *Lecture Notes in Mathematics*, pages 292–315. Springer-Verlag, 1977.

[117] V. Rokhlin. Rapid solutions of integral equations of classical potential theory. *J. Comput. Phys.*, 60:187–207, 1985.

[118] V. Rokhlin. Rapid solutions of integral equations of scattering theory in two dimensions. *J. Comput. Phys.*, 86(2):414–439, 1990.

[119] Z. S. Sacks, D. M. Kinsland, R. Lee, and J. F. Lee. A perfectly matched anisotropic absorber for use as an absorbing boundary condition. *IEEE Trans. Antennas Propagat.*, 43:1460–1463, 1995.

[120] V. K. Saul'ev. Solution of certain boundary-value problems on high-speed computers by the fictitious-domain method. *Sibirsk. Mat. Ž.*, 4:912–925, 1963.

[121] J. B. Schneider and K. L. Shlager. FDTD simulations of TEM horns and the implications for staircased representations. *IEEE Trans. Antennas Propagat.*, 45:1830–1838, December 1997.

[122] G. Strang. On the construction and comparison of difference schemes. *SIAM J. Num. Anal.*, 5:506–517, 1968.

[123] A. Taflove. Application of the finite difference time domain method to sinusoidal steady-state electromagnetic penetration problems. *IEEE Trans. Electromagn. Compat.*, 22:191–202, 1980.

[124] A. Taflove. *Advances in Computational Electrodynamics: The Finite-Difference Time-Domain method*. Artech House, Norwood, MA, 1998.

[125] F. Teixeria and W. Chew. A general approach to extend Berenger's absorbing boundary condition to anistropic and dispersive media. *IEEE Trans. Antennas Propagat.*, 46(9):1386–1387, September 1998.

[126] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM Rev.*, 24(2):113–136, April 1982.

[127] S. V. Tsynkov. Numerical solution of problems on unbounded domains. A review. Absorbing boundary conditions. *Appl. Numer. Math.*, 27(4):465–532, 1998.

[128] E. Turkel. Introduction to the special issue on absorbing boundary conditions. *Appl. Numer. Math.*, 27(4):327–329, 1998.

[129] E. Turkel and A. Yefet. Absorbing PML boundary layers for wave-like equations. *Appl. Numer. Math.*, 27:533–557, 1998.

[130] R. Vichnevetsky and J. Bowles. *Fourier Analysis of Numerical Approximations of Hyperbolic Equations*. SIAM Studies in Applied Mathematics. SIAM, 1982.

[131] G. S. Warren and W. R. Scott. An investigation of numerical dispersion in the vector finite element method using quadrilateral elements. *IEEE Trans. Antennas Propagat.*, 42(11):1502–1508, November 1994.

[132] T. Weiland. Comment on "a simple way to model curved metal boundaries in FDTD algorithm avoiding staircase approximation". *IEEE Microwave Guided Wave Lett.*, 6(4):183, April 1996.

[133] N. Yanenko. *The Method of Fractional Steps*. Springer-Verlag, Berlin, 1971.

[134] K. S. Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans. Antennas Propagat.*, 14:302–307, 1966.

[135] L. Zhao and A. Cangellaris. A general approach for the development of unsplit-field time-domain implementations of perfectly matched layers for FDTD grid truncation. *IEEE Microwave Guided Wave Lett.*, 6(5):209–211, May 1996.