Finite element approximation of Helmholtz problems with application to seismic wave propagation

Théophile Chaumont Frelet

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN

Normandie Université

THESE

Pour obtenir le grade de Docteur délivré par l'

INSA Rouen

Spécialité : Mathématiques Appliquées

Titre de la thèse :

Approximation par éléments finis de problèmes d'Helmholtz pour la propagation d'ondes sismiques

VERSION du 11/12/2015

Théophile CHAUMONT FRELET

Jury		
Civilité / Prénom NOM	Grade / Fonction / Statut / Lieu d'exercice	Rôle (Président, Rapporteur, Membre)
M. Rémi ABGRALL	Professeur, Université de Zurich	Rapporteur
M. Grégoire ALLAIRE	Professeur, Ecole Polytechnique	Rapporteur
Mme Hélène BARUCQ	Directeur de Recherche, INRIA	Directrice de Thèse
M. Henri CALANDRA	Ingénieur Docteur TOTAL	Examinateur
M. Christian GOUT	Professeur, INSA Rouen	Directeur de Thèse
Mme Carole LE GUYADER	Professeur INSA Rouen	Examinatrice
M. Jens Markus MELENK	Professeur, TU Wien	Rapporteur
M. Serge NICAISE	Professeur, Uni. de Valenciennes	Président du Jury

SPMII ECOLE DOCTORALE

informatics mathematics Inria

TOTAL

Depth Imaging Partnership DIP

LMI - EA 3226

Contents

General introduction

From radar or sonar detection to medical and seismic imaging, wave propagation is a complex physical phenomenon which is involved in a large number of applications : noninvasive methods have been developed since the end of the 19th century (works of Roentgen or Becquerel -X rays- in 1895 and 1896, works of Galton in 1883 -ultrasons- and so on), and these methods are always being developed, especially in medical imaging, seismic imaging, radar imaging, global seismology or in many other engineering applications.

Numerical simulations of waves deserve particular attention because they require applying advanced numerical methods in particular when the propagation domain is heterogeneous, as considered in this work.

Our main motivation is actually the design of advanced propagators which are meant to be used for solving seismic inverse problems. This is a very challenging purpose which is of great interest for oil companies. This PhD thesis has been prepared within the Inria project-team Magique-3D from Inria Bordeaux Sud-Ouest center in the framework of the joint Inria-Total research program DIP (Depth Imaging Partnership). Thus, the general context of this work is the development of an efficient solution methodology for computing Helmholtz solution in highly heterogeneous media as the Earth can be. Obviously, any progress in this topic may have nice consequences on other applications like for instance medial imaging but herein, we are interested in geophysical applications which provide us synthetic data for validating numerical methods.

Depending on the application, the simulation of high-frequency waves can be required, especially for high-resolution imaging. In this context, it is of course crucial to propose efficient numerical algorithms corresponding to (rigorous) PDE based modeling in order to work on 3D applications. Although realistic 3D full wave simulations become possible with the increase of computational power. It is worth noting that it is sometimes impossible to get numerical simulation on very complex cases because it requires too much computing capacity (even considering high performance computing).

In this work, we focus on efficient numerical algorithms for time-harmonic wave propagation in heterogeneous propagation media, modelled by the heterogeneous Helmholtz equation. We especially focus on the robustness of the algorithm with the respect to the frequency and to small scale heterogeneities in the propagation medium.

The lack of robustness of standard discretization methods (like low order finite element and finite difference methods) with respect to the frequency is common knowledge. Indeed, these methods fail to reproduce oscillations of high frequency solutions, unless a very fine

mesh is used, which is often unaffordable. This phenomenon is known as the "pollution effect" and has been extensively studied for wave propagation in homogeneous media.

However, though several solutions have been proposed to reduce the pollution effect in homogeneous media, much less work dealing with heterogeneous media is available. Beside, most of the ideas proposed to solve the problem in homogeneous media are not easy to extend to the heterogeneous case.

Among the methodologies developed to reduce the pollution effect in homogeneous media, we mention high order polynomial Finite Element methods (FEm) and Plane Wave methods (PWm). In both cases, highly heterogeneous media are not easily handled: high order polynomial FEm are based on meshes with a great number of degrees of freedom per cell. For this reason, these methods are usually based on rather coarse meshes with large cells. In this case, the parameters describing highly heterogeneous media can not be considered to be constant inside each cell as they might feature important variations.

The philosophy of PWm is to take advantage of a priori knowledge of the solution, which is expected to be a wave. In this regard, the discrete basis functions are taken as plane wave (or possibly Bessel functions and/or evanescent waves) instead of polynomials. More generally, the idea is to take homogeneous solutions of the PDE inside each cell as basis functions. This approach is very fruitful in homogeneous media, because analytical solutions are available as soon as the medium parameters are constant inside each mesh cell. Unfortunately, the requirement that the medium parameters are constant inside each mesh cell is sometime too restrictive when the medium is highly heterogeneous.

In this work, extensions of high order FEm and PWm for highly heterogeneous media have been considered:

- We propose to use high order FEm together with a second-level strategy to take into account small-scale heterogeneities. We call this approach the Multiscale Medium Approximation method (MMAm) [20, 50].

- We are also investigating (work in progress) a multiscale approach, where the basis functions are taken as local solutions of the PDE. This work is based on a multiscale method called the Multiscale Hybrid Mixed method (MHMm) originally developed for Darcy flow problems in highly heterogeneous media. Because the medium parameters are supposed to vary inside each mesh cell, there is no analytical expression of the basis functions, and a second-level numerical method is used to approximate the basis functions. This part of the work can be considered as an extension of PWm because the basis functions are taken as local solutions to the PDE. The difference lies in the fact that because the parameters are allowed to vary inside each mesh cell, the basis functions have no analytical expression and are approximated through a second-level approximation.

In this manuscript, we focus on the MMAm since it is very suitable to the industrial framework we are working with.

The problem of taking into account small scales heterogeneities on coarse meshes has been already investigated in [9, 13, 14, 17, 19, 57]. The problem is considered in the context

of elliptic equations applied to composite materials or porous media. In this context, the model problem is a second-order elliptic equation, with a highly oscillating heterogeneity parameter. It has been observed that for this model problem, standard polynomial discretizations fail to handle small scale variations of the heterogeneity parameter. Multiscale methods have thus been developed where the shape functions are taken as local solutions to the PDE. These methods include in particular the so-called multiscale finite element method [9,57], operator-based upscaling [14] and the Multiscale Hybrid Mixed method [13].

Another viewpoint is given by homogenization theory. The idea of the homogenization theory is to replace the original highly heterogeneous medium, by a simpler "homogenized" equivalent medium. While the method is very elegant and powerful, it relies on restrictive assumptions like scale separation and periodicity of the medium [86]. It is worth noting that homogenization can work under simpler assumptions (for instance: stochastic homogenization [64], two-scale convergence [8] or unfolding operators [33]). However, it turns out that these methodologies are not suited for our main application: seismic wave propagation.

Operator-based upscaling as well as multiscale finite elements have been recently applied to transient wave problems [2, 63, 101, 102]. Homogenization theory has been used in Geophysics as well to homogenize finely layered media [18,31]. More recently, a framework of "non-periodic" homogenization has been developed [27–29], again in the context of geophysical applications.

Based on the amount of works available for elliptic and transient wave propagation problems, we first though that it was mandatory to use a multiscale strategy, with special shape functions or some kind of homogenization strategy. However, if it is clear from the literature that standard polynomial discretizations can not handle highly heterogeneous media in the context of elliptic problems, we believe that it might not be the case for wave propagation problems.

It turns out that wave propagation problems are really different from elliptic problems. Indeed the error is mostly due to the pollution effect in the first case while the restrictive factor is the best-approximation error in the latter. Also, we have mostly focused on the acoustic wave equation with constant density as a model problem. In this model, the heterogeneities are "outside" the divergence operator and can be expected to be easier to handle than for the model Laplace problem, where the heterogeneous parameter is "inside" the divergence operator.

We believe that these two points are the reason why it is possible to use simple polynomial shape functions (leading to such good results), instead of local solutions.

After introducing the MMAm algorithm we propose, we will focus on carefully justifying the method both from the theoretical and numerical point of view. Theoretical analysis of the MMAm involves error estimates in 1D and 2D. The frequency appears explicitly in all the constants involved in error estimates. Numerical experiments include analytical test-cases in 1D and 2D and geophysical benchmarks in 2D and 3D. In order to tackle realistic geophysical test-cases, especially in 3D, an efficient and parallel algorithm of the method has been implemented. The most demanding numerical examples have been run on the CRIHAN super-computer facility.

In order to derive precise error-estimates for the MMAm, it is mandatory to show fine properties of the continuous solution. Thus, our work also includes a mathematical analysis of our model problem: the acoustic Helmholtz equation in heterogeneous media in 1D and 2D.

The main achievements of this work are:

- the derivation of frequency-explicitly stability estimates for heterogeneous Helmholtz problems,
- the design and the convergence analysis of the MMAm,
- the parallel implementation of the algorithm with openMPI,
- the validation of the method for the targeted applications.

The manuscript is organized as follows: a first chapter is devoted to the general setting of the seismic imaging framework, the derivation of the elastodynamic equation is given with suitable boundary conditions. The corresponding variational formulation is also presented. In Chapter 2, we focus on the analysis of the problem in 1D: the analysis of the continuous problem is precisely studied, then the discretization using the MMAm approach is given. In Chapter 3, we present the 2D problem since it is not possible to easily extend the results obtained in the 1D case. We first present a stability analysis for the Helmholtz problem set in an heterogeneous medium. We then show that it is possible to take the discontinuities of the velocity c into account on a coarse mesh by considering an approximation c_ϵ. We show that c_ϵ can be chosen to obtain a quadrature-like formula that can be mastered to ensure that the construction of the discrete system is cheap, and computational costs are then given. In Chapter 4, we give both analytical and geophysical numerical examples in 1D, 2D and 3D. In particular, we illustrate that the MMAm outperforms standard finite element approximations in highly heterogeneous media. A section in this chapter is also devoted to a comparison with homogenization.

We end up with a conclusion delivering some perspectives for the future.

Chapter 1

General setting

This work has been launched in the framework of a collaborative research program that is maintained between Inria and the oil company Total for the design of advanced software packages focusing on seismic imaging.

In a nutshell, the goal of seismic imaging is to produce a map of the underground based on surface data acquisition and using seismic wave reflections. It is worth noting that the concept is in fact general and not limited to underground imaging. For instance, radar imaging, medical imaging, or non-destructive testing are based on the same ideas.

The physical phenomenon that makes seismic imaging possible (or more generally, wave imaging) is the reflection of waves: when a wave front impinges an obstacle or a discontinuity, a part of the energy is scattered back to the emitter. This property makes it possible to emit waves from one location and receive information back in the same spot.

As a result, a crucial ingredient in seismic imaging is a good model to represent waves, taking into account reflection and refraction accurately. Several models have been proposed to represent seismic waves, ranging from asymptotic ray theory to poro-elasticity. Of course, there is a trade off between the accuracy of the model, and the computational requirements. Nowadays, full-wave modeling is possible, but still very costly to solve.

In the context of this PhD we propose to work on time-harmonic full-wave propagation in 2D and 3D isotropic acoustic and elastic media. We are focusing on optimizing numerical methods for the case of highly heterogeneous problems.

This chapter is devoted to a general setting of the key ingredient describing the context of this work. We give a brief overview of what seismic imaging is in Section 1.1, in particular, we illustrate two important seismic data processing tools in Sections 1.1.2 and 1.1.3: reverse-time migration and full waveform inversion.

Section 1.2 is devoted to the presentation of the two wave equation models considered in this work: acoustic and elastic isotropic wave propagation. For that purpose, we recall important notions of continuous mechanics in subsections 1.2.1, 1.2.2 and 1.2.3. We discuss how elasticity of the Earth can be represented using the generalized Hooke's law in subsection 1.2.4. The isotropic elastic wave equation is introduced in 1.2.5 and we derive its acoustic approximation in 1.2.6.

We present the boundary conditions used to close the problem in Section 1.3. In

particular, we give some details about non-reflecting boundary conditions in Subsection 1.3.2.

Section 1.4 ens up with the mathematical problems considered in the remaining of this work.

1.1 Seismic Imaging

The purpose of seismic imaging is to take advantage of seismic wave propagation to analyse the Earth subsurface. To start with, seismic data are gathered during a seismic campaign. Seismic waves are generated either by a vibrator truck for land acquisition or by an airgun for marine acquisition. Wave reflections are recorded by geophones or hydrophones located on a line containing the seismic source. Figure 1.1 illustrates the process.

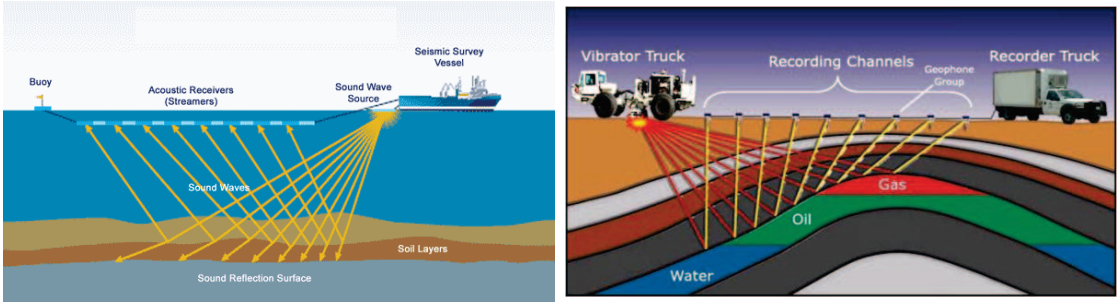


Figure 1.1: Offshore (left) and onshore (right) seismic acquisition

At this point, the huge amount of data recorded during the acquisition must be processed using high performance computing techniques. The aim is therefore to convert recorded seismic traces (see figures 1.2 and 1.4) into information about the subsurface (see figures 1.3 and 1.7). In the following, we present two examples of seismic data processing techniques: reverse time migration (RTM) and full waveform inversion (FWI). But before, let us say a few words about seismic wave modeling.

1.1.1 Seismic wave modeling

It is well-known that pseudo-elastic models are accurate to represent wave propagation inside the Earth. However, modeling a full elastic wavefield is computationally expensive. If realistic 3D elastic wave propagations become possible with nowadays computers, it is still a tricky task. This is the reason why several simplifications were introduced to help process the data.

The first simplification introduced is the acoustic approximation in which the Earth is considered as a fluid. Beside being less computationally demanding, the acoustic approximation is interesting because it only represents P-waves and is therefore easier to analyse.

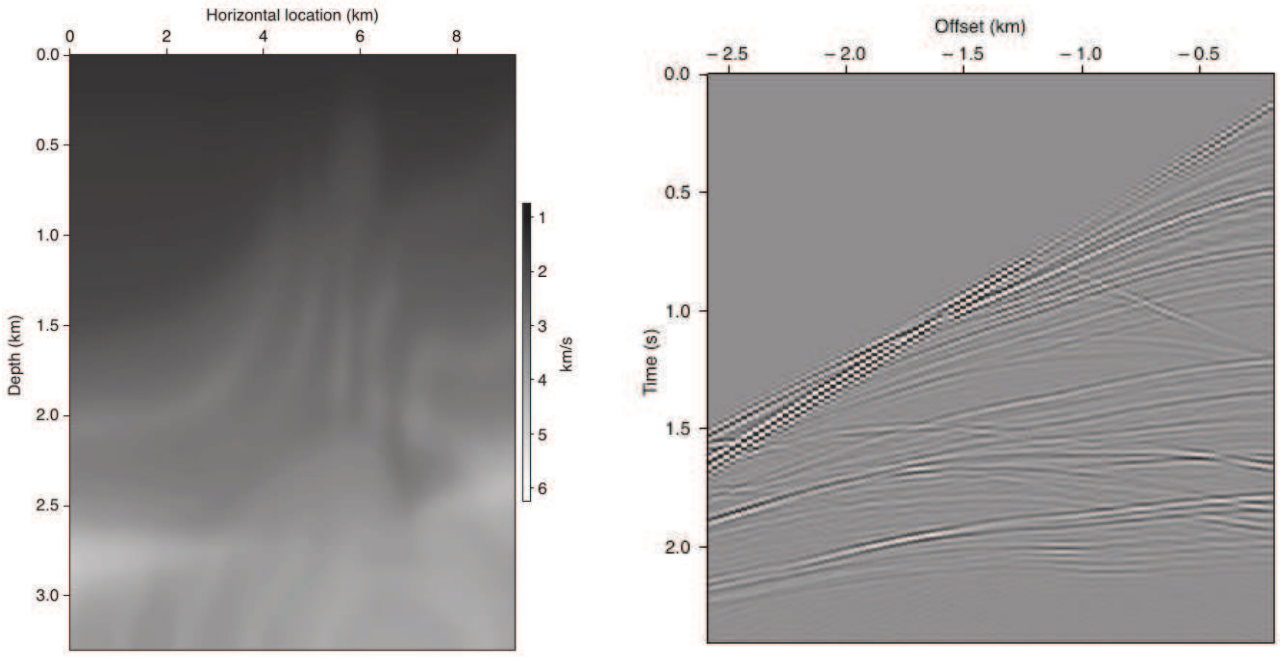


Figure 1.2: Input model (left) and example of seismic trace (right)

While 3D elastic modeling is still under development, 3D acoustic propagation is nowadays commonly used in industry. In this work, we develop optimized algorithms for elastic and acoustic wave propagation problems in their time-harmonic form.

In the past, fullwave modeling was considered to be too costly for practical purpose and additional simplifications have been introduced. For instance, ray methods are based on high-frequency approximations [45] and one-way methods use a splitting of the wave operators [83]. In the following, we focus on full-wave methods only because they are more accurate in heterogeneous media.

1.1.2 Reverse time migration

The concept of reverse time migration has been introduced by Clearbout [38]. It is a direct procedure which hints at recovering reflexivity of the underground based on a smooth initial model. We show an example presented in [93] from which we have extracted Figures 1.2 and 1.3.

Figure 1.2 shows the data available before the migration process. It consists of the seismic traces recorded in the acquisition and an input model generally obtained by travel-time analysis.

The output reflexivity model is compared to the ideal reflexivity model on Figure 1.7.

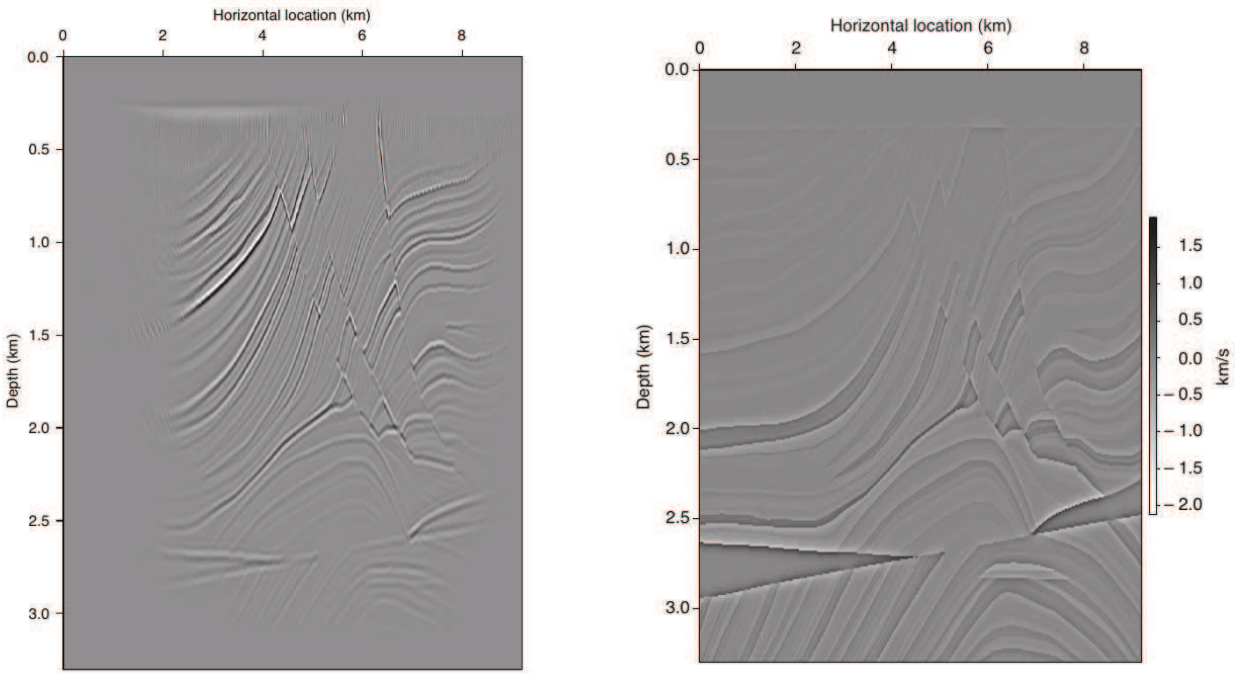


Figure 1.3: Output of the RTM (left) and true reflexivity (right)

1.1.3 Full waveform inversion

Full waveform inversion is an optimization procedure based on the minimization of a misfit function between the observation (seismic traces) and the simulated data. It is viewed as an inverse problem and for which wave propagation is the direct problem.

While RTM only aims at finding the location reflectors, full properties of the subsurface can, in theory, be recovered using FWI. We present an example from [91] (Figures 1.4, 1.5, 1.6 and 1.7).

In this example, seismic traces have been simulated using the Marmousi model [71] (see Figures 1.4 and 1.5).

The algorithm uses as input the seismic traces (Figure 1.4) and a smooth initial model constructed from travel-time analysis (Figure 1.6). The resulting velocity model is presented on Figure 1.7. We refer the reader to [94] for an illustration of initial model building using travel-time tomography.

1.2 Derivation of the elastodynamic equation

The main objective of this section is to introduce the time-harmonic wave equations which will be under study in the following. After presenting briefly some notions of continuous Mechanics, we derive the elastodynamic equations and then introduce their acoustic simplification. We also present the boundary conditions used to close the problems. In particular, we detail how non-reflecting boundary conditions are derived.

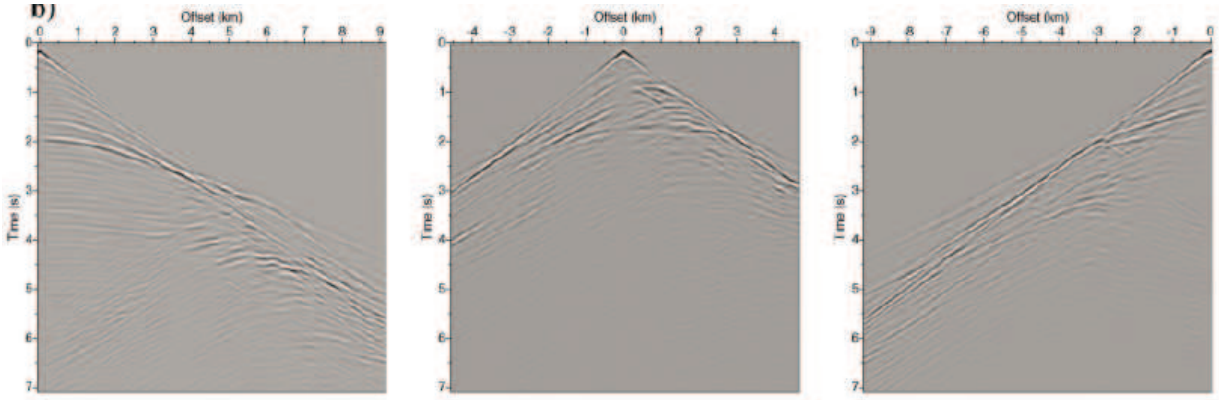


Figure 1.4: Example of seismic traces used in FWI

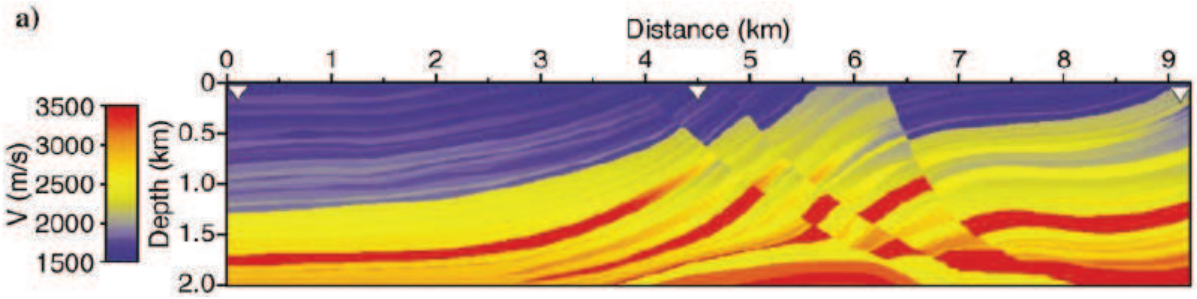


Figure 1.5: True model used to simulate traces

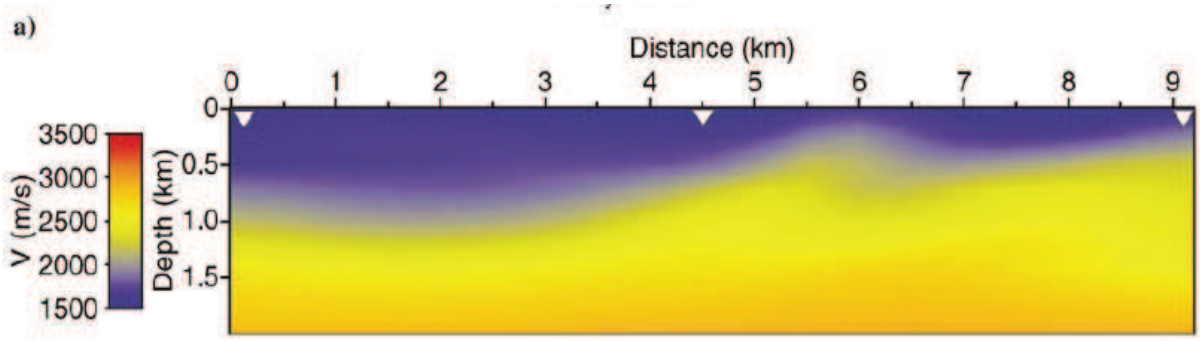


Figure 1.6: Input model for FWI

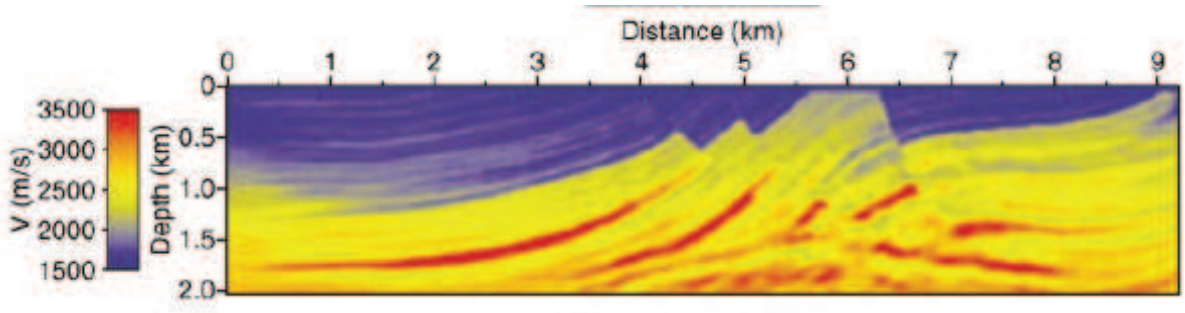


Figure 1.7: Output model from FWI

1.2.1 Lagrangian description of the underground

Seismic waves propagation is usually represented in the context of continuum Mechanics [48]. We describe the underground as a subset $\Omega \subset \mathbb{R}^3$ where the matter is represented as a continuum of particles $\mathbf{x} \in \Omega$. We call Ω the reference configuration of the underground.

To analyse the motion of the subsurface, we adopt the Lagrangian description, where we follow the particles. Hence, the space variable $\mathbf{x} \in \Omega$ denotes the position of a particle in the reference configuration Ω, and every particle in the subsurface is identified by a point $\mathbf{x} \in \Omega$.

We introduce the time variable t and assume that when $t = 0$, the Earth is in its reference configuration Ω. Under the action of external forces, the particles in Ω can move. Hence, at the time $t > 0$, a particle $\mathbf{x} \in \Omega$ can be located to the position $\mathbf{x}' \in \mathbb{R}^3$. To represent the motion of the particle, we introduce the displacement $\mathbf{u}$. Consider a particle $\mathbf{x} \in \Omega$. At the time t, the position $\mathbf{x}' \in \mathbb{R}^3$ of the particle $\mathbf{x}$ is given by

$$\mathbf{x}' = \mathbf{x} + \mathbf{u}(\mathbf{x}, t).$$

Hence, the path of a given particle $\mathbf{x} \in \Omega$ is given by the function $t \to \mathbf{x} + \mathbf{u}(\mathbf{x}, t)$. On the other hand, at a given time t the state of the ground is represented by the displacement of all particles, that is the function $\mathbf{x} \to \mathbf{x} + \mathbf{u}(\mathbf{x}, t)$.

Since we adopt the Lagrangian description, the position $\mathbf{x} \in \Omega$ identifies a given particle in the subsurface. It is natural to associate to each point $\mathbf{x} \in \Omega$ physical properties related to the nature of the rock. In the following we will see that the interesting quantities are the density ρ of the rock and its elastic properties defined by the stiffness tensor $\mathbf{C}$. Since these properties depend on the nature of the rock, they are naturally associated with the particle $\mathbf{x} \in \Omega$ rather than with the spatial position $\mathbf{x}' \in \mathbb{R}^3$.

Furthermore, we can assume that the density and the stiffness tensor do not depend on the time. Actually, these properties might change in time, but along a geological time scale which is much longer that the time scale of a wave propagation. Hence, the density in the Earth subsurface is represented by the function $\rho : \Omega \to \mathbb{R}$ and the elastic properties by the function $\mathbf{C} : \Omega \to \mathbb{R}^{36}$. If $\mathbf{x} \in \Omega$, $\rho(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$ represent then the density and the elastic properties of the particle $\mathbf{x}$.

To sum up, the properties of the Earth are characterized by the function $\rho : \Omega \to \mathbb{R}$ and $\mathbf{C} : \Omega \to \mathbb{R}^{36}$ and the motions of the ground are fully represented by the displacement function $\mathbf{u} : \Omega \times (0, T) \to \mathbb{R}^3$. Ω is the reference configuration of the Earth and the space variable $\mathbf{x} \in \Omega$ denotes a position in the reference configuration at $t = 0$ which identifies a single particle.

1.2.2 Representation of the deformation: the strain tensor

As explained in the previous section, the state of the underground at time t is represented by the displacement of all particles $\mathbf{x} \to \mathbf{u}(\mathbf{x}, t)$. In this section, we explain how we can relate the displacement of the particle to the deformation of the underground.

We first need to introduce the notion of rigid motion. A rigid motion is a displacement of the underground which does not deform it. To clarify this notion, consider two points $\mathbf{x}, \mathbf{y} \in \Omega$ and a fixed time $t > 0$. At time t, the particles $\mathbf{x}$ and $\mathbf{y}$ have moved to the positions $\mathbf{x}', \mathbf{y}' \in \mathbb{R}^3$ respectively, and we have

$$\mathbf{x}' = \mathbf{x} + \mathbf{u}(\mathbf{x}, t), \quad \mathbf{y}' = \mathbf{y} + \mathbf{u}(\mathbf{y}, t).$$

The distance between the particles $\mathbf{x}$ and $\mathbf{y}$ in the reference configuration Ω is given by $|\mathbf{x} - \mathbf{y}|$ and is modified to $|\mathbf{x}' - \mathbf{y}'| = |(\mathbf{x} - \mathbf{y}) + (\mathbf{u}(\mathbf{x}, t) - \mathbf{u}(\mathbf{y}, t))|$ at time t. At time t, we say that the ground has been subjected to a rigid motion if the distances between all particles have been preserved, that is

$$|\mathbf{x} - \mathbf{y}| = |(\mathbf{x} - \mathbf{y}) + (\mathbf{u}(\mathbf{x}, t) - \mathbf{u}(\mathbf{y}, t))|, \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Mathematically, it means that the function $\mathbf{x} \to \mathbf{x} + \mathbf{u}(\mathbf{x}, t)$ is an isometry.

It is intuitive to understand that rigid motions do not induce any deformation. Hence, we wish to distinguish between rigid motions, and true deformations. We introduce the Green-Saint Venant strain tensor $\mathbf{E}$ defined from the displacement $\mathbf{u}$ as

$$\mathbf{E}(\mathbf{u}) = \frac{1}{2}\left(\nabla \mathbf{u} + \nabla \mathbf{u}^T + \nabla \mathbf{u} \nabla \mathbf{u}^T\right). \tag{1.1}$$

It is possible to show that $\mathbf{E}(\mathbf{u}) = 0$ if and only if the function $\mathbf{x} \to \mathbf{u}(\mathbf{x}, t)$ is a rigid motion. We see that the Green-Saint Venant strain tensor $\mathbf{E}$ "filters out" the rigid motion. The strain tensor can be considered as a measure of the deformation. More precisely at the time t, the second order tensor $\mathbf{E}(\mathbf{x}, t)$ represents the deformation undergone by the underground in a neighborhood of the particle $\mathbf{x}$.

Remark 1. *It is possible to be more precise about how the distance and angle are locally modified. Distance and angle transformation are related to the Right Cauchy-Green deformation tensor [48]:*

$$\mathbf{C}(\mathbf{u}) = (\mathbf{I} + \nabla \mathbf{u})(\mathbf{I} + \nabla \mathbf{u})^T.$$

The formula

$$\mathbf{E}(\mathbf{u}) = \frac{1}{2}(\mathbf{C}(\mathbf{u}) - \mathbf{I})$$

shows that the Cauchy-Saint Venant tensor measures how much the distances and angle are modified.

Remark 2. *The Cauchy-Saint Venant strain tensor is not the only possible choice to measure the deformations. Other definitions of strain are possible. However, they all result in the same linearized strain tensor, and are equivalent for small deformations. For more details about the different measures of strain, we refer the reader to [56] and [48].*

Because of the last term in (1.1) the relation between the displacement $\mathbf{u}$ and the Cauchy-Saint Venant strain tensor is non-linear. This non-linearity makes the analysis of the strain very complex. Fortunately, if we assume that the deformations are small enough, we can linearize (1.1) and define the linearized strain tensor

$$\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2}\left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right). \tag{1.2}$$

The hypothesis of small deformations is reasonable for the applications we are targeting. Therefore, we use the linearized strain tensor $\boldsymbol{\varepsilon}(\mathbf{u})$ as a measure of deformation.

1.2.3 Representation of the internal forces: the stress tensor

When an external force is applied to the Earth, internal forces appear as a reaction. These forces are represented by a contact force characterized by a stress vector $\mathbf{t}$. If we consider a portion $A \subset \Omega$ of the underground, the stress in A at time t can be represented by

$$\int_{\partial A} \mathbf{t}_{\partial A}(\mathbf{s}, t) d\mathbf{s},$$

where the stress vector $\mathbf{t}_{\partial A} : \partial A \to \mathbb{R}^3$ depends on the shape of A.

The dependency of $\mathbf{t}_{\partial A}$ on ∂A can be fairly general. To simplify the situation, we introduce the Cauchy-Euler postulate and assume that the stress vector depends on the shape of A only through the unit normal vector. Hence, we have $\mathbf{t}_{\partial A}(\mathbf{s}, t) = \mathbf{t}(\mathbf{s}, t, \mathbf{n}_A)$, where $\mathbf{n}_A$ is the unit normal vector on ∂A. Under the Cauchy-Euler postulate, the stress can be represented by a function, $\mathbf{t} : \Omega \times (0, T) \times \mathbb{S}^2 \to \mathbb{R}^3$, called a Cauchy stress vector, such that the stress in A is

$$\int_{\partial A} \mathbf{t}(\mathbf{s}, t, \mathbf{n}_A) d\mathbf{s},$$

for $A \subset \Omega$. Furthermore, it is possible to show that there exists a second order symmetric tensor $\boldsymbol{\sigma}$, called the Cauchy stress tensor such that

$$\mathbf{t}(\mathbf{x}, t, \mathbf{n}) = \boldsymbol{\sigma}(\mathbf{x}, t)\mathbf{n}$$

Remark 3. *The existence of the Cauchy stress tensor $\boldsymbol{\sigma}$ is the result of the Cauchy fundamental theorem. The demonstration uses Newton's second law of motion and the conservation of moment [48].*

We can interpret the stress as a volumic force using the Stoke's formula. Indeed, if $A \subset \Omega$, we have

$$\int_{\partial A} \boldsymbol{\sigma}(\mathbf{s},t)\mathbf{n}_A d\mathbf{s} = \int_A \operatorname{div} \boldsymbol{\sigma}(\mathbf{x},t) d\mathbf{x}. \tag{1.3}$$

Since (1.3) is valid for any regular subset $A \subset \Omega$, we conclude that the stress can be represented by the force density $\mathbf{f}_{int} : \Omega \times (0,T) \to \mathbb{R}^3$:

$$\mathbf{f}_{int} = \operatorname{div} \boldsymbol{\sigma}. \tag{1.4}$$

1.2.4 Constitutive equation: the generalized Hooke's law

We have explained that in the context of continuum Mechanics, small deformations of the Earth are represented by the linearized strain tensor $\boldsymbol{\varepsilon}(\mathbf{u})$. These deformations result in internal forces represented by the stress tensor $\boldsymbol{\sigma}$

This framework is quite general and makes it possible to represent different behaviours of the considered material (namely plasticity, viscosity, and elasticity) depending on how the stress tensor $\boldsymbol{\sigma}$ is related to the displacement $\mathbf{u}$ [48]. The equation relating the stress to the displacement is called the constitutive equation. Of course, the constitutive equation depends on the considered material, but also on which feature we want to model.

In the context of off-shore seismic campaigns, the materials we have to consider are water and water-saturated porous rock. An accurate poro-elastic model for wave propagation has been developed by Biot [23, 24]. The poro-elastic constitutive equation of Biot takes into account elasticity, viscosity and porosity of the rocks. This model can be simplified and, for most target applications, we only need to consider elastic and viscous properties of rocks [53, 54]. Also sound wave in the water can be treated as a special case of elasticity.

In the following, we consider the elasticity of the rocks only, it means that the stress $\boldsymbol{\sigma}(\mathbf{x},t)$ at the particle $\mathbf{x}$ at time t only depends on the strain $\boldsymbol{\varepsilon}(\mathbf{x},t)$ at the particle $\mathbf{x}$ at time t. This hypothesis prevents nonlocal and memory effects as well as absorption.

Furthermore, since we are interested in small deformations, we can focus on the simplest case of elasticity: linear elasticity. In this context, the relation between the stress and strain is linear and the constitutive equation is the so-called generalized Hooke's law. The stress tensor is determined from the strain by the stiffness tensor $\mathbf{C}$:

$$\boldsymbol{\sigma}(\mathbf{x},t) = \mathbf{C}(\mathbf{x})\boldsymbol{\varepsilon}(\mathbf{x},t). \tag{1.5}$$

Remark 4. *Anelastic properties of the Earth, in particular attenuation, are shown to be of importance in seismic data processing [54, 62]. Attenuation is complex to represent in time domain modeling. In time-domain numerical simulations, dedicated "memory variables" [30] need to be added to numerical schemes which increase the computational cost. However, attenuation is naturally handled in frequency domain formulations by considering a complex frequency [43] or complex Lamé parameters [53]. Since we are focusing on frequency domain simulations, we do not consider attenuation as a major difficulty in terms of numerical approximation and drop it to simplify.*

The elastic properties of the Earth at the particle $\mathbf{x} \in \Omega$ are completely determined by the stiffness tensor $\mathbf{C}(\mathbf{x})$. Since the strain and stress tensor are symmetric, they both have 6 independent coefficients. Therefore, in the general case, the stiffness tensor is characterized by $6 \times 6 = 36$ coefficients. To simplify the rest of the presentation, we adopt the Voigt notation. The stiffness tensor is represented as a 6×6 matrix $\mathbf{C} = \{\mathbf{C}_{\alpha\beta}\}_{\alpha,\beta=1}^{6}$, where $\mathbf{C}_{ijkl} = \mathbf{C}_{\alpha\beta}$ with the convention

$$\begin{array}{ccc} ij & \leftrightarrow & \alpha \\ 11 & \leftrightarrow & 1 \\ 22 & \leftrightarrow & 2 \\ 33 & \leftrightarrow & 3 \\ 32 = 23 & \leftrightarrow & 4 \\ 31 = 13 & \leftrightarrow & 5 \\ 21 = 12 & \leftrightarrow & 6. \end{array}$$

Though the 36 coefficients are independent in the general case, it is possible to greatly simplify the stiffness tensor under the assumption of isotropy. A material is said to be isotropic if it has the same elastic properties in every direction. It means that waves propagate with the same speed in every direction. In the isotropic case, they are only two independent coefficients C_{33}, C_{44} and we have

$$\mathbf{C} = \begin{bmatrix} C_{33} & (C_{33}-2C_{44}) & (C_{33}-2C_{44}) & 0 & 0 & 0 \\ (C_{33}-2C_{44}) & C_{33} & (C_{33}-2C_{44}) & 0 & 0 & 0 \\ (C_{33}-2C_{44}) & (C_{33}-2C_{44}) & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{bmatrix}. \tag{1.6}$$

We can express the coefficients C_{33} and C_{44} in terms of the Lamé parameters λ and μ. In this case, we have

$$C_{33} = \lambda + 2\mu$$

and

$$C_{44} = \mu.$$

For isotropic media, we can rewrite the Hooke's law in a simpler form with the Lamé parameters:

$$\boldsymbol{\sigma} = \lambda \operatorname{div} \mathbf{u}\mathbf{I} + 2\mu\boldsymbol{\varepsilon}(\mathbf{u}). \tag{1.7}$$

It is also possible to express the stiffness tensor as

$$\mathbf{C} = \rho \begin{bmatrix} c_p^2 & c_p^2-2c_s^2 & c_p^2-2c_s^2 & 0 & 0 & 0 \\ c_p^2-2c_s^2 & c_p^2 & c_p^2-2c_s^2 & 0 & 0 & 0 \\ c_p^2-2c_s^2 & c_p^2-2c_s^2 & c_p^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_s^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_s^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & c_s^2 \end{bmatrix}, \tag{1.8}$$

where

$$c_p = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad c_s = \sqrt{\frac{\mu}{\rho}}, \tag{1.9}$$

are the compressional and shear wave velocities.

Remark 5. *The hypothesis of isotropy is not always sufficient to represent the elastic properties of the Earth. This might be for two main reasons. First some rocks have an intrinsic anisotropy, because of a preferred orientation of minerals [100]. Second, when a large enough wavelength travels through a finely layered medium, it appears as an homogeneous anisotropic, transverse isotropic, medium [18, 31]. More generally, homogenization of (not necessary layered) isotropic media usually results in anisotropic effective representation [27–29].*

When the medium is anisotropic, the wavespeed varies along the direction of propagation. Hence, wavefronts in homogeneous anisotropic media are not circles, but can have general shapes. In geophysical applications, the rocks have special form of anisotropy. In the worst case, there are orthotropic, but they are currently mostly represented as transverse isotropic media [25].

In transverse isotropic media, there exist one axis of symmetry, which is a preferred direction. When the preferred direction is the depth, we speak about vertical transverse isotropic (VTI) medium. In a VTI medium, the stiffness tensor contains five independent coefficients $C_{11}, C_{33}, C_{44}, C_{66}, C_{13}$ and has the following shape

$$\mathbf{C} = \begin{bmatrix} C_{11} & (C_{11}-2C_{66}) & C_{13} & 0 & 0 & 0 \\ (C_{11}-2C_{66}) & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{bmatrix}. \tag{1.10}$$

The coefficients of the stiffness tensor are usually expressed using the vertical P and S velocity c_p, c_s and three measures of anisotropy proportion ϵ, γ and δ introduced by Thomsen [98].

The general transverse isotropic case, or tilted transverse isotropic case (TTI), is obtained from the VTI case through a rotation.

1.2.5 Elastodynamic equations

We are now ready to derive the elastodynamic equations by applying Newton's second law of motion. We assume that an external force of volumic density $\mathbf{f}(\mathbf{x}, t)$ is applied to the Earth. Recalling (1.4) and (1.5), we also take into account the internal forces which are considered locally as

$$\mathbf{f}_{int} = \operatorname{div} \boldsymbol{\sigma} = \operatorname{div} \mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}). \tag{1.11}$$

Since the density $\rho = \rho(\mathbf{x})$ is supposed to be constant in time, Newton's second law of motion in its local form yields

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \mathbf{f}_{int} + \mathbf{f},$$

and we immediately obtain the elastodynamic equations formulated in displacement by applying (1.11)

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - \operatorname{div} \mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) = \mathbf{f}. \tag{1.12}$$

Equation (1.12) is widely used in Geophysics to model the propagation of seismic waves [30,103]. It is also used in other fields such as structural vibration analysis, non-destructive testing and fluid-solid interaction.

Remark 6. *It is possible to obtain other formulations of the elastodynamic equations by substituting variables. For instance, the velocity-stress formulation is very popular for time domain algorithms [103]. However, equation* (1.12) *usually leads to numerical schemes with less degrees of freedom.*

In the isotropic case, equation (1.12) leads to

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - \operatorname{div} \left(\lambda \operatorname{div}(\mathbf{u})\mathbf{I} + 2\mu\boldsymbol{\varepsilon}(\mathbf{u})\right) = \mathbf{f}. \tag{1.13}$$

Remark 7. *We can easily characterize P and S waves assuming that the medium is homogeneous and that there is no external sources. If the medium is homogeneous, ρ, c_p and c_s do not depend on the space variable* $\mathbf{x}$ *and* (1.13) *leads to*

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - (\lambda + 2\mu)\nabla \operatorname{div} \mathbf{u} + \mu \operatorname{curl} \operatorname{curl} \mathbf{u} = 0. \tag{1.14}$$

If we further write $p = \operatorname{div} \mathbf{u}$ and $S = \operatorname{curl} \mathbf{u}$ and divide (1.14) *by ρ, we obtain*

$$\frac{\partial^2 \mathbf{u}}{\partial t^2} - \frac{\lambda + 2\mu}{\rho}\nabla p - \frac{\mu}{\rho} \operatorname{curl} S = 0. \tag{1.15}$$

Taking the divergence of (1.15)*, we see that*

$$\frac{\partial^2 p}{\partial t^2} - \frac{\lambda + 2\mu}{\rho}\Delta p = 0. \tag{1.16}$$

On the other hand, taking the rotational of (1.15)*, we obtain*

$$\frac{\partial^2 S}{\partial t^2} - \frac{\mu}{\rho} \operatorname{curl} \operatorname{curl} S = 0,$$

further more, since $S = \operatorname{curl} \mathbf{u}$, $\operatorname{div} S = \operatorname{div} \operatorname{curl} \mathbf{u} = 0$. Therefore

$$\operatorname{curl} \operatorname{curl} S = \nabla \operatorname{div} S - \Delta S = -\Delta S,$$

and

$$\frac{\partial^2 S}{\partial t^2} - \frac{\mu}{\rho}\Delta S = 0,$$

or

$$\frac{\partial^2 S_n}{\partial t^2} - \frac{\mu}{\rho}\Delta S_n = 0, \tag{1.17}$$

for $1 \leq n \leq 3$.

We identify the P and S wave velocities in the wave equations (1.16) *and* (1.17) *respectively:*

$$c_p^2 = \frac{\lambda + 2\mu}{\rho}, \quad c_s^2 = \frac{\mu}{\rho}.$$

A plane wave analysis easily reveals the polarization of each mode. P waves are longitudinal (and therefore, compressional) waves while S waves are transverse (and therefore, shear) waves.

1.2.6 Acoustic approximation

The propagation of both P and S waves is modeled by elastodynamic equation (1.12). Since P waves are travelling faster than S waves, they bring information first. Some applications, like first time arrival computations, require P waves only. Another example is the reverse time migration where P and S modes require to be filtered to obtain noise-free images. It is therefore of interest to introduce a simpler model, called the acoustic approximation, in which P waves are considered only. Acoustic approximation is interesting because it is simpler to analyse and cheaper to compute than the elastic model.

Actually, the elastic model requires the computation of the displacement, which is a vectorial unknown. In the acoustic approximation, it is possible to introduce the pressure as an auxiliary unknown. The pressure is a scalar quantity, so that three times less unknowns are required for the acoustic approximation on the same mesh.

Besides, the speed difference between P and S waves also plays an important role because numerical methods are conditioned by the smallest wavelength in the simulation. As described in [32], the ratio c_p/c_s ranges from 1.5 to 1.7 in sandstones. In clays, the ratio is at least 2, and can be much greater. An example of values in water-saturated Berea sandstone is $c_p = 3888$ m.s^{-1} and $c_s = 2302$ m.s^{-1} ($c_p/c_s \simeq 1.69$) [32]. Another example of velocity values for silty clay is $c_p = 1519$ m.s^{-1} and $c_s = 287$ m.s^{-1} ($c_p/c_s \simeq 5.29$) [53]. As a result, S wavelength is at least 1.5 times smaller than P wavelength and therefore, the elastic model requires a finer mesh.

We refer the reader to [7, 80, 108] for a more general discussion on the usability and computational trade-off of the acoustic approximation.

Remark 8. *In the following, we will only consider acoustic approximation of isotropic media. However, it is possible to derive acoustic approximation of anisotropic media [7, 80, 108].*

In the isotropic case the resulting equation (1.24) *is the "true" equation that represents acoustic waves in an ideal fluid (for instance, water or air can be considered as ideal fluids depending on the application). The equation is second order in time and space and has the pressure as unique scalar unknown. In this case, we can say the Earth is considered as an ideal fluid, since equation* (1.24) *actually models sound waves in ideal fluid. We also speak of "acoustic Earth".*

On the other hand, acoustic approximations of anisotropic media do not describe a physical phenomenon, and are very different from acoustic wave equation (1.24)*. Actually, acoustic media can not be anisotropic by nature [7]. Furthermore, the resulting equation is more than second order in time and space [7], or has at least two unknowns [80, 108].*

In the isotropic case, we simply obtain the acoustic approximation by removing shear waves. This is done by setting the shear wave velocity to zero. Hence, we consider that $c_s = 0$ in (1.8). We introduce the bulk modulus $\kappa = \rho c_p^2$ and obtain

$$\mathbf{C} = \rho \begin{bmatrix} c_p^2 & c_p^2 & c_p^2 & 0 & 0 & 0 \\ c_p^2 & c_p^2 & c_p^2 & 0 & 0 & 0 \\ c_p^2 & c_p^2 & c_p^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \kappa & \kappa & \kappa & 0 & 0 & 0 \\ \kappa & \kappa & \kappa & 0 & 0 & 0 \\ \kappa & \kappa & \kappa & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{1.18}$$

We can greatly simplify Hooke's law using (1.18). Indeed, since $\mathbf{C}$ has a simple form, it holds that

$$\boldsymbol{\sigma} = \mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u}) = \kappa \operatorname{div}(\mathbf{u})\mathbf{I}. \tag{1.19}$$

At this point, it is clear that the stress tensor has one single independent component: the pressure $p = \kappa \operatorname{div}(\mathbf{u})$. We obtain a simple expression for $\boldsymbol{\sigma}$ and the internal forces,

$$\boldsymbol{\sigma} = p\mathbf{I}, \quad \operatorname{div} \boldsymbol{\sigma} = \nabla p \tag{1.20}$$

and (1.12) simplifies to

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} - \nabla p = \mathbf{f}. \tag{1.21}$$

We obtain the acoustic formulation in pressure by substituting the displacement $\mathbf{u}$ for the pressure in (1.21). We first divide (1.21) by ρ and take the divergence:

$$\operatorname{div} \frac{\partial^2 \mathbf{u}}{\partial t^2} - \operatorname{div} \left(\frac{1}{\rho} \nabla p \right) = \operatorname{div} \left(\frac{1}{\rho} \mathbf{f} \right). \tag{1.22}$$

We can now substitute the displacement for the pressure using the Schwarz' theorem on cross derivatives. By definition of the pressure p, it holds that

$$\operatorname{div} \frac{\partial^2 \mathbf{u}}{\partial t^2} = \frac{1}{\kappa} \frac{\partial^2 p}{\partial t^2}. \tag{1.23}$$

We obtain the acoustic pressure formulation by plugging (1.23) in (1.22):

$$\frac{1}{\kappa}\frac{\partial^2 p}{\partial t^2} - \operatorname{div}\left(\frac{1}{\rho}\nabla p\right) = \operatorname{div}\left(\frac{1}{\rho}\mathbf{f}\right). \tag{1.24}$$

Acoustic approximation (1.24) is widely used in geophysical applications. It is of particular importance in the context of inverse problems because there are less parameters to identify in the model. For instance, equation (1.24) is the model studied by Tarantola in [95] in the context of full waveform inversion or by Clayton et Al. in [37] for Born-WKBJ inversion. An acoustic approximation for TTI media is also used as a direct propagator by Virieux et al. [51]. Acoustic approximations are also used for reverse-time migration because kinematic information only is important for this application [80] (see Remark 9). We refer the reader to the work of Zhang et al. [107] for reverse time migration using a TTI acoustic approximation.

Equation (1.24) is also the natural equation to model sound waves in ideal fluids. In ideal fluids, there is no shear, so that expression (1.18) of the stiffness tensor is directly obtained from physical consideration (while it is obtained from an approximation in Geophysics). Hence equation (1.24) is used in Geophysics to model sound waves in the water during off-shore campaign.

Remark 9. *In Remark 7, we observed that P and S waves can be easily separated in homogeneous media. Furthermore, P-wave equation* (1.16) *is actually nothing but equation* (1.24)*. Hence, in the acoustic approximation of homogeneous media, S waves are filtered out and compressional waves are computed using the pressure p as the unknown.*

Unfortunately, this is not the case in heterogeneous media. From the mathematical point of view, the simplification (1.14) *does not hold if λ and μ depend on the space variable $\mathbf{x}$. If we consider piecewise constant Lamé parameters λ and μ, it is still possible to write* (1.14) *locally and we can derive compatibility conditions at the interface between each subdomain. It turns out that these compatibility conditions couple P and S waves at each interface.*

From the physical point of view, we speak about PS converted waves: when a reflection occurs, conversions between P and S modes happen (for instance, see [40]). For instance, when a P wave front is reflected in an elastic medium, the energy is separated into P and S fronts.

In the acoustic simplification, when a reflection occurs, the whole P energy is conserved and there is no energy conversion. As a result, the acoustic approximation yields the correct first arrival travel time (because most of P wave fronts are correctly represented) but amplitudes might be wrong because the whole energy is maintained at every reflection. Therefore, it is usually said that acoustic approximations correctly represent the kinematic of P waves, but do not preserve the amplitudes [80].

Acoustic approximation are widely used in reverse time migration algorithms, because the kinematic is the only information used by this application [80]. Actually, special dedicated imaging conditions where P and S mode are separated need to be applied for elastic imaging (see [69] \$3.1 or [92]).

Acoustic approximation (1.24) is sometimes further simplified by assuming that the density is constant. Multiplying (1.24) by (the constant value) ρ, we obtain

$$\frac{1}{c_p^2}\frac{\partial^2 p}{\partial t^2} - \Delta p = \operatorname{div} \mathbf{f}. \tag{1.25}$$

As an example of application, equation (1.25) is used as a propagator in [93] to achieve reverse time migration. We will mostly focus on the time-harmonic version of equation (1.25) in Chapter 2 and 3.

Remark 10. *Like in Remark 9, we can say that approximation* (1.25) *preserves kinematic but changes amplitudes. For this reason, constant density acoustic equation is used in the context of migration techniques.*

1.2.7 Time-harmonic formulation

Wave equations (1.12) and (1.24) are formulated in time-domain. It means that the unknown depends on the space variable $\mathbf{x}$ and the time variable t. By introducing the Fourier transform $\mathcal{F}$, we can describe equations (1.12) and (1.24) in the frequency domain, where the unknown depends on the space variable $\mathbf{x}$ and the dual variable of time t: the so-called (angular) frequency ω.

If ϕ is a sufficiently regular function, we introduce its Fourier transform $\mathcal{F}(\phi)$ defined by

$$\mathcal{F}(\phi)(\mathbf{x},\omega) = \int_{-\infty}^{+\infty} \phi(\mathbf{x},t) e^{-\mathbf{i}\omega t} dt$$

An important property of the Fourier transform is that if ϕ is regular enough, we have

$$\mathcal{F}\left(\frac{\partial \phi}{\partial t}\right) = -\mathbf{i}\omega \mathcal{F}(\phi). \tag{1.26}$$

Consider a fix angular frequency $\omega > 0$ and define $\hat{\mathbf{u}}_\omega(\mathbf{x}) = \mathcal{F}(\mathbf{u})(\mathbf{x},\omega)$, $\hat{p}_\omega(\mathbf{x}) = \mathcal{F}(p)(\mathbf{x},\omega)$ and $\hat{\mathbf{f}_\omega}(\mathbf{x}) = \mathcal{F}(\mathbf{f})(\mathbf{x},\omega)$. Then, because of (1.26), $\hat{\mathbf{u}}_\omega$ and $\hat{p}_\omega$ are solutions to the time-harmonic equations

$$-\omega^2 \rho \hat{\mathbf{u}}_\omega - \operatorname{div} \mathbf{C}\boldsymbol{\varepsilon}(\hat{\mathbf{u}}_\omega) = \hat{\mathbf{f}_\omega}, \tag{1.27}$$

and

$$-\frac{\omega^2}{\kappa}\hat{p}_\omega - \operatorname{div}\left(\frac{1}{\rho}\nabla \hat{p}_\omega\right) = \operatorname{div}\left(\frac{1}{\rho}\hat{\mathbf{f}_\omega}\right). \tag{1.28}$$

Time harmonic formulations have several applications. First, in the case where the source $f(\mathbf{x},t) = g(\mathbf{x})e^{\mathbf{i}\omega t}$ is time harmonic, the solution becomes time harmonic when $t \to \infty$ and the time harmonic solution is given by $\mathbf{u}(\mathbf{x},t) = \hat{\mathbf{u}}_\omega(\mathbf{x})e^{\mathbf{i}\omega t}$ or $p(\mathbf{x},t) = \hat{p}_\omega(\mathbf{x})e^{\mathbf{i}\omega t}$. This is often the case for electromagnetic scattering or vibration analysis.

Second, it is possible to recover the time-domain solution by Fourier synthesis. In this case, equation (1.27) or (1.28) is solved for a finite number of frequencies ω_j and the time-domain solution is obtained by an approximate inverse Fourier transform

$$\mathbf{u}(\mathbf{x},t) \simeq \sum_j \mathbf{u}_{\omega_j}(\mathbf{x}) e^{\mathbf{i}\omega_j t}.$$

This approach is also called frequency-domain treatment of the wave equation. One of the advantage is that attenuation can be naturally taken into account. Also, this approach is naturally parallel since the computation of each frequency is independent. We refer the reader to [43] for more information.

The possibility of frequency-domain treatment of 3D seismic wave propagation problems is currently under study: Operto and his collaborators have proposed a feasibility study using acoustic approximation (1.28) and a massively parallel direct solver (the software package MUMPS [11]) in [79].

Third, even when the source is not time periodic, it is often of interest to consider only some frequency content of the solution. This is the case in Geophysics for some applications like full waveform inversion. In the context of full waveform inversion, multiscale frequency-domain techniques are used to mitigate the non-linearity of the problem. The algorithm requires only one or a few frequencies at each iteration [84, 104]. Furthermore, carefully choosing the frequencies makes it possible to speed up full waveform inversion algorithm as observed by Sirgue and Pratt [91]. The authors also mention that their argument might also apply to imaging method. 3D frequency-domain full waveform inversion has been carried out on synthetic examples by Ben Hadj Ali et al. [6] and on real datasets by Plessix [82].

1.3 Boundary conditions

In the context of seismic imaging, we want to simulate the propagation of waves inside a domain of interest $\Omega \subset \mathbb{R}^3$. Without loss of generality, we assume that the domain of interest Ω has the following shape,

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^3 \quad \middle| \quad \begin{array}{c} x_1^m < x_1 < x_1^M \\ x_2^m < x_2 < x_2^M \\ S(x_1,x_2) < x_3 < x_3^M \end{array} \right\}$$

where $S : \mathbb{R}^2 \to \mathbb{R}$ locally describe the surface of the Earth (see figure 1.8).

In order to constrain the wave propagation problem in Ω, equations (1.27) and (1.28) need to be coupled with boundary conditions on $\partial\Omega$.

First, a boundary condition is required to take into account the surface of the ground that we denote by Γ_F. This boundary condition is called free-surface boundary condition. The free surface boundary condition is valid on the Earth surface.

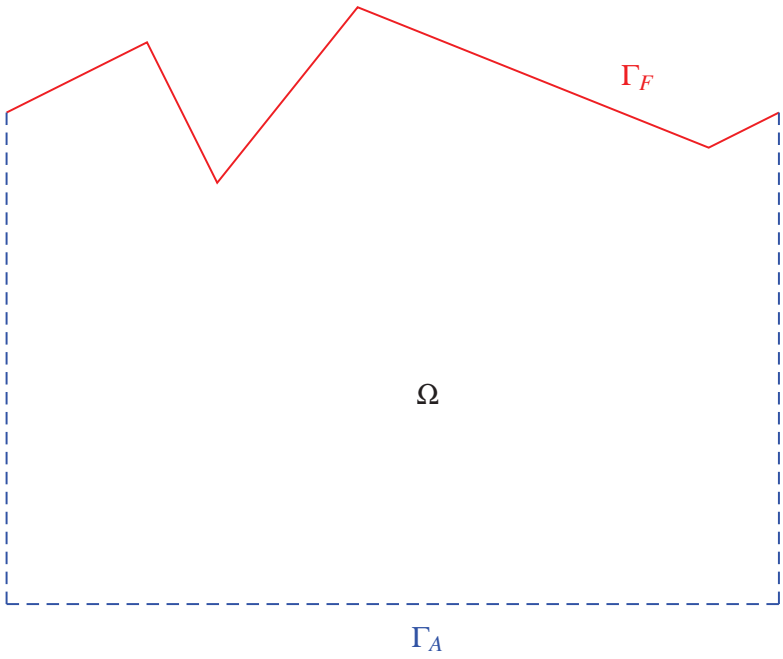


Figure 1.8: Domain of interest with boundary conditions

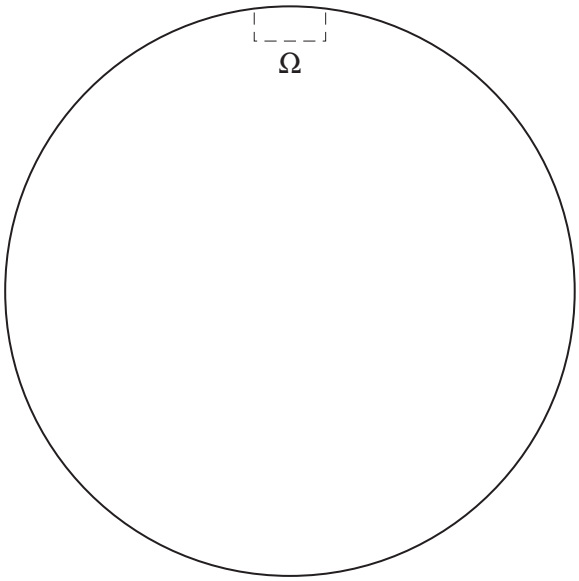


Figure 1.9: Domain of interest compared to the whole Earth

Second, as depicted by figure 1.9, the domain of interest is much smaller than the size of the Earth. Therefore, it is relevant to impose artificial "non-reflecting" boundary conditions on Γ_A to close the problem. In contrast with the free surface Γ_F, the boundary Γ_A is artificial: it just represents the end of the zone of interest, but has no physical justification.

We assume that there is no energy source outside of the domain of interest. Hence, we set a boundary condition that let the wave coming from the domain of interest travel freely through the artificial boundary, but filters out waves coming from outside. Such a boundary condition is called a non-reflecting, or transparent, boundary condition.

The free surface boundary condition has to be satisfied on

$$\Gamma_F : x_3 = S(x_1, x_2),$$

while the non-reflecting boundary condition needs to be imposed on the boundaries

$$x_1 = x_1^m, \quad x_1 = x_1^M, \quad x_2 = x_2^m, \quad x_2 = x_2^M, \quad x_3 = x_3^M.$$

1.3.1 Free surface condition

The free-surface boundary condition on Γ_F is easy to write. It is characterized by the fact that that the traction vanishes on Γ_F: $\boldsymbol{\sigma}\mathbf{n} = 0$. We can write this condition on the displacement $\mathbf{u}$ using Hooke's law (1.5):

$$\mathbf{C}\boldsymbol{\varepsilon}(\hat{\mathbf{u}}_\omega)\mathbf{n} = 0 \text{ on } \Gamma_F, \tag{1.29}$$

which is a Neumann-like boundary condition.

In the acoustic approximation, free surface condition (1.29) leads to:

$$\hat{p}_\omega = 0 \text{ on } \Gamma_F, \tag{1.30}$$

which is a Dirichlet boundary condition.

1.3.2 Non-reflecting boundary conditions

Non-reflecting boundary conditions define a complex topic which is still under development. In the following we consider two types of non-reflecting boundary conditions: absorbing boundary conditions, and perfectly match layers. We refer the reader to the PhD thesis of J. Diaz [42] for a general presentation of this conditions.

Absorbing boundary conditions

The aim of absorbing boundary conditions is to simulate an infinite domain of propagation by imposing a local boundary condition on the boundary Γ_A. Pioneering works on absorbing boundary conditions include the results of Engquist and Majda [44]. The main

ingredient is that there exists a non-local pseudo-differential operator $\mathcal{T}$ such that the boundary condition

$$\frac{1}{\rho}\nabla\hat{p}_\omega \cdot \mathbf{n} = \mathcal{T}(\hat{p}_\omega) \text{ on } \Gamma_A, \tag{1.31}$$

is non-reflecting. Unfortunately, since the operator $\mathcal{T}$ is nonlocal, its discretization is costly. Therefore one wishes to approximate the operator $\mathcal{T}$ by a local expression.

As shown, for instance, in [52] or [42], several expansions of the operator $\mathcal{T}$ are possible. In the following, we will use the simplest first order approximation given by Engquist and Majda [44]:

$$\frac{1}{\rho}\nabla\hat{p}_\omega \cdot \mathbf{n} = \frac{\mathbf{i}\omega}{\sqrt{\rho\kappa}}\hat{p}_\omega \text{ on } \Gamma_A. \tag{1.32}$$

A similar approach for isotropic elastic media is presented by Clayton and Engquist in [36]. However, we will only use perfectly matched layers for elastic waves. We also refer the reader to the recent work of Boillot et al. [25] for the derivation of an absorbing boundary condition for transverse isotropic media.

Perfectly Matched Layers

Another approach, called Perfectly Matched Layers (PML), was proposed by Bérenger for electromagnetic waves [22]. Instead of imposing a local boundary condition at the boundary of the domain interest, the domain of interest is surrounded by an additional layer, in which outgoing waves are absorbed. This is done by introducing artificial dispersion inside the layer.

To simplify, suppose we want to impose an artificial boundary at $x_2 = 0$ to reduce a simulation in $\mathbb{R}^2$ to $\mathbb{R}^2_+$ only. We introduce an additional layer $\mathbb{R} \times (0, -L)$ of length $L > 0$ under the domain. In the layer, the derivative operator with respect to x_2 variable is formally replaced as follows:

$$\frac{\partial}{\partial x_2} \longrightarrow \frac{\mathbf{i}\omega}{\mathbf{i}\omega + \sigma}\frac{\partial}{\partial x_2},$$

where $\sigma > 0$ is an arbitrary positive constant. A Dirichlet boundary condition is applied on the external boundary of the layer to close the problem.

The absorbing properties of the layer depend on length L and the constant σ. When L and σ increase, the layer is more efficient, but a finer and longer mesh is required inside the layer.

Figure 1.10 shows the domain of interest surrounded by perfectly matched layers in 2D. The domain is surrounded by PML of size L, resulting in the computational domain

$$\Omega = \left\{ \mathbf{x} \in \mathbb{R}^3 \quad \middle| \quad \begin{array}{c} x_1^m - L < x_1 < x_1^M + L \\ x_2^m - L < x_2 < x_2^M + L \\ S(x_1, x_2) < x_3 < x_3^M + L \end{array} \right\}$$

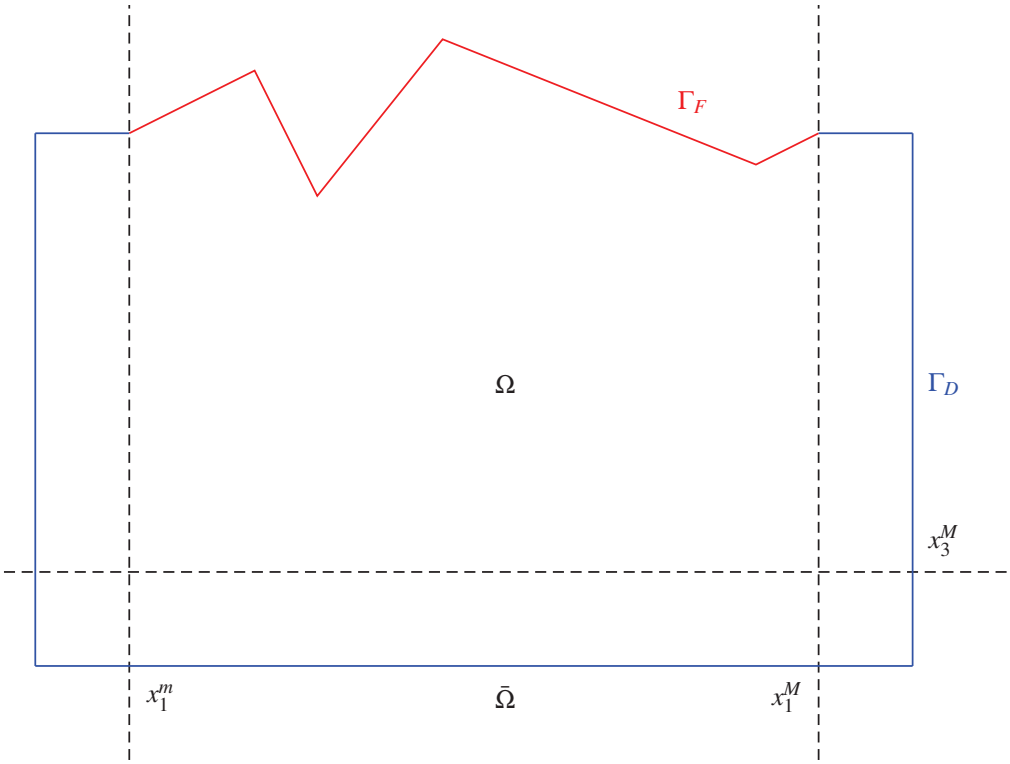


Figure 1.10: Domain of interest surrounded by PML

We introduce the functions $\nu_j : \bar{\Omega} \to \mathbb{C}$ defined by

$$\nu_j(x) = \begin{cases} 1 & \text{if } x_j^m \leq x_j \leq x_j^M \\ \dfrac{\mathbf{i}\omega}{\mathbf{i}\omega + \sigma} & \text{otherwise.} \end{cases}$$

Applying a PML condition simply consists in formally transforming the derivative operators

$$\frac{\partial}{\partial x_j} \longrightarrow \nu_j \frac{\partial}{\partial_j}.$$

In the acoustic case, we can rewrite the PML equation under a compact form:

$$-\frac{\omega^2}{a\kappa}\hat{p}_\omega - \operatorname{div}\left(\frac{1}{\rho} B \nabla \hat{p}_\omega\right) = \operatorname{div}\left(\frac{1}{\rho}\hat{\mathbf{f}}\right),$$

where

$$a = \prod_j \nu_j, \quad B_{ij} = \frac{\nu_i \nu_j}{a}.$$

The condition $\hat{p}_\omega = 0$ is imposed on the external boundary Γ_D of the PML.

For isotropic elastic media, we define the fourth order tensor $\tilde{\mathbf{C}}$ as

$$\left(\tilde{\mathbf{C}}\nabla\boldsymbol{\phi}\right)_{nm} = \delta_{mn}\lambda \sum_k \frac{\nu_n \nu_k}{a}\frac{\partial \phi_k}{\partial x_k} + \mu\left(\frac{\nu_m^2}{a}\frac{\partial \phi_n}{\partial x_m} + \frac{\nu_n \nu_m}{a}\frac{\partial \phi_m}{\partial x_n}\right).$$

We have $\tilde{\mathbf{C}}\nabla\mathbf{u} = \mathbf{C}\boldsymbol{\varepsilon}(\mathbf{u})$ inside Ω and the elastic wave equation reads

$$-\frac{\rho\omega^2}{a}\hat{\mathbf{u}}_\omega - \operatorname{div}\left(\tilde{\mathbf{C}}\nabla\hat{\mathbf{u}}_\omega\right) = \mathbf{f},$$

together with the condition that $\hat{\mathbf{u}}_\omega = 0$ on Γ_D.

It has been observed that PMLs might be instable in anisotropic media [21]. For this reason, new methodologies are currently developed to obtain non-reflecting boundary conditions in general anisotropic case (for instance, see [99]).

1.4 Boundary value problems and their variational formulation

We have derived the elastodynamic equation and its acoustic approximation, together with appropriate boundary conditions in the framework of continuous Mechanics. This task has been carried out formally, without precise mathematical statement. We fill this lack in this section by stating the considered problems in precise mathematical formulation.

It is well known that boundary value problems usually do not admit regular solutions in the general case, so that a precise notation of "derivative" is required. We briefly introduce notions of measure and distributions theory to solve this problem. We refer the reader to the book of W. Rudin [85] for a general presentation of measure theory and to L. Schwartz [89] for distributions theory.

From now on, we assume that $\Omega \subset \mathbb{R}^N$ is a regular open simply-connected domain. We require the definition of the spaces $L^p(\Omega,\mathbb{C})$ for $1 \leq p \leq +\infty$. We refer the reader to [85] for the definition of these spaces. We will also use the Sobolev spaces $W^{m,p}(\Omega,\mathbb{C})$ with $m \in \mathbb{N}^\star$ and $1 \leq p \leq +\infty$. We define in particular $H^m(\Omega,\mathbb{C}) = W^{m,2}(\Omega,\mathbb{C})$. $H^1_0(\Omega,\mathbb{C})$ is the space of functions of $H^1(\Omega,\mathbb{C})$ which vanish on $\partial\Omega$. Finally, if $\Gamma \subset \partial\Omega$, $H^1_\Gamma(\Omega,\mathbb{C})$ is the space of functions in $H^1(\Omega,\mathbb{C})$ vanishing on Γ. We refer the reader to the book of H. Brezis [26] for a presentation of the Sobolev spaces $W^{m,p}(\Omega,\mathbb{C})$.

With the introduction of the Sobolev spaces, we are now ready to establish the boundary value problems considered afterward, and their variational formulations. The acoustic wave propagation problem with ABC consists in finding a scalar unknown, the pressure, denoted here by $u : \Omega \to \mathbb{C}$, solution to

$$\left\{\begin{array}{rcll} -\dfrac{\omega^2}{\kappa}u - \operatorname{div}\left(\dfrac{1}{\rho}\nabla u\right) &=& f & \text{in } \Omega, \\ u &=& 0 & \text{on } \Gamma_F, \\ \dfrac{1}{\rho}\nabla u\cdot\mathbf{n} - \dfrac{\mathbf{i}\omega}{\sqrt{\kappa\rho}}u &=& 0 & \text{on } \Gamma_A, \end{array}\right. \tag{1.33}$$

where $\kappa^{-1}, \rho^{-1} \in L^\infty(\Omega,\mathbb{R})$ describe the acoustic medium and $f \in L^2(\Omega,\mathbb{C})$ represents the seismic source. The variational formulation of problem (1.33) is to find $u \in H^1_{\Gamma_D}(\Omega,\mathbb{C})$

such that

$$-\omega^2 \int_\Omega \frac{1}{\kappa} u\bar{v} - \mathbf{i}\omega \int_{\partial\Omega} \frac{1}{\sqrt{\kappa\rho}} u\bar{v} + \int_\Omega \frac{1}{\rho} \nabla u \cdot \nabla \bar{v} = \int_\Omega f\bar{v}, \tag{1.34}$$

for all $v \in H^1_{\Gamma_D}(\Omega, \mathbb{C})$.

The acoustic wave propagation problem with PML requires the introduction of artificial coefficients a and B. From the definition given in the previous section, it is clear that $a^{-1} \in L^\infty(\Omega, \mathbb{C})$ and $B_{ij} \in L^\infty(\Omega, \mathbb{C})$ $(1 \leq i, j \leq N)$. The scalar unknown $u : \Omega \to \mathbb{C}$ represents the pressure and must satisfy:

$$\left\{ \begin{array}{rcll} -\dfrac{\omega^2}{\kappa a} u - \operatorname{div}\left(\dfrac{1}{\rho} B \nabla u\right) & = & f & \text{in } \Omega, \\ u & = & 0 & \text{on } \partial\Omega, \end{array} \right. \tag{1.35}$$

In the variational version of (1.35), we seek $u \in H^1_0(\Omega, \mathbb{C})$ such that

$$-\omega^2 \int_\Omega \frac{1}{\kappa a} u\bar{v} + \int_\Omega \frac{1}{\rho} B \nabla u \cdot \nabla \bar{v} = \int_\Omega f\bar{v} \tag{1.36}$$

for all $v \in H^1_0(\Omega, \mathbb{C})$.

We also consider elastic wave propagation problems where PML are used to simulate a semi-infinite propagation medium. In this case, the unknown $\mathbf{u} : \Omega \to \mathbb{C}^N$ is vectorial and represents the displacement. The unknown must satisfy:

$$\left\{ \begin{array}{rcll} -\dfrac{\rho\omega^2}{a} \mathbf{u} - \operatorname{div}\left(\tilde{\mathbf{C}} \nabla \mathbf{u}\right) & = & \mathbf{f} & \text{in } \Omega \\ \tilde{\mathbf{C}} \nabla \mathbf{u} \cdot \mathbf{n} & = & 0 & \text{on } \Gamma_F, \\ \mathbf{u} & = & 0 & \text{on } \Gamma_D \end{array} \right. \tag{1.37}$$

where $\rho/a \in L^\infty(\Omega, \mathbb{C})$ and each coefficient $C_{\alpha,\beta}$ of $\tilde{\mathbf{C}}$ is such that $C_{\alpha,\beta} \in L^\infty(\Omega, \mathbb{C})$. The load term is assumed to square integrable: $\mathbf{f} \in L^2(\Omega, \mathbb{C})^N$. The variational version of problem (1.37) is to find $\mathbf{u} \in H^1_{\Gamma_D}(\Omega, \mathbb{C})^N$ such that

$$-\omega^2 \int_\Omega \frac{\rho}{a} \mathbf{u} \cdot \bar{\mathbf{v}} + \int_\Omega \tilde{\mathbf{C}} \nabla \mathbf{u} : \nabla \bar{\mathbf{v}} = \int_\Omega \mathbf{f} \cdot \bar{\mathbf{v}}, \tag{1.38}$$

for all $\mathbf{v} \in H^1_{\Gamma_D}(\Omega, \mathbb{C})^N$.

Chapter 2

Analysis of the problem in one dimension

2.1 Analysis of the continuous problem

The convergence analysis of numerical schemes is based on stability properties of the PDE to be discretized. Since we aim at solving Helmholtz problems in heterogeneous media, there is a need for studying its stability. Indeed, most of the results have been obtained for homogeneous media. The only result dealing with the stability of variational formulation of he Helmholtz equation in heterogeneous media that we are aware of dates back from 1988, and is restricted to the case of a one dimensional problems with a smoothly varying wavespeed [16]. More recent investigations are available, namely the works of Cummings and Feng [39] and Hetmaniuk [55], but they are restricted to the analysis of the homogeneous equation (with a constant wavespeed).

Claeys and Hiptmair have developed a theory for acoustic (and even electromagnetic) scattering by composite structures, described by piecewise constant medium parameters, in the context of integral equations [34, 35]. They show the coercivity of their formulation, but the dependency of the stability constant with respect to the frequency is not explicit (see Theorem 10.4 of [34]). Hence, it is not useful for frequency-explicit convergence study.

In our 1D contributions, we consider an acoustic medium defined by piecewise constant parameters. We give a new results demonstrated in the variational framework. First, we derive a stability estimate for the one dimensional case with mixed boundary conditions and arbitrary piecewise constant parameters κ and ρ. This result can be generalized to a class of two-dimensional layered propagation domains.

We deal with the variational formulation of the heterogeneous Helmholtz equation coupled with standard boundary conditions like Dirichlet, Neumann and Fourier-Robin conditions. Our objective is to establish stability estimates depending on the frequency explicitly.

We go beyond the standard estimates of the solution H^1 norm. Indeed, we give an estimate in the $W^{1,\infty}$ norm. Besides, we are able to give stability estimates not only with

respect to the right-hand-side, but also with respect to the model parameters κ and ρ. Furthermore, we will take advantage of these results for the numerical analysis to bound the jumps in the solution derivative by a constant explicitly depending on the frequency.

2.1.1 Preliminary information

One might think that the Helmholtz equation, in particular in 1D, is simple. Indeed, it is a linear PDE related to the Laplace operator. In the simplest case, we can write the operator as $-\Delta - \omega^2 I$. However, if the operator has a simple expression, it behaves very differently from the Laplace operator. It is because the perturbation $-\omega^2 I$ has a negative sign, so that the coercivity of the Laplace operator is lost. Furthermore, the loss of coercivity is related to the frequency ω, and becomes more important at high frequencies. For this reason, classical arguments based on the Lax-Milgram theorem can not apply to the Helmholtz equation (see for example \$2.4 of [58] or \$2.10 of [87]).

If the Helmholtz operator is indefinite, it still keeps an interesting form. Indeed, we see that $-\Delta - \omega^2 I = -\Delta(I + \omega^2 \Delta^{-1})$, so that we can think about the operator as a compact perturbation $\omega^2 \Delta^{-1}$ of the identity. Hence, we can study the stability of the operator in the context of the Fredholm alternative theory: the problem of well-posedness resumes to the problem of uniqueness of the solution. For a demonstration of the Fredholm alternative, we refer the reader to the book of Brezis [26]. The application to variational formulations of Helmholtz problems is given, for instance, in [58] or [87].

In the context of interior problems, the Fredholm alternative explains very well the existence of resonance frequencies. They are the values of ω for which the homogeneous equation admits non-trivial solutions. For exterior problems, or geophysical applications, the Sommerfeld radiation condition, or its approximation, prevents from resonances, and we are able to show uniqueness for all frequencies $\omega > 0$. Therefore, we are able to show the well-posedness of the Helmholtz equation using the Fredholm alternative.

However, if the Fredholm alternative provides a nice tool to show existence, uniqueness and stability with respect to the right-hand-side, the stability constant remains implicit. In particular, the behaviour of the stability constant (and therefore, the behaviour of the solution) with respect to the frequency is not specified. Furthermore, in the context of heterogeneous media, the influence of the medium parameters is unknown as well. Frequency-explicit stability estimates are crucial for several reasons:

- From the physical point of view, frequency-explicit stability estimates describe the behaviour of the energy depending on the frequency. For example, they indicate how the amplitude of the solution depends on the frequency.

- In the context of numerical analysis, frequency-explicit stability estimates are also of importance. Indeed, when analysing finite element schemes, regularity of the solution is involved. In particular, it is well known that the bound on the finite element error involves the semi-norm of the solution in suitable Sobolev spaces.

Since we are especially concerned with the design of efficient numerical methods for high frequency regime, the behaviour of the solution derivatives at high frequency will take a crucial place in the forth-coming analysis.

This need for additional informations has motivated several developments. To the best of our knowledge, the first frequency-explicit stability estimate available in the literature has been demonstrated by Aziz et al. [16]. They consider a smooth velocity parameter c and the boundary value problem

$$\left\{\begin{array}{rcl} -\dfrac{\omega^2}{c^2}\mathbf{u} - \mathbf{u}'' &=& f, \\ \mathbf{u}(0) &=& u_0, \\ \mathbf{u}'(1) - \dfrac{\mathbf{i}\omega}{c_\infty}\mathbf{u}(1) &=& 0. \end{array}\right.$$

Their idea is to use a special test function of the form $v = \alpha\mathbf{u} + \beta\mathbf{u}'$, where α and β are smooth functions. Since the velocity parameter c is smooth, it is possible to pick α and β to be the solutions of the ODE system

$$\left\{\begin{array}{rcl} 2\alpha + \beta' &=& 1 \\ 2c^{-2}\alpha - (c^{-2}\beta)' &=& c^{-2}, \end{array}\right.$$

yielding estimates depending explicitly on the frequency.

This methodology has two main drawbacks:

- It can not be generalized to non-smooth velocity parameters. Indeed, the result they obtain is only valid for a high enough frequency range $\omega \geq \omega_0(c)$. It turns out that $\omega_0(c)$ increases with the derivative of c, so that if we try to approximate a non-smooth velocity parameter by a smooth one, we will not be able to extend the proof.

- The other issue is that the method of Aziz et al. do not precise how the solution depends on the velocity. Indeed, their stability constant depends explicitly on the frequency, but implicitly on the velocity. This is because the constant depends on the derivatives of α and β, which are implicitly defined through an ODE.

In the context of homogeneous propagation media, a simpler methodology has been developed by Douglas et al. [43]. Indeed, in a one dimensional homogeneous medium, the solution can be expressed as a convolution with the Green function (which is analytically available):

$$\mathbf{u}(x) = \frac{\mathbf{i}}{2\omega}\int_0^1 f(\xi)\exp\left(\mathbf{i}\omega|x-\xi|\right),$$

and stability estimates follow from simple computations. The same approach is adopted by Babuška and Ihlenburg with different boundary conditions [59]. But if this approach is straightforward, it is limited to the one dimensional homogeneous case.

Makridakis, Ihlenburg and Babuška have also considered the problem of a fluid-solid interaction [70]. A solid body is immersed in a fluid and waves are propagating. They have

used special test functions of the form $v(x) = (x - \bar{x})\mathbf{u}'(x)$, where $\bar{x}$ is a carefully selected point and their approach came to our attention. In the following we propose to extend the methodology of Makridakis, Ihlenburg and Babuška to the case of a heterogeneous acoustic one dimensional medium. We assume that the medium is determined by piecewise constant parameters, so that the problem can be understood as a transmission problem. Our method differs from Makridakis, Ihlenburg and Babuška. Indeed, in every layer, an acoustic Helmholtz equation is considered and the transmission conditions are different. Besides, we do not restrict our model to 3 layers, but to an arbitrary number of layers L. Furthermore, the parameters of the medium can be chosen freely in each layer.

Though our 1D proof can be generalized to the case of two dimensional layered media, it is not valid for general geometries. This is the reason why we will propose another methodology for 2D problems in Chapter 2.

2.1.2 Description of the propagation medium

We consider a 1D acoustic propagation medium defined by a bulk modulus κ and a density ρ. In order to simplify the notations in the following, we need to introduce a set M of admissible models, or equivalently, a set of admissible couple (κ, ρ).

Definition 1. *We will say that a couple $(\kappa, \rho) \in L^\infty(0, Z, \mathbb{R}) \times L^\infty(0, Z, \mathbb{R})$ is an admissible propagation medium, and we will note $(\kappa, \rho) \in M$ if the following conditions are satisfied:*

1. *There exist an integer L, a partition*

$$0 = \mathbf{z}_0 < \mathbf{z}_1 < \dots \mathbf{z}_{L-1} < \mathbf{z}_L = Z$$

and two sets of values $\{\kappa_l\}_{l=1}^L$, $\{\rho_k\}_{l=1}^L$ such that

$$\kappa|_{(\mathbf{z}_{l-1}, \mathbf{z}_l)} = \kappa_l, \quad \rho|_{(\mathbf{z}_{l-1}, \mathbf{z}_l)} = \rho_l,$$

for all $l \in \{1, \dots, L\}$.

2. *Each parameter is bounded*

$$\kappa_\star \leq \kappa_l \leq \kappa^\star, \quad \rho_\star \leq \rho_l \leq \rho^\star,$$

for all $L \in \{1, \dots, L\}$.

3. *We define $\mathbf{h}_l = \mathbf{z}_l - \mathbf{z}_{l-1}$ for all $l \in \{1, \dots, L\}$ and assume that $\mathbf{h}_l \geq \mathbf{h}_\star$ for all $l \in \{1, \dots, L\}$.*

4. *We define the constant*

$$\mathcal{M} = \prod_{l=1}^{L-1} \max\left(\frac{\kappa_{l+1}}{\kappa_l}, \frac{\rho_l}{\rho_{l+1}}, 1\right), \tag{2.1}$$

and assume that $\mathcal{M} \leq \mathcal{M}^\star$,

where the constants $\kappa_\star$, $\kappa^\star$, $\rho_\star$, $\rho^\star$, $\mathbf{h}_\star$ and $\mathcal{M}^\star$ are fixed. We also assume that $\kappa_\star, \rho_\star, \mathbf{h}_\star > 0$.

Furthermore, if $(\kappa, \rho) \in M$, we note $c = \sqrt{\kappa/\rho} \in L^\infty(0, Z, \mathbb{C})$ the wavespeed. We also note $c_l = \sqrt{\kappa_l/\rho_l}$ for every layer $l \in \{1, \ldots, L\}$.

In the remaining of Chapter 2, we will use the notations $||.||$ and $||.||_\infty$ for the $L^2(0, Z, \mathbb{C})$ and $L^\infty(0, Z, \mathbb{C})$ norms. Theses norms are defined by

$$||v||^2 = \int_0^Z |v(z)|^2 dz, \quad \forall v \in L^2(0, Z, \mathbb{C})$$

and

$$||v||_\infty = \text{esssup}_{z\in(0,Z)} |v(z)|, \quad \forall v \in L^\infty(0, Z, \mathbb{C}).$$

2.1.3 Problem setting

Let $f \in L^2(0, 1, \mathbb{C})$ be a load term, $(\kappa, \rho) \in M$ be an admissible model and $c = \sqrt{\kappa, \rho}$ be the associated wavespeed. We seek a function $\mathbf{u} \in L^2(0, Z, \mathbb{C})$ satisfying the following conditions:

- In each layer $l \in \{1, \ldots, L\}$, $\mathbf{u}$ satisfies

$$-\frac{\omega^2}{c_l^2}\mathbf{u} - \mathbf{u}'' = \rho_l f \text{ in } (\mathbf{z}_{l-1}, \mathbf{z}_l), \tag{2.2}$$

  in the sense of distribution $\mathcal{D}'(\mathbf{z}_{l-1}, \mathbf{z}_l, \mathbb{C})$.

- For each interface $l \in \{1, \ldots, L-1\}$, the following compatibility conditions hold:

$$\mathbf{u}(\mathbf{z}_l^-) = \mathbf{u}(\mathbf{z}_l^+) \tag{2.3}$$

$$\frac{1}{\rho_l}\mathbf{u}'(\mathbf{z}_l^-) = \frac{1}{\rho_{l+1}}\mathbf{u}'(z_l^+) \tag{2.4}$$

- $\mathbf{u}$ satisfies the boundary conditions

$$-\mathbf{u}'(0) = 0 \tag{2.5}$$

$$\mathbf{u}'(Z) - \frac{\mathbf{i}\omega}{c_L}\mathbf{u}(Z) = 0. \tag{2.6}$$

Our first task is to show that transmission problem (2.2-2.6) is equivalent to variational problem (2.7). Establishing a variational formulation is crucial, because Galerkin discretizations (including finite elements) are not based on transmission problem (2.2-2.6), but rather on variational formulation (2.7). Furthermore, several important properties of the solution are easier to show in the variational framework.

Theorem 1. *A function* $\mathbf{u} \in L^2(0, Z, \mathbb{C})$ *satisfies the boundary value problem* (2.2) *to* (2.6) *if and only if* $\mathbf{u} \in H^1(0, Z, \mathbb{C})$ *and*

$$B_{\omega,\kappa,\rho}(\mathbf{u}, v) = \int_0^Z f(z)\overline{v(z)}dz, \quad \forall v \in H^1(0, Z, \mathbb{C}), \tag{2.7}$$

where

$$B_{\omega,\kappa,\rho}(w, v) = -\omega^2 \int_0^Z \frac{1}{\kappa(z)} w(z)\overline{v(z)}dz - \frac{\mathbf{i}\omega}{\sqrt{\kappa_l \rho_l}} w(Z)\overline{v(z)} + \int_0^Z \frac{1}{\rho(z)} w'(z)\overline{v'(z)}dz. \tag{2.8}$$

Proof. Let us first assume that $\mathbf{u} \in L^2(0, Z, \mathbb{C})$ satisfies (2.2) to (2.6). We are going to show that $\mathbf{u} \in H^1(0, Z, \mathbb{C})$ and that (2.7) holds.

Since $\mathbf{u}$ satisfies (2.2), it is clear that

$$\mathbf{u}'' = -\rho_l f - \frac{\omega^2}{c_l^2}\mathbf{u} \in L^2(\mathbf{z}_{l-1}, \mathbf{z}_l, \mathbb{C}),$$

for $l \in \{1, \dots, L\}$. It follows that $\mathbf{u} \in H^2(\mathbf{z}_{l-1}, \mathbf{z}_l, \mathbb{C})$ for every layer $l \in \{1, \dots, L\}$ and compatibility condition (2.3) implies that $\mathbf{u} \in H^1(0, Z, \mathbb{C})$.

To show that (2.7) holds, we consider a test function $\phi \in H^1(0, Z, \mathbb{C})$. Multiplying (2.2) by $\rho_l^{-1}\overline{\phi}$ we obtain

$$-\frac{\omega^2}{\kappa_l}\mathbf{u}(z)\overline{\phi(z)} - \frac{1}{\rho_l}\mathbf{u}''(z)\overline{\phi(z)} = f(z)\overline{\phi(z)}, \quad z \in (\mathbf{z}_{l-1}, \mathbf{z}_l). \tag{2.9}$$

We integrate (2.9) on each layer $(\mathbf{z}_{l-1}, \mathbf{z}_l)$ and sum over all layers $l \in \{1, \dots, L\}$. It follows

$$\sum_{l=1}^{L} \left\{ -\frac{\omega^2}{\kappa_l} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z)\overline{\phi(z)}dz - \frac{1}{\rho_l} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z)\overline{\phi(z)}dz \right\} = \sum_{l=1}^{L} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi(z)}dz. \tag{2.10}$$

By Chasles relation, we have

$$\sum_{l=1}^{L} \left\{ -\frac{\omega^2}{\kappa_l} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z)\overline{\phi(z)}dz \right\} = -\omega^2 \int_0^Z \frac{1}{\kappa(z)}\mathbf{u}(z)\overline{\phi(z)}dz, \tag{2.11}$$

and

$$\sum_{l=1}^{L} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi(z)}dz = \int_0^Z f(z)\overline{\phi(z)}dz. \tag{2.12}$$

Using integration by parts, compatibility condition (2.4), boundary conditions (2.5) and (2.6), and Chasles relation, we end up with

$$\sum_{l=1}^{L} \left\{ -\frac{1}{\rho_l} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z)\overline{\phi(z)}dz \right\} = -\frac{\mathbf{i}\omega}{\sqrt{\rho_l \kappa_l}}\mathbf{u}(Z)\overline{\phi(Z)} + \int_0^Z \frac{1}{\rho(z)}\mathbf{u}'(z)\overline{\phi'(z)}dz. \tag{2.13}$$

Inserting (2.11), (2.12) and (2.13) in (2.10), we have

$$B_{\omega,\kappa,\rho}(\mathbf{u},\phi) = \int_0^Z f(z)\overline{\phi(z)}dz, \tag{2.14}$$

and (2.7) holds, since (2.14) is true for all $\phi \in H^1(0,Z,\mathbb{C})$.

We now assume that $\mathbf{u} \in H^1(0,Z,\mathbb{C})$ and that (2.7) holds. We are going to show that $\mathbf{u}$ satisfies (2.2) to (2.6).

We consider a test function $\phi \in \mathcal{D}(0,Z,\mathbb{C})$, and integrate by parts. We have

$$B_{\omega,\kappa,\rho}(\mathbf{u},\phi) = \int_0^Z \left\{ -\frac{\omega^2}{\kappa(z)}\mathbf{u}(z) - \left(\frac{1}{\rho}\mathbf{u}'\right)'(z) \right\} \overline{\phi(z)}dz,$$

and, since (2.7) holds

$$\int_0^Z \left(\frac{1}{\rho}\mathbf{u}'\right)'(z)\overline{\phi(z)}dz = \int_0^Z \left(-f(z) - \frac{\omega^2}{\kappa(z)}\mathbf{u}(z)\right)\overline{\phi(z)}dz.$$

It follows that

$$\left(\frac{1}{\rho}\mathbf{u}'\right)' = f - \frac{\omega^2}{\kappa}\mathbf{u}, \tag{2.15}$$

in the sense of distribution and therefore $(\rho^{-1}\mathbf{u}')' \in L^2(0,Z,\mathbb{C})$. We thus have $\rho^{-1}\mathbf{u}' \in H^1(0,Z,\mathbb{C})$.

Since $\mathbf{u}, \rho^{-1}\mathbf{u}' \in H^1(0,Z,\mathbb{C})$, we have $\mathbf{u}, \rho^{-1}\mathbf{u}' \in C^0(0,Z,\mathbb{C})$. It follows that compatibility conditions (2.3) and (2.4) hold.

Furthermore, if we consider a given layer $(\mathbf{z}_{l-1},\mathbf{z}_l)$ for $l \in \{1,\dots,L\}$, $\kappa(z) = \kappa_l$ and $\rho(z) = \rho_l$ for $z \in (\mathbf{z}_{l-1},\mathbf{z}_l)$. Thus, we can multiply (2.15) by ρ_l to obtain (2.2).

In order to show that boundary conditions (2.5) and (2.6) are satisfied, we consider a test function $\phi \in H^1(0,Z,\mathbb{C})$ and integrate by parts

$$\int_0^Z \left\{ -\frac{\omega^2}{\kappa(z)}\mathbf{u}(z) - \left(\frac{1}{\rho}\mathbf{u}'\right)'(z) \right\} \overline{\phi(z)}dz - \frac{1}{\rho_1}\mathbf{u}'(0)\overline{\phi(0)} + \frac{1}{\rho_L}\mathbf{u}'(Z)\overline{\phi(Z)} - \frac{\mathbf{i}\omega}{\sqrt{\kappa_L\rho_L}}\mathbf{u}(Z)\overline{\phi(Z)} = \int_0^Z f(z)\overline{\phi(z)}dz,$$

and we obtain

$$-\frac{1}{\rho_1}\mathbf{u}'(0)\overline{\phi(0)} + \left(\frac{1}{\rho_L}\mathbf{u}'(Z) + \frac{\mathbf{i}\omega}{\sqrt{\kappa_L\rho_L}}\mathbf{u}(Z)\right)\overline{\phi(Z)} = 0 \tag{2.16}$$

using (2.15). Selecting the test functions $\phi(z) = z/Z$ and $\phi(z) = (Z-z)/Z$, we obtain boundary conditions (2.5) and (2.6). □

2.1.4 Well-posedness

This section is devoted to the existence and uniqueness of a solution $\mathbf{u}$ to problem (2.2-2.6) (or equivalently, problem (2.7)).

We follow standard paths for demonstration: since the sesquilinear form (2.8) of the variational problem satisfies a Gårding inequality, it is sufficient to show uniqueness of the solution to obtain well-posedness in the sense of Hadamard. We refer the reader to Chapter 2 of [87], in particular Corollary 2.1.61.

The theory around the Gårding inequality is based on the Fredholm alternative. Note that if it provides a simple tool to show the well-posedness of our problem, it provides no information about how the solution depends upon the parameters. Namely, the stability constant is not explicit (see for instance Proposition 8.1.3 of [73]). We will overcome this lack of information in the next section.

To begin with, let us take advantage of the 1D setting to depict the solution as a function of f:

Lemma 1. *Let $\mathbf{u} \in L^2(0, Z, \mathbb{C})$ be solution to (2.2). Then there exist $2L$ complex constants $\alpha_l, \beta_l \in \mathbb{C}$, $1 \leq l \leq L$ such that*

$$\mathbf{u}(z) = \alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} z\right) + \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} z\right) - \frac{\mathbf{i} c_l \rho_l}{2\omega} \int_0^Z f(\xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|z-\xi|\right) d\xi, \qquad (2.17)$$

for a.e. $z \in (\mathbf{z}_{l-1}, \mathbf{z}_l)$ and for $1 \leq l \leq L$.

Proof. Consider a given layer $l \in \{1, \ldots, L\}$ inside which $\mathbf{u}$ is solution to the linear ODE

$$-\frac{\omega^2}{c_l^2}\mathbf{u} - \mathbf{u}'' = \rho_l f.$$

One can verify by hand that

$$-\frac{\mathbf{i} c_l \rho_l}{2\omega} \int_0^Z f(\xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|z-\xi|\right) d\xi,$$

is a particular solution and that

$$\exp\left(\frac{\mathbf{i}\omega}{c_l} z\right), \quad \exp\left(-\frac{\mathbf{i}\omega}{c_l} z\right),$$

are two independent homogeneous solutions.

Since (2.1.4) is a second order linear ODE, it follows from classical theory that there exist two constants $\alpha_l, \beta_l \in \mathbb{C}$ such that (2.17) holds. □

We are now ready to prove uniqueness of the solution to problem (2.2-2.6). Using Lemma 1, we apply mathematical induction to prove that any solution $\mathbf{u}$ vanishes in each layer.

Lemma 2. *Assume* $\mathbf{u} \in L^2(0, Z, \mathbb{C})$ *satisfies boundary value problem (2.2-2.6) with* $f = 0$. *Then* $\mathbf{u} = 0$.

Proof. Because $f = 0$, Lemma 2 yields that for each layer $1 \leq l \leq L$,

$$\mathbf{u}(z) = \alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} z\right) + \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} z\right), \quad z \in (\mathbf{z}_{l-1}, \mathbf{z}_l),$$

for some complex constants $\alpha_l, \beta_l \in \mathbb{C}$. Now, since (2.7) holds, we have in particular

$$\operatorname{Im} B_{\omega,\kappa,\rho}(\mathbf{u}, \mathbf{u}) = -\frac{\mathbf{i}\omega}{\sqrt{\kappa_L \rho_L}} |\mathbf{u}(Z)|^2 = 0,$$

so that $\mathbf{u}(Z) = 0$. Furthermore, $\mathbf{u}$ satisfies boundary condition (2.6) and we deduce that $\mathbf{u}(Z) = \mathbf{u}'(Z) = 0$. It follows that

$$\begin{cases} \exp\left(\frac{\mathbf{i}\omega Z}{c_l}\right) \alpha_L + \exp\left(-\frac{\mathbf{i}\omega Z}{c_l}\right) \beta_L = 0 \\ \frac{\mathbf{i}\omega}{c_l} \exp\left(\frac{\mathbf{i}\omega Z}{c_l}\right) \alpha_L - \frac{\mathbf{i}\omega}{c_l} \exp\left(-\frac{\mathbf{i}\omega Z}{c_l}\right) \beta_L = 0, \end{cases}$$

and $\alpha_L = \beta_L = 0$.

Assume that $\alpha_l = \beta_l = 0$ for $l_\star \leq l \leq L$. We are going to show that $\alpha_{l_\star - 1} = \beta_{l_\star - 1} = 0$. Since $\alpha_{l_\star} = \beta_{l_\star} = 0$, is it clear that $\mathbf{u}(\mathbf{z}_{l_\star}^+) = \mathbf{u}'(\mathbf{z}_{l_\star}^+) = 0$, and transmission conditions (2.3) and (2.4) ensure that $\mathbf{u}(\mathbf{z}_{l_\star}^-) = \mathbf{u}'(\mathbf{z}_{l_\star}^-) = 0$. It follows that

$$\begin{cases} \exp\left(\frac{\mathbf{i}\omega \mathbf{z}_{l_\star}}{c_{l_\star - 1}}\right) \alpha_{l_\star - 1} + \exp\left(-\frac{\mathbf{i}\omega \mathbf{z}_{l_\star}}{c_{l_\star - 1}}\right) \beta_{l_\star - 1} = 0 \\ \frac{\mathbf{i}\omega}{c_l} \exp\left(\frac{\mathbf{i}\omega \mathbf{z}_{l_\star}}{c_{l_\star - 1}}\right) \alpha_{l_\star - 1} - \frac{\mathbf{i}\omega}{c_l} \exp\left(-\frac{\mathbf{i}\omega \mathbf{z}_{l_\star}}{c_{l_\star - 1}}\right) \beta_{l_\star - 1} = 0, \end{cases}$$

and $\alpha_{l_\star - 1} = \beta_{l_\star - 1} = 0$.

Using mathematical induction, we obtain that $\alpha_l = \beta_l = 0$ for all $l \in \{1, \ldots, Ł\}$, and hence, $\mathbf{u} = 0$. □

Uniqueness being established, we can now turn to the proof of existence. We recall that the M is the set of admissible propagation media defined in Definition 1.

Theorem 2. *Consider an admissible model* $(\kappa, \rho) \in M$ *and* $f \in L^2(0, Z, \mathbb{C})$. *Then there exists a unique function* $\mathcal{S}_{\omega,\kappa,\rho} f \in L^2(0, Z, \mathbb{C})$ *satisfying the transmission problem (2.2-2.6).*

Proof. Recalling Theorem 1, a function $\mathbf{u} \in L^2(0, Z, \mathbb{C})$ is solution to the transmission problem (2.2-2.6) iff $\mathbf{u} \in H^1(0, Z, \mathbb{C})$ and $\mathbf{u}$ is solution to the variational problem (2.7).

It is clear that sesquilinear form (2.8) satisfies a Gårding inequality. Indeed, for all $\phi \in H^1(0, Z, \mathbb{C})$, it holds that

$$\operatorname{Re} B_{\omega,\kappa,\rho}(\phi, \phi) = -\omega^2 \int_0^Z \frac{1}{\kappa(z)} |\phi(z)|^2 dz + \int_0^Z \frac{1}{\rho(z)} |\phi'(z)|^2 \geq \frac{1}{\rho^\star} ||\phi'||^2 - \frac{\omega^2}{\kappa_\star} ||\phi||^2.$$

Following the classical theory (see for, instance Corollary 2.1.61 of [87]), it is then sufficient to show uniqueness of the solution, and we conclude thanks to Lemma 2. □

2.1.5 Frequency-explicit stability estimates

In the previous section, we have demonstrated the existence and uniqueness of the solution to problem (2.2-2.6). Our proofs are based on the Fredholm alternative theory, and actually, they include stability of the solution with respect to the right hand side, that is

$$||\mathcal{S}_{\omega,\kappa,\rho}f|| + ||(\mathcal{S}_{\omega,\kappa,\rho}f)'|| \leq C||f||, \tag{2.18}$$

for some constant $C \in \mathbb{R}$.

Obviously, regarding the extreme role played by the frequency in Helmholtz problems, stability estimate (2.18) is not sufficient, because the constant C depends implicitly on all the parameters of the problem, including ω, κ and ρ.

In this section, we introduce fully-explicit stability estimates. We say that our estimates are fully explicit, because we obtain an explicit formula for all constants that occur in the estimates. More precisely, the constants depend (explicitly) on the length of the domain Z, the frequency ω, the maximum values $\kappa^\star$ and $\rho^\star$ of κ and ρ, the values κ_L and ρ_L of κ and ρ in the last layer, and the constant

$$\mathcal{M} = \prod_{l=1}^{L-1} \max\left(\frac{\kappa_{l+1}}{\kappa_l}, \frac{\rho_l}{\rho_{l+1}}, 1\right).$$

Note that κ_L and ρ_L play a special role because they are involved in the boundary condition (2.6). In fact, only the value c_L matters.

The constant $\mathcal{M}$ appears naturally in the above demonstrations. It can be interpreted as a measure of the variations of the propagation medium. In the worst case, it can be bounded by a constant depending on the extreme values of the parameters $\kappa_\star, \kappa^\star, \rho_\star, \rho^\star$ and the number of interfaces $L-1$

$$\mathcal{M} \leq \left(\frac{\kappa^\star\rho^\star}{\kappa_\star\rho_\star}\right)^{L-1}.$$

If κ is increasing and ρ is decreasing (hence, c is increasing), we have a simpler bound, independent of the number of layers

$$\mathcal{M} \leq \frac{\kappa^\star\rho^\star}{\kappa_\star\rho_\star}.$$

Finally, if κ is decreasing and ρ is increasing (that is, c is decreasing) we can neglect the constant $\mathcal{M}$ since

$$\mathcal{M} \leq 1.$$

Theorem 3. *Let $f \in L^2(0, Z, \mathbb{C})$ and $(\kappa, \rho) \in M$. The following stability estimate holds*

$$\left|\left|\kappa^{-1/2}(\mathcal{S}_{\omega,\kappa,\rho}f)\right|\right| \leq C_s\omega^{-1}\,||f||\,, \quad \left|\left|\rho^{-1/2}(\mathcal{S}_{\omega,\kappa,\rho}f)'\right|\right| \leq C_s\,||f||\,, \tag{2.19}$$

where

$$C_s = \rho^{\star 1/2}\left(1 + \frac{\kappa^\star}{\kappa_\star}\right)^{1/2}\mathcal{M}^\star Z \tag{2.20}$$

Theorem 3 is a classical stability result for the Helmholtz equation, revisited to the case of piecewise constant coefficients κ and ρ (we recall that M is the set of admissible propagation media defined in Definition 1). Note that only the values $\{\kappa_l, \rho_l\}_{l=1}^L$ of the parameters and the length Z of the domain are involved in the stability constant. In particular, the length $\{\mathbf{h}_l\}_{l=1}^L$ does not matter.

Theorem 4 gives additional information on the solution norm at the interfaces. This result is important because the amplitude of the jumps in the solution derivative are directly related to these quantities. In particular, Theorem 4 is a key point in the analysis of finite element schemes.

In Theorem 4, we handle the traces of $\mathbf{u}$ and its derivative on the interfaces $\mathbf{z}_l$. Because of the transmission condition (2.4), the traces of the derivative are multi-valued: the trace is not the same depending on which side we look at. In Definition 2, we introduce auxiliary notations to avoid confusions.

Definition 2. *For $l \in \{1, \dots, L-1\}$, we define the traces $\mathbf{v}_l$ and $\mathbf{w}_l$ by*

$$\mathbf{v}_l = (\mathcal{S}_{\omega,\kappa,\rho} f)(\mathbf{z}_l) = (\mathcal{S}_{\omega,\kappa,\rho} f)(\mathbf{z}_l^+) = (\mathcal{S}_{\omega,\kappa,\rho} f)(\mathbf{z}_l^-),$$

and

$$\mathbf{w}_l = \frac{1}{\rho_{l+1}} (\mathcal{S}_{\omega,\kappa,\rho} f)'(\mathbf{z}_l^+) = \frac{1}{\rho_l} (\mathcal{S}_{\omega,\kappa,\rho} f)'(\mathbf{z}_l^-).$$

We further define $\mathbf{v}_0, \mathbf{w}_0$, $\mathbf{v}_L$ and $\mathbf{w}_L$ by

$$\mathbf{v}_0 = (\mathcal{S}_{\omega,\kappa,\rho} f)(0), \quad \mathbf{v}_L = (\mathcal{S}_{\omega,\kappa,\rho} f)(Z),$$

and

$$\mathbf{w}_0 = \frac{1}{\rho_1} (\mathcal{S}_{\omega,\kappa,\rho} f)'(0), \quad \mathbf{w}_L = \frac{1}{\rho_L} (\mathcal{S}_{\omega,\kappa,\rho} f)'(Z).$$

Theorem 4. *At each interface $l \in \{1, \dots, L\}$, we have*

$$\kappa_l^{-1/2} |\mathbf{v}_l| \leq \frac{C_s}{\mathbf{h}_\star} \omega^{-1} \, ||f|| \,, \quad \rho_l^{1/2} |\mathbf{w}_l| \leq \frac{C_s}{\mathbf{h}_\star} \, ||f|| \,. \tag{2.21}$$

The demonstration of Theorems 3 and 4 is rather technical and is presented in a separate section.

Basically, the main ideas of the proof of Theorem 3 draw their inspiration from the proof of Makridakis, Ihlenburg and Babuška for fluid-solid interaction [70]. The difference is that we consider an arbitrary number of layers L and that we have different transmission conditions.

As a direct consequence of Theorem 3 we can bound the H^2 semi norm of the solution when the density $\rho = 1$ is constant.

Corollary 1. *Assume that $\rho = 1$. Then $\mathcal{S}_{\omega,\kappa,\rho} f \in H^2(0, Z, \mathbb{C})$ and we have*

$$||(\mathcal{S}_{\omega,\kappa,\rho} f)''|| \leq C_{s,2} \omega ||f||, \tag{2.22}$$

with

$$C_{s,2} = 1 + \frac{C_s}{\kappa_\star}$$

Proof. Since $\rho = 1$ is constant, it is clear that it holds that

$$-\frac{\omega^2}{\kappa} - u'' = f$$

in $(0, Z)$. It follows that

$$u'' = -f - \frac{\omega^2}{\kappa} u,$$

and thus $u'' \in L^2(0, Z, \mathbb{C})$ and

$$||u''|| \leq ||f|| + \frac{\omega^2}{\kappa_\star} ||u||. \tag{2.23}$$

Hence, (2.22) follows from Theorem 3. □

As a direct consequence of Theorem 3, we are able to derive an inf-sup condition for the sesquilinear form (2.8). Inf-sup condition (2.24) is equivalent to the well-posedness of the Helmholtz operator that we show in Theorem 2. Additionally, as observed by Ihlenburg and Babuška [59], we will see that its proof is a direct application of the stability estimate (Theorem 3).

Before giving Theorem 5, we need the following notation.

Definition 3. *Let $v \in H^1(0, Z, \mathbb{C})$. We introduce the norm (equivalent to the standard $H^1(0, Z, \mathbb{C})$ norm)*

$$||v||_{\omega,\kappa,\rho} = \left(\omega^2 ||\kappa^{1/2} v||^2 + ||\rho^{1/2} v'||^2\right)^{1/2}.$$

Theorem 5. *The following inf-sup condition holds*

$$\inf_{u \in H^1(0,Z,\mathbb{C})} \sup_{v \in H^1(0,Z,\mathbb{C})} \frac{\operatorname{Re} B_{\omega,\kappa,\rho}(u, v)}{||u||_{\omega,\kappa,\rho} ||v||_{\omega,\kappa,\rho}} \geq \frac{1}{1 + 2C_s\omega}. \tag{2.24}$$

Note that our inf-sup condition is frequency-explicit (and actually, explicit with respect to all parameters), because we were able to derive a frequency-explicit stability estimate. In contrast, the inf-sup constant obtained by Claeys and Hiptmair in [34] depends implicitly on all parameters.

We are now ready to establish our inf-sup condition. Observe that our result is a generalization of the inf-sup condition demonstrated by Ihlenburg and Babuška [59] for the homogeneous case. They showed that their result is optimal in the sense that the exponent on ω in the inf-sup condition is as small as possible.

Proof. Consider an arbitrary $w \in H^1(0, Z, \mathbb{C})$. We define $\phi \in H^1(0, Z, \mathbb{C})$ by

$$\phi = 2\omega^2 \overline{\mathcal{S}_{\omega,\kappa,\rho}(\kappa^{-1}\overline{w})}. \tag{2.25}$$

Definition (2.25) of ϕ ensures that

$$\begin{aligned} B_{\omega,\kappa,\rho}(\xi,\phi) &= B_{\omega,\kappa,\rho}(\overline{\phi},\overline{\xi}) \\ &= 2\omega^2 B_{\omega,\kappa,\rho}\left(\mathcal{S}_{\omega,\kappa,\rho}(\kappa^{-1}\overline{w}),\overline{\xi}\right) \\ &= 2\omega^2 \int_0^Z \frac{1}{\kappa(z)}\xi(z)\overline{w(z)}dz, \end{aligned}$$

for all $\xi \in H^1(0,Z,\mathbb{C})$. It is thus clear that

$$B_{\omega,\kappa,\rho}(w,\phi) = 2\omega^2 \int_0^Z \frac{1}{\kappa(z)}|w(z)|^2 dz = 2\omega^2 ||\kappa^{-1/2}w||^2,$$

and it follows

$$\operatorname{Re} B_{\omega,\kappa,\rho}(w,w+\phi) = \operatorname{Re} B_{\omega,\kappa,\rho}(w,\phi) - \omega^2||\kappa^{-1/2}w||^2 + ||\rho^{-1/2}w'||^2 = ||w||^2_{\omega,\kappa,\rho}.$$

Furthermore, Theorem 3 ensures that

$$||\phi||_{\omega,\kappa,\rho} \leq 2C_s\omega^2||\kappa^{-1/2}w||,$$

and therefore

$$||w+\phi||_{\omega,\kappa,\rho} \leq (1+2C_s\omega)||w||_{\omega,\kappa,\rho}.$$

Setting $\eta = w + \phi \in H^1(0,Z,\mathbb{C})$, we conclude that

$$\operatorname{Re} B_{\omega,\kappa,\rho}(w,\eta) \geq \frac{1}{1+2C_s\omega}||w||_{\omega,\kappa,\rho}||\eta||_{\omega,\kappa,\rho}.$$

□

We close our stability analysis with respect to the right-hand-side f, with Theorem 6, where we introduce additional stability estimates in L^∞ norm.

Proposition 1. *Let $c \in \mathbb{R}_+^*$ and $z \in \mathbb{R}$. Then the matrix*

$$M_{c,z} = \begin{pmatrix} \exp\left(\dfrac{\mathbf{i}\omega}{c}z\right) & \exp\left(-\dfrac{\mathbf{i}\omega}{c}z\right) \\ \exp\left(\dfrac{\mathbf{i}\omega}{c}z\right) & -\exp\left(-\dfrac{\mathbf{i}\omega}{c}z\right) \end{pmatrix},$$

is invertible.

$$M_{c,z}^{-1} = \frac{-1}{2}\begin{pmatrix} -\exp\left(-\dfrac{\mathbf{i}\omega}{c}z\right) & -\exp\left(-\dfrac{\mathbf{i}\omega}{c}z\right) \\ -\exp\left(\dfrac{\mathbf{i}\omega}{c}z\right) & \exp\left(\dfrac{\mathbf{i}\omega}{c}z\right) \end{pmatrix},$$

and

$$||M_{c,z}^{-1}||_\infty \leq 1.$$

Lemma 3. *Consider a given layer $l \in \{1, \dots, L\}$. There exist constants $\alpha_l, \beta_l \in \mathbb{C}$ such that*

$$(\mathcal{S}_{\omega,\kappa,\rho} f)(z) = \alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} z\right) + \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} z\right) - \frac{\mathbf{i} c_l \rho_l}{2\omega} \int_0^Z f(\xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|z-\xi|\right) d\xi,$$

for $z \in (\mathbf{z}_{l-1}, \mathbf{z}_l)$. Furthermore

$$|\alpha_l|, |\beta_l| \leq C_{\alpha,\beta} \omega^{-1} ||f||, \tag{2.26}$$

where

$$C_{\alpha,\beta} = \max_{l \in \{1,\dots,L\}} \left(\frac{C_s \kappa_l^{-1/2}}{\mathbf{h}_\star} + \frac{c_l \rho_l \sqrt{Z}}{2} \right).$$

Proof. The first part of the Lemma is a direct consequence of Lemma 1. Therefore, we focus on proving (2.26). Recalling Definition 2, we have

$$\alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} \mathbf{z}_l\right) + \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} \mathbf{z}_l\right) - \frac{\mathbf{i} c_l \rho_l}{2\omega} \int_0^Z f(\xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|\mathbf{z}_l - \xi|\right) d\xi = \mathbf{v}_l,$$

and

$$\frac{\mathbf{i}\omega}{c_l} \alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} \mathbf{z}_l\right) - \frac{\mathbf{i}\omega}{c_l} \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} \mathbf{z}_l\right) - \frac{\rho_l}{2} \int_0^Z f(\xi) \operatorname{sign}(\mathbf{z}_l - \xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|\mathbf{z}_l - \xi|\right) d\xi = \mathbf{w}_l.$$

Therefore, we have $M_{c_l,\mathbf{z}_l} A = B$, where $M_{c_l,\mathbf{z}_l}$ is defined in Proposition 1, $A = (\alpha_l, \beta_l)^T$ and

$$B = \begin{pmatrix} \mathbf{v}_l + \dfrac{\mathbf{i} c_l \rho_l}{2\omega} \displaystyle\int_0^Z f(\xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|\mathbf{z}_l - \xi|\right) d\xi \\ \dfrac{c_l}{\mathbf{i}\omega} \mathbf{w}_l + \dfrac{c_l \rho_l}{2\mathbf{i}\omega} \displaystyle\int_0^Z f(\xi) \operatorname{sign}(\mathbf{z}_l - \xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l}|\mathbf{z}_l - \xi|\right) d\xi \end{pmatrix}.$$

Recalling Proposition 1, $||M||_\infty \leq 1$, and therefore $|\alpha_l|, |\beta_l| \leq ||A||_\infty \leq ||B||_\infty$.

It remains to bound $||B||_\infty$. We have

$$\begin{aligned} |B_1| &\leq |\mathbf{v}_l| + \frac{c_l \rho_l}{2\omega} ||f||_1 \\ &\leq \frac{C_s}{\mathbf{h}_\star} \omega^{-1} \kappa_l^{-1/2} ||f|| + \frac{c_l \rho_l}{2\omega} ||f||_1 \\ &\leq \left(\frac{C_s}{\mathbf{h}_\star} \kappa_l^{-1/2} + \frac{c_l \rho_l \sqrt{Z}}{2} \right) \omega^{-1} ||f||, \end{aligned}$$

and

$$\begin{aligned} |B_2| &\leq \frac{c_l}{2\omega} |\mathbf{w}_l| + \frac{c_l \rho_l}{2\omega} ||f||_1 \\ &\leq \frac{c_l}{2\omega} \frac{C_s}{\mathbf{h}_\star} \rho_l^{-1/2} ||f|| + \frac{c_l \rho_l}{2\omega} ||f||_1 \\ &\leq \left(\frac{C_s}{2\mathbf{h}_\star} \kappa_l^{-1/2} + \frac{c_l \rho_l \sqrt{Z}}{2} \right) \omega^{-1} ||f||, \end{aligned}$$

so that

$$||B||_\infty \leq \left(\frac{C_s}{\mathbf{h}_\star} \kappa_l^{-1/2} + \frac{c_l \rho_l \sqrt{Z}}{2} \right) \omega^{-1} ||f||.$$

□

Theorem 6. *We have*

$$||\mathcal{S}_{\omega,\kappa,\rho} f||_\infty \leq C_\infty \omega^{-1} ||f||, \quad ||(\mathcal{S}_{\omega,\kappa,\rho} f)'||_\infty \leq C'_\infty ||f||, \tag{2.27}$$

with

$$C_\infty = 2C_{\alpha,\beta}, \quad C'_\infty = \frac{C_\infty}{\min_{l \in \{1,\dots,L\}} c_l}.$$

Proof. Consider an arbitrary layer $l \in \{1, \dots, L\}$. Using Lemma 3, there exist two complex constants $\alpha_l, \beta_l \in \mathbb{C}$ such that

$$\mathbf{u}(z) = \alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} z\right) + \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} z\right) - \frac{\mathbf{i} c_l}{2\omega} \int_0^Z f(\xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l} |z - \xi|\right) d\xi,$$

for $z \in (\mathbf{z}_{l-1}, \mathbf{z}_l)$. Furthermore, we have

$$\mathbf{u}'(z) = \frac{\mathbf{i}\omega}{c_l} \alpha_l \exp\left(\frac{\mathbf{i}\omega}{c_l} z\right) - \frac{\mathbf{i}\omega}{c_l} \beta_l \exp\left(-\frac{\mathbf{i}\omega}{c_l} z\right) - \frac{1}{2} \int_0^Z f(\xi) \operatorname{sign}(\mathbf{z}_l - \xi) \exp\left(-\frac{\mathbf{i}\omega}{c_l} |z - \xi|\right) d\xi,$$

for $z \in (\mathbf{z}_{l-1}, \mathbf{z}_l)$.

It follows that, for $z \in (\mathbf{z}_{l-1}, \mathbf{z}_l)$,

$$|\mathbf{u}(z)| \leq |\alpha_l| + |\beta_l| + \frac{c_l}{2\omega} \int_0^Z |f(z)| dz,$$

and

$$|\mathbf{u}'(z)| \leq \frac{\omega}{c_l} |\alpha_l| + \frac{\omega}{c_l} |\beta_l| + \frac{1}{2} \int_0^Z |f(z)| dz.$$

Hence, (2.27) follows from (2.26), since

$$\int_0^Z |f(z)| dz \leq \sqrt{Z} ||f||.$$

□

2.1.6 Stability with respect to the medium parameters

Stability estimates with respect to the medium parameters are rarely tackled in the literature even if they have an important role to play.

First, in the context of inverse problems, the dependency of the solution with respect to medium parameters is at stake.

Second, in the context of multiscale numerical methods, it might be of interest to approximate the medium parameters. For instance, in this work, we use a piecewise constant approximation of the velocity in order to simplify the computations of integral coefficients related to finite element approximation. In recent developments of plane-wave methods, the so-called generalized plane-waves have been introduced [61, 96]. In homogeneous media, plane waves are homogeneous solutions to the Helmholtz equation and they might be used as finite element shape functions to approximate the solution. The idea of generalized plane-wave is to locally approximate the wavespeed (by a linearization in [96] or a higher order approximation in [61]) in such a way that analytical homogeneous solutions are available. These homogeneous solutions are the so-called generalized plane-waves.

In order to analyse such numerical methods, it is clear that the approximation of the wavespeed has a major impact. Indeed the generalized plane wave are not exactly homogeneous solutions to the Helmholtz equation with the real wavespeed. As a matter of fact, it turns out that frequency-explicit convergence analysis of generalized plane-wave method is not available yet.

In Theorem 7 we provide a frequency-explicit stability result for the 1D case. For this purpose, we first establish Proposition 2. We recall that M is the set of admissible propagation media defined in Definition 1.

Proposition 2. *Let $(\kappa, \rho) \in M$. Then, for all $v \in H^1(0, Z, \mathbb{C})$, we have*

$$||v||_\infty \leq C_{tr}\omega^{-1/2}||v||_{\omega,\kappa,\rho}, \tag{2.28}$$

with

$$C_{tr} = \max\left\{1, \sqrt{\left(\frac{1}{Z} + \rho^\star\right)\kappa^\star}\right\}.$$

Proof. First, for all $x, y \in (0, Z)$ it holds that

$$\begin{aligned}
|v(x)|^2 - |v(y)|^2 &= \int_y^x \frac{d}{dz}\left(|v|^2\right)(\xi)d\xi \\
&= 2\,\mathrm{Re}\int_y^x v(\xi)\overline{v'(\xi)}d\xi \\
&= 2\,\mathrm{Re}\int_y^x \left(\rho^{1/2}v(\xi)\right)\left(\rho^{-1/2}\overline{v'(\xi)}\right)d\xi \\
&\leq 2||\rho^{1/2}v||\;||\rho^{-1/2}v'||.
\end{aligned}$$

We thus have

$$|v(x)|^2 \leq |v(y)|^2 + 2\sqrt{\rho^\star}||v||\;||\rho^{-1/2}v'||.$$

for all $x, y \in (0, Z)$. Integrating with respect to y over $(0, Z)$, and dividing by Z, we get

$$|v(x)|^2 \leq \frac{1}{Z}||v||^2 + 2\sqrt{\rho^\star}||v||\;||\rho^{-1/2}v'||,$$

Using the algebraic inequality $2ab \leq \omega a^2 + \omega^{-1}b^2$, and the assumption that $\omega \geq 1$, we can conclude:

$$\begin{aligned}
||v||_\infty^2 &\leq \frac{1}{Z}||v||^2 + \rho^\star\omega||v||^2 + \omega^{-1}||\rho^{-1/2}v'||^2 \\
&\leq \omega^{-1}\left\{\left(\frac{\omega^{-1}}{Z} + \rho^\star\right)\omega^2||v||^2 + ||\rho^{-1/2}v'||^2\right\} \\
&\leq \omega^{-1}\left\{\left(\frac{\omega^{-1}}{Z} + \rho^\star\right)\omega^2\kappa^\star||\kappa^{-1/2}v||^2 + ||\rho^{-1/2}v'||^2\right\} \\
&\leq \max\left\{1, \kappa^\star\left(\frac{1}{Z} + \rho^\star\right)\right\}\omega^{-1}||v||_{\omega,\kappa,\rho}^2.
\end{aligned}$$

□

Theorem 7. *Consider two propagation media defined by the parameters $(\kappa_1, \rho_1) \in M$ and $(\kappa_2, \rho_2) \in M$. Then*

$$\begin{aligned}
&||\mathcal{S}_{\omega,\kappa_1,\rho_1}f - \mathcal{S}_{\omega,\kappa_2,\rho_2}f|| \\
&\leq C_{s,m}||f||\left(||\kappa_1^{-1} - \kappa_2^{-1}|| + ||\rho_1^{-1} - \rho_2^{-1}|| + \omega^{-1/2}|\kappa_1(Z)\rho_1(Z))^{-1/2} - (\kappa_2(Z)\rho_2(Z))^{-1/2}|\right),
\end{aligned}$$

and

$$\begin{aligned}
&||(\mathcal{S}_{\omega,\kappa_1,\rho_1}f)' - (\mathcal{S}_{\omega,\kappa_2,\rho_2}f)'|| \\
&\leq C_{s,m}\omega||f||\left(||\kappa_1^{-1} - \kappa_2^{-1}|| + ||\rho_1^{-1} - \rho_2^{-1}|| + \omega^{-1/2}|\kappa_1(Z)\rho_1(Z))^{-1/2} - (\kappa_2(Z)\rho_2(Z))^{-1/2}|\right),
\end{aligned}$$

where $C_{s,m}$ is a constant independent of ω, (κ_1, ρ_1), (κ_2, ρ_2) and f.

Proof. To simplify the notations, let us write $\mathbf{u}_1 = \mathcal{S}_{\omega,\kappa_1,\rho_1}f$ and $\mathbf{u}_2 = \mathcal{S}_{\omega,\kappa_2,\rho_2}f$. First, since $\mathbf{u}_1, \mathbf{u}_2$ satisfy problem (2.7), it is clear that

$$B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_1, v) = B_{\omega,\kappa_2,\rho_2}(\mathbf{u}_2, v)$$

for all $v \in H^1(0, Z, \mathbb{C})$. Hence, we have

$$B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_1 - \mathbf{u}_2, v) = B_{\omega,\kappa_2,\rho_2}(\mathbf{u}_2, v) - B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_2, v), \tag{2.29}$$

for all $v \in H^1(0, Z, \mathbb{C})$.

Using Theorem 5, we can lower bound the left-hand-side of (2.29) by the norm of $\mathbf{u}_1 - \mathbf{u}_2$. Focusing on the right-hand-side of (2.29), we have

$$\begin{aligned}
B_{\omega,\kappa_2,\rho_2}(\mathbf{u}_2, v) - B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_2, v) = &- \omega^2\int_0^Z \left(\kappa_2^{-1} - \kappa_1^{-1}\right)(z)\mathbf{u}_2(z)v(z)dz \\
&- \mathbf{i}\omega\left(\kappa_2(Z)\rho_2(Z))^{-1/2} - (\kappa_1(Z)\rho_1(Z))^{-1/2}\right)\mathbf{u}_2(Z)v(Z) \\
&+ \int_0^Z \left(\rho_2^{-1} - \rho_1^{-1}\right)(z)\mathbf{u}_2'(z)v'(z)dz,
\end{aligned}$$

and

$$\begin{aligned}
|B_{\omega,\kappa_2,\rho_2}(\mathbf{u}_2,v)-B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_2,v)| &\leq \omega^2||\kappa_2^{-1}-\kappa_1^{-1}||\,||\mathbf{u}_2||_\infty\,||v|| \\
&+ \omega|\kappa_2(Z)\rho_2(Z))^{-1/2}-(\kappa_1(Z)\rho_1(Z))^{-1/2}|\,||\mathbf{u}_2||_\infty\,||v||_\infty \\
&+ ||\rho_2^{-1}-\rho_1^{-1}||\,||\mathbf{u}_2'||_\infty\,||v'|| \\
&\leq C_\infty\omega||\kappa_2^{-1}-\kappa_1^{-1}||\,||f||\,||v|| \\
&+ C_\infty|\kappa_2(Z)\rho_2(Z))^{-1/2}-(\kappa_1(Z)\rho_1(Z))^{-1/2}|\,||f||\,||v||_\infty \\
&+ C'_\infty||\rho_2^{-1}-\rho_1^{-1}||\,||f||\,||v'|| \\
&\leq (C_\infty||\kappa_2^{-1}-\kappa_1^{-1}|| \\
&+ C_\infty C_{tr}\omega^{-1/2}|\kappa_2(Z)\rho_2(Z))^{-1/2}-(\kappa_1(Z)\rho_1(Z))^{-1/2}| \\
&+ C'_\infty||\rho_2^{-1}-\rho_1^{-1}||)\,||f||\,||v||_{\omega,\kappa_1,\rho_1}.
\end{aligned}$$

Using Proposition 2, we have

$$\frac{|B_{\omega,\kappa_2,\rho_2}(\mathbf{u}_2,v)-B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_2,v)|}{||v||_{\omega,\kappa_1,\rho_1}} \leq \max((1+C_{tr})C_\infty, C'_\infty)\times$$
$$(||\kappa_2^{-1}-\kappa_1^{-1}||+\omega^{-1/2}|(\kappa_2(Z)\rho_2(Z))^{-1/2}-(\kappa_1(Z)\rho_1(Z))^{-1/2}|+||\rho_2^{-1}-\rho_1^{-1}||)||f||. \quad (2.30)$$

We conclude with Theorem 5. We have,

$$||\mathbf{u}_1-\mathbf{u}_2||_{\omega,\kappa_1,\rho_1} \leq (1+2C_{s,1}\omega)\sup_{v\in H^1(0,Z,\mathbb{C})}\frac{|B_{\omega,\kappa_1,\rho_1}(\mathbf{u}_1-\mathbf{u}_2,v)|}{||v||_{\omega,\kappa_1,\rho_1}},$$

and the proof follows from (2.29) and (2.30). □

2.1.7 Proof of Theorems 3 and 4

The main tool of the demonstration of Theorems 3 and 4 is a set of discontinuous test functions $\{\phi_l\}_{l=1}^L$. Each ϕ_l has a support contained in $[\mathbf{z}_l,\mathbf{z}_{l-1}]$ and depends on a parameter $\hat{\mathbf{z}}_l$ that will be set in the end of demonstration.

Definition 4. *Let $l\in\{1,\ldots,L\}$, we define $\phi_l:(\mathbf{z}_l,\mathbf{z}_{l-1})\to\mathbb{C}$ by*

$$\phi_l(z)=(z-\hat{\mathbf{z}}_l)(\mathcal{S}_{\omega,\kappa,\rho}f)'(z),$$

where $\hat{\mathbf{z}}_l\in\mathbb{R}$ is a real constant. Observe that, since $\mathcal{S}_{\omega,\kappa,\rho}f\in H^2(\mathbf{z}_{l-1},\mathbf{z}_l,\mathbb{C})$, $\phi_l\in H^1(\mathbf{z}_{l-1},\mathbf{z}_l,\mathbb{C})$.

For ease of presentation we will write $\mathbf{u}=\mathcal{S}_{\omega,\kappa,\rho}f$. We define, for each layer $l\in\{1,\ldots,L\}$, $\mathcal{H}_l\mathbf{u}\in L^2(\mathbf{z}_{l-1},\mathbf{z}_l,\mathbb{C})$ by

$$\big(\mathcal{H}_l\mathbf{u}\big)(z)=-\frac{\omega^2}{\kappa_l}\mathbf{u}(z)-\frac{1}{\rho_l}\mathbf{u}''(z),\quad z\in(\mathbf{z}_{l-1},\mathbf{z}_l).$$

We also introduce

$$\mathcal{H}\mathbf{u}=-\frac{\omega^2}{\kappa}\mathbf{u}-\left(\frac{1}{\rho}\mathbf{u}'\right)'.$$

The proof is divided in 3 distinct parts.

Part I

The first step of the demonstration is to obtain identity (2.31) recorded in Proposition 3.

Proposition 3. *For any $f \in L^2(0, Z, \mathbb{C})$, we have*

$$\begin{aligned}
&\int_0^Z \left(\frac{\omega^2}{\kappa(z)} |\mathbf{u}(z)|^2 + \frac{1}{\rho(z)} |\mathbf{u}'(z)|^2 \right) dz \\
&+ \omega^2 \sum_{l=1}^{L-1} \left\{ \frac{1}{\kappa_{l+1}} (\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \frac{1}{\kappa_l} (\mathbf{z}_l - \hat{\mathbf{z}}_l) \right\} |\mathbf{v}_l|^2 + \sum_{l=1}^{L-1} \left\{ \rho_{l+1} (\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \rho_l (\mathbf{z}_l - \hat{\mathbf{z}}_l) \right\} |\mathbf{w}_l|^2 \\
&= (Z - \hat{\mathbf{z}}_L) \left(\frac{\omega^2}{\kappa_L} |\mathbf{v}_L|^2 + \rho_L |\mathbf{w}_L|^2 \right) - (0 - \hat{\mathbf{z}}_1) \left(\frac{\omega^2}{\kappa_1} |\mathbf{v}_0|^2 + \rho_1 |\mathbf{w}_0|^2 \right) + 2 \operatorname{Re} \left\{ \sum_{l=1}^{L} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z) \overline{\phi_l(z)} dz \right\}.
\end{aligned} \tag{2.31}$$

We obtain identity (2.31) by applying each test function ϕ_l against $\mathcal{H}_l \mathbf{u}$. The proof is based on the Rellich identity (which is equivalent to integrating by parts in 1D) and the fact that $\mathcal{H}\mathbf{u} = f$. The proof is straightforward, but requires a lot of hand computations. We thus introduce three preliminary Lemmas.

In Lemma 4 we obtain identities (2.33) and (2.34) using integration by parts together with the derivation rule

$$\left(|\psi|^2 \right)' = 2 \operatorname{Re} \left(\psi' \overline{\psi} \right), \tag{2.32}$$

for complex functions $\psi \in H^1$.

Lemma 4. *For $l \in \{1, \ldots, L\}$, it holds that*

$$2 \operatorname{Re} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z) \overline{\phi_l(z)} dz = - \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} |\mathbf{u}(z)|^2 dz + (\mathbf{z}_l - \hat{\mathbf{z}}_l) |\mathbf{v}_l|^2 - (\mathbf{z}_{l-1} - \hat{\mathbf{z}}_l) |\mathbf{v}_{l-1}|^2, \tag{2.33}$$

and

$$2 \operatorname{Re} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z) \overline{\phi_l(z)} dz = - \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} |\mathbf{u}'(z)|^2 dz + (\mathbf{z}_l - \hat{\mathbf{z}}_l) \rho_l^2 |\mathbf{w}_l|^2 - (\mathbf{z}_{l-1} - \hat{\mathbf{z}}_l) \rho_{l-1}^2 |\mathbf{w}_{l-1}|^2. \tag{2.34}$$

Proof. As a starting point, we use the fact that $z - \hat{\mathbf{z}}_l$ is real to write

$$\begin{aligned}
2 \operatorname{Re} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z) \overline{\phi_l(z)} dz &= 2 \operatorname{Re} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z) \overline{(z - \hat{\mathbf{z}}_l) \mathbf{u}'(z)} dz \\
&= \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} (z - \hat{\mathbf{z}}_l) 2 \operatorname{Re} \left(\mathbf{u}(z) \overline{\mathbf{u}'(z)} \right) dz,
\end{aligned}$$

and

$$\begin{aligned}
2 \operatorname{Re} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z) \overline{\phi_l(z)} dz &= 2 \operatorname{Re} \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z) \overline{(z - \hat{\mathbf{z}}_l) \mathbf{u}'(z)} dz \\
&= \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} (z - \hat{\mathbf{z}}_l) 2 \operatorname{Re} \left(\mathbf{u}''(z) \overline{\mathbf{u}'(z)} \right) dz.
\end{aligned}$$

The chain rule (2.32) enables us to derive

$$2\,\mathrm{Re}\left(\mathbf{u}(z)\overline{\mathbf{u}'(z)}\right) = \frac{d}{dz}\left(|\mathbf{u}|^2\right)(z), \quad 2\,\mathrm{Re}\left(\mathbf{u}''(z)\overline{\mathbf{u}'(z)}\right) = \frac{d}{dz}\left(|\mathbf{u}'|^2\right)(z).$$

Lemma 4 follows from integration by parts:

$$\begin{aligned} 2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z)\overline{\phi_l(z)}dz &= \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} (z-\hat{\mathbf{z}}_l)\frac{d}{dz}\left(|\mathbf{u}|^2\right)(z)dz. \\ &= -\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \frac{d}{dz}\left(z-\hat{\mathbf{z}}_l\right)|\mathbf{u}(z)|^2 dz + \left[(z-\hat{\mathbf{z}}_l)|\mathbf{u}(z)|^2\right]_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \\ &= -\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} |\mathbf{u}(z)|^2 dz + (\mathbf{z}_l-\hat{\mathbf{z}}_l)|\mathbf{u}(\mathbf{z}_l^-)|^2 - (\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)|\mathbf{u}(\mathbf{z}_{l-1}^+)|^2, \end{aligned}$$

and

$$\begin{aligned} 2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z)\overline{\phi_l(z)}dz &= \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} (z-\hat{\mathbf{z}}_l)\frac{d}{dz}\left(|\mathbf{u}'|^2\right)(z)dz. \\ &= -\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \frac{d}{dz}\left(z-\hat{\mathbf{z}}_l\right)|\mathbf{u}'(z)|^2 dz + \left[(z-\hat{\mathbf{z}}_l)|\mathbf{u}'(z)|^2\right]_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \\ &= -\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} |\mathbf{u}'(z)|^2 dz + (\mathbf{z}_l-\hat{\mathbf{z}}_l)|\mathbf{u}'(\mathbf{z}_l^-)|^2 - (\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)|\mathbf{u}'(\mathbf{z}_{l-1}^+)|^2. \end{aligned}$$

The results follow by the definition of $\mathbf{v}_{l-1}, \mathbf{v}_l, \mathbf{w}_{l-1}$ and $\mathbf{w}_l$. □

In Lemma 5, we test $\mathcal{H}_l\mathbf{u}$ upon the test functions ϕ_l. We use the identities (2.33) and (2.34) to simplify the expression, yielding identity (2.35).

Lemma 5. *For $l \in \{1, \ldots, L\}$, we have*

$$\begin{aligned} 2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} (\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz &= \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz \\ &+ (\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_{l-1}|^2 + \rho_l|\mathbf{w}_{l-1}|^2\right) \\ &- (\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) \end{aligned} \tag{2.35}$$

Proof. By definition of $\mathcal{H}_l\mathbf{u}$, we have

$$2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} (\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz = -\frac{\omega^2}{\kappa_l}\left(2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}(z)\overline{\phi_l(z)}dz\right) - \frac{1}{\rho_l}\left(2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} \mathbf{u}''(z)\overline{\phi_l(z)}dz\right).$$

We can simplify the right-hand-side thanks to identities (2.33), (2.34) and by grouping the terms

$$\begin{aligned}
2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz &= -\frac{\omega^2}{\kappa_l}\left(-\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}|\mathbf{u}(z)|^2dz+(\mathbf{z}_l-\hat{\mathbf{z}}_l)|\mathbf{v}_l|^2-(\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)|\mathbf{v}_{l-1}|^2\right)\\
&\quad -\frac{1}{\rho_l}\left(-\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}|\mathbf{u}'(z)|^2dz+(\mathbf{z}_l-\hat{\mathbf{z}}_l)\rho_l^2|\mathbf{w}_l|^2-(\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)\rho_l^2|\mathbf{w}_{l-1}|^2\right)\\
&= \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}\left(\frac{\omega^2}{\kappa_l}|\mathbf{u}(z)|^2+\frac{1}{\rho_l}|\mathbf{u}'(z)|^2\right)dz\\
&+ (\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_{l-1}|^2+\rho_l|\mathbf{w}_{l-1}|^2\right)\\
&- (\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2+\rho_l|\mathbf{w}_l|^2\right).
\end{aligned}$$

□

Lemma 6. *We have*

$$\begin{aligned}
2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz\right\} &= \int_0^Z\left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2+\frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right)dz\\
&+ \omega^2\sum_{l=1}^{L-1}\left\{\frac{1}{\kappa_{l+1}}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})-\frac{1}{\kappa_l}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\right\}|\mathbf{v}_l|^2\\
&+ \sum_{l=1}^{L-1}\left\{\rho_{l+1}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})-\rho_l(\mathbf{z}_l-\hat{\mathbf{z}}_l)\right\}|\mathbf{w}_l|^2\\
&+ (0-\hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2+\rho_1|\mathbf{w}_0|^2\right)\\
&- (Z-\hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2+\rho_l|\mathbf{w}_L|^2\right)
\end{aligned}$$

Proof. First, as a direct consequence of (2.35), we obviously have

$$\begin{aligned}
2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz\right\} &= \sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}\left(\frac{\omega^2}{\kappa_l}|\mathbf{u}(z)|^2+\frac{1}{\rho_l}|\mathbf{u}'(z)|^2\right)dz\\
&+ \sum_{l=1}^{L}(\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_{l-1}|^2+\rho_l|\mathbf{w}_{l-1}|^2\right)\\
&- \sum_{l=1}^{L}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2+\rho_l|\mathbf{w}_l|^2\right).
\end{aligned}$$

Then the first term simplifies like

$$\begin{aligned}\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}\left(\frac{\omega^2}{\kappa_l}|\mathbf{u}(z)|^2+\frac{1}{\rho_l}|\mathbf{u}'(z)|^2\right)dz &= \sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}\left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2+\frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right)dz\\ &= \int_0^Z\left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2+\frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right)dz.\end{aligned}$$

We can simplify the second term as

$$\begin{aligned}\sum_{l=1}^{L}(\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_{l-1}|^2+\rho_l|\mathbf{w}_{l-1}|^2\right) &= (\mathbf{z}_0-\hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2+\rho_1|\mathbf{w}_0|^2\right)\\ &+ \sum_{l=2}^{L}(\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_{l-1}|^2+\rho_l|\mathbf{w}_{l-1}|^2\right)\\ &= (0-\hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2+\rho_1|\mathbf{w}_0|^2\right)\\ &+ \sum_{l=l}^{L-1}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})\left(\frac{\omega^2}{\kappa_{l+1}}|\mathbf{v}_l|^2+\rho_{l+1}|\mathbf{w}_l|^2\right).\end{aligned}$$

We rewrite the last term as

$$\begin{aligned}\sum_{l=1}^{L}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2+\rho_l|\mathbf{w}_l|^2\right) &= \sum_{l=1}^{L-1}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2+\rho_l|\mathbf{w}_l|^2\right)\\ &+ (\mathbf{z}_L-\hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2+\rho_l|\mathbf{w}_L|^2\right)\\ &= \sum_{l=1}^{L-1}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2+\rho_l|\mathbf{w}_l|^2\right)\\ &+ (Z-\hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2+\rho_l|\mathbf{w}_L|^2\right).\end{aligned}$$

Regrouping the three terms, we get

$$
\begin{aligned}
2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz\right\} &= \int_0^Z\left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2+\frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right)dz\\
&+ (0-\hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2+\rho_1|\mathbf{w}_0|^2\right)\\
&+ \sum_{l=1}^{L-1}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})\left(\frac{\omega^2}{\kappa_{l+1}}|\mathbf{v}_l|^2+\rho_{l+1}|\mathbf{w}_l|^2\right)\\
&- \sum_{l=1}^{L-1}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2+\rho_l|\mathbf{w}_l|^2\right)\\
&- (Z-\hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2+\rho_l|\mathbf{w}_L|^2\right)\\
&= \int_0^Z\left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2+\frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right)dz\\
&+ \omega^2\sum_{l=1}^{L-1}\left\{\frac{1}{\kappa_{l+1}}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})-\frac{1}{\kappa_l}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\right\}|\mathbf{v}_l|^2\\
&+ \sum_{l=1}^{L-1}\left\{\rho_{l+1}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})-\rho_l(\mathbf{z}_l-\hat{\mathbf{z}}_l)\right\}|\mathbf{w}_l|^2\\
&+ (0-\hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2+\rho_1|\mathbf{w}_0|^2\right)\\
&- (Z-\hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2+\rho_l|\mathbf{w}_L|^2\right).
\end{aligned}
$$

□

We are now ready to establish Proposition 3. We obtain identity (2.31) using the fact that $\mathbf{u}$ satisfies (2.2) together with Lemma 6.

Proof of Proposition 3. Since $\mathbf{u}$ satisfies (2.2), we have $(\mathcal{H}\mathbf{u})(z) = f(z)$ for a.e. $z \in (0, Z)$. Therefore

$$\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz = \int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz,$$

for all $l \in \{1, \ldots, L\}$. Then

$$2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(\mathcal{H}_l\mathbf{u})(z)\overline{\phi_l(z)}dz\right\} = 2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz.\right\},$$

and the result follows from Lemma 6. □

Part II

In Definition 4 of the test functions ϕ_l, we have left the definition of the coefficients $\hat{\mathbf{z}}_l$ undefined. The second step of demonstration is to derive Proposition 4 by carefully selecting the coefficients $\hat{\mathbf{z}}_l$.

Proposition 4. *We have*

$$\frac{1}{2}\int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz + \sum_{l=1}^{L-1} \mathbf{h}_l \left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) \leq 18\mathcal{M}^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2 \left|\left|f\right|\right|^2. \quad (2.36)$$

Again, the proof is based on a lot of hand computations. After giving the definition of the coefficients $\hat{\mathbf{z}}_l$, we demonstrate Proposition 4 using 4 preliminary Lemmas. To simplify the notations, we write

$$\mathbf{h}_l = \mathbf{z}_l - \mathbf{z}_{l-1}, \quad \hat{\mathbf{h}}_l = \mathbf{z}_{l-1} - \hat{\mathbf{z}}_l.$$

Definition 5. *We define the sequence* $\hat{\mathbf{h}} = \{\hat{\mathbf{h}}_l\}_{l=1}^L$ *recursively by* $\hat{\mathbf{h}}_0 = 0$ *and*

$$\hat{\mathbf{h}}_{l+1} = \max\left(\frac{\kappa_{l+1}}{\kappa_l}, \frac{\rho_l}{\rho_{l+1}}, 1\right)\left(\hat{\mathbf{h}}_l + 2\mathbf{h}_l\right).$$

Furthermore, we define $\hat{\mathbf{z}}$ *as*

$$\hat{\mathbf{z}}_l = \mathbf{z}_l - \hat{\mathbf{h}}_l, \quad l \in \{1, \dots, L\}.$$

Lemma 7. $\hat{\mathbf{h}}$ *is an increasing sequence. Furthermore, we have*

$$\hat{\mathbf{h}}_L \leq 2\mathcal{M}Z$$

and

$$\frac{1}{\kappa_{l+1}}(\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \frac{1}{\kappa_l}(\mathbf{z}_l - \hat{\mathbf{z}}_l) \geq \frac{\mathbf{h}_l}{\kappa_l}, \quad \rho_{l+1}(\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \rho_l(\mathbf{z}_l - \hat{\mathbf{z}}_l) \geq \rho_l\mathbf{h}_l.$$

Proof. It is clear that $\hat{\mathbf{h}}$ is an increasing sequence. Indeed, for a given $l \in \{1, \dots, L-1\}$, we have

$$\hat{\mathbf{h}}_{l+1} = \max\left(\frac{\kappa_{l+1}}{\kappa_l}, \frac{\rho_l}{\rho_{l+1}}, 1\right)\left(\hat{\mathbf{h}}_l + 2\mathbf{h}_l\right) \geq \hat{\mathbf{h}}_l + 2\mathbf{h}_l \geq \hat{\mathbf{h}}_l,$$

since $\mathbf{h}_l > 0$.

Then, by recurrence, we have

$$\hat{\mathbf{h}}_l = 2\sum_{n=1}^{l-1}\left\{\prod_{m=1}^{n}\max\left(\frac{\kappa_{m+1}}{\kappa_m}, \frac{\rho_m}{\rho_{m+1}}, 1\right)\right\}\mathbf{h}_n.$$

It follows that

$$\begin{aligned}
\hat{\mathbf{h}}_L &= 2\sum_{n=1}^{L-1}\left\{\prod_{m=1}^{n}\max\left(\frac{\kappa_{m+1}}{\kappa_m},\frac{\rho_m}{\rho_{m+1}},1\right)\right\}\mathbf{h}_n \\
&\leq 2\left\{\prod_{m=1}^{L-1}\max\left(\frac{\kappa_{m+1}}{\kappa_m},\frac{\rho_m}{\rho_{m+1}},1\right)\right\}\sum_{n=1}^{L-1}\mathbf{h}_n \\
&\leq 2\left\{\prod_{m=1}^{L-1}\max\left(\frac{\kappa_{m+1}}{\kappa_m},\frac{\rho_m}{\rho_{m+1}},1\right)\right\}Z \\
&\leq 2\mathcal{M}Z.
\end{aligned}$$

Finally, we have

$$\hat{\mathbf{h}}_{l+1} = \max\left(\frac{\kappa_{l+1}}{\kappa_l},\frac{\rho_l}{\rho_{l+1}},1\right)\left(\hat{\mathbf{h}}_l + 2\mathbf{h}_l\right),$$

therefore

$$\hat{\mathbf{h}}_{l+1} \geq \frac{\kappa_{l+1}}{\kappa_l}\left(\hat{\mathbf{h}}_l + 2\mathbf{h}_l\right), \quad \hat{\mathbf{h}}_{l+1} \geq \frac{\rho_l}{\rho_{l+1}}\left(\hat{\mathbf{h}}_l + 2\mathbf{h}_l\right),$$

so that

$$\frac{1}{\kappa_{l+1}}\hat{\mathbf{h}}_{l+1} - \frac{1}{\kappa_l}(\hat{\mathbf{h}}_l + \mathbf{h}_l) \geq \frac{1}{\kappa_l}\mathbf{h}_l, \quad \rho_{l+1}\hat{\mathbf{h}}_{l+1} - \rho_l(\hat{\mathbf{h}}_l + \mathbf{h}_l) \geq \rho_l\mathbf{h}_l,$$

and we conclude the proof of Lemma 7, since by definition of $\hat{\mathbf{h}}$, $\hat{\mathbf{h}}_{l+1} = \mathbf{z}_l - \hat{\mathbf{z}}_{l+1}$ and $\hat{\mathbf{h}}_l + \mathbf{h}_l = \mathbf{z}_l - \hat{\mathbf{z}}_l$. □

Lemma 7 ensures that every term in the left-hand-side of (2.31) is positive. More precisely, we have

$$\begin{aligned}
\mathbf{h}_l\left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) &\leq \omega^2\left\{\frac{1}{\kappa_{l+1}}(\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \frac{1}{\kappa_l}(\mathbf{z}_l - \hat{\mathbf{z}}_l)\right\}|\mathbf{v}_l|^2 \\
&+ \left\{\rho_{l+1}(\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \rho_l(\mathbf{z}_l - \hat{\mathbf{z}}_l)\right\}|\mathbf{w}_l|^2.
\end{aligned}$$

Then, it remains to bound the right-hand-side of (2.31) to conclude the demonstration of Proposition 4. The right-hand-side reads

$$(Z-\hat{\mathbf{z}}_L)\left(\frac{1}{\kappa_L}|\mathbf{v}_L|^2 + \rho_L|\mathbf{w}_L|^2\right) - (0-\hat{\mathbf{z}}_1)\left(\frac{1}{\kappa_1}|\mathbf{v}_0|^2 + \rho_L|\mathbf{w}_0|^2\right) + 2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz\right\},$$

and can be rewritten as

$$(Z - \hat{\mathbf{z}}_L)\left(\frac{1}{\kappa_L}|\mathbf{v}_L|^2 + \rho_L|\mathbf{w}_L|^2\right) + 2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz\right\},$$

since $0 - \hat{\mathbf{z}}_1 = \hat{\mathbf{h}}_1 = 0$.

The term involving f is bounded in Lemma 8 with simple algebraic arguments. The bound for the other term is derived in Lemmas 9 and 10. The boundary conditions (2.5) and (2.6) at $z = 0$ and $z = Z$ are involved in this proof.

Lemma 8. *For any $f \in L^2(0, Z, \mathbb{C})$, we have*

$$2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(v)\overline{\phi_l(z)}dz\right\} \leq 2\rho^\star(Z+\hat{\mathbf{h}}_L)^2\,||f||^2 + \frac{1}{2}\int_0^Z \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2dz. \tag{2.37}$$

Proof. First, let us consider $l \in \{1,\dots,L\}$. We have

$$\begin{aligned} 2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz &= 2\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}(z-\hat{\mathbf{z}}_l)\,\mathrm{Re}\left(f(z)\overline{\mathbf{u}'(z)}\right)dz \\ &\leq 2\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}|z-\hat{\mathbf{z}}_l||f(z)||\mathbf{u}'(z)|dz. \end{aligned}$$

Then, for all $z \in (\mathbf{z}_{l-1}, \mathbf{z}_l)$, we have

$$|z-\hat{\mathbf{z}}_l| \leq \max(|\mathbf{z}_{l-1}-\hat{\mathbf{z}}_l|, |\mathbf{z}_l-\hat{\mathbf{z}}_l|) = \max(|\hat{\mathbf{h}}_l|, |\mathbf{h}_l+\hat{\mathbf{h}}_l|) = (\mathbf{h}_l+\hat{\mathbf{h}}_l).$$

Hence, we get

$$\begin{aligned} 2|z-\hat{\mathbf{z}}_l||f(z)||\mathbf{u}'(z)| &\leq 2(\mathbf{h}_l+\hat{\mathbf{h}}_l)|f(z)||\mathbf{u}'(z)| \\ &\leq 2\left(\sqrt{2\rho_l}(\mathbf{h}_l+\hat{\mathbf{h}}_l)|f(z)|\right)\left(\frac{1}{\sqrt{2\rho_l}}|\mathbf{u}'(z)|\right) \\ &\leq 2\rho_l(\mathbf{h}_l+\hat{\mathbf{h}}_l)^2|f(z)|^2 + \frac{1}{2\rho_l}|\mathbf{u}'(z)|^2. \end{aligned}$$

Since we know that $\hat{\mathbf{h}}$ is an increasing sequence, we have $\hat{\mathbf{h}}_l \leq \hat{\mathbf{h}}_L$. Furthermore, $\mathbf{h}_l \leq Z$ and $\rho_l \leq \rho^\star$, so that

$$2|z-\hat{\mathbf{z}}_l||f(z)||\mathbf{u}'(z)| \leq 2\rho^\star(Z+\hat{\mathbf{h}}_L)^2|f(z)|^2 + \frac{1}{2\rho_l}|\mathbf{u}'(z)|^2.$$

It follows that

$$2\,\mathrm{Re}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz \leq 2\rho^\star(Z+\hat{\mathbf{h}}_L)^2\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}|f(z)|^2dz + \frac{1}{2}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}\frac{1}{\rho_l}|\mathbf{u}'(z)|^2dz,$$

for each $l \in \{1,\dots,L\}$.

We obtain the desired result by summation over l. □

In Lemma 9, we prepare the proof of Lemma 10 using the absorbing boundary condition at $z = Z$ and the Neumann or Dirichlet boundary condition at $z = 0$.

Lemma 9. *We have*

$$\mathrm{Im}\int_0^Z(\mathcal{H}\mathbf{u})(z)\overline{\mathbf{u}(z)}dz = -\frac{\omega}{\sqrt{\kappa_L\rho_L}}|\mathbf{u}(Z)|^2. \tag{2.38}$$

Proof. By integrating by parts, we have

$$\begin{aligned}\int_0^Z (\mathcal{H}\mathbf{u})(z)\overline{\mathbf{u}(z)}dz &= \int_0^Z \left(-\frac{\omega^2}{\kappa(z)}\mathbf{u}(z) - \left(\frac{1}{\rho}\mathbf{u}'\right)'(z)\right)\overline{\mathbf{u}(z)}dz \\ &= \int_0^Z -\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2dz - \int_0^Z \left(\frac{1}{\rho}\mathbf{u}'\right)'(z)\overline{\mathbf{u}(z)}dz \\ &= \int_0^Z -\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2dz + \int_0^Z \frac{1}{\rho(z)}\mathbf{u}'(z)\overline{\mathbf{u}'(z)}dz - \left[\frac{1}{\rho(z)}\mathbf{u}'(z)\overline{\mathbf{u}(z)}\right]_0^Z \\ &= \int_0^Z -\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2dz + \int_0^Z \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2dz - \left[\frac{1}{\rho(z)}\mathbf{u}'(z)\overline{\mathbf{u}(z)}\right]_0^Z \quad (2.39)\end{aligned}$$

It is clear that the first two terms of (2.39) are real. Therefore

$$\operatorname{Im}\int_0^Z (\mathcal{H}\mathbf{u})(z)\overline{\mathbf{u}(z)}dz = -\operatorname{Im}\left[\frac{1}{\rho(z)}\mathbf{u}'(z)\overline{\mathbf{u}(z)}\right]_0^Z = \frac{1}{\rho_1}\operatorname{Im}\left(\mathbf{u}'(0)\overline{\mathbf{u}(0)}\right) - \frac{1}{\rho_L}\operatorname{Im}\left(\mathbf{u}'(Z)\overline{\mathbf{u}(Z)}\right).$$

Because of the boundary condition (2.5), $\mathbf{u}'(0) = 0$, and because of the boundary condition (2.6)

$$\frac{1}{\rho_L}\mathbf{u}'(Z)\overline{\mathbf{u}(Z)} = \frac{1}{\sqrt{\rho_L}}\left(\frac{1}{\sqrt{\rho_L}}\mathbf{u}'(Z)\right)\overline{\mathbf{u}(Z)} = \frac{1}{\sqrt{\rho_L}}\left(\frac{\mathbf{i}\omega}{\sqrt{\kappa_L}}\mathbf{u}(Z)\right)\overline{\mathbf{u}(Z)} = \frac{\mathbf{i}\omega}{\sqrt{\kappa_L\rho_L}}|\mathbf{u}(Z)|^2.$$

□

Lemma 10. *For any $f \in L^2(0, Z, \mathbb{C})$, we have*

$$(Z - \hat{\mathbf{z}}_L)\left(\frac{1}{\kappa_L}|\mathbf{v}_L|^2 + \rho_L|\mathbf{w}_L|^2\right) \leq 2(Z + \hat{\mathbf{h}}_L)^2\frac{\rho_L}{\kappa_L}\kappa^\star\,||f||^2 + \frac{1}{2}\int_0^Z \frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2dz.$$

Proof. Since $\mathbf{u}$ satisfies the absorbing boundary condition

$$\frac{1}{\sqrt{\rho_L}}\mathbf{u}'(Z) - \frac{\mathbf{i}\omega}{\sqrt{\kappa_L}}\mathbf{u}(Z) = 0,$$

we have

$$\frac{1}{\rho_L}|\mathbf{u}'(Z)|^2 = \frac{\omega^2}{\kappa_L}|\mathbf{u}(Z)|^2,$$

and therefore

$$\frac{1}{\kappa_L}|\mathbf{v}_L|^2 + \rho_L|\mathbf{w}_L|^2 = \frac{\omega^2}{\kappa_L}|\mathbf{u}(Z)|^2.$$

Now using Lemma 9, we have

$$-\frac{\omega}{\sqrt{\kappa_L\rho_L}}|\mathbf{u}(Z)|^2 = \operatorname{Im}\int_0^Z (\mathcal{H}\mathbf{u})(z)\overline{\mathbf{u}(z)}dz = \operatorname{Im}\int_0^Z f(z)\overline{\mathbf{u}(z)}dz,$$

so that

$$\begin{aligned}
(Z-\hat{\mathbf{z}}_L)\left(\frac{1}{\kappa_L}|\mathbf{v}_L|^2+\rho_L|\mathbf{w}_L|^2\right) &= 2(Z-\hat{\mathbf{z}}_L)\frac{\omega^2}{\kappa_L}|\mathbf{u}(Z)|^2 \\
&= 2(Z-\hat{\mathbf{z}}_L)\omega\frac{\sqrt{\rho}_L}{\sqrt{\kappa_L}}\left(\frac{\omega}{\sqrt{\kappa_L\rho_L}}|\mathbf{u}(Z)|^2\right) \\
&\leq 2(Z-\hat{\mathbf{z}}_L)\omega\frac{\sqrt{\rho}_L}{\sqrt{\kappa_L}}\left|\int_0^Z f(z)\overline{\mathbf{u}(z)}dz\right| \\
&\leq 2\int_0^Z\left(\sqrt{\frac{2\rho_L\kappa(z)}{\kappa_L}}(Z-\hat{\mathbf{z}}_L)|f(z)|\right)\left(\frac{\omega}{\sqrt{2\kappa(z)}}|\mathbf{u}(z)|\right)dz \\
&\leq 2\int_0^Z\left(\frac{\rho_L\kappa(z)}{\kappa_L}(Z-\hat{\mathbf{z}}_L)^2|f(z)|^2dz+\frac{1}{2}\int_0^Z\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2\right)dz \\
&\leq 2(Z-\hat{\mathbf{z}}_L)^2\frac{\rho_L}{\kappa_L}\int_0^Z\kappa(z)|f(z)|^2dz+\frac{1}{2}\int_0^Z\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2dz \\
&\leq 2(Z-\hat{\mathbf{z}}_L)^2\frac{\rho_L}{\kappa_L}\kappa^\star||f||^2+\frac{1}{2}\int_0^Z\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2dz
\end{aligned}$$

We conclude by observing that

$$\begin{aligned}
Z-\hat{\mathbf{z}}_L &= \mathbf{z}_L-\hat{\mathbf{z}}_L \\
&= \mathbf{z}_L-\mathbf{z}_{L-1}+\mathbf{z}_{L-1}-\hat{\mathbf{z}}_L \\
&= \mathbf{h}_L+\hat{\mathbf{h}}_L \\
&\leq Z+\hat{\mathbf{h}}_L.
\end{aligned}$$

□

We are now ready to give a demonstration of Proposition 4.

Proof of Proposition 4. We begin with using Proposition 3

$$\begin{aligned}
&\int_0^Z\left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2+\frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right)dz \\
&+\omega^2\sum_{l=1}^{L-1}\left\{\frac{1}{\kappa_{l+1}}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})-\frac{1}{\kappa_l}(\mathbf{z}_l-\hat{\mathbf{z}}_l)\right\}|\mathbf{v}_l|^2+\sum_{l=1}^{L-1}\left\{\rho_{l+1}(\mathbf{z}_l-\hat{\mathbf{z}}_{l+1})-\rho_l(\mathbf{z}_l-\hat{\mathbf{z}}_l)\right\}|\mathbf{w}_l|^2 \\
&=(Z-\hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2+\rho_L|\mathbf{w}_L|^2\right)-(0-\hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2+\rho_1|\mathbf{w}_0|^2\right)+2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l}f(z)\overline{\phi_l(z)}dz\right\}.
\end{aligned}$$

Then, Lemma 7 enables us to write

$$\int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz + \sum_{l=1}^{L-1} \mathbf{h}_l \left(\frac{1}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right)$$
$$\leq \int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz + \omega^2 \sum_{l=1}^{L-1} \left\{\frac{1}{\kappa_{l+1}}(\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \frac{1}{\kappa_l}(\mathbf{z}_l - \hat{\mathbf{z}}_l)\right\} |\mathbf{v}_l|^2$$
$$+ \sum_{l=1}^{L-1} \left\{\rho_{l+1}(\mathbf{z}_l - \hat{\mathbf{z}}_{l+1}) - \rho_l(\mathbf{z}_l - \hat{\mathbf{z}}_l)\right\} |\mathbf{w}_l|^2 \quad (2.40)$$

On the other hand, Lemmas 8 and 10 together with the fact that $0 - \hat{\mathbf{z}}_1 = \hat{\mathbf{h}}_1 = 0$ yield

$$(Z - \hat{\mathbf{z}}_L)\left(\frac{\omega^2}{\kappa_L}|\mathbf{v}_L|^2 + \rho_L|\mathbf{w}_L|^2\right) - (0 - \hat{\mathbf{z}}_1)\left(\frac{\omega^2}{\kappa_1}|\mathbf{v}_0|^2 + \rho_1|\mathbf{w}_0|^2\right) + 2\,\mathrm{Re}\left\{\sum_{l=1}^{L}\int_{\mathbf{z}_{l-1}}^{\mathbf{z}_l} f(z)\overline{\phi_l(z)}dz\right\}$$
$$\leq 2(Z + \hat{\mathbf{h}}_L)^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right)||f||^2 + \frac{1}{2}\int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz \quad (2.41)$$

Combining (2.40) and (2.41), we obtain

$$\frac{1}{2}\int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz + \sum_{l=1}^{L-1} \mathbf{h}_l \left(\frac{1}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) \leq 2(Z + \hat{\mathbf{h}}_L)^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right)||f||^2.$$

To conclude the proof, we use Lemma 7 again to derive

$$Z + \hat{\mathbf{h}}_L \leq Z + 2\mathcal{M}Z \leq 3\mathcal{M}Z,$$

since $\mathcal{M} \geq 1$. □

Part III

We are now able to prove Theorems 3 and 4 as direct consequences of Proposition 4.

Proof of Theorem 3. We first recall stability estimate (2.36) from Proposition 4. We have

$$\frac{1}{2}\int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz + \sum_{l=1}^{L-1} \mathbf{h}_l \left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right)$$
$$\leq 18\mathcal{M}^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2 ||f||^2.$$

In particular, it holds that

$$\frac{1}{2}\int_0^Z \frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 \leq 18\mathcal{M}^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2 ||f||^2, \quad (2.42)$$

and

$$\frac{1}{2}\int_0^Z \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2 \leq 18\mathcal{M}^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2\,||f||^2\,. \tag{2.43}$$

We obtain Theorem 3 by taking the square roots of (2.42) and (2.43). □

Proof of Theorem 4. We recall again stability estimate (2.36) from Proposition 4. We have

$$\frac{1}{2}\int_0^Z \left(\frac{\omega^2}{\kappa(z)}|\mathbf{u}(z)|^2 + \frac{1}{\rho(z)}|\mathbf{u}'(z)|^2\right) dz + \sum_{l=1}^{L-1} \mathbf{h}_l \left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) \leq 18\mathcal{M}^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2\,||f||^2\,,$$

and observe that we have in particular

$$\sum_{l=1}^{L-1} \mathbf{h}_l \left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) \leq 18\mathcal{M}^2\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2\,||f||^2\,.$$

Since, $\mathbf{h}_l \geq \mathbf{h}_\star$ for $l \in \{1,\ldots,L\}$, it is clear that

$$\sum_{l=1}^{L-1} \left(\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 + \rho_l|\mathbf{w}_l|^2\right) \leq \frac{18\mathcal{M}^2}{\mathbf{h}_\star}\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2\,||f||^2\,,$$

and it follows in particular that

$$\frac{\omega^2}{\kappa_l}|\mathbf{v}_l|^2 \leq \frac{18\mathcal{M}^2}{\mathbf{h}_\star}\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2\,||f||^2\,, \tag{2.44}$$

and

$$\rho_l|\mathbf{w}_l|^2 \leq \frac{18\mathcal{M}^2}{\mathbf{h}_\star}\left(\rho^\star + \frac{\kappa^\star}{\kappa_L}\rho_L\right) Z^2\,||f||^2\,, \tag{2.45}$$

for $l \in \{1,\ldots,L\}$.

We obtain Theorem 4 by taking the square roots of (2.44) and (2.45). □

2.2 Some results in the homogeneous case

Before we tackle the discretization of 1D heterogeneous Helmholtz problems, we would like to introduce important results about the discretization of homogeneous Helmholtz problem.

As already stated in this document, the frequency plays a very important role both in the mathematical property of the equation and its numerical approximation. In particular, high frequency solutions are very expensive to compute. As a result, error analysis must be handled very carefully by taking into account the frequency parameter explicitly in every constant.

When using standard variational theory, based on the Schatz argument [88], one obtains the so-called "asymptotic" error-estimates. A quasi-optimal error-estimate for the finite element solution is established under the condition that the mesh step is small enough.

A quasi-optimal error-estimate is an error-estimate of the form

$$||u - u_h||_\star \leq C \inf_{v_h \in V_h} ||u - v_h||_\star, \tag{2.46}$$

where $V_h \subset H^1$ is the finite element space, u_h is the finite element solution and $||.||_\star$ is a norm of interest. This type of error-estimate is called "quasi-optimal" because, up to a constant, the finite element solution is as accurate as the best approximation provided by the discretization space. If (2.46) holds, we also say that the finite element solution is quasi-optimal.

In this spirit, the pioneering work of Aziz and Kellogg [15] establishes quasi-optimality of the finite element solution under the assumption that $h \leq C(\omega)$ where the constant C depends implicitly on ω. They extend this result in the late 80's and show that the solution is quasi-optimal under the condition that $\omega^2 h \leq C$ [16], where the condition on h depends explicitly on ω. Asymptotic error-estimates for linear elements are further investigated by Douglas et al. [43] and Ihlenburg and Babuška [59] in the 90's. Ihlenburg and Babuška also generalize asymptotic error-estimates for Lagrangian finite elements of arbitrary order p [60] in the context of one-dimensional domains. Melenk and Sauter have recently further extended the theory to general 3D problems with rough right-hand-sides $f \in L^2$ [75, 76].

As we are going to explain in the next subsections, if asymptotic error-estimates provide sufficient and necessary conditions to ensure quasi-optimality of the finite element solution, they are not fully satisfactory to quantify the error of the finite element solution. More precisely, they are only valid in a certain asymptotic range $h \in (0, h_0)$ where h_0 is going to zero as ω increases. Hence, the asymptotic error-estimates do not provide any information in the pre-asymptotic range $h \geq h_0$. This is troublesome, especially because the requirements on the mesh step h_0 are usually too restrictive to cover practical applications.

This has motivated the design of so-called "pre-asymptotic" error-estimates which are valid as soon as a given number of points per wavelength is achieved ($\omega h < C$). Ihlenburg and Babuška show in [59, 60] that in the pre-asymptotic range, the finite element solution is not quasi-optimal, and is polluted by an additional term in the error estimate. Though their work is limited to 1D problems with smooth data, their results are optimal and valid for Lagrangian elements of arbitrary degree p. They are also able to make a link between the so-called "pollution term" and the phase lag of the finite element solution.

Another powerful tool to study numerical methods for the Helmholtz equation is dispersion analysis. This methodology is less rigorous in the sense that it does not yield any error-estimate. Also, it is limited to cartesian grids (or space-periodic meshes). However, it is simpler to use than pre-asymptotic error-estimates and very powerful to analyse and design new schemes. Actually, if error-estimates are not proven, the results are usually very accurate when compared with numerical experiments.

The main idea is that if the right-hand-side vanishes and the boundary conditions are

omitted, the general solution of the Helmholtz equation is

$$u(x) = Ae^{\mathbf{i}\omega x} + Be^{-\mathbf{i}\omega x}, \tag{2.47}$$

where $A, B \in \mathbb{C}$. It can be shown that (2.47) has a discrete counterpart, and the discrete solution satisfies

$$u_h(x_j) = A_h e^{\mathbf{i}\omega_h x_j} + B_h e^{-\mathbf{i}\omega_h x_j},$$

for all mesh vertices $x_j = jh$, where $A_h, B_h \in \mathbb{C}$ and $\omega_h > 0$. Dispersion analysis then gives an estimate of the phase lag $\omega - \omega_h$.

It turns out that among the various methodologies used to analyse wave propagation schemes, dispersion analysis was the first to be used. Pioneering works focus on the transcient wave equation and include the work of Krieg and Key [65], Mullen and Betyschko [78], and Abboud and Pinsky [1].

An optimal bound on the phase lag in Helmholtz discretizations by Lagrangian finite elements of degree $1 \leq p \leq 3$ is first given by Thompson and Pinsky [97] on a one dimensional problem. Their result is generalized by Babuška and Ihlenburg to arbitrary order p in one dimension [60]. These results are obtained under the assumption that enough points per wavelength are used ($\omega h \leq C$).

To the best of our knowledge, Ainsworth's paper [3] is a reference work to dispersion analysis. It provides an estimation of the phase lag for arbitrary order p on cartesian grid in one, two or three dimensions. The analysis of Ainsworth is not limited to the condition $\omega h \leq C$ and the phase lag is provided for any value of h.

In the next subsections, we try to make a link between asymptotic error-estimates, dispersion analysis and pre-asymptotic error-estimates. We also give a quick overview of how the proofs are derived and refer the reader to the literature for more details. We discuss the possibility of generalizations to heterogeneous problems as well.

To clarify, given $f \in L^2(0,1,\mathbb{C})$, we consider the simple homogeneous 1D problem of finding $u = \mathcal{S}_\omega f \in H^1(0,1,\mathbb{C})$ such that

$$\left\{ \begin{array}{rcll} -\omega^2 u(z) - u''(z) & = & f(z), & z \in (0,1), \\ -u'(0) - \mathbf{i}\omega u(0) & = & 0, & \\ u'(1) - \mathbf{i}\omega u(1) & = & 0. & \end{array} \right. \tag{2.48}$$

We discretize problem (2.48) with Lagrangian finite element of degree $p \geq 1$. Introducing the variational formulation of problem (2.48), the continuous and discretized problems read: find $\mathcal{S}_\omega f \in H^1(0,1,\mathbb{C})$ and $\mathcal{S}_\omega^{h,p} f \in V_{h,p}$ such that

$$B(\mathcal{S}_\omega f, v) = \int_0^1 f(z)\overline{v(z)}dz, \ \forall v \in H^1(0,1,\mathbb{C}), \qquad B(\mathcal{S}_\omega^{h,p} f, v_h) = \int_0^1 f(z)\overline{v_h(z)}dz, \ \forall v_h \in V_{h,p},$$

where

$$B(w,v) = -\omega^2 \int_0^1 w(z)\overline{v}(z)dz - i\omega w(0)\overline{v(0)} - i\omega w(1)\overline{v(1)} + \int_0^1 w'(z)\overline{v'(z)}dz, \ \forall w, v \in H^1(0,1,\mathbb{C}),$$

and

$$V_{h,p} = \left\{ v_h \in C^0(0,1,\mathbb{C}) \quad | \quad v_h|_{((j-1)h,jh)} \in \mathcal{P}_p((j-1)h, jh); \quad j \in \{1,\dots,n = j/h\} \right\}.$$

We also introduce the adjoint operator $\mathcal{S}_\omega^\star : L^2(0,1,\mathbb{C}) \to H^1(0,1,\mathbb{C})$ where $\mathcal{S}_\omega^\star g$ is defined for $g \in L^2(0,1,\mathbb{C})$ as $\mathcal{S}_\omega^\star g = \overline{\mathcal{S}_\omega \bar{g}}$ and is the unique solution to

$$B(w, \mathcal{S}_\omega^\star g) = \int_0^1 w(z)\overline{g(z)}dz.$$

Finally, the $V_{h,p}$-interpolant of a function $v \in H^1(0,1,\mathbb{C})$ is defined as the unique function $\Pi_{h,p} v \in V_{h,p}$ such that

$$(\Pi_{h,p} v)(jh + \frac{kh}{p}) = v(jh + \frac{kh}{p}), \quad 0 \le j \le n, \quad 0 \le k \le p.$$

For ease of presentation, we will write $u = \mathcal{S}_\omega f$ and $u_{h,p} = \mathcal{S}_\omega^{h,p} f$. We will also use the notations $||.||_k$ and $|.|_k$ for the $H^k(0,1,\mathbb{C})$ norm and semi-norm ($k \in \mathbb{N}$):

$$||v||_k^2 = \sum_{j=0}^{k} \int_0^1 |v^{(j)}(z)|^2 dz, \quad |v|_k^2 = \int_0^1 |v^{(k)}(z)|^2 dz, \quad \forall v \in H^k(0,1,\mathbb{C}).$$

We introduce the essential result of Babuška and Ihlenburg [60] concerning the analysis of finite element error: assuming that $\omega h \le C$, we have

$$\omega |u - u_{h,p}|_0 + |u - u_{h,p}|_1 \le C_1 \omega h + C_2 \omega^{2p+1} h^{2p}.$$

We see that the finite element error is decomposed into two terms. The first term of order $\omega^p h^p$ is called the best approximation error term and is directly related to the approximation properties of the finite element space. The second term of order $\omega^{2p+1} h^{2p}$ is called the "pollution" term. It is worth noting that even when $\omega^p h^p$ is small, the term of order $\omega^{2p+1} h^{2p} = \omega(\omega h)^{2p}$ can be important if the frequency is high.

In standard finite element analysis of elliptic problem, the finite element scheme is quasi-optimal, and the finite element error is bounded by the best approximation error. The pollution term is thus typical of wave problems which are undefinite.

It is also clear that if we consider a given frequency ω, we can define two regions in which each term in the error bound is predominant. Asymptotically, if $h \to 0$, the best approximation term $\omega^p h^p$ is more important than the pollution term $\omega^{2p+1} h^{2p}$. On the other hand, especially for a high frequency ω, the pollution term is greater than the best approximation term for large mesh steps h. We can therefore define two regions:

- The pre-asymptotic range where the pollution term is dominant. It is characterized by $\omega^p h^p \le \omega^{2p+1} h^{2p}$.
- The asymptotic range where the best approximation term is dominant. It is characterized by $\omega^p h^p \ge \omega^{2p+1} h^{2p}$.

In the following, we focus on conforming Lagrangian finite element discretizations, but it is worth mentioning that dispersion and convergence analysis are also available for other popular discretization methods:

- In the context of polynomial Discontinuous Galerkin methods, we can mention the convergence analysis of the Local Discontinuous Galerkin (LDGm) and Internal Penalty Discontinuous Galerkin (IPDGm) methods [46, 47]. Wu and collaborators also recently introduced the so-called Continuous Internal Penalty method (CIPm) which is analysed in [106, 109].

- Melenk and collaborators have proposed a convergence analysis for general Discontinuous Galerkin methods, where functions of the discretization space are assumed to satisfy an inverse trace inequality [74]. Their convergence results apply to polynomial discretizations, but also to plane wave based discretizations. In particular, the Ultra-Weak Variational Formulation method (UWVFm) is handled.

- Among popular "plane wave" methods, we can also cite the Discontinuous Enrichment method (DEm). A convergence analysis for the lowest-order DEm elements is avaible [10].

- In the context of dispersion analysis, we mention the work of Ainsworth and collaborators on the Spectral Element method and Discontinuous Galerkin discretizations [4, 5].

We now focus on standard conforming Lagrangian elements. We will first talk about asymptotic error-estimates which are valid in the asymptotic range only. Then, we will present dispersion analysis and introduce the notion of phase-lag, which can be linked to the pollution term. We will close this introduction with pre-asymptotic error-estimates, which are valid in the pre-asymptotic range.

2.2.1 Asymptotic error-estimates

Asymptotic error-estimates give a bound of the finite element error under the condition that the mesh is sufficiently refined. If the mesh step $h \in (0, h_0)$ lies in a so-called asymptotic range bounded by a given $h_0 > 0$, then the finite element solution is quasi-optimal, and the finite element error is bounded by the best approximation error up to constant.

If we are using finite elements of order p and assuming that the right-hand-side $f \in H^{p-1}(0, Z, \mathbb{C})$, the asymptotic quasi-optimality and error-estimate for the Helmholtz equation can be stated in three points. We refer the reader to [16, 43, 59] for the analysis of the linear element case and to [60] for the general p-version in one-dimension. The linear case is treated in two dimensions in Melenk's PhD [73]. First, the best approximation error is bounded explicitly in frequency:

$$\omega|u - \Pi_{h,p}u|_0 + |u - \Pi_{h,p}u|_1 \leq C\omega^p h^p ||f||_{p-1}. \tag{2.49}$$

Second, under the condition that $\omega^{p+1}h^p \leq C$, the finite element solution is quasi-optimal:

$$\omega|u-u_{h,p}|_0+|u-u_{h,p}|_1 \leq C\left(\omega|u-\Pi_{h,p}u|_0+|u-\Pi_{h,p}u|_1\right). \tag{2.50}$$

Third, we can summarize the first two points as an asymptotic error-estimate. If we assume that $h \in (0,h_0)$ with $h_0 = C\omega^{-1-1/p}$, the following error-estimate holds

$$\omega|u-u_{h,p}|_0+|u-u_{h,p}|_1 \leq C\omega^p h^p||f||_{p-1}. \tag{2.51}$$

We see that the best approximation error is controlled by the number of points per wavelength $N_\lambda \simeq (\omega h)^{-1}$:

$$\omega|u-\Pi_{h,p}u|_0+|u-\Pi_{h,p}u|_1 \leq \frac{C||f||_{p-1}}{N_\lambda^p}.$$

It is worth noting that the best approximation error is bounded independently of the frequency when the number of points per wavelength is constant. It is not the case of the finite element error, and it is precisely what the pollution effect is.

We will analyse the pollution effect more precisely later. However we can already state that the finite element error increases linearly with the frequency when the number of points per wavelength N_λ is kept constant. We see that the same idea applies to the asymptotic range. In the asymptotic range, h must satisfies $\omega^{p+1}h^p \leq C$, that is $N_\lambda \leq C\omega^{-1/p}$. We can summarize the pollution effect as follows: the number of points per wavelength N_λ must increase with the frequency for the finite element solution to remain quasi-optimal.

Another way to describe the pollution effect is that if $N_\lambda \geq C$ is bounded below, we have

$$\lim_{\omega\to+\infty}\frac{|u-u_{h,p}|_1}{|u-\Pi_{h,p}u|_1}=+\infty.$$

Another important comment is that the approximation order is p, which is quite natural since we assume the right-hand-side to be in H^{p-1}. Hence, the solution is in H^{p+1}. Melenk and Sauter [75,76] have developed an asymptotic theory when the right-hand-side is in L^2 only. In this context, the best approximation converges to the best solution only at order 1. Hence, (2.49) does not hold and we only have

$$\omega|u-\Pi_{h,p}u|_0+|u-\Pi_{h,p}u|_1 \leq C\omega h|f|_0.$$

However the asymptotic range in which quasi-optimality holds is the same: $\omega^{p+1}h^p \leq C$. The main idea here is that having a p-order convergence by itself is not so important. Indeed, the best approximation error is always controlled by $\omega h \simeq N_\lambda^{-1}$ and h is always selected so that $\lim_{\omega\to+\infty}\omega h = 0$ to ensure that the solution is precise enough. The important point here is the region in which the quasi-optimality is valid, which is the same if $f \in L^2$ only.

In the context of highly heterogeneous media, $u \in H^{p+1}$ does not hold for two reasons. Indeed, we assume the right-hand-side to be in L^2 only, so that we can not expect more than $u \in H^2$ in the general case. Besides, there are jumps in the medium parameters

which lower the regularity of the solution. In the general case, we can not even expect the solution to be in H^2. However, if we assume that the density is constant (which is what we will do afterwards) the solution remains in H^2. It is worth noting that even if we assume that $f \in H^{p-1}$, $u \in H^{p+1}$ does not hold because of the jumps in the medium parameters.

Assuming that the density is constant, $u \in H^2$ and we obtain a result similar in spirit to Melenk and Sauter [75, 76]: when the velocity parameter is rough, the convergence order of the best approximation is not improved when p increases, but the asymptotic region $h \in (0, h_0)$ in which the solution is quasi-optimal is larger. To the best of our knowledge, this asymptotic error-estimate for heterogeneous media is new and will be presented afterwards.

The asymptotic results we have stated are optimal: it is mandatory that $\omega^{p+1}h^p \leq C$ to achieve quasi-optimality. Indeed, in the pre-asymptotic range ($h \geq h_0$) the finite element solution is "polluted" and is much less accurate than the best approximation. However, if the result is optimal in terms of quasi-optimality, we can make a simple comment about its quality as an error-estimate. If we apply the result as-is, we need to select h in the asymptotic range, that is $h \leq h_0 \simeq \omega^{-1-1/p}$. Hence, we have

$$\omega|u - u_{h,p}|_0 + |u - u_{h,p}|_1 \leq C\omega^p h^p \simeq \omega^{-1},$$

and the error is decreasing to 0 as ω is increasing. It means that condition $\omega^{p+1}h^p \leq C$ is not a good strategy to obtain a constant error independently of the frequency: the condition is too restrictive.

This observation has motivated the study of the behaviour of the finite element error in the so-called "pre-asymptotic" range, without assuming that $h \leq h_0$. We will state important results about pre-asymptotic analysis in the next subsections, but before, we give an overview of how the asymptotic error-estimates are derived.

The first issue is to estimate the best approximation error. From standard finite element theory, the interpolation error can be bounded by a semi-norm of the solution in an appropriate Sobolev space. Therefore, the problem of estimating the best approximation error simply amounts to estimating the norm of the solution derivatives.

We would like to point out that this is one of the reasons why it is really important to derive frequency-explicit stability estimates. Indeed, having frequency-explicit estimates for the continuous problem immediately yields frequency-explicit bounds of the best approximation error. For our purpose of briefly explaining how asymptotic errors are established, we recall that we have

$$|\mathcal{S}_\omega f|_2 \leq C\omega|f|_0, \quad |\mathcal{S}_\omega^\star f|_2 \leq C\omega|f|_0, \tag{2.52}$$

as a particular case of Theorem 5. We are now able to show that the best approximation error is controlled by the number of points per wavelength. We record this result in Lemma 11.

Lemma 11. *Let $f \in L^2(0, Z, \mathbb{C})$. The following estimates hold:*

$$\omega|\mathcal{S}_\omega f - \Pi_h \mathcal{S}_\omega f|_0 + |\mathcal{S}_\omega f - \Pi_h \mathcal{S}_\omega f|_1 \leq C\omega h|f|_0, \tag{2.53}$$

and

$$\omega|\mathcal{S}_\omega^\star f - \Pi_h \mathcal{S}_\omega^\star f|_0 + |\mathcal{S}_\omega^\star f - \Pi_h \mathcal{S}_\omega^\star f|_1 \leq C\omega h |f|_0. \tag{2.54}$$

Proof. The proof relies on classical approximation properties of polynomials and the frequency-explicit stability estimate of the Helmholtz equation. To simplify the notations, we will write $u_f = \mathcal{S}_\omega f$. Using approximation theory, we have

$$\omega|u_f - \Pi_h u_f|_0 + |u_f - \Pi_h u_f|_1 \leq C(\omega h^2 + h)|u_f|_2.$$

The second step uses the frequency-explicit stability estimate. We have

$$|u_f|_2 \leq C\omega |f|_0,$$

so that

$$\omega|u_f - \Pi_h u_f|_0 + |u_f - \Pi_h u_f|_1 \leq C(\omega^2 h^2 + \omega h)|f|_0.$$

We obtain (2.53) since $\omega h \leq 1$. The demonstration of (2.54) follows the same guidelines than for (2.53) and is thus omitted here. □

We now show that the finite element solution is asymptotically quasi-optimal. Most of asymptotic error-estimates for the Helmholtz equation are based on the so-called Schatz argument [88]. Actually, the Schatz argument is valid for any continuous sesquilinear form satisfying a Gårding inequality, that is a sesquilinear form B such that

$$\mathrm{Re}\, B(u,u) \geq \alpha ||u||_1^2 - \mu |u|_0^2, \quad \forall u \in H^1, \tag{2.55}$$

and

$$|B(u,v)| \leq M||u||_1||v||_1, \quad \forall u, v \in H^1,$$

for some positive constants α, μ and M.

It is well-known that variational formulations of Fredholm operator fall into this category. For the case of the Helmholtz equation, it is simple enough to show that our sesquilinear form B is continuous and satisfies the Gårding inequality with the constants $M = \omega^2$, $\alpha = 1$ and $\mu = \omega^2 + 1$. More precisely, we have:

Lemma 12. *Let $w, v \in H^1(0, Z, \mathbb{C})$. Then, we have*

$$\begin{aligned} |B(w,v)| &\leq C(\omega|w|_0 + |w|_1)(\omega|v|_0 + |v|_1) \\ \mathrm{Re}\, B(v,v) &\geq |v|_1^2 - \omega^2 |v|_0^2. \end{aligned}$$

The Schatz argument can be viewed as a generalization of Céa's Lemma. Céa's Lemma is only valid for coercive forms, and states that the finite element solution is quasi-optimal for any h. In contrast, the Schatz argument is valid under a weaker assumption of Gårding inequality but ensures the quasi optimality only for small enough $h < h_0$, where h_0 depends on the constants α and μ of Gårding inequality (2.55).

In Theorem 8, we demonstrate that the linear finite element solution is quasi-optimal in the asymptotic range using the Schatz argument. We also derive the corresponding error-estimate.

Theorem 8. *Assume that $\omega^2 h \leq C$. Then the solution $\mathcal{S}_\omega^{h,1} f$ is quasi optimal:*

$$\omega|\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,1} f|_0 + |\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,1} f|_1 \leq C\left(\omega|\mathcal{S}_\omega f - \Pi_h(\mathcal{S}_\omega f)|_0 + |\mathcal{S}_\omega f - \Pi_h(\mathcal{S}_\omega f)|_1\right), \quad (2.56)$$

and the following error-estimate holds

$$\omega|\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,1} f|_0 + |\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,1} f|_1 \leq C\left(\omega h + \omega^3 h^2\right)|f|_0. \quad (2.57)$$

Proof. We will use the notations $u = \mathcal{S}_\omega f$ and $u_h = \mathcal{S}_\omega^{h,1} f$. The first part of the proof is the usual Aubin-Nitsche duality trick. We introduce the Riesz representation of the error $z \in H^1(0,1,\mathbb{C})$, defined as the unique function of $H^1(0,1,\mathbb{C})$ satisfying

$$B(w,z) = (w, u - u_h) \quad \forall w \in H^1(0,1,\mathbb{C}).$$

In particular, it holds that $|u-u_h|_0^2 = B(u-u_h, z)$, and by Galerkin orthogonality together with the continuity of B, we have

$$|u-u_h|_0^2 = B(u-u_h, z-\Pi_h z) \leq C(\omega|u-u_h|_0 + |u-u_h|_1)(\omega|z-\Pi_h z|_0 + |z-\Pi_h z|_1)$$

According to the definition of z, Lemma 11 yields

$$(\omega|z-\Pi_h z|_0 + |z-\Pi_h z|_1) \leq C\omega h|u-u_h|_0,$$

hence

$$|u-u_h|_0^2 \leq C\omega h(\omega|u-u_h|_0 + |u-u_h|_1)|u-u_h|_0,$$

and

$$\left(1 - C\omega^2 h\right)|u-u_h|_0^2 \leq C\omega h|u-u_h|_1|u-u_h|_0.$$

Assuming that $\omega^2 h$ is small enough and dividing by $|u-u_h|_0$, we obtain

$$|u-u_h|_0 \leq C\omega h|u-u_h|_1. \quad (2.58)$$

Note that (2.58) is the usual bound obtained with the Aubin-Nitsche duality trick: the order of convergence is one order higher in the L^2 norm than in the H^1 semi-norm. When dealing with elliptic problems, the H^1 semi-norm is bounded using Céa's Lemma and the L^2 error bound follows. Here, (2.58) is just a step in the convergence proof.

The second part of the proof is similar to Céa's Lemma. We use Galerkin orthogonality and the Gårding inequality (remark that in Céa's Lemma, coercivity of the sesquilinear form is used instead of the Gårding inequality). We have $B(u-u_h, u-u_h) = B(u-u_h, u-\Pi u_h)$, and

$$|u-u_h|_1^2 - \omega^2|u-u_h|_0^2 \leq \mathrm{Re}\, B(u-u_h, u-\Pi_h u) \leq C(\omega|u-u_h|_0 + |u-u_h|_1)(\omega|u-\Pi_h u|_0 + |u-\Pi_h u|_1).$$

With estimate (2.58), we have that

$$\left(1 - C\omega^4 h^2\right)|u-u_h|_1^2 \leq C(1+\omega^2 h)|u-u_h|_1(\omega|u-\Pi_h u|_0 + |u-\Pi_h u|_1),$$

and, dividing by $|u-u_h|_1$, we get the quasi-optimality in the H^1 semi-norm if we assume that $\omega^2 h$ is small enough:

$$\left(1 - C\omega^4 h^2\right) |u-u_h|_1 \leq C(1+\omega^2 h)(\omega|u-\Pi_h u|_0 + |u-\Pi_h u|_1),$$

Because we assume that $\omega^2 h$ is small enough and that $\omega h < 1$, we can use (2.58) to derive

$$\omega|u-u_h|_0 + |u-u_h|_1 \leq C(1+\omega^2 h)(\omega|u-\Pi_h u|_0 + |u-\Pi_h u|_1),$$

which is (2.56) using again that $\omega^2 h$ is small.

We now turn to the proof of error-estimate (2.57). A direct application of Lemma 11 yields

$$\omega|u-\Pi_h u|_0 + |u-\Pi_h u|_1 \leq \omega h|f|_0,$$

and therefore, inserting the previous estimate in (2.56), we have

$$\omega|u-u_h|_0 + |u-u_h|_1 \leq C(1+\omega^2 h)\omega h|f|_0,$$

which simplifies to (2.57). □

Remark that we keep the term $\omega^3 h^2$ in the above estimates because it appears naturally in the proofs. However, since we assume that $\omega^2 h \leq C$, we obviously have $\omega^3 h^2 \leq C\omega h$ so that it can be omitted to recover (2.51) with $p=1$. As we are going to see in the dispersion analysis, this term plays an important role, so that we keep it in the estimates.

2.2.2 Dispersion relations

In the previous subsection, we observed that the condition $\omega^{p+1}h^p \leq C$ is mandatory to achieve quasi-optimality. We also stated that this condition on h is not satisfactory, because the finite element error is going to zero when the frequency increases if it is applied as-is. We also demonstrated that the quality of the best approximation only depends on the number of points per wave length N_λ and is bounded if $\omega^p h^p \leq C$. However, we have stated that this condition is not sufficient to guarantee the quality of the numerical solution.

Therefore, to select h, the rule $h \simeq \omega^{-1}$ where the number of points per wavelength is constant is not enough. On the other hand, the asymptotic-range restriction $h \simeq \omega^{-1-1/p}$ is to restrictive. One would wish to obtain a rule, say an exponent θ such that the finite element error is constant if $h \simeq \omega^\theta$ for all frequencies $\omega \geq 1$. Dispersion analysis is a nice tool to achieve this goal.

Strictly speaking, dispersion analysis does not provide any error-estimate. Actually, it gives a measure of the phase-lag between the numerical solution and the continuous solution. The basic idea comes from the fact that if we neglect boundary conditions, homogeneous solutions to the homogeneous 1D Helmholtz equation reads

$$u(x) = Ae^{i\omega x} + Be^{-i\omega x} \quad \forall x \in \mathbb{R}, \tag{2.59}$$

where $A, B \in \mathbb{C}$ are two complex constants. If the problem is discretized on a regular infinite grid, (2.60) has a discrete counterpart. It can be shown that the values of the solution at the nodal points $x_n = nh$ satisfy

$$u_{h,p}(x_n) = A_h e^{i\omega_{h,p} x_n} + B_h e^{-i\omega_{h,p} x_n} \quad \forall n \in \mathbb{N}, \tag{2.60}$$

for two constants $A_h, B_h \in \mathbb{C}$ and the so-called "discrete pulsation" $\omega_{h,p} \in \mathbb{C}$.

The aim of the dispersion analysis is to estimate the phase lag, that is the difference $|\omega - \omega_{h,p}|$ between the continuous and the discrete pulsations. Babuška and Ihlenburg analysed the dispersion of Lagrangian finite elements in [59, 60] and proved that

$$|\omega - \omega_{h,p}| \leq C\omega^{2p+1} h^{2p+1}, \tag{2.61}$$

provided that there are enough number of points per wave length ($\omega h \leq 1$). Note that the condition $\omega h \leq 1$ is much less restrictive than the asymptotic condition $\omega^{p+1} h^p \leq C$ when the frequency is high. Furthermore, since the best approximation error depends on the number of points per wavelength, it is fairly natural to assume that $\omega h \leq 1$.

Although dispersion analysis is not a rigorous error analysis (there is no right-hand-side and boundary conditions are not taken into account), it turns out that (2.61) exactly fits numerical experiments in most situations and there is actually a theoretical reason, that we will present in the next subsection. Thus dispersion analysis is a very popular technique to analyse finite element schemes for wave propagation problems. The results are optimal (even though there is no rigorous justification), easy to derive (compared to pre-asymptotic error-estimates) and have a clear physical meaning: the phase-lag.

Dispersion analysis is not limited to the one-dimensional case. The important hypothesis is that the numerical scheme must be periodic in space. Dispersion analysis in two and three dimensions includes the work of Mullen and Belytschko [78], Abboud and Pinsky [1] and Ainsworth [3] for cartesian grids. An analysis for more general schemes is given by Deraemaeker, Babuška and Bouillard [41].

Unfortunately, finite element schemes are not all space-periodic. This is the case of Lagrangian elements on unstructured meshes which are interesting in Geophysics for their h-adaptivity. More importantly, dispersion analysis can not be applied in heterogeneous media (apart from the case where the medium is periodic, and the period of the medium is a multiple of the mesh step h or simple configurations like two layered media).

The advantage of dispersion analysis over pre-asymptotic error-estimates however, is that the requirement $\omega h \leq C$ is not mandatory and that all computations can be made explicitly. That way, Ainsworth obtained [3] a closed formula for the discrete pulsation which is valid for all ω, h and p:

$$\cos(\omega_{h,p} h) = R_p(\omega h),$$

where R_p is a rational function which is explicitly defined using Padé approximants.

We close this subsection with Theorem 9 where we estimate the phase-lag of linear elements in one dimension. In order to state and demonstrate Theorem 9, we need to

introduce an Helmholtz problem without boundary conditions, set on the whole real line $\mathbb{R}$, together with a "finite element approximation". To this end, we introduce additional notations. We assume $h > 0$ is given. For all $n \in \mathbb{Z}$, we define $x_n = nh$ and $\phi_n : \mathbb{R} \to \mathbb{R}$

$$\phi_n(x) = \begin{cases} \dfrac{x - x_{n-1}}{h}, & \text{if } x \in (x_{n-1}, x_n), \\ \dfrac{x_{n+1} - x}{h}, & \text{if } x \in (x_n, x_{n+1}), \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $j \in \mathbb{Z}$ with $|j| \geq 2$, the Lebesgue measure of $\mathrm{supp}(\phi_n) \cap \mathrm{supp}(\phi_{j+n})$ vanishes. Hence,

$$u = \sum_{n \in \mathbb{Z}} u_n \phi_n \in W^{1,\infty}_{loc}(\mathbb{R}, \mathbb{C}),$$

is well defined for any sequence $(u_n)_{n \in \mathbb{Z}} \subset \mathbb{C}$.

For $u \in W^{1,\infty}$ and $\phi \in W^{1,1}$, we define

$$B_\infty(u, \phi) = -\omega^2 \int_{-\infty}^{+\infty} u(x)\overline{\phi(x)}dx + \int_{-\infty}^{+\infty} u'(x)\overline{\phi'(x)}dx.$$

In particular,

$$B_\infty\left(\sum_{n \in \mathbb{Z}} u_n \phi_n, \phi_m\right)$$

is well defined for all $m \in \mathbb{Z}$.

Theorem 9. *Assume that $\omega h < \sqrt{12}$ and*

$$u = \sum_{n \in \mathbb{Z}} u_n \phi_n,$$

satisfies

$$B(u, \phi_m) = 0, \quad \forall m \in \mathbb{Z}.$$

Then there exist two constants $A, B \in \mathbb{C}$ and a discrete pulsation $\omega_h \in \mathbb{R}_+$ such that

$$u_n = u(x_n) = A \exp\left(\mathbf{i}\omega_h x_n\right) + B \exp\left(-\mathbf{i}\omega_h x_n\right), \quad \forall n \in \mathbb{Z}. \tag{2.62}$$

The discrete pulsation ω_h is defined by the equation

$$\cos(\omega_h h) = \frac{6 - 2\omega^2 h^2}{6 + \omega^2 h^2}, \tag{2.63}$$

and we have

$$\omega - \omega_h = \frac{\omega^3 h^2}{24} + o(\omega^5 h^4). \tag{2.64}$$

Proof. Consider a given $m \in \mathbb{Z}$. Since the basis functions ϕ_n and ϕ_m have disconnected supports when $n+2 \leq m$ or $m \leq n+2$, we have

$$B_\infty(u, \phi_m) = B_\infty\left(\sum_{n\in\mathbb{Z}} u_n\phi_n, \phi_m\right) = B_\infty\left(\sum_{n=m-1}^{m+1} u_n\phi_n, \phi_m\right) = 0.$$

Using sesquilinearity of B_∞, we obtain a recurrence relation:

$$u_{m+1}B_\infty(\phi_{m+1}, \phi_m) + u_m B_\infty(\phi_m, \phi_m) + u_{m-1}B_\infty(\phi_{m-1}, \phi_m) = 0, \quad \forall m \in \mathbb{Z}. \tag{2.65}$$

We now use the key feature: space periodicity. We have

$$B_\infty(\phi_{m+1}, \phi_m) = b_+, \quad B_\infty(\phi_m, \phi_m) = b_0, \quad B_\infty(\phi_{m-1}, \phi_m) = b_-,$$

for all $m \in \mathbb{Z}$, where $b_+, b_0, b_- \in \mathbb{R}$ are real constants independent of the position m. Furthermore, direct computations yield that

$$b_0 = \frac{2}{h} - \frac{2\omega^2 h}{3}, \quad b_- = b_+ = -\frac{1}{h} - \frac{\omega^2 h}{6}.$$

Hence, multiplying (2.65) by $-h$, we obtain a linear recurrence relation for $(u_m)_{m\in\mathbb{Z}}$:

$$\left(\frac{\omega^2h^2}{6} + 1\right)u_{m+1} + 2\left(\frac{\omega^2h^2}{3} - 1\right)u_m + \left(\frac{\omega^2h^2}{6} + 1\right)u_{m-1} = 0 \quad \forall m \in \mathbb{Z}, \tag{2.66}$$

and we can compute a formula for any solution up to two complex constants by finding the roots of the characteristic polynomial

$$P(r) = \left(\frac{\omega^2h^2}{6} + 1\right)r^2 + \left(\frac{\omega^2h^2}{3} - 1\right)r + \left(\frac{\omega^2h^2}{6} + 1\right)$$

To start with, we compute the discriminant:

$$\Delta = 4\left(\frac{\omega^2h^2}{3} - 1\right)^2 - 4\left(\frac{\omega^2h^2}{6} + 1\right)^2 = \omega^2h^2\left(\frac{\omega^2h^2}{3} - 4\right) = 2\omega^2h^2\left(\frac{\omega^2h^2}{12} - 1\right);$$

Since $\omega h > 0$, we deduce that $\Delta < 0$ iff $\omega h < \sqrt{12}$. Assuming that $\Delta < 0$ from now on, P has two conjugate roots $r_\pm \in \mathbb{C}$ given by

$$r_\pm = \frac{1}{2\left(\frac{\omega^2h^2}{6} + 1\right)}\left(-\left(\frac{\omega^2h^2}{6} + 1\right) \pm \mathbf{i}\sqrt{-\Delta}\right),$$

and there exist two constants $A, B \in \mathbb{C}$ such that

$$u_m = Ar_+^m + Br_-^m, \quad \forall m \in \mathbb{Z}.$$

The expression of $r_\pm$ is hard to handle. Therefore, we are going to identify $r_\pm$ by inserting $u_m = e^{\mathbf{i}\theta m}$ into recurrence relation (2.66). We obtain

$$\left(\frac{\omega^2 h^2}{6} + 1\right) e^{\mathbf{i}(m+1)\theta} + 2\left(\frac{\omega^2 h^2}{3} - 1\right) e^{\mathbf{i}m\theta} + \left(\frac{\omega^2 h^2}{6} + 1\right) e^{\mathbf{i}(m-1)\theta} = 0.$$

Dividing by $e^{\mathbf{i}m\theta}$ and using De Moivre identity, we have

$$\cos(\theta) = \frac{6 - 2\omega^2 h^2}{6 + \omega^2 h^2}. \tag{2.67}$$

We define $\omega_h = \theta/h$. Then (2.67) directly yields (2.63). We also obtain (2.62) easily. Indeed, since $x_m = mh$, we have

$$u(x_m) = u_m = Ae^{\mathbf{i}m\theta} + Be^{-\mathbf{i}m\theta} = Ae^{\mathbf{i}\omega_h x_m} + Be^{-\mathbf{i}\omega_h x_m}.$$

Following Ihlenburg and Babuška [59], we prove (2.64) using (2.67) and Taylor expansions. We have

$$\omega_h h = \arccos\left(\frac{6 - 2\omega^2 h^2}{6 + \omega^2 h^2}\right) = \omega h - \frac{\omega^3 h^3}{24} + o(\omega^5 h^5),$$

and we obtain (2.64) by dividing by h. □

2.2.3 Pre-Asymptotic error-estimates

We have defined two types of results: asymptotic error-estimates which give a necessary condition on h for the finite element solution to be quasi-optimal and dispersion analysis which gives an explicit expression of the phase-lag. We recall that asymptotic error-estimates are only valid if h satisfies $\omega^{p+1}h^p \leq C$. Thus, asymptotic error-estimates can only be applied in an asymptotic range $h \in (0, h_0]$ where $h_0 \simeq \omega^{-1-1/p}$. The main problem is that this condition is too restrictive at high frequency. Indeed, suppose we want to achieve a given accuracy ϵ. If asymptotic error-estimate are to be applied, the largest possible mesh step is $h = h_0 \simeq \omega^{-1-1/p}$ which leads to

$$\omega|u - u_{h,p}|_0 + |u - u_{h,p}|_1 \leq \omega^p h^p \simeq \omega^{-1}. \tag{2.68}$$

Assume we want to solve for a high frequency ω such that $\omega^{-1} << \epsilon$. Then, regarding (2.68), if we apply the asymptotic condition $h \leq h_0$, the solution is more precise than needed. Hence, to achieve the desired precision exactly, we need to select $h > h_0$, but this case is not covered by asymptotic error-estimates.

Dispersion analysis is limited as well because it applies to simplified models and space-periodic schemes. This has motivated the development of pre-asymptotic error-estimates. Compared to asymptotic error-estimates, pre-asymptotic error-estimates are valid as soon as there are enough points per wavelength (or equivalently, the number of points per

wavelength N_λ is bounded above, i.e. $\omega h \leq C$). Furthermore, pre-asymptotic error-estimates apply to realistic models including boundary conditions and right-hand-side when dispersion analysis can not be carried out.

Against this background, an essential result has been established by Ihlenburg and Babuška in a pair of papers [59, 60]. Assuming that $\omega h \leq C$, the finite element error is bounded by the best-approximation error plus a "pollution" term:

$$\omega|\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,p} f|_0 + |\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,p} f|_1 \leq C_1 \omega^p h^p + C_2 \omega^{2p+1} h^{2p}, \tag{2.69}$$

where $C_1, C_2 > 0$ are two constants independent of ω, h.

The first term in the right-hand-side of (2.69) is the best approximation error. It is the same term than in asymptotic error-estimates. The second is called the pollution term and has the same order than the phase-lag.

Actually, pre-asymptotic error-estimates are a proper generalization of asymptotic error-estimates. Indeed, if we assume that h lies in the asymptotic range, $\omega^{p+1} h^p \leq C$, then $\omega^{2p+1} h^{2p} \leq C \omega^p h^p$ and (2.69) becomes

$$\omega|\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,p} f|_0 + |\mathcal{S}_\omega f - \mathcal{S}_\omega^{h,p} f|_1 \leq (C_1 + CC_2)\omega^p h^p,$$

which is (2.51). Hence, pre-asymptotic error-estimate (2.69) is equivalent to asymptotic error-estimate (2.51) in the asymptotic range where $\omega^{p+1} h^p \leq C$.

More precisely, we see that the quasi-optimality condition $\omega^{p+1} h^p \leq C$ corresponds to the zone where the phase-lag is of the same order than the best approximation. The quasi-optimality condition thus defines two different regimes:

- In the pre-asymptotic range, $\omega^{p+1} h^p >> 1$. Therefore $\omega^p h^p << \omega^{2p+1} h^{2p}$ and the phase-lag error, or pollution error, is dominant.
- In the asymptotic range, $\omega^{p+1} h^p << 1$, then $\omega^p h^p >> \omega^{2p+1} h^{2p}$ and the best-approximation error is dominant: the solution is quasi-optimal.

It is not obvious to extend the proof of (2.69) to higher dimensions or to heterogeneous media. This proof is the work of Babuška and Ihlenburg [60] and it involves subtle arguments including stability estimates in dual norms and specific interpolants.

A simpler method to obtain pre-asymptotic error-estimates has been recently developed by Wu and Zhu [106, 109]. It consists in defining an elliptic projection of the Riesz representation of the error and it turns out that the proof is based on an extension of the Schatz argument. The method work for 2D and 3D problems, however, the results are not optimal as compared to (2.69).

2.3 Discretization using the Multiscale Medium Approximation method

In this section, we consider the discretization of heterogeneous Helmholtz problems by finite elements.

2.3.1 Problem statement

We are focusing on the 1D Helmholtz problem with constant density. Hence, we assume that $\rho = 1$, and $\kappa = c^2$, where c is the wavespeed.

It follows that the problem reads

$$\left\{ \begin{array}{rcl} -\dfrac{\omega^2}{c(z)^2}u(z) - u''(z) &=& f(z), \quad z \in (0, Z) \\ u'(0) &=& 0 \\ u'(Z) - \dfrac{\mathbf{i}\omega}{c(Z)}u(Z) &=& 0, \end{array} \right. \tag{2.70}$$

We assume that c is chosen so that $(c^2, 1) \in M$. It means in particular that c is a piecewise constant parameter satisfying $c_{min} \leq c(z) \leq c_{max}$ for $z \in (0, Z)$.

Since ρ is set, we write $S_{\omega,c} = S_{\kappa,\rho,c}$, $S^\star_{\omega,c} = S^\star_{\kappa,\rho,c}$ and $B_{\omega,c} = B_{\kappa,\rho,c}$. We also introduce the weighted norm

$$||v||^2_{\omega,c} = \omega^2 ||c^{-1}v||^2 + ||v'||^2,$$

for $v \in H^1(0, Z, \mathbb{C})$.

2.3.2 Finite element discretization

In this section, we recall the theory developed by Melenk and Sauter [75, 76] to derive asymptotic stability estimates. The main idea is that if the discrete space is sufficiently rich, the scheme is quasi-optimal.

For the sake of simplicity, we will consider a uniform decomposition of the domain $(0, Z)$ together with polynomial basis functions with constant degree. To this end, we consider a discretization step $h = 1/n_h$ with $n_h \in \mathbb{N}$, and the associated decomposition $t_j = jh$, for $j \in \{0, \ldots, n_h\}$. Then we define the discretization space as

$$V^{h,p} = \left\{ v \in H^1(0, Z) \quad | \quad v|_{(t_{j-1},t_j)} \in \mathcal{P}_p,\ 0 \leq j \leq n_h \right\}, \tag{2.71}$$

where $1 \leq p \leq 3$ is a given integer and $\mathcal{P}_p$ stands for the space of polynomials of degree smaller or equal to p.

In the following, we want to quantify the ability of the discretization space $V^{h,p}$ to approximate solutions to the Helmholtz equation. In this regard, we introduce

$$\eta^{h,p}_{\omega,c} = \sup_{f \in L^2(0,Z)} \inf_{v_h \in V^{h,p}} \frac{||\mathcal{S}^\star_{\omega,c} f - v_h||_{\omega,c}}{||f||} = \sup_{f \in L^2(0,Z)} \inf_{v_h \in V^{h,p}} \frac{||\mathcal{S}_{\omega,c} f - v_h||_{\omega,c}}{||f||}. \tag{2.72}$$

We first show that the sesquilinear form B is bounded by a constant independent of ω for the norm $||.||_{\omega,c}$.

Proposition 5. *For all $u, v \in H^1(0, Z)$ we have*

$$|B_{\omega,c}(u, v)| \leq C_b ||u||_{\omega,c} ||v||_{\omega,c}, \tag{2.73}$$

with

$$C_b = 1 + \frac{C_{tr}^2}{c_{min}}.$$

Proof. First, by application of the Schwarz inequality, we have

$$|B_{\omega,c}(u,v)| \leq \omega^2 ||c^{-1}u||\ ||c^{-1}v|| + ||u'||\ ||v'|| + \frac{\omega}{c(Z)}|u(Z)|\ |v(Z)|.$$

Then, we have

$$\omega^2 ||c^{-1}u||\ ||c^{-1}v|| + ||u'||\ ||v'|| \leq ||u||_{\omega,c}||v||_{\omega,c},$$

and, by application of Proposition 2,

$$\frac{\omega}{c(Z)}|u(Z)||v(Z)| \leq \frac{1}{c_{min}}\left(\omega^{-1/2}||u||_\infty\right)\left(\omega^{-1/2}||v||_\infty\right) \leq \frac{C_{tr}^2}{c_{min}}||u||_{\omega,c}||v||_{\omega,c}.$$

It follows that

$$|B_{\omega,c}(u,v)| \leq (1 + \frac{C_{tr}^2}{c_{min}})||u||_{\omega,c}||v||_{\omega,c}.$$

□

In Lemma 13, we recall an important result due to Melenk and Sauter [75, 76]: if $\omega\eta_{\omega,c}^{h,p}$ is properly controlled, the accuracy of the numerical scheme is ensured by approximation properties of the discretization space $V^{h,p}$.

Lemma 13. *Assume that $u_h \in V^{h,p}$ satisfies*

$$B_{\omega,c}(u_h, v_h) = \int_0^Z f(z)\overline{v_h(z)}dz, \tag{2.74}$$

for all $v_h \in V^{h,p}$. Then, if we assume that $\omega\eta_{\omega,c}^{h,p} \leq \alpha$, we have

$$||\mathcal{S}f - u_h||_{\omega,c} \leq C_e\eta_{\omega,c}^{h,p}||f||, \tag{2.75}$$

with

$$\alpha = \frac{c_{min}}{2C_b}, \quad C_e = \frac{2C_b c_{max}}{c_{min}}.$$

Proof. To simplify the notation, let us write $u = \mathcal{S}_{\omega,c}f$ and $s^\star = \mathcal{S}_{\omega,c}^\star(c^{-2}(u-u_h))$. Let $s_h^\star \in V^{h,p}$ denote the best approximation of $s^\star$ in the $||.||_{\omega,c}$ norm. Then by Galerkin orthogonality and by definition of $\eta_{\omega,c}^{h,p}$ (2.72), we have

$$\begin{aligned} ||c^{-1}(u-u_h)||^2 &= \mathrm{Re}\, B_{\omega,c}(u-u_h, s^\star - s_h^\star) \\ &\leq C_b||u-u_h||_{\omega,c}||s^\star - s_h^\star||_{\omega,c} \\ &\leq \frac{C_b\eta_{\omega,c}^{h,p}}{c_{min}}||u-u_h||_{\omega,c}||c^{-1}(u-u_h)||. \end{aligned}$$

Furthermore, it holds that

$$\begin{aligned}\operatorname{Re} B_{\omega,c}(u-u_h,u-u_h) &= ||u-u_h||^2_{\omega,c} - 2\omega^2||c^{-1}(u-u_h)||^2 \\ &\geq \left(1-2\left(\frac{\omega C_b \eta^{h,p}_{\omega,c}}{c_{min}}\right)^2\right)||u-u_h||^2_{\omega,c},\end{aligned}$$

and we can conclude that

$$\begin{aligned}\left(1-2\left(\frac{\omega C_b \eta^{h,p}_{\omega,c}}{c_{min}}\right)^2\right)||u-u_h||^2_{\omega,c} &\leq \operatorname{Re} B_{\omega,c}(u-u_h,u-u_h) \\ &\leq \operatorname{Re}\int_0^Z f(z)\overline{(u-u_h)(z)}dz \\ &\leq ||cf||||c^{-1}(u-u_h)|| \\ &\leq \frac{C_b c_{max}}{c_{min}}\eta^{h,p}_{\omega,c}||f||||u-u_h||_{\omega,c}.\end{aligned}$$

□

As a direct consequence of Lemma 13, the numerical scheme corresponding to $V^{h,p}$ is well-posed under the condition that $\omega\eta^{h,p}_{\omega,c} \leq \alpha$ and the discrete solution is quasi-optimal. To simplify the notations, we also introduce the discrete solution operator $\mathcal{S}^{h,p}_{\omega,c}$.

Theorem 10. *Assume that $\omega\eta^{h,p}_{\omega,c} \leq \alpha$. Then for all $f \in L^2(0,Z)$, there exists a unique element $\mathcal{S}^{h,p}_{\omega,c}f \in V^{h,p}$ such that*

$$B_{\omega,c}(\mathcal{S}^{h,p}_{\omega,c}f, v_h) = \int_0^Z f(z)\overline{v_h(z)}dz, \tag{2.76}$$

for all $v \in V^{h,p}$. Furthermore, we have

$$||\mathcal{S}_{\omega,c}f - \mathcal{S}^{h,p}_{\omega,c}f||_{\omega,c} \leq C_e\eta^{h,p}_{\omega,c}||f||. \tag{2.77}$$

Proof. Error estimate (3.23) directly follows from Theorem 13. Therefore, we only need to show existence and uniqueness of $\mathcal{S}^{h,p}_{\omega,c}f$ and, since $\mathcal{S}^{h,p}_{\omega,c}f$ is defined as the solution to a finite dimensional linear system, proving uniqueness is sufficient. But uniqueness directly follows from (2.75) with $f = 0$, and the proof is thus complete. □

2.3.3 Approximation properties

In the previous section, we have clarified how the quality of the best approximation is crucial to obtain the quasi-optimality of the scheme: the condition $\omega\eta^{h,p}_{\omega,c} \leq \alpha$ is required (the constant α being defined in Lemma 13). The aim of this section is to bound $\eta^{h,p}_{\omega,c}$ explicitly with respect to ω, h and p. This is achieved by building a good approximation

of $\mathcal{S}_{\omega,c}f$ for a given $f \in L^2(0,Z)$. In the context of homogeneous media with a non-regular right-hand-side f in L^2, Melenk and Sauter have proposed a frequency-splitting of the solution to build such an approximation [75, 76].

Here, we provide a methodology to construct the best approximation in the context of highly heterogeneous media. We are not aware of previous work dealing with discretization of Helmholtz problems with non-matching interfaces inside the mesh. Hence, we believe this result is new.

We are considering polynomials of degree $1 \leq p \leq 3$. For the case $p > 1$, standard approximation theory requires the solution to be more regular than H^2 to achieve optimal convergence rate. We propose to isolate "non-regular parts" of the solution which are H^2 only. We call them non-regular, because they are not regular enough to apply standard approximation theory with polynomial of degree $p > 1$. The key point in our analysis is to construct special approximants for these non-regular parts.

For the sake of simplicity, we assume in the remaining of the Chapter that $\omega h \leq 1$. Note that this is a rational assumption which means that the number of discretization points per wavelength is bounded below.

We start with a standard approximation property of polynomials of degree p in the norm $||.||_{\omega,c}$.

Proposition 6. *Let $1 \leq p \leq 3$. For all $v \in H^{p+1}(0,Z)$, there exists an element $v_h \in V^{h,p}$ such that*

$$||v - v_h||_{\omega,c} \leq C_a h^p ||v^{(p+1)}||, \tag{2.78}$$

where

$$C_a = 2\hat{C}\max(1, \frac{1}{c_{min}}),$$

and $\hat{C}$ is a numeric constant independent of all parameters.

Proof. Since $v \in H^{p+1}(0,Z)$, classical approximation theory ensures that there exists $v_h \in V^{h,p}$ such that

$$||v - v_h|| \leq \hat{C}h^{p+1}||v_h^{(p+1)}||, \quad ||(v - v_h)'|| \leq \hat{C}h^p||v_h^{(p+1)}||,$$

and therefore

$$\begin{aligned} ||v - v_h||_{\omega,c} &\leq \hat{C}\left(\frac{\omega h^{p+1}}{c_\star} + h^p\right)||v^{(p+1)}|| \\ &\leq \hat{C}\max(1, \frac{1}{c_\star})(1 + \omega h)h^p||v^{(p+1)}|| \\ &\leq 2\hat{C}\max(1, \frac{1}{c_\star})h^p||v^{(p+1)}||. \end{aligned}$$

□

Since we do not assume more than $L^2(0,Z)$ regularity for the right-hand-side, we might not expect the solution to be more than in $H^2(0,Z)$. In Lemma 14, we isolate this non-regular part of the solution and define its approximation.

Lemma 14. *Let $f \in L^2(0,Z)$. There exists a function $\phi \in H_0^1(0,Z) \cap H^2(0,Z)$ such that $\phi'' = f$. Furthermore, there exists an element $\phi_h \in V_h^1$ such that*

$$||\phi - \phi_h||_{\omega,c} \leq C_a h ||f||.$$

Proof. Since the Laplace operator (nothing but second derivative in 1D), is elliptic, it is clear that there exists a unique function $\phi \in H_0^1(0,Z)$, such that $-\phi'' = -f$. Furthermore, the definition of ϕ ensures that $\phi \in H^2(0,Z)$ and $||\phi''|| = ||f||$. We conclude the proof with Proposition 6. □

Lemma 15. *Consider the function $W_l^p \in L^2(0,Z)$ defined by*

$$W_l^p(z) = \frac{1}{p!}(z - z_l)^p 1_{z>z_l}.$$

We have $(W_l^p)^{(p+1)} = \delta_{z_l}$, and there exists a function $v_{l,h}^p \in V^{h,p}$ such that

$$||W_l^p - v_{l,h}^p||_{\omega,c} \leq C_{a,w} h^{p-1/2}, \tag{2.79}$$

with

$$C_{a,w} = 2\max(1, \frac{1}{c_{min}}).$$

Proof. First, it is clear that if $z_l = t_j$ for some integer j, then $W_l^p \in V^{h,p}$ and the Lemma is trivial. Therefore, assume that $z_l \neq t_j$. There exists a unique integer $j_\star$ such that $t_{j\star} < z_l < t_{j_\star+1}$. We define

$$v_{l,h}^p|_{(t_{j-1},t_j)}(z) = \begin{cases} 0, & j < j_\star \\ \frac{1}{p!}(t_{j_\star} - z_l)^p \frac{(z - t_{j_\star-1})^p}{(t_{j_\star} - t_{j_\star-1})^p}, & j = j_\star \\ \frac{1}{p!}(z - z_l)^p, & j > j_\star. \end{cases}$$

One can easily verify that $v_{l,h}^p \in V^{h,p}$. Furthermore, we can show that $v_{l,h}^p$ satisfies (2.79) by direct computations. □

That the solution is non-regular is also due to the velocity parameter c which is discontinuous. Lemma 16 presents one way to isolate those irregularities together with an approximation.

Lemma 16. *For $f \in L^2(0,Z)$, we define*

$$\mu_2 = \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)(z_l) W_l^2, \quad \mu_3 = \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)'(z_l) W_l^3.$$

Then we have

$$\mu_2^{(3)} = \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)(z_l) \delta_{z_l}, \quad \mu_3^{(4)} = \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)'(z_l) \delta_{z_l}.$$

Furthermore, there exist $\mu_{2,h} \in V_h^2$ *and* $\mu_{3,h} \in V_h^3$ *such that*

$$||\mu_2 - \mu_{2,h}||_{\omega,c} \leq C_{a,2}\omega h^{3/2}, \quad ||\mu_3 - \mu_{3,h}||_{\omega,c} \leq C_{a,3}\omega^2 h^{5/2},$$

with

$$C_{a,2} = C_{a,w}C_\infty, \quad C_{a,3} = C_{a,w}C'_\infty \left(\frac{1}{c_{min}^2} - \frac{1}{c_{max}^2}\right).$$

Proof. The first part of Lemma 16 is a direct consequence of the definition of the Dirac distribution δ. Therefore, let us focus on the construction of $\mu_{2,h}$ and $\mu_{3,h}$. We define

$$\mu_{2,h} = \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)(z_l) v_{l,h}^2.$$

In view of (2.79), it is clear that

$$||\mu_2 - \mu_{2,h}||_{\omega,c} \leq C_{a,w} \left(\sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l |(\mathcal{S}_{\omega,c} f)(z_l)|\right) \omega^2 h^{3/2}.$$

Therefore, using (2.27), we obtain

$$||\mu_2 - \mu_{2,h}||_{\omega,c} \leq C_{a,w}C_\infty \omega h^{3/2} ||f||.$$

We define

$$\mu_{3,h} = \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)'(z_l) v_{l,h}^3,$$

and using (2.79), we have

$$\begin{aligned} ||\mu_3 - \mu_{3,h}||_{\omega,c} &\leq C_{a,w} \left(\sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)'(z_l)\right) \omega^2 h^{5/2} \\ &\leq C_{a,w} \left(\sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l\right) ||(\mathcal{S}_{\omega,c} f)'||_\infty \omega^2 h^{5/2}. \end{aligned}$$

Therefore, according to (2.27), we have

$$||\mu_3 - \mu_{3,h}||_{\omega,c} \leq C_{a,w}C'_\infty \left(\frac{1}{c_{min}^2} - \frac{1}{c_{max}^2}\right) \omega^2 h^{5/2} ||f||.$$

□

Theorem 11. *Let* $1 \leq p \leq 3$. *We have*

$$\eta_{\omega,c}^{h,p} \leq C_{\eta,p}\omega h, \tag{2.80}$$

with

$$C_{\eta,1} = C_a C_{s,2}, \quad C_{\eta,2} = C_a \max\left\{1, C_{a,2}, \frac{C_a C_s}{c_{min}^2}\right\}, \quad C_{\eta,3} = C_a \max\left\{1, C_{a,2}, C_{a,3}, \frac{C_a C_{s,2}}{c_{min}^2}\right\}.$$

Furthermore, if

$$\omega^{p+1} h^p \leq \alpha_p, \tag{2.81}$$

where α_p is a constant depending on α and $C_{\eta,p}$ only, then the condition $\omega\eta_{h,p} \leq \alpha$ is satisfied.

Proof. The case of linear approximation is easy. Indeed, reminding estimate (2.22), it is clear that there exists an element $v_{h,1} \in V_h^1$ such that

$$||\mathcal{S}_{\omega,c} f - v_{h,1}||_{\omega,c} \leq C_a h ||(\mathcal{S}_{\omega,c} f)''|| \leq C_a C_{s,2} \omega h ||f||.$$

Therefore $\eta_{\omega,c}^{h,1} \leq C_a C_{s,a} \omega h$ and (2.80) and (2.81) immediately follow for $p = 1$.

We now consider the case $p > 1$. Since $\mathcal{S}_{\omega,c} f$ is a weak solution to (2.70), we have

$$(\mathcal{S}_{\omega,c} f)'' = -f - \frac{\omega^2}{c^2} \mathcal{S}_{\omega,c} f,$$

in the sense of distributions. Hence, defining $\phi \in H^2(0, Z)$ as in Lemma 14, we have

$$(\mathcal{S}_{\omega,c} f - \phi)'' = -\frac{\omega^2}{c^2} \mathcal{S}_{\omega,c} f. \tag{2.82}$$

Differentiating (2.82) in the sense of distributions, we obtain

$$(\mathcal{S}_{\omega,c} f - \phi)^{(3)} = -\frac{\omega^2}{c^2} (\mathcal{S}_{\omega,c} f)' - \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l (\mathcal{S}_{\omega,c} f)(z_l) \delta_{z_l},$$

so that, defining μ_2 as in Lemma 16

$$(\mathcal{S}_{\omega,c} f - \phi - \mu_2)^{(3)} = -\frac{\omega^2}{c^2} (\mathcal{S}_{\omega,c} f)'. \tag{2.83}$$

To conclude on the case $p = 2$, we define $\theta_2 = \mathcal{S}_{\omega,c} f - \phi - \mu_2 \in H^3(0, Z)$. According to (2.83) and (2.19), we have

$$||\theta_2^{(3)}|| \leq \frac{\omega^2}{c_{min}^2} ||(\mathcal{S}_{\omega,c} f)'|| \leq \frac{C_s}{c_{min}^2} \omega^2 ||f||.$$

Therefore, there exists a function $\theta_{2,h} \in V_h^2$ such that

$$||\theta_2 - \theta_{2,h}||_{\omega,c} \leq \frac{C_a C_s}{c_{min}^2} \omega^2 h^2 ||f||.$$

We define $\phi_h \in V_h^1$ and $\mu_{2,h} \in V_h^2$ as in Lemmas 14-16 and $v_{h,2} = \phi_h + \mu_{2,h} + \theta_{2,h}$. Since $\mathcal{S}_{\omega,c} f = \phi + \mu_2 + \theta_2$, we obtain

$$||\mathcal{S}_{\omega,c} f - v_{h,2}||_{\omega,c} \leq \left(C_a h + C_{a,2} \omega h^{3/2} + \frac{C_a C_s}{c_{min}^2} \omega^2 h^2 \right) ||f||,$$

and (2.80) follows for $p = 2$ because $\omega \geq 1$ and $\omega h \leq 1$. We now establish (2.81) for $p = 2$ and we get

$$\begin{aligned} \omega \eta_{\omega,c}^{h,2} &\leq C_{\eta,2} \left(\omega h + \omega^2 h^{3/2} + \omega^3 h^2 \right) \\ &\leq C_{\eta,2} \left(\omega^{-1/2} (\omega^3 h^2)^{1/2} + (\omega h)^{1/2} (\omega^3 h^2)^{1/2} + \omega^3 h^2 \right). \end{aligned}$$

Thus, since $\omega^{-1/2} \leq 1$ and $(\omega h)^{1/2} \leq 1$ assuming that $\omega^3 h^2 \leq \alpha_2$, we have

$$\omega \eta_{\omega,c}^{h,2} \leq C_{\eta,2} \left(2\alpha_2^{1/2} + \alpha_2 \right),$$

and selecting

$$\alpha_2 = \left(\left(\frac{\alpha}{C_{\eta,2}} + 1 \right)^{1/2} - 1 \right)^2,$$

we have $\omega \eta_{\omega,c}^{h,2} \leq \alpha$.

We now tackle the case $p = 3$. We differentiate (2.83) again and obtain

$$(\mathcal{S}_{\omega,c} f - \phi - \mu_2)^{(4)} = -\frac{\omega^2}{c^2} (\mathcal{S}_{\omega,c} f)'' - \omega^2 \sum_{l=1}^{L-1} \left[\frac{1}{c^2} \right]_l (\mathcal{S}_{\omega,c} f)'(z_l) \delta_{z_l},$$

so that, defining μ_3 as in Lemma 16, we have

$$(\mathcal{S}_{\omega,c} f - \phi - \mu_2 - \mu_3)^{(4)} = -\frac{\omega^2}{c^2} (\mathcal{S}_{\omega,c} f)''. \tag{2.84}$$

To conclude, we define $\theta_3 = \mathcal{S}_{\omega,c} f - \phi - \mu_2 - \mu_3 \in H^4(0, Z)$, so that $\mathcal{S}_{\omega,c} f = \phi + \mu_2 + \mu_3 + \theta_3$ and

$$||\theta_3^{(4)}|| \leq \frac{\omega^2}{c_{min}^2} ||(\mathcal{S}_{\omega,c} f)''|| \leq \frac{C_{s,2}}{c_{min}^2} \omega^3 ||f||.$$

Let $\theta_{3,h} \in V_h^3$ be the best approximation to θ_3 and define $\mu_{3,h} \in V_h^3$ as in Lemma 16. We set $v_{h,3} = \phi_h + \mu_{2,h} + \mu_{3,h} + \theta_{3,h} \in V_h^3$, and then we have

$$||\mathcal{S}_{\omega,c} f - v_{h,3}||_{\omega,c} \leq \left(C_a h + C_{a,2} \omega h^{3/2} + C_{a,3,h} \omega^2 h^{5/2} + \frac{C_a C_{s,2}}{c_{min}^2} \omega^3 h^3 \right) ||f||,$$

and we obtain (2.80) and (2.81) for $p = 3$ using the same arguments than for $p = 2$. □

Our analysis is limited to the case $p \leq 3$. We have obtained that when $p \geq 2$, $\omega ||\mu_2 - \mu_{2,h}||_{\omega,c}$ is bounded by $(\omega^4 h^3)^{1/2}$. Hence, even for $p > 3$, the best estimate for $\omega \eta_{\omega,c}^{h,p}$ is $\omega^4 h^3$.

Now that $\eta_{\omega,c}^{h,p}$ has been estimated, we are able to deliver frequency-explicit stability conditions and error-estimates.

Corollary 2. *Let $1 \leq p \leq 3$. Assume that $\omega^{p+1}h^p \leq \alpha_p$. Then the discrete problem admits a unique numerical solution $\mathcal{S}^{h,p}_{\omega,c}f$ and the following error-estimate holds*

$$||\mathcal{S}_{\omega,c}f - \mathcal{S}^{h,p}_{\omega,c}f||_{\omega,c} \leq C_{e,p}\omega h||f||,$$

where α_p is defined in Theorem 11 and

$$C_{e,3} = C_e C_{\eta,p}.$$

Proof. The proof is a direct application of Theorem 10 and Theorem 11. Indeed, assuming $\omega^{p+1}h^p \leq \alpha_p$, (2.81) yields $\omega\eta^{h,p}_{\omega,c} \leq \alpha$ and we can use Theorem 10. The result directly follows from (3.23) because we also have $\eta^{h,p}_{w,c} \leq C_{\eta,p}\omega h$ from (2.80). □

2.3.4 Multiscale medium approximation

In the previous subsections, we have derived frequency-explicit stability conditions and asymptotic error-estimates assuming that we are able to solve problem (2.76) exactly. It requires thus to compute the coefficients of the linear system, including the integrals

$$\int_0^Z \frac{1}{c^2(z)}\phi_h(z)\overline{\psi_h(z)}dz, \tag{2.85}$$

for all basis functions $\phi_h, \psi_h \in V^{h,p}$ of $V^{h,p}$. Of course, in a one-dimensional space, it is always possible to evaluate integral (2.85) analytically, since it can be decomposed into several intervals where c is constant. However, this is not the case in two-dimensional domains. Furthermore, even when the analytical formula is available (for example, if we assume that the interfaces defining c are polygons in 2D), it might be expensive to compute, since the quadrature scheme to be used will be different in each cell.

We propose a different approach which consists in approximating c by another parameter c_ϵ designed so that the integrals (2.85) are always cheap to compute numerically. We construct c_ϵ so that $(c_\epsilon^2, 1) \in M$, where M is the set of admissible propagation media (κ, ρ) defined in Definition 1. Hence, Theorem 7 will ensure the quality of the numerical approximation. We call this process the Multiscale Medium Approximation method (MMAm), just because the scale ϵ of the medium approximation is independent of the scale h of the finite element approximation.

We now tackle the construction of c_ϵ. We suppose that the velocity parameter c is such that $(c^2, 1) \in M$ and

$$\min_{l\in\{1,\dots,L\}} z_l - z_{l-1} > 2\mathbf{h}_\star. \tag{2.86}$$

We use the discretization space (2.71) with $n \in \mathbb{N}^\star$ cells. We consider $m \in \mathbb{N}^\star$ subdivisions of each cell of the mesh. Then, for $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$

$$c_\epsilon|_{(t_i^{j-1}, t_i^j)} = \sup_{(t_i^{j-1}, t_i^j)} c$$

where $t_i^j = t_i + j\epsilon h$, and $\epsilon = 1/m$. Note that, because of (2.86), it is clear that both $(c^2, 1)$ and $(c_\epsilon^2, 1)$ belong to M as soon as $h\epsilon < \mathbf{h}_\star$. In Lemma 17, we show that the medium is properly approximated if $h\epsilon$ is small enough.

Lemma 17. *Assume that $h\epsilon < \mathbf{h}_\star$. Then we have*

$$|c^{-2} - c_\epsilon^{-2}|_1 \leq (c_{min}^{-2} - c_{max}^{-2})\epsilon h. \tag{2.87}$$

Proof. For each $l \in \{1, \dots, L-1\}$, there exists a unique pair $i_l \in \{1, \dots, n\}$ and $j_l \in \{1, \dots, m\}$ such that $z_l \in [t_{i_l}^{j_l - 1}, t_{i_l}^{j_l})$. Furthermore, since $h\epsilon < \mathbf{h}_\star$, $i_l \neq i_k$ and $j_l \neq j_k$ if $j \neq k$. If $i \neq i_l$ and $j \neq j_l$ for all $l \in \{1, \dots, L-1\}$, c is constant on (t_i^{j-1}, t_i^j) and therefore $c_\epsilon = c$ on this interval. It follows that

$$|c^{-2} - c_\epsilon^{-2}|_1 = \sum_{l=1}^{L-1} \int_{t_{i_l}^{j_l - 1}}^{t_{i_l}^{j_l}} |c^{-2}(z) - c_\epsilon^{-2}(z)| dz \leq h\epsilon \sum_{l=1}^{L-1} \left[\frac{1}{c^2}\right]_l \leq h\epsilon \left(c_{min}^{-2} - c_{max}^{-2}\right).$$

□

In Theorem 12, we conclude about the convergence of the MMAm.

Theorem 12. *Assume that $h\epsilon < \mathbf{h}_\star$ and $\omega^{p+1} h^p \leq \alpha_p$. Then for all $f \in L^2(0, Z)$, there exists a unique element $\mathcal{S}_{\omega, c_\epsilon}^{h,p} f \in V^{h,p}$ such that*

$$B_{\omega, c_\epsilon}(\mathcal{S}_{\omega, c_\epsilon}^{h,p} f, v) = \int_0^Z f(z)\overline{v(z)} dz, \tag{2.88}$$

for all $v \in V^{h,p}$. Furthermore, it holds that

$$||\mathcal{S}_{\omega,c} f - \mathcal{S}_{\omega, c_\epsilon}^{h,p} f||_{\omega,c} \leq C_{e,p,\epsilon}(\omega h + \omega^2 h\epsilon)||f||, \tag{2.89}$$

with

$$C_{e,p,\epsilon} = \frac{c_{max}}{c_{min}} \max\left\{(c_{min}^{-2} - c_{max}^{-2}) C_{s,m} Z, C_{e,p}\right\}.$$

Proof. Using Lemma 17, the result directly follows from Theorem 7, the bounds (2.80) and (2.81) from Corollary 2 and the error-estimate (3.23) of Theorem 10:

$$\begin{aligned}
||\mathcal{S}_{\omega,c} f - \mathcal{S}_{\omega, c_\epsilon}^{h,p} f||_{\omega,c} &\leq ||\mathcal{S}_{\omega,c} f - \mathcal{S}_{\omega, c_\epsilon} f||_{\omega,c} + ||\mathcal{S}_{\omega, c_\epsilon} f - \mathcal{S}_{\omega, c_\epsilon}^{h,p} f||_{\omega,c} \\
&\leq ||\mathcal{S}_{\omega,c} f - \mathcal{S}_{\omega, c_\epsilon} f||_{\omega,c} + \frac{c_{max}}{c_{min}} ||\mathcal{S}_{\omega, c_\epsilon} f - \mathcal{S}_{\omega, c_\epsilon}^{h,p} f||_{\omega, c_\epsilon} \\
&\leq \frac{c_{max} C_{s,m}}{c_{min}} \omega^2 ||c^{-2} - c_\epsilon^{-2}||_1 ||f|| + \frac{c_{max} C_{e,p}}{c_{min}} \omega h ||f|| \\
&\leq \left(\frac{c_{max} C_{s,m}}{c_{min}} (c_{min}^{-2} - c_{max}^{-2}) Z \omega^2 h\epsilon + \frac{c_{max} C_{e,p}}{c_{min}} \omega h\right) ||f||.
\end{aligned}$$

□

Chapter 3

Analysis of the problem in two dimensions

This chapter is devoted to the analysis of the problem in 2D. We are not able to extend all the results obtained in 1D to the two-dimensional case. Like in the 1D case, we start by deriving frequency-explicit stability estimates for the continuous problem in Section 3.1. We then turn to the convergence analysis of the MMAm for 2D problems in Section 3.2.

3.1 Analysis of the continuous problem

3.1.1 Background

Makridakis, Ihlenburg and Babuška consider the problem of a one dimensional fluid-solid interaction [70]. A solid body is immersed in a fluid and waves are propagating. They use special test functions of the form $v(x) = (x - \bar{x})\mathbf{u}'(x)$, where $\bar{x}$ is a carefully selected point.

The method of Makridakis, Ihlenburg and Babuška is generalized to two dimensional problems in star-shaped domains by Melenk during his PhD [73]. The test function is generalized to $v(x) = x \cdot \nabla \mathbf{u}(x)$ in two dimensions.

The method is further extended to three dimensional homogeneous media and to elastic waves (but the results are not optimal for the elastic case) by Cummings and Feng [39]. The case of mixed boundary conditions is handled with the same test-functions by Hetmaniuk [55].

The test function $v(x) = x \cdot \nabla \mathbf{u}(x)$, is called a "Morawetz multiplier". This multiplier was first introduced by Morawetz in [77] to analyse a non-linear time-domain wave equation and is commonly used in the analysis of Helmholtz and Schrödinger problems (see for instance [67, 81]). A general presentation of how Morawetz multipliers can be applied to Helmholtz problems is given in [81]. In particular, the analysis presented in [81] includes heterogeneous propagation media.

In the following, we use a Morawetz multiplier to show frequency-explicit stability-estimates for Helmholtz problems in heterogeneous propagation media where the density

is constant and the wavespeed satisfies a monotonicity hypothesis. We consider bounded propagation media surrounded by an absorbing condition.

The results presented in this section are strongly linked to [81]. In particular, our monotonicity hypothesis is similar to the assumptions of [81]. The main difference between our results and [81] is the boundary conditions. In [81], the authors consider an Helmholtz problem set in the whole space $\mathbb{R}^N$ and the uniqueness of the solution is ensured by the limiting absorption principle. In contrast, we propose to analyse the problem in a bounded domain of $\mathbb{R}^2$ surrounded by a first-order radiation condition.

3.1.2 Problem statement

We now consider acoustic wave propagation in two dimensions. We consider problem (1.33) with two simplifications. First, we assume that the propagation domain $\Omega = (0, L_1) \times (0, L_2) \subset \mathbb{R}^2$, is a rectangle. Second we assume that the density ρ is constant (see (1.25)). We further assume that there is no free surface, and the whole domain is surrounded by an absorbing condition. Finally, we are using a simplified absorbing boundary condition compared to (1.33). As a result, the propagation of harmonic seismic waves is governed by the equation:

$$\left\{ \begin{array}{rcll} -k^2 u - \Delta u & = & f & \text{in } \Omega \\ \nabla u \cdot \mathbf{n} - ik_{max} u & = & 0 & \text{on } \partial\Omega. \end{array} \right. \tag{3.1}$$

The wave number k is defined from the pulsation ω and the velocity c through the relation $k = \omega/c$. The pulsation is a given positive constant and the velocity varies in the whole domain. Since we are especially concerned with high frequency waves, we consider pulsations ω higher than a given minimum ω_0. The field f is a given distributed source. To get into the right condition of numerical experiments, we assume that the domain of interest is limited by an absorbing boundary. We thus set the simplest outgoing radiation condition on the boundary of Ω. We use the coefficient $k_{max} = \sup_\Omega k$ to define this radiation condition.

We assume that $c \in L^\infty(\Omega)$ is piecewise constant and the values of c are distributed as follows. The velocity model is composed of R subdomains Ω_r enclosed in Ω and in each Ω_r, the velocity is $c_r = c|_{\Omega_r} \in \mathbb{R}^{+*}$ with $c_{min} = \min_r c_r$, $c_{max} = \max_r c_r$ and we assume that $c_{min} > 0$. It is worth noting that $k_{max} = \omega/c_{min}$. We further assume that there exists a point $x_0 \in \Omega$ such that

$$\frac{\mathbf{n}_r \cdot (x - x_0)}{c_r^2} + \frac{\mathbf{n}_l \cdot (x - x_0)}{c_l^2} < 0 \quad \forall x \in \partial\Omega_r \cap \partial\Omega_l, \tag{3.2}$$

for all $r, l \in \{1, \ldots, R\}$ such that $\Omega_r \cap \Omega_l \neq \emptyset$, where $\mathbf{n}_r$ and $\mathbf{n}_l$ stand for the unit outward normal vectors to $\partial\Omega_r$ and $\partial\Omega_l$ respectively. Examples of velocity models satisfying (3.2) are given in Figures 3.1 and 3.2.

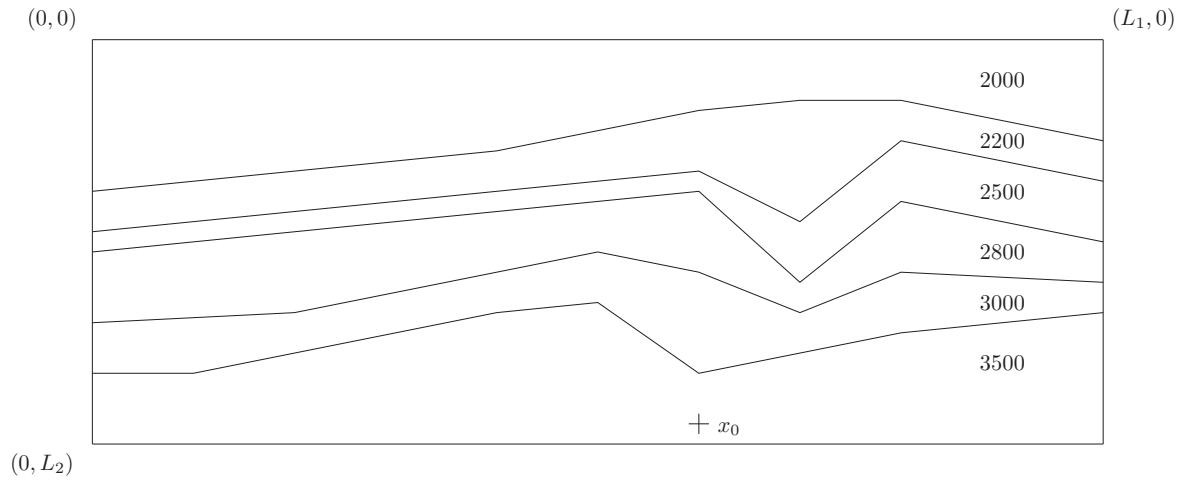


Figure 3.1: A stratified velocity parameter

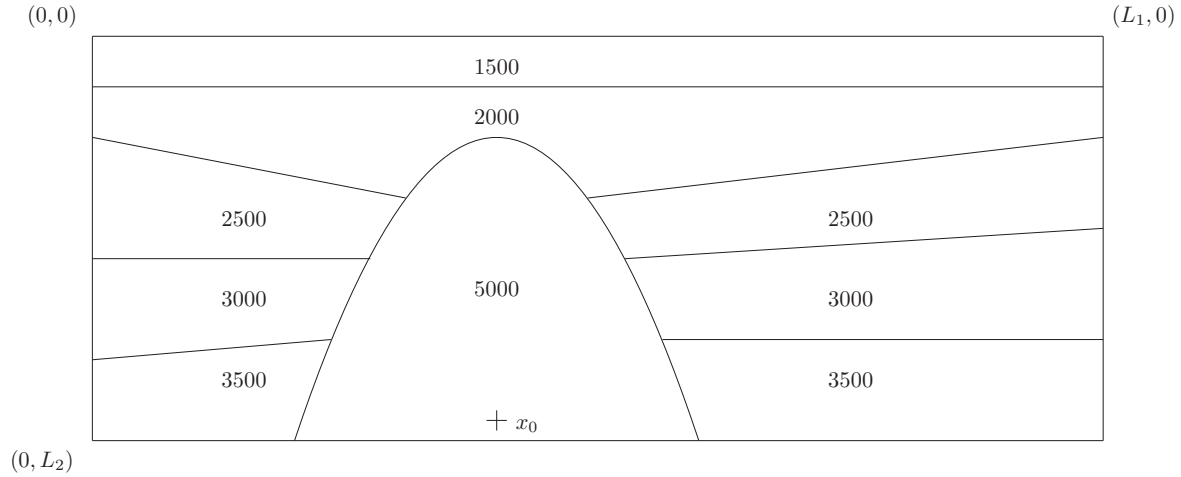


Figure 3.2: A velocity parameter with a salt body

In the following, we employ the notation $k_r = \omega/c_r$.

It is well-known that $u \in H^1(\Omega, \mathbb{C})$ is solution to (3.1) in a weak sense if and only if u satisfies the variational equation

$$B(u, v) = -\int_\Omega k^2 u\bar{v} - ik_{max}\int_{\partial\Omega} u\bar{v} + \int_\Omega \nabla u \cdot \nabla\bar{v} = \int_\Omega f\bar{v}, \tag{3.3}$$

for all $v \in H^1(\Omega, \mathbb{C})$, where $B : H^1(\Omega, \mathbb{C}) \times H^1(\Omega, \mathbb{C}) \to \mathbb{C}$ is the sesquilinear form associated with Problem (3.1). The proof is similar to the one-dimensional case and is not repeated here.

Proposition 7. *Let $u \in H^1(\Omega, \mathbb{C})$ be any solution to (3.3). Then $u \in H^2(\Omega, \mathbb{C})$ and there exists a constant $C := C(\Omega, c_{min})$ such that*

$$|u|^2_{2,\Omega} \leq C\big(|f|^2_{0,\Omega} + (\omega^2 + \omega^4)|u|^2_{0,\Omega} + \omega^2|u|^2_{1,\Omega}\big).$$

Proof. u being a solution to (3.3), it satisfies

$$\int_\Omega \nabla u \cdot \nabla \bar{v} = \int_\Omega F\bar{v} + \int_{\partial\Omega} G\bar{v} \qquad \forall v \in H^1(\Omega, \mathbb{C}),$$

with $F = f + k^2 u$ and $G = ik_{max}u$. Since $k \in L^\infty(\Omega)$ and $u \in H^1(\Omega, \mathbb{C})$, we have $F \in L^2(\Omega, \mathbb{C})$ and $G \in H^{1/2}(\partial\Omega, \mathbb{C})$. Since Ω is convex, the classical theory for the homogeneous Laplace operator implies that there exists a constant C depending on Ω only such that

$$|u|^2_{2,\Omega} \leq C\big(|F|^2_{0,\Omega} + ||G||^2_{1/2,\partial\Omega}\big).$$

Furthermore, regarding norms $|F|^2_{0,\Omega}$ and $||G||^2_{1/2,\partial\Omega}$, we have

$$\begin{aligned} |F|^2_{0,\Omega} &= |f + k^2 u|^2_{0,\Omega} \\ &\leq |f|^2_{0,\Omega} + k^4_{max}|u|^2_{0,\Omega} \\ &\leq C\big(|f|^2_{0,\Omega} + \omega^4|u|^2_{0,\Omega}\big), \end{aligned}$$

with $C = \max(1, 1/c^4_{min})$. Moreover,

$$\begin{aligned} ||G||^2_{1/2,\partial\Omega} &= ||ik_{max}u||^2_{1/2,\partial\Omega} \\ &= k^2_{max}||u||^2_{1/2,\partial\Omega} \\ &\leq C\omega^2||u||^2_{1/2,\partial\Omega}, \end{aligned}$$

with $C = \max(1, 1/c^2_{min})$. We end the proof thanks to the following trace inequality

$$||u||^2_{1/2,\Omega} \leq C\big(|u|^2_{0,\Omega} + |u|^2_{1,\Omega}\big),$$

where C is a constant depending on Ω only. □

Before turning to stability in the $L^2(\Omega, \mathbb{C})$ norm, we need two identities. It is worth mentioning that Lemma 18 is only valid in 2D. We refer the reader to [55] for a 3D version of Lemma 18.

Lemma 18. *For all $w \in H^2(\Omega, \mathbb{C})$,*

$$2Re \int_\Omega \nabla w \cdot \nabla(\mathbf{x} \cdot \nabla \bar{w}) = \int_{\partial\Omega} |\nabla w|^2 \mathbf{x} \cdot \mathbf{n}. \tag{3.4}$$

For all $w \in H^1(\Omega, \mathbb{C})$,

$$2Re \int_\Omega k^2 w\mathbf{x}\cdot\nabla\bar{w} = -2\int_\Omega k^2|w|^2 - \sum_{r,l=1}^{R} \int_{\Omega_r \cap \Omega_l} (k_r^2 \mathbf{x}\cdot n_r + k_l^2 \mathbf{x}\cdot n_l)|w|^2 + \int_{\partial\Omega} k^2|w|^2\mathbf{x}\cdot\mathbf{n}, \tag{3.5}$$

where $\mathbf{x} = x - x_0$.

Proof. In the proof of (3.4) and (3.5), we use the identity

$$2\text{Re } v\partial_j\bar{v} = \partial_j|v|^2, \quad \forall v \in H^1(\Omega,\mathbb{C}),\ j=1,2. \tag{3.6}$$

To demonstrate (3.4), we first develop the expression:

$$\begin{aligned}
\partial_j w\partial_j(\mathbf{x}\cdot\nabla\bar{w}) &= \sum_{k=1}^{2}\partial_j w\partial_j(\mathbf{x}_k\partial_k\bar{w}) \\
&= \sum_{k=1}^{2}\partial_j w(\partial_j\mathbf{x}_k\partial_k\bar{w}+\mathbf{x}_k\partial_{jk}\bar{w}) \\
&= \sum_{k=1}^{2}\delta_{jk}\partial_j w\partial_k\bar{w}+\sum_{k=1}^{2}\mathbf{x}_k\partial_j w\partial_{jk}\bar{w} \\
&= |\partial_j w|^2+\sum_{k=1}^{2}\mathbf{x}_k\partial_j w\partial_k(\partial_j\bar{w})
\end{aligned}$$

Using (3.6) with $v=\partial_j w$, we get

$$\begin{aligned}
2\text{Re }\partial_j w\partial_j(\mathbf{x}\cdot\nabla\bar{w}) &= 2|\partial_j w|^2+\sum_{k=1}^{2}\mathbf{x}_k\partial_k|\partial_j w|^2 \\
&= 2|\partial_j w|^2+\mathbf{x}\cdot\nabla|\partial_j w|^2.
\end{aligned}$$

We shall now integrate and then use a Green formula:

$$\begin{aligned}
2\text{Re }\int_\Omega\partial_j w\partial_j(\mathbf{x}\cdot\nabla\bar{w}) &= 2\int_\Omega|\partial_j w|^2+\int_\Omega\mathbf{x}\cdot\nabla|\partial_j w|^2 \\
&= 2\int_\Omega|\partial_j w|^2-\int_\Omega\operatorname{div}\mathbf{x}|\partial_j w|^2+\int_{\partial\Omega}\mathbf{x}\cdot\mathbf{n}|\partial_j w|^2 \\
&= 2\int_\Omega|\partial_j w|^2-\int_\Omega 2|\partial_j w|^2+\int_{\partial\Omega}\mathbf{x}\cdot\mathbf{n}|\partial_j w|^2 \\
&= \int_{\partial\Omega}\mathbf{x}\cdot\mathbf{n}|\partial_j w|^2
\end{aligned}$$

We demonstrate (3.4) by summing over j. We now turn to (3.5).

$$\begin{aligned}
2\mathrm{Re}\int_\Omega k^2 w\mathbf{x}\cdot\nabla\bar{w} &= \int_\Omega k^2\mathbf{x}\cdot\nabla|w|^2 \\
&= \sum_{r=1}^{R} k_r^2\int_{\Omega_r}\mathbf{x}\cdot\nabla|w|^2 \\
&= \sum_{r=1}^{R} k_r^2\left\{-\int_{\Omega_r}\mathrm{div}\,\mathbf{x}|w|^2+\int_{\partial\Omega_r}\mathbf{x}\cdot\mathbf{n}_r|w|^2\right\} \\
&= -2\int_\Omega k^2|w|^2+\sum_{r=1}^{R} k_r^2\int_{\partial\Omega_r}\mathbf{x}\cdot\mathbf{n}_r|w|^2. \\
&= -2\int_\Omega k^2|w|^2+\sum_{r,l=1}^{R}\int_{\Omega_r\cap\Omega_l}(k_r^2\mathbf{x}\cdot\mathbf{n}_r+k_l^2\mathbf{x}\cdot\mathbf{n}_l)|w|^2+\int_{\partial\Omega}k^2\mathbf{x}\cdot\mathbf{n}|w|^2
\end{aligned}$$

□

We are now ready to introduce our stability result in the $L^2(\Omega,\mathbb{C})$ norm. We point out that Proposition 8 is valid for all pulsations $\omega\geq\omega_0$.

Proposition 8. *Let $u\in H^1(\Omega,\mathbb{C})$ be any solution to (3.3). Then there exists a constant $C:=C(\Omega,c_{max},c_{min},x_0,\omega_0)$ such that*

$$|u|_{0,\Omega}\leq\frac{C}{\omega}|f|_{0,\Omega}.$$

Proof. According to Proposition 7, $u\in H^2(\Omega,\mathbb{C})$ and $v=\mathbf{x}\cdot\nabla u$ is regular enough to be used as a test function in the variational equation (3.3). Recalling (3.4) and (3.5), then taking the real part of (3.3), we have

$$\begin{aligned}
2\int_\Omega k^2|u|^2-\sum_{r,l=1}^{R}\int_{\Omega_r\cap\Omega_l}(k_r^2\mathbf{x}\cdot n_r+k_l^2\mathbf{x}\cdot n_l)|u|^2+\int_{\partial\Omega}|\nabla u|^2\mathbf{x}\cdot\mathbf{n} \\
=2\mathrm{Re}\int_\Omega f\mathbf{x}\cdot\nabla\bar{u}+2\mathrm{Re}\;ik_{max}\int_{\partial\Omega}u\mathbf{x}\cdot\nabla\bar{u}+\int_{\partial\Omega}k^2|u|^2\mathbf{x}\cdot\mathbf{n}.
\end{aligned}$$

Ω being a rectangle, it is strictly star-shaped with respect to x_0, and there exists a constant $\gamma>0$ depending on Ω and x_0 only such that $\mathbf{x}\cdot\mathbf{n}\geq\gamma$ on $\partial\Omega$. Since c satisfies (3.2), we have $(k_r^2\mathbf{x}\cdot\mathbf{n}_r+k_l^2\mathbf{x}\cdot\mathbf{n}_l)\leq 0$. Then, observing that $|\mathbf{x}|\leq\mathrm{diam}\;\Omega=(L_1^2+L_2^2)^{1/2}=L$, it follows

$$\begin{aligned}
2k_{min}^2|u|^2+\gamma|\nabla u|_{0,\partial\Omega}^2 &\leq 2L|f|_{0,\Omega}|u|_{1,\Omega}+2Lk_{max}|u|_{0,\partial\Omega}|\nabla u|_{0,\partial\Omega}+Lk_{max}^2|u|_{0,\partial\Omega}^2 \\
&\leq \frac{L^2}{\epsilon}|f|_{0,\Omega}^2+\epsilon|u|_{1,\Omega}^2+\frac{L^2k_{max}^2}{\gamma}|u|_{0,\partial\Omega}^2+\gamma|\nabla u|_{0,\Omega}^2+Lk_{max}^2|u|_{0,\partial\Omega}^2.
\end{aligned}$$

We then get that for any $\epsilon > 0$

$$2k_{min}|u|^2_{0,\Omega} \leq \frac{L^2}{\epsilon}|f|^2_{0,\Omega} + \epsilon|u|^2_{1,\Omega} + \left(\frac{L^2}{\gamma} + L\right) k^2_{max}|u|^2_{0,\partial\Omega}. \tag{3.7}$$

We complete the proof by deriving estimates for $|u|_{1,\Omega}$ and $|u|_{0,\partial\Omega}$. This is carried out by picking $v = u$ as a test function in (3.3) and considering the real and imaginary parts separately. We start by studying $|u|_{1,\Omega}$. We have:

$$\text{Re } B(u,u) = -\int_\Omega k^2|u|^2 + \int_\Omega |\nabla u|^2 = \text{Re } \int_\Omega f\bar{u} \leq |f|_{0,\Omega}|u|_{0,\Omega}.$$

It follows that

$$|u|^2_{1,\Omega} \leq |f|_{0,\Omega}|u|_{0,\Omega} + k^2_{max}|u|^2_{0,\Omega} \leq \frac{1}{4k^2_{max}}|f|^2_{0,\Omega} + 2k^2_{max}|u|^2_{0,\Omega}.$$

Then selecting $\epsilon = k^2_{min}/4k^2_{max}$, we obtain

$$\frac{L^2}{\epsilon_0}|f|_{0,\Omega} + \epsilon_0|u|^2_{1,\Omega} \leq \left(4L^2\frac{k^2_{max}}{k^2_{min}} + \frac{k^2_{min}}{k^4_{max}}\right)|f|^2_{0,\Omega} + \frac{k^2_{min}}{2}|u|^2_{0,\Omega}. \tag{3.8}$$

We now move on estimating $|u|_{0,\partial\Omega}$. We have

$$\text{Im } B(u,u) = -k_{max}|u|^2_{0,\partial\Omega} = \text{Im } \int_\Omega f\bar{u}.$$

It follows that

$$\begin{aligned}\left(\frac{L^2}{\gamma} + L\right) k^2_{max}|u|^2_{0,\partial\Omega} &\leq \left(\frac{L^2}{\gamma} + L\right) k_{max}|f|_{0,\Omega}|u|_{0,\Omega} \\ &\leq \frac{1}{2}\left(\frac{L^2}{\gamma} + L\right)^2 \frac{k^2_{max}}{k^2_{min}}|f|^2_{0,\Omega} + \frac{k^2_{min}}{2}|u|^2_{0,\Omega}. \end{aligned} \tag{3.9}$$

Combining (3.7), (3.8) with (3.9), we get

$$k^2_{min}|u|^2 \leq \left\{\left(4L^2 + \frac{1}{2}\left(\frac{L^2}{\gamma} + L\right)^2\right)\frac{k^2_{max}}{k^2_{min}} + \frac{k^2_{min}}{k^4_{max}}\right\}|f|^2_{0,\Omega},$$

so that the proposition holds with

$$C = c_{max}\sqrt{\left(4L^2 + \frac{1}{2}\left(\frac{L^2}{\gamma} + L\right)^2\right)\frac{c^2_{min}}{c^2_{max}} + \frac{c^2_{max}}{c^4_{min}}\frac{1}{\omega^2_0}}.$$

□

We end this section by a full statement of the results obtained in the section.

Theorem 13. *For all $f \in L^2(\Omega, \mathbb{C})$ and for all $\omega \geq \omega_0$, problem (3.3) admits a unique solution $\mathcal{S}_{\omega,c} f \in H^1(\Omega, \mathbb{C})$. Furthermore, $\mathcal{S}_{\omega,c} f \in H^2(\Omega, \mathbb{C})$, and there exists a constant $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ such that*

$$|\mathcal{S}_{\omega,c} f|_{0,\Omega} \leq \frac{C}{\omega} |f|_{0,\Omega}, \quad |\mathcal{S}_{\omega,c} f|_{1,\Omega} \leq C |f|_{0,\Omega}, \quad |\mathcal{S}_{\omega,c} f|_{2,\Omega} \leq C\omega |f|_{0,\Omega}.$$

Proof. Regarding existence and uniqueness, observe that the sesquilinear form B satisfies a Gårding inequality. Indeed for all $v \in H^1(\Omega, \mathbb{C})$, we have

$$\text{Re } B(v, v) = -\int_\Omega k^2 |v|^2 + \int_\Omega |\nabla v|^2 \geq -k_{max}^2 |v|_{0,\Omega}^2 + |v|_{1,\Omega}^2.$$

Therefore, it follows that we can apply the Fredholm alternative and thus focus on uniqueness (see Chapter 2 of [87]). But Proposition 8 applied to (3.3) with $f = 0$ implies that $u = 0$, which proves uniqueness and thus existence.

Problem (3.3) admits thus a unique solution $\mathcal{S}_{\omega,c} f \in H^1(\Omega, \mathbb{C})$. Now, Proposition 8 implies that

$$|\mathcal{S}_{\omega,c} f|_{0,\Omega} \leq \frac{C_0}{\omega} |f|_{0,\Omega},$$

with a suitable constant C_0. Moreover, if we write $u = \mathcal{S}_{\omega,c} f$, we have

$$\text{Re } B(u, u) = -\int_\Omega k^2 |u|^2 + \int_\Omega |\nabla u|^2 = \text{Re } \int_\Omega f \bar{u}.$$

which implies that

$$\begin{aligned}
|u|_{1,\Omega}^2 &\leq |f|_{0,\Omega} |u|_{0,\Omega} + k_{max}^2 |u|_{0,\Omega}^2 \\
&\leq \frac{1}{4k_{max}^2} |f|_{0,\Omega}^2 + 2k_{max}^2 |u|_{0,\Omega}^2 \\
&\leq \left(\frac{c_{min}^2}{4\omega_0^2} + 2\frac{C_0^2}{c_{min}^2} \right) |f|_{0,\Omega}^2 \\
&\leq C_1^2 |f|_{0,\Omega}^2.
\end{aligned}$$

The demonstration of the theorem is then ended since Proposition 7 allows to write the estimates:

$$\begin{aligned}
|u|_{2,\Omega}^2 &\leq C(\Omega)\big(|f|_{0,\Omega}^2 + (\omega^2 + \omega^4)|u|_{0,\Omega}^2 + \omega^2 |u|_{1,\Omega}^2\big) \\
&\leq C(\Omega)\big(1 + (1 + \omega^2)C_0^2 + \omega^2 C_1^2\big)|f|_{0,\Omega}^2 \\
&\leq C(\Omega)\big(\frac{1 + C_0^2}{\omega^2} + (C_0^2 + C_1^2)\big)\omega^2 |f|_{0,\Omega}^2 \\
&\leq C(\Omega)\big(\frac{1 + C_0^2}{\omega_0^2} + (C_0^2 + C_1^2)\big)\omega^2 |f|_{0,\Omega}^2 \\
&\leq C_2^2 \omega^2 |f|_{0,\Omega}^2.
\end{aligned}$$

□

Corollary 3. *For all $g \in L^2(\Omega, \mathbb{C})$ and for all $\omega \geq \omega_0$, there exists a unique solution $\mathcal{S}^\star_{\omega,c} g \in H^1(\Omega, \mathbb{C})$ such that*

$$B(w, \mathcal{S}^\star_{\omega,c} g) = \int_\Omega w\bar{g}, \quad \forall w \in H^1(\omega, \mathbb{C}).$$

Furthermore, $\mathcal{S}^\star_{\omega,c} g \in H^2(\Omega, \mathbb{C})$, and there exists a constant $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ such that

$$|\mathcal{S}^\star_{\omega,c} g|_{0,\Omega} \leq \frac{C}{\omega}|g|_{0,\Omega}, \quad |\mathcal{S}^\star_{\omega,c} g|_{1,\Omega} \leq C|g|_{0,\Omega}, \quad |\mathcal{S}^\star_{\omega,c} g|_{2,\Omega} \leq C\omega|g|_{0,\Omega}.$$

Proof. To start with, by definition of $\mathcal{S}_{\omega,c}\bar{g}$, we have

$$B(\mathcal{S}_{\omega,c}\bar{g}, v) = \int_\Omega \bar{g}\bar{v}, \quad \forall v \in H^1(\Omega, \mathbb{C}). \tag{3.10}$$

Then, we easily remark that $B(\mathcal{S}_{\omega,c}\bar{g}, v) = B(\bar{v}, \overline{\mathcal{S}_{\omega,c}\bar{g}})$. Hence setting $w = \bar{v}$ in (3.10), we obtain

$$B(w, \overline{\mathcal{S}_{\omega,c}\bar{g}}) = \int_\Omega \bar{g}w, \quad \forall w \in H^1(\Omega, \mathbb{C}).$$

Therefore, $\mathcal{S}^\star_{\omega,c} g = \overline{\mathcal{S}_{\omega,c}\bar{g}}$. Hence, existence uniqueness and stability estimates follow from Theorem 13. □

3.2 Discretization using the Multiscale Medium Approximation method

3.2.1 Background

We explained in detail the importance of frequency-explicit convergence analysis of finite element schemes in the 1D case. Let us give a brief comparison of the results available in the literature between 1D and 2D.

We explained that asymptotic error-estimates are purely based on finite element techniques using the Schatz argument. Therefore, they are easily applicable to 2D problems. For instance, the case of linear elements is treated in Melenk's PhD [73]. Also, Melenk and Sauter have developed an asymptotic theory for arbitrary polynomial degree when the right hand side is in L^2 only [75, 76].

Dispersion relations can also be extended to 2D in some cases. As we already mention, the key ingredient is the space periodicity of the mesh. Dispersion analysis in two and three dimensions includes the work of Mullen and Belytschko [78], Abboud and Pinsky [1] and Ainsworth [3] for cartesian grids. An analysis for more general schemes is given by Deraemaeker, Babuška and Bouillard [41].

Concerning asymptotic error-estimate, an optimal results has been given by Ihlenburg and Babuška [60] for 1D homogeneous problems: provided that $\omega h \leq C$, the finite element solution u_h satisfies

$$\omega|u - u_h|_0 + |u - u_h|_1 \leq C_1 \omega h + C_2 \omega^{2p+1} h^{2p}, \tag{3.11}$$

where C_1 and C_2 are two constants independent of ω and h.

The proof of (3.11) given by Ihlenburg and Babuška [60] is tricky to extend to the higher dimensions or to heterogeneous media. It involves subtle arguments including stability estimates in dual norms and specific interpolants. As a result, no rigorous generalization of their work was available until very recently.

A simpler method to obtain pre-asymptotic error-estimates has been recently developed by Wu and Zhu [106, 109]. It consists in defining an elliptic projection of the Riesz representation of the error and we can think about the proof as an extension of the Schatz argument.

The argument of Wu and Zhu applies to 2D and 3D and to rough right hand sides $f \in L^2$. However, the resulting pre-asymptotic error-estimates are not optimal compared to (2.69). Their result reads: if $\omega^{p+2} h^{p+1} \leq C$ then

$$\omega|u - u_h|_0 + |u - u_h|_1 \leq C_1 \omega^p h^p + C_2 \omega^{p+2} h^{p+1}. \tag{3.12}$$

In pre-asymptotic estimate (3.12), the error is decomposed into the best approximation error and a pollution term like in (3.11). The difference is that it is only valid in a given range where $\omega^{p+2} h^{p+1} \leq C$. Also the order of the pollution term is not optimal as compared to (3.11).

The main achievement of this section is the derivation of a pre-asymptotic error-estimate for the MMAm with linear Lagrangian elements using the method of Wu and Zhu in the context of heterogeneous media.

3.2.2 Problem statement

In this section, we pertain to a finite element discretization of problem (3.3) and we study its convergence with respect to the pulsation ω and the maximum size h of cells forming the mesh.

We propose convergence estimates which are based on the analysis of Zhu and Wu [106, 109]. Our proof is elaborated for a 2D heterogeneous domain and its main ingredient is the construction of an approximate propagation medium by the mean of an approximate velocity c_ϵ. We are then able to extend the optimal convergence result for linear elements in homogeneous media providing that $\omega^3 h^2$ and $\omega \mathcal{M}_{h,\epsilon}$ are small enough. The quantity $\mathcal{M}_{h,\epsilon}$ which involves two parameters h and ϵ, stands for the approximation error of c by c_ϵ (see Definition 7). As abovementioned, h denotes the discretization step related to the finite element mesh while ϵ represents the size of the local submesh that is used to represent the approximate velocity c_ϵ.

Let then $\mathcal{T}_h$ be a regular mesh of Ω and its associated conforming discrete space $V_h \subset H^1(\Omega, \mathbb{C})$. Since Theorem 13 indicates that $\mathcal{S}_{\omega,c} f \in H^2(\Omega, \mathbb{C})$, we may expect a linear convergence in the $H^1(\Omega, \mathbb{C})$ norm when using linear or bilinear elements.

We now tackle the issue of computing the entries of the linear system associated with V_h. Indeed, even when using piecewise polynomials, we must integrate quantities involving c. In fact, if we assume that each interface $\Omega_r \cap \Omega_l$ is polygonal, we could accurately mesh it with a finer mesh $\mathcal{T}_\epsilon$ where ϵ has already been introduced with the approximate velocity c_ϵ. We could then perform an exact integration on $\mathcal{T}_\epsilon$. But this is not fully satisfactory since it requires to build an auxiliary mesh and we prefer to avoid any superfluous mesh with a view to reduce the implementation time. Furthermore, if we accept the idea of constructing an auxiliary mesh, the quadrature scheme induced by the fine mesh $\mathcal{T}_\epsilon$ is different in each coarse cell, making integration of linear system entries very costly. Finally, for realistic applications, the interfaces $\Omega_l \cap \Omega_r$ are not given explicitly and the parameter c is rather given as a set of sampling values. It seems thus difficult to introduce $\mathcal{T}_\epsilon$. We have to cope with a technical difficulty and for that purpose, we propose to construct an approximation c_ϵ of c such that the entries of the linear system are both cheap and easy to compute. This is what we are doing in the next subsection, but before, we focus on proving that the finite element scheme we apply is stable when c is replaced by its approximation. More precisely, we demonstrate that if c_ϵ converge to c when ϵ goes to zero (in a sense to be defined), the numerical solution converges to the analytical solution as both h and ϵ go to zero.

In the following, we will assume that $\omega h \leq 1$. It is worth mentioning that this hypothesis is not restrictive, since the final results are obtained under the stronger condition that $\omega^3 h^2$ is small enough.

We start by requiring approximation properties on the discretization space and we introduce the quantity $\mathcal{M}_{h,\epsilon}$ in Definition 6 and 7. Note that the conditions given in Definition 6 are fulfilled, for instance, by $\mathcal{P}_1$ Lagrangian polynomials.

Definition 6. *We consider a partition $\mathcal{T}_h$ of Ω. We assume that each cell $K \in \mathcal{T}_h$ is the image of a reference cell $\hat{K} \subset \mathbb{R}^2$ through an invertible affine map $\mathcal{F}_K \in \mathcal{L}(\mathbb{R}^2)$. We also consider a (finite dimensional) reference discretization space $\hat{P} \subset C^\infty(\hat{K})$, and define the discretization space V_h by*

$$V_h = \left\{ v_h \in H^1(\Omega, \mathbb{C}) \mid v_h|_K \circ \mathcal{F}_K \in \hat{P} \quad \forall K \in \mathcal{T}_h \right\}.$$

We further assume that there is a projection operator $\Pi_h \in \mathcal{L}(H^1(\Omega, \mathbb{C}), V_h)$ satisfying

$$|w - \Pi_h w|_{0,\Omega} \leq C h^2 |w|_{2,\Omega}, \quad |w - \Pi_h w|_{1,\Omega} \leq C h |w|_{2,\Omega}, \quad \forall w \in H^2(\Omega, \mathbb{C}),$$

where C is a constant depending on Ω, $\hat{K}$ and $\hat{P}$. Note that the multiplicative trace inequality ensures that

$$|w - \Pi_h w|_{0,\partial\Omega} \leq C h^{3/2} |w|_{2,\Omega},$$

where C is a constant depending on Ω, $\hat{K}$ and $\hat{P}$.

The construction of c_ϵ is depicted at Section 4. In this section, assume that $c_\epsilon \in L^\infty(\Omega)$ and $c_{min} \leq c_\epsilon \leq c_{max}$. We also define the quantity $\mathcal{M}_{h,\epsilon}$:

Definition 7. *The velocity approximation error is defined by*

$$\mathcal{M}_{h,\epsilon} = \max_{K\in\mathcal{T}_h} \frac{1}{|K|}\int_K \left|\frac{1}{c^2} - \frac{1}{c_\epsilon^2}\right|,$$

where $|K|$ is the Lebesgue measure of the cell K.

3.2.3 Convergence analysis

In the following, we assume that $\mathcal{M}_{h,\epsilon}$ converges to zero as h and ϵ go to zero. To simplify the notations we will consider a given $f \in L^2(\Omega,\mathbb{C})$ and define $u = \mathcal{S}_{\omega,c}f$. V_h and c_ϵ being defined, we now introduce the discrete finite element problem. We write $k_\epsilon = \omega/c_\epsilon$. The discrete equation consists in finding $u_h \in V_h$ such that

$$B_\epsilon(u_h, v_h) = -\int_\Omega k_\epsilon^2 u_h\bar{v}_h - ik_{max}\int_{\partial\Omega} u_h\bar{v}_h + \int_\Omega \nabla u_h\cdot\nabla\bar{v}_h = \int_\Omega f\bar{v}_h, \quad \forall v_h \in V_h. \quad (3.13)$$

In the remaining of this section $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ denotes a constant independent of ω, h and ϵ.

Proposition 9. *There exists a constant $C > 0$ such that*

$$|B(w,v)| \le C\big(\omega|w|_{0,\Omega} + |w|_{1,\Omega}\big)\big(\omega|v|_{0,\Omega} + |v|_{1,\Omega}\big), \quad \forall w, v \in H^1(\Omega,\mathbb{C}),$$

and

$$|B_\epsilon(w_h, v_h)| \le C\big(\omega|w_h|_{0,\Omega} + |w_h|_{1,\Omega}\big)\big(\omega|v_h|_{0,\Omega} + |v_h|_{1,\Omega}\big), \quad \forall w_h, v_h \in V_h.$$

Proof. Since the proofs are similar for B and B_ϵ, we focus on the first case only. Consider $w, v \in H^1(\Omega,\mathbb{C})$. It is obvious that

$$\begin{aligned} |B(w,v)| &\le k_{max}^2|w|_{0,\Omega}|v|_{0,\Omega} + k_{max}|w|_{0,\partial\Omega}|v|_{0,\partial\Omega} + |w|_{1,\Omega}|v|_{1,\Omega} \\ &\le \big(k_{max}|w|_{0,\Omega} + |w|_{1,\Omega}\big)\big(k_{max}|v|_{0,\Omega} + |v|_{1,\Omega}\big) + k_{max}|w|_{0,\partial\Omega}|v|_{0,\partial\Omega}. \end{aligned}$$

Moreover, for all $\mu \in H^1(\Omega,\mathcal{C})$, we have

$$\begin{aligned} k_{max}|\mu|_{0,\partial\Omega}^2 &\le C(\Omega)k_{max}\big(|\mu|_{0,\Omega}^2 + |\mu|_{0,\Omega}|\mu|_{1,\Omega}\big) \\ &\le C(\Omega)k_{max}\big(|\mu|_{0,\Omega}^2 + k_{max}|\mu|_{0,\Omega}^2 + \frac{1}{k_{max}}|\mu|_{1,\Omega}^2\big) \\ &\le C(\Omega,\omega_0,c_{min})\big(k_{max}^2|\mu|_{0,\Omega}^2 + |\mu|_{1,\Omega}^2\big) \\ &\le C(\Omega,\omega_0,c_{min})\big(k_{max}|\mu|_{0,\Omega} + |\mu|_{1,\Omega}\big)^2, \end{aligned}$$

and the result follows since $k_{max} = \omega/c_{min}$. □

We now give a result concerning the error induced by the approximation of the velocity parameter between the two sesquilinear forms B and B_h in Proposition 10.

Proposition 10. *There exists a constant $C := C(\hat{K}, \hat{P})$ such that*

$$|B(u_h, v_h) - B_\epsilon(u_h, v_h)| \leq C\omega^2 \mathcal{M}_{h,\epsilon} |u_h|_{0,\Omega} |v_h|_{0,\Omega}, \quad \forall u_h, v_h \in V_h.$$

Proof. Consider $u_h, v_h \in V_h$. We have

$$\begin{aligned} |B(u_h, v_h) - B_\epsilon(u_h, v_h)| &= \left| \int_\Omega (k^2 - k_\epsilon^2) u_h v_h \right| \\ &\leq \omega^2 \sum_{K \in \mathcal{T}_h} \int_K \left| \frac{1}{c^2} - \frac{1}{c_\epsilon^2} \right| |u_h| |v_h| \\ &\leq \omega^2 \sum_{K \in \mathcal{T}_h} |u_h|_{0,\infty,K} |v_h|_{0,\infty,K} \int_K \left| \frac{1}{c^2} - \frac{1}{c_\epsilon^2} \right|. \end{aligned} \tag{3.14}$$

Furthermore, for any cell $K \in \mathcal{T}_h$, $w_h \circ \mathcal{F}_K$ belongs to the finite dimensional space $\hat{P}$ if $w_h \in V_h$ and there exists a constant $\hat{C}$ depending on $\hat{P}$ only, such that

$$|w_h|_{0,\infty,K} = |w_h \circ \mathcal{F}_K|_{0,\infty,\hat{K}} \leq \hat{C} |w_h \circ \mathcal{F}_K|_{0,\hat{K}}.$$

We can thus derive

$$|w_h \circ \mathcal{F}_K|_{0,\hat{K}}^2 = \int_{\hat{K}} |w_h \circ \mathcal{F}_K|^2 = \text{Det } J_{\mathcal{F}_K}^{-1} \int_K |w_h|^2 = \frac{|\hat{K}|}{|K|} |w_h|_{0,K}^2,$$

so that

$$|w_h|_{0,\infty,K} \leq \hat{C} \sqrt{\frac{|\hat{K}|}{|K|}} |w_h|_{0,K}. \tag{3.15}$$

We can conclude by using (3.15) with $w_h = u_h, v_h$ in (3.14).

$$\begin{aligned} |B(u_h, v_h) - B_\epsilon(u_h, v_h)| &\leq \hat{C}^2 |\hat{K}| \omega^2 \sum_{K \in \mathcal{T}_h} \frac{|u_h|_{0,K} |v_h|_{0,K}}{|K|} \int_K \left| \frac{1}{c^2} - \frac{1}{c_\epsilon^2} \right| \\ &\leq \hat{C}^2 |\hat{K}| \omega^2 \mathcal{M}_{h,\epsilon} \sum_{K \in \mathcal{T}_h} |u_h|_{0,K} |v_h|_{0,K} \\ &\leq \hat{C}^2 |\hat{K}| \omega^2 \mathcal{M}_{h,\epsilon} |u_h|_{0,\Omega} |v_h|_{0,\Omega}. \end{aligned}$$

□

Before we establish our convergence result, we need three additional Lemma. In Lemma 19, we define the Riesz representation of the error z together with its elliptic projection z_h. We use the Riesz representation and its elliptic projection in Lemma 20 to bound the finite element error in the L^2 norm. Lemma 21 is a technical result required to prove the convergence in the H^1 norm in Theorem 14.

The proof of our error-estimate is based on the theory of Zhu and Wu [106, 109] who establishes in particular Proposition 11.

Proposition 11. *Let a be the sesquilinear form*

$$a(w,v) = \int_\Omega \nabla w \cdot \nabla \bar{v} - ik_{max} \int_{\partial\Omega} w\bar{v}, \quad \forall w, v \in H^1(\Omega, \mathbb{C}).$$

For all $z \in H^1(\Omega, \mathbb{C})$, there exists a unique $z_h \in V_h$ such that

$$a(w_h, z_h) = a(w_h, z), \quad \forall w_h \in V_h,$$

and we have

$$\begin{aligned} |z - z_h|_{0,\Omega} &\leq Ch^2 |z|_{2,\Omega}, \\ |z - z_h|_{1,\Omega} &\leq Ch |z|_{2,\Omega}, \\ |z - z_h|_{0,\partial\Omega} &\leq Ch^{3/2} |z|_{2,\Omega}. \end{aligned}$$

Lemma 19. *Let $u_h \in V_h$. Then there exists a unique element $z \in H^1(\Omega, \mathbb{C})$ such that*

$$|u - u_h|^2 = B(u - u_h, z). \tag{3.16}$$

Furthermore, there exists an element $z_h \in V_h$ such that

$$\frac{|B(u - u_h, z - z_h)|}{|u - u_h|_{0,\Omega}} \leq C \left(\omega^3 h^2 |u - u_h|_{0,\Omega} + \omega^2 h^2 |f|_{0,\Omega} \right). \tag{3.17}$$

Proof. According to Corollary 3, it is clear that there exists a unique $z \in H^1(\Omega, \mathbb{C})$ such that

$$B(w,z) = \int_\Omega w \overline{(u - u_h)}, \quad \forall w \in H^1(\Omega, \mathbb{C}).$$

In particular, picking $w = u - u_h$ yields (3.16).

Using Proposition 11, there exists an element $z_h \in V_h$ such that

$$a(u - u_h, z - z_h) = a(u - \Pi_h u, z - z_h)$$

It follows that

$$\begin{aligned} B(u - u_h, z - z_h) &= -\int_\Omega k^2 (u - u_h) \overline{(z - z_h)} + a(u - u_h, z - z_h) \\ &= -\int_\Omega k^2 (u - u_h) \overline{(z - z_h)} + a(u - \Pi_h u, z - z_h) \end{aligned}$$

Hence,

$$\begin{aligned} |B(u - u_h, z - z_h)| &\leq k_{max}^2 |u - u_h|_{0,\Omega} |z - z_h|_{0,\Omega} + k_{max} |u - \Pi_h u|_{0,\partial\Omega} |z - z_h|_{0,\partial\Omega} \\ &+ |u - \Pi_h u|_{1,\Omega} |z - z_h|_{1,\Omega} \\ &\leq C \left(k_{max}^2 h^2 |u - u_h|_{0,\Omega} |z|_{2,\Omega} + k_{max} h^3 |u|_{2,\Omega} |z|_{2,\Omega} + h^2 |u|_{2,\Omega} |z|_{2,\Omega} \right) \\ &\leq C \left(\omega^2 h^2 |u - u_h|_{0,\Omega} |z|_{2,\Omega} + \omega h^3 |u|_{2,\Omega} |z|_{2,\Omega} + h^2 |u|_{2,\Omega} |z|_{2,\Omega} \right) \end{aligned}$$

Now, using Corollary 3 again, we have

$$|z|_{2,\Omega} \leq C\omega |u - u_h|_{0,\Omega},$$

and therefore

$$\frac{|B(u-u_h, z-z_h)|}{|u-u_h|_{0,\Omega}} \leq C\left(\omega^3 h^2 |u-u_h|_{0,\Omega} + \omega^2 h^3 |u|_{2,\Omega} + \omega h^2 |u|_{2,\Omega}\right).$$

We conclude thanks to Theorem 13. We have

$$|u|_{2,\Omega} \leq C\omega |f|_{0,\Omega},$$

and the proof follows since $\omega h \leq 1$. □

Lemma 20. *Let $u_h \in V_h$ be any solution to problem (3.13). Then if $\omega^3 h^2$ and $\omega \mathcal{M}_{h,\epsilon}$ are small enough, there exists a constant C such that*

$$|u - u_h|_{0,\Omega} \leq C\left(\omega^2 h^2 + \mathcal{M}_{h,\epsilon}\right) |f|_{0,\Omega}.$$

Proof. Recalling (3.16) from Lemma 19, there exits an element $z \in H^1(\Omega, \mathbb{C})$ such that

$$|u - u_h|^2_{0,\Omega} = B(u - u_h, z).$$

We then introduce $z_h \in V_h$ defined as in Lemma 19. Since u and u_h solve (3.3) and (3.13) respectively, we have

$$\begin{aligned} B(u-u_h, z) &= B(u-u_h, z-z_h) + B(u-u_h, z_h) \\ &= B(u-u_h, z-z_h) + B_\epsilon(u_h, z_h) - B(u_h, z_h), \end{aligned}$$

and therefore

$$|u-u_h|_{0,\Omega} \leq \frac{|B(u-u_h, z-z_h)|}{|u-u_h|_{0,\Omega}} + \frac{|B_\epsilon(u_h, z_h) - B(u_h, z_h)|}{|u-u_h|_{0,\Omega}}. \tag{3.18}$$

We bound the first term in the right hand side of (3.18) using Lemma 19. To deal with the second term, we recall Proposition 10: there holds

$$|B_\epsilon(u_h, z_h) - B(u_h, z_h)| \leq C\omega^2 \mathcal{M}_{h,\epsilon} |u_h|_{0,\Omega} |z_h|_{0,\Omega},$$

but we have

$$\begin{aligned} |z_h|_{0,\Omega} &\leq |z|_{0,\Omega} + |z - z_h|_{0,\Omega} \\ &\leq C\left(\omega^{-1} |u-u_h|_{0,\Omega} + h^2 |z|_{2,\Omega}\right) \\ &\leq C\left(\omega^{-1} |u-u_h|_{0,\Omega} + \omega h^2 |u-u_h|_{0,\Omega}\right) \\ &\leq C\omega^{-1}(1 + \omega^2 h^2) |u-u_h|_{0,\Omega} \\ &\leq C\omega^{-1} |u-u_h|_{0,\Omega}, \end{aligned}$$

and

$$\begin{aligned} |u_h|_{0,\Omega} &\leq |u|_{0,\Omega} + |u-u_h|_{0,\Omega} \\ &\leq C\omega^{-1}|f|_{0,\Omega} + |u-u_h|_{0,\Omega}, \end{aligned}$$

so that

$$\frac{|B_\epsilon(u_h,z_h) - B(u_h,z_h)|}{|u-u_h|_{0,\Omega}} \leq C\left(\mathcal{M}_{h,\epsilon}|f|_{0,\Omega} + \omega\mathcal{M}_{h,\epsilon}|u-u_h|_{0,\Omega}\right).$$

Recalling (3.17) from Lemma 19, we obtain

$$|u-u_h|_{0,\Omega} \leq C\left(\omega^3h^2|u-u_h|_{0,\Omega} + \omega^2h^2|f|_{0,\Omega} + \mathcal{M}_{h,\epsilon}|f|_{0,\Omega} + \omega\mathcal{M}_{h,\epsilon}|u-u_h|_{0,\Omega}\right).$$

It follows that

$$\left(1 - C\omega^3h^2 - C\omega\mathcal{M}_{h,\epsilon}\right)|u-u_h|_{0,\Omega} \leq C\left(\omega^2h^2 + \mathcal{M}_{h,\epsilon}\right)|f|_{0,\Omega},$$

and we get Lemma 20 by assuming that ω^3h^2 and $\omega\mathcal{M}_{h,\epsilon}$ are small enough. □

Lemma 21. *The following estimate holds*

$$|u_h - \Pi_h u|^2_{1,\Omega} \leq C\big(\omega^2|u-u_h|^2_{0,\Omega} + (\mathcal{M}^2_{h,\epsilon} + \omega^2h^2)|f|^2_{0,\Omega}\big),$$

where u is the solution to (3.3) *and u_h is any solution to* (3.13).

Proof. First, the following relation holds

$$|u_h - \Pi_h u|^2_{1,\Omega} = \text{Re } B_\epsilon(u_h - \Pi_h u, u_h - \Pi_h u) + \int_\Omega k_\epsilon^2 |u_h - \Pi_h u|^2_{0,\Omega}. \tag{3.19}$$

Developing the first term of the right-hand-side in the above equation leads to:

$$\begin{aligned} B_\epsilon(u_h - \Pi_h u, u_h - \Pi_h u) &= B_\epsilon(u_h, u_h - \Pi_h u) - B_\epsilon(\Pi_h u, u_h - \Pi_h u) \\ &= B(u, u_h - \Pi_h u) - B_\epsilon(\Pi_h u, u_h - \Pi_h u) \\ &= B(u - \Pi_h u, u_h - \Pi_h u) + B(\Pi_h u, u_h - \Pi_h u) - B_\epsilon(\Pi_h u, u_h - \Pi_h u) \end{aligned}$$

It follows that

$$\begin{aligned} &\text{Re } B_\epsilon(u_h - \Pi_h u, u_h - \Pi_h u) \\ &\qquad \leq |B(u - \Pi_h u, u_h - \Pi_h u)| + |B(\Pi_h u, u_h - \Pi_h u) - B_\epsilon(\Pi_h u, u_h - \Pi_h u)|. \end{aligned} \tag{3.20}$$

Then, using Proposition 9 and Theorem 13, we have:

$$\begin{aligned} |B(u - \Pi_h u, u_h - \Pi_h u)| &\leq C\big(\omega|u - \Pi_h u|_{0,\Omega} + |u - \Pi_h u|_{1,\Omega}\big)\big(\omega|u_h - \Pi_h u|_{0,\Omega} + |u_h - \Pi_h u|_{1,\Omega}\big) \\ &\leq \frac{1}{2}|u_h - \Pi_h u|^2_{1,\Omega} + C\left(\omega^2|u_h - \Pi_h u|^2_{0,\Omega} + \omega^2|u - \Pi_h u|^2_{0,\Omega} + |u - \Pi_h u|^2_{1,\Omega}\right) \\ &\leq \frac{1}{2}|u_h - \Pi_h u|^2_{1,\Omega} + C\left(\omega^2|u_h - \Pi_h u|^2_{0,\Omega} + \omega^2h^4|u|^2_{2,\Omega} + h^2|u|^2_{2,\Omega}\right) \\ &\leq \frac{1}{2}|u_h - \Pi_h u|^2_{1,\Omega} + C\left(\omega^2|u_h - \Pi_h u|^2_{0,\Omega} + \omega^4h^4|f|^2_{0,\Omega} + \omega^2h^2|f|^2_{0,\Omega}\right) \\ &\leq \frac{1}{2}|u_h - \Pi_h u|^2_{1,\Omega} + C\left(\omega^2|u_h - \Pi_h u|^2_{0,\Omega} + \omega^2h^2|f|^2_{0,\Omega}\right), \end{aligned} \tag{3.21}$$

where we have used algebraic inequalities "$2ab \leq \eta a^2 + \eta^{-1} b^2$" in the second line and the hypothesis that $\omega h \leq 1$ in the last line.

Moreover, Proposition 10 implies that

$$\begin{aligned}
|B(\Pi_h u, u_h - \Pi_h u) - B_\epsilon(\Pi_h u, u_h - \Pi_h u)| &\leq C\omega^2 \mathcal{M}_{h,\epsilon} |\Pi_h u|_{0,\Omega} |u_h - \Pi_h u|_{0,\Omega} \\
&\leq C\omega^2 \big(\mathcal{M}_{h,\epsilon}^2 |\Pi_h u|_{0,\Omega}^2 + |u_h - \Pi_h u|_{0,\Omega}^2 \big) \\
&\leq C\omega^2 \big(\mathcal{M}_{h,\epsilon}^2 (|u|_{0,\Omega}^2 + |u - \Pi_h u|_{0,\Omega}^2) + |u_h - \Pi_h u|_{0,\Omega}^2 \big) \\
&\leq C\omega^2 \big(\mathcal{M}_{h,\epsilon}^2 (|u|_{0,\Omega}^2 + h^4 |u|_{2,\Omega}^2) + |u_h - \Pi_h u|_{0,\Omega}^2 \big) \\
&\leq C\omega^2 \big(\mathcal{M}_{h,\epsilon}^2 (\frac{1}{\omega^2} + \omega^2 h^4) |f|_{0,\Omega}^2 + |u_h - \Pi_h u|_{0,\Omega}^2 \big) \\
&\leq C \big(\mathcal{M}_{h,\epsilon}^2 (1 + \omega^4 h^4) |f|_{0,\Omega}^2 + \omega^2 |u_h - \Pi_h u|_{0,\Omega}^2 \big). \quad (3.22)
\end{aligned}$$

Now, since $k_\epsilon^2 \leq C\omega^2$, we have

$$|u_h - \Pi_h u|_{1,\Omega}^2 \leq \text{Re } B_\epsilon(u_h - \Pi_h u, u_h - \Pi_h u) + C\omega^2 |u_h - \Pi_h u|_{0,\Omega}^2$$

Plugging (3.21) and (3.22) in (3.20) implies that

$$\frac{1}{2} |u_h - \Pi_h u|_{1,\Omega}^2 \leq C \big\{ \omega^2 |u_h - \Pi_h u|_{0,\Omega}^2 + (\mathcal{M}_{h,\epsilon}^2 (1 + \omega^4 h^4) + \omega^2 h^2) |f|_{0,\Omega}^2 \big\}.$$

Then, since $\omega h < 1$, we end up with

$$|u_h - \Pi_h u|_{1,\Omega}^2 \leq C \big\{ \omega^2 |u_h - \Pi_h u|_{0,\Omega}^2 + (\mathcal{M}_{h,\epsilon}^2 + \omega^2 h^2) |f|_{0,\Omega}^2 \big\}.$$

We end the demonstration by observing that

$$\begin{aligned}
\omega^2 |u_h - \Pi_h u|_{0,\Omega}^2 &\leq \omega^2 |u - u_h|_{0,\Omega}^2 + \omega^2 |u - \Pi_h u|_{0,\Omega}^2 \\
&\leq \omega^2 |u - u_h|_{0,\Omega}^2 + C\omega^2 h^4 |u|_{2,\Omega}^2 \\
&\leq \omega^2 |u - u_h|_{0,\Omega}^2 + C\omega^4 h^4 |f|_{0,\Omega}^2.
\end{aligned}$$

□

We now establish a convergence result under the assumption that $\omega^3 h^2$ and $\omega \mathcal{M}_{h,\epsilon}$ can be made arbitrarily small.

Theorem 14. *Assume that $\omega^3 h^2$ and $\omega \mathcal{M}_{h,\epsilon}$ are small enough. Then problem (3.13) has a unique solution $u_h \in V_h$. Furthermore, u_h satisfies*

$$\omega |u - u_h|_{0,\Omega} + |u - u_h|_{1,\Omega} \leq C \left(\omega \mathcal{M}_{h,\epsilon} + \omega h + \omega^3 h^2 \right) |f|_{0,\Omega}, \quad (3.23)$$

where $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ denotes a constant independent of ω, h and ϵ.

Proof. Let us first show existence end uniqueness of u_h. Since V_h is a finite dimensional space, (3.13) is equivalent to a linear system with size $\left(\dim V_h \times \dim V_h\right)$. Therefore, we only need to prove uniqueness. Assume then that $f = 0$ in the discrete and continuous problem (3.3) and (3.13). According to Theorem 13, the corresponding continuous solution u is $u = 0$. Then, from Lemma 20, we deduce that

$$|u_h|_{0,\Omega} = |u - u_h|_{0,\Omega} \leq C\left(\omega^2 h^2 + \mathcal{M}_{h,\epsilon}\right)|f|_{0,\Omega} = 0,$$

so that $u_h = 0$ and uniqueness occurs.

We now turn to the proof of error-estimate (3.23). Recalling Lemma 20, it is clear that

$$\omega|u - u_h|_{0,\Omega} \leq C\left(\omega^3 h^2 + \omega\mathcal{M}_{h,\epsilon}\right)|f|_{0,\Omega},$$

and it remains to show that

$$|u - u_h|_{1,\Omega} \leq C\left(\omega\mathcal{M}_{h,\epsilon} + \omega h + \omega^3 h^2\right)|f|_{0,\Omega}.$$

To start with, it is clear that

$$\begin{aligned} |u - u_h|_{1,\Omega} &\leq |u - \Pi_h u|_{1,\Omega} + |u_h - \Pi_h u|_{1,\Omega} \\ &\leq Ch|u|_{2,\Omega} + |u_h - \Pi_h u|_{1,\Omega} \\ &\leq C\omega h|f|_{0,\Omega} + |u_h - \Pi_h u|_{1,\Omega}, \end{aligned}$$

but recalling Lemma 21, we have

$$\begin{aligned} |u_h - \Pi_h u|^2_{1,\Omega} &\leq C\left(\omega^2|u - u_h|^2_{0,\Omega} + (\mathcal{M}^2_{h,\epsilon} + \omega^2 h^2)|f|^2_{0,\Omega}\right) \\ &\leq C\left(\omega^6 h^4 + \omega^2\mathcal{M}^2_{h,\epsilon} + \mathcal{M}^2_{h,\epsilon} + \omega^2 h^2\right)|f|^2_{0,\Omega}, \end{aligned}$$

hence

$$|u_h - \Pi_h u|_{1,\Omega} \leq C\left(\omega^3 h^2 + \omega\mathcal{M}_{h,\epsilon} + \mathcal{M}_{h,\epsilon} + \omega h\right)|f|_{0,\Omega},$$

and the result follows since $\mathcal{M}_{h,\epsilon} \leq \omega_0^{-1}\omega\mathcal{M}_{h,\epsilon}$. □

3.2.4 Approximation of c

So far, we have studied the convergence of the method, assuming that the velocity parameter c is replaced by an approximation c_ϵ, where ϵ is a small parameter describing the convergence of c_ϵ to c. As indicated by Theorem 14, the convergence has to be understood in the sense $\mathcal{M}_{h,\epsilon} \to 0$ as $\epsilon \to 0$ (the quantity $\mathcal{M}_{h,\epsilon}$ is introduced in Definition 7).

In this section we discuss how to pick an approximation c_ϵ of c which is both accurate and easy to compute. For the sake of simplicity, we restrict our study to the case of flat interfaces. We also show that the entries of the linear system related to (3.13) are easy to compute.

The approximation of c is based upon the following procedure. Let $\mathcal{T}_\epsilon$ be a given partition of the reference cell $\hat{K}$. We can map this partition to each actual cell $K \in \mathcal{T}_h$ and

thus obtain a partition $\mathcal{T}_{h,\epsilon}^K$ of the cell K. Finally, gathering all the partitions associated to each cell $K \in \mathcal{T}_h$ together, we obtain a (possibly non-conforming) partition $\mathcal{T}_{h,\epsilon}$ of Ω (see Figure 3.3 which illustrates this process). The approximate velocity parameter is defined as follow:

Definition 8. *Let $c \in L^\infty(\Omega)$ be the global velocity supposed to satisfy assumption 3.2. Let $x_A \in A$ be the barycenter of $A \in \mathcal{T}_{h,\epsilon}$. If x_A does not belong to an interface, we set $c_\epsilon|_A = c(x_A)$, otherwise we define $c_\epsilon|_A = \sup_A c$.*

Our definition of c_ϵ corresponds to a P_0-interpolation of c. Recalling Definition 7, it is clear that other choices are possible and covered by our convergence analysis. However we consider P_0-interpolation only. Indeed, since we consider a piecewise constant parameter, it is not clear that higher order approximations might bring additional precision. Furthermore, difficulties can arise when defining high order approximation of c. For instance, it is shown in [72] that c_ϵ can take negative values if it is defined as a P_2-interpolation of c.

We now show that in the simple case of flat interfaces, the quantity $\mathcal{M}_{h,\epsilon}$ goes to zero as ϵ goes to zero uniformly with respect to h. Figure 3.3 is helpful to figure out different quantities used in the demonstration.

Proposition 12. *Assume that the interfaces of the partition (Ω_r) are flat and that the medium approximation submesh $\mathcal{T}_\epsilon$ is regular. Then there exists a constant C depending on the reference cell $|\hat{K}|$ only such that*

$$\mathcal{M}_{h,\epsilon} \leq CR\epsilon \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right|.$$

Proof. Consider a given cell $K \in \mathcal{T}_h$. Then, K is crossed by at-most R straight interfaces and, since the submesh $\mathcal{T}_{h,\epsilon}^K$ is regular, there exists a constant C such that the number of subcells $A \in \mathcal{T}_{h,\epsilon}^K$ crossed by each interface is less than C/ϵ. Then the total number of subcells of K crossed by an interface is less than CR/ϵ.

We can easily upper-bound the measure $|A|$ of each subcell $A \in \mathcal{T}_\epsilon$ like $|A| \leq C\epsilon^2$. Since the submesh $\mathcal{T}_{h,\epsilon}^K$ is constructed from a linear mapping, it follows that for all $A \in \mathcal{T}_{h,\epsilon}^K$

$$|A| \leq C \frac{|K|}{|\hat{K}|} \epsilon^2.$$

Let $\mathcal{A}_c \subset \mathcal{T}_{h,\epsilon}^K$ be the set of all subcells crossed by an interface. The total measure of the crossed subcells is then satisfying

$$\sum_{A \in \mathcal{A}_c} |A| \leq CR \frac{|K|}{|\hat{K}|} \epsilon.$$

Next, let $\mathcal{A}_e = \mathcal{T}_{h,\epsilon}^K \setminus \mathcal{A}_c$ be the set of subcells which are not crossed by any interface. Then, the approximation of c by c_ϵ is exact on each cell $A \in \mathcal{A}_e$. Therefore, we have

$$\int_K \left| \frac{1}{c^2} - \frac{1}{c_\epsilon^2} \right| \leq \sum_{A \in \mathcal{A}_c} \int_A \left| \frac{1}{c^2} - \frac{1}{c_\epsilon^2} \right| \leq \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right| \sum_{A \in \mathcal{A}_c} |A| \leq CR \frac{|K|}{|\hat{K}|} \epsilon \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right|.$$

But by definition of $\mathcal{M}_{h,\epsilon}$, we have

$$\mathcal{M}_{h,\epsilon} = \max_{K\in\mathcal{T}_{h,\epsilon}} \frac{1}{|K|}\int_K \left|\frac{1}{c^2} - \frac{1}{c_\epsilon^2}\right| \leq \frac{C}{|\hat{K}|} R\epsilon \left|\frac{1}{c_{min}^2} - \frac{1}{c_{max}^2}\right|.$$

which concludes the proof of Proposition 12. □

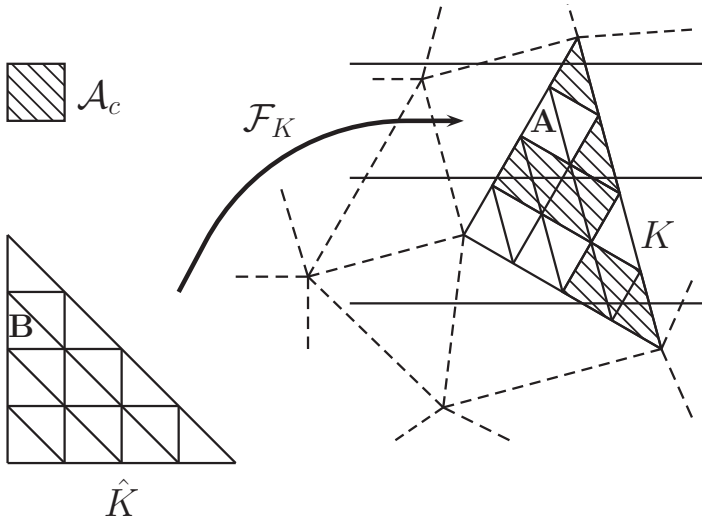


Figure 3.3: Mapping of the reference submesh

3.2.5 Computational cost

To end up with this section, we discuss on the computational cost of the proposed method. The corresponding linear system reads nearly as the one related to the classical FEM, except that the coefficients of the discrete system are weighted differently just because c_ϵ is different. Therefore, only the construction of the linear system is more expensive. To compute the entries of the linear system, we first compute reference integrals on each subcell $B \in \mathcal{T}_\epsilon$. This is done once and for all at the beginning of the simulation (or directly hard-coded, if the mesh $\mathcal{T}_\epsilon$ is known before execution) and it corresponds thus to a pre-processing step. Next, the mapping $\mathcal{F}_K$ is used to compute the coefficients associated with each cell K.

Let $\{\hat{\varphi}_i\}_{i=1}^D$ be a basis of $\hat{P}$. Note that if $\hat{P} = \mathcal{P}_k(\hat{K})$, $D = (p+1)(p+2)/2$. On each cell $K \in \mathcal{T}_h$, one has to compute

$$\int_K k_\epsilon^2 \hat{\varphi}_i \circ \mathcal{F}_K^{-1} \varphi_j \circ \mathcal{F}_K^{-1} = \sum_{A\in\mathcal{T}_{h,\epsilon}^K} k_\epsilon^2 \int_A \hat{\varphi}_i \circ \mathcal{F}_K^{-1} \varphi_j \circ \mathcal{F}_K^{-1} = \text{Det } J_{\mathcal{F}_K} \sum_{B\in\mathcal{T}_\epsilon} k_\epsilon^2 \int_B \hat{\varphi}_i \hat{\varphi}_j.$$

It should be noted that the last integral is independent of the given cell K. Therefore, we may compute the reference integrals

$$M_{ij}^B = \int_B \hat{\varphi}_i \hat{\varphi}_j, \quad \forall B \in \mathcal{T}_\epsilon.$$

once and for all independently of the number of coarse cells. The corresponding computational cost is thus insignificant. Now, for a given cell K, we have to compute

$$\text{Det } J_{\mathcal{F}_K} \sum_{B \in \mathcal{T}_h} k_\epsilon^2 M_{ij}^B.$$

If N_ϵ is the number of cell in $\mathcal{T}_\epsilon$, we thus need to perform N_ϵ multiplications, $N_\epsilon - 1$ additions, and one multiplication by the Jacobian, which comes to $2N_\epsilon$ operations for each coefficient. Now, arguing the symmetry of the system, we only need to compute $D(D+1)/2$ coefficients, which requires $N_\epsilon(D+1)$ operations per cell. Then, if we assume that the mesh $\mathcal{T}_\epsilon$ is regular, $N_\epsilon \leq C/\epsilon^2$ and the number of operations per cell is of $\mathcal{O}(D(D+1)/\epsilon^2)$ operations. Another way to think about it, is that if we are using N_ϵ subcells, the computational cost of the matrix assembly is multiplied by N_ϵ. Note that only the cost of the assembly is increased, since the linear system keeps the same size and stencil.

Chapter 4

Numerical examples

The aim of this chapter is to illustrate the MMAm approach presented in Chapters 2 and 3 on numerical examples. We first focus on analytical test-cases: Section 4.1 is devoted to one dimensional examples while we consider two dimensional layered media with plane wave solutions in Section 4.2.

Then, we numerically compare the MMAm with other methods in terms of performance and accuracy. In Section 4.3 we compare the MMAm with an homogenization procedure in a periodic layered medium. We analyse the behaviour of the MMAm on geophysical benchmarks in Section 4.4. A comparison with the standard FEm coupled with parameter averaging is included.

4.1 Analytical test-cases in 1D

4.1.1 Model problem

In this section, we consider a one dimensional Helmholtz problem set in the domain $\Omega = (0, 1000)$. We will consider different velocity parameters in each experiment. These velocity parameters are defined as piecewise constant functions $c : (0, 1000) \to \mathbb{R}$ which are bounded above and below as $1000 \leq c(z) \leq 5000$, for $z \in (0, 1000)$. The Helmholtz equation is coupled with absorbing boundary conditions at $z = 0, 1000$. We use an inhomogeneous boundary condition at $z = 0$ to represent an incoming wave travelling from the surface to depth. The homogeneous boundary condition at $z = 1000$ ensures that no wave enters the domain from depth.

Hence, the model problem reads

$$\left\{\begin{array}{rcll} -\dfrac{\omega^2}{c^2(z)}u(z) - u''(z) & = & 0, & z \in (0, 1000), \\ -u'(0) - \dfrac{\mathbf{i}\omega}{c(0)}u(0) & = & 1, & \\ u'(1000) - \dfrac{\mathbf{i}\omega}{c(1000)}u(1000) & = & 0, & \end{array}\right. \tag{4.1}$$

4.1.2 Analytical solution

An analytical solution for problem (4.1) is available. Since we consider piecewise constant parameter $c : (0, 1000) \to \mathbb{R}$ we can write c as

$$c = \sum_{l=1}^{L} c_l \mathbf{1}_{(z_{l-1}, z_l)}. \tag{4.2}$$

where $0 = z_0 \leq z_1 \leq \ldots z_L = 1000$, and the values $\{c_l\}_{l=1}^{L}$ are such that $c_{min} \leq c_l \leq c_{max}$.

It is therefore clear that in each layer (z_{l-1}, z_l) the solution u satisfies the following ODE:

$$-\frac{\omega^2}{c_l^2} u - u'' = 0, \tag{4.3}$$

and there exist two constants $A_l, B_l \in \mathbb{C}$ such that

$$u(z) = A_l \exp\left(\frac{\mathbf{i}\omega z}{c_l}\right) + B_l \exp\left(-\frac{\mathbf{i}\omega z}{c_l}\right), \tag{4.4}$$

for all $z \in (z_{l-1}, z_l)$.

We can obtain the analytical expression of u by retrieving the constant A_l, B_l for $1 \leq l \leq L$. This is done by solving the $2L \times 2L$ linear system define from the equations given by the boundary conditions of problem (4.1) and the C^1 compatibility conditions at the interface between each two consecutive layers. These conditions reads:

- Boundary condition at $z = 0$:
$$A_0 = 1.$$

- Boundary condition at $z = 1000$:
$$B_L = 0.$$

- C^0 compatibility condition at z_l $(0 < l < L)$:
$$A_l \exp\left(\frac{\mathbf{i}\omega z_l}{c_l}\right) + B_l \exp\left(-\frac{\mathbf{i}\omega z_l}{c_l}\right) = A_{l+1} \exp\left(\frac{\mathbf{i}\omega z_l}{c_{l+1}}\right) + B_{l+1} \exp\left(-\frac{\mathbf{i}\omega z_l}{c_{l+1}}\right).$$

- C^0 compatibility condition of the derivative at z_l $(0 < l < L)$:
$$A_l \frac{\mathbf{i}\omega}{z_l} c_l \exp\left(\frac{\mathbf{i}\omega z_l}{c_l}\right) - B_l \frac{\mathbf{i}\omega}{z_l} c_l \exp\left(-\frac{\mathbf{i}\omega z_l}{c_l}\right) = A_{l+1} \frac{\mathbf{i}\omega}{z_l} c_{l+1} \exp\left(\frac{\mathbf{i}\omega z_l}{c_{l+1}}\right) - B_{l+1} \frac{\mathbf{i}\omega}{z_l} c_{l+1} \exp\left(-\frac{\mathbf{i}\omega z_l}{c_{l+1}}\right).$$

4.1.3 Numerical experiments

We consider six different velocity models represented by $c^{(i)}$ for $i \in \{1, \dots, 6\}$. We point out that the length ν^i of the thinest layer of the model is decreasing from one experiment to the other. The medium is thus more complicated to handle from one experiment to another which justifies the use of the MMAm.

For $i \in \{2, \dots, 6\}$, we have constructed the velocity model $c^{(i)}$ so that the length of each layer is different and can not be fitted by a regular mesh. It ensures that we do not accidentally obtain an exact approximation of the velocity parameter when using the MMAm.

Let m be the number of subdivisions that are used inside a cell. We then introduce $\epsilon = 1/m$ as the small parameter representing the second scale used to approximate the medium. We thus have c_ϵ as approximated velocity use in the MMAm.

For each experiment, we tabulate the relative L^2 error $|u - u_h|/|u|$ for different frequencies ω, mesh steps $h = 1/n$ and multiscale subdivisions m. We also note *ndf* the number of degrees of freedom in the finite element space.

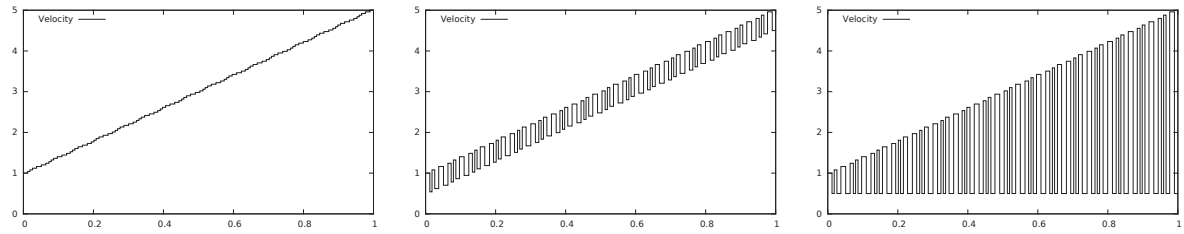


Figure 4.1: Velocity parameters used in experiments 2, 3 and 4

Experiment 1: Two-layered gradient

In experiment 1, we focus on a simple two-layered medium:

$$c^{(1)}(z) = \begin{cases} 1 & \text{if } z < 500 \\ 2 & \text{if } z > 500. \end{cases}$$

We consider (4.1) with the parameter $c^{(1)}$ and solve it with three different methodologies. First, we use a mesh with an even number of cells which fits the interface of the velocity parameter. Then we use a mesh with an odd number of cells, so that there is a velocity contrast in the middle cell of the mesh. With the non-conforming mesh, we first run without multiscale medium approximation: $c_\epsilon^{(1)}$ is taken constant in each cell ($m = 1$), and the medium is approximated in the middle cell. Then, we use two subcells per cell ($m = 2$) to approximate the medium, so that the medium is perfectly represented. We refer the reader to Figures 4.2-4.4.

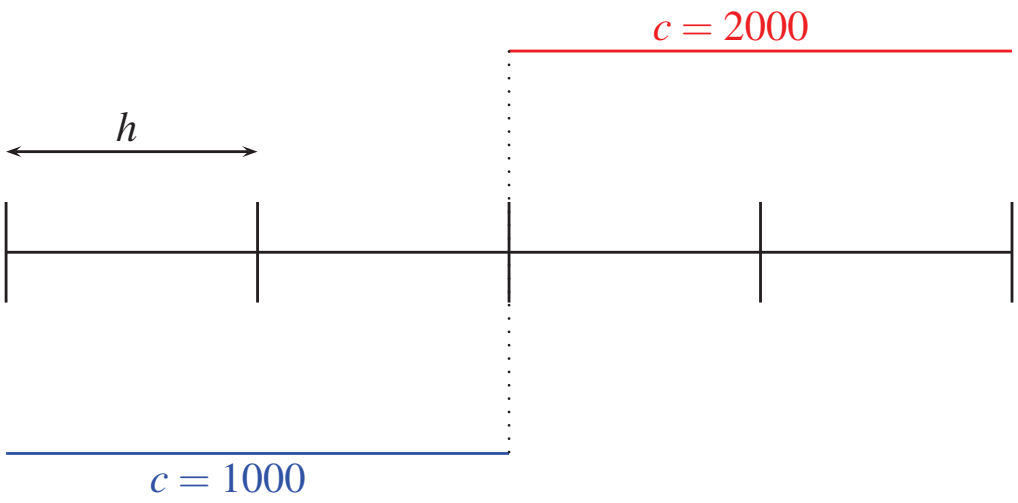


Figure 4.2: Fitting mesh with an even number of cells

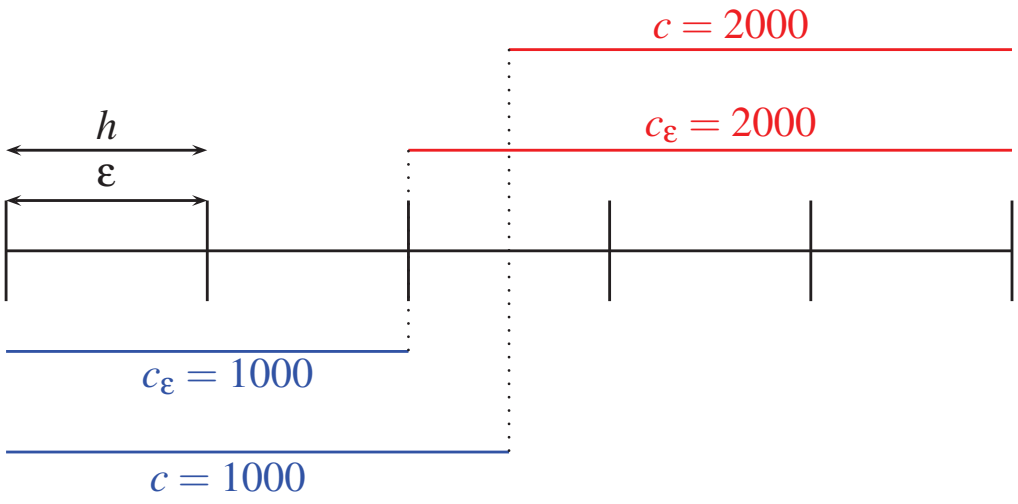


Figure 4.3: Non-fitting mesh with an odd number of cells

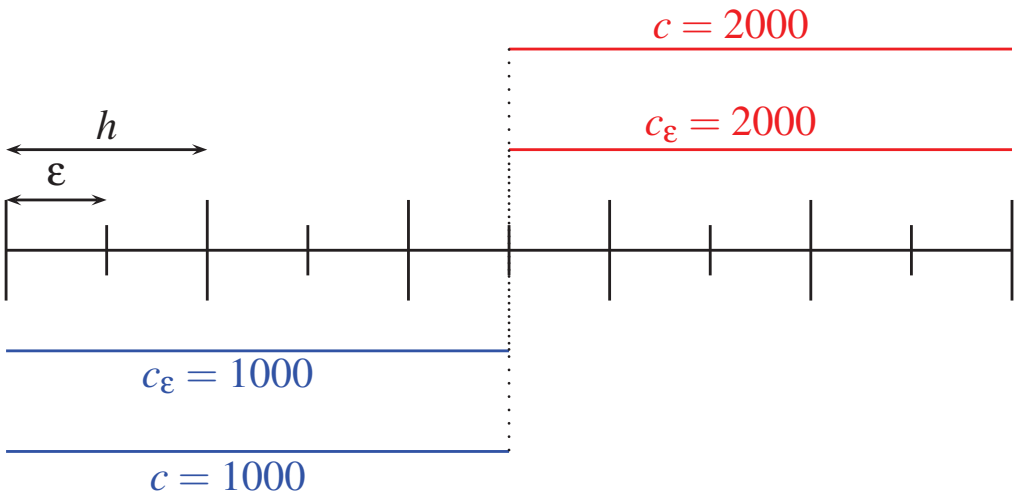


Figure 4.4: Non-fitting mesh with an odd number of cells and two subcells

We wish to understand if the condition $\omega^{2p+1}h^{2p} \leq C$ is sufficient to ensure that the finite element error remains bounded. We run our experiment with p ranging from 1 to 6 for different frequencies. For each p, we start with a low frequency ω_0 and select a mesh step h_0 so that the finite element relative error is less than 5% (this is done by trying different values of h_0). We then use the rule

$$\omega^{2p+1}h^{2p} = \omega_0^{2p+1}h_0^{2p} \tag{4.5}$$

to select the mesh steps h to solve for higher frequencies $\omega \geq \omega_0$.

In Table 4.1, we present the relative error on the numerical solution, for the three different techniques: "even" refers to the conforming mesh, "odd1" to the non-conforming with $m = 1$ and "odd2" to the non-conforming mesh with $m = 2$. We use the condition (4.5) to select the mesh steps.

It is clear that the meshing strategy $\omega^{2p+1}h^{2p} \leq C$ is optimal since the error remains constant when ω is increasing for all tables.

Besides, the tables show that for a given frequency, high order methods require less degrees of freedom for a given accuracy both when the fitting mesh is used, or when the MMAm is used with $m = 2$ on the non-fitting mesh.

Actually, we see that when the subquadrature technique is used, the results obtained with the non-fitting mesh are comparable to those obtained with the conforming one when the frequency is high enough. For the cases $p = 1, 2$ and 3, the results are similar for the fitting and non-fitting meshes for all frequencies. For the cases $p = 4, 5$ and $p = 6$, the MMAm solution is less precise than the solution obtained on the fitting mesh for the lowest frequency $\omega = 100$. However, the results are similar for higher frequencies. Finally, apart from the case of $p = 1$, the results with the non-conforming mesh are always improved when

using $m = 2$ subcells. It shows that for high order method, the medium approximation error can be larger than the best approximation error.

$\mathcal{P}_1$					
ω	n	ndf	even	odd1	odd2
100	1001	1000	1.97e-02	1.55e-02	1.98e-02
500	11181	11180	1.98e-02	1.15e-02	1.98e-02
1000	31623	31622	1.98e-02	1.33e-02	1.98e-02
2000	89443	89442	1.98e-02	1.49e-02	1.98e-02
5000	353554	353552	1.98e-02	1.66e-02	1.98e-02
$\mathcal{P}_2$					
ω	n	ndf	even	odd1	odd2
100	238	78	7.89e-02	2.84e-01	8.61e-02
500	1774	590	7.90e-02	1.71e-01	8.06e-02
1000	4216	1404	7.97e-02	1.36e-01	8.05e-02
2000	10030	3342	8.01e-02	1.07e-01	8.05e-02
5000	31534	10510	8.05e-02	7.72e-02	8.07e-02
$\mathcal{P}_3$					
ω	n	ndf	even	odd1	odd2
100	131	26	9.47e-02	3.40e-02	4.18e-02
500	881	176	1.91e-02	4.35e-02	3.92e-02
1000	1976	395	7.81e-02	4.79e-02	4.31e-02
2000	4436	887	3.06e-02	4.08e-02	4.47e-02
5000	12921	2584	6.09e-02	5.13e-02	4.65e-02

$\mathcal{P}_4$					
ω	n	ndf	even	odd1	odd2
100	176	24	5.03e-02	8.23e-01	8.39e-02
500	1086	154	4.11e-02	7.06e-01	4.93e-02
1000	2367	338	4.10e-02	6.61e-01	4.64e-02
2000	5167	738	4.24e-02	6.18e-01	4.61e-02
5000	14498	2070	4.41e-02	5.62e-01	4.64e-02
$\mathcal{P}_5$					
ω	n	ndf	even	odd1	odd2
100	172	18	3.27e-02	1.00e+00	6.34e-02
500	1045	116	1.66e-02	8.61e-01	2.15e-02
1000	2242	248	1.79e-02	8.26e-01	2.12e-02
2000	4807	534	1.81e-02	7.90e-01	2.03e-02
5000	13177	1464	1.91e-02	7.45e-01	2.04e-02
$\mathcal{P}_6$					
ω	n	ndf	even	odd1	odd2
100	177	16	6.00e-03	1.13e+00	3.32e-02
500	1024	92	9.70e-03	9.97e-01	1.71e-02
1000	2168	196	1.11e-02	9.54e-01	1.45e-02
2000	4599	418	1.23e-02	9.16e-01	1.26e-02
5000	12420	1128	1.49e-02	8.74e-01	1.16e-02

Table 4.1: Relative L^2 error in experiment 1

Experiment 2: 100 layered gradient

In this experiment, we consider the velocity parameter $c^{(2)}$ defined by

$$c^{(2)}(z) = \sum_{l=1}^{L} c_l^{(2)} \mathbf{1}_{(z_{l-1}, z_l)}, \tag{4.6}$$

where

$$z_0 = 0, \quad z_l = 1000 \frac{l + 0.4 \cos l}{L} \quad (0 < l < L), \quad z_L = 1000,$$

with $L = 100$, and

$$c_l^{(2)} = 1000 + 4000 \frac{l - 1}{99}.$$

The definition of the velocity parameter $c^{(2)}$ is motivated by three different aspects. First, we are using a cosine in the definition of the layers in order to make sure that the layers do not define a regular partition of $(0, 1000)$. That way, it is hard to integrate exactly the coefficients of the linear system if we use a regular non-fitting mesh and it makes sense to use the MMAm. Second, the size of each layer is in the order of 100 m, which is reasonable considering geophysical applications. Third, the values of the velocity are chosen to linearly increase from 1000 m.s^{-1} to 5000 m.s^{-1} with depth, which is also representative of geophysical applications.

We solve the problem with finite elements ranging from $p = 1$ to $p = 6$ together with the meshing condition $\omega^{2p+1}h^{2p} \leq C$ (more precisely (4.5)). The results are presented with different number of subcells $m = 1, 10$ and 100.

To start with, we can make the same comment than the previous experiment. The meshing strategy $\omega^{2p+1}h^{2p} \leq C$ is enough to ensure the precision of the method. Furthermore, for a given frequency, higher order methods give an equivalent result with less degrees of freedom.

In most cases, increasing the number of subcells improve the precision of the numerical scheme as expected. However, for $p = 3, 4, 5$ especially at high frequency, this is not the case. This can be explained by the fact that in those cases, the mesh is fine enough to capture the variations of the velocity, and improving the approximation of the medium does not improve the accuracy of the numerical solution. The error is then increasing a little because of numerical error due to finite precision arithmetic.

$\mathcal{P}_1$					
ω	n	ndf	$m = 1$	$m = 10$	$m = 100$
100	500	501	1.43e-02	1.63e-02	1.63e-02
500	5590	5591	1.52e-02	1.60e-02	1.63e-02
1000	15811	15812	1.93e-02	1.64e-02	1.64e-02
2000	44721	44722	1.62e-02	1.64e-02	1.65e-02
5000	176776	176777	1.63e-02	1.62e-02	1.63e-02
$\mathcal{P}_2$					
ω	n	ndf	$m = 1$	$m = 10$	$m = 100$
100	52	157	5.16e-02	4.87e-02	4.66e-02
500	394	1183	5.73e-02	4.79e-02	4.75e-02
1000	937	2812	7.10e-02	4.97e-02	4.83e-02
2000	2229	6688	9.29e-02	4.97e-02	4.89e-02
5000	7007	21022	8.92e-02	4.81e-02	4.95e-02
$\mathcal{P}_3$					
ω	n	ndf	$m = 1$	$m = 10$	$m = 100$
100	26	131	9.47e-02	3.40e-02	4.18e-02
500	176	881	1.91e-02	4.35e-02	3.92e-02
1000	395	1976	7.81e-02	4.79e-02	4.31e-02
2000	887	4436	3.06e-02	4.08e-02	4.47e-02
5000	2584	12921	6.09e-02	5.13e-02	4.65e-02

$\mathcal{P}_4$					
ω	n	ndf	$m = 1$	$m = 10$	$m = 100$
100	17	120	2.75e-01	3.49e-02	3.53e-02
500	108	757	1.44e-01	5.49e-02	4.05e-02
1000	237	1660	2.43e-01	4.57e-02	4.21e-02
2000	517	3620	2.64e-01	3.99e-02	4.11e-02
5000	1449	10144	1.20e-01	5.09e-02	4.54e-02
$\mathcal{P}_5$					
ω	n	ndf	$m = 1$	$m = 10$	$m = 100$
100	13	118	2.39e-01	5.02e-02	2.40e-02
500	77	694	8.29e-02	7.39e-02	3.50e-02
1000	166	1495	3.97e-01	5.64e-02	3.72e-02
2000	356	3205	1.49e-01	5.80e-02	4.33e-02
5000	976	8785	3.49e-01	4.54e-02	4.67e-02
$\mathcal{P}_6$					
ω	n	ndf	$m = 1$	$m = 10$	$m = 100$
100	10	111	1.21e-01	2.34e-02	2.87e-02
500	59	650	7.86e-02	3.66e-02	3.98e-02
1000	127	1398	1.78e-01	5.04e-02	3.86e-02
2000	269	2960	1.42e-01	4.57e-02	4.26e-02
5000	726	7987	8.35e-02	7.05e-02	4.93e-02

Table 4.2: Relative L^2 error in experiment 2

Experiment 3: 100 layered gradient with perturbations

The velocity parameter $c^{(2)}$ is representative of some applications in the sense that it is increasing with depth from 1000 m.s^{-1} to 5000 m.s^{-1}. However, since the velocity is increasing linearly, it does not feature high contrasts.

Thus, in experiment 3, we propose to perturbate the parameter $c^{(2)}$ every other layer

to create a more complex medium (see figure 4.1). The values

$$c_l^{(3)} = \begin{cases} 1000 + 4000\dfrac{l-1}{99}, & l \text{ odd} \\ 500 + 4000\dfrac{l-1}{99}, & l \text{ even} \end{cases}$$

define the velocity parameter

$$c^{(3)}(z) = \sum_{l=1}^{L} c_l^{(3)} \mathbf{1}_{(z_{l-1}, z_l)}$$

with $L = 100$ and

$$z_0 = 0, \quad z_l = 1000\frac{l + 0.4\cos l}{L} \quad (0 < l < L), \quad z_L = 1000,$$

The parameter $c^{(3)}$ has been designed so that the velocity contrast between two consecutive layers is approximately 500 m.s^{-1}. Numerical results are displayed in Table 4.3.

We use the meshing condition $\omega^{2p+1}h^{2p} \leq C$ and observe that for $1 \leq p \leq 6$, the error on the numerical solution remains bounded. However, we see than the error is not "constant", but varies from one frequency to another, more than in the previous experiments. We observe again than for a given frequency, higher order discretization require less degrees of freedom for a given accuracy. In particular, the case $\omega = 500$ is detailed on Table 4.4 and Figure 4.5. If one wishes to obtain a 5% accuracy, the best choice is $p = 4$.

$\mathcal{P}_1$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	1000	1001	2.84e-02	1.16e-02	1.29e-02
200	2828	2829	1.50e-02	9.19e-03	9.48e-03
500	11180	11181	1.03e-02	1.22e-02	1.24e-02
1000	31622	31623	7.09e-03	1.02e-02	1.03e-02
2000	89442	89443	4.40e-03	5.20e-03	5.26e-03
5000	353553	353554	1.23e-02	1.25e-02	1.25e-02
$\mathcal{P}_2$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	105	316	1.36e-01	3.13e-02	3.37e-02
200	250	751	5.32e-02	3.06e-02	3.60e-02
500	788	2365	7.77e-02	3.30e-02	4.27e-02
1000	1874	5623	7.77e-02	2.67e-02	2.55e-02
2000	4458	13375	4.16e-02	1.03e-02	1.25e-02
5000	14014	42043	2.70e-02	3.33e-02	2.83e-02
$\mathcal{P}_3$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	53	266	1.03e-01	2.27e-02	2.94e-02
200	120	601	1.47e-01	3.63e-02	4.50e-02
500	352	1761	1.25e-01	6.95e-02	5.16e-02
1000	790	3951	2.23e-01	3.89e-02	2.61e-02
2000	1774	8871	9.04e-02	1.05e-02	1.30e-02
5000	5168	25841	5.21e-02	2.61e-02	2.97e-02

$\mathcal{P}_4$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	44	309	1.04e-01	8.10e-02	5.71e-02
200	96	673	9.27e-02	1.77e-02	2.57e-02
500	271	1898	2.40e-01	3.18e-02	2.26e-02
1000	592	4145	2.40e-01	2.99e-02	8.64e-03
2000	1293	9052	5.56e-02	5.22e-03	3.95e-03
5000	3624	25369	1.92e-01	9.70e-03	9.22e-03
$\mathcal{P}_5$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	31	280	1.59e-01	9.84e-02	7.11e-02
200	67	604	3.05e-01	4.92e-02	3.33e-02
500	186	1675	1.19e-01	2.64e-02	2.73e-02
1000	399	3592	1.86e-01	2.63e-02	2.10e-02
2000	855	7696	9.52e-02	1.44e-02	1.45e-02
5000	2343	21088	6.52e-02	5.45e-03	1.34e-02
$\mathcal{P}_6$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	24	265	4.53e-01	5.12e-02	5.98e-02
200	51	562	1.95e-01	1.08e-02	2.21e-02
500	139	1530	4.29e-01	3.40e-02	4.93e-02
1000	296	3257	3.54e-01	2.40e-02	2.16e-02
2000	628	6909	7.96e-02	4.73e-02	2.05e-02
5000	1694	18635	2.11e-01	5.32e-02	1.73e-02

Table 4.3: Relative L^2 error in experiment 3

p	n	ndf	m	err
1	5500	5501	1000	4.96e-02
2	700	2101	1000	4.95e-02
3	370	1851	1000	4.90e-02
4	217	1520	1000	4.34e-02
5	180	1621	1000	4.94e-02
6	142	1563	1000	4.72e-02

Table 4.4: Comparison of different orders of discretization for Experiment 3

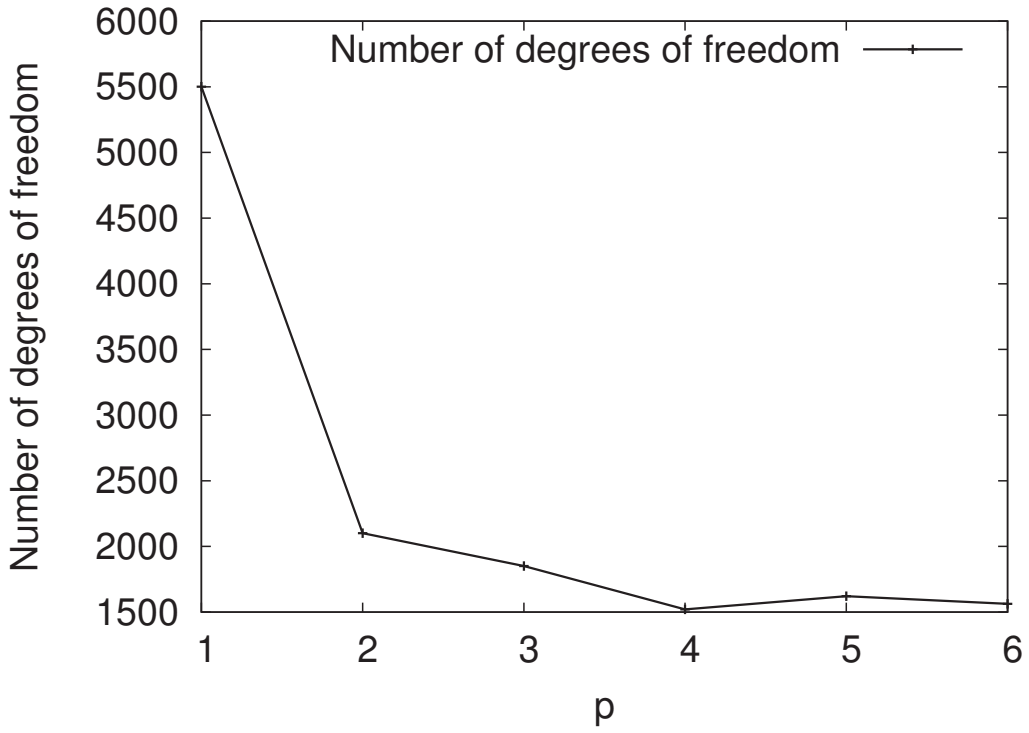


Figure 4.5: Number of degrees of freedom required for 5% accuracy in experiment 3

Experiment 4: 100 layered gradient with rough perturbations

We now further perturbate the velocity parameter $c^{(2)}$ in order to incorporate higher velocity contrasts in the model. The velocity parameter $c^{(4)}$ is defined by

$$c^{(4)}(z) = \sum_{l=1}^{L} c_l^{(4)} \mathbf{1}_{(z_{l-1}, z_l)}$$

where $L = 100$,

$$z_0 = 0, \quad z_l = 1000 \frac{l + 0.4 \cos l}{L} \quad (0 < l < L), \quad z_L = 1000,$$

and

$$c_l^{(4)} = \begin{cases} 1000 + 4000\dfrac{l-1}{99}, & l \text{ odd} \\ 500, & l \text{ even.} \end{cases}$$

The velocity parameter $c^{(5)}$ includes velocity contrasts ranging from 500 m.s^{-1} to over 4500 m.s^{-1}. Numerical results are presented in Table 4.5.

We see that the error is varying a lot depending on the frequency when $\omega^{2p+1}h^{2p}$ is kept constant. In particular, we observe an outstanding error peak for $\omega = 2000$ for all polynomial order on Table 4.5. This might indicate that the condition $\omega^{2p+1}h^{2p} \leq C$ is not sufficient to guarantee a constant error independently of the frequency when the medium is heterogeneous.

$\mathcal{P}_1$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	2000	2001	2.72e-03	1.85e-03	1.91e-03
200	5656	5657	1.92e-03	1.16e-03	1.53e-03
500	22360	22361	1.46e-02	8.59e-03	9.17e-03
1000	63245	63246	5.46e-02	3.51e-02	3.66e-02
2000	178885	178886	9.33e-02	8.28e-02	8.39e-02
5000	707106	707107	1.19e-03	1.21e-03	1.23e-03
$\mathcal{P}_2$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	316	949	1.39e-02	3.29e-03	8.49e-04
200	752	2257	1.91e-02	1.64e-03	8.46e-04
500	2364	7093	5.28e-02	1.34e-02	5.09e-03
1000	5623	16870	2.26e+00	1.68e-02	1.64e-02
2000	13374	40123	9.29e-02	5.77e-02	3.02e-02
5000	42044	126133	3.00e-03	8.17e-04	2.71e-04
$\mathcal{P}_3$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	107	536	5.26e-02	5.63e-03	4.00e-03
200	241	1206	1.65e-02	2.45e-03	3.98e-03
500	704	3521	6.81e-01	3.22e-02	1.51e-02
1000	1581	7906	1.14e+01	1.00e-01	6.82e-02
2000	3549	17746	3.37e-01	2.09e-01	9.57e-02
5000	10337	51686	2.36e-02	3.73e-03	8.81e-04

$\mathcal{P}_4$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	88	617	5.76e-02	1.15e-02	4.73e-03
200	193	1352	4.68e-02	6.95e-03	3.98e-03
500	543	3802	3.31e-01	1.72e-02	1.38e-02
1000	1185	8296	1.62e-01	1.28e-02	2.95e-02
2000	2586	18103	6.38e+00	1.00e-01	4.95e-02
5000	7249	50744	1.86e-02	4.50e-03	4.89e-04
$\mathcal{P}_5$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	63	568	1.61e-01	1.30e-02	5.14e-03
200	135	1216	1.51e-01	1.45e-02	7.97e-03
500	372	3349	2.54e-01	4.33e-02	2.03e-02
1000	798	7183	8.46e-01	7.32e-02	9.86e-02
2000	1710	15391	1.50e+00	2.66e-01	1.75e-01
5000	4687	42184	5.16e-02	3.93e-03	1.05e-03
$\mathcal{P}_6$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	48	529	1.17e-01	1.34e-02	8.92e-03
200	103	1134	6.94e-02	1.41e-02	4.84e-03
500	279	3070	3.74e-01	5.01e-02	2.07e-02
1000	592	6513	8.66e-01	2.06e-01	5.67e-02
2000	1256	13817	3.05e+00	3.31e-01	1.39e-01
5000	3389	37280	6.98e-02	4.47e-03	1.57e-03

Table 4.5: Relative L^2 error in experiment 4

If the condition $\omega^{2p+1}h^{2p} \leq C$ is not satisfactory here, we can still investigate which polynomial degree p is the best to obtain a given accuracy. In this regard, we consider the frequency $\omega = 500$ and compare the number of degrees of freedom required to achieve a 5% accuracy depending on p.

Based on the results presented in Table 4.6 and Figure 4.6, we claim that increasing the polynomial degree, at least up to $p = 6$, reduce the size of the linear system for a given accuracy.

p	n	ndf	m	err
1	10000	10001	10	4.67e-02
2	1200	3601	100	5.40e-02
3	610	3051	100	5.17e-02
4	375	2626	300	4.00e-02
5	275	2476	1000	4.61e-02
6	220	2421	1000	4.77e-02

Table 4.6: Comparison of different orders of discretization for Experiment 4

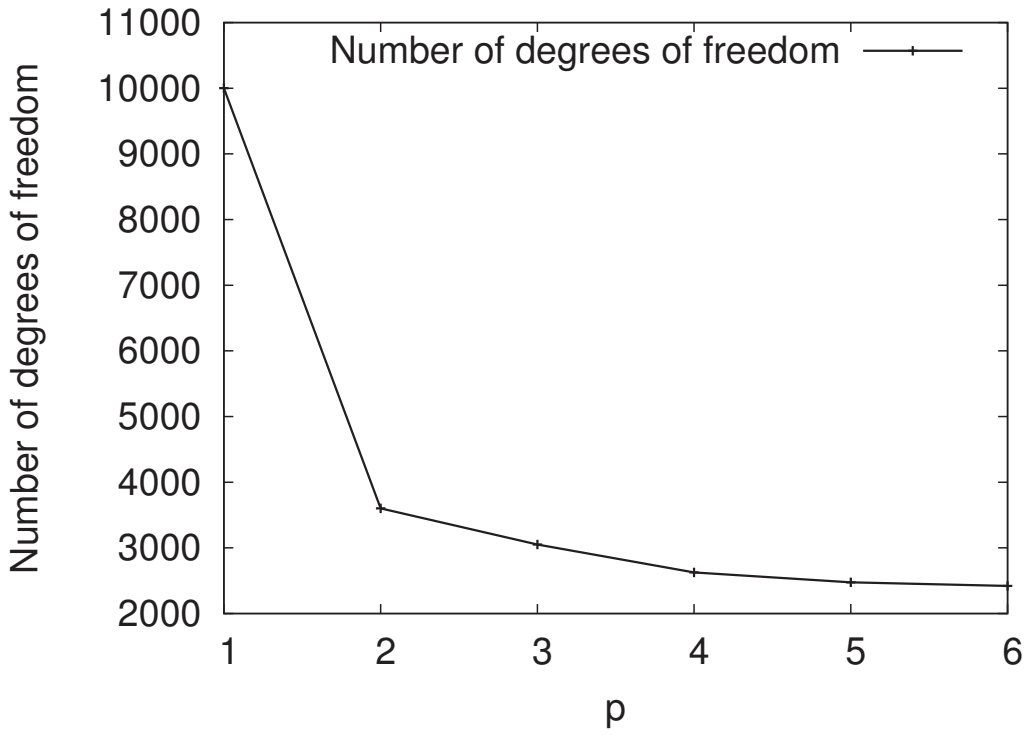


Figure 4.6: Number of degrees of freedom required for 5% accurracy in experiment 4

Experiment 5: 1000 layered gradient with perturbations

We consider here a velocity parameter $c^{(5)}$ similar to $c^{(3)}$, $c^{(5)}$ featuring 1000 layers. It is defined by

$$c^{(5)}(z) = \sum_{l=1}^{L} c_l^{(5)} \mathbf{1}_{(z_{l-1}, z_l)}$$

where $L = 1000$,

$$z_0 = 0, \quad z_l = 1000 \frac{l + 0.4 \cos l}{L} \quad (0 < l < L), \quad z_L = 1000,$$

and

$$c_l^{(5)} = \begin{cases} 1000 + 4000\dfrac{l-1}{999}, & l \text{ odd} \\ 500 + 4000\dfrac{l-1}{999}, & l\text{even}. \end{cases}$$

Like in experiment 4, we observe that the condition $\omega^{2p+1}h^{2p} \leq C$ is not satisfactory. Indeed, as shown in Table 4.7, the error level is varying a lot when ω changes.

$\mathcal{P}_1$					
ω	n	ndf	$m=10$	$m=100$	$m=1000$
100	1000	1001	1.97e-02	1.82e-02	1.25e-02
200	2828	2829	2.64e-03	1.29e-02	1.31e-02
500	11180	11181	2.28e-02	1.26e-02	1.52e-02
1000	31622	31623	5.19e-03	3.02e-03	2.95e-03
2000	89442	89443	1.33e-01	3.02e-02	3.61e-02
5000	353553	353554	1.51e-02	1.61e-02	1.69e-02
$\mathcal{P}_2$					
ω	n	ndf	$m=10$	$m=100$	$m=1000$
100	158	475	7.53e-02	5.67e-03	2.63e-03
200	376	1129	4.25e-02	3.19e-02	2.32e-02
500	1182	3547	3.07e-02	3.65e-02	3.06e-02
1000	2811	8434	5.65e-02	1.12e-02	6.03e-03
2000	6687	20062	2.06e-01	1.51e-01	7.45e-02
5000	21022	63067	4.02e-01	5.46e-02	1.91e-02
$\mathcal{P}_3$					
ω	n	ndf	$m=10$	$m=100$	$m=1000$
100	53	266	4.55e-01	1.15e-02	8.83e-03
200	120	601	1.10e-01	5.51e-02	3.24e-02
500	352	1761	4.08e-01	1.20e-01	1.99e-01
1000	790	3951	1.83e-01	3.49e-02	4.22e-02
2000	1774	8871	1.91e-01	1.04e+00	1.10e+00
5000	5168	25841	3.04e-01	1.55e-01	1.77e-01

$\mathcal{P}_4$					
ω	n	ndf	$m=10$	$m=100$	$m=1000$
100	44	309	9.87e-02	6.31e-02	6.74e-03
200	96	673	3.10e-01	3.84e-02	2.90e-02
500	271	1898	4.52e-01	1.96e-01	1.79e-01
1000	592	4145	2.48e-01	3.15e-02	3.38e-02
2000	1293	9052	3.70e-01	2.55e+00	5.34e-01
5000	3624	25369	6.67e-01	2.72e-01	1.34e-01
$\mathcal{P}_5$					
ω	n	ndf	$m=10$	$m=100$	$m=1000$
100	31	280	6.53e-01	4.98e-02	7.32e-03
200	67	604	6.58e-01	4.12e-02	2.83e-02
500	186	1675	7.73e-01	2.35e-01	2.29e-01
1000	399	3592	3.84e-01	7.75e-02	5.02e-02
2000	855	7696	9.76e-01	1.06e+00	3.06e+00
5000	2343	21088	5.77e-01	2.53e-01	2.43e-01
$\mathcal{P}_6$					
ω	n	ndf	$m=10$	$m=100$	$m=1000$
100	36	397	3.60e-01	2.95e-02	5.09e-03
200	77	848	1.08e-01	4.01e-02	2.48e-02
500	209	2300	4.30e-01	1.54e-01	1.33e-01
1000	444	4885	3.18e-01	2.91e-02	2.25e-02
2000	942	10363	2.54e+00	4.05e-01	3.10e-01
5000	2541	27952	2.63e-01	4.70e-02	7.06e-02

Table 4.7: Relative L^2 error in experiment 5

Nevertheless, though the meshing condition is not satisfactory, we are still able to show the advantage of higher order methods. To this end, we consider the problem of obtaining an error of 5% for a given frequency of $\omega = 500$. Table 4.8 and Figure 4.7 show that for higher discretization orders, the size of the linear system is smaller. Here, the best choice for experiment 5 seems to be $p = 5$.

p	n	ndf	m	err
1	5500	5501	1	4.01e-02
2	1000	3001	1	4.95e-02
3	550	2751	50	3.25e-02
4	370	2591	100	4.00e-02
5	285	2566	100	4.33e-02
6	270	2971	300	4.80e-02

Table 4.8: Comparison of different orders of discretization for Experiment 5

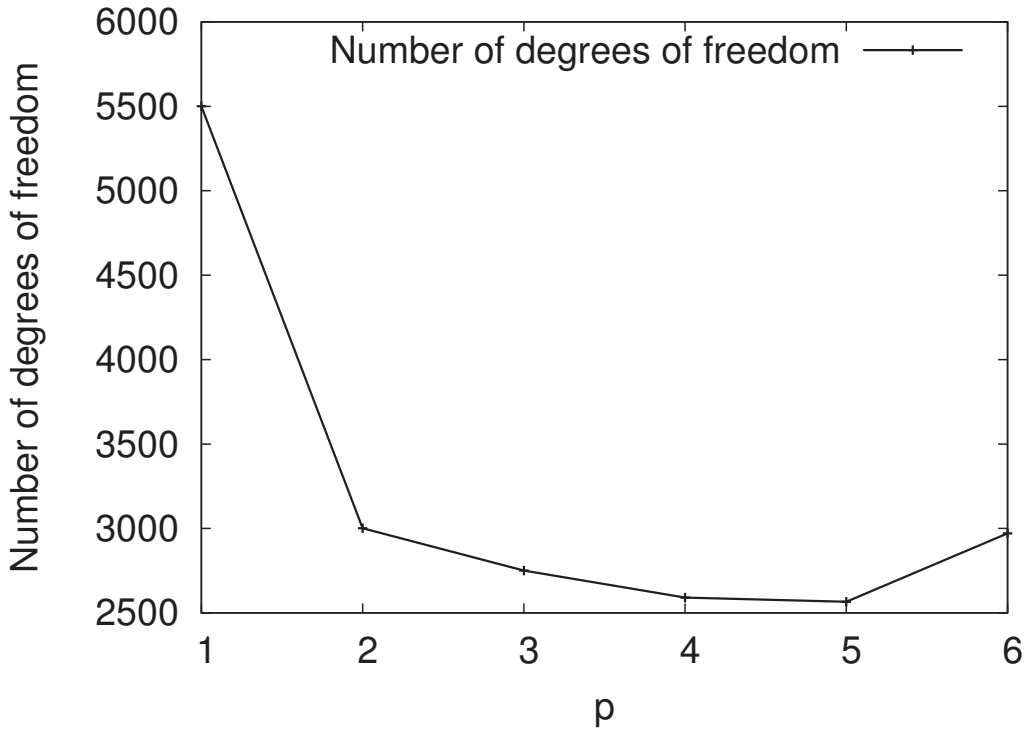


Figure 4.7: Number of degrees of freedom required for 5% accuracy in experiment 5

Experiment 6: 1000 layered gradient with rough perturbations

We consider a velocity parameter like $c^{(4)}$, but with 1000 layers. We thus combine strong velocity contrasts with a high number of value changes. The velocity parameter $c^{(6)}$ is defined by

$$c^{(6)}(z) = \sum_{l=1}^{L} c_l^{(6)} \mathbf{1}_{(z_{l-1}, z_l)}$$

where $L = 1000$,

$$z_0 = 0, \quad z_l = \frac{l + 0.4 \cos l}{L} \quad (0 < l < L), \quad z_L = 1,$$

and

$$c_l^{(6)} = \begin{cases} 1000 + 4000 \dfrac{l-1}{999}, & l \text{ odd} \\ 500, & l \text{ even.} \end{cases}$$

Like in experiments 4 and 5, we do not obtain a constant level of accuracy by using the rule $\omega^{2p+1} h^{2p} \leq C$. This is depicted in Table 4.9. This is the reason why we focus on comparing different orders of discretization for the frequency $\omega = 500$. The results are presented in Table 4.10 and Figure 4.8. They show that high order methods require a smaller linear system to achieve the same precision.

$\mathcal{P}_1$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	2000	2001	2.82e-02	1.93e-02	2.94e-02
200	5656	5657	3.81e-02	3.91e-02	3.90e-02
500	22360	22361	1.32e-02	6.61e-02	5.47e-02
1000	63245	63246	6.73e-04	2.58e-04	1.91e-04
2000	178885	178886	2.25e-03	2.62e-03	2.29e-03
5000	707106	707107	1.36e-02	5.50e-03	6.25e-03
$\mathcal{P}_2$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	316	949	1.05e-01	7.61e-02	6.97e-02
200	752	2257	1.39e-01	2.05e-01	1.36e-01
500	2364	7093	1.63e-01	6.48e-02	1.08e-01
1000	5623	16870	3.82e-03	6.51e-04	1.34e-04
2000	13374	40123	4.04e-02	2.94e-03	2.73e-03
5000	42044	126133	1.31e-01	7.66e-03	4.55e-03
$\mathcal{P}_3$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	430	2151	2.20e-01	2.80e-02	2.55e-02
200	967	4836	1.13e-01	3.47e-02	1.33e-02
500	2817	14086	1.07e-01	3.24e-02	1.11e-02
1000	6324	31621	7.70e-03	1.00e-03	1.03e-04
2000	14198	70991	6.47e-02	1.98e-03	6.13e-04
5000	41351	206756	1.51e-01	2.63e-02	1.74e-03

$\mathcal{P}_4$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	88	617	8.29e-01	1.37e-01	8.23e-02
200	193	1352	9.72e-01	5.26e-01	4.79e-01
500	543	3802	9.94e-01	4.33e-01	4.92e-01
1000	1185	8296	4.37e-02	5.57e-03	2.27e-03
2000	2586	18103	3.99e-01	2.29e-02	1.69e-02
5000	7249	50744	2.62e-01	5.38e-02	2.39e-02
$\mathcal{P}_5$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	63	568	1.76e+00	7.11e-02	4.96e-02
200	135	1216	1.02e+00	5.99e-01	5.38e-01
500	372	3349	1.44e+00	1.04e+00	7.41e-01
1000	798	7183	5.57e-02	1.13e-02	2.96e-03
2000	1710	15391	4.49e-01	2.63e-02	1.23e-02
5000	4687	42184	4.65e-01	1.66e-01	4.11e-02
$\mathcal{P}_6$					
ω	n	ndf	$m = 10$	$m = 100$	$m = 1000$
100	48	529	1.05e+00	1.54e-01	4.37e-02
200	103	1134	1.38e+00	4.80e-01	5.60e-01
500	279	3070	9.63e-01	8.71e-01	1.02e+00
1000	592	6513	6.36e-02	8.46e-03	4.75e-03
2000	1256	13817	1.51e+00	8.54e-02	3.21e-02
5000	3389	37280	7.15e-01	1.07e-01	6.29e-02

Table 4.9: Relative L^2 error in experiment 6

p	n	ndf	m	err
1	25000	25001	100	4.58e-02
2	3000	9001	200	2.83e-02
3	1700	8501	300	2.45e-02
4	1150	8051	300	2.18e-02
5	900	8101	1000	5.25e-02
6	700	7701	1000	5.32e-02

Table 4.10: Comparison of different orders of discretization for Experiment 6

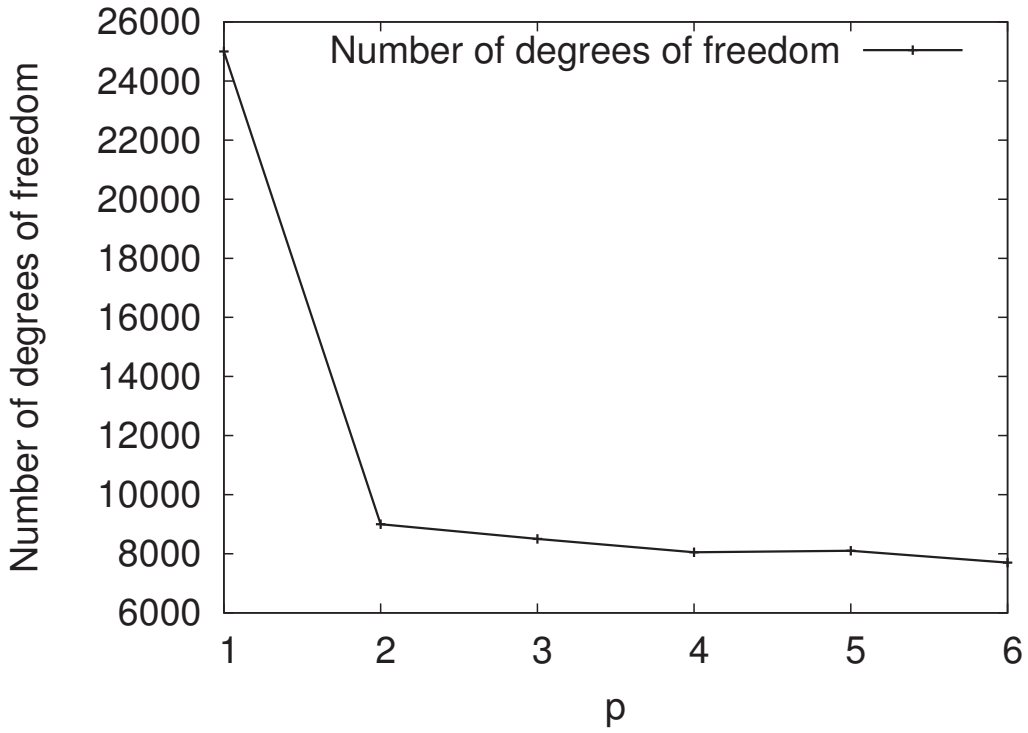


Figure 4.8: Number of degrees of freedom required for 5% accuracy in experiment 6

Conclusion

First, we have investigated explicit frequency requirements for meshing. The three first test-cases show when the medium is slowly varying, the pre-asymptotic error-estimate in $\mathcal{O}(\omega^{2p+1}h^{2p})$ known in the homogeneous case is still valid and optimal, provided that the medium is properly approximated (see Tables 4.1 to 4.3). However experiment 4 shows that this is no longer the case for more complex media (see Table 4.5).

The other aim of our study were to figure out which order of discretization is the cheapest for a given accuracy and frequency. In the three first experiments, we see that for a given accuracy and frequency, the number of degrees of freedom required decreases when the order is increasing, showing the efficiency of high order methods. For the other experiments, the situation is more complex and is described in Tables 4.6, 4.8 and 4.10. We have observed that for a given accuracy, the number of degrees of freedom required for $p = 5$ is always less than for $p = 1, 2, 3$ and 4.

We can make an additional comment which does not directly apply to the method in higher dimensions. We can apply static condensation on the degrees of freedom inside one cell (they only depend on the values at the vertices, see [60] or [105]) and reduce the size of the global linear system to $n \times n$. In this situation, high order methods look even more attractive, since the number of cells required for a given accuracy is clearly decreasing when the order is increasing.

Our experiments also confirms the interest of the MMAm second level approximation

strategy. Indeed, we have observed that the order of discretization that yields the smallest linear system is always $p \geq 4$. In this case, the mesh cells are rather large and the solution obtain by the standard FEm without MMAm subcells ($m = 1$) is not accurate. It is therefore of interest to use the MMAm with $m = 100$ or 1000 to obtain an accurate solution. Using an important number of subcells is not a problem, since it corresponds to a pre-processing step in the computations which can be easily parallelized. Furthermore, the linear system size and stencil remain the same for any choice of m.

4.2 Analytical test-cases in 2D

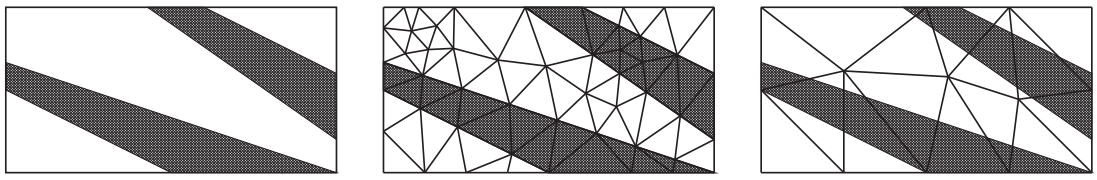
The objective of this section is to illustrate the MMAm approach on 2D analytical solutions. We base our analysis on artificial stratified media in which we have a plane wave analytical solution. In particular, we illustrate how the MMAm performs well even when the velocity is strongly varying and does not satisfy the technical assumption 3.2. The performance of the method is measured from the values of the $L^2(\Omega)$ norm relative error, that is

$$E = \frac{\int_\Omega |u - u_{h,\epsilon}|^2 dx}{\int_\Omega |u|^2 dx} \tag{4.7}$$

where u denotes the exact (analytical) solution and $u_{h,\epsilon}$ is the numerical solution. We recall that h stands for the size of the finite element mesh cells while ϵ denotes the size of the second-level subcells used to approximate the medium.

The numerical results are depicted by the mean of the solution profile, that is the graph of $x_2 \to u_{h,\epsilon}(500, x_2)$.

All along this section, we use two kinds of meshes as depicted in figure 4.9. Some are constructed so that the velocity is constant inside each cell. We then speak about fitting meshes in contrast to non-fitting meshes which are composed of cells inside which the velocity may vary. Obviously, the MMAm must be used on non-fitting meshes to take into account subcells velocity variations. Standard FEM, or other usual methods, are rather used on fitting meshes.



(a) (b) (c)

Figure 4.9: Velocity model (a), fitting (b) and non-fitting (c) meshes

Herein, we will also consider the standard FEM on non-fitting meshes. In this case, we transform the velocity parameter so that it is constant in each cell of the mesh. We

use two different strategies. The first idea is to select the value of the velocity parameter in the center of the cell. It corresponds to using the MMAm with only one subcell. The other strategy is to average the velocity parameter on the cell and choosing the value

$$\frac{1}{c_K^2} = \int_K \frac{1}{c^2}. \tag{4.8}$$

Remark 11. *We will justify the averaged value* (4.8) *in Section 4.3, where we focus on the comparison between the MMAm and homogenization techniques.*

When analysing MMAm results, we will distinguish between the FEM approximation error and the medium approximation error. The FEM approximation error is defined as the error of the best approximation, i.e.

$$E_{FEM} = \inf_{v_h \in V_h} |u - v_h|_{0,\Omega},$$

while the medium approximation error is defined as $E_{MED} = \mathcal{M}_{h,\epsilon}$ (the quantity $\mathcal{M}_{h,\epsilon}$ is defined in Definition 7 of Chapter 3). We observe that for a given mesh (i.e. h is fixed), the FEM approximation error is fixed but the medium approximation error can be reduced by refining the submesh (i.e. ϵ goes to zero).

In each of the following examples, we consider a fixed propagation medium together with a given mesh and an approximation order. We present the results obtained for different values of ω and ϵ. In particular, we show that in the case where the dominant part of the error is due to the medium approximation, the quality of the numerical solution can be slightly improved by increasing the number of subcells.

For the computations, we use triangular Lagrangian finite elements. The medium approximation submesh is obtained through a homothety of the reference triangle, as shown in Figure 4.10. Note that those meshes are obviously regular and satisfy the hypothesis of Proposition 12.

4.2.1 Analytical solution

To construct an analytical solution, we introduce an auxiliary 1D problem, that is to find $v \in C^1([0, Z])$ such that

$$\begin{cases} -\dfrac{\omega^2}{c^2(z)} v(z) - v''(z) &= 0 \quad \text{for } z \in (0, Z) \\ -v'(0) &= 1 \\ v'(Z) - i\dfrac{\omega}{c(Z)} v(Z) &= 0, \end{cases}$$

where c is piecewise constant on a partition $0 = z_0 < z_1 < \cdots < z_m = Z$. We obtain an analytical solution using the methodology presented in Section 4.1.1.

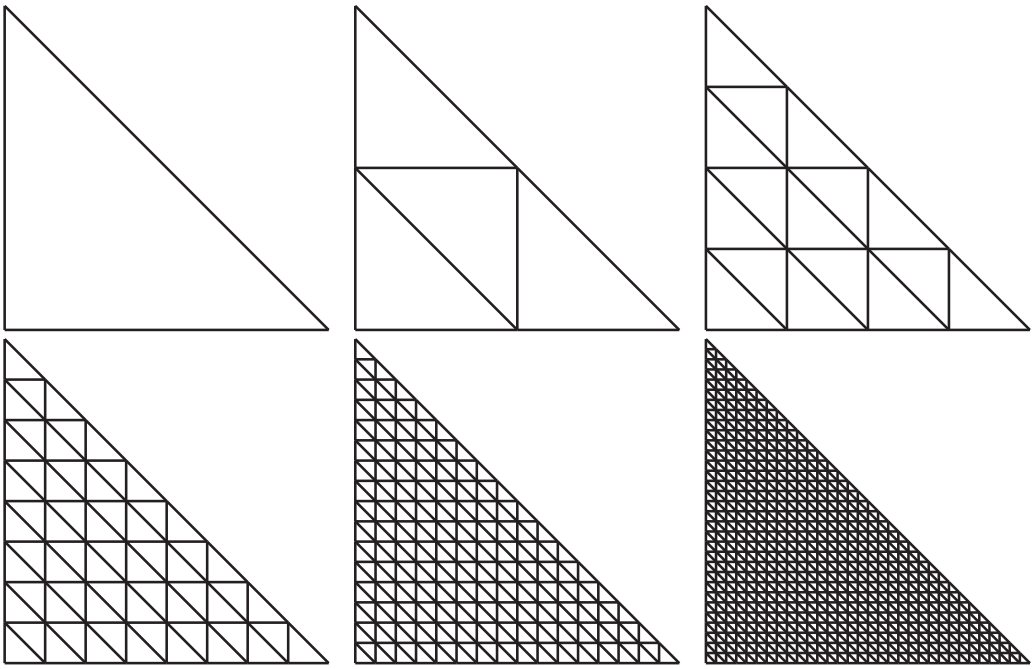


Figure 4.10: Velocity approximation schemes for $\epsilon = 1, 0.5, 0.25, 0.125, 0.0625$ and 0.03125.

We then get a two dimensional problem by setting $\Omega = (0, 1000) \times (0, Z)$ and $k \in L^\infty(\Omega)$ is defined as $k \in L^\infty(\Omega)$, $k|_{\Omega_j} = \omega/c_j$ where $\Omega_j = (0, 1000) \times (x_{j-1}, x_j)$. Then $u(x_1, x_2) = v(x_2)$ is the unique solution to

$$\begin{cases} -k^2 u - \Delta u &= 0 & \text{in } \Omega \\ \nabla u \cdot \mathbf{n} &= 1 & \text{on } (0, 1000) \times \{0\} \\ \nabla u \cdot \mathbf{n} - ik_L u &= 0 & \text{on } (0, 1000) \times \{Z\} \\ \nabla u \cdot \mathbf{n} &= 0 & \text{on } \{0\} \times (0, Z) \\ \nabla u \cdot \mathbf{n} &= 0 & \text{on } \{1000\} \times (0, Z). \end{cases}$$

4.2.2 A two-layered media

We begin with evaluating the medium approximation error as a function of ϵ. For that purpose, we consider the case of a two-layered medium composed of two homogeneous layers. In this case, the use of a fitting mesh is obviously relevant and this case gives us a way to measure the effect of MMAm on the accuracy of the solution.

We set $x_0 = 0, x_1 = 500, x_2 = Z = 1000$, $c_1 = 1000$ and $c_2 = 2000$. In order to quantify the error coming from the medium approximation we use both a fitting and a non-fitting meshes. When using the fitting mesh, the medium is perfectly represented, since the coefficient c is constant in each cell of the finite element mesh. On the other hand, when using the non-fitting mesh, c must be approximated by c_ϵ since it may vary inside an element. The experiment then shows that when the velocity approximation is refined, the solution error obtained with the non-fitting mesh is getting closer to the error

obtained on the fitting mesh.

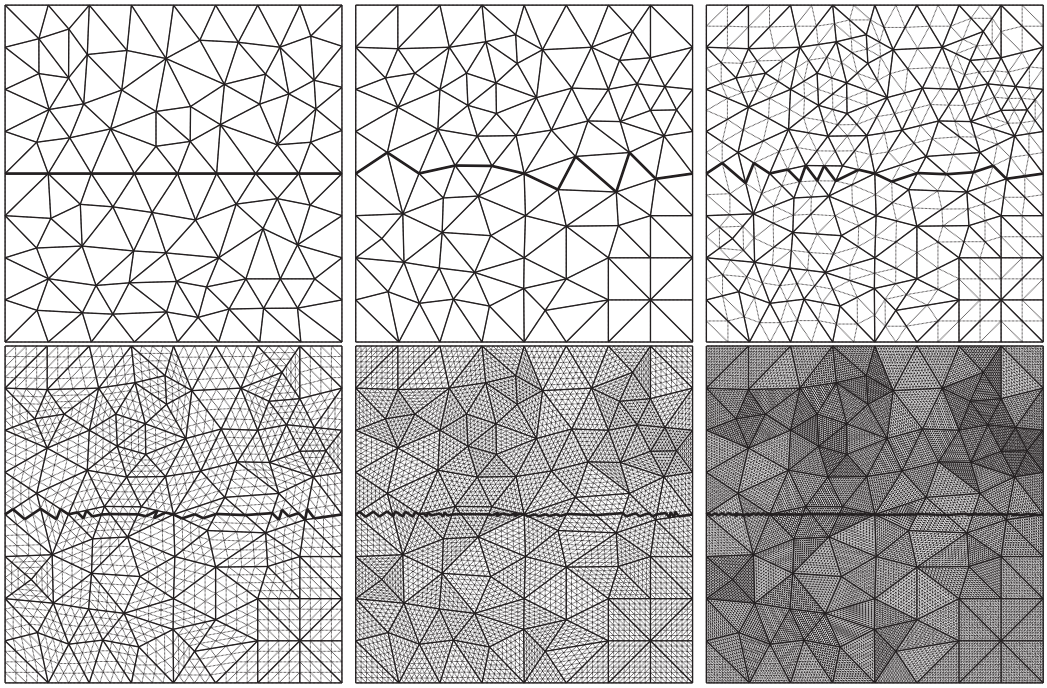


Figure 4.11: Evolution of the interface: fitting mesh (top-left) and non fitting mesh with $\epsilon = 1, 0.5, 0.25, 0.125$ and 0.0625.

The non-fitting mesh contains 164 cells and the fitting mesh contains 166 cells. We start with $\mathcal{P}_2$ elements and the corresponding results are represented in the Table 4.11.

In the first column, the integer numbers indicate the number of subcells that are used to approximate the velocity inside each cell of the non-fitting mesh. The last line stands for the results obtained by using the standard $\mathcal{P}_2$ FEM with the fitting mesh.

$\mathcal{P}_2$	$\omega = 2\pi$	$\omega = 4\pi$	$\omega = 6\pi$
1	9.76×10^{-2}	2.38×10^{-1}	9.11×10^{-1}
4	2.26×10^{-2}	7.92×10^{-2}	3.24×10^{-1}
16	1.18×10^{-2}	4.62×10^{-2}	2.02×10^{-1}
64	5.20×10^{-3}	3.76×10^{-2}	2.05×10^{-1}
256	3.05×10^{-3}	3.61×10^{-2}	2.09×10^{-1}
1024	2.59×10^{-3}	3.59×10^{-2}	2.11×10^{-1}
fitting	1.81×10^{-3}	3.78×10^{-2}	2.65×10^{-1}

Table 4.11: $\mathcal{P}_2$ elements

We can observe that for each value of ω, the error decreases when letting ϵ go to 0. Moreover, when comparing with the last line of the table, we can see that the MMAm reaches the same level of accuracy that the standard $\mathcal{P}_2$ FEM. When the frequency is

increasing, the two methods result in the same level of accuracy and MMAm accuracy seems to reach a plateau. We believe that the medium approximation error becomes so small that quickly the values of the error describe the finite element approximation only.

Table 4.12 represents the results obtained when using $\mathcal{P}_4$ elements. The same conclusions hold except that due to a highest degree of approximation, the medium approximation error stabilizes itself on a plateau for $\omega = 10\pi$ only. It is worth noting than when ω is less than 10π, the convergence is super linear which illustrates well Section 3 results.

$\mathcal{P}_4$	$\omega = 2\pi$	$\omega = 4\pi$	$\omega = 6\pi$	$\omega = 8\pi$	$\omega = 10\pi$
1	9.67×10^{-2}	2.25×10^{-1}	3.42×10^{-1}	4.81×10^{-1}	5.01×10^{-1}
4	2.22×10^{-2}	6.59×10^{-2}	1.42×10^{-1}	4.03×10^{-1}	1.90×10^{-1}
16	1.22×10^{-2}	3.75×10^{-2}	6.65×10^{-2}	2.37×10^{-1}	8.94×10^{-2}
64	4.70×10^{-3}	1.44×10^{-2}	2.74×10^{-2}	9.81×10^{-2}	4.50×10^{-2}
256	1.47×10^{-3}	4.91×10^{-3}	1.13×10^{-2}	4.54×10^{-2}	2.94×10^{-2}
1024	5.25×10^{-4}	1.54×10^{-3}	4.58×10^{-3}	1.67×10^{-2}	2.52×10^{-2}
fitting	2.62×10^{-6}	8.80×10^{-5}	8.10×10^{-4}	5.76×10^{-3}	2.44×10^{-2}

Table 4.12: $\mathcal{P}_4$ elements

The behaviour of the discrete solution as $\epsilon \to 0$ is depicted on Figures 4.12 and 4.13 for $\mathcal{P}_2$ and $\mathcal{P}_4$ elements respectively. We see that the MMAm solution is as accurate as the solution computed on the fitting mesh when ϵ is sufficiently small.

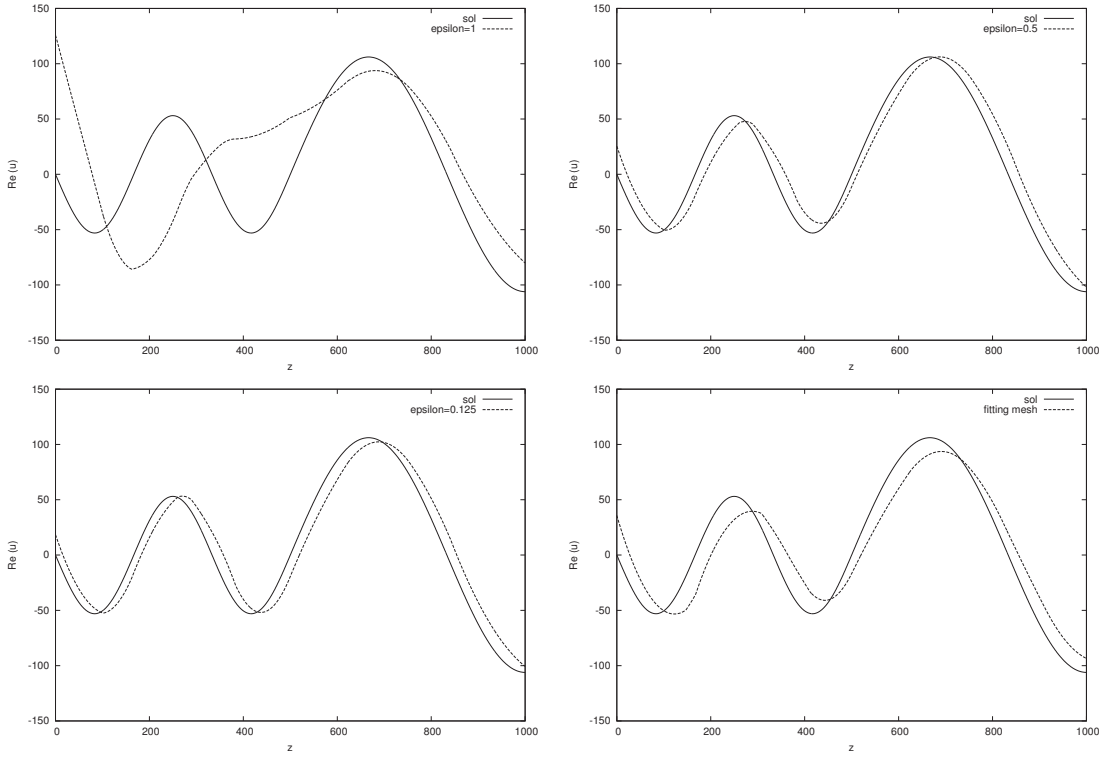


Figure 4.12: Solution profile for $\mathcal{P}_2$ elements, $\omega = 6\pi$

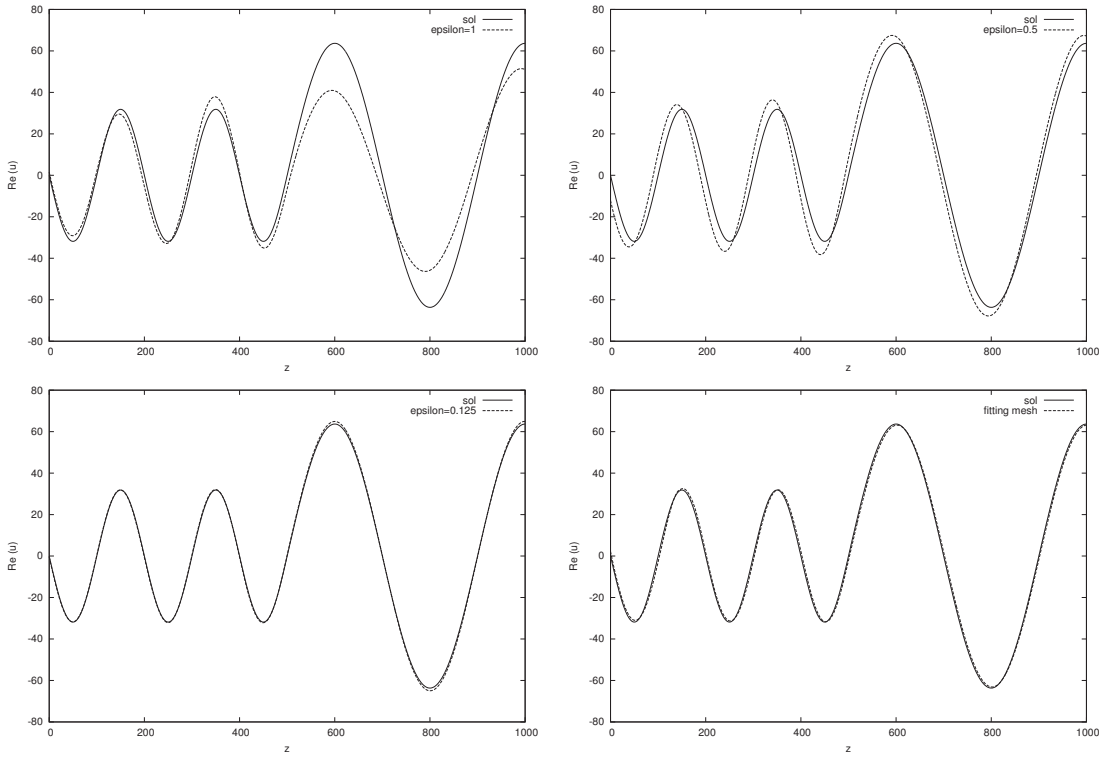


Figure 4.13: Solution profile for $\mathcal{P}_4$ elements, $\omega = 10\pi$

4.2.3 Multi-layered medium

We now set $Z = 3000$. We decompose the propagation domain into 1000 layers of 3 meters each. We set $c_{min} = 1500$, $c_{max} = 5500$. The velocity parameter varies linearly from $c_1 = c_{min}$ to $c_{1000} = c_{max}$. We use $\mathcal{P}_6$ elements on a 1033 cells mesh. We carry out simulations for different values of ϵ. To compare with parameter averaging methods, we perform simulations for $k_\epsilon^2|_K$ given as the mean value of k^2 on the cell K as explained above (4.8).

On table 4.13, we present the results that we have obtained by discretizing with $\mathcal{P}_6$ Lagrangian elements. We can draw the same conclusion than in the previous test case. It is interesting to note that the MMAm results are always better then when the standard FEM is used with the mean value of the wavenumber in each cell. This example shows that the subscheme quadrature strategy of the MMAm is superior to a simple averaging of the wavenumber, as depicted by the first line of Table 4.13.

It is also clear that for a given pulsation, reducing the approximation step ϵ reduces the solution error. For the lowest pulsation $\omega = 20\pi$, the convergence is super linear, which is consistent with the results of section 3. For higher pulsations, the part of the error due to finite element approximation is much larger, so that the linear convergence is not observed anymore.

For the lowest frequency $\omega = 20\pi$, the standard finite element solution is accurate

($\epsilon = 1$), but it is mandatory use the MMAm to obtain an accurate solution for higher frequencies ($\epsilon \leq 0.25$ for $\omega = 40\pi$ for instance). This observation is in accordance with the error-estimate of Theorem 14. Indeed, the term related to medium approximation in the error-estimate behaves as $\mathcal{O}(\omega \mathcal{M}_{h,\epsilon})$, which let us guess that the medium approximation error has more impact for high frequencies ω.

Figure 4.14 illustrates how the finite element solution depends on ϵ in the case $\omega = 60\pi$.

$\mathcal{P}_6$	$\omega = 20\pi$	$\omega = 30\pi$	$\omega = 40\pi$	$\omega = 50\pi$	$\omega = 60\pi$
mean	4.38×10^{-2}	1.70×10^{-1}	6.44×10^{-1}	1.90×10^{-1}	2.33×10^{0}
1	4.19×10^{-2}	1.61×10^{-1}	5.04×10^{-1}	1.87×10^{-1}	1.19×10^{0}
4	7.27×10^{-3}	2.39×10^{-2}	4.83×10^{-1}	1.02×10^{-1}	4.47×10^{-1}
16	2.12×10^{-3}	7.06×10^{-3}	5.97×10^{-2}	6.63×10^{-2}	3.52×10^{-1}
64	1.02×10^{-3}	3.76×10^{-3}	3.64×10^{-2}	6.33×10^{-2}	3.34×10^{-1}
256	4.93×10^{-4}	1.74×10^{-3}	3.52×10^{-2}	6.26×10^{-2}	3.40×10^{-1}
1024	2.00×10^{-4}	9.40×10^{-4}	3.69×10^{-2}	6.19×10^{-2}	3.37×10^{-1}

Table 4.13: Multi-layered medium

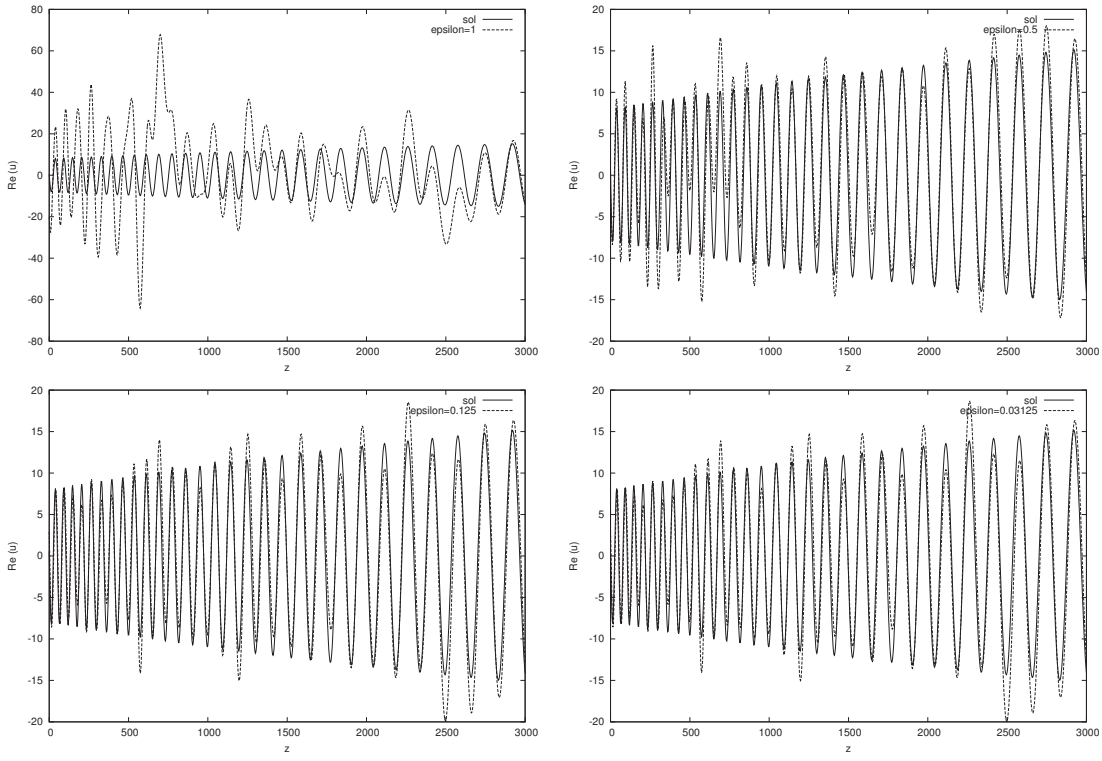


Figure 4.14: Solution profile in gradient domain for $\mathcal{P}_6$ elements, $\omega = 60\pi$

4.2.4 Multi-layered medium: Highly heterogeneous

We consider here the case where the velocity does not satisfy the technical condition (3.2). The velocity model is now constructed by modifying the previous one as follows. Between 0 and 1500 meters and between 2000 and 3000 meters, the velocity is decreased by 500 every other layer, and increased by 500 in the remaining layers. Between 1500 and 2000 meters, the velocity is 500 in every other layer. We use an adaptive mesh, which is more refined between 1500 and 2000 meters in order to correctly fit the small wavelength in this area. The mesh is made of 4838 cells and is represented in Figure 4.15.

$\mathcal{P}_6$	$\omega = 20\pi$	$\omega = 30\pi$	$\omega = 40\pi$	$\omega = 50\pi$
mean	1.06×10^{0}	6.73×10^{-1}	1.17×10^{0}	2.76×10^{0}
1	9.99×10^{-1}	1.81×10^{0}	7.44×10^{0}	3.20×10^{0}
4	7.41×10^{-1}	3.84×10^{0}	1.71×10^{0}	1.88×10^{0}
16	3.41×10^{-1}	6.79×10^{-1}	3.34×10^{0}	2.68×10^{0}
64	3.12×10^{0}	1.86×10^{-1}	4.45×10^{-1}	1.05×10^{0}
256	8.40×10^{-2}	6.60×10^{-2}	1.03×10^{-1}	2.77×10^{-1}
1024	6.23×10^{-2}	3.63×10^{-2}	7.00×10^{-2}	2.12×10^{-1}

Table 4.14: Highly heterogeneous multi-layered medium

The results of the experiment are presented on Table 4.14 and some solution profiles are plotted on Figure 4.16. We observe that the MMAm solution is accurate as soon as ϵ is less that 0.0625, which means that we need to use at least 256 subcells to compute the entries of the matrix. This is not surprising because we consider a velocity model including very strong contrasts. It is indeed composed of very thin layers and the variations of the velocity are important.

It is also clear that the averaging method is not satisfying for this experiment. Indeed, as shown by the first line of Table 4.14, the solution obtained with the averaged parameter is not accurate, even for the lowest frequency $\omega = 20\pi$.

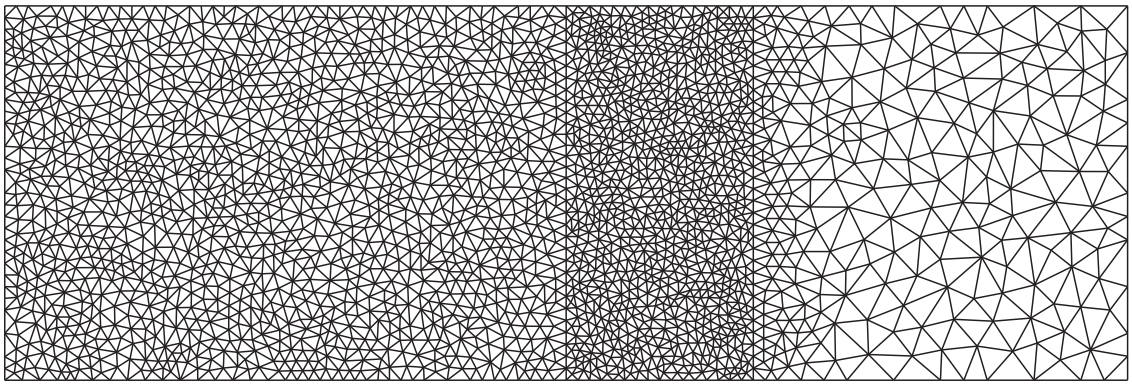


Figure 4.15: Adaptive mesh (90 degrees rotation)

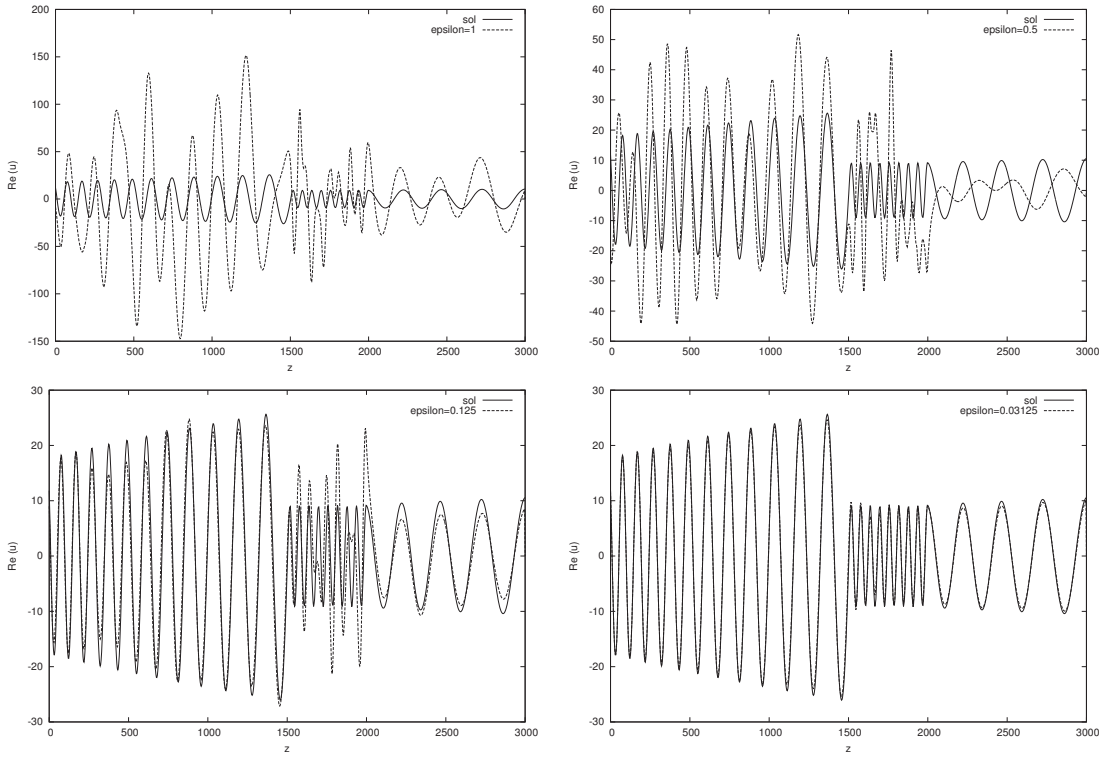


Figure 4.16: Solution profile in highly heterogeneous domain for $\mathcal{P}_6$ elements, $\omega = 40\pi$

4.2.5 High order MMAm VS fitting mesh based method

In the previous numerical tests, we have shown the efficiency of the MMAm and we have concluded that when using enough subcells we obtain accurate results even in highly heterogeneous media. In particular, the first experiment showed that when a fitting mesh is available, the accuracy of the MMAm on a non-fitting mesh of the same size is comparable to the standard FEM on the fitting mesh.

In this section, we investigate the reduction of the computational cost offered by the MMAm to obtain a 5% relative error on the previous velocity model at the frequency $\omega = 40\pi$.

We use regular meshes based on cartesian grids of different sizes and different polynomial degrees. As a starting point, we discretize the problem with the coarsest possible fitting mesh. The mesh steps are given by $h_x = 3.33m$, $h_z = 3m$. The z step is chosen to be exactly the length of a layer, so that the mesh is fitting, and the x step is chosen so that the grid cells are nearly squares. Hence, the mesh is formed by a regular grid of 300×1000 squares, each square being divided into two triangles.

If we use $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ elements on the fitting mesh, we obtain relative L^2 errors of 1.79×10^{-1}, 3.19×10^{-4} and 1.21×10^{-6}. We thus have that the $\mathcal{P}_1$ solution is not precise enough regarding the level of accuracy we target and the $\mathcal{P}_2$ and $\mathcal{P}_3$ solutions are very precise but in the same time very expensive to compute. For example, the computation of

the $\mathcal{P}_2$ solution requires to invert a system with 1.20×10^6 degrees of freedom and 1.26×10^7 non-zero elements in the matrix.

We now focus on the size of the cells which obviously impacts the size of the corresponding linear system. It turns out that if p is greater than 2, the MMAm delivers 5% relative error on a much coarser (and non-fitting) mesh than the fitting mesh as shown in Table 4.15. We see that when p is greater than 2, we can use a coarse non-fitting mesh and use less than 1.20×10^6 degrees of freedom to get 5% of accuracy. We conclude that the MMAm enables to reduce the computational cost compared to the standard FEM on fitting meshes.

p	err	h_z	ndf	nz
1	5.5×10^{-2}	1.5	12.0×10^5	14.4×10^6
2	4.0×10^{-2}	6	3.01×10^5	3.15×10^6
3	5.8×10^{-2}	12	1.70×10^5	2.06×10^6
4	5.9×10^{-2}	18.75	1.24×10^5	1.84×10^6
5	5.9×10^{-2}	20	1.70×10^5	3.12×10^6
6	5.5×10^{-2}	24	1.67×10^5	3.76×10^6

Table 4.15: Comparison of different p to obtain a 5% accuracy

To give a comparison with another fitting mesh method, consider the coarsest fitting cartesian grid made of 300×1000 squares. It includes 6.01×10^5 edges, which means that lowest order discontinuous Galerkin plane wave method would require at least 6.01×10^5 degrees of freedom to solve (see, for example [10]). On the other hand, the $\mathcal{P}_4$ solution is computed on a 64×160 non-fitting cartesian grid. This grid is much coarser than the 300×1000 fitting grid and the number of degrees of freedom required to obtain the $\mathcal{P}_4$ solution is 1.24×10^5 (4.8 times less than for the plane wave method).

4.3 Comparison with homogenization

Periodic homogenization techniques have been applied to upscale fine scale properties of the Earth. They permit to obtain an "effective" propagation medium that can be easily meshed. Under restrictive assumptions of periodicity, the homogenization process is well understood and convergence analysis is available [8,33]. In Geophysics, the idea dates back to 1962: Backus showed that fine scale isotropic layers can be upscaled into a homogeneous anisotropic medium [18]. Periodic homogenization has two main drawbacks:

- in the standard mathematical setting, even if the periodicity hypothesis can be weakened, the medium parameters are expected either to belong to a family of parameters which converge in some weak sense (for instance, two-scale convergence [8] or unfolding methods [33]), or to be randomly distributed according to a specific distribution [64]. Unfortunately, these hypothesis seem hard to apply to geophysical models.

- a separation of scale between the wavelength and the spatial period of the heterogeneities is required. This point has been quantitatively analysed by Carcione et al. [31]. They showed that the anisotropic approximation of finely layered media is valid only if the wavelength is at least five times larger than the spatial period of the heterogeneities.

In recent developments, Capdeville and collaborators have introduced the so-called "non-periodic homogenization" framework [27–29]. The procedure is based on a user defined parameter λ_0 which separates the microscopic and macroscopic scales. The homogenization procedure includes a low-pass filtering to upscale properties of the medium under the wavelength λ_0.

The advantage of the non-periodic homogenization procedure of Capdeville et al. over "standard" homogenization methods is that there is no requirement on the medium parameters. However, the scale separation between the wavelength λ and the low-pass filtering parameter λ_0 is still mandatory to obtain accurate results. Also, though the non-periodic homogenization procedure of Capdeville et al. is numerically efficient, it lacks a strong mathematical justification.

Because the non-periodic homogenization procedure developed by Capdeville et al. is an attractive solution for wave propagation in highly heterogeneous media, we propose a comparison. We focus on the simplest case of a periodic layered medium with constant density. In Subsection 4.3.1 we briefly present the principle of periodic homogenization and derive the formula for the homogenized coefficients. We carry out simulations using the homogenized parameters in Subsection 4.3.2 and compare the results with the MMAm in Subsection 4.3.3.

4.3.1 Principle of periodic homogenization

The aim of this subsection is to introduce the concept of periodic homogenization. We consider that the medium of propagation is periodic, with a small period ϵ. We consider the "cell" $Y = (0,1)^N$. Then periodic medium parameters κ_ϵ and ρ_ϵ can be defined as

$$\kappa_\epsilon(x) = \kappa(\epsilon^{-1}x), \quad \rho_\epsilon(x) = \rho(\epsilon^{-1}x)$$

where κ and ρ are $Y-$periodic functions. For a given period ϵ, the wave equation reads

$$\mathcal{H}_\epsilon u_\epsilon = -\frac{\omega^2}{\kappa_\epsilon}u_\epsilon - \operatorname{div}\left(\frac{1}{\rho_\epsilon}\nabla u_\epsilon\right) = f. \tag{4.9}$$

Problem (4.9) is actually defined for each $\epsilon > 0$. Hence, if we denote by $\mathcal{H}_\epsilon$ the Helmholtz operator for the period ϵ, we have, for any $\epsilon > 0$ a function $u_\epsilon \in H^1(\Omega, \mathbb{C})$ such that

$$\mathcal{H}_\epsilon u_\epsilon = f.$$

It is actually possible to show that when ϵ tends toward 0, u_ϵ weakly converges to a function u^0 in $H^1(\Omega)$. The aim of the homogenization process is to identify this weak limit.

An important result is that u^0 can be characterized as the solution of

$$\hat{\mathcal{H}}u^0 = f,$$

where $\hat{\mathcal{H}}$ is the so-called "homogenized" operator.

The homogenization procedure has two main interests. First, we obtain u^0 which might be used as an approximation of u_ϵ for small ϵ. Second, and more importantly, the homogenized operator $\hat{\mathcal{H}}$ describe the macroscopic representation of microscopic scales. In $\hat{\mathcal{H}}$ the dependency of ϵ is gone and there is no microscopic scale.

In the context of wave propagation, the homogenized operator $\hat{\mathcal{H}}$ is still a wave propagation operator and it has the following shape

$$\hat{\mathcal{H}}u = -\frac{\omega^2}{\hat{\kappa}}u - \operatorname{div}\left(\hat{B}\nabla u\right), \tag{4.10}$$

where $\hat{\kappa} \in \mathbb{R}$ and $\hat{B} \in M_2(\mathbb{R})$ are constant parameters. First, we see that $\hat{\mathcal{H}}$ is different from the operators $\mathcal{H}_\epsilon$ in the sense that it can be anisotropic because ρ_ϵ^{-1} has been changed by a matrix. Second, the small scales have actually disappeared since the parameters are now constant.

From a numerical point of view, the homogenized operator is interesting because it is much simpler to discretize than the original operator. Indeed, though it is anisotropic, the parameters are constant so that there is no restriction on the mesh due to small scale heterogeneities.

From a physical point of view, we can consider that an isotropic periodic medium with a small-period behaves macroscopically like an homogeneous, but possibly anisotropic, medium.

The expressions of the homogenized parameters $\hat{\kappa}$ and $\hat{B}$ are available in the literature (see [8,86] for instance). The coefficients of the matrix $\hat{B}$ are obtained as the mean value of the solution to a PDE set in the reference cell with periodic boundary conditions. Hereafter, we will focus on the simplest case where the density is constant. Hence, it is only required to homogenize the parameter κ and the corresponding value is given by

$$\frac{1}{\hat{\kappa}} = \int_Y \frac{1}{\kappa}. \tag{4.11}$$

4.3.2 Experiments with the homogenized parameters

We consider a layered propagation medium $\Omega = (0, 1000) \times (0, 1000)$ with constant density ρ. Each horizontal layer has the same length ϵ. The wavespeed c takes two different values c_1 and c_2 every other layer. In accordance with (4.11), we also consider the homogenized version where the wavespeed is constant and takes the value

$$\hat{c} = \frac{c_1 c_2}{\sqrt{c_1^2 + c_2^2}}.$$

We propose to solve the original problem

$$\begin{cases} -\dfrac{\omega^2}{c^2}u - \Delta u &= f \quad \text{in } \Omega \\ \nabla u \cdot \mathbf{n} - \dfrac{\mathbf{i}\omega}{c}u &= 0 \quad \text{on } \partial\Omega, \end{cases}$$

with the MMAm and the homogenized problem

$$\begin{cases} -\dfrac{\omega^2}{\hat{c}^2}u - \Delta u &= f \quad \text{in } \Omega \\ \nabla u \cdot \mathbf{n} - \dfrac{\mathbf{i}\omega}{\hat{c}}u &= 0 \quad \text{on } \partial\Omega, \end{cases}$$

with the standard FEm.

In every experiments, we keep $c_2 = 3000$ m.s^{-1}. We try different velocity contrasts by choosing $c_1 = 1000, 2000$ or 2500 m.s^{-1}. We experiment with frequencies ranging from 1 to 20Hz. We also experiment with different numbers of layer $m = 21, 51, 101$ or 201 and set $\epsilon = 1000/m$. Hence ϵ approximately ranges from 50 m to 5 m. We choose an odd number of layers so that the Dirac source, located at $(500, 500)$ always lies within the middle layer.

We carry out simulations on a fitting cartesian grid based mesh (210×210 for $m = 21$, 204×204 for $m = 51$, 202×202 for $m = 101$ and 201×201 for $m = 201$) with $p = 3$ elements. The results are presented from Figure 4.17 to Figure 4.20. The error axis are limited to 50% of relative L^2 error. The curve 2000, refers to the experiment where $c_1 = 1000$ and the velocity contrast is 2000. The curve 1000 refers to the case where $c_1 = 2000$ and the curve 500 to the case where $c_1 = 2500$.

Let us first consider the case $c_1 = 1000$, where the velocity contrast is the strongest. We see that the homogenized solution has more that 20% relative error for all frequencies, even when there is a high number of layers. Furthermore, the homogenized solution is completely inaccurate at high frequency for the test cases $m = 21, 51$ and 101.

The homogenization method performs better for the other velocity contrasts. For the cases with the higher number of layers $m = 101$ and $m = 201$ we obtain a constant approximation error of 10% for $c_1 = 2000$ and 5% for $c_2 = 2500$ for all frequencies.

In the case where there are 51 layers, we see that homogenization works well with the lowest velocity contrast for all frequencies. However, apart from the case $c_2 = 2500$, the quality of the approximation is decreasing when the frequency increases.

For the hardest case where there are only 21 layers, we see that the homogenized solution is good only for low frequencies and small velocity contrasts.

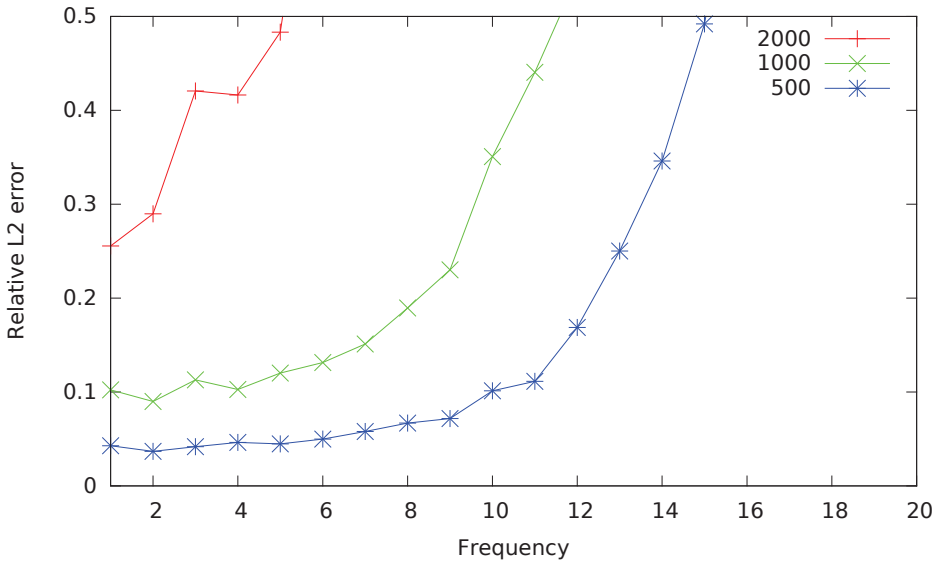


Figure 4.17: Comparison of different velocity contrast for 21 layers

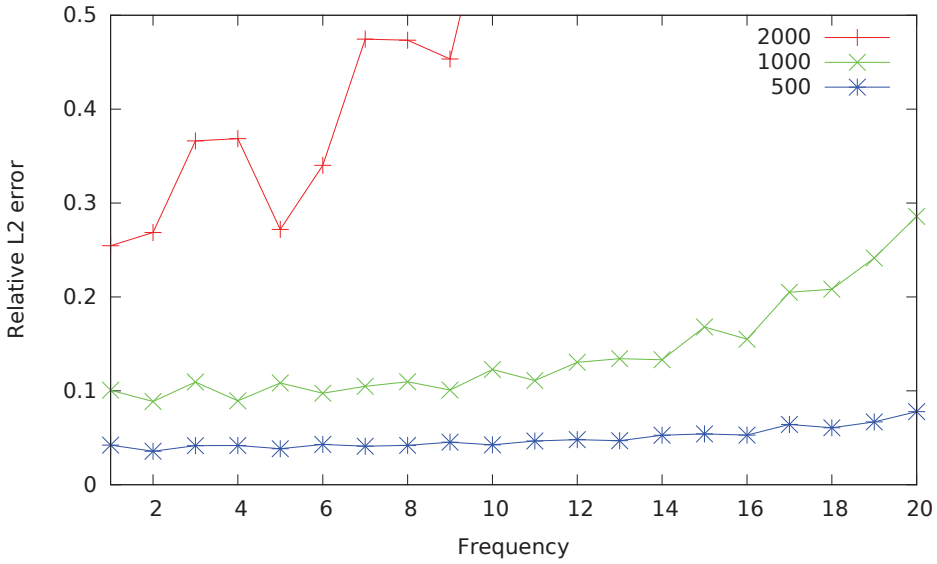


Figure 4.18: Comparison of different velocity contrast for 51 layers

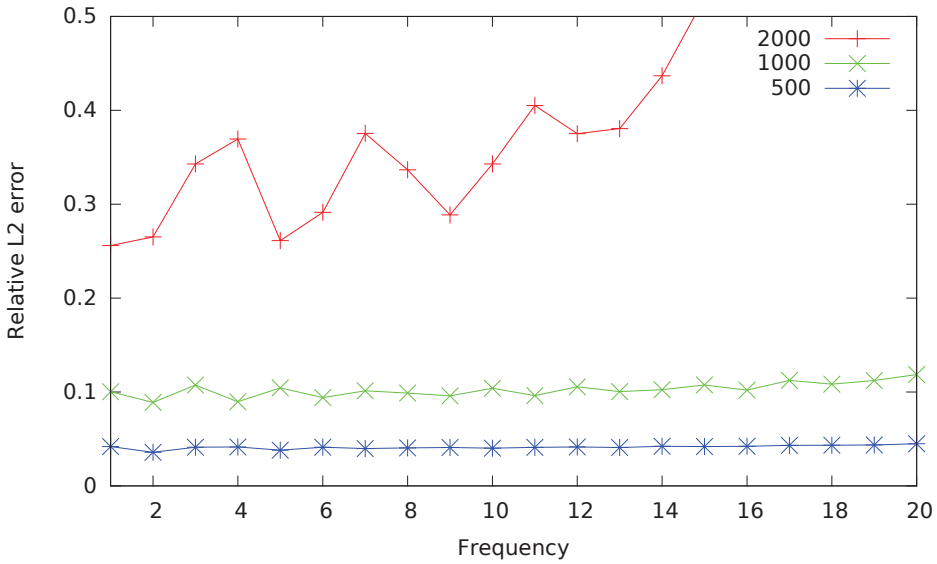


Figure 4.19: Comparison of different velocity contrast for 101 layers

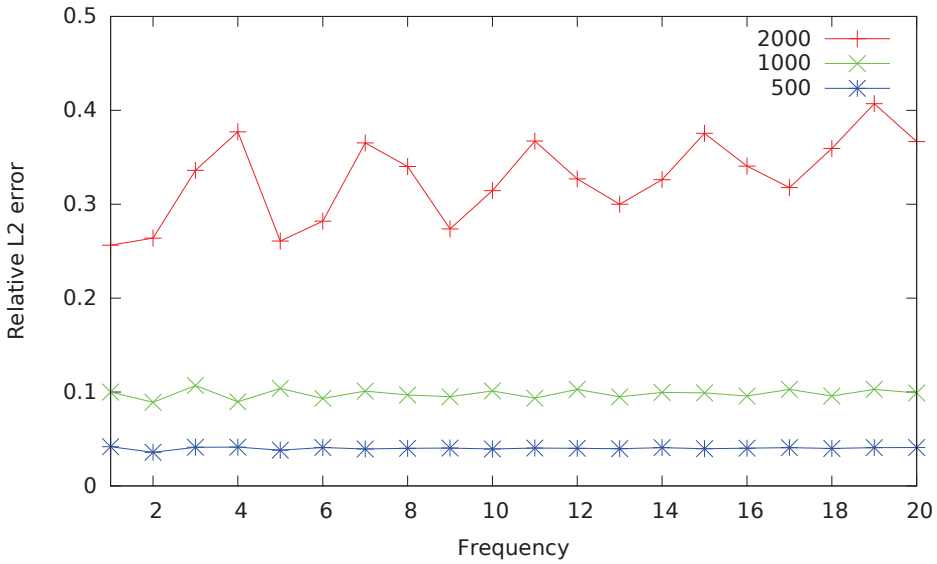


Figure 4.20: Comparison of different velocity contrast for 201 layers

4.3.3 Comparison with the MMAm

We solve the heterogeneous problem using the MMAm on a 10×10 grid with $p = 6$ elements. We use 1024 subcells for the multiscale medium approximation. We compare the MMAm solution to the solution obtained on the fitting mesh with $p = 3$. Results are presented from Figure 4.21 to Figure 4.32.

For the cases with 101 and 201 layers, the results obtained with the MMAm are equivalent to the result obtained in the homogenized medium.

In the experiment with 21 and 51 layers, the MMAm is more accurate than the homogenized solution. The highest contrast case is correctly handle for low frequency $f < 10Hz$ only. When the contrast is smaller ($c_1 = 2000, 25000$), the MMAm solution is accurate even if they are several layers per mesh cell.

These few examples shows that in the simplest case where the density is constant, the MMAm outperforms periodic homogenization techniques, especially when the layers are relatively large. This is because no scale separation is assumed in the method.

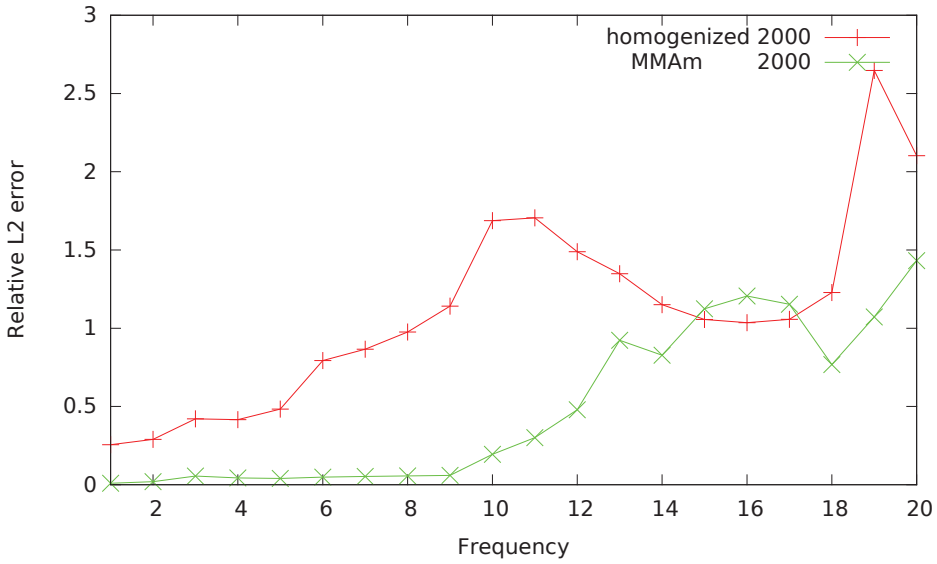


Figure 4.21: Homogenization vs MMam: 21 layers, 2000 m.s^{-1} contrast

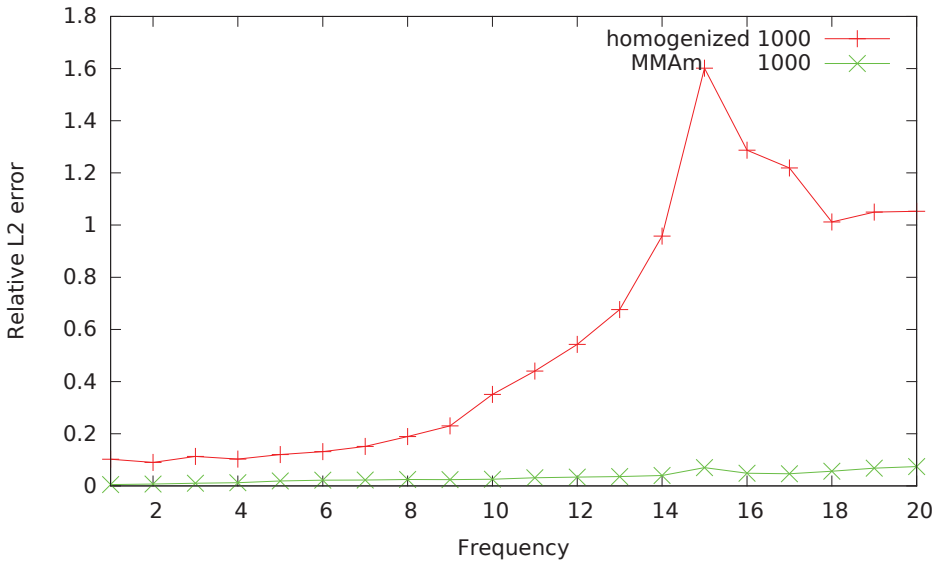


Figure 4.22: Homogenization vs MMam: 21 layers, 1000 $\mathrm{m.s^{-1}}$ contrast

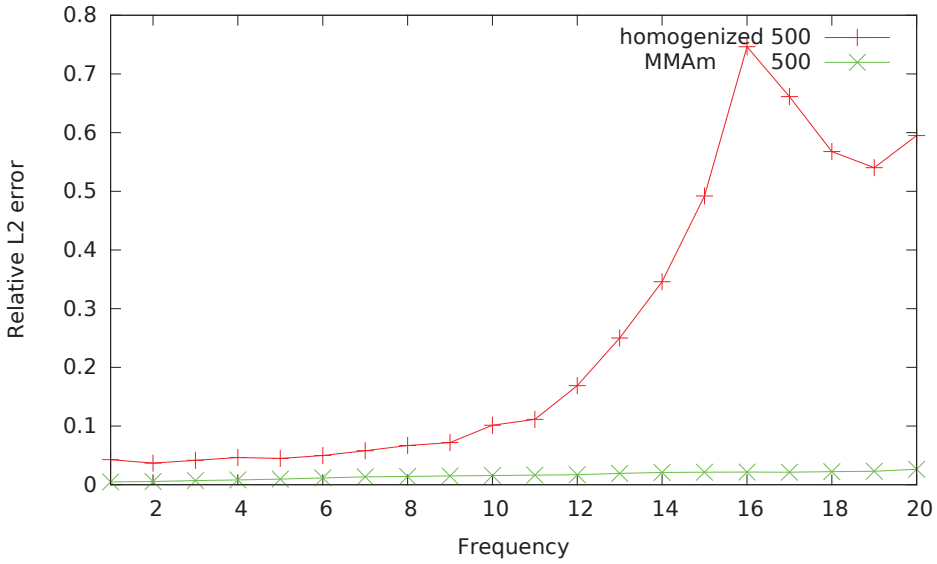


Figure 4.23: Homogenization vs MMam: 21 layers, 500 $\mathrm{m.s^{-1}}$ contrast

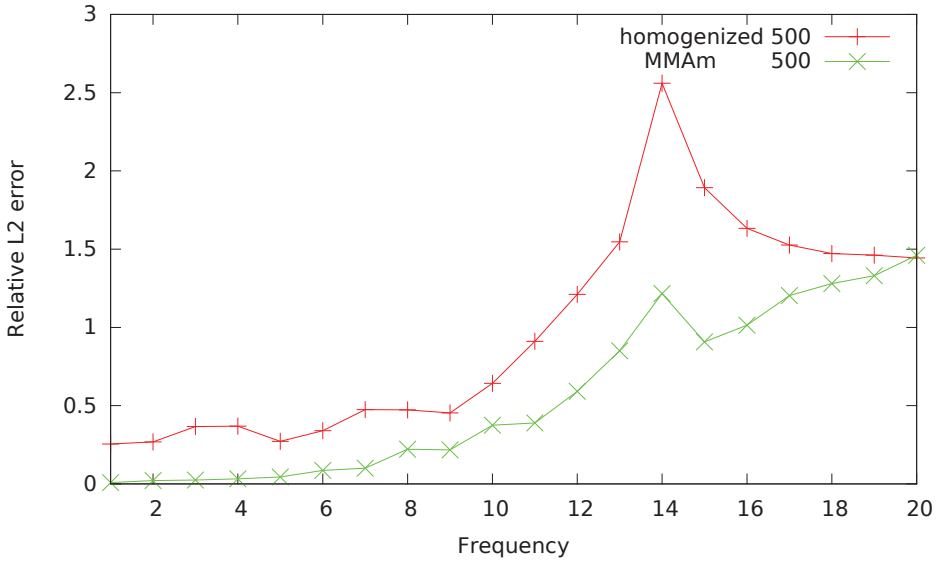


Figure 4.24: Homogenization vs MMam: 51 layers, 2000 m.s^{-1} contrast

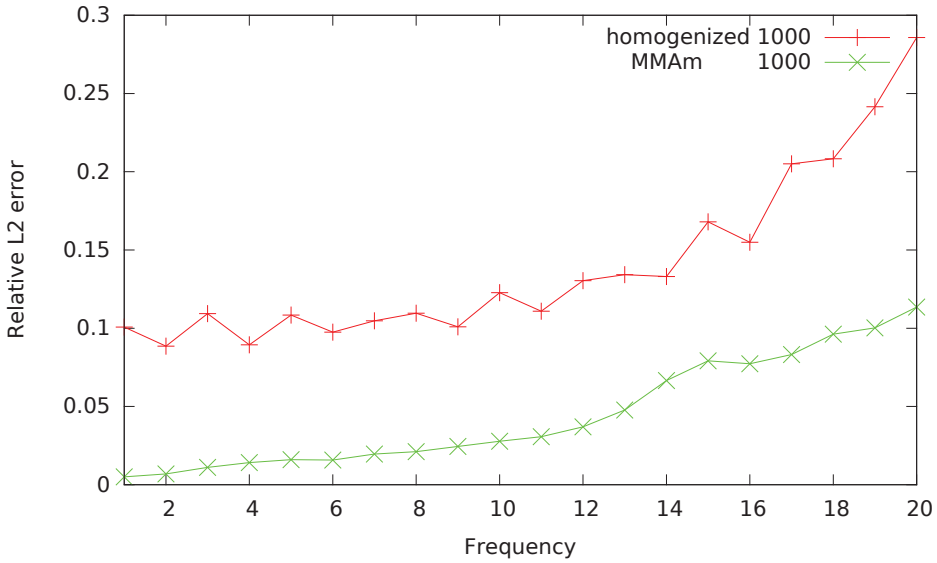


Figure 4.25: Homogenization vs MMam: 51 layers, 1000 m.s^{-1} contrast

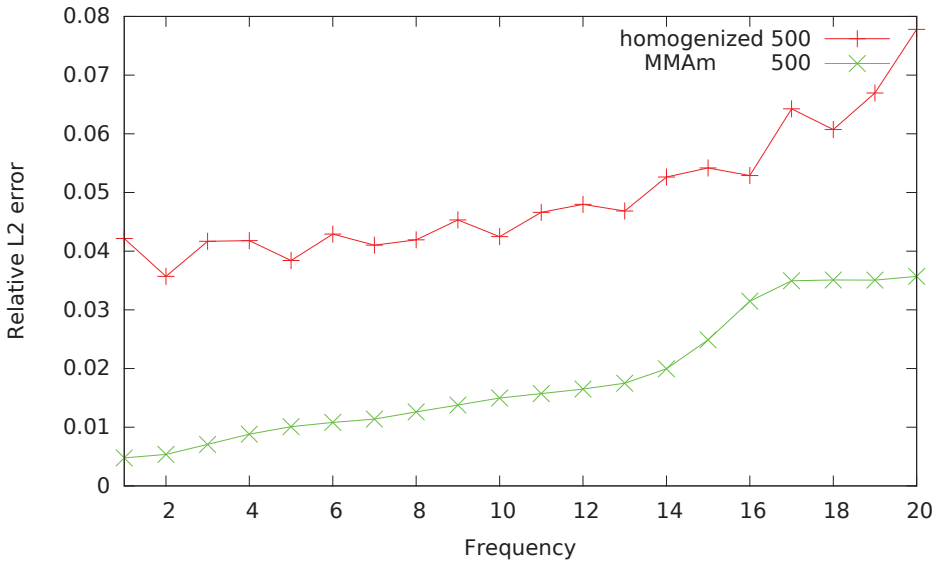


Figure 4.26: Homogenization vs MMam: 51 layers, 500 $\mathrm{m.s^{-1}}$ contrast

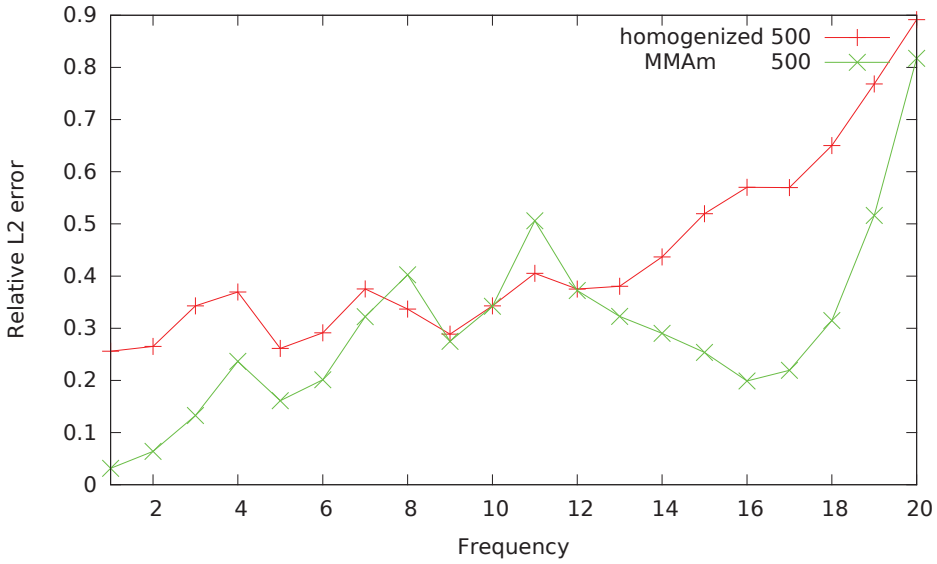


Figure 4.27: Homogenization vs MMam: 101 layers, 2000 $\mathrm{m.s^{-1}}$ contrast

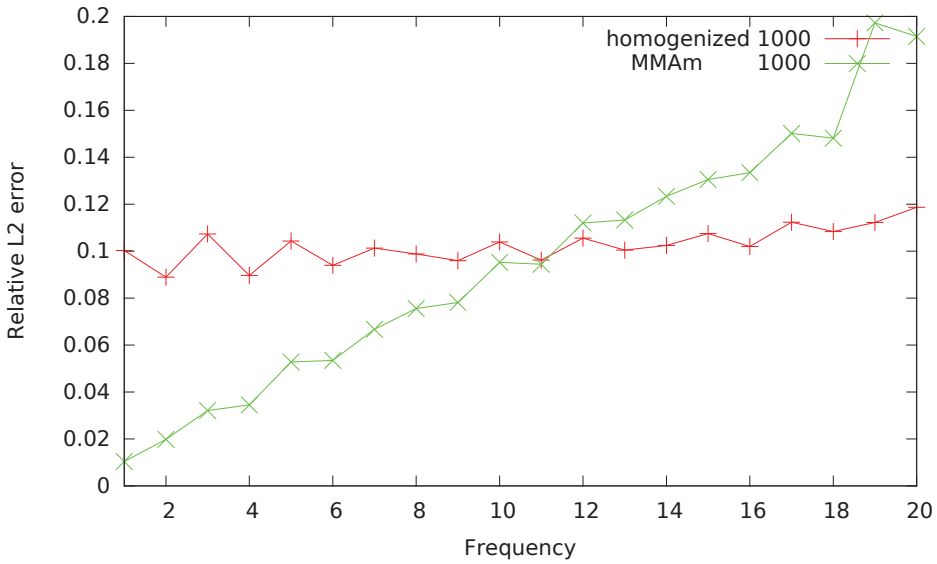


Figure 4.28: Homogenization vs MMam: 101 layers, 1000 m.s^{-1} contrast

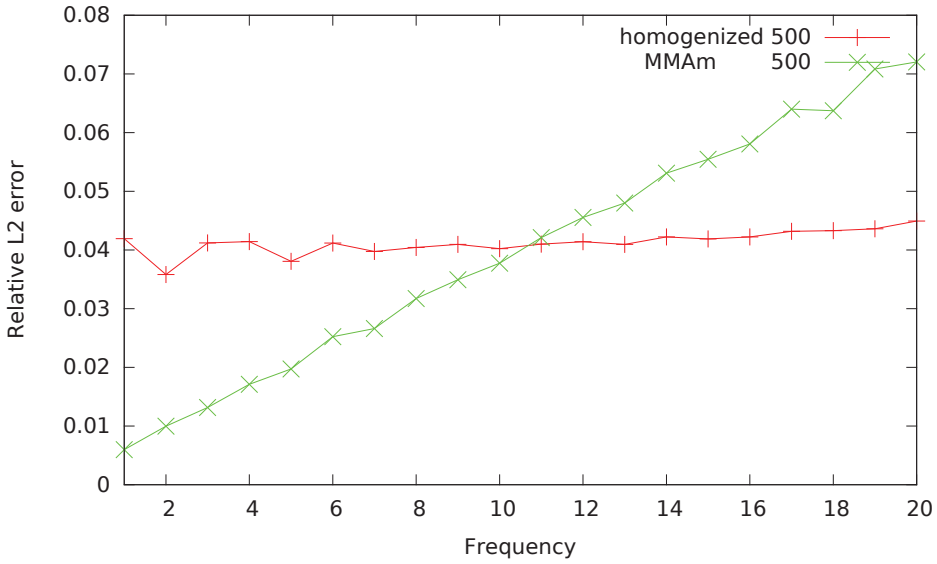


Figure 4.29: Homogenization vs MMam: 101 layers, 500 m.s^{-1} contrast

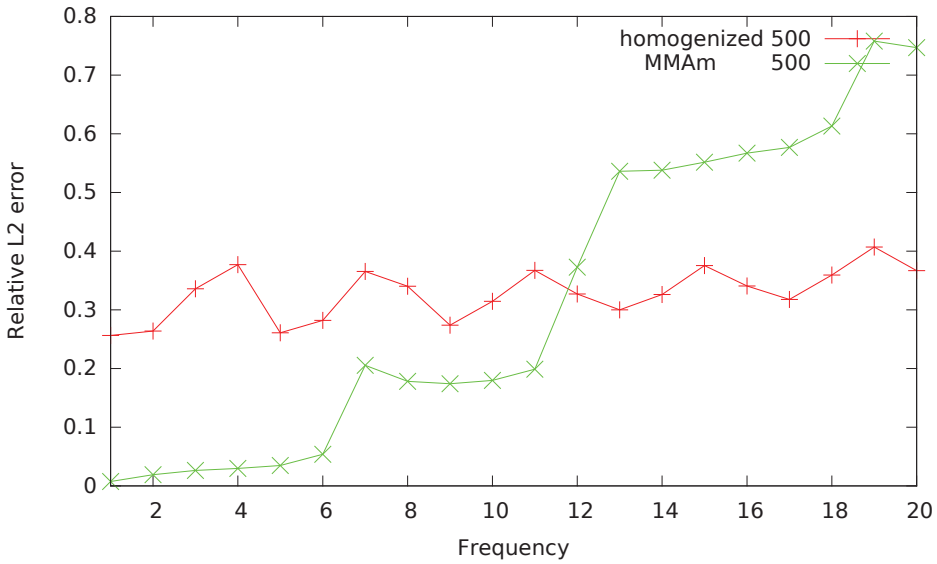


Figure 4.30: Homogenization vs MMam: 201 layers, 2000 m.s^{-1} contrast

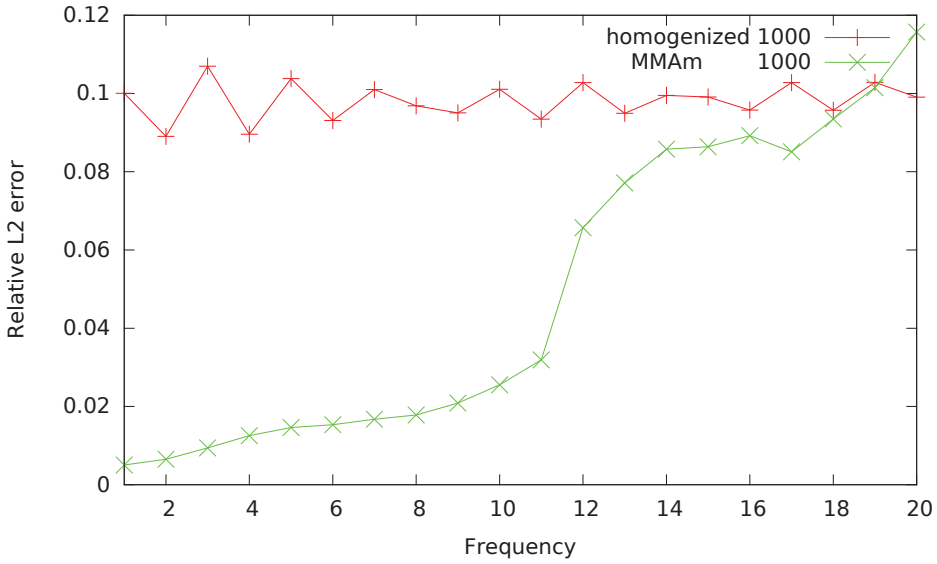


Figure 4.31: Homogenization vs MMam: 201 layers, 1000 m.s^{-1} contrast

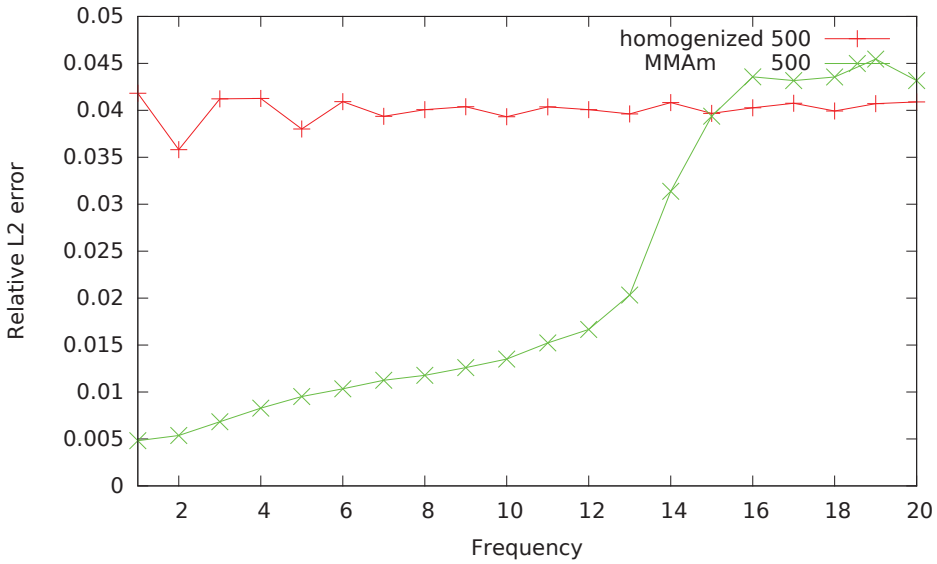


Figure 4.32: Homogenization vs MMam: 201 layers, 500 m.s^{-1} contrast

4.4 Geophysical test-cases

In this section, we consider standard geophysical benchmark models. These geophysical models represent Earth subsurface. In particular, they feature strong contrasts in the medium parameters.

While strongly contrasted media are a good opportunity to validate the MMAm, they might not be completely representative of actual applications. The selected benchmarks can be considered as good Earth models. However, in geophysical applications, wave propagation methods are used in the context of imaging or inverse problems. In this situation, the actual Earth model is not known and the simulations take place either in a "background" model for imaging or in an approximation of the true model for inversion. These models are usually smooth (or at least, smoother than the real Earth) and therefore, easier to handle numerically.

In order to give a fair comparison between the MMAm and the standard FEm, we also include smooth propagation media in our experiments. These smoothed propagation media are obtained by applying a low-pass filter to the original media. This approach is used, for example, in [49]. One can also compare our smooth version of the Marmousi II model (Figure 4.47) with the FWI output model of Sirgue and Pratt [91] depicted on Figure 1.7.

For 2D experiments with $p \geq 3$ we will use static condensation to remove the internal degrees of freedom of each cell K. The process is explained in details by Wilson [105].

4.4.1 Methodology

In the following, we consider geophysical test-cases where the medium parameters are defined by sample values on a cartesian grid. We have arranged the test cases so that the grids dimensions are integer powers of 2 and can be subdivided easily.

For the 2D test-cases, we consider 100 different right hand sides. The right hand sides are Dirac sources located at $z = 50$ m depth and different offsets $\{x_s\}_{s=1}^{100}$. In the elastic case, the Dirac impulsion is placed on the vertical component of the displacement. Hence, the s^{th} right hand side $f^{(s)}$ $(1 \leq s \leq 100)$ is given by

$$\langle f^{(s)}, v\rangle = v(x_s, z), \quad \forall v \in H^1(\Omega, \mathbb{C}), \tag{4.12}$$

in the acoustic case and by

$$\langle f^{(s)}, v\rangle = v_2(x_s, z), \quad \forall v = (v_1, v_2) \in H^1(\Omega, \mathbb{C}^2), \tag{4.13}$$

in the elastic case.

We also consider a 3D acoustic medium. In this case the source is a Gaussian right hand side centered at $\bar{x} = (500, 500, 50)$ represented by the density

$$f(x) = \exp\left(-\frac{|x - \bar{x}|^2}{\sigma}\right), \tag{4.14}$$

with $\sigma = 50$.

Numerical solutions are computed on different meshes with different orders of discretization. After solving the linear system, the finite element solutions are projected onto a cartesian grid of fixed size $n_x \times n_z$ ($n_x \times n_y \times n_y$ in 3D) for comparison. We use the solution computed on the finer mesh with the higher polynomial degree as reference to compute finite element error. For a given finite element configuration, we denote by $u_{fem}^{(s)}$ the numerical solution obtained for the right hand side number s. Considering the reference solution $u_{ref}^{(s)}$ for each right hand side $1 \leq s \leq 100$, we evaluate the precision of a finite element configuration by

$$E = \sqrt{\frac{\sum_{s=1}^{100}\sum_{i=1}^{n_x}\sum_{j=1}^{n_z}|u_{ref}^s(x_i, z_j) - u_{fem}^s(x_i, z_j)|^2}{\sum_{s=1}^{100}\sum_{i=1}^{n_x}\sum_{j=1}^{n_z}|u_{ref}^s(x_i, z_j)|^2}}. \tag{4.15}$$

In the 3D example, there is a single right hand side so that the finite element error is measured by

$$E = \sqrt{\frac{\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_z}|u_{ref}^s(x_i, y_j, z_k) - u_{fem}^s(x_i, y_j, z_k)|^2}{\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_z}|u_{ref}^s(x_i, y_j, z_k)|^2}}. \tag{4.16}$$

In each test cases, we start by simulating the wave propagation on a fitting mesh (i.e., if the medium is given by a $n_x \times n_z$ grid, we use a mesh based on a $n_x \times n_z$ cartesian grid ($n_x \times n_y \times n_z$ in 3D)). In this case, the medium parameters are constant in each cell, so that the MMAm is not required and we use the classical FEM. The reference solution is the solution computed on the fitting mesh with the highest possible polynomial degree our computational resources allow.

4.4.2 2D Acoustic simulations with constant density: Overthrust model

We consider a vertical slice of SEG/EAGE Overthrust 3D velocity model [12]. The density $\rho = 1$ is assumed to be constant. The sound velocity is given as a 512×128 grid and illustrated on Figure 4.33.

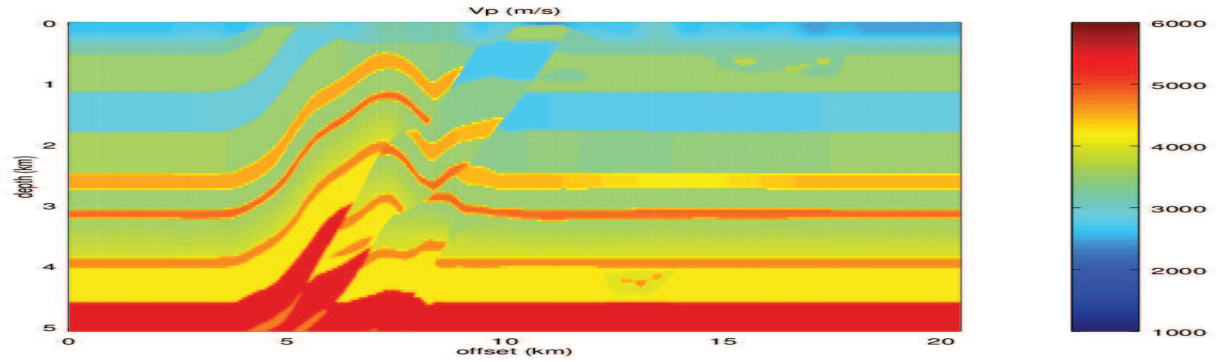


Figure 4.33: Overthrust P-velocity model

We also consider a smoothed version of the model, where a 200 m low-pass filter have been applied to the velocity model. The smoothed version of c_p is represented on Figure 4.34.

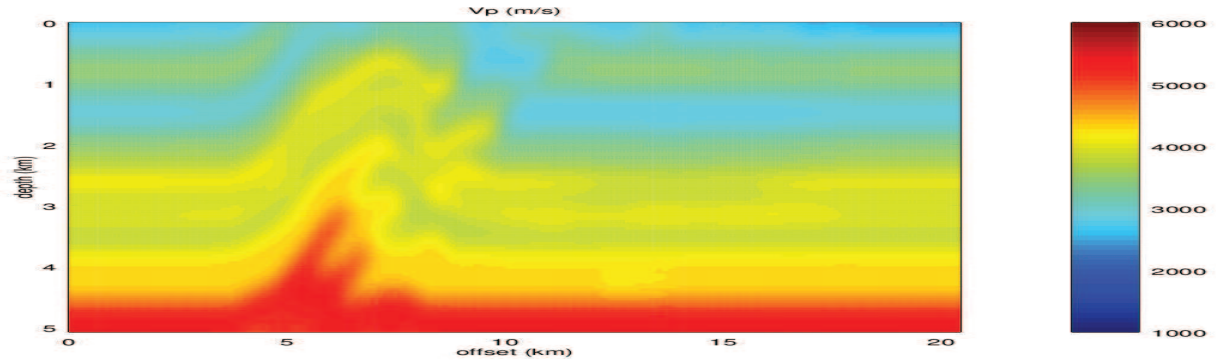


Figure 4.34: Smoothed Overthrust P-velocity model

We run acoustic simulations for frequencies $f = 5, 10, 15$ and 20Hz in the original and smoothed models. We use the numerical solution computed on the $h = 40$ m mesh with $p = 6$ as a reference solution to compute relative errors. Numerical solutions obtained on coarser mesh and/or with lower order polynomial, are compared to the reference solution from Tables 4.16 to Table 4.19.

We pay special attention to the case $f = 10$Hz which is detailed afterward.

We observe that, in terms of relative error, the results are very similar in the original and smoothed models. This is in strong agreement with the fact that high contrast in piecewise constant velocity model can be handle on non-fitting mesh without difficulty by the MMAm. Indeed, since the error levels are the same in the smoothed and original media, we can conclude that the error is mostly caused by dispersion.

Furthermore, Tables 4.16 to 4.19 confirm the interest of the MMAm. Indeed, it is clear that when high order polynomials are used, it is possible to obtain an accurate solution on a non-fitting mesh. When high order polynomials are used, we can obtain accurate results with a mesh step 8 times larger than the parameter grid for $f = 5$Hz, 4 times larger for $f = 10$Hz and $f = 15$Hz and 2 times larger for $f = 20$Hz.

h	p=1	p=2	p=3	p=4	p=5	p=6
40	60.6	0.49	0.32	0.22	0.16	0
80	119	3.76	0.69	0.46	0.42	0.41
160	116	41.9	2.41	1.44	0.98	0.73
320	107	119	46.6	5.67	2.38	1.91

h	p=1	p=2	p=3	p=4	p=5	p=6
40	48.0	0.49	0.33	0.23	0.17	0
80	110	2.81	0.71	0.48	0.44	0.44
160	116	31.2	2.19	1.50	1.03	0.76
320	106	110	32.7	3.97	2.41	1.99

Table 4.16: Relative error (%) for $f = 5$Hz in the original (left) and smoothed (right) acoustic Overthrust models

h	p=1	p=2	p=3	p=4	p=5	p=6
40	127	6.72	0.25	0.17	0.12	0
80	120	74.8	2.55	0.39	0.32	0.32
160	115	119	78.3	7.91	1.06	0.61
320	100	107	124	127	77.0	19.2

h	p=1	p=2	p=3	p=4	p=5	p=6
40	120	4.83	0.27	0.18	0.13	0
80	123	57.8	1.69	0.39	0.35	0.35
160	107	113	58.1	4.82	0.93	0.64
320	100	108	123	116	51.4	9.95

Table 4.17: Relative error (%) for $f = 10$Hz in the original (left) and smoothed (right) acoustic Overthrust models

h	p=1	p=2	p=3	p=4	p=5	p=6
40	129	45.8	0.72	0.18	0.13	0
80	112	124	33.3	1.75	0.38	0.35
160	101	121	124	99.5	18.8	2.62
320	104	102	118	125	123	120

h	p=1	p=2	p=3	p=4	p=5	p=6
40	131	34.2	0.52	0.18	0.13	0
80	112	116	22.5	0.93	0.35	0.35
160	101	118	120	77.5	10.5	1.28
320	107	102	121	133	119	111

Table 4.18: Relative error (%) for $f = 15$Hz in the original (left) and smoothed (right) acoustic Overthrust models

h	p=1	p=2	p=3	p=4	p=5	p=6
40	124	122	5.15	0.23	0.15	0
80	113	132	121	16.1	1.30	0.56
160	100	110	126	125	115	36.8
320	100	100	106	110	125	125

h	p=1	p=2	p=3	p=4	p=5	p=6
40	123	108	3.47	0.22	0.16	0
80	113	134	106	9.32	0.64	0.41
160	100	114	128	116	95.0	20.1
320	100	100	107	109	126	122

Table 4.19: Relative error (%) for $f = 20$Hz in the original (left) and smoothed (right) acoustic Overthrust models

Frequency $f = 10$Hz

We detail the case $f = 10$Hz. Figures 4.35 to 4.38 present convergence curves of the MMAm with a high number of subcells and of the standard FEm with an averaged parameter. We use the same submesh to compute the MMAm coefficient and to compute the averaged parameter. Hence, if the submesh of the cell K is denoted by $\mathcal{T}_\epsilon = \{K_l\}_{l=1}^m$ we have

$$\int_K \frac{1}{c^2}\phi_i\phi_j \simeq \sum_{l=1}^m \frac{1}{c_l^2}\int_{K_l}\phi_i\phi_j,$$

for the MMAm and

$$\int_K \frac{1}{c^2}\phi_i\phi_j \simeq \left(\frac{1}{|K|}\int_K \frac{1}{c^2}\right)\int_K \phi_i\phi_j, \quad \int_K \frac{1}{c^2} \simeq \sum_{l=1}^m \frac{1}{c_l^2}|K_l|$$

for the FEm.

We observe that the MMAm is overconvergent for large h. Indeed, the fit in h^α always give an $\alpha > p+1$ for the MMAm both in the original and smoothed media. This result is well-known for the standard FEm in homogeneous media: in the transition zone between the pre-asymptotic and the asymptotic range, the finite element solution is less accurate than the best approximation, but has an improved convergence rate. We refer the reader to the numerical examples of Babuška and Ihlenburg [59].

On the other hand, the convergence rate of the FEm solution is always lower than expected, even in the smoothed medium: the fit in h^α gives $\alpha < p+1$. We interpret that this loss of convergence is due to the poor approximation of the medium parameter by the averaging technique. In the smoothed medium with $p = 3$ and $p = 4$ the convergence rate is between 3 and 4. In all other test cases, the convergence rate is less than 2. We interpret this convergence rate by the fact that the convergence of the averaged parameter to the true parameter is only linear and is a limitation for the optimal convergence of the FEm.

In terms of accuracy, the MMAm and FEm provide comparable results only in the smoothed medium when $p = 3$. In this case, the mesh step is relatively small and the averaged parameter is a good approximation of the smoothed parameter. For all the other cases, the MMAm provides a slightly improved accuracy. In the smoothed medium, the MMAm solution, is between 2 to 20 times more accurate than the FEm solution on the

same mesh. In the original medium, we are never able to obtain less than 10% accuracy with the FEm on all meshes consider, while a 1% accuracy is achieved by the MMAm.

In terms of performance, the MMAm and the FEm give similar results for $p = 3$ in the smoothed medium. In every other configuration, the MMAm outperforms the FEm. Assuming that static condensation is used, we can give the following numbers:

- In the smoothed medium, the FEm is able to achieve a 10% accuracy for $p = 3, 4$ and 5. For $p = 3$, the number of degrees of freedom required is similar. However, for $p = 4$, the number of degrees of freedom required is reduced from 51×10^3 to 33×10^3. The number of degrees of freedom is reduced from 59×10^3 to 21×10^3 for $p = 5$.

- In the original medium, the only case where the FEm achieves an accuracy close to 10% is $p = 3$. In this configuration, the number of degrees of freedom required is 201×10^3 for the FEm against only 66×10^3 for the MMAm.

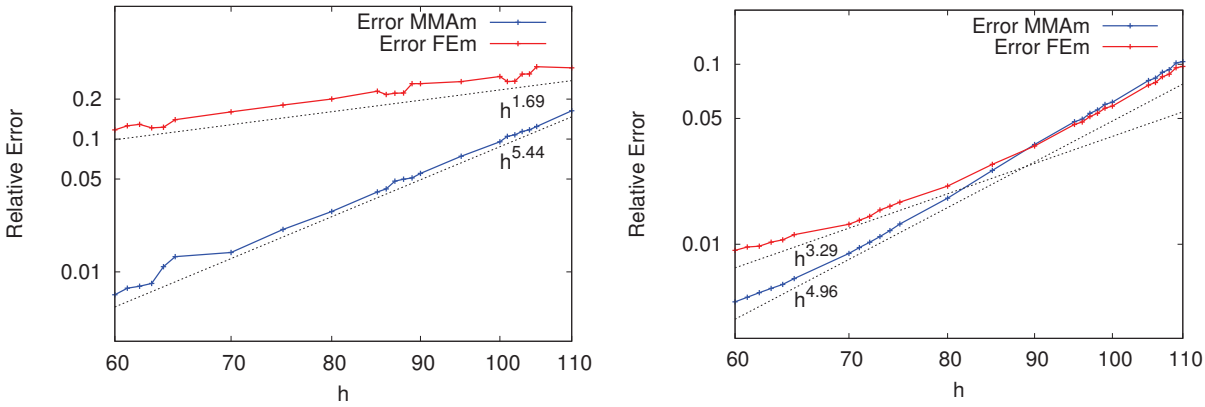


Figure 4.35: Convergence curves for $p = 3$ in the original (left) and smoothed (right) media

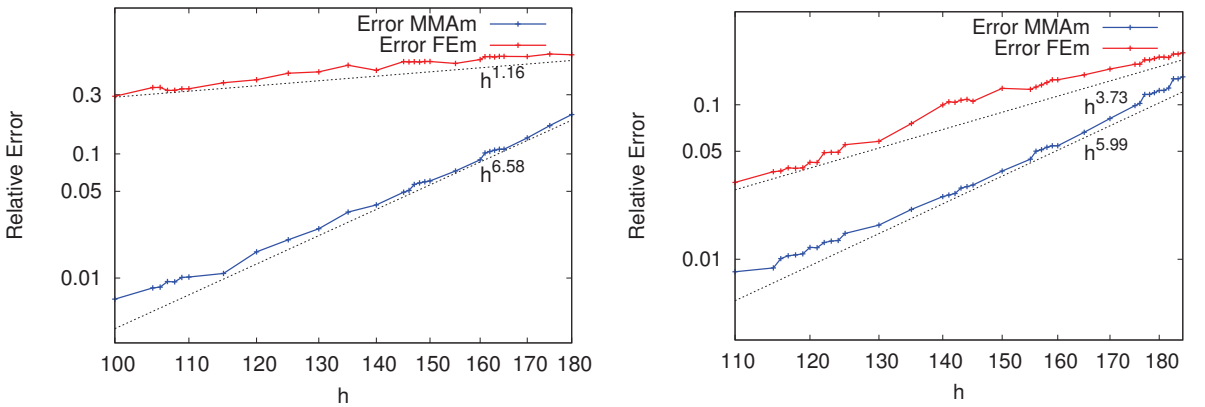


Figure 4.36: Convergence curves for $p = 4$ in the original (left) and smoothed (right) media

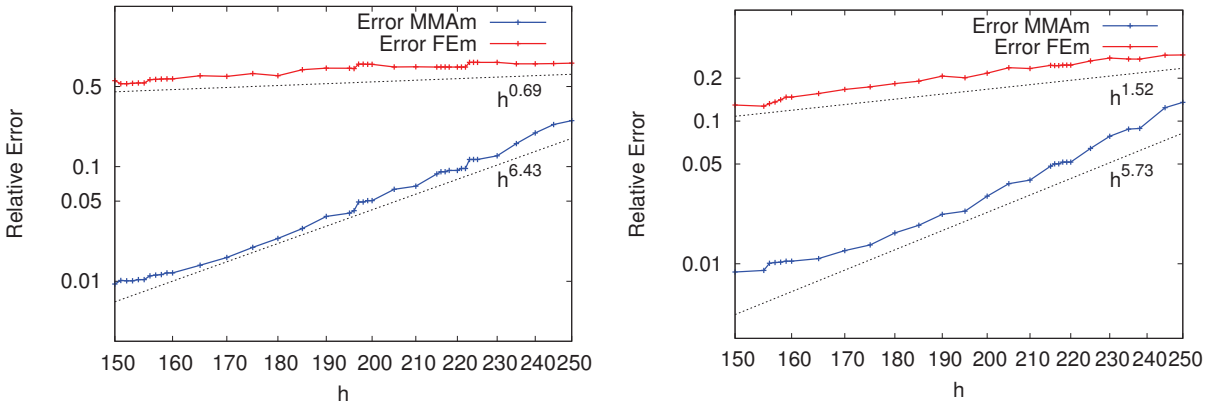


Figure 4.37: Convergence curves for $p = 5$ in the original (left) and smoothed (right) media

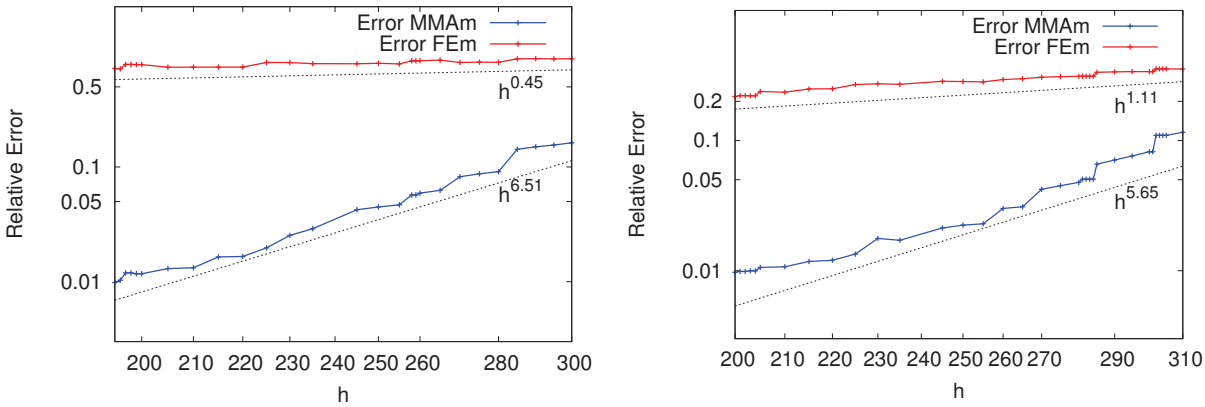


Figure 4.38: Convergence curves for $p = 6$ in the original (left) and smoothed (right) media

Figures 4.39 and 4.40 present the number of degrees of freedom required to obtain different levels of accuracy. The number of floating-point numbers required to represent the finite element matrix is also represented (since the matrix is symmetric, it is the number of non-zero elements in the upper triangle).

When static condensation is not used, the number of degrees of freedom required is always reduced when p increases from 2 to 6. However, if the number of non-zero elements reduces from $p = 2$ to 4, it increases from $p = 4$ to 6. This observation holds for all accuracy levels considered.

When static condensation is used, the numbers of degrees of freedom and non-zero elements are reduced when p is increased.

We conclude that high order MMAm is an interesting method for constant density cases if static-condensation is used. In the case where static condensation is not used, the efficiency of high order polynomials might depend on the sparse linear solver used. However, increasing the order of discretization seems to be interesting at least up to $p = 4$.

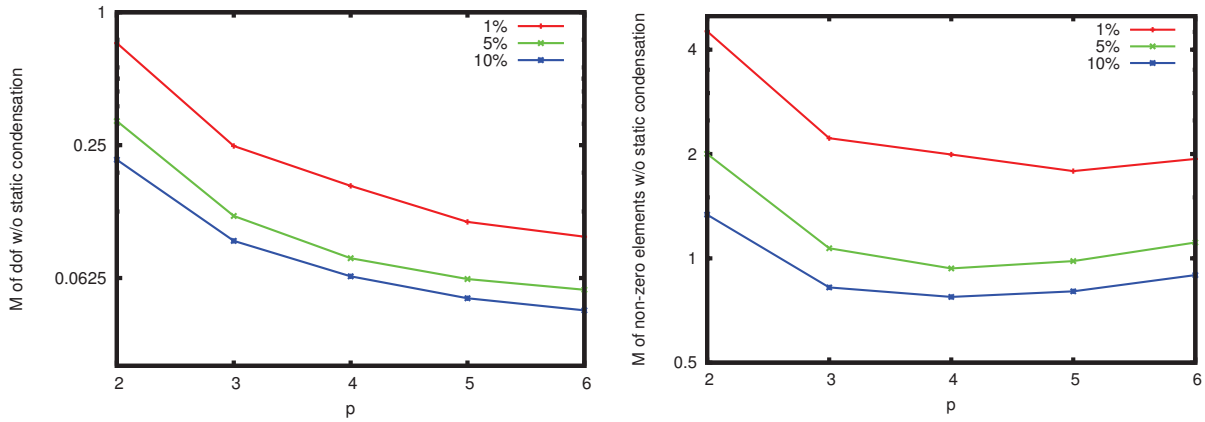


Figure 4.39: Number of degrees of freedom and non-zero elements in the linear system without static condensation for $f = 10$Hz

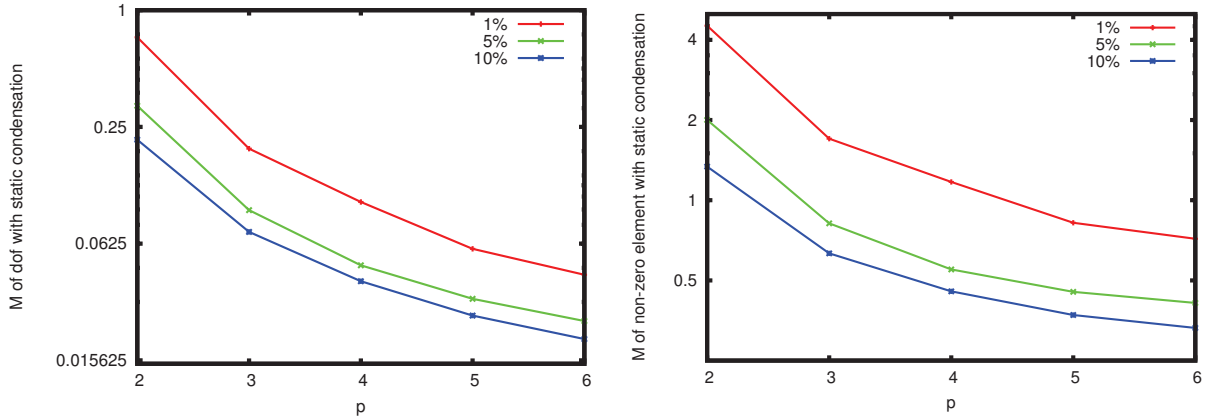


Figure 4.40: Number of degrees of freedom and non-zero elements in the linear system with static condensation for $f = 10$Hz

Adaptive meshes

One advantage of finite element discretizations over finite difference discretizations is the possibility to locally adapt the mesh. Indeed, finite element discretizations do not have to rely on a cartesian grid but are rather based on a mesh made of triangles. Finite element methods can thus naturally handle different discretization steps in the same simulation as long as it is possible to produce a mesh.

This is of particular importance in Geophysics because it is possible to guess where the mesh needs to be refined in advance. Indeed, it is well-known that the crucial requirement for the finite element method to be accurate is to ensure that they are enough points per wavelength. The wavelength of the solution is directly linked to the wave velocity, which is known before the simulation. Since the velocity can vary from 1000 ms^{-1} in shallow

regions to 5000 m.s^{-1} in depth, it might be of interest to have a mesh 5 times more refined near the surface than in depth.

More precisely, assume we have a mesh $\mathcal{T}_h$ and that the velocity is constant in each cell. A cell $K \in \mathcal{T}_h$ is associated with a wave velocity c_K. If h_K is the diameter of K, the number of points per wavelength when using finite elements of degree p can be expressed as

$$N_\lambda^K \simeq \frac{pc_K}{fh_K}. \tag{4.17}$$

If $h_K = h$ is constant (for example, using a cartesian grid), the number of points per wavelength varies from one cell to another, depending on the velocity value c_K. On the other hand, if one wishes to keep the number of points per wavelength N_λ to a constant value in all cells, the cell diameters need to be selected so that

$$h_K \simeq \frac{pc_K}{fN_\lambda}. \tag{4.18}$$

Thus, the ideal cell size at the point $x \in \Omega$ would be

$$h(x) \simeq \frac{pc(x)}{fN_\lambda}. \tag{4.19}$$

Actually (4.19) defines a "metric" than can be used to produce adaptive meshes [68]. In this section, we propose to use the software BL2D [66] to produce and use such meshes. In order to compare with cartesian grid results, we define a "minimum step" h_{min} and define the mesh metric by

$$h(x) = \frac{c(x)}{c_{min}} h_{min}, \quad x \in \Omega. \tag{4.20}$$

Definition (4.20) ensures that the number of points per wavelength is the same for all cells. Furthermore, the smallest cell in the mesh should be of length h_{min}.

In the following, we compare the adaptive mesh with minimum mesh step h_{min} to the cartesian mesh with constant mesh step $h = h_{min}$. The idea behind this choice is that the strongest requirement on the mesh is where the velocity is minimal. When using a cartesian mesh, the mesh step is the same everywhere and the mesh is unnecessary refined where the wavespeed is higher. We thus expect that the adaptive and cartesian meshes yield the same precision, the adaptive mesh being optimal and therefore, less costly to use.

Figures 4.41 to 4.44 show adaptive and cartesian meshes for $h_{min} = 80$ and 160 m. The number of cells in the adaptive meshes is approximately divided by two as compared to the cartesian meshes.

Of course, since adaptive meshes are non-cartesians, they are obviously non-fitting (since the velocity parameter is given on a cartesian grid). We therefore use the MMAm with 1024 subcells to approximate the medium.

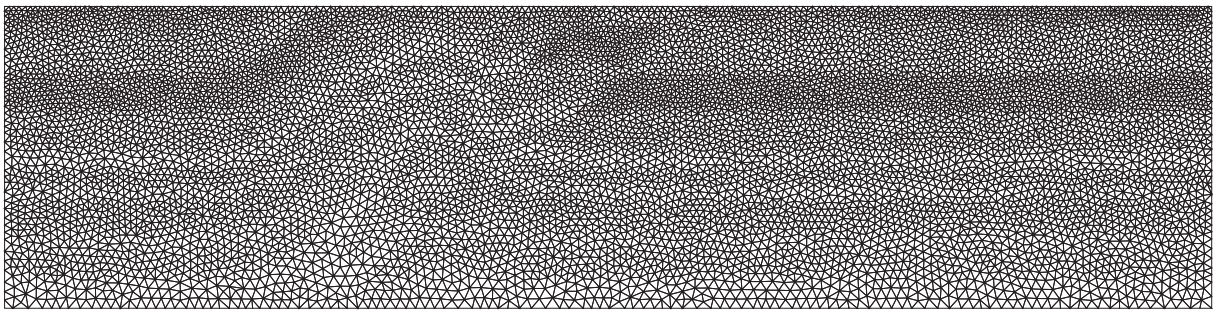


Figure 4.41: Adaptive mesh with $h_{min} = 80$ m (17383 triangles)

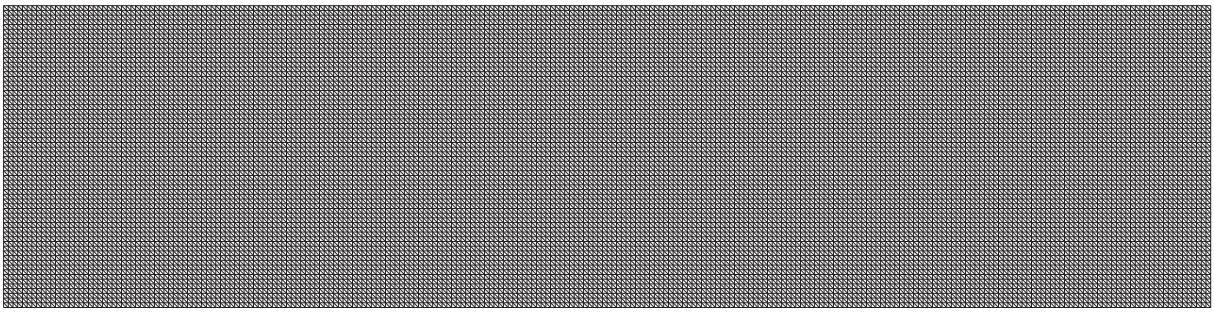


Figure 4.42: Cartesian mesh with $h = 80$ m (32768 triangles)

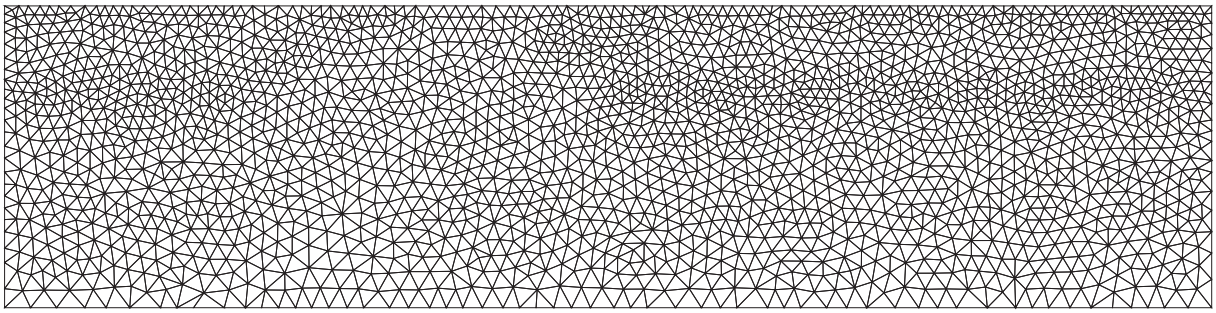


Figure 4.43: Adaptive mesh with $h_{min} = 160$ m (4315 triangles)

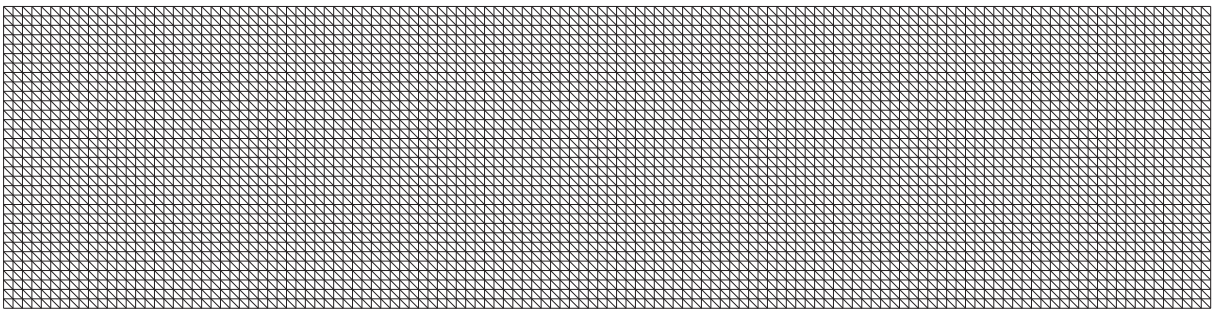


Figure 4.44: Cartesian mesh with $h = 160$ m (8192 triangles)

Numerical results are presented in Tables 4.20 to 4.23. In every test case, the error level is comparable for the adaptive and the cartesian meshes. It confirms the idea that the cartesian mesh is over-refined in some regions and that the bottlenecks are the regions with minimal velocity.

In terms of performance, if we compare the case $h_{min} = 320$ m and $p = 6$ (2.03% error for $f = 5$Hz) with static condensation, the number of degrees of freedom is 15589 for the cartesian mesh against only 8844 for the adaptive mesh. The number of non-zero elements in the finite element matrix is also reduced from 256312 to 144500 by using the adaptive mesh.

For the case $h_{min} = 80$ m and $p = 2$ (4.62% error for $f = 5$Hz), the number of degrees of freedom is reduced from 64897 to 35200 and the number of non-zero elements is reduce from 403216 to 218371.

The number of cells is thus almost halved when adaptive meshes are used. It follows that the number of degrees of freedom and the filling of the linear system are divided by a factor close to 2 as well.

In order to show the interest of the MMAm, let us consider again the cases $h_{min} = 320$ m with $p = 6$ and $h_{min} = 80$ m with $p = 2$ for $f = 5$Hz. If we use an averaged parameter with the standard FEm, the relative error is 64.7% for $h_{min} = 320$ m and $p = 6$ (against 2.03% using the MMAm). For the case $h_{min} = 80$ m and $p = 2$, the relative error is 10.7% for the standard FEm with an averaged parameter (against 4.62% using the MMAm).

h_{min}	p=1	p=2	p=3	p=4	p=5	p=6	h	p=1	p=2	p=3	p=4	p=5	p=6
40	72.2	0.56	0.34	0.27	0.19	0.14	40	60.6	0.49	0.32	0.22	0.16	0
80	123	4.62	0.70	0.41	0.37	0.36	80	119	3.76	0.69	0.46	0.42	0.41
160	120	54.2	2.93	1.44	1.00	0.76	160	116	41.9	2.41	1.44	0.98	0.73
320	110	126	72.3	9.13	2.68	2.03	320	107	119	46.6	5.67	2.38	1.91

Table 4.20: Relative error (%) for $f = 5$Hz with the adaptive (left) and cartesian (right) meshes

h_{min}	p=1	p=2	p=3	p=4	p=5	p=6	h	p=1	p=2	p=3	p=4	p=5	p=6
40	135	9.48	0.29	0.22	0.17	0.14	40	127	6.72	0.25	0.17	0.12	0
80	118	90.5	3.56	0.44	0.36	0.35	80	120	74.8	2.55	0.39	0.32	0.32
160	125	135	101	11.3	1.60	0.95	160	115	119	78.3	7.91	1.06	0.61
320	100	122	131	140	111	35.8	320	100	107	124	127	77.0	19.2

Table 4.21: Relative error (%) for $f = 10$Hz with the adaptive (left) and cartesian (right) meshes

h_{min}	p=1	p=2	p=3	p=4	p=5	p=6	h	p=1	p=2	p=3	p=4	p=5	p=6
40	125	59.7	1.18	0.29	0.24	0.22	40	129	45.8	0.72	0.18	0.13	0
80	118	142	43.1	2.22	0.65	0.58	80	112	124	33.3	1.75	0.38	0.35
160	100	122	138	120	25.6	3.94	160	101	121	124	99.5	18.8	2.62
320	101	105	122	132	133	137	320	104	102	118	125	123	120

Table 4.22: Relative error (%) for $f = 15$Hz with the adaptive (left) and cartesian (right) meshes

h_{min}	p=1	p=2	p=3	p=4	p=5	p=6	h	p=1	p=2	p=3	p=4	p=5	p=6
40	124	131	7.44	0.49	0.35	0.32	40	124	122	5.15	0.23	0.15	0
80	164	143	132	19.6	1.84	0.99	80	113	132	121	16.1	1.30	0.56
160	100	154	139	131	132	46.0	160	100	110	126	125	115	36.8
320	100	100	112	127	141	132	320	100	100	106	110	125	125

Table 4.23: Relative error (%) for $f = 20$Hz with the adaptive (left) and cartesian (right) meshes

4.4.3 2D Acoustic simulations with non-constant density: Marmousi II model

In this subsection, we consider the Marmousi II model [71]. The P-velocity and the density are represented by a 2048×512 grid with 5 m step as depicted on Figures 4.45 and 4.46. We consider a smoothed version of the model by applying a 50 m low-pass filter. The smoothed versions of the P-velocity and density are given on Figures 4.47 and 4.48.

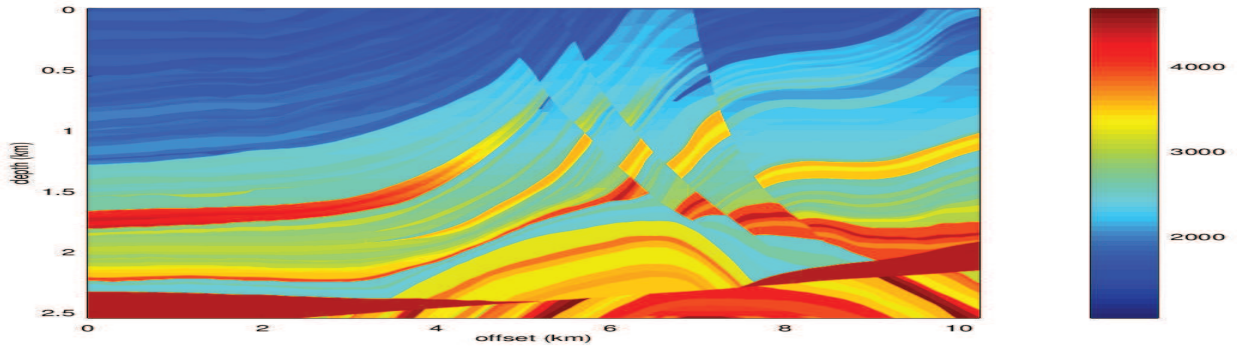


Figure 4.45: Marmousi II P-velocity model

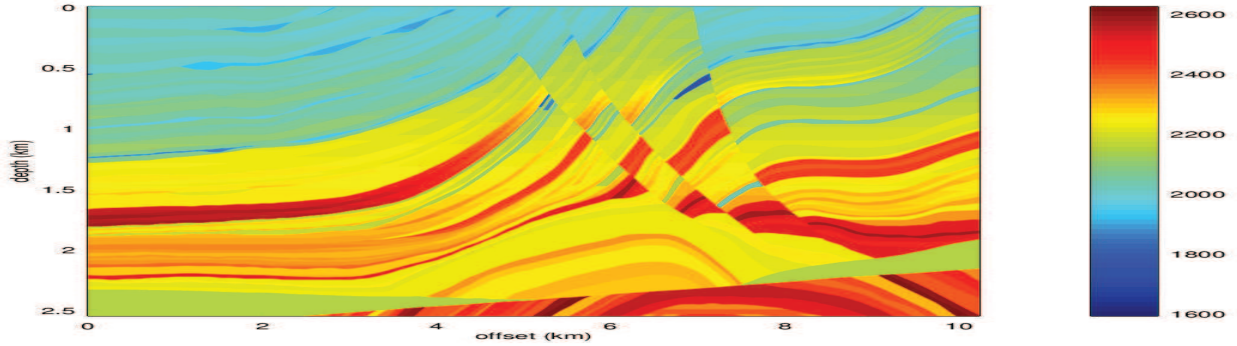


Figure 4.46: Marmousi II density model

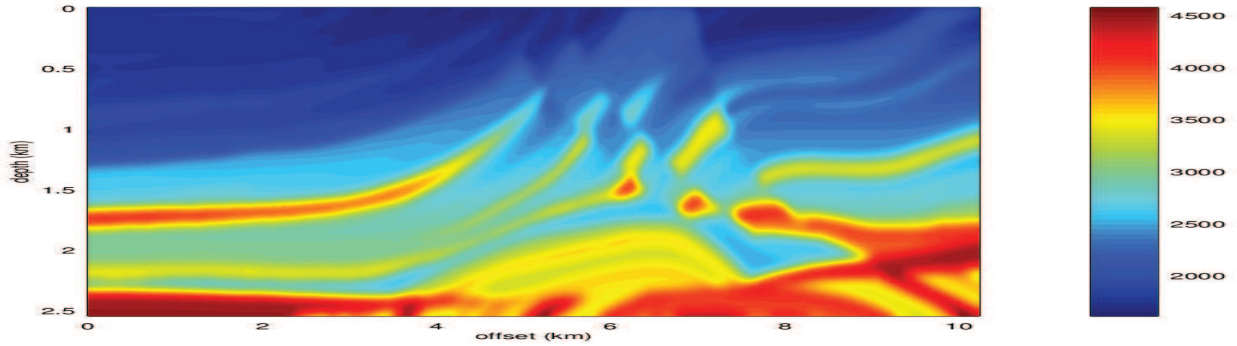


Figure 4.47: Marmousi II smoothed P-velocity model

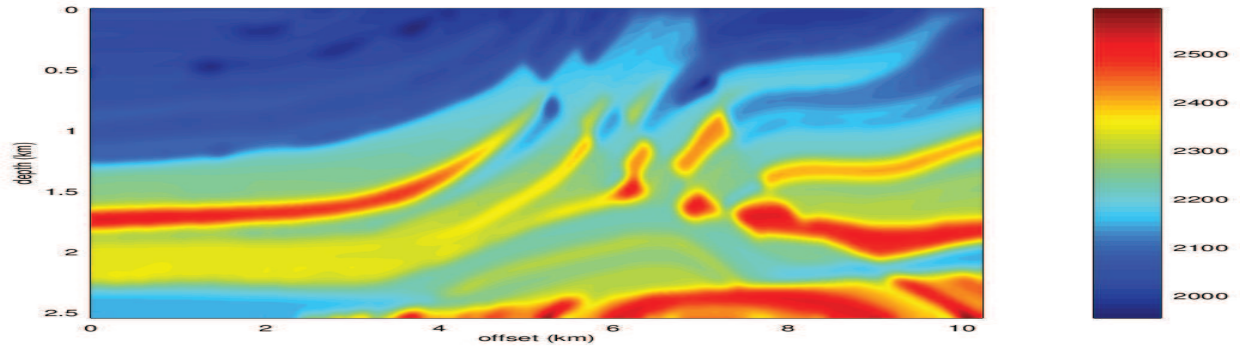


Figure 4.48: Marmousi II smoothed density model

Tables 4.24 to 4.28 present results for frequencies $f = 5, 10, 15, 20$ and 30Hz. For all frequencies, we are able to obtain accurate results on non-fitting meshes if high order polynomials are used. Like in the constant-density case, we obtain similar results in the original and smoothed media. This might be explained by the fact that the error is mostly due to dispersion, and that the impact of small-scale heterogeneities is small.

h	p=1	p=2	p=3	p=4	p=5	p=6
5	2.55	0.13	0.00	x	x	x
10	10.1	0.33	0.19	0.11	x	x
20	38.7	0.73	0.59	0.44	x	x
40	104	2.09	0.89	0.75	0.66	x
80	119	20.0	1.66	1.25	1.05	0.96
160	109	112	19.0	2.93	1.80	1.53
320	101	112	123	68.2	14.5	4.39

h	p=1	p=2	p=3	p=4	p=5	p=6
5	2.34	0.13	0.00	x	x	x
10	9.18	0.33	0.19	0.11	x	x
20	34.6	0.73	0.59	0.44	x	x
40	95.7	1.79	0.89	0.75	0.65	x
80	117	17.5	1.54	1.25	1.05	0.96
160	110	105	15.6	2.45	1.75	1.52
320	100	112	115	57.9	11.2	3.19

Table 4.24: Relative error (%) for $f = 5$Hz in the original (left) and smoothed (right) Marmousi II models

h	p=1	p=2	p=3	p=4	p=5	p=6
5	22.6	0.12	0.00	x	x	x
10	77.8	0.40	0.19	0.11	x	x
20	129	3.46	0.58	0.43	x	x
40	129	43.4	1.42	0.75	0.64	x
80	109	127	38.1	2.97	1.15	0.98
160	100	116	122	105	27.3	5.28
320	100	101	106	118	125	120

h	p=1	p=2	p=3	p=4	p=5	p=6
5	19.3	0.13	0.00	x	x	x
10	68.7	0.37	0.19	0.11	x	x
20	122	2.76	0.59	0.44	x	x
40	129	36.5	1.11	0.76	0.66	x
80	109	120	30.4	1.98	1.07	0.98
160	100	116	115	97.0	20.9	3.15
320	101	101	107	118	119	114

Table 4.25: Relative error (%) for $f = 10$Hz in the original (left) and smoothed (right) Marmousi II models

h	p=1	p=2	p=3	p=4	p=5	p=6
5	76.3	0.21	0.00	x	x	x
10	135	1.90	0.27	0.15	x	x
20	135	26.0	0.90	0.62	x	x
40	123	129	14.6	1.28	0.94	x
80	110	129	128	51.6	5.97	1.76
160	150	112	117	125	124	107
320	100	100	105	107	112	120

h	p=1	p=2	p=3	p=4	p=5	p=6
5	70.6	0.21	0.00	x	x	x
10	134	1.64	0.28	0.16	x	x
20	133	22.7	0.89	0.65	x	x
40	123	130	12.1	1.15	0.97	x
80	116	131	125	42.6	4.03	1.49
160	107	112	117	121	121	99.8
320	100	100	105	107	111	123

Table 4.26: Relative error (%) for $f = 15$Hz in the original (left) and smoothed (right) Marmousi II models

h	p=1	p=2	p=3	p=4	p=5	p=6
5	127	0.55	0.00	x	x	x
10	134	7.71	0.37	0.20	x	x
20	127	90.2	2.41	0.85	x	x
40	121	134	80.3	5.23	1.34	x
80	101	124	132	124	57.4	9.06
160	101	106	116	117	128	122
320	100	102	100	104	111	108

h	p=1	p=2	p=3	p=4	p=5	p=6
5	126	0.50	0.00	x	x	x
10	132	6.75	0.39	0.22	x	x
20	128	83.7	2.04	0.91	x	x
40	123	133	72.5	3.93	1.37	x
80	101	126	130	124	46.8	5.80
160	196	101	119	118	127	123
320	100	274	100	105	109	108

Table 4.27: Relative error (%) for $f = 20$Hz in the original (left) and smoothed (right) Marmousi II models

h	p=1	p=2	p=3	p=4	p=5	p=6
5	139	3.35	0.00	x	x	x
10	138	47.9	0.66	0.21	x	x
20	116	135	26.2	1.23	x	x
40	113	128	131	82.6	9.10	x
80	101	117	118	134	126	123
160	100	100	102	109	114	124
320	100	100	100	100	104	103

h	p=1	p=2	p=3	p=4	p=5	p=6
5	135	2.65	0.00	x	x	x
10	141	38.4	0.52	0.21	x	x
20	116	128	20.1	1.00	x	x
40	112	128	122	67.2	6.14	x
80	101	114	116	130	119	116
160	100	100	102	106	115	121
320	100	100	100	100	103	104

Table 4.28: Relative error (%) for $f = 30$Hz in the original (left) and smoothed (right) Marmousi II models

Frequency $f = 20$Hz

We focus on the case $f = 20$Hz. In order to provide a comparison with the standard FEm, we propose a method where an average bulk modulus $\tilde{\kappa}$ is used instead of the original value. We do not homogenize the density. Indeed, homogenizing the density is more complicated than just averaging. We thus keep the original value and use the MMAm formula for the density.

Hence, we use the following formula to compute the coefficients in the MMAm:

$$\int_K \frac{1}{\kappa}\phi_i\phi_j \simeq \sum_{l=1}^{m} \frac{1}{\kappa_l}\int_{K_l} \phi_i\phi_j, \quad \int_K \frac{1}{\rho}\nabla\phi_i\cdot\nabla\phi_j \simeq \sum_{l=1}^{m} \frac{1}{\rho_l}\int_{K_l} \nabla\phi_i\cdot\nabla\phi_j.$$

To compare with the standard FEm, we use the following quadrature formula to compute the mass matrix coefficient

$$\int_K \frac{1}{\kappa}\phi_i\phi_j \simeq \left(\frac{1}{|K|}\int_K \frac{1}{\kappa}\right)\int_K \phi_i\phi_j, \quad \int_K \frac{1}{\kappa} \simeq \sum_{l=1}^{m} \frac{1}{\kappa_l}|K_l|,$$

and keep the MMAm formula for the stiffness matrix:

$$\int_K \frac{1}{\rho}\nabla\phi_i\cdot\nabla\phi_j \simeq \sum_{l=1}^{m} \frac{1}{\rho_l}\int_{K_l} \nabla\phi_i\cdot\nabla\phi_j.$$

Convergence curves for $p = 3, 4, 5$ and 6 are presented on Figures 4.49 to 4.52. In every considered configuration, the MMAm is more accurate than the FEm.

We provide two fits in h^α for every convergence curve. The first fit is carried out on the larger mesh steps and gives $\alpha > p + 1$. The second fit is carried out in small mesh steps and gives $\alpha < p + 1$. We interpret these two different convergence rates as the separation between the pre-asymptotic range where the method is over-convergent and the asymptotic range where the convergence rate is suboptimal. Again, a similar behaviour has already been observed by Ihlenburg and Babuška [59] for 1D homogeneous problem with the standard FEm. The difference here is that since the solution is not H^2, the rate of convergence of the best approximation is suboptimal (less than $p + 1$ in the L^2 norm). This is the reason why the convergence rate is not optimal in the asymptotic range.

In particular, for the cases $p = 5$ and 6, the asymptotic convergence rate is less than 1. This can be explained by the fact that the density varies quickly or is discontinuous. When the density is discontinuous, the solution does not belong to $H^2(\Omega)$ which explains that the convergence rate is not even 1. On the other hand, the convergence is good in the pre-asymptotic range.

The MMAm slightly outperforms the FEm in the original medium. The FEm never reaches 10% accuracy on all considered meshes, while the MMAm is able to achieve 1% relative L^2 error.

The MMAm and the FEm gives similar error levels in the smoothed medium for the case $p = 3$ (see Figure 4.49). For $p > 3$, the MMAm solution is at least four times more accurate than the FEm solution on the same mesh.

We can say that the MMAm always improves the quality of the numerical solution compared to the FEm while the linear system to solve has exactly the same size and stencil. It is also clear that the MMAm reduces the computational cost for a given accuracy, and we can give similar numbers than in the constant density case. On the other hand, the asymptotic convergence rate of high order elements is very poor, which can be explained

by the fact that the density is discontinuous or highly oscillating. We also make this observation in performance assessments.

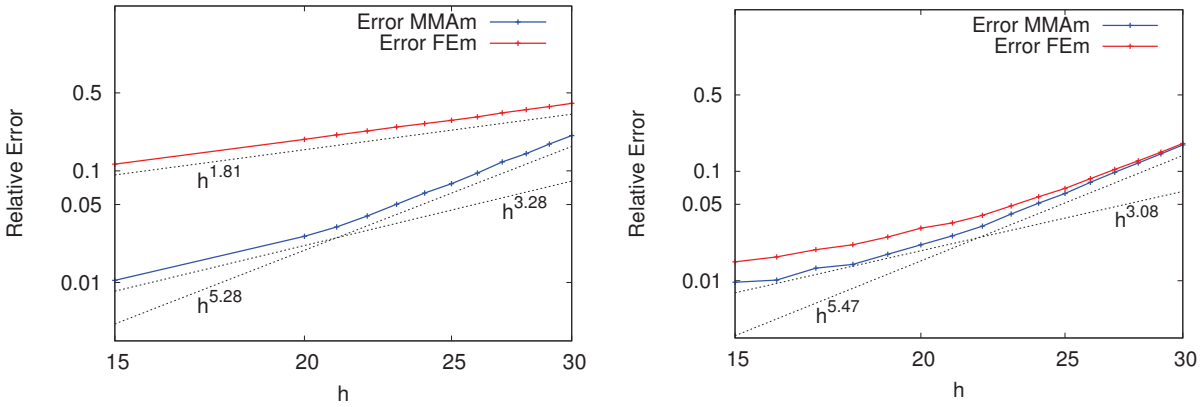


Figure 4.49: Convergence curves for $p = 3$ in the original (left) and smoothed (right) media

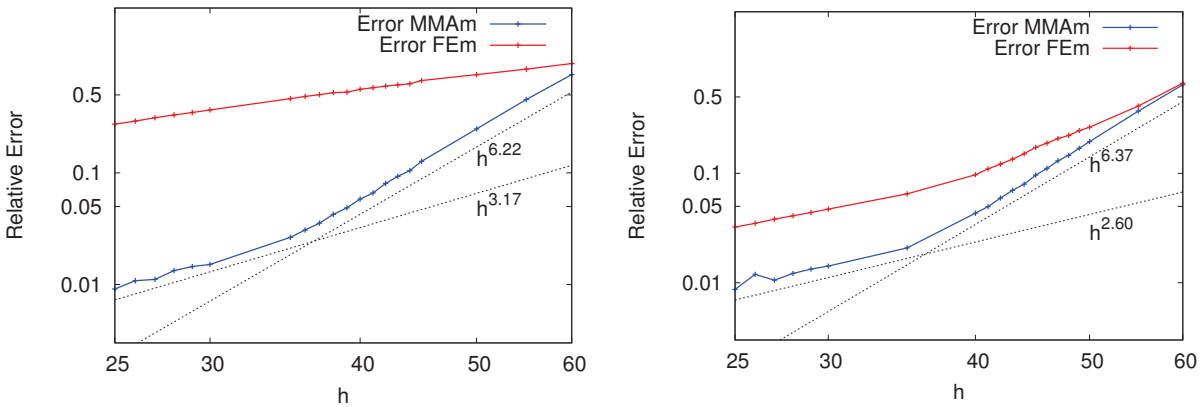


Figure 4.50: Convergence curves for $p = 4$ in the original (left) and smoothed (right) media

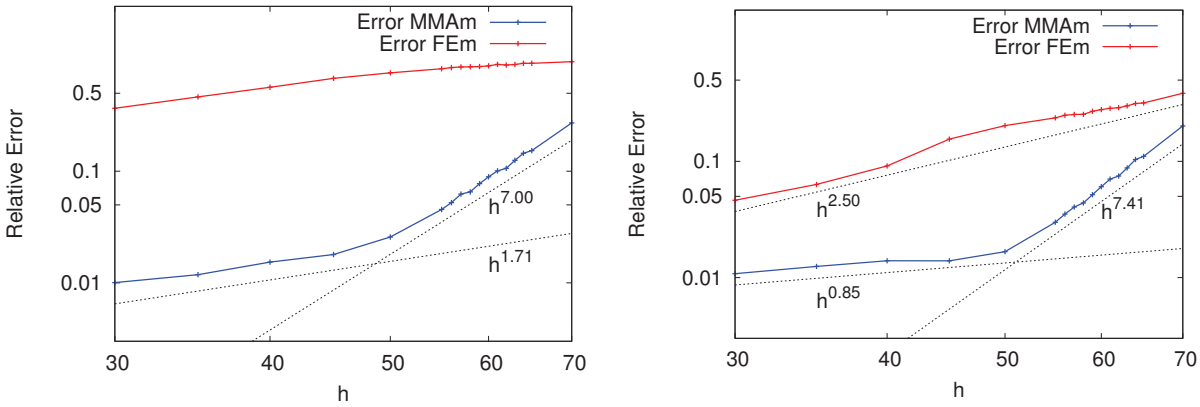


Figure 4.51: Convergence curves for $p = 5$ in the original (left) and smoothed (right) media

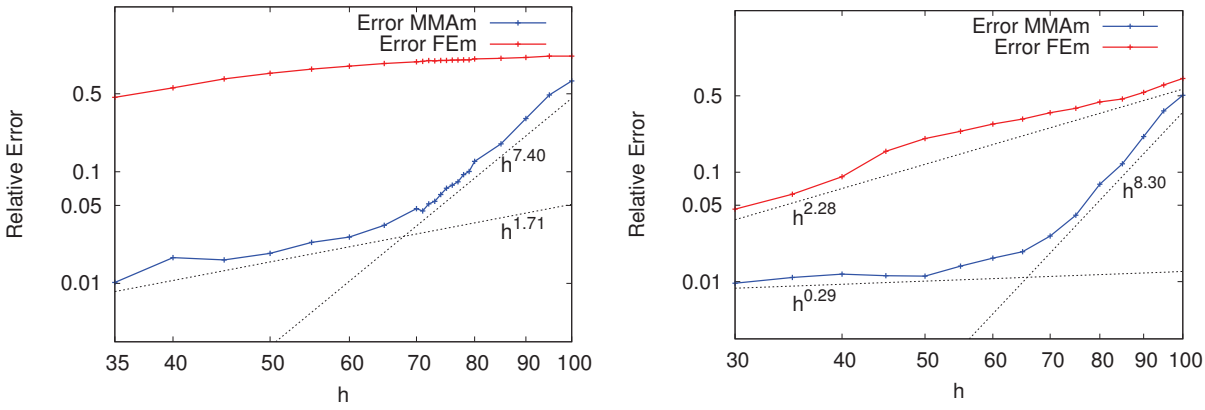


Figure 4.52: Convergence curves for $p = 6$ in the original (left) and smoothed (right) media

Performance assessments of the MMAm are depicted on Figures 4.53 and 4.54. We see that the size and the filling of the linear system are reduced for $p = 4$ compared to $p = 3$ (it is also smaller than for $p = 1$ and 2 though this is not represented here). The case of high order discretizations with $p \geq 5$ needs to be discussed more precisely.

We see that the size and filling of the linear system are reduced for $p = 5$ and 6 if static condensation is used and the desired accuracy level is 5 or 10%.

However, it is clear that $p = 5$ and $p = 6$ elements are more costly than $p = 4$ elements if the desired accuracy is 1%. This is agreement with the observation that the asymptotic convergence rate is poor for $p = 5$ and 6 (see Figures 4.51 and 4.52). Also, it can be explained by the fact that since the density is either discontinuous or fastly variating, high order methods do not bring additional precision to the best approximation. Actually, the same mesh step needs to be used to obtain a 1% accuracy for $p = 4, 5$ and 6, so that $p = 4$ discretization is cheaper.

As a conclusion, we might say that $p = 4$ elements seem to be the best choice for the non constant-density test-case considered here. In this case, the MMAm makes it possible

to use a coarse mesh, while preserving the quality of the numerical solution. If the standard FEm with parameter averaging is used, we might expect the solution to be at least 4 times less accurate.

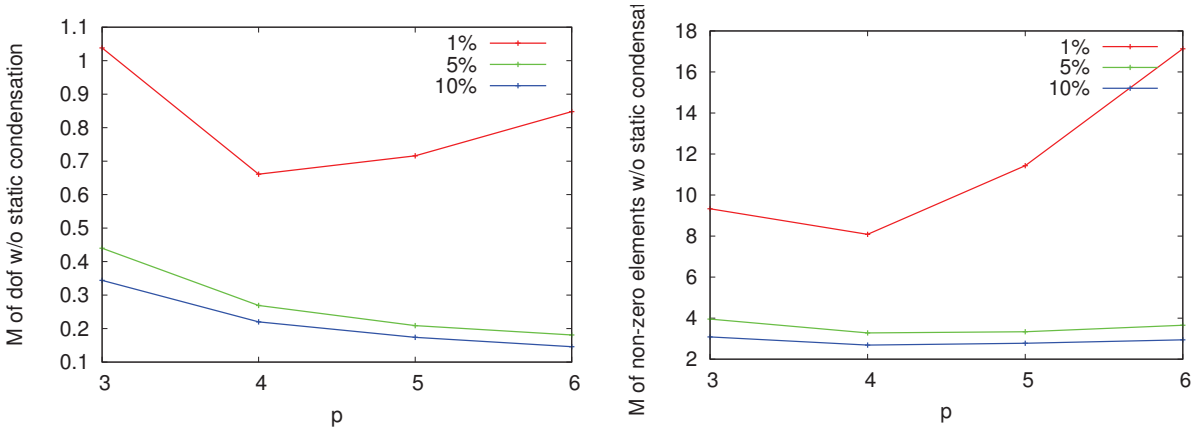


Figure 4.53: Number of degrees of freedom and non-zero elements in the linear system without static condensation for $f = 20$hz

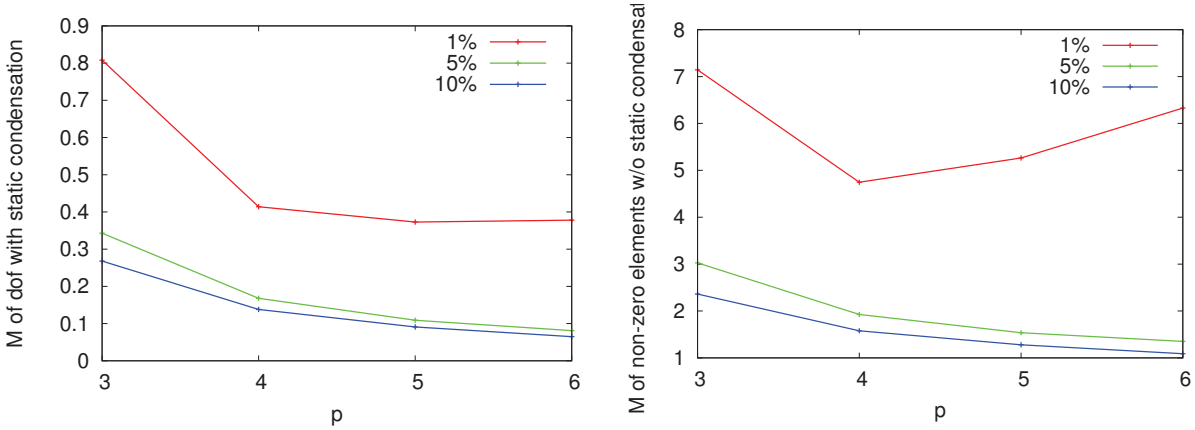


Figure 4.54: Number of degrees of freedom and non-zero elements in the linear system with static condensation for $f = 20$hz

4.4.4 3D Acoustic simulations with constant density: Louro Model

In this subsection, we consider a 3D velocity model. The density $\rho = 1$ is assumed to be constant. The velocity is given by a $64 \times 64 \times 32$ cartesian grid with 20 m step which is depicted on Figure 4.55.

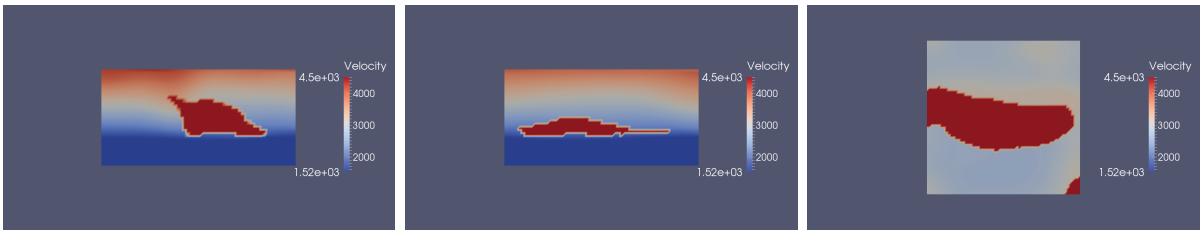


Figure 4.55: x, y and z cuts of the velocity model

A Dirichlet boundary condition is imposed as a free surface condition and PML layers of 160 m are used to bound the computation domain. A Gaussian source located at $y = (500m, 500m, 50m)$ with $\sigma = 50$ m is used as right hand side:

$$f(x) = \exp\left(-\frac{|x-y|^2}{\sigma^2}\right), \quad x \in \Omega.$$

Numerical meshes are based on cartesian grids, each cube of the grid being subdivided into 6 tetrahedron. The mesh steps are ranging from $h = 20$ m (fitting mesh) to $h = 160$ m. Numerical solutions are evaluated on a $128 \times 128 \times 64$ cartesian grid for comparison. We use single precision arithmetic and the single precision complex version of MUMPS as linear solver [11]. The MMAm code has been run on CRIHAN (HPC center). We are limited to 1024 Gb of memory (64 MPI processes with 16 Gb each), and the "out of core" option of MUMPS makes it possible to compute a reference solution using 13×10^6 of degrees of freedom and 766×10^6 of non-zero elements in the linear system using $p = 4$ elements on a fitting mesh (we are not using static condensation).

The MMAm solutions are computed using a 569 tetrahedron reference submesh. The submesh has been generated by Tetgen [90]. We compare MMAm solutions with FEm solutions. FEm solutions are computed using a constant value for the velocity in each cell. This value is selected in the grid using the barycenter of the cell (remark that no averaging is done here).

Numerical results for the frequency $f = 10$Hz are presented on Table 4.29. The reference solution is computed on a fitting mesh ($h = 20$ m) with $p = 4$ elements. We see that it is possible to obtain accurate solutions on non-fitting meshes if high order elements are used with the MMAm. In particular, it is possible to use a mesh step 8 times larger than the medium grid ($h = 160$ m) if polynomials of degree $p = 6$ are used. The MMAm solution is 10 times more accurate than the FEm solution in this case.

h	p=1	p=2	p=3	p=4	p=5	p=6
20	66.0	2.80	2.50	0	x	x
40	113	15.3	3.40	3.77	3.53	x
80	135	84.8	17.1	4.89	4.62	4.65
160	108	123	93.9	51.6	17.4	6.53

h	p=1	p=2	p=3	p=4	p=5	p=6
20	66.0	2.80	2.50	0	x	x
40	114	21.4	15.3	11.7	11.5	x
80	131	91.1	37.8	30.9	30.6	30.8
160	107	128	101	77.3	68.2	61.7

Table 4.29: Relative error in % for $f = 10$Hz with the MMAm (left) and the FEm (right)

The size and filling of the linear system are presented on Figure 4.56. Except for the case $p = 5$, the size and the filling of the linear system are reduced when p increases. We would like to point out that we do not obtain a "perfect" curve as in the Overthrust test case because we only take mesh size of the form 20×2^j. We think we could obtain a smaller linear system with less filling for $p = 5$ by letting the mesh step increase between 80 m and 160 m.

In order to compare with the standard FEm, we can consider the $p = 2$ solution on the fitting mesh as a reference. Then, the number of degrees of freedom required to obtain less that 10% of relative L^2 error is reduced by a factor 6 for $p = 4$ and by a factor 14 for $p = 6$. The filling is reduced by a factor 4 for $p = 4$ and by a factor 6 for $p = 6$. We would like to point out, however, that this result needs to be mitigate by the fact that the $p = 4$ is 1.74 less times precise and the $p = 6$ solution is 2.33 less times precise. Again, this is because the mesh size are of the form 20×2^j and we think we could obtain a similar result (less in favor of the MMAm) with the same 5% precision for $p = 2, 4$ and 6 by selecting h more freely.

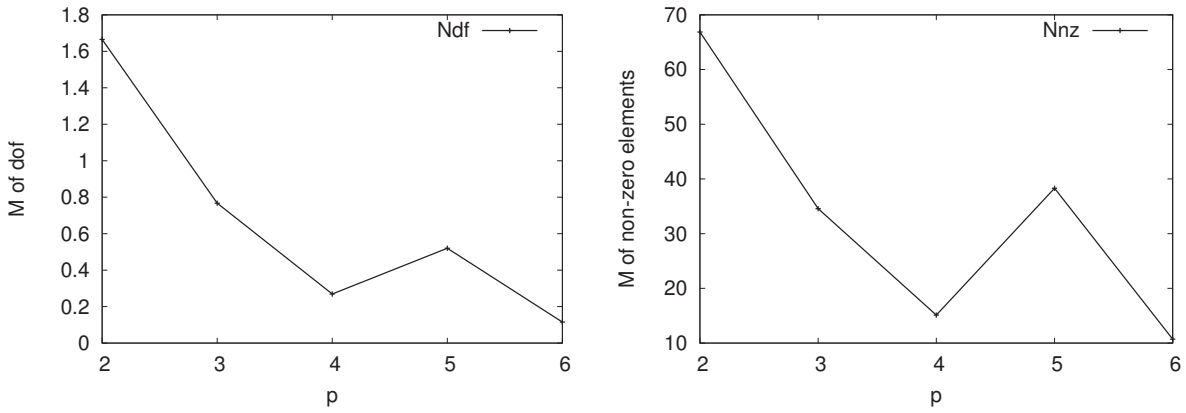


Figure 4.56: Size and filling of the linear system for a $< 10\%$ accuracy

We now provide additional information on Figure 4.57 concerning the computational time and memory usage using 16 MPI process with the linear solver CMUMPS 5.0.0 [11].

In all test cases, the time required to compute the matrix coefficients is of the order of the second, and is negligible compared to the time required to solve the linear system. This result strongly confirms the interest of the MMAm. We would like to point out that the result would be even more impressive if several right hand side were used.

It is clear that the computational time and the memory usage are reduced when p increases, at least up to $p = 4$. If $p = 2$ is taken as a reference for the standard FEm, the computational time and memory usage are reduced by 15 and 11 for $p = 4$ and by 65 and 32 for $p = 6$ with the MMAm.

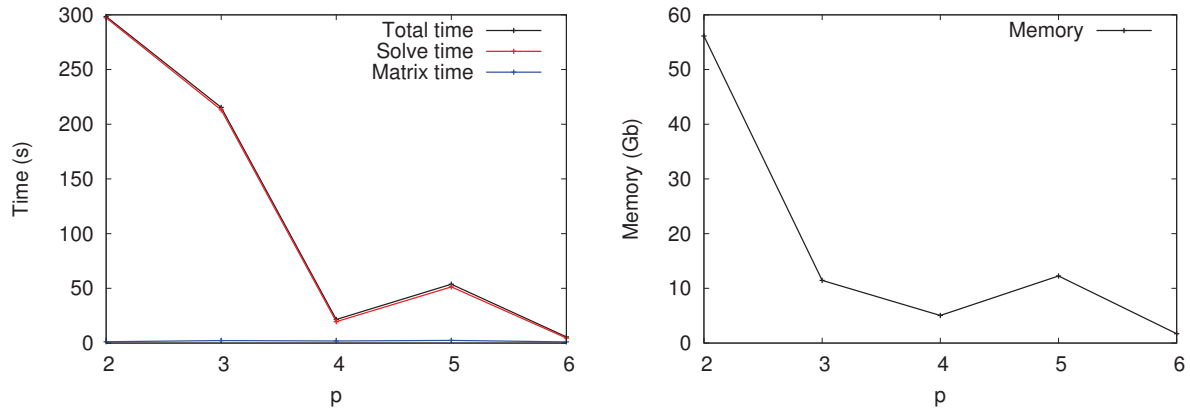


Figure 4.57: Computational times and memory usages (CMUMPS 5.0.0 with 16 MPI processes)

The results are impressive when comparing different polynomial degrees to obtain an accuracy level close to 5%. However, the results would have been less impressive if we had compared for a 3% accuracy for example. In particular, we see that when h is small the precision is not greatly improved when p is increased. For instance, we have the same precision for $p = 4, 5$ and 6 when $h = 80$. However, for larger h this is not the case and the precision is improved when p ranges from 4 to 6 and $h = 160$ m. We think that this might be because we are using simple precision arithmetic and that the matrix is ill-conditioned. Unfortunately, we did not compute the condition number of the matrix while doing the experiment (the option is available is MUMPS, however, it requires more memory and computational time). Also, the difference between the reference solution $p = 4$ and the closest solution $p = 3$ computed on the fitting mesh is of the order of 1% which might be to much to compare for a 3% level of accuracy.

We conclude that the MMAm seems very interesting for 3D acoustic experiments with constant density if an accuracy level of 10 or 5% is required. Further investigations are required to discuss the case of 1% accuracy.

4.4.5 2D Elastic simulations with constant density: Overthrust model

We close this section with 2D elastic experiment. We consider again the Overthrust model. As illustrated on Figure 4.58, we obtain a shear velocity model by applying the rule from Castagna et al. [32]:

$$c_p = 1.16 c_s + 1360.$$

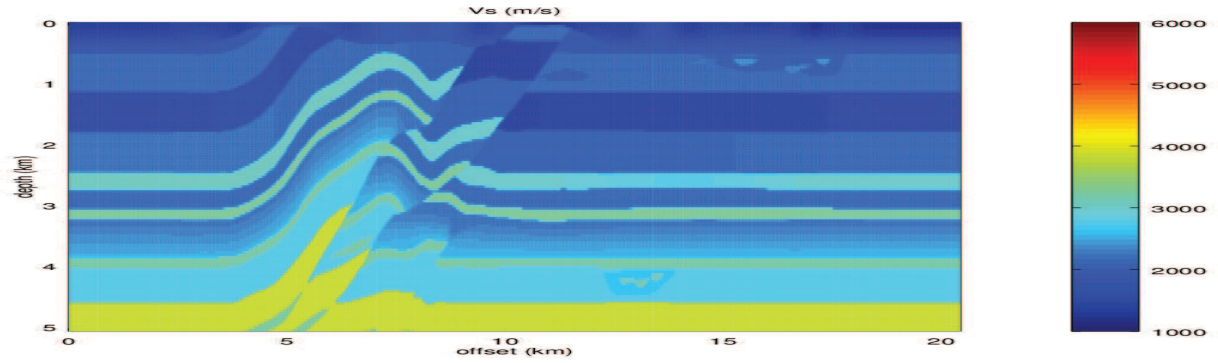


Figure 4.58: Overthrust S-velocity model

We also consider a smoothed version of the model, where a 200 m low-pass filter has been applied to the velocity model. The smoothed version of c_s is represented on Figure 4.59.

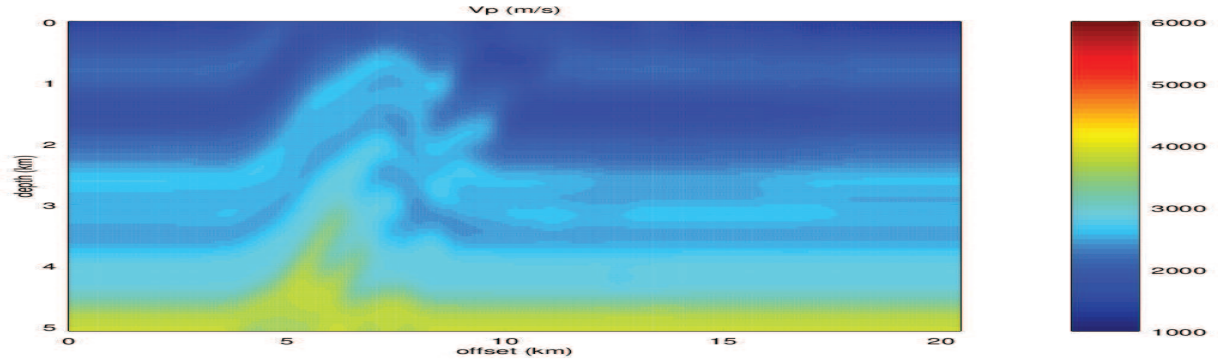


Figure 4.59: Smoothed Overthrust S-velocity model

A Neumann free-surface boundary condition is imposed on the top of the domain. 640 m PML layers are used to simulate an infinite propagation medium.

Tables 4.30 to 4.32 present numerical results where $f = 2, 5$ and 10Hz. We see that, unlike the acoustic case, there is an important accuracy difference between the original and smoothed media. As an example, when $p = 6$, the MMAm solution is always at least 5 times more accurate in the smoothed medium, than in the original medium.

We explain the difference between the acoustic and the elastic cases by the fact than in the elastic case, the velocity parameters c_p and c_s are "inside" the divergence operator. Therefore, the velocity contrasts have a stronger impact on the solution. Yet, the MMAm is still able to provide accurate solutions on non-fitting meshes, even in the original medium at low frequencies $f = 2$ and 5Hz.

However, as shown by Table 4.32, we are not able to obtain accurate solutions on non-fitting meshes in the original medium for $f = 10$Hz.

The results are promising in the smoothed medium, where accurate solutions are obtain on non-fitting meshes, even for the frequency $f = 10$Hz.

We conclude that the MMAm provide improved results in the smoothed medium for all frequencies considered and in the original medium at low frequency. From our experiment, it is not clear that the MMAm might improve the performance of the standard FEm in highly contrasted media for high frequencies.

h	p=1	p=2	p=3	p=4	p=5	p=6
40	91.7	0.64	0.14	0.09	0.06	0
80	121	8.53	2.34	1.40	1.29	0.95
160	114	64.7	5.73	3.38	2.75	2.29
320	110	117	61.3	11.5	5.76	4.50

h	p=1	p=2	p=3	p=4	p=5	p=6
40	82.9	0.43	0.14	0.09	0.06	0
80	121	5.06	0.45	0.28	0.24	0.19
160	116	51.1	1.83	0.69	0.53	0.44
320	111	116	42.3	3.55	1.25	0.93

Table 4.30: Relative error (%) for $f = 2$Hz in the original (left) and smoothed (right) elastic Overthrust models

h	p=1	p=2	p=3	p=4	p=5	p=6
40	126	40.2	0.58	0.16	0.10	0
80	116	126	24.2	4.20	3.42	2.56
160	125	123	124	60.3	12.5	7.61
320	105	110	112	128	121	88.8

h	p=1	p=2	p=3	p=4	p=5	p=6
40	125	34.9	0.39	0.14	0.09	0
80	116	125	15.6	0.80	0.46	0.35
160	111	180	123	40.3	3.45	0.97
320	103	108	129	123	120	58.8

Table 4.31: Relative error (%) for $f = 5$Hz in the original (left) and smoothed (right) elastic Overthrust models

h	p=1	p=2	p=3	p=4	p=5	p=6
40	122	135	44.8	1.42	0.20	0
80	122	136	133	120	58.8	40.0
160	104	111	115	128	131	128
320	100	104	108	111	113	120

h	p=1	p=2	p=3	p=4	p=5	p=6
40	121	135	33.5	0.87	0.17	0
80	119	146	134	82.4	7.18	0.99
160	103	109	188	128	132	108
320	100	102	105	109	113	122

Table 4.32: Relative error (%) for $f = 10$Hz in the original (left) and smoothed (right) elastic Overthrust models

Conclusion

This work focus on numerical approximation of the Helmholtz equation in heterogeneous media. The numerical solution methodology has been considered in the context of seismic imaging applications where accuracy must be achieved with a limited computational burden to have any chance of solving the corresponding inverse problem. Regarding its practical impact, the main contribution of this PhD is the full design of the Multiscale Medium Approximation method (MMAm).

Because special properties of the continuous solution were required to analyse the MMAm properly, we have started with the mathematical analysis of the continuous problem. This work has been presented in Chapter 2 for 1D Helmholtz problems and in Chapter 3 for 2D problems. Our main achievements are then the derivation of frequency-explicit stability estimates for the solution of 1D Helmholtz problem in general acoustic media. In 2D, our results are only valid if the density is constant and if the velocity satisfies a monotonous hypothesis. In both cases, our results are proper generalization of the stability estimates available for homogeneous Helmholtz problems and our bounds are optimal with respect to the frequency.

Concerning the analysis of the MMAm, we have derived a pre-asymptotic error-estimate for 2D problems when linear Lagrangian elements are used. We believe that the crucial result of our analysis is our asymptotic error-estimates for 1D problems. Indeed, we were able to show that using quadratic and cubic finite elements is very interesting, even if the solution is H^2 only because of the discontinuities of the velocity parameter. Theses results have been applied under the assumption that the density is constant.

From the numerical point of view, we have been able to validate the MMAm with analytical test-cases featuring strong velocity contrasts. These numerical results are in accordance with our theoretical analysis. The MMAm has also been tested with geophysical benchmarks, chosen to be representative of seismic imaging. These tests show that the method is accurate and efficient and outperforms the standard finite element method for the targeted application. In particular, numerical experiments on benchmarks with non-constant density show that the method is still efficient even if this case is not covered by our analysis. We have also investigated how the method works for elastic wave propagation. The results are promising, but the interest of the method might be limited to the low frequency regime.

Future developments should focus on the analysis and the possible extensions of the method for the non-constant densities both in acoustic and elastic cases. Indeed, these

results are not covered by our analysis and have been observed to be tricky to discretize from the numerical point of view. A natural extension of the MMAm for layered media with non-constant density could be the "special finite element method" of Babuška et al. [17].

In the "special finite element method" of Babuška et al., special shape functions are obtained from Lagrangian polynomials through a change of variables. This change of variables follows the variations of the density, which are supposed to be one-dimensional. The method has been tested for linear Lagrangian element. One of the drawbacks of the method is that the conformity of the original elements is lost. Since this might be a problem for higher order elements, we would propose to use a penalization technique to ensure the stability of the resulting scheme.

A promising approach for complex propagation media with non-constant density is the MHMm proposed by F. Valentin et al. [13]. In the framework of the HOSCAR project (`www-sop.inria.fr/hoscar/`), I had the opportunity to meet F. Valentin and we started working on this topic. Using the MHMm is thus an on-going work and preliminary results make us confident in the capability of MHMm to handle complex media.

Bibliography

[1] N.N. Abboud and P.M. Pinsky, *Finite element dispersion analysis for the three-dimensional second-order scalar wave equation*, International Journal for Numerical Methods in Engineering **35** (1992), 1183–1218.

[2] A. Abdulle and M.J. Grote, *Finite element heterogeneous multiscale method for the wave equation*, Multiscale Model. Simul. **9** (2011), no. 2, 766–792.

[3] M. Ainsworth, *Discrete dipersion relation for hp-version finite element approximation at high wave number*, SIAM J. Numer. Anal. **42** (2004), no. 2, 553–575.

[4] M. Ainsworth, P. Monk, and W. Muniz, *Dispersive and dissipative properties of discontinuous galerkin finite element methods for the second-order wave equation*, J. Sci. Comput. **27** (2006), 5–40.

[5] M. Ainsworth and H. Wajid, *Dispersive and dissipative behavior of the spectral element method*, SIAM J. Numer. Anal. **47** (2009), no. 5, 3910–3937.

[6] H. Ben Hadj Ali, S. Operto, and J. Virieux, *Velocity model building by 3d frequency-domain full-waveform inversion of wide-aperture seismisc data*, Geophysics **73** (2008), no. 5, VE101.

[7] T. Alkhalifah, *An acoustic wave equation for anisotropic media*, Geophysics **65** (2000), no. 4, 1239–1250.

[8] G. Allaire, *Homogenization and two-scale convergence*, SIAM J. Math. Anal. **23** (1992), no. 6, 1482–1518.

[9] G. Allaire and R. Brizzi, *A multiscale finite element method for numerical homogenization*, Multiscale Model. Simul. **4** (2005), no. 3, 790–812.

[10] M. Amara, R. Djellouli, and C. Farhat, *Convergence analysis of a discontinuous galerkin method with plane waves and lagrange multipliers for the solution of helmholtz problems*, SIAM J. Numer. Anal. **47** (2009), 1038–1066.

[11] P.R. Amestoy, I.S. Duff, and J.Y L'excellent, *Multifrontal parallel distributed symmetric and unsymmetric solvers*, Comput. Methods Appl. Engrg. **184** (2000), 501–520.

[12] F. Aminzadeh, B. Jean, N. Burkhard, J. Long, T. Kunz, and P. Duclos, *Three dimensional seg/eaeg models – an update*, The Leading Edge **15** (1996), no. 2, 131–134.

[13] R. Araya, C. Harder, D. Paredes, and F. Valentin, *Mutliscale hybrid-mixed method*, SIAM J. Numer. Anal. **51** (2013), no. 6, 3505–3531.

[14] T. Arbogast, S.E. Minkoff, and P.T. Keenan, *An operator-based approach to upscaling the pressure equation*, Computational Methods in Water Resources XII, Vol 1: Computational Methods in Contamination and Remediation of Water Resources, V.N. Burganos et al., eds., Computational Mechanics Publications, Southamption, U.K. (1998), 405–412.

[15] A.K. Aziz and R.B. Kellogg, *Finite element analysis of a scattering problem*, Mathematics of Computation **37** (1981), no. 156, 261–272.

[16] A.K. Aziz, R.B. Kellogg, and A.B. Stephens, *A two point boundary value problem with a rapidly oscillating solution*, Numerische Mathematik **53** (1988), 107–121.

[17] I. Babuška, G. Caloz, and E. Osborn, *Special finite element methods for a class of second order elliptic problems with rough coefficients*, SIAM J. Numer. Anal. **31** (1994), 510.

[18] G.E. Backus, *Long-wave anisotropy produced by horizontal layering*, J. Geophys. Res. **67** (1962), 4427–4440.

[19] H. Barucq, T. Chaumont Frelet, J. Diaz, and V. Péron, *Upscaling for the laplace problem using a discontinuous galerkin method*, J. of Comp. and Appl. Math. **240** (2013), 192–203.

[20] H. Barucq, T. Chaumont Frelet, and C. Gout, *Stability analysis of heterogeous helmholtz problems and finite element solution based on propagation media approximation*, in revision (2015).

[21] E. Bécache, S. Fauqueux, and P. Joly, *Stability of perfectly matched layers, group velocities and anisotropic waves*, Journal of Computational Physics **188** (2006), no. 2, 399–433.

[22] J.P. Bérenger, *A perfectly matched layer for the absorption of electromagnetic waves*, Journal of Computational Physics **114** (1994), 185–200.

[23] M.A. Biot, *Theory of propagation of elastic waves in a fluid-saturated porous solid. i. low-frequency range*, The journal of the acoustical society of america **28** (1956), no. 2, 168–178.

[24] ———, *Theory of propagation of elastic waves in a fluid-saturated porous solid. ii. higher frequency range*, The journal of the acoustical Society of America **28** (1956), no. 2, 179–191.

[25] L. Boillot, *Contributions à la modélisation mathématique et à l'algorithmique parallèle pour l'optimisation d'un propagateur d'ondes élastiques en milieu anisotrope*, PhD Thesis (2014).

[26] H. Brezis, *Functional analysis, sobolev spaces and partial differential equations*, Springer, 1983.

[27] Y. Capdeville, L. Guillot, and J.-J. Marigo, *1-d non-periodic homogenization for the seismisc wave equation*, Geophys. J. Int. **181** (2010), 897–910.

[28] ———, *2-d non-periodic homogenization of the elastic wave equation: Sh case*, Geophys. J. Int. **182** (2010), 1438–1454.

[29] ———, *2-d non-periodic homogenization to upscale elastic media for p-sv waves*, Geophys. J. Int. **182** (2010), 903–922.

[30] J.M. Carcione, *Seismic modeling in viscoelastic media*, Geophysics **58** (1993), no. 1, 110–120.

[31] J.M. Carcione, D. Kosloff, and A. Behle, *Long-wave anisotropy in stratified media: A numerical test*, Geophysics **56** (1991), no. 2, 245–254.

[32] J.P. Castagna, M.L. Batzle, and R.L. Eastwood, *Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks*, Geophysics **50** (1985), no. 4, 571–581.

[33] D. Cioranescu, A. Damlamian, and G. Griso, *The periodic unfolding method in homogenization*, SIAM J. Math. Anal. **40** (2008), no. 4, 1585–1620.

[34] X. Claeys and R. Hiptmair, *Boundary integral formulation of the first kind for acoustic scattering by composite structures*, http://onlinelibrary.wiley.com/doi/10.1002/cpa.21462/abstract (2011).

[35] ———, *Electromagnetic scattering at composite objects: A novel multi-trace boundary integral formulation*, https://eudml.org/doc/222174 (2011).

[36] R. Clayton and B. Engquist, *Absorbing boundary conditions for acoustic and elastic wave equations*, Bulletin of the Seismological Society of America **67** (1977), no. 6, 1529–1540.

[37] R.W. Clayton and R.H. Stolt, *A born-wkbj inversion method for acoustic reflection data*, Geophysics **46** (1981), no. 11, 1559–1567.

[38] J.F. Clearboot, *Toward a unified theory of reflector mapping*, Geophysics **36** (1971), no. 3, 467–481.

[39] P. Cummings and X. Feng, *Sharp regularity coefficient estimates for complex-valued acoustic and elastic helmholtz equations*, Mathematical Models and Methods in Applied Sciences **16** (2006), no. 1, 139–160.

[40] P.F. Daley and F. Hron, *Reflection and transmission coefficients for tranversly isotropic media*, Bulletin of the Seismological Society of America **67** (1977), no. 3, 661–675.

[41] A. Deraemaeker, I. Babuška, and P. Bouillard, *Dispersion and pollution of the fem solution for the helmholtz equation in one, two and tree dimensions*, International Journal for Numerical Methods in Engineering **46** (1999), 471–499.

[42] J. Diaz, *Approches analytiques et numériques de problèmes de transmision en propagation d'ondes en régime transitoire. application au couplage fluide-structure et aux méthodes de couches parfaitement adaptées*, PhD Thesis, HAL id: tel-00008708 (2005).

[43] J. Douglas, J.E. Santos, D. Sheen, and L.S. Bennethum, *Frequency domain treatment of one-dimensional scalar waves*, Mathematical Models and Methods in Applied Sciences **3** (1993), no. 2, 171–194.

[44] B. Engquist and A. Majda, *Absorbing boundary conditions for numerical simulation of waves*, Proc. Natl. Acad. Sci. USA **74** (1977), no. 5, 1765–1766.

[45] V. Farra, *Ray tracing in complex media*, Journal of Applied Geophysics **30** (1993), 55–73.

[46] X. Feng and H. Wu, *hp-discontinuous galerkin methods for the helmholtz equation with large wave number*, Mathematics of Computation **80** (2011), 1997–2024.

[47] X. Feng and Y. Xing, *Absolutely stable local discontinuous galerkin methods for the helmholtz equation with large wave number*, Mathematics of Computation **82** (2013), 1269–1296.

[48] S. Forest, M. Amestoy, G. Damamme, S. Kruch, V. Maurel, and M. Maziere, *Mécanique des milieux continus*, Ecole des mines de paris, available online: http://mms2.ensmp.fr/mmc_paris/poly/MMC.pdf (2009).

[49] E. Forgues, E. Scala, and R.G. Pratt, *High resolution velocity model estimation from refraction and reflection data*, 68th Annual International Meeting of Exploration Geophysicists, Expanded Abstracts (1998), 1211–1214.

[50] T. Chaumont Frelet, *On high order methods for the helmholtz equation in highly heterogeneous media*, submitted (2015).

[51] Y. Gholami, R. Brossier, S. Operto, A. Ribodetti, and J. Virieux, *Which parametrization is suitable for acoustic vertical transverse isotropic full waveform inversion? part 1: Sensitivity and trade-off analysis*, Geophysics **78** (2013), no. 2, R81–R105.

[52] D. Givoli, *High-order local non-reflection boundary conditions: a review*, Wave Motion **39** (2004), 319–326.

[53] E.L. Hamilton, *Elastic properties of marine sediments*, Journal of geophysical research **76** (1971), no. 2.

[54] ———, *Compressional wave attenuation in marine sediments*, Geophysics **37** (1972), no. 4, 620–646.

[55] U. Hetmaniuk, *Stability estimates for a class of helmholtz problems*, Commun. Math. Sci. **5** (2007), no. 3, 665–678.

[56] R. Hill, *On constitutitive inequalities for simple materials i*, J. Mech. Phys. Slids **16** (1968), 229–242.

[57] T.Y. Hou and X.-H. Wu, *A mutiscale finite element method for elliptic problems in composite materials and porous media*, Journal of Computational Physics **134** (1997), 169–189.

[58] F. Ihlenburg, *Finite element analysis of acoustic scattering*, Springer, 1998.

[59] F. Ihlenburg and I. Babuška, *Finite element solution of the helmholtz equation with high wave number part i: the h-version of the fem*, Computers Math. Applic. **30** (1995), no. 9, 9–37.

[60] ———, *Finite element solution of the helmholtz equation with high wave number part ii: the h-p version of the fem*, SIAM J. Numer. Anal. **34** (1997), no. 1, 315–358.

[61] L.-M. Imbert-Gérard and B. Després, *A generalized plane wave numerical method for smooth non constant coefficients*, IMA Journal of Numerical Analysis **34** (2014), 1072–1103.

[62] S.I. Karato, *Importance of anelasticity in the interpretation of seismic tomography*, Geophysical research letters **20** (1993), no. 15, 1623–1626.

[63] O. Korostyshevskaya and S.E. Minkoff, *A matrix analysis of operator-based upscaling for the wave equation*, SIAM J. Numer. Anal. **44** (2006), no. 2, 586–612.

[64] S.M. Kozlov, *Averaging of random operators*, Math. U.S.S.R. Sbornik **37** (1980), 167–180.

[65] R.D. Krieg and S.W. Key, *Transient shell response by numerical time integration*, International journal for numerical methods in engineering **7** (1973), 273–286.

[66] P. Laug and H. Borouchaki, *The bl2d mesh generator: Beginner's guide, user's and programmer's manual*, HAL Id: inria-00069977 (2006).

[67] J.E. Lin and W.A. Strauss, *Decay and scattering of solutions of a nonlinear schrödinger equation*, Journal of Functional Analysis **30** (1978), 245–263.

[68] A. Loseille and F. Alauzet, *Continuous mesh framework part i: well-posed continuous interpolation error*, SIAM J. Numer. Anal. **49** (2011), no. 1, 38–60.

[69] J. Luquel, *Imagerie de milieux complexes par equations d'ondes elastiques*, PhD Thesis (2015).

[70] Ch. Makridakis, F. Ihlenburg, and I. Babuška, *Analysis and finite element methods for a fluid-solid interaction problem in one dimesion*, Techical Note BN-1183 (1995).

[71] G.S. Martin, R. Wiley, and J. Manfurt, *Marmousi2: An elastic upgrade for marmousi*, The Leading Edge **25** (2006), no. 2, 156–166.

[72] V. Mattesi, H. Barucq, and J. Diaz, *Prise en compte de vitesses de propation polynomiales dans un code de simulation galerkine discontinue*, hal-01176854v1 (2015).

[73] J.M. Melenk, *On generalized finite element methods*, Ph.D. thesis, University of Maryland, 1995.

[74] J.M. Melenk, A. Parsania, and S. Sauter, *General dg-methods for highly indefinite helmholtz problems*, J. of Sci. Comp. **57** (2013), no. 3, 536–581.

[75] J.M. Melenk and S. Sauter, *Convergence analysis for finite element discretizations of the helmoltz equation with dirichlet-to-neumann boundary conditions*, Mathematics of Computation **79** (2010), no. 272, 1871–1914.

[76] ———, *Wavenumber explicit convergence analysis for galerkin discretizations of the helmholtz equation*, SIAM J. Numer. Anal. **49** (2011), no. 3, 1210–1243.

[77] C.S. Morawetz, *Time decay for the nonlinear klein-gordon equation*, Prol. Roy. Soc. London A. **306** (1968), 291–296.

[78] R. Mullen and T. Belytschko, *Dispersion analysis of finite element semidiscretizations of the two-dimensional wave equation*, International journal for numerical methods in engineering **18** (1982), 11–29.

[79] S. Operto, J. Virieux, P. Amestoy, J.Y. L'Excellent, L. Giraud, and H. Ben Hadj Ali, *3d finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study*, Geophysics **72** (2007), no. 5, SM195–SM511.

[80] S. Operto, J. Virieux, A. Ribodetti, and J.E. Anderson, *Finite-difference frequency-domain modeling of viscoacoustic wave propagation in 2d tilted transversly isotropic (tti) media*, Geophysics **74** (2009), no. 5, 75–95.

[81] B. Perthame and L. Vega, *Morrey-campanato estimates for helmholtz equations*, Journal of Functional Analysis **164** (1999), 340–355.

[82] R.E. Plessix, *Three-dimensional frequency-domain full-waveform inversion with an iterative solver*, Geophysics **74** (2009), no. 6, WCC149–WCC157.

[83] R.E. Plessix and W.A. Mulder, *A comparison between one-way and two-way wave-equation migration*, Geophysics **69** (2004), 1491–1504.

[84] R. Gerhard Pratt, Changsoo Shin, and G.J. Hicks, *Gauss-newton and full newton methods in frequency-space seismic waveform inversion*, Geophysics. J. Int. (1998), 341–362.

[85] W. Rudin, *Real and complex analysis*, Mcgraw Hill Higher Education, 1987.

[86] E. Sanchez-Palencia, *Non-homogeneous media and vibration theory*, Springer, 1980.

[87] S.A. Sauter and C. Schwab, *Boundary element methods*, Springer, 2011.

[88] A.H. Schatz, *An observation concerning ritz-galerkin methods with indefinite bilinear forms*, Mathematics of Computation **28** (1974), no. 128, 959–962.

[89] L. Schwartz, *Théorie des distributions*, Hermann, 1966.

[90] H. Si, *Tetgen – a quality tetrahedral mesh generator and three-dimensional delaunay triangulator*, Web Intelligence and Agent Systems: An International Journal - WIAS 75 (2007).

[91] L. Sirgue and R.G. Pratt, *Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies*, Geophysics **69** (2003), no. 1, 231–248.

[92] R. Sun and G.A. McMechan, *Scalar reverse-time depth migration of prestack elastic seismic data*, Geophysics **66** (2001), no. 5, 1519–1527.

[93] W.W. Symes, *Reverse time migration with optimal checkpointing*, Geophysics **72** (2007), no. 5, SM213–SM221.

[94] C. Taillandier, M. Noble, H. Churais, and H. Calandra, *First-arrival traveltime tomography based on the adjoint state method*, Geophysics **74** (2009), no. 6, WCB57–WCB66.

[95] A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation*, Geophysics **48** (1984), no. 8, 1259–1266.

[96] R. Tezaur, I. Kalashnikova, and C. Farhat, *The discontinuous enrichment method for medium-frequency helmholtz problems with a spatially variable wavenumber*, Comput. Methods Appl. Mech. Engrg. **268** (2013), 126–140.

[97] L.L. Thompson and P.M. Pinsky, *Complex wavenumber fourier analysis of the p-version finite element method*, Computational Mechanics **13** (1994), 255–275.

[98] L. Thosmen, *Weak elastic anisotropy*, Geophysics **51** (1986), no. 10, 1954–1966.

[99] A. Tonnoir, *Conditions transparentes pour la diffraction d'ondes en milieu élastique anisotrope*, PhD Thesis (2015).

[100] S.L.A. Valcke, M. Caseu, G.E. Lloyd, J.M Kendall, and G.J. Fisher, *Lattice preferred orientation and seismic anisotropy in sedimenary rocks*, Geophys. J. Int. **166** (2006), 652–666.

[101] T. Vdovina and S.E. Minkoff, *An a priori error analysis of operator upcaling for the acoustic wave equation*, International journal of numerical analysis and modeling **5** (2008), no. 4, 543–569.

[102] T. Vdovina, S.E. Minkoff, and O. Korostyshevskaya, *Operator upscaling for the acoustic wave equation*, Multiscale Model. Simul. **4** (2005), no. 4, 1305–1338.

[103] J. Virieux, *P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method*, Geophysics **51** (1986), no. 4.

[104] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics*, Geophysics **74** (2009), no. 6, 127–152.

[105] E.L. Wilson, *The static condensation algorithm*, International Journal for Numerical Methods in Engineering **8** (1974), no. 1, 198–203.

[106] H. Wu, *Pre-asymptotic error analysis of cip-fem and fem for the helmholtz equation with high wave number. part i: linear version*, IMA Journal of Numerical Analysis **34** (2013), 1266–1288.

[107] Y. Zhang, H. Zhang, and G. Zhang, *A stable tti reverse time migration and its implementation*, Geophysics **76** (2011), no. 3, WA3–WA11.

[108] H Zhou, G. Zhang, and R. Bloor, *An anisotropic acoustic wave equation for modeling and migration in 2d tti media*, An anisotropic acoustic wave equation for modeling and migration in 2D TTI media: 76th Annual International Meeting, SEG, Expanded Astracts (2006).

[109] L. Zhu and H. Wu, *Pre-asymptotic error analysis of cip-fem and fem for the helmholtz equation with high wave number. part ii: hp version*, SIAM J. Numer. Anal. **51** (2013), no. 3, 1828–1852.