Numerical study of time-harmonic acoustic problems in layered media using partition of unity finite element methods

Author: Paula M. López Pérez

PhD thesis, University of A Coruña / 2017

Supervisors: Luis Hervella Nieto, Andrés Prieto Aneiros

PhD program in Mathematical Modelling and Numerical Simulation in Engineering and Applied Science

UNIVERSIDADE DA CORUÑA

Don Luis Hervella Nieto, profesor titular do Departamento de Matemáticas da Universidade da Coruña e Don Andrés Prieto Aneiros, profesor contratado doutor do Departamento de Matemáticas da Universidade da Coruña, informan que a memoria titulada:

NUMERICAL STUDY OF TIME-HARMONIC ACOUSTIC PROBLEMS IN LAYERED MEDIA USING PARTITION OF UNITY FINITE ELEMENT METHODS

foi realizada baixo a súa dirección por Dona Paula María López Pérez, estimando que a interesada atópase en condicións de optar ao grao de Doutora en Matemáticas, polo que solicitan que sexa admitida a trámite para a súa lectura e defensa pública.

En A Coruña, a 20 de abril de 2017.

Os directores:

Prof. Dr. Luis Hervella Nieto

Prof. Dr. Andrés Prieto Aneiros

A doutoranda: Paula María López Pérez

Firma de los miembros del tribunal de la tesis doctoral, leída en A Coruña a 22 de septiembre de 2017.

Presidente:

Dr. Philippe Destuynder

Vocal:

Dr. Pablo Ortiz Rossini

Secretaria:

Dra. Maria Elena Vázquez Cendón

Contents

List of Figures

List of Tables

Preface

The modelling of acoustic wave propagation can be applied to several problems such as noise reduction, medical ultrasonics, seismic exploration, underwater acoustics or non destructive testing. In this context, it emerges the need of solving diverse and challenging acoustic propagation problems that can not be solved with classic mathematical methods. Tests with prototypes are often used to assure the accuracy of the proposed technologies. But the high cost of its fabrication makes necessary that the tests are carried out in an advanced stage of design, with a proposal that is close to the final solution. Numerical simulation is a determinant technique to analyse and design acoustic systems in a short time and with competitive costs.

The mathematical diversity of the acoustic propagation problems makes necessary the employment of a wide variety of numerical models and the application of advanced numerical computation techniques. Between all these models, the Helmholtz equation is widely used as the reference model in time-harmonic acoustic propagation problems. At middle and high frequency regime, its numerical approximation, computed by a nodal Finite Element Method (FEM), differs significantly from the exact solution, due to the so-called "pollution" effect (see [14]). So, the accuracy and reliability of the Helmholtz numerical approximations are based on pollution-free discrete methods, which should have a robust behaviour with respect to the wave number.

The Partition of Unity Finite Element Method (PUFEM), introduced by Babuška and Ihlenburg in 1996 (see [36]), has been considered in this thesis among the pollution-free methods. Computational advantages and implementation drawbacks of the PUFEM discretization have been shown numerically in several works (see for example [35]), but as far as the author knowledge goes, there is not any PUFEM error estimate in terms of the wave number available in the literature.

The goal of the first chapter of this thesis will be to deduce an error estimate in terms of the wave number for a PUFEM discretization based on a plane wave enrichment, applied to a one-dimensional Helmholtz problem. The second chapter is devoted to the numerical approximation of time-harmonic acoustic one and two-dimensional problems in bi-layered media, developing PUFEM techniques that take into account the reflection and the transmission occurred at the interface between subdomains. Finally, the last chapter of this thesis proposes a novel PUFEM discretization that involves Love waves as a tool in non-destructive testing.

A more detailed description of the content of each chapter is discussed below:

Chapter 1. Error estimates for partition of unity finite element solutions of the Helmholtz equation

In the first chapter, *a priori* error estimates for a PUFEM discretization of a one-dimensional Helmholtz problem are deduced. First, the one-dimensional Helmholtz problem is posed, considering Dirichlet and Robin boundary conditions. Then, the variational formulation is derived and the Ladyzhenskaya-Babuška-Brezzi (LBB) *inf-sup* continuous condition and the stability of the weak solution respect to the data of the problem are stated. A PUFEM discretization of the Helmholtz one-dimensional problem, based on a plane wave enrichment, is described in terms of exponential and trigonometric functions. An additional perturbation parameter δ is introduced in the wave number of the basis functions, in order to reproduce situations where the exact solution is not known in closed form, and to try to reflect the problems to approximate the exact solution that will be found in two-dimensional Helmholtz problems or in problems with variable wave number. After that, two interpolation estimates are stated and the combination of both leads to an interpolant-like procedure which approximates accurately the H^1 projection in the PUFEM discrete space. An LBB discrete condition and a stability condition for the PUFEM approximate solution respect to the source function are demonstrated, and the error estimate in terms of the wave number k, the mesh size h and an additional perturbation parameter δ is deduced. Finally, some numerical results illustrate the second-order accuracy of the PUFEM approximation with respect to the mesh size and respect to the additional perturbation parameter. It can be checked that the PUFEM relative error does not depend on the wave number values.

Chapter 2. A partition of unity finite element method for layered media

The second chapter deals with several Helmholtz problems. Firstly, a one-dimensional Helmholtz problem in a layered media is considered. After posing the model problem, with Dirichlet and Robin boundary conditions and being the wave number a strictly positive piecewise constant function, the variational formulation is deduced. Four different kinds of PUFEM are explained in parallel: a global average method, a local element-wise method, a local average method (based on the approach introduced by Pablo Ortiz [41]) and the discretization proposed in this thesis: the transmission-reflection method, that takes into account the transmissions and reflections that occur in each element. After describing the matrix discrete problem, some numerical results show that an exact solution defined differently in each subdomain as a linear combination of plane waves, is fully recovered by the transmission-reflection method.

The second problem considered in this chapter is a two-dimensional Helmholtz problem with constant wave number. After introducing the model problem, with Neumann boundary conditions, and deducing the weak problem, the constant PUFEM discretization is described in detail and the discrete problem and the matrix discrete system are posed. The integration techniques used, in order to deal with integrands that oscillate in terms of two variables, are detailed. The numerical results in this section illustrate the accuracy of the method, the exponential decay of the relative error when the number of plane waves used in the PUFEM discretization is increased and the behaviour of the relative error with respect to the wave number.

Finally, the last problem that this chapter deals with is a Helmholtz problem in a bilayered medium. The model problem is posed, with Neumann boundary conditions and piecewise constant wave number. Coupling conditions are imposed over the interface between media. After posing the weak problem, the novel transmission-reflection PUFEM discretization is described in detail and some integration techniques used, that involve an affine change of reference to the reference triangle, are explained. The numerical results show the accuracy and efficiency of the transmission-reflection PUFEM method to approximate a two-dimensional Helmholtz problem with Neumann conditions and piecewise constant wave number.

Chapter 3. A modal-based partition of unity finite element method for layered wave propagation problems

The last chapter gives a numerical tool for the non destructive testing. The problem posed can be applied to the modelling of the transversal section of a pipe with a coating, where the internal media is a thin layer (austenitic material) and the external media a thicker one (ferritic material). The non destructive testing goal is to detect a crack on the interface between this two layers, and in order to do that, the knowledge of the solution of the problem in a domain without a crack has a vital importance. This chapter proposes a novel PUFEM discretization involving Love waves to approximate the solution of these problems without crack. After posing the model problem, an exhaustive spectral analysis is carried out, before of describing the modal-based PUFEM method in detail. A wide battery of numerical results illustrate the accuracy of the proposed modal-based PUFEM method with just Love waves and with both, internal and Love waves, the deterioration of the numerical results due to the high condition numbers of the discrete matrix and its potential mitigation using regularization techniques, and the accuracy of the modal-based method for solutions which are close to the constant-valued eigenmodes (which are not included in the modal enrichment)

The last part of the thesis is devoted to the discussion on some future research lines and open problems. A summary of this dissertation thesis in Spanish language is enclosed too.

Notation

The notation that will be used all over the thesis is introduced here (see [39], [20], [12] and [26] for more details). Consider the interval $(0,1) \subset \mathbb{R}$. The space $\mathrm{L}^2(0,1)$ is defined as

$$\mathrm{L}^2(0,1) := \left\{ f : (0,1) \to \mathbb{C};\ \int_0^1 |f(x)|^2 \mathrm{d}x < +\infty \right\},$$

with the scalar product associated

$$\langle f, g\rangle_0 := \int_0^1 f(x)\bar{g}(x)\mathrm{d}x \qquad \forall\, f, g \in \mathrm{L}^2(0,1),$$

and the corresponding norm

$$\|f\|_0 := \left(\int_0^1 |f(x)|^2 \mathrm{d}x \right)^{1/2} \qquad \forall\, f \in \mathrm{L}^2(0,1).$$

Let $m > 0$ an integer. The Sobolev space $\mathrm{H}^m(0,1)$ can be defined as

$$\mathrm{H}^m(0,1) := \left\{ v \in \mathrm{L}^2(0,1);\ \partial^\alpha v \in \mathrm{L}^2(0,1)\ \forall\, \alpha = 0, 1, \ldots, m \right\}.$$

In particular, the Sobolev space of order one in $(0,1)$ is

$$\mathrm{H}^1(0,1) := \left\{ v \in \mathrm{L}^2(0,1);\ v' \in \mathrm{L}^2(0,1) \right\},$$

whose scalar product associated is

$$\langle u, v\rangle_1 := \langle u, v\rangle_0 + \langle u', v'\rangle_0 \qquad \forall\, u, v \in \mathrm{H}^1(0,1),$$

and the seminorm and norm associated are, respectively

$$|v|_1 := \left(\int_0^1 |v'(x)|^2 \mathrm{d}x \right)^{1/2} \qquad \forall\, v \in \mathrm{H}^1(0,1),$$

and

$$\|v\|_1 := \left(\|v\|_0^2 + |v|_1^2 \right)^{1/2} \qquad \forall\, v \in \mathrm{H}^1(0,1).$$

Let Ω an open bounded domain in $\mathbb{R}^2$ and let $\partial\Omega$ its boundary. Let $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$. The space of Lebesgue measurable and square-integrable complex-valued functions is denoted

$$\mathrm{L}^2(\Omega) := \left\{ f : \Omega \to \mathbb{C};\ \int_\Omega |f(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} < +\infty \right\},$$

with the scalar product associated

$$\langle f, g \rangle_{0,\Omega} := \int_\Omega f(\boldsymbol{x}) \bar{g}(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \qquad \forall\, f, g \in \mathrm{L}^2(\Omega),$$

and the corresponding norm

$$\|f\|_{0,\Omega} := \left(\int_\Omega |f(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2} \qquad \forall\, f \in \mathrm{L}^2(\Omega).$$

The Sobolev space H^m for $m \in \mathbb{N}$ is defined by

$$\mathrm{H}^m(\Omega) := \left\{ v \in \mathrm{L}^2(\Omega);\ \partial^{\boldsymbol{\alpha}} v \in \mathrm{L}^2(\Omega)\ \forall\, \boldsymbol{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{N}^2 \text{ with } \alpha_1 + \alpha_2 \leq m \right\},$$

where the derivative

$$\partial^{\boldsymbol{\alpha}} v(\boldsymbol{x}) := \frac{\partial^{\alpha_1+\alpha_2} v(\boldsymbol{x})}{\partial^{\alpha_1} x_1 \partial^{\alpha_2} x_2},$$

is interpreted in the sense of distributions.

In particular, if $s = 1$, the Sobolev space of order one is denoted

$$\mathrm{H}^1(\Omega) := \left\{ v \in \mathrm{L}^2(\Omega);\ \frac{\partial v}{\partial \boldsymbol{x}_j} \in \mathrm{L}^2(\Omega),\ \forall\, j = 1, 2 \right\},$$

whose scalar product associated is

$$\langle u, v \rangle_{1,\Omega} := \langle u, v \rangle_{0,\Omega} + \langle \frac{\partial u}{\partial \boldsymbol{x}_1}, \frac{\partial v}{\partial \boldsymbol{x}_1} \rangle_{0,\Omega} + \langle \frac{\partial u}{\partial \boldsymbol{x}_2}, \frac{\partial v}{\partial \boldsymbol{x}_2} \rangle_{0,\Omega} \qquad \forall\, u, v \in \mathrm{H}^1(\Omega),$$

and the seminorm and norm associated are, respectively

$$|v|_{1,\Omega} := \left(\int_\Omega |\nabla v(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} \right)^{1/2} \qquad \forall\, v \in \mathrm{H}^1(\Omega),$$

and

$$\|v\|_{1,\Omega} := \left(\|v\|_{0,\Omega}^2 + |v|_{1,\Omega}^2 \right)^{1/2} \qquad \forall\, v \in \mathrm{H}^1(\Omega).$$

The Sobolev space of order one that has homogeneous Dirichlet conditions over a particular part of the boundary $\Gamma \subset \partial\Omega$ is defined

$$\mathrm{H}^1_\Gamma(\Omega) := \left\{ v \in \mathrm{H}^1(\Omega);\ v = 0 \text{ over } \Gamma \right\}.$$

The Sobolev space $\mathrm{H}^s(\Omega)$, being $s = m + \sigma$, $m \in \mathbb{N}$ and $\sigma \in (0,1)$, is defined as

$$\mathrm{H}^s(\Omega) := \left\{ v \in \mathrm{H}^m(\Omega);\ \int_\Omega \int_\Omega \frac{|\partial^{\boldsymbol{\alpha}} v(\boldsymbol{x}) - \partial^{\boldsymbol{\alpha}} v(\boldsymbol{y})|^2}{|\boldsymbol{x} - \boldsymbol{y}|^{2+2\sigma}} \, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \right\}.$$

A function is said to be of class $\mathcal{C}^\infty$ if it has continuous derivatives of all orders. The space $\mathrm{H}(\mathrm{div}, \Omega)$ is defined as follows

$$\mathrm{H}(\mathrm{div}, \Omega) := \left\{ \boldsymbol{\varphi} \in \left(\mathrm{L}^2(\Omega)\right)^2;\ \mathrm{div}\boldsymbol{\varphi} \in \mathrm{L}^2(\Omega) \right\}.$$

Let γ the trace operator, $\gamma : \mathrm{H}^1(\Omega) \to \mathrm{L}^2(\partial\Omega)$. Then, the space $\mathrm{H}^{1/2}(\partial\Omega)$ can be defined

$$\mathrm{H}^{1/2}(\partial\Omega) := \left\{ u \in \mathrm{L}^2(\partial\Omega);\ \exists\, v \in \mathrm{H}^1(\Omega) \text{ with } u = \gamma(v) \right\},$$

and its dual space is $\mathrm{H}^{-1/2}(\partial\Omega)$.

The space $\mathrm{H}^{1/2}_{00}(\Gamma)$ with $\Gamma \subset \partial\Omega$ is defined as

$$\mathrm{H}^{1/2}_{00}(\Gamma) := \left\{ v \in \mathrm{L}^2(\Gamma);\ \widetilde{v} \in \mathrm{H}^{1/2}(\partial\Omega) \right\},$$

being $\widetilde{v}$ the extension by zero of v,

$$\widetilde{v} = \begin{cases} v(\boldsymbol{x}) & \text{if } \boldsymbol{x} \in \Gamma, \\ 0 & \text{if } \boldsymbol{x} \in \partial\Omega \backslash \Gamma. \end{cases}$$

Chapter 1

Error estimates for partition of unity finite element solutions of the Helmholtz equation

Contents

1.1 Introduction

Boundary-value problems for the Helmholtz equation arise in a number of physical applications, in particular in problems of wave scattering in Acoustics, Optics and Electromagnetism. It is well known (see, for instance, [24]) that to obtain accurate results, the mesh size h for finite element and finite difference computations should depend on the wave number k, usually following a “rule of the thumb”, which ensures a minimum number of nodes per wavelength. In problems where the typical size of the computational domain has the same order of magnitude as the wavelength of the harmonic motion, this criterion leads to accurate results. However, the quality of the numerical approximation deteriorates if the computational domain or the wave number are large enough. Under certain assumptions on the magnitude hk, it has been shown in [27] that the H^1-relative error of the FEM solution e_{fe} can be bounded by $e_{\mathrm{fe}} \leq C_1 kh + C_2 k^3 h^2$, where the second term on the right-hand side is the so-called numerical pollution error.

In [36], the Partition of the Unity Finite Element Method was proposed with the aim of mitigating the pollution effect of standard FEM approximations. The computational advantages of this method have been illustrated by a variety of numerical results (see e.g. [35]). In this chapter, a PUFEM discretization based on a plane wave enrichment is applied to the one-dimensional Helmholtz equation.

The one-dimensional Helmholtz boundary-value problem with Dirichlet and Robin boundary conditions is described in Section 1.2, as well as its variational formulation and the associated Ladyzhenskaya-Babuška-Brezzi (LBB) *inf-sup* condition. The PUFEM discretization chosen is discussed in detail in Section 1.3. The remainder of the chapter is organized as follows: two interpolation estimates and an approximation result of the H^1 projection of the weak solution is shown for the PUFEM discrete space in Section 1.4. Then, a discrete *inf-sup* condition is proved and the existence and uniqueness of the discrete solution and its stability with respect to the boundary data are obtained in Section 1.5. An error estimate for the PUFEM is deduced in Section 1.6. Finally, some numerical results are presented in Section 1.7.

1.2 Model problem

The time-harmonic wave propagation in isotropic homogeneous compressible media is modelled linearly by means of the Helmholtz equation. Throughout this chapter, a one-dimensional model will be considered. Without loss of generality it will be assumed the interval $(0,1)$ as computational domain (otherwise, a change of scale could be performed to transform the domain to the unit interval). Analogously to the model problem used in [27] for the FEM analysis, the following boundary-value problem will be considered

$$\left\{\begin{array}{rcl} -u'' - k^2 u &=& f \quad \text{in } (0,1), \\ u(0) &=& u_0, \\ u'(1) - iku(1) &=& u_1, \end{array}\right. \tag{1.1}$$

where u and f are complex-valued functions. The source term f is assumed independent of k. The boundary data $u_0, u_1 \in \mathbb{C}$ and the wave number $k > 0$ (k strictly positive and lower bounded far from zero) are constant. From an acoustic point of view, u could be understood as the complex-valued time-harmonic amplitude of the pressure field in a compressible fluid at a fixed wave number k. Since at $x = 0$, a Dirichlet boundary condition is assumed and a complex-valued Robin condition is imposed at $x = 1$, it is straightforward to check that the model problem has a unique solution. The proof is based on the classical *inf-sup* condition. In what follows, the variational formulation and the result of existence and uniqueness of solution will be recalled.

In the model problem (1.1), the Dirichlet and the Robin data u_0 and u_1 can be lifted by a smooth function and then it can be used to translate the solution u. In this manner, the boundary data u_0 and u_1 can be considered null without loss of generality. Hence, to write the variational formulation, the solution will be sought in the space

$$\mathrm{V} = \left\{v \in \mathrm{H}^1(0,1);\ v(0) = 0\right\} = \mathrm{H}^1_{(0}(0,1),$$

and the variational formulation of problem (1.1) is written as follows:

$$\begin{cases} \text{Given } f \in \mathrm{L}^2(0,1), \text{ find } u \in \mathrm{V} \text{ such that} \\ B_k(u,v) - iku(1)\bar{v}(1) = \displaystyle\int_0^1 f(x)\bar{v}(x)\,\mathrm{d}x \qquad \forall v \in \mathrm{V}, \end{cases} \tag{1.2}$$

where the sesquilienar form $B_k : \mathrm{V} \times \mathrm{V} \to \mathbb{C}$ is defined by

$$B_k(u,v) = \int_0^1 \big(u'(x)\bar{v}'(x) - k^2 u(x)\bar{v}(x)\big)\,\mathrm{d}x, \qquad u, v \in \mathrm{V}. \tag{1.3}$$

The *inf-sup* condition of the sesquilinear form $(u,v) \mapsto B_k(u,v) - ik(1)u(1)\bar{v}(1)$ can be obtained explicitly in terms of the wave number k (see [27] for the proof details).

Theorem 1.2.1. *Let B_k be the sesquilinear form defined in* (1.3)*. The following inf-sup condition holds*

$$\gamma = \inf_{u\in V}\sup_{v\in V} \frac{|B_k(u,v) - iku(1)\bar{v}(1)|}{|u|_1|v|_1} > 0.$$

Moreover, there exist two positive constants C_1, C_2, not depending on k, such that

$$\frac{C_1}{k} \leq \gamma \leq \frac{C_2}{k}. \tag{1.4}$$

The *inf-sup* condition stated above ensures the existence and uniqueness of solution for the problem (1.2) in V. In addition, it guarantees the continuous dependence of the solution with respect to the source term and the boundary data.

Corollary 1.2.2. *If $u \in \mathrm{V}$ is the solution of the variational problem* (1.2) *then u depends continuously on the source term f and the boundary data u_1 satisfying the stability estimate*

$$|u|_1 \leq Ck\left(||f||_0 + |u_1|\right), \tag{1.5}$$

being $C > 0$ a constant independent on k.

Proof. Applying the *inf-sup* estimate (1.4) to the solution u of the variational problem (1.2), there exists $v \in \mathrm{V}$ such that

$$\frac{C_1}{k} \leq \frac{|B_k(u,v) - iku(1)\bar{v}(1)|}{|u|_1|v|_1} = \frac{\left|\int_0^1 f(x)\bar{v}(x)\,\mathrm{d}x + u_1\bar{v}(1)\right|}{|u|_1|v|_1}.$$

Using a Poincare inequality, this is, for $v \in \mathrm{V}$ it is immediate to check that $||v||_0 \leq |v|_1$ and the trace inequality $|v(1)| \leq |v|_1$, it is possible to bound the numerator of the last quotient in the estimate above and so it leads to

$$\frac{C_1}{k} \leq \frac{||f||_0|v|_1 + |u_1||v|_1}{|u|_1|v|_1} = \frac{||f||_0 + |u_1|}{|u|_1},$$

from which the estimate (1.5) is obtained with $C = 1/C_1$. □

1.3 PUFEM discrete problem

The Partition of Unity Finite Element Method (PUFEM) is introduced in this section in the same manner as Babuška and Melenk [36] described it: PUFEM can be understood as a Galerkin method where a kind of specialized functions (related to the model problem to be solved) are multiplied by a partition of unity of the computational domain. So, two main ingredients have to be considered in the PUFEM discretization: the partition of unity and the set of the problem-related functions.

To define the partition of unity, an equispaced mesh $\mathcal{T}_h = \{x_j = hj : j = 0, \ldots, n\} \subset [0,1]$ of $n+1$ nodes with mesh size $h = 1/n$ is considered. On this mesh, a standard Lagrange $\mathbb{P}_1$ (piecewise linear) finite element basis $\{\varphi\}_{j=0}^n$ will be used as the set of elements of the partition of unity. In fact, since $\varphi_j(x_l) = \delta_{jl}$ for $j, l = 0, \ldots, n$, being δ_{jl} the Kronecker delta, it is easy to check the partition of unity property

$$\sum_{j=0}^{n} \varphi_j(x) = 1,$$

taking into account the nodal interpolation properties of the finite element basis and due to the continuous piecewise $\mathbb{P}_1$-polynomial expressions, in particular, using the fact that the constant functions are in the discrete space generated by $\{\varphi\}_{j=0}^n$.

The second key component in the PUFEM discretization is the set of problem-related functions. As it has been devised by other authors [35, 41] for the Helmholtz equation, plane wave solutions of the homogeneous Helmholtz equation can be used for this purpose. However, in the present chapter, instead of working with exact solutions of the Helmholtz equation, and additional perturbation parameter δ will be introduced in the problem-related functions to reproduce a lack of knowledge on the exact solution or to mimic a situation where the exact solution is not completely known in closed form. Hence, the perturbed plane waves used to describe the PUFEM discrete space are $e^{i(k+\delta)x}$ and $e^{-i(k+\delta)x}$. Consequently, the PUFEM estimates derived throughout this work will depend on three parameters: the mesh size h, the wave number k and the perturbation parameter δ.

1.3.1 Exponential discrete basis

Once these two components have been chosen, the functions involved in the PUFEM discretization, $\{\psi_j^-\}_{j=0}^n \cup \{\psi_j^+\}_{j=0}^n$, are the products of the perturbed planewave functions multiplied by each element of the partition of unity, this is,

$$\psi_j^-(x) = \varphi_j(x)e^{-i(k+\delta)(x-x_j)}, \qquad \psi_j^+(x) = \varphi_j(x)e^{+i(k+\delta)(x-x_j)} \qquad \text{for } j = 0,\dots,n.$$

In this manner, if $\mathrm{X}_h = \langle\{\psi_j^-\}_{j=0}^n \cup \{\psi_j^+\}_{j=0}^n\rangle$ then the PUFEM discrete space where the discrete solution will be sought is given by $\mathrm{V}_h = \{v_h \in \mathrm{X}_h;\ v_h(0) = 0\} = \mathrm{X}_h \cap \mathrm{H}^1_{(0}(0,1)$. Hence, the discrete PUFEM approximation u_h is defined as the solution of the following linear problem:

$$\begin{cases} \text{Given } f \in \mathrm{L}^2(0,1), \text{ find } u_h \in \mathrm{V}_h \text{ such that} \\ B_k(u_h, v_h) - iku_h(1)\bar{v}_h(1) = \displaystyle\int_0^1 f(x)\bar{v}_h(x)\,\mathrm{d}x \qquad \forall v_h \in \mathrm{V}_h. \end{cases} \tag{1.6}$$

In the following sections, the numerical properties of this discrete problem will be analysed in terms of approximability, stability and dispersion.

Since a basis has been fixed for the PUFEM discrete space V_h, it is possible to write the linear problem (1.6) in matrix form. Since the homogeneous Dirichlet condition must be satisfied for any element of the basis, it is straightforward to check that the set $\{\psi_0^+ - \psi_0^-, \psi_1^-, \psi_1^+, \dots, \psi_n^-, \psi_n^+\}$ is a basis for V_h. Hence, any function v_h can be written as

$$v_h = v_0^+(\psi_0^+ - \psi_0^-) + \sum_{j=1}^n v_j^-\psi_j^- + \sum_{j=1}^n v_j^+\psi_j^+,$$

where $(v_0^+, v_1^-, v_1^+, \dots, v_n^-, v_n^+)^t$ is the coordinates vector of v_h with respect to this basis. This coordinates can be considered as the degrees of freedom of the PUFEM discretization. Hence, problem (1.6) can be written in terms of the amplitudes $u_j^\pm$ of the approximate solution $u_h = u_0^+(\psi_0^+ - \psi_0^-) + \sum_{j=1}^n u_j^-\psi_j^- + \sum_{j=1}^n u_j^+\psi_j^+ \in \mathrm{V}_h$ as follows: Given the Robin boundary data $u_1 \in \mathbb{C}$ and $f \in \mathrm{L}^2(0,1)$, find $(u_0^+, u_1^-, u_1^+, \dots, u_n^-, u_n^+)^t \in \mathbb{C}^{2n+1}$ such that

$$\left\{\begin{array}{ll} (b_4 - b_3)u_0^+ + (b_2 - b_1)u_1^+ + (b_1 - \bar{b}_2)u_1^- = f_0^+ - f_0^-, & \\ (b_2 - b_1)u_0^+ + b_3u_1^+ + b_1u_2^+ + b_4u_1^- + \bar{b}_2u_2^- = f_1^-, & \\ (b_1 - \bar{b}_2)u_0^+ + b_4u_1^+ + b_2u_2^+ + b_3u_1^- + b_1u_2^- = f_1^+, & \\ b_1u_{j-1}^+ + b_3u_j^+ + b_1u_{j+1}^+ + b_2u_{j-1}^- + b_4u_j^- + \bar{b}_2u_{j+1}^- = f_j^-, & j = 1,\dots,n-1, \\ \bar{b}_2u_{j-1}^+ + b_4u_j^+ + b_2u_{j+1}^+ + b_1u_{j-1}^- + b_3u_j^- + b_1u_{j+1}^- = f_j^+, & j = 1,\dots,n-1, \\ b_1u_{n-1}^+ + (b_3/2 - ik)u_n^+ + b_2u_{n-1}^- + (b_4/2 - ik)u_n^- = f_n^- + u_1, & \\ \bar{b}_2u_{n-1}^+ + (b_4/2 - ik)u_n^+ + b_1u_{n-1}^- + (b_3/2 - ik)u_n^- = f_n^+ + u_1, & \end{array}\right. \tag{1.7}$$

where $f_j^\pm = \int_0^1 f(x)\psi_j^\pm(x)\,\mathrm{d}x$, $j = 0,\dots,n$. Taking into account that $\psi_j^+(x) = \overline{\psi_j^-(x)}$ for $j = 0,\dots,n$, and, since the mesh is uniform (with mesh size h), $\psi_j^-(x) = \psi_0^-(x - jh)$ and

$\psi_j^+(x) = \psi_0^+(x - jh)$ for $j = 0, \dots, n$, it holds

$$B_k(\psi_j^+, \psi_l^-) = \overline{B_k(\psi_j^-, \psi_l^+)}, \qquad B_k(\psi_j^+, \psi_l^+) = \overline{B_k(\psi_j^-, \psi_l^-)} \qquad \text{for all } 0 \le j, l \le n, \tag{1.8}$$

$$B_k(\psi_{j+m}^+, \psi_{l+m}^+) = B_k(\psi_j^+, \psi_l^+), \qquad B_k(\psi_{j+m}^-, \psi_{l+m}^-) = B_k(\psi_j^-, \psi_l^-) \qquad \text{for all } 0 \le j, l \le n, \tag{1.9}$$

such that $0 \le j + m,\ l + m \le n$. In addition, since the sesquilinear form B_k is hermitian, a direct computation of the matrix coefficients in (1.7) reveals that they are given by

$$\begin{aligned} b_1 = B_k(\psi_{j-1}^-, \psi_j^+) = B_k(\psi_1^-, \psi_2^+) = & \frac{-1}{2h^2(k+\delta)} \left((k+\delta)h\cos((k+\delta)h) + \sin((k+\delta)h)\right) \\ & + \frac{k^2}{2h^2(k+\delta)^3} \left((k+\delta)h\cos((k+\delta)h) - \sin((k+\delta)h)\right) \end{aligned} \tag{1.10}$$

$$b_2 = B_k(\psi_{j-1}^+, \psi_j^+) = B_k(\psi_1^+, \psi_2^+) = e^{i(k+\delta)h} \left(\frac{-1}{h} + i(k+\delta) + \frac{h\delta}{6}(2k+\delta) \right), \tag{1.11}$$

$$\begin{aligned} b_3 = B_k(\psi_j^-, \psi_j^+) = B_k(\psi_1^-, \psi_1^+) = & \frac{-1}{4h^2(k+\delta)} \left(-4h(k+\delta) - 2\sin(2(k+\delta)h)\right) \\ & + \frac{k^2}{4h^2\,(k+\delta)^3} \left(-4h(k+\delta) + 2\sin(2(k+\delta)h)\right), \end{aligned} \tag{1.12}$$

$$b_4 = B_k(\psi_j^+, \psi_j^+) = B_k(\psi_1^+, \psi_1^+) = \frac{2}{h} + \frac{2h\delta}{3}(2k+\delta), \tag{1.13}$$

for $j = 1, \dots, n-1$. These expressions and the symmetry of the basis functions $\psi_0^\pm$ and $\psi_n^\pm$ with respect to $x = x_0 = 0$ and $x = x_n = 1$ can also be used to the corresponding coefficients obtained for the two first and two last rows of the linear system (1.7). Notice that, despite the PUFEM basis consists in a set a complex-valued functions, matrix coefficients b_1, b_3 and b_4 in (1.7) are real.

Obviously, as in any Galerkin method applied to the Helmholtz equation, the linear system (1.7) admits a matrix representation in terms of the stiffness and mass matrices: given $\vec{f}_h = (f_0^-, f_0^+, \dots, f_{n-1}^-, f_{n-1}^+, f_n^- + u_1, f_n^+ + u_1)^t$, find $\vec{u}_h = (u_0^-, u_0^+, \dots, u_n^-, u_n^+)^t$ such that

$$(-k^2\mathcal{M}_h - ik\mathcal{R}_h + \mathcal{K}_h)\vec{u}_h = \vec{f}_h, \tag{1.14}$$

under the restriction $u_0^- + u_0^+ = 0$, where the components of the stiffness and mass matrices are given by

$$[\mathcal{M}_h]_{\tilde{j}^\pm \tilde{l}^\pm} = \int_0^1 \psi_j^\pm(x)\bar{\psi}_l^\pm(x)\,\mathrm{d}x, \qquad [\mathcal{K}_h]_{\tilde{j}^\pm \tilde{l}^\pm} = \int_0^1 (\psi_j^\pm)'(x)(\bar{\psi}_l^\pm)'(x)\,\mathrm{d}x, \tag{1.15}$$

for all $0 \le j, l \le n$ (with $\tilde{j}^\pm = 4(j + 3 \pm 1)/2$), and the matrix $\mathcal{R}_h$ associated to the Robin condition has all its coefficients null except $[\mathcal{R}_h]_{jl} = 1$ for all $j, l = 2n+1, 2(n+1)$.

Despite this discrete basis and the associated linear system (1.14) have been used in the computer implementation of this PUFEM method, the basis description presented above is far from being adequate to analyse numerically the *a priori* error due to the PUFEM discretization. In what follows, a more convenient description of the PUFEM discrete spaces X_h and V_h will be introduced.

1.3.2 Trigonometric discrete basis

Since the elements of the partition of unity (piecewise linear functions) are multiplied by $\exp(\pm i(k+\delta)x)$, it is clear that any function $v_h \in \mathrm{X}_h$ can be rewritten as

$$v_h(x) = \sum_{j=0}^{n} \left(v^{\mathfrak{p}}_{hj} \varphi_j(x) \cos((k+\delta)(x-x_j)) + v^{\mathfrak{b}}_{hj} \varphi_j(x) \sin((k+\delta)(x-x_j)) \right),$$

where $\{\phi_j\}_{j=0}^n$ are the canonical basis functions of the $\mathbb{P}_1$-Lagrange discrete space. As it is usual in a finite element framework, the description of the finite element discrete space is made by means of the writing the discrete functions in terms of the local expressions in the element of reference.

Due to the uniform partition mesh, the j-th finite element is defined by $T_j = [x_{j-1}, x_j]$ for $j = 1, \ldots, n$ and the affine transformation from the reference element $\hat{T} = [0,1]$ onto the finite element T_j, $F_j : \hat{T} = [0,1] \to T_j$, given by $F_j(\hat{x}) = h\hat{x} + x_{j-1}$ with $\hat{x} \in [0,1]$. With these concepts in mind, it is only necessary to introduce the shape of the functional basis in the element of reference to define the global discrete space. In this manner, the definition of the standard polynomial $\mathbb{P}_1$-Lagrange finite element space is given by

$$\mathrm{V}_h^{\mathrm{fe}} = \{ v \in \mathrm{H}^1_{(0}(0,1) : v|_{T_j} \circ F_j^{-1} \in \mathbb{P}^1(\mathbb{C}) \text{ for } j = 1, \ldots, n \}$$

or equivalently, if the local shape functions $\hat{\theta}_1^{\mathrm{fe}}(\hat{x}) = \hat{x}$ and $\hat{\theta}_2^{\mathrm{fe}}(\hat{x}) = 1 - \hat{x}$ are considered, then

$$\mathrm{V}_h^{\mathrm{fe}} = \{ v \in \mathrm{H}^1_{(0}(0,1) : v|_{T_j} \circ F_j^{-1} \in \langle \hat{\theta}_1^{\mathrm{fe}}, \hat{\theta}_2^{\mathrm{fe}} \rangle \text{ for } j = 1, \ldots, n \}.$$

From these finite local shape functions, the canonical basis $\{\varphi_j\}_{j=0}^n$ (consisting of the so-called *hat* functions) can be defined by

$$\varphi_j = \begin{cases} \hat{\theta}_1^{\mathrm{fe}} \circ F_j^{-1} & \text{in } T_j, \\ \hat{\theta}_2^{\mathrm{fe}} \circ F_{j+1}^{-1} & \text{in } T_{j+1}, \\ 0 & \text{otherwise}, \end{cases}$$

for $j = 0, \ldots, n$. The FEM discrete basis for $j = 0$ and $j = n$ are defined from the definitions written above but only taking into account their expressions in T_1 and T_n, respectively. Obviously, as it has been stated previously φ_j is a continuous piecewise linear function satisfying $\varphi_j(x_l) = \delta_{jl}$ for $j, l = 0, \ldots, n$.

To mimic this description in the case of the PUFEM discretization, first notice that the restrictions on T_j and T_{j+1} of the piecewise linear functions multiplied by sine and cosine functions and composed respectively with F_j^{-1} and F_{j+1}^{-1} lead to the PUFEM shape functions $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_1^{\mathfrak{b}}$, and $\hat{\theta}_2^{\mathfrak{b}}$ given by

$$\varphi_j(F_j(\hat{x}))\cos((k+\delta)(F_j(\hat{x})-x_j)) = \hat{x}\cos((k+\delta)h(1-\hat{x})) = \hat{\theta}_1(\hat{x}), \tag{1.16}$$

$$\varphi_j(F_{j+1}(\hat{x}))\cos((k+\delta)(F_{j+1}(\hat{x})-x_j)) = (1-\hat{x})\cos((k+\delta)h\hat{x}) = \hat{\theta}_2(\hat{x}), \tag{1.17}$$

$$\frac{1}{(k+\delta)h}\varphi_j(F_j(\hat{x}))\sin((k+\delta)(F_j(\hat{x})-x_j)) = \hat{x}\frac{\sin((k+\delta)h(\hat{x}-1))}{(k+\delta)h} = \hat{\theta}_1^{\mathfrak{b}}(\hat{x}), \tag{1.18}$$

$$\frac{1}{(k+\delta)h}\varphi_j(F_{j+1}(\hat{x}))\sin((k+\delta)(F_{j+1}(\hat{x})-x_j)) = (1-\hat{x})\frac{\sin((k+\delta)h\hat{x})}{(k+\delta)h} = \hat{\theta}_2^{\mathfrak{b}}(\hat{x}), \tag{1.19}$$

for $\hat{x} \in \hat{T} = [0,1]$ and for any fixed $j = 1, \dots, n-1$. Taking into account these four local shape functions, it can be defined a discrete basis $\{\psi_j^{\mathfrak{v}}\}_{j=0}^n \cup \{\psi_j^{\mathfrak{b}}\}_{j=0}^n$ for the PUFEM space X_h where (applying (1.16)-(1.17))

$$\psi_j^{\mathfrak{v}}(x) = \begin{cases} \hat{\theta}_1 \circ F_j^{-1}(x) & \text{for } x \in T_j, \\ \hat{\theta}_2 \circ F_{j+1}^{-1}(x) & \text{for } x \in T_{j+1}, \\ 0 & \text{otherwise.} \end{cases}$$

$$= \left\{ \begin{array}{ll} F_j^{-1}(x)\cos((k+\delta)h(1-F_j^{-1}(x))) & \text{for } x \in T_j, \\ (1-F_{j+1}^{-1}(x))\cos((k+\delta)hF_{j+1}^{-1}(x)) & \text{for } x \in T_{j+1}, \\ 0 & \text{otherwise.} \end{array} \right\} = \varphi_j(x)\cos((k+\delta)h(x-x_j)),$$

for $j = 1, \dots, n$. Analogous computations also show that applying (1.18)-(1.19), it holds

$$\psi_j^{\mathfrak{b}}(x) = \left\{ \begin{array}{ll} \hat{\theta}_1 \circ F_j^{-1}(x) & \text{for } x \in T_j, \\ \hat{\theta}_2 \circ F_{j+1}^{-1}(x) & \text{for } x \in T_{j+1}, \\ 0 & \text{otherwise.} \end{array} \right\} = \frac{1}{(k+\delta)h}\varphi_j(x)\sin((k+\delta)h(x-x_j)).$$

The PUFEM discrete basis for $j = 0$ and $j = n$ are defined from the definitions written above but only taking into account their expressions in T_1 and T_n, respectively.

Consequently, the PUFEM discrete space $\mathrm{V}_h = \mathrm{X}_h \cap \mathrm{H}^1_{(0}(0,1)$ can be written as the direct sum $\mathrm{V}_h = \mathrm{V}_h^{\mathfrak{v}} \oplus \mathrm{V}_h^{\mathfrak{b}}$ where

$$\begin{aligned} \mathrm{V}_h^{\mathfrak{v}} &= \{v \in \mathrm{H}^1_{(0}(0,1) : \ v|_{T_j} \circ F_j^{-1} \in \langle \hat{\theta}_1, \hat{\theta}_2 \rangle \text{ for } j = 1, \dots, n\} \\ &= \{v \in \mathrm{H}^1_{(0}(0,1) : \ v|_{T_j} \in \langle \psi_{j-1}^{\mathfrak{v}}, \psi_j^{\mathfrak{v}} \rangle \text{ for } j = 1, \dots, n\} \\ &= \{v \in \mathrm{H}^1_{(0}(0,1) : \ v \in \langle \{\psi_j^{\mathfrak{v}}\}_{j=0}^n \rangle\} = \langle \{\psi_j^{\mathfrak{v}}\}_{j=1}^n \rangle \end{aligned} \tag{1.20}$$

and

$$\mathrm{V}_h^{\mathfrak{b}} = \{v \in \mathrm{H}^1_{(0}(0,1) : \ v \in \langle \{\psi_j^{\mathfrak{b}}\}_{j=0}^n \rangle\} = \langle \{\psi_j^{\mathfrak{b}}\}_{j=0}^n \rangle. \tag{1.21}$$

The first equality in the definition of $\mathrm{V}_h^{\mathfrak{v}}$ in (1.20) is straightforward since $\psi_{j-1}^{\mathfrak{v}}|_{T_j} \circ F_j^{-1} = \hat{\theta}_2$ and $\psi_j^{\mathfrak{v}}|_{T_j} \circ F_j^{-1} = \hat{\theta}_1$. The second equality is also deduced immediately from the continuity of the functions belonging to $\mathrm{V}_h^{\mathfrak{v}}$ and the fact that $\psi_j^{\mathfrak{v}}(x_l) = \delta_{kl}$ for $k, l = 1, \dots, n$.

***Remark* 1.3.1** (Vertex-value space). *In addition, due to the definition of $\{\psi_j^{\mathfrak{v}}\}_{j=1}^n$, the degrees of freedom associated to those functions v_h in $\mathrm{V}_h^{\mathfrak{v}}$ are the values on the vertices of the mesh. In fact, it holds*

$$v_h = \sum_{j=1}^{n} v_h(x_j)\psi_j^{\mathfrak{v}}.$$

for all $v_h \in \mathrm{V}_h^{\mathfrak{v}}$. This is the reason because of the discrete space $\mathrm{V}_h^{\mathfrak{v}}$ will be called vertex-valued discrete space.

***Remark* 1.3.2** (Twin-bubble space). *On the contrary to the case of the vertex-value space $\mathrm{V}_h^{\mathfrak{v}}$, the discrete space $\mathrm{V}_h^{\mathfrak{v}}$ defined in* (1.21) *cannot be described in terms of the local shape functions $\hat{\theta}_1^{\mathfrak{b}}, \hat{\theta}_2^{\mathfrak{b}}$. More precisely, it is easy to check that*

$$\mathrm{V}_h^{\mathfrak{b}} \subset \{v \in \mathrm{H}^1_{(0}(0,1) : \ v|_{T_j} \circ F_j^{-1} \in \langle \hat{\theta}_1^{\mathfrak{b}}, \hat{\theta}_2^{\mathfrak{b}} \rangle \text{ for } j = 1, \ldots, n\}.$$

Additionally, to check that the equality of the two discrete spaces written above does not hold, consider the discrete function z_J which is null in all the elements except at element T_J where $z_J|_{T_J} \circ F_J^{-1} = \hat{\theta}_1^{\mathfrak{b}}$. Obviously, z_J is continuous since $z_J(x_J) = z_J(x_{J+1}) = 0$ but $z_J \notin \mathrm{V}_h^{\mathfrak{b}}$ since it is not a linear combination of functions in the span of $\{\psi_j^{\mathfrak{b}}\}_{j=0}^n$.

In fact, despite the functions $v_h^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}$ hold that their values are null at the vertices of the mesh, they do not behave as typical bubble functions in standard piecewise continuous $\mathbb{P}^p$-finite elements with $p \geq 2$. In that case, the polynomial bubbles have support on an unique element mesh T_j. However, in this PUFEM discretization the bubbles functions $\{\psi_j^{\mathfrak{b}}\}_{j=0}^n$ extend their support to two adjacent elements $T_j \cup T_{j+1}$, and at the interior of each element, its local shape resembles the classical polynomial bubbles (with opposite sign in each element). That is the reason because through the rest of this work, the discrete space $\mathrm{V}_h^{\mathfrak{b}}$ will be called as twin-bubble discrete space.

***Remark* 1.3.3** ($\mathbb{P}^2$-finite element limit). *If it is taken into account the expressions of the local shape functions* (1.16)-(1.19), *if the parameter $h(k+\delta)$ tends to zero, $\hat{\theta}_1(x)|_{h(k+\delta)=0} = \hat{x}$, $\hat{\theta}_2(x)|_{h(k+\delta)=0} = 1 - \hat{x}$, and $\lim_{h(k+\delta)\to 0} \hat{\theta}_1^{\mathfrak{b}}(x) = \lim_{h(k+\delta)\to 0} \hat{\theta}_1^{\mathfrak{b}}(x) = \hat{x}(1-\hat{x})$. Hence, in the limit case, $\mathrm{V}_h^{\mathfrak{v}} = \mathrm{V}_h^{\mathrm{fe}}$ and $\mathrm{V}_h^{\mathfrak{b}}$ is the the classical bubble space associated to piecewise continuous $\mathbb{P}^2$-finite elements. In conclusion, in the limit when $h(k+\delta) \to 0$, the local discrete matrices associated to the PUFEM discretization coincides with those ones computed with the standard $\mathbb{P}^2$-finite element method.*

1.4 Interpolation estimates

A typical error estimation analysis for a Galerkin method, and in particular, for a finite element approximation, requires the use of an interpolation error, which mimics the error behaviour of the projection operator in the discrete space. Usually, the standard piecewise $\mathbb{P}_n$-finite element methods uses piecewise local polynomial interpolants in each element and an error estimation for the interpolant function is obtained by using a Taylor expansion.

However, in the case of enriched methods where the discrete basis is not polynomial locally, analogous arguments could lead to unsharp estimates and consequently to inaccurate error estimates for the global method. In addition, for enriched methods, there exists an extra difficulty coming from the fact that the degrees of freedom cannot be identified as nodal values of the discrete function but simply as amplitude coefficients of each discrete basis function.

In the case of the PUFEM method, the derivation of an approximability result in the PUFEM discrete space X_h should to overcome these challenges. With this aim, the design an accurate interpolant-like operator involves two strategies, which are going to be combined to obtain an accurate and computational efficient discrete approximation of a given function. First, an interpolant-like operator will be defined valid for any mesh size h, which will be qualified as pre-asymptotic. Second, for h small enough, a $\mathbb{P}_2$-based interpolant will be recast for the PUFEM discrete space, which will be identified as an asymptotic interpolant. Finally, the combination of both approximations leads to an interpolant-like procedure which approximates accurately the H^1-projection in the PUFEM discrete space.

***Remark* 1.4.1** (Oscillatory solutions). *Usually, for any polynomial-based finite element method the most challenging functions to be accurately approximated are those ones which are highly oscillatory. In the context of the Helmholtz equation, and in particular, in the present work, it will be qualified as oscillatory solutions those functions which are solution of the homogeneous Helmholtz equation. In what follows, all the interpolation estimates will be focused on this kind of oscillatory solutions.*

1.4.1 Pre-asymptotic interpolant-like operator

The first step to analyse the error in the PUFEM discretization consists in the derivation of an approximability result in X_h. To define the first interpolant-like operator, any smooth function v will be split in two parts attending to the intensity orientation, trying to identify which part of the function will be accurate approximated by planewaves which travel to the right (and whose intensity vector points towards the positive axis) and by those ones which travel to the left (and whose intensity vector points towards the negative axis). Such spitting is given by the differential operators involved in the Sommerfeld radiation condition, this is, for $v \in \mathcal{C}^1(0,1)$

$$v = \frac{1}{2ik}\left(v' + ikv\right) - \frac{1}{2ik}\left(v' - ikv\right), \tag{1.22}$$

and hence, the first term will be approximated by the discrete functions $\{\psi_j^+\}_{j=0}^n$ and the second one will be approximated by $\{\psi_j^-\}_{j=0}^n$. In consequence, given $v \in \mathrm{H}^2(0,1)$, the interpolant-like $\mathcal{I}_p v \in \mathrm{X}_h$ is defined by

$$\mathcal{I}_p v = \frac{1}{2ik}\sum_{j=0}^{n}\left(\left(v'(x_j) + ikv(x_j)\right)\psi_j^+ - \left(v'(x_j) - ikv(x_j)\right)\psi_j^-\right). \tag{1.23}$$

From (1.23), it is immediate to check that

$$
\begin{aligned}
(\mathcal{I}_p v)(x) &= \sum_{j=0}^{n} \left(\frac{v'(x_j)}{k} \varphi_j(x) \sin((k+\delta)(x-x_j)) + v(x_j)\varphi_j(x)\cos((k+\delta)(x-x_j)) \right) \\
&= \sum_{j=0}^{n} \left(\frac{v'(x_j)}{k} \psi_j^{\mathfrak{b}}(x) + v(x_j)\psi_j^{\mathfrak{v}}(x) \right) \qquad (1.24)
\end{aligned}
$$

and hence and it trivially holds $\mathcal{I}_p v(x_j) = x_j$. However, despite of use the point-wise values of the derivatives at the mesh nodes $v'(x_j)$, $(\mathcal{I}_p v)'(x_j) \neq v'(x_j)$, and, moreover, $\mathcal{I}_p v$ does not belong to $\mathcal{C}^1(0,1)$ and so the point wise evaluation of the derivative of $\mathcal{I}_p v$ is not well-defined.

An unusual feature of interpolant-like operator $\mathcal{I}_p : \mathrm{H}^2(0,1) \to \mathrm{X}_h$ is that $\mathcal{I}_p v_h \neq v_h$ for all $v \in \mathrm{X}_h$. To check this fact, it is enough to consider $v_h(x) = e^{+i(k+\delta)x}$, which can be written in the PUFEM discrete basis as

$$
v_h = \sum_{j=0}^{n} e^{i(k+\delta)x_j} \psi_j^{+}.
$$

If the above expression is compared with $\mathcal{I}_p v_h$, which is given by (from (1.23))

$$
\mathcal{I}_p v_h = \sum_{j=0}^{n} \left(\left(1 + \frac{\delta}{2k}\right) e^{i(k+\delta)x_j} \psi_j^{+} - \frac{\delta}{2k} e^{i(k+\delta)x_j} \psi_j^{-} \right),
$$

it is deduced immediately that the coefficients of the basis representation is different from $\delta \neq 0$ and only if $\delta = 0$ then $\mathcal{I}_p v_h = v_h$. Anyway, it is fulfilled that $\mathcal{I}_p(\mathrm{X}_h) \subseteq \mathrm{X}_h$ and $\mathcal{I}_p(\mathrm{V}_h) \subseteq \mathrm{V}_h$.

Despite this atypical behaviour for a interpolant-like operator, the approximation computed when this interpolant procedure is applied to a the linear combination of planewaves (solution of the homogeneous Helmholtz equation $-u'' - k^2 u = 0$) is highly accurate. In what follows, it will be assumed that $\delta/k \leq 1$ to ensure that the wave number perturbation δ introduces a relative error with respect to the exact wave number k smaller than 100% in the PUFEM discretization.

Lemma 1.4.2. *Given $\epsilon > 1$, if $u \in \mathrm{V}$ is a solution of the homogeneous Helmholtz equation with wave number $k > \epsilon$ and assuming $\delta/k \leq 1$, then there exists the interpolant-like discrete function $u_{\mathrm{I}} = \mathcal{I}_p u \in X_h$ defined by* (1.23) *satisfies*

$$
\inf_{v_h \in \mathrm{X}_h} \|u - v_h\|_0 \leq \|u - \mathcal{I}_p u\|_0 \leq C h^2 \delta^2 \|u\|_0, \qquad (1.25)
$$

$$
\inf_{v_h \in \mathrm{X}_h} |u - v_h|_1 \leq |u - \mathcal{I}_p u|_1 \leq \left(\frac{Ch}{k} + \hat{C} h^2 \right) \delta^2 |u|_1, \qquad (1.26)
$$

where the positive constants C and $\hat{C}$ only depend on ϵ.

Proof. Firstly, notice that $u \in \mathrm{H}^2(0,1)$ since u is the solution of the Helmholtz equation with null right-hand side and so $\mathcal{I}_p u$ is well-defined. Now, since $u_{\mathrm{I}} = \mathcal{I}_p u \in X_h$, any restriction of u_{I} to the mesh element $[x_j, x_{j+1}]$ should be written as a linear combination of basis functions in X_h, which are not null on that element, this is, those functions multiplied by φ_j and φ_{j+1}. Hence, it is satisfied

$$\begin{aligned} u_{\mathrm{I}}(x) =& \alpha_{1j}\psi_j^+(x) + \alpha_{2j}\psi_{j+1}^+(x) + \alpha_{3j}\psi_j^-(x) + \alpha_{4j}\psi_{j+1}^-(x) \\ =& \alpha_{1j}\varphi_j(x)e^{i(k+\delta)(x-x_j)} + \alpha_{2j}\varphi_{j+1}(x)e^{i(k+\delta)(x-x_{j+1})} \\ &+ \alpha_{3j}\varphi_j(x)e^{-i(k+\delta)(x-x_j)} + \alpha_{4j}\varphi_{j+1}(x)e^{-i(k+\delta)(x-x_{j+1})}, \qquad \text{for } x \in [x_j, x_{j+1}]. \end{aligned}$$

Since the exact solution for the homogeneous Helmholtz equation is given by $u(x) = Ae^{ikx} + Be^{-ikx}$, from (1.23) it is easy to check that $\alpha_{1j} = Ae^{ikx_j}$, $\alpha_{2j} = Ae^{ikx_{j+1}}$, $\alpha_{3j} = Be^{-ikx_j}$ and $\alpha_{4j} = Be^{-ikx_{j+1}}$ for $j = 1, \ldots, n$, and so it holds

$$\begin{aligned} \|u - u_{\mathrm{I}}\|_0^2 \le& |A|^2 \sum_{j=1}^n \int_{x_j}^{x_{j+1}} \left|e^{ikx} - e^{-i\delta x_j}\varphi_j(x)e^{i(k+\delta)x} - e^{-i\delta x_{j+1}}\varphi_{j+1}(x)e^{i(k+\delta)x}\right|^2 \mathrm{d}x \\ &+ |B|^2 \sum_{j=1}^n \int_{x_j}^{x_{j+1}} \left|e^{-ikx} - e^{i\delta x_j}\varphi_j(x)e^{-i(k+\delta)x} - e^{i\delta x_{j+1}}\varphi_{j+1}(x)e^{-i(k+\delta)x}\right|^2 \mathrm{d}x. \end{aligned}$$

Both integrals can be computed explicitly and it is immediate to check that they are identical (the integrands are the square modulus of conjugate expressions) and independent of the mesh element j and the wave number k. For instance, in the case of the first integral, if $\delta = 0$ then the integral value is null. Otherwise, for $\delta \neq 0$, it holds

$$\begin{aligned} &\int_{x_j}^{x_{j+1}} \left|e^{ikx} - e^{-i\delta x_j}\varphi_j(x)e^{i(k+\delta)x} - e^{-i\delta x_{j+1}}\varphi_{j+1}(x)e^{i(k+\delta)x}\right|^2 \mathrm{d}x \\ &= \int_{x_j}^{x_{j+1}} \left|e^{ikx}\left(1 - \varphi_j(x)e^{i\delta(x-x_j)} - \varphi_{j+1}(x)e^{i\delta(x-x_{j+1})}\right)\right|^2 \mathrm{d}x \\ &= \int_0^h \left|1 - \frac{h-x}{h}e^{i\delta x} - \frac{x}{h}e^{i\delta(x-h)}\right|^2 \mathrm{d}x = \frac{5}{3}h - \frac{4}{\delta^2 h} + \left(\frac{h}{3} + \frac{4}{\delta^2 h}\right)\cos(\delta h). \end{aligned}$$

Taking into account the identical contribution of the integrals in each element of the mesh, it is obtained

$$\|u - u_{\mathrm{I}}\|_0^2 \le \left(|A|^2 + |B|^2\right)\frac{1}{h}\left(\frac{5}{3}h - \frac{4}{\delta^2 h} + \left(\frac{h}{3} + \frac{4}{\delta^2 h}\right)\cos(\delta h)\right).$$

Now, using the bound derived in (1.96) for the expression written above between parenthesis (see Appendix 1.A for further details), it holds

$$\|u - u_{\mathrm{I}}\|_0^2 \le= \left(|A|^2 + |B|^2\right)\frac{17}{360}\delta^4 h^4 \le \frac{17C_\epsilon}{360}\delta^4 h^4 \|u\|_0^2. \tag{1.27}$$

To obtain (1.27), it has been used

$$|A|^2 + |B|^2 \leq C_\epsilon \|u\|_0^2, \tag{1.28}$$

with $C_\epsilon > 0$ depending only on ϵ. This bound can be derived immediately since the exact solution is given by $u(x) = Ae^{ikx} + Be^{-ikx}$ and so

$$\begin{aligned}\|u\|_0^2 &= \int_0^1 |Ae^{ikx} + Be^{-ikx}|^2 \,\mathrm{d}x = \int_0^1 \left(|A|^2 + |B|^2 + 2\mathrm{Re}(A\bar{B}e^{2kix})\right) \,\mathrm{d}x \\ &= |A|^2 + |B|^2 - \mathrm{Re}(A\bar{B})\frac{\sin(2k)}{k} - \mathrm{Im}(A\bar{B})\frac{\cos(2k)-1}{k} \\ &\geq |A|^2 + |B|^2 - \frac{2}{k}|A||B| \geq \left(1 - \frac{1}{k}\right)(|A|^2 + |B|^2) \geq \left(1 - \frac{1}{\epsilon}\right)(|A|^2 + |B|^2),\end{aligned}$$

where it has been used that $1 - 1/k$ is a monotonically decreasing function, bounded in $[\epsilon, +\infty)$, and so satisfying the estimate (1.28) with $C_\epsilon = 1 - 1/\epsilon$.

To obtain estimate (1.26), analogous arguments can be used to bound the H^1-seminorm of $u - u_\mathrm{I}$. In fact, for $\delta \neq 0$, straightforward computations on the integral contribution at each mesh element and taking into account the bound derived in (1.97) (see Appendix 1.A for details) leads to

$$\begin{aligned}|u - u_\mathrm{I}|_1^2 \leq &(|A|^2 + |B|^2) \sum_{j=1}^{n} \left(\frac{2}{h} + hk^2 + \frac{2}{3}h(k+\delta)^2 - \frac{4k^2}{\delta^2 h} - 2(k+\delta)\sin(\delta h)\right. \\ &\left. + 2\cos(\delta h)\left(-\frac{1}{h} + \frac{h}{6}(k+\delta)^2 + \frac{2k^2}{\delta^2 h}\right)\right) \\ \leq &(|A|^2 + |B|^2)\frac{1}{h}\left(\frac{1}{12}\delta^4 h^3 + \frac{1}{30}\delta^6 h^5 + \frac{2}{45}\delta^5 k h^5 + \frac{7}{360}\delta^4 k^2 h^5\right) \\ \leq &\frac{C_\epsilon}{12}\delta^4 \frac{h^2}{k^2}|u|_1^2 + \frac{7C_\epsilon}{72}\delta^4 h^4 |u|_1^2.\end{aligned} \tag{1.29}$$

In the last inequality written above, it has been used that $\delta/k \leq 1$ and an analogous estimate to (1.28) for the H^1-seminorm, this is, $k^2(|A|^2 + |B|^2) \leq C_\epsilon |u|_1^2$. Hence, estimate (1.26) is obtained, with positive constants $C = C_\epsilon/12$ and $\hat{C} = 7C_\epsilon/72$ independent of h, δ and k. □

To illustrate that the approximation estimates (1.25) and (1.26) are sharp, the L^2 and H^1-errors have been computed for the particular case of the Helmholtz exact solution $u(x) = \sin(kx)$. Figures 1.1 and 1.2 show respectively the L^2 and H^1-error curves varying the mesh size for different values of the wave number k and the perturbation parameter δ.

1.4.2 Asymptotic $\mathbb{P}_2$-based interpolant

The second ingredient in the definition of a global accurate interpolant procedure consists in a truly interpolant procedure, which will be defined for $h(k+\delta) \leq \alpha < \pi$ and

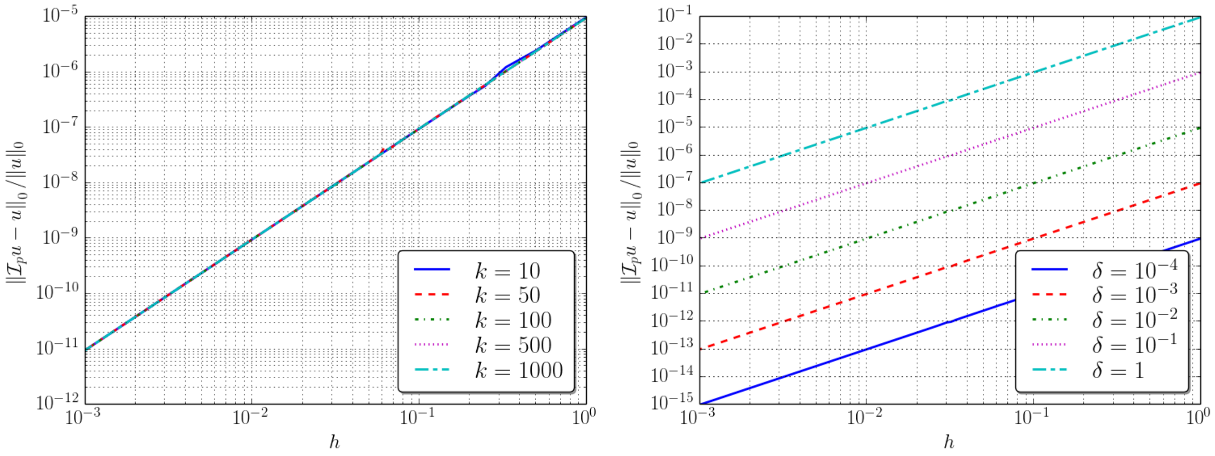

Figure 1.1: L^2-approximation errors of the interpolant-like $\mathcal{I}_p$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

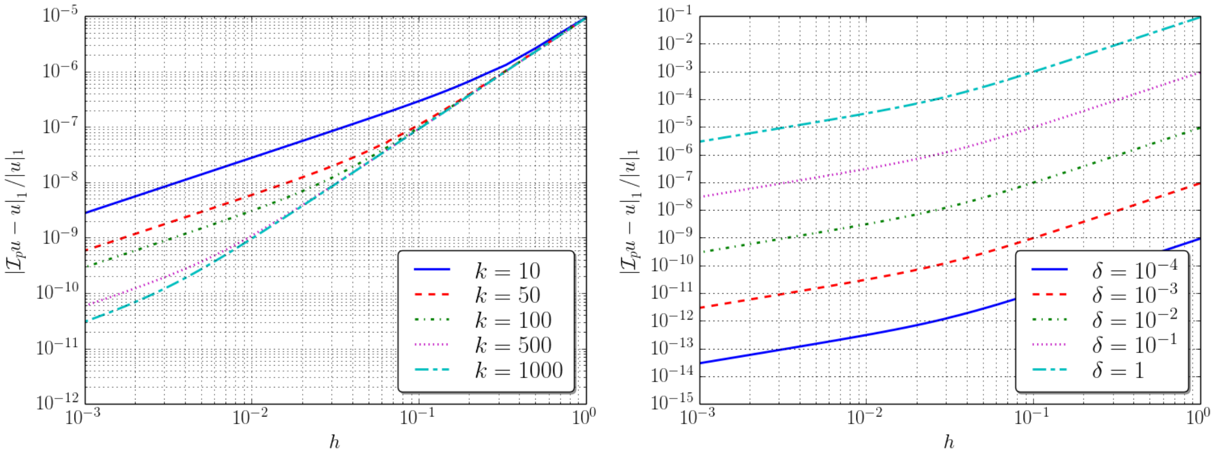

Figure 1.2: H^1-approximation errors of the interpolant-like $\mathcal{I}_p$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

this fact will be the reason because of it is qualified as asymptotic. This new interpolant $\mathcal{I}_2 : H^1(0,1) \rightarrow X_h$ reassembles the definition of the standard $\mathbb{P}_2$-interpolant since it is defined as follows: given $v \in \mathrm{H}^1(0,1)$,

$$\begin{aligned}\mathcal{I}_2 v(x) &= \sum_{j=0}^{n} \left(\gamma_j^{\mathfrak{b}} \psi_j^{\mathfrak{b}}(x) + \gamma_j^{\mathfrak{v}} \psi_j^{\mathfrak{v}}(x)\right) \\ &= \sum_{j=0}^{n} \left(\gamma_j^{\mathfrak{b}} \varphi_j(x) \sin((k+\delta)(x-x_j)) + \gamma_j^{\mathfrak{v}} \varphi_j(x) \cos((k+\delta)(x-x_j))\right), \quad (1.30)\end{aligned}$$

will be defined by imposing the conditions

$$\mathcal{I}_2 v(x_j) = v(x_j) \qquad \text{for } j = 0, \ldots, n, \tag{1.31}$$

$$\mathcal{I}_2 v\left(\frac{x_j + x_{j+1}}{2}\right) = v\left(\frac{x_j + x_{j+1}}{2}\right) \text{ for } j = 0, \ldots, n-1. \tag{1.32}$$

Since these conditions forms a set of $2n+1$ linear equations (in terms of the coefficients $\gamma_j^{\mathfrak{v}}, \gamma_j^{\mathfrak{b}}$ of $\mathcal{I}_2 v$ in the discrete PUFEM basis $\{\psi_j^{\mathfrak{v}}, \psi_j^{\mathfrak{b}}\}_{j=0}^{n}$), an additional equation is required to have a well-posed linear problem with an unique solution (see Remark 1.6.2 for a detailed discussion). Arbitrarily, it will be fixed that $\gamma_0^{\mathfrak{b}} = 0$. From (1.30)-(1.31), it is clear that $\gamma_j^{\mathfrak{v}} = v(x_j)$. So, since $\gamma_0^{\mathfrak{b}} = 0$, it is only necessary to compute the coefficients $\gamma_j^{\mathfrak{b}}$ for $j = 1, \ldots, n-1$. Taking into account (1.32) for $j = 1, \ldots, n$, it is obtained

$$v\left(\frac{x_j + x_{j+1}}{2}\right) = \gamma_j^{\mathfrak{b}} \varphi_j\left(\frac{x_j + x_{j+1}}{2}\right) \sin\left((k+\delta)\frac{h}{2}\right) - \gamma_{j+1}^{\mathfrak{b}} \varphi_{j+1}\left(\frac{x_j + x_{j+1}}{2}\right) \sin\left((k+\delta)\frac{h}{2}\right)$$

and it leads to

$$\gamma_{j+1}^{\mathfrak{b}} = \frac{-2}{\sin\left((k+\delta)\dfrac{h}{2}\right)} \left(v\left(\frac{x_j + x_{j+1}}{2}\right) - \frac{\gamma_j^{\mathfrak{b}}}{2} \sin\left((k+\delta)\frac{h}{2}\right)\right) \text{ for } j = 1, \ldots, n-1.$$

Notice that since $h(k+\delta) < \pi$ then the expression $\sin((k+\delta)h/2)$ will be always strictly positive and the coefficients $\gamma_j^{\mathfrak{b}}$ for $j = 0, \ldots, n$ are always well-defined.

Lemma 1.4.3. *If* $v \in \mathrm{H}^3(0,1)$ *is solution of the Hemlholtz equation then the interpolant discrete function* $\mathcal{I}_2 v \in \mathrm{X}_h$ *defined by* (1.30)-(1.32)*, then there exists a constant* C *independent of* h*,* k *and* δ *such that it holds*

$$\|v - \mathcal{I}_2 v\|_0 \le C h^3 k^3 \|v\|_0, \tag{1.33}$$

$$|v - \mathcal{I}_2 v|_1 \le C h^2 k^2 |v|_1. \tag{1.34}$$

Proof. Let $\mathcal{I}_{\mathbb{P}^2}$ be the continuous piecewise $\mathbb{P}^2$ interpolant. If $v \in \mathrm{H}^3(0,1)$, the order of approximation of this polynomial interpolant is optimal in the sense that (see [28, Section 1.5])

$$\|v - \mathcal{I}_{\mathbb{P}^2} v\|_0 + h|v - \mathcal{I}_{\mathbb{P}^2} v|_1 \le C h^3 |v|_3, \tag{1.35}$$

where C is a constant independent of h. Direct but cumbersome computations analogous to those ones used in Lemma 1.4.2 (taking into account that v is a linear combination of sine and cosine functions and using the Taylor approximations of the PUFEM local basis functions defined in (1.16)-(1.19)) show the same kind of inequality to (1.35) but now involving the PUFEM interpolant $\mathcal{I}_2$:

$$\|v - \mathcal{I}_2 v\|_0 + h|v - \mathcal{I}_2|_1 \leq C h^3 |v|_3,$$

from which (1.33) and (1.34) follows by using that v is an oscillatory solution (and hence $|v|_3 \leq C k^3 \|v\|_0$ and $|v|_3 \leq C k^2 |v|_1$, being C a positive constant independent of k). □

To illustrate that the approximation estimates (1.33) and (1.34) are sharp, the L^2 and H^1-errors have been computed for the particular case of the Helmholtz exact solution $u(x) = \sin(kx)$. Figures 1.3 and 1.4 show respectively the L^2 and H^1-error curves varying the mesh size for different values of the wave number k and the perturbation parameter δ.

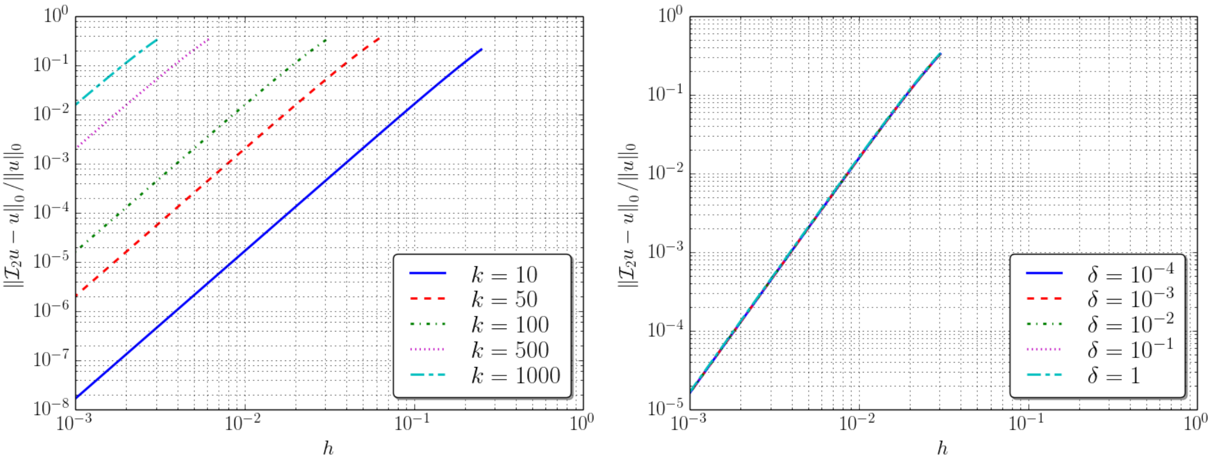

Figure 1.3: L^2-approximation errors of the interpolant-like $\mathcal{I}_2$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

1.4.3 An accurate global interpolation procedure

Finally, the combination of the pre-asymptotic interpolant-like operator $\mathcal{I}_p$ and the asymptotic interpolant $\mathcal{I}_2$ leads to an accurate global interpolant $\mathcal{I}_h$, which will have similar approximation properties to those ones exhibit by the discrete projection operators. These new global operator $\mathcal{I}_h : \mathrm{H}^1(0,1) \to \mathrm{X}_h$ is given by

$$\mathcal{I}_h v = \begin{cases} \mathcal{I}_p v & \text{if } h(k+\delta) \geq \pi, \\ \mathcal{I}_p v + \mathcal{I}_2 v - \mathcal{I}_2 \mathcal{I}_p v & \text{if } h(k+\delta) < \pi. \end{cases} \tag{1.36}$$

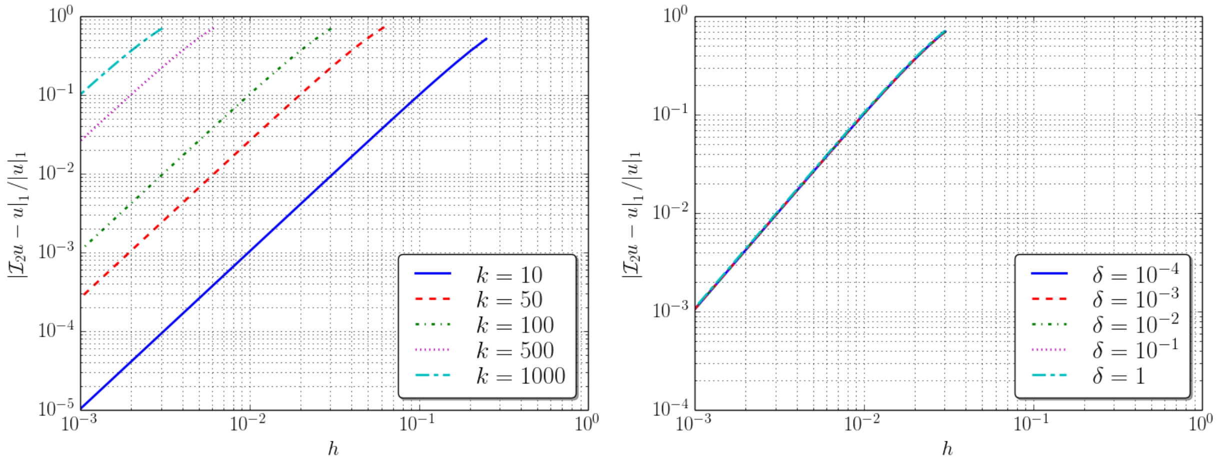

Figure 1.4: H^1-approximation errors of the interpolant-like $\mathcal{I}_2$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

Due to the approximation properties of both interpolant operators, it is easy to obtain estimates for this new operator $\mathcal{I}_h$.

Lemma 1.4.4. *Given $\epsilon > 1$, if $u \in \mathrm{V}$ is a solution of the homogeneous Helmholtz equation with wave number $k > \epsilon$ and assuming $\delta/k \leq 1$, then it holds for $h(k+\delta) > \pi$*

$$\|u - \mathcal{I}_h u\|_0 \leq Ch^2\delta^2\|u\|_0, \tag{1.37}$$

$$|u - \mathcal{I}_h u|_1 \leq \hat{C}h^2\delta^2|u|_1, \tag{1.38}$$

and for $h(k+\delta) < 2\pi$, it holds

$$\|u - \mathcal{I}_h u\|_0 \leq Ch^3k\delta^2\|u\|_0, \tag{1.39}$$

$$|u - \mathcal{I}_h u|_1 \leq \hat{C}h^2\delta^2|u|_1, \tag{1.40}$$

where the positive constants C and $\hat{C}$ do not depend on h, k, and δ.

Proof. Firstly, the case $hk \geq \pi$ will be considered. Since $\mathcal{I}_h$ coincides with $\mathcal{I}_p$, estimate (1.37) follows immediately from (1.25). The H^1-error estimate also is implied by (1.26) since for $hk \geq \pi$ (which is equivalent to $\pi/k < h$), it holds

$$|u - \mathcal{I}_h u|_1 \leq \left(\frac{Ch}{k} + \hat{C}h^2\right) \leq (C + \hat{C})h^2\delta^2|u|_1.$$

Second, the estimates (1.39)-(1.40) will be shown for $h(k+\delta) < 2\pi$. Using the fact that $2\pi/k$ is an upper bound of the mesh size h, it is clear from (1.25) that

$$\|u - \mathcal{I}_p\|_0 \leq Ch^2\delta^2\|u\|_0 \leq \frac{C\delta^2}{k^2}\|u\|_0. \tag{1.41}$$

Then, taking into account the definition of the operator $\mathcal{I}_h$ and estimates (1.33) and (1.41), it holds

$$\begin{aligned}\|u-\mathcal{I}_h u\|_0 &= \|u-(\mathcal{I}_p u+\mathcal{I}_2 u-\mathcal{I}_2\mathcal{I}_p u)\|_0 = \|(u-\mathcal{I}_p u)-\mathcal{I}_2(u-\mathcal{I}_p u)\|_0\\ &\leq Ch^3k^3\|u-\mathcal{I}_p u\|_0 \leq Ch^3k^3\frac{C\delta^2}{k^2}\|u\|_0 = Ch^3k\delta^2\|u\|_0,\end{aligned}$$

and hence (1.39) is obtained. Analogously, from (1.26) and since $h \leq 2\pi/k$, it is deduced that

$$|u-\mathcal{I}_p|_1 \leq Ch^2\delta^2|u|_1 \leq \frac{C\delta^2}{k^2}|u|_1. \tag{1.42}$$

Then, taking into account the definition of the operator $\mathcal{I}_h$ and estimates (1.33) and (1.42), it holds

$$\begin{aligned}|u-\mathcal{I}_h u|_1 &= |u-(\mathcal{I}_p u+\mathcal{I}_2 u-\mathcal{I}_2\mathcal{I}_p u)|_1 = |(u-\mathcal{I}_p u)-\mathcal{I}_2(u-\mathcal{I}_p u)|_1\\ &\leq Ch^2k^2|u-\mathcal{I}_p u|_1 \leq Ch^2k^2\frac{C\delta^2}{k^2}|u|_1 = Ch^2\delta^2|u|_1,\end{aligned}$$

and consequently (1.40) follows. □

***Remark* 1.4.5.** *In view of the arguments used in the proof described above, the operator $\mathcal{I}_h$ can be read as a correction of the interpolant-like operator $\mathcal{I}_p$ using the $\mathcal{I}_2$-interpolant of its approximation error, this is, if e_h denotes the interpolation error made by $\mathcal{I}_p u$ (i.e., $e_h = u - \mathcal{I}_p u$) then the value of the global interpolation for $h(k+\delta) < 2\pi$ is given by $\mathcal{I}_h u = \mathcal{I}_p u + \mathcal{I}_2 e_h$. Hence, it is shown that $\mathcal{I}_h$ is a interpolant operator in V_h at the mesh vertices $\{x_j\}_{j=0}^n$ since*

$$(\mathcal{I}_h u)(x_j) = (\mathcal{I}_p u)(x_j) + (\mathcal{I}_2 e_h)(x_j) = (\mathcal{I}_p u)(x_j) + e_h(x_j) = u(x_j).$$

To illustrate that the approximation estimates (1.37)-(1.40) are sharp, the L^2 and H^1-errors have been computed for the particular case of the Helmholtz exact solution $u(x) = \sin(kx)$. Figures 1.5 and 1.6 show respectively the L^2 and H^1-error curves varying the mesh size for different values of the wave number k and the perturbation parameter δ. The plots in both figures confirm the orders in the parameters h, k, and δ shown in estimates (1.37)-(1.40).

1.4.4 Comparison between projections and interpolants

Finally, to illustrate that the approximability estimates based on the global interpolant procedure defined in (1.36), it will be checked numerically that the error estimates (1.37)-(1.40) coincide with those ones computed for the projection operators onto the PUFEM discrete space. With this purpose, the L^2 and H^1-distances between the exact solution u of a Hemlholtz problem and their projections in the PUFEM discrete space V_h have been computed.

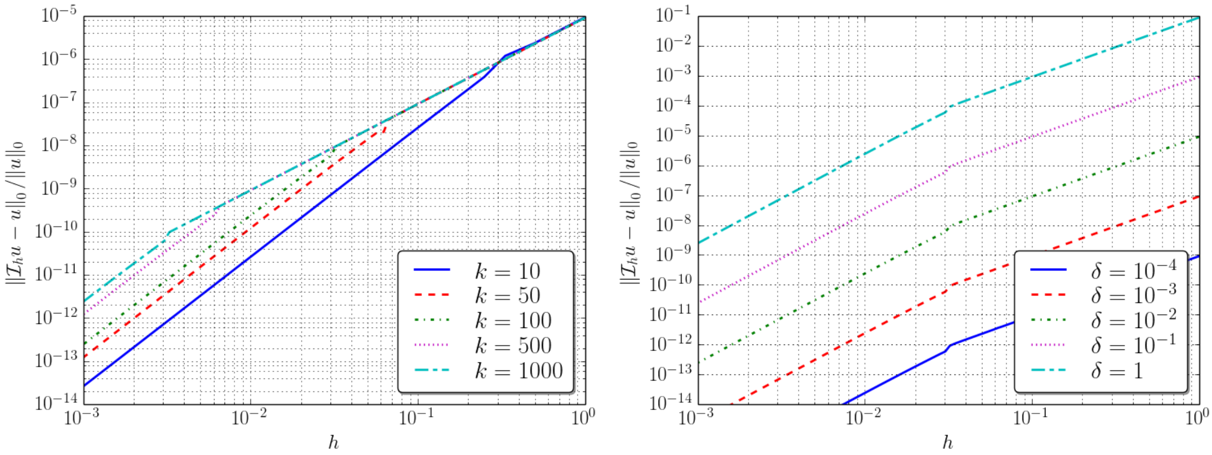

Figure 1.5: L^2-approximation errors of the global interpolant procedure $\mathcal{I}_h$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

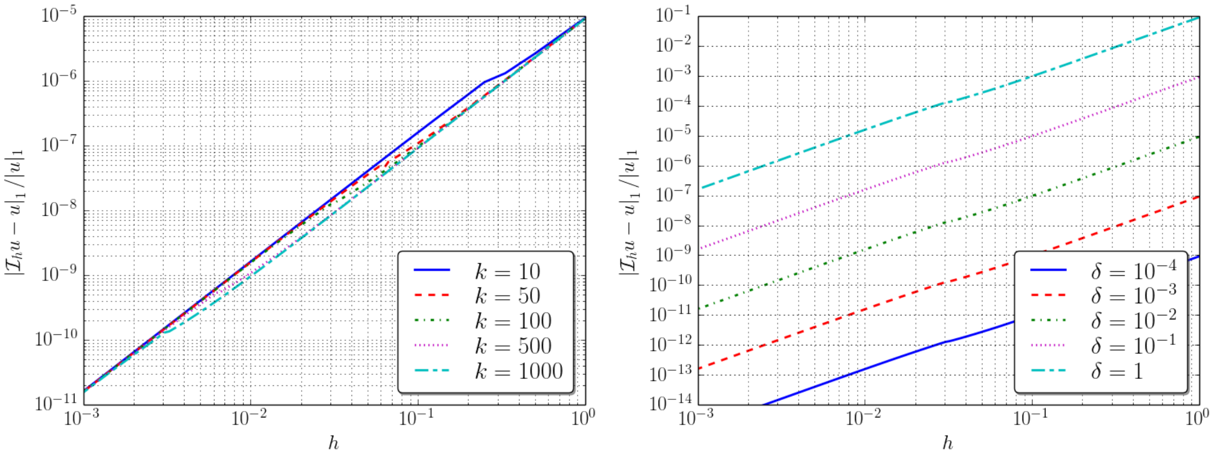

Figure 1.6: H^1-approximation errors of the global interpolant procedure $\mathcal{I}_h$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

Firstly, the L^2 and H^1-projections have been computed numerically. In what follows, the definition of the projection operators are described in detail, highlighting the discrete matrices involved in those computations. Given a function $g \in L^2(0,1)$, its L^2-projection onto the PUFEM discrete space V_h is defined as the solution of the following discrete problem:

$$\begin{cases}\text{Given } g \in \mathrm{L}^2(0,1), \text{ find } \Pi^h_{\mathrm{L}^2} g = r_h \in \mathrm{V}_h \text{ such that} \\ \displaystyle\int_0^1 r_h(x)\bar{v}_h(x)\,\mathrm{d}x = \int_0^1 g(x)\bar{v}_h(x)\,\mathrm{d}x \quad \forall v_h \in \mathrm{V}_h,\end{cases} \tag{1.43}$$

Since $r_h \in \mathrm{V}_h$ then it admits the discrete basis representation $r_h = \sum_{j=0}^n (r_j^+\psi_j^+ + r_j^-\psi_j^-)$ and it is straightforward to check that the associated vector $\vec{r}_h = (r_0^-, r_0^+, \dots, r_n^-, r_n^+)^t$ is the solution of the linear system $\mathcal{M}_h\vec{r}_h = \vec{g}_h$ being $\vec{g}_h = (g_0^-, g_0^+, \dots, g_n^-, g_n^+)^t$ with $g_j^\pm = \int_0^1 g(x)\bar{\psi}_j^\pm(x)\,\mathrm{d}x$ for $j = 0, \dots, n$, under the restriction $r_0^- + r_0^+ = 0$ (to impose that $r_h(0) = 0$).

Analogously, the H^1-projection of $g \in \mathrm{V}$ onto V_h is defined as the solution s_h of the PUFEM problem

$$\int_0^1 s_h'(x)\bar{v}_h'(x)\,\mathrm{d}x = \int_0^1 g'(x)\bar{v}_h'(x)\,\mathrm{d}x \quad \forall v_h \in \mathrm{V}_h. \tag{1.44}$$

In particular, if $g \in \mathrm{V}$ is assumed to be the solution of the Helmholtz equation, it holds $-g'' = f + k^2 g$ with $f \in \mathrm{L}^2(0,1)$ and $g(0) = 0$. Hence, by standard elliptic regularity results for second-order differential operators with constant coefficients, $g \in \mathrm{H}^2(0,1)$. In consequence, by integrating by parts the right-hand side of the weak form (1.44), the definition of the H^1-projection s_h can be rewritten in terms of a variation problem stated in V_h (involving only the derivative of g evaluated at $x = 1$, which is well-defined since g' is continuous) as follows: Given $g \in \mathrm{V}$ solution of the Helmholtz equation with right-hand side f,

$$\begin{cases}\text{Find } \Pi^h_{\mathrm{H}^1} g = s_h \in \mathrm{V}_h \text{ such that} \\ \displaystyle\int_0^1 s_h'(x)\bar{v}_h'(x)\,\mathrm{d}x = \int_0^1 (f(x) + k^2 g(x))\bar{v}_h(x)\,\mathrm{d}x + g'(1)\bar{v}_h(1) \quad \forall v_h \in \mathrm{V}_h.\end{cases}$$

Again, since $s_h \in \mathrm{V}_h$ then it admits the discrete basis representation $s_h = \sum_{j=0}^n (s_j^+\psi_j^+ + s_j^-\psi_j^-)$ and it is straightforward to check that the associated vector of coefficients $\vec{s}_h = (s_0^-, s_0^+, \dots, s_n^-, s_n^+)^t$ is the solution of the linear system $\mathcal{K}_h\vec{s}_h = \vec{g}_h$ where the right-hand side is given by $\vec{g}_h = (g_0^-, g_0^+, \dots, g_n^-, g_n^+)^t$ with

$$\begin{aligned} g_j^\pm &= \int_0^1 (f(x) + k^2 g(x))\bar{\psi}_j^\pm(x)\,\mathrm{d}x \qquad \text{for } j = 0, \dots, n-1, \\ g_n^\pm &= \int_0^1 (f(x) + k^2 g(x))\bar{\psi}_n^\pm(x)\,\mathrm{d}x + g'(1), \end{aligned}$$

under the restriction $s_1 + s_2 = 0$. Notice that the vector $\vec{g}_h = \mathcal{M}_h\vec{s}_h + (0, \dots, 0, g'(1), g'(1))^t$ where $\vec{s}_h$ is the coefficient vector associated to the L^2-projection of $f + k^2 g$.

To compute the errors between the projections (or the interpolants) in V_h and the exact solution of the model problem in V, special care should be paid to not introduce further quadrature errors which could distort the numerical behaviour with respect to the parameters h, k, and δ. With this aim in mind, the computation of the L^2-distance, $\|v - v_h\|_0$ between a function $v \in \mathrm{V}$ and another one in $v_h \in \mathrm{V}_h$ is computed as follows:

$$\begin{aligned}\|v - v_h\|_0^2 &= \int_0^1 (v(x) - v_h)(\bar{v}(x) - \bar{v}_h(x))\,\mathrm{d}x = \int_0^1 \big(|v(x)|^2 + |v_h(x)|^2 - 2\mathrm{Re}(\bar{v}(x)v_h(x))\big)\mathrm{d}x\\ &= \|v\|_0^2 + \vec{v}_h^*\mathcal{M}_h\vec{v}_h - 2\mathrm{Re}(\vec{r}_h^*\mathcal{M}_h\vec{v}_h) = \|v\|_0^2 + \vec{v}_h^*\mathcal{M}_h\vec{v}_h - 2\mathrm{Re}(\vec{g}_h^*\vec{v}_h),\end{aligned}$$

where $\vec{r}_h$ and $\vec{g}_h$ are the coefficient vectors involved in the computation of the L^2-projection $\Pi^h_{\mathrm{L}^2}v$, which satisfy $\mathcal{M}_h\vec{r}_h = \vec{g}_h$, and $\|u\|_0$ is computed exactly in closed form.

Similarly, to compute the H^1-distance, $|v - v_h|_1$ between a function $v \in \mathrm{V}$ and another one in $v_h \in \mathrm{V}_h$, it is used the following expression:

$$\begin{aligned}|v - v_h|_1^2 &= \int_0^1 (v'(x) - v_h')(\bar{v}'(x) - \bar{v}_h'(x))\,\mathrm{d}x = \int_0^1 \big(|v'(x)|^2 + |v_h'(x)|^2 - 2\mathrm{Re}(\bar{v}'(x)v_h'(x))\big)\mathrm{d}x\\ &= |v|_1^2 + |v_h|_1^2 - 2\mathrm{Re}\left(\int_0^1 v'(x)\bar{v}_h'(x)\,\mathrm{d}x\right).\end{aligned}$$

Additionally, if it is assumed that $v \in \mathrm{V}$ is a solution of the Hemlholtz equation $-v'' = f + k^2 v$ satisfying $v(0) = 0$ then, integrating by parts, it is obtained

$$\int_0^1 v'(x)\bar{v}_h'(x)\,\mathrm{d}x = -\int_0^1 v''\bar{v}_h(x)\,\mathrm{d}x + v'(1)\bar{v}_h(1) = \int_0^1 (f + k^2 v)\bar{v}_h(x)\,\mathrm{d}x + v'(1)\bar{v}_h(1),$$

and hence the H^1-distance can be computed by means of

$$|v - v_h|_1^2 = |v|_1^2 + \vec{v}_h^*\mathcal{K}_h\vec{v}_h - 2\mathrm{Re}(\vec{s}_h^*\mathcal{K}_h\vec{v}_h) = |v|_1^2 + \vec{v}_h^*\mathcal{M}_h\vec{v}_h - 2\mathrm{Re}(\vec{g}_h^*\vec{v}_h),$$

where $\vec{s}_h$ and $\vec{g}_h$ are the coefficient vectors involved in the computation of the H^1-projection $\Pi^h_{\mathrm{H}^1}v$, which satisfy $\mathcal{K}_h\vec{s}_h = \vec{g}_h$, and $|u|_1$ is computed exactly in closed form.

Numerical results

Taking into account these projection operators, the relative L^2-distance of the the exact solution $u(x) = \sin(kx)$ to V_h has been computed by means of the projection, i.e., $\|\Pi^h_{\mathrm{L}^2}u - u\|_0/\|u\|_0$. Similarly, the relative H^1-distance of u to the PUFEM discrete space is computed by $|\Pi^h_{\mathrm{H}^1}u - u|_1/|u|_1$. Plots on Figures 1.7 and 1.8 show the dependence of both distances with respect to parameters h, δ, and k. In the case of the L^2 projection it is observed roughly that

$$\frac{\|\Pi^h_{\mathrm{L}^2}u - u\|_0}{\|u\|_0} \le \begin{cases} Ch^2\delta^2 & \text{if } hk > \pi,\\ Ch^3k\delta^2 & \text{if } hk < \pi.\end{cases}$$

In the same manner, from Figure 1.8 it can be deduced approximately that the overall numerical behaviour of the H^1 projection is given by

$$\frac{|\Pi^h_{\mathrm{H}^1}u - u|_1}{|u|_1} \le \begin{cases} Ch^2\delta^2 & \text{if } hk > 2\pi,\\ Ch\sqrt{k}\delta^2 & \text{if } \pi < hk < 2\pi,\\ Ch^2\delta^2 & \text{if } hk < \pi.\end{cases}$$

Notice that the transition region in $\pi < hk < 2\pi$ could be identified as those mesh size where the role that the enrichment and the mesh size interchange their role: for h large enough, the accuracy of the method is ruled by the exponentials expressions in he discrete basis, however, for a mesh size h small enough, it is the mesh size which determines the accuracy of the projection.

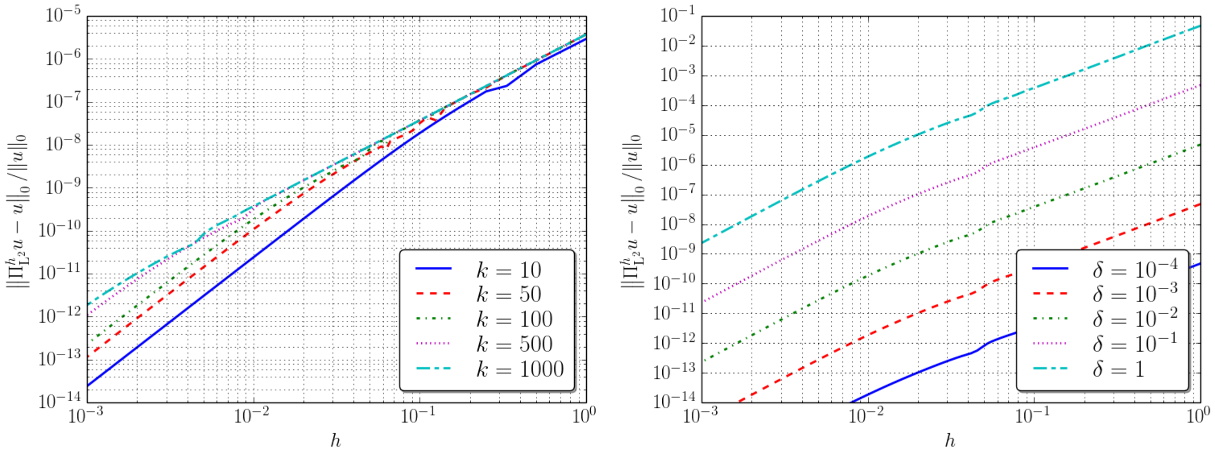

Figure 1.7: L^2-approximation errors of the projection $\Pi^h_{\mathrm{L}^2}$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

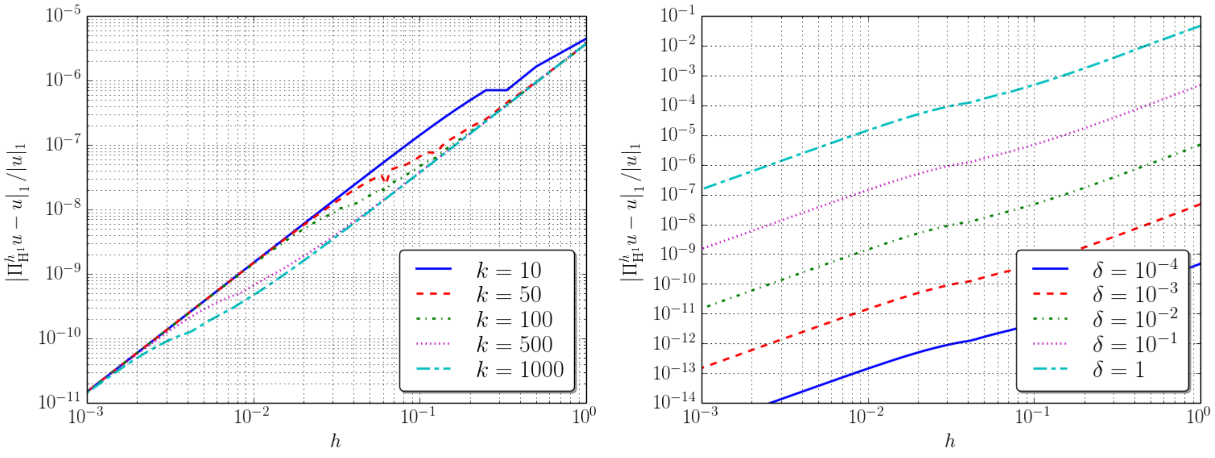

Figure 1.8: H^1-approximation errors of the projection $\Pi^h_{\mathrm{H}^1}$ applied to the exact solution $u(x) = \sin(kx)$, plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

In conclusion, if the numerical errors corresponding to both projections are compared with respect to the approximation errors obtained by using the interpolant $\mathcal{I}_h$ (see Figures 1.5 and 1.6), it can be checked that the projection distances exhibit exactly the same numerical behaviour. Indeed, the interpolant errors in H^1-norm presents a similar transition region for intermediate values of the mesh size.

1.5 Existence and uniqueness of the discrete solution

Before the derivation of *a priori* error estimates of the PUFEM method, it should be ensured the existence and uniqueness of solution of the discrete problem (1.6) (or equivalently, its matrix formulation (1.14)). It has been mentioned previously in Section 1.3.1, that the exponential-type PUFEM discrete basis is not suitable for a typical numerical error analysis. An intuitive idea of the possible drawbacks that the use of this basis could imply are illustrated when hk tends to zero. In fact, in the limit case, if it is assumed formally that $hk = 0$ (either because $k = 0$ or since h is taken formally equal to zero), then the PUFEM basis of $2(n+1)$ functions collapse and becomes linear dependent since ψ_j^+ coincides with ψ_j^- and formally again both functions could be identify with the finite element basis function φ_j.

To avoid partially these drawbacks, the trigonometric PUFEM basis was introduced in Section 1.3.2. Despite this basis does not suffer from the severe linear-dependence issue described above, it cannot be easily used for the study of the existence and uniqueness of the discrete problem (1.6). In the following sections, the space of the twin-bubble functions and the vertex-valued functions will be used separately to decouple the discrete problem in two independent discrete problems.

1.5.1 Global condensation procedure

To mimic the local condensation procedure used in $\mathbb{P}^p$-finite elements (see [28]), a similar orthogonal procedure will be applied to the discrete space V_h. However, due to the non-empty intersection between the supports of the basis functions in the twin-bubble space $\mathrm{V}_h^{\mathfrak{b}}$ it is not possible to compute this orthogonalization locally (in the interior of each element T_j). On the contrary, the condensation procedure will be related to a global problem stated in the whole domain $(0,1)$. To discuss properly this global condensation procedure an unusual functional framework must be introduced.

H^1-bubble space

Let $\mathrm{H}^1_{\mathcal{T}_h}(0,1)$ be the subset of H^1 functions which are null on the mesh vertices $\mathcal{T}_h = \{x_j = hj\}_{j=0}^n$, this is, it is defined as follows:

$$\mathrm{H}^1_{\mathcal{T}_h}(0,1) = \{v \in \mathrm{H}^1(0,1) : \; v(x) = 0 \text{ for all } x \in \mathcal{T}_h\}.$$

Analogously to the $\mathrm{H}^1_{(0}(0,1)$ or $\mathrm{H}^1_0(0,1)$, the space $\mathrm{H}^1_{\mathcal{T}_h}(0,1)$, which will be called H^1-bubble space, is a Hilbert space endowed with inner product associated to the H^1-seminorm $|\cdot|_1$.

Taking into account the definition written above, it is immediately to obtain an coercive result on the form B_k defined in (1.3).

Lemma 1.5.1. *If $hk \leq \alpha < \pi$ and B_k is defined by* (1.3) *then the sesquilinear form given by $(u, v) \mapsto B_k(u, v) - iku(1)\bar{v}(1)$ for all $u, v \in \mathrm{H}^1_{\mathcal{T}_h}(0, 1)$ is continuous, hermitian and coercive.*

Proof. Firstly, since $u(1) = v(1) = 0$ for any $u, v \in \mathrm{H}^1_{\mathcal{T}_h}(0, 1)$ then it is clear that the sesquilinear form defined in the statement of the lemma coincides with B_k and hence it is hermitian ($B(u, v) = \overline{B(v, u)}$ for all $u, v \in \mathrm{H}^1_{\mathcal{T}_h}(0, 1)$).

The continuity of B_k in $\mathrm{H}^1_{\mathcal{T}_h}(0, 1)$ follows directly from the continuity of B_k in $\mathrm{H}^1_{(0}(0, 1)$. However, a sharper continuity constant (smaller than $1 + k^2$) can be obtained as follows. If it is introduced $\hat{v}_j = v|_{T_j} \circ F_j^{-1}$ defined in $(0, 1)$ then for any fixed $u, v \in \mathrm{H}^1_{\mathcal{T}_h}(0, 1)$, it holds

$$\begin{aligned}
|B_k(u, v)| &= \left| \sum_{j=1}^{n} \int_{T_j} \big(u'(x)\bar{v}'(x) - k^2 u(x)\bar{v}(x)\big)\, \mathrm{d}x \right| \\
&\leq \sum_{j=1}^{n} \left| \int_0^1 \left(\frac{1}{h} \hat{u}_j'(\hat{x})\bar{\hat{v}}_j'(\hat{x}) - k^2 h \hat{u}_j(\hat{x})\bar{\hat{v}}(\hat{x}) \right) \mathrm{d}\hat{x} \right| \\
&\leq \sum_{j=1}^{n} \frac{1}{h} \int_0^1 \left| \hat{u}_j'(\hat{x})\bar{\hat{v}}_j'(\hat{x}) - (kh)^2 \hat{u}_j(\hat{x})\bar{\hat{v}}(\hat{x}) \right| \mathrm{d}\hat{x} \leq \frac{1 + (kh)^2}{h} \sum_{j=1}^{n} \|\hat{u}_j\|_{\mathrm{H}^1_0(0,1)} \|\hat{v}_j\|_{\mathrm{H}^1_0(0,1)} \\
&\leq \sqrt{2} \frac{1 + (kh)^2}{h} \sum_{j=1}^{n} |\hat{u}_j|_{\mathrm{H}^1_0(0,1)} |\hat{v}_j|_{\mathrm{H}^1_0(0,1)} = \sqrt{2}(1 + (kh)^2) \sum_{j=1}^{n} |u|_{\mathrm{H}^1_0(T_j)} |v|_{\mathrm{H}^1_0(T_j)} \\
&\leq \sqrt{2}(1 + (kh)^2) \sum_{j=1}^{n} |u|_{\mathrm{H}^1_0(T_j)} |v|_{\mathrm{H}^1_0(T_j)} \\
&\leq \sqrt{2}(1 + (kh)^2) \left(\sum_{j=1}^{n} |u|^2_{\mathrm{H}^1_0(T_j)} \right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} |v|^2_{\mathrm{H}^1_0(T_j)} \right)^{\frac{1}{2}} \\
&= \sqrt{2}(1 + (kh)^2) |u|_{\mathrm{H}^1_0(0,1)} |v|_{\mathrm{H}^1_0(0,1)} \leq \sqrt{2}(1 + \alpha^2) |u|_{\mathrm{H}^1_0(0,1)} |v|_{\mathrm{H}^1_0(0,1)},
\end{aligned}$$

where it has been used the H^1-Cauchy-Schwarz estimate in the third inequality, the Poincare estimate $\|u\|_1 \leq \sqrt{1 + 1/\pi^2}|u|_1 < \sqrt{2}|u|_1$ in the fourth inequality (see [28, Lemma 2.2]), and the n-dimensional Cauchy-Schwarz estimate in the last inequality. Notice also that $u'(x)$ denotes $\mathrm{d}u/\mathrm{d}x$ for $x \in T_j$ whereas $\hat{u}_j'$ denotes $\mathrm{d}\hat{u}_j/\mathrm{d}\hat{x}$ for $\hat{x} \in (0, 1)$ and any $j = 1, \ldots, n$.

The coercivity of form B_k also is deduced using similar arguments. More precisely, it holds

$$\begin{aligned}
B_k(u, u) &= \sum_{j=1}^{n} \int_{T_j} \big(|u'(x)|^2 - k^2 |u(x)|^2\big)\, \mathrm{d}x = \sum_{j=1}^{n} \int_0^1 \frac{1}{h} \big(|\hat{u}_j'(\hat{x})|^2 - (kh)^2 |\hat{u}_j'(\hat{x})|^2\big)\, \mathrm{d}\hat{x} \\
&\geq \frac{\pi^2 - \alpha^2}{\pi^2} \frac{1}{h} \sum_{j=1}^{n} \int_0^1 |\hat{u}_j'(\hat{x})|^2 \mathrm{d}\hat{x} = \frac{\pi^2 - \alpha^2}{\pi^2} \sum_{j=1}^{n} |u|^2_{\mathrm{H}^1_0(T_j)} = \frac{\pi^2 - \alpha^2}{\pi^2} |u|^2_{\mathrm{H}^1_0(0,1)},
\end{aligned}$$

where it has been used that $(kh)^2 < \alpha^2$ is smaller than π^2, which is the smallest eigenvalue of the second-order derivative $-\mathrm{d}^2/\mathrm{d}\hat{x}^2$ in $(0,1)$ (see [28, Lemma 2.2]). □

In addition to the result stated above, the Lax-Milgram lemma also ensures the existence and uniqueness of solution of the variational problem: fixed $hk \leq \alpha < \pi$ and given $f \in \mathrm{L}^2(0,1)$, find $v \in \mathrm{H}^1_{\mathcal{T}_h}(0,1)$ such that

$$B_k(v,\phi) = \langle f,\phi\rangle_{\mathrm{L}^2(0,1)} \text{ for all } \phi \in \mathrm{H}^1_{\mathcal{T}_h}(0,1). \tag{1.45}$$

From the coercivity of B_k and the Poincare inequality $\|v\|_0 \leq |v|_1$, it is straightforward the estimate

$$|v|_1 \leq \frac{\pi^2}{\pi^2-\alpha^2}\|f\|_0. \tag{1.46}$$

It is also clear from the definition (1.21) of the twin-bubble space that $\mathrm{V}_h^{\mathfrak{b}} \subset \mathrm{H}^1_{\mathcal{T}_h}(0,1)$. Since B_k is coercive in $\mathrm{H}^1_{\mathcal{T}_h}(0,1)$, it will be also coercive in $\mathrm{V}_h^{\mathfrak{b}}$ and, in fact, B_k defines an inner product in both spaces, equivalent to the product associated to the seminorm $|\cdot|_1$. Hence, the analogous discrete version of the problem (1.45), this is, fixed $hk \leq \alpha < \pi$ and given $f \in \mathrm{L}^2(0,1)$, find $v_{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}$ such that

$$B_k(v_{\mathfrak{b}},\phi_{\mathfrak{b}}) = \langle f,\phi_{\mathfrak{b}}\rangle_{\mathrm{L}^2(0,1)} \text{ for all } \phi_{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}, \tag{1.47}$$

has an unique solution and it also holds

$$|v_{\mathfrak{b}}|_1 \leq \frac{\pi^2}{\pi^2-\alpha^2}\|f\|_0. \tag{1.48}$$

Despite the previous estimates (in the continuous and discrete variational problems guarantees the well-posedness of both problems), the estimate (1.46) is not sharp and it can be improved as follows taking into account that $\mathrm{H}^1_{\mathcal{T}_h}(0,1) = \bigoplus_{j=1}^n \mathrm{H}^1_0(T_j)$ (understanding that the inclusion of $\mathrm{H}^1_0(T_j)$ in $\mathrm{H}^1_0(0,1)$ is made by the extension by zero of those functions defined in $T_j \subset (0,1)$.

Lemma 1.5.2. *Fixed $hk \leq \alpha < \pi$ and given $f \in \mathrm{L}^2(0,1)$, $v \in \mathrm{H}^1_{\mathcal{T}_h}(0,1)$ is solution of problem* (1.45) *if and only if $v|_{T_j} = v_j$ is solution of the problem*

$$B_k(v_j,\phi) = \langle f|_{T_j},\phi\rangle_{\mathrm{L}^2(T_j)} \textit{ for all } \phi \in \mathrm{H}^1_0(T_j). \tag{1.49}$$

with $j = 1,\dots,n$. In addition, it holds

$$|v|_1 \leq \frac{\pi^2}{\pi^2-\alpha^2}h\|f\|_0. \tag{1.50}$$

Proof. The equivalence between problem (1.45) and (1.49) is immediate. If $v_j = v|_{T_j}$ and taking test functions ϕ with compact support in T_j and then substituted in problem (1.45) then (1.49) is obtained. Reciprocally, if each v_j is extended by zero to the exterior of T_j, and then these extensions $\chi_{T_j}v_j$ are summed up, then $v = \sum_{j=1}^n \chi_{T_j}v_j$ is the solution of problem (1.45). To check this claim, it is enough to use that any $\phi \in \mathrm{H}^1_{\mathcal{T}_h}(0,1)$ can be rewritten as $\phi = \sum_{j=1}^n \chi_{T_j}\phi_j$ with $\phi_j \in \mathrm{H}^1_0(T_j)$ and add the variational formulations (1.49) from $j=1$ to n.

To obtain the sharper estimate, the variational problem (1.49) is rewritten in the reference element $(0,1)$. Hence, it is obtained that $\hat{v}_j = v|_{T_j} \circ F_j^{-1}$ is solution of the variational problem

$$\int_0^1 \left(\hat{v}_j'(\hat{x})\bar{\hat{\phi}}'(\hat{x})\,\mathrm{d}\hat{x} - (kh)^2\hat{v}_j(\hat{x})\bar{\hat{\phi}}(\hat{x})\right)\mathrm{d}\hat{x} = h^2\int_0^1 (f|_{T_j}\circ F_j^{-1})(\hat{x})\bar{\hat{\phi}}(\hat{x})\,\mathrm{d}\hat{x}$$

for all $\phi \in \mathrm{H}^1_0(T_j)$ with $j = 1,\dots,n$. The analogous estimate to (1.46), but now applied to a problem stated in T_j, leads to

$$|\hat{v}_j|_{\mathrm{H}^1_0(0,1)} \le \frac{\pi^2}{\pi^2-\alpha^2}h^2\|f|_{T_j}\circ F_j^{-1}\|_{\mathrm{L}^2(0,1)}$$

and coming back to T_j it is obtained

$$|\hat{v}_j|_{\mathrm{H}^1_0(T_j)} \le \frac{\pi^2}{\pi^2-\alpha^2}h\|f|_{T_j}\|_{\mathrm{L}^2(T_j)}.$$

Estimate (1.50) follows adding the squares of the left and right-hand side in the inequality written above from $j=1$ to n. □

As it has been discussed previously in Remark 1.3.2, since $\mathrm{V}_h^{\mathfrak{b}}$ cannot be rewritten as a direct sum of the space of bubbles functions with support in each finite element T_j. Consequently, the proof of Lemma 1.5.2 cannot be replicated for the discrete problem (1.47). However, despite this drawback, the estimate (1.48) for the discrete solution can also be improved by using that the error $v - v_{\mathfrak{b}}$ is orthogonal to $\mathrm{V}_h^{\mathfrak{b}}$ with respect to the inner product B_k.

Lemma 1.5.3. *Fixed $hk \le \alpha < \pi$ and given $f \in \mathrm{L}^2(0,1)$, if $v_{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}$ is the solution of problem* (1.47) *then it holds*

$$|v_{\mathfrak{b}}|_1 \le \sqrt{2}(1+\alpha^2)\left(\frac{\pi^2}{\pi^2-\alpha^2}\right)^2 h\|f\|_0. \tag{1.51}$$

Proof. From variational problems (1.45) and (1.47), it is clear that $B_k(v - v_{\mathfrak{b}}, \phi_{\mathfrak{b}}) = 0$ for all $\phi_{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}$, or equivalently, $B_k(v_{\mathfrak{b}}, \phi_{\mathfrak{b}}) = B_k(v, \phi_{\mathfrak{b}})$. If $\phi_{\mathfrak{b}} = v_{\mathfrak{b}}$, taking into account the coercivity and the continuity of B_k (see Lemma 1.5.1), and also estimate (1.50), it holds

$$\begin{aligned}\frac{\pi^2-\alpha^2}{\pi^2}|v_{\mathfrak{b}}|_1^2 &\le |B_k(v_{\mathfrak{b}}, v_{\mathfrak{b}})| = |B_k(v, v_{\mathfrak{b}})| \\ &\le \sqrt{2}(1+\alpha^2)|v_{\mathfrak{b}}|_1|v|_1 \le \sqrt{2}(1+\alpha^2)\frac{\pi^2}{\pi^2-\alpha^2}h|v_{\mathfrak{b}}|_1\|f\|_0.\end{aligned}$$

Since it can be supposed that $|v_{\mathfrak{b}}|_1 > 0$ (otherwise $f = 0$ and the lemma follows immediately), the expression above leads to (1.51), simplifying the factor $|v_{\mathfrak{b}}|_1$ at the most right and most left term of the inequalities written above. □

PUFEM partially orthogonal basis

For the subsequent parts of the proof the existence and uniqueness results and the *a priori* error analysis, it will be useful to split the PUFEM discrete space as the direct sum $\mathrm{V}_h = \tilde{\mathrm{V}}_h^{\mathfrak{v}} \oplus \mathrm{V}_h^{\mathfrak{b}}$ where the orthogonality is computed by means of the inner product induced by form B_k. With this purpose, for each $\psi_j^{\mathfrak{b}}$, it will be defined $\tilde{\psi}_j^{\mathfrak{v}} = \psi_j^{\mathfrak{b}} + \xi_j^{\mathfrak{b}}$ such that it is satisfied the orthogonal relation

$$B_k(\tilde{\psi}_j^{\mathfrak{v}}, \phi^{\mathfrak{b}}) = 0 \text{ for all } \phi^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}},$$

or equivalently, find $\xi_j^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}$ such that

$$B_k(\xi_j^{\mathfrak{b}}, \phi^{\mathfrak{b}}) = -B_k(\psi_j^{\mathfrak{v}}, \phi^{\mathfrak{b}}) \text{ for all } \phi^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}. \tag{1.52}$$

Since B_k is a coercive form in $\mathrm{H}^1_{\mathcal{T}_h}(0,1)$ and also in $\mathrm{V}_h^{\mathfrak{b}}$. Hence, the application of the Lax-Milgram lemma guarantees the existence and uniqueness of solution of problem (1.52) and the estimate (1.51) with $f = \psi_j^{\mathfrak{v}} \in \mathrm{L}^2(0,1)$ reads

$$|\xi_j^{\mathfrak{v}}|_1 \leq \sqrt{2}(1+\alpha^2)\left(\frac{\pi^2}{\pi^2-\alpha^2}\right)^2 h\|\psi_j^{\mathfrak{v}}\|_0. \tag{1.53}$$

In conclusion, instead of using the original trigonometric discrete basis $\{\psi_j^{\mathfrak{v}}\}_{j=1}^n \cup \{\psi_j^{\mathfrak{b}}\}_{j=0}^n$ (described in Section 1.3.2), which generates the writing of V_h as the direct sum $\mathrm{V}_h^{\mathfrak{v}} \oplus \mathrm{V}_h^{\mathfrak{b}}$, the discrete PUFEM problem will be represented in terms of the partially orthogonal basis $\{\tilde{\psi}_j^{\mathfrak{v}}\}_{j=1}^n \cup \{\psi_j^{\mathfrak{b}}\}_{j=0}^n$, which induces the representation $\mathrm{V}_h = \tilde{\mathrm{V}}_h^{\mathfrak{v}} \oplus \mathrm{V}_h^{\mathfrak{b}}$.

***Remark* 1.5.4** (Invariant translation)**.** *Notice that since the mesh is uniform (all the elements have the same length h), any discrete basis function in the trigonometric basis $\{\psi_j^{\mathfrak{v}}\}_{j=1}^n \cup \{\psi_j^{\mathfrak{b}}\}_{j=0}^n$ is invariant under translation, i.e., $\psi_j^{\mathfrak{v}}(x) = \psi_m^{\mathfrak{v}}(x - h(j-m))$ and $\psi_j^{\mathfrak{b}}(x) = \psi_m^{\mathfrak{b}}(x - h(j-m))$. Consequently, also the partial orthogonal basis $\{\tilde{\psi}_j^{\mathfrak{v}}\}_{j=1}^n \cup \{\psi_j^{\mathfrak{b}}\}_{j=0}^n$ shares the same property since its functions are linear combination of the trigonometric basis functions. In addition, it is important to realize that $\tilde{\psi}_j^{\mathfrak{v}}$ is symmetric with respect to $x = x_j$, i.e., $\tilde{\psi}_j^{\mathfrak{v}}(x_j + s) = \tilde{\psi}_j^{\mathfrak{v}}(x_j - s)$ for $0 \leq s \leq h$. Such symmetry property does not hold for the twin-bubble functions $\psi_j^{\mathfrak{b}}$.*

Now, taking into account this orthogonal relation between the different functions of the basis and its invariant translation property, problem (1.6) admits the matrix representation

$$\mathcal{L}_h \vec{u}_h = \vec{f}_h, \tag{1.54}$$

where $\vec{u}_h = (\vec{u}_h^{\mathfrak{v}}, \vec{u}_h^{\mathfrak{b}}) = (u_1^{\mathfrak{v}}, \ldots, u_n^{\mathfrak{v}}, u_0^{\mathfrak{b}}, \ldots, u_n^{\mathfrak{b}})^t$ are the coefficients of the discrete solution $u_h \in \mathrm{V}_h$, given by

$$u_h = \sum_{j=1}^n u_j^{\mathfrak{v}} \tilde{\psi}_j^{\mathfrak{v}} + \sum_{j=0}^n u_j^{\mathfrak{b}} \psi_j^{\mathfrak{b}},$$

the matrix $\mathcal{L}_h$ (of size $(2n+1)\times(2n+1)$) is defined by blocks as follows:

$$\mathcal{L}_h = \begin{pmatrix} \mathcal{L}^{\mathfrak{v}}_h & 0_{n\times(n+1)} \\ 0_{(n+1)\times n} & \mathcal{L}^{\mathfrak{b}}_h \end{pmatrix}$$

where the $n \times n$ matrix $\mathcal{L}^{\mathfrak{v}}_h$ and the $(n+1)\times(n+1)$ matrix $\mathcal{L}^{\mathfrak{b}}_h$ are given by

$$\mathcal{L}^{\mathfrak{v}}_h = \begin{pmatrix} 2S_h & R_h & & & \\ R_h & 2S_h & R_h & & \\ & \ddots & \ddots & \ddots & \\ & & R_h & 2S_h & R_h \\ & & & R_h & S_h - ik \end{pmatrix}, \; \mathcal{L}^{\mathfrak{b}}_h = \begin{pmatrix} S^{\mathfrak{b}}_{1h} & R^{\mathfrak{b}}_h & & & \\ R^{\mathfrak{b}}_h & S^{\mathfrak{b}}_{1h} + S^{\mathfrak{b}}_{2h} & R^{\mathfrak{b}}_h & & \\ & \ddots & \ddots & \ddots & \\ & & R^{\mathfrak{b}}_h & S^{\mathfrak{b}}_{1h} + S^{\mathfrak{b}}_{2h} & R^{\mathfrak{b}}_h \\ & & & R^{\mathfrak{b}}_h & S^{\mathfrak{b}}_{2h} \end{pmatrix}, \tag{1.55}$$

being $S_h = B_k(\tilde{\psi}^{\mathfrak{v}}_j, \tilde{\psi}^{\mathfrak{v}}_j)/2$ and $R_h = B_k(\tilde{\psi}^{\mathfrak{v}}_j, \tilde{\psi}^{\mathfrak{v}}_{j-1})$ for any $j = 1, \ldots, n-1$, and analogously $S^{\mathfrak{b}}_{1h} + S^{\mathfrak{b}}_{2h} = B_k(\psi^{\mathfrak{b}}_j, \psi^{\mathfrak{b}}_j)/2$ and $R_h = B_k(\psi^{\mathfrak{b}}_j, \psi^{\mathfrak{b}}_{j-1})$ for any $j = 1, \ldots, n-1$ (where S_{1h} is the contribution from the element T_j and S_{h2} is that one coming from T_{j+1}). The right-hand side $\vec{f} = (\vec{f}^{\mathfrak{v}}_h, \vec{f}^{\mathfrak{v}}_h) = (f^{\mathfrak{v}}_1, \ldots, f^{\mathfrak{v}}_n, f^{\mathfrak{b}}_0, \ldots, f^{\mathfrak{b}}_n)^t$ in (1.54) is given by the projection of f on each element of the discrete basis, i.e.,

$$f^{\mathfrak{v}}_j = \int_0^1 f \tilde{\psi}^{\mathfrak{v}}_j \,\mathrm{d}x, \qquad f^{\mathfrak{b}}_j = \int_0^1 f \psi^{\mathfrak{b}}_j \,\mathrm{d}x \qquad \text{for } j = 0, \ldots, n.$$

Hence, the solution of the linear system (1.54) can be decoupled in the two linear systems $\mathcal{L}^{\mathfrak{v}}_h \vec{u}^{\mathfrak{v}}_h = \vec{f}^{\mathfrak{v}}_h$ and $\mathcal{L}^{\mathfrak{b}}_h \vec{u}^{\mathfrak{b}}_h = \vec{f}^{\mathfrak{b}}_h$. The latter one is the matrix description of the discrete variational problem (1.45) and hence applying Lemma 1.5.3 and more precisely, the estimate (1.51), there exists an unique solution $u^{\mathfrak{b}}_h \in \mathrm{V}^{\mathfrak{b}}_h$ and it holds

$$|u^{\mathfrak{b}}_h|_1 \leq Ch\|f\|_0, \tag{1.56}$$

where C is a positive constant independent of k, δ, and h, once it is satisfied $hk \leq \alpha < \pi$.

The first linear system $\mathcal{L}^{\mathfrak{v}}_h \vec{u}^{\mathfrak{v}}_h = \vec{f}^{\mathfrak{v}}_h$ is equivalent to the following variational problem: given $f \in \mathrm{L}^2(0,1)$, find $v_{\mathfrak{b}} \in \tilde{V}^{\mathfrak{v}}_h$ such that

$$B_k(u^{\mathfrak{v}}_h, \tilde{\phi}_{\mathfrak{v}}) - iku^{\mathfrak{v}}_h(1)\overline{\overline{\tilde{\phi}}}_{\mathfrak{v}}(1) = \langle f, \tilde{\phi}_{\mathfrak{v}} \rangle_{\mathrm{L}^2(0,1)} \text{ for all } \tilde{\phi}_{\mathfrak{b}} \in \tilde{V}^{\mathfrak{v}}_h. \tag{1.57}$$

The following subsections will be devoted to ensure the existence and uniqueness of the discrete problem (1.57). With this purpose, it will be analysed the discrete dispersion relation associated to this discrete problem, the discrete Green's function, and finally, it will be shown the discrete inf-sup condition, which guarantees the well-posedness of problem (1.57).

Finally, it will be useful for the derivation of estimates by means of the Green's function to establish a relation of equivalence between the standard finite element norms for continuous piecewise $\mathcal{P}^1$-finite elements (defined with respect to its point-wise values) and the corresponding L^2 and H^1-norms using the vector of point-wise values of the PUFEM space $\tilde{V}^{\mathfrak{v}}_h$. More precisely, the L^2-finite element norm $\|\cdot\|_{0,\mathrm{fe}}$ and H^1-finite element seminorm $|\cdot|_{1,\mathrm{fe}}$

for a vector of point-wise values $\vec{v} = (v_1, \ldots, v_n)^t$ associated to a finite element function $v_{\text{fe}} = \sum_{j=1}^{n} v_j \varphi_h$ is defined as follows:

$$\|v_{\text{fe}}\|_{\mathrm{L}^2(0,1)} = \|\vec{v}\|_{0,\mathrm{V}_h^{\text{fe}}} = \left(h \sum_{j=1}^{n} |v_j|^2 \right)^{\frac{1}{2}}, \qquad |v_{\text{fe}}|_{\mathrm{H}^1(0,1)} = |\vec{v}|_{1,\mathrm{V}_h^{\text{fe}}} = \left(h \sum_{j=1}^{n} \left| \frac{v_j - v_{j-1}}{h} \right|^2 \right)^{\frac{1}{2}}, \tag{1.58}$$

where it is assumed that $v_0 = 0$ (due to the homogeneous Dirichlet condition at $x = 0$). Analogously, the PUFEM norms associated to a function $v_h = \sum_{j=1}^{n} v_j \tilde{\psi}_j^{\mathfrak{v}} \in \tilde{V}_h^{\mathfrak{v}}$ associated to its point-wise value vector $\vec{v} = (v_1, \ldots, v_n)$ is defined by

$$\|v_h\|_{\mathrm{L}^2(0,1)} = \|\vec{v}\|_{0,\tilde{V}_h^{\mathfrak{v}}} = \left(\vec{v}^* \tilde{\mathcal{M}}_h \vec{v} \right)^{\frac{1}{2}}, \qquad |v_h|_{\mathrm{H}^1(0,1)} = |\vec{v}|_{1,\tilde{V}_h^{\mathfrak{v}}} = \left(\vec{v}^* \tilde{\mathcal{K}}_h \vec{v} \right)^{\frac{1}{2}}, \tag{1.59}$$

where $[\tilde{\mathcal{M}}_h]_{jl} = \int_0^1 \tilde{\psi}_j^{\mathfrak{v}} \overline{\tilde{\psi}_l^{\mathfrak{v}}} \, \mathrm{d}x$ and $[\tilde{\mathcal{K}}_h]_{jl} = \int_0^1 (\tilde{\psi}_j^{\mathfrak{v}})' (\overline{\tilde{\psi}_l^{\mathfrak{v}}})' \, \mathrm{d}x$, and again it has been assumed that $v_0 = 0$. Despite any pair of norms are equivalent in an finite-dimensional space, the following lemma states the equivalence constants independently of h, k, and δ.

Lemma 1.5.5. *Assume $h(k + \delta) \leq \alpha < \pi$, if $v_h = \sum_{j=1}^{n} v_j \tilde{\psi}_j^{\mathfrak{v}} \in \tilde{V}_h^{\mathfrak{v}}$ and $\vec{v} = (v_1, \ldots, v_n)$ is its point-wise value vector then it holds*

$$C_1 \|\vec{v}\|_{0,\mathrm{V}_h^{\text{fe}}} \leq \|v_h\|_{\mathrm{L}^2(0,1)} \leq C_2 \|\vec{v}\|_{0,\mathrm{V}_h^{\text{fe}}}, \qquad C_1 |\vec{v}|_{1,\mathrm{V}_h^{\text{fe}}} \leq |v_h|_{\mathrm{H}^1(0,1)} \leq C_2 |\vec{v}|_{1,\mathrm{V}_h^{\text{fe}}}, \tag{1.60}$$

where C_1 and C_2 are positive constant functions independent of h, k, and δ (depending only on α).

Proof. It will be followed a slight modification of the steps used in [21, Lemma 9.7] to proof the equivalence of norms in polynomial finite element spaces between the discrete functions and its point-wise value vectors.

Clearly, if $v_h \in \tilde{V}_h^{\mathfrak{v}}$ then $\hat{v}_j = v_h|_{T_j} \circ F_j^{-1}$ for any fixed $j = 1, \ldots, n$ belongs to the span of local functions $\langle \tilde{\psi}_{j-1}^{\mathfrak{v}}|_{T_j} \circ F_j^{-1}, \tilde{\psi}_j^{\mathfrak{v}}|_{T_j} \circ F_j^{-1} \rangle$ defined by the partially orthogonal procedure described in Section 1.5.1. Hence, $\hat{v}_j$ defined in $\hat{T}$ is represented by the $\mathbb{C}^2$-coordinate basis vector $\vec{v}_j = (v_{j-1}, v_j)^t$ A direct inspection reveals that the two local functions $\{\tilde{\psi}_{j-1}^{\mathfrak{v}}|_{T_j} \circ F_j^{-1}, \tilde{\psi}_j^{\mathfrak{v}}|_{T_j} \circ F_j^{-1}\}$ depend continuously on the parameter $h(k + \delta) \in (0, \alpha]$. Moreover, for $h(k + \delta) = 0$ these two local functions coincides with the local condensation basis of the piecewise $\mathbb{P}^2$-finite element (see Remark 1.3.3 for further details). Consequently, if $\hat{\mathcal{K}}_{\text{loc}}$ and $\hat{\mathcal{M}}_{\text{loc}}$ denote the local stiffness and mass matrices defined in $\hat{T}$ with respect to this local PUFEM basis, then the coefficients of these matrices also depend continuously on the parameter $h(k + \delta)$. In addition, if $\hat{\mathcal{K}}_{\text{loc}}^{\text{fe}}$ and $\hat{\mathcal{M}}_{\text{loc}}^{\text{fe}}$ denote the analogous local stiffness and diagonal lumped mass matrices ($\hat{\mathcal{M}}_{\text{loc}}^{\text{fe}}$ is equal to the 2×2 identity matrix) with respect to this local standard $\mathbb{P}^1$-FEM basis, then it holds

$$\lambda_{\min}(h(k+\delta)) \leq \frac{\vec{v}_j^* \hat{\mathcal{M}}_{\text{loc}} \vec{v}_j}{\vec{v}_j^* \hat{\mathcal{M}}_{\text{loc}}^{\text{fe}} \vec{v}_j} \leq \lambda_{\max}(h(k+\delta)), \quad \mu_{\min}(h(k+\delta)) \leq \frac{\vec{v}_j^* \hat{\mathcal{K}}_{\text{loc}} \vec{v}_j}{\vec{v}_j^* \hat{\mathcal{M}}_{\text{loc}}^{\text{fe}} \vec{v}_j} \leq \mu_{\max}(h(k+\delta)),$$

where $\lambda_{\min}(h(k+\delta))$ and $\lambda_{\max}(h(k+\delta))$ are respectively the minimum and maximum eigenvalues of the symmetric generalized eigenvalue problem $\hat{\mathcal{K}}_{\text{loc}}\vec{v} = \lambda\hat{\mathcal{K}}_{\text{loc}}^{\text{fe}}\vec{v}$, and $\mu_{\min}(h(k+\delta))$ and $\mu_{\max}(h(k+\delta))$ are respectively the minimum and maximum eigenvalues of the symmetric generalized eigenvalue problem $\hat{\mathcal{M}}_{\text{loc}}\vec{v} = \lambda\hat{\mathcal{M}}_{\text{loc}}^{\text{fe}}\vec{v}$. In both cases, their eigensolutions also depend continuously on the parameter $h(k+\delta)$. Hence, the maps $h(k+\delta) \mapsto \lambda_{\min}(h(k+\delta))$ and $hk \mapsto \lambda_{\max}(h(k+\delta))$ are continuous functions defined in a non-empty compact domain $[0,\alpha]$. So, using the Weierstrass theorem, both continuous functions reaches respectively a minimum $\lambda_{\min}$ and a maximum value $\lambda_{\max}$ (possibly depending on α). The same argument should be applied to bound the eigenvalues $\mu_{\min}(h(k+\delta))$ and $\mu_{\max}(h(k+\delta))$.

Now, taking into account that $\vec{v}_j^*\hat{\mathcal{M}}_{\text{loc}}^{\text{fe}}\vec{v}_j = |v_{j-1}|^2 + |v_j|^2$ and $\vec{v}_j^*\hat{\mathcal{K}}_{\text{loc}}^{\text{fe}}\vec{v}_j = |v_j - v_{j-1}|^2$ and the fact that $\|\hat{v}_j\|^2_{\mathrm{L}^2(\hat{T})} = \vec{v}_j^*\hat{\mathcal{M}}_{\text{loc}}\vec{v}_j$, and $|\hat{v}_j|^2_{\mathrm{H}^1(\hat{T})} = \vec{v}_j^*\hat{\mathcal{K}}_{\text{loc}}\vec{v}_j$, it holds

$$\begin{gathered}\lambda_{\min}(|v_{j-1}|^2 + |v_j|^2) \leq \|\hat{v}_j\|^2_{\mathrm{L}^2(\hat{T})} \leq \lambda_{\max}(|v_{j-1}|^2 + |v_j|^2),\\ \mu_{\min}|v_j - v_{j-1}|^2 \leq |\hat{v}_j|^2_{\mathrm{H}^1(\hat{T})} \leq \mu_{\max}|v_j - v_{j-1}|^2.\end{gathered}$$

and coming back to element T_j by applying the affine transform F_j, using that $|v_h|_{T_j}|^2_{\mathrm{H}^1(T_j)} = |\hat{v}_j|^2_{\mathrm{H}^1(\hat{T})}/h$ and $\|v_h|_{T_j}\|^2_{\mathrm{L}^2(T_j)} = h\|\hat{v}_j\|^2_{\mathrm{L}^2(\hat{T})}$, the estimate written above leads to

$$\begin{gathered}\lambda_{\min}h(|v_{j-1}|^2 + |v_j|^2) \leq \|v_h|_{T_j}\|^2_{\mathrm{L}^2(T_j)} \leq \lambda_{\max}h(|v_{j-1}|^2 + |v_j|^2),\\ \mu_{\min}\frac{|v_j - v_{j-1}|^2}{h} \leq |v_h|_{T_j}|^2_{\mathrm{H}^1(T_j)} \leq \mu_{\max}\frac{|v_j - v_{j-1}|^2}{h}.\end{gathered}$$

If the terms in the previous inequality are added from $j = 1$ to n and the root square is computed, estimates (1.60) are obtained. □

1.5.2 Discrete dispersion relation

The derivation of discrete dispersion relations for the whole linear system (1.14) can be made identifying those Bloch discrete waves in V_h, which are homogeneous solutions of the discrete Helmholtz problem on the uniform mesh. This Bloch analysis is described in detail in Section 1.B. However, that analysis is not helpful to obtain the discrete wave numbers which should be involved in the definition of the discrete Green's function in $\tilde{V}_h^{\mathfrak{p}}$. In what follows, estimates of the difference between the continuous and the discrete wave number will be derived using analogous arguments to those one described in [28].

To write the discrete Green's function associated to the discrete PUFEM problem (1.57), the first step consists in the estimation of the discrete wave number, this is, to compare the wave number associated with the exact solution of the homogeneous Helmholtz equation with those solutions who satisfy the row equations of the tridiagonal matrix $\mathcal{L}_h^{\mathfrak{p}}$, stated in an uniform mesh extended throughout the whole real line.

With this comparative aim, first a exact tridiagonal stencil $\mathcal{L}_h^{\text{ex}}$ will be computed is such a manner that the Bloch planewaves with exact wave number k satisfy this *exact* stencil. So, instead of using the discrete basis $\{\tilde{\psi}_j^{\mathfrak{p}}\}_{j=1}^n$ in $\tilde{V}_h^{\mathfrak{p}} \subset \mathrm{H}^1_{(0}(0,1)$, the set of linearly independent functions $\{u_j\}_{j=1}^n$ in $\mathrm{H}^1_{(0}(0,1)$ is considered, which are defined as the unique solution of the continuous Helmholtz problem:

$$\begin{cases} -u_j'' - k^2 u_j = 0 & \text{in } T_{j-1} \cup T_j = [x_{j-1}, x_{j+1}], \\ u(x_{j-1}) = 0, \qquad u(x_j) = 1, & u(x_{j+1}) = 0. \end{cases} \tag{1.61}$$

Inserting this set of functions in the variational problem (1.2) (without taking into account the boundary conditions), the tridiagonal stencil, which is obtained for the interior nodes (for $j = 1, \dots, n-1$), satisfies

$$R_{\mathrm{ex}} u_{\mathrm{ex}}(x_{j+1}) + 2S_{\mathrm{ex}} u_{\mathrm{ex}}(x_j) + R_{\mathrm{ex}} u_{\mathrm{ex}}(x_{j-1}), \tag{1.62}$$

where u_{ex} is an exact solution of the homogeneous Helmholtz equation and S_{ex} and R_{ex} are given by

$$2S_{\mathrm{ex}} = B_k(u_j, u_j), \qquad R_{\mathrm{ex}} = B_k(u_j, u_{j+1}) = B_k(u_j, u_{j-1}). \tag{1.63}$$

Since $u_{\mathrm{ex}}(x) = Ae^{ikx} + Be^{-ikx}$, the fundamental Bloch solutions of (1.62) are

$$u_h^+(x) = \sum_{j \in \mathbb{Z}} u_j(x) e^{ikx_j}, \qquad u_h^-(x) = \sum_{j \in \mathbb{Z}} u_j(x) e^{-ikx_j}$$

and consequently

$$\cos(kh) = -\frac{S_{\mathrm{ex}}}{R_{\mathrm{ex}}}. \tag{1.64}$$

The next step in the derivation of the discrete wave number for the PUFEM discretization in $\tilde{V}_h^{\mathfrak{p}}$ consists in the statement of an equivalent variational formulation associated to problem (1.61). Since u_j is defined piecewise in each element T_j and T_{j+1} it can be rewritten as the addition of a basis function in $\mathrm{V}_h^{\mathfrak{p}}$ plus a function of the H^1-bubble space. Hence, given $\psi_j^{\mathfrak{p}} \in \mathrm{V}_h^{\mathfrak{p}}$, the exact solution $u_j = \psi_j^{\mathfrak{p}} + \xi_j \in \mathrm{V}_h^{\mathfrak{p}} \oplus \mathrm{H}^1_{\mathcal{T}_h}(0,1)$ is determined by means of the solution of the variational problem

$$B_k(\xi_j, \phi) = -B_k(\psi_j^{\mathfrak{p}}, \phi) \text{ for all } \phi \in \mathrm{H}^1_{\mathcal{T}_h}(0,1). \tag{1.65}$$

Using Lemma 1.5.1, if $hk \leq \alpha < \pi$ then the problem stated above has an unique solution since B_k is continuous and coercive in $\mathrm{H}^1_{\mathcal{T}_h}(0,1)$. It should be remarked that the variational problem (1.65) is the continuous version of the discrete variational problem (1.52), where the partially orthogonal basis $\{\tilde{\psi}_j^{\mathfrak{p}}\}_{j=0}^n$ was defined by means of the computation of $\{\xi_j^{\mathfrak{b}}\}_{j=0}^n \subset \mathrm{V}_h^{\mathfrak{p}} \subset \mathrm{H}^1_{\mathcal{T}_h}(0,1)$. It should be also notice that the form B_k has real-valued coefficients and hence, since the right-hand side of problems (1.65) and (1.52) is defined by real-valued functions (as in the case of functions $\{\psi_j^{\mathfrak{p}}\}_{j=0}^n$), then the solution of these variational problems are also real-valued.

Lemma 1.5.6. *If $hk \leq \alpha < \pi$ and $\tilde{u}_j$ and $\tilde{\psi}_j^{\mathfrak{p}}$ are defined by the variational problems* (1.65) *and* (1.52)*, then*

$$B_k(u_j - \tilde{\psi}_j^{\mathfrak{p}}, \tilde{u}_l - \tilde{\psi}_l^{\mathfrak{p}}) = B_k(\tilde{\psi}_j^{\mathfrak{p}}, \tilde{\psi}_l^{\mathfrak{p}}) - B_k(u_j, \tilde{u}_l), \tag{1.66}$$

for all $j = 0, \dots, n$.

Proof. The arguments used here are completely analogous to those ones used in [28, Lemma 3.1]. Firstly, it is clear from (1.65) that $B_k(\psi_j^{\mathfrak{p}}, \xi_l) = -B_k(\xi_j, \xi_l)$ and analogously $B_k(\psi_j^{\mathfrak{p}}, \xi_l^{\mathfrak{b}}) = -B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}})$, and since $\mathrm{V}_h^{\mathfrak{b}} \subset \mathrm{H}^1_{\mathcal{T}_h}(0,1)$, it holds $B_k(\psi_j^{\mathfrak{p}}, \xi_l^{\mathfrak{b}}) = -B_k(\xi_j, \xi_l^{\mathfrak{b}})$ for any arbitrary value of j and l. In addition, the error between the variational solutions is orthogonal to $\mathrm{V}_h^{\mathfrak{b}}$ with respect to B_k and so $B_k(\xi_j, \xi_l^{\mathfrak{b}}) = B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}})$. Finally, since $\{\xi_j\}_{j=0}^n$ and $\{\xi_l^{\mathfrak{b}}\}_{j=0}^n$ are real-valued functions then the form B_k applied to any pair of these two sets behaves like a real-valued symmetric bilinear form. Straightforward computations show that the left hand side in (1.66) can be rewritten as follows:

$$\begin{aligned}
B_k(u_j - \tilde{\psi}_j^{\mathfrak{p}}, \tilde{u}_l - \tilde{\psi}_l^{\mathfrak{p}}) &= B_k(\psi_j^{\mathfrak{p}} + \xi_j - \psi_j^{\mathfrak{p}} - \xi_j^{\mathfrak{b}}, \psi_l^{\mathfrak{p}} + \xi_l - \psi_l^{\mathfrak{p}} - \xi_l^{\mathfrak{b}}) \\
&= B_k(\xi_j, \xi_l) - B_k(\xi_j, \xi_l^{\mathfrak{b}}) - B_k(\xi_j^{\mathfrak{b}}, \xi_l) + B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}}) \\
&= B_k(\xi_j, \xi_l) - 2B_k(\xi_j^{\mathfrak{b}}, \xi_l) + B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}}) = B_k(\xi_j, \xi_l) - B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}}),
\end{aligned}$$

where it has been used that $B_k(\xi_j, \xi_l^{\mathfrak{b}}) = B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}}) = B_k(\xi_j^{\mathfrak{b}}, \xi_l)$. A direct computation of the right-hand side in 1.66 shows that

$$\begin{aligned}
B_k(\tilde{\psi}_j^{\mathfrak{p}}, \tilde{\psi}_l^{\mathfrak{p}}) - B_k(u_j, \tilde{u}_l) =& B_k(\psi_j^{\mathfrak{p}} + \xi_j^{\mathfrak{b}}, \psi_l^{\mathfrak{p}} + \xi_l^{\mathfrak{b}}) - B_k(\psi_j^{\mathfrak{p}} + \xi_j, \psi_l^{\mathfrak{p}} + \xi_l) \\
=& B_k(\psi_j^{\mathfrak{p}}, \psi_l^{\mathfrak{p}}) + 2B_k(\psi_j^{\mathfrak{p}}, \xi_l^{\mathfrak{b}}) + B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}}) \\
&- B_k(\psi_j^{\mathfrak{p}}, \psi_l^{\mathfrak{p}}) - 2B_k(\psi_j^{\mathfrak{p}}, \xi_l) - B_k(\xi_j, \xi_l) \\
=& B_k(\xi_j, \xi_l) - B_k(\xi_j^{\mathfrak{b}}, \xi_l^{\mathfrak{b}}),
\end{aligned}$$

where it has been used that all functions are real-valued and hence B_k behaves like a symmetric form and the relations derived from variational forms stated above. □

Now, the attention must be focused on the discrete problem associated to the variational problem (1.57) stated in $\tilde{V}_h^{\mathfrak{p}}$. In that case, since the discrete functions $\tilde{V}_h^{\mathfrak{p}}$ are determined by its point-wise values at the vertices of the mesh, the tridiagonal stencil which is formally satisfied for a Bloch wave $u_h(x) = \sum_{j\in\mathbb{Z}} \tilde{\psi}_j^{\mathfrak{p}}(x) e^{ik'x_j}$ leads to

$$R_h u_h(x_{j+1}) + 2S_h u_h(x_j) + R_h u_h(x_{j-1}) = 0, \tag{1.67}$$

where recall that S_h and R_h are given by

$$2S_h = B_k(\tilde{\psi}_j^{\mathfrak{p}}, \tilde{\psi}_j^{\mathfrak{p}}), \qquad R_h = B_k(\tilde{\psi}_j^{\mathfrak{p}}, \tilde{\psi}_{j-1}^{\mathfrak{p}}).$$

Hence, from (1.67) it is obtained the discrete dispersion relation

$$\cos(k'h) = -\frac{S_h}{R_h}, \tag{1.68}$$

being k' the so-called discrete wave number associated to the PUFEM discretization in $\tilde{V}_h^{\mathfrak{p}}$.

Theorem 1.5.7. *Assume that there exist strictly positive constants α and β such that $hk \leq \alpha < 1$ and $\delta^4 h^4 < (1-\beta)/(\sqrt{2}\hat{C})$ being $\hat{C}$ the approximation constant involved in* (1.40)*. If k' is the discrete wave number defined in* (1.68) *then it holds*

$$|\cos(k'h) - \cos(kh)| \leq C\delta^4 h^4, \tag{1.69}$$

$$|k' - k| \leq C\frac{\delta^4 h^2}{k}, \tag{1.70}$$

where C is a positive constant independent of h, k and δ (and only dependent on α and β).

Proof. The arguments used in this proof are almost identical to those ones used in [28, Theorem 3.2]. Firstly, straightforward computations show that u_j are defined by

$$u_j(x) = \begin{cases} -\cot(kh)\sin(k(x-x_j)) + \cos(k(x-x_j)) & \text{for } x \in T_{j+1} \\ \cot(kh)\sin(k(x-x_j)) + \cos(k(x-x_j)) & \text{for } x \in T_j, \\ 0 & \text{otherwise.} \end{cases}$$

Direct computations show that

$$\|u_j\|_0^2 = h\left(\frac{2}{3} + \mathcal{O}(k^2h^2)\right), \qquad |u_j|_1^2 = \frac{1}{h}\left(2 + \mathcal{O}(k^2h^2)\right), \tag{1.71}$$

and also

$$2S_{\text{ex}} = B_k(u_j, u_j) = \frac{1}{h}\left(2 + \mathcal{O}(k^2h^2)\right), \qquad R_{\text{ex}} = B_k(u_j, u_{j+1}) = \frac{1}{h}\left(-1 + \mathcal{O}(k^2h^2)\right), \tag{1.72}$$

where $\mathcal{O}(k^2h^2)$ must be read as a tailored expression bounded by $C_1k^2h^2 + C_2k^4h^4 + \ldots$, where C_1, $C_2, \ldots$ are positive constants independent of k and h.

Using the discrete dispersion relations (1.64) and (1.67)

$$|\cos(kh) - \cos(k'h)| = \left|\frac{S_h}{R_h} - \frac{S_{\text{ex}}}{R_{\text{ex}}}\right| = \left|\frac{S_hR_{\text{ex}} - S_{\text{ex}}R_h}{R_hR_{\text{ex}}}\right|. \tag{1.73}$$

In consequence, to estimate the difference between both cosines in the expression above it is enough to obtain an upper bound on the numerator and a positive lower bound for the denominator.

First, it should be taken into account that $u_j|_{T_j}$ is the exact solution of an analogous variational to problem (1.65) but test functions in $\mathrm{H}_0^1(T_j)$. In the same manner, the PUFEM approximation $\tilde{\psi}_j^{\mathfrak{v}}|_{T_j}$ is the exact solution of an analogous variational to problem (1.52) but test functions used are in the discrete space $\mathrm{W}_j^{\mathfrak{b}} = \{\phi^{\mathfrak{b}}|_{T_j} : \phi^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}\}$. Hence, estimates in Lemma 1.4.4 can be applied for $h(k+\delta) < 2\pi$, taking into account the discretization space $\{\psi_j^{\mathfrak{v}}|_{T_j}\} \cup \mathrm{W}_j^{\mathfrak{b}}$. Due to the interpolatory properties of $\mathcal{I}_h$ (see Remark 1.4.5), it holds that

$\mathcal{I}_h(u_j|_{T_j}) = \psi_j^{\mathfrak{v}}|_{T_j} + \phi_{\mathcal{I}}^{\mathfrak{b}}|_{T_j}$ with some $\phi_{\mathcal{I}}^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}$. In this manner, utilizing again Cea's lemma (see [11]) applied to the variational problems (1.65) and (1.52) rewritten in $\mathrm{H}^1(\hat{T})$, it holds

$$\begin{aligned}
\left| u_j|_{T_j} - \tilde{\psi}_j^{\mathfrak{v}}|_{T_j} \right|_{\mathrm{H}^1(T_j)} &\leq \frac{\sqrt{2}\pi^2(1+\alpha^2)}{\pi^2-\alpha^2} \inf_{\phi \in \mathrm{V}_h^{\mathfrak{b}}} \left| u_j|_{T_j} - \left(\psi_j^{\mathfrak{v}}|_{T_j} + \phi^{\mathfrak{b}}|_{T_j}\right) \right|_{\mathrm{H}^1(T_j)} \\
&\leq \frac{\sqrt{2}\pi^2(1+\alpha^2)}{\pi^2-\alpha^2} \left| u_j|_{T_j} - \left(\psi_j^{\mathfrak{v}}|_{T_j} + \phi_{\mathcal{I}}^{\mathfrak{b}}|_{T_j}\right) \right|_{\mathrm{H}^1(T_j)} \\
&= \frac{\sqrt{2}\pi^2(1+\alpha^2)}{\pi^2-\alpha^2} \left| u_j|_{T_j} - \mathcal{I}_h\left(u_j|_{T_j}\right) \right|_{\mathrm{H}^1(T_j)} \\
&\leq C\delta^2 h^2 \left| u_j|_{T_j} \right|_{\mathrm{H}^1(T_j)}.
\end{aligned}$$

Adding the analogous estimation in T_{j+1}, it is obtained

$$|u_j - \tilde{\psi}_j^{\mathfrak{v}}|_1 \leq C\delta^2 h^2 |u_j|_1 \text{ for all } j = 0, \dots, n, \tag{1.74}$$

being C a positive constant independent of h, k and δ.

Second, the numerator in (1.73) can be estimated using Lemma 1.5.6, the continuity of B_k in $\mathrm{H}^1_{\mathcal{T}_h}$, the computations (1.72), and the estimate (1.74) as follows:

$$\begin{aligned}
|S_h R_{\mathrm{ex}} - S_{\mathrm{ex}} R_h| = |B_k(u_j,u_j)B_k(\tilde{\psi}_j^{\mathfrak{v}}, \tilde{\psi}_{j+1}^{\mathfrak{v}}) - B_k(u_j,u_{j+1})B_k(\tilde{\psi}_j^{\mathfrak{v}}, \tilde{\psi}_j^{\mathfrak{v}})|& \\
=& |B_k(u_j,u_j)(B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_{j+1} - \tilde{\psi}_{j+1}^{\mathfrak{v}}) - B_k(u_j,u_{j+1})) \\
&- B_k(u_j,u_{j+1})(B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_j - \tilde{\psi}_j^{\mathfrak{v}}) - B_k(u_j,u_j))| \\
=& |B_k(u_j,u_j)B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_{j+1} - \tilde{\psi}_{j+1}^{\mathfrak{v}}) - B_k(u_j,u_{j+1})B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_j - \tilde{\psi}_j^{\mathfrak{v}})| \\
\leq& |B_k(u_j,u_j)||B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_{j+1} - \tilde{\psi}_{j+1}^{\mathfrak{v}})| + |B_k(u_j,u_{j+1})B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_j - \tilde{\psi}_j^{\mathfrak{v}})| \\
\leq& \frac{1}{h}(2 + \mathcal{O}(h^2k^2))\sqrt{2}(1 + (h^2k^2))\hat{C}^2\delta^4h^4|u_j|_1|u_{j+1}|_1 \\
&+ \frac{1}{h}(1 + \mathcal{O}(h^2k^2))\sqrt{2}(1 + (h^2k^2))\hat{C}^2\delta^4h^4|u_j|_1^2,
\end{aligned}$$

where $\hat{C}$ is the positive constant involved in (1.40) Now, using (1.71) and taking into account that $hk < h(k+\delta) \leq \alpha < 1$, it holds

$$|S_h R_{\mathrm{ex}} - S_{\mathrm{ex}} R_h| \leq C\delta^4 h^2, \tag{1.75}$$

where C is a positive constant independent of h, k and δ (only dependent on α).

The denominator in (1.73) can be rewritten as

$$\begin{aligned}
|R_h R_{\mathrm{ex}}| =& |B_k(u_j,u_{j+1})||B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_{j+1} - \tilde{\psi}_{j+1}^{\mathfrak{v}}) + B_k(u_j,u_{j+1})| \\
\geq& \left|\frac{1}{h}(-1 + \mathcal{O}(h^2k^2))\right| \left| |B_k(u_j - \tilde{\psi}_j^{\mathfrak{v}}, u_{j+1} - \tilde{\psi}_{j+1}^{\mathfrak{v}})| - \left|\frac{1}{h}\left(-1 + \mathcal{O}(k^2h^2)\right)\right| \right| \\
\geq& \frac{1}{h}(1 + \mathcal{O}(h^2k^2))\left(\frac{1}{h}\left(1 + \mathcal{O}(k^2h^2)\right) - \sqrt{2}(1 + (h^2k^2))\hat{C}^2\delta^4h^4\frac{1}{h}\left(2 + \mathcal{O}(h^2k^2)\right)\right).
\end{aligned}$$

Once it is assumed that $1-\sqrt{2}\hat{C}^2\delta^4h^4 \geq \beta > 0$ the expression between large parenthesis in the last term of the inequality written above is strictly positive and lower bounded by β and hence

$$|R_h R_{\mathrm{ex}}| \geq \frac{C}{h^2}, \tag{1.76}$$

being C a positive constant independent of h, k, and δ (only dependent on α and β). Finally, inserting estimates (1.75) and (1.76) in (1.73), it holds (1.69). Analogous arguments to those ones used in [28, Theorem 3.2], estimate (1.69) leads to (1.70) straightforwardly. □

1.5.3 Discrete Green's function and inf-sup condition

Once the discrete dispersion relation have been studied, now it is possible to deduce the discrete Green's function associated to the discrete sub-problem with matrix $\mathcal{L}_h^{\mathfrak{v}}$ defined in (1.55). To write this discrete Green's function, it be followed analogous arguments to those ones used in [46] in continuous case (see also Appendix 1.C) and [27, Section 3.2]. Analogously to the continuous case the expression of the discrete Green's function $G_h(x_j, x_l)$ will be written in terms of two discrete functions α_h and β_h as follows:

$$G_h(x_j, x_l) = \begin{cases} \dfrac{\alpha_h(x_j)\beta_h(x_l)}{\Delta_h} & \text{if } x_j \leq x_m, \\ \dfrac{\alpha_h(x_l)\beta_h(x_j)}{\Delta_h} & \text{if } x_j \geq x_m, \end{cases} \tag{1.77}$$

where Δ_h is a quantity, which is fixed to satisfy

$$R_h G_h(x_{m-1}, x_m) + 2S_h G(x_m, x_m) + R_h G(x_{m+1}, x_m) = \frac{1}{h}. \tag{1.78}$$

In the case of α_h, any homogeneous solution of the discrete variational problem with matrix $\mathcal{L}_h^{\mathfrak{v}}$ is given has linear combination of the Bloch-type waves

$$\alpha_h(x) = A\sum_{j=0}^{n} \tilde{\psi}_j^{\mathfrak{v}}(x)\cos(jk'h) + B\sum_{j=0}^{n} \tilde{\psi}_j^{\mathfrak{v}}(x)\sin(jk'h). \tag{1.79}$$

Since the vector given by the values of α_h at the mesh vertices should satisfies the first row of the matrix $\mathcal{L}_h^{\mathfrak{v}}$, this is, $2S\alpha_h(x_1) + R\alpha_h(x_2) = 0$. It is straightforward to check that it is equivalent to satisfy $\alpha_h(x_0) = 0$ and hence $A = 0$ being B any non-null constant.

The computation of β_h is analogous. Since it can be defined by

$$\beta_h(x) = C\left(\sum_{j=0}^{n} \tilde{\psi}_j^{\mathfrak{v}}(x)\cos(jk'h) + D\sum_{j=0}^{n} \tilde{\psi}_j^{\mathfrak{v}}(x)\sin(jk'h)\right), \tag{1.80}$$

with C a non-null constant, it only should be checked that the last row of the linear system involving $\mathcal{L}_h^{\mathfrak{v}}$ is satisfied. In this case, it must be verified that $R_h\beta_h(x_{n-1}) + (S_h - ik)\beta_h(x_n) = 0$, or equivalently,

$$R_h(\cos(k'h(n-1)) + D\sin(k'h(n-1))) + (S_h - ik)(\cos(k'hn) + D\sin(k'nh)) = 0.$$

A direct computation from the equation written above shows that

$$D = \frac{\sin(k')\cos(k')(R_h^2\sin^2(k'h) - k^2) - ikR_h\sin(k'h)}{R_h^2\sin(k'h)\cos^2(k'h) + k^2\sin^2(k')}.$$

Additionally, straightforward computations also show that if it is taken $B = 1$ in (1.79) and $C = 1$ in (1.80) then

$$R_h\alpha_h(x_{m-1})\beta_h(x_m) + 2S_h\alpha_h(x_m)\beta_h(x_m) + R_h\beta_h(x_{m+1})\alpha_h(x_m) = -R_h\sin(k'h).$$

Hence, to satisfy (1.78) then $\Delta_h = -R_h h\sin(k'h)$. It should be remarked that since $kh \leq \alpha < \pi$, estimate (1.70) leads to $k'h < \alpha + C\delta^4h^3$ which is smaller than π for δ and h small enough. In addition, from (1.72) and the estimate (1.76) it is guaranteed that R_h is lower bounded by a positive constant far from being null. Consequently, the Green's function given by (1.77) is well-defined.

The most attractive feature of the Green's function is that it allows to write explicitly the inverse of the matrix $\mathcal{L}_h^{\mathfrak{v}}$, or equivalently, to write in closed form the solution of the linear system $\mathcal{L}_h^{\mathfrak{v}}\vec{u}_h^{\mathfrak{v}} = \vec{f}_h^{\mathfrak{v}}$. Using (1.77) and taking into account that the PUFEM discrete functions $u_h^{\mathfrak{v}}$ in $V_h^{\mathfrak{v}}$ is determined by the vector $\vec{u}_h^{\mathfrak{v}}$ of its point-wise values, it holds

$$\vec{u}_h^{\mathfrak{v}}(x_l) = [\vec{u}_h^{\mathfrak{v}}]_l = h\sum_{j=1}^{n} G_h(x_l, x_j)[\vec{f}_h^{\mathfrak{v}}]_j. \tag{1.81}$$

From the equation written above it is immediately to deduce that the coefficients of the inverse matrix of $\mathcal{L}_h^{\mathfrak{v}}$ (in the case of being uniquely defined) are given by $[(\mathcal{L}_h^{\mathfrak{v}})^{-1}]_{lj} = hG_h(x_l, x_j)$.

Lemma 1.5.8. *Given the source data $f \in \mathrm{L}^2(0,1)$ and assuming that $hk \leq \alpha < 1$ and $\delta^4h^4 < (1-\beta)/(\sqrt{2}\hat{C})$ being $\hat{C}$ the approximation constant involved in* (1.40)*, if $u_h^{\mathfrak{v}} \in \tilde{V}_h^{\mathfrak{v}}$ is a solution of the discrete variational problem* (1.57) *then*

$$|u_h^{\mathfrak{v}}|_1 \leq C\|f\|_0, \tag{1.82}$$

where C is a positive constant independent of h, k and δ (depending only on α and β).

Proof. The proof is entirely analogous to those one shown in [27, Lemma 3]. The estimate (1.82) is obtained by using the equivalence of norms between stated in Lemma 1.5.5. □

Since the explicit computation of the Green's function and its well-posedness for $h(k+\delta) \leq \alpha < 1$ can be read as the proof of existence of solution for the discrete problem (1.57), the following theorem guarantees the uniqueness of solution by means of the discrete *inf-sup* condition (see [8] for a detailed discussion).

Lemma 1.5.9. *If it is assumed $h(k+\delta) \leq \alpha < 1$ and $\delta^4 h^4 < (1-\beta)/(\sqrt{2}\hat{C})$, being $\hat{C}$ the approximation constant involved in* (1.40)*, then it holds the inf-sup condition*

$$\inf_{u_h \in \tilde{V}_h^{\mathfrak{v}}} \sup_{v_h \in \tilde{V}_h^{\mathfrak{v}}} \frac{|B_k(u_h, v_h) - iku_h(1)\bar{v}_h(1)|}{|u_h|_1 \, |v_h|_1} \geq \frac{C}{k}, \tag{1.83}$$

where C is a positive constants independent of h, k, and δ (depending only on α).

Proof. The same kind of arguments used in [27, Appendix B] will be followed. Inequality (1.83) is equivalent to show that

$$\sup_{v_h \in \tilde{V}_h^{\mathfrak{v}}} \frac{|B_k(u_h, v_h) - iku_h(1)\bar{v}_h(1)|}{|u_h|_1 \, |v_h|_1} \geq \frac{C}{k}|u_h|_1 \qquad \text{for all } u_h \in \tilde{V}_h^{\mathfrak{v}}.$$

With the aim of proof the inequality written above, fix an arbitrary $u_h \in \tilde{V}_h^{\mathfrak{v}}$ and define $v_h = u_h + z_h$, being z_h the solution of the auxiliary problem

$$B_k(w_h, z_h) - ikw_h(1)\bar{z}_h(1) = k^2 \langle w_h, z_h \rangle_{\mathrm{L}^2(0,1)} \qquad \text{for all } w_h \in \tilde{V}_h^{\mathfrak{v}}.$$

Since $k^2 u_h \in \mathrm{L}^2(0,1)$, this problem has at least a solution given by the application of the discrete Green's function. The arguments used in [27, Appendix B] in combination with the equivalence of norms stated in Lemma 1.5.5 shows that

$$|z_h|_1 \leq Ck \left| \frac{k}{k'} \right| \|u_h'\|_0,$$

with a positive constant C independent of h, k, and δ. From the estimation (1.70) and the assumptions of the present lemma, it is immediate to check that k/k' is bounded independently of h, k, and δ and hence it holds

$$|z_h|_1 \leq Ck|u_h|_1. \tag{1.84}$$

Coming back to the numerator in the *inf-sup* condition ans using the expression of $v_h = u_h + z_h$, it is satisfied

$$\begin{aligned} B_k(u_h, v_h) - iku_h(1)\bar{v}_h(1) &= B_k(u_h, u_h + z_h) - iku_h(1)(\bar{u}_h(1) + \bar{z}_h(1) \\ &= B_k(u_h, u_h) - ik|u_h(1)|^2 + B_k(u_h, z_h) - iku_h(1)\bar{z}_h(1) \\ &= B_k(u_h, u_h) - ik|u_h(1)|^2 + k^2 \langle u_h, u_h \rangle_{\mathrm{L}^2(0,1)} \\ &= |u_h|_1^2 - ik|u_h(1)|^2, \end{aligned}$$

and so, using (1.84), $|v_h|_1 \leq (1 + Ck)|u_h|_1$. In consequence, it holds

$$\sup_{v_h \in \tilde{V}_h^{\mathfrak{v}}} \frac{|B_k(u_h, v_h) - iku_h(1)\bar{v}_h(1)|}{|u_h|_1 \, |v_h|_1} \geq \frac{|u_h|_1^2}{|v_h|_1} \geq \frac{1}{1 + Ck}|u_h|_1,$$

which leads to (1.83) since k is strictly positive lower bounded far from zero. □

***Remark* 1.5.10** (Dual norm estimate)**.** *Using an standard argument, the inf-sup condition also provides automatically a stability estimate in terms of the dual norm in* $(\mathrm{H}^1_{(0}(0,1))'$ *(see [28, Section 2.1]). More precisely, using* (1.83)*,*

$$\frac{C}{k}|u_h|_1 \le \sup_{v_h \in \tilde{V}_h^{\mathfrak{v}}} \frac{|B_k(u_h, v_h) - iku_h(1)\bar{v}_h(1)|}{|v_h|_1}$$

$$= \sup_{v_h \in \tilde{V}_h^{\mathfrak{v}}} \frac{|\langle f, v_h\rangle_{\mathrm{L}^2}|}{|v_h|_1} \le \sup_{v \in \mathrm{H}^1_{(0}(0,1)} \frac{|\langle f, v\rangle_{\mathrm{L}^2}|}{|v|_1} = \|f\|_{\mathrm{H}^1_{(0}(0,1)'}$$

and hence $|u_h|_1 \le Ck\|f\|_{\mathrm{H}^1_{(0}(0,1)'}$ *(notice that implicitly it has been used the continuous embedding of* $\mathrm{L}^2(0,1)$ *in* $(\mathrm{H}^1_{(0}(0,1))'$*).*

Finally, combining the stability estimates for the sub-problems stated in $\tilde{V}_h^{\mathfrak{v}}$ and $\mathrm{V}_h^{\mathfrak{b}}$, it can be stated an stability result for the whole discrete problem stated in the PUFEM discrete space V_h.

Theorem 1.5.11. *If it is assumed* $h(k+\delta) \le \alpha < 1$ *and* $\delta^4 h^4 < (1-\beta)/(\sqrt{2}\hat{C})$*, being* $\hat{C}$ *the approximation constant involved in* (1.40)*, then there exists an unique solution* $u_h \in \mathrm{V}_h$ *of the discrete PUFEM problem* (1.6)*. In addition, it holds the stability estimate*

$$|u_h|_1 \le C\|f\|_0, \tag{1.85}$$

where C *is a positive constant independent of* h*,* k*, and* δ*.*

Proof. The existence and uniqueness result comes from straightforwardly from the existence and uniqueness solution of both sub-problems (1.57) and (1.47) defined in $\tilde{V}_h^{\mathfrak{v}}$ and $\mathrm{V}_h^{\mathfrak{b}}$, respectively. In addition, since $u_h = u_h^{\mathfrak{v}} + u_h^{\mathfrak{v}}$, estimate (1.85) is obtained combining (1.56) and (1.82). □

1.6 A priori error estimate

Finally, this section is devoted to write sharp error estimates for the PUFEM discretization. More precisely, it will be estimate the H^1-distance between oscillatory functions, which are exact solutions of the Helmholtz problem and the PUFEM approximations. The main two ingredients to obtain such estimates are the stability of the discrete PUFEM variational problem (stated in the previous section), the interpolant estimates for oscillatory solutions described in Section 1.4, and its relation with the projections of the exact solution in the PUFEM discrete space.

Firstly, to highlight the difficulties of passing to the limit when $h(k+\delta)$ tends to zero, it will be shown that the functions belonging to the PUFEM space on uniform meshes satisfy an inverse inequality, once the limit case $h(k+\delta) = 0$ is avoided.

Lemma 1.6.1 (Inverse inequality). *For any fixed $\varepsilon > 0$, if $\varepsilon \le h(k+\delta) \le \alpha < \pi$, then there exist constants C_0 and C_1 independent of h, k and δ (only dependent of ε and α) such that*

$$\frac{C_0}{h}\|v_h\|_0 \le |v_h|_1 \le \frac{C_1}{h}\|v_h\|_0, \tag{1.86}$$

for all $v_h \in \mathrm{V}_h$.

Proof. It will be followed a slight modification of the steps used in [11, Chapter 3] to proof the classical inverse inequality in standard polynomial spaces, for instance, for any continuous piecewise $\mathbb{P}^1$-discrete function $v_h^{\mathrm{fe}} \in \langle\{\varphi_j\}_{j=0}^n\rangle$ defined on an one-dimensional equispaced mesh where it is satisfies $|v_h^{\mathrm{fe}}|_1 \le C/h\|v_h^{\mathrm{fe}}\|_0$.

Clearly, if $v_h \in \mathrm{V}_h$ then $\hat{v}_j = v_h|_{T_j} \circ F_j^{-1}$ belongs to the span of the local shape functions $\langle\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_1^{\mathfrak{b}}, \hat{\theta}_2^{\mathfrak{b}}\rangle$ defined in (1.16)-(1.19), this is, v_j is represented by its coordinate basis vector $\vec{v} = (v_1, v_2, v_1^{\mathfrak{b}}, v_2^{\mathfrak{b}})^t$ A direct inspection reveals that all of these shape functions depend continuously on the parameter $h(k+\delta) \in [\varepsilon, \alpha]$. In addition, if $\mathcal{K}_{\mathrm{loc}}$ and $\mathcal{M}_{\mathrm{loc}}$ denotes the local stiffness and mass matrices with respect to this local shape basis, the coefficients of these matrices also depend continuously on the parameter $h(k+\delta)$ and it holds

$$\lambda_{\min}(h(k+\delta)) \le \frac{\vec{v}^*\mathcal{K}_{\mathrm{loc}}\vec{v}}{\vec{v}^*\mathcal{M}_{\mathrm{loc}}\vec{v}} \le \lambda_{\max}(h(k+\delta))$$

where $\lambda_{\min}(h(k+\delta))$ and $\lambda_{\max}(h(k+\delta))$ are respectively the minimum and maximum eigenvalues of the symmetric generalized eigenvalue problem $\mathcal{K}_{\mathrm{loc}}\vec{v} = \lambda\mathcal{M}_{\mathrm{loc}}\vec{v}$, whose eigensolutions also depend continuously on the parameter $h(k+\delta)$. Hence, the maps $h(k+\delta) \mapsto \lambda_{\min}(h(k+\delta))$ and $h(k+\delta) \mapsto \lambda_{\max}(h(k+\delta))$ are continuous functions defined in a non-empty compact domain $[\varepsilon, \alpha]$. So, using the Weierstrass theorem, both continuous functions reaches respectively a minimum $\lambda_{\min}$ and a maximum value $\lambda_{\max}$ (possibly depending on ε and α). Hence, taking into account that $|\hat{v}_j|^2_{\mathrm{H}^1(\hat{T})} = \vec{v}^*\mathcal{K}_{\mathrm{loc}}\vec{v}$ and $\|\hat{v}_j\|^2_{\mathrm{L}^2(\hat{T})} = \vec{v}^*\mathcal{M}_{\mathrm{loc}}\vec{v}$, it holds

$$\lambda_{\min}\|\hat{v}_j\|^2_{\mathrm{L}^2(\hat{T})} \le |\hat{v}_j|^2_{\mathrm{H}^1(\hat{T})} \le \lambda_{\max}\|\hat{v}_j\|^2_{\mathrm{L}^2(\hat{T})}$$

or equivalently, coming back to the element T_j,

$$\frac{1}{h^2}\lambda_{\min}\|v_h|_{T_j}\|^2_{\mathrm{L}^2(T_j)} \le |v_h|_{T_j}|^2_{\mathrm{H}^1(T_j)} \le \frac{1}{h^2}\lambda_{\max}\|v_h|_{T_j}\|^2_{\mathrm{L}^2(T_j)}.$$

If the terms in the previous inequality are added and the root square is computed, estimates (1.86) are obtained. □

***Remark* 1.6.2.** *The assumption of $\varepsilon < h(k+\delta)$ is essential to avoid the limit case $h(k+\delta) = 0$. However, it does not suppose any restriction on the error analysis since ε can be chosen as small as it would be desired independently of k and δ. As it has been discussed previously at the beginning of Section 1.5, if formally it is considered the limit case $h(k+\delta) = 0$, the PUFEM subspace $\mathrm{V}_h^{\mathfrak{p}}$ will be identical to the standard continuous piecewise $\mathbb{P}^1$-finite elements and the restrictions of functions of $\mathrm{V}_h^{\mathfrak{b}}$ at each element coincides with the $\mathbb{P}^2$-bubble functions. In the classical polynomial bubble space, the number of bubbles coincides* *with the number of element, i.e., n. However, the number of basis elements in* $\mathrm{V}_h^{\mathfrak{b}}$ *coincides with the number of vertices* $n+1$*. So, in the limit case of* $h(k+\delta) = 0$*, the twin-bubble basis of* $\mathrm{V}_h^{\mathfrak{b}}$ *collapses and a function of this discrete basis should be removed to avoid a linear dependency.*

Despite of it will not be used throughout the present work, for the sake of completeness, the difference between the L^2 and H^1-projections will be estimated in terms of the H^1-norm.

Lemma 1.6.3 (Projection differences)**.** *Let* $u \in \mathrm{H}^1(0,1)$ *be the exact solution of the variational problem* (1.2)*. Fixed* $0 < \varepsilon \le h(k+\delta) \le \alpha < \pi$*, if* $\Pi^h_{\mathrm{L}^2}u$ *and* $\Pi^h_{\mathrm{H}^1}u$ *are respectively the* L^2 *and* H^1*-projections in the PUFEM discrete space given by* (1.43) *and* (1.4.4)*, it holds*

$$\left|\Pi^h_{\mathrm{L}^2}u - \Pi^h_{\mathrm{H}^1}u\right|_1 \le Ch^2\delta^2(|u|_1 + k\|u\|_0) \tag{1.87}$$

where C is a positive constant independent of h, k and δ (only dependent on ε and α).

Proof. Firstly, using successive triangular inequalities and the inverse inequality (1.86), it holds

$$\begin{aligned}
|\Pi^h_{\mathrm{L}^2}u - \Pi^h_{\mathrm{H}^1}u|_1 &\le |\Pi^h_{\mathrm{L}^2}u - \mathcal{I}_hu|_1 + |\mathcal{I}_hu - \Pi^h_{\mathrm{H}^1}u|_1 \\
&\le \frac{C_1}{h}\|\Pi^h_{\mathrm{L}^2}u - \mathcal{I}_hu\|_0 + |\mathcal{I}_hu - u|_1 + |u - \Pi^h_{\mathrm{H}^1}u|_1 \\
&\le \frac{C_1}{h}\|\Pi^h_{\mathrm{L}^2}u - \mathcal{I}_hu\|_0 + \frac{C_1}{h}\|\Pi^h_{\mathrm{L}^2}u - u\|_0 + \|u - \mathcal{I}_hu\| + 2|\mathcal{I}_hu - u|_1 \\
&\le \frac{2C_1}{h}\|\mathcal{I}_hu - u\|_0 + 2|\mathcal{I}_hu - u|_1
\end{aligned}$$

Estimates (1.39) and (1.38) for the interpolant $\mathcal{I}_hu$ in Lemma 1.4.4, once they are inserted in the inequality written above, lead straightforwardly to (1.87). □

Finally, since all the ingredients are being introduced in the previous sections, it is possible to conclude an *a priori* error estimate for the approximation computed by means of the PUFEM discretization.

Theorem 1.6.4. *Let* $u \in \mathrm{H}^1(0,1)$ *be a solution of the variational problem* (1.2) *and let* $u_h \in \mathrm{V}_h$ *be the solution of the PUFEM discrete problem defined in* (1.6)*. If it is assumed* $h(k+\delta) \le \alpha < 1$ *and* $\delta^4h^4 < (1-\beta)/(\sqrt{2}\hat{C})$*, being* $\hat{C}$ *the approximation constant involved in* (1.40)*, then it holds*

$$|u - u_h|_1 \le Ck|u - \mathcal{I}_hu|_1, \tag{1.88}$$

where C is a positive constant independent of h, δ and k.

Proof. Firstly, since $u - u_h$ is orthogonal to V_h with respect to the sesquilinear form of the variational problem (1.6), it holds

$$B_k(u - u_h, v_h) - ik(u(1) - u_h(1))\bar{v}_h(1) = 0$$

for all $v_h \in \mathrm{V}_h$. Hence, if $z_h = u_h - \mathcal{I}_h u \in \mathrm{V}_h$ then, since $z_h = (u_h - u) + (u - \mathcal{I}_h u$, the discrete function z_h is the solution of the following variational problem:

$$B_k(z_h, v_h) - ikz(1)\bar{v}_h(1) = -B_k(u - \mathcal{I}_h u, v_h)$$

for all $v_h \in \mathrm{V}_h$. Since $\mathrm{V}_h = \tilde{V}_h^{\mathfrak{v}} \oplus \mathrm{V}_h^{\mathfrak{b}}$, the variational equality written above is satisfied independently for test functions in the vertex-value space $\tilde{V}_h^{\mathfrak{v}}$ and in the twin-bubble space $\mathrm{V}_h^{\mathfrak{b}}$. Same kind of considerations can be applied to split z_h, this is, $z_h = z_h^{\mathfrak{v}} + z_h^{\mathfrak{b}}$. Due the orthogonality relation between the discrete spaces $\tilde{V}_h^{\mathfrak{v}}$ and $\mathrm{V}_h^{\mathfrak{b}}$, each of these functions, $z_h^{\mathfrak{v}}$ and $z_h^{\mathfrak{b}}$ are respectively solution of the variational problems

$$\begin{aligned} &B_k(z_h^{\mathfrak{v}}, v_h^{\mathfrak{v}}) - ikz_h^{\mathfrak{v}}(1)\bar{v}_h^{\mathfrak{v}}(1) = -B_k(u - \mathcal{I}_h u, v_h^{\mathfrak{v}}) && \text{for all } v_h^{\mathfrak{v}} \in \tilde{V}_h^{\mathfrak{v}}, \\ &B_k(z_h^{\mathfrak{b}}, v_h^{\mathfrak{b}}) = -B_k(u - \mathcal{I}_h u, v_h^{\mathfrak{b}}) && \text{for all } v_h^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}. \end{aligned}$$

In addition, due to the linearity of these two problems, their solutions can be rewritten as the sum of two new discrete functions, $z_h^{\mathfrak{v}} = z_{1h}^{\mathfrak{v}} + z_{2h}^{\mathfrak{v}}$ and $z_h^{\mathfrak{b}} = z_{1h}^{\mathfrak{b}} + z_{2h}^{\mathfrak{b}}$ where each addend is solution respectively of the following variational problems:

$$B_k(z_{1h}^{\mathfrak{v}}, v_h^{\mathfrak{v}}) - ikz_{1h}^{\mathfrak{v}}(1)\bar{v}_h^{\mathfrak{v}}(1) = -\langle (u - \mathcal{I}_h u)', (v_h^{\mathfrak{v}})' \rangle_{\mathrm{L}^2(0,1)} \qquad \text{for all } v_h^{\mathfrak{v}} \in \tilde{V}_h^{\mathfrak{v}}, \tag{1.89}$$

$$B_k(z_{2h}^{\mathfrak{v}}, v_h^{\mathfrak{v}}) - ikz_{2h}^{\mathfrak{v}}(1)\bar{v}_h^{\mathfrak{v}}(1) = k^2 \langle u - \mathcal{I}_h u, v_h^{\mathfrak{v}} \rangle_{\mathrm{L}^2(0,1)} \qquad \text{for all } v_h^{\mathfrak{v}} \in \tilde{V}_h^{\mathfrak{v}}, \tag{1.90}$$

$$B_k(z_{1h}^{\mathfrak{b}}, v_h^{\mathfrak{b}}) = -\langle (u - \mathcal{I}_h u)', (v_h^{\mathfrak{b}})' \rangle_{\mathrm{L}^2(0,1)} \qquad \text{for all } v_h^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}, \tag{1.91}$$

$$B_k(z_{2h}^{\mathfrak{b}}, v_h^{\mathfrak{b}}) = k^2 \langle u - \mathcal{I}_h u, v_h^{\mathfrak{b}} \rangle_{\mathrm{L}^2(0,1)} \qquad \text{for all } v_h^{\mathfrak{b}} \in \mathrm{V}_h^{\mathfrak{b}}. \tag{1.92}$$

For each one of the solutions of the discrete variational problems stated above, it can be applied some of the estimates written in the previous sections. More precisely, if the arguments described in Remark 1.5.10, an analogous derivation to those one used to obtain (1.84), a coercive estimate similar to (1.48), and (1.51) are applied respectively to the solutions of problems (1.89)-(1.92), then it is satisfied

$$\begin{aligned} |z_{1h}^{\mathfrak{v}}|_1 &\leq Ck|u - \mathcal{I}_h u|_1, \\ |z_{2h}^{\mathfrak{v}}|_1 &\leq Ck \left| \frac{k}{k'} \right| |u - \mathcal{I}_h u|_1, \\ |z_{1h}^{\mathfrak{b}}|_1 &\leq C|u - \mathcal{I}_h u|_1, \\ |z_{2h}^{\mathfrak{b}}|_1 &\leq Chk^2|u - \mathcal{I}_h u|_1, \end{aligned}$$

where C is a positive constant independent of h, k, and δ (depending only on α and β).

Finally, collecting all these estimates and using the fact that $z_h = z_{1h}^{\mathfrak{v}} + z_{2h}^{\mathfrak{v}} + z_{1h}^{\mathfrak{b}} + z_{2h}^{\mathfrak{b}}$, then

$$|z_h|_1 \leq C(1 + k)|u - \mathcal{I}_h u|_1 + Ck\|u - \mathcal{I}_h u\|_0,$$

from which (1.88) is concluded applying a Poincare inequality and due to k is strictly positive lower bounded far from zero. □

Corollary 1.6.5. *Let* $u \in \mathrm{H}^1(0,1)$ *be an oscillatory solution of the variational problem* (1.2) *and let* $u_h \in \mathrm{V}_h$ *be the solution of the PUFEM discrete problem defined in* (1.6)*. If it is assumed* $0 < \varepsilon \leq h(k+\delta) \leq \alpha < 1$ *and* $\delta^4 h^4 < (1-\beta)/(\sqrt{2}\hat{C})$*, being* $\hat{C}$ *the approximation constant involved in* (1.40)*, then it holds*

$$|u - u_h|_1 \leq Ckh^2\delta^2|u|_1, \tag{1.93}$$

where C *is a positive constant independent of* h*,* δ *and* k*.*

Proof. The combination of estimates (1.88) and (1.40) leads to (1.93). □

As it will be checked in the following section, this estimate can be improved for oscillatory solutions. In fact, since they are solutions of the Helmholtz equation with smooth right-hand side $f \in \mathrm{H}^l(0,1)$ with $l \geq 1$, duality stability estimates (analogous to that one described in [28, Theorem 3.2]) should be used to obtain a more accurate estimate (possibly independent of k).

1.7 Numerical Results

In this section, some numerical results are shown to illustrate how the PUFEM discrete errors depend on the mesh size, the wave number and the perturbation parameter. For this purpose, the boundary data (or equivalently the source term) is chosen in problem (1.1) to obtain $u(x) = \sin(kx)$ as the exact solution. The relative error for the PUFEM discretization has been computed in terms of the L^2-norm and H^1-seminorm. Plots on Figures 1.9 and 1.10 illustrate the second-order accuracy of the PUFEM approximation with respect to the mesh size. In addition, it confirms the error estimate (1.88) once $hk < \pi$.

Moreover, Figures 1.9 and 1.10 also show the dependence of PUFEM relative errors on the wave number k. Overall it can be checked that the PUFEM relative error does not depend on the wave number values. The convergent second-order behaviour of the perturbation parameter δ that holds in (1.88) for the PUFEM discretization can be checked for both L^2 and H^1-error curves.

1.8 Conclusions

In this chapter, a one-dimensional Helmholtz problem and its weak formulation have been posed. The LBB continuous ans discrete conditions have been demonstrated. A plane wave based PUFEM discretization in terms of exponential and trigonometrical functions have been described. Two interpolation estimates have been proved and from that, an *a priori* error estimate for the approximation computed by means of the PUFEM discretization has been deduced. The numerical results confirm the second order of accuracy of the PUFEM approximation with respect to the mesh size h and the additional perturbation parameter δ that is stated in the error estimate. The independence on the wave number

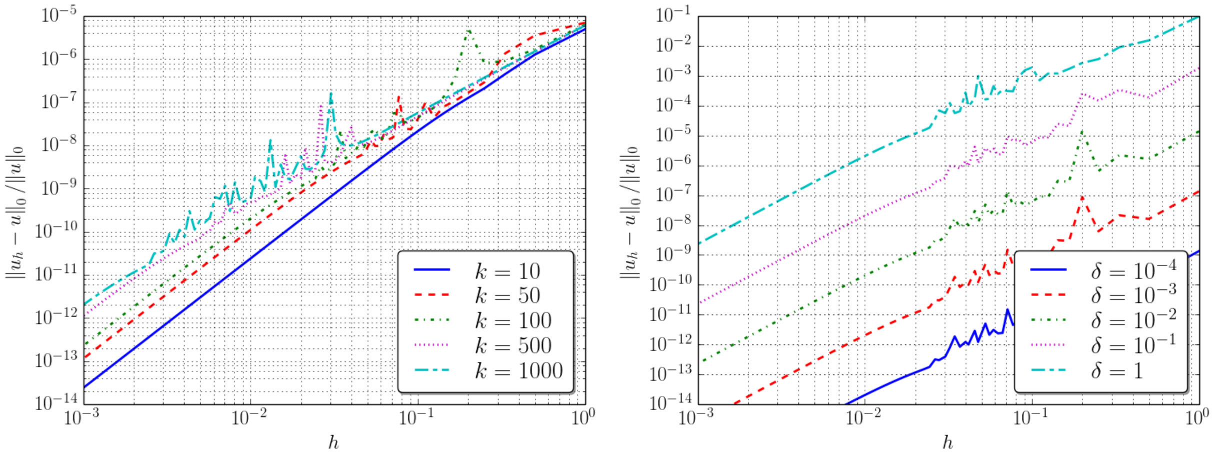

Figure 1.9: L^2-approximation errors of the PUFEM solution (computed when the exact solution is given by $u(x) = \sin(kx)$), plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

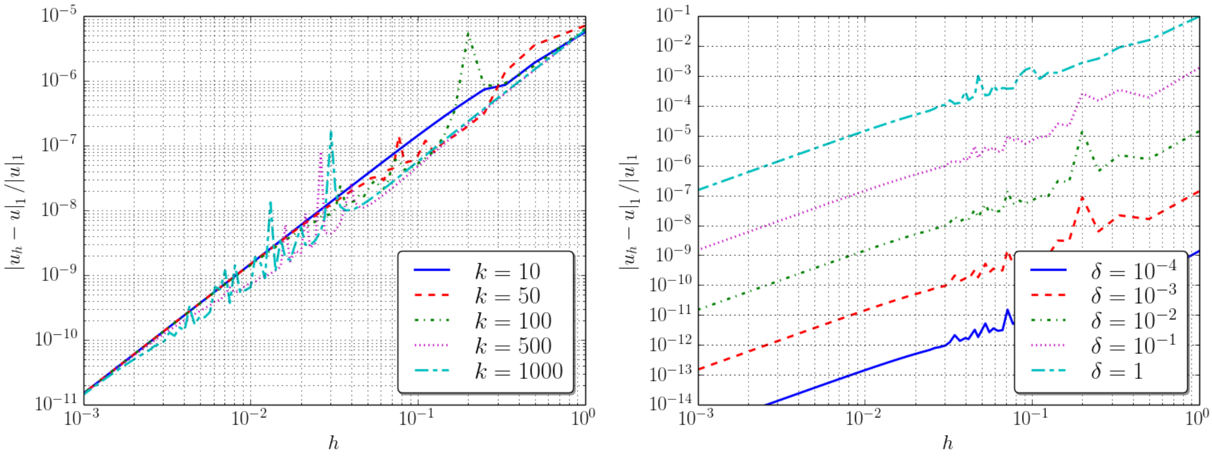

Figure 1.10: H^1-approximation errors of the PUFEM solution (computed when the exact solution is given by $u(x) = \sin(kx)$), plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^{-2}$ (left) or the wave number $k = 100$ (right).

that it is observed on the figures was not obtained on the error estimate, so the estimate can be improved by increasing the smoothness of the right hand side.

Appendices

1.A Maclaurin expansions to obtain estimates (1.27)-(1.29)

The L^2 and H^1-distance between the solution of the Helmholtz equation and its interpolant in the PUFEM discrete space (derived in Section 1.4) depends on a sophisticated manner through rational and polynomial expressions on parameters h, δ, k and the trigonometric functions either $\cos(\delta h)$ or $\sin(\delta h)$. More precisely, to obtain the interpolatory estimate (1.27), it must be bounded the function

$$f(h,\delta) = \frac{5}{3}h + \frac{h}{3}\cos(\delta h) + \frac{4}{\delta^2 h}\left(\cos(\delta h) - 1\right) \tag{1.94}$$

and analogously, to have the estimate (1.29), it has to be considered

$$\begin{aligned} g(h,\delta,k) = \frac{2}{h} + hk^2 + \frac{2}{3}h(k+\delta)^2 - \frac{4k^2}{\delta^2 h} - 2(k+\delta)\sin(\delta h) \\ + 2\cos(\delta h)\left(-\frac{1}{h} + \frac{2k^2}{\delta^2 h}\right) + \frac{h}{3}\cos(\delta h)(k+\delta)^2. \end{aligned} \tag{1.95}$$

In both cases, Maclaurin polynomials in h applied to $\cos(\delta h)$ or $\sin(\delta h)$ will be used with different orders at each occurrence in the expressions of $f(h,\delta)$ and $g(h,\delta,k)$.

To bound (1.94), the expression $\cos(\delta h)$ is replaced by a Maclaurin polynomial of third order in the second addend and by a fifth order polynomial in the third one. So, it holds

$$\begin{aligned} f(h,\delta) = \frac{5}{3}h + \frac{h}{3}\left(1 - \frac{1}{2!}\delta^2 h^2 + \frac{1}{4!}\delta^4\cos(\delta\xi_1)h^4\right) \\ + \frac{4}{\delta^2 h}\left(-\frac{1}{2!}\delta^2 h^2 + \frac{1}{4!}\delta^4 h^4 - \frac{1}{6!}\delta^6\cos(\delta\xi_2)h^6\right) = \delta^4 h^5\left(\frac{\cos(\delta\xi_1)}{24} - \frac{\cos(\delta\xi_2)}{180}\right) \leq \frac{17}{360}\delta^4 h^5, \end{aligned} \tag{1.96}$$

where ξ_1 and ξ_2, involved in the Cauchy remainders, belong to $[0,h]$. In an analogous manner, if the expression $\sin(\delta h)$ in (1.95) is replaced by a fourth-order Maclaurin polynomial and the occurrences of $\cos(\delta h)$ in the last two addends of (1.95) are replaced respectively by a fifth- and third-order polynomials, it yields

$$
\begin{aligned}
g(h,\delta,k) =& \frac{2}{h} + hk^2 + \frac{2}{3}h(k+\delta)^2 - \frac{4k^2}{\delta^2 h} - 2(k+\delta)\left(\delta h - \frac{1}{3!}\delta^3 h^3 + \frac{1}{5!}\delta^5\cos(\delta\xi_3)h^5\right) \\
&+ \left(1 - \frac{1}{2!}\delta^2 h^2 + \frac{1}{4!}\delta^4 h^4 - \frac{1}{6!}\delta^6\cos(\delta\xi_4)h^6\right)\left(-\frac{2}{h} + \frac{4k^2}{\delta^2 h}\right) \\
&+ \left(1 - \frac{1}{2!}\delta^2 h^2 + \frac{1}{4!}\delta^4\cos(\delta\xi_5)h^4\right)\frac{h}{3}(k+\delta)^2 \\
=& \frac{1}{12}d^4h^3 + \delta^6 h^5\left(-\frac{1}{60}\cos(\delta\xi_3) + \frac{1}{360}\cos(\delta\xi_4) + \frac{1}{72}\cos(\delta\xi_5)\right) \\
&+ d^5h^5k\left(-\frac{1}{60}\cos(\delta\xi_3) + \frac{1}{36}\cos(\delta\xi_5)\right) + d^4h^5k^2\left(-\frac{1}{180}\cos(\delta\xi_4) + \frac{1}{72}\cos(\delta\xi_5)\right) \\
\leq& \frac{1}{12}\delta^4 h^3 + \frac{1}{30}\delta^6 h^5 + \frac{2}{45}\delta^5 k h^5 + \frac{7}{360}\delta^4 k^2 h^5 \qquad (1.97)
\end{aligned}
$$

where ξ_3, ξ_4, and ξ_5, involved in the Cauchy remainders, belong to $[0,h]$.

1.B Discrete dispersion equations

The main aim in the analysis of the numerical dispersion relation of the PUFEM discretization consists in the identification of those equations satisfied by the discrete wave number k_d in terms of the parameters h, k and δ. With that purpose, following analogous ideas to [2], a linear combination of Bloch waves with wave number k_d involving the discrete PUFEM functions,

$$
U_h(x) = \alpha\sum_{m\in\mathbb{Z}} e^{ik_\mathrm{d}mh}\varphi_m(x)e^{i(k+\delta)(x-x_m)} + \beta\sum_{m\in\mathbb{Z}} e^{-ik_\mathrm{d}mh}\varphi_m(x)e^{-i(k+\delta)(x-x_m)}, \qquad \alpha,\beta\in\mathbb{C}, \qquad (1.98)
$$

will be imposed on the discrete variational problem in a infinite uniform mesh of size h with vertices $x_j = jh$, $j\in\mathbb{Z}$. It is clear that each addend in (1.98) is a Bloch wave. More precisely, if $U_h^+(x) = \sum_{m\in\mathbb{Z}} e^{ik_\mathrm{d}mh}\varphi_m(x)e^{i(k+\delta)(x-x_m)}$, it holds $U_h^+(x+lh) = e^{ik_\mathrm{d}lh}U_h^+(x)$ as follows:

$$
\begin{aligned}
U_h^+(x+lh) = \sum_{m\in\mathbb{Z}} e^{ik_\mathrm{d}mh}\varphi_m(x+lh)e^{i(k+\delta)(x+lh-x_m)} &= \sum_{m\in\mathbb{Z}} e^{ik_\mathrm{d}mh}\varphi_{m-l}(x)e^{i(k+\delta)(x-x_{m-l})} \\
&= e^{ik_\mathrm{d}lh}\sum_{m\in\mathbb{Z}} e^{ik_\mathrm{d}(m-l)h}\varphi_{m-l}(x)e^{i(k+\delta)(x-x_{m-l})} = e^{ik_\mathrm{d}lh}U_h^+(x),
\end{aligned}
$$

where it has been used the translation property of the finite element basis on an equispaced mesh $\varphi_m(x+lh) = \varphi_{m-l}(x)$.

To deduce the dispersion equations satisfied by the discrete wave number k_d, the Bloch wave is imposed on the variational formulation, which can be written as a non-linear eigenvalue problem: Find $k_\mathrm{d}\in\mathbb{C}$ and a non-null vector $(\alpha,\beta)^t\in\mathbb{C}^2$ such that

$$
B_k(U_h,\psi_{2j-1}) = 0, \quad B_k(U_h,\psi_{2j}) = 0, \qquad \text{for all } j\in\mathbb{Z}. \qquad (1.99)
$$

Due to the invariance of the PUFEM discrete functions under translation of integer multiples of the mesh size, it is enough to use as PUFEM trial functions $\psi_{-1}(x) = \varphi_0(x)e^{-i(k+\delta)x}$ and $\psi_0(x) = \varphi_0(x)e^{i(k+\delta)x}$ in the equations written above. Hence, inserting the expression of U_h in the eigenvalue problem (1.99), and taking into account that the compact support of ψ_{-1} and ψ_0 only intersect the support of $\{\psi\}_{j=-3}^{2}$, $(\alpha,\beta)^t$ and k_d must satisfy the dispersion equations

$$\begin{aligned}
&\alpha\left(e^{-ik_\mathrm{d}h}B_k(\psi_{-2},\psi_{-1}) + B_k(\psi_0,\psi_{-1}) + e^{ik_\mathrm{d}h}B_k(\psi_2,\psi_{-1})\right)\\
&\quad+\beta\left(e^{ik_\mathrm{d}h}B_k(\psi_{-3},\psi_{-1}) + B_k(\psi_{-1},\psi_{-1}) + e^{-ik_\mathrm{d}h}B_k(\psi_1,\psi_{-1})\right) = 0,\\
&\alpha\left(e^{-ik_\mathrm{d}h}B_k(\psi_{-2},\psi_0) + B_k(\psi_0,\psi_0) + e^{ik_\mathrm{d}h}B_k(\psi_2,\psi_0)\right)\\
&\quad+\beta\left(e^{ik_\mathrm{d}h}B_k(\psi_{-3},\psi_0) + B_k(\psi_{-1},\psi_0) + e^{-ik_\mathrm{d}h}B_k(\psi_1,\psi_0)\right) = 0.
\end{aligned}$$

Equivalently, taking into account the definition of the matrix coefficients (1.10)-(1.13), the form properties (1.8)-(1.9) and b_1, b_3 and b_4 are real valued, it is satisfied

$$\alpha\left(e^{-ik_\mathrm{d}h}b_1 + b_3 + e^{ik_\mathrm{d}h}b_1\right) + \beta\left(e^{ik_\mathrm{d}h}\bar{b}_2 + b_4 + e^{-ik_\mathrm{d}h}b_2\right) = 0, \tag{1.100}$$

$$\alpha\left(e^{-ik_\mathrm{d}h}b_2 + b_4 + e^{ik_\mathrm{d}h}\bar{b}_2\right) + \beta\left(e^{-ik_\mathrm{d}h}b_1 + b_3 + e^{ik_\mathrm{d}h}b_1\right) = 0. \tag{1.101}$$

To admit a non-null solution $(\alpha,\beta)^t$, the linear system above should have multiple solutions. So, the determinant of the associated real-valued matrix in the linear system written above should be null, this is,

$$\begin{aligned}
\det&\begin{pmatrix} b_3 + 2b_1\cos(k_\mathrm{d}h) & b_4 + 2\mathrm{Re}\left(e^{-ik_\mathrm{d}h}b_2\right)\\ b_4 + 2\mathrm{Re}\left(e^{-ik_\mathrm{d}h}b_2\right) & b_3 + 2b_1\cos(k_\mathrm{d}h)\end{pmatrix}\\
&= (b_3 + 2b_1\cos(k_\mathrm{d}h))^2 - (b_4 + 2\mathrm{Re}(e^{-ik_\mathrm{d}h}b_2))^2 = 0.
\end{aligned}\tag{1.102}$$

Hence, to have a non-null solution $(\alpha,\beta)^t$ in the dispersion equations, the equation written above leads to two cases. More precisely, case (a):

$$b_4 + 2\mathrm{Re}(e^{-ik_\mathrm{d}h}b_2) = b_3 + 2b_1\cos(k_\mathrm{d}h)$$

and case (b):

$$b_4 + 2\mathrm{Re}(e^{-ik_\mathrm{d}h}b_2) = -b_3 - 2b_1\cos(k_\mathrm{d}h).$$

Let us describe the procedure used to solve the dispersion equation in case (a). Firstly, straightforward computations allow us to rewrite the dispersion equation in case (a) in terms of $\cos(k_\mathrm{d}h)$ and $\sin(k_\mathrm{d}h)$,

$$\begin{aligned}
&\left(\left(-\frac{1}{h} + \frac{h\delta}{6}(2k+\delta)\right)\cos((k+\delta)h) - (k+\delta)\sin((k+\delta)h) - b_1\right)\cos(k_\mathrm{d}h)\\
&\quad+\left(\left(-\frac{1}{h} + \frac{h\delta}{6}(2k+\delta)\right)\sin((k+\delta)h) + (k+\delta)\cos((k+\delta)h)\right)\sin(k_\mathrm{d}h) = \frac{b_3 - b_4}{2}.
\end{aligned}$$

Introducing the auxiliary variable $y = \cos(k_{\mathrm{d}}h)$, it is possible to rewrite the equation stated above as a second-order polynomial on the variable y with real-valued coefficients. Applying similar arguments to case (b), fourth different roots y_j, $0 \leq j \leq 3$ and in consequence, four different discrete wave numbers $k_{\mathrm{d}j} = \arccos(y_j)/h$ are obtained. After some cumbersome algebraic manipulations, a direct inspection on the Taylor expansions computed[1] for the expressions of y_j reveals that for $hk \leq \pi$ it holds

$$\begin{array}{ll} \left|k_{\mathrm{d}0} - \dfrac{h^2}{60}(k^3 + 4k^2\delta)\right| \leq Ch^2\delta^2 k + \hat{C}h^4k^5, & \left|k_{\mathrm{d}1} - \left(k + \dfrac{\delta^2}{2(k+\delta)}\right)\right| \leq Ch^2\delta^2 k, \\ \left|k_{\mathrm{d}2} - \left(k - \dfrac{h^2}{3}(k^3 + 2k^2\delta)\right)\right| \leq Ch^2\delta^2 k + \hat{C}h^4k^5, & |k_{\mathrm{d}3} - k| \leq Ch^2\delta^2 k, \end{array} \tag{1.103}$$

where C and $\hat{C}$ are positive constants independent of k, h, and δ. To illustrate numerically the dispersion estimates stated above, plots on Figures 1.B.1-1.B.4 show the dependency of the different discrete wave numbers in terms of parameters k, h, and δ.

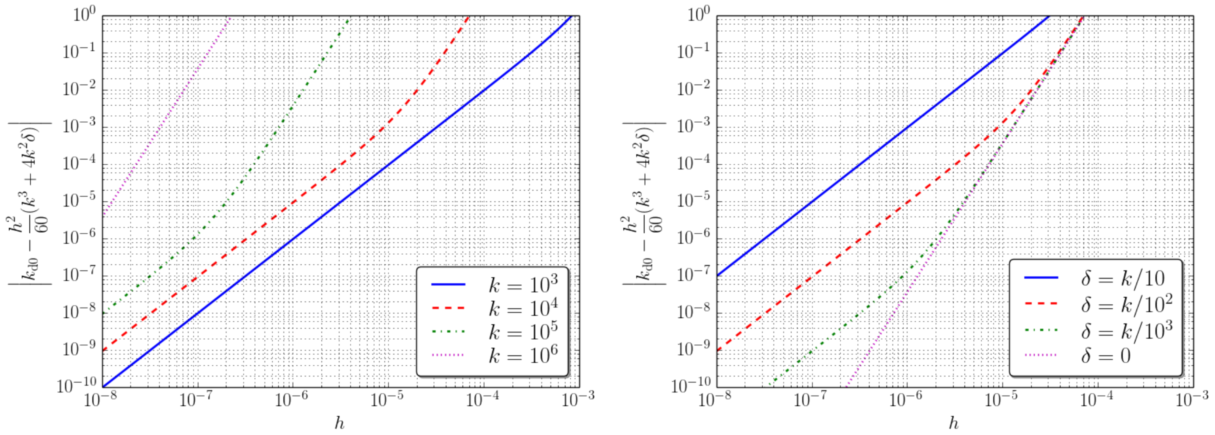

Figure 1.B.1: Difference between the discrete wave number $k_{\mathrm{d}0}$ and its asymptotic expression given by (1.103), plotted with respect to the mesh size but fixing the value of the perturbation parameter δ (left) or the wave number (right).

Inserting each one of the four discrete wave numbers on the linear system (1.100)-(1.100), the wave numbers $k_{\mathrm{d}0}$ and $k_{\mathrm{d}1}$ derived from case (a) leads to $\alpha = -\beta$ whereas the solution for $k_{\mathrm{d}2}$ and $k_{\mathrm{d}3}$ obtained from case (b) is given by $\alpha = \beta$. Consequently, from (1.98), four

[1] by means of the symbolic Python package Sympy

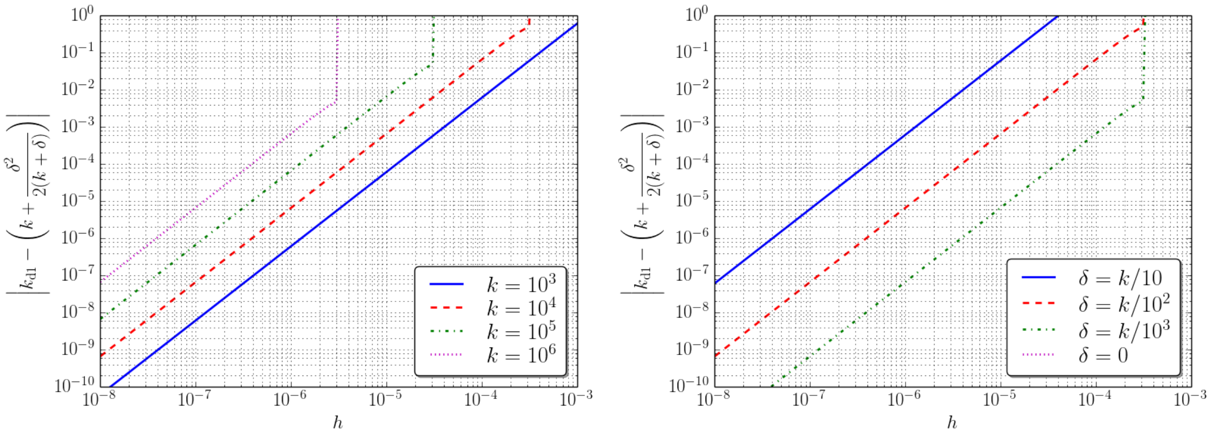

Figure 1.B.2: Difference between the discrete wave number $k_{\mathrm{d}1}$ and its asymptotic expression given by (1.103), plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^2$ (left) or the wave number $k = 10^4$ (right).

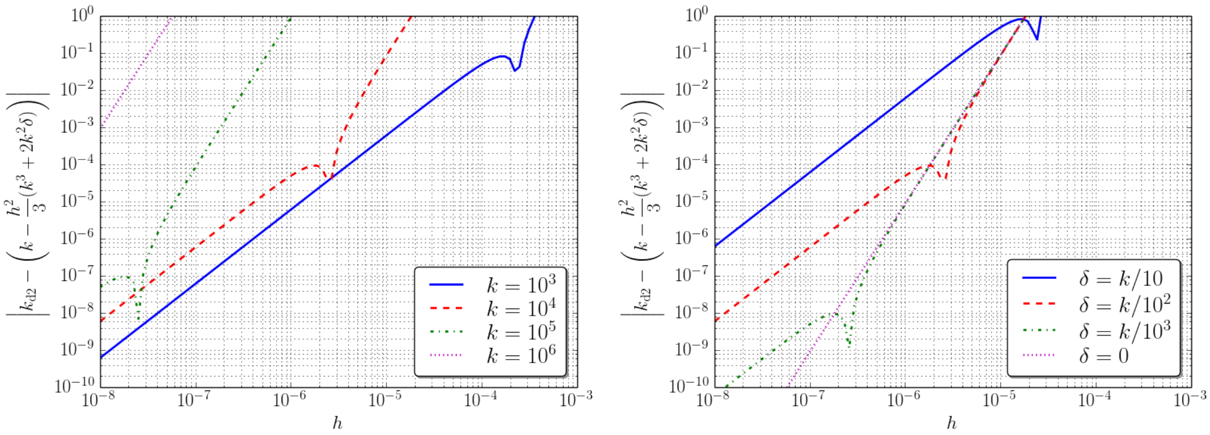

Figure 1.B.3: Difference between the discrete wave number $k_{\mathrm{d}2}$ and its asymptotic expression given by (1.103), plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^2$ (left) or the wave number $k = 10^4$ (right).

different types of Bloch waves can be deduced for the PUFEM discretization:

$$U_{h0}(x) = 2i\alpha \sum_{m\in\mathbb{Z}} \varphi_m(x) \sin((k+\delta)(x - x_m) + k_{\mathrm{d}0} x_m), \tag{1.104}$$

$$U_{h1}(x) = 2i\alpha \sum_{m\in\mathbb{Z}} \varphi_m(x) \sin((k+\delta)(x - x_m) + k_{\mathrm{d}1} x_m), \tag{1.105}$$

$$U_{h2}(x) = 2\alpha \sum_{m\in\mathbb{Z}} \varphi_m(x) \cos((k+\delta)(x - x_m) + k_{\mathrm{d}2} x_m), \tag{1.106}$$

$$U_{h3}(x) = 2\alpha \sum_{m\in\mathbb{Z}} \varphi_m(x) \cos((k+\delta)(x - x_m) + k_{\mathrm{d}3} x_m). \tag{1.107}$$

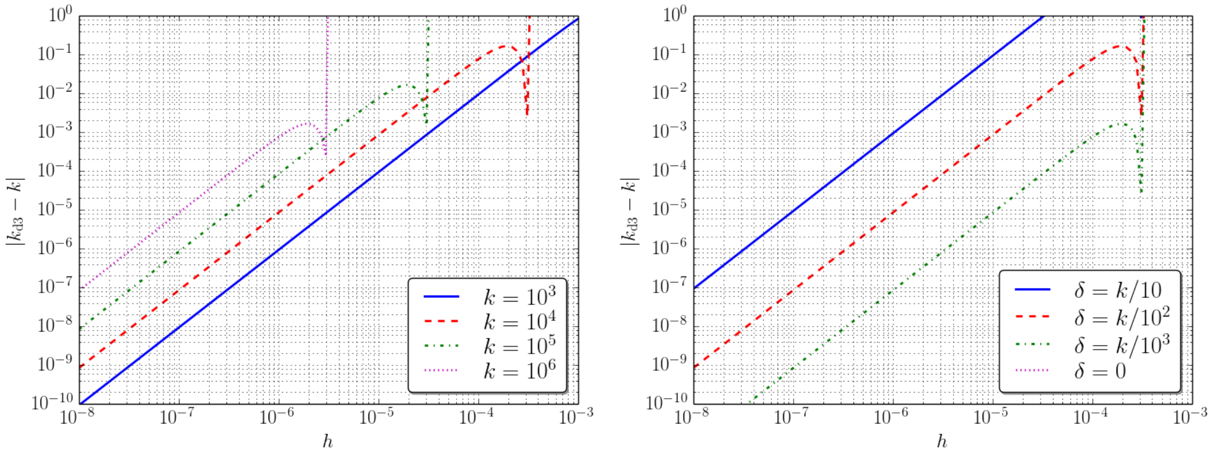

Figure 1.B.4: Difference between the discrete wave number k_{d3} and its asymptotic expression given by (1.103), plotted with respect to the mesh size but fixing the value of the perturbation parameter $\delta = 10^2$ (left) or the wave number $k = 10^4$ (right).

Figure 1.B.5 shows the plots of the Bloch waves (1.104)-(1.107). For comparison purposes, in the case of U_{h0} and U_{h1}, it has been fixed $\alpha = 1/2i$ and both Bloch waves are compared with the exact Helmholtz solution $\sin(kx)$. Analogously, in the case of U_{h2} and U_{h3}, they have been compared with $\cos(kx)$, taking into account $\alpha = 1/2$.

To stress the dispersion properties of the PUFEM discretization, it has been fixed $h = 10^{-5}$, $k = 10^5$, and $\delta = 5 \times 10^4$, what ensures $hk = 1$ and a perturbation error of 50%. It can be observed that only dispersion wave numbers k_{d1} and k_{d3} lead to dispersionless results for large values of k with a phase leakage due to the large value of δ. Otherwise, spurious oscillations are present on the Bloch waves U_{h0} and U_{h2}.

1.B.1 Null wave number perturbation

If the wave number perturbation parameter δ is assumed null, the dispersion equations are simplified, recovering the so-claimed dispersionless character of the PUFEM discretization, which can be observed straightforwardly on the expressions of the PUFEM Bloch waves. More precisely, if $\delta = 0$ then from (1.103) it is straightforward to check that the discrete wave numbers, k_{dj}, $0 \leq j \leq 3$ satisfy

$$|k_{d0}| \leq Ch^2k^3+\hat{C}h^4k^5, \qquad k_{d1} = k, \qquad |k_{d2} - k| \leq Ch^2k^3+\hat{C}h^4k^5, \qquad k_{d3} = k, \tag{1.108}$$

where C and $\hat{C}$ are positive constants independent of h, k, and δ. In fact, using the difference between the asymptotic expressions of k_{d0} and k_{d2} in (1.103), plots in Figure 1.B.6 illustrate numerically that the second terms in their estimates depend on $\mathcal{O}(h^4k^5)$, which can be identified as a pollution term.

Hence, taking into account (1.108), two of the discrete Bloch waves, U_{h1} and U_{h3}, are dispersionless since k_d coincides with k. Despite this is an unusual feature for classical

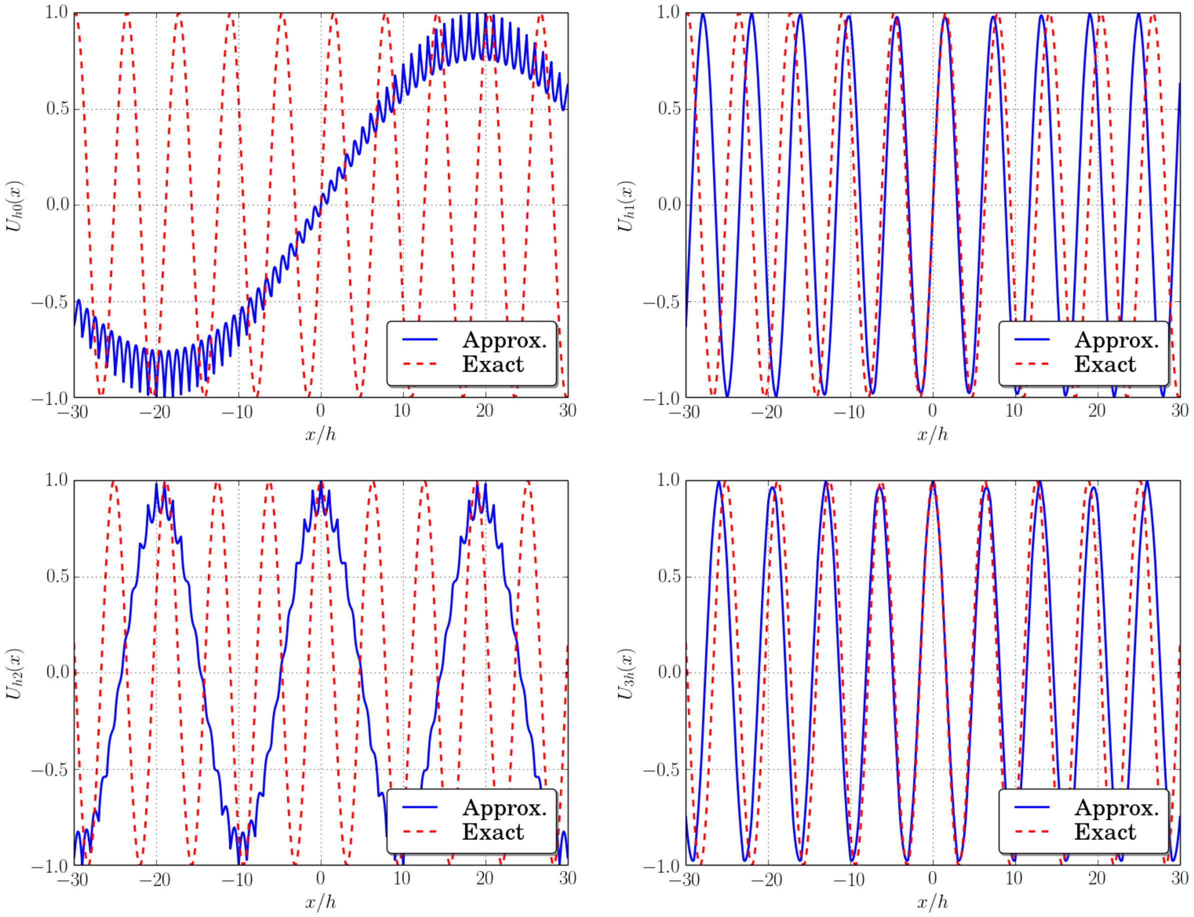

Figure 1.B.5: Bloch waves U_{hj} associated to each discrete wave number $k_{\mathrm{d}j}$ for $0 \leq 3$ (given by expressions (1.104)-(1.107)) for $h = 10^{-5}$, $k = 10^5$, $\delta = 5 \times 10^4$. Upper-left: U_{h0}, upper-right: U_{h1}, lower-left: U_{h2}, and lower-right: U_{h3}. The approximated waves U_{hj} are compared with respect to the exact waves $\cos(kx)$ for $j = 0, 1$ and $\sin(kx)$ for $j = 2, 3$.

polynomial-based finite element methods, the basis functions of the PUFEM discretization already include the right oscillatory dependency when $\delta = 0$. In consequence, using that $\{\varphi_m\}_{m\in\mathbb{Z}}$ is a partition of unity on the real line, in the case $\delta = 0$, (1.105) and (1.107) lead to $U_{h1}(x) = 2i\alpha \sin(kx)$ and $U_{h3}(x) = 2\alpha \cos(kx)$, which are exact solutions of the homogeneous Helmholtz equation. This result could be also derived from the fact that, since for $\delta = 0$, the plane waves solutions $e^{\pm kx}$ are included in the PUFEM discrete space V_h.

The other two Bloch waves, U_{h0} and U_{h2}, since they are based on the discrete wave numbers $k_{\mathrm{d}0}$ and $k_{\mathrm{d}2}$, which converges respectively to zero and k, they do not able to reproduce accurate approximations. In the case of U_{h0} is far from be a solution of the Helmholtz equation. On the contrary, U_{h2} converges to the exact solution $\sin(kx)$ but it suffer from a severe phase leakage when the pollution term $\mathcal{O}(h^4k^5)$ is not controlled (as it

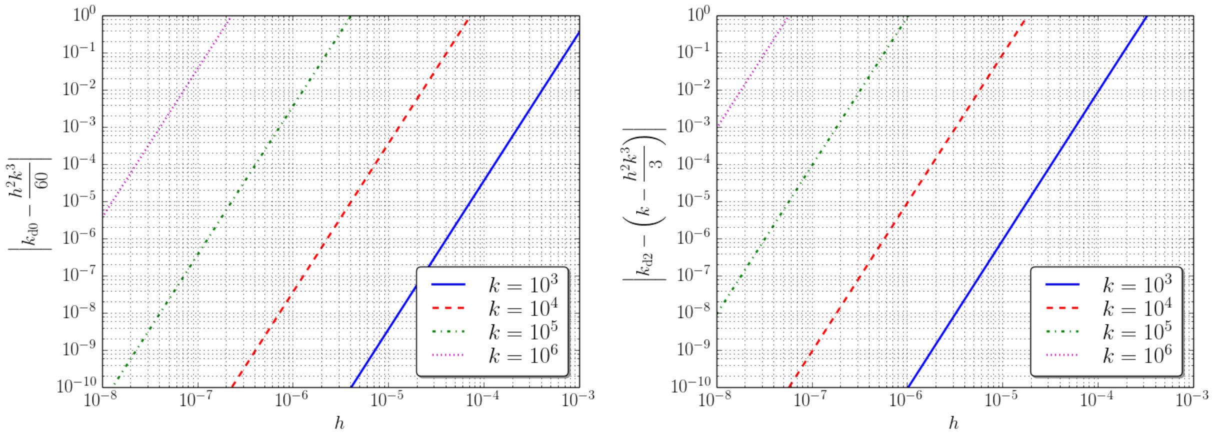

Figure 1.B.6: Difference between the discrete wave numbers $k_{\mathrm{d}0}$ (left) and $k_{\mathrm{d}2}$ (right) with respect its asymptotic expression given by (1.108), plotted with respect to the mesh size but fixing $\delta = 0$.

can be observed in Figure 1.B.7).

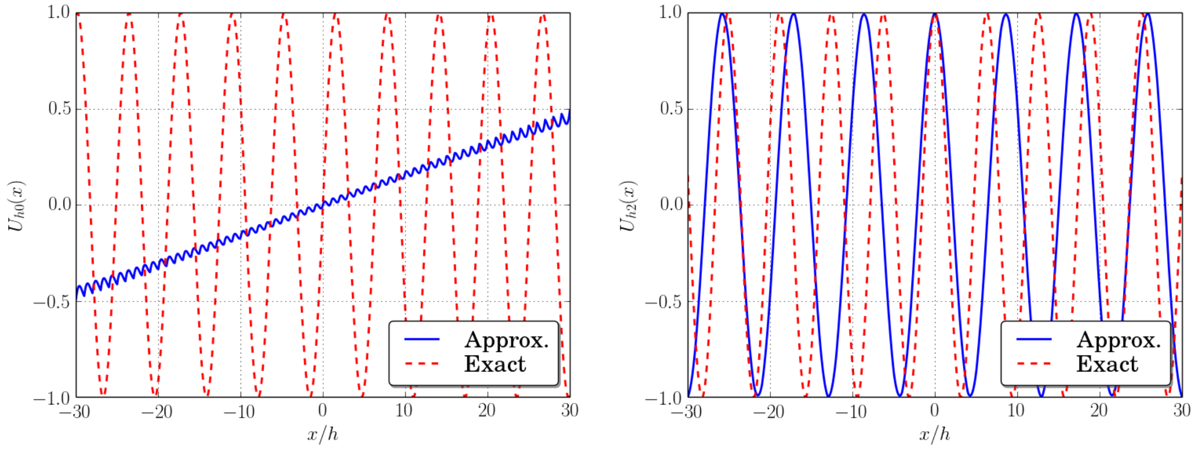

Figure 1.B.7: Bloch waves U_{h0} (left plot) and U_{h2} (right plot) associated to each discrete wave number $k_{\mathrm{d}0}$ and $k_{\mathrm{d}2}$ for $h = 10^{-5}$, $k = 10^5$, $\delta = 0$. The approximated waves U_{h0} is compared with respect to $\sin(kx)$ and U_{h2} is compared using $\cos(kx)$ as reference.

1.C Continuous Green's function

Following the ideas of described for the Laplacian problem in [46], the continuous Green function will be derived. A general procedure for the computation of Green's function with general boundary conditions and high-order differential equations has been studied in [9].

The case of interest of the present work involves a boundary value problem whose solution must satisfy the Helmholtz equation

$$\begin{array}{rcl} -u''(x) - k^2u(x) & = & f(x) \quad \text{in } (0,1), \\ u(0) & = & 0, \\ u'(1) - iku(1) & = & 0, \end{array}$$

where the source term satisfies $f \in L^2(0,1)$, and the wave number k is assumed positive. The Green's function of this boundary value problem is given by

$$G(x,s) = \frac{1}{k}\left\{ \begin{array}{ll} \sin(kx)e^{iks} & x \leq s, \\ \sin(ks)e^{ikx} & x \geq s. \end{array} \right. \tag{1.109}$$

In fact, following the ideas described in [46], this Green's function can be written in terms of $\alpha(x)$ and $\beta(x)$

$$G(x,s) = \left\{ \begin{array}{ll} \dfrac{\alpha(x)\beta(s)}{\beta(0)} & x \leq s, \\ \dfrac{\alpha(s)\beta(x)}{\beta(0)} & x \geq s. \end{array} \right. \tag{1.110}$$

The functions $\alpha(x)$ and $\beta(x)$ are respectively solutions of the Cauchy problems

$$\begin{array}{rcl} \alpha''(x) + k^2\alpha(x) & = & 0 \quad \text{in } (0,1), \\ \alpha(0) & = & 0, \\ \alpha'(0) & = & 1, \end{array}$$

and

$$\begin{array}{rcl} \beta''(x) + k^2\beta(x) & = & 0 \quad \text{in } (0,1), \\ \beta'(1) - ik\beta(1) & = & 0, \\ \beta(1) & = & 1. \end{array}$$

Direct computations show that $\alpha(x) = \sin(kx)/k$ and $\beta(x) = e^{-ik}e^{ikx}$. An analogy strategy will be also followed to compute the discrete Green's function in Section 1.5.3.

Chapter 2

A partition of unity finite element method for layered media

Contents

2.1 Introduction

The previous chapter dealt with one-dimensional Helmholtz problems in one media. This chapter will propose and describe some partition of unity finite element methods to approximate the solution of several Helmholtz problems: a one-dimensional problem in two media (with constant piecewise wave number), a two-dimensional problem in one media and finally, a two-dimensional Helmholtz problem in bi-layered media.

The modelling of acoustic wave propagation can be applied to several problems, like medical ultrasonics, seismic exploration or underwater acoustics. In particular, most of the physical environment of interest in underwater acoustics involve heterogeneous media. The spacial variability of these media depend on quantities such as the temperature, the salinity, the water depth or the presence of biological components. Several numerical methods can be used to approximate the solutions of that kind of problems. The goal of this chapter will be to propose and describe a partition of unity finite element method to approximate a two-dimensional Helmholtz problem in a bi-layered domain, having into account the reflection and the transmission occurred at the interface between media.

The first step (Section 2.2) will be to propose a PUFEM discretization to approximate a one-dimensional Helmholtz problem in two media. In subsection 2.2.1, the model problem is described. Several discretizations for this problem are described in subsection 2.2.2, and some numerical results to illustrate the discretization proposed can be observed in subsection 2.2.4.

The next step (Section 2.3) will be to study a two-dimensional Helmholtz problem in one media. The model problem and its variational formulation is introduced in 2.3.1. The PUFEM discretization proposed and some integration techniques are described in subsections 2.3.2 and 2.3.3 respectively. Finally for this section, some numerical results are presented in subsection 2.3.4.

After that two previous steps, Section 2.4 will focus on the novel PUFEM discretization of a two-dimensional Helmholtz problem in a bi-layered domain. The model problem is stated in subsection 2.4.2, and its variational formulation posed in subsection 2.4.1. The PUFEM discretization of this problem is proposed and described in subsection 2.4.3. Some integration techniques applied to the matrix system are explained in 2.4.4, and several numerical results, for two particular problems, can be observed in subsection 2.4.5. The conclusions for this chapter are exposed in Section 2.5.

2.2 PUFEM for layered media in one-dimensional problems

The Helmholtz equation with variable wave number is required for the resolution of some acoustic propagation problems stated in heterogeneous layered (or stratified) media [6], where the sound speed of each layer can be constant but different between any pair of layers.

To illustrate the variety of PUFEM approaches which could be followed to introduce the variable (but piecewise constant) profile of the sound speed, it will be considered an one-dimensional Helmholtz problem analogous to those one studied in Chapter 1, but in this case k must be read as a piecewise constant wave number $k(x)$. Even in this more general setting, the existence and uniqueness of solution is guaranteed by using a *inf-sup* condition.

Despite of the variety of alternative procedures to obtain a PUFEM method which could be a potential pollution-free discretization, it will be checked that only that one based on a transmission-reflection planewave enrichment leads to accurate results. Finally, once the PUFEM strategy has been selected, a brief overview about the numerical results obtained with this PUFEM method shows how its relative error depends on the mesh size and on the wave number.

2.2.1 Model problem

Firstly, for the sake of clarity in the exposition of the present chapter (and despite most of its features are common with those one presented in Chapter 1), the model problem for layered media is introduced in detail. The time-harmonic wave propagation in isotropic homogeneous compressible media is modelled linearly by means of the Helmholtz equation. Throughout this work, a one-dimensional model will be considered. Without loss of generality it will be assumed the interval $(0,1)$ as computational domain (otherwise, a change of scale could be performed to transform the domain to the unit interval). The following boundary-value problem will be considered

$$\left\{\begin{array}{rcl} -u'' - k^2 u & = & f \quad \text{in } (0,1), \\ u(0) & = & u_0, \\ u'(1) - ik(1)u(1) & = & u_1, \end{array}\right. \tag{2.1}$$

where u and f are complex-valued functions. The source term f is assumed independent of k. The boundary data $u_0, u_1 \in \mathbb{C}$ and the wave number k is a strictly positive piecewise constant function, lower bounded far from zero. From an acoustic point of view, u could be understood as the complex-valued time-harmonic amplitude of the pressure field in a compressible fluid in a layered media with constant sound speed in each layer and driven at a fixed frequency. Since at $x = 0$, a Dirichlet boundary condition is assumed and a complex-valued Robin condition is imposed at $x = 1$, it is straightforward to check that the model problem has a unique solution. The proof is based on the classical *inf-sup* condition (see Section 1.2). In what follows, the variational formulation and the result of existence and uniqueness of solution will be recalled.

In the model problem (1.1), the Dirichlet and the Robin data u_0 and u_1 can be lift by a smooth function and then it can be used to translate the solution u. In this manner, the boundary data u_0 and u_1 can be considered null without loss of generality. Hence, to write the variational formulation, the solution will be sought in the space

$$\mathrm{V} = \left\{ v \in \mathrm{H}^1(0,1) : \ v(0) = 0 \right\} = \mathrm{H}^1_{(0}(0,1),$$

and the variational formulation of problem (1.1) is written as follows:

$$\begin{cases} \text{Given } u_1 \in \mathbb{C} \text{ and } f \in \mathrm{L}^2(0,1), \text{ find } u \in \mathrm{V} \text{ such that} \\ B_k(u,v) - ik(1)u(1)\bar{v}(1) = \displaystyle\int_0^1 f(x)\bar{v}(x)\,\mathrm{d}x \qquad \forall v \in \mathrm{V}, \end{cases} \tag{2.2}$$

where the sesquilinear form $B_k : \mathrm{V} \times \mathrm{V} \to \mathbb{C}$ is defined by

$$B_k(u,v) = \int_0^1 \big(u'(x)\bar{v}'(x) - k(x)^2 u(x)\bar{v}(x)\big)\,\mathrm{d}x, \qquad u,v \in \mathrm{V}. \tag{2.3}$$

The *inf-sup* condition of the sesquilinear form $(u,v) \mapsto B_k(u,v) - ik(1)u(1)\bar{v}(1)$ can be obtained explicitly in terms of the wave number k (see Section 1.2 for a detailed discussion).

2.2.2 Discretization

As in the constant case, any PUFEM discretization, and in particular, that one which will be applied to the layered Helmholtz equation, is based on the partition of unity and the set of the problem-related functions, which are selected as close related to the exact solution of the problem to be solved.

In the same manner as it has been introduced in Section 1.3, to define the partition of unity, an equispaced mesh $\mathcal{T}_h = \{x_j = hj : j = 0,\ldots,n\} \subset [0,1]$ of $n+1$ nodes with mesh size $h = 1/n$ is considered. On this mesh, a standard Lagrange $\mathbb{P}_1$ (piecewise linear) finite element basis $\{\varphi\}_{j=0}^n$ will be used as the elements of the partition of unity. The second key component in the PUFEM discretization are the problem-related functions. As it has been devised by other authors [35, 41] for the Helmholtz equation with a constant wave number, planewave solutions of the homogeneous Helmholtz equation can be used for this purpose, this is, functions of type $e^{\pm ik(x-x_j)}$. However, if the wave number k is piecewise constant, replacing the constant wave number by a variable profile in the exponential functions written above do not provide solutions of the differential equation in (2.1) globally stated in $(0,1)$.

To overcome this difficulty, there exist different variations of the PUFEM method for a Helmholtz problem with constant wave number (see Section 1.3 for further details). In what follows, four different strategies are going to be described, all of them potential candidates for being a free-pollution numerical method. All these strategies lead to different discrete spaces but has a common feature: the hat-functions, which are the canonical basis of the standard piecewise linear finite element space, are multiplied by exponential-type expressions of the form $e^{\pm i\lambda(x)(x-x_j)}$ (or linear combinations of these exponential-type functions), being the computation of the function λ which will make the difference between each method and another. In any case, the PUFEM discrete space will be denoted by $\mathrm{X}_h = \langle \{\psi_j^-\}_{j=0}^n \cup \{\psi_j^+\}_{j=0}^n \rangle$ where

$$\psi_j^-(x) = \varphi_j(x) e^{-i\lambda(x)(x-x_j)}, \quad \psi_j^+(x) = \varphi_j(x) e^{+i\lambda(x)(x-x_j)} \qquad \text{for } j = 0,\ldots,n.$$

The restriction of functions of X_h to those functions which satisfy the homogeneous Dirichlet boundary condition defines the discrete trial and test space of the PUFEM discretization:

$$\mathrm{V}_h = \{v \in \mathrm{X}_h\,;\ v(0) = 0\} = \langle \psi_0^- - \psi_0^+, \{\psi_j^-\}_{j=1}^n \cup \{\psi_j^+\}_{j=1}^n \rangle.$$

In what follows, the definition of each discrete basis will be described attending to the four different numerical strategies. In any case, it will be assumed that the mesh which defines the partition of unity is conformal with the points where the piecewise constant k is discontinuous, this is, it is assumed that the jump points $\{y_p\}_{p=0}^m$ of k coincide with some of the mesh vertices $\{x_j\}_{j=0}^n$, or equivalently, it is assumed that $\{y_p\}_{p=0}^m \subset \mathcal{T}_h$.

Global average method. An approximated constant wave number $k_{\rm gl}$ is chosen as the global average of the variable wave number k,

$$k_{\rm gl} = \int_0^1 k(x)\,\mathrm{d}x,$$

and hence, since $\lambda(x) = k_{\rm g}$ the discrete PUFEM basis is given by

$$\psi_j^+(x) = \varphi_j(x)e^{+ik_{\rm gl}(x-x_j)}, \qquad \psi_j^-(x) = \varphi_j(x)e^{-ik_{\rm gl}(x-x_j)}, \tag{2.4}$$

for $j = 0, \dots, n$. In consequence, with this choice of the PUFEM functions, the standard PUFEM discrete space, associated to the Helmholtz equation with constant wave number $k_{\rm gl}$, is being used to approximate the solutions of the Helmholtz problem (2.2) with variable (piecewise constant) wave number k.

Local element-wise method. In this approach, planewaves are written replacing formally the typical constant wave number by the corresponding value given by the $k|_{T_j}$ (different in each element $T_j \in \mathcal{T}_h$). Hence, $\lambda(x) = k(x)$ and the discrete PUFEM basis is given by

$$\psi_j^+(x) = \varphi_j(x)e^{+ik(x)(x-x_j)}, \qquad \psi_j^-(x) = \varphi_j(x)e^{-ik(x)(x-x_j)}, \tag{2.5}$$

Obviously, the expressions $e^{\pm}ik(x)x$ are not solutions of the Helmholtz equation involved in (2.1) and stated in $(0, 1)$. However, in the interior of each element T_j, the exponential-type expressions are local planewave solutions of the Helmholtz equation of the constant wave number Helholtz problem in the interior of each layer.

Notice also that, despite of being k discontinuous, every discrete basis function belongs to $\mathrm{H}^1(0, 1)$. In fact, it is clear that $\psi_j^\pm$ are continuous at any point of the mesh since either φ_j is null on the vertices or it is null the argument of the exponentials.

Local average method. This method is mainly based on the approach introduced by Ortiz in [40, 41] for two-dimensional problems. For each one of the finite element hat functions φ_j, a local average wave number is computed in the compact support of φ_j, this is, the local average of k is compute in $T_j \cup T_{j+1}$,

$$\begin{aligned}
k_0 &= \frac{1}{h}\int_{x_0}^{x_1} k(x)\,\mathrm{d}x, \\
k_j &= \frac{1}{2h}\int_{x_{j-1}}^{x_{j+1}} k(x)\,\mathrm{d}x, \qquad \text{for all } j = 1, \ldots, n-1, \\
k_n &= \frac{1}{h}\int_{x_{n-1}}^{x_n} k(x)\,\mathrm{d}x.
\end{aligned}$$

Consequently, λ is a multi-valued function, whose value depends on the finite element function which is modifying, this is, in the element T_j, $\lambda|_{T_j} = k_j$ if the exponential-type function is multiplying to φ_j but simultaneously $\lambda|_{T_j} = k_{j+1}$ if the exponential-type function is multiplying to φ_{j+1}. In conclusion, the PUFEM discrete basis is given by

$$\psi_j^+(x) = \varphi_j(x)e^{+ik_j(x-x_j)}, \qquad \psi_j^-(x) = \varphi_j(x)e^{-ik_j(x-x_j)}, \tag{2.6}$$

On the contrary, to the other strategies described above, the exponential-type expressions used with this approach are neither global nor local solutions of the Helmholtz equation with piecewise constant wave number.

Transmission-reflection method. The three approaches described above are essentially based on the use of a sort of planewaves with different sign (i.e., which can be read as signals travelling from and to $+\infty$). The main drawback in the three cases is shared among them: the exponential-type expressions used to define the PUFEM discrete space are not exact solutions of the homogeneous Helmholtz equation. So, it seems natural to replace these expressions by fully exact planewave solutions of the layered Helmholtz problem. With this aim, two planewaves with opposite direction of propagation (one which can be read as a signal coming from $+\infty$ and another one understood as a signal going to $+\infty$).

Following this strategy, two functions w_j^- and w_j^+ are defined in the support of each finite element basis function φ_j. In the first case, $w_j^- \in \mathrm{H}^2(x_{j-1}, x_{j+1})$ is the solution of the layered Helmholtz problem

$$\left\{\begin{array}{rcll}
-(w_j^-)'' - k|_{T_j}^2 w_j^- &=& 0 & \text{in } T_j = (x_{j-1}, x_j), \\
-(w_j^-)'' - k|_{T_{j+1}}^2 w_j^- &=& 0 & \text{in } T_{j+1} = (x_j, x_{j+1}), \\
w_j^-|_{T_j}(x_j) &=& w_j^-|_{T_{j+1}}(x_j), & \\
(w_j^-)'|_{T_j}(x_j) &=& (w_j^-)'|_{T_{j+1}}(x_j), & \\
(w_j^-)'(x_{j-1}) - ik|_{T_j} w_j^-(x_{j-1}) &=& -2ike^{ik|_{T_j}h}, & \\
(w_j^-)'(x_{j+1}) + ik|_{T_{j+1}} w_j^-(x_{j+1}) &=& 0, &
\end{array}\right.$$

which is given by

$$w_j^-(x) = \begin{cases} \dfrac{2k|_{T_j}}{k|_{T_j} + k|_{T_{j+1}}} e^{-ik|_{T_j}(x-x_j)} & \text{in } (x_{j-1}, x_j], \\ \dfrac{k|_{T_{j+1}} - k|_{T_j}}{k|_{T_j} + k|_{T_{j+1}}} e^{ik|_{T_{j+1}}(x-x_j)} + e^{-ik|_{T_{j+1}}(x-x_j)} & \text{in } (x_j, x_{j+1}), \end{cases}$$

Notice that the last two boundary conditions are designed as radiation Sommerfeld-like conditions, which ensure that a planewave proportional to $e^{-ik|_{T_j}x}$ is impinging with unity amplitude the left endpoint and an exact radiation condition at $x = x_{j+1}$. The two coupling conditions at $x = x_j$ ensures that w_j^- is globally a strong solution of the Helmholtz problem with piecewise constant wave number. Clearly, if $k|_{T_j} = k|_{T_{j+1}}$ then $w_j^- = e^{-ik|_{T_j}(x-x_j)}$ in $T_j \cup T_{j+1}$. It should be also remarked that the quotients in the expression of w_j^- can be understood as the reflection and the transmission coefficients of the planewaves solution of the Helmholtz problem in each layer.

Analogously, $w_j^+ \in \mathrm{H}^2(x_{j-1}, x_{j+1})$ is the solution of the layered Helmholtz problem

$$\begin{cases} -(w_j^+)'' - k|_{T_j}^2 w_j^+ &= 0 \qquad \text{in } T_j = (x_{j-1}, x_j), \\ -(w_j^+)'' - k|_{T_{j+1}}^2 w_j^+ &= 0 \qquad \text{in } T_{j+1} = (x_j, x_{j+1}), \\ w_j^+|_{T_j}(x_j) &= w_j^+|_{T_{j+1}}(x_j), \\ (w_j^+)'|_{T_j}(x_j) &= (w_j^+)'|_{T_{j+1}}(x_j), \\ (w_j^+)'(x_{j+1}) + ik|_{T_j} w_j^+(x_{j+1}) &= 2ike^{ik|_{T_j}h}, \\ (w_j^+)'(x_{j-1}) - ik|_{T_{j-1}} w_j^+(x_{j-1}) &= 0, \end{cases}$$

which is given by

$$w_j^+(x) = \begin{cases} e^{ik|_{T_{j+1}}(x-x_j)} + \dfrac{k|_{T_{j+1}} - k|_{T_j}}{k|_{T_j} + k|_{T_{j+1}}} e^{-ik|_{T_{j+1}}(x-x_j)} & \text{in } (x_j, x_{j+1}), \\ \dfrac{2k|_{T_j}}{k|_{T_j} + k|_{T_{j+1}}} e^{ik|_{T_j}(x-x_j)} & \text{in } (x_{j-1}, x_j], \end{cases}$$

Again the last two boundary conditions are designed as radiation Sommerfeld-like conditions, which ensure that a planewave proportional to $e^{ik|_{T_j}x}$ is impinging with unity amplitude the right endpoint and an exact radiation condition at $x = x_{j-1}$. The two coupling conditions at $x = x_j$ ensures that w_j^+ is globally a strong solution of the Helmholtz problem with piecewise constant wave number. Once again, if $k|_{T_j} = k|_{T_{j+1}}$ then $w_j^+ = e^{ik|_{T_j}(x-x_j)}$ in $T_j \cup T_{j+1}$. It should be also remarked that the quotients in the expression of w_j^+ can be understood as the reflection and the transmission coefficients of the planewaves solution of the Helmholtz problem in each layer (and both of them coincides with the reflection and transmission coefficients of w_j^-). In conclusion, the PUFEM discrete space will be generated as the span of the basis function

$$\psi_j^+(x) = \varphi_j(x) w_j^+(x), \qquad \psi_j^-(x) = \varphi_j(x) w_j^-(x). \tag{2.7}$$

***Remark* 2.2.1.** *Despite of the different definition of the PUFEM discrete basis functions, the four approaches share also a common feature: if the wave number is constant, this is, if an unique layer is involved in the Helmholtz problem all the approaches described above recover the PUFEM discretization for the constant case analyzed in detail in Chapter 1.*

2.2.3 Matrix description

Once the PUFEM discrete space is defined, the discrete PUFEM approximation u_h is defined as the solution of the following linear problem:

$$\begin{cases} \text{Given } u_1 \in \mathbb{C} \text{ and } f \in \mathrm{L}^2(0,1), \text{ find } u_h \in \mathrm{V}_h \text{ such that} \\ B_k(u_h, v_h) - ik(1)u_h(1)\bar{v}_h(1) = \displaystyle\int_0^1 f(x)\bar{v}_h(x)\,\mathrm{d}x \qquad \forall v_h \in \mathrm{V}_h. \end{cases} \tag{2.8}$$

In the following sections, the numerical properties of this discrete problem will be analysed in terms of approximability, stability and dispersion.

Since a basis has been fixed for the PUFEM discrete space V_h, it is possible to write the linear problem (2.8) in matrix form. Since the homogeneous Dirichlet condition must be satisfied for any element of the basis, it is straightforward to check that the set $\{\psi_0^+ - \psi_0^-, \psi_1^-, \psi_1^+, \dots, \psi_n^-, \psi_n^+\}$ is a basis for V_h. Hence, any function v_h can be written as

$$v_h = v_0^+(\psi_0^+ - \psi_0^-) + \sum_{j=1}^{n} v_j^- \psi_j^- + \sum_{j=1}^{n} v_j^+ \psi_j^+,$$

where $(v_0^+, v_1^-, v_1^+, \dots, v_n^-, v_n^+)^t$ is the coordinates vector of v_h with respect to this basis. This coordinates can be considered as the degrees of freedom of the PUFEM discretization. Obviously, as in any Galerkin method applied to the Helmholtz equation, the linear variational problem (2.8) admits a matrix representation in terms of the stiffness and mass matrices: given $\vec{f_h} = (f_0^-, f_0^+, \dots, f_{n-1}^-, f_{n-1}^+, f_n^- + u_1, f_n^+ + u_1)^t$, find $\vec{u}_h = (u_0^-, u_0^+, \dots, u_n^-, u_n^+)^t$ such that

$$(-k(1)^2\mathcal{M}_h - ik(1)\mathcal{R}_h + \mathcal{K}_h)\vec{u}_h = \vec{f_h}, \tag{2.9}$$

under the restriction $u_0^- + u_0^+ = 0$, where the components of the stiffness and mass matrices are given by

$$[\mathcal{M}_h]_{\tilde{j}^\pm \tilde{l}^\pm} = \int_0^1 \frac{k^2(x)}{k^2(1)} \psi_j^\pm(x)\bar{\psi}_l^\pm(x)\,\mathrm{d}x, \qquad [\mathcal{K}_h]_{\tilde{j}^\pm \tilde{l}^\pm} = \int_0^1 (\psi_j^\pm)'(x)(\bar{\psi}_l^\pm)'(x)\,\mathrm{d}x, \tag{2.10}$$

for all $0 \le j, l \le n$ (with $\tilde{j}^\pm = 4(j + 3 \pm 1)/2$), and the matrix $\mathcal{R}_h$ associated to the Robin condition has all its coefficients null except $[\mathcal{R}_h]_{jl} = 1$ for all $j, l = 2n+1, 2(n+1)$.

2.2.4 Numerical results

To evaluate the accuracy of each approach to design the discrete PUFEM basis, a simple bi-layered material has been taking into account. This numerical example will be used to illustrate the different dispersion behaviour (numerical phase leakage) of each discrete basis choice for a Helmholtz problem with variable wave number. Finally, in the case of the transmission-reflection method, the H^1-relative error will be analysed in detail to determine its dependency with respect to the mesh size h, to the wave number magnitude, and the perturbation parameter δ (as it has been introduced in the analysis described in Chapter 1 for the constant wave number case).

The piecewise constant profile of the wave number in the bi-layered material has been fixed to $k(x) = k(1)/4$ for $0 < x \leq 0.5$ whereas $k(x) = k(1)$ for $0.5 < x < 1$, being $k(1) = 150$. The boundary data are chosen in problem (2.1) such that the exact solution is given by

$$u(x) = \begin{cases} e^{ik(1)x/4} + \dfrac{3}{5}e^{-ik(1)x/4} & \text{for } \ 0 < x \leq 0.5 \\ \dfrac{2}{5}e^{ik(1)x} & \text{for } \ 0.5 < x < 1. \end{cases}$$

Notice that despite it is not written explicitly in problem (2.1), the coupling conditions, which ensure the continuity of u and u' at the point $x = 0.5$ (where the wave number $k(x)$ has a jump discontinuity), have been used to compute the exact solution.

In Figure 2.1, the approximate PUFEM solution computed with each one of the discretization approaches are plotted and compared with respect to the exact solution. As it is mentioned previously $k = 150$ and a finite element mesh with $n = 30$ elements has been considered. These plots reveals clearly how inaccurate the global average procedure is (comparing it with respect the rest of schemes). Local approaches (both local average and local element-wise methods) reach similar numerical results. As it could be predicted by the definition of the discrete PUFEM basis, the exact solution is fully recovered (without error) by the transmission-reflection method since the exact solution belongs to the discrete PUFEM space.

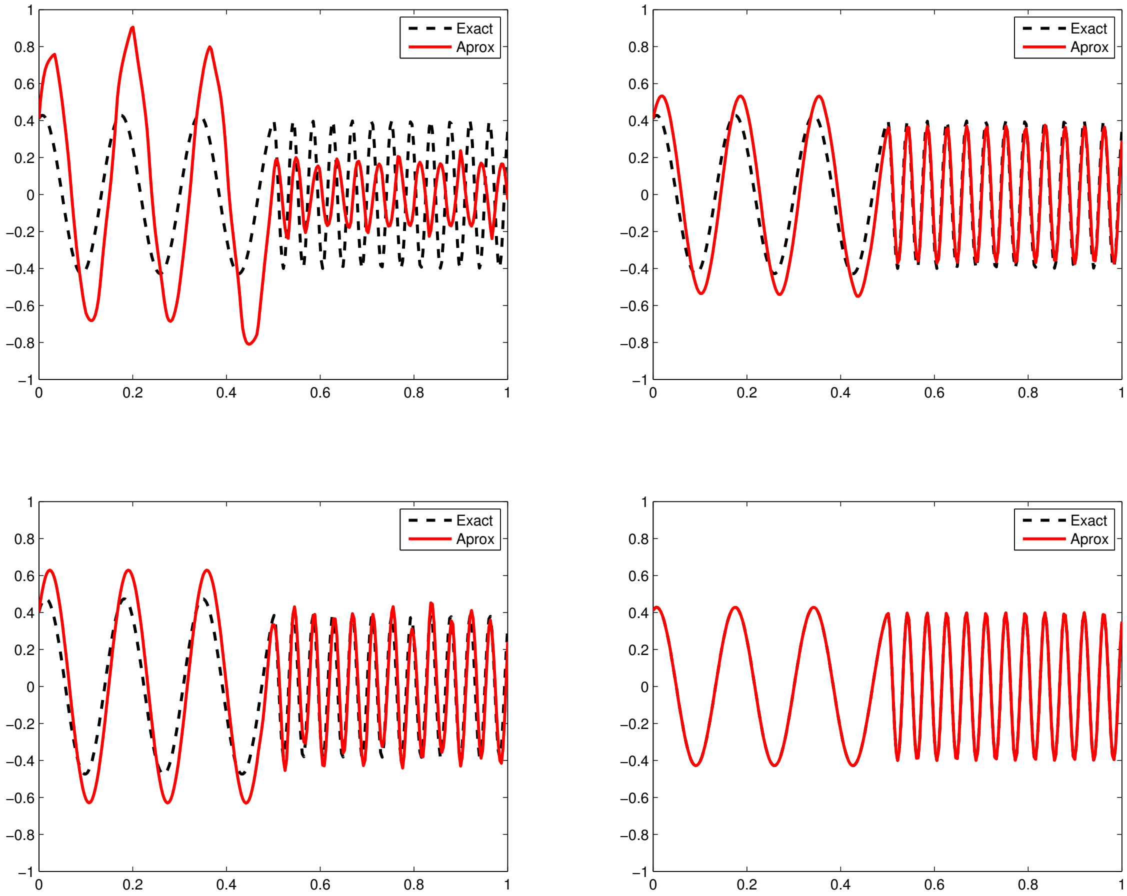

Figure 2.1: Numerical comparison of the PUFEM approximation (continuous red line) with respect to the exact solution (dashed black line) for $k = 150$ and $n = 30$ with different approaches: global average (upper-left), local element-wise average (upper-right), local average (lower-left) and transmission-reflection (lower-right).

To perform an analogous analysis of sensitivity with respect to errors introduced in the wave number values used to define the discrete PUFEM basis, a parameter $\delta \leq 0$ has been introduced in each element of the discrete basis. In this manner, in the definition of w_j^{pm}, the occurrences of k will be replaced by $k + \delta$. Obviously, if this parameter δ is strictly positive, the exact solution is not longer in the discrete PUFEM space. If $\delta = 10^{-2}$, the plots in Figure 2.2 show the dependence of the H^1-relative error for the transmission-reflection PUFEM method with respect to the mesh size and the wave number $k(1)$. The relative error is computed as the relative difference $|u_h - u|_1/|u|_1$. In the case of the mesh size dependency, the second order of accuracy is observed in the variable approximation, which is the same order of accuracy obtained by the PUFEM approximation of a Helmholtz problem with constant wave number. The independent behaviour of the PUFEM relative error with respect to the wave number is also observed in the right plot of Figure 2.2.

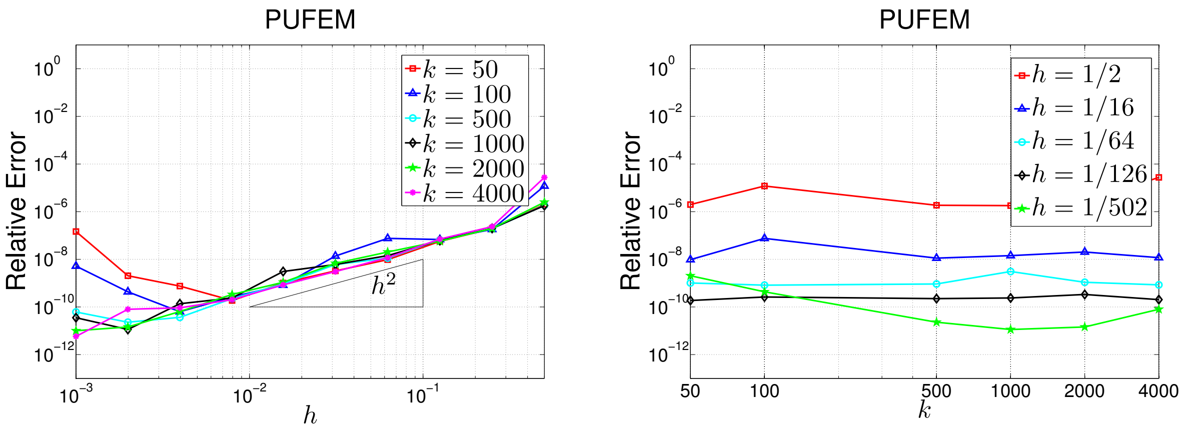

Figure 2.2: H^1-relative error of the approximation computed with the transmission-reflection PUFEM method (using the perturbation parameter $\delta = 10^{-2}$), plotted with respect to the mesh size h (left) and to the wave number $k(1)$ (right).

In conclusion, from the numerical examples considered above, it seems natural to extend the transmission-reflection PUFEM approach to a two-dimensional Helmholtz problem. The first step in this extension will be the description of the PUFEM method applied to the Helmholtz problem with constant wave number. Once the PUFEM has been described in this simpler framework, the transmission-reflection PUFEM will be described in two-dimensions for a problem with two layers.

2.3 PUFEM for two-dimensional problems with constant wave number

The previous step of introducing the PUFEM method in layered media consists in the full description of the standard PUFEM method applied to the Helmholtz equation with constant wave number (using planewaves to enrich the piecewise linear finite element space). Throughout the rest of this chapter, both for the constant and piecewise constant wave number the same model problem will be used: the Helmholtz equation stated in a bounded regular domain with Neumann boundary conditions.

In this chapter only Cartesian coordinates with respect to the canonical basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ will be used. Hence, an arbitrary two-dimensional point $\boldsymbol{x}$ will be identify with its Cartesian coordinates $\boldsymbol{x} = (x_1, x_2)$. Abusing on the notation, the vector position of a point $\boldsymbol{x}$ will be also denoted as $\boldsymbol{x}$. In the same manner, any vector $\vec{a} = a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2$ will be identified with its Cartesian coordinates column vector, this is, $\vec{a} = (a_1, a_2)^t$.

2.3.1 Model problem

Let Ω be an open regular domain and $\partial\Omega$ its boundary. If $u : \Omega \to \mathbb{C}$ is the unknown acoustic field and $k > 0$ is the constant wave number, the two-dimensional Helmholtz problem considered can be read as follows:

$$-\Delta u - k^2 u = 0 \qquad \text{in } \Omega, \tag{2.11}$$

$$\frac{\partial u}{\partial \boldsymbol{n}} = g \qquad \text{on } \partial\Omega, \tag{2.12}$$

being $\boldsymbol{n}$ the outward unit normal vector to $\partial\Omega$ and g the Neumann data on the boundary.

To write the variational formulation of the strong problem (2.11)-(2.12), let $v : \Omega \to \mathbb{C}$ be a test function regular enough. Multiplying (2.11) by the complex-conjugate of a test function v and integrating in the domain Ω, it can be written

$$-\int_\Omega \Delta u \, \bar{v} \, \mathrm{d}\boldsymbol{x} - k^2 \int_\Omega u \, \bar{v} \, \mathrm{d}\boldsymbol{x} = 0. \tag{2.13}$$

Using now an standard Green's formula applied to (2.13) and taking into account the Neumann boundary condition, the weak formulation of the model problem consists in: given $g \in \mathrm{L}^2(\partial\Omega)$, find $u \in \mathrm{H}^1(\Omega)$ such as

$$\int_\Omega \nabla u \cdot \nabla \bar{v} \, \mathrm{d}\boldsymbol{x} - k^2 \int_\Omega u \, \bar{v} \, \mathrm{d}\boldsymbol{x} = \int_{\partial\Omega} g \, \bar{v} \, \mathrm{d}\sigma, \tag{2.14}$$

for all $v \in \mathrm{H}^1(\Omega)$. Classical arguments based on the Fredholm's alternative theory [7] and the fact that the resolvent operator associated to this problem is a self-adjoint compact operator show that the variational problem (2.14) has an unique solution except for an infinite sequence of real wave number values $\{k_j\}_{j=0}^{\infty}$, which should be understood as the resonances of the mechanical system associated to this model problem. Throughout the rest of this section, all the wave number values will be selected such are not coincident with any of the resonance values where the solution is not unique.

2.3.2 Constant wave number PUFEM discretization

To avoid any error coming for the triangular mesh, it will be assumed that the domain Ω is a two-dimensional polygon. In this manner, if $\mathcal{T}_h$ is a regular triangulation of the domain, where the mesh size h is defined as the maximum diameter in the triangulation, it holds

$$\Omega = \bigcup_{T \in \mathcal{T}_h} T, \qquad h = \max_{T \in \mathcal{T}_h} d_T,$$

being d_T the diameter of the triangle T (the diameter of the circle circumscribed in the triangle T [18]). Each node of the triangulation is denoted by $\boldsymbol{x}_j$, for all $n = 1, \ldots, N_{\text{fe}}$, being N_{fe} the total number of nodes. Let $\{\varphi_n\}_{n=1}^{N_{\text{fe}}}$ denote the standard Lagrange $\mathbb{P}_1$ two-dimensional finite element basis, where $\varphi_n(\boldsymbol{x}_m) = \delta_{mn}$, being δ_{mn} the Kronecker's delta.

The discrete PUFEM space X_h is defined by multiplying each finite element basis function by a certain number of planewave functions, whose directions of the wave number vector are evenly distributed in the plane. Let N_{pw} be the number of planewave functions considered, and let $\theta_j = 2\pi(j-1)/N_{\rm pw}$, for all $j = 1, \dots, N_{\rm pw}$, the angles which defines the direction of propagation of the planewave functions. Then, the basis functions $\psi_{n,j}$ of the discrete PUFEM space can be written

$$\psi_{nj}(\boldsymbol{x}) = \varphi_n(\boldsymbol{x})e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)} = \varphi_n(\boldsymbol{x})e^{i\vec{k}_j\cdot\boldsymbol{x}}, \quad 1 \le n \le N_{\rm fe},\ 1 \le j \le N_{\rm pw}, \tag{2.15}$$

where $\vec{k}_j = k(\cos\theta_j, \sin\theta_j)^t$ for $j = 1, \dots, N_{\rm pw}$. Hence, the discrete PUFEM space is given by $\mathrm{X}_h = \langle\{\psi_{n1}\}_{n=1}^{N_{\rm fe}} \cup \ldots \cup \{\psi_{nN_{\rm pw}}\}_{n=1}^{N_{\rm fe}}\rangle$. To highlight the definition of the discrete PUFEM space, Figures 2.1, 2.2 and 2.3 show, if the number of plane waves chosen is eight ($N_{\rm pw} = 8$), the direction of propagation of the first, second and third of these plane waves (i.e., the unit vectors which have the same direction and orientation as the wave number vectors $\vec{k}_j$). The blue straight lines in the left plots of these figures mark the direction of propagation of the eight planewaves used in the definition of the discrete PUFEM space. The red arrow in the left plots and the black arrow in the right plots mark the direction and orientation of the wave number vector for each planewave.

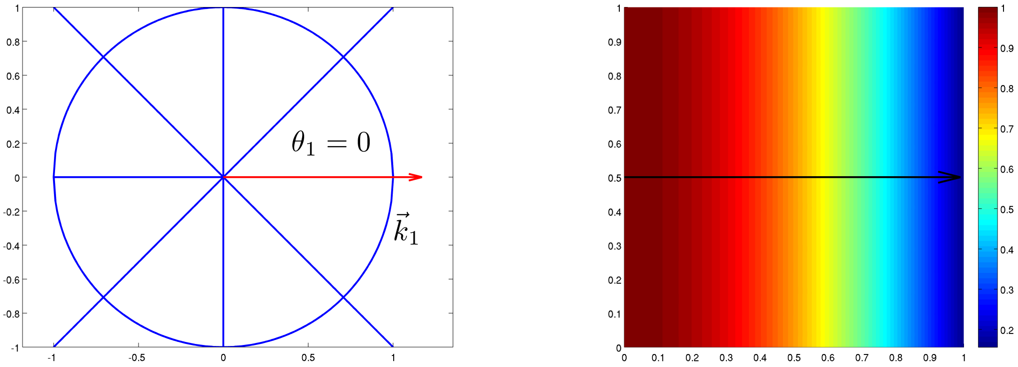

Figure 2.1: Real part of the first planewave $e^{i\vec{k}_1\cdot\boldsymbol{x}}$ used in a PUFEM discretization with $N_{\rm pw} = 8$ (right plot), being the direction of propagation $\vec{k}_1 = k(\cos\theta_1, \sin\theta_1) = k(1,0)$ (left plot).

Taking into account the definition of the discrete PUFEM space, given by the span of basis functions (2.15), the discrete PUFEM problem is described as follows: fixed $k > 0$ and given $g \in \mathrm{L}^2(\partial\Omega)$, find $u_h \in \mathrm{X}_h$ such that

$$\int_\Omega \nabla u_h \cdot \nabla \bar{v}_h \,\mathrm{d}\boldsymbol{x} - k^2 \int_\Omega u_h\, \bar{v}_h \,\mathrm{d}\boldsymbol{x} = \int_{\partial\Omega} g\, \bar{v}_h \,\mathrm{d}\sigma, \tag{2.16}$$

for all $v_h \in \mathrm{X}_h$.

The discrete PUFEM solution u_h can be written in terms of the basis functions in X_h,

$$u_h(\boldsymbol{x}) = \sum_{n=1}^{N_{\rm fe}} \sum_{j=1}^{N_{\rm pw}} u_{nj}\psi_{nj}(\boldsymbol{x}) = \sum_{n=1}^{N_{\rm fe}} \sum_{j=1}^{N_{\rm pw}} u_{nj}\varphi_n(\boldsymbol{x})e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)}, \tag{2.17}$$

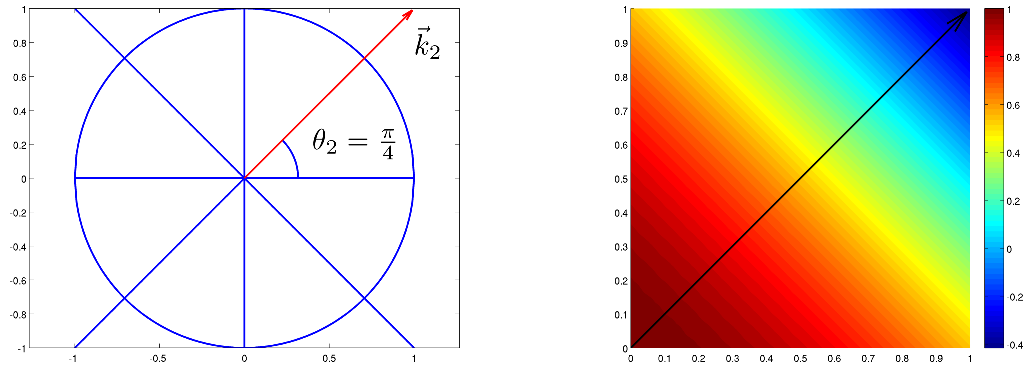

Figure 2.2: Real part of the second planewave $e^{i\vec{k}_2\cdot\boldsymbol{x}}$ used in a PUFEM discretization with $N_{\rm pw}=8$ (right plot), being the direction of propagation $\vec{k}_2 = k(\cos\theta_2, \sin\theta_2) = k(1,1)/\sqrt{2}$ (left plot).

where $\vec{u}_h(u_{11}, u_{21}, \dots, u_{N_{\rm fe}1}, \dots, u_{1N_{\rm pw}}, \dots, u_{N_{\rm fe}N_{\rm pw}})^t \in \mathbb{C}^{N_{\rm fe}N_{\rm pw}}$ is the complex vector of coefficients of the discrete PUFEM function $\boldsymbol{u}_h$. The discrete problem can be written in matrix form as

$$(\mathcal{K}_h - k^2\mathcal{M}_h)\vec{u}_h = \vec{g}_h, \tag{2.18}$$

where the mass matrix $\mathcal{M}_h$ and the stiffness matrix $\mathcal{K}_h$ are defined by

$$[\mathcal{K}_h]_{nj,ml} = \int_\Omega \nabla\left(\varphi_n(\boldsymbol{x})e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)}\right)\cdot\nabla\left(\varphi_m(\boldsymbol{x})e^{-ik(x_1\cos\theta_l + x_2\sin\theta_l)}\right)\,\mathrm{d}\boldsymbol{x}, \tag{2.19}$$

$$[\mathcal{M}_h]_{nj,ml} = \int_\Omega \varphi_n(\boldsymbol{x})e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)}\varphi_m(\boldsymbol{x})e^{-ik(x_1\cos\theta_l + x_2\sin\theta_l)}\,\mathrm{d}\boldsymbol{x}, \tag{2.20}$$

for all $1 \le n, m \le N_{\rm fe}$ and $1 \le j, l \le N_{\rm pw}$ (it should be remarked that the ordering of the matrix coefficients is given by the ordering induced by the coefficient order of the unknown vector $\vec{u}_h$). Analogously, each coefficient of the right-hand side vector $\vec{g}_h$ has the projection of the boundary data g onto the discrete PUFEM basis, this is,

$$[\vec{g}_h]_{nj} = \int_{\partial\Omega} g(\boldsymbol{x})\varphi_n(\boldsymbol{x})e^{-ik(x_1\cos\theta_j + x_2\sin\theta_j)}\,\mathrm{d}\sigma. \tag{2.21}$$

Since the integrals stated above are highly oscillatory if the wavelength of the planewaves $2\pi/k$ is much smaller than the typical size h of the support of the finite element functions, standard numerical quadrature rules (for instance, based on Gauss-Legendre with a reduced number of points) lead to inaccurate results. The following section describe in detail how these oscillatory integrals are computed in closed form.

2.3.3 Integration techniques

Firstly, the right hand side is computed locally on each edge s of the boundary $\partial\Omega$, and then assembled globally. To compute each coefficient of the right hand side $\vec{g}_h$ locally,

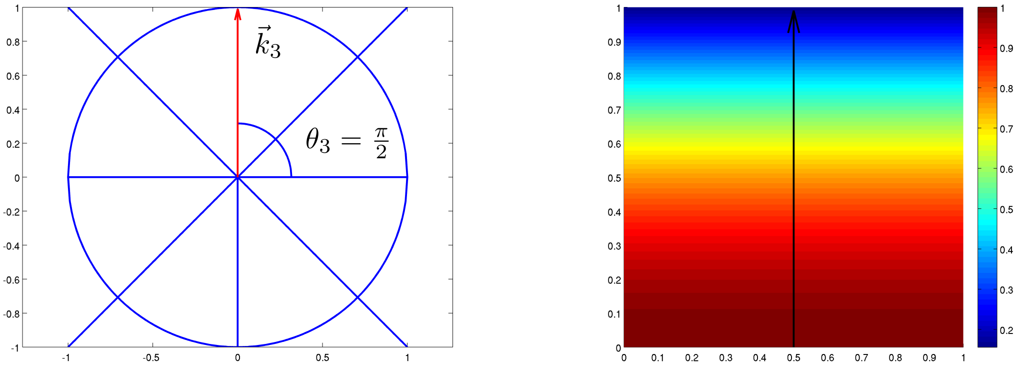

Figure 2.3: Real part of the third planewave $e^{i(\vec{k}_3\cdot\boldsymbol{x})}$ used in a PUFEM discretization with $N_{\text{pw}} = 8$ (right plot), being the direction of propagation $\vec{k}_3 = k(\cos\theta_3, \sin\theta_3) = k(0,1)$ (left plot).

exact integration in closed form is used in each edge s, performing a change of variable to the interval which allows to parametrize each edge s onto $[0, |s|]$, being $|s|$ the length of the edge. More precisely, if the edge has endpoints (a_1, a_2) and (b_1, b_2), the mapping $\xi \in [0, |s|] \mapsto (x_1(\xi), x_2(\xi)) \in s$, given by

$$\left\{ \begin{aligned} x_1(\xi) = &\quad (b_1 - a_1)\xi + a_1, \\ x_2(\xi) = &\quad (b_2 - a_2)\xi + a_2, \end{aligned} \right. \tag{2.22}$$

is a bijective function which parametrizes the edge. Notice that the Neumann function g can be integrated in closed form if its expression is known also in close form. In the present work, piecewise constant functions and exponential-type functions will be considered (see the numerical examples in Section 2.3.4). Taking in mind these kind of expressions for g, a simple integration by parts leads to the exact integration formulas for the contribution of the right-hand side $\vec{g}_h$.

In the same manner, the mass and the stiffness matrices are computed locally over each triangle of the mesh and then assembled globally. In each triangle, it must be integrated functions highly oscillatory functions with respect to x_1 and x_2. In order to perform these computations, some techniques can be found in the bibliography: numerical integration based on Gauss-Legendre two-dimensional formulas with a high number of quadrature nodes (see [38]), semi-analytical integration formulas (see [5]) or closed-form integration procedures, which reduces the two-dimensional integrals in triangles to the simplest computation of integrals stated on the edges (see [19]). In this work, the coefficients of the mass and stiffness matrices will be computed by using an exact integration method based in the rotation technique described by Ortiz in [41] and [40].

In what follows, the proposed procedure to obtain exact integration formulas is explained in detail, applied to the computation of a fixed coefficient of the mass matrix $\mathcal{M}_h$. Those computations to obtain the coefficients of stiffness matrix the calculations are completely analogous. First, the integration in the whole computational domain Ω is rewritten as the sum of the integrals over every triangle of the mesh, this is,

$$[\mathcal{M}_h]_{nj,ml} = \int_\Omega \varphi_n(\boldsymbol{x}) e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)} \varphi_m(\boldsymbol{x}) e^{-ik(x_1\cos\theta_l + x_2\sin\theta_l)}\,\mathrm{d}\boldsymbol{x} = \sum_{T\in\mathcal{T}_h} \int_T \varphi_n(\boldsymbol{x}) e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)} \varphi_m(\boldsymbol{x}) e^{-ik(x_1\cos\theta_l + x_2\sin\theta_l)}\,\mathrm{d}\boldsymbol{x}. \quad (2.23)$$

Clearly, this sum over $T \in \mathcal{T}_h$ is reduced to the addition of those integrals stated on the triangles with have the nodes $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$ as vertices.

To integrate over each triangle, a translation rotation will be used, in order to rewrite an integrand that oscillates respect to the Cartesian coordinates x_1 and x_2 as an integrand that oscillates respect to just one spatial variable. With this aim, consider a triangle $T \in \mathcal{T}_h$ with vertices $\boldsymbol{a} = (a_1, a_2)$, $\boldsymbol{b} = (b_1, b_2)$, and $\boldsymbol{c} = (c_1, c_2)$. Let the affine mapping be given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \quad (2.24)$$

If this change of variables is applied to the triangle T, a new triangle $\widetilde{T}$ is obtained. More precisely, T is rotated by an angle α and translated such that the first vertex $\boldsymbol{a}$ is mapped into the origin of coordinates. The angle α is chosen in such a way that, after the change of variable $(\xi, \eta) \in \widetilde{T} \mapsto (x_1, x_2) \in T$, the integrand only oscillates with respect to the new spatial coordinate ξ.

Applying this change of variable to (2.23), the integral over T is rewritten in an integral stated on $\widetilde{T}$, and so it is obtained

$$\int_T \varphi_n(\boldsymbol{x}) e^{ik(x_1\cos\theta_j + x_2\sin\theta_j)} \varphi_m(\boldsymbol{x}) e^{-ik(x_1\cos\theta_l + x_2\sin\theta_l)}\,\mathrm{d}\boldsymbol{x} = e^{ik\left(C_{jl}a_1 + D_{jl}a_2\right)} \int_{\widetilde{T}} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta) e^{ik\xi(C\cos\alpha - D\sin\alpha)} e^{ik\eta(C_{jl}\sin\alpha + D_{jl}\cos\alpha)}\,\mathrm{d}\xi\,\mathrm{d}\eta, \quad (2.25)$$

where $C_{jl} = \cos\theta_j - \cos\theta_l$ and $D_{jl} = \sin\theta_j - \sin\theta_l$. Notice that it has been used that the Jacobian matrix of the affine mapping (2.24) is a rotation and so its determinant is equal to one. As it has been mentioned previously, the angle α is then chosen such as $C_{jl}\sin\alpha + D_{jl}\cos\alpha = 0$, this is,

$$\alpha = \alpha_{jl} = \frac{\theta_j + \theta_l - \pi}{2}, \quad (2.26)$$

and so the integrand in (2.25) can be written now as a polynomial multiplied by a function that only oscillates with respect to ξ,

$$[\mathcal{M}_h]_{nj,ml} = e^{ik\left(C_{jl}a_1 + D_{jl}a_2\right)} \int_{\widetilde{T}} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta) e^{ik\xi(C_{jl}\cos\alpha_{jl} - D_{jl}\sin\alpha_{jl})}\,\mathrm{d}\xi\,\mathrm{d}\eta. \quad (2.27)$$

At this point, in order to integrate exactly the previous expression (2.27), six cases must be taken into account depending on the position of the vertices of the transformed triangle $\widetilde{T}$. It will be denoted by ξ_1, ξ_2 and ξ_3 the coordinates in the ξ-direction of the three vertices of $\widetilde{T}$, where the vertex are sorted in counter-clock wise order. In addition, using the same ordering, it will be assumed that the edges will be parametrized by the equations $\eta = \eta_1(\xi)$, $\eta = \eta_2(\xi)$, and $\eta = \eta_3(\xi)$. To compute the integral (2.27), six cases are distinguished (labelled as a)-f)):

Case a) If $\xi_1 = 0 < \xi_2 \leq \xi_3$ (see left plot in Figure 2.4),

$$\begin{aligned}[\mathcal{M}_h]_{nj,ml} &= e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_1}^{\xi_2} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_1(\xi)}^{\eta_3(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta \right) \mathrm{d}\xi \\ &+ e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_2}^{\xi_3} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_2(\xi)}^{\eta_3(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta \right) \mathrm{d}\xi. \quad (2.28)\end{aligned}$$

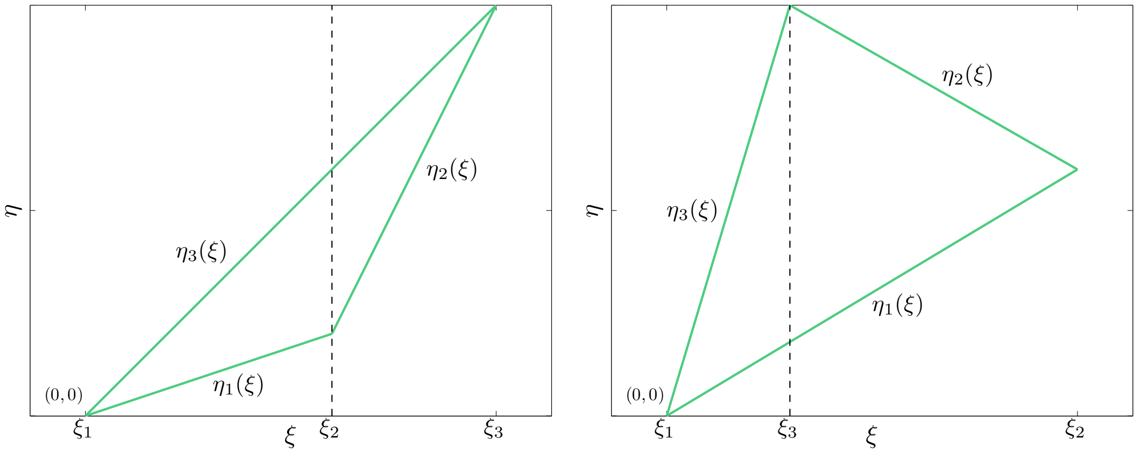

Figure 2.4: Splitting of the triangles for the exact integration: case a) if $\xi_1 = 0 < \xi_2 \leq \xi_3$ (left) and case b) if $\xi_1 = 0 \leq \xi_3 < \xi_2$ (right).

Case b) If $\xi_1 = 0 \leq \xi_3 < \xi_2$ (see right plot in Figure 2.4),

$$\begin{aligned}[\mathcal{M}_h]_{nj,ml} &= e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_1}^{\xi_3} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_1(\xi)}^{\eta_3(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta \right) \mathrm{d}\xi \\ &+ e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_3}^{\xi_2} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_1(\xi)}^{\eta_2(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta \right) \mathrm{d}\xi. \quad (2.29)\end{aligned}$$

Case c) If $\xi_3 \leq \xi_2 < \xi_1 = 0$ (see left plot in Figure 2.5),

$$[\mathcal{M}_h]_{nj,ml} = e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_3}^{\xi_2} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_3(\xi)}^{\eta_2(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta\right)\mathrm{d}\xi$$
$$+ e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_2}^{\xi_1} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_3(\xi)}^{\eta_1(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta\right)\mathrm{d}\xi. \quad (2.30)$$

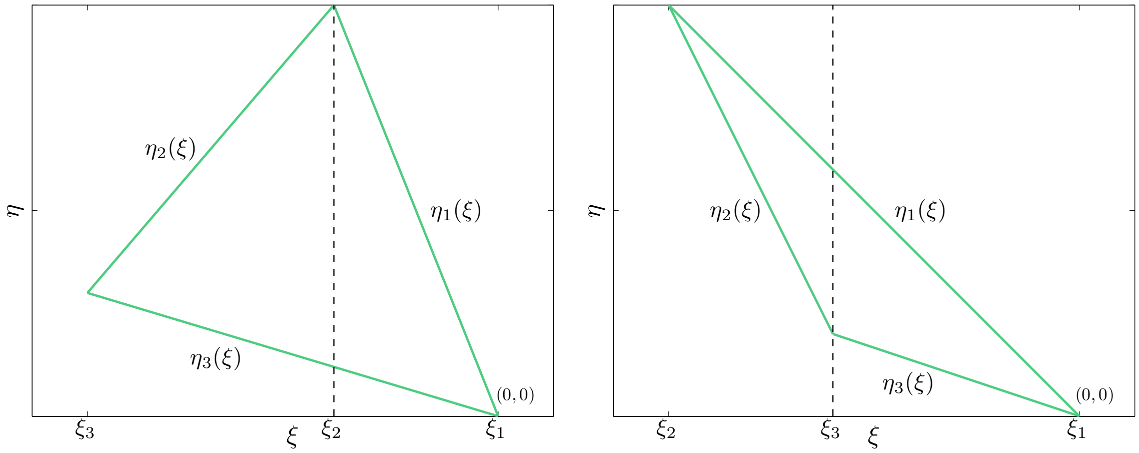

Figure 2.5: Splitting of the triangles for the exact integration: case c) if $\xi_3 \leq \xi_2 < \xi_1 = 0$ (left) and case d) if $\xi_2 < \xi_3 \leq \xi_1 = 0$ (right).

Case d) If $\xi_2 < \xi_3 \leq \xi_1 = 0$ (see right plot in Figure 2.5),

$$[\mathcal{M}_h]_{nj,ml} = e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_2}^{\xi_3} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_2(\xi)}^{\eta_1(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta\right)\mathrm{d}\xi$$
$$+ e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_3}^{\xi_1} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_3(\xi)}^{\eta_1(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta\right)\mathrm{d}\xi. \quad (2.31)$$

Case e) If $\xi_3 < \xi_1 = 0 \leq \xi_2$ (see left plot in Figure 2.6),

$$[\mathcal{M}_h]_{nj,ml} = e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_3}^{\xi_1} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_3(\xi)}^{\eta_2(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta\right)\mathrm{d}\xi$$
$$+ e^{ik\left(C_{jl}a_1+D_{jl}a_2\right)} \int_{\xi_1}^{\xi_2} e^{ik\xi(C_{jl}\cos\alpha_{jl}-D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_1(\xi)}^{\eta_2(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta\right)\mathrm{d}\xi. \quad (2.32)$$

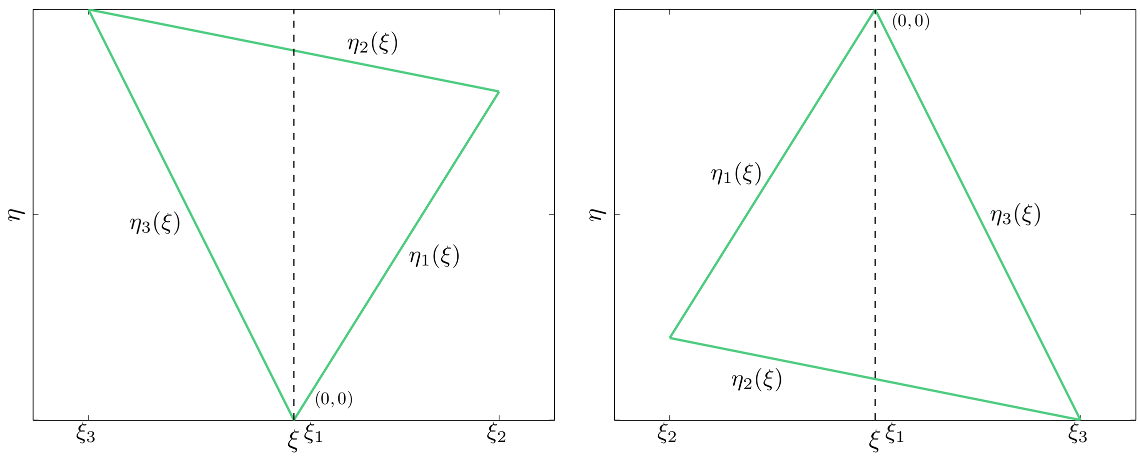

Figure 2.6: Division of the triangles for the exact integration: case e) if $\xi_3 < \xi_1 = 0 \leq \xi_2$ (left) and case f) if $\xi_2 \leq \xi_1 = 0 < \xi_3$ (right).

Case f) If $\xi_2 \leq \xi_1 = 0 < \xi_3$ (see right plot in Figure 2.6),

$$[\mathcal{M}_h]_{nj,ml} = e^{ik\left(C_{jl}a_1 + D_{jl}a_2\right)} \int_{\xi_2}^{\xi_1} e^{ik\xi(C_{jl}\cos\alpha_{jl} - D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_2(\xi)}^{\eta_1(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta \right) \mathrm{d}\xi$$
$$+ e^{ik\left(C_{jl}a_1 + D_{jl}a_2\right)} \int_{\xi_1}^{\xi_3} e^{ik\xi(C_{jl}\cos\alpha_{jl} - D_{jl}\sin\alpha_{jl})} \left(\int_{\eta_2(\xi)}^{\eta_3(\xi)} \varphi_n(\xi,\eta)\varphi_m(\xi,\eta)\mathrm{d}\eta \right) \mathrm{d}\xi. \quad (2.33)$$

Once the integral expressions only dependent on η have been computed, an integration by parts procedure is used to obtain the closed-form expressions for the integrals depending on ξ, which involves polynomial and exponential factors. However, some considerations must be taken into account to apply this strategy in any general case to compute (2.33)-(2.28). If the polynomial factor has degree p, the result of the integral depending on ξ is an expression involving some terms of order $\mathcal{O}\left((k(C_{jl}\cos\alpha_{jl} - D_{jl}\sin\alpha_{jl}))^{-r}\right)$, with $r = 1, \ldots, p+1$. Obviously, this terms could be potentially inaccurate evaluated due to round-off errors when $k(C_{jl}\cos\alpha_{jl} - D_{jl}\sin\alpha_{jl})$ is small enough or even worse, they are not well-defined in the case of $k = 0$ or $C_{jl}\cos\alpha_{jl} - D_{jl}\sin\alpha_{jl} = 0$ (which it holds when $j = l$). In these limit cases, the exponentials (with a small value in its coefficient) are approximated by a 5-th order Taylor expansion around the origin, and hence in these limits cases, the computation of the integrals (2.33)-(2.28) is reduced to a simple integration of polynomials (also computed in closed form).

2.3.4 Numerical results

To illustrate the accuracy of the PUFEM method based on planewaves applied to the Helmholtz problem with constant wave number, an extend variety of numerical results are presented. First, it will be considered a problem where the exact solution is given by a planewave in Section 2.3.4. Then, the behaviour of the relative error in L^∞-norm will be studied with respect to the mesh size h, the wave number k, and the number of planewaves N_{pw} used in the discretization. Finally, the analogous analysis will be performed by considering a problem with constant Neumann boundary conditions in Section 2.3.4.

Throughout this section about the numerical results of the PUFEM method, recall that the mesh chosen for the discretization has N_{fe} nodes $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_{\text{fe}}}\}$ and its elements have a maximum diameter h. The condition number of the matrix $\mathcal{K}_h - k^2\mathcal{M}_h$ involved in the linear system (2.18) is denoted by κ. The relative error field in L^∞-norm e_h and its norm will be computed on the mesh vertices as follows:

$$e_h(\boldsymbol{x}) = \frac{u(\boldsymbol{x}) - u_h(\boldsymbol{x})}{\max\limits_{j=1,\dots,N_{\text{fe}}} |u(\boldsymbol{x}_j)|}, \qquad \text{relative error} = \max_{j=1,\dots,N_{\text{fe}}} |e_h(\boldsymbol{x}_j)| \tag{2.34}$$

where u is the exact solution of the Helmholtz problem (2.11)-(2.12) and u_h is the discrete PUFEM approximation computed by means of (2.16).

Helmholtz problem with a planewave solution

In this first subsection, it is studied the behaviour of the PUFEM method applied to a Helmholtz problem (2.11)-(2.12) with a constant wave number, and whose exact solution is given by a unique planewave. The computational domain Ω is the unit square $(0,1)\times(0,1)$. The Neumann boundady data is settled by an exponential-type expression in such a manner that the exact solution is given by

$$u(\boldsymbol{x}) = e^{ik(x_1 \cos\beta + x_2 \sin\beta)}, \tag{2.35}$$

where β is the incident angle of the planewave measured with respect to the x_1-axis.

The first test consists in taking ten exact solutions with different incident angles β, some of them in the discrete space X_h, and study how the discrete PUFEM solution approximates them. The incident angles for the exact solutions are taken $\beta = 2\pi(j-1)/10$, for all $j = 1, \dots, 10$. In Figure 2.7, the relative error e_h for each one of these ten exact solutions is shown (left plot). In this numerical test, the number of planewaves used in the discretization is fixed to $N_{\text{pw}} = 5$. Different values of the wave number k have been considered for a mesh of size $h = 1.7 \times 10^{-1}$. Notice that for the values $\beta = j\pi/5$ with $j = 0, 2, 4, 6, 8$, the exact solution belongs to the discrete space X_h, so the relative error reaches the typical round-off errors of double precision floating-point arithmetic around $\mathcal{O}(10^{-15})$.

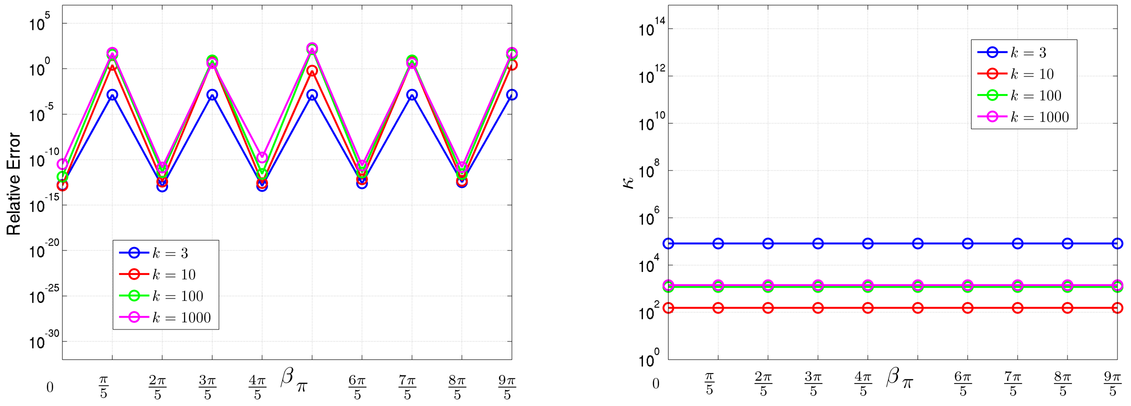

Figure 2.7: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k. The PUFEM discretization involves $N_{\text{pw}} = 5$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using double precision floating-point arithmetic.

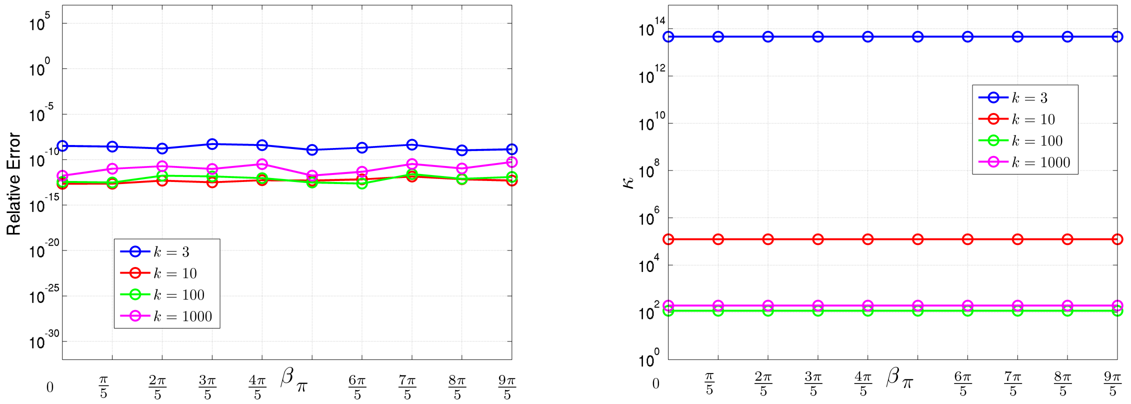

Figure 2.8: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k. The PUFEM discretization involves $N_{\text{pw}} = 10$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using double precision floating-point arithmetic.

If the number of planewaves is now taken $N_{\text{pw}} = 10$ and $N_{\text{pw}} = 20$, the ten exact solutions fall into the discrete space X_h. Figures 2.8 and 2.9 show the numerical results obtained by using double precision floating-point arithmetic. The relative error (left plots) for these ten solutions and for different values of k is higher than the order expected $\mathcal{O}(10^{-15})$ because of the high condition number (right plots). This fact can be explained observing the condition number κ of the PUFEM matrix system $\mathcal{K}_h - k^2\mathcal{M}_h$ (right plots). In fact, the accuracy of the approximated solutions computed with the PUFEM method based on planewaves is potentially pretty sensitive to the condition number. Notice that in these simulations performed with double precision floating-point arithmetic, in some cases the condition number has a magnitude larger than $\mathcal{O}(10^{14})$.

In order to check if the PUFEM method can overcome this potentially lack of accuracy, the simulations are repeated in quadruple precision floating point arithetic (this is, using 32 digits of precision). Figures 2.11 and 2.12 show the analogous numerical results to those ones of Figures 2.8 and 2.9 but now using the Matlab toolbox Advanpix [34] using quadruple precision. In these cases, the relative error (left plots) is close or smaller than $\mathcal{O}(10^{-15})$ even if the condition number is high (right plots). Notice that when the number of planewaves used in the discretization is increased, the condition number of the PUFEM matrix system increases too.

In Figure 2.10, the results for the numerical simulation with $N_{\text{pw}} = 5$ have been repeated for quadruple precision too, and it can be observed that the relative error (left plot) has a magnitude close to $\mathcal{O}(10^{-30})$ for those exact solutions which belong to X_h.

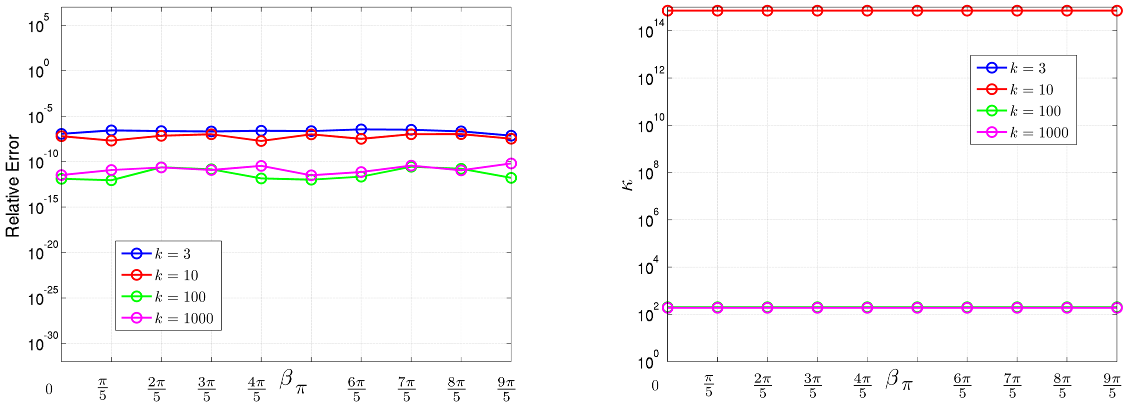

Figure 2.9: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k. The PUFEM discretization involves $N_{\text{pw}} = 20$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using double precision floating-point arithmetic.

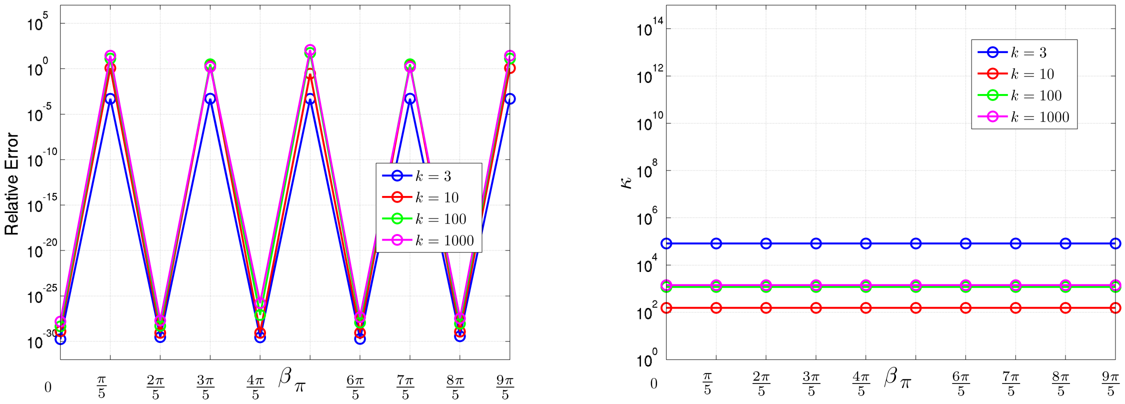

Figure 2.10: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k. The PUFEM discretization involves $N_{\text{pw}} = 5$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using quadruple precision floating-point arithmetic..

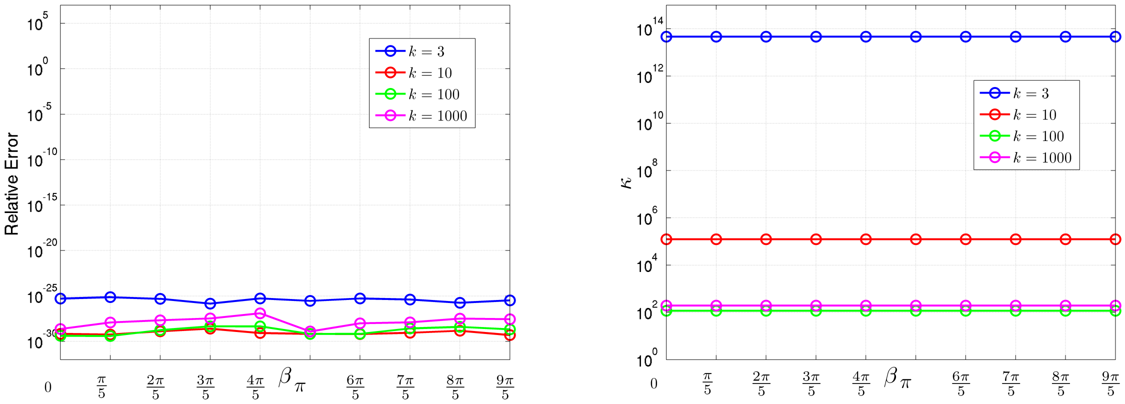

Figure 2.11: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k. The PUFEM discretization involves $N_{\text{pw}} = 10$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using quadruple precision floating-point arithmetic.

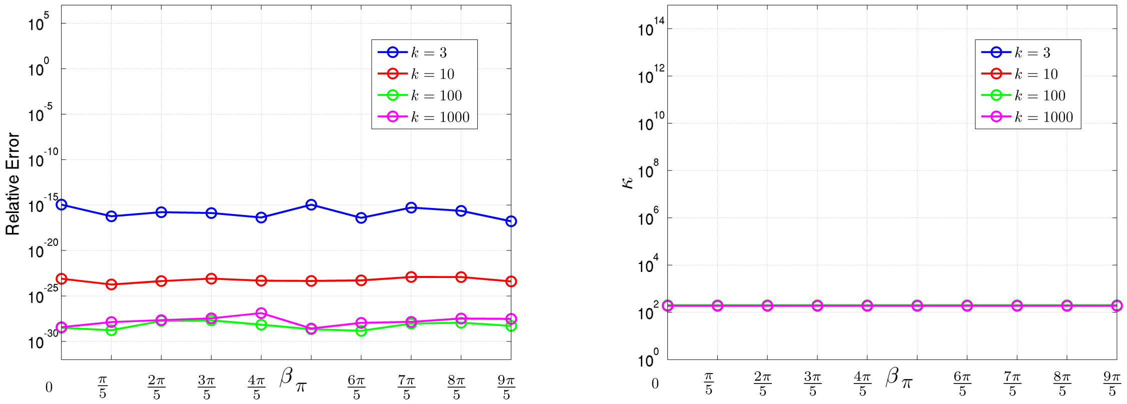

Figure 2.12: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k. The PUFEM discretization involves $N_{\text{pw}} = 20$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using quadruple precision floating-point arithmetic.

For the second test in this subsection, the exact planewave solution is fixed with an incident angle of $\beta = 2\pi/21$. Once again, the PUFEM discretization uses the same mesh with size $h = 1.7 \times 10^{-1}$ considered in the numerical tests described above. To analyze the behaviour of the relative error with respect to the number of planewaves used in the discretization, N_{pw} has been varied between 2 and 20, considering different values of the wave number k. Figures 2.13 (using double precision floating-point arithmetic) and 2.14 (using quadruple precision floating-point arithmetic) show that the relative error (left plots) decays exponentially when the number of planewaves used in the PUFEM discretization is increased. This convergence behaviour is pretty sensitive to the wave number values.

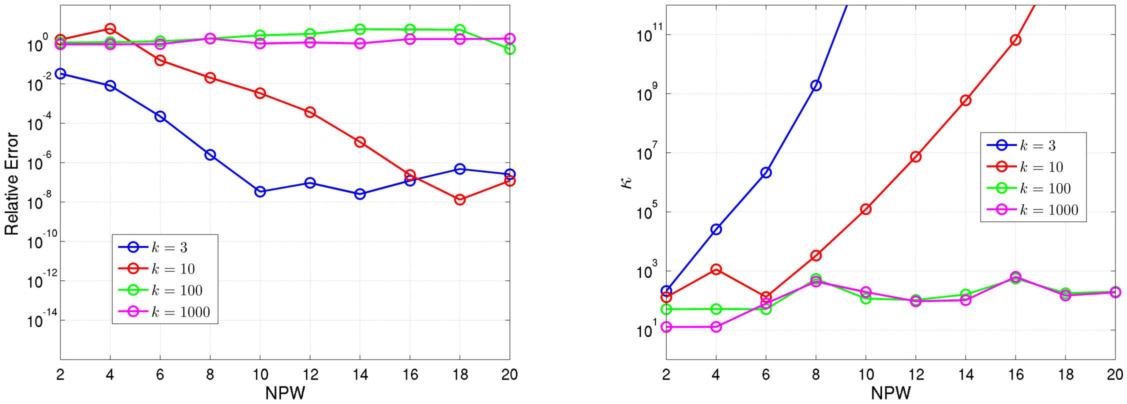

Figure 2.13: Relative error and condition number κ for an exact planewave solution with incident angle $\beta = 2\pi/21$, plotted with respect to the number of planewaves N_{pw} used in the discretization and the wave number k considered in the Helmholtz problem. The PUFEM discretization involves a triangular mesh with size $h = 1.7 \times 10^{-1}$ and an implementation using double precision floating-point arithmetic.

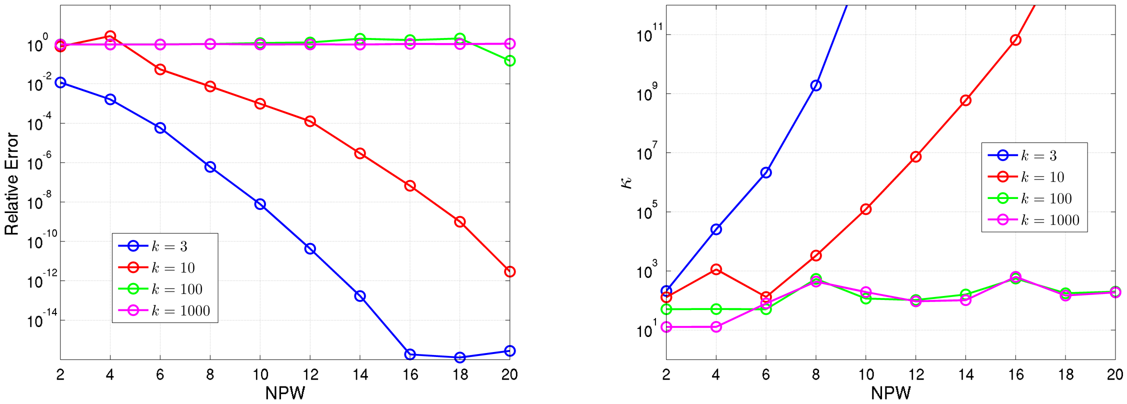

Figure 2.14: Relative error and condition number κ for an exact planewave solution with incident angle $\beta = 2\pi/21$, plotted with respect to the number of planewaves N_{pw} used in the discretization and the wave number k considered in the Helmholtz problem. The PUFEM discretization involves a triangular mesh with size $h = 1.7 \times 10^{-1}$ and an implementation using quadruple precision floating-point arithmetic.

Problem with constant Neumann boundary data

In this subsection, once again it will be illustrated the behaviour of the relative error with respect to the mesh size, the wave number and the number of planewaves used in the PUFEM discretization. However, in this case, the Helmholtz problem (2.11)-(2.12) (stated again in $\Omega = (0,1) \times (0,1)$) has been settled with constant Neumann boundary data and consequently the exact solution is not given by a unique planewave.

More precisely, the Neumann data g has been fixed throughout the rest of this subsection as follows:

$$g(x_1, x_2) = \begin{cases} 1 & \text{for } x_1 = 0 \text{ or } x_2 = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.36}$$

In consequence, straightforward computations show that the exact solution is given by

$$u(\boldsymbol{x}) = \frac{1}{ik\left(e^{2ik} - 1\right)} \left(e^{2ik}\left(e^{-ikx_1} + e^{-ikx_2}\right) + e^{ikx_1} + e^{ikx_2}\right). \tag{2.37}$$

This numerical example also is going to be utilized to highlight that the choice of the angle of incidence is pretty arbitrarily. For instance, in this case the planewave incident angles are not taken $\theta_j = 2\pi(j-1)/N_{\text{pw}}$ as it has been used previously. For this numerical test, it has been considered $\theta_j = 2\pi(j-1)/N_{\text{pw}} + 2\pi/2500$, for all $j = 1, \ldots, N_{\text{pw}}$. Notice that if the angles $\{\theta_j\}_{j=1}^{N_{\text{pw}}}$ wwre selected as in the previous numerical test, the exact solution would always belong to X_h, once it holds $N_{\text{pw}} \geq 4$ (and hence the relative error would be around the double precision round-off errors $\mathcal{O}(10^{-15})$ (see Figure 2.15)).

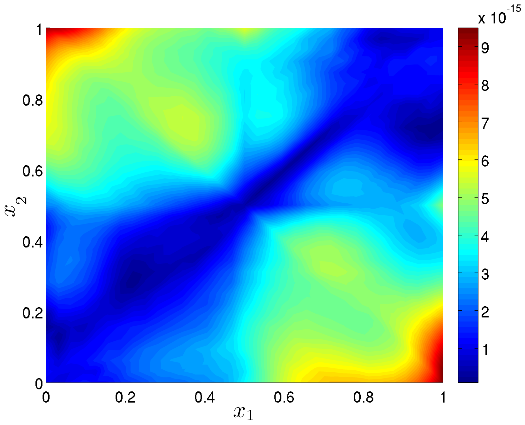

Figure 2.15: Modulus of the relative error e_h for a problem with wave number $k = 3$, using a discretization with $N_{\text{pw}} = 4$, mesh size $h = 1.7 \times 10^{-1}$ and choosing $\theta_j = 2\pi(j-1)/N_{\text{pw}}$, for $j = 1, \ldots, N_{\text{pw}}$ as angle of incidence in the PUFEM planewave discretization.

The behaviour of the relative error with respect to the mesh size h is illustrated in Figures 2.16 (using double precision floating-point arithmetic) and in 2.17 (using quadruple precision floating-point arithmetic). The numerical results in this test with smooth solution seems to indicate that the PUFEM method converges with an order of $\mathcal{O}(h^{3/2})$. Figures 2.18 and 2.19 show the real part of the approximated PUFEM solution and its corresponding modulus of the relative error, computed by using a fine mesh with $h = 8.5 \times 10^{-2}$, for two different wave number values, $k = 3$ and $k = 10$.

The behaviour of the relative error with respect to the wave number can be infer from the results shown in Figure 2.20 (using double precision floating-point arithmetic) and Figure 2.21 (using quadruple precision floating-point arithmetic). In both cases, it could be deduced that an overall $\mathcal{O}(k^2)$ order is reached by the PUFEM method.

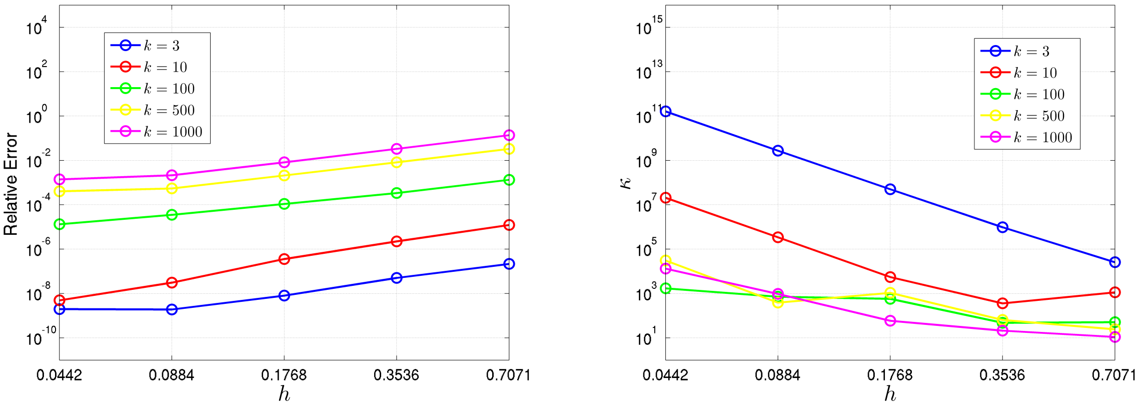

Figure 2.16: Relative error (left) and condition number κ (right) plotted with respect to the mesh size h for different values of the wave number k. the PUFEM discretization uses $N_{\text{pw}} = 4$ and it has been implemented using double precision floating-point arithmetic.

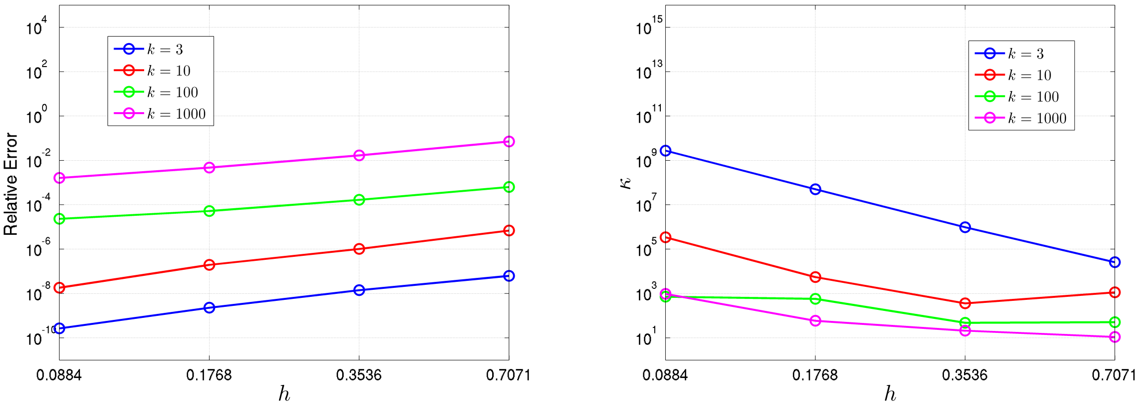

Figure 2.17: Relative error (left) and condition number κ (right) plotted with respect to the mesh size h for different values of the wave number k. The PUFEM discretization uses $N_{\text{pw}} = 4$ and it has been implemented using quadruple precision floating-point arithmetic.

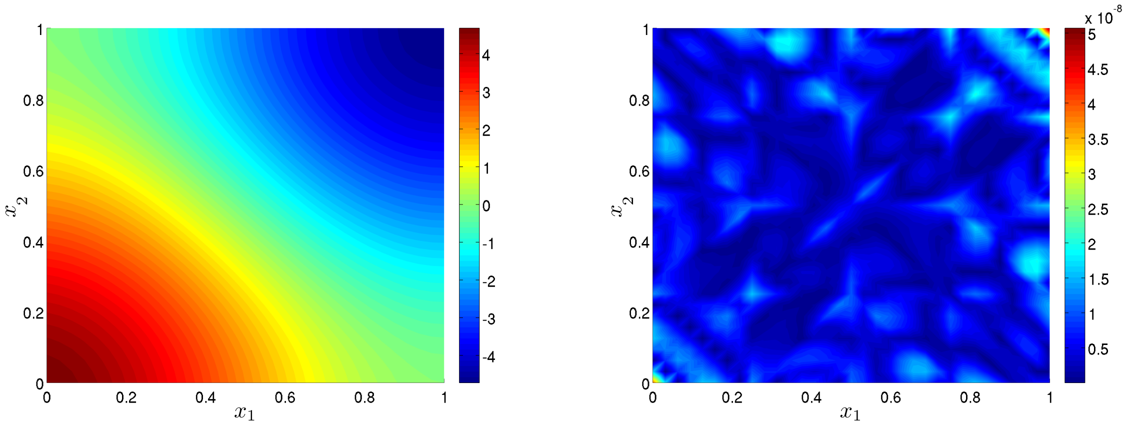

Figure 2.18: Real part of the PUFEM approximate solution (left) and modulus of the relative error computed for a discretization with $N_{\text{pw}} = 4$, mesh size $h = 8.5 \times 10^{-2}$ and wave number $k = 3$ (using double precision floating-point arithmetic).

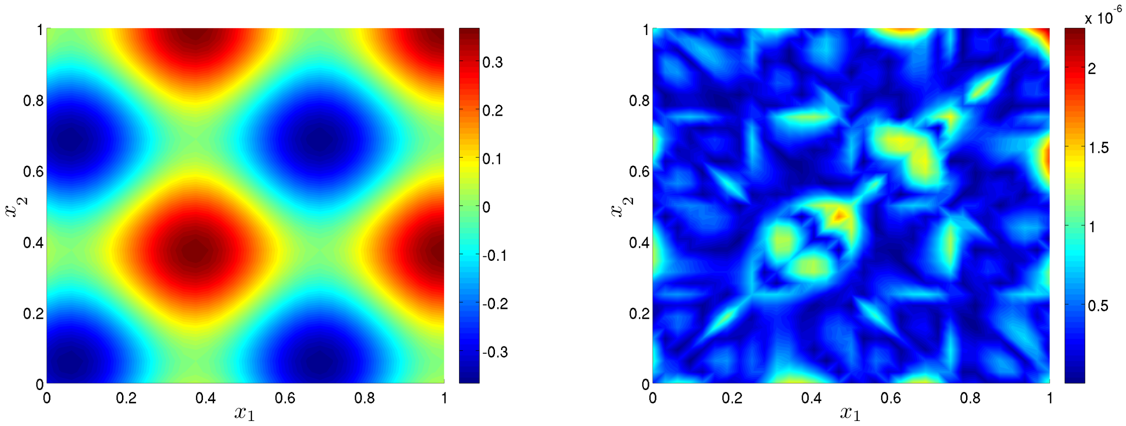

Figure 2.19: Real part of the PUFEM approximate solution (left) and modulus of the relative error computed for a discretization with $N_{\text{pw}} = 4$, mesh size $h = 8.5 \times 10^{-2}$ and wave number $k = 10$ (using double precision floating-point arithmetic).

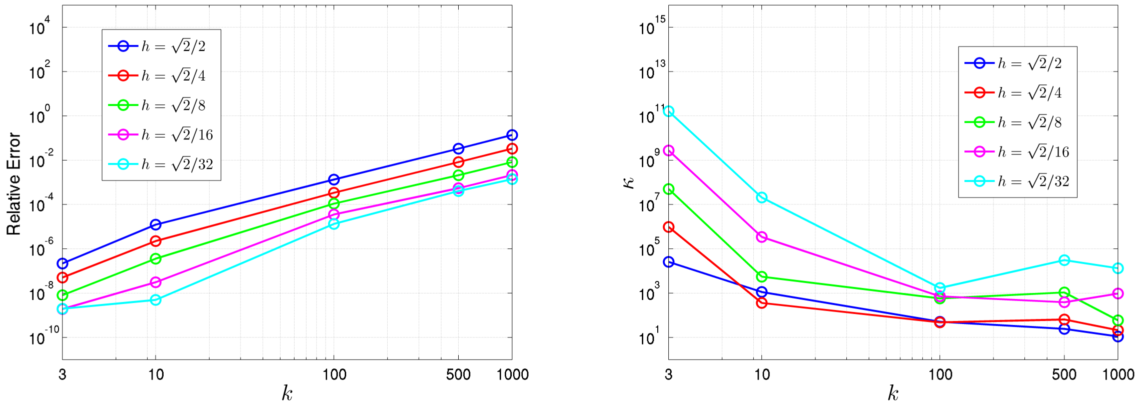

Figure 2.20: Relative error (left) and condition number κ (right) plotted with respect to the wave number k. The PUFEM discretization uses $N_{\text{pw}} = 4$ and it has been implemented using double precision floating-point arithmetic.

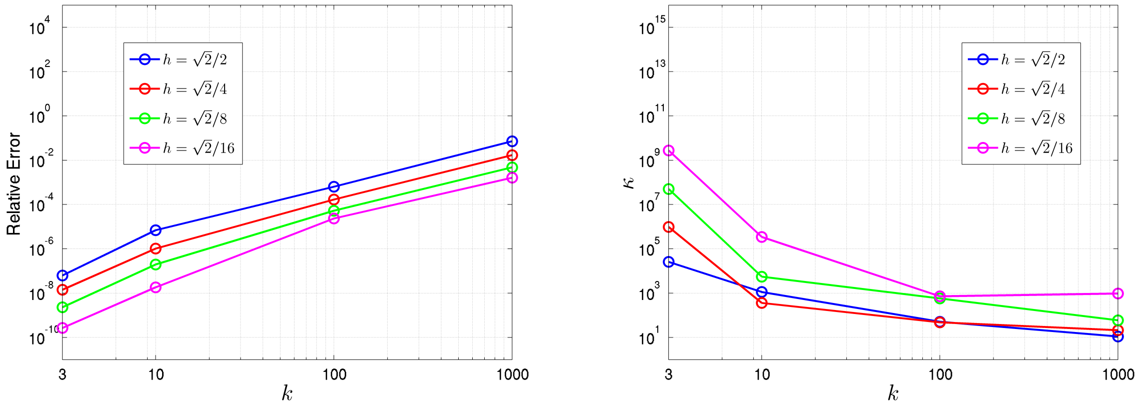

Figure 2.21: Relative error (left) and condition number κ (right) plotted with respect to the wave number k. The PUFEM discretization uses $N_{\text{pw}} = 4$ and it has been implemented using quadruple precision floating-point arithmetic.

2.4 Two-dimensional Helmholtz problem with piecewise constant wave number

In this section, the PUFEM method will be applied to the two-dimensional Helmholtz problem with piecewise constant wave number, which is already introduced in Section 2.4.1. The proposed PUFEM method consists in the two-dimensional extension of the transmission-reflection PUFEM discretization introduced previously in the one-dimensional case. As it has been made in the section above, the discrete PUFEM space and the integration techniques are described in detail (see 2.4.3 and 2.4.4, respectively). Finally, a variety of numerical test are carried out in Section 2.4.5.

For sake of simplicity in the exposition of the main ideas of the proposed transmission-reflection PUFEM method in two dimensions, the PUFEM discretization will be described only for Helmholtz problem involving a computational with two layers. Clearly, analogous arguments can be reproduced for a problem stated in a multilayered domain with piecewise constant wave number. So, the first step in the description of the transmission-reflection PUFEM method will consist in recasting the original Helmholtz problem with variable wave number as a coupled problem involving two layers.

2.4.1 Bi-layered Helmholtz problem

Let Ω_+ and Ω_- be two open bounded regular domains in $\mathbb{R}^2$, with Σ a planar interface between them, which without loss of generality, it will be assumed that it is located on $\Sigma = \partial\Omega_+ \cap \partial\Omega_- \subset \{\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2;\ x_2 = H\}$, being H a fix value. Let Ω the interior of the compact set $\overline{\Omega}_+ \cup \overline{\Omega}_-$, which will be considered as the global computational domain and $\partial\Omega$ its boundary. Let Γ_+ and Γ_- be the part of the boundaries of Ω_+ and Ω_-, respectively, which lie on the boundary of Ω, this is, $\Gamma_+ = \partial\Omega \cap \partial\Omega_+$ and $\Gamma_- = \partial\Omega \cap \partial\Omega_-$. If $u : \Omega \to \mathbb{C}$ is the global unknown of the Helmholtz problem, it can be split in two new unknown fields:

$$u = \left\{ \begin{array}{ll} u_+ & \text{in } \Omega_+, \\ u_- & \text{in } \Omega_-. \end{array} \right. \tag{2.38}$$

In the same manner, the (strictly positive) piecewise constant wave number can be defined independently for each layer as follows:

$$k = \left\{ \begin{array}{ll} k_+ & \text{in } \Omega_+, \\ k_- & \text{in } \Omega_-, \end{array} \right. \tag{2.39}$$

being k_+ and k_- two strictly positive constants. Consequently, the two-dimensional bi-layered Helmholtz problem consists in: fixed k_+, $k_- > 0$ and given Neumann boundary data g_+, g_-, find u_+, u_- such that it is satisfied

$$-\Delta u_+ - k_+^2 u_+ = 0 \qquad \text{in } \ \Omega_+, \tag{2.40}$$

$$-\Delta u_- - k_-^2 u_- = 0 \qquad \text{in } \ \Omega_-, \tag{2.41}$$

$$\frac{\partial u_+}{\partial \boldsymbol{n}_+} = g_+ \qquad \text{on } \ \Gamma_+, \tag{2.42}$$

$$\frac{\partial u_-}{\partial \boldsymbol{n}_-} = g_- \qquad \text{on } \ \Gamma_-, \tag{2.43}$$

$$u_+ = u_- \qquad \text{on } \ \Sigma, \tag{2.44}$$

$$\frac{\partial u_+}{\partial \boldsymbol{\nu}} = \frac{\partial u_-}{\partial \boldsymbol{\nu}} \qquad \text{on } \ \Sigma, \tag{2.45}$$

where $\boldsymbol{n}_+$ is the unit normal vector outward Ω_+ on boundary Γ_+, $\boldsymbol{n}_-$ is the unit normal vector outward Ω_- on boundary Γ_-, and $\boldsymbol{\nu} = (1,0)$ the unit normal vector outward Ω_- on interface boundary Σ.

2.4.2 Variational formulation

Let $v : \Omega \to \mathbb{C}$ be a test function regular enough. Multiplying equations (2.40) and (2.41) by the complex-conjugated of a test function and integrating over Ω_+ and Ω_- respectively, the following equations are obtained:

$$-\int_{\Omega_+} \Delta u_+ \, \bar{v} \, \mathrm{d}\boldsymbol{x} - k_+^2 \int_{\Omega_+} u_+ \, \bar{v} \, \mathrm{d}\boldsymbol{x} = 0, \tag{2.46}$$

$$-\int_{\Omega_-} \Delta u_- \, \bar{v} \, \mathrm{d}\boldsymbol{x} - k_-^2 \int_{\Omega_-} u_- \, \bar{v} \, \mathrm{d}\boldsymbol{x} = 0. \tag{2.47}$$

Notice that since v is smooth enough then the coupling condition (2.44) on Σ is satisfied automatically (in fact, it is enough to assume that $v \in \mathrm{H}^1(\Omega)$).

Applying now a standard Green's formula over (2.46) and (2.47), and having into account the boundary and coupling conditions (2.42)-(2.45) satisfied by u_-, u_+ and the test function v, the variational formulation for the bi-layered two-dimensional Helmholtz problem chosen is stated as follows: given $g_+ \in \mathrm{L}^2(\Gamma_+)$ and $g_- \in \mathrm{L}^2(\Gamma_-)$, find $u_+ \in \mathrm{H}^1(\Omega_+)$, $u_- \in \mathrm{H}^1(\Omega_-)$ such that

$$\int_{\Omega_+} \nabla u_+ \cdot \nabla \bar{v} \, \mathrm{d}\boldsymbol{x} + \int_{\Omega_-} \nabla u_- \cdot \nabla \bar{v} \, \mathrm{d}\boldsymbol{x} - k_+^2 \int_{\Omega_+} u_+ \, \bar{v} \, \mathrm{d}\boldsymbol{x} \\ - k_-^2 \int_{\Omega_-} u_- \, \bar{v} \, \mathrm{d}\boldsymbol{x} = \int_{\Gamma_+} g_+ \, \bar{v} \, \mathrm{d}\sigma + \int_{\Gamma_-} g_- \, \bar{v} \, \mathrm{d}\sigma,$$

for all $v \in \mathrm{H}^1(\Omega)$, or equivalently, taking into account (2.38) and (2.39), the variational problem consists in finding $u \in \mathrm{H}^1(\Omega)$ such as

$$\int_{\Omega} \nabla u \cdot \nabla \bar{v} \, \mathrm{d}\boldsymbol{x} - \int_{\Omega} k^2 u \, \bar{v} \, \mathrm{d}\boldsymbol{x} = \int_{\Gamma_+} g_+ \, \bar{v} \, \mathrm{d}\sigma + \int_{\Gamma_-} g_- \, \bar{v} \, \mathrm{d}\sigma, \tag{2.48}$$

for all $v \in \mathrm{H}^1(\Omega)$.

Classical arguments based on the Fredholm's alternative theory [7] and the fact that the resolvent operator associated to this problem is a self-adjoint compact operator show that the variational problem (2.48) has an unique solution except for an infinite sequence of real wave number values $\{\tilde{k}_j\}_{j=0}^{\infty}$, which should be understood as the resonances of the mechanical system associated to this bi-layered model problem. Throughout the rest of this section, all the wave number values will be selected such are not coincident with any of the resonance values where the solution is not unique.

2.4.3 Transmission-reflection PUFEM discretization

Similarly to the PUFEM discretization applied to the Helmholtz problem with constant wave number, in order to define a basis of the PUFEM discrete space with a variable wave number, the standard Lagrange $\mathbb{P}_1$ two-dimensional finite element functions will be multiplied by some kind of planewave expressions. Four alternatives have been studied in the bi-layered one-dimensional problem (see Section 2.2) and it was concluded that the transmission-reflection PUFEM discretization was the best candidate to obtain the most accurate results in two dimensions.

In conclusion, since the Helmholtz problem involves two layers with different wave number, this fact must be taken into account and hence the reflections and transmissions that occur on the interface should be included in the PUFEM basis. More precisely, when a certain planewave that propagates with an incident angle θ_I (see Figure 2.1) impinges on the interface between the two layers (possibly with different wave numbers in each layer), then a reflected wave with amplitude R is produced (this new planewave contribution can be understood as a signal travelling in the direction given by the angle $2\pi - \theta_I$, i.e., an angle that has the same cosine as θ_I and opposite sine), and a transmitted wave with amplitude T, whose angle of propagation β depends on the Snell's law:

$$k_+ \cos\theta_I = k_- \cos\beta. \tag{2.49}$$

In the case of $k_- < k_+$, the transmission planewave in Ω_- is a so-called evanescence wave (due its exponential decay) once it holds $\theta_I \in [\pi, 2\pi]$ and

$$(\cos\theta_I)^2 > \left(\frac{k_-}{k_+}\right)^2. \tag{2.50}$$

This evanescent behaviour causes that the most part of the contribution of the incident planewave reflects into Ω_+ and the an exponential decay of the wave propagation in Ω_- mentioned above. Since the transmission-reflection PUFEM discretization will use fully planewave solutions, this kind of phenomena will be naturally included in the PUFEM discretization. In fact, the proposed PUFEM method use propagative and evanescent planewave functions.

As it is already assumed for the Helmholtz problem with constant wave number, to avoid any error coming for the triangular mesh, it will be assumed that the domain Ω is a two-dimensional polygon. Additionally, in this bi-layered problem, it will be required that the mesh is conformal with the coupling boundary Σ. In this manner, if $\mathcal{T}_h$ is a regular triangulation of the domain, where the mesh size h is defined as the maximum diameter in the triangulation, it holds

$$\Omega = \bigcup_{T\in\mathcal{T}_h} T, \qquad h = \max_{T\in\mathcal{T}_h} d_T,$$

being d_T the diameter of the triangle T (the diameter of the circle circumscribed in the triangle T [18]). Each node of the triangulation is denoted by $\boldsymbol{x}_j$, for all $n = 1, \ldots, N_{\text{fe}}$,

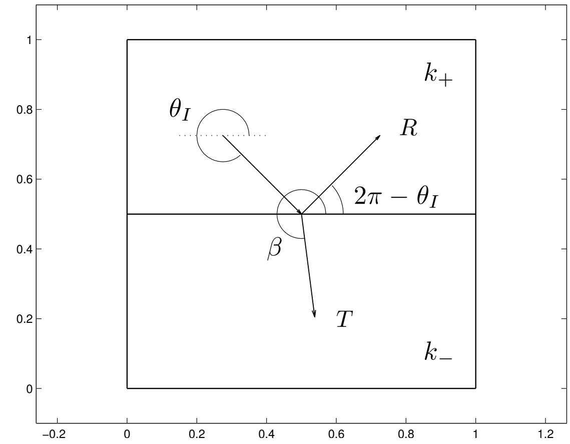

Figure 2.1: Scheme of the reflection and transmission planewaves on the interface between two layers (possibly with different wave number): A planewave is impiging on the interface boundary with an incident angle θ_I. If the condition (2.50) over θ_I is satisfied, an evanescence arises in Ω_- and the transmitted wave decays exponentially.

being N_{fe} the total number of nodes. Let $\{\varphi_n\}_{n=1}^{N_{\mathrm{fe}}}$ denote the standard Lagrange $\mathbb{P}_1$ two-dimensional finite element basis, where $\varphi_n(\boldsymbol{x}_m) = \delta_{mn}$, being δ_{mn} the Kronecker's delta.

The transmission-reflection PUFEM discrete space X_h will be defined by multiplying each finite element basis function by a certain number of planewave solutions $\{w_j\}_{j=1}^{N_{\mathrm{pw}}}$, being N_{pw} the number of these solutions computed in closed form. Let the incident angles θ_j of the plane wave functions be chosen evenly distributed in the plane, this is, $\theta_j = 2\pi(j-1)/N_{\mathrm{pw}}$. Under the considerations written above, the planewave solutions of the bi-layered Helmholtz problem can be computed by means of the following closed-form expressions:

$$w_j(\boldsymbol{x}) = \begin{cases} e^{ik_j^0 x_1}(T_j^+ e^{ik_j^+ x_2} + R_j^+ e^{-ik_j^+ x_2}) & \text{if } \boldsymbol{x} \in \Omega_+, \\ e^{ik_j^0 x_1}(T_j^- e^{ik_j^- x_2} + R_j^- e^{-ik_j^- x_2}) & \text{if } \boldsymbol{x} \in \Omega_-, \end{cases} \tag{2.51}$$

where $R_j^\pm$ and $T_j^\pm$ are respectively, the reflection and transmission coefficients of the planewaves, and k_j^0 ensures that the Snell's law is fulfilled, and it is defined as

$$k_j^0 = \begin{cases} k_+ \cos\theta_j & \text{if } \theta_j \in [\pi, 2\pi], \\ k_- \cos\theta_j & \text{if } \theta_j \in (0, \pi), \end{cases}$$

and

$$k_j^+ = -\sqrt{(k_+)^2 - (k_j^0)^2}, \qquad k_j^- = -\sqrt{k_-^2 - (k_j^0)^2}.$$

Note that, if the angle $\theta_j \in [\pi, 2\pi]$ and $(\cos\theta_j)^2 < (k_-/k_+)^2$, the coefficient k_j^- will be a pure imaginary number and hence the transmitted planewave will be evanescent (it will decay exponentially with respect to the distance to the coupling interface Σ).

In order to compute the transmission and reflection coefficients for each $j = 1, \dots, N_{\text{pw}}$, a system with four unknowns and four equations is posed. These equations impose the continuity condition on the coupling interface for the function (2.51),

$$T_j^+ e^{ik_j^+ H} + R_j^+ e^{-ik_j^+ H} = T_j^- e^{ik_j^- H} + R_j^- e^{-ik_j^- H}, \tag{2.52}$$

and the continuity of its normal derivative on the coupling interface, leading to

$$k_j^+ \left(T_j^+ e^{ik_j^+ H} - R_j^+ e^{-ik_j^+ H} \right) = k_j^- \left(T_j^- e^{ik_j^- H} - R_j^- e^{-ik_j^- H} \right). \tag{2.53}$$

Additionally, the linear system is completed with some radiation conditions with impose certain values of the transmission and reflection coefficients. These two conditions are settled depending on the angle of incidence:

$$\begin{cases} R_j^- = 1, \quad T_j^+ = 0 & \text{if } \theta_j \in (0, \pi), \\ R_j^- = 0, \quad T_j^+ = 1 & \text{if } \theta_j \in (\pi, 2\pi), \\ R_j^+ = 0, \quad T_j^+ = 1 & \text{if } \theta_j = 0, \pi, 2\pi. \end{cases}$$

Consequently, once the planewave solutions are computed, then the basis functions $\psi_{n,j}$ of the discrete PUFEM space can be written

$$\psi_{nj}(\boldsymbol{x}) = \varphi_n(\boldsymbol{x}) w_j(\boldsymbol{x}), \quad 1 \leq n \leq N_{\text{fe}}, \ 1 \leq j \leq N_{\text{pw}}. \tag{2.54}$$

Hence, the discrete PUFEM space is given by $\mathrm{X}_h = \langle \{\psi_{n1}\}_{n=1}^{N_{\text{fe}}} \cup \ldots \cup \{\psi_{nN_{\text{pw}}}\}_{n=1}^{N_{\text{fe}}} \rangle$. Taking into account the definition of the discrete PUFEM space, given by the span of basis functions (2.54), the discrete PUFEM problem is described as follows: fixed $k > 0$ and given $g_\pm \in \mathrm{L}^2(\Gamma_\pm)$, find $u_h \in \mathrm{X}_h$ such that

$$\int_\Omega \nabla u_h \cdot \nabla \bar{v}_h \, \mathrm{d}\boldsymbol{x} - \int_\Omega k^2 u_h \, \bar{v}_h \, \mathrm{d}\boldsymbol{x} = \int_{\Gamma_+} g_+ \, \bar{v}_h \, \mathrm{d}\sigma + \int_{\Gamma_-} g_- \, \bar{v}_h \, \mathrm{d}\sigma, \tag{2.55}$$

for all $v_h \in \mathrm{X}_h$. The discrete PUFEM solution u_h can be written in terms of the basis functions in X_h,

$$u_h(\boldsymbol{x}) = \sum_{n=1}^{N_{\text{fe}}} \sum_{j=1}^{N_{\text{pw}}} u_{nj} \psi_{nj}(\boldsymbol{x}) = \sum_{n=1}^{N_{\text{fe}}} \sum_{j=1}^{N_{\text{pw}}} u_{nj} \varphi_n(\boldsymbol{x}) w_j(\boldsymbol{x}), \tag{2.56}$$

where $\vec{u}_h(u_{11}, u_{21}, \ldots, u_{N_{\text{fe}}1}, \ldots, u_{1N_{\text{pw}}}, \ldots, u_{N_{\text{fe}}N_{\text{pw}}})^t \in \mathbb{C}^{N_{\text{fe}}N_{\text{pw}}}$ is the complex vector of coefficients of the discrete PUFEM function $\boldsymbol{u}_h$. The discrete problem can be written in matrix form as

$$(\mathcal{K}_h - k_+^2 \mathcal{M}_h) \vec{u}_h = \vec{g}_h, \tag{2.57}$$

where the mass matrix $\mathcal{M}_h$ and the stiffness matrix $\mathcal{K}_h$ are defined by

$$[\mathcal{K}_h]_{nj,ml} = \int_{\Omega_+} \nabla\left(\varphi_n w_j\right) \cdot \nabla\left(\varphi_m \bar{w}_l\right) \mathrm{d}\boldsymbol{x} + \int_{\Omega_-} \nabla\left(\varphi_n w_j\right) \cdot \nabla\left(\varphi_m \bar{w}_l\right) \mathrm{d}\boldsymbol{x}, \tag{2.58}$$

$$[\mathcal{M}_h]_{nj,ml} = \int_{\Omega_+} \varphi_n w_j \varphi_m \bar{w}_l \, \mathrm{d}\boldsymbol{x} + \int_{\Omega_-} \frac{k_-^2}{k_+^2} \varphi_n w_j \varphi_m \bar{w}_l \, \mathrm{d}\boldsymbol{x}, \tag{2.59}$$

for all $1 \leq n, m \leq N_{\text{fe}}$ and $1 \leq j, l \leq N_{\text{pw}}$ (it should be remarked that the ordering of the matrix coefficients is given by the ordering induced by the coefficient order of the unknown vector $\vec{u}_h$). Analogously, each coefficient of the right-hand side vector $\vec{g}_h$ has the L^2-projection of the boundary data g onto the discrete PUFEM basis, this is,

$$[\vec{g}_h]_{nj} = \int_{\Gamma_+} g_+ \varphi_n(\boldsymbol{x}) \bar{w}_j \, \mathrm{d}\sigma + \int_{\Gamma_-} g_- \varphi_n(\boldsymbol{x}) \bar{w}_j \, \mathrm{d}\sigma. \tag{2.60}$$

Since the integrals stated above are highly oscillatory if the wavelength of the planewaves $2\pi/k$ is much smaller than the typical size h of the support of the finite element functions, standard numerical quadrature rules (for instance, based on Gauss-Legendre with a reduced number of points) lead to inaccurate results. The following section describe in detail how these oscillatory integrals are computed in closed form.

2.4.4 Integration techniques

Since a certain number of PUFEM planewave basis functions are composed by evanescent planewaves, the application of the integration technique described for the PUFEM discretization with constant wave number (section 2.3.3) becomes inadequate, mainly because of the affine mapping which rotate the triangles is possibly complex-valued, the nand so the new rotated triangle could be lying on $\mathbb{C}^2$. This section will be devoted to describe the alternative exact integration procedure used for the computation of the local contributions to the integrals (2.58) and (2.59).

Analogously to the PUFEM applied to problems with constant wave number, the matrix $\mathcal{K}_h - k_+^2 \mathcal{M}_h$ and the right hand side $\vec{g}_h$ are both computed locally, respectively in each triangle of the mesh or on each edge of the boundary and then assembled globally. To compute the matrix coefficients of $\mathcal{K}_h - k_+^2 \mathcal{M}_h$ locally, exact integration is used by making an affine change of variable to the reference triangle $\hat{T}$ (with vertices $(0,0)$, $(1,0)$ and $(0,1)$), this is, if $\boldsymbol{a} = (a_1, a_2)$, $\boldsymbol{b} = (b_1, b_2)$, and $\boldsymbol{c} = (c_1, c_2)$ are the vertices of the triangle T (in counter-clock wise order), the affine mapping to the triangle of reference $F_T : \hat{\boldsymbol{x}} = (\hat{x}_1, \hat{x}_2) \in \hat{T} \mapsto (x_1, x_2) \in T$ is given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 - a_1 & c_1 - a_1 \\ b_2 - a_2 & c_2 - a_2 \end{pmatrix} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} + \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{2.61}$$

To compute the right-hand side $\vec{g}$ locally, exact integration is used too, by making the same parametrization described for the Helmholtz problem with constant wave number (see (2.22)). Notice that the Neumann functions $g_\pm$ can be integrated in closed form if its expression is known also in close form. In the present work, piecewise constant functions and exponential-type functions will be considered. Taking in mind these kind of expressions for $g_\pm$ and the expressions in the integrands of matrices (2.58) and (2.59), a simple integration by parts (now performed in both variables x_1 and x_2) leads to the exact integration formulas for the contribution of the right-hand side $\vec{g}_h$.

2.4.5 Numerical results

To illustrate the accuracy and efficiency of the transmission-reflection PUFEM method to approximate the solution of a two-dimensional Helmholtz problem with piecewise constant wave number and Neumann boundary conditions, some numerical results have been carried out. Some problems with exact solutions that can easily be computed analytically, resulting plane waves that impinge on the interface with different angles, will be first studied. Then, a comparison of the variable partition of unity finite element method with a standard Lagrange $\mathbb{P}_1$ finite element method for a two-dimensional Helmholtz problem with constant Neumann boundary conditions will be described in Section 2.4.5. It should be remarked that the computer code has been implemented and run in MATLAB.

Throughout this section about the numerical results of the PUFEM method, recall that the mesh chosen for the discretization has N_{fe} nodes $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_{\text{fe}}}\}$ and its elements have a maximum diameter h. The condition number of the matrix $\mathcal{K}_h - k_+^2 \mathcal{M}_h$ involved in the linear system (2.57) is denoted by κ. The relative error field in L^∞-norm e_h and its norm will be computed on the mesh vertices as follows:

$$e_h(\boldsymbol{x}) = \frac{u(\boldsymbol{x}) - u_h(\boldsymbol{x})}{\max\limits_{j=1,\ldots,N_{\text{fe}}} |u(\boldsymbol{x}_j)|}, \qquad \text{relative error} = \max_{j=1,\ldots,N_{\text{fe}}} |e_h(\boldsymbol{x}_j)| \tag{2.62}$$

where u is the exact solution of the Helmholtz problem (2.40)-(2.45) and u_h is the discrete PUFEM approximation computed by means of (2.55).

Approximation of planewave-type solutions

In this first subsection, it will be analysed the behaviour of the PUFEM variable method applied to a Helmholtz problem, where the solution is given by a linear combination of plane waves (defined differently in each layer). Let $\Omega = (0,1) \times (0,1)$ be the unit square, with upper subdomain given by $\Omega_+ = (0,1) \times (0.5,1)$ and the lower subdomain $\Omega_- = (0,1) \times (0,0.5)$. So, the coupling interface Σ is located at the line $x_2 = H = 1/2$ (see Figure 2.2).

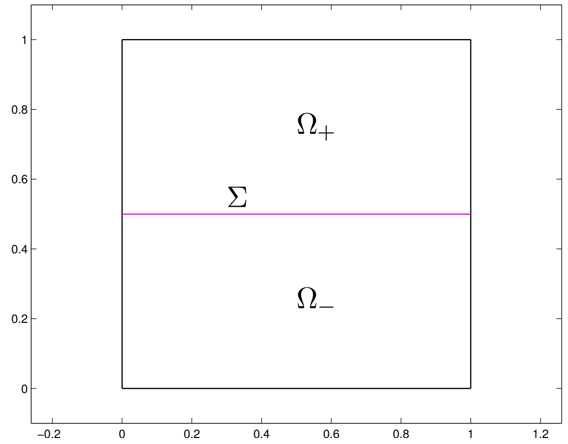

Figure 2.2: Domain considered for the bi-layered Helmholtz problem used in the numerical test described in Section 2.4.5. The interface Σ is lying on $x_2 = 1/2$ (pink segment).

In this first test, the bi-layered Helmholtz problem (2.40)-(2.45) is settled with the variable wave number

$$k(\boldsymbol{x}) = \begin{cases} k_+ & \text{in } \Omega_+, \\ k_+/4 & \text{in } \Omega_-, \end{cases}$$

and considering the Neumann boundary data $g_\pm$ such that the exact solution is given by the transmission and reflection planewaves generated by an incident planewave with angle β, i.e.,

$$u(\boldsymbol{x}) = \begin{cases} e^{ik_1x_1}(T^+e^{ik_2^+x_2} + R^+e^{-ik_2^+x_2}) & \text{if } \boldsymbol{x} \in \ \Omega_+, \\ e^{ik_1x_1}(T^-e^{ik_2^-x_2} + R^-e^{-ik_2^-x_2}) & \text{if } \boldsymbol{x} \in \ \Omega_-, \end{cases} \tag{2.63}$$

with

$$k_1 = \begin{cases} k_+ \cos\beta & \text{if } \beta \in [\pi, 2\pi], \\ \dfrac{k_+ \cos\beta}{4} & \text{if } \beta \in (0, \pi), \end{cases}$$

and where the reflection and transmission coefficients and the components of the wave number vector satisfies the jump conditions (2.52)-(2.53) and the Snell's law (2.49). Obviously, from the definition of the exact solution if β coincides with an angle of incidence θ_j used in the PUFEM discretization, the exact solution u will belong to the PUFEM discrete space and hence the error will be theoretically null.

The first test consists in taking ten exact solutions with different incident angles β, some of them in the discrete space X_h, and study how the discrete PUFEM solution approximates them. The incident angles for the exact solutions are taken $\beta = 2\pi(j-1)/10$, for all $j = 1, \dots, 10$. In Figure 2.3, the relative error e_h for each of these ten exact solutions is shown (left plot), choosing the number of plane wave functions in the discretization $N_{\text{pw}} = 5$, for different values of the wave number k_+ and in a mesh with $h = 1.7 \times 10^{-1}$. Taking into account the values $\beta = j\pi/5$ for $j = 0, 2, 4, 6, 8$, the exact solution belongs to the discrete space X_h, so the relative error should have order similar to $\mathcal{O}(10^{-15})$. The fact that it does not can be explained observing the condition number κ of the PUFEM matrix system $\mathcal{K}_h - k_+^2\mathcal{M}_h$ (right plot). The PUFEM variable method is potentially very sensitive to the condition number, and for these simulations in double precision it has order between $\mathcal{O}(10^8)$ and $\mathcal{O}(10^{25})$. So in order to see if the method works, the simulations are repeated in quadruple precision (32 digits). Figure 2.6 shows the results for these simulations. It can be observed that using quadruple precision the relative error (left plot) has order close to $\mathcal{O}(10^{-30})$ for the exact solutions in X_h. Note that for the quadruple precision, the Matlab toolbox Advanpix has been used [34].

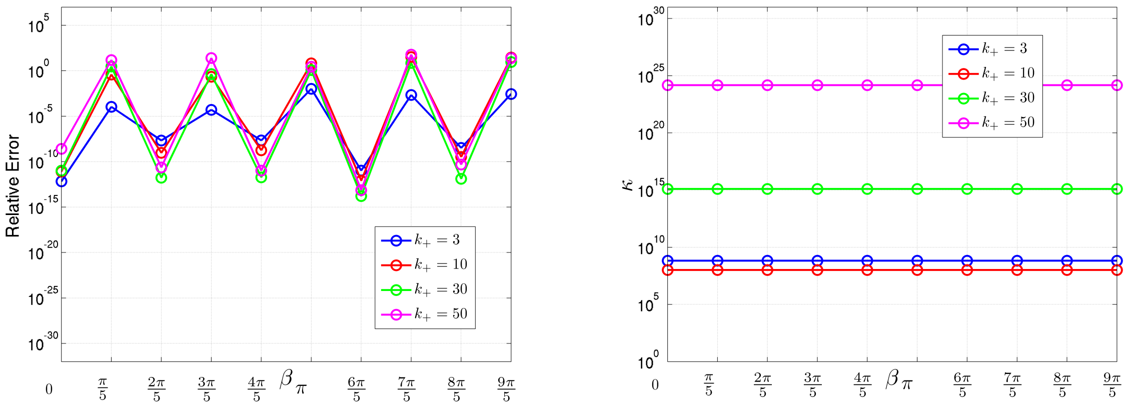

Figure 2.3: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k_+. The PUFEM discretization involves $N_{\mathrm{pw}} = 5$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using double precision floating-point arithmetic.

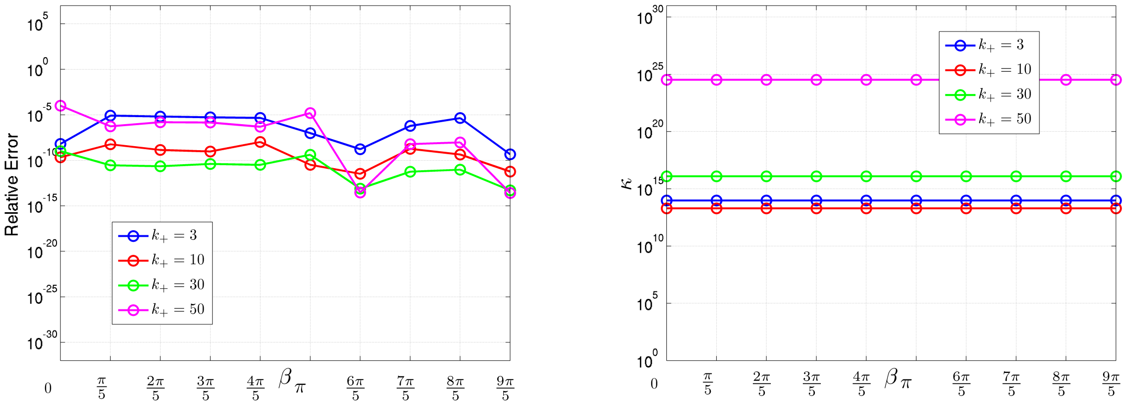

Figure 2.4: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k_+. The PUFEM discretization involves $N_{\text{pw}} = 10$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using double precision floating-point arithmetic.

If the number of planewave functions is now taken $N_{\text{pw}} = 10$ and $N_{\text{pw}} = 20$, the ten exact solutions fall into the discrete space X_h. Figures 2.4 and 2.5 show that in double precision, the relative error (left plots) for these ten solutions and for several values of k_+ is higher than the order expected $\mathcal{O}(10^{-15})$ because of the high condition number (right plots). The same simulations are then repeated with quadruple precision in Figures 2.7 and 2.8. In these cases, the relative error (left plots) is close or smaller than $\mathcal{O}(10^{-15})$ even if the condition number is high (right plots). Notice that when the number of planewave functions used in the discretization grows, the condition number of the PUFEM matrix system grows too.

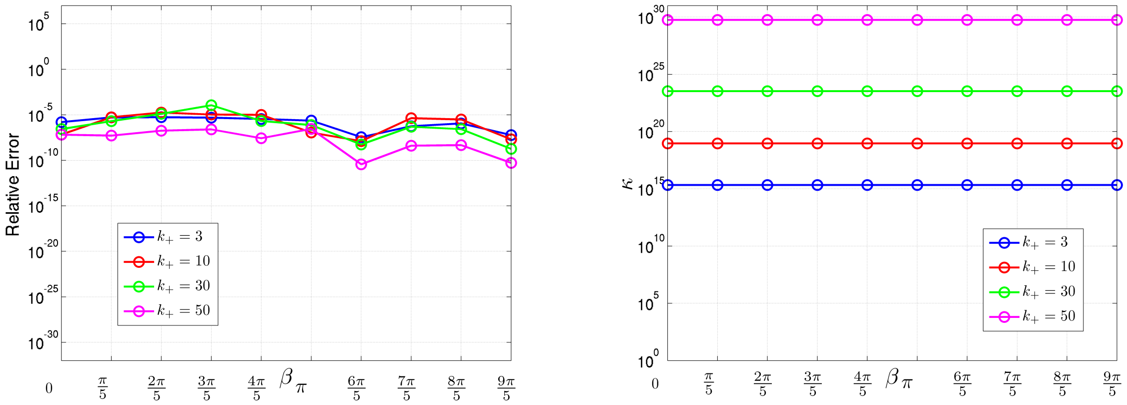

Figure 2.5: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k_+. The PUFEM discretization involves $N_{\text{pw}} = 20$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using double precision floating-point arithmetic.

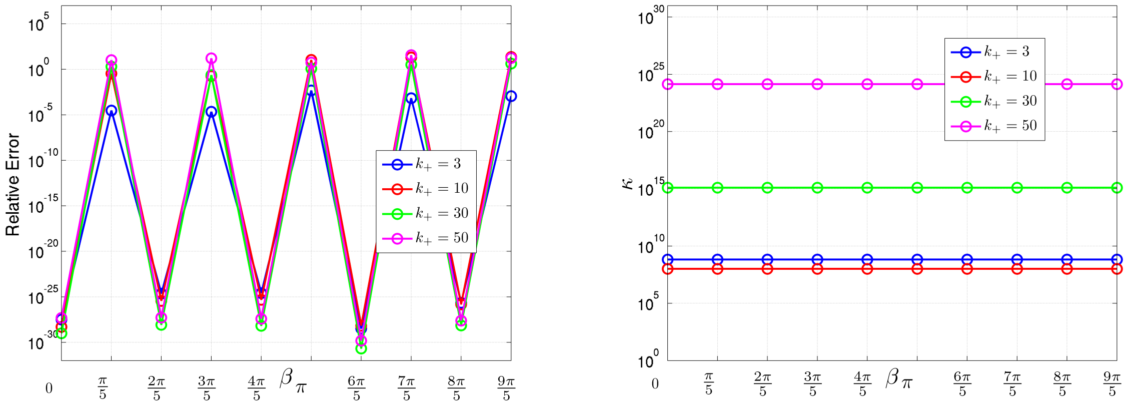

Figure 2.6: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k_+. The PUFEM discretization involves $N_{\text{pw}} = 5$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using quadruple precision floating-point arithmetic.

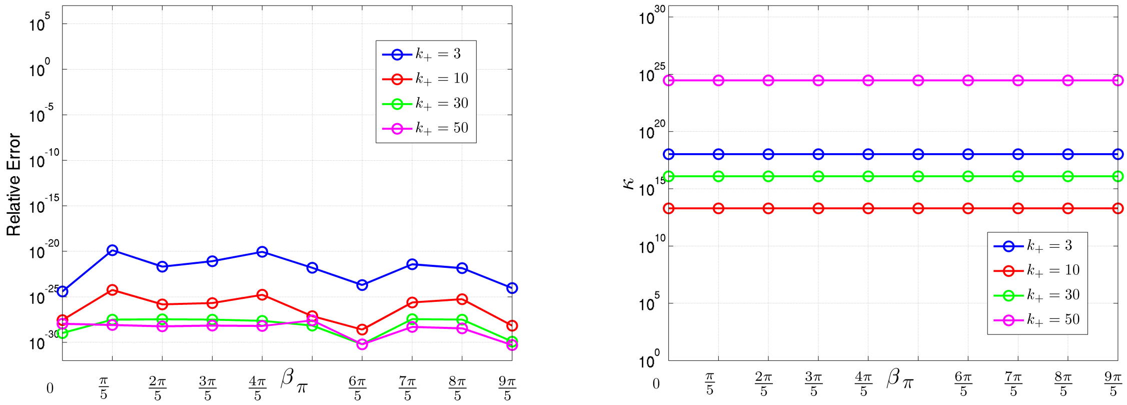

Figure 2.7: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k_+. The PUFEM discretization involves $N_{\text{pw}} = 10$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using quadruple precision floating-point arithmetic.

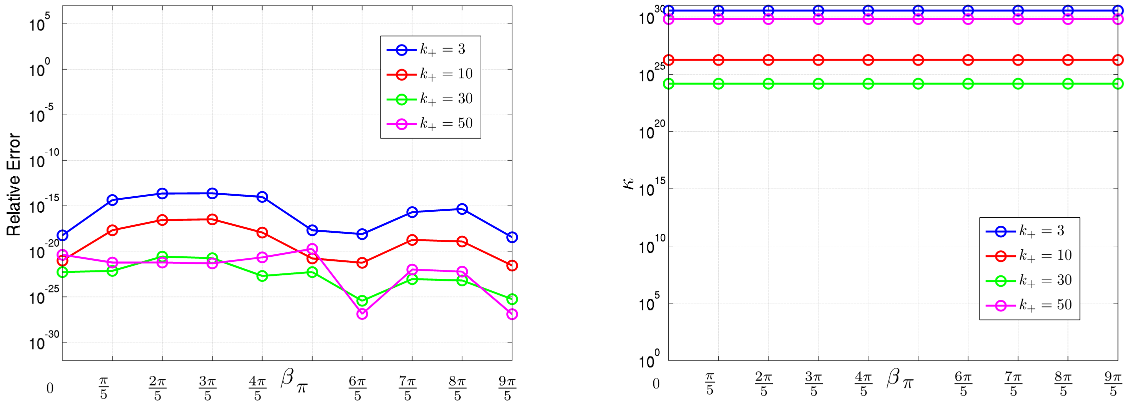

Figure 2.8: Relative error and condition number κ for a variety of exact planewave solutions varying the angle of incident β of the exact solution and the wave number k_+. The PUFEM discretization involves $N_{\text{pw}} = 20$ planewaves, a triangular mesh with size $h = 1.7 \times 10^{-1}$, and an implementation using quadruple precision floating-point arithmetic.

For the second test, an exact solution is fixed, with incident angle $\beta = 2\pi/21$. Some PUFEM discretizations are used in a mesh with $h = 1.7 \times 10^{-1}$, each of them with a different number of planewave functions N_{pw}, between 2 and 20. For different values of the wave number k_+, Figures 2.9 (in double precision) and 2.10 (quadruple precision) show that relative error (left plots) decays exponentially when the number of planewave functions chosen in the variable PUFEM discretization is increased.

The last test in this subsection consists in taking two different exact solutions, one evanescent and the other one propagative, and study the convergence of the relative error

e_h in terms of the mesh size and with respect to the wave number k_+. The chosen PUFEM discretization has eight planewave functions, $N_{\mathrm{pw}} = 8$. All the results in this test use double precision floating-point arithmetic.

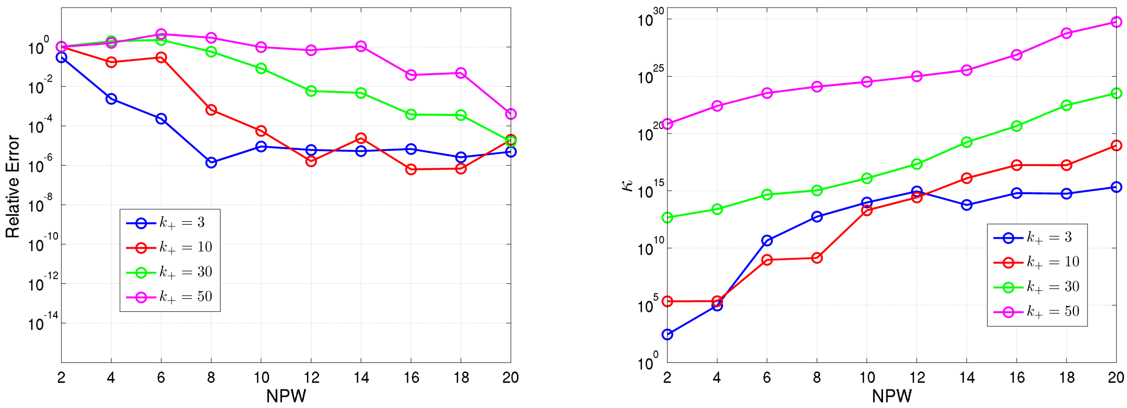

Figure 2.9: Relative error and condition number κ for an exact solution with incident angle $\beta = 2\pi/21$ and for different values of the wave number k_+. All the PUFEM approximations use a mesh with $h = 1.7 \times 10^{-1}$. Simulations have been run in double precision.

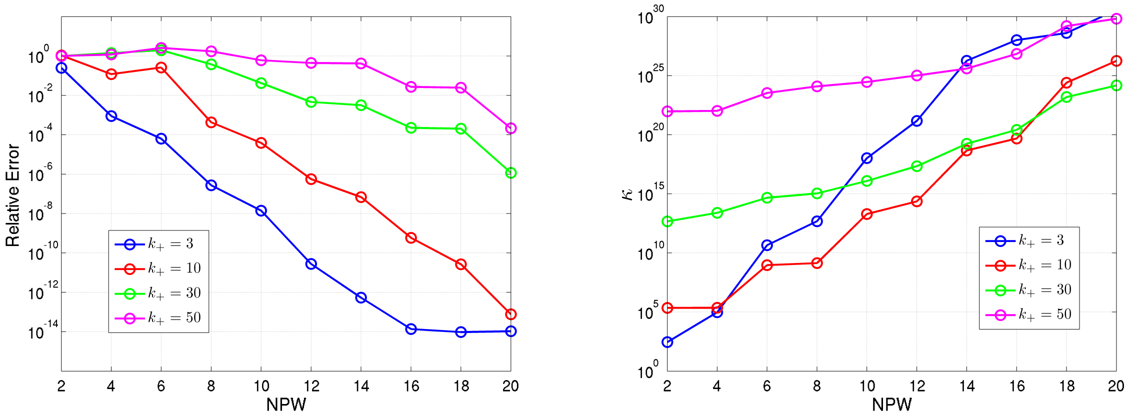

Figure 2.10: Relative error and condition number κ for an exact solution with incident angle $\beta = 2\pi/21$ and for different values of the wave number k_+. All the PUFEM approximations use a mesh with $h = 1.7 \times 10^{-1}$. Simulations have been run in quadruple precision.

The first exact solution considered has incident angle $\beta = \pi + 2\pi/9$ (Figure 2.11). When the wave impinges on the interface, an evanescent planewave is generated in Ω_- and nearly all the incident wave returns to Ω_+. Figure 2.12 shows the behaviour of the relative error (left plot) when the mesh size decreases for different values of the wave number. Figure 2.13 shows the behaviour of the relative error (left plot) respect to the wave number, considering different mesh sizes.

The second exact solution considered is a propagative wave that has incident angle $\beta = 2\pi/8 + 2\pi/9$ (Figure 2.14). The wave propagates to Ω_+ after it impinges on the interface. Figure 2.15 shows the behaviour of the relative error (left plot) when the mesh size decreases for different values of the wave number. Figure 2.16 shows the behaviour of the relative error (left plot) respect to the wave number, considering different mesh sizes.

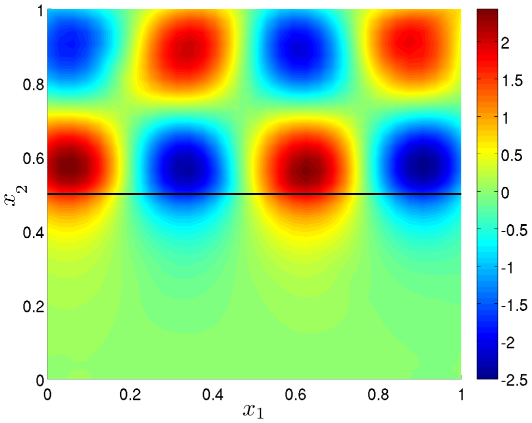

Figure 2.11: Real part of the exact solution computed with an incident angle $\beta = \pi + 2\pi/9$. A evanescent planewave is produced in Ω_- once the wave impinges on the coupling interface.

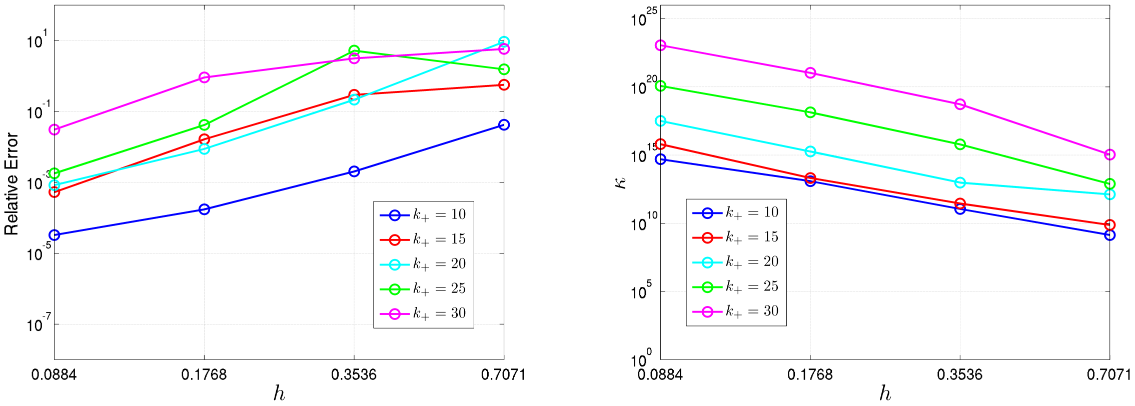

Figure 2.12: Relative error plotted with respect to the mesh size h (left) and condition number κ (right) for different values of the wave number k_+. The PUFEM discretization is settled with $N_{\text{pw}} = 8$, and the exact solution is an evanescent wave with incident angle $\beta = \pi + 2\pi/9$.

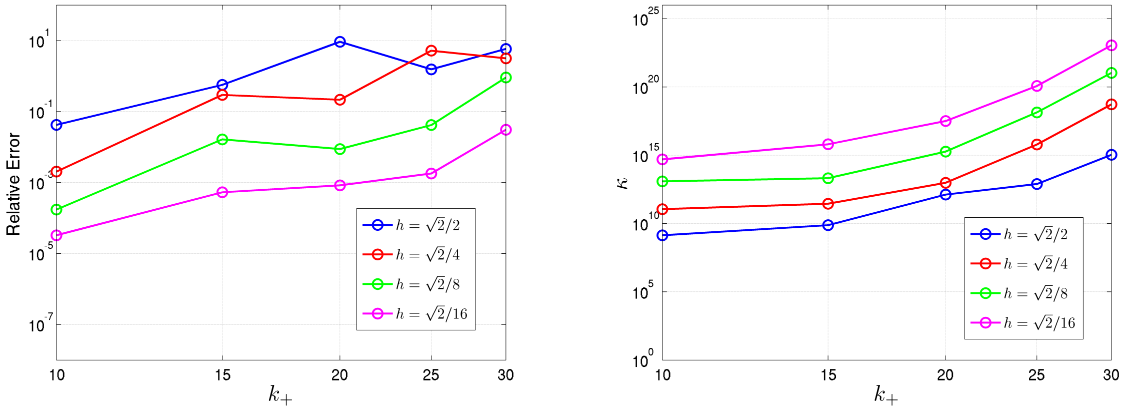

Figure 2.13: Relative error plotted with respect to the wave number k_+ (left) and condition number κ (right) for different values of the mesh size h. The PUFEM discretization is settled with $N_{\text{pw}} = 8$, and the exact solution is an evanescent wave with incident angle $\beta = \pi + 2\pi/9$.

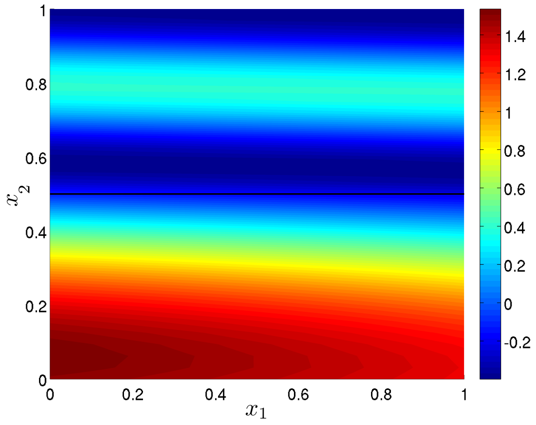

Figure 2.14: Real part of the exact solution with an incident angle $\beta = 2\pi/8 + 2\pi/9$. The wave propagates to Ω_+ after the incident wave impinges on the interface.

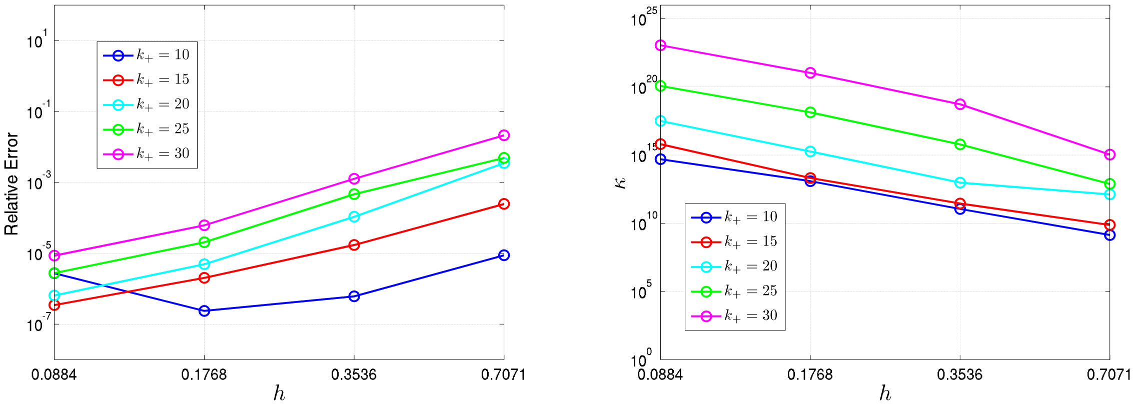

Figure 2.15: Relative error plotted with respect to the mesh size h (left) and condition number κ (right) for different values of the wave number k_+. The PUFEM discretization is settled with $N_{\text{pw}} = 8$, and the exact solution is a propagative wave with incident angle $\beta = 2\pi/8 + 2\pi/9$.

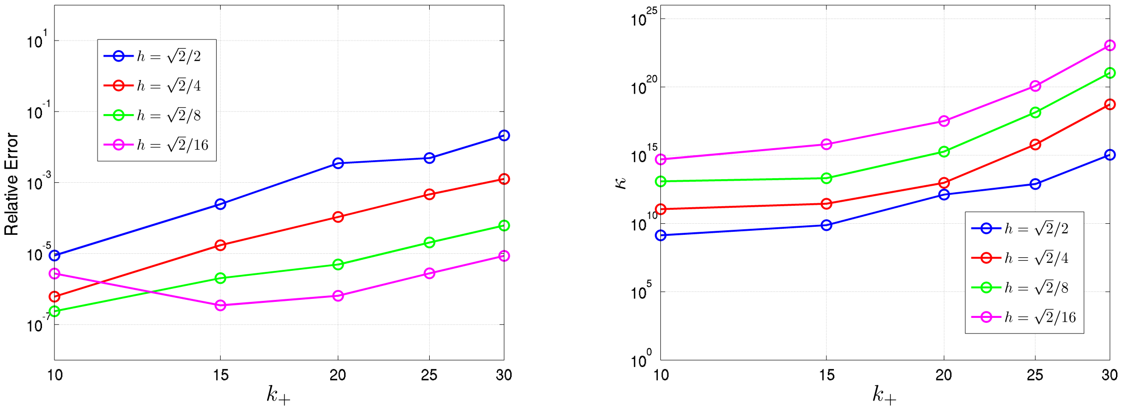

Figure 2.16: Relative error plotted with respect to the wave number k_+ (left) and condition number κ (right) for different values of the mesh size h. The PUFEM discretization is settled with $N_{\text{pw}} = 8$, and the exact solution is a propagative wave with incident angle $\beta = 2\pi/8 + 2\pi/9$.

Numerical comparison with a two-dimensional finite element method

Consider the computational domain depicted in Figure 2.17, which is split in two disjoint subdomains Ω_+ and Ω_-. The common boundary between both subdomains will be denoted by Σ (highlighted in pink in Figure 2.17). The upper boundary of the domain (highlighted in blue) is located on

$$\Gamma_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^2;\ x_1 \in \left(\frac{1}{4}, \frac{13}{40} \right),\ x_2 = \frac{23}{40} \right\}$$

. Only in this portion of the boundary the Neumann data will be not null, more precisely, $g_+ = 1$. On the rest of the boundary the Neumann data will be null.

The variable two-dimensional problem (2.40)-(2.45) is settled with the coupling interface Σ lying on $H = 0$ and the variable wave number k given by

$$k = \begin{cases} 32 & \text{in } \Omega_+, \\ 8 & \text{in } \Omega_-. \end{cases}$$

Under these conditions, the exact solution is not known in closed form. So, in order to check the accuracy of the PUFEM method, a numerical comparison will be made with respect to a standard piecewise linear finite element method. The finite element solution computed in a fine mesh will be used as a reference solution to be compared with those PUFEM approximations computed in coarse meshes.

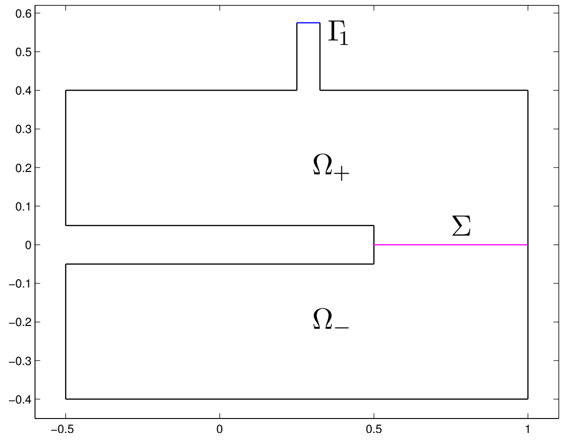

Figure 2.17: Polygonal domain considered for a two-dimensional Helmholtz problem with variable wave number and constant Neumann boundary conditions.

The approximate Lagrange $\mathbb{P}_1$ finite element solution u_{fe} is computed using a fine mesh, whose maximum diameter is 5.3×10^{-3} On the contrary, the two-dimensional PUFEM approximation is computed using a coarser mesh with larger maximum diameter h and hence with a reduced number of nodes. The relative difference in L^∞-norm between the finite element and the PUFEM approximation is computed as follows:

$$d_h(\boldsymbol{x}) = \frac{u_{\text{fe}}(\boldsymbol{x}) - u_h(\boldsymbol{x})}{\displaystyle\max_{j=1,\ldots,N_{\text{fe}}} |u_{\text{fe}}(\boldsymbol{x}_j)|}, \qquad \text{relative difference} = \|d_h\|_\infty = \max_{j=1,\ldots,N_{\text{fe}}} |d_h(\boldsymbol{x}_j)|, \tag{2.64}$$

where $\{\boldsymbol{x}_j\}_{j=1}^{N_{\text{fe}}}$ are the nodes of the coarse mesh used in the PUFEM discretization. It should be remarked that all the computations in this section have been carried out in double precision.

To approximate an exact solution that it is not a planewave, the PUFEM variable method behaves in a similar way as the finite element method, but using a much more coarse mesh and with a low number of planewaves N_{pw} chosen for the discretization. The behaviour of the PUFEM variable method in terms of the mesh size and in terms of the number of planewaves chosen for the PUFEM approximate solution will be studied on the following tables.

	$h = 1.7 \times 10^{-1}$		$h = 8.5 \times 10^{-2}$	
N_{pw}	$\|d_h\|_\infty$	κ	$\|d_h\|_\infty$	κ
4	2.40×10^{0}	3.3×10^{13}	2.06×10^{-1}	1.1×10^{15}
6	1.73×10^{-1}	8.1×10^{16}	1.76×10^{-1}	2.2×10^{18}
8	3.06×10^{-2}	3.6×10^{18}	2.07×10^{-2}	2.1×10^{21}
10	4.69×10^{-2}	3.4×10^{21}	2.43×10^{-2}	4.5×10^{23}
12	2.91×10^{-2}	7.9×10^{24}	2.75×10^{-2}	1.1×10^{26}
14	4.76×10^{-2}	7.4×10^{23}	4.64×10^{-2}	7.3×10^{25}
16	2.13×10^{-2}	6.0×10^{25}	5.24×10^{-2}	8.1×10^{27}

Table 2.1: Relative difference $\|d_h\|_\infty$ and condition number κ of the PUFEM discrete matrix for a variety of number of planewaves N_{pw} and two different meshes.

Table 2.1 shows the behaviour of the PUFEM variable method when the approximated PUFEM solution is computed with two meshes with maximum diameter $h = 1.7 \times 10^{-1}$ and $h = 8.5 \times 10^{-2}$, and for different choices of the number of planewaves N_{pw}. Although the condition number is larger than $\mathcal{O}(10^{13})$, the approximated PUFEM solution behaves reasonably similar to the finite element approximation even for the more coarse mesh and with $N_{\text{pw}} = 8$.

In Figures 2.18 and 2.19, the approximated PUFEM solution u_h and the relative difference d_h for two particular cases in Table 2.1 are illustrated. More precisely, for the discretization with $N_{\text{pw}} = 4$ and mesh size 8.5×10^{-2}, and for that one which uses $N_{\text{pw}} = 8$ and mesh size 1.7×10^{-1}. Only for plotting purposes, the relative difference has been evaluated in a fine structured mesh, with maximum diameter 5.3×10^{-3}.

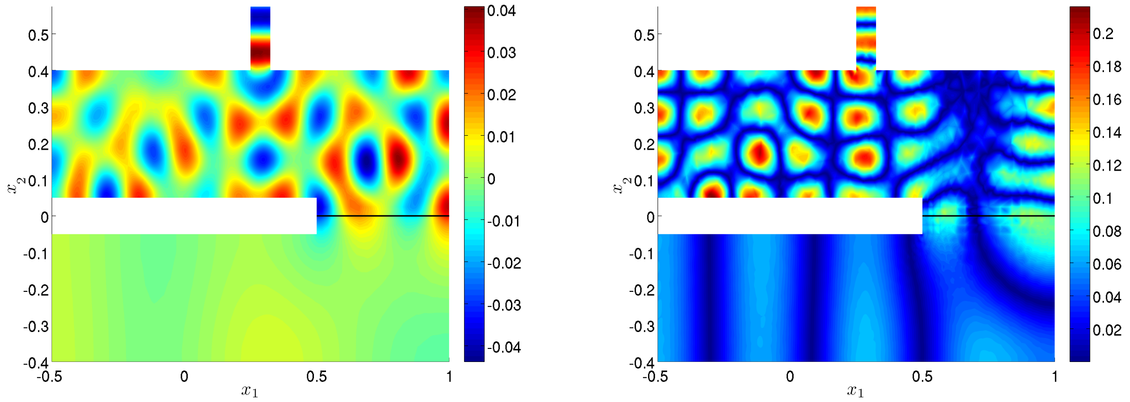

Figure 2.18: Real part of the approximated PUFEM solution u_h (left) and modulus of the relative difference d_h computed with the transmission-reflection PUFEM discretization using $N_{\mathrm{pw}} = 4$ and mesh size 8.5×10^{-2}.

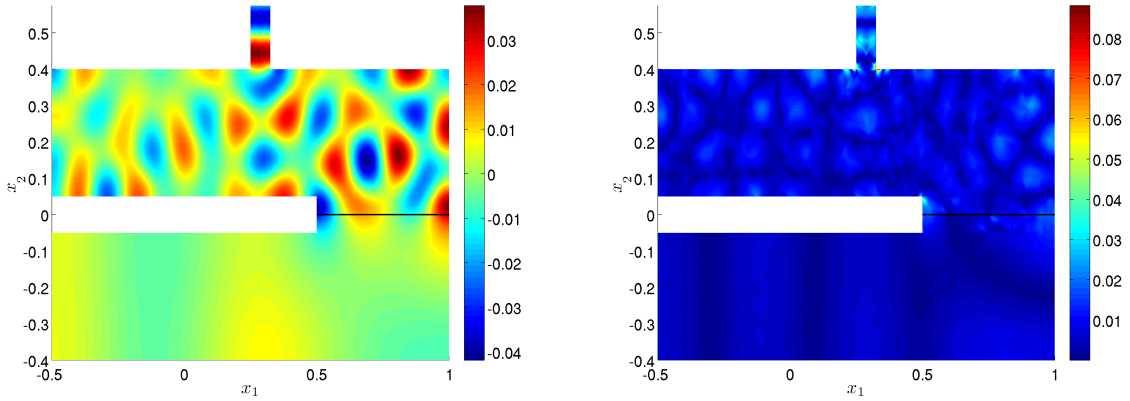

Figure 2.19: Real part of the approximated PUFEM solution u_h (left) and modulus of the relative difference d_h computed with the transmission-reflection PUFEM discretization using $N_{\mathrm{pw}} = 8$ and mesh size 1.7×10^{-1}.

If the number of planewaves at the PUFEM variable discretization is fixed at $N_{\mathrm{pw}} = 5$, and the mesh size varies between 1.7×10^{-1} and 2.1×10^{-2} (Table 2.2), the relative difference between the PUFEM approximate solution and the finite element approximation reaches a value around $\mathcal{O}(10^{-2})$. Figure 2.20 illustrates the PUFEM approximation u_h and the relative difference d_h for a particular case shown in Table 2.2, more precisely, that one corresponding to $N_{\mathrm{pw}} = 5$ and mesh size 1.7×10^{-1}. In this case, the relative difference has been computed in a fine structured mesh, with maximum diameter 5.3×10^{-3}.

h	e_h	κ
1.7×10^{-1}	1.03×10^{0}	3.9×10^{18}
8.5×10^{-2}	1.21×10^{-1}	2.4×10^{20}
4.3×10^{-2}	1.73×10^{-2}	2.1×10^{22}
2.1×10^{-2}	3.23×10^{-2}	3.0×10^{24}

Table 2.2: Relative difference $\|d_h\|_\infty$ and condition number κ for the transmission-reflection PUFEM approximation computed with $N_{\text{pw}} = 5$ and mesh size h.

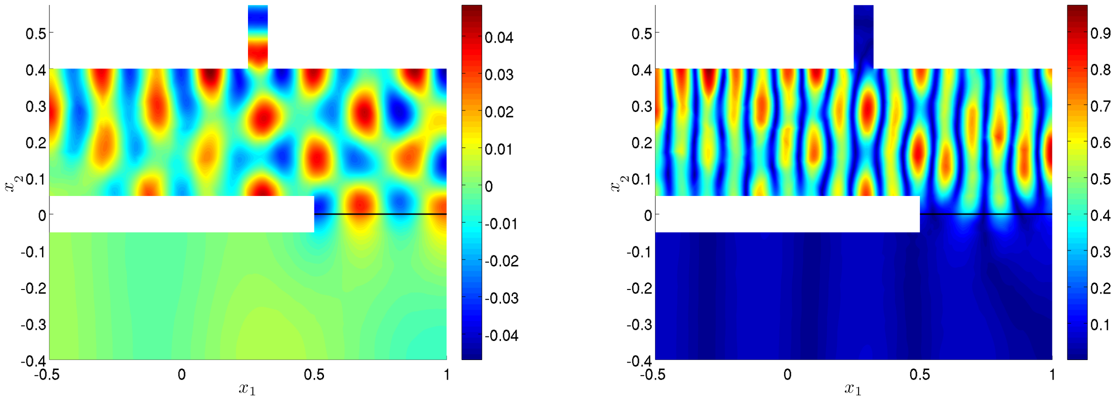

Figure 2.20: Real part of the approximated PUFEM solution u_h (left) and modulus of the relative difference d_h computed with the transmission-reflection PUFEM discretization using $N_{\text{pw}} = 5$ and mesh size 1.7×10^{-1}.

2.5 Conclusions

In this chapter, a novel PUFEM discretization for a one-dimensional Helmholtz problem in two media has been proposed. It has been found more accurate that other PUFEM discretizations described. A standard plane wave PUFEM discretization of a two-dimensional Helmholtz problem in one media has been described and some particular integration techniques have been introduced. Finally, a novel PUFEM discretization to approximate the solution of a two-dimensional problem in a layered media has been proposed. This method has into account the transmission and reflection that occurs at the interface. The accuracy of the method has been showed in some numerical results. Compared with a standard finite element method, this PUFEM discretization has relative errors of the same order but with much less degrees of freedom.

Chapter 3

A modal-based partition of unity finite element method for layered wave propagation problems

Contents

3.1 Introduction

In previous chapters, the partition of unity finite element method has been applied to approximate solutions of some one and two-dimensional Helmholtz problems, by multiplying the finite element basis functions by some plane wave functions. In this chapter, a different PUFEM discretization will be proposed, in order to solve some problems in bi-layered media that are used in non destructive testing.

The development of techniques to find cracks at the interface between two materials it is important to the early detection of defects in some structures like pipes with a coating. Ultrasonic testing and Foucault currents that propagate transversally to the interface are the more often used techniques. But they are both limited to cases where the source is close to the crack. The possibility of using Love waves was suggested recently (see [17]) to find a defect that is far from the source. It is basic for these detections with Love waves to know a priori the solution of the problem without a crack. In order to give a tool to approximate the solutions of these non destructive testing problems in bi-layered media without crack, a PUFEM method that involve Love waves will be proposed in this chapter.

The outline of this section is as follows: The model problem is presented in Section 3.2 as well as its variational formulation. In Section 3.3, the spectral analysis of the problem is described in detail. The PUFEM enrichment proposed, the discrete problem and its matrix description, and an analysis of the condition number are explained in Section 3.4. The Section 3.5 includes a wide battery of numerical tests in order to illustrate the behaviour of the proposed modal-based PUFEM method. Finally, in the last section, some conclusions are exposed.

3.2 Model problem

Under the assumptions of small perturbations of the displacement field and the stress tensor, the mechanical vibrations of bi-layered structures can be modelled with a linear elastic model. In particular, if the modelling interest is focused on the transverse displacement components and the geometry is invariant in one of the parallel directions to the interface of the layered structure, a two-dimensional time-dependent model problem can be assumed.

Let $\Omega_0 = \mathbb{R} \times (-a, H) \times \mathbb{R}$ an unbounded domain, and let $\Gamma_+^0 = \mathbb{R} \times \{H\} \times \mathbb{R}$ and $\Gamma_-^0 = \mathbb{R} \times \{-a\} \times \mathbb{R}$, where $a, H \in \mathbb{R}$. Let the strain tensor

$$\mathcal{E}(\boldsymbol{U}) = \frac{\nabla \boldsymbol{U} + \nabla \boldsymbol{U}^t}{2},$$

and the stress tensor

$$\sigma(\mathcal{E}(\boldsymbol{U})) = \lambda \mathrm{tr}(\mathcal{E}(\boldsymbol{U}))I + 2\mu\mathcal{E}(\boldsymbol{U}),$$

being λ and μ the Lamé coefficients. If $\boldsymbol{U} : \Omega_0 \to \mathbb{R}$ is the three-dimensional displacement, and ρ is the mass density, the elasticity model can be written

$$\rho\frac{\partial^2 \boldsymbol{U}}{\partial t^2} - \operatorname{div}\sigma(\mathcal{E}(\boldsymbol{U})) = \boldsymbol{F} \qquad \text{in } \Omega_0 \times [0,T], \tag{3.1}$$

$$\sigma(\mathcal{E}(\boldsymbol{U}))\boldsymbol{\nu} = \boldsymbol{G} \qquad \text{on } \Gamma_+^0 \cup \Gamma_-^0, \tag{3.2}$$

$$\boldsymbol{U}|_{t=0} = \boldsymbol{U}_0 \qquad \text{in } \ \Omega_0, \tag{3.3}$$

$$\left.\frac{\partial \boldsymbol{U}}{\partial t}\right|_{t=0} = \boldsymbol{V}_0 \qquad \text{in } \ \Omega_0, \tag{3.4}$$

where $\boldsymbol{\nu}$ is the unit normal vector along the boundary $\Gamma_+^0 \cup \Gamma_-^0$ and outwards the domain Ω_0, the source term is given by $\boldsymbol{F} : \Omega_0 \times [0,T] \to \mathbb{R}$ and the function $\boldsymbol{G}$ defines the boundary load on $\Gamma_+^0 \cup \Gamma_-^0$. The initial conditions involve the initial displacement $\boldsymbol{U}_0$ and the initial velocity field $\boldsymbol{V}_0$ at the initial time $t = 0$.

Assuming that the displacement $\boldsymbol{U}$, the velocity field, the source term $\boldsymbol{F}$ and the boundary load $\boldsymbol{G}$ can be written in terms of its transversal components as

$$\boldsymbol{U}(x_1,x_2,x_3) = U(x_1,x_2)\boldsymbol{e}_3, \quad \frac{\partial \boldsymbol{U}}{\partial t}(x_1,x_2,x_3) = V(x_1,x_2)\boldsymbol{e}_3,$$

$$\boldsymbol{F}(x_1,x_2,x_3) = F(x_1,x_2)\boldsymbol{e}_3, \quad \boldsymbol{G}(x_1,x_2,x_3) = G(x_1,x_2)\boldsymbol{e}_3,$$

considering the speed of sound of shear waves

$$c = \sqrt{\frac{\mu}{\rho}},$$

and abusing the notation in the boundary load, that is G/ρ, the problem (3.1)-(3.4) can be rewritten as follows (see [1] for more details).

Consider a two-dimensional domain, Ω, divided into two parts which represent two layers of different elastic nature, $\Omega = \Omega_+ \cup \Omega_-$. It is assumed that the boundary of Ω splits into a Neumann and a Robin boundary. In Figure 3.1 a particular case of this situation is depicted (in fact, this kind rectangular domains will be used for computational purposes):

$$\Omega = (0,1) \times (-a,H), \quad \Omega_+ = (0,1) \times (0,H), \text{ and } \Omega_- = (0,1) \times (-a,0).$$

In this case, it is assumed that the Neumann boundary conditions are placed over $\Gamma_+ \cup \Gamma_-$, and the Robin boundary conditions are placed over $\Gamma_e \cup \Gamma_s$, where

$$\Gamma_+ = [0,L] \times \{H\}, \quad \Gamma_- = [0,L] \times \{-a\}, \quad \Gamma_e = \{0\} \times [-a,H], \text{ and } \Gamma_s = \{L\} \times [-a,H].$$

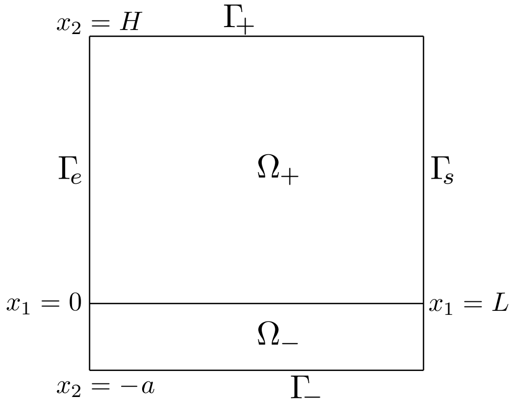

Figure 3.1: Computational domain of the bi-layered elastic material.

The transverse displacement $U : \Omega \times [0, T] \to \mathbb{R}$ satisfies the following governing equations:

$$\begin{aligned} \frac{\partial^2 U}{\partial t^2} - \operatorname{div}\left(c^2 \nabla U\right) &= F && \text{in } \Omega \times [0,T], && (3.5)\\ c^2 \frac{\partial U}{\partial \boldsymbol{\nu}} &= G && \text{on } \{\Gamma_+ \cup \Gamma_-\} \times [0,T], && (3.6)\\ \beta \frac{\partial U}{\partial t} + c \frac{\partial U}{\partial \boldsymbol{\nu}} &= R && \text{on } \{\Gamma_e \cup \Gamma_s\} \times [0,T], && (3.7)\\ U|_{t=0} &= U_0 && \text{in } \Omega, && (3.8)\\ \left.\frac{\partial U}{\partial t}\right|_{t=0} &= V_0 && \text{in } \Omega, && (3.9) \end{aligned}$$

where $\boldsymbol{\nu}$ is the unit normal vector along the boundary $\partial\Omega$ and outwards the domain Ω, the source term is given by $F : \Omega \times [0, T] \to \mathbb{R}$ and the functions G and R define the boundary loads on $\Gamma_+ \cup \Gamma_-$ and $\Gamma_e \cup \Gamma_s$, respectively. The initial conditions in (3.8)-(3.9) involve the initial displacement U_0 and the initial velocity field V_0 at the initial time $t = 0$. The parameter β in the boundary condition (3.7) could be null, to model boundary load conditions (Neumann conditions), or any other non-null value (for instance, $\beta = 1$ to reproduce a first-order absorbing boundary condition on the right and left boundaries of the computational domain).

Since the linear model (3.5)-(3.9) characterizes the mechanical behaviour of a bi-layered material, the transverse speed of sound c is defined as a piecewise-constant function given by

$$c(\boldsymbol{x}) = \begin{cases} c_+ & \text{if } \boldsymbol{x} \in \Omega_+, \\ c_- & \text{if } \boldsymbol{x} \in \Omega_-, \end{cases}$$

where $0 < c_- < c_+$. Due to the discontinuity of the speed of sound, the governing equations in (3.8)-(3.9) implicitly assume the following coupling conditions on the interface boundary

$\Gamma_I = \overline{\Omega}_+ \cap \overline{\Omega}_-$:

$$\begin{aligned} U|_{\Omega_-} &= U|_{\Omega_+} && \text{on } \Gamma_I \times [0,T], \\ c_-^2 \left.\frac{\partial U}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} &= c_+^2 \left.\frac{\partial U}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} && \text{on } \Gamma_I \times [0,T], \end{aligned}$$

where $\boldsymbol{\nu}$ is the unit normal vector outwards to Ω_-.

3.2.1 Time-harmonic problem

To study the time-harmonic behaviour of the mechanical system, it will be assumed that the source term and the boundary loads are time-harmonic functions, this is, formally it is supposed that there exist spatial-dependent functions f, g and r such that

$$F = \Re\left(fe^{-i\omega t}\right), \quad G = \Re\left(g\, e^{-i\omega t}\right), \quad R = \Re\left(re^{-i\omega t}\right),$$

being ω the angular frequency of the time-harmonic excitations. In this case, due to the linearity of the model problem (3.5)-(3.7), the long-time behaviour of the transverse displacement admits also a time-harmonic representation given by $U = \Re\left(ue^{-i\omega t}\right)$, being u the complex-valued displacement field at time-harmonic regime. Hence, the time-harmonic problem can be stated as follows: find the complex-valued transverse displacement field $u : \Omega \to \mathbb{C}$ such that it holds

$$\begin{aligned} -\omega^2 u - \operatorname{div}\left(c^2 \nabla u\right) &= f && \text{in } \Omega_+ \cup \Omega_-, && (3.10) \\ c^2 \frac{\partial u}{\partial \boldsymbol{\nu}} &= g && \text{on } \Gamma_+ \cup \Gamma_-, && (3.11) \\ -i\omega\beta u + c\frac{\partial u}{\partial \boldsymbol{\nu}} &= r && \text{on } \Gamma_e \cup \Gamma_s, && (3.12) \\ u|_{\Omega_-} &= u|_{\Omega_+} && \text{on } \Gamma_I, && (3.13) \\ c_-^2 \left.\frac{\partial u}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} &= c_+^2 \left.\frac{\partial u}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} && \text{on } \Gamma_I. && (3.14) \end{aligned}$$

We recall that here f is the complex-valued source term, g is the complex-valued boundary term associated to the Neumann boundary condition and r is the complex-valued term arising in the right-hand side of the Robin boundary condition.

3.2.2 Variational formulation

In order to derive the weak problem associated to the time-harmonic problem (3.10)-(3.14), a classical Green's formula plays a key role. More precisely, if Ω is a regular domain with piecewise Lipschitz boundary and it is assumed a complex-valued vector function $\boldsymbol{\varphi} \in \mathrm{H}(\operatorname{div}, \Omega)$, then the following Green's formula holds for any complex-valued function $\phi \in \mathrm{H}^1(\Omega)$:

$$\int_\Omega \operatorname{div} \boldsymbol{\varphi}\, \bar{\phi}\, \mathrm{d}\boldsymbol{x} + \int_\Omega \boldsymbol{\varphi} \cdot \nabla \bar{\phi}\, \mathrm{d}\boldsymbol{x} = \int_{\partial\Omega} (\boldsymbol{\varphi} \cdot \boldsymbol{\nu})\, \bar{\phi}\, \mathrm{d}\sigma,$$

where $\boldsymbol{\nu}$ is the unit normal vector on boundary $\partial\Omega$ and outwards Ω, and $\mathrm{d}\boldsymbol{x}$ and $\mathrm{d}\sigma$ indicates area integration in a two-dimensional domain and integration on a boundary, respectively. Taking into account the Green's formula stated above, if the Helmholtz equation (3.10) is multiplied by a test function $\phi \in \mathrm{H}^1(\Omega_\pm)$ and considering $\boldsymbol{\varphi} = c_\pm^2 \nabla u$, it is obtained, after integrating in Ω_+ and Ω_- separately,

$$\int_{\Omega_+} c_+^2 \nabla u \cdot \nabla \bar{\phi}\,\mathrm{d}\boldsymbol{x} - \omega^2 \int_{\Omega_+} u\bar{\phi}\,\mathrm{d}\boldsymbol{x} - \int_{\partial\Omega_+} c_+^2 \frac{\partial u}{\partial \boldsymbol{\nu}} \bar{\phi}\,\mathrm{d}\sigma = \int_{\Omega_+} f\bar{\phi}\,\mathrm{d}\boldsymbol{x},$$
$$\int_{\Omega_-} c_-^2 \nabla u \cdot \nabla \bar{\phi}\,\mathrm{d}\boldsymbol{x} - \omega^2 \int_{\Omega_-} u\bar{\phi}\,\mathrm{d}\boldsymbol{x} - \int_{\partial\Omega_-} c_-^2 \frac{\partial u}{\partial \boldsymbol{\nu}} \bar{\phi}\,\mathrm{d}\sigma = \int_{\Omega_-} f\bar{\phi}\,\mathrm{d}\boldsymbol{x}.$$

Now, if the boundary and coupling conditions are used to rewrite the boundary terms arising in the left hand side of the two equations written above and then both equations are added, the variational formulation of the time-harmonic problem is given by

$$\int_{\Omega} c^2 \nabla u \cdot \nabla \bar{\phi}\,\mathrm{d}\boldsymbol{x} - \omega^2 \int_{\Omega} u\bar{\phi}\,\mathrm{d}\boldsymbol{x} - i\omega\beta \int_{\Gamma_e \cup \Gamma_s} c\, u\bar{\phi}\,\mathrm{d}\sigma$$
$$= \int_{\Omega} f\bar{\phi}\,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_+ \cup \Gamma_-} g\bar{\phi}\,\mathrm{d}\sigma + \int_{\Gamma_e \cup \Gamma_s} c\, r\bar{\phi}\,\mathrm{d}\sigma.$$

Hence, taking into account an adequate functional space setting for the data and for the unknown field in order to obtain a well-posed weak problem, it is necessary to introduce the bounded sesquilinear form $A_\beta : \mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega) \to \mathbb{C}$ associated to the variational formulation and defined by

$$A_\beta(z, \phi) = \int_{\Omega} c^2 \nabla z \cdot \nabla \bar{\phi}\,\mathrm{d}\boldsymbol{x} - i\omega\beta \int_{\Gamma_e \cup \Gamma_s} c\, z\bar{\phi}\,\mathrm{d}\sigma, \qquad z, \phi \in \mathrm{H}^1(\Omega), \tag{3.15}$$

and given $f \in \mathrm{L}^2(\Omega)$, $cr \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, and $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$ it can be defined the linear functional $\ell : \mathrm{H}^1(\Omega) \to \mathbb{C}$ by

$$\ell(\phi) = \int_{\Omega} f\bar{\phi}\,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_+ \cup \Gamma_-} g\bar{\phi}\,\mathrm{d}\sigma + \int_{\Gamma_e \cup \Gamma_s} c\, r\bar{\phi}\,\mathrm{d}\sigma, \qquad \phi \in \mathrm{H}^1(\Omega), \tag{3.16}$$

where the integral notation has been abused since the second and third integrals should be understood as a duality pair product between elements in $\mathrm{H}^{\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$ and $\mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ spaces and its corresponding dual spaces $\mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$ and $\mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, respectively. For instance, on $\Gamma_e \cup \Gamma_s$ the boundary term should be read

$$\int_{\Gamma_e \cup \Gamma_s} \varphi\bar{\phi}\,\mathrm{d}\sigma = \langle \varphi, \phi \rangle_{\mathrm{H}^{-\frac{1}{2}}(\Gamma_e), \mathrm{H}^{\frac{1}{2}}(\Gamma_e)} + \langle \varphi, \phi \rangle_{\mathrm{H}^{-\frac{1}{2}}(\Gamma_s), \mathrm{H}^{\frac{1}{2}}(\Gamma_s)}.$$

***Remark* 3.2.1.** *The unusual condition $cr \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ cannot be replaced by $r \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ since c is a piecewise constant function with a jump discontinuity on Γ_e and*

Γ_s. In fact, if it is simply assumed $r \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ then the duality product $\int_{\Gamma_e} cr\bar{\phi}\,\mathrm{d}\sigma$ would be well defined only for those functions $\phi \in \mathrm{H}^1(\Omega)$ whose traces hold $c\phi|_{\Gamma_e} \in \mathrm{H}^{\frac{1}{2}}(\Gamma_e)$. In fact, a sufficient condition to ensure that $c\phi|_{\Gamma_e} \in \mathrm{H}^{\frac{1}{2}}(\Gamma_e)$ would be that $\phi|_{\Gamma_e \cap \partial\Omega_+} \in \mathrm{H}^{\frac{1}{2}}_{00}(\Gamma_e \cap \partial\Omega_+)$ and $\phi|_{\Gamma_e \cap \partial\Omega_-} \in \mathrm{H}^{\frac{1}{2}}_{00}(\Gamma_e \cap \partial\Omega_-)$ (see [20] for a detailed discussion).

Hence, the transverse displacement field is the solution of the following weak problem: Given the source term $f \in \mathrm{L}^2(\Omega)$, and the boundary loads $cr \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ and $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$, find $u \in \mathrm{H}^1(\Omega)$ such that

$$A_\beta(u,\phi) - \omega^2\langle u,\phi\rangle_{0,\Omega} = \ell(\phi) \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.17}$$

From the definition of the form A_β, it could be defined the associated linear operator $\mathcal{A}_\beta : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ as follows:
given $z \in \mathrm{L}^2(\Omega)$, its image $\mathcal{A}_\beta z$ is defined as the solution of the variational problem

$$A_\beta(\mathcal{A}_\beta z,\phi) + \langle \mathcal{A}_\beta z,\phi\rangle_{0,\Omega} = \langle z,\phi\rangle_{0,\Omega} \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.18}$$

Clearly, this operator $\mathcal{A}_\beta$ is bounded due to the coercivity of the sesquilinear form $A_\beta(\cdot,\cdot) + \langle\cdot,\cdot\rangle_{0,\Omega}$, but, in general, it is not self-adjoint since the sesquilinear form A_β does not satisfy $A_\beta(z,v) = \overline{A_\beta(v,z)}$. Only when $\beta = 0$, the sesquilinear form A_0 is hermitian, and so the bounded operator $\mathcal{A}_0$ is self-adjoint. In addition, since the solution of the variational problem (3.18) belongs to $\mathrm{H}^1(\Omega)$, the operator $\mathcal{A}_0 : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ is compact due to the compact inclusion of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^2(\Omega)$.

Consequently, for $\beta = 0$, the combination of the classical Fredholm's alternative theorem (see [44]) and the spectral decomposition of self-adjoint compact operators can be applied to deduce the existence and uniqueness of solution of the weak problem, except for an infinitely countable set of frequencies, which corresponds to the eigenvalues of finite multiplicity of the operator $\mathcal{A}_0$ (see Section 3.3.2 for further details).

However, for the other cases where $\beta > 0$, it is necessary to characterize the spectrum of a quadratic eigenvalue problem to derive such uniqueness and existence of solution in the source problem. With this aim, the following section is devoted to the study of the spectral analysis in these two cases separately (for $\beta = 0$ and $\beta > 0$).

3.3 Spectral analysis

The spectrum of operators associated to the source problem (3.17) is clearly of different nature attending to the value of β. If $\beta = 0$, classical results on linear eigenvalue problems lead to the conclusion that its spectrum is only given by the discrete spectrum (consisting on a sequence of isolated eigenvalues of finite algebraic multiplicity). On the contrary, in the case $\beta > 0$, the spectrum of the operator $\mathcal{A}_\beta$, associated to the quadratic eigenvalue problem related to (3.17), contains both the discrete spectrum but also a non-empty essential spectrum. In what follows, the characterization of both spectra are described in detail.

3.3.1 Spectral characterization for $\beta > 0$

The strong formulation of the spectral problem associated to the source problem (3.10)-(3.14) consists in finding the eigenpairs (w, λ), $w \neq 0$, such that

$$\lambda^2 u - \operatorname{div}\left(c^2 \nabla u\right) = 0 \qquad \text{in } \Omega_+ \cup \Omega_-, \tag{3.19}$$

$$c^2 \frac{\partial u}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \Gamma_+ \cup \Gamma_-, \tag{3.20}$$

$$\lambda \beta u + c \frac{\partial u}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \Gamma_e \cup \Gamma_s, \tag{3.21}$$

$$u|_{\Omega_-} = u|_{\Omega_+} \qquad \text{on } \Gamma_I, \tag{3.22}$$

$$c_-^2 \left.\frac{\partial u}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} = c_+^2 \left.\frac{\partial u}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} \qquad \text{on } \Gamma_I. \tag{3.23}$$

Obviously, due to the presence of terms multiplied by λ^2 and λ, problem (3.19)-(3.23) is an example of the so-called quadratic eigenvalue problems (see [7]). Consequently, the spectral analysis cannot be based on the classical results regarding linear eigenvalue problems. The analysis of the quadratic problem will be made following the guidelines described in [31] and, more precisely, it has been developed a similar mathematical analysis to the one described in [4], but adapted to the present functional framework.

In this way, instead of characterizing directly the spectrum of the quadratic problem (3.19)-(3.23), it will be analysed the spectrum of a more general spectral quadratic problem, which will be rewritten as a linear generalized eigenvalue problem. In what follows, each one of these steps are described in detail.

Firstly, the new general quadratic spectral problem is introduced. Let $\alpha \geq 0$ be a constant parameter, which is introduced in the original quadratic problem (3.19)-(3.23) on the Robin boundary condition as follows:

$$\lambda^2 u - \operatorname{div}\left(c^2 \nabla u\right) = 0 \qquad \text{in } \Omega_+ \cup \Omega_-, \tag{3.24}$$

$$c^2 \frac{\partial u}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \Gamma_+ \cup \Gamma_-, \tag{3.25}$$

$$(\alpha + \lambda \beta) u + c \frac{\partial u}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \Gamma_e \cup \Gamma_s, \tag{3.26}$$

$$u|_{\Omega_-} = u|_{\Omega_+} \qquad \text{on } \Gamma_I, \tag{3.27}$$

$$c_-^2 \left.\frac{\partial u}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} = c_+^2 \left.\frac{\partial u}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} \qquad \text{on } \Gamma_I. \tag{3.28}$$

Using similar arguments to those described in Section 3.2.2, the weak formulation of the perturbed quadratic spectral problem for α, $\beta \geq 0$ is stated as follows: Find $\lambda \in \mathbb{C}$ and $u \in \mathrm{H}^1(\Omega)$, $u \neq 0$, such that

$$\int_\Omega c^2 \nabla u \cdot \nabla \bar{\phi} \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma_e \cup \Gamma_s} \alpha c u \bar{\phi} \, \mathrm{d}\sigma + \lambda \int_{\Gamma_e \cup \Gamma_s} \beta c u \bar{\phi} \, \mathrm{d}\sigma + \lambda^2 \int_\Omega u \bar{\phi} \, \mathrm{d}\boldsymbol{x} = 0 \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.29}$$

It is easy to check that any non-null eigensolution ($\lambda \neq 0$) of (3.29) satisfies $\operatorname{Re}\lambda < 0$.

Proposition 3.3.1. *Let $\lambda \in \mathbb{C}$ and $0 \neq u \in \mathrm{H}^1(\Omega)$ solution of the quadratic problem* (3.29)*. If $\alpha \geq 0$ and $\beta > 0$ then either $\lambda = 0$ or $\operatorname{Re}\lambda < 0$.*

Proof. If the eigenpair (λ, u) is solution of the quadratic problem (3.29), taking into account $\phi = u$, it holds $A\lambda^2 + B\lambda + C = 0$ with

$$A = \int_\Omega |u|^2 \, \mathrm{d}\boldsymbol{x} > 0, \; B = \int_{\Gamma_e \cup \Gamma_s} c\beta |u|^2 \, \mathrm{d}\sigma \geq 0, \text{ and } C = \int_\Omega c^2 |\nabla u|^2 \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma_e \cup \Gamma_s} c\alpha |u|^2 \, \mathrm{d}\sigma \geq 0.$$

Firstly, notice that for $\alpha > 0$ it is easy to check that $\lambda \neq 0$, since C is an equivalent $\mathrm{H}^1(\Omega)$-norm of the function u (see Remark 3.3.3). Since $\lambda = (-B \pm \sqrt{B^2 - 4AC})/(2A)$, it is straightforward to check that if $C > 0$ and $B > 0$ then $\operatorname{Re}\lambda < 0$. To show it, two cases must be distinguished: (i) if $B^2 - 4AC \geq 0$ and since $C > 0$ and $A > 0$ then $B^2 - 4AC < B^2$ and so $-B + \sqrt{B^2 - 4AC} < 0$ (in addition, it is trivial that $-B - \sqrt{B^2 - 4AC} < 0$); (ii) if $B^2 - 4AC < 0$ then $\operatorname{Re}\lambda = -B/(2A) < 0$.

Finally, the limit cases $C = 0$ and $B = 0$ must be analysed. In the first case, (iii) if $C = 0$ then $\lambda = 0$ or $\lambda = -B/A$. For $B > 0$, the statement is proved and the case $B = 0$ again implies $\lambda = 0$. The last case, (iv) if $B = 0$ then $u = 0$ on $\Gamma_e \cup \Gamma_s$. From the spectral problem (3.29) using as testing functions $\phi \in \mathcal{C}^\infty(\bar{\Omega}) \subset \mathrm{H}^1(\Omega)$, it holds

$$\int_\Omega c^2 \nabla u \cdot \nabla \bar{\phi} \, \mathrm{d}\boldsymbol{x} + \lambda^2 \int_\Omega u \bar{\phi} \, \mathrm{d}\boldsymbol{x} = 0 \qquad \text{for all } \phi \in \mathcal{C}^\infty(\bar{\Omega}).$$

and hence, from a distributional sense, it leads to the conclusion that u is solution of the following problem:

$$\lambda^2 u - \operatorname{div}\left(c^2 \nabla u\right) = 0 \qquad \text{in } \Omega, \tag{3.30}$$

$$\frac{\partial u}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \partial\Omega, \tag{3.31}$$

$$u = 0 \qquad \text{on } \Gamma_e \cup \Gamma_s. \tag{3.32}$$

Consequently, u should be an eigenfunction of a second-order coercive elliptic operator (with L^∞-bounded coefficients in a two-dimensional smooth domain Ω) satisfying simultaneously homogeneous Dirichlet and Neumann boundary conditions on $\Gamma_s \cup \Gamma_e$. Using the Uniqueness Principle of Continuation for local Cauchy data (see for instance [45]), the unique solution of (3.30)-(3.32) is given by $u = 0$, what is not possible since $u \neq 0$ is an eigenfunction of problem (3.29). □

***Remark* 3.3.2.** *Despite the previous result for $\alpha \geq 0$ and $\beta > 0$, it cannot be guaranteed the existence and uniqueness of solution of the source problem associated to the eigenvalue problem* (3.24)*-*(3.28) *for complex-valued eigenvalues $\lambda = -i\omega$ with $\omega > 0$ due to the possible presence of accumulation points in its spectra (part of the essential spectrum of the associated operator) on the imaginary axis* $\operatorname{Re}\lambda = 0$. *This is the main reason because of a detailed analysis of the spectrum must be made.*

***Remark* 3.3.3.** *If $\alpha > 0$ then $\lambda = 0$ is not an eigenvalue of the quadratic problem* (3.29) *since*

$$\|u\|_\alpha^2 = \int_\Omega c^2 |\nabla u|^2 \,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_e \cup \Gamma_s} \alpha c |u|^2 \,\mathrm{d}\sigma \tag{3.33}$$

is an equivalent norm to the standard $\mathrm{H}^1(\Omega)$*-norm. In fact, the first integral in the expression above is the classical semi-norm in* $\mathrm{H}^1(\Omega)$ *and the second boundary integral is a continuous semi-norm in* $\mathrm{H}^1(\Omega)$*, whose value is null only for the null constant function (among the polynomials of order zero). Hence, from [3, Theorem 7.3.12], it is ensured that* $\|\cdot\|_\alpha$ *is an equivalent norm to the standard* $\mathrm{H}^1(\Omega)$*-norm.*

Using an standard procedure, the quadratic eigenvalue problem (3.29) can be rewritten as a linear eigenvalue problem doubling the size of the spectral problem. With this aim, there exist multiple ways to write such an equivalent linear eigenvalue problem. Following the ideas presented in [4], an equivalent linear eigenvalue problem is rewritten as follows: The linear eigenvalue problem is introduced by considering the new unknown function $v = \lambda u$. So the original weak formulation of the spectral problem can be stated: Find $\lambda \in \mathbb{C}$ and $(u, v) \in \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$ with $(u, v) \neq (0, 0)$ such that

$$\int_\Omega c^2 \nabla u \cdot \nabla \bar{\phi} \,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_e \cup \Gamma_s} \alpha c u \bar{\phi} \,\mathrm{d}\sigma = \lambda \left(- \int_{\Gamma_e \cup \Gamma_s} \beta c u \bar{\phi} \,\mathrm{d}\sigma - \int_\Omega v \bar{\phi} \,\mathrm{d}\boldsymbol{x} \right), \tag{3.34}$$

$$\int_\Omega v \bar{\psi} \,\mathrm{d}\boldsymbol{x} = \lambda \int_\Omega u \bar{\psi} \,\mathrm{d}\boldsymbol{x}, \tag{3.35}$$

for all $(\phi, \psi) \in \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$. Now, the next step consists in showing the equivalence between the linear spectral problem (3.34)-(3.35) and the perturbed quadratic problem (3.29).

Lemma 3.3.4. *For $\alpha \geq 0$, the pair $(\lambda, u) \in \mathbb{C} \times \mathrm{H}^1(\Omega)$ is an eigenpair of the quadratic problem* (3.29) *if and only if $(\lambda, (u, v)) \in \mathbb{C} \times (\mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega))$ is an eigenpair of the linear problem* (3.34)-(3.35)*.*

Proof. Firstly, if an eigensolution (λ, u) of (3.29) is fixed, v can be defined as $v = \lambda u \in \mathrm{H}^1(\Omega) \subset \mathrm{L}^2(\Omega)$ (it holds from (3.35) due to the density of $\mathrm{H}^1(\Omega)$ in $\mathrm{L}^2(\Omega)$). Then, if the expression of v is inserted in (3.34), the resulting expression coincides with (3.29), and hence u is also solution of (3.34). Reciprocally, if $(\lambda, (u, v))$ is an eigenpair of the linear problem (3.34)-(3.35), inserting (3.35) in (3.34), it holds straightforwardly the original quadratic problem (3.29). □

Due to this equivalence between the quadratic problem and this linear spectral problem, studying the eigensolutions of the quadratic problem will be equivalent to analysing the spectrum of the linear operator associated to (3.34)-(3.35). Such analysis will be different from $\alpha = 0$ and $\alpha > 0$. So, despite its similarities, for completeness in the description of the mathematical analysis, both cases will be considered separately.

Spectrum for α, $\beta > 0$

If $\alpha > 0$ then $\lambda = 0$ is not an eigenvalue of problem (3.34)-(3.35), since from (3.34) with $\phi = u$ it is deduced that $\|u\|_\alpha = 0$ and from (3.35) it holds $v = 0$, so $(u, v) = (0, 0)$. Now, the operators associated to the perturbed linear problem (3.34)-(3.35) will be introduced. For this purpose, consider the sesquilinear form $B : \mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega) \to \mathbb{C}$ defined by

$$B(u,\phi) = \int_\Omega c^2 \nabla u \cdot \nabla \bar{\phi} \, \mathrm{d}\boldsymbol{x} + \int_{\Gamma_e \cup \Gamma_s} \alpha c u \bar{\phi} \, \mathrm{d}\sigma \qquad \text{for all } u, \phi \in \mathrm{H}^1(\Omega),$$

and the sesquilinear forms $\tilde{B}$, $\tilde{D} : \mathrm{V} \times \mathrm{V} \to \mathbb{C}$ with $\mathrm{V} = \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$ given by

$$\tilde{B}((u,v),(\phi,\psi)) = B(u,\phi) + \int_\Omega v \bar{\psi} \, \mathrm{d}\boldsymbol{x}, \tag{3.36}$$

$$\tilde{D}((u,v),(\phi,\psi)) = -\int_{\Gamma_e \cup \Gamma_s} \beta c u \bar{\phi} \, \mathrm{d}\sigma - \int_\Omega v \bar{\phi} \, \mathrm{d}\boldsymbol{x} + \int_\Omega u \bar{\psi} \, \mathrm{d}\boldsymbol{x}, \tag{3.37}$$

for all (u, v), $(\phi, \psi) \in \mathrm{V}$. From the definitions written above, if $\alpha > 0$ then it is trivial to check that form B is $\mathrm{H}^1(\Omega)$-coercive (in fact, it is the inner product associated to the $\mathrm{H}^1(\Omega)$ norm $\| \cdot \|_\alpha$). Consequently, it also holds that $\tilde{B}$ is V-coercive. Now, let be the bounded linear operator $\mathcal{B} : \mathrm{V} \to \mathrm{V}$ defined such that $\mathcal{B}(f, g) = (u, v)$ if and only if

$$\tilde{B}((u,v),(\phi,\psi)) = \tilde{D}((f,g),(\phi,\psi)) \qquad \text{for all } (\phi,\psi) \in \mathrm{V}. \tag{3.38}$$

Taking into account this definition and those tests functions with $\phi = 0$, it is clear $v = f$ and hence u is solution of the variational problem

$$B(u,\phi) = -\int_{\Gamma_e \cup \Gamma_s} \beta c f \bar{\phi} \, \mathrm{d}\sigma - \int_\Omega g \bar{\psi} \, \mathrm{d}\boldsymbol{x} \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega),$$

which has an unique solution due the $\mathrm{H}^1(\Omega)$-coercivity of B (using the Lax-Milgram theorem). Hence, the operator $\mathcal{B}$ is well-defined.

Lemma 3.3.5. *For $\alpha > 0$, $(\mu, (u, v))$ is an eigenpair of $\mathcal{B}$ with $\mu \neq 0$ if and only if and $(1/\mu, (u, v))$ is an eigensolution of* (3.34)-(3.35)*.*

Proof. If $(\mu, (u, v))$ is an eigenpair of $\mathcal{B}$ with $\mu \neq 0$ then, from (3.38), it holds

$$\tilde{B}((u,v),(\phi,\psi)) = \frac{1}{\mu} \tilde{D}((u,v),(\phi,\psi)) \qquad \text{for all } (\phi,\psi) \in \mathrm{V}. \tag{3.39}$$

and hence similar arguments to those ones used to split the definition of each component of the image of $\mathcal{B}$ (taking as test functions those ones with $\phi = 0$) leads to $v = u/\mu \in \mathrm{H}^1(\Omega)$ and consequently (3.35) holds with $\lambda = 1/\mu$. Inserting the expression of v in (3.39), it again results (3.35) with $\lambda = 1/\mu$. Hence, $(1/\mu, (u, v))$ is an eigensolution of (3.34)-(3.35). Conversely, let $(1/\mu, (u, v))$ and eigensolution of (3.34)-(3.35), adding both equations it is obtained (3.39). □

Since $\mathcal{B}$ is a bounded operator in V, in general its spectrum $\sigma(\mathcal{B})$ could be formed by the discrete spectrum (set of isolated eigenvalues of finite algebraic multiplicity) and the essential spectrum (the set of eigenvalues of infinite algebraic multiplicity and the accumulation points of $\sigma(\mathcal{B})$). To characterize the spectrum of $\mathcal{B}$, the ideas introduced in [31] (and, in particular, in [4]) will be followed.

With this aim, two new bounded operators $\mathcal{B}_1 : \mathrm{H}^1(\Omega) \to \mathrm{H}^1(\Omega)$ and $\mathcal{B}_2 : \mathrm{L}^2(\Omega) \to \mathrm{H}^1(\Omega)$ are considered:

$$\mathcal{B}_1 f = u_1 \in \mathrm{H}^1(\Omega): \quad B(u_1, \phi) = \int_{\Gamma_e \cup \Gamma_s} \beta c f \bar{\phi} \, \mathrm{d}\sigma \quad \text{for all } \phi \in \mathrm{H}^1(\Omega), \tag{3.40}$$

$$\mathcal{B}_2 g = u_2 \in \mathrm{H}^1(\Omega): \quad B(u_2, \phi) = \int_{\Omega} g \bar{\phi} \, \mathrm{d}\sigma \quad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.41}$$

Both operators are well defined since the sesquilinear form is $\mathrm{H}^1(\Omega)$-coercive. In addition, since B is hermitian they are self-adjoint and due to the Lax-Milgram theorem, the solution of the variational problems (3.40)-(3.41) depends continuously on the data and consequently $\mathcal{B}_1$ and $\mathcal{B}_2$ are bounded operators in $\mathrm{H}^1(\Omega)$ and $\mathrm{L}^2(\Omega)$, respectively. Moreover, due to the regularity of the solution of the elliptic problem with piecewise constant coefficients in an smooth domain and $\mathrm{L}^2(\Omega)$ source data, then $u \in \mathrm{H}^{1+s}(\Omega)$ for some $s > 0$ (due to the presence of a cross-point on the boundary, see [20, 30] for further details). Hence, using the compact embedding of $\mathrm{H}^{1+s}(\Omega)$ in $\mathrm{H}^1(\Omega)$, it is clear that $\mathcal{B}_2$ is compact. In addition, $\mathcal{B}_2$ is positive definite with respect to the inner product $B(\cdot, \cdot)$ since

$$B(\mathcal{B}_2 g, g) = B(u_2, g) = \int_{\Omega} |g|^2 \, \mathrm{d}\boldsymbol{x} > 0 \qquad \text{for all } g \in \mathrm{H}^1(\Omega), \; g \neq 0.$$

Taking into account the definitions (3.40)-(3.41) of bounded operators $\mathcal{B}_1$ and $\mathcal{B}_2$, the operator $\mathcal{B}$ acting on V can be rewritten in terms of a block operator matrix acting on $\mathrm{V} = \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$ as follows:

$$\mathcal{B} = \begin{pmatrix} -\mathcal{B}_1 & -\mathcal{B}_2 \\ \mathcal{I} & 0 \end{pmatrix}. \tag{3.42}$$

Since $\mathcal{B}_2$ is compact and positive definite, it admits the computation of its square root operator $\mathcal{B}_2^{\frac{1}{2}}$ (by using the projections onto its spectral basis [29]). If the operators $\mathcal{S}$, $\mathcal{U}$ and $\mathcal{H}$ are defined by

$$\mathcal{S} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & \mathcal{B}_2^{\frac{1}{2}} \end{pmatrix}, \qquad \mathcal{U} = \begin{pmatrix} -\mathcal{B}_1 & -\mathcal{B}_2^{\frac{1}{2}} \\ \mathcal{I} & 0 \end{pmatrix}, \qquad \text{and} \qquad \mathcal{H} = \begin{pmatrix} -\mathcal{B}_1 & -\mathcal{B}_2^{\frac{1}{2}} \\ \mathcal{B}_2^{\frac{1}{2}} & 0 \end{pmatrix}.$$

It is straightforward to show that $\mathcal{SB} = \mathcal{HS}$, $\mathcal{B} = \mathcal{US}$, $\mathcal{H} = \mathcal{SU}$, and $\mathcal{UH} = \mathcal{BU}$. In addition, due to the positive definite character, $\mathcal{B}_2$ is invertible ($0 \notin \sigma(\mathcal{B}_2)$) and hence the operators $\mathcal{S}$, $\mathcal{U}$, and $\mathcal{H}$ are also invertible and the following result follows.

Proposition 3.3.6. *The spectrum of operator $\mathcal{B}$ and $\mathcal{H}$ coincides.*

Proof. See [4] for a detailed proof, where it is shown that the eigenvalues of $\mathcal{B}$ and $\mathcal{H}$ and their algebraic multiplicities coincide. The proof is based on the analysis of the Jordan chains associated to each eigenvalue. □

Since the operator $\mathcal{H}$ can be written as the sum of a self-adjoint operator $\mathcal{E}$ and a compact operator $\mathcal{C}$ as follows

$$\mathcal{H} = \mathcal{E} + \mathcal{C} \quad \text{with } \mathcal{E} = \begin{pmatrix} -\mathcal{B}_1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } \mathcal{C} = \begin{pmatrix} 0 & -\mathcal{B}_2^{\frac{1}{2}} \\ \mathcal{B}_2^{\frac{1}{2}} & 0 \end{pmatrix} \tag{3.43}$$

then it is trivial to check using the Weyl's theorem (see for instance [44]) that $\mathcal{H}$ and $\mathcal{B}$ share the same essential spectrum, and hence

$$\sigma_{\text{ess}}(\mathcal{H}) = \sigma_{\text{ess}}(\mathcal{B}) = \sigma_{\text{ess}}(\mathcal{B}_1) \cup \{0\}, \qquad \sigma_{\text{disc}}(\mathcal{H}) = \sigma(\mathcal{H}) \setminus \sigma_{\text{ess}}(\mathcal{H}).$$

Lemma 3.3.7. *For $\alpha > 0$, $\sigma_{\text{ess}}(\mathcal{B}_1) = \{0\}$.*

Proof. It is clear from the definition of operator $\mathcal{B}_1$ that $\lambda = 0$ is an eigenvalue of infinite algebraic multiplicity since any function in $v \in \mathrm{H}^1_{\Gamma_e \cup \Gamma_s}(\Omega)$ satisfies $\mathcal{B}_1 v = 0$. Straightforward computation also show that $\lambda = \beta/\alpha$ is one of its isolated eigenvalues of finite algebraic multiplicity whose eigenfunctions are the constant functions. Due to the self-adjoint character of $\mathcal{B}_1$, the rest of the eigenfunctions are in the orthogonal space of the direct sum of the subspace generated for both the eigenfunctions associated to both eigenvalues. Hence, its orthogonal complement will be computed with respect to the inner product B, this is

$$\mathrm{X} = \{w \in \mathrm{H}^1(\Omega) : \ B(w, \phi) = 0, \quad \text{for all } \phi \in \mathrm{H}^1_{\Gamma_e \cup \Gamma_s}(\Omega) \text{ and } \phi = 1\},$$

and it holds $\sigma(\mathcal{B}_1) = \{0, \beta/\alpha\} \cup \sigma(\mathcal{B}_1|_{\mathrm{X}})$. Since $B(w, \phi) = \int_\Omega c^2 \nabla w \cdot \nabla \phi \, \mathrm{d}\boldsymbol{x} = 0$ for all $\phi \in \mathrm{H}^1_{\Gamma_e \cup \Gamma_s}(\Omega)$, using $\phi \in \mathcal{C}^\infty(\bar{\Omega})$ such that $\phi|_{\Gamma_e \cup \Gamma_s} = 0$, any $w \in \mathrm{X}$ is solution of the following problem in the sense of the distributions:

$$-\text{div}(c^2 \nabla w) = 0 \qquad \text{in } \Omega, \tag{3.44}$$

$$w = g \qquad \text{on } \Gamma_e \cup \Gamma_s, \tag{3.45}$$

$$c^2 \frac{\partial w}{\partial \boldsymbol{n}} = 0 \qquad \text{on } \Gamma_+ \cup \Gamma_-, \tag{3.46}$$

where the Dirichlet boundary data $g \in \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ holds the orthogonality condition $\int_{\Gamma_e \cup \Gamma_s} cg \, \mathrm{d}\sigma = 0$. Let $\mathcal{T} : \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{H}^1(\Omega)$ be the operator defined by $w = \mathcal{T}g$ being w the solution of the variational problem associated to (3.44)-(3.46). The coercive character of this problem and the use of the Lax-Milgram theorem ensures that $\mathcal{T}$ is well-defined. In addition, a standard Green's formula shows that if $w = \mathcal{T}g$ then

$$\int_\Omega c^2 \nabla w \cdot \nabla \bar{\phi} \, \mathrm{d}\boldsymbol{x} = \int_{\Gamma_e \cup \Gamma_s} c^2 \frac{\partial w}{\partial \boldsymbol{n}} \bar{\phi} \, \mathrm{d}\sigma = 0 \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.47}$$

Analogously, since $\lambda = 0 \notin \sigma(\mathcal{B}_1|_{\mathrm{X}})$, the spectral problem $\mathcal{B}_1|_{\mathrm{X}} w = \lambda w$ admits the variational formulation

$$\int_\Omega c^2 \nabla u \cdot \nabla \bar{\phi} \, \mathrm{d}\boldsymbol{x} = \frac{\beta - \lambda\alpha}{\lambda} \int_{\Gamma_e \cup \Gamma_s} c u \bar{\phi} \, \mathrm{d}\sigma = 0 \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.48}$$

Comparing the variational terms in (3.47) and (3.48), it follows that the spectral problem restricted to the subspace X can be rewritten as

$$c \frac{\partial}{\partial \boldsymbol{n}} \mathcal{T} g = \frac{\beta - \lambda\alpha}{\lambda} g, \tag{3.49}$$

for those $g \in \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ satisfying $\int_{\Gamma_e \cup \Gamma_s} c g \, \mathrm{d}\sigma = 0$. Clearly, $\partial_{\boldsymbol{n}} \mathcal{T} : \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ is a linear bounded operator due to the boundedness character of the normal derivative of the solution of a second-order coercive elliptic problem stated in an smooth domain (see [20]). In addition, $\partial_{\boldsymbol{n}} \mathcal{T}$ has a bounded inverse $(\partial_{\boldsymbol{n}} \mathcal{T})^{-1} : \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ defined as follows: if $f \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ $(\partial_{\boldsymbol{n}} \mathcal{T})^{-1} f$ is defined as the trace on $\Gamma_e \cup \Gamma_s$ of the solution v of the problem

$$-\mathrm{div}(c^2 \nabla z) = 0 \qquad \text{in } \Omega, \tag{3.50}$$

$$c^2 \frac{\partial z}{\partial \boldsymbol{n}} = f \qquad \text{on } \Gamma_e \cup \Gamma_s, \tag{3.51}$$

$$c^2 \frac{\partial z}{\partial \boldsymbol{n}} = 0 \qquad \text{on } \Gamma_+ \cup \Gamma_-. \tag{3.52}$$

Due to the existence and uniqueness solution of the Laplace like problem, the inverse operator is well-defined and since the solution of this second-elliptic problem depends continuously with respect to the Neumann boundary data f, then $(\partial_{\boldsymbol{n}} \mathcal{T})^{-1}$ is a bounded operator. To check that it is actually the inverse of $\partial_{\boldsymbol{n}} \mathcal{T}$, it is enough to consider $f = \partial_{\boldsymbol{n}} w$ on $\Gamma_e \cup \Gamma_s$ in (3.50)-(3.52) being u the weak solution of (3.44)-(3.46) and consider the trace of z on $\Gamma_e \cup \Gamma_s$, i.e., $g = z|_{\Gamma_e \cup \Gamma_s}$, in (3.44)-(3.46) being w the weak solution of (3.50)-(3.52). In both cases, due to the existence and uniqueness of solutions of both Laplace problems, it is obtained that z (the solution of problem (3.50)-(3.52)) is solution of (3.44)-(3.46) and reciprocally, w (the solution of problem (3.44)-(3.46)) coincides with the solution of (3.50)-(3.52). This fact shows that $(\partial_{\boldsymbol{n}} \mathcal{T})(\partial_{\boldsymbol{n}} \mathcal{T})^{-1}$ is the identity in $\mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ and $(\partial_{\boldsymbol{n}} \mathcal{T})^{-1}(\partial_{\boldsymbol{n}} \mathcal{T})$ is the identity in $\mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$.

Finally, applying $(\partial_{\boldsymbol{n}} \mathcal{T})^{-1}$ in (3.49) and taking into account that $\lambda = \alpha/\beta \notin \sigma(\mathcal{B}_1|_{\mathrm{X}})$, the following spectral problem is obtained: find $(\mu, g) \in \mathbb{C} \times \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, $g \neq 0$ satisfying $\int_{\Gamma_e \cup \Gamma_s} c g \, \mathrm{d}\sigma = 0$, such that

$$\left((\partial_{\boldsymbol{n}} \mathcal{T})^{-1} \circ \mathrm{i}^* \right) \frac{g}{c} - \mu g = 0 \qquad \text{with } \mu = \frac{\lambda}{\beta - \lambda\alpha}, \tag{3.53}$$

where i^* is the dual continuous embedding operator from $\mathrm{L}^2(\Gamma_e \cup \Gamma_s)$ to $\mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$. Since the continuous embedding $\mathrm{i} : \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{L}^2(\Gamma_e \cup \Gamma_s)$ is compact (and using the Riesz identification of $\mathrm{L}^2(\Gamma_e \cup \Gamma_s)$ with its dual space) then also i^* is compact and consequently the composition operator $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1} \circ \mathrm{i}^*$ is compact. Hence, the spectral decomposition theorem for compact operators can be applied to show that there exists only an isolated countable discrete set of eigenvalues for the spectral problem (3.53). Hence, it is concluded that $\sigma(\mathcal{B}_1|_\mathrm{X})$ is discrete and so $\sigma_{\mathrm{ess}}(\mathcal{B}_1|_\mathrm{X}) = \emptyset$. Hence, using again that $\sigma(\mathcal{B}_1) = \{0, \beta/\alpha\} \cup \sigma(\mathcal{B}_1|_\mathrm{X})$, and since α/β is an isolated eigenvalue of finite multiplicity and only $\lambda = 0$ has infinite multiplicity, it is obtained that $\sigma_{\mathrm{ess}}(\mathcal{B}_1) = \{0\}$. □

In summary, the spectrum of $\mathcal{H}$ (and hence the spectrum of $\mathcal{B}$ too) is composed by the null value (the essential spectrum composed by an eigenvalue of infinite multiplicity) and a countable discrete (isolated) set of eigenvalues with finite algebraic multiplicity. Since Lemmas 3.3.4 and 3.3.10 ensure the equivalence between the spectrum $\{\mu_j\}_{j\in\mathbb{N}} \cup \{0\}$ of operator $\mathcal{B}$ and eigenvalues $\{\lambda_j = 1/\mu_j\}_{j\in\mathbb{N}}$ of the perturbed quadratic problem (3.29) with α, $\beta > 0$, then the eigensolutions of the perturbed quadratic problem are given by the union of a discrete set of pairs $\{(\lambda_j, u_j)\}_{j\in\mathbb{N}}$ with finite algebraic multiplicity, with $+\infty$ as the unique accumulation point, which could be formally associated to an eigenvalue of infinite algebraic multiplicity.

Finally, this spectral characterization leads to the following existence and uniqueness result on the solution of a perturbed source problem.

Theorem 3.3.8. *For any $\omega > 0$ and α, $\beta > 0$ and given $f \in \mathrm{L}^2(\Omega)$, $cr \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, and $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$, there exists an unique solution $u \in \mathrm{H}^1(\Omega)$ of the variational problem associated to the time-harmonic source problem* (3.10)-(3.14)*, where the first Robin term in* (3.12)*, $-i\omega\beta u$, has been replaced by $\alpha - i\omega\beta u$.*

Proof. Due the discrete character of the eigensolutions $\{(\lambda_j, u_j)\}_{j\in\mathbb{N}}$ of the associated spectral quadratic problem (3.29) and since $\operatorname{Re}\lambda_j > 0$ with an unique accumulation point at $+\infty$, all the complex values $\lambda = -i\omega$ with $\omega > 0$ do not belong to the spectrum of the quadratic problem. Now, taking into account that the variational problem (3.17) associated to the source problem (with the modification in the Robin boundary condition) can be rewritten as

$$\tilde{B}((u,v),(\phi,\psi)) + i\omega\tilde{D}((u,v),(\phi,\psi)) = \ell(\phi) \qquad \text{for all } (\phi,\psi) \in \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega),$$

with the linear form ℓ defined by (3.16), then the Fredholm's alternative applied to operator $\mathcal{B}$ ensures that the solution of the source problem exists and it is unique. □

Spectrum for $\alpha = 0$ and $\beta > 0$

If the quadratic problem is stated in the case of $\alpha = 0$ and $\beta > 0$ then $\lambda = 0$ is an eigenvalue since the constant functions satisfies the quadratic problem (3.19)-(3.23) with a null eigenvalue. In that case, following [4], the variational formulation of the quadratic problem cannot be stated in whole space $\mathrm{H}^1(\Omega)$ and it should be restricted to an subspace which should contain all the eigenfunctions of the quadratic problem different from the constant ones. To define such subspace, consider an eigenmode $((u,v),\lambda) \in (\mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)) \times \mathbb{C}$ with $\lambda \neq 0$ of the equivalent linear problem (3.34)-(3.35) with $\alpha = 0$. Taking into account that the $(u,v) = (1,0)$ is the eigenmode associated to $\lambda = 0$, if the test function is chosen as $(\phi,\psi) = (1,0)$ then from (3.34)-(3.35), it holds

$$\int_{\Gamma_e \cup \Gamma_s} \beta c u \,\mathrm{d}\sigma + \int_\Omega v \,\mathrm{d}\boldsymbol{x} = 0,$$

which is equivalent to the orthogonality condition

$$\langle u, 1\rangle_\beta + \langle v, 1\rangle_{\mathrm{L}^2(\Omega)} = 0, \tag{3.54}$$

being $\langle \cdot,\cdot\rangle_\beta$ the $\mathrm{H}^1(\Omega)$-inner product

$$\langle \phi,\psi\rangle_\beta = \int_\Omega c^2 \nabla\phi \cdot \nabla\bar{\psi} \,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_e \cup \Gamma_s} \beta c \phi \bar{\psi} \,\mathrm{d}\sigma. \tag{3.55}$$

This inner product is equivalent to the usual $\mathrm{H}^1(\Omega)$ inner product using the arguments described in Remark 3.3.3 (replacing there the role of α and β play in the norm $\|\cdot\|_\alpha$).

In conclusion, the spectral analysis of the linear eigenvalue problem (3.34)-(3.35) will be restricted to the subspace

$$\mathrm{V} = \{(\phi,\psi) \in \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega) : \ \langle \phi, 1\rangle_\beta + \langle \psi, 1\rangle_{\mathrm{L}^2(\Omega)} = 0\}. \tag{3.56}$$

Regarding the definition of V, it will be useful to rewrite this orthogonal complement of the one-dimensional space $\mathrm{K} = \langle 1\rangle \subset \mathrm{H}^1(\Omega)$ in terms of different subspaces, which take into account the orthogonal restriction of each component in $\mathrm{H}^1(\Omega)$ and $\mathrm{L}^2(\Omega)$ separately.

Lemma 3.3.9. *Let* $\mathrm{V} \subset \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$ *be defined by* (3.56) *and* $\mathrm{K} = \langle 1\rangle$*, it holds:*

(i) If $\mathrm{K}^\perp$ *is the orthogonal complement of* K *in* $\mathrm{L}^2(\Omega)$ *with respect to the standard* $\mathrm{L}^2(\Omega)$*-inner product, and* $|\Omega|$ *and* $|\Gamma_e \cup \Gamma_s|$ *denote the Lebesgue measures of domains* Ω *and* $\Gamma_e \cup \Gamma_s$ *using the measures* $\mathrm{d}\boldsymbol{x}$ *and* $\beta c\,\mathrm{d}\sigma$*, respectively, then*

$$\mathrm{V} = \left(\mathrm{K}^{\perp_\beta} \times \{0\}\right) \oplus \left(\{0\} \times \mathrm{K}^\perp\right) \oplus \langle (|\Omega|, -|\Gamma_e \cup \Gamma_s|)\rangle\,, \tag{3.57}$$

(ii) If $\mathrm{W} \subset \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$ *is defined by*

$$\mathrm{W} = \left\{ \left(-\frac{1}{|\Gamma_e \cup \Gamma_s|} \int_\Omega \psi \,\mathrm{d}\boldsymbol{x}, \psi \right) : \quad \psi \in \mathrm{L}^2(\Omega) \right\} \tag{3.58}$$

then $\mathrm{V} = \left(\mathrm{K}^{\perp_\beta} \times \{0\}\right) \oplus \mathrm{W}$.

Proof. (i) From the subspace defined in the right-hand side of (3.57), it is clear that each subspace $\mathrm{K}^{\perp_\beta} \times \{0\}$, $\{0\} \times \mathrm{K}^\perp$, and $(|\Omega|, -|\Gamma_e \cup \Gamma_s|)$ satisfy the orthogonal condition stated in (3.56). Hence, its direct sum is contained in V. From (3.56), it is also clear that $\mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega) = \mathrm{V} \oplus \langle(1,1)\rangle$, so to conclude (3.57), it is enough to check that it holds

$$\mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega) = \left(\mathrm{K}^{\perp_\beta} \times \{0\}\right) \oplus \left(\{0\} \times \mathrm{K}^{\perp}\right) \oplus \langle(|\Omega|, -|\Gamma_e \cup \Gamma_s|)\rangle \oplus \langle(1,1)\rangle\,. \tag{3.59}$$

Since the measures of domains Ω and $\Gamma_e \cup \Gamma_s$ are strictly positive, the direct sum of the two last subspaces in (3.59) generate the two-dimensional constant space $\{(\gamma,\delta) :\ \gamma,\delta \in \mathbb{C}\}$. Hence, since any $(\phi,\psi) \in \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega)$ can be rewritten as follows

$$\begin{aligned}(\phi,\psi) = \left(\phi - \frac{1}{|\Gamma_e \cup \Gamma_s|}\int_{\Gamma_e\cup\Gamma_s} c\beta\phi \,\mathrm{d}\sigma, 0\right) + \left(0, \psi - \frac{1}{|\Omega|}\int_\Omega \psi \,\mathrm{d}\boldsymbol{x}\right)& \\ + \left(\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_{\Gamma_e\cup\Gamma_s} c\beta\phi, \frac{1}{|\Omega|}\int_\Omega \psi \,\mathrm{d}\boldsymbol{x}\right)&,\end{aligned}$$

then (3.59) holds since the first term belongs to $\mathrm{K}^{\perp_\beta}$ and the second one is in $\mathrm{K}^{\perp}$. Consequently (3.57) is verified.

(ii) To show (3.58), it is equivalent to show $\mathrm{W} = \left(\{0\} \times \mathrm{K}^{\perp}\right) \oplus \langle(|\Omega|, -|\Gamma_e \cup \Gamma_s|)\rangle$ by using (3.57). For any $\psi \in \mathrm{L}^2(\Omega)$, it is clear that

$$\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \psi \,\mathrm{d}\boldsymbol{x}, \psi\right) = \underbrace{\left(0, \psi - \frac{1}{|\Omega|}\int_\Omega \psi \,\mathrm{d}\boldsymbol{x}\right)}_{\in\{0\}\times\mathrm{K}^\perp} - \left(\frac{1}{|\Gamma_e \cup \Gamma_s||\Omega|}\int_\Omega \psi \,\mathrm{d}\boldsymbol{x}\right)(|\Omega|, -|\Gamma_e \cup \Gamma_s|)$$

what leads to $\mathrm{W} \subseteq \left(\{0\} \times \mathrm{K}^{\perp}\right) \oplus \langle(|\Omega|, -|\Gamma_e \cup \Gamma_s|)\rangle$. Conversely, consider any element of $\left(\{0\} \times \mathrm{K}^{\perp}\right) \oplus \langle(|\Omega|, -|\Gamma_e \cup \Gamma_s|)\rangle$. It is given by $(\gamma|\Omega|, \varphi - \gamma|\Gamma_e \cup \Gamma_s|)$ with $\varphi \in \mathrm{K}^\perp$ and $\gamma \in \mathbb{C}$. It is obvious that $\psi = \varphi - \gamma|\Gamma_e \cup \Gamma_s|$ belongs to $\mathrm{L}^2(\Omega)$ and moreover

$$-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \psi \,\mathrm{d}\boldsymbol{x} = -\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega (\varphi - \gamma|\Gamma_e \cup \Gamma_s|) \,\mathrm{d}\boldsymbol{x} = \gamma|\Omega|,$$

since $\int_\Omega \varphi \,\mathrm{d}\boldsymbol{x} = 0$. Hence, $\left(\{0\} \times \mathrm{K}^{\perp}\right) \oplus \langle(|\Omega|, -|\Gamma_e \cup \Gamma_s|)\rangle \subseteq \mathrm{W}$ and consequently (3.58) is obtained. □

In the case of $\alpha = 0$, for those eigenfunctions in V all the eigenvalues of (3.34)-(3.35) are not null and satisfy $\operatorname{Re}\lambda < 0$. Now, the operators associated to the perturbed linear problem (3.34)-(3.35) will be introduced. For this purpose, consider the sesquilinear form $B : \mathrm{K}^{\perp_\beta} \times \mathrm{K}^{\perp_\beta} \to \mathbb{C}$ defined by

$$B(u,\phi) = \int_\Omega c^2 \nabla u \cdot \nabla\bar{\phi} \,\mathrm{d}\boldsymbol{x} \qquad \text{for all } u, \phi \in \mathrm{K}^{\perp_\beta},$$

and the sesquilinear forms $\tilde{B}$, $\tilde{D} : \mathrm{V} \times \mathrm{V} \to \mathbb{C}$ with $\mathrm{V} = \left(\mathrm{K}^{\perp_\beta} \times \{0\}\right) \oplus \mathrm{W}$ given by

$$\tilde{B}((u,v),(\phi,\psi)) = B(u,\phi) + \int_\Omega v\bar{\psi} \,\mathrm{d}\boldsymbol{x}, \tag{3.60}$$

$$\tilde{D}((u,v),(\phi,\psi)) = -\int_{\Gamma_e\cup\Gamma_s} \beta c u\bar{\phi} \,\mathrm{d}\sigma - \int_\Omega v\bar{\phi} \,\mathrm{d}\boldsymbol{x} + \int_\Omega u\bar{\psi} \,\mathrm{d}\boldsymbol{x}, \tag{3.61}$$

for all (u,v), $(\phi,\psi)\in \mathrm{V}$. From the definitions written above, it is trivial to check that form B is $\mathrm{K}^{\perp_\beta}$-coercive (since the sesquilinear form B coincides with the inner product $\langle\cdot,\cdot\rangle_\beta$ in $\mathrm{K}^{\perp_\beta}$). Consequently, it also holds that $\tilde{B}$ is V-coercive. Now, define the bounded linear operator $\mathcal{B}:\mathrm{V}\to\mathrm{V}$ such that $\mathcal{B}(f,g)=(u,v)$ if and only if

$$\tilde{B}((u,v),(\phi,\psi))=\tilde{D}((f,g),(\phi,\psi)) \qquad \text{for all } (\phi,\psi)\in \mathrm{V}. \tag{3.62}$$

Taking into account this definition and those tests functions with $\phi=0$, it is clear $v=f$ and hence u is solution of the variational problem

$$B(u,\phi)=-\int_{\Gamma_e\cup\Gamma_s}\beta c f\bar{\phi}\,\mathrm{d}\sigma-\int_\Omega g\bar{\psi}\,\mathrm{d}\boldsymbol{x} \qquad \text{for all } \phi\in \mathrm{K}^{\perp_\beta},$$

which has an unique solution due the $\mathrm{K}^{\perp_\beta}$-coercivity of B (using the Lax-Milgram theorem). Hence, the operator $\mathcal{B}$ is well-defined.

Lemma 3.3.10. *For $\alpha=0$ and $\beta>0$, $(\mu,(u,v))$ is an eigenpair of $\mathcal{B}$ with $\mu\neq 0$ if and only if and $(1/\mu,(u,v))$ is an eigensolution of* (3.34)-(3.35).

Proof. If $(\mu,(u,v))$ is an eigenpair of $\mathcal{B}$ with $\mu\neq 0$ then, from (3.62), it holds

$$\tilde{B}((u,v),(\phi,\psi))=\frac{1}{\mu}\tilde{D}((u,v),(\phi,\psi)) \qquad \text{for all } (\phi,\psi)\in \mathrm{V}. \tag{3.63}$$

and hence similar arguments to those ones used to split the definition of each component of the image of $\mathcal{B}$ (taking as test functions those ones in W) leads to $v=u/\mu\in \mathrm{L}^2(\Omega)$ and consequently (3.35) holds with $\lambda=1/\mu$. Inserting the expression of v in (3.63), it again results (3.35) with $\lambda=1/\mu$ for those test functions in V. To show that (3.34) is is satisfied for any test function in $\mathrm{H}^1(\Omega)\times \mathrm{L}^2(\Omega)$, it is easy to check (3.63) holds inserting $(\phi,\psi)=(1,1)$. In that case, since $(u,v)\in \mathrm{V}$, this choice of test functions leads to

$$\int_\Omega v\,\mathrm{d}\boldsymbol{x}=\frac{1}{\mu}\int_\Omega u\,\mathrm{d}\boldsymbol{x},$$

which is verified since $v=u/\mu$. Hence, $(1/\mu,(u,v))$ is an eigensolution of (3.34)-(3.35). Conversely, let $(1/\mu,(u,v))$ be an eigensolution of (3.34)-(3.35), adding both equations it is obtained (3.63) for any test function in $\mathrm{H}^1(\Omega)\times \mathrm{L}^2(\Omega)$, and, in particular, in V. To show that $(u,v)\in \mathrm{V}$ for $\mu\neq 0$, it is only necessary to recall that V has been defined by using the orthogonality condition (3.54), coming from (3.34)-(3.35) with test function $(1,0)$ (the eigenfunction associated to the null eigenvalue). □

Since $\mathcal{B}$ is a bounded operator in V, in general its spectrum $\sigma(\mathcal{B})$ could be formed by the discrete spectrum (set of isolated eigenvalues of finite algebraic multiplicity) and the essential spectrum (the set of eigenvalues of infinite algebraic multiplicity and the accumulation points of $\sigma(\mathcal{B})$). To characterize the spectrum of $\mathcal{B}$, the ideas introduced in [31] (and, in particular, in [4]) will be followed.

With this aim, the operator $\mathcal{B}$ will be written as a matrix of operators acting on $\mathrm{K}^{\perp_\beta} \times \mathrm{L}^2(\Omega)$. Notice that this identification between $\mathcal{B}$ and its matrix rewriting can be done since $\mathrm{V} = \left(\mathrm{K}^{\perp_\beta} \times \{0\}\right) \oplus \mathrm{W}$ and $\mathrm{K}^{\perp_\beta} \times \{0\}$ is isometric to $\mathrm{K}^{\perp_\beta}$ and W is isometric to $\mathrm{L}^2(\Omega)$. In that manner, four operators should be considered using different variational problems restricted to $\mathrm{K}^{\perp_\beta} \times \{0\}$ and W. The first problem defines the operator $\mathcal{B}_1$ as follows: given $f \in \mathrm{K}^{\perp_\beta}$, find $u_1 = \mathcal{B}_1 f \in \mathrm{K}^{\perp_\beta}$ such that it is the solution of the variational problem

$$\tilde{B}((u_1, 0), (\phi, 0)) = -\tilde{D}((f, 0), (\phi, 0)) \qquad \text{for all } \phi \in \mathrm{K}^{\perp_\beta}.$$

Due to the orthogonality condition on f and $B(\phi, 1) = 0$ for all $\phi \in \mathrm{H}^1(\Omega)$, the variational problem stated above is equivalent to find $u_1 \in \mathrm{K}^{\perp_\beta}$ such that

$$\mathcal{B}_1 f = u_1 \in \mathrm{K}^{\perp_\beta} : \qquad B(u_1, \phi) = \int_{\Gamma_e \cup \Gamma_s} \beta c f \bar{\phi} \, \mathrm{d}\sigma \quad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.64}$$

Since the sesquilinear form B is coercive in $\mathrm{K}^{\perp_\beta}$ then the operator $\mathcal{B}_1 : \mathrm{K}^{\perp_\beta} \to \mathrm{K}^{\perp_\beta}$ defined by $\mathcal{B}_1 f = u_1$ is well-posed and bounded. The second problem defines the operator $\mathcal{B}_2$ as follows: given $g \in \mathrm{L}^2(\Omega)$, find $u_2 = \mathcal{B}_2 f \in \mathrm{K}^{\perp_\beta}$ such that it is the solution of the variational problem

$$\tilde{B}((u_2, 0), (\phi, 0)) = -\tilde{D}\left(\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g \, \mathrm{d}\boldsymbol{x}, g\right), (\phi, 0)\right) \qquad \text{for all } \phi \in \mathrm{K}^{\perp_\beta}.$$

Due to the orthogonality condition satisfied by the test functions the null average functions in $\mathrm{L}^2(\Omega)$, the variational problem stated above is equivalent to find $u_2 \in \mathrm{K}^{\perp_\beta}$ such that

$$\mathcal{B}_2 g = u_2 \in \mathrm{K}^{\perp_\beta} : \qquad B(u_2, \phi) = \int_\Omega \left(g - \frac{1}{|\Omega|}\int_\Omega g \, \mathrm{d}\boldsymbol{x}\right) \bar{\phi} \, \mathrm{d}\boldsymbol{x} \quad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.65}$$

Again, since the sesquilinear form B is coercive in $\mathrm{K}^{\perp_\beta}$ then the operator $\mathcal{B}_2 : \mathrm{L}^2(\Omega) \to \mathrm{K}^{\perp_\beta}$ defined by $\mathcal{B}_2 g = u_2$ is well-posed and bounded. The third problem is stated as follows: given $f \in \mathrm{K}^{\perp_\beta}$, find $v_1 \in \mathrm{L}^2(\Omega)$ such that it is the solution of the variational problem

$$\begin{aligned} &\tilde{B}\left(\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega v_1 \, \mathrm{d}\boldsymbol{x}, v_1\right), \left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \psi \, \mathrm{d}\boldsymbol{x}, \psi\right)\right) \\ &\qquad\qquad = -\tilde{D}\left((f, 0), \left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \psi \, \mathrm{d}\boldsymbol{x}, \psi\right)\right) \qquad \text{for all } \psi \in \mathrm{L}^2(\Omega). \end{aligned}$$

Due to the orthogonality condition satisfied by f and since $B(1, 1) = 0$, the variational problem stated above is equivalent to find $v_1 \in \mathrm{L}^2(\Omega)$ such that

$$\int_\Omega v_1 \bar{\psi} \, \mathrm{d}\boldsymbol{x} = \int_\Omega f \bar{\psi} \, \mathrm{d}\boldsymbol{x} \quad \text{for all } \psi \in \mathrm{L}^2(\Omega),$$

which leads to the identity operator $\mathcal{I}$ since $v_1 = \mathcal{I} f = f$. Rigorously, $\mathcal{I}$ should be understood as the compact embedding of $\mathrm{K}^{\perp_\beta}$ in $\mathrm{L}^2(\Omega)$. However, since such compactness character will not be used throughout the rest of this section then the composition with this compact embedding will be omitted and denoted by the identity operator $\mathcal{I}$. Finally, the fourth problem is stated as follows: given $g \in \mathrm{L}^2(\Omega)$, find $v_2 \in \mathrm{L}^2(\Omega)$ such that it is the solution of the variational problem

$$\begin{aligned}\tilde{B}&\left(\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega v_2\,\mathrm{d}\boldsymbol{x}, v_2\right), \left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \psi\,\mathrm{d}\boldsymbol{x}, \psi\right)\right)\\ &= -\tilde{D}\left(\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}, g\right), \left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \psi\,\mathrm{d}\boldsymbol{x}, \psi\right)\right) \qquad \text{for all } \psi \in \mathrm{L}^2(\Omega).\end{aligned}$$

Since $B(1,1) = 0$, the variational problem stated above is equivalent to find $v_2 \in \mathrm{L}^2(\Omega)$ such that

$$\begin{aligned}\int_\Omega v_2\bar{\psi}\,\mathrm{d}\boldsymbol{x} = &-\int_{\Gamma_e\cup\Gamma_s} \beta c\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}\right)\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \bar{\psi}\,\mathrm{d}\boldsymbol{x}\right)\mathrm{d}\sigma\\ &-\int_\Omega g\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega \bar{\psi}\,\mathrm{d}\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x} + \int_\Omega\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}\right)\bar{\psi}\,\mathrm{d}\boldsymbol{x}\\ &= -\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}\int_\Omega \bar{\psi}\,\mathrm{d}\boldsymbol{x} = \int_\Omega\left(-\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}\right)\bar{\psi}\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

for all $\psi \in \mathrm{L}^2(\Omega)$ and hence the one-rank (and so compact) operator $\mathcal{B}_3 : \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ is a constant function given by

$$\mathcal{B}_3 g = v_2 \in \mathrm{L}^2(\Omega): \qquad v_2 = -\frac{1}{|\Gamma_e \cup \Gamma_s|}\int_\Omega g\,\mathrm{d}\boldsymbol{x} \tag{3.66}$$

In addition, since B is hermitian, operator $\mathcal{B}_1$ is self-adjoint. Due to the Lax-Milgram theorem, the solution of the variational problems (3.64)-(3.65) depends continuously on the data and consequently $\mathcal{B}_1$ and $\mathcal{B}_2$ are bounded operators. Moreover, due to the regularity of the solution of the elliptic problem with piecewise constant coefficients in an smooth domain and $\mathrm{L}^2(\Omega)$ source data, then the variational problem (3.65) stated in $\mathrm{H}^1(\Omega)$ admits the infinity family of solution of type $u_2 + \gamma$ with $\gamma \in \mathbb{C}$ and $u_2 \in \mathrm{H}^{1+s}(\Omega)$ for some $s > 0$ (due to the presence of a cross-point on the boundary, see [20, 30] for further details). Hence, using the compact embedding of $\mathrm{H}^{1+s}(\Omega)$ in $\mathrm{H}^1(\Omega)$ and fixing the value of γ with the orthogonality condition of belonging to $\mathrm{K}^{\perp_\beta}$, it is concluded that $\mathcal{B}_2$ is compact. In addition, $\mathcal{B}_2$ is positive definite with respect to the inner product $B(\cdot,\cdot)$ since

$$B(\mathcal{B}_2 g, g) = B\left(u_2, g - \frac{1}{|\Omega|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}\right) = \int_\Omega\left|g - \frac{1}{|\Omega|}\int_\Omega g\,\mathrm{d}\boldsymbol{x}\right|^2\mathrm{d}\boldsymbol{x} > 0,$$

for all $g \in \mathrm{K}^{\perp_\beta}$, $g \neq 0$.

Taking into account the definitions (3.64)-(3.66) of bounded operators $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{B}_3$, the operator $\mathcal{B}$ acting on V can be rewritten in terms of a block operator matrix acting on $\mathrm{V} = \mathrm{K}^{\perp_\beta} \times \mathrm{L}^2(\Omega)$ as follows:

$$\mathcal{B} = \begin{pmatrix} -\mathcal{B}_1 & -\mathcal{B}_2 \\ \mathcal{I} & \mathcal{B}_3 \end{pmatrix} \tag{3.67}$$

Since $\mathcal{B}_2$ is compact and positive definite, it admits the computation of its square root operator $\mathcal{B}_2^{\frac{1}{2}}$ (by using the projections onto its spectral basis [29]). If the operators $\mathcal{S}$, $\mathcal{U}$ and $\mathcal{H}$ are defined by

$$\mathcal{S} = \begin{pmatrix} -\mathcal{I} & 0 \\ 0 & \mathcal{B}_2^{\frac{1}{2}} \end{pmatrix}, \qquad \mathcal{U} = \begin{pmatrix} -\mathcal{B}_1 & -\mathcal{B}_2^{\frac{1}{2}} \\ \mathcal{I} & \mathcal{B}_3\mathcal{B}_2^{-\frac{1}{2}} \end{pmatrix}, \qquad \text{and} \quad \mathcal{H} = \begin{pmatrix} -\mathcal{B}_1 & -\mathcal{B}_2^{\frac{1}{2}} \\ \mathcal{B}_2^{\frac{1}{2}} & \mathcal{B}_2^{\frac{1}{2}}\mathcal{B}_3\mathcal{B}_2^{-\frac{1}{2}} \end{pmatrix}.$$

It is straightforward to show that $\mathcal{S}\mathcal{B} = \mathcal{H}\mathcal{S}$, $\mathcal{B} = \mathcal{U}\mathcal{S}$, $\mathcal{H} = \mathcal{S}\mathcal{U}$, and $\mathcal{U}\mathcal{H} = \mathcal{B}\mathcal{U}$. In addition, due to the positive definite character, $\mathcal{B}_2$ is invertible ($0 \notin \sigma(\mathcal{B}_2)$) and hence the operators $\mathcal{S}$, $\mathcal{U}$, and $\mathcal{H}$ are also invertible and the following result follows.

Proposition 3.3.11. *The spectrum of operator $\mathcal{B}$ and $\mathcal{H}$ coincides.*

Proof. See [4] for a detailed proof, where it is shown that the eigenvalues of $\mathcal{B}$ and $\mathcal{H}$ and their algebraic multiplicities coincide. The proof is based on the analysis of the Jordan chains associated to each eigenvalue. □

Since the operator $\mathcal{H}$ can be written as the sum of a self-adjoint operator $\mathcal{E}$ and a compact operator $\mathcal{C}$ as follows

$$\mathcal{H} = \mathcal{E} + \mathcal{C} \quad \text{with } \mathcal{E} = \begin{pmatrix} -\mathcal{B}_1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } \mathcal{C} = \begin{pmatrix} 0 & -\mathcal{B}_2^{\frac{1}{2}} \\ \mathcal{B}_2^{\frac{1}{2}} & \mathcal{B}_2^{\frac{1}{2}}\mathcal{B}_3\mathcal{B}_2^{-\frac{1}{2}} \end{pmatrix} \tag{3.68}$$

then it is trivial to check using the Weyl's theorem (see for instance [44]) that $\mathcal{H}$ and $\mathcal{B}$ share the same essential spectrum, and hence

$$\sigma_{\text{ess}}(\mathcal{H}) = \sigma_{\text{ess}}(\mathcal{B}) = \sigma_{\text{ess}}(\mathcal{B}_1) \cup \{0\}, \qquad \sigma_{\text{disc}}(\mathcal{H}) = \sigma(\mathcal{H}) \setminus \sigma_{\text{ess}}(\mathcal{H}).$$

Lemma 3.3.12. *For $\beta > 0$, $\sigma_{\text{ess}}(\mathcal{B}_1) = \{0\}$.*

Proof. It is clear from the definition of operator $\mathcal{B}_1$ that $\lambda = 0$ is an eigenvalue of infinite algebraic multiplicity since any function in $v \in \mathrm{H}^1_{\Gamma_e\cup\Gamma_s}(\Omega) \subset \mathrm{K}^{\perp_\beta}$ satisfies $\mathcal{B}_1 v = 0$. Due to the self-adjoint character of $\mathcal{B}_1$, the rest of the eigenfunctions are in the orthogonal space of the direct sum of the subspace generated for both the eigenfunctions associated to both eigenvalues. Hence, its orthogonal complement will be computed with respect to the inner product $\langle , \rangle_\beta$, this is

$$\mathrm{X} = \{w \in \mathrm{K}^{\perp_\beta} : \langle w, \phi\rangle_\beta = 0 \quad \text{for all } \phi \in \mathrm{H}^1_{\Gamma_e\cup\Gamma_s}(\Omega)\},$$

and it holds $\sigma(\mathcal{B}_1) = \{0\} \cup \sigma(\mathcal{B}_1|_{\mathrm{X}})$. Since $\langle w, \phi\rangle_\beta = B(w,\phi) = \int_\Omega c^2 \nabla w \cdot \nabla\phi \,\mathrm{d}\boldsymbol{x} = 0$ for all $\phi \in \mathrm{H}^1_{\Gamma_e\cup\Gamma_s}(\Omega)$, using $\phi \in \mathcal{C}^\infty(\bar{\Omega})$ such that $\phi|_{\Gamma_e\cup\Gamma_s} = 0$, any $w \in \mathrm{X}$ is solution of the following problem in the sense of the distributions:

$$-\mathrm{div}(c^2\nabla w) = 0 \qquad \text{in } \Omega, \tag{3.69}$$

$$w = g \qquad \text{on } \Gamma_e \cup \Gamma_s, \tag{3.70}$$

$$c^2\frac{\partial w}{\partial \boldsymbol{n}} = 0 \qquad \text{on } \Gamma_+ \cup \Gamma_-, \tag{3.71}$$

where the Dirichlet boundary data $g \in \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ holds the orthogonality condition $\int_{\Gamma_e\cup\Gamma_s} cg\, \mathrm{d}\sigma = 0$. Let $\mathcal{T} : \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{H}^1(\Omega)$ be the operator defined by $w = \mathcal{T}g$ being w the solution of the variational problem associated to (3.69)-(3.71). The coercive character of this problem and the use of the Lax-Milgram theorem ensures that $\mathcal{T}$ is well-defined. In addition, a standard Green's formula shows that if $w = \mathcal{T}g$ then

$$\int_\Omega c^2 \nabla w \cdot \nabla\bar{\phi}\, \mathrm{d}\boldsymbol{x} = \int_{\Gamma_e\cup\Gamma_s} c^2 \frac{\partial w}{\partial \boldsymbol{n}} \bar{\phi}\, \mathrm{d}\sigma = 0 \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.72}$$

Analogously, since $\lambda = 0 \notin \sigma(\mathcal{B}_1|_{\mathrm{X}})$, the spectral problem $\mathcal{B}_1|_{\mathrm{X}} w = \lambda w$ admits the variational formulation

$$\int_\Omega c^2 \nabla u \cdot \nabla\bar{\phi}\, \mathrm{d}\boldsymbol{x} = \frac{\beta}{\lambda} \int_{\Gamma_e\cup\Gamma_s} cu\bar{\phi}\, \mathrm{d}\sigma = 0 \qquad \text{for all } \phi \in \mathrm{H}^1(\Omega). \tag{3.73}$$

Comparing the variational terms in (3.72) and (3.73), it follows that the spectral problem restricted to the subspace X can be rewritten as

$$c\frac{\partial}{\partial \boldsymbol{n}}\mathcal{T}g = \frac{\beta - \lambda\alpha}{\lambda} g, \tag{3.74}$$

for those $g \in \mathrm{H}^{\frac{1}{2}}(\Gamma_e\cup\Gamma_s)$ satisfying $\int_{\Gamma_e\cup\Gamma_s} cg\, \mathrm{d}\sigma = 0$. Clearly, $\partial_{\boldsymbol{n}}\mathcal{T} : \mathrm{H}^{\frac{1}{2}}(\Gamma_e\cup\Gamma_s) \to \mathrm{H}^{-\frac{1}{2}}(\Gamma_e\cup \Gamma_s)$ is a linear bounded operator due to the boundedness character of the normal derivative of the solution of a second-order coercive elliptic problem stated in an smooth domain (see [20]). In addition, $\partial_{\boldsymbol{n}}\mathcal{T}$ has a bounded inverse $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1} : \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ defined as follows: if $f \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1}f$ is defined as the trace on $\Gamma_e \cup \Gamma_s$ of the solution v of the problem

$$-\mathrm{div}(c^2\nabla z) = 0 \qquad \text{in } \Omega, \tag{3.75}$$

$$c^2\frac{\partial z}{\partial \boldsymbol{n}} = f \qquad \text{on } \Gamma_e \cup \Gamma_s, \tag{3.76}$$

$$c^2\frac{\partial z}{\partial \boldsymbol{n}} = 0 \qquad \text{on } \Gamma_+ \cup \Gamma_-. \tag{3.77}$$

Due to the existence and uniqueness solution of the Laplace like problem, the inverse operator is well-defined and since the solution of this second-elliptic problem depends continuously with respect to the Neumann boundary data f, then $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1}$ is a bounded operator. To check that it is actually the inverse of $\partial_{\boldsymbol{n}}\mathcal{T}$, it is enough to consider $f = \partial_{\boldsymbol{n}}w$ on $\Gamma_e \cup \Gamma_s$ in (3.75)-(3.77) being u the weak solution of (3.69)-(3.71) and consider the trace of z on $\Gamma_e \cup \Gamma_s$, i.e., $g = z|_{\Gamma_e\cup\Gamma_s}$, in (3.69)-(3.71) being w the weak solution of (3.75)-(3.77). In both cases, due to the existence and uniqueness of solutions of both Laplace problems, it is obtained that z (the solution of problem (3.75)-(3.77)) is solution of (3.69)-(3.71) and reciprocally, w (the solution of problem (3.69)-(3.71)) coincides with the solution of (3.75)-(3.77). This fact shows that $(\partial_{\boldsymbol{n}}\mathcal{T})(\partial_{\boldsymbol{n}}\mathcal{T})^{-1}$ is the identity in $\mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ and $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1}(\partial_{\boldsymbol{n}}\mathcal{T})$ is the identity in $\mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$.

Finally, applying $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1}$ in (3.74) and taking into account that $\beta > 0$, the following spectral problem is obtained: find $(\mu, g) \in \mathbb{C}\times \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, $g \neq 0$ satisfying $\int_{\Gamma_e \cup \Gamma_s} cg \, \mathrm{d}\sigma = 0$, such that

$$\left((\partial_{\boldsymbol{n}}\mathcal{T})^{-1} \circ \mathrm{i}^*\right) \frac{g}{c} - \mu g = 0 \qquad \text{with } \mu = \frac{\lambda}{\beta}, \tag{3.78}$$

where i^* is the dual continuous embedding operator from $\mathrm{L}^2(\Gamma_e \cup \Gamma_s)$ to $\mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$. Since the continuous embedding $\mathrm{i} : \mathrm{H}^{\frac{1}{2}}(\Gamma_e \cup \Gamma_s) \to \mathrm{L}^2(\Gamma_e \cup \Gamma_s)$ is compact (and using the Riesz identification of $\mathrm{L}^2(\Gamma_e \cup \Gamma_s)$ with its dual space) then also i^* is compact and consequently the composition operator $(\partial_{\boldsymbol{n}}\mathcal{T})^{-1} \circ \mathrm{i}^*$ is compact. Hence, the spectral decomposition theorem for compact operators can be applied to show that there exists only an isolated countable discrete set of eigenvalues for the spectral problem (3.78). Hence, it is concluded that $\sigma(\mathcal{B}_1|_\mathrm{X})$ is discrete and so $\sigma_{\mathrm{ess}}(\mathcal{B}_1|_\mathrm{X}) = \emptyset$. Hence, using again that $\sigma(\mathcal{B}_1) = \{0\} \cup \sigma(\mathcal{B}_1|_\mathrm{X})$, it is obtained that $\sigma_{\mathrm{ess}}(\mathcal{B}_1) = \{0\}$. □

In summary, the spectrum of $\mathcal{H}$ (and hence the spectrum of $\mathcal{B}$ too) is composed by the null value (the essential spectrum composed by an eigenvalue of infinite multiplicity) and a countable discrete (isolated) set of eigenvalues with finite algebraic multiplicity. Since Lemmas 3.3.4 and 3.3.10 ensure the equivalence between the spectrum $\{\mu_j\}_{j\in\mathbb{N}} \cup \{0\}$ of operator $\mathcal{B}$ and eigenvalues $\{\lambda_j = 1/\mu_j\}_{j\in\mathbb{N}}$ of the perturbed quadratic problem (3.29) with $\beta > 0$, then the eigensolutions of the quadratic problem are given by the union of a discrete set of pairs $\{(\lambda_j, u_j)\}_{j\in\mathbb{N}}$ with finite algebraic multiplicity, with $+\infty$ as the unique accumulation point, which could be formally associated to an eigenvalue of infinite algebraic multiplicity.

Finally, this spectral characterization leads to the following existence and uniqueness result on the solution of a perturbed source problem.

Theorem 3.3.13. *For any $\omega > 0$ and $\beta > 0$ and given $f \in \mathrm{L}^2(\Omega)$, $cr \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, and $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$, there exists an unique solution $u \in \mathrm{H}^1(\Omega)$ of the variational problem associated to the time-harmonic source problem* (3.10)-(3.14)*.*

Proof. Due the discrete character of the eigensolutions $\{(\lambda_j, u_j)\}_{j\in\mathbb{N}}$ of the associated spectral quadratic problem (3.29) and since $\operatorname{Re}\lambda_j > 0$ with an unique accumulation point at $+\infty$, all the complex values $\lambda = -i\omega$ with $\omega > 0$ do not belong to the spectrum of the quadratic problem. Now, taking into account that the variational problem (3.17) associated to the source problem can be rewritten as

$$\tilde{B}((u, v), (\phi, \psi)) + i\omega\tilde{D}((u, v), (\phi, \psi)) = \ell(\phi) \qquad \text{for all } (\phi, \psi) \in \mathrm{H}^1(\Omega) \times \mathrm{L}^2(\Omega),$$

with the linear form ℓ defined by (3.16), then the Fredholm's alternative applied to operator $\mathcal{B}$ ensures that the solution of the source problem exists and it is unique. □

3.3.2 Spectral characterization for $\beta = 0$

Since the proposed methodology of a modal-based PUFEM method requires the combination of a spectral basis with the partition of unity finite element method, a complete

Hilbert basis should be considered. The most suitable candidate for such spectral basis can be computed from the time-harmonic model chosen $\beta = 0$. In this case, since $\mathcal{A}_0$ is self-adjoint and compact in $\mathrm{L}^2(\Omega)$, it is guaranteed that the spectral problem associated to (3.17) defines a complete Hilbert basis in $\mathrm{L}^2(\Omega)$.

More precisely, the strong formulation of the spectral problem consists in finding the eigenpairs (w, λ), $w \neq 0$, such that

$$\lambda w - \operatorname{div}\left(c^2 \nabla w\right) = 0 \qquad \text{in } \Omega, \tag{3.79}$$

$$\frac{\partial w}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \partial\Omega, \tag{3.80}$$

$$w|_{\Omega_-} = w|_{\Omega_+} \qquad \text{on } \Gamma_I, \tag{3.81}$$

$$c_-^2 \left.\frac{\partial w}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} = c_+^2 \left.\frac{\partial w}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} \qquad \text{on } \Gamma_I. \tag{3.82}$$

and requiring the standard normalization in L^2-norm, i.e., $\|w\|_{0,\Omega} = 1$. As it is also described in the section above, the weak formulation of the spectral problem relies on the adequate functional setting and the self-adjoint compact operator $\mathcal{A}_0$. More precisely, it is stated as follows: Find the eigenpairs $(w, \lambda) \in \mathrm{L}^2(\Omega) \times \mathbb{C}$, $w \neq 0$, such that

$$\mathcal{A}_0 w = \lambda w. \tag{3.83}$$

The standard spectral theory (see the theorem of spectral decomposition of self-adjoint compact operators [44] poses that spectral problem (3.83) has an infinite countable family of eigenpair solutions $\{(w_n, \lambda_n)\}_{n\in\mathbb{N}} \subset \mathrm{L}^2(\Omega) \times \mathbb{R}$ such that $\lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \ldots \lambda_n \leq \ldots < +\infty$, and $\{w_n\}_{n\in\mathbb{N}}$ is a Hilbert basis in $\mathrm{L}^2(\Omega)$. Furthermore, the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ tends to infinity and the multiplicity of each eigenvalue is finite.

Theorem 3.3.14. *For any $\omega > 0$ and $\beta = 0$ and given $f \in \mathrm{L}^2(\Omega)$, $cr \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$, and $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$, there exists an unique solution $u \in \mathrm{H}^1(\Omega)$ of the variational problem associated to the time-harmonic source problem* (3.10)-(3.14)*, except for a infinity numerable set of resonance frequencies $\{\omega_j\}_{j\in\mathbb{N}}$, which tends to infinity.*

Proof. Taking into account the definition of $\mathcal{A}_0$ and the variational problem (3.17) associated to the source problem, the Fredholm's alternatives theorem ensures that the operator $\mathcal{A}_0 - \lambda\mathcal{I}$ is invertible for those frequencies such that $\lambda = \omega^2 + 1 \neq \lambda_j$ for all $j \in \mathbb{N}$. Since $\lambda_j \in \mathbb{R}$ with an unique accumulation point at $+\infty$, the uniqueness and existence of solution for the source problem is guaranteed except for a infinity numerable set of frequencies $\omega_j = \sqrt{\lambda_j - 1}$, $j \in \mathbb{N}$. □

Modal decomposition for $\beta = 0$

Obviously, in the case $\beta = 0$, the solution of the source problem (3.10)-(3.14) can be written in terms of the eigenfunctions w_n. To deduce an explicit series representation of the solution of the source problem in terms of the Hilbert basis, first let us consider a lift function $z \in \mathrm{H}^1(\Omega)$ from the boundary data g and r. More precisely, since $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ and $r \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$, the existence of a continuous lift operator from this boundary data is ensured by solving the Laplace problem with this boundary data on the smooth domain Ω: the lift function satisfies

$$\begin{aligned}
-\mathrm{div}\left(c^2 \nabla z\right) &= 0 && \text{in } \Omega, && (3.84)\\
c^2 \frac{\partial z}{\partial \boldsymbol{\nu}} &= g && \text{on } \Gamma_+ \cup \Gamma_-, && (3.85)\\
c \frac{\partial z}{\partial \boldsymbol{\nu}} &= r && \text{on } \Gamma_e \cup \Gamma_s, && (3.86)\\
z|_{\Omega_-} &= z|_{\Omega_+} && \text{on } \Gamma_I, && (3.87)\\
c_+^2 \left.\frac{\partial z}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} &= c_-^2 \left.\frac{\partial z}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} && \text{on } \Gamma_I. && (3.88)
\end{aligned}$$

In fact, from the coercivity of the weak formulation of the Laplace problem, it is straightforward to show that there exists a constant $C > 0$ only dependent on Ω and c such that

$$\|z\|_{1,\Omega} \leq C \left(\|g\|_{\mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)} + \|r\|_{\mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)} \right).$$

Once the lift function z has been computed, the solution of the problem (3.10)-(3.14) is translated in order to obtain a new rewriting version of the original source problem but now with homogeneous boundary conditions. With that purpose, a new unknown field y is defined such that $u = y + z$. In this manner, inserting this relation for u in (3.10)-(3.14) and taking into account that z is solution of problem (3.84)-(3.88), the unknown function y satisfies

$$\begin{aligned}
-\omega^2 y - \mathrm{div}\left(c^2 \nabla y\right) &= f + \omega^2 z && \text{in } \Omega, && (3.89)\\
c^2 \frac{\partial y}{\partial \boldsymbol{\nu}} &= 0 && \text{on } \Gamma_+ \cup \Gamma_-, && (3.90)\\
c \frac{\partial y}{\partial \boldsymbol{\nu}} &= 0 && \text{on } \Gamma_e \cup \Gamma_s, && (3.91)\\
y|_{\Omega_-} &= y|_{\Omega_+} && \text{on } \Gamma_I, && (3.92)\\
c_+^2 \left.\frac{\partial y}{\partial \boldsymbol{\nu}}\right|_{\Omega_-} &= c_-^2 \left.\frac{\partial y}{\partial \boldsymbol{\nu}}\right|_{\Omega_+} && \text{on } \Gamma_I. && (3.93)
\end{aligned}$$

Now, if it is assumed that $f \in \mathrm{L}^2(\Omega)$ then it is clear that $f + \omega^2 z$ belongs also to $\mathrm{L}^2(\Omega)$. Hence, it this case, this function admits the series representation in the Hilbert basis

$$f + \omega^2 z = \sum_{n=0}^{\infty} \langle f + \omega^2 z, w_n \rangle_{\mathrm{L}^2(\Omega)} w_n.$$

To compute the solution y of the translated problem (3.89)-(3.93), the series representation $y = \sum_{n=0}^{\infty} y_n w_n$ is inserted in the weak formulation of the time-harmonic problem. Using the orthogonality of the Hilbert basis and using the elements of the basis as test functions, it is deduced from $-\omega^2\langle y, w_n\rangle_{\mathrm{L}^2(\Omega)} + A_0(y, w_n) = \langle f + \omega^2 z, w_n\rangle_{\mathrm{L}^2(\Omega)}$ and the definition of the operator $\mathcal{A}_0$ that

$$-(\omega^2+1)y_n + \lambda_n y_n = \langle f + \omega^2 z, w_n\rangle_{\mathrm{L}^2(\Omega)}$$

and hence the solution y is given by

$$y = \sum_{n=0}^{\infty} \frac{\langle f + \omega^2 z, w_n\rangle_{\mathrm{L}^2(\Omega)}}{\lambda_n - (\omega^2+1)} w_{n,j}.$$

Finally, the solution of the original source problem (3.10)-(3.14) is given by

$$u = y + z = \sum_{n=0}^{\infty} \frac{\langle f + \lambda_n z, w_n\rangle_{\mathrm{L}^2(\Omega)}}{\omega_n^2 - \omega^2} w_n,$$

where $\omega_n = \sqrt{\lambda_n - 1}$ for $n \in \mathbb{N}$ are the resonance frequencies where the existence and uniqueness of solution is not ensured.

From standard results on the regularity of the solution u with respect to the source term and the boundary data, the rate of convergence of this series depends on the smoothness of the right hand side $f + \omega^2 z$, and in the particular case of being in the domain of $\mathcal{A}_0^{-m}$ (in this case $\mathrm{H}^{2m}(\Omega)$), the truncation error of the series representation to N terms is of order $\mathcal{O}((\omega_{N+1}^2 - \omega^2)^{-m})$ (see for example [13]) and hence it is enough to truncate the series only using the very first terms to obtain accurate approximations of the exact solution of problem (3.10)-(3.14) with $\beta = 0$.

However, if the same Hilbert basis is used to represent the solution of the source problem (3.10)-(3.14) with $\beta \neq 0$, then the convergence is very slow due to the elements of the basis do not satisfies the Robin boundary conditions on $\Gamma_e \cup \Gamma_s$. To deal with this drawback, the partition of unity finite element method it is used to replace the factor depending on the x_2 coordinate in each eigenfunction w_n. In this way, it will be the PUFEM approximation which will have to handle the Robin boundary conditions on $\Gamma_e \cup \Gamma_s$. The following sections describes in detail this novel numerical procedure and the modal basis used in the proposed discretization.

3.3.3 Dispersion equations for $\beta = 0$ in a rectangular domain

The analytic computation of these eigenpairs is straightforward using a classical separation of variables procedure (see for instance [15]). For completeness, the computations related to the separation of variables are described in detail. Firstly, let us consider the spectral problem in its strong form (3.83) and assume that the non-null eigenfunctions are smooth enough and are given by the product $w = q \otimes p$, i.e., $w(x_1, x_2) = q(x_1)p(x_2)$. In this case, since the profile of the speed of sound c is piecewise constant, the Helmholtz equation in $\Omega_+ = (0, L) \times (0, H)$ is rewritten as $-c_+^2 q''p - c_+^2 p''q = \lambda qp$. Since q and p only depends on an unique spatial variable, x_1 and x_2 respectively, then there exists a constant μ such that

$$q'' + \mu q = 0 \quad \text{in } (0, L) \qquad \text{and} \qquad - c_+^2 p'' - (\lambda - c_+^2 \mu) p = 0 \quad \text{in } (0, H).$$

Similar arguments can be performed for $\Omega_- = (0, L) \times (-a, 0)$. In both cases, the differential equation satisfied by q is given by $q'' = \mu q$, which is completed with the homogeneous Neumann boundary conditions at $x_1 = 0$ and $x_1 = L$. Straightforward computations show that there exists a sequence of eigenpairs $\{(\mu_n, q_n)\}_{n\in\mathbb{N}}$ (normalized with respect the $\mathrm{L}^2(0, L)$ norm) defined by

$$q_0(x_1) = \sqrt{\frac{1}{L}}, \qquad \mu_0 = 0, \tag{3.94}$$

$$q_n(x_1) = \sqrt{\frac{2}{L}} \cos\left(\sqrt{\mu_n} x_1\right), \qquad \mu_n = \left(\frac{n\pi}{L}\right)^2, \qquad n \in \mathbb{N},\ n \neq 0. \tag{3.95}$$

For each eigenpair (μ_n, q_n), the x_2-dependent factor $p = p_n$ must be computed. If the differential equation satisfied by p_n is completed with the homogeneous Neumann boundary conditions at $x_2 = -a$ and $x_2 = H$, p_n satisfies

$$- \left(c^2 p_n'\right)' - \left(\lambda_n - c^2 \mu_n\right) p_n = 0 \qquad \text{in } (-a, 0) \cup (0, H), \tag{3.96}$$

$$p_n'(-a) = p_n'(H) = 0, \tag{3.97}$$

$$p_n(0^+) = p_n(0^-), \tag{3.98}$$

$$c_+^2 p_n'(0^+) = c_-^2 p_n'(0^-). \tag{3.99}$$

For each fixed value of $n \in \mathbb{N}$, there exist a sequence of eigenpairs $\{(\lambda_{n,j}, p_{n,j})\}$ which are solution of the spectral differential problem (3.96)-(3.99). To describe them, two different cases should be considered. First, those eigenmodes which can be understood as interface waves (the so-called Love waves) which satisfy $\mu_n c_-^2 < \lambda_{n,j} < \mu_n c_+^2$. Having into account this condition, the solutions of equation (3.96), can be written

$$p_{n,j}(x_2) = \begin{cases} C_1 \cos(K_-^{n,j}(x_2 + a)) + C_2 \sin(K_-^{n,j}(x_2 + a)) & \text{if } x_2 \in (-a, 0), \\ D_1 \cosh(K_+^{n,j}(x_2 - H)) + D_2 \sinh(K_+^{n,j}(x_2 - H)) & \text{if } x_2 \in [0, H), \end{cases}$$

being C_1, C_2, D_1 and D_2 constants to be determined and where the positive wave numbers in each subdomain are given by

$$K_-^{n,j} = \sqrt{\mu_n \left(\left(\frac{\xi_{n,j}}{c_-}\right)^2 - 1\right)}, \quad K_+^{n,j} = \sqrt{\mu_n \left(1 - \left(\frac{\xi_{n,j}}{c_+}\right)^2\right)}, \tag{3.100}$$

where $\xi_{n,j} = \sqrt{\lambda_{n,j}/\mu_n}$ and hence variables $\xi_{n,j} \in (c_-, c_+)$. The Neumann boundary conditions (3.97) give that $C_2 = D_2 = 0$. Applying now the interface conditions (3.98)-(3.99) in order to find C_1 and D_1, the next system has to be solved

$$\left\{\begin{aligned} C_1\cos(K_-^{n,j}a) &= D_1\cosh(K_+^{n,j}H),\\ C_1K_-^{n,j}\sin(K_-^{n,j}a) &= D_1K_+^{n,j}\sinh(K_+^{n,j}H). \end{aligned}\right. \tag{3.101}$$

To assure that the determinant of the matrix in system (3.101) is null, the following dispersion equation must be fulfilled

$$\frac{c_+}{c_-}\sqrt{\frac{c_+^2-(\xi_{n,j})^2}{(\xi_{n,j})^2-c_-^2}}\tanh\left(H\sqrt{\mu_n\left(1-\left(\frac{\xi_{n,j}}{c_+}\right)^2\right)}\right)=\tan\left(a\sqrt{\mu_n\left(\left(\frac{\xi_{n,j}}{c_-}\right)^2-1\right)}\right). \tag{3.102}$$

As the system is indeterminate, it is chosen that $C_1=\cos(K_-^{n,j}a)^{-1}$ and then, from the first equation in (3.101), it can be deduced that $D_1=\cosh(K_+^{n,j}H)^{-1}$. So in this case, the eigenfunctions $p_{n,j}$ (normalized to satisfy $p_{n,j}(0)=1$) are given by

$$p_{n,j}(x_2)=\left\{\begin{array}{ll} \dfrac{\cos(K_-^{n,j}(x_2+a))}{\cos(K_-^{n,j}a)} & \text{if } x_2\in(-a,0),\\[2ex] \dfrac{\cosh(K_+^{n,j}(x_2-H))}{\cosh(K_+^{n,j}H)} & \text{if } x_2\in[0,H). \end{array}\right. \tag{3.103}$$

The second type of eigenmodes are the so-called interior waves. They correspond to those eigenmodes whose eigenvalue satisfies $\lambda_{n,j}>\mu_n c_+^2$. Having into account this condition, the solutions of equation (3.96) for this case can be written

$$p_{n,j}(x_2)=\left\{\begin{array}{ll} \widetilde{C}_1\cos(\widetilde{K}_-^{n,j}(x_2+a))+\widetilde{C}_2\sin(\widetilde{K}_-^{n,j}(x_2+a)) & \text{if } x_2\in(-a,0),\\ \widetilde{D}_1\cos(\widetilde{K}_+^{n,j}(x_2-H))+\widetilde{D}_2\sin(\widetilde{K}_+^{n,j}(x_2-H)) & \text{if } x_2\in[0,H), \end{array}\right.$$

being $\widetilde{C}_1$, $\widetilde{C}_2$, $\widetilde{D}_1$ and $\widetilde{D}_2$ constants to be determined and where the positive wave numbers in each subdomain are given by

$$\widetilde{K}_-^{n,j}=\sqrt{\mu_n\left(\left(\frac{\zeta_{n,j}}{c_-}\right)^2-1\right)},\qquad \widetilde{K}_+^{n,j}=\sqrt{\mu_n\left(\left(\frac{\zeta_{n,j}}{c_+}\right)^2-1\right)}. \tag{3.104}$$

In the expressions written above $\zeta_{n,j}=\sqrt{\lambda_{n,j}/\mu_n}$ and hence $\zeta_{n,j}\in(c_+,+\infty)$. The Neumann boundary conditions (3.97) give that $\widetilde{C}_2=\widetilde{D}_2=0$. Applying now the interface conditions (3.98)-(3.99) in order to find $\widetilde{C}_1$ and $\widetilde{D}_1$, the next system has to be solved

$$\left\{\begin{aligned} \widetilde{C}_1\cos(\widetilde{K}_-^{n,j}a) &= \widetilde{D}_1\cos(\widetilde{K}_+^{n,j}H),\\ -\widetilde{C}_1\widetilde{K}_-^{n,j}\sin(\widetilde{K}_-^{n,j}a) &= \widetilde{D}_1\widetilde{K}_+^{n,j}\sin(\widetilde{K}_+^{n,j}H). \end{aligned}\right. \tag{3.105}$$

To assure that the determinant of the matrix in system (3.105) is null, the following dispersion equation must be fulfilled

$$\frac{c_+}{c_-}\sqrt{\frac{(\zeta_{n,j})^2-c_+^2}{(\zeta_{n,j})^2-c_-^2}}\tan\left(H\sqrt{\mu_n\left(\left(\frac{\zeta_{n,j}}{c_+}\right)^2-1\right)}\right)+\tan\left(a\sqrt{\mu_n\left(\left(\frac{\zeta_{n,j}}{c_-}\right)^2-1\right)}\right)=0. \tag{3.106}$$

As the system is indeterminate, it is chosen that $\widetilde{C}_1=\cos(\widetilde{K}_-^{n,j}a)^{-1}$ and then, from the first equation in (3.105), it can be deduced that $\widetilde{D}_1=\cos(\widetilde{K}_+^{n,j}H)^{-1}$. So in this case, the eigenfunctions $p_{n,j}$ (normalized to satisfy $p_{n,j}(0)=1$) are given by

$$p_{n,j}(x_2)=\begin{cases}\dfrac{\cos(\widetilde{K}_-^{n,j}(x_2+a))}{\cos(\widetilde{K}_-^{n,j}a)} & \text{if } x_2\in[-a,0],\\[2ex] \dfrac{\cos(\widetilde{K}_+^{n,j}(x_2-H))}{\cos(\widetilde{K}_+^{n,j}H)} & \text{if } x_2\in[0,H].\end{cases} \tag{3.107}$$

Figure 3.1 illustrates two eigenmodes, a Love wave on the left plot and an internal wave on the right one. The speed of sound has been taken $c_-=1/2$ in Ω_- and $c_+=1$ in Ω_+. The geometrical dimensions of the computational domain are given in this example by $L=1$, $a=0.2$ and $H=0.8$. The eigenmode Love wave, described by equation (3.103) for $n=15$ and $j=5$, has an oscillatory behaviour in $(-a,0)$ and decays exponentially in $(0,H)$, as it can be observed in the left plot. The right plot illustrates the eigenmode internal wave, described by equation (3.107) for $n=15$ and $j=5$. It has an oscillatory behaviour in the whole domain, although the oscillation changes when the wave crosses the interface at $x_2=0$.

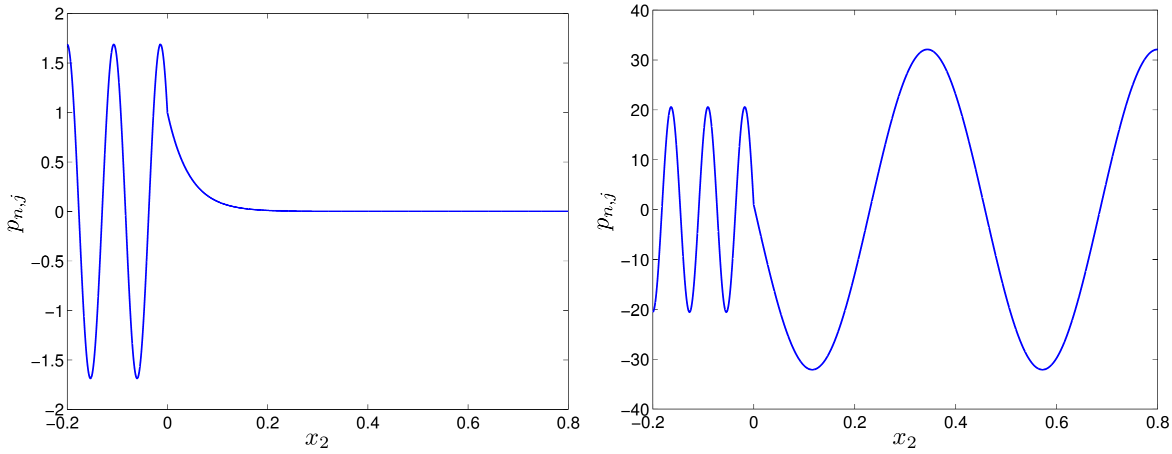

Figure 3.1: Love wave $p_{n,j}$ from equation (3.103) (left) and internal wave $p_{n.j}$ from equation (3.107) (right) plotted with respect to x_2, for $n=15$ and $j=5$. It can be observed the exponential decay of the Love wave and the oscillatory behaviour of the internal wave in $(0,H)$, being $H=0.8$.

Note that the eigenmodes whose eigenvalue satisfies $\lambda_{n,j} < \mu_n c_-^2$ are not considered, because the dispersion equation that should be fulfilled is

$$\frac{c_+}{c_-}\sqrt{\frac{c_+^2 - (\xi_{n,j})^2}{c_-^2 - (\xi_{n,j})^2}} \tanh\left(H\sqrt{\mu_n\left(1 - \left(\frac{\xi_{n,j}}{c_+}\right)^2\right)}\right) + \tanh\left(a\sqrt{\mu_n\left(1 - \left(\frac{\xi_{n,j}}{c_-}\right)^2\right)}\right) = 0. \tag{3.108}$$

for $\xi_{n,j} = \sqrt{\lambda_{n,j}/\mu_n} \in (-\infty, c_-)$, which it is impossible as the arguments of the hyperbolic tangents are strictly positive.

In conclusion, the eigenpairs $\{(\lambda_{n,j}, w_{n,j})\}_{n,j\in\mathbb{N}}$ of the spectral problem (3.79)-(3.80) are given by $w_{n,j}(x_1, x_2) = q_n(x_1)p_{n,j}(x_2)$, where q_n are defined by (3.94)-(3.95) and $p_{n,j}$ are given by (3.103)-(3.106). To distinguish those eigenpairs which correspond to internal waves from those ones which are associated to Love waves, for each index $n \in \mathbb{N}$, which fixes the mode q_n with the x_1-dependency, the indexes $j \in \mathbb{N}$ will be split in two disjoint sorted subsets: $w_{n,j}$ with $j \in \mathcal{I}_n \subset \mathbb{N}$ will be considered internal modes whereas if $j \in \mathcal{L}_n \subset \mathbb{N}$ will denote Love eigenpairs. The ordering of subsets $\mathcal{L}_n$ and $\mathcal{I}_n$ are given by the natural ascending order with respect to the magnitude of their associated eigenvalues $\lambda_{n,j}$.

***Remark* 3.3.15.** *Despite the spectral problems with $\beta = 0$ and $\beta > 0$ share similar variational formulations, the change of nature on the boundary condition type (from Robin to Neumann boundary condition on $\Gamma_e \cup \Gamma_s$) implies that the eigenfunctions of the spectral problem* (3.79)-(3.82) *are not eigenfunctions of problem* (3.19)-(3.23)*. Moreover, even in the case of constant functions, it is straightforward to show that the spectral problem* (3.19)-(3.23) *for $\beta > 0$ does not admit eigenfunctions of type $w(x_1, x_2) = p(x_2)$ since the non-null constant functions do not satisfied the Robin conditions* (3.19).

3.4 Modal-based PUFEM method

The partition of unity finite element method [35] is considered as an enriched method where the standard discretization of a classical finite element method is used as partition of unity and hence, every local polynomial basis is multiplied by a exact solution of the problem to be solved numerically. Usually, in the case of the Helmholtz equation stated in two dimensions, this enrichment procedure involves plane waves [36, 42], radial solutions (written in terms of Bessel functions) [36] or two-dimensional eigenfunctions [10], which are multiplied by piecewise polynomials functions defined on a two-dimensional triangular mesh.

3.4.1 Modal-based enrichment

However, in the present approach, the PUFEM method is only used in the x_1-axis and hence requiring only an inexpensive one-dimensional domain, keeping the expressions $p_{n,j}$ of modal decomposition with the x_2 dependency. The second main difference with respect to the standard PUFEM discretization lies on the fact that the enrichment in the x_1-axis is not based on plane waves with a fixed wave number. At the contrary, it will be ensured that the modal contributions of every q_n with the x_1-dependency belongs to the PUFEM discrete space.

Clearly, if expressions $q_n(x_1)$ were used directly to define the enrichment of the PUFEM space, since q_n are computed to satisfy homogeneous Neumann boundary conditions (obtained with $\beta = 0$), a lack of convergence will again arise mainly around the boundaries where the Robin conditions (with $\beta \neq 0$) were considered. To avoid this kind of drawbacks, q_n is rewritten in terms of complex exponential of different sign, i.e.,

$$q_n(x_1) = C_0 q_n^+(x_1) + C_1 q_n^-(x_1), \qquad n \in \mathbb{N},\ n \neq 0,$$

where $C_0 = C_1 = \sqrt{2/L}/2$, $q_n^+(x_1) = \exp(i\sqrt{\mu_n}x_1)$ and $q_n^-(x_1) = \exp(-i\sqrt{\mu_n}x_1)$. Taking this new rewriting of the modes q_n using complex exponential expressions, if both functions q_n^+ and q_n^- are involved separately in the PUFEM enrichment, it is guaranteed that any boundary condition at $x_1 = 0$ and $x_1 = L$ could be satisfied by a linear combination of type $C_0 q_n^+ + C_1 q_n^-$ with adequate constants C_0 and C_1.

The modal contribution for $n = 0$ will be treated in a different way. Since q_0 is a constant function, it belongs to the standard piecewise linear polynomial finite element space and, hence, it does not add any new feature to the classical discrete FEM approximation. In addition, for the case $n = 0$, it can be deduced from the dispersion equation (3.102) that there does not exist any Love eigenmodes associated to $n = 0$. In fact, as it has been discussed in Remark 3.3.15, there does not exist any eigenfunction, solution of the quadratic eigenvalue problem (3.19)-(3.23), which depends only on the x_2 spatial coordinate. Hence, the proposed numerical method approximates the solution of the source problem without the contribution of mode associated to $n = 0$ in the PUFEM modal enrichment. It is, the eigenfunctions (internal waves) $w_{0,j}$ with $j \in \mathcal{I}_0$, will not be used in the discretization method.

Obviously, as it has been already discussed in the section above, for each $n \in \mathbb{N}$, the eigenmodes associated to the internal waves are infinite (but countable) and for discretization purposes, such set modes must be truncated and so only considering a finite number of eigenmodes with the smallest eigenvalues. The truncated set of indexes for the interior modes will be denoted by $\mathcal{I}_n^{J_n}$, being J_n the number of internal modes used in the discretization. The criterion to truncate the infinite sequence of internal modes corresponds to keep in the discretization only those internal eigenvalues $\lambda_{n,j}$ which satisfy

$$c_+^2 \mu_n \leq \lambda_{n,j} \leq c_0^2 \mu_n \qquad \text{for } n = 0, \ldots, N,$$

and where c_0 is the maximum value allowed for the solutions $\zeta_{n,j}$ of the dispersion equation (3.106). In the case of the eigenpairs associated to Love waves, its dispersion equation only admits a finite number of solutions and so, for a fixed value of $n \in \mathbb{N}$, all the Love eigenmodes are considered in the discretization. The number of Love eigenmodes included in the subset $\mathcal{L}_n$ will be denoted by L_n. Using this notation, if $\lambda_{n,j}$ is a eigenvalue of the spectral problem then there exists a k-th family of eigenmodes such that the pair of indexes

$(n,j) \in \{k\} \times (\mathcal{L}_k \cup \mathcal{I}_k^{J_k})$, this is,

$$n = k \qquad \text{and} \qquad j \in \{\underbrace{1,\ldots,L_k}_{j\in\mathcal{L}_k}, \underbrace{L_k+1,\ldots,L_k+J_k}_{j\in\mathcal{I}_k^{J_k}}\}, \qquad \text{with } 0 \le k \le N.$$

To describe precisely the proposed modal-based PUFEM method, an one-dimensional finite element mesh must be introduced. For simplicity, an uniform mesh of size h will be used throughout the rest of the present work, this is, a mesh with M elements and whose nodes are given by $\{y_m = hm : m = 0,\ldots,M\} \subset [0,L]$. Clearly, such mesh has $M+1$ nodes and a mesh size $h = L/M$. In addition, it has been chosen as local polynomial basis $\{\varphi_m\}_{m=0}^M$ the standard Lagrange $\mathbb{P}_1$ (piecewise linear) finite element basis, defined by the nodal relation $\varphi_m(y_l) = \delta_{lm}$, where δ_{lm} is the Kronecker's delta. Hence, the discrete space X_h will be defined by the span

$$\mathrm{X}_h = \Big\langle \Big\{ (\varphi_m q_n^+) \otimes p_{n,j},\ (\varphi_m q_n^-) \otimes p_{n,j},\ m = 0,\ldots,M, \\ (n,j) \in \{k\} \times (\mathcal{L}_k \cup \mathcal{I}_k^{J_k}) \text{ for } k = 1,\ldots,N \Big\} \Big\rangle, \quad (3.109)$$

where recall that $[(\varphi_m q_n^\pm) \otimes p_{n,j}](x_1,x_2) = \varphi_m(x_1) q_n^\pm(x_1) p_{n,j}(x_2)$ and the ordering of indexes (n,j) in the subsets $\mathcal{L}_k$ and $\mathcal{I}_k^{J_k}$ are given by the natural ascending order with respect to the magnitude of their associated eigenvalues $\lambda_{n,j}$.

From the definition of X_h and since $\{\varphi_m\}_{m=0}^M$ is a partition of unity of the interval $[0,L]$, i.e., $\sum_{m=0}^M \varphi_m(x_1) = 1$, it is clearly deduced that

$$w_{n,j} = \sqrt{\frac{1}{2L}} \sum_{m=0}^{M} (\varphi_m q_n^+ + \varphi_m q_n^-) \otimes p_{n,j},$$

with $(n,j) \in \{k\} \times (\mathcal{L}_k \cup \mathcal{I}_k^{J_k})$ and any $k = 1,\ldots,N$, belongs to the discrete space X_h. Due to this fact, the proposed discretization inherits potentially the spectral convergence of the modal basis approximations (see Section 3.5 for the illustration of the numerical behaviour of the proposed method). Simultaneously, due to the use of a partition of unity, the functions used for the enrichment in the discrete space has not to satisfy all the boundary conditions of the source problem, what increase the flexibility of choice for the modal basis. In addition, due to the compact support of the finite element basis $\{\varphi_m\}_{m=0}^M$, the matrix of the discrete problem will be sparse, what decreases the computational storage requirements for a typical modal discretization which uses full discrete matrices.

Since the modal-based PUFEM enrichment is flexible enough to select only a part of the spectral basis, the impact in the accuracy of considering only Love waves in the discrete space has been analysed in the numerical results shown in Section 3.5.1. In this case, the discrete space is defined by

$$\mathrm{X}_h^{\mathcal{L}} = \big\langle \big\{ (\varphi_m q_n^+) \otimes p_{n,j},\ (\varphi_m q_n^-) \otimes p_{n,j},\ m = 0,\ldots,M, \\ (n,j) \in \{k\} \times \mathcal{L}_k \text{ for } k = 1,\ldots,N \big\} \big\rangle, \quad (3.110)$$

The numerical features of the proposed modal-based PUFEM discretization with these two discrete spaces are described in detail in the following two sections.

3.4.2 Discrete problem

To write the matrix description of the variational problem using the discrete space X_h (and analogously $\mathrm{X}_h^{\mathcal{L}}$), each term of the variational formulation associated to the sesquilinear form A_β, the L^2 -inner product and the source and boundary data contributions are computed for unknown and test functions belonging to the discrete space. Hence, the discrete variational formulation can be stated as follows: Given the source term $f \in \mathrm{L}^2(\Omega)$, and the boundary loads $r \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_e \cup \Gamma_s)$ and $g \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_+ \cup \Gamma_-)$, find $u_h \in \mathrm{X}_h$ such that

$$A_\beta(u_h, v_h) - \omega^2 \langle u_h, v_h \rangle_{\mathrm{L}^2(\Omega)} = \ell(v_h) \qquad \text{for all } v_h \in \mathrm{X}_h. \tag{3.111}$$

Clearly, any function $u_h \in \mathrm{X}_h$ is determined by their coordinate vectors

$$\begin{aligned} \vec{u}_h = ((u^+_{mnj}, u^-_{mnj})_{j\in\mathcal{L}_n\cup\mathcal{I}_n^{J_n}})_{m,n=0}^{M,N} \\ = (u^+_{011}, u^-_{011}, u^+_{012}, u^-_{012}, \ldots, u^+_{01L_1+J_1}, u^-_{01L_1+J_1}, \ldots, \\ u^+_{0NL_N+J_N}, u^-_{0NL_N+J_N}, u^+_{111}, u^-_{111}, \ldots, u^+_{MNL_N+J_N}, u^-_{MNL_N+J_N}), \end{aligned} \tag{3.112}$$

and so these coordinate coefficients define the discrete function

$$u_h = \sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{j=1}^{L_n+J_n} \left(u^+_{mnj}(\varphi_m q_n^+) \otimes p_{n,j} + u^-_{mnj}(\varphi_m q_n^-) \otimes p_{n,j} \right). \tag{3.113}$$

The coordinate ordering in (3.112) has been chosen to reduce as much as possible the bandwidth of the sparse matrices involved in the discretization. In fact, since the degrees of freedom related to the same finite element basis φ_m are stored consecutively, it is straightforward to show that due to the compact support of the one-dimensional finite element basis, the bandwidth of the matrix description is given by $6\max_{1\leq n\leq N}(L_n + J_n)$.

Taking into account this basis representation in X_h, the discrete variational formulation (3.111) admits the matrix description

$$-\omega^2 \mathcal{M}\vec{u}_h - i\omega\beta\mathcal{C}\vec{u}_h + \mathcal{K}\vec{u}_h = \vec{b}_h, \tag{3.114}$$

where the coefficients of the matrix $\mathcal{M}$, $\mathcal{C}$, and $\mathcal{K}$ (with respect to the coordinates $u^\pm_{mnj}$ induced by the basis of X_h) are given by the following expressions: taking into account the expression of the sesquilinear form (3.17), the mass matrix M is defined by

$$[\mathcal{M}]^{++}_{mnj,\,lki} = \int_\Omega (\varphi_m q_n^+) \otimes p_{n,j}\overline{(\varphi_l q_k^+) \otimes p_{k,i}}\,\mathrm{d}\boldsymbol{x} = \left(\int_0^L \varphi_m\varphi_l q_n^+ q_k^- \,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j}p_{k,i}\,\mathrm{d}x_2\right),$$

$$[\mathcal{M}]^{+-}_{mnj,\,lki} = \int_\Omega (\varphi_m q_n^+) \otimes p_{n,j}\overline{(\varphi_l q_k^-) \otimes p_{k,i}}\,\mathrm{d}\boldsymbol{x} = \left(\int_0^L \varphi_m\varphi_l q_n^+ q_k^+ \,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j}p_{k,i}\,\mathrm{d}x_2\right),$$

$$[\mathcal{M}]^{-+}_{mnj,\,lki} = \int_\Omega (\varphi_m q_n^-) \otimes p_{n,j}\overline{(\varphi_l q_k^+) \otimes p_{k,i}}\,\mathrm{d}\boldsymbol{x} = \left(\int_0^L \varphi_m\varphi_l q_n^- q_k^- \,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j}p_{k,i}\,\mathrm{d}x_2\right),$$

$$[\mathcal{M}]^{--}_{mnj,\,lki} = \int_\Omega (\varphi_m q_n^-) \otimes p_{n,j}\overline{(\varphi_l q_k^-) \otimes p_{k,i}}\,\mathrm{d}\boldsymbol{x} = \left(\int_0^L \varphi_m\varphi_l q_n^- q_k^+ \,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j}p_{k,i}\,\mathrm{d}x_2\right),$$

the damping matrix C is given by

$$
\begin{aligned}
[\mathcal{C}]^{++}_{mnj,\,lki} &= \int_{\Gamma_e \cup \Gamma_s} c\,(\varphi_m q_n^+) \otimes p_{n,j} \overline{(\varphi_l q_k^+) \otimes p_{k,i}}\,\mathrm{d}\sigma \\
&= \Big((\varphi_m \varphi_l q_n^+ q_k^-)\big|_{x_1=0} + (\varphi_m \varphi_l q_n^+ q_k^-)\big|_{x_1=L} \Big) \left(\int_{-a}^{H} p_{n,j} p_{k,i}\,\mathrm{d}x_2 \right), \\
[\mathcal{C}]^{+-}_{mnj,\,lki} &= \int_{\Gamma_e \cup \Gamma_s} c\,(\varphi_m q_n^+) \otimes p_{n,j} \overline{(\varphi_l q_k^-) \otimes p_{k,i}}\,\mathrm{d}\sigma \\
&= \Big((\varphi_m \varphi_l q_n^+ q_k^+)\big|_{x_1=0} + (\varphi_m \varphi_l q_n^+ q_k^+)\big|_{x_1=L} \Big) \left(\int_{-a}^{H} p_{n,j} p_{k,i}\,\mathrm{d}x_2 \right), \\
[\mathcal{C}]^{-+}_{mnj,\,lki} &= \int_{\Gamma_e \cup \Gamma_s} c\,(\varphi_m q_n^-) \otimes p_{n,j} \overline{(\varphi_l q_k^+) \otimes p_{k,i}}\,\mathrm{d}\sigma \\
&= \Big((\varphi_m \varphi_l q_n^- q_k^-)\big|_{x_1=0} + (\varphi_m \varphi_l q_n^- q_k^-)\big|_{x_1=L} \Big) \left(\int_{-a}^{H} p_{n,j} p_{k,i}\,\mathrm{d}x_2 \right), \\
[\mathcal{C}]^{--}_{mnj,\,lki} &= \int_{\Gamma_e \cup \Gamma_s} c\,(\varphi_m q_n^-) \otimes p_{n,j} \overline{(\varphi_l q_k^-) \otimes p_{k,i}}\,\mathrm{d}\sigma \\
&= \Big((\varphi_m \varphi_l q_n^- q_k^+)\big|_{x_1=0} + (\varphi_m \varphi_l q_n^- q_k^+)\big|_{x_1=L} \Big) \left(\int_{-a}^{H} p_{n,j} p_{k,i}\,\mathrm{d}x_2 \right),
\end{aligned}
$$

and the stiffness matrix is defined by

$$
\begin{aligned}
[\mathcal{K}]^{++}_{mnj,\,lki} &= \int_{\Omega} c^2 \int_{\Omega} \nabla((\varphi_m q_n^+) \otimes p_{n,j}) \cdot \nabla(\overline{(\varphi_l q_k^+) \otimes p_{k,i}})\,\mathrm{d}\boldsymbol{x} \\
&= \left(\int_0^L (\varphi_m q_n^+)'(q_k^- \varphi_l)'\,\mathrm{d}x_1 \right) \left(\int_{-a}^{H} p_{n,j} p_{k,i}\,\mathrm{d}x_2 \right) \\
&\quad + \left(\int_0^L \varphi_m q_n^+ q_k^- \varphi_l\,\mathrm{d}x_1 \right) \left(\int_{-a}^{H} p'_{n,j} p'_{k,i}\,\mathrm{d}x_2 \right),
\end{aligned}
$$

$$
\begin{aligned}
[\mathcal{K}]^{+-}_{mnj,\,lki} &= \int_\Omega c^2 \int_\Omega \nabla((\varphi_m q_n^+) \otimes p_{n,j}) \cdot \nabla(\overline{(\varphi_l q_k^-) \otimes p_{k,i}})\,\mathrm{d}\boldsymbol{x} \\
&= \left(\int_0^L (\varphi_m q_n^+)'(q_k^+ \varphi_l)'\,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j} p_{k,i}\,\mathrm{d}x_2\right) \\
&\quad + \left(\int_0^L \varphi_m q_n^+ q_k^+ \varphi_l\,\mathrm{d}x_1\right)\left(\int_{-a}^H p'_{n,j} p'_{k,i}\,\mathrm{d}x_2\right), \\
[\mathcal{K}]^{-+}_{mnj,\,lki} &= \int_\Omega c^2 \int_\Omega \nabla((\varphi_m q_n^-) \otimes p_{n,j}) \cdot \nabla(\overline{(\varphi_l q_k^+) \otimes p_{k,i}})\,\mathrm{d}\boldsymbol{x} \\
&= \left(\int_0^L (\varphi_m q_n^-)'(q_k^- \varphi_l)'\,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j} p_{k,i}\,\mathrm{d}x_2\right) \\
&\quad + \left(\int_0^L \varphi_m q_n^- q_k^- \varphi_l\,\mathrm{d}x_1\right)\left(\int_{-a}^H p'_{n,j} p'_{k,i}\,\mathrm{d}x_2\right), \\
[\mathcal{K}]^{--}_{mnj,\,lki} &= \int_\Omega c^2 \int_\Omega \nabla((\varphi_m q_n^-) \otimes p_{n,j}) \cdot \nabla(\overline{(\varphi_l q_k^-) \otimes p_{k,i}})\,\mathrm{d}\boldsymbol{x} \\
&= \left(\int_0^L (\varphi_m q_n^-)'(q_k^+ \varphi_l)'\,\mathrm{d}x_1\right)\left(\int_{-a}^H p_{n,j} p_{k,i}\,\mathrm{d}x_2\right) \\
&\quad + \left(\int_0^L \varphi_m q_n^- q_k^+ \varphi_l\,\mathrm{d}x_1\right)\left(\int_{-a}^H p'_{n,j} p'_{k,i}\,\mathrm{d}x_2\right).
\end{aligned}
$$

It should be noted that all the integrals stated below have been computed using one-dimensional exact integration with closed form integral formulas (without requiring the use of quadrature formulas). Such exact integration strategy has been applied also to the right-hand side term $\vec{b}_h$ but restricted to five-order polynomial source function $f = f_1 \otimes f_2$ and boundary functions g and r.

Analogous considerations are applied to the computation of the coefficients of the right-hand side in the linear system (3.114), which are given by

$$
\begin{aligned}
[\vec{b}_h]^+_{mnj} &= \int_\Omega f\overline{(\varphi_m q_n^+) \otimes p_{n,j}}\,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_+ \cup \Gamma_-} g\overline{(\varphi_m q_n^+) \otimes p_{n,j}}\,\mathrm{d}\sigma + \int_{\Gamma_e \cup \Gamma_s} c\,r\overline{(\varphi_m q_n^+) \otimes p_{n,j}}\,\mathrm{d}\sigma \\
&= \left(\int_0^L f_1 \varphi_m q_n^-\,\mathrm{d}x_1\right)\left(\int_{-a}^H f_2 p_{n,j}\,\mathrm{d}x_2\right) + p_{n,j}(-a)\int_0^L g|_{x_2=-a}\varphi_m q_n^-\,\mathrm{d}x_1 \\
&\quad + p_{n,j}(H)\int_0^L g|_{x_2=H}\varphi_m q_n^-\,\mathrm{d}x_1 + \left.(\varphi_m q_n^-)\right|_{x_1=0}\int_{-a}^H cr|_{x_1=0} p_{n,j}\,\mathrm{d}x_2 \\
&\quad + \left.(\varphi_m q_n^-)\right|_{x_1=L}\int_{-a}^H cr|_{x_1=L} p_{n,j}\,\mathrm{d}x_2,
\end{aligned}
$$

and

$$
\begin{aligned}
[\vec{b}_h]^{-}_{mnj} &= \int_{\Omega} f\,\overline{(\varphi_m q_n^-)\otimes p_{n,j}}\,\mathrm{d}\boldsymbol{x} + \int_{\Gamma_+\cup\Gamma_-} g\,\overline{(\varphi_m q_n^-)\otimes p_{n,j}}\,\mathrm{d}\sigma + \int_{\Gamma_e\cup\Gamma_s} c\,r\,\overline{(\varphi_m q_n^-)\otimes p_{n,j}}\,\mathrm{d}\sigma \\
&= \left(\int_0^L f_1\varphi_m q_n^+\,\mathrm{d}x_1\right)\left(\int_{-a}^H f_2 p_{n,j}\,\mathrm{d}x_2\right) + p_{n,j}(-a)\int_0^L g|_{x_2=-a}\varphi_m q_n^+\,\mathrm{d}x_1 \\
&\quad + p_{n,j}(H)\int_0^L g|_{x_2=H}\varphi_m q_n^+\,\mathrm{d}x_1 + \left.(\varphi_m q_n^+)\right|_{x_1=0}\int_{-a}^H cr|_{x_1=0}p_{n,j}\,\mathrm{d}x_2 \\
&\quad + \left.(\varphi_m q_n^+)\right|_{x_1=L}\int_{-a}^H cr|_{x_1=L}p_{n,j}\,\mathrm{d}x_2.
\end{aligned}
$$

Obviously, from the hermitian character of the L^2-inner product and the sesquilinear form A_β for $\beta = 0$, both matrices $\mathcal{M}$ and $\mathcal{K}$ are hermitian matrices. A direct inspection on coefficients of damping matrix C also reveals that it is an hermitian matrix.

3.4.3 Analysis of the condition number

It is well known in the scientific computing literature that the enriched methods and, in particular, those ones which are based on partition of unity and the use of plane-waves suffer from a poor conditioning. The proposed modal-based partition of unity method also share this kind of conditioning drawbacks even with the PUFEM discretization is restricted to a one-dimensional discretization in the x_1 axis.

To check the origin of this conditioning problem, the condition number $\kappa(\mathcal{M})$ of the mass matrix $\mathcal{M}$ will be analysed in a very simple case: it has been considered the pure Neumann problem (with $\beta = 0$) for a one-layer material (i.e. $c^+ = c^-$). Similar arguments could be also applied to the stiffness and damping matrix $\mathcal{K}$ and $\mathcal{C}$ in the linear system (3.114). To highlight the different order of magnitude of conditioning in PUFEM methods, it will be compared with those condition numbers coming from an standard finite element discretization.

First, notice that the condition number of the mass matrix is not an issue in a standard piecewise linear finite element discretization (in one-dimension with a uniform mesh). In this case, for the finite element mass matrix, its condition number is upper bounded independently of the mesh size, this is, $\kappa(\mathcal{M}) = \mathcal{O}(1)$. On the contrary, the condition number of the finite element stiffness matrix grows as $\mathcal{O}(h^{-2})$ (see [18] for further details). In what follows, it will be checked that the condition number of the modal-based PUFEM mass matrix increases when the number of eigenmodes is enlarged and simultaneously a refined finite element mesh is used in the partition of unity). In fact, it will be shown that $\kappa(\mathcal{M}) = \mathcal{O}(h^{-2})$.

Firstly, in the simple case of $\beta = 0$ and $c^- = c^+$, the modal basis solution of the spectral problem is given by $w_{n,j} = q_n \otimes p_j$ where recall that q_n, $n \in \mathbb{N}$, $n \neq 0$ are defined by (3.95) and p_j, $j \in \mathbb{N}$ are given as follows:

$$p_0(x_1) = \sqrt{\frac{1}{a+H}}, \tag{3.115}$$

$$p_j(x_1) = \sqrt{\frac{2}{a+H}} \cos\left(\frac{j\pi x_1}{a+H}\right), \ j \in \mathbb{N}, \ j \neq 0. \tag{3.116}$$

Notice that $\{p_j\}_{j\in\mathbb{N}}$ is a Hilbert basis in $\mathrm{L}^2(-a, H)$.

Since the discretization space X_h admits a discrete basis where the elements are tensor products of functions (with independent factors), i.e., it holds

$$\mathrm{X}_h = \left\langle \left\{ (\varphi_m q_n^+) \otimes p_j, \ (\varphi_m q_n^-) \otimes p_j, \quad m = 0, \dots, M, \quad n, j = 0, \dots, N, \qquad n \neq 0 \right\} \right\rangle, \tag{3.117}$$

then, the complex-valued mass matrix $\mathcal{M}$ of size $2N(N+1)(M+1) \times 2N(N+1)(M+1)$ also inherits this separation of variables and after a reordering (permutation of rows and columns), it can be written as a Kronecker product of matrices $\mathcal{M} = \mathcal{A} \otimes \mathcal{B}$ (where the size of $\mathcal{A}$ is $2(M+1)N \times 2(M+1)N$ and the size of $\mathcal{B}$ is $(N+1) \times (N+1)$) being

$$[\mathcal{A}]_{mn,\,lk}^{++} = \int_0^L \varphi_m \varphi_l q_n^+ q_k^- \,\mathrm{d}x_1, \qquad [\mathcal{A}]_{mn,\,lk}^{+-} = \int_0^L \varphi_m \varphi_l q_n^+ q_k^+ \,\mathrm{d}x_1, \tag{3.118}$$

$$[\mathcal{A}]_{mn,\,lk}^{-+} = \int_0^L \varphi_m \varphi_l q_n^- q_k^- \,\mathrm{d}x_1, \qquad [\mathcal{A}]_{mn,\,lk}^{--} = \int_0^L \varphi_m \varphi_l q_n^- q_k^+ \,\mathrm{d}x_1, \tag{3.119}$$

for $0 \le m, l \le M$, $1 \le n, k \le N$, and

$$[\mathcal{B}]_{i,j} = \int_{-a}^{H} p_j p_i \,\mathrm{d}x_2,$$

for $0 \le i, j \le N$. Trivially, from the orthogonality of the basis $\{p_j\}_{j\in\mathbb{N}}$, it is obtained that $\mathcal{B}$ is the identity matrix $\mathcal{I}$. Hence, in the simple case considered here, $\mathcal{M} = \mathcal{A} \otimes \mathcal{I}$. Classical linear algebra results show that the spectrum of $\mathcal{M}$ and $\mathcal{A}$ coincides (see [32]) and so their condition number also coincides.

Lemma 3.4.1. *Let $\mathcal{A}$ be the matrix defined by* (3.118)-(3.119). *If there exists $\tilde{N} \in \{1, \dots, N\}$ such that the size of the finite element mesh satisfies $h < L/(2(\tilde{N}+1))$ then it holds*

$$\kappa(\mathcal{A}) \le Ch^{-2}, \tag{3.120}$$

where C is a positive constant independent of M and N, only dependent of L and $\tilde{N}$.

Proof. With the aim of estimating $\kappa(\mathcal{A})$, some estimates will be computed from the numerical range of $\mathcal{A}$. Firstly, fixed $m \in \{0, \dots, M\}$ and $n = \tilde{N}$, consider the coordinate vector $\vec{v} \in \mathbb{C}^{2(M+1)N}$ associated to the function of the discrete space X_h given by $v(x_1) = \varphi_m \sin(n\pi(x_1 - mh)/L)$, which corresponds to the linear combination of basis functions

$$v = \frac{\varphi_m}{2i} \left(\frac{q_n^+}{q_n^+(mh)} - \frac{q_n^-}{q_n^-(mh)} \right) \in \mathrm{X}_h. \tag{3.121}$$

It holds

$$\vec{v}^*\mathcal{A}\vec{v} = \int_0^L v(x_1)\bar{v}(x_1)\,\mathrm{d}x_1 = \int_0^L \varphi_m^2(x_1)\sin^2\left(\frac{n\pi(x_1-mh)}{L}\right)\mathrm{d}x_1$$
$$= 2\int_0^h \frac{s^2}{h^2}\sin^2\left(\frac{n\pi s}{L}\right)\mathrm{d}s \le C\int_0^h \frac{s^4}{h^2}\,\mathrm{d}s \le Ch^3,$$

where each occurrence of constant C could denote a different value independent of h (only dependent on L and $n = \tilde{N}$). To obtain the estimate above, it has been used the first order Taylor polynomial approximation of the sine function around the origin. Now, it is straightforward to show from (3.121) that the unique non-null coefficients of $\vec{v}$ are given by $(1/(2iq_n^+(mh)), -1/(2iq_n^-(mh)))$ and hence

$$\vec{v}^*\vec{v} = \left|\frac{1}{2i}e^{-i\frac{n\pi}{L}mh}\right|^2 + \left|\frac{1}{2i}e^{+i\frac{n\pi}{L}mh}\right|^2 = \frac{1}{2},$$

and consequently, it has been shown that there exists $\vec{v} \neq \vec{0}$ such that

$$\frac{\vec{v}^*\mathcal{A}\vec{v}}{\vec{v}^*\vec{v}} \le 2Ch^3.$$

Secondly, a different vector coordinate $\vec{v}$ is taken into account. In this case, fixed $m \in \{0,\dots,M\}$ and $n = \tilde{N}$, consider the coordinate vector $\vec{v} \in \mathbb{C}^{2(M+1)N}$ associated to the function of the discrete space X_h given by $v(x_1) = \varphi_m\cos(n\pi(x_1 - mh)/L)$, which corresponds to the linear combination of basis functions

$$v = \frac{\varphi_m}{2}\left(\frac{q_n^+}{q_n^+(mh)} + \frac{q_n^-}{q_n^-(mh)}\right) \in \mathrm{X}_h. \tag{3.122}$$

It holds

$$\vec{v}^*\mathcal{A}\vec{v} = \int_0^L v(x_1)\bar{v}(x_1)\,\mathrm{d}x_1 = \int_0^L \varphi_m^2(x_1)\cos^2\left(\frac{n\pi(x_1-mh)}{L}\right)\mathrm{d}x_1$$
$$= 2\int_0^h \frac{s^2}{h^2}\cos^2\left(\frac{n\pi s}{L}\right)\mathrm{d}s \ge \tilde{C}\int_0^h \frac{s^2}{h^2}\,\mathrm{d}s \ge \tilde{C}h, \tag{3.123}$$

where each occurrence of constant $\tilde{C}$ could denote a different value independent of h (only dependent on L and $n = \tilde{N}$). To obtain the estimate above, it has been used a strictly positive lower bound for the cosine function around in the compact interval $[0,h] \subset [0, L/(2(n+1))]$ (where it is ensured that $\cos(n\pi s/L)$ is strictly positive for any n). Now, it is straightforward to show from (3.122) that the unique non-null coefficients of $\vec{v}$ are given by $(1/(2q_n^+(mh)), 1/(2q_n^-(mh)))$ and hence

$$\vec{v}^*\vec{v} = \left|\frac{1}{2}e^{-i\frac{n\pi}{L}mh}\right|^2 + \left|\frac{1}{2}e^{+i\frac{n\pi}{L}mh}\right|^2 = \frac{1}{2},$$

and consequently, it has been shown that there exists $\vec{v} \neq \vec{0}$ such that

$$\frac{\vec{v}^* \mathcal{A} \vec{v}}{\vec{v}^* \vec{v}} \geq 2\tilde{C}h. \tag{3.124}$$

Now, if $\lambda_{\min}$ and $\lambda_{\max}$ are respectively the largest and smallest eigenvalues of matrix $\mathcal{A}$, using the classical property of the Rayleigh quotient for hermitian complex-valued matrices (which ensures that the numerical range is a real interval with eigenvalues as endpoints [49]), it holds

$$\lambda_{\min} \leq \frac{\vec{v}^* \mathcal{A} \vec{v}}{\vec{v}^* \vec{v}} \leq \lambda_{\max} \quad \text{for all } \vec{v} \neq \vec{0}.$$

Then, from (3.124) and (3.123), there exist two positive constants C and $\tilde{C}$, independent of M and N (and hence also independent of h) such that

$$2\tilde{C}h \leq \lambda_{\max} \qquad \text{and} \qquad \lambda_{\min} \leq 2Ch^3.$$

Consequently, since $\mathcal{A}$ is a positive definite hermitian matrix (it is associated to the L^2-inner product in X_h), it is satisfied

$$\kappa(\mathcal{A}) = \frac{\lambda_{\max}}{\lambda_{\min}} \geq \frac{\tilde{C}}{C} h^{-2}, \tag{3.125}$$

and hence (3.120) is obtained. □

Consequently, from Lemma 3.4.1, since the spectrum of $\mathcal{A}$ and $\mathcal{M}$ coincides, it is obtained that $\kappa(\mathcal{M}) = \mathcal{O}(h^{-2})$, what implies an increasing behaviour of the condition number as soon as the finite element mesh is refined. This high condition number (compared with respect to the low conditioning of standard finite element methods) in the mass matrix could indicate the numerical mechanism because of the the matrix of the linear system (3.114) suffers for high condition numbers (in comparison with an standard finite element discretization). As it is reported in the following section, to mitigate as much as possible the conditioning issues, the finite element meshes have been kept as coarse as possible in most of the numerical test.

3.5 Numerical results

A wide battery of numerical test has been considered to illustrate the performance and the potential drawbacks of the proposed modal-based PUFEM method. With this aim, Section 3.5.1 includes some numerical tests are done to show the accuracy of the modal-based PUFEM method but with a discrete space that only involves Love waves rather than the complete modal space. Next, in Section 3.5.3 the numerical results illustrate the different numerical performance obtained with a discrete space that only involves Love waves or with another space that includes both, Love and interior waves.

Last Section 3.5.4 includes the numerical results obtained with the modal-based PUFEM where a complete modal based with Love and interior waves. The goals of this section are focused on three topics: (i) the illustration of the accuracy of the method for smooth and non-smooth solutions, (ii) the deterioration of the numerical results due to the high condition numbers of the discrete matrix and its potential mitigation using regularization techniques, and (iii) the accuracy of the modal-based method for solutions which are close to the constant-valued eigenmode (which is not included in the modal enrichment). In those numerical simulations where internal modes are involved, the eigenmodes used in the modal-based PUFEM discretization hold the condition

$$c_+^2 \mu_n \leq \lambda_{n,j} \leq c_0^2 \mu_n \qquad \text{for } n = 0, \dots, N,$$

with $c_0 = 2c_+$, or equivalently, it is only considered those solutions $\zeta_{n,j}$ of the dispersion equation (3.106) which belong to the interval $(c_+, 2c_+)$.

Throughout this section, the relative errors are computed using a point-wise L^∞-norm on an 5×5 equispaced grid of points $\{y_{jk}\}_{j,k=1}^5$ in the rectangular domain $[0, L] \times [-a, H]$. More precisely, the relative error is given by

$$\epsilon_h = \frac{\max\limits_{1 \leq j,k, \leq 5} |u(y_{jk}) - u_h(y_{jk})|}{\max\limits_{1 \leq j,k, \leq 5} |u(y_{jk})|},$$

where u is the exact solution of the source problem and u_h is the approximated solution computed with the proposed modal-based PUFEM method. Notice that the grid points are either on the exterior boundary $\partial\Omega$ or in the interior of the computational domain, but in any case they are not lying on the coupling interface Γ_I. Other finer grids with a larger number of points have been also considered leading to similar relative errors. To illustrate the approximated solution computed by means of the modal-based PUFEM method, for every test the real part of the approximation is plotted on a 17×17 equispaced grid of points $\{y_{jk}\}_{j,k=1}^{17}$ in the rectangular domain $[0, L] \times [-a, H]$. Additionally, the point-wise relative error with respect to L^∞-norm is also plotted in the computational domain Ω.

3.5.1 Accuracy of the method for x_1-dependent problems

In order to check the accuracy of the method, first only Love eigenmodes have been considered in the discretization, i.e, the discrete space used in the numerical test presented in this section is given by $\mathrm{X}_h^{\mathcal{L}}$, which only contains Love waves. Despite the Loves waves does not form a Hilbert basis, the combination of the PUFEM method in the x_1-direction with this modal enrichment allows to reach accurate results for certain solutions that only depend on the x_1 spatial coordinate.

For this first numerical test, the problem (3.10)-(3.14) has been settled with constant speed of sound $c = 1$ in Ω, the angular frequency $\omega = \pi$ and the parameter $\beta = 1$ (absorbing boundary conditions on $\Gamma_e \cup \Gamma_s$). The geometric dimensions of the computational domain $\Omega = (0, L) \times (-a, H)$ are given by $a = 0.2$, $H = 0.8$, $L = 1$.

In this test, instead of computing the solution of the dispersion equation for Love waves, it is assumed that there exists an eigenmode almost independent of the x_2-direction (i.e., except for a round-off error ε). More precisely, it is used the discrete space $\mathrm{X}_h^{\mathcal{L}}$ with only the Love wave $w_{1,1}$ (what implies that $N = 1$) and where the wave numbers are given by $K_-^{1,1} = K_+^{1,1} = \varepsilon = 10^{-13}$. The boundary data in problem (3.10)-(3.14) has been chosen such that the exact solution is $u(x_1, x_2) = e^{-i\omega x_1/c}$. Obviously, if $K_-^{1,1} = K_+^{1,1} = 0$ then the exact solution u would belong to the discrete space $\mathrm{X}_h^{\mathcal{L}}$ and consequently the error would be null theoretically. However, due to the round-off error introduced in the modal basis element $w_{1,1}$, the relative errors should be expected of order ε since the exact solution coincides with the expression of $q_1^-(x_1)$.

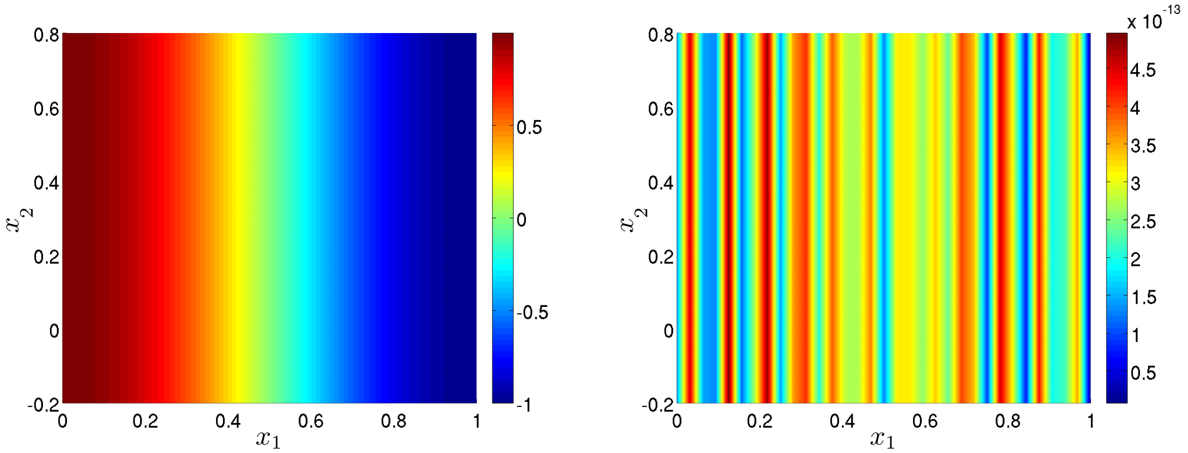

Figure 3.1: Real part of the approximate solution (left) and relative error (right) obtained from the modal-based PUFEM method with a one-dimensional mesh of ten elements (i.e. $M = 10$), for a discretization involving only the first Love mode $w_{1,1}$, where it has been settled $K_-^{1,1} = K_+^{1,1} = 10^{-13}$. The exact solution coincides with q_1^-.

Table 3.1 show the relative errors and the condition number obtained with three different finite element meshes of $M = 1$, 10 and 100 elements. For the first two meshes, the relative error has order $\mathcal{O}(\delta)$ as it would be expected. In addition, it can be observed that the condition number grows with the mesh size, what also produces a growth on the relative error. Figure 3.1 illustrates the real part of the approximated solution and the relative error for $M = 10$.

M	dof	ϵ_h	κ
1	4	3.72×10^{-16}	5.2×10^{0}
10	22	1.40×10^{-13}	8.1×10^{5}
100	202	1.50×10^{-11}	7.6×10^{11}

Table 3.1: Relative error ϵ_h and condition number κ for different finite element mesh with M elements and obtained for a discretization involving only the first Love mode $w_{1,1}$, where it has been settled $K_-^{1,1} = K_+^{1,1} = 10^{-13}$. The exact solution coincides with q_1^-.

Now, an additional perturbation parameter $\delta > 0$ on the wave number is introduced in the expression of the exact solution in order to avoid that it coincides with $q_1^-(x_1)$, this is, the boundary condition of problem (3.10)-(3.14) are chosen such that the exact solution is given by $u(\boldsymbol{x}) = e^{-i(\omega/c+\delta)x_1}$. From Table 3.2 it can be concluded that the proposed modal-based PUFEM recovers the same order of convergence $\mathcal{O}(\delta^2 h^2)$ as it has been analysed for a planewave-based PUFEM method in one-dimensional problems (see [25] for further details). Figure 3.2 illustrates the real part of the approximated solution and the relative error for $M = 10$.

M	δ	dof	ϵ_h	κ
	10^{-3}	4	2.08×10^{-8}	5.2×10^{0}
1	10^{-2}	4	2.08×10^{-6}	5.2×10^{0}
	10^{-1}	4	2.12×10^{-4}	5.2×10^{0}
	10^{-3}	22	1.15×10^{-10}	8.1×10^{5}
10	10^{-2}	22	1.10×10^{-8}	8.1×10^{5}
	10^{-1}	22	1.12×10^{-6}	8.1×10^{5}
	10^{-3}	202	4.13×10^{-9}	7.6×10^{11}
100	10^{-2}	202	3.28×10^{-10}	7.6×10^{11}
	10^{-1}	202	1.46×10^{-9}	7.6×10^{11}

Table 3.2: Relative error ϵ_h and condition number κ for different finite element mesh with M elements and obtained for a discretization involving only the first Love mode $w_{1,1}$, where it has been settled $K_-^{1,1} = K_+^{1,1} = 10^{-13}$. The exact solution is a perturbation of q_1^- multiplying it by the factor $e^{i\delta x_1}$.

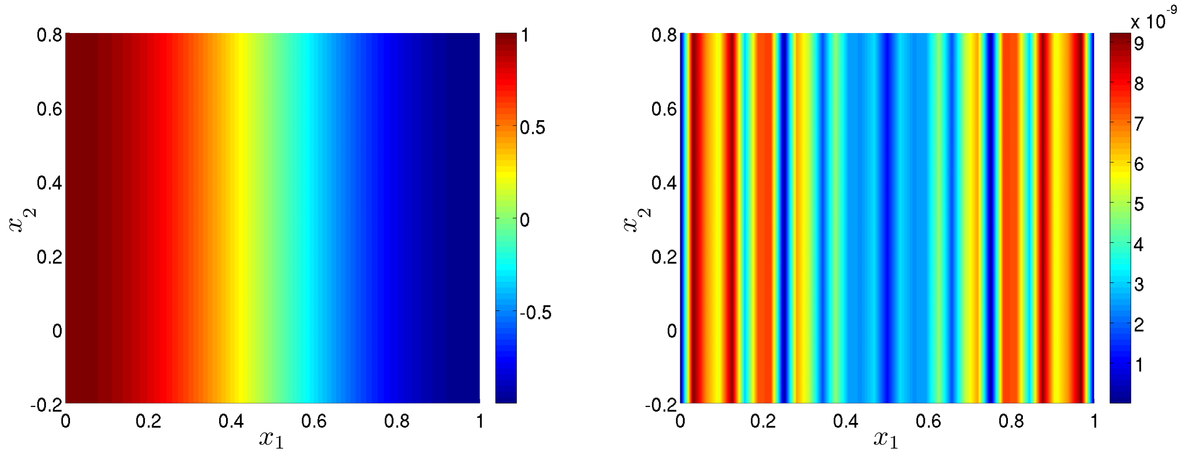

Figure 3.2: Real part of the approximate solution (left) and relative error (right) obtained from the modal-based PUFEM method with a one-dimensional mesh of ten elements (i.e. $M = 10$), for a discretization involving only the first Love mode $w_{1,1}$, where it has been settled $K_-^{1,1} = K_+^{1,1} = 10^{-13}$. The exact solution is a perturbation of q_1^- multiplying it by the factor $e^{i\delta x_1}$, with $\delta = 10^{-2}$.

Finally, to illustrate the numerical behaviour of the proposed modal-based PUFEM method to handle x_1-dependent solutions, it has been considered the adequate boundary functions g and r and source term f to obtain as exact solution the following expression:

$$u(x_1, x_2) = \frac{1}{2\omega^2}(e^{-i\omega(x_1-1)/c} + e^{i\omega x_1/c} - 2). \tag{3.126}$$

In particular, due to the constant term in the exact solution, $f(x_1, x_2) = 1$. As it can be observed in Table 3.3 the typical order of convergence of a linear finite element method $\mathcal{O}(h^2)$ is recovered (it should be notice that the piecewise linear finite element basis is been used in the present method as partition of unity). Figure 3.3 illustrates the real part of the approximated solution and the relative error for $M = 10$.

M	dof	ϵ_h	κ
1	4	4.13×10^{-2}	5.2×10^{0}
10	22	2.48×10^{-4}	8.1×10^{5}
100	202	2.73×10^{-7}	7.6×10^{11}

Table 3.3: Relative error ϵ_h and condition number κ for different finite element mesh with M elements and obtained for a discretization involving only the first Love mode $w_{1,1}$, where it has been settled $K_-^{1,1} = K_+^{1,1} = 10^{-13}$. The exact solution consists in the addition of one-dimensional plane waves and a constant term (3.126).

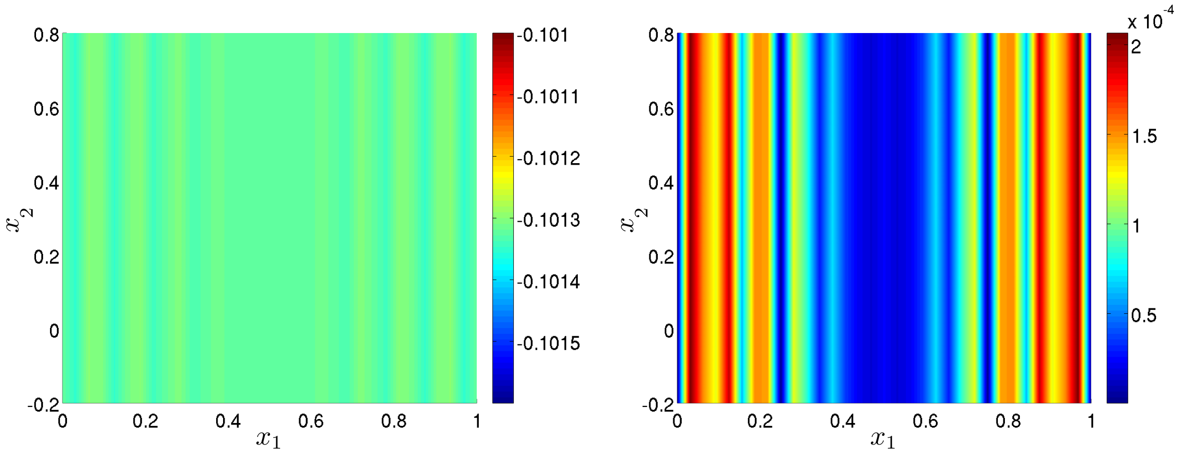

Figure 3.3: Real part of the approximate solution (left) and relative error (right) obtained from the modal-based PUFEM method with a one-dimensional mesh of ten elements (i.e. $M = 10$), for a discretization involving only the first Love mode $w_{1,1}$, where it has been settled $K_-^{1,1} = K_+^{1,1} = 10^{-13}$. The exact solution consists in the addition of one-dimensional plane waves and a constant term (3.126).

3.5.2 Numerical comparison with a two-dimensional finite element method

Instead of considering homogeneous speed of sound profiles as it has been used previously, now the problem (3.10)-(3.14) is settled such as $c_- = 1/2$ and $c_+ = 1$ in $\Omega_- = (0, L) \times (-a, 0)$ and $\Omega_+ = (0, L) \times (0, H)$, respectively. Again, the angular frequency is given by $\omega = \pi$ and Robin boundary conditions are assumed in the left and right boundaries ($\beta = 1$) on the square domain $\Omega = (0, L) \times (-a, H)$, being $a = 0.2$, $H = 0.8$, $L = 1$.

For this numerical test, the source term has been chosen as $f = 0$, the boundary terms are given by $g = 0$ on $\Gamma_e \cup \Gamma_s$, $r = 0$ on Γ_- and $r = 1$ on Γ_+. With these boundary conditions, the exact solution is not known in closed form and hence it is not possible to compute the relative error. As alternative, the numerical comparison of the approximated solution u_h computed with the modal-based PUFEM method has been made with respect to the piecewise linear finite element approximation u_{fem} in two dimensions.

For this purpose, the PUFEM solution u_h is computed taking into account a one-dimensional mesh with $M = 10$ elements and the first ten families of Love eigenmodes, this is, the definition of $\mathrm{X}_h^{\mathcal{L}}$ involves the Love modes $w_{n,j}$ with $(n, j) \in \{k\} \times \mathcal{L}_k$ for $k = 1, \ldots, N = 10$. To compare both approximations, the relative difference d_h of the PUFEM solution u_h and the FEM approximation u_{FEM} is computed as the maximum norm of the difference between both functions evaluated at the nodes of the two-dimensional finite element mesh used to compute u_{fem} and normalized with respect to the maximum point-wise value of u_h on those nodes.

A variety of two-dimensional regular triangular meshes with different maximum diameter $h_{\max}$ has been used. Table 3.4 illustrates that the relative difference in L^∞-norm behaves like $\mathcal{O}(h_{\max})$ as it is predicted for the approximation of piecewise linear finite element methods applied for regular solutions in two dimensions (see [11, Chapter 3, Section 3.3]). Such behaviour indicates that the relative difference is dominated by the finite element error which is larger than the error coming from the PUFEM approximation even when the value of $M = 10$ is fixed.

$h_{\max}$	d_h
1.2×10^{-1}	4.43×10^{-2}
6.2×10^{-1}	1.16×10^{-2}
3.1×10^{-2}	2.30×10^{-3}
1.5×10^{-2}	9.19×10^{-4}
7.7×10^{-3}	1.35×10^{-3}
3.8×10^{-3}	1.43×10^{-3}
1.9×10^{-3}	1.45×10^{-3}

Table 3.4: Relative difference d_h between the modal-based PUFEM approximation u_h computed with $M = 10$, $N = 10$ and the standard piecewise linear finite element approximation with different mesh sizes $h_{\max}$. The exact solution is not known in closed form.

3.5.3 Numerical comparison of discrete spaces with or without internal waves

To illustrate the relevance of including the internal waves on the discrete space (and consequently use the whole Hilbert basis in the enrichment of the PUFEM method), a detailed comparison between the modal-based method have being carried out using the discrete spaces $\mathrm{X}_h^{\mathcal{L}}$ (only considering Love eigenmodes) and X_h (using Love and internal eigenmodes).

As it has been used previously, problem (3.10)-(3.14) is settled with angular frequency $\omega = \pi$, but now pure Neumann boundary conditions have been considered in the whole boundary $\partial\Omega$ (taking $\beta = 0$). The square domain $\Omega = (0, L) \times (-a, H)$ with $a = 0.2$, $H = 0.8$, $L = 1$ is split in two subdomains where the speed of sound is given by $c_- = 1/2$ in $\Omega_- = (0, L) \times (-a, 0)$ and $c_+ = 1$ in $\Omega_+ = (0, L) \times (0, H)$. The source term is given by

$$f(x_1, x_2) = \begin{cases} 1 & \text{for } (x_1, x_2) \in \Omega_+, \\ x_2 & \text{for } (x_1, x_2) \in \Omega_-, \end{cases}$$

and the boundary functions are fixed to $g = 0$ and $r = 0$. Under these boundary conditions and this source term, it is straightforward to compute the exact solution in closed form. More precisely, the exact solution is given by

$$u(x_1, x_2) = \begin{cases} A_+ e^{-i\omega x_2/c_+} + B_+ e^{i\omega x_2/c_+} - \dfrac{1}{\omega^2} & \text{if } (x_1, x_2) \in \Omega_+, \\ A_- e^{-i\omega x_2/c_-} + B_- e^{i\omega x_2/c_-} - \dfrac{x_2}{\omega^2} & \text{if } (x_1, x_2) \in \Omega_-, \end{cases} \tag{3.127}$$

being A_+, B_+, A_-, B_- coefficients that are determined solving the system

$$\begin{pmatrix} 0 & 0 & e^{-i\omega H/c_+} & -e^{i\omega H/c_+} \\ e^{i\omega a/c_-} & -e^{-i\omega a/c_-} & 0 & 0 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -c_+/c_- & c_+/c_- \end{pmatrix} \begin{pmatrix} A_- \\ B_- \\ A_+ \\ B_+ \end{pmatrix} = \begin{pmatrix} 0 \\ ic_-/\omega^3 \\ -1/\omega^2 \\ ic_-/\omega^3 \end{pmatrix}.$$

that results from applying the boundary conditions and the coupling conditions.

Table 3.5 shows the relative error for both, an approximated solution in the discrete space $\mathrm{X}_h^{\mathcal{L}}$ involving only Love eigenmodes and an approximated solution in the discrete space X_h with Love and internal eigenmodes. As it is expected, if internal and Love waves are included in the discretization (this is, if a whole Hilbert basis is used in the discretization) in the discretization then the approximated PUFEM solutions are much more accurate than those computed with only Love waves. This conclusion is valid for any value of mesh size M and any number of eigenmodes N as it can be checked in Table 3.5. Figures 3.4 and 3.5 illustrate the real part of the approximated solution and the relative error for $M = 4$ and $N = 10$ computed using the discrete space $\mathrm{X}_h^{\mathcal{L}}$ and X_h, respectively.

		$\mathrm{X}_h^{\mathcal{L}}$ (without internal waves)			X_h (with internal waves)		
M	N	dof	ϵ_h	κ	dof	ϵ_h	κ
1	1	4	1.26×10^{0}	4.1×10^{0}	12	9.42×10^{-2}	1.5×10^{2}
	2	8	1.35×10^{0}	2.3×10^{1}	32	1.25×10^{-2}	3.8×10^{6}
	3	16	1.65×10^{0}	1.1×10^{3}	60	2.14×10^{-3}	1.5×10^{8}
	4	24	2.40×10^{0}	8.5×10^{3}	96	1.11×10^{-4}	1.5×10^{11}
	5	32	2.95×10^{0}	5.0×10^{4}	140	3.00×10^{-5}	3.1×10^{13}
	10	100	1.87×10^{0}	9.4×10^{8}	500	1.55×10^{-5}	1.2×10^{19}
2	1	6	1.25×10^{0}	4.8×10^{1}	18	4.19×10^{-2}	2.2×10^{3}
	2	12	1.81×10^{0}	3.2×10^{2}	48	6.21×10^{-3}	6.7×10^{7}
	3	24	2.70×10^{0}	1.3×10^{4}	90	4.80×10^{-4}	2.5×10^{10}
	4	36	3.69×10^{0}	1.8×10^{5}	144	3.28×10^{-5}	4.9×10^{12}
	5	48	9.51×10^{0}	1.5×10^{6}	210	2.23×10^{-5}	2.5×10^{15}
	10	150	3.44×10^{-2}	1.6×10^{11}	750	8.00×10^{-6}	2.1×10^{20}
3	1	8	1.25×10^{0}	5.9×10^{2}	24	1.44×10^{-2}	2.2×10^{4}
	2	16	1.95×10^{0}	2.7×10^{3}	64	4.60×10^{-3}	8.8×10^{7}
	3	32	3.21×10^{0}	1.5×10^{5}	120	2.43×10^{-4}	1.3×10^{11}
	4	48	1.93×10^{2}	2.4×10^{6}	192	3.11×10^{-5}	3.3×10^{13}
	5	64	7.30×10^{-1}	1.5×10^{7}	280	1.88×10^{-5}	6.7×10^{15}
	10	200	7.12×10^{-3}	1.0×10^{11}	1000	5.47×10^{-6}	1.9×10^{20}
4	1	10	1.25×10^{0}	3.4×10^{3}	30	1.34×10^{-2}	1.2×10^{5}
	2	20	2.38×10^{0}	1.5×10^{4}	80	2.63×10^{-3}	5.4×10^{8}
	3	40	3.53×10^{0}	1.0×10^{5}	150	7.28×10^{-5}	2.7×10^{12}
	4	60	1.92×10^{0}	2.4×10^{7}	240	1.86×10^{-5}	4.4×10^{14}
	5	80	4.45×10^{-1}	1.1×10^{8}	350	1.72×10^{-5}	5.2×10^{16}
	10	250	3.30×10^{-3}	1.3×10^{14}	1250	3.88×10^{-6}	1.0×10^{23}
5	1	12	1.25×10^{0}	1.3×10^{4}	36	1.28×10^{-2}	4.5×10^{5}
	2	24	2.48×10^{0}	5.6×10^{4}	96	1.31×10^{-3}	3.2×10^{9}
	3	48	4.52×10^{0}	4.5×10^{6}	180	4.88×10^{-5}	2.1×10^{13}
	4	72	9.20×10^{-1}	1.6×10^{8}	288	1.95×10^{-5}	3.0×10^{15}
	5	96	2.60×10^{-1}	6.8×10^{8}	420	1.80×10^{-5}	4.1×10^{18}
	10	300	1.51×10^{-3}	6.9×10^{14}	1550	5.43×10^{-6}	7.3×10^{19}
10	1	22	1.25×10^{0}	8.1×10^{5}	66	1.11×10^{-2}	2.7×10^{7}
	2	44	2.61×10^{0}	3.6×10^{6}	176	1.83×10^{-4}	1.0×10^{12}
	3	88	5.17×10^{0}	3.5×10^{8}	330	2.73×10^{-5}	3.6×10^{15}
	4	132	3.76×10^{-1}	8.8×10^{10}	528	1.44×10^{-5}	1.2×10^{17}
	5	176	6.62×10^{-2}	4.1×10^{11}	770	7.74×10^{-6}	2.0×10^{18}
	10	550	3.69×10^{-4}	6.2×10^{16}	2750	1.97×10^{-5}	2.2×10^{19}

Table 3.5: Comparison of the relative error ϵ_h and the condition number κ for two different approximated PUFEM solutions, one in the discrete space $\mathrm{X}_h^{\mathcal{L}}$ that only includes Love waves, and another one belonging to the discrete space X_h with both Love and interior eigenmodes. The numerical results are shown for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation.

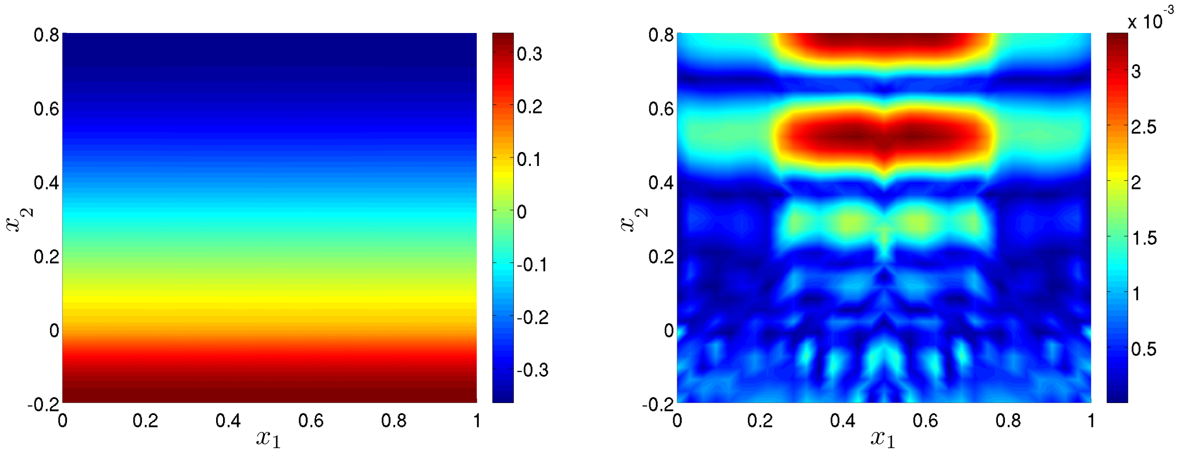

Figure 3.4: Real part of the approximate solution (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 10\} \times \{\mathcal{L}_n\}$ (i.e. $N = 10$). The exact solution is given by (3.127).

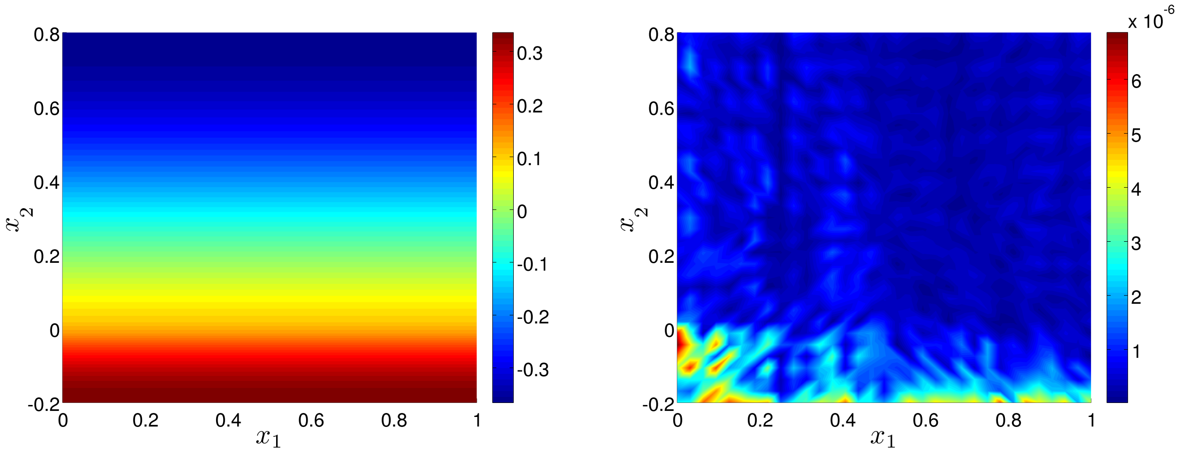

Figure 3.5: Real part of the approximate solution (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 10\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 10$). The exact solution is given by (3.127).

It is also relevant that, if the relative errors obtained with both discrete spaces are compared for similar values of degrees of freedom (and hence with almost similar computational cost), the numerical results reached with the discrete X_h outperforms those results obtained with only Love waves in $\mathrm{X}_h^{\mathcal{L}}$. In conclusion, the numerical results described throughout the rest of this section, will take into account both Love and interior eigenmodes and hence the modal-based PUFEM discretization will always use the discrete space X_h.

3.5.4 Consistency of the modal-based PUFEM method with Love and internal waves

In order to check the consistency of the PUFEM method, using the discrete space with both internal and Love waves, the relative error has been analysed in two numerical tests where the exact solutions belong to the discrete space X_h. In the first case the exact solution is given by a Love wave and in the second case, an internal wave is imposed as exact solution.

For the first numerical test, problem (3.10)-(3.14) is settled with an angular frequency $\omega = \pi$, and assuming pure homogeneous Neumann boundary conditions (this is, the parameter β is null and the boundary functions are fixed to $g = 0$ and $r = 0$. As in previous cases described above, the speed of sound is given by $c_- = 1/2$ in $\Omega_- = (0, L) \times (-a, 0)$ and $c_+ = 1$ in $\Omega_+ = (0, L) \times (0, H)$ being $a = 0.2$, $H = 0.8$, $L = 1$.

The first solution considered is the Love eigenmode with the lowest eigenvalue λ_{1,j_1}, i.e., $u = w_{1,j_1}$ being j_1 the first index in the sorted set $\mathcal{L}_1$. To obtain such exact solution, the source term in (3.10) has been fixed to $f = (\lambda_{1,j_1} - \omega^2)w_{1,j_1}$. Obviously, from a theoretical point of view, since the exact solution belongs to the discrete space, the numerical approximation error should be null. However, due to the round-off errors introduced by the double precision arithmetic representation, the relative errors shown in the first rows of Table 3.6 are of magnitude $\mathcal{O}(10^{-13})$. The numerical results of Table 3.6 also illustrate how the relative errors are increased as the one-dimensional mesh is refined (M is increased) and more families of eigenmodes are involved in the discrete space X_h (value of N is increased). In both cases, since the condition number of the linear system grows, the relative errors are also increased. Despite of this well-known phenomena, it should be remarked that five digits of accuracy are kept even in those numerical approximations where the condition number is as high as $\mathcal{O}(10^{18})$. Figure 3.6 illustrates the real part of the approximated solution and the relative error for $M = 10$ and $N = 3$.

M	N	dof	ϵ_h	κ
1	1	12	2.43×10^{-15}	1.5×10^{2}
	2	32	2.01×10^{-15}	3.8×10^{6}
	3	60	1.46×10^{-14}	1.5×10^{8}
	4	96	4.05×10^{-13}	1.5×10^{11}
	5	140	4.74×10^{-12}	3.1×10^{13}
10	1	66	2.83×10^{-13}	2.7×10^{7}
	2	176	4.65×10^{-12}	1.0×10^{12}
	3	330	1.22×10^{-10}	3.6×10^{15}
	4	528	3.10×10^{-9}	1.2×10^{17}
	5	770	1.25×10^{-8}	2.0×10^{18}
100	1	606	2.01×10^{-11}	2.5×10^{13}
	2	1616	7.11×10^{-9}	2.0×10^{16}
	3	3030	5.45×10^{-6}	1.4×10^{18}
	4	4848	1.16×10^{-7}	5.2×10^{18}
	5	7070	3.56×10^{-6}	9.4×10^{18}

Table 3.6: Relative error ϵ_h and the condition number κ for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation. The exact solution is given by the Love wave with the lowest eigenvalue.

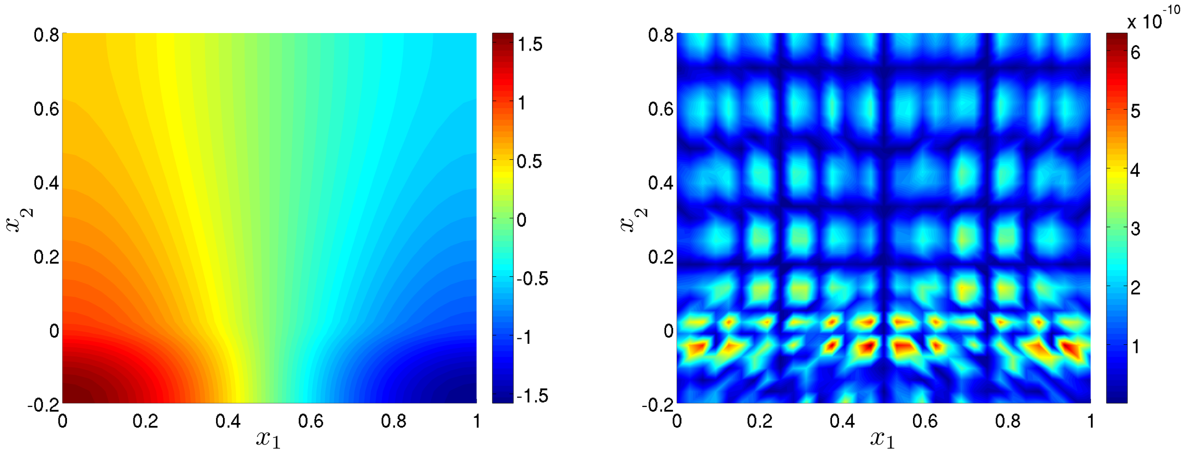

Figure 3.6: Real part of the approximate solution (left) and relative error (right) obtained from the modal-based PUFEM method with a one-dimensional mesh of ten elements (i.e. $M = 10$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, 2, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is the Love mode associated to the lowest eigenvalue.

For the second numerical test, the exact solution is fixed to be the non-constant internal eigenmode with the lowest associated eigenvalue, this is $u = w_{1,k_1}$ being k_1 the first index in the sorted subset $\mathcal{I}_1^{J_1}$. To obtain such exact solution, the source term is given by $f = (\lambda_{1,k_1} - \omega^2)w_{1,k_1}$. As it has been discussed previously, since the exact solution is contained in the discrete space X_h then the relative error should be null from a theoretically point of view. As in the previous numerical test, Table 3.7 shows that the first rows corresponding to coarse meshes and a reduced number of eigenmode families in the discrete space X_h, the relative errors are of magnitude $\mathcal{O}(10^{-13})$. Again, such non-null relative errors are produced by the amplification of round-off errors in the numerical solution of a linear system with large condition numbers. Figure 3.7 illustrates the real part of the approximated solution and the relative error for $M = 10$ and $N = 3$.

M	N	dof	ϵ_h	κ
1	1	12	1.49×10^{-15}	1.5×10^{2}
	2	32	4.92×10^{-15}	3.8×10^{6}
	3	60	3.81×10^{-14}	1.5×10^{8}
	4	96	5.99×10^{-13}	1.5×10^{11}
	5	140	6.60×10^{-12}	3.1×10^{13}
10	1	66	3.74×10^{-13}	2.7×10^{7}
	2	176	1.04×10^{-11}	1.0×10^{12}
	3	330	7.28×10^{-10}	3.6×10^{15}
	4	528	6.01×10^{-9}	1.2×10^{17}
	5	770	9.30×10^{-8}	2.0×10^{18}
100	1	606	2.78×10^{-11}	2.5×10^{13}
	2	1616	1.11×10^{-8}	2.0×10^{16}
	3	3030	1.06×10^{-5}	1.4×10^{18}
	4	4848	3.55×10^{-7}	5.1×10^{18}
	5	7070	4.92×10^{-6}	9.5×10^{18}

Table 3.7: Relative error ϵ_h and the condition number κ for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation. The exact solution is given by the interior wave with the lowest eigenvalue.

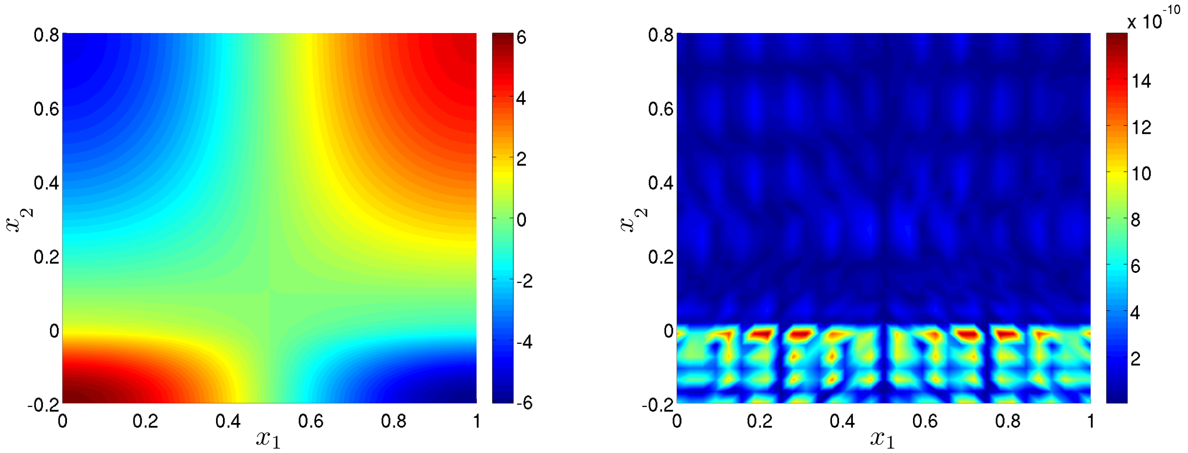

Figure 3.7: Real part of the approximate solution (left) and relative error (right) obtained from the modal-based PUFEM method with a one-dimensional mesh of one element (i.e. $M = 10$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, 2, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is the interior mode associated to the lowest eigenvalue.

3.5.5 Influence of the condition number on the numerical results

In previous subsections, it has been reported that the modal-based PUFEM method suffers for large condition numbers in the linear systems to be solved. Such issue represents a potential drawback in the use of direct LU-based solvers. From the numerical results described in the sections above, this conditioning problem is more relevant as soon as the one-dimensional finite element mesh is refined and the number of eigenmodes involved in the discrete space is increased. However, there exists a number of methodologies to deal with high condition numbers and try to mitigate the amplification of the round-off errors on the solution of the linear systems. In the present section, three different regularization techniques are used: a naive damped strategy, the classical Tikhonov filtering, and the truncated singular value decomposition method. The latter has been already used for solving linear systems with large condition numbers in the context of two-dimensional PUFEM discretizations (see [16]).

The LU-based solver and the three regularization techniques that are used in this subsection are described as follows. For simplicity in the notation, let $Aw = b$ the system that has to be solved, being A a square matrix of size n, b the right hand side vector and w the solution vector. The LU-based solver used consists in writing the matrix A as $A = LU$, where L is a permutation of a lower triangular matrix and U is an upper triangular matrix. Then, the vector w is calculated by solving two permuted triangular systems, $w = U\backslash(L\backslash b)$. The approximate solution in e_h has been calculated with this LU solver.

The naive damped strategy used in this subsection consists in solving $(A + \lambda_\mathrm{d} I)w = b$ instead of $Aw = b$, being λ_d a parameter. To solve the modified system, the LU solver described above is used. The approximate solution in e_d has been calculated with this strategy.

The classical Tikhonov filtering (see [43], [47], [48]) consists in finding

$$\min\left\{\|Aw-b\|_2^2-\lambda_{\mathrm{t}}\|w-w^*\|_2^2\right\},$$

where w^* is an initial estimate of the solution and λ_{t} is the regularization parameter. The approximate solutions in e_{t} have been calculated with this filtering.

The truncated SVD procedure (see [22], [23], [50] for more detail) consists in writing the matrix A as

$$A=UDV^t=\sum_{j=1}^{n}\boldsymbol{u}_j\sigma_j\boldsymbol{v}_j^t,$$

being $U=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)$ and $V=(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n)$ matrices with orthonormal columns, $U^tU=V^tV=I_n$, and where $D=\operatorname{diag}(\sigma_1,\ldots,\sigma_n)$ has non-negative diagonal elements appearing in non-increasing order such that $\sigma_1\geq\ldots\geq\sigma_n\geq 0$. The numbers σ_j are the singular values of A, while the vectors $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are the left and right singular vectors of A, respectively, for all $j=1,\ldots,n$. Then, the truncation consists in choosing a parameter λ_{SVD} and consider just the singular values that are larger than λ_{SVD}, i.e., in considering the truncated matrix

$$A_T=\sum_{j=1}^{k}\boldsymbol{u}_j\sigma_j\boldsymbol{v}_j^t,$$

for $\sigma_1\geq\ldots\sigma_k>\lambda_{\mathrm{SVD}}$. The approximate solutions in e_{SVD} have been calculated with this procedure.

To choose the regularization parameters, several techniques can be applied. In this work, the L-curve and the generalized cross validation techniques have been used. The so-called L-curve is a plot of the norm $\|w_{\mathrm{reg}}\|_2$ of the regularized solution (with any of the three techniques) versus the corresponding residual norm $\|Aw_{\mathrm{reg}}-b\|_2$. For discrete ill-posed problems (see [33], [37] for more details) the L-curve, when plotted in log-log scale, has a characteristic L-shaped appearance, with a distinct corner separating the vertical and the horizontal parts of the curve. This corner gives the optimal value for the regularization parameter. The generalized cross validation (GCV) technique has into account the fact that the choice of the regularization parameter should be independent of an orthogonal transformation of the right hand side b (see [51] for more details).

As it has been used previously, problem (3.10)-(3.14) is settled with angular frequency $\omega=\pi$, but now pure Neumann boundary conditions have been considered in the whole boundary $\partial\Omega$ (taking $\beta=0$). The square domain $\Omega=(0,L)\times(-a,H)$ with $a=0.2$, $H=0.8$, $L=1$ is split in two subdomains where the speed of sound is given by $c_-=1/2$ in $\Omega_-=(0,L)\times(-a,0)$ and $c_+=1$ in $\Omega_+=(0,L)\times(0,H)$. Now, to avoid those exact solutions which could belong to X_h, the source term is given by

$$f(x_1,x_2)=\begin{cases}\cos\left(\dfrac{3\pi x_1}{L}\right) & \text{for } (x_1,x_2)\in\Omega_+,\\[2ex] (1+x_2)\cos\left(\dfrac{3\pi x_1}{L}\right) & \text{for } (x_1,x_2)\in\Omega_-,\end{cases}$$

and the boundary functions are fixed to $g = 0$ and $r = 0$. Under these boundary conditions and this source term, it is straightforward to compute the exact solution in closed form and it is clear that it does not belong to X_h. More precisely, the exact solution is given by

$$u(x_1, x_2) = \cos\left(\frac{3\pi x_1}{L}\right) \begin{cases} A_+ e^{-i\alpha_+ x_2} + B_+ e^{i\alpha_+ x_2} - \dfrac{1}{c_+^2 \alpha_+^2} & \text{if } (x_1, x_2) \in \Omega_+, \\ A_- e^{-i\alpha_- x_2} + B_- e^{i\alpha_- x_2} - \dfrac{1 + x_2}{c_-^2 \alpha_-^2} & \text{if } (x_1, x_2) \in \Omega_-, \end{cases} \tag{3.128}$$

where

$$\alpha_+ = \sqrt{\frac{\omega^2}{c_+^2} - \frac{9\pi^2}{L^2}}, \qquad \alpha_- = \sqrt{\frac{\omega^2}{c_-^2} - \frac{9\pi^2}{L^2}},$$

and being A_+, B_+, A_-, B_- coefficients that are determined solving the system

$$\begin{pmatrix} 0 & 0 & -e^{-i\alpha_+ H} & e^{i\alpha_+ H} \\ -i\alpha_- e^{i\alpha_- a} & i\alpha_- e^{-i\alpha_- a} & 0 & 0 \\ -1 & -1 & 1 & 1 \\ ic_-^2 \alpha_- & -ic_-^2 \alpha_- & -ic_+^2 \alpha_+ & ic_+^2 \alpha_+ \end{pmatrix} \begin{pmatrix} A_- \\ B_- \\ A_+ \\ B_+ \end{pmatrix} = \begin{pmatrix} 0 \\ 1/c_-^2 \alpha_-^2 \\ 1/c_+^2 \alpha_+^2 - 1/c_-^2 \alpha_-^2 \\ -1/\alpha_-^2 \end{pmatrix}.$$

that results from applying the boundary conditions and the coupling conditions.

Firstly, Table 3.8 shows the comparison of the relative errors obtained with a LU-based direct solver and with the naive damped algorithm (adding a damping coefficient λ_{d} on the diagonal entries of the matrix). It can be observed that both methodologies lead to similar relative errors without any significant advantage between both methods. Analogous conclusions can be deduced from the numerical results reported for the truncated singular value decomposition (see Tables 3.11 and 3.12), and the Tikhonov filtering technique (see Tables 3.9, and 3.10), both in the case where the regularization parameter is chosen using a L-curve strategy and in that one where the generalized cross validation (GCV) method is utilized.

M	N	dof	λ_{d}	e_r	e_{d}	κ
1	1	12	8.7×10^{-7}	2.98×10^{-1}	2.98×10^{-1}	1.4×10^{2}
	2	32	8.7×10^{-7}	2.09×10^{-2}	2.09×10^{-2}	3.8×10^{6}
	3	60	8.7×10^{-7}	1.13×10^{-3}	1.14×10^{-3}	1.5×10^{8}
	4	96	1.0×10^{-8}	7.76×10^{-5}	8.27×10^{-5}	1.5×10^{11}
	5	140	5.3×10^{-10}	1.24×10^{-4}	1.25×10^{-4}	3.1×10^{13}
	10	500	2.3×10^{-12}	1.05×10^{-4}	3.89×10^{-5}	1.2×10^{19}
2	1	18	8.7×10^{-7}	9.88×10^{-2}	9.88×10^{-2}	2.2×10^{3}
	2	48	8.7×10^{-7}	3.44×10^{-3}	3.45×10^{-3}	6.7×10^{7}
	3	90	8.7×10^{-7}	2.27×10^{-4}	3.15×10^{-4}	2.5×10^{10}
	4	144	8.1×10^{-10}	1.06×10^{-4}	1.06×10^{-4}	4.9×10^{12}
	5	210	1.1×10^{-10}	8.56×10^{-5}	9.09×10^{-5}	2.5×10^{15}
	10	750	1.1×10^{-12}	3.94×10^{-5}	3.58×10^{-5}	1.3×10^{20}
3	1	24	8.7×10^{-7}	7.89×10^{-2}	7.89×10^{-2}	2.2×10^{4}
	2	64	8.7×10^{-7}	5.76×10^{-3}	5.77×10^{-3}	8.8×10^{7}
	3	120	8.7×10^{-7}	2.03×10^{-4}	2.92×10^{-4}	1.3×10^{11}
	4	192	1.4×10^{-9}	1.03×10^{-4}	1.07×10^{-4}	3.3×10^{13}
	5	280	8.7×10^{-11}	5.57×10^{-5}	6.73×10^{-5}	6.7×10^{15}
	10	1000	3.5×10^{-12}	4.63×10^{-5}	3.04×10^{-5}	4.4×10^{19}
4	1	30	8.7×10^{-7}	2.27×10^{-2}	2.27×10^{-2}	1.2×10^{5}
	2	80	5.0×10^{-7}	3.48×10^{-3}	3.47×10^{-3}	5.4×10^{8}
	3	150	2.8×10^{-7}	6.79×10^{-5}	8.33×10^{-5}	2.7×10^{12}
	4	240	9.3×10^{-10}	1.01×10^{-4}	1.06×10^{-4}	4.4×10^{14}
	5	350	2.3×10^{-10}	4.70×10^{-5}	6.61×10^{-5}	5.2×10^{16}
	10	1250	2.3×10^{-12}	4.70×10^{-5}	2.90×10^{-5}	3.9×10^{20}
5	1	36	8.7×10^{-7}	2.33×10^{-2}	2.33×10^{-2}	1.4×10^{5}
	2	96	7.1×10^{-8}	1.69×10^{-3}	1.69×10^{-3}	3.8×10^{9}
	3	180	9.3×10^{-8}	8.12×10^{-5}	7.69×10^{-5}	1.5×10^{13}
	4	288	1.2×10^{-9}	1.01×10^{-4}	1.07×10^{-4}	1.5×10^{15}
	5	420	4.3×10^{-11}	6.03×10^{-5}	5.63×10^{-5}	3.1×10^{18}
	10	1500	5.3×10^{-12}	5.22×10^{-5}	2.92×10^{-5}	1.2×10^{20}
10	1	66	8.7×10^{-7}	2.37×10^{-2}	2.37×10^{-2}	1.4×10^{7}
	2	176	5.7×10^{-7}	7.92×10^{-4}	7.46×10^{-4}	3.8×10^{12}
	3	330	3.8×10^{-9}	1.33×10^{-4}	1.22×10^{-4}	1.5×10^{15}
	4	528	1.3×10^{-12}	9.88×10^{-5}	1.01×10^{-4}	1.5×10^{17}
	5	770	1.0×10^{-12}	5.47×10^{-5}	5.71×10^{-5}	3.1×10^{18}
	10	2750	1.1×10^{-12}	4.94×10^{-5}	2.53×10^{-5}	1.2×10^{19}

Table 3.8: Comparison of the relative error ϵ_h computed from solving the discrete linear system using a LU-based direct solver and the relative error e_{d} obtained from the approximated solution using a naive damped method. The relative errors and the condition number κ are reported for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation.

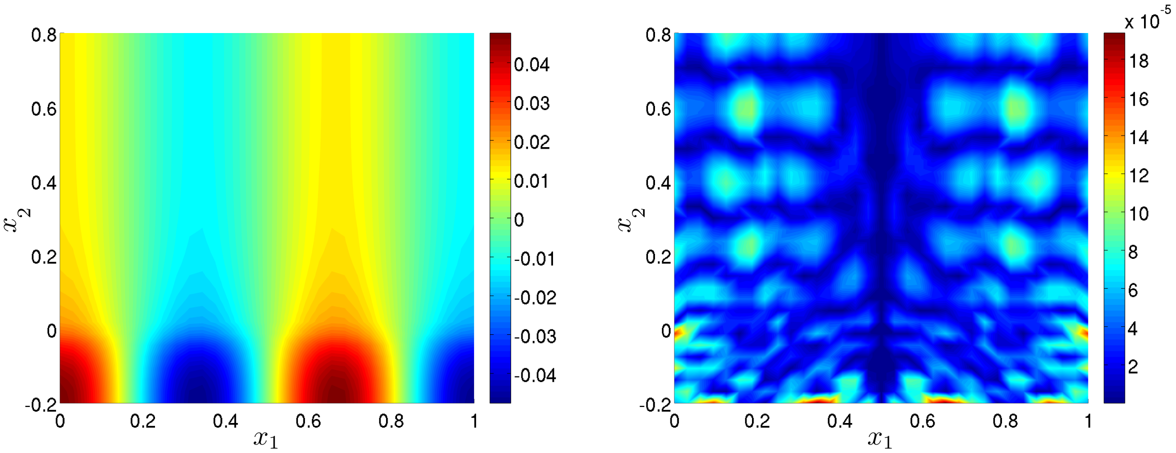

Figure 3.8: Real part of the approximate solution (using a LU-based direct solver) (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is given by equation (3.128).

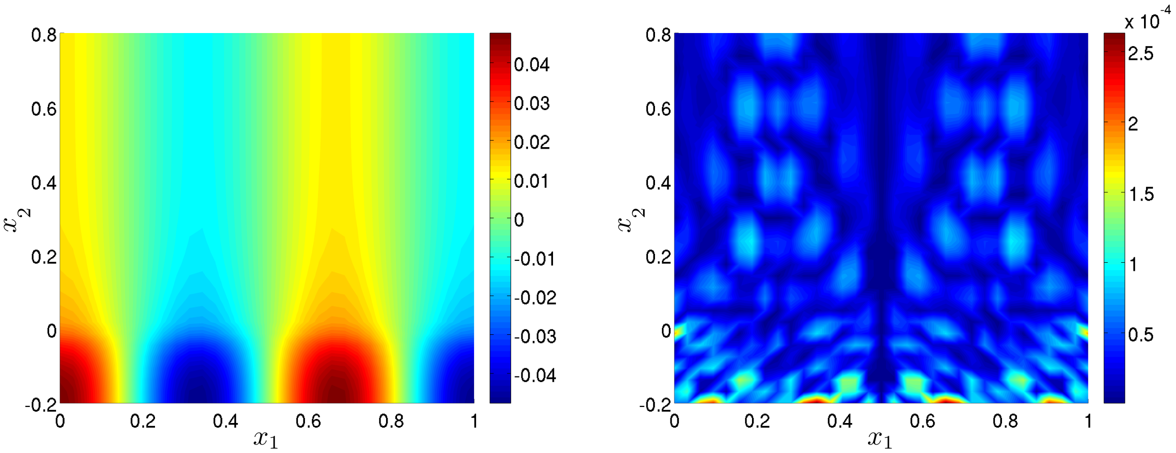

Figure 3.9: Real part of the approximate solution (using a naive damped method) (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is given by equation (3.128).

M	N	dof	λ_{t} (L-curve)	e_r	e_{t} (L-curve)	κ
1	1	12	6.1×10^{0}	2.98×10^{-1}	1.18×10^{0}	1.4×10^{2}
	2	32	8.2×10^{-1}	2.09×10^{-2}	1.31×10^{-1}	3.8×10^{6}
	3	60	6.3×10^{-5}	1.13×10^{-3}	1.31×10^{-3}	1.5×10^{8}
	4	96	7.2×10^{-8}	7.76×10^{-5}	9.40×10^{-5}	1.5×10^{11}
	5	140	6.2×10^{-9}	1.24×10^{-4}	1.22×10^{-4}	3.1×10^{13}
	10	500	4.8×10^{-8}	1.05×10^{-4}	6.79×10^{-5}	1.2×10^{19}
2	1	18	2.7×10^{-1}	9.88×10^{-2}	1.03×10^{-1}	2.2×10^{3}
	2	48	1.9×10^{-1}	3.44×10^{-3}	4.26×10^{-2}	6.7×10^{7}
	3	90	2.9×10^{-5}	2.27×10^{-4}	1.30×10^{-3}	2.5×10^{10}
	4	144	6.1×10^{-9}	1.06×10^{-4}	1.08×10^{-4}	4.9×10^{12}
	5	210	3.5×10^{-9}	8.56×10^{-5}	9.84×10^{-5}	2.5×10^{15}
	10	750	1.0×10^{-9}	3.94×10^{-5}	4.38×10^{-5}	1.3×10^{20}
3	1	24	5.5×10^{-2}	7.89×10^{-2}	1.33×10^{-1}	2.2×10^{4}
	2	64	1.2×10^{-4}	5.76×10^{-3}	5.89×10^{-3}	8.8×10^{7}
	3	120	7.0×10^{-5}	2.03×10^{-4}	1.10×10^{-3}	1.3×10^{11}
	4	192	1.1×10^{-8}	1.03×10^{-4}	1.15×10^{-4}	3.3×10^{13}
	5	280	7.1×10^{-9}	5.57×10^{-5}	8.09×10^{-5}	6.7×10^{15}
	10	1000	$4.7 \times 10^{+2}$	4.63×10^{-5}	9.28×10^{-1}	4.4×10^{19}
4	1	30	1.2×10^{-2}	2.27×10^{-2}	2.27×10^{-2}	1.2×10^{5}
	2	80	1.6×10^{-5}	3.48×10^{-3}	3.47×10^{-3}	5.4×10^{8}
	3	150	3.6×10^{-6}	6.79×10^{-5}	6.79×10^{-5}	2.7×10^{12}
	4	240	$2.4 \times 10^{+2}$	1.01×10^{-4}	1.01×10^{-4}	4.4×10^{14}
	5	350	2.0×10^{-9}	4.70×10^{-5}	4.70×10^{-5}	5.2×10^{16}
	10	1250	$4.0 \times 10^{+2}$	4.70×10^{-5}	4.70×10^{-5}	3.9×10^{20}
5	1	36	4.3×10^{-3}	2.33×10^{-2}	2.33×10^{-2}	1.4×10^{5}
	2	96	5.3×10^{-3}	1.69×10^{-3}	1.69×10^{-3}	3.8×10^{9}
	3	180	9.8×10^{-7}	8.12×10^{-5}	8.12×10^{-5}	1.5×10^{13}
	4	288	$2.2 \times 10^{+2}$	1.01×10^{-4}	1.01×10^{-4}	1.5×10^{15}
	5	420	$3.0 \times 10^{+2}$	6.03×10^{-5}	6.03×10^{-5}	3.1×10^{18}
	10	1500	$4.0 \times 10^{+2}$	5.22×10^{-5}	5.22×10^{-5}	1.2×10^{20}
10	1	66	$4.4 \times 10^{+1}$	2.37×10^{-2}	2.37×10^{-2}	1.4×10^{7}
	2	176	2.4×10^{-2}	7.92×10^{-4}	7.92×10^{-4}	3.8×10^{12}
	3	330	7.9×10^{-5}	1.33×10^{-4}	1.33×10^{-4}	1.5×10^{15}
	4	528	5.3×10^{-8}	9.88×10^{-5}	9.88×10^{-5}	1.5×10^{17}
	5	770	2.6×10^{-11}	5.47×10^{-5}	5.47×10^{-5}	3.1×10^{18}
	10	2750	6.8×10^{-9}	4.94×10^{-5}	4.94×10^{-5}	1.2×10^{19}

Table 3.9: Comparison of the relative error ϵ_h computed from solving the discrete linear system using a LU-based direct solver and the relative error e_{t} obtained by using the Tikhonov filtering technique (whose regularization parameter has been chosen by the L-curve). The relative errors and the condition number κ are reported for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation.

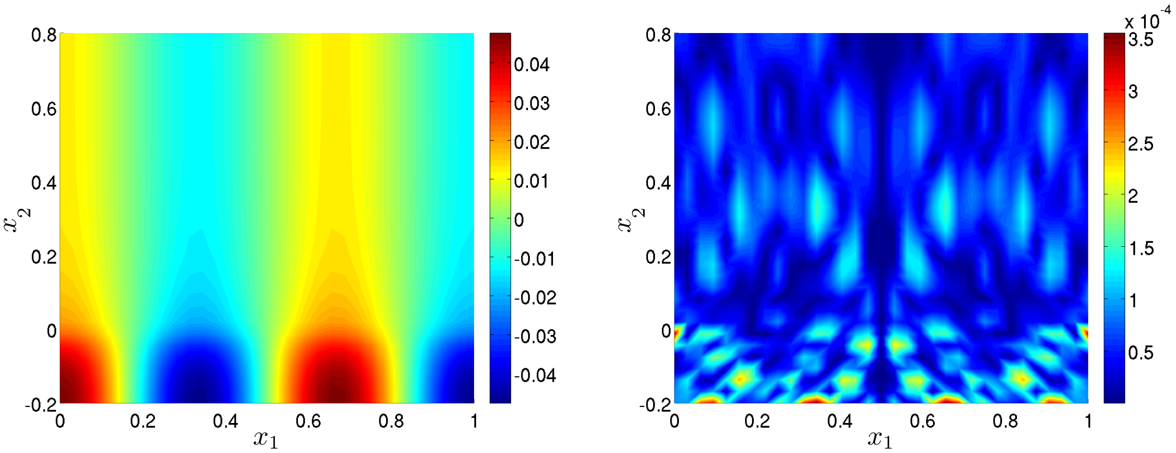

Figure 3.10: Real part of the approximate solution (using the Tikhonov filtering technique whose regularization parameter has been chosen by the L-curve) (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is given by equation (3.128).

M	N	dof	λ_{t} (GCV)	e_r	e_{t} (GCV)	κ
1	1	12	$1.1 \times 10^{+1}$	2.98×10^{-1}	1.11×10^{0}	1.4×10^{2}
	2	32	2.5×10^{-2}	2.09×10^{-2}	2.25×10^{-2}	3.8×10^{6}
	3	60	9.2×10^{-5}	1.13×10^{-3}	1.24×10^{-3}	1.5×10^{8}
	4	96	1.4×10^{-7}	7.76×10^{-5}	9.25×10^{-5}	1.5×10^{11}
	5	140	2.2×10^{-10}	1.24×10^{-4}	1.25×10^{-4}	3.1×10^{13}
	10	500	1.8×10^{-10}	1.05×10^{-4}	4.54×10^{-5}	1.2×10^{19}
2	1	18	5.5×10^{-1}	9.88×10^{-2}	1.10×10^{-1}	2.2×10^{3}
	2	48	1.3×10^{-5}	3.44×10^{-3}	3.52×10^{-3}	6.7×10^{7}
	3	90	8.5×10^{-8}	2.27×10^{-4}	2.33×10^{-4}	2.5×10^{10}
	4	144	1.3×10^{-8}	1.06×10^{-4}	1.11×10^{-4}	4.9×10^{12}
	5	210	2.1×10^{-11}	8.56×10^{-5}	8.24×10^{-5}	2.5×10^{15}
	10	750	1.8×10^{-10}	3.94×10^{-5}	3.94×10^{-5}	1.3×10^{20}
3	1	24	1.3×10^{-1}	7.89×10^{-2}	1.25×10^{-1}	2.2×10^{4}
	2	64	2.4×10^{-5}	5.76×10^{-3}	5.88×10^{-3}	8.8×10^{7}
	3	120	5.2×10^{-8}	2.03×10^{-4}	2.04×10^{-4}	1.3×10^{11}
	4	192	9.2×10^{-11}	1.03×10^{-4}	1.04×10^{-4}	3.3×10^{13}
	5	280	1.9×10^{-11}	5.57×10^{-5}	5.93×10^{-5}	6.7×10^{15}
	10	1000	1.6×10^{-10}	4.63×10^{-5}	3.45×10^{-5}	4.4×10^{19}
4	1	30	2.9×10^{-2}	2.27×10^{-2}	2.52×10^{-2}	1.2×10^{5}
	2	80	4.0×10^{-5}	3.48×10^{-3}	3.42×10^{-3}	5.4×10^{8}
	3	150	6.0×10^{-10}	6.79×10^{-5}	6.78×10^{-5}	2.7×10^{12}
	4	240	9.7×10^{-12}	1.01×10^{-4}	1.01×10^{-4}	4.4×10^{14}
	5	350	1.7×10^{-11}	4.70×10^{-5}	5.52×10^{-5}	5.2×10^{16}
	10	1250	1.3×10^{-10}	4.70×10^{-5}	3.49×10^{-5}	3.9×10^{20}
5	1	36	9.8×10^{-3}	2.33×10^{-2}	2.89×10^{-2}	1.4×10^{5}
	2	96	2.1×10^{-7}	1.69×10^{-3}	1.69×10^{-3}	3.8×10^{9}
	3	180	1.3×10^{-10}	8.12×10^{-5}	8.13×10^{-5}	1.5×10^{13}
	4	288	9.5×10^{-12}	1.01×10^{-4}	1.01×10^{-4}	1.5×10^{15}
	5	420	1.6×10^{-11}	6.03×10^{-5}	5.70×10^{-5}	3.1×10^{18}
	10	1500	1.2×10^{-10}	5.22×10^{-5}	3.43×10^{-5}	1.2×10^{20}
10	1	66	3.3×10^{-4}	2.37×10^{-2}	2.37×10^{-2}	1.4×10^{7}
	2	176	1.4×10^{-9}	7.92×10^{-4}	7.92×10^{-4}	3.8×10^{12}
	3	330	7.9×10^{-12}	1.33×10^{-4}	1.33×10^{-4}	1.5×10^{15}
	4	528	2.9×10^{-11}	9.88×10^{-5}	1.01×10^{-4}	1.5×10^{17}
	5	770	2.4×10^{-11}	5.47×10^{-5}	5.88×10^{-5}	3.1×10^{18}
	10	2750	8.1×10^{-11}	4.94×10^{-5}	2.74×10^{-5}	1.2×10^{19}

Table 3.10: Comparison of the relative error ϵ_h computed from solving the discrete linear system using a LU-based direct solver and the relative error e_{t} obtained by using the Tikhonov filtering technique (whose regularization parameter has been chosen by the generalized cross validation technique). The relative errors and the condition number κ are reported for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation.

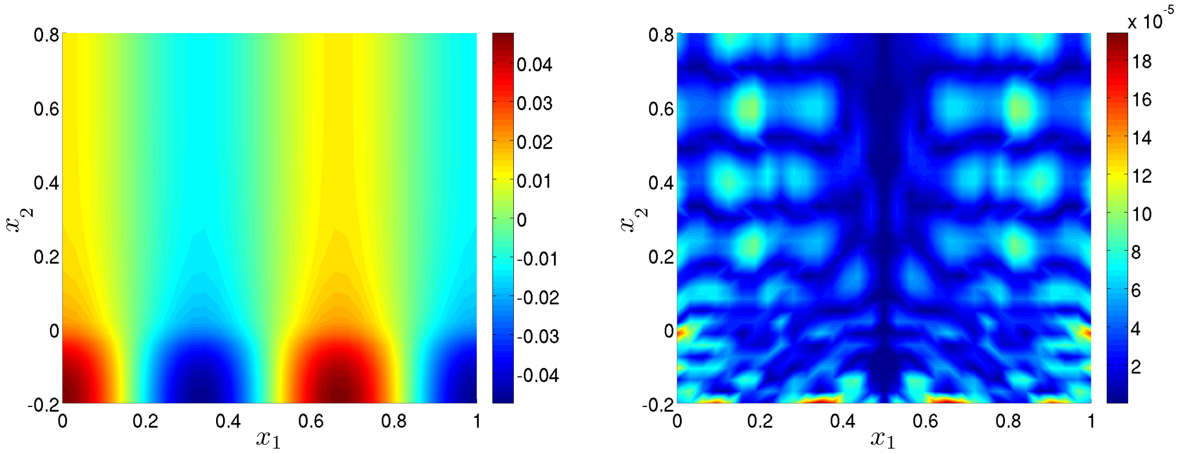

Figure 3.11: Real part of the approximate solution (using the Tikhonov filtering technique whose regularization parameter has been chosen by the generalized cross validation technique) (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is given by equation (3.128).

M	N	dof	λ_{svd} (L-curve)	e_r	e_{svd} (L-curve)	κ
1	1	12	4	2.98×10^{-1}	1.25×10^{0}	1.4×10^{2}
	2	32	21	2.09×10^{-2}	2.75×10^{-1}	3.8×10^{6}
	3	60	23	1.13×10^{-3}	1.53×10^{-1}	1.5×10^{8}
	4	96	17	7.76×10^{-5}	9.50×10^{-1}	1.5×10^{11}
	5	140	119	1.24×10^{-4}	1.39×10^{-4}	3.1×10^{13}
	10	500	94	1.05×10^{-4}	8.94×10^{-1}	1.2×10^{19}
2	1	18	8	9.88×10^{-2}	1.47×10^{-0}	2.2×10^{3}
	2	48	14	3.44×10^{-3}	1.07×10^{-0}	6.7×10^{7}
	3	90	10	2.27×10^{-4}	1.18×10^{-0}	2.5×10^{10}
	4	144	58	1.06×10^{-4}	1.09×10^{-2}	4.9×10^{12}
	5	210	24	8.56×10^{-5}	9.42×10^{-1}	2.5×10^{15}
	10	750	410	3.94×10^{-5}	4.67×10^{-5}	1.3×10^{20}
3	1	24	1	7.89×10^{-2}	1.12×10^{-0}	2.2×10^{4}
	2	64	7	5.76×10^{-3}	1.11×10^{-0}	8.8×10^{7}
	3	120	20	2.03×10^{-4}	1.23×10^{-0}	1.3×10^{11}
	4	192	61	1.03×10^{-4}	6.35×10^{-2}	3.3×10^{13}
	5	280	238	5.57×10^{-5}	8.06×10^{-5}	6.7×10^{15}
	10	1000	77	4.63×10^{-5}	9.25×10^{-1}	4.4×10^{19}
4	1	30	17	2.27×10^{-2}	2.89×10^{-1}	1.2×10^{5}
	2	80	60	3.48×10^{-3}	9.50×10^{-3}	5.4×10^{8}
	3	150	119	6.79×10^{-5}	8.94×10^{-4}	2.7×10^{12}
	4	240	36	1.01×10^{-4}	1.09×10^{0}	4.4×10^{14}
	5	350	94	4.70×10^{-5}	1.35×10^{-2}	5.2×10^{16}
	10	1250	538	4.70×10^{-5}	4.08×10^{-5}	3.9×10^{20}
5	1	36	13	2.33×10^{-2}	1.34×10^{0}	1.4×10^{5}
	2	96	30	1.69×10^{-3}	1.22×10^{0}	3.8×10^{9}
	3	180	150	8.12×10^{-5}	8.39×10^{-4}	1.5×10^{13}
	4	288	130	1.01×10^{-4}	9.74×10^{-4}	1.5×10^{15}
	5	420	305	6.03×10^{-5}	1.10×10^{-4}	3.1×10^{18}
	10	1500	255	5.22×10^{-5}	1.02×10^{-2}	1.2×10^{20}
10	1	66	29	2.37×10^{-2}	1.26×10^{0}	1.4×10^{7}
	2	176	34	7.92×10^{-4}	1.17×10^{0}	3.8×10^{12}
	3	330	39	1.33×10^{-4}	1.19×10^{0}	1.5×10^{15}
	4	528	168	9.88×10^{-5}	1.40×10^{-2}	1.5×10^{17}
	5	770	678	5.47×10^{-5}	5.89×10^{-5}	3.1×10^{18}
	10	2750	568	4.94×10^{-5}	2.50×10^{-4}	1.2×10^{19}

Table 3.11: Comparison of the relative error ϵ_h computed from solving the discrete linear system using a LU-based direct solver and the relative error e_{svd} obtained by using the truncated singular value decomposition method (whose regularization parameter has been chosen by the L-curve technique). The relative errors and the condition number κ are reported for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation.

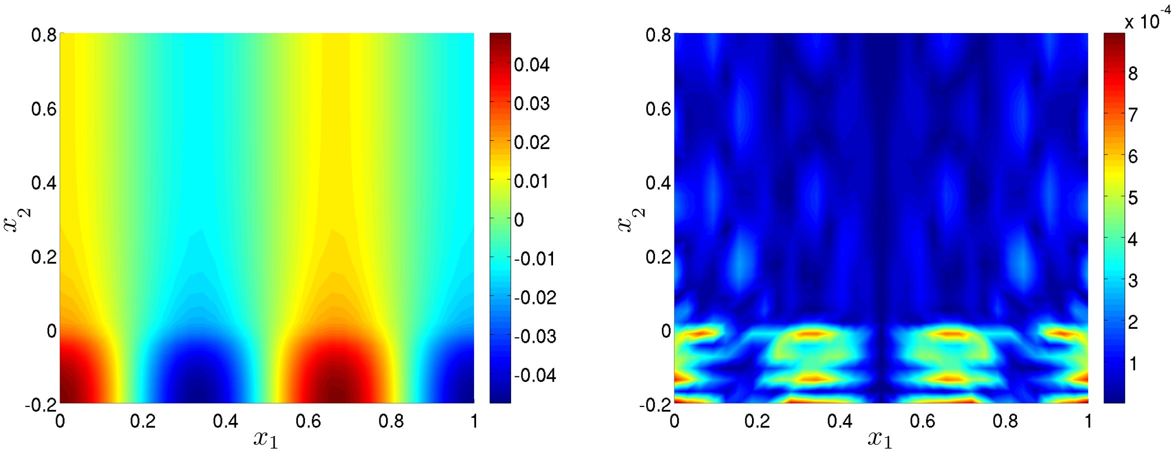

Figure 3.12: Real part of the approximate solution (using the truncated singular value decomposition method whose regularization parameter has been chosen by the L-curve technique) (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \ldots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is given by equation (3.128).

M	N	dof	λ_{svd} (GCV)	e_r	e_{svd} (GCV)	κ
1	1	12	4	2.98×10^{-1}	1.25×10^{0}	1.4×10^{2}
	2	32	29	2.09×10^{-2}	2.15×10^{-2}	3.8×10^{6}
	3	60	59	1.13×10^{-3}	1.25×10^{-3}	1.5×10^{8}
	4	96	94	7.76×10^{-5}	9.72×10^{-5}	1.5×10^{11}
	5	140	139	1.24×10^{-4}	1.24×10^{-4}	3.1×10^{13}
	10	500	468	1.05×10^{-4}	3.53×10^{-5}	1.2×10^{19}
2	1	18	14	9.88×10^{-2}	1.26×10^{-1}	2.2×10^{3}
	2	48	46	3.44×10^{-3}	3.60×10^{-3}	6.7×10^{7}
	3	90	89	2.27×10^{-4}	2.38×10^{-4}	2.5×10^{10}
	4	144	139	1.06×10^{-4}	1.08×10^{-4}	4.9×10^{12}
	5	210	209	8.56×10^{-5}	8.50×10^{-5}	2.5×10^{15}
	10	750	641	3.94×10^{-5}	3.46×10^{-5}	1.3×10^{20}
3	1	24	22	7.89×10^{-2}	1.35×10^{-1}	2.2×10^{4}
	2	64	58	5.76×10^{-3}	5.89×10^{-3}	8.8×10^{7}
	3	120	117	2.03×10^{-4}	2.04×10^{-4}	1.3×10^{11}
	4	192	190	1.03×10^{-4}	1.04×10^{-4}	3.3×10^{13}
	5	280	278	5.57×10^{-5}	5.61×10^{-5}	6.7×10^{15}
	10	1000	856	4.63×10^{-5}	2.51×10^{-5}	4.4×10^{19}
4	1	30	28	2.27×10^{-2}	2.44×10^{-2}	1.2×10^{5}
	2	80	74	3.48×10^{-3}	3.42×10^{-3}	5.4×10^{8}
	3	150	149	6.79×10^{-5}	6.78×10^{-5}	2.7×10^{12}
	4	240	238	1.01×10^{-4}	1.00×10^{-4}	4.4×10^{14}
	5	350	345	4.70×10^{-5}	5.06×10^{-5}	5.2×10^{16}
	10	1250	1017	4.70×10^{-5}	2.79×10^{-5}	3.9×10^{20}
5	1	36	31	2.33×10^{-2}	2.80×10^{-2}	1.4×10^{5}
	2	96	93	1.69×10^{-3}	1.68×10^{-3}	3.8×10^{9}
	3	180	179	8.12×10^{-5}	8.13×10^{-5}	1.5×10^{13}
	4	288	285	1.01×10^{-4}	1.01×10^{-4}	1.5×10^{15}
	5	420	406	6.03×10^{-5}	5.76×10^{-5}	3.1×10^{18}
	10	1500	1262	5.22×10^{-5}	2.72×10^{-5}	1.2×10^{20}
10	1	66	61	2.37×10^{-2}	2.36×10^{-2}	1.4×10^{7}
	2	176	175	7.92×10^{-4}	7.92×10^{-4}	3.8×10^{12}
	3	330	327	1.33×10^{-4}	1.33×10^{-4}	1.5×10^{15}
	4	528	526	9.88×10^{-5}	9.84×10^{-5}	1.5×10^{17}
	5	770	735	5.47×10^{-5}	5.73×10^{-5}	3.1×10^{18}
	10	2750	2500	4.94×10^{-5}	2.46×10^{-5}	1.2×10^{19}

Table 3.12: Comparison of the relative error ϵ_h computed from solving the discrete linear system using a LU-based direct solver and the relative error e_{svd} obtained by using the truncated singular value decomposition method (whose regularization parameter has been chosen by the generalized cross validation technique). The relative errors and the condition number κ are reported for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation.

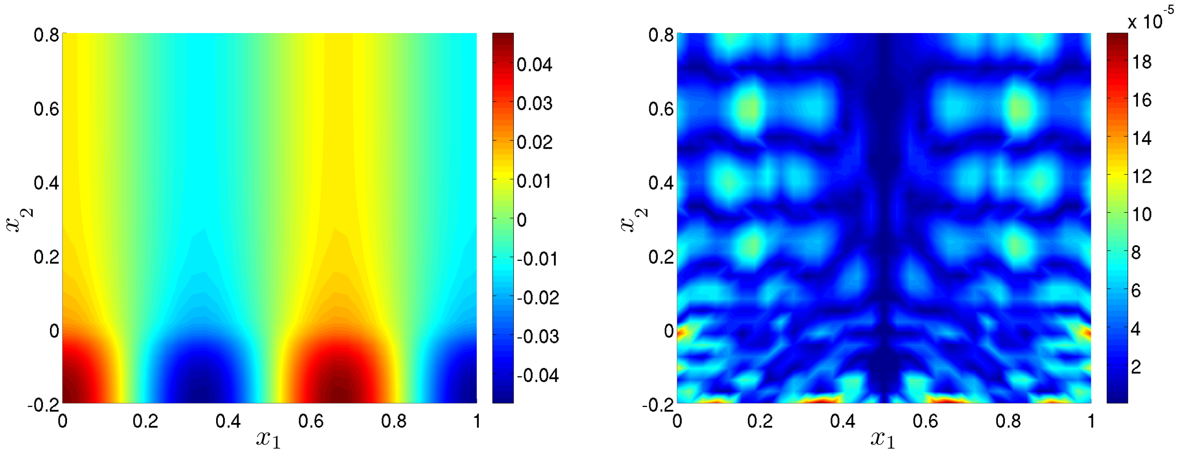

Figure 3.13: Real part of the approximate solution (using the truncated singular value decomposition method whose regularization parameter has been chosen by the generalized cross validation technique) (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of four elements (i.e. $M = 4$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \dots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is given by equation (3.128).

3.5.6 Accurate approximation of eigenmodes not included in the discrete space

It has been discussed in previous sections that those eigenpairs corresponding to constant factors $q_0(x_1)$ are not considered to define the enrichment of the discrete space X_h. So, it could be natural to conclude that the exact solutions of the problem (3.10)-(3.14) could be inaccurately approximated by the modal-based PUFEM method. On the contrary, the numerical test described in this section illustrates the high accuracy of the method even for the constant case.

With this aim, as it has been considered in the previous sections, problem (3.10)-(3.14) is settled with angular frequency $\omega = \pi$, and pure Neumann boundary conditions have been considered in the whole boundary $\partial\Omega$ (taking $\beta = 0$). The square domain $\Omega = (0, L) \times (-a, H)$ with $a = 0.2$, $H = 0.8$, $L = 1$ is split in two subdomains where the speed of sound is given by $c_- = 1/2$ in $\Omega_- = (0, L) \times (-a, 0)$ and $c_+ = 1$ in $\Omega_+ = (0, L) \times (0, H)$.

The source term has been fixed to $f = 1$ to obtain as exact solution $u = -1/\omega^2$. Table 3.13 shows that, even if the eigenmodes which are independent of the x_1 spatial coordinate are not included in the modal-based discretization, the constant exact solution can be approximated accurately with similar relative errors to those other solutions which does not belongs to X_h (see for instance, the similar relative errors reported in Table 3.5).

M	N	dof	ϵ_h	κ
1	1	12	1.21×10^{-1}	1.5×10^{2}
	2	32	1.99×10^{-2}	3.8×10^{6}
	3	60	2.31×10^{-3}	1.5×10^{8}
	4	96	2.13×10^{-4}	1.5×10^{11}
	5	140	1.61×10^{-5}	3.1×10^{13}
10	1	66	9.42×10^{-3}	2.7×10^{7}
	2	176	1.56×10^{-4}	1.0×10^{12}
	3	330	1.04×10^{-5}	3.6×10^{15}
	4	528	5.02×10^{-7}	1.2×10^{17}
	5	770	1.31×10^{-7}	2.0×10^{18}
100	1	606	9.38×10^{-3}	2.5×10^{13}
	2	1616	4.72×10^{-5}	2.0×10^{16}
	3	3030	1.66×10^{-5}	1.4×10^{18}
	4	4848	1.54×10^{-6}	5.1×10^{18}
	5	7070	1.26×10^{-5}	8.2×10^{18}

Table 3.13: Relative error ϵ_h and the condition number κ for different values of the mesh size M, the number of eigenpair families considered in the discretization N, and the degrees of freedom (dof) of the discrete approximation. The exact solution is constant.

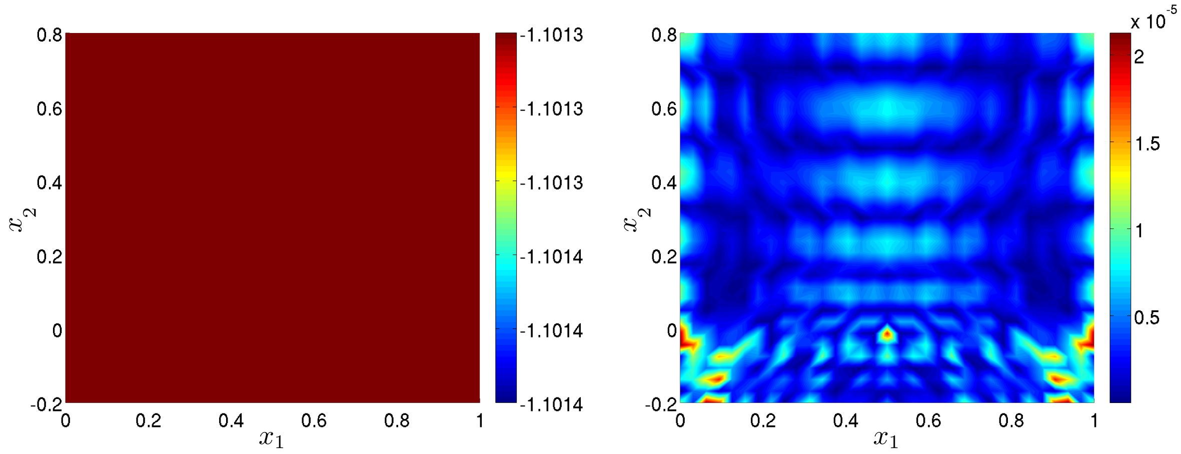

Figure 3.14: Real part of the approximate solution (left) and relative error (right), obtained from the modal-based PUFEM method with a one-dimensional mesh of ten elements (i.e. $M = 10$) and considering the family of Love and interior waves $w_{n,j}$ with $(n, j) \in \{1, \dots, 3\} \times \{\mathcal{L}_n \cup \mathcal{I}_n^{J_n}\}$ (i.e. $N = 3$). The exact solution is $u = -1/\omega^2$.

3.6 Conclusions

In this chapter, a non destructive testing problem in a bi-layered domain without a crack has been studied. The existence and uniqueness of the weak problem, together with its spectral analysis have been deduced. A modal-based partition of unity finite element method, using Love and internal waves to approximate the solution of the problem, has been proposed and described in detail. The bad conditioning of the matrix have been studied. Finally, some numerical results have been presented, in order to illustrate the accuracy of the method, the deterioration of the results due to the high condition number and some regularization techniques.

Further research

To finish this PhD dissertation, some of the research lines that could be explored are briefly described. They can be divided in three large blocks, attending to the aims of each one: the numerical analysis of partition of unity finite element methods in one, two and three dimensions, the numerical enhancement of the partition of unity finite element methods applied to heterogeneous media and the challenges of the application of PUFEM methods in non destructive testing.

- The first future research line is first devoted to obtain a more accurate error estimate for the approximation computed with a PUFEM discretization of a one-dimensional Helmholtz problem. This estimate should confirm the independence on the wave number observed in the numerical results. The path to follow in order to achieve this accuracy on the error estimate should be to increase the smoothness of the source term f. The oscillatory solutions of the variational problem (1.2) considered in the first chapter of this thesis, are solutions too of the Helmholtz equation with smooth right-hand side $f \in \mathrm{H}^l(0,1)$ with $l \geq 1$, and duality stability estimates (analogous to that one described in [28, Theorem 3.2]) should be used to obtain a more accurate estimate (possibly independent of k). Then, the possibility of the extension of the results to two and three-dimensional Helmholtz problems should be studied.

- The second block of open problems arises around the application of PUFEM methods to acoustic problems in heterogeneous media. A first approach could be to dive into regularization techniques, in order to find a way to deal with the high condition number that deteriorates the numerical results (as it happened in chapters 2 and 3). Then, the extension of the transmission-reflection PUFEM methods to problems where the wave number is not piecewise constant but a variable function, defined differently in each media, should be considered.

- The last block of future work is the deepening on the non destructive testing problems. After having applied PUFEM discretizations (by means of involving Love waves in the enrichment of the FEM functions) to problems in bi-layered media without a crack on the interface, the natural step would be to continue the study applying partition of unity finite element methods to approximate the solution of problems with a defect on the interface between media. And, eventually, to solve an inverse problem in order to know if a certain object, a pipe with a coating for example, has a defect in the interface between its layers, after having sent Love waves.

Resumen en castellano

Son numerosos y variados los problemas de la física y de la ingeniería cuyos modelos matemáticos involucran la propagación de ondas acústicas. Por citar algunos, se podrían mencionar problemas como la reducción de ruido (las propiedades acústicas son ya un criterio emergente de calidad tanto en instalaciones como en productos industriales), la exploración seísmica, la acústica submarina, los ultrasonidos en medicina o los ensayos no destructivos. En este contexto, surge la necesidad de resolver problemas de propagación acústica cada vez más complejos, que no pueden ser resueltos por técnicas basadas en métodos matemáticos clásicos. Es habitual la construcción de prototipos para asegurar que las tecnologías utilizadas sean precisas, pero su alto coste de fabricación hace necesario que los test se lleven a cabo en fases avanzadas del diseño y con propuestas muy próximas a la solución final. La simulación numérica es una técnica determinante para analizar y diseñar sistemas acústicos en poco tiempo y con costes competitivos.

La diversidad matemática de los problemas de propagación acústica hace necesario el empleo de una amplia variedad de modelos numéricos, y la aplicación de técnicas de computación numéricas avanzadas. De entre todos esos modelos, la ecuación de Helmholtz es ampliamente utilizada como modelo de referencia en problemas de propagación acústica armónicos en tiempo. En regímenes de altas y medias frecuencias, su aproximación numérica calculada con un método de elementos finitos (FEM) nodal difiere significativamente de la solución exacta, debido al llamado efecto de la polución (véase [14]). Por lo tanto, la precisión de las aproximaciones numéricas de los problemas de Helmholtz se basa en métodos discretos libres de polución, que deberían tener un comportamiento robusto respecto al número de onda.

El método de partición de la unidad basado en elementos finitos (PUFEM), introducido por Babŭska e Ihlenburg en 1996 (véase [36]) será el método libre de polución, escogido de entre todos ellos, utilizado a lo largo de esta tesis. Las ventajas computacionales y los inconvenientes en la implementación de las discretizaciones de tipo PUFEM han sido demostrados de forma numérica en varios trabajos (por ejemplo en [35]), pero no se ha encontrado en la literatura ninguna estimación de error para PUFEM en términos del número de onda.

El objetivo del primer capítulo de esta tesis será deducir una estimación de error, en términos del número de onda, para una discretización de tipo PUFEM, basada en un enriquecimiento de las funciones base de elementos finitos con ondas planas, aplicado a un problema de Helmholtz unidimensional. El segundo capítulo se dedicará aproximar de forma numérica problemas acústicos armónicos en tiempo uni y bidimensionales, en un dominio dividido en dos capas. El método PUFEM desarrollado en este capítulo tendrá en cuenta la transmisión y reflexión que tiene lugar en el interfaz entre subdominios. Finalmente, el último capítulo de esta tesis propone una novedosa discretización PUFEM que involucra ondas de Love, como herramienta en ensayos no destructivos.

A continuación, se describe con detalle cada capítulo:

Capítulo 1. Estimaciones de error para soluciones aproximadas de la ecuación de Helmholtz utilizando un método de partición de la unidad basado en elementos finitos

Los problemas de Helmholtz con diversas condiciones de contorno surgen de varias aplicaciones físicas. Para obtener resultados precisos en la aproximación numérica de éstos problemas (véase [24]), el tamaño de la malla escogida h utilizando métodos de elementos finitos o diferencias finitas, debe depender del número de onda k, habitualmente según una regla "rule of the thumb", que asegura un mínimo número de nodos por longitud de onda. En problemas donde el dominio computacional tiene el mismo orden de magnitud que la longitud de onda del movimiento armónico, este criterio lleva a precisión en los resultados. Sin embargo, la calidad de las aproximaciones numéricas con dichos métodos se deteriora si el dominio computacional o el número de onda son suficientemente grandes. Nuestra atención se centrará en problemas de propagación acústica en el régimen de medias a altas frecuencias, en el que que la discretización mediante un método de partición de la unidad basado en elementos finitos es una de las pocas posibilidades de resolver este tipo de problemas de forma poco costosa desde un punto de vista computacional.

En este primer capítulo, se deducirán estimaciones de error *a priori* para una discretización PUFEM sobre un problema de Helmholtz unidimensional. En primer lugar, se planteará el problema modelo en el intervalo $(0, 1)$, que consta de la ecuación de Helmholtz unidimensional con un segundo miembro f en $\mathrm{L}^2(0, 1)$, y condiciones de contorno de tipo Dirichlet en el extremo izquierdo y Robin en el derecho. Se deduce su formulación variacional y se demuestra la condición *inf-sup* continua y un resultado de estabilidad de la solución del problema débil con respecto de los datos. La discretización PUFEM para dicho problema de Helmholtz unidimensional se basa en un enriquecimiento con ondas planas de las funciones base de elementos finitos. Dicha discretización se describe en términos de funciones exponenciales y de funciones trigonométricas, siendo esta última descripción más adecuada para el análisis del error. El número de onda de las funciones base del espacio discreto, se modifica añadiéndole un parámetro de perturbación adicional δ, para reproducir situaciones en las que la solución exacta no se conoce o tratar de reflejar los problemas para aproximar la solución exacta que aparecen en problemas de Helmholtz bidimensionales o en problemas con número de onda variable. Notar que, si este parámetro δ no se introduce, la solución exacta cae dentro del espacio de discretización. Después de esto, se deducen varias estimaciones de interpolación y se demuestran la condición LBB discreta y un resultado de estabilidad de la solución aproximada con respecto a los datos. En estas condiciones, se puede demostrar una estimación de error en términos del número de onda k, el tamaño de malla h y el parámetro de perturbación adicional δ. Los resultados numéricos ilustran el orden de convergencia para el tamaño de malla h y el parámtro de perturbación δ, que coinciden con la estimación obtenida. La independencia del error relativo con respecto al número de onda que puede observarse en los resultados numéricos no aparece sin embargo en la estimación, por lo que ésta podría ser mejorada si se añade regularidad a la función segundo miembro f.

Capítulo 2. Métodos de partición de la unidad basados en elementos finitos en medios multicapa

Muchos de los problemas acústicos de interés tienen lugar en medios heterogéneos. Es el caso, por ejemplo, de la acústica submarina, donde las diferentes capas de agua bajo la superficie de la mar tienen distintos grados de salinidad, de temperatura, diferente profundidad y mayor o menor cantidad de componentes biológicos. Esto hace necesario plantear problemas aproximados que resuelvan problemas acústicos en medios con varias capas.

En este segundo capítulo, se trabaja con varios problemas de Helmholtz. En primer lugar, se considera un problema de Helmholtz unidimensional en un dominio multicapa. El problema modelo se plantea en el dominio $(0, 1)$, con condiciones de contorno de tipo Dirichlet en el extremo izquierdo, Robin en el derecho, y un número de onda k constante a trozos y estrictamente positivo. Tras deducir el problema continuo, se plantean cuatro posibles discretizaciones, en términos de la elección del número de onda utilizado en las funciones base PUFEM. El primero de ellos consiste en calcular la media global del número de onda variable, y utilizar esa media como número de onda en cada función base. A continuación, se describe un método local, que considera el número de onda en cada elemento de la malla a la hora de definir el número de onda de las funciones base. El tercer método está basado en las aproximaciones introducidas por Pablo Ortiz [41], y consiste en calcular una media local del número de onda variable en cada elemento. Y finalmente el método propuesto de transmisión-reflexión, que tiene en cuenta cada transmisión y reflexión ocurridas en cada elemento de la malla. Notar que, en caso de que el número de onda fuera constante, las cuatro discretizaciones darían lugar al método PUFEM propuesto en el capítulo 1. Tras plantear el problema discreto en notación matricial, se muestran varios resultados numéricos, que confirman que el método de transmisión-reflexión recupera totalmente la solución exacta.

El segundo problema considerado en este capítulo es un problema de Helmholtz bidimensional con número de onda constante. Tras introducir el problema modelo, formado por una ecuación de Helmholtz homogénea bidimensional y condiciones de contorno de tipo Neumann, y deducir su formulación variacional, se describe en detalle la discretización considerada en este caso, consistente en escoger como espacio discreto el subespacio generado por las funciones base Lagrange $\mathbb{P}_1$ de elementos finitos en dimensión dos, multiplicadas por ondas planas. Tras plantear el problema discreto y su notación matricial, se especifican algunas de las técnicas de integración utilizadas para calcular la solución aproximada. En las matrices tanto de masa como de rigidez, aparecen integrandos que están formados por polinomios multiplicados por exponenciales, que oscilan con respecto a x_1 y x_2. Para tratar estas integrales, se seguirá la técnica de rotación introducida por Pablo Ortiz, que permitirá reescribir el integrando de tal forma que ahora sólo oscile respecto de una sola variable. Además, la integración sobre cada tríangulo tras el cambio de variable es detallada de forma exhaustiva en seis casos. Los resultados numéricos de esta parte del segundo capítulo ilustran la precisión del método PUFEM descrito, el decrecimiento exponencial del error relativo en norma L^∞ cuando en número de ondas planas utilizadas en la discretización se incrementa, y el comportamiento del error relativo con respecto al número de onda.

Finalmente, el último problema considerado en este capítulo se ocupa de problemas de Helmholtz en medios bicapa. El problema modelo se plantea considerando la ecuación de Helmholtz homogénea, condiciones de contorno de tipo Neumann, condiciones de acople en el interfaz (continuidad de la solución y de su derivada normal) y un número de onda constante a trozos. Tras plantear la formulación variacional del problema, una novedosa discretización PUFEM para este tipo de problemas es descrita con detalle. Dicha discretización consiste en, de forma similar al método PUFEM utilizado en el problema de Helmholtz bidimensional con número de onda constante, definir el espacio discreto como el subespacio generado por funciones Lagrange $\mathbb{P}_1$ FEM standard, multiplicadas por funciones tipo onda plana, y de forma similar al método PUFEM de transmisión y reflexión utilizado en el problema unidimensional con número de onda variable, tener en cuenta a la hora de definir esas funciones tipo onda plana la transmisión y reflexión ocurrida en cada trángulo de la malla. Notar que, dependiendo del ángulo de incidencia en esas funciones de tipo onda plana, puede aparecer evanescencia en alguno de los medios. Tras plantear el problema discreto y su notación matricial, se explican las técnicas de integración utilizadas en este caso. Debido al fenómeno de evanescencia que aparece en algunas ondas planas de la discretización, las técnicas utilizadas en el problema bidimensional con número de onda constante no son aplicables en este caso, por lo que se aplicará un cambio afín al triángulo de referencia. Los resultados numéricos muestran la precisión y eficiencia del método PUFEM propuesto para aproximar problemas de Helmholtz bidimensionales con condiciones de contorno Neumann y número de onda constante a trozos.

Capítulo 3. Un método de partición de la unidad modal basado en elementos finitos para problemas de propagación de ondas en dominios bicapa

El desarrollo de técnicas para encontrar cracks en el interfaz entre dos materiales es de vital importancia en la detección temprana de posibles defectos en estructuras como tuberías con revestimientos. Las técnicas más utilizadas en ensayos no destructivos son los ultrasonidos y las corrientes de Foucault que se propagan de forma transversal al interfaz. Pero ambas están limitadas a problemas donde la fuente desde la que se envía la onda está cerca del crack. La posibilidad de utilizar ondas de Love para encontrar un defecto que esté lejos de la fuente ha sido sugerida recientemente (véase [17]). Es básico en este tipo de detecciones conocer a priori la solución del problema sin crack.

El objetivo de este capítulo es ofrecer una herramienta para aproximar la solución de estos problemas sin crack en medios bicapa. Para ello, se propone un método PUFEM que involucra ondas de Love. Tras plantear el problema modelo y su formulación débil, se lleva a cabo un exhaustivo análisis espectral. En el caso de tener un dominio cuadrangular y condiciones de contorno de tipo Neumann, se han descrito las ecuaciones de dispersión tanto para ondas de Love como para ondas internas. En las figuras que las describen, el decaimiento exponencial de las ondas de Love y el comportamiento oscilatorio de las ondas internas pueden apreciarse. La discretización propuesta, junto con el problema discreto y su notación matricial se describen en detalle, junto con el análisis del número de condicionamiento de la matriz del sistema. Una amplia batería de resultados numéricos ilustran la precisión del método PUFEM modal propuesto, tanto para el caso de incluir sólamente ondas de Love como para el caso de considerar tanto ondas de Love como ondas internas. Describen además el deterioro de los resultados numéricos debido al alto número de condicionamiento de la matriz discreta y su potencial desaparición aplicando técnicas de regularización (como estrategias de amortigüación, filtrado clásico de Tikhonov o métodos de descomposición en valores singulares truncados).

Como posibles líneas futuras, se describen aquí tres bloques, cada uno con un objetivo diferente: el análisis numérico de métodos de partición de la unidad basados en elementos finitos en una, dos y tres dimensiones, mejoras de tipo numérico en los métodos PUFEM sobre medios heterogéneos y la continuación de la aplicación de métodos PUFEM en ensayos no destructivos.

- La primera línea de trabajo futuro se dedicaría a obtener una estimación de error más precisa para aproximaciones calculadas con una discretización de un problema de Helmholtz unidimensional. Esta estimación debería confirmar la independencia respecto al número de onda que se observa en los resultados numéricos. Para conseguir esta precisión, deberían considerarse términos fuente con mayor regularidad. Una vez hecho esto, parece natural trabajar en la extensión de estos resultados a problemas bi y tridimensionales.

- El segundo bloque de problemas abiertos se ocuparía de la aplicación de métodos de partición de la unidad basados en elementos finitos a problemas acústicos en medios heterogéneos. Una primera aproximación podría ser estudiar la aplicación de nuevos métodos de regularización, para conseguir manejar el alto condicionamiento que estropea los resultados numéricos. Además de eso, la extensión de los métodos de transmisión y reflexión PUFEM a problemas donde el número de onda no es constante a trozos sino una función definida de forma distinta en cada medio, debería ser considerada.

- El último bloque de trabajo futuro sería la profundización en problemas de ensayos no destructivos. Tras haber aplicado discretizaciones de tipo PUFEM (incluyendo ondas de Love en el enriquecimiento de las funciones base de elementos finitos) a problemas en medios bicapa sin crack en su interfaz, el siguiente paso consistiría en continuar el estudio aplicando métodos de partición de la unidad modales basados en elementos finitos, a problemas con un defecto entre los dos medios. Y finalmente, resolver problemas inversos aplicados a problemas de la ingeniería, para saber por ejemplo si una tubería con revestimiento interno presenta un defecto en el interfaz entre sus dos capas, analizando la onda reflejada tras enviar una onda de Love.

Agradecimientos

Son muchas las personas a las que me gustaría dar las gracias por su apoyo en estos cinco años.

A mi director, Andrés Prieto. Nuestra relación director-doctoranda ha pasado, como todas las relaciones, por altos y bajos. Pero haciendo balance de estos años, puedo decir que no me equivoqué al proponerle que fuera co-director de mi tesis. Es una persona con una capacidad de trabajo inmensa, que ha dedicado a esta tesis muchas horas de su tiempo y que nunca me ha dejado en la estacada. No se ha arrugado ante ninguna de las tareas que conlleva dirigir una tesis. Lo mismo borda una demostración de pizarra, que corrige al detalle un documento, que se remanga sin miramientos para buscar contigo el gazapo en un código. Y en este último tramo, cuando más falta hacía, cada frase de ánimo ayudó a llevar a buen puerto esta tesis.

A mi director, Luis Hervella. Al que tengo que agradecerle que me propusiera para el que fue mi primer trabajo, como investigadora en el proyecto SLA. Aprendí mucho trabajando a su lado en esa etapa previa al doctorado. Estuvo ahí desde el primer minuto en que se barajó la idea de realizar esta tesis. Y a pesar de que es decano de la facultad de informática desde 2013, con todo el trabajo que eso conlleva, no ha descuidado sus tareas como director de tesis en ningún momento. Además, su sentido del humor ha ayudado a aligerar algunos momentos en los que el ánimo decaía.

Al profesor Philippe Destuynder, que dirigió mi estancia en París, me gustaría agradecerle su cercanía y su buena acogida en el CNAM. Su puerta siempre estuvo abierta para cualquier duda que tuviera, y junto con Jose, compartimos inspiradoras reuniones llenas de demostraciones en pizarra con música clásica de fondo. Aprendí mucho de él en esos tres meses y medio.

Al profesor Jose Orellana, con el que pude colaborar en mi estancia en París. Por haber estado ahí desde el mismo momento en el que llegué al aeropuerto hasta el día en que me fui. Tanto para discutir alguna demostración en la pizarra del despacho, como para enseñarme lo básico para moverme por una cuidad tan grande como es París.

A Francoise Santi, Olivier Wilk, Bertrand Mercier, Alan Sabathe, Iraj Mortazavi ,Thierry Horsin, Marco Caponigro, Giorgio Russolillo y Juanjo López, del CNAM, por su amabilidad y hospitalidad conmigo y por hacer del CNAM un estupendo lugar de trabajo. Y por supuesto a Claire, que fue compañera en el CNAM.

A Iñigo Arregui, por compartir su experiencia en el CNAM antes de que me fuera de estancia, por todos sus consejos y por su amabilidad.

A María González Taboada, porque además de ser mi profesora en el máster y estar más que dispuesta a resolver cuaquier duda que pudiera tener, desde que llegué a la UDC me acogió de asturiana a asturiana.

A María José Souto, por haber compartido sus experiencias doctorales y postdoctorales conmigo, y por ser una maravillosa persona que nunca me trató como una alumna de doctorado sino como a una más.

A mis compañeras y compañeros de laboratorio, primero en la FIC y después en el CITIC. A María, Carmiña, Jose, Alejandro, Marta, Paula, Aldana, Álvaro, Manu, Dani, Miguel y Javier, y a mis otros compañeros de la FIC, Jezú, Brais, Miguelón, Porta, Sasi, Dani, por todos esos cafés, cañas y comidas a la canasta compartidas.

A cada persona con la que me he cruzado en la FIC o en el CITIC que me ha sacado una sonrisa o me ha dirigido palabras de ánimo.

A las chicas de PlatoyZapato. Por esas excursiones que, además de mostrarme esta bonita tierra en la que llevo ya unos años, ayudaron a relajarme en momentos de mucho estrés.

A Merce, Cris, Espe, Pau, Alma, Nor, Haris, Alex, Miguel, Fran y a mis compañeros y compañeras de la resi. Mis amigas y amigos. Por sacarme una sonrisa en los malos momentos, por apoyarme y sobre todo por ser personas con las que siempre merece la pena hablar. A Iván, que me ha apoyado sin condiciones y ha tenido paciencia infinita conmigo.

Y por supuesto a mi familia. A mi madre, un modelo de mujer fuerte, independiente y maravillosa, y a mi padre, que siempre me ha apoyado en todo y ha estado orgulloso de mi, a ellos le debo todo lo que soy. A mis abuelas y abuelos, a toda mi familia.

Se cierra una etapa, y comienza una nueva. A todos y todas, GRACIAS.

Fundings

- This work has been partially supported by the Xunta de Galicia project EM2013/052, cofunded with European Regional Development Funds (ERDF).

- In addition, the PhD candidate has been supported by "Ayuda CN 2011/004", cofunded with European Regional Development Funds (ERDF).

- Furthermore, the last part of the thesis was made in an stage in France founded by the grant INDITEX-UDC 2016.

Bibliography

[1] J.D. Achenbach. *Wave Propagation in Elastic Solids.* North Holland Publishing Company, 1984.

[2] M. Ainsworth. Discrete dispersion relation for hp-version finite element approximation at high wave number. *SIAM Journal on Numerical Analysis*, 42(2):553–575, 2004.

[3] K. Atkinson and W. Han. *Theoretical Numerical Analysis: A Functional Analysis Framework*, volume 39. Springer, 2005.

[4] A. Bermúdez, R. G. Durán, R. Rodríguez, and J. Solomin. Finite element analysis of a quadratic eigenvalue problem arising in dissipative acoustics. *SIAM Journal on Numerical Analysis*, 38(1):267–291, 2000.

[5] P. Bettess, J. Shirron, O. Laghrouche, B. Peseux, R. Sugimoto, and J. Trevelyan. A numerical integration scheme for special finite elements for the helmholtz equation. *International journal for numerical methods in engineering*, 56:531–552, 2003.

[6] L. Brekhovskikh. *Waves in layered media*, volume 16. Elsevier, 2012.

[7] H. Brezis. *Analyse Fonctionnelle. Théorie et Applications.* Collection Mathématiques Appliquées pour la Maıtrise. Masson, Paris, 1983.

[8] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag New York, 1991.

[9] A. Cabada. *Green's Functions in the Theory of Ordinary Differential Equations.* Springer, New York, 2013.

[10] G. Capuano, M. Ruzzene, and J. J. Rimoli. Modal-based finite elements for efficient wave propagation analysis. In *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* American Society of Mechanical Engineers, 2013.

[11] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*, volume 40. SIAM, 2002.

[12] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*, volume 2. Springer-Verlag Berlin, 1988-93.

[13] E. B. Davies. *Linear Operators and Their Spectra*, volume 106. Cambridge University Press, 2007.

[14] A. Deraemaeker, I. Babuška, and P. Bouillard. Dispersion and pollution of the FEM solution for the Helmholtz equation in one, two and three dimensions. *International Journal for Numerical Methods in Engineering*, 46(4):471–499, 1999.

[15] P. Destuynder and C. Fabre. Can we hear the echos of cracks? *Journal of Elasticity*, 2017. doi:10.1007/s10659-017-9632-7.

[16] G. Diwan. *Partition of unity boundary element and finite element method: overcoming nonuniqueness and coupling for acoustic scattering in heterogeneous media.* PhD thesis, Durham University, 2014.

[17] J. C. Dumont-Fillon. Contrŏle non destructif par les ondes de love et lamb. *Editions techniques de l'ingénieur*, 2012.

[18] A. Ern and J. L. Guermond. *Theory and Practice of Finite Elements.* Applied Mathematical Sciences 159. Springer-Verlag New York, 1 edition, 2004.

[19] G. Gabard. Exact integration of polynomial-exponential products with application to wave-based numerical methods. *Comunications in numerical methods in engineering*, 25:237–246, 2009.

[20] P. Grisvard. *Elliptic Problems in Nonsmooth Domains.* SIAM, 1985.

[21] J. L. Guermond and A. Ern. *Theory and Practice of Finite Elements.* Springer, New York, 2004.

[22] P. C. Hansen. The truncated svd as a method for regularization. (27):543–553, 1987.

[23] P. C. Hansen. Truncated svd solutions to discrete ill-posed problems with ill-determined numerical rank. *Journal on Scientific and Statistical Computing*, 11:503–518, 1990.

[24] I. Harari and T. Hughes. Finite element methods for the helmholtz equation in an exterior domain: model problems. *Computer Methods in Applied Mechanics and Engineering*, 87(1):59–96, 1991.

[25] L. Hervella-Nieto, P. M. López-Pérez, and A. Prieto. Error estimates for partition of unity finite element solutions of the helmholtz equation. In preparation.

[26] F. Ihlenburg. *Finite Element Analysis of Acoustic Scattering.* Springer-Verlag New York, 1998.

[27] 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 & Mathematics with Applications*, 30(9):9–37, 1995.

[28] F. Ihlenburg and I. Babuška. Finite element solution of the helmholtz equation with high wave number. part ii: the hp version of the fem. *SIAM Journal on Numerical Analysis*, 34(1):315–358, 1997.

[29] T. Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, 2013.

[30] R. B. Kellogg. On the poisson equation with intersecting interfaces. *Applicable Analysis*, 4(2):101–129, 1974.

[31] M. G. Krein, H. Langer, and R. Troelstra. On some mathematical principles in the linear theory of damped oscillations of continua i. *Integral Equations and Operator Theory*, 1(3):364–399, 1978.

[32] A. J. Laub. *Matrix Analysis for Scientists and Engineers.* Siam, 2005.

[33] C. L. Lawson and R. J. Hanson. *Solving Least Square Problems.* Prentice-Hall, Englewood Cliffs, 1974.

[34] Advanpix LLC. Multiprecision computing toolbox for matlab.

[35] J. M. Melenk. *On Generalized Finite Element Methods.* PhD thesis, University of Maryland, 1995.

[36] J. M. Melenk and I. Babuška. The partition of unity finite element method: basic theory and applications. *Computer Methods in Applied Mechanics and Engineering*, 139(1):289–314, 1996.

[37] K. Miller. Least squares methods for ill-posed problems with a prescribed bound. *Journal on Mathematical Analysis*, 1:52–74, 1970.

[38] M. S. Mohamed, O. Laghrouche, and A. El-Kacimi. Some numerical aspects of the pufem for efficient solution of 2d helmholtz problems. *Computers and Structures*, 88:1484–1491, 2010.

[39] J. Nečas. *Les méthodes directes en théorie des équations elliptiques.* Masson, Paris, 1967.

[40] P. Ortiz. Finite elements using a plane-wave basis for scattering of surface water waves. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 362(1816):525–540, 2004.

[41] P. Ortiz and E. Sánchez. An improved partition of unity finite element model for diffraction problems. *International Journal for Numerical Methods in Engineering*, 50(12):2727–2740, 2001.

[42] E. Perrey-Debain, O. Laghrouche, P. Bettess, and J. Trevelyan. Plane-wave basis finite elements and boundary elements for three-dimensional wave scattering. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 362(1816):561–577, 2004.

[43] D. L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the ACM*, 9:84–97, 1962.

[44] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: Analysis of Operators, vol. IV.* New York, Academic Press, 1978.

[45] M. Salo. Unique continuation for elliptic equations. Technical report, University of Jyväskylä, Department of Mathematics and Statistics, 2014.

[46] A. A. Samarskii. *The Theory of Difference Schemes*, volume 240. CRC Press, 2001.

[47] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady*, 4:1035–1038, 1963.

[48] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* Winston and Sons, Washington D.C., 1977.

[49] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra: the Behavior of Nonnormal Matrices and Operators.* Princeton University Press, 2005.

[50] J. M. Varah. On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems. *Journal on Numerical Analysis*, 10:257–267, 1973.

[51] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* SIAM, Philadelphia, 1990.