POLITECNICO DI MILANO

Facoltà di Ingegneria Industriale e dell'Informazione

Dipartimento di Scienze e Tecnologie Aerospaziali

Corso di Laurea Magistrale in Ingegneria Aeronautica

A POD reduced order method for parameterized Maxwell's equations and applications

Relatore: Prof. Ing. Stefano MICHELETTI
Correlatore: Prof. Ing. Gianluigi ROZZA (SISSA)

Tesi di laurea di:
Emiliano CANGEMI
Matr. 787329

Anno Accademico 2014 - 2015

a Maria, Stefano e Milo,
sempre al mio fianco.

Vorrei ringraziare prima di tutto il Prof. Gianluigi Rozza, il quale, oltre ad avermi concesso l'opportunità di affacciarmi al mondo della riduzione di basi, è sempre stato gentile e pronto ad ogni mia richiesta di consulto. La sua disponibilità e puntualità nelle risposte, nonostante i mille impegni, sono un esempio formidabile di serietà e rispetto. Inoltre, la fiducia e il costante ottimismo mostrati nei momenti più difficili hanno svolto un ruolo fondamentale nel portare a termine questa Tesi.
Ringrazio il Prof. Stefano Micheletti per le sue critiche e la sua competenza. Il continuo scambio di pareri e punti di vista hanno migliorato di volta in volta questo lavoro, permettendo di raggiungere un livello più che degno per un produzione scientifica.
Ringrazio mia cugina Diana Agostini e il mio caro amico Andrea Corbetta che generosamente hanno donato parte del loro tempo per aiutarmi nella stesura e redazione dello scritto in lingua inglese.
Un doveroso e sentito grazie va ai Dr. Stefano Lorenzi, Dr. Alessandro Lucantonio e Dr. Andrea Mola per il supporto tecnico e i consigli pratici nel utilizzo dei software. Inoltre, ringrazio il Prof. Vincente Torre e la Dr. Monica Mazzolini della SISSA, Scuola Internazionale Superiore di Studi Avanzati, per la documentazione introduttiva sul loro lavoro con le fibre ottiche e, il Dr. Remo Proietti Zaccaria e il Dr. Salvatore Tuccio dello IIT, Istituto Italiano di Tecnologia, per avermi aiutato a capire i fenomeni fisici presenti in questa applicazione. Vorrei ringraziare anche il Prof. Olivier Pironneau per aver reso disponibile il materiale sul profilo alare stealth.
Un ringraziamento va anche a tutti i compagni incontrati durante il percorso di studi. La sincera amicizia che ci lega, non solo ha reso le fatiche universitarie più facili da affrontare, ma è qualcosa che porterò per sempre con me.
Infine vorrei dire "Grazie" alle tre persone che mi sono state più vicine in questi duri mesi di lavoro. Come me, hanno sofferto per ogni simulazione andata male, e con me hanno sperato che i risultati fossero quelli sperati. Le immancabili parole di sostegno, l'incrollabile pazienza, i sacrifici e gli sforzi nell'essere sempre amorevoli e comprensivi nei miei confronti sono stati la forza senza la quale nulla sarebbe stato possibile.

Grazie Sara,
Grazie Maria,
Grazie Stefano.

Abstract

The reduced order method (ROM) based on the proper orthogonal decomposition (POD) will be described in this thesis. It will be applied to Maxwell's equations for two typical electromagnetic problems: a scattering problem for an airfoil and a far field problem of an optical fiber wave-guide.
Numerical simulations were obtained using the COMSOL [7] software, which is based on the finite element method (FEM). The post processing (application of POD method) has been instead carried out using Matlab [21], with the creation *ad hoc* of scripts for both problems.
For the test case of the scattering problem, the results obtained are in good agreement with those available in the literature [23]. Even for the problem of the optical fiber, the results are to be considered physically possible and in good accordance with what is shown in the work [22].
As regards the innovative part of the thesis, that is the verification that the POD method is applicable to both the problems mentioned above with a relative computational advantage over the FEM methodology, the results confirm its excellent universality and applicability. These conclusions should be considered as a starting point for the development of a series of analysis for the planning and optimization of real applications related to both problems.

Keywords: scattering problem, airfoil, far field problem, optical fiber, reduced order method.

Sommario

In questa tesi viene descritto il metodo di riduzione di modello (ROM) basato sulla Proper Orthogonal Decomposition (POD). Esso verrà applicato alle equazioni di Maxwell per due problemi tipici elettromagnetici: un problema di riflessione-dispersione per un profilo alare e un problema di campo lontano per una fibra ottica.
Le simulazioni numeriche sono state svolte mediante il software COMSOL [7], il quale si basa sul metodo degli elementi finiti (FEM). Il post-processing (applicazione del metodo POD) è stato invece effettuato in ambiente Matlab [21], con la creazione *ad hoc* di script per i due diversi problemi.
Per il caso test del problema di scattering sono stati ottenuti risultati in buon accordo con quanto già presente in letteratura [23]. Anche per il problema della fibra guidante i risultati sono da considerarsi fisicamente plausibili e in buon accordo con quanto esposto nel lavoro [22].
Per quanto concerne la parte innovativa della tesi, ossia la verifica che il metodo POD sia applicabile ad entrambi i problemi sopra citati e il relativo vantaggio computazionale rispetto alla metodologia FEM, i risultati hanno confermano l'ottima generalità e applicabilità. Queste conclusioni vanno considerate come punto di partenza per lo sviluppo di analisi per le fasi di progettazione e ottimizzazione per applicazioni reali legate ad entrambi i problemi.

Parole chiave: problema di riflessione-dispersione, profilo alare, problema di campo lontano, fibra ottica, metodo di riduzione di modello.

Contents

Introduction

In every field of daily life, the physical laws rule every action of our lives. Each object, be it manufactured or not, is subject to the rules dictated by physics. Thanks to mathematics, these are translated into formulas and made available to physicists, chemists and engineers.
Generally, every physical process is described by partial differential equations that are able to predict the temporal and spatial evolution of any variable that they describe. The two macro-trends that nowadays are solved by this extraordinary resource can be divided into:

- *design problems*: once both the governing equations that model the process with unknown variables (*state*) and the physical and geometrical values/parameters (*inputs*) are given, we need a solution that allows us to describe the behavior of the state in order to obtain the quantities (*outputs*) useful to the designer to determine whether the sizing of the system conforms to the requirements. We can immediately understand that, before finding a definite configuration of inputs, the attempts are numerous (and often unsuccessful) and that only a gradual path is the solution to get a good result;

- *control problems*: to impose a certain behavior to specific quantities *outputs* through operations on the dynamic system. Now this objective is achieved thanks to the *optimal control* theory, whose basic idea is to define and minimize the appropriate *cost functions* formed by a combination of inputs and outputs. Since there is no *magic formula* that can explicit this relation so as to obtain the predetermined behavior, even in this context the only possible way remains to proceed by trial and error.

In both cases, if we want to ensure a certain quality in the results, we would have to make countless attempts, so we will need to reduce as much as possible the computational cost for each single test. It is just to satisfy this need that the *reduced order modeling* is able to provide a valid answer. The basic idea

is to project the solution into a space with contained size, built according to the nature of the problem, thus being able to reduce the computational costs. So, instead of having a solution with a very large space ($\mathbb{V}_h$ on the order N_h), like the one obtained using the *Finite Element Method* (FEM), we will obtain a solution with a smaller space ($\mathbb{V}_{rb}$ on the order N_{rb}) while preserving the same precision, accuracy and efficiency.
The idea of representing a physical model that maintains the reliability and accuracy of the results, despite a fewer number of degrees of freedom, is implemented thanks to the subdivision in two stages of the process:

- *offline phase*, where we get as many solutions as parameter configurations to test out. Thanks to these, the bases forming the reduced space will be built. Despite this part being quite heavy on the computational point of view, the advantage lies in having to carry it out only once;

- *online phase*, where a solution is obtained for a specific parameter configuration, without solving the algebraic problem in space $\mathbb{V}_h$, but rather solving it in space $\mathbb{V}_{rb}$. This part is so efficient in terms of computational costs that the resolution can take place in *real time*, thus providing a huge advantage from the analysis point of view compared to the same FEM approach.

This thesis deals with the application of the reduced order modeling method in the field of electromagnetism, in particular performing a parameterization of Maxwell's equations in the frequency domain. In the literature, there are many works that demonstrate how these equations are suitable for the approximation with reduction method ([12], [32], [11]), whereas there is no work focused on the two problems here analyzed (*scattering problem* and *waveguide problem*). In addition to a required theoretical introduction of the physical context, we formalized both the physical and geometrical parametrization of the equations, showing the results obtained in terms of reliability and efficiency.

The following describes in general terms the contents of this thesis work.
After this short introduction, in Chapter 1 we will introduce Maxwell's equations in the time domain and the mathematical model used for the description of the fundamental quantities. We will lay the groundwork for the abstract formalization of the problems that will be analyzed later, in particular the development of both the strong and weak formulation and the suitable functional space in which to search the solution. Finally, we will pass to frequency domain to show which physical quantities appear recursively and which properties they are based upon.

In Chapter 2, we will start with the formalization of a generic numerical approximation of Maxwell's equations in frequency domain before specializing it to the finite element method. A simple one-dimensional example of the Poisson problem will show how to compute the stiffness and mass matrices. At the end, we will describe the main methods of solving algebraic problems.

In Chapter 3, we will illustrate the theoretical and practical aspects of the reduced basis method. In the first part, we will describe the purely formal aspects along the way leading to the definition of the problem in weak form of Maxwell's parameterized equations. Secondly, we will introduce the criterion that defines and chooses the bases that compose the approximating space, according to the POD methodology (Proper Orthogonal Decomposition) and the matrix implementation method.

In Chapter 4, we will study the *scattering problem* for a NACA profile. We will resume the formalizations of Chapter 1 and Chapter 3 and we will apply them to the case at hand. First, we will verify the model by comparing a test case - symmetric profile - with the relative solution shown in the literature [23]. Then we will proceed with a parameterized approximation of a second problem with an asymmetric profile, showing both the difference between the FEM solution and the POD solution and the relative percentage error for different values of the parameter. Finally, we will make some considerations on the computational savings.

In Chapter 5, we will study the *waveguide problem* for a *Tapered Optical Fiber* (TOF). After an initial introduction and the analytical setup of the problem, the considerations on the application of the method and its results will be the same as in Chapter 4.

In Chapter 6, we will summarize the results obtained in this work and we will formulate some considerations about possible developments and future improvements on the two topics treated in Chapters 4 and 5.

At the end, an extract of the work will be presented in Italian.

The two softwares used to derive all the simulations and the results were COMSOL Multi-physics [7] and Matlab [21].

This work has been carried out in the framework of a collaboration between MOX center of Politecnico di Milano and SISSA mathLab, with the opportunity of an internship at SISSA, International School for Advanced Studies, supported by Neurosciences Area of SISSA (Prof. Vincent Torre's group) for the case study of Chapter 5, in collaboration with IIT (Italian Institute of Technology, Genova) and CNR (Italian Research Council).

Milano and Trieste, April 2016.

1 | Maxwell's equations: mathematical model

In this chapter, we deal with Maxwell's equations, a set of four partial differential equations, two vector and two scalar, which allow the complete determination in a spatial domain and in a specific time period of five vector fields - electric intensity, electric flux density, magnetic intensity, magnetic flux intensity, electric current density - and one scalar field - electric charge density. These fields are functions of space and time, and completely describe the electro-magnetic phenomena. More information can be found in [16] and [24].

1.1 Maxwell's equations

The unknowns of the equations are:

- $\mathbf{E}(\mathbf{r},\mathrm{t})$: electric field intensity in vacuum (volt/meter),
- $\mathbf{D}(\mathbf{r},\mathrm{t})$: electric flux density (coulomb/meter^2),
- $\mathbf{B}(\mathbf{r},\mathrm{t})$: magnetic flux density (weber/meter^2),
- $\mathbf{H}(\mathbf{r},\mathrm{t})$: magnetic field intensity in vacuum (ampère/meter),
- $\mathbf{J}(\mathbf{r},\mathrm{t})$: electric current density (ampère/meter^2),
- $\rho(\mathbf{r},t)$: electric charge density (coulomb/meter^3).

The system of Maxwell's equations is:

$$\begin{cases} \nabla \cdot \mathbf{D}(\mathbf{r},t) = \rho(\mathbf{r},t) \quad \text{Gauss' law (electric)}, \\ \dfrac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} = -\nabla \times \mathbf{E}(\mathbf{r},t) \quad \text{Faraday-Neumann-Lenz's law}, \\ \nabla \cdot \mathbf{B}(\mathbf{r},t) = 0 \quad \text{Gauss' law (magnetic)}, \\ \dfrac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} = \nabla \times \mathbf{H}(\mathbf{r},t) - \mathbf{J}(\mathbf{r},t) \quad \text{Ampère-Maxwell's law}. \end{cases} \tag{1.1}$$

The Maxwell equations can be extended with the addition of another equation. This equation represents the charge conservation, and it can be obtained independently of Maxwell's equations.
The formulation is the following:

$$\frac{\partial \rho(\mathbf{r},t)}{\partial t} + \nabla \cdot \mathbf{J}(\mathbf{r},t) = 0 \quad \text{charge conservation.} \tag{1.2}$$

To solve any system of equations it is generally necessary that the numbers of independent equations be equal to the number of unknowns. A quick analysis shows that the unknowns are 16 (the three components of the fields $\mathbf{E}$, $\mathbf{D}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{J}$ and the scalar ρ) against only 7 equations (the three vector ones from Faraday-Neumann-Lenz's law and Ampère-Maxwell's law, and the scalar ones given by the charge conservation).

1.2 Constitutive equations

To overcome the lack of information, and ensure that the balance of equations and unknowns be the same, we have to introduce the "constitutive equations" in order to link some unknowns to other ones to reduce their overall number. As Maxwell's equations must apply in every medium, we will somehow have to enter the information of the medium in which one is solving the system. To do this, we introduce, in a general way, some tensors able to determine the type of medium. Before determining these tensors, we introduce the following constitutive relations:

- $\mathbf{J}(\mathbf{r},\mathrm{t}) = \mathbf{J}(\mathbf{E}(\mathbf{r},\mathrm{t}))$,
- $\mathbf{D}(\mathbf{r},\mathrm{t}) = \mathbf{D}(\mathbf{E}(\mathbf{r},\mathrm{t}))$,
- $\mathbf{B}(\mathbf{r},\mathrm{t}) = \mathbf{B}(\mathbf{H}(\mathbf{r},\mathrm{t}))$.

The first one, under the hypothesis of Ohmic material [16], can be expressed in the form:

$$\mathbf{J}(\mathbf{r},t) = \underline{\underline{\sigma}}(\mathbf{r},t)\mathbf{E}(\mathbf{r},t) + \mathbf{J}_N(\mathbf{r},t), \tag{1.3}$$

which is equivalent to Ohm's law in local form. The term $\underline{\underline{\sigma}}(\mathbf{r},t)$ represents the conductivity tensor of the medium, while the term $\mathbf{J}_N$ corresponds to an imposed flow of current.
A slightly different remark needs to be done for the other two equations. Both can be expressed according to Taylor expansion having as independent variables $\mathbf{E}$ and $\mathbf{H}$, in the neighbourhood of point $\mathbf{E}_0$ and $\mathbf{H}_0$, respectively.

$$\begin{cases} \mathbf{D} & = \mathbf{D}(\mathbf{E}_0) + \dfrac{\partial \mathbf{D}}{\partial \mathbf{E}}(\mathbf{E}-\mathbf{E}_0) + \dfrac{1}{2!}\dfrac{\partial^2 \mathbf{D}}{\partial \mathbf{E}^2}(\mathbf{E}-\mathbf{E}_0)^2 + \dots \\ & \quad + \dfrac{1}{n!}\dfrac{\partial^n \mathbf{D}}{\partial \mathbf{E}^n}(\mathbf{E}-\mathbf{E}_0)^n + \dots, \\ \mathbf{B} & = \mathbf{B}(\mathbf{H}_0) + \dfrac{\partial \mathbf{B}}{\partial \mathbf{H}}(\mathbf{H}-\mathbf{H}_0) + \dfrac{1}{2!}\dfrac{\partial^2 \mathbf{B}}{\partial \mathbf{H}^2}(\mathbf{H}-\mathbf{H}_0)^2 + \dots \\ & \quad + \dfrac{1}{n!}\dfrac{\partial^n \mathbf{B}}{\partial \mathbf{H}^n}(\mathbf{H}-\mathbf{H}_0)^n + \dots . \end{cases} \tag{1.4}$$

Let us define

$$\begin{aligned} &\mathbf{D} - \mathbf{D}(\mathbf{E}_0) = \Delta\mathbf{D}; \\ &\mathbf{E} - \mathbf{E}_0 = \Delta\mathbf{E}; \\ &\mathbf{B} - \mathbf{B}(\mathbf{H}_0) = \Delta\mathbf{B}; \\ &\mathbf{H} - \mathbf{H}_0 = \Delta\mathbf{H}. \end{aligned} \tag{1.5}$$

We can insert (1.5) in (1.4):

$$\begin{cases} \Delta\mathbf{D} = \dfrac{\partial \mathbf{D}}{\partial \mathbf{E}}\Delta\mathbf{E} + \dfrac{1}{2!}\dfrac{\partial^2 \mathbf{D}}{\partial \mathbf{E}^2}\Delta\mathbf{E}^2 + \dots + \dfrac{1}{n!}\dfrac{\partial^n \mathbf{D}}{\partial \mathbf{E}^n}\Delta\mathbf{E}^n + \dots, \\ \Delta\mathbf{B} = \dfrac{\partial \mathbf{B}}{\partial \mathbf{H}}\Delta\mathbf{H} + \dfrac{1}{2!}\dfrac{\partial^2 \mathbf{B}}{\partial \mathbf{H}^2}\Delta\mathbf{H}^2 + \dots + \dfrac{1}{n!}\dfrac{\partial^n \mathbf{B}}{\partial \mathbf{H}^n}\Delta\mathbf{H}^n + \dots . \end{cases} \tag{1.6}$$

Our attention is focused only on the behavior in the neighborhood of the linearization point, so we can consider all the quantities with Δ as the true expressions of $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{H}$. In this way, (1.6) becomes:

$$\begin{cases} \mathbf{D} = \dfrac{\partial \mathbf{D}}{\partial \mathbf{E}}\mathbf{E} + \dfrac{1}{2!}\dfrac{\partial^2 \mathbf{D}}{\partial \mathbf{E}^2}\mathbf{E}^2 + \dots + \dfrac{1}{n!}\dfrac{\partial^n \mathbf{D}}{\partial \mathbf{E}^n}\mathbf{E}^n + \dots, \\ \mathbf{B} = \dfrac{\partial \mathbf{B}}{\partial \mathbf{H}}\mathbf{H} + \dfrac{1}{2!}\dfrac{\partial^2 \mathbf{B}}{\partial \mathbf{H}^2}\mathbf{H}^2 + \dots + \dfrac{1}{n!}\dfrac{\partial^n \mathbf{B}}{\partial \mathbf{H}^n}\mathbf{H}^n + \dots . \end{cases} \tag{1.7}$$

Now some further physical hypothesis is necessary to write the expression of (1.7) that will be used in the thesis:

- if the electric field and the magnetic field are of relatively low intensity, then the higher-order terms can be neglected in a first approximation without significant errors of simplification. This way, the media treated are called linear, and they are subject to the relation:

$$\begin{cases} \mathbf{D} = \dfrac{\partial \mathbf{D}}{\partial \mathbf{E}}\mathbf{E} \\ \mathbf{B} = \dfrac{\partial \mathbf{B}}{\partial \mathbf{H}}\mathbf{H}. \end{cases} \tag{1.8}$$

  It is possible to define:

$$\underline{\underline{\epsilon}}(\mathbf{r},t) = \frac{\partial \mathbf{D}}{\partial \mathbf{E}}(\mathbf{r},t) \quad \text{electric permittivity tensor} \tag{1.9}$$

$$\underline{\underline{\mu}}(\mathbf{r},t) = \frac{\partial \mathbf{B}}{\partial \mathbf{H}}(\mathbf{r},t) \quad \text{magnetic permeability tensor;} \tag{1.10}$$

- if the media are isotropic then tensors become scalars, as the features do not depend on the direction of the medium, while they still depend on place and time. If the media are invariant through time, then the time dependence disappears. If the media are homogeneous, then the dependence on place disappears. If the media are non-dispersive, then the electrical and magnetic characteristics do not depend on the frequency of variation of the fields. So the following constitutive equations will be assumed:

$$\begin{cases} \mathbf{J}(\mathbf{r},t) = \sigma \mathbf{E}(\mathbf{r},t) + \mathbf{J}_N(\mathbf{r},t) \\ \mathbf{D}(\mathbf{r},t) = \epsilon \mathbf{E}(\mathbf{r},t) \\ \mathbf{B}(\mathbf{r},t) = \mu \mathbf{H}(\mathbf{r},t), \end{cases} \tag{1.11}$$

  with σ (siemens/meter), ϵ and μ constant. The permittivity and the permeability of a homogeneous media are usually given relative to that of free space, as a relative permittivity ϵ_r and a relative permeability μ_r defined:

$$\epsilon_r = \frac{\epsilon}{\epsilon_0}, \text{ with } \epsilon_0 = 8.8542 \cdot 10^{-12} \text{ (farads/meter)}, \tag{1.12}$$

$$\mu_r = \frac{\mu}{\mu_0}, \text{ with } \mu_0 = 4\pi \cdot 10^{-7} \text{(henries/meter)}. \tag{1.13}$$

1.3 Mathematical formulation

Now, let us plug the last three constitutive relations in (1.1). In this way, we can obtain the complete Maxwell equations with the same number of equations and unknowns:

$$\begin{cases} \epsilon \nabla \cdot \mathbf{E}(\mathbf{r},t) = \rho(\mathbf{r},t) \quad \text{Gauss' law (electric)}, \\ \mu \dfrac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t} = -\nabla \times \mathbf{E}(\mathbf{r},t) \quad \text{Faraday-Neumann-Lenz's law}, \\ \nabla \cdot \mathbf{H}(\mathbf{r},t) = 0 \quad \text{Gauss' law (magnetic)}, \\ \epsilon \dfrac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} = \nabla \times \mathbf{H}(\mathbf{r},t) - \sigma \mathbf{E}(\mathbf{r},t) - \mathbf{J}_N(\mathbf{r},t) \quad \text{Ampère-Maxwell's law}. \end{cases} \tag{1.14}$$

This is a coherent formulation of Maxwell's equations, but it is not very useful for computation purposes. As a matter of fact, the complete system is very difficult to solve in this form. We have four different equations, with different intrinsic structures: the two divergence equations are PDEs only with spatial derivative and the other two curl equations are space-time PDEs. Thus, we have to manipulate the entire system, in order to have only space-time dependent PDEs equations.

1.3.1 Governing equation for $\mathbf{E}(\mathbf{r},t)$

Let us choose the Faraday-Neumann-Lenz law first. The first step is to apply the curl operator on this one:

$$\nabla \times (\mu \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}) \quad \Rightarrow \quad \mu \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) = -\nabla \times \nabla \times \mathbf{E}. \tag{1.15}$$

The Ampère-Maxwell law is used for substituting the magnetic field curl and yields:

$$\mu \frac{\partial}{\partial t}(\epsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} + \mathbf{J}_N) = -\nabla \times \nabla \times \mathbf{E}. \tag{1.16}$$

All the terms with the electric field dependence are left on the left-hand side:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) + \mu\epsilon \frac{\partial^2 \mathbf{E}(\mathbf{r},t)}{\partial^2 t} + \mu\sigma \frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} = -\mu \frac{\partial \mathbf{J}_N(\mathbf{r},t)}{\partial t}. \tag{1.17}$$

1.3.2 Governing equation for $\mathbf{H}(\mathbf{r}, t)$

Now, we can do the same operation with the Ampére-Maxwell law, thus we apply first the curl operator and then substitute the Faraday-Neumann-Lenz law into the resulting expression:

$$\begin{aligned} &\nabla \times (\epsilon \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \sigma \mathbf{E} - \mathbf{J}_N) \\ \Rightarrow \quad &\epsilon \frac{\partial}{\partial t}(\nabla \times \mathbf{E}) = \nabla \times \nabla \times \mathbf{H} - \sigma \nabla \times \mathbf{E} - \nabla \times \mathbf{J}_N. \end{aligned} \tag{1.18}$$

The Faraday-Neumann-Lenz law is used for substituting the electric field curl and gives:

$$-\mu\epsilon \frac{\partial^2 \mathbf{H}}{\partial^2 t} = \nabla \times \nabla \times \mathbf{H} + \mu\sigma \frac{\partial \mathbf{H}}{\partial t} - \nabla \times \mathbf{J}_N. \tag{1.19}$$

All the terms with the magnetic field dependence are left on the left-hand side:

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r}, t) + \mu\epsilon \frac{\partial^2 \mathbf{H}(\mathbf{r}, t)}{\partial^2 t} + \mu\sigma \frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} = \nabla \times \mathbf{J}_N(\mathbf{r}, t). \tag{1.20}$$

We can notice that the two equations are de-coupled from each other. This is very useful because we can solve independently the one for $\mathbf{E}(\mathbf{r}, t)$ and the other one for $\mathbf{H}(\mathbf{r}, t)$. There is a only way to have a useful formulation to do that, which is the writing of the weak formulation for both unknown fields.

1.3.3 From governing equations to the weak formulation of $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$

One of the most commonly used methods for solving differential problems is the Galerkin method, which belongs to the family of weighted residual methods [16]. Fundamentally, this method seeks an approximate solution of the problem, by searching in a subspace of a linear space.
The method consists in multiplying a vector or a scalar test function, with compact support, with the differential equation of the problem. Later it will be integrated across the spatial domain, thus obtaining the remaining boundary conditions. This yields the so-called weak formulation of the problem. To

complete this formulation, we have to choose the right function space where the weak formulation solution belongs to. This operation will be done after completing the introduction of the boundary conditions, thus, for the moment, the next operations will be carried out without any further mathematical rigour. The importance of this formulation is found in the class of equations that they solve: in fact, it should be noted that the solutions belong to a function space that requires a lower degree in the derivatives, compared to the solutions of the strong formulation of the same problem.

Now, we proceed with the formal steps to find the weak formulation. Equation (1.17) is multiplied by a vector test function $\mathbf{v}$ and then integrated (by parts) in the spatial domain V. In a similar way we operate on (1.20), multiplying it by the vector test function $\mathbf{z}$, then integrating (by parts) in the spatial domain V. It is to be observed that the integration over time - in the real field $\mathbb{R}^+$ - will be performed after the discretization of the spatial domain. Multiplying by $\mathbf{v}$ and $\mathbf{z}$ yields the following equations for the electric and magnetic fields, respectively:

$$\int_V \mathbf{v} \cdot (\nabla \times \nabla \times \mathbf{E}) \; + \int_V \mathbf{v} \cdot (\epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{E}}{\partial t}) \; = - \int_V \mathbf{v} \cdot \mu \frac{\partial \mathbf{J}_N}{\partial t} \quad (1.21)$$

$$\int_V \mathbf{z} \cdot (\nabla \times \nabla \times \mathbf{H}) \; + \int_V \mathbf{z} \cdot (\epsilon\mu \frac{\partial^2 \mathbf{H}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{H}}{\partial t}) \; = \int_V \mathbf{z} \cdot \nabla \times \mathbf{J}_N. \quad (1.22)$$

Before proceeding with the development of the terms in the weak form, it is necessary to fix the behavior of the unknown fields on the boundary. To do this, the different characteristics associated with the different types of boundary which can be found in the problem will be specially specified. There are three types of boundary, in particular:

- boundary called S_1: all three components of the vector are assigned, so

$$\begin{aligned} &\mathbf{E} = \mathbf{E}_{S_1} \text{on } S_1 \\ &\text{and} \\ &\mathbf{H} = \mathbf{H}_{S_1} \text{on } S_1; \end{aligned} \quad (1.23)$$

- boundary called S_2: only the normal component to the boundary, namely the component parallel to the normal vector that characterizes $\hat{\mathbf{n}}$ is known, so

$$\begin{aligned} &\hat{\mathbf{n}} \cdot \mathbf{E} = E_{S_2} \text{ on } S_2 \\ &\text{and} \\ &\hat{\mathbf{n}} \cdot \mathbf{H} = H_{S_2} \text{ on } S_2, \end{aligned} \quad (1.24)$$

where $\tilde{\mathbf{n}}$ is the unit normal to the boundary;

- boundary called S_3: the tangential components to the normal vector that characterizes $\hat{\mathbf{n}}$ are enforced, so

$$\begin{aligned} &\hat{\mathbf{n}} \times \mathbf{E} = \mathbf{E}_{S_3} \text{ on } S_3 \\ &\text{and} \\ &\hat{\mathbf{n}} \times \mathbf{H} = \mathbf{H}_{S_3} \text{ on } S_3; \end{aligned} \tag{1.25}$$

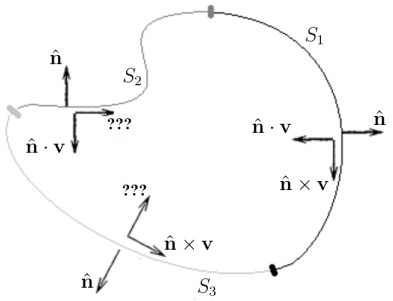

FIGURE 1.1. The three boundary types and the relative known and unknown components.

Combining the following classification of the boundary and the properties of the test functions with compact support, it is possible to fix the following boundary conditions, for both the test function $\mathbf{v}$ and $\mathbf{z}$:

$$\begin{cases} \mathbf{v} = \mathbf{0} \quad \text{on } S_1 \\ \hat{\mathbf{n}} \cdot \mathbf{v} = 0 \quad \text{on } S_2 \\ \hat{\mathbf{n}} \times \mathbf{v} = \mathbf{0} \quad \text{on } S_3 \end{cases} \tag{1.26}$$

and

$$\begin{cases} \mathbf{z} = \mathbf{0} \quad \text{on } S_1 \\ \hat{\mathbf{n}} \cdot \mathbf{z} = 0 \quad \text{on } S_2 \\ \hat{\mathbf{n}} \times \mathbf{z} = \mathbf{0} \quad \text{on } S_3. \end{cases} \tag{1.27}$$

The development of the weak form is based on the Gauss divergence theorem, on a particular identity vector, and on certain properties of the mixed product of vectors:

$$\int_V \nabla \cdot \mathbf{F}\, dV = \oint_S \hat{\mathbf{n}} \cdot \mathbf{F}\, dS \quad \text{divergence theorem} \tag{1.28}$$

$$\nabla \cdot (\mathbf{v} \times \boldsymbol{\phi}) = (\nabla \times \mathbf{v}) \cdot \boldsymbol{\phi} - \mathbf{v} \cdot (\nabla \times \boldsymbol{\phi}) \quad \text{vector identity} \tag{1.29}$$

$$\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c} \quad \text{triple product property} \tag{1.30}$$

$$\mathbf{a} \times \mathbf{b} \cdot \mathbf{c} = (\mathbf{c} \times \mathbf{a}) \cdot \mathbf{b} = -(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b} \quad \text{triple product property.} \tag{1.31}$$

Weak formulation for the electric field $\mathbf{E}(\mathbf{r}, t)$

Let us start from the analysis of equation (1.21). The only term to be further developed is the one involving $\nabla \times \nabla \times \mathbf{E}$. The latter, imposing $\nabla \times \mathbf{E} = \boldsymbol{\phi}$ becomes:

$$\int_V \mathbf{v} \cdot \nabla \times \boldsymbol{\phi} \; . \tag{1.32}$$

Thanks to (1.29), this term can be written as :

$$\int_V \mathbf{v} \cdot \nabla \times \boldsymbol{\phi} \; = \int_V \nabla \times \mathbf{v} \cdot \boldsymbol{\phi} \; - \int_V \nabla \cdot (\mathbf{v} \times \boldsymbol{\phi}) \; . \tag{1.33}$$

The last term of (1.33) can be expressed by using a surface integral thanks to (1.28). Rewriting $\boldsymbol{\phi} = \nabla \times \mathbf{E}$ and using (1.30), the integral initially taken under consideration will be expressed as:

$$\int_V \mathbf{v} \cdot (\nabla \times \nabla \cdot \mathbf{E}) \; = \int_V (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) \; - \oint_S (\hat{\mathbf{n}} \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) \; . \tag{1.34}$$

The integral on the surface can be further simplified. In fact, the surface consists of three different parts such that $S = S_1 \cup S_2 \cup S_3$, as seen before. Thanks to the properties (1.26), the only part different from zero turns out to be the one relative to the boundary S_2, that is:

$$\int_V \mathbf{v} \cdot (\nabla \times \nabla \cdot \mathbf{E}) \; = \int_V (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) \; - \int_{S_2} (\hat{\mathbf{n}} \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) \; . \tag{1.35}$$

Using (1.31), this becomes:

$$\int_V \mathbf{v} \cdot (\nabla \times \nabla \cdot \mathbf{E}) \; = \int_V (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) \; + \int_{S_2} (\mathbf{v} \cdot \hat{\mathbf{n}}) \times (\nabla \times \mathbf{E}) \; . \tag{1.36}$$

Thus, after the appropriate manipulations and simplifications, (1.21) becomes:

$$\int_V (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) + \int_{S_2} (\hat{\mathbf{n}} \times \nabla \times \mathbf{E}) \cdot \mathbf{v} + \int_V \mathbf{v} \cdot (\epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{E}}{\partial t}) \; = - \int_V \mathbf{v} \cdot \mu \frac{\partial \mathbf{J}_N}{\partial t} \; . \tag{1.37}$$

Weak formulation for the magnetic field $\mathbf{H}(\mathbf{r},t)$

The whole process can be repeated for equation (1.22), obtaining the analogous result. The equation for the magnetic induction field $\mathbf{H}$ will be:

$$\int_V (\nabla\times\mathbf{z})\cdot(\nabla\times\mathbf{H}) + \int_{S_2} (\hat{\mathbf{n}}\times\nabla\times\mathbf{H})\cdot\mathbf{z} + \int_V \mathbf{z}\cdot(\epsilon\mu\frac{\partial^2\mathbf{H}}{\partial t^2}+\sigma\mu\frac{\partial\mathbf{H}}{\partial t}) \;=\; = \int_V \mathbf{z}\cdot\nabla\times\mathbf{J}_N \,. \tag{1.38}$$

1.3.4 Completing the boundary conditions

In general, for each type of boundary surface in the domain, each component of the unknown vector should be known. The definition given earlier of the boundaries did assume the lack of tangential component on S_2, and of normal component on S_3.
The problem was partly solved as the weak form has allowed the construction in a natural way only of the tangential component missing on S_2, while it was not sufficient to complete the boundary S_3.
The latter, however, is in full accord with the nature of the Maxwell equation. To explain the motivation, we have to consider (1.1). As a matter of fact, we can show that two Gauss' flux theorems (one for the electric field and the other one for the magnetic field) can be obtained with the appropriate manipulation of two curl equations. Let us apply the divergence operator on both curl equations. Thus we obtain:

$$\begin{cases} \dfrac{\partial}{\partial t}(\nabla\cdot\mathbf{B}(\mathbf{r},t)) = -\nabla\cdot(\nabla\times\mathbf{E}(\mathbf{r},t)) \\ \dfrac{\partial}{\partial t}(\nabla\cdot\mathbf{D}(\mathbf{r},t)) = \nabla\cdot(\nabla\times\mathbf{H}(\mathbf{r},t)) - \nabla\cdot\mathbf{J}(\mathbf{r},t)). \end{cases} \tag{1.39}$$

It is simple to observe that all the $\nabla\cdot(\nabla\times ...)$ terms are identically equal to zero [3]. Let us insert (1.2) into the second equation, so as to obtain that:

$$\begin{cases} \dfrac{\partial}{\partial t}(\nabla\cdot\mathbf{B}(\mathbf{r},t)) = 0 \\ \dfrac{\partial}{\partial t}(\nabla\cdot\mathbf{D}(\mathbf{r},t) - \rho(\mathbf{r},t)) = 0 \end{cases} \quad\Rightarrow\quad \begin{cases} \nabla\cdot\mathbf{B}(\mathbf{r},t) = C_1(\mathbf{r}) \\ \nabla\cdot\mathbf{D}(\mathbf{r},t) - \rho(\mathbf{r},t)) = C_2(\mathbf{r}). \end{cases} \tag{1.40}$$

Both equations represent the two Gauss' up to a constant. But both these constants are equal to zero. In fact, for the magnetic field $\mathbf{B}(\mathbf{r},t)$, not being

able to exist magnetic monopoles, the divergence is certainly equal to zero; vice-versa, for the electric flux field $\mathbf{D}(\mathbf{r}, t)$, being able to exist single charges in nature, in general, the divergence is equal to the charge distribution. Thus:

$$\begin{cases} \nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0 \\ \nabla \cdot \mathbf{D}(\mathbf{r}, t) = \rho(\mathbf{r}, t). \end{cases} \tag{1.41}$$

The independence of the two equations containing the curl operator has been proved. Now, let us consider a domain with two or more different media. It is necessary that Maxwell's equations at every discontinuity interface be satisfied. It is possible to prove that for every unknown field there is a certain interface condition linking different media across the discontinuity.
We can define the left electromagnetic characteristic media as μ_1 and ϵ_1, and the right μ_2 and ϵ_2, respectively (Figure 1.2).

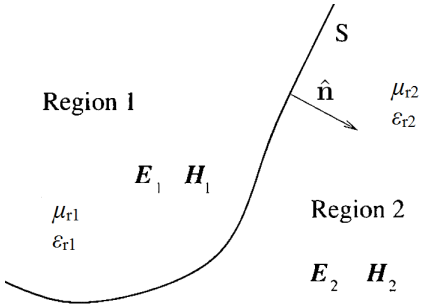

FIGURE 1.2. Geometry of the interface region and the electromagnetic quantities used in the interface conditions.

The correct relations, called *interface conditions*, are:

$$\begin{cases} (\mathbf{D}_2 - \mathbf{D}_1) \cdot \hat{\mathbf{n}} = \rho_s, \\ (\mathbf{B}_2 - \mathbf{B}_1) \cdot \hat{\mathbf{n}} = 0, \\ \hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{J}_s, \\ \hat{\mathbf{n}} \times (\mathbf{E}_2 - \mathbf{E}_1) = \mathbf{0}, \end{cases} \quad \text{on S} \tag{1.42}$$

with $\hat{\mathbf{n}}$ the unit normal vector entering the media tagged with 2, while ρ_s and $\mathbf{J}_s$ are the surface charge distribution and the density current distribution on the media discontinuity surface S, respectively. It is possible to demonstrate that the first relations of (1.42) can be derived from the two Gauss' laws, while the last two are derived from the equation with the curl operator. Thanks

to the previous proof, we will consider only two out of these ones for future goals, and in particular:

$$\begin{cases} \hat{\mathbf{n}} \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{J}_s, \\ \hat{\mathbf{n}} \times (\mathbf{E}_2 - \mathbf{E}_1) = \mathbf{0}. \end{cases} \quad \text{on S} \tag{1.43}$$

We can observe that these relations involve only the *tangential component* on the discontinuity S. The conditions of the interface are integral relations, valid at each point of the domain, not necessarily at the turn of a discontinuity. Thanks to this, it is also possible to use them on the domain boundary.
Because of this, on all edges previously defined, S_1, S_2, S_3, it suffices to know *only* the tangential component to have a well-defined problem. The fundamental result is $S_1 = S_3$ and so, the problem to complete the knowledge of the components on S_3 is solved.
Thus, the boundary conditions relating to each boundary for the electric field and the magnetic field are:

$$\begin{cases} \hat{\mathbf{n}} \times \mathbf{E} = \mathbf{E}_{S_1} = \mathbf{E}_{S_3} = \mathbf{a}_E \quad \text{on } S_1 \\ \hat{\mathbf{n}} \times \nabla \times \mathbf{E} = \mathbf{E}_{S_2} \quad \text{on } S_2, \end{cases} \tag{1.44}$$

$$\begin{cases} \hat{\mathbf{n}} \times \mathbf{H} = \mathbf{H}_{S_1} = \mathbf{H}_{S_3} = \mathbf{a}_H \quad \text{on } S_1 \\ \hat{\mathbf{n}} \times \nabla \times \mathbf{H} = \mathbf{H}_{S_2} \quad \text{on } S_2. \end{cases} \tag{1.45}$$

respectively, where $\mathbf{a}_E$ and $\mathbf{a}_H$ are assigned functions.

This procedure has allowed to demonstrate the boundary conditions which are necessary to obtain a well-defined problem. Thanks to this, will be chosen such boundary conditions for a complete writing of the strong formulation.

1.3.5 The function space H(curl,V)

In this short paragraph, some information about function spaces is given, and the actual space will be chosen.

Let us introduce two definitions about two function spaces. More information and proofs can be found in [24]:

- H(div,V): all the solutions that belong to it have the normal component on ∂V well-defined; it is defined as $\mathrm{H}(\mathrm{div},V) = \{\mathbf{v} \in (L^2(V))^3 : \nabla \cdot \mathbf{v} \in L^2(V)\}$;

- H(curl,V): all the solutions belong to it have the tangential component on ∂V well-defined; it is defined as H(curl,V) = $\{\mathbf{v} \in (L^2(V))^3 : \nabla \times \mathbf{v} \in (L^2(V))^3\}$.

For the considerations in the previous paragraph about the independence of two curl equations and about the boundary conditions, we choose H(curl,V) as the actual space where to find the solution of (1.37) and (1.38).
Finally, the complete weak formulations are:

$$\begin{aligned}&\text{for any } t \in (0,T] \text{ find } \mathbf{E}(\mathbf{r},t) \in \text{H(curl,V) such that}\\ &\int_V (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{E}) \; + \int_{S_2} (\hat{\mathbf{n}} \times \nabla \times \mathbf{E}) \cdot \mathbf{v} \; + \int_V \mathbf{v} \cdot (\epsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{E}}{\partial t}) \; = \\ &= -\int_V \mathbf{v} \cdot \mu \frac{\partial \mathbf{J}_N}{\partial t} \quad \forall \mathbf{v} \in \text{ H(curl,V);}\end{aligned} \tag{1.46}$$

and

$$\begin{aligned}&\text{for any } t \in (0,T] \text{ find } \mathbf{H}(\mathbf{r},t) \in \text{H(curl,V) such that}\\ &\int_V (\nabla \times \mathbf{z}) \cdot (\nabla \times \mathbf{H}) \; + \int_{S_2} (\hat{\mathbf{n}} \times \nabla \times \mathbf{H}) \cdot \mathbf{z} \; + \int_V \mathbf{z} \cdot (\epsilon\mu \frac{\partial^2 \mathbf{H}}{\partial t^2} + \sigma\mu \frac{\partial \mathbf{H}}{\partial t}) \; = \\ &= \int_V \mathbf{z} \cdot \nabla \times \mathbf{J}_N \quad \forall \mathbf{z} \in \text{ H(curl,V).}\end{aligned} \tag{1.47}$$

1.3.6 The strong formulation of $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{H}(\mathbf{r},t)$

The analysis carried out in the previous paragraph, showed which components of the unknown vector should be fixed to make sure that the problem can be solved. We also found out that not all of three types of boundaries are used for the completion of the boundary conditions.
The problems in strong form are now written in full form, but first, we introduce the initial and the boundary conditions useful for the thesis.

Initial conditions and boundary conditions

Equations (1.17) and (1.20) have no unique solution without specifying the initial and boundary conditions. Thus, it is necessary to impose, in particular, two time-initial conditions and boundary conditions on the surfaces representing a finite-dimension object and at infinity:

- **initial conditions**: we have to fix the unknown field value at a specific time $t = \tilde{t}$ for all domain spatial coordinates. One usually chooses $\tilde{t}$ as the instant equal to zero. In this situation, we are dealing with an initial values problem. In Maxwell's strong form system, we have two unknown vector fields and the maximum derivative order is the second. Thus, we have to impose two temporal conditions at $\tilde{t} = 0$, for each unknown field, and their time derivative:

$$\begin{cases} \mathbf{E}(\mathbf{r}, 0) = \mathbf{E}(\mathbf{r})_0, \\ \left.\dfrac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t}\right|_0 = \mathbf{E}'(\mathbf{r})_0, \end{cases} \tag{1.48}$$

and

$$\begin{cases} \mathbf{H}(\mathbf{r}, 0) = \mathbf{H}(\mathbf{r})_0, \\ \left.\dfrac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t}\right|_0 = \mathbf{H}'(\mathbf{r})_0. \end{cases} \tag{1.49}$$

with $\mathbf{E}(\mathbf{r})_0$, $\mathbf{E}'(\mathbf{r})$, $\mathbf{H}(\mathbf{r})_0$ and $\mathbf{H}'(\mathbf{r})_0$ assigned functions.

- **boundary condition on a finite-dimension object**: we have to fix the unknown field in a specific domain region $\mathbf{r}_b$ for all time domain interval. There are some conditions to be able to apply on a finite-dimension object [24] and [16].

  The first one is a simplified case of the interface conditions. In fact, when a material on one side of the discontinuity surface is considered an *electric perfect conductor*, we can obtain the *electric perfect conductor condition* with a simple limit operation. Let us consider Ohm's local law $\mathbf{J} = \sigma\mathbf{E}$. In an electric perfect conductor $\sigma \to \infty$ and if the current density has to remain bounded the field $\mathbf{E} \to 0$. This suggests that the electric field in a perfect conductor vanishes. If we consider the media labeled with 1 as an electric perfect conductor, the condition becomes:

$$\hat{\mathbf{n}} \times \mathbf{E}_2 = 0 \qquad \text{on } \mathbf{r}_b. \tag{1.50}$$

  A similar consideration can be applied for a *magnetic perfect conductor*, with the limit operation carried out on the magnetic permeability μ.

Thus, if we consider medium labeled with 1 a magnetic perfect conductor, we obtain the *magnetic perfect conductor condition*:

$$\hat{\mathbf{n}} \times \mathbf{H}_2 = \mathbf{J}_{\mathbf{r}_b} \qquad \text{on } \mathbf{r}_b. \tag{1.51}$$

This type of condition is a special case of (1.25) and it is called *Dirichlet boundary condition*. For this reason it will be applied on the boundary type S_2, which is re-defined S_2^{fd} because applied on a finite-dimensional object.

The second condition imposes the curl of the unknown field. In fact, for the uniqueness of the solution in a domain V, one can fix, on different boundaries, both the field and its derivatives.
Thus, the conditions are:

$$\hat{\mathbf{n}} \times \nabla \times \mathbf{E} = \mathbf{E}_{S_2} = \mathbf{b}_E \quad \text{on } \mathbf{r}_b, \tag{1.52}$$

and

$$\hat{\mathbf{n}} \times \nabla \times \mathbf{H} = \mathbf{H}_{S_2} = \mathbf{b}_H \quad \text{on } \mathbf{r}_b, \tag{1.53}$$

where $\mathbf{b}_E$ and $\mathbf{b}_H$ are assigned functions.
This kind of condition, called *Neumann boundary condition*, is the same emerged during the development of the weak formulation. For this reason, it will be applied to the edge of type S_1^{fd}.

- **boundary condition at infinity**: we have to fix the unknown field in a specific domain region $\mathbf{r}_b \to \infty$ for all time domain interval. For this situation, we need a special condition, called *Silver-Muller condition* ([24], [16]), where the quantity Z_0, which will be analysed in the section about the frequency domain, is the *vacuum impedance*:

$$\lim_{r\to\infty} r^{\frac{(n-1)}{2}}[\mathbf{H}(\mathbf{r},t) - \frac{1}{Z_0}\mathbf{r} \times \mathbf{E}(\mathbf{r},t)] = 0 \quad \text{on } \mathbf{r}_b \to \infty. \tag{1.54}$$

It is possible to write the same condition with the field $\mathbf{H}$ replaced with $\mathbf{E}$. Let us post cross multiply (1.54) by the vector $\mathbf{r}$, thus:

$$\lim_{r\to\infty} r^{\frac{(n-1)}{2}}[\mathbf{H} \times \mathbf{r} - \frac{1}{Z_0}(\mathbf{r} \times \mathbf{E}) \times \mathbf{r}] = 0 \quad \text{on } \mathbf{r}_b \to \infty. \tag{1.55}$$

The following vector identity is used for every vector:

$$\mathbf{A} = (\mathbf{r}\cdot\mathbf{A})\mathbf{r} + (\mathbf{r}\times\mathbf{A})\times\mathbf{r} \quad \Rightarrow \quad (\mathbf{r}\times\mathbf{A})\times\mathbf{r} = \mathbf{A} - (\mathbf{r}\cdot\mathbf{A})\mathbf{r}. \tag{1.56}$$

Let us use (1.56) on the electric field and substitute in the previous equation. Then, let us pre-cross multiply again by $\mathbf{r}$:

$$\lim_{r\to\infty} r^{\frac{(n-1)}{2}}[\mathbf{r}\times\mathbf{H}\times\mathbf{r} - \frac{1}{Z_0}\mathbf{r}\times(\mathbf{E} - (\mathbf{r}\cdot\mathbf{E})\mathbf{r})] = 0, \tag{1.57}$$

where the last left-hand term is equal to zero for (1.56). Thus, we can collect the quantity -$\mathbf{r}\times$ on the left-hand side and obtain another version of the Silver-Muller condition:

$$\lim_{r\to\infty} r^{\frac{(n-1)}{2}}[\mathbf{E}(\mathbf{r},t) - Z_0\mathbf{H}(\mathbf{r},t)\times\mathbf{r}] = 0 \quad \text{on } \mathbf{r}_b \to \infty. \tag{1.58}$$

We can replace the vector $\mathbf{r}$ with the unit normal vector $\mathbf{n}$ as their direction to infinity are parallel. This operation also allows to not consider the limit operator anymore. In this way, the two Silver-Muller conditions became:

$$\mathbf{H}(\mathbf{r},t) - \frac{1}{Z_0}\mathbf{r}\times\mathbf{E}(\mathbf{r},t) = 0, \tag{1.59}$$

and

$$\mathbf{E}(\mathbf{r},t) - Z_0\mathbf{H}(\mathbf{r},t)\times\mathbf{r} = 0. \tag{1.60}$$

Now, let us apply the time-derivative to (1.59) in order to obtain:

$$\frac{\partial\mathbf{H}}{\partial t} - \frac{1}{Z_0}\frac{\partial}{\partial t}(\hat{\mathbf{n}}\times\mathbf{E}) = \mathbf{0}. \tag{1.61}$$

The Faraday-Neumann-Lenz law allows us to replace the left-hand side of the equation by:

$$-\frac{1}{\mu_0}\nabla\times\mathbf{E} - \frac{1}{Z_0}\frac{\partial}{\partial t}(\hat{\mathbf{n}}\times\mathbf{E}) = \mathbf{0}. \tag{1.62}$$

Multiplying the latter by $\hat{\mathbf{n}}$, we obtain the condition on the boundary:

$$\hat{\mathbf{n}} \times (\nabla \times \mathbf{E}) = -\frac{\mu_0}{Z_0}\frac{\partial}{\partial t}(\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{E}). \tag{1.63}$$

This boundary condition will be used to solve the problem of the electric field, since this is the only unknown field appearing in the relation.

Then, let us apply the time-derivative to (1.60) in order to obtain:

$$\frac{\partial \mathbf{E}}{\partial t} - Z_0 \mathbf{H} \times \hat{\mathbf{n}} = \mathbf{0}. \tag{1.64}$$

The Maxwell-Ampère law allows us to substitute the first term of the relation, yielding:

$$\frac{1}{\epsilon_0}\nabla \times \mathbf{H} - Z_0 \mathbf{H} \times \hat{\mathbf{n}} = \mathbf{0}. \tag{1.65}$$

Multiplying the latter by $\hat{\mathbf{n}}$, we obtain the condition on the boundary:

$$\hat{\mathbf{n}} \times \nabla \times \mathbf{H} = -\epsilon_0 Z_0 \frac{\partial}{\partial t}(\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \mathbf{H}). \tag{1.66}$$

This boundary condition will be used to solve the problem of the magnetic field, since this is the only unknown field to appear in the relation.

Again, we can immediately notice that both terms on the left side of (1.63) and (1.66) are in the form that appears in (1.44) and (1.45). For this reason, both will be applied on a boundary type S_2, re-defined S_2^∞ as open edge. The only difference is that E_{S_2} and H_{S_2} are not known functions but contain the unknown field. This condition is called *Robin boundary condition*.

It is important to note that not all the previous conditions have to hold on the boundary: the simultaneous existence will create the vanishing of the unknown fields $\mathbf{E}$ and $\mathbf{H}$. The possible solution is to fix all the conditions relative to the electric field or the magnetic field; another choice is to divide the boundary into two separate zones where to apply conditions on the electric field or the magnetic field.

Strong formulation for the electric field $\mathbf{E}(\mathbf{r},t)$

The problem for the electric field consists of equation (1.17) with the addition of the initial conditions (1.48) and the boundary conditions in the form previously found:

$$\begin{cases} \nabla\times\nabla\times\mathbf{E}(\mathbf{r},t)+\mu\epsilon\dfrac{\partial^2\mathbf{E}(\mathbf{r},t)}{\partial^2 t}+\mu\sigma\dfrac{\partial\mathbf{E}(\mathbf{r},t)}{\partial t}=-\mu\dfrac{\partial\mathbf{J}_N(\mathbf{r},t)}{\partial t} \quad \text{in } V \\ \mathbf{E}(\mathbf{r},0)=\mathbf{E}(\mathbf{r})_0 \\ \left.\dfrac{\partial\mathbf{E}(\mathbf{r},t)}{\partial t}\right|_0=\mathbf{E}'(\mathbf{r})_0 \\ \hat{\mathbf{n}}\times\mathbf{E}=\mathbf{a}_E \quad \text{on } S_1^{fd} \\ \hat{\mathbf{n}}\times\nabla\times\mathbf{E}=\mathbf{b}_E \quad \text{on } S_2^{fd} \\ \hat{\mathbf{n}}\times\nabla\times\mathbf{E}=-\dfrac{\mu_0}{Z_0}\dfrac{\partial}{\partial t}(\hat{\mathbf{n}}\times\hat{\mathbf{n}}\times\mathbf{E}) \quad \text{on } S_2^{\infty}, \end{cases} \tag{1.67}$$

Strong formulation for the magnetic field $\mathbf{H}(\mathbf{r},t)$

The problem for the magnetic field comprises equation (1.20) with the addition of the initial conditions (1.49) and the boundary conditions in the form previously found:

$$\begin{cases} \nabla\times\nabla\times\mathbf{H}(\mathbf{r},t)+\mu\epsilon\dfrac{\partial^2\mathbf{H}(\mathbf{r},t)}{\partial^2 t}+\mu\sigma\dfrac{\partial\mathbf{H}(\mathbf{r},t)}{\partial t}=\nabla\times\mathbf{J}_N(\mathbf{r},t) \quad \text{in } V \\ \mathbf{H}(\mathbf{r},0)=\mathbf{H}(\mathbf{r})_0 \\ \left.\dfrac{\partial\mathbf{H}(\mathbf{r},t)}{\partial t}\right|_0=\mathbf{H}'(\mathbf{r})_0 \\ \hat{\mathbf{n}}\times\mathbf{H}=\mathbf{a}_H \quad \text{on } S_1^{fd} \\ \hat{\mathbf{n}}\times\nabla\times\mathbf{H}=\mathbf{b}_H \quad \text{on } S_2^{fd} \\ \hat{\mathbf{n}}\times\nabla\times\mathbf{H}=-\epsilon_0 Z_0\dfrac{\partial}{\partial t}(\hat{\mathbf{n}}\times\hat{\mathbf{n}}\times\mathbf{H}) \quad \text{on } S_2^{\infty}. \end{cases} \tag{1.68}$$

The notation previously adopted was used only to provide an overview of all the cases: in fact, on the boundary type S_2, we can impose either the Neumann condition only, or the Robin condition only, or both of them.

1.4 Orthogonality, wavenumber and impedance

If the quantities of the problem have a harmonic behavior in time, then we can use the representation of Maxwell equations in the frequency domain. This step is allowed only if all of the inputs, i.e., the forcing terms, exhibit harmonic behavior and if the dynamic system is linear-time-invariant (LTI) [34], as the outputs, in particular the system's response, will also be harmonic, but with different amplitude and phase.
The quantities are expressed in the frequency domain thanks to the Fourier transform, which is defined in this way:

$$\tilde{\mathbf{F}}(\mathbf{r},\omega) = \int_{-\infty}^{+\infty} \mathbf{F}(\mathbf{r},t) e^{-j\omega t}\, dt, \tag{1.69}$$

with temporal frequency $\omega > 0$ as the main angular frequency contained in the time-dependent electromagnetic phenomena. Thus, if we apply (1.69) to (1.14), we obtain the frequency-dependent Maxwell system [24]:

$$\begin{cases} \epsilon \nabla \cdot \tilde{\mathbf{E}} = \tilde{\rho}, \\ j\omega\mu \tilde{\mathbf{H}} = -\nabla \times \tilde{\mathbf{E}}, \\ \nabla \cdot \tilde{\mathbf{H}} = 0, \\ j\omega\epsilon \tilde{\mathbf{E}} = \nabla \times \tilde{\mathbf{H}} - \sigma \tilde{\mathbf{E}} - \tilde{\mathbf{J}}_N. \end{cases} \tag{1.70}$$

This preliminary form of the system of equations is practically useless for the numerical solution of the problem. The usefulness of this formalization is that, with some simplifications and manipulations, many important properties and definitions will be found. These ones will later be exploited to set up the specific problems concerning this thesis.

Now it is very useful to proceed with the wavenumber analysis. For this reason, we define the space-frequency-dependent unknown parts as $\tilde{\mathbf{F}}(\mathbf{r},\omega) = \tilde{\mathbf{F}}(\omega) e^{(-\mathbf{s}\cdot\mathbf{r})} = \mathbf{F}_a e^{(-\mathbf{s}\cdot\mathbf{r})}$, where $\mathbf{F}_a$ is the unknown magnitude and $\mathbf{s} = \mathbf{a} + j\mathbf{k}$ is the complex propagation vector defined by the attenuation $\mathbf{a}$ and the wavenumber $\mathbf{k}$. Let us suppose that the known current density $\tilde{\mathbf{J}}_N$ and the charge density $\tilde{\rho}$ are zero to simplify the following calculation without loss of validity of the argument.

System (1.70) can be written as:

$$\begin{cases} -\epsilon(\mathbf{s}\cdot\mathbf{E}_a)=0 \\ j\omega\mu\mathbf{H}_a=\mathbf{s}\times\mathbf{E}_a \\ -(\mathbf{s}\cdot\mathbf{H}_a)=0 \\ j\omega\epsilon\mathbf{E}_a=-\mathbf{s}\times\mathbf{H}_a-\sigma\mathbf{E}_a \end{cases} \Rightarrow \begin{cases} \mathbf{s}\cdot\mathbf{E}_a=0 \\ \mathbf{H}_a=\dfrac{\mathbf{s}\times\mathbf{E}_a}{j\omega\mu} \\ \mathbf{s}\cdot\mathbf{H}_a=0 \\ \mathbf{E}_a=-\dfrac{\mathbf{s}\times\mathbf{H}_a}{j\omega\epsilon+\sigma}. \end{cases} \tag{1.71}$$

First conclusion: the first and third equations in (1.71) show that the electric and the magnetic fields are perpendicular to the propagation vector. The second equation shows how the cross product between the propagation vector and the electric field has the magnetic field as a result up to a multiplying factor. In the same way, the fourth equation shows the same result for the electric field, with the minus sign. This means that the propagation vector $\mathbf{s}$, electric field $\mathbf{E}$ and magnetic field $\mathbf{H}$ are orthogonal one to the others.

Let us now compute the cross product between $\mathbf{E}_a = E_a\mathbf{i}_{E_a}$ and $\mathbf{H}_a = H_a\mathbf{i}_{H_a}$, with $\mathbf{i}_{E_a}$ and $\mathbf{i}_{H_a}$ the unit vectors such that $\mathbf{i}_{E_a}\times\mathbf{i}_{H_a}=\mathbf{i}_s$:

$$\begin{aligned} \mathbf{E}_a\times\mathbf{H}_a &= \left(-\frac{\mathbf{s}\times\mathbf{H}_a}{\sigma+j\epsilon\omega}\right)\times\left(\frac{\mathbf{s}\times\mathbf{E}_a}{j\mu\omega}\right) \\ &= \left(\frac{sH_a}{\sigma+j\omega\epsilon}\mathbf{i}_{E_a}\right)\times\left(\frac{sE_a}{j\omega\mu}\mathbf{i}_{H_a}\right) \\ &= \frac{s^2H_aE_a}{(\sigma+j\omega\epsilon)(j\omega\mu)}\mathbf{i}_s. \end{aligned} \tag{1.72}$$

The last cross product is equal to $E_aH_a\mathbf{i}_s$, thus:

$$s^2=-\omega^2\mu\epsilon+j\omega\mu\sigma. \tag{1.73}$$

The vector $\mathbf{s}$ is multiplied with itself, getting

$$\mathbf{s}\cdot\mathbf{s}=s^2=|\mathbf{a}|^2-|\mathbf{k}|^2+j2\mathbf{a}\cdot\mathbf{k}. \tag{1.74}$$

Being(1.74) and (1.73) the same quantity, the real and imaginary parts must be the same, which leads to the writing of two conditions:

$$\begin{cases} |\mathbf{a}|^2-|\mathbf{k}|^2=-\omega^2\mu\epsilon \\ \mathbf{a}\cdot\mathbf{k}=\omega\mu\sigma \end{cases} \Rightarrow \begin{cases} a^2-k^2=-\omega^2\mu\epsilon \\ ak\cos(\theta)=\omega\mu\sigma, \end{cases} \tag{1.75}$$

with $0 \leqslant \theta < \pi/2$ the angle between the vector $\mathbf{a}$ and $\mathbf{k}$. Leaving out some mathematics steps, the two equations can be manipulated in order to obtain a single one:

$$\left(\frac{k}{a}\right)^2 - \frac{2\epsilon\omega\cos(\theta)}{\sigma}\left(\frac{k}{a}\right) - 1 = 0. \tag{1.76}$$

The solution in terms of (k/a) is computed and, having used the equations (1.75), the expressions of $|\mathbf{k}|^2$ and $|\mathbf{a}|^2$ are obtained:

$$\begin{aligned} |\mathbf{k}|^2 &= \frac{\omega^2\mu\epsilon}{2}\left[\sqrt{1+\left(\frac{\sigma}{\omega\epsilon\cos(\theta)}\right)^2}+1\right], \\ |\mathbf{a}|^2 &= \frac{\omega^2\mu\epsilon}{2}\left[\sqrt{1+\left(\frac{\sigma}{\omega\epsilon\cos(\theta)}\right)^2}-1\right]. \end{aligned} \tag{1.77}$$

Second conclusion: depending on the value assumed by σ, the behavior of the unknown fields can be analized when the electromagnetic phenomena takes place into two fundamental materials:

- for $\sigma \neq 0$ the medium is said *conductor*, i.e. the charge distribution is possible and thus also a density current distribution. In this case, $|\mathbf{k}|^2$ and $|\mathbf{a}|^2$ depend on the angle θ and assume finite values. The fields $\mathbf{E}_a$ and $\mathbf{H}_a$ are characterized by an exponential attenuation in space;
- for $\sigma = 0$ the medium is said *dielectric* and the flow of current is not possible. In this case, the two quantities result

$$\begin{cases} |\mathbf{k}| = \omega\sqrt{\mu\epsilon} \\ |\mathbf{a}| = 0, \end{cases} \tag{1.78}$$

  which implies the absence of a spatial decay for the fields $\mathbf{E}_a$ and $\mathbf{H}_a$ because of $\mathbf{s} = j\mathbf{k}$. The orthogonality relationship is shown in Figure 1.3.

Let us consider the second equation in (1.71):

$$H_a\mathbf{i}_{H_a} = -\frac{sE_a}{\omega\mu}\mathbf{i}_{H_a}. \tag{1.79}$$

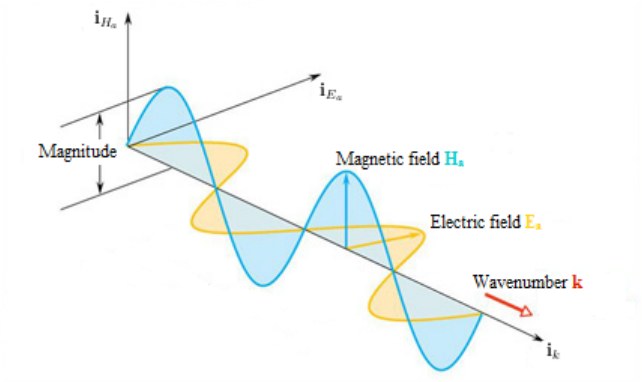

FIGURE 1.3. The electromagnetic wave, composed by the electric and magnetic field, travels in the direction of wavenumber vector conserving the orthogonality among three vector $\mathbf{E}_a$, $\mathbf{H}_a$ and $\mathbf{k}$ when the medium is a dielectric material.

We can substitute (1.73) in (1.79) and find the relation among E_a and H_a:

$$\frac{E_a}{H_a} = \sqrt{\frac{\mu}{\epsilon}\frac{1+j\dfrac{\sigma}{\omega\epsilon}}{1+\left(\dfrac{\sigma}{\omega\epsilon}\right)^2}}. \tag{1.80}$$

Third conclusion: these developments lead to the definition of *impedance*. It is indicated with Z, when it is related to a generic medium, or with Z_0, when it is related to vacuum [16] and [24]

$$Z = \sqrt{\frac{\mu_0}{\epsilon_0}}\sqrt{\frac{\mu_r}{\epsilon_r}\frac{1+j\dfrac{\sigma}{\omega\epsilon}}{1+\left(\dfrac{\sigma}{\omega\epsilon}\right)^2}}, \text{ with } Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}. \tag{1.81}$$

Now, using (1.73) and (1.81), it is possible to rewrite the expression of electric and magnetic magnitude after some algebraic manipulation in the following way:

$$\begin{cases} \mathbf{H}_a = \dfrac{1}{Z}\mathbf{i}_s \times \mathbf{E}_a \\ \mathbf{E}_a = -Z\mathbf{i}_s \times \mathbf{H}_a. \end{cases} \tag{1.82}$$

Fourth conclusion: let us consider a spherical volume and its normal vector

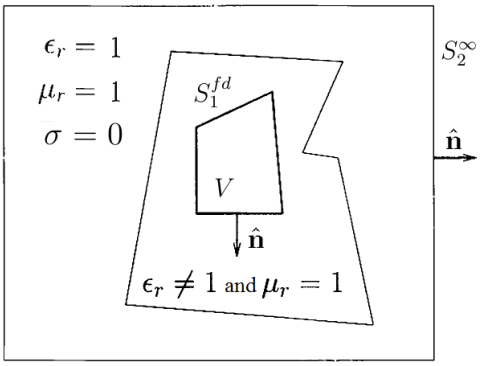

FIGURE 1.4. Geometry and boundaries for a problem with a Silver-Muller condition on the boundary of fictitious domain

$\hat{\mathbf{n}}$. Now an electromagnetic plane wave is considered. When the radius r$\rightarrow \infty$, the electromagnetic plane propagation vector will be aligned with the external normal. In this case, the two last formulas are equivalent to the argument, under limit operator, of the Silver-Muller condition (1.54) and (1.58) only if the medium considered is a dielectric. In fact, the Silver-Muller condition is valid only at infinity when the electromagnetic medium is the vacuum. So, this approximate condition is used only if it is applied on a boundary of a volume with vacuum electromagnetic characteristics ($\sigma = 0$, $\epsilon_r = 1$ and $\mu_r = 1$). To reach this constraint, a fictitious numerical domain, as shown in Figure 1.4, will be added around the external real numerical domain and, on its boundary, we will fix the Silver-Muller approximate condition. However, if the medium is air, the electromagnetic characteristics can be expressed as the vacuum ones and the fictitious domain will not be used [16].

2 | Numerical approximation of Maxwell's equations

Generally, a system of partial differential equations (PDEs) can not be solved in an analytically and relatively simple manner. Exceptions are characterized by a very particular geometric domains and without a relevant practical utility. This explains why the problem is solved by looking for approximate numerical solutions.
This chapter primarily describes the approximate weak formulation of Maxwell equations, by taking into consideration decomposition of variables, in the frequency domain. Once described the formulation, we will introduce a particular method for spatial discretization – the Finite Element Method (FEM). FEM will be illustrated by showing its practical application to a model problem. We will then close the chapter with a description of two methods of numerical solution to problems arising from spatial discretization.

2.1 The method of separation of variables

The method of separation of variables is widely used in the solution of approximate numerical problems, in particular, when the approximate solution, also known as numerical one, depends on several variables, such as space and time. The solution is expanded as a linear combination of functions depending only on space coordinates, and functions depending only on time. The main advantage of this approach, is that ones the functions depending on space are defined, the problem in weak formulation will turn into an ordinary differential problem.

In this thesis, however, thanks to the nature of the problems discussed in Chapters 4 and 5, it will be more convenient to use a formulation in the frequency domain.

This has considerable advantages:

- in the strong formulation the initial conditions will not use anymore;
- the problem will become elliptical and *Lax-Milgram Lemma* [26] will guarantee the existence and uniqueness of the approximate solution;
- the problem in weak formulation will turn into an algebraic system, which is relatively easier to solve rather than a system of differential equations.

Thanks to the hypothesis of separation of variables, we can write explicitly the unknown vector as:

$$\mathbf{E}(\mathbf{r},\omega) = \underline{\underline{\mathbf{N}}}_E(\mathbf{r})\ \underline{\mathbf{u}}_E(\omega) \tag{2.1}$$

and

$$\mathbf{H}(\mathbf{r},\omega) = \underline{\underline{\mathbf{N}}}_H(\mathbf{r})\ \underline{\mathbf{u}}_H(\omega). \tag{2.2}$$

For the electric field and the magnetic field, angular frequency dependent functions are now the new unknowns of the problem, or *degrees of freedom* (*DoFs*), placed in a vector $N_{tot} \times 1$, where N_{tot} is the number of unknown variables. The functions depending on space are called *shape functions*, which are collected within matrices $3 \times N_{tot}$, called *shape matrices*.

After the new unknown vectors have been redefined, also the vectors used as test functions must be defined. Those functions can be searched in the same function approximate space where the unknown fields are to be found. Now the test functions will be redefined as linear combinations of the same shape functions - which form the basis of the function space sought - and special coefficients that, according to the numerical method adopted, will take different values. Thus,

$$\mathbf{v} = \underline{\underline{\mathbf{N}}}_E(\mathbf{r})\ \underline{\mathbf{v}} \tag{2.3}$$

and

$$\mathbf{z} = \underline{\underline{\mathbf{N}}}_H(\mathbf{r})\ \underline{\mathbf{z}}. \tag{2.4}$$

Obviously, the dimensions of (2.3) and (2.4) are the same of $\mathbf{u}_E(\omega)$ and $\mathbf{u}_H(\omega)$.

In the further paragraphs, for a lighter notation, the dependence of the variables, the matrices and the vectors on their underlines will not be shown. Now that all the notation has been defined, the last relationships are included in the weak formulation, expressed in the frequency domain.

2.1.1 Approximate weak formulation in frequency domain for $\mathbf{E}(\mathbf{r},\omega)$

Definition (1.69) is applied to (1.46). After some computations, the weak formulation is obtained:

$$\begin{aligned}&\text{find } \tilde{\mathbf{E}}(\mathbf{r},\omega) \in \text{H(curl,V) such that}\\ &\int_V (\nabla\times\mathbf{v})\cdot(\nabla\times\tilde{\mathbf{E}}) \; + \int_{S_2} (\hat{\mathbf{n}}\times\nabla\times\tilde{\mathbf{E}})\cdot\mathbf{v} \; + \int_V \mathbf{v}\cdot(-\omega^2\epsilon\mu\tilde{\mathbf{E}} + j\omega\sigma\mu\tilde{\mathbf{E}}) =\\ &= -j\int_V \mathbf{v}\cdot\mu\omega\tilde{\mathbf{J}_N} \quad \forall\mathbf{v}\in \text{ H(curl,V)}.\end{aligned} \tag{2.5}$$

The scalar products can be replaced with the transposed sign. This operation, despite not affecting the result in any way, is very practical to show the structure of each single term in the equation, and to exploit the properties of the weak form. In fact, a weak form applies to each test function belonging to the function space chosen. This property is essential to write the system of algebraic equations. This yields:

$$\begin{aligned}&\int_V (\nabla\times(\mathbf{N}_E\mathbf{v}))^T\nabla\times(\mathbf{N}_E\mathbf{u}_E) \; +\\ &\int_V (\mathbf{N}_E\mathbf{v})^T(-\omega^2\epsilon\mu\mathbf{N}_E\mathbf{u}_E + j\omega\sigma\mu\mathbf{N}_E\mathbf{u}_E) = -\int_{S_2^{fd}} (\mathbf{N}_E\mathbf{v})^T\mathbf{b}_E \; -\\ &\int_{S_2^{\infty}} (\hat{\mathbf{n}}\times(\mathbf{N}_E\mathbf{v}))^T j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}\times(\mathbf{N}_E\mathbf{u}_E)) \; - j\int_V (\mathbf{N}_E\mathbf{v})^T\omega\mu\mathbf{J}_N.\end{aligned} \tag{2.6}$$

Making explicit the product of transposed vectors, gives:

$$\begin{aligned}&\int_V \mathbf{v}^T(\nabla\times\mathbf{N}_E)^T\nabla\times(\mathbf{N}_E\mathbf{u}_E) \; +\\ &\int_V \mathbf{v}^T\mathbf{N}_E^T(-\omega\epsilon\mu\mathbf{N}_E\mathbf{u}_E + j\omega\sigma\mu\mathbf{N}_E\mathbf{u}_E) = -\int_{S_2^{fd}} \mathbf{v}^T\mathbf{N}_E^T\mathbf{b}_E \; -\\ &\int_{S_2^{\infty}} \mathbf{v}^T(\hat{\mathbf{n}}\times\mathbf{N}_E)^T j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}\times(\mathbf{N}_E\mathbf{u}_E)) - j\int_V \mathbf{v}^T\mathbf{N}_E^T\omega\mu\mathbf{J}_N.\end{aligned} \tag{2.7}$$

Now, we are allowed to carry out of the integral all of the vectors containing the degrees of freedom, as these are independent of the spatial coordinates. Then the simplified system will be:

$$\begin{aligned}
&\mathbf{v}^T \int_V (\nabla \times \mathbf{N}_E)^T (\nabla \times \mathbf{N}_E)\ \mathbf{u}_E - \\
&\mathbf{v}^T \omega^2 \int_V \mathbf{N}_E^T \epsilon\mu \mathbf{N}_E\ \mathbf{u}_E + \mathbf{v}^T j\omega \int_V \mathbf{N}_E^T \sigma\mu \mathbf{N}_E\ \mathbf{u}_E = -\mathbf{v}^T \int_{S_2^{fd}} \mathbf{N}_E^T \mathbf{b}_E - \\
&\mathbf{v}^T j\omega \int_{S_2^{\infty}} (\hat{\mathbf{n}} \times \mathbf{N}_E)^T \frac{\mu_0}{Z_0} (\hat{\mathbf{n}} \times \mathbf{N}_E)\ \mathbf{u}_E - \mathbf{v}^T j\omega \int_V \mathbf{N}_E^T \mu \mathbf{J}_N.
\end{aligned} \tag{2.8}$$

It is now possible to define the new matrices $N_{tot} \times N_{tot}$ [16] :

$$\mathbf{T}_E = \int_V \mathbf{N}_E^T \epsilon\mu \mathbf{N}_E, \tag{2.9}$$

$$\mathbf{R}_E = \int_V \mathbf{N}_E^T \sigma\mu \mathbf{N}_E, \tag{2.10}$$

$$\mathbf{Q}_E = \int_{S_2^{\infty}} (\hat{\mathbf{n}} \times \mathbf{N}_E)^T \frac{\mu_0}{Z_0} (\hat{\mathbf{n}} \times \mathbf{N}_E), \tag{2.11}$$

$$\mathbf{S}_E = \int_V (\nabla \times \mathbf{N}_E)^T (\nabla \times \mathbf{N}_E), \tag{2.12}$$

and the new vector $N_{tot} \times 1$ [16]:

$$\mathbf{f}_E = -\int_{S_2^{fd}} \mathbf{N}_E^T \mathbf{b}_E - j\omega \int_V \mathbf{N}_E^T \mu \mathbf{J}_N. \tag{2.13}$$

Thanks to the generality, independence and compliance with the constraints of the test functions, from the previous scalar equation (2.8), we obtain a system of equations which have the degrees of freedom as unknown:

$$\mathbf{v}^T \rightarrow \qquad [-\omega^2 \mathbf{T}_E + j\omega(\mathbf{R}_E + \mathbf{Q}_E) + \mathbf{S}_E]\ \mathbf{u}_E = \mathbf{f}_E. \tag{2.14}$$

2.1.2 Approximate weak formulation in frequency domain for $\mathbf{H}(\mathbf{r}, \omega)$

Applying again (1.69) to (1.47), the weak formulation is obtained:

find $\tilde{\mathbf{H}}(\mathbf{r}, \omega) \in$ H(curl,V) such that

$$\int_V (\nabla \times \mathbf{z}) \cdot (\nabla \times \tilde{\mathbf{H}})\ + \int_{S_2} (\hat{\mathbf{n}} \times \nabla \times \tilde{\mathbf{H}}) \cdot \mathbf{z}\ + \int_V \mathbf{z} \cdot (-\omega^2 \epsilon\mu \tilde{\mathbf{H}} + j\omega\sigma\mu \tilde{\mathbf{H}}) =$$
$$= \int_V \mathbf{z} \cdot (\nabla \times \tilde{\mathbf{J}_N}) \quad \forall \mathbf{z} \in \text{ H(curl,V)}.$$

(2.15)

Similarly operations performed in the previous section can be applied to equation (2.15), in order to obtain:

$$\begin{aligned}&\mathbf{z}^T \int_V (\nabla \times \mathbf{N}_H)^T (\nabla \times \mathbf{N}_H)\ \mathbf{u}_H - \\ &\mathbf{z}^T \omega^2 \int_V \mathbf{N}_H^T \epsilon\mu \mathbf{N}_H\ \mathbf{u}_H + \mathbf{z}^T j\omega \int_V \mathbf{N}_H^T \sigma\mu \mathbf{N}_H\ \mathbf{u}_H = -\mathbf{z}^T \int_{S_2^{fd}} \mathbf{N}_H^T \mathbf{b}_H\ - \\ &\mathbf{z}^T j\omega \int_{S_2^\infty} (\hat{\mathbf{n}} \times \mathbf{N}_H)^T \epsilon_0 Z_0 (\hat{\mathbf{n}} \times \mathbf{H})\ \mathbf{u}_H + \mathbf{z}^T \int_V \mathbf{N}_H^T \nabla \times \mathbf{J}_N.\end{aligned} \tag{2.16}$$

It is also possible to define the new matrices $N_{tot} \times N_{tot}$ [16] :

$$\mathbf{T}_H = \int_V \mathbf{N}_H^T \epsilon\mu \mathbf{N}_H, \tag{2.17}$$

$$\mathbf{R}_H = \int_V \mathbf{N}_H^T \sigma\mu \mathbf{N}_H, \tag{2.18}$$

$$\mathbf{Q}_H = \int_{S_2^\infty} (\hat{\mathbf{n}} \times \mathbf{N}_H)^T \epsilon_0 Z_0 (\hat{\mathbf{n}} \times \mathbf{N}_H), \tag{2.19}$$

$$\mathbf{S}_H = \int_V (\nabla \times \mathbf{N}_H)^T (\nabla \times \mathbf{N}_H), \tag{2.20}$$

and the new vector $N_{tot} \times 1$ [16]:

$$\mathbf{f}_H = -\int_{S_2^{fd}} \mathbf{N}_H^T \mathbf{b}_H\ + \int_V \mathbf{N}_H^T \nabla \times \mathbf{J}_N, \tag{2.21}$$

same considerations previously made on the test function yield [16]:

$$\mathbf{z}^T \rightarrow \qquad [-\omega^2 \mathbf{T}_H + j\omega(\mathbf{R}_H + \mathbf{Q}_H) + \mathbf{S}_H]\ \mathbf{u}_H = \mathbf{f}_H. \tag{2.22}$$

The presence of the frequency could generate different types of solutions:

- eigenvalues and eigenfunctions: if the frequency is the unknown and the forcing term vanishes, the solutions are the resonance frequencies of the dynamic system (eigenvalues) and their modes (eigenfunctions).
- response: if the frequency is given and the forcing term is assigned, the solution is the time-response of the dynamic system.

In this thesis we will consider the second one.

2.2 Spatial discretization: finite element method

The finite element method is a technique used to solve PDEs when the spatial is complex, and analytical solutions are not known. In general, the continuous domain where the solution is to be found, is divided into elements. In every single sub-region, the solution is approximated by a piecewise polynomial function with a specified degree (i.e., a linear polynomial on tetrahedra).
Our purpose is to approximate the domain V into a discrete domain V_h characterized by a partition $\mathcal{T}_h$. This partition consists of K_j elements which depends on a spatial parameter, called h, linked with the K_j single grid element spatial dimensions. The mathematical definition is $h = max_{K_j \in \mathcal{T}_h} h_K$, where $h_K = diam(K)$, for each $K \in \mathcal{T}_h$, with $diam(K) = max_{P,Q \in K}|P - Q|$ is the diameter of K.
This parameter is important to define the grid regularity. We can define ρ_K as the radius of the maximum sphere contained in K. A triangulation $\mathcal{T}_h$, with $h > 0$, is defined *regular* if

$$\frac{h_K}{\rho_k} \leqq \delta \quad \forall K \in \mathcal{T}_h, \text{ for some } \delta > 0. \tag{2.23}$$

The functional analysis [24] shows that the right space where to find the Maxwell equations' solution is the Hilbert space H(curl,V). The functions belong to H(curl,V) space have the tangential component continuous, so we can define a special family of spaces with the following characteristics:

$$\mathcal{X}_h^r = \{u_h \in C^0(V) : u_h|_{K_j} \in \mathbb{P}_r, \text{ for each } K_j \in \mathcal{T}_h\} \quad with \quad \text{r=1,2,...} \tag{2.24}$$

where $\mathbb{P}_r$ is the polynomial space with degree equal to or less than r. Now, we are able to start to construct the approximate solution. Fore more information, the reader can consult [26], [15], [19].

2.2.1 Finite Element definition

Generally, a finite element is completely defined when a triple (K,$\mathbb{P}_r$,Σ) is specified:

- K is the geometric segment domain where the single element is defined. In one dimension, it is a line; in two dimensions it could be a triangle or a quadrilateral; in three dimensions it could be a prism, a tetrahedron or a hexahedron. The domain portion characteristic is a parameter h that identifies size and regularity;

- $\mathbb{P}_r$ is the polynomial function space, defined on K, having a $\mathbb{P}_r$ basis $\{\varphi_i(\mathbf{r})\}_{i=0}^r$. The important aspect is that this space has to be a finite dimensional, vector space;

- Σ is a set of linear functionals on $\mathbb{P}_r$, $\Sigma = \{\gamma_i : \mathbb{P}_r \to \mathbb{R}\}_{i=0}^r$, such that $\gamma_i(\varphi_j(\mathbf{r})) = \delta_{ij}$ with δ_{ij} the Kronecker symbol. These functions are able to identify the coefficients set $\{\alpha_j\}_{j=0}^r$ that fix univocally the polynomial $p = p(\mathbf{r}) = \sum_{j=0}^r \alpha_j \varphi_j(\mathbf{r}) \in \mathbb{P}_r$. These coefficients are called *degrees of freedom*, with the acronym DoFs.

The definition of the degrees of freedom is not univocal: as a matter of fact it depends on the finite element type chosen. Some of them are very well known, intensively used, and called *Lagrange finite elements*. The degrees of freedom are the values at the points where the polynomial $p(\mathbf{r})$ is evaluated. These points belong to the sub-domain K, and are called *nodes*. So, another specified finite element characteristic can be defined:

- nodes: specific points belonging to K that satisfy the relation $\alpha_j = p(\mathbf{r}_j)$ with $j = 0, 1, ...r$.

For general polynomial functions used like basis, there is a critical number of points necessary to define the DoFs. This number is defined by the following formula:

$$\dim_{K \in \mathbb{R}^d} \mathbb{P}_r = \frac{\prod_{i=1}^d (r+i)}{\prod_{i=1}^d i}. \tag{2.25}$$

Definition 1. 1 *The finite element defined by the triple (K,*$\mathbb{P}_{r,K}, \Sigma_K$*) is called unisolvent if the set* $\{\boldsymbol{a}_j\}_{j=0}^{dim} \subset K$*, defines uniquely a function* $p(\boldsymbol{r}) \in \mathbb{P}_r$ *such that* $p(\boldsymbol{a}_j) = \alpha_j$*, with j=0,1,...dim.*

Once the unisolvent property has been established, we can use the DoFs found to define the basis of the space. A general vector space with this basis is denominated $\mathcal{V}_h$.
Another feature which is very important to ensure is defined by the following property:

Property 2. 1 *A sufficient condition for which a function u could belong to* $H^1(V)$ *is that* $u \in C^0(V)$ *and u belongs to* $H^1(K)$ $\forall K \in \mathcal{T}_h$.

2.2.2 Finite Element implementation

A Poisson multidimensional model problem will be briefly introduced in order to obtain the general expressions of the mass and stiffness matrices - the relative of **T** and **S** in (2.14) and (2.22) - of a fundamental element. The same process will be followed for a one-dimensional problem in which, however, the approximating polynomial function will be chosen. These descriptions will be followed by an introduction to some methods for solving integrals formulations. For more information, consult [26], [15], [19].

Multi-dimensional scalar case

Let us consider a model problem, in particular a Poisson multi-dimensional problem, with physical dimension $d = i,$ for $i = 1, 2, 3$, defined in a region $\Omega = (0, 1)^d$ resulting in:

$$\begin{cases} -\nabla^2 u = f(\mathbf{r}) \quad \text{in } \Omega \\ u|_\Gamma = 0 \ , \end{cases} \tag{2.26}$$

with Γ the boundary of Ω, whose weak formulation is:

$$\begin{aligned} &\text{find } u \in X_0^N \text{ such that:} \\ &\int_\Omega (\nabla v)^T (\nabla u) = \int_\Omega v^T f \quad \forall v \in X_0^N. \end{aligned} \tag{2.27}$$

The approximate solution of this problem belongs to the finite-dimensional function space $X_0^N = \{N_j(\mathbf{r}),\ N_j(\mathbf{r})|_\Gamma = \mathbf{0} \text{ for } j = 1, ..., N_h\}$, which specifies the solution behavior on the boundary Γ, in this case the value is zero, the polynomial degree, in this case N, and the shape functions $N_j(\mathbf{r})$.
As it has been described in the first paragraph of this chapter, the solution $u(\mathbf{r})$, the test function $v(\mathbf{r})$, and the forcing function $f(\mathbf{r})$ will be decomposed as linear combinations of basis function as follows:

$$u(\mathbf{r}) = \underline{\mathbf{N}}(\mathbf{r})\underline{u}, \qquad v(\mathbf{r}) = \underline{\mathbf{N}}(\mathbf{r})\underline{v}, \qquad f(\mathbf{r}) = \underline{\mathbf{N}}(\mathbf{r})\underline{f}. \tag{2.28}$$

The problem in the weak formulation, thanks to the generality of the test functions, will take the form of an algebraic linear problem:

$$\underline{\underline{K}}\ \underline{u} = \underline{\underline{M}}\ \underline{f}\ , \tag{2.29}$$

where $\underline{\underline{K}}$ e $\underline{\underline{M}}$ are the stiffness and mass matrices, whose elements are defined

as

$$K_{i,j} = \int_{\Omega} (\nabla N_i(\mathbf{r}))^T (\nabla N_j(\mathbf{r})) d\Omega \quad \text{and} \quad M_{i,j} = \int_{\Omega} N_i(\mathbf{r})^T N_j(\mathbf{r}) d\Omega. \tag{2.30}$$

Preliminarily we will find the relationships, called mappings, between the coordinates of the *elementary physical element* K, which are defined as $\mathbf{r} = [r_1\ r_2\ r_3]^T$ for example, and the coordinates of the *elementary reference element* $\hat{K}$, which are defined as $\hat{\mathbf{r}} = [\hat{r}_1\ \hat{r}_2\ \hat{r}_3]^T$. For further calculations, it is useful to introduce the nabla operator defined in two different reference frames. Thus, we define $\nabla_{\mathbf{r}} = [\partial/\partial r_1\ \partial/\partial r_2\ \partial/\partial r_3]^T$ and $\nabla_{\hat{\mathbf{r}}} = [\partial/\partial \hat{r}_1\ \partial/\partial \hat{r}_2\ \partial/\partial \hat{r}_3]^T$. The relationships are expressed as:

$$\hat{\mathbf{r}}(\mathbf{r}) = \underline{\underline{G}}\ \mathbf{r} + \underline{g}, \tag{2.31}$$

and

$$\nabla_{\mathbf{r}} = \frac{\partial \hat{r}_i}{\partial r_j} \frac{\partial}{\partial \hat{r}_i} = G_{ij} \frac{\partial}{\partial \hat{r}_i} = \underline{\underline{G}}\ \nabla_{\hat{\mathbf{r}}}. \tag{2.32}$$

where $\underline{\underline{G}} \in \mathbb{R}^{d \times d}$ is the rotation and stretching matrix and $\underline{g} \in \mathbb{R}^{d \times 1}$ is the translation vector. The $(d \times d)$ elements constituting $\underline{\underline{G}}$ and the other ones d belong to $\underline{g}$ are functions of the nodes coordinates elements of both K and $\hat{K}$. The previously exposed properties are useful to choose a suitable set of number of nodes, used to evaluate the shape functions. In general, some of these nodes belong necessarily to the boundary, and some other are inside of K and $\hat{K}$. The number of nodes and their locations are dependent on the selection of the polygon composing the discretized domain.
The physical element and the reference element will have the same polygonal form, and the nodes are chosen in the same relative place, as in the example in Figure 2.1.

This is a standard procedure used to find the stiffness and mass matrices in a simple and efficient manner.
Thus, the stiffness and mass elements are:

$$\begin{aligned} \underline{\underline{K}}_{K_j} &= \int_{\Omega_j} (\nabla_{\mathbf{r}} \underline{N}_j(\hat{\mathbf{r}}(\mathbf{r})))^T (\nabla_{\mathbf{r}} \underline{N}_j(\hat{\mathbf{r}}(\mathbf{r}))) \ d\Omega(\mathbf{r}) \\ &= \int_{\Omega_j} (\underline{\underline{G}}\ \nabla_{\hat{\mathbf{r}}} \underline{N}_j(\hat{\mathbf{r}}))^T (\underline{\underline{G}}\ \nabla_{\hat{\mathbf{r}}} \underline{N}_j(\hat{\mathbf{r}})) \ d\Omega(\mathbf{r})) \end{aligned} \tag{2.33}$$

and

$$\underline{\underline{M}}_{K_j} = \int_{\Omega_j} (\underline{N}_j(\hat{\mathbf{r}}(\mathbf{r})))^T (\underline{N}_j(\hat{\mathbf{r}}(\mathbf{r}))) \ d\Omega(\mathbf{r}). \tag{2.34}$$

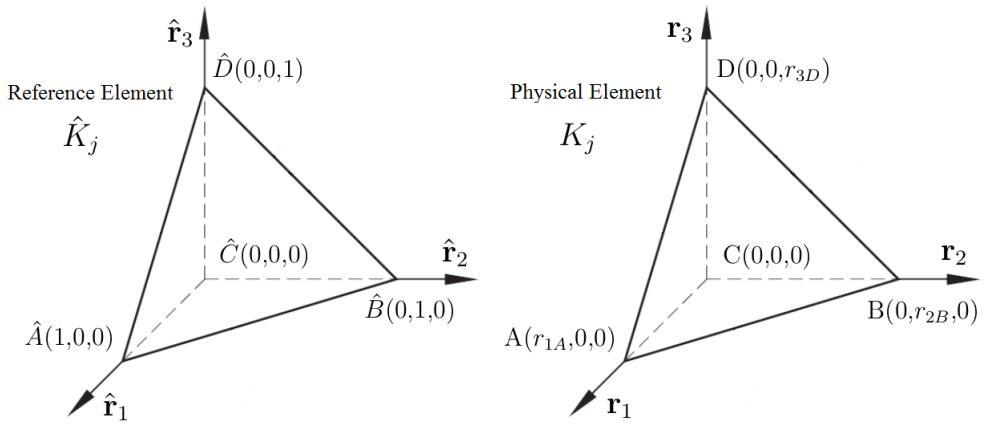

FIGURE 2.1. The reference element $\hat{K}_j$ is represented in the left figure. The dimensions are chosen to simplify the integral calculation; in fact, all the three vertexes are unitary. The physical element K_j is represented on the right side. The dimensions depend on the meshing and so, every single element has different vertex coordinates.

From here, we can draw the relation between the differentials appearing in the integrals. In fact, we have:

$$\begin{cases} d\Omega(\mathbf{r}) = dr_1\ dr_2\ dr_3 \\ d\hat{\Omega}(\hat{\mathbf{r}}) = d\hat{r}_1\ d\hat{r}_2\ d\hat{r}_3 \\ \dfrac{d\hat{r}_1\ d\hat{r}_2\ d\hat{r}_3}{dr_1\ dr_2\ dr_3} = |\underline{\underline{J}}(\hat{\mathbf{r}};\mathbf{r})| \end{cases} \quad \Rightarrow \quad d\Omega(\mathbf{r}) = |\underline{\underline{J}}(\hat{\mathbf{r}};\mathbf{r})|^{-1} d\hat{\Omega}(\hat{\mathbf{r}}). \tag{2.35}$$

At this point, the integrals depending on the reference variables only can be rewritten as follows:

$$\underline{\underline{K}}_{\hat{K}_j} = \int_{\hat{\Omega}_j} (\underline{\underline{G}}\ \nabla_{\hat{\mathbf{r}}} \underline{N}_j(\hat{\mathbf{r}}))^T (\underline{\underline{G}}\ \nabla_{\hat{\mathbf{r}}} \underline{N}_j(\hat{\mathbf{r}}))\ |\underline{\underline{J}}(\hat{\mathbf{r}};\mathbf{r})|^{-1} d\hat{\Omega}(\hat{\mathbf{r}}) \tag{2.36}$$

and

$$\underline{\underline{M}}_{\hat{K}_j} = \int_{\hat{\Omega}_j} (\underline{N}_j(\hat{\mathbf{r}}))^T (\underline{N}_j(\hat{\mathbf{r}}))\ |\underline{\underline{J}}(\hat{\mathbf{r}};\mathbf{r})|^{-1} d\hat{\Omega}(\hat{\mathbf{r}}), \tag{2.37}$$

where $\hat{\Omega}$ is the domain of the elementary reference element.
The versatility of this method is found in the simplicity of the calculation of these integrals: in fact, once the reference integral is computed, the only quantities depending on the position of Ω_j are $\underline{\underline{G}}$ and $\underline{\underline{J}}$. The last quantity, physically, is the volume, the area or the length of element Ω_j according to the dimension d of the problem.

One-dimensional linear shape function scalar case

Let us consider a model problem, in particular the one-dimensional Poisson problem defined in region L = (0, 1):

$$\begin{cases} -\dfrac{d^2u}{dx^2} = f(x) \quad \text{in } L \\ u(0) = u(1) = 0 \ , \end{cases} \tag{2.38}$$

whose weak form is: find u $\in X_0^N$ such that

$$\int_0^1 \frac{dv}{dx}\frac{du}{dx}\,dx = \int_0^1 v f\,dx \quad \forall v \in \mathcal{X}_0^N. \tag{2.39}$$

With the same procedure as in the multi-dimensional case, the problem in weak form is transformed, through the arbitrary test functions, in an algebraic problem of the form:

$$\underline{\underline{K}}\ \underline{u} = \underline{\underline{M}}\ \underline{f} \ , \tag{2.40}$$

whose corresponding coefficients are given by:

$$K_{i,j} = \int_0^1 N_i(x)' N_j(x)'\,dx \quad \text{and} \quad M_{i,j} = \int_0^1 N_i(x) N_j(x)\,dx. \tag{2.41}$$

To solve the algebraic problem two more steps have to be taken in order to decide which shape function should be adopted and to compute the integrals that will define the coefficients of the two matrices.

A 1D problem and his relative spatial discretization are considered. The integration domain is the interval (a,b). A partition $\mathcal{T}_h$, depending on h parameter such that $a = x_0 < x_1 < x_2 < ... < x_N < x_{N+1} = b$ with $h_{j+1|j} = x_{j+1} - x_j$ is introduced. The strategy is to consider only one $\hat{K}$ element, to construct the shape function on it, and then to specialize it on every single K_j with a mapping ϕ_j.
The definition of space X_h^1 requires the functions $\hat{\varphi}(\xi)$ to be continuous within the domain of integration L, but also in each single K_j, and consequently in $\hat{K}$. To fulfil this requirement, it is necessary for some of the points useful for the construction of the shape function, to be located on the boundary of $\hat{K}$. When applying (2.25) it becomes evident that two points are needed for a *linear* function of the type $\hat{\varphi}(\xi) = a_1\xi + a_0$, both placed at the ends of $\hat{K}$. Let

us take $\hat{a} = 0$ and $\hat{b} = 1$ as points, and let us impose that $\hat{\varphi}(\xi = \hat{a}) = 1$ and $\hat{\varphi}(\xi = \hat{b}) = 0$. A 2x2 linear system is to be solved in order to find coefficients $[a_0, a_1]$, and the first shape function. The same operation is repeated, but imposing $\hat{\varphi}(\xi = \hat{a}) = 0$ and $\hat{\varphi}(\xi = \hat{b}) = 1$. For such choices there will be another 2x2 linear system to be solved to find coefficients $[a_0, a_1]$ of the second shape function. This operation allows us to find $\hat{\varphi}_a(\xi)$ and $\hat{\varphi}_b(\xi)$, where the subscripts indicate the point where the unit value was imposed to the shape function:

$$\hat{\varphi}_a(\xi) = 1 - \xi \quad \text{and} \quad \hat{\varphi}_b(\xi) = \xi. \tag{2.42}$$

The same operation is repeated in the physical space, in a generic interval $[x_{i+1}, x_i]$:

$$\varphi_{x_i}(x) = \frac{x - x_i}{x_{i+1} - x_i} \quad \text{and} \quad \varphi_{x_{i+1}}(x) = -\frac{x - x_{i+1}}{x_{i+1} - x_i}. \tag{2.43}$$

Now, having found $\hat{\varphi}_a(\xi)$, $\hat{\varphi}_b(\xi)$ and $\varphi_{x_i}(x)$, $\varphi_{x_{i+1}}(x)$, equality between $\hat{\varphi}_a(\xi)$ and $\varphi_{x_i}(x)$ will be imposed, in order to find the relationship between ξ and x, i.e, the relationship allowing to map $\hat{K}_j$ to K_j, which is:

$$\xi = \frac{x - x_i}{x_{i+1} - x_i}. \tag{2.44}$$

To find the $\underline{\underline{K}}$ matrix and the $\underline{\underline{M}}$ matrix of the K_j domain, the spatial discretization of the quantities on this domain will be imposed, resulting in:

$$\begin{aligned} u_{i+1|i} &= \underline{N}_{i+1|i}(x)\underline{u} \\ &= [\varphi_{x_i}((x)) \quad \varphi_{x_{i+1}}((x))] \begin{bmatrix} u_i \\ u_{i+1} \end{bmatrix} \\ &= [\hat{\varphi}_a(\xi(x)) \quad \hat{\varphi}_b(\xi(x))] \begin{bmatrix} u_i \\ u_{i+1} \end{bmatrix} \\ &= [1 - \xi \quad \xi] \begin{bmatrix} u_i \\ u_{i+1} \end{bmatrix} \end{aligned} \tag{2.45}$$

and

$$
\begin{aligned}
\underline{\underline{K}}_{K_{i+1|i}} &= \int_{x_i}^{x_{i+1}} \underline{N}'^{T}_{i+1|i}(\xi(x))\ \underline{N}'_{i+1|i}(\xi(x))\, dx \\
&= \int_{x_i}^{x_{i+1}} \frac{d\xi}{dx} \frac{\underline{N}^{T}_{i+1|i}(\xi(x))}{d\xi} \frac{\underline{N}_{i+1|i}(\xi(x))}{d\xi} \frac{d\xi}{dx}, dx \\
&= \int_{x_i}^{x_{i+1}} \begin{bmatrix} 1-\xi \\ \xi \end{bmatrix}' [1-\xi \quad \xi]' \left(\frac{d\xi}{dx}\right)^2 dx \\
&= \int_{x_i}^{x_{i+1}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} [-1 \quad 1] \left(\frac{d\xi}{dx}\right)^2 dx\ ,
\end{aligned}
\tag{2.46}
$$

$$
\begin{aligned}
\underline{\underline{M}}_{K_{i+1|i}} &= \int_{x_i}^{x_{i+1}} \underline{N}^{T}_{i+1|i}(x) \underline{N}_{i+1|i}(x)\, dx \\
&= \int_{x_i}^{x_{i+1}} \begin{bmatrix} 1-\xi \\ \xi \end{bmatrix} [1-\xi \quad \xi]\, dx.
\end{aligned}
\tag{2.47}
$$

At this point the integration domain is brought into the reference domain coordinate, then the integration extremes and the integration variable will change:

$$
\begin{cases}
x_i = 0 \\
x_{i+1} = 1 \\
dx = (x_{i+1} - x_i) d\xi \\
\dfrac{d\xi}{dx} = \dfrac{1}{(x_{i+1} - x_i)}
\end{cases}
\tag{2.48}
$$

$$\Downarrow$$

$$
\underline{\underline{K}}_{K_{i+1|i}} = \int_0^1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} \frac{1}{(x_{i+1} - x_i)} [-1 \quad 1]\, d\xi \tag{2.49}
$$

and

$$
\underline{\underline{M}}_{K_{i+1|i}} = \int_0^1 \begin{bmatrix} 1-\xi \\ \xi \end{bmatrix} (x_{i+1} - x_i) [1-\xi \quad \xi]\, d\xi. \tag{2.50}
$$

Numerical integral solution: quadrature methods

The Lagrangian approximation representation allows us to approximate any function $f(x)$ once given a set of points belonging to it.

To find the coefficients of the resulting linear system, the resolution of integrals extended to a reference interval $\hat{K}_j$ is required. Generally, to compute an approximate integral over an interval $\hat{L}$, such integral is expressed using the following formula:

$$\int_{\hat{L}} f(x)\,dx = \int_{\hat{L}} \prod_N f(x)\,dx \quad = \sum_{i=0}^{N} w_i f(x_i), \tag{2.51}$$

where $\prod_N f(x)$ is the N-degree interpolating polynomial of the function $f(x)$, $\{w_i\}$ are real numbers called *weights* and $\{x_i\}$ are points called *nodes*.
The utility of this approximation is based on the choice of points, which are points that allow to use the quadrature according to the Gauss-Legendre formula (GL) or the Gauss-Legendre-Lobatto one (GLL). Moreover, this yields an excellent conditioning of the linear problem obtained [26].
An interval $\hat{L}$ = (-1,1) is considered. The above mentioned type of points is distinguished in the following way:

- Gauss-Legendre nodes: an interpolating Lagrange polynomial of degree N is given. These points are the zeros of the polynomial. These nodes are all within the interval $\hat{L}$ and therefore can not guarantee the continuity of the solution between two different intervals. If the shape function has degree M, then this type of nodes will allow the exact computation of the integrals that contain, in their expression, shape functions with degree *M<2N+1* ([26], [15], [19]).

- Gauss-Legendre-Lobatto nodes: an interpolating Lagrange polynomial of degree N is given. These points will be the end points $\hat{L}$ and the points where the first derivative of the polynomial is equal to zero. This choice implies that N> 0. These nodes will be both on the border of $\hat{L}$ and on inside, thus they will guarantee the continuity of the solution between two different intervals. If the shape function has degree M, then this type of nodes will allow the exact computation of the integrals that contain, in their expression, shape functions with degree *M<2N-1* ([26], [15], [19]).

Now that all the arrays relating to each individual domain $\hat{K}_j$ are calculated, it will be necessary to assemble the latter ones in order to build the global matrices of the problem. This operation is performed in a relatively simple manner, starting from the matrices of element $\hat{K}_j$, then adding up the items in i, j place when the same node - so the same unknown - is part of the domain K_j. To complete the matrix, it is required an expansion of the matrix until reaching the size $N_{tot} \times N_{tot}$ and completing the empty spaces with

zeros. The fundamental operation in this technique is to remove the rows and the columns belonging to the DoFs on the Dirichlet boundary. For example, given the 3×3 stiffness matrix for the first two physical elements whit two points in common

$$\underline{\underline{K}}_{K^{(1)}} = \begin{bmatrix} K_{11}^{(1)} & K_{12}^{(1)} & 0 \\ K_{21}^{(1)} & K_{22}^{(1)} & 0 \\ 0 & 0 & K_{33}^{(1)} \end{bmatrix} \quad \text{and} \quad \underline{\underline{K}}_{K^{(2)}} = \begin{bmatrix} K_{11}^{(2)} & K_{12}^{(2)} & 0 \\ K_{21}^{(2)} & K_{22}^{(2)} & 0 \\ 0 & 0 & K_{33}^{(2)} \end{bmatrix}, \tag{2.52}$$

the global stiffness matrix will be:

$$\underline{\underline{K}} = \begin{bmatrix} K_{11}^{(1)} & K_{12}^{(1)} & 0 & 0 & 0 & \cdots & 0 \\ K_{21}^{(1)} & K_{22}^{(1)} + K_{11}^{(2)} & K_{12}^{(2)} & 0 & 0 & \cdots & 0 \\ 0 & K_{21}^{(2)} & K_{33}^{(1)} + K_{22}^{(2)} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & K_{33}^{(2)} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{2.53}$$

This process is continued until all the physical stiffness matrices are listed and assembled inside the global one. The final result can be, after the incorporation of boundary conditions, as shown in Figure 2.2.

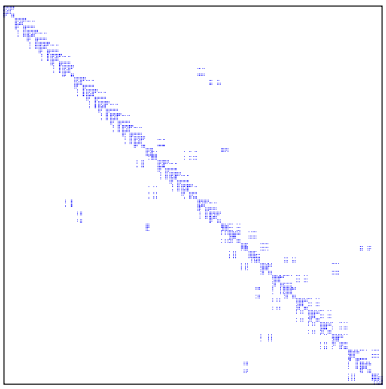

FIGURE 2.2. Example of a sparsity pattern stiffness matrix. The blue points indicate the non-null values. We can note the sparsity of the matrix built with FEM.

2.3 Solution of the algebraic system

In many fields of applied science, including the engineering field, the final result reached from a finite element method approach, is obtained by solving an algebraic linear or linearized system. In fact, defining a matrix $A \in \mathbb{C}^{N_h \times N_h}$, a vector of known terms $b \in \mathbb{C}^{N_h \times 1}$ and a vector of unknowns $x \in \mathbb{C}^{N_h \times 1}$, the finite element method provides the following system:

$$Ax = b. \tag{2.54}$$

In a real problem in engineering, the number of degrees of freedom, that is the unknowns, can be on the order of hundred thousands, if not a million. For this reason, it is not possible to solve system (2.54) with *Cramer's rule*, for this would result in the waste of hours, maybe years (if not centuries), of calculations.
For this reason, in order to solve (2.54), other strategies are needed, that are, not only more effective, but also, and most importantly, more rapid. The two most commonly used classes of methods for this purpose will be described below in their fundamental principles, i.e. *direct methods*, based on *LU factorization*, and the *iterative methods*, in particular the *Conjugate Gradient Method*. Further details are contained in [27].

2.3.1 Direct methods: LU decomposition

The LU factorization of a square matrix $A \in \mathbb{C}^{N_h \times N_h}$ consists in finding two matrices, respectively a lower triangular matrix $L \in \mathbb{C}^{N_h \times N_h}$, and an upper triangular matrix $U \in \mathbb{C}^{N_h \times N_h}$, whose multiplication LU is PA, where the matrix $P \in \mathbb{R}^{N_h \times N_h}$ is called *pivoting matrix*, and whose purpose will be explained afterwards. In this way, to solve (2.54) leads to the computation of two triangular systems

$$\begin{cases} Ly = Pb \\ Ux = y. \end{cases} \tag{2.55}$$

For a matrix $N_h \times N_h$, the procedure below can be followed:

- the elements of the two matrices L and U satisfy the following non-linear system:

$$\sum_{r=1}^{\min(i,j)} l_{ir} u_{rj} = a_{i,j} \ , \ i,j = 1,2,...,N_h; \tag{2.56}$$

- Since the system (2.56) is under-determined, having ${N_h}^2 + N_h$ unknowns but N_h^2 equations, let us impose that the elements on the diagonal of L are equal to 1, i.e.,

$$l_{i,i} = 1 \ , \ i = 1, 2, ..., N_h, \tag{2.57}$$

  so as to reduce the unknowns to N_h^2;

- let us calculate the remaining terms - the ones out of the diagonal - with the following procedure:

$$\begin{aligned}
&\text{for } k = 1, ..., N_h - 1, \\
&\quad \text{for } i = k + 1, ..., N_h, \\
&\qquad \text{find } \tilde{r} \text{ such that } |a^{(k)}_{\tilde{r},k}| = \max_{r=k,...,n} |a^{(k)}_{r,k}|, \\
&\qquad \text{interchange the k row with the } \tilde{r} \text{ row}, \\
&\qquad l_{ik} = \frac{a^{(k)}_{ik}}{a^{(k)}_{kk}}, \\
&\qquad \text{for } j = k + 1, .., N_h, \\
&\qquad\quad a^{k+1}_{ij} = a^{(k)}_{ij} - l_{ik} a^{(k)}_{kj}.
\end{aligned} \tag{2.58}$$

  This sequence shows the usefulness of matrix P as, whenever the element of maximum absolute value is found, the matrix interchanges the two involved rows automatically. So P has the task to *permute* the rows;

- now that every term of L has been calculated, they can be inserted in (2.56) to find every term in U too.

In the case of the first system $Ly = Pb$, the solution will be found by following the algorithm of the *forward substitution*:

$$\begin{aligned}
y_1 &= \frac{1}{l_{11}} \sum_{j=1}^{N_h} P_{1j} bj, \\
y_i &= \frac{1}{l_{ii}} \left(\sum_{j=1}^{i-1} P_{ij} b_j - \sum_{j=1}^{i-1} l_{ij} y_j \right), \ i = 2, ..., N_h;
\end{aligned} \tag{2.59}$$

and in a similar manner, the solution of the second system $Ux = y$ will be

found by following the algorithm of the *backward substitution*:

$$\begin{aligned} x_{N_h} &= \frac{1}{u_{N_h N_h}} y_{N_h}, \\ x_i &= \frac{1}{u_{ii}} \left(y_i - \sum_{j=i+1}^{N_h} l_{ij} y_j \right), \ i = N_h - 1, ..., 1. \end{aligned} \tag{2.60}$$

The cascade resolution of the two systems allows one to reduce the computational time needed to solve the original system. This is the fundamental property that makes this strategy more feasible rather than the resolution by Cramer's rule.

2.3.2 Iterative methods: Conjugate Gradient method

An iterative method for solving (2.54) type is, starting from an initial guess vector $x^{(0)} \in \mathbb{R}^{N_h \times 1}$, entails the construction of a sequence of vectors $\{x^{(k)} \in \mathbb{R}^{N_h \times 1}, k \geq 0\}$. These vectors are required to converge as the number of iterations k grows, to the solution vector x^*, i.e.,

$$\lim_{k \to \infty} x^{(k)} = x^*. \tag{2.61}$$

In practice, an iterative method will never reach the exact solution, but at infinity. To stop the sequence, an a priori defined tolerance will be needed. Obviously, the solution at the k-th step will be close to the real one as long as the tolerance will be small.
A general technique to build an iterative method consists in the *splitting* of the system matrix A. This is $A = P - (P - A)$, with $P \in \mathbb{R}^{N_h \times N_h}$ called *preconditioning matrix*, whence:

$$Px = (P - A)x + b. \tag{2.62}$$

When assuming that the solution at step $k+1$ is $x^{(k+1)}$ and the one at step k is $x^{(k)}$, then (2.62) will be:

$$\begin{aligned} &Px^{(k+1)} = (P - A)x^{(k)} + b, \\ &\text{from which} \\ &x^{(k+1)} = x^{(k)} + \alpha^{(k)} P^{-1}(b - Ax^{(k)}), \end{aligned} \tag{2.63}$$

where $b - Ax^{(k)} = r^{(k)}$, with $r^{(k)} \in \mathbb{R}^{N_h \times 1}$ representing the *residual* at k-th

iteration.
Now that the basic parts of the problem have been described, the method of preconditioned conjugate gradient will be explained.
Let us define the quadratic functional

$$Q(x) = \frac{1}{2}x^T Ax - x^T b. \tag{2.64}$$

If we want to calculate the stationary point, the gradient will have to be equal to zero, thus obtaining:

$$\begin{cases} \nabla Q(x) = Ax - b \\ \nabla Q(x) = 0 \end{cases} \quad \Rightarrow \quad Ax = b. \tag{2.65}$$

The functional $Q(x)$ represents a convex paraboloid, whose minimum point coincides with the solution of (2.54). Geometrically speaking, an iterative method is none other than the descent towards the minimum point of the paraboloid, starting from any point having $x^{(0)}$ as coordinate. The conjugate gradient method involves the following definitions:

- two vectors in the set $\{\underline{d}_i\}_1^k$, with $\underline{d} \in \mathbb{R}^{N_h \times k}$, are called A - conjugated if they satisfy $\underline{d}_i^T A \underline{d}_j = 0 \ \forall \ i, j = 1, ..., k$;
- if two vectors are A - conjugated, then these two vectors are linearly independent;
- if two vectors are linearly independent, then these two vectors can span a space basis, that every vector can be expressed as $\underline{x} = \underline{\underline{d}} \, \underline{\alpha}$, with $\underline{\alpha} \in \mathbb{R}^{k \times 1}$.

At this point, we can express the quadratic functional as a function of the new basis and minimum can be compared with respect to α^T:

$$Q(\alpha) = \frac{1}{2}\alpha^T d^T A d\alpha - \alpha^T d^T b; \tag{2.66}$$

$$\begin{cases} \nabla Q(\alpha) = d^T A d\alpha - d^T b \\ \nabla Q(\alpha) = 0 \end{cases} \quad \Rightarrow \alpha = \frac{d^T b}{d^T A d}. \tag{2.67}$$

As we are in the context of iterative methods, at the k-th step the exact solution will not be known. So, should we insert it in (2.54), the result would

not be a zero vector, but a residual $\underline{r}$. Projecting (2.54) in space with $\underline{d}$ basis, the results is:

$$d^T Ax - d^T b = -d^T r; \tag{2.68}$$

Considering that x is a linear combination of d, the first term of (2.68) will be equal to zero because of the A - conjugation, so we will have $d^T b = d^T r$. Substituting the latter one in the expression of α, yields

$$\alpha = \frac{d^T r}{d^T Ad}. \tag{2.69}$$

To obtain the desired $\alpha^{(k)}$, let us consider the relative k-th vectors, that is

$$\alpha^{(k)} = \frac{{d^{(k)}}^T r^{(k)}}{{d^{(k)}}^T Ad^{(k)}}. \tag{2.70}$$

To force the method to follow the directions dictated by d, let us use the *Gram-Schmidt method*, which, starting from a given vector, builds its A - conjugate vector. In this way, at each iteration, we will find the A-conjugate directions on which to make the descent proceed.
The method is as follows:

- let us build the desired vector through the sum of two vectors, so $d^{(k+1)} = \tilde{d}^{(k+1)} - \beta^{(k)} d^{(k)}$. Then, the vector will be pre-multiplied by an A - conjugate vector, so ${d^{(k)}}^T A d^{(k+1)} = {d^{(k)}}^T A \tilde{d}^{(k+1)} - \beta^{(k)} {d^{(k)}}^T A d^{(k)}$. The left-side term is equal to zero, because A - conjugation, while putting $\tilde{d}^{(k)} = r^{(k+1)}$ to impose the direction along the residual between the solution $x^{(k)}$ and $x^{(k+1)}$. This results in:

$$\beta^{(k)} = \frac{{d^{(k)}}^T A r^{(k+1)}}{{d^{T(k)}} A d^{(k)}}. \tag{2.71}$$

Now that we have defined everything we needed, the method can be implemented using the following algorithm:

$$\begin{aligned} &\text{inizialization: for } k = 0 \quad x^{(0)} = \text{any vector}; \\ &\qquad\qquad\qquad\qquad\quad\; r^{(0)} = b - Ax^{(0)}; \\ &\qquad\qquad\qquad\qquad\quad\; d^{(0)} = r^{(0)}; \end{aligned} \tag{2.72}$$

$$\text{iteratively, for } k>0: \begin{cases} \alpha^{(k)} = \dfrac{{d^{(k)}}^T r^{(k)}}{{d(k)}^T A d^{(k)}}; \\ x^{(k+1)} = x^{(k)} + \alpha^{(k)} P^{-1} d^{(k)}; \\ r^{(k+1)} = b - A x^{(k+1)}; \\ \beta^{(k)} = \dfrac{{d^{(k)}}^T A r^{(k+1)}}{{d^{T(k)}} A d^{(k)}}; \\ d^{(k+1)} = r^{(k+1)} - \beta^{(k)} d^{(k)}. \end{cases} \tag{2.73}$$

For a generic matrix A, it is not possible to establish a preconditioning matrix that ensures optimum balance between convergence and computational effort to calculate P^{-1}. Therefore, diagonal, triangular or tridiagonal matrices will be the most appropriate for this purpose.

3 | Reduced Order Methods for Maxwell's equations

3.1 Introduction

With target applications characterized by computationally intensive parameterized problems that require repeated evaluation, it is clear that we need to seek alternatives to simply solving the full problem many times. This is exactly the place where reduced methods can be effective. When introducing reduced models it is unavoidable to familiarize with the notion of a parametric solution manifold. A parametric solution manifold is the set of all solutions to the parameterized problem, under variation of the parameter. The final goal of a Reduced Order Method (ROM) is to approximate any member of this solution manifold with a low number of basis function - N_{rb} - properly selected and computed. This set of basis functions is denoted as the reduced basis. This methodology consists of two different computational steps:

- offline stage: during the possible very expensive offline stage, the solution manifold is empirically explored to construct a reduced basis that approximates any member of the solution manifold to within a prescribed accuracy. As this involves the solution of at least M approximated problems, each with N_h degrees of freedom, the cost can be very high. This results in the identification of a linear N_{rb} - dimensional reduced basis;

- online stage: the online stage consists of a Galerkin projection, using the parameterized bilinear form $a(\cdot, \cdot; \mu)$ with a varying parameter value μ, onto the space spanned by the reduced basis. During this stage, the parameter space can be explored at a substantially quatly reduced cost, ideally at a cost independent of N_h [14].

3.2 Parameterized weak formulation for Maxwell's equations

This chapter will briefly introduce the concept of parameterization and its mathematical formulation.
A parameterized problem expressed in terms of partial derivatives, consists of an equation set describing a particular mathematical model. The solution depends on a set of parameters, which have a direct influence:

- on the problem geometry (the set of parameters changes the geometry of the integration domain);
- on the problem data (then initial conditions and boundary conditions, external forces terms and, in general, the quantities that change coefficients of the problem equations).

We construct the space in which the vector-valued field variable shall reside as the Cartesian product $\mathbb{V} = \mathbb{V}_1 \times \ldots \mathbb{V}_{d_v}$ where d_v denotes the dimension of the field variable; a typical element of $\mathbb{V}$ is denoted $v = (v_1, \ldots, v_{d_v})$. We equip $\mathbb{V}$ with an inner product $(w, v)_{\mathbb{V}}, \forall w, v \in \mathbb{V}$, and the induced norm $||v||_{\mathbb{V}} = (v, v)_{\mathbb{V}}, \forall v \in \mathbb{V}$.
We are given parameterized linear forms $f : \mathbb{V} \times \mathbb{P} \to \mathbb{R}$, where $\mathbb{P}$ is the parameter space and a parameterized bilinear form $a : \mathbb{V} \times \mathbb{V} \times \mathbb{P} \to \mathbb{R}$ where the bi-linearity is with respect to the first two variables. The abstract formulation reads: given $\mu \in \mathbb{P}$, we seek $\tau(\mu) \in \mathbb{V}$ such that

$$a(\tau(\mu), v; \mu) = f(v; \mu), \quad \forall v \in \mathbb{V}. \tag{3.1}$$

3.2.1 Well-posedness of the parametric weak formulation

The well-posedness of the abstract problem formulation (3.1) can be established by the Lax-Milgram theorem [26]. In order to state a well-posed problem for all parameter values $\mu \in \mathbb{P}$, in addition to the bi-linearity and the linearity of the parameterized forms $a(\cdot, \cdot; \mu)$ and $f(\cdot; \mu)$, respectively, it is assumed that:

- $a(\cdot, \cdot; \mu)$ is coercive and continuous for all $\mu \in \mathbb{P}$ with respect to the norm $||\cdot||_{\mathbb{V}}$, i.e., for every $\mu \in \mathbb{P}$, there exists a positive constant $\alpha(\mu) \geq \alpha > 0$ and a finite constant $\gamma(\mu) \leq \gamma < \infty$ such that

$$a(v, v; \mu) \geq \alpha(\mu)||v||^2_{\mathbb{V}} \quad \text{and} \quad a(w, v; \mu) \leq \gamma(\mu)||w||_{\mathbb{V}}||v||_{\mathbb{V}}, \forall w, v \in \mathbb{V};$$

(3.2)

- $f(\cdot;\mu)$ is continuous for all $\mu \in \mathbb{P}$ with respect to the norm $||\cdot||_{\mathbb{V}}$, i.e., for every $\mu \in \mathbb{P}$, there exists a constant $\delta(\mu) \leq \delta < \infty$ such that:

$$f(v;\mu) \leq \delta(\mu)||v||_{\mathbb{V}}, \forall v \in \mathbb{V}. \tag{3.3}$$

The coercivity and continuity constants of $a(\cdot,\cdot;\mu)$ over $\mathbb{V}$ are, respectively, defined as:

$$\alpha(\mu) = \inf_{v\in\mathbb{V}} \frac{a(v,v;\mu)}{||v||^2_{\mathbb{V}}} \quad \text{and} \quad \gamma(\mu) = \sup_{w\in\mathbb{V}} \sup_{v\in\mathbb{V}} \frac{a(w,v;\mu)}{||w||_{\mathbb{V}}||v||_{\mathbb{V}}}, \forall \mu \in \mathbb{V}. \tag{3.4}$$

3.2.2 Discretization techniques

This section supplies an abstract framework of the discrete approximations of the parametric weak formulation (3.1) for conforming approximations, i.e., there is a discrete approximation space $\mathbb{V}_h$ in which the approximate solution is sought. This is a subset of $\mathbb{V}$, i.e., $\mathbb{V}_h \subset \mathbb{V}$.
We denote the dimension of the discrete space $\mathbb{V}_h$ by $N_h = dim(\mathbb{V}_h)$ and equip $\mathbb{V}_h$ with a basis $\{N_i\}_{i=1}^{N_h}$. For each $\mu \in \mathbb{P}$, the discrete problem consists of finding $\tau_h(\mu) \in \mathbb{V}_h$ such that:

$$a(\tau_h(\mu), v_h;\mu) = f(v_h;\mu), \forall v_h \in \mathbb{V}_h. \tag{3.5}$$

The stiffness matrix and the right-hand side of the approximate problem are denoted by $K_h(\mu) \in \mathbb{C}^{N_h\times N_h}$ and $f_h(\mu) \in \mathbb{C}^{N_h}$, respectively. Further, we denote by $M_h \in \mathbb{R}^{N_h\times N_h}$ the matrix associated with the inner product $(\cdot,\cdot)_{\mathbb{V}}$ of $\mathbb{V}_h$, defined as:

$$\begin{aligned} M_{h_{i,j}} &= (N_j, N_i)_{\mathbb{V}} \ , \\ K_{h_{i,j}}(\mu) &= a(N_j, N_i;\mu)_{\mathbb{V}} \ , \\ f_{h_i}(\mu) &= f(N_i;\mu) \ , \end{aligned} \tag{3.6}$$

for all $1 \leq i,j \leq N_h$. We recall that $\{N_i\}_{i=1}^{N_h}$ is a basis $\mathbb{V}_h$. Then, the approximate problem reads: for each $\mu \in \mathbb{P}$, find $\tau_h(\mu) \in \mathbb{C}^{N_h}$ such that

$$K_h(\mu)\tau_h(\mu) = f_h(\mu). \tag{3.7}$$

The size of the unknown vector is N_h.
Due to the coercivity and continuity of the bilinear form, and the conformity of the approximation space, the *Galerkin orthogonality* [26] is ensured:

$$a(\tau(\mu) - \tau_h(\mu), v_h; \mu) = 0 \quad \forall v_h \in \mathbb{V}_h. \tag{3.8}$$

3.3 The solution manifold and the reduced basis approximation

We first introduce the solution of the parametric exact problem given as: find $\tau(\mu) \in \mathbb{V}$ such that

$$a(\tau(\mu), v; \mu) = f(v; \mu), \quad \forall v \in \mathbb{V}. \tag{3.9}$$

This is referred to as the exact solution.
Let us introduce the notion of solution manifold comprising all solutions of the parametric problem under variation of the parameters, as [14]:

$$\mathcal{M} = \{\tau(\mu) \mid \mu \in \mathbb{P}\} \subset \mathbb{V}, \tag{3.10}$$

where each $\tau(\mu) \in \mathbb{V}$ corresponds to the solution of the exact problem.
In the cases of interest, the exact solution is not available in an analytic or any other simple manner, and an approximate solution is found by approximating $\tau_h(\mu) \in \mathbb{V}_h$ such that

$$a(\tau_h(\mu), v_h; \mu) = f(v_h; \mu), \quad \forall v_h \in \mathbb{V}_h. \tag{3.11}$$

We define this solution as truth solution.
Following the definition for the continuous problem, the discrete version of the solution manifold is defined as ([30], [14]):

$$\mathcal{M}_h = \{\tau_h(\mu) \mid \mu \in \mathbb{P}\} \subset \mathbb{V}_h, \tag{3.12}$$

where each $\tau_h(\mu) \in \mathbb{V}_h$ corresponds to the solution of the parametric approximated problem (3.5).
Throughout the subsequent discussion we assume that $||\tau(\mu) - \tau_h(\mu)||_{\mathbb{V}}$ can be made arbitrarily small for any given parameter value, $\mu \in \mathbb{P}$. We assume

that a computational model is available to solve the truth problem, in order to approximate the exact solution at any required accuracy. This accuracy requirement also implies that the computational cost of evaluating the exact model may be very high, and depend directly on $N_h = \dim(\mathbb{V}_h)$.

A central assumption in the development of any reduced order method is that the solution manifold is characterized by a low dimension. This would assume that the span of a low number of appropriately chosen basis functions represents the solution manifold with a very small error. We shall call these basis functions the *reduced basis* and it will allow the representation of the approximate solution, $\tau_h(\mu)$ based on an N_h-dimensional subspace $\mathbb{V}_{rb}$ of $\mathbb{V}_h$. Let us initially assume that an N_{rb}-dimensional reduced basis, denoted as $\{\xi\}_{=1}^{N_{rb}} \subset \mathbb{V}_h$, is available, then, the associated reduced basis space is given by:

$$\mathbb{V}_{rb} = \text{span}\{\xi_1, ..., \xi_{N_{rb}}\} \subset \mathbb{V}_h. \tag{3.13}$$

The assumption of the low dimensionality of the solution manifold implies that $N_{rb} \ll N_h$. Given the N_{rb}-dimensional reduced basis space $\mathbb{V}_{rb}$, the reduced basis approximation is sought as: for any given $\mu \in \mathbb{P}$, find $\tau_{rb}(\mu) \in \mathbb{V}_{rb}$ such that

$$a(\tau_{rb}(\mu), v_{rb}; \mu) = f(v_{rb}; \mu), \quad \forall v_{rb} \in \mathbb{V}_{rb}. \tag{3.14}$$

3.4 Reduced basis space generation

A discrete and finite-dimensional point-set $\mathbb{P}_h \subset \mathbb{P}$ in parameter domain is introduced and, for example, it can consist of a regular lattice or a randomly generated point-set intersecting with $\mathbb{P}$. The following set can be introduced

$$\mathcal{M}_h(\mathbb{P}_h) = \tau_h(\mu) \quad | \quad \mu \in \mathbb{P}_h, \tag{3.15}$$

of cardinality $M = |\mathbb{P}_h|$. It holds that $\mathcal{M}_h(\mathbb{P}_h) \subset \mathcal{M}_h$ as $\mathbb{P}_h \subset \mathbb{P}$ but if $\mathbb{P}_h$ is fine enough, $\mathcal{M}_h(\mathbb{P}_h)$ is also a good representation of $\mathcal{M}_h$ [14].

3.4.1 Proper Orthogonal Decomposition (POD)

While there are several strategies for generating reduced basis spaces (greedy, Monte Carlo, etc), we shall focus on the Proper Orthogonal Decomposition

(POD) construction in the following [6]. Proper Orthogonal Decomposition (POD) is an explore-and-compress strategy in which the parameter space is sampled, the corresponding truth solutions is computed at all sample points, and, following compression, only the essential information is retained [14]. The N_{rb}-dimensional POD-space is the space that minimizes the functional:

$$J = \sqrt{\frac{1}{M}\sum_{\mu\in\mathbb{P}} \inf_{v_{rb}\in\mathbb{V}_{rb}} ||\tau_h(\mu) - v_{rb}||^2_{\mathbb{V}}}, \tag{3.16}$$

over all N_{rb}-dimensional subspaces $\mathbb{V}_{rb}$ of the span $\mathbb{V}_{\mathcal{M}} = \text{span}\{s_h(\mu) \mid \mu \in \mathbb{P}_h\}$ of the elements of $\mathcal{M}_h(\mathbb{P}_h)$. We introduce an ordering $\mu_1, ..., \mu_M$ of the parameters values in $\mathbb{P}_h$, hence inducing an ordering $\tau_h(\mu_1), ..., \tau_h(\mu_M)$ of the elements of $\mathcal{M}_h(\mathbb{P}_h)$. The m-th solution is denoted as $\tau_h(\mu_m) = \tau_{h_m}$. To construct the POD-space, let us define the *correlation* marix $C \in \mathbb{C}^{M\times M}$ defined by:

$$\begin{aligned} C_{i,j} &= \frac{1}{M}(\tau_{h_i})^T(\tau_{h_j}) \\ &= \frac{1}{M}(u_{h_i})^T(N(\mathbf{r})^T N(\mathbf{r}))(u_{h_j}) \\ &= \frac{1}{M}(u_{h_i})^T(N(\mathbf{r}), N(\mathbf{r}))_{\mathbb{V}}(u_{h_j}) \\ &= \frac{1}{M}(u_{h_i})^T M_{h_{i,j}}(u_{h_j}), \end{aligned} \tag{3.17}$$

where $M_{h_{i,j}} \in \mathbb{R}^{N_h\times N_h}$ is the mass matrix in $\mathbb{V}$ space. Consider now the eigenvalue-eigenfunction pairs (λ_n, r_n) of the matrix C with normalization constraint $||r_n||_{\mathbb{V}} = 1$ satisfying:

$$C \cdot r_n = \lambda_n \cdot r_n \quad , \quad \text{with} \quad 1 \leqslant n \leqslant M. \tag{3.18}$$

It is assumed that the eigenvalues are sorted in descending order $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_M$.
The orthogonal POD basis functions $\{\xi_1, ..., \xi_N\}$ span the POD-space $\mathcal{X}_{POD} = \text{span}\{\xi_1, ..., \xi_N\}$ and are given by the linear combinations:

$$\begin{aligned} \xi_n(\mathbf{r}) &= \frac{1}{\sqrt{M}}\sum_{m=1}^{M}(r_n)_m \tau_h^m \\ &= \frac{1}{\sqrt{M}}\sum_{m=1}^{M}(r_n)_m \underline{N}^m(\mathbf{r})\underline{u}_h^m, \quad \text{with} \quad 1 \leq n \leq N_{rb}, \end{aligned} \tag{3.19}$$

where $(r_n)_m$ denotes the m-th coefficient of the eigenvector r_n. If the coefficients of the m-th solution are stored in an $N_h \times M$ matrix called $U \in \mathbb{C}^{N_h \times M}$, the eigenvector of the C matrix are stored in a matrix called $R \in \mathbb{C}^{M \times N_{rb}}$ and the n-th reduced basis $\xi_n(\mathbf{r})$ is stored in a row vector $\xi(\mathbf{r}) \in \mathbb{C}^{1 \times N_{rb}}$ in same analogy with vector of bases $\underline{N}(\mathbf{r})$, the previous equation can be written as:

$$\begin{aligned} U &= \left[u_h(\mu_1) \;\middle|\; ... \;\middle|\; u_h(\mu_M) \right], \\ R &= \left[r_1 \;\middle|\; ... \;\middle|\; r_M \right], \\ \underline{\xi}(\mathbf{r}) &= \underline{N}(\mathbf{r}) \frac{1}{\sqrt{M}} \underline{\underline{U}} \cdot \underline{\underline{R}}. \end{aligned} \tag{3.20}$$

We can define $B \in \mathbb{C}^{N_h \times N_r b}$ as the matrix:

$$B = \frac{1}{\sqrt{M}} UR. \tag{3.21}$$

If the basis is truncated and only the first N_{rb} functions $\xi_1, ..., \xi_{N_{rb}}$ are considered, they span the N_{rb}-dimensional space $\mathbb{V}_{POD}$ that satisfies the optimality criterion (3.16). The choice of bases which form the space $\mathbb{V}_{rb}$ is imposed by the fulfilment of a condition on the eigenvalues module of the correlation matrix C. Given a tolerance, all the eigenvectors r_n associated with eigenvalues such that $|\lambda_n / \sum_{k=1}^{M} \lambda_k| > tol$ for $n = 1, ..., M$ are used for the construction of the reduced basis. The *tol* value is not unique and depends on the problem, and on the precision that is wanted for τ_{rb}.
Now that the relationship between the reduced basis and the shape functions has been found, the linear system that will allow to find the approximate solution in the reduced basis space can be solved. Let us define the solution with the choice of both the basis functions:

$$\tau_h(\mathbf{r}; \mu) = \underline{N}(\mathbf{r}) \underline{u}_h(\mu) \quad \text{and} \quad \tau_{rb}(\mathbf{r}; \mu) = \underline{\xi}(\mathbf{r}) \underline{u}_{rb}(\mu). \tag{3.22}$$

The two solutions are identical, therefore:

$$\begin{cases} \underline{N}(\mathbf{r}) \underline{u}_h(\mu) = \underline{\xi}(\mathbf{r}) \underline{u}_{rb}(\mu) \\ \underline{\xi}(\mathbf{r}) = \underline{N}(\mathbf{r}) \underline{\underline{B}} \end{cases} \quad \Rightarrow \quad \underline{u}_h(\mu) = \underline{\underline{B}} \; \underline{u}_{rb}(\mu). \tag{3.23}$$

The system solving the approximated problem is:

$$\underline{\underline{K}}_h(\mu) \underline{u}_h(\mu) = \underline{f}_h(\mu). \tag{3.24}$$

Pre-multiplying the left equation for the vector $\underline{u}_h^T$ and using (3.23), a new linear system will be constructed having as unknown the coefficients of the solution in reduced basis space:

$$\underline{\underline{B}}^T \underline{\underline{K}}_h(\mu) \underline{\underline{B}}\ \underline{u}_{rb}(\mu) = \underline{\underline{B}}^T \underline{f}_h(\mu) \quad \text{or} \quad \underline{\underline{K}}_{rb}(\mu)\ \underline{u}_{rb}(\mu) = \underline{f}_{rb}(\mu). \tag{3.25}$$

The abstract form of (3.25) is (3.14). Since $\mathbb{V}_{rb} \subset \mathbb{V}_h$ we have:

$$a(\tau_h(\mu), v_{rb}, \mu) = f(v_{rb}; \mu) \quad \forall v_{rb} \in \mathbb{V}_{rb}. \tag{3.26}$$

Substracting (3.26) to (3.14) we get:

$$a(\tau_h(\mu) - \tau_{rb}(\mu), v_{rb}) = 0 \quad \forall v_{rb} \in \mathbb{V}_{rb}, \tag{3.27}$$

from that, due to the bilinearity of the form $a(\cdot, \cdot; \mu)$, we get (3.8). This proof shows that the Galerkin method is an *orthogonal projection*. To better understand its meaning, a geometric interpretation is adopted. In fact, if $a(\cdot, \cdot; \mu)$ was a scalar Euclidean product, $\tau_h(\mu)$ and $\tau_{rb}(\mu)$ two vectors and $\mathbb{V}_{rb}$ a sub-space of $\mathbb{V}_h$, (3.8) would express the error $e = \tau_h(\mu) - \tau_{rb}(\mu)$ orthogonality with respect to the sub-space $\mathbb{V}_h$ [26].

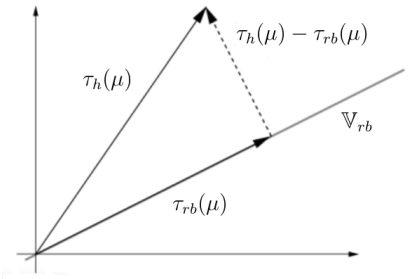

FIGURE 3.1. Geometric interpretation of Galerkin method.

In this sense, as shown in Figure 3.1, it can be stated that the POD method solution $\tau_{rb}(\mu)$ is the projection on $\mathbb{V}_{rb}$ of the truth $\tau_h(\mu)$. Thus, it is, among all the elements belonging to $\mathbb{V}_{rb}$, the only one that minimizes the distance from the truth solution $\tau_h(\mu)$ in the induced norm by the scalar product $a(\cdot, \cdot; \mu)$ [26]

$$||\tau_h(\mu) - \tau_{rb}(\mu)||_a = \sqrt{a(\tau_h(\mu) - \tau_{rb}(\mu), \tau_h(\mu) - \tau_{rb}(\mu))}. \tag{3.28}$$

4 | Model reduction of Maxwell's equations: the Stealth problem

In this chapter, we will deal with a first application: the problem of the reflection of the electromagnetic waves due to the presence of an object, called the *scattering problem*. In particular, we will deal with the reflection given by an airfoil hit by a known incident wave. Initially, some technical applications will be briefly introduced; then the problem will be formulated both in the weak and the strong form, and then it will be applied to a test case based on a NACA 0012 profile - as a verification -, and then on a NACA 4312 profile [1]. In the second case, we will proceed with the parameterization of the frequency of the incident wave and with the application of a model reduction method for parameterized equations, called Proper Orthogonal Decomposition (POD), and with the relative verifications of accuracy and reliability.

4.1 Introduction

Scattering is a physical phenomenon that causes waves - both scalar (e.g. acoustic), or vector (e.g. electromagnetic) - to be reflected and diffused when meeting an object along their path. Scattering applies to a number of different practical utilizations including medical imaging [17], subsurface geophysical prospecting [8], non-destructive diagnostics of materials [20].
Another important area in which scattering is used is radar visualization, which will be the focus of the thesis. Radar main use is security, such as the assistance during the aircraft landing maneuver [28] or the sounding of the seabed to avoid collisions in ships navigation [13]. These applications have contributed to a significant development of stealth technology. That is happened in order to respond to an increasingly high demand, in the military sector, to design aircraft, ships, and overall strategic resources most invisible

as possible. The concept of visibility needs clarification: an object is *visible* when it can be detected by sensors through heat, radio or sound. The key element to keep in consideration is to able to reflect as little as possible the signal sent by the detector. As a matter of fact, the return signal allows the spatial localization of the object and other important information about it. This type of problem is called *scattering problem*. The attention will be devoted to a 2D radio problem, in order to study the electromagnetic response of an airfoil hit by a plane radar wave, having fixed amplitude, wavelength and direction.
The unknown vector, for a scattering problem, is the sum of two fields:

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E}^s(\mathbf{r},t) + \mathbf{E}^i(\mathbf{r},t) = \begin{bmatrix} E_x^s(\mathbf{r},t) \\ E_y^s(\mathbf{r},t) \\ E_z^s(\mathbf{r},t) \end{bmatrix} + \begin{bmatrix} E_x^i(\mathbf{r},t) \\ E_y^i(\mathbf{r},t) \\ E_z^i(\mathbf{r},t) \end{bmatrix}, \tag{4.1}$$

with $\mathbf{E}^s(\mathbf{r},t)$ the scattering electric field and $\mathbf{E}^i(\mathbf{r},t)$ the incident electric field. The incident field is a given element of the problem, so the unknown is $\mathbf{E}^s(\mathbf{r},t)$ ([16], [24]).

4.2 Problem formulation

This section introduces the mathematical formalization of the problem, in particular both the weak form and the strong form will be written.
The assumptions underlying the model are:

- bi-dimensional problem: the reference frame is Cartesian, axis $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ lying on the same plane of the domain. The origin point O(0,0) is placed on the airfoil leading edge (Figure 4.1 and Figure 4.6). These assumptions simplify some components of the vector:

$$\hat{\mathbf{n}} = \begin{bmatrix} n_x \\ n_y \\ 0 \end{bmatrix} \quad \text{with} \quad |\hat{\mathbf{n}}| = 1, \quad \nabla = \begin{bmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial y} \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}; \tag{4.2}$$

- dielectric media: there is no possibility of a distribution of free charge, then $\rho = 0$; moreover $\sigma = 0$. This involves the simplification of the curl-curl operator term in (1.67):

$$\begin{cases} \nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \\ \nabla \cdot \mathbf{E} = 0 \end{cases} \quad \Rightarrow \quad \nabla \times \nabla \times \mathbf{E} = -\nabla^2 \mathbf{E}; \tag{4.3}$$

- absence of source: the terms of flow source density in (1.67) are null, then:

$$\mathbf{J}_N = \mathbf{0}; \tag{4.4}$$

- single wavelength: the source wave will have a single wavelength, therefore the associated frequency, ω, will be the parameter of the problem. Due to the previous two hypothesis, the relation used is (1.78), i.e.,

$$|\mathbf{k}| = k = \omega\sqrt{\mu\epsilon}. \tag{4.5}$$

  Furthermore, regarding the vacuum impedance, equation (1.81) will become:

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}; \tag{4.6}$$

- electric field imposed: the airfoil is considered as a perfect conductor, so its boundary will be a S_1^{fb} surface type, where we have:

$$\hat{\mathbf{n}} \times \mathbf{E} = 0; \tag{4.7}$$

- *Silver-Muller* condition: it is applied on the S_2^{∞} boundary, so the computational domain is modeled as an open domain. Furthermore, in a scattering problem, Silver-Muller condition applies directly to the scattered field ([16], [24]).

The equation and its boundary conditions that describe this problem are collected in (1.67). But in order to follow the guidelines shown in Chapter 2, the problem is written in the frequency domain. Then, the equation (1.69) is applied and thanks to equations (4.3), (4.4), (4.5),(4.6) and (4.7), (1.67) is simplified in the following way:

$$\begin{cases} \nabla^2\tilde{\mathbf{E}} + \mu\epsilon\omega^2\tilde{\mathbf{E}} = \mathbf{0} \quad \text{in } V \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}} = \mathbf{0} \quad \text{on } S_1^{fb} \\ \hat{\mathbf{n}} \times \nabla \times \tilde{\mathbf{E}} + j\omega\dfrac{\mu_0}{Z_0}(\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \tilde{\mathbf{E}}) = \\ \quad = \hat{\mathbf{n}} \times \nabla \times \tilde{\mathbf{E}}^i + j\omega\dfrac{\mu_0}{Z_0}(\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \tilde{\mathbf{E}}^i) \quad \text{on } S_2^{\infty}. \end{cases} \tag{4.8}$$

Due to the subdivision of vector $\mathbf{E}(\mathbf{r},t)$ in (4.1), we obtain:

$$\begin{cases} \nabla^2\tilde{\mathbf{E}}^s + \mu\epsilon\omega^2\tilde{\mathbf{E}}^s = -(\nabla^2\tilde{\mathbf{E}}^i + \mu\epsilon\omega^2\tilde{\mathbf{E}}^i) \quad \text{in } V \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}}^s = -(\hat{\mathbf{n}} \times \tilde{\mathbf{E}}^i) \quad \text{on } S_1^{fb} \\ \hat{\mathbf{n}} \times \nabla \times \tilde{\mathbf{E}}^s = -j\omega\dfrac{\mu_0}{Z_0}(\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \tilde{\mathbf{E}}^s) \quad \text{on } S_2^{\infty}. \end{cases} \tag{4.9}$$

The field $\tilde{\mathbf{E}}^i$ fulfils the indefinite equation

$$\nabla^2\tilde{\mathbf{E}}^i + \mu\epsilon\omega^2\tilde{\mathbf{E}}^i = 0, \tag{4.10}$$

that is a wave extending, hypothetically, without attenuation throughout the domain. This electromagnetic wave travels along the direction of vector $\mathbf{k}$, while the electric field and the magnetic field develop orthogonally between them. These properties allow to simplify the vector problem (4.9) into a scalar problem, requiring that vector $\mathbf{k}$ and vector $\mathbf{H}$ belong to domain V - a plane, in other words - and that vector $\mathbf{E}$, as a result, is along the $\hat{z}$ direction. The vector of the incident wave will be:

$$\tilde{\mathbf{E}}_i(x,y) = \begin{bmatrix} 0 \\ 0 \\ \tilde{E}_z^i(x,y) \end{bmatrix}. \tag{4.11}$$

The incident electric field $\tilde{E}_z^i(\mathbf{r})$ is written like:

$$\tilde{E}_z^i(\mathbf{r}) = \tilde{E}_{za}^i e^{-j\mathbf{k}\cdot\mathbf{r}} = \tilde{E}_{za}^i e^{-j|\mathbf{k}|(x\cos\gamma + y\sin\gamma)}, \tag{4.12}$$

with γ the angle of orientation and $\tilde{E}_{za}$ the wave amplitude.
The boundary conditions to be inserted in (4.9) relating to $\tilde{\mathbf{E}}^i(\mathbf{r},t)$ are its evaluations on the boundary S_1^{fb}:

$$\hat{\mathbf{n}} \times \tilde{\mathbf{E}}^i\Big|_{S_1^{fb}} = \tilde{\mathbf{c}}_E. \tag{4.13}$$

Accordingly, the unknown vector will also be:

$$\tilde{\mathbf{E}}_s(x,y) = \begin{bmatrix} 0 \\ 0 \\ \tilde{E}_z^s(x,y) \end{bmatrix}. \tag{4.14}$$

The scalar equation in strong form is obtained by applying the hypothesis of a bi-dimensional problem to (4.9), so:

$$\begin{cases} \nabla^2 \tilde{E}_z^s + \mu\epsilon\omega^2 \tilde{E}_z^s = 0 \quad \text{in } V \\ \tilde{E}_z^s = -\tilde{c}_z \quad \text{on } S_1^{fb} \\ \dfrac{\partial \tilde{E}_z^s}{\partial \hat{n}} = -j\omega \dfrac{\mu_0}{Z_0} |\hat{\mathbf{n}}|^2 \tilde{E}_z^s \quad \text{on } S_2^{\infty}. \end{cases} \tag{4.15}$$

We introduce the function space H(curl,V) defined in Chapter 1.
The weak formulation for the electric field (2.5) is used in its scalar form. The hypothesis listed above are applied to it, and the resulting weak formulation is:

$$\begin{aligned} &\text{find } \tilde{E}_z^s \in \text{H(curl,}V\text{) such that} \\ &\int_V \left(\frac{\partial v}{\partial x}\frac{\partial \tilde{E}_z^s}{\partial x} + \frac{\partial v}{\partial y}\frac{\partial \tilde{E}_z^s}{\partial y} - v(\omega^2 \epsilon\mu \tilde{E}_z^s) \right) \\ &+ \int_{S_2^{\infty}} v(j\omega \frac{\mu}{Z} |\hat{\mathbf{n}}|^2 \tilde{E}_z^s) \ = 0 \quad \forall v \in \text{ H(curl,V)}. \end{aligned} \tag{4.16}$$

Introducing a bilinear form:

$$a(\tilde{E}_z^s, v) = \int_V \left(\frac{\partial v}{\partial x}\frac{\partial \tilde{E}_z^s}{\partial x} + \frac{\partial v}{\partial y}\frac{\partial \tilde{E}_z^s}{\partial y} - v(\omega^2 \epsilon\mu \tilde{E}_z^s) \right) + \int_{S_2^{\infty}} v(j\omega \frac{\mu}{Z} |\hat{\mathbf{n}}|^2 \tilde{E}_z^s), \tag{4.17}$$

and a linear functional:

$$f(v) = 0, \tag{4.18}$$

the problem (4.16) is written as:

$$\begin{aligned} &\text{find } \tilde{E}_z^s \in \text{H(curl,}V\text{) such that} \\ &a(\tilde{E}_z^s, v) = f(v) \quad \forall v \in \text{H(curl,}V\text{).} \end{aligned} \tag{4.19}$$

The coercivity and continuity of the bilinear form $a(\tilde{E}_z^s, v)$ and the continuity of the linear functional $f(v)$ can be established. As a result, the Lax-Milgram theorem ensures the existence and uniqueness of the solution [26].
The unknown field $\tilde{E}_z$ will be indicated as E to make the notation lighter.

4.3 Test case: a NACA 0012 airfoil

A validation problem will be preliminarily solved, using the COMSOL solver [7], in order to have a comparison with the results obtained by a work presented by O. Pironneau [23]. We will consider an airfoil NACA 0012 with angle of attack equal to $\alpha = 0°$ and length $l = 1$ m, hit by a known electromagnetic wave having direction $\gamma = 15°$, intensity $E_a = 1$ V/m and angular velocity $\omega = 9.163 \cdot 10^9$ rad/s. The wave travels through air, with $\epsilon_r = 1$ and $\mu_r = 1$.
The analytical problem solved is (4.15), having chosen S_1^{fb} as the boundary profile where to apply the condition of perfect conductor, while the perimeter of the rectangle was chosen as S_2^{∞}, where the condition of Silver-Muller was applied. The problem in weak formulation (4.16) has been solved using the finite elements method, as introduced in Chapter 2. A class of $\mathbb{P}^3$ function was chosen as an approximating shape function. The motivation is based on the Shannon sampling theorem because, with a class of $\mathbb{P}^2$, the spatial grid would not be thin enough to well approximate the solution variations. The linear system associated has been characterized by 251049 DoFs, and a MUltifrontal Massively Parallel Sparse direct Solver (MUMPS) [4] solver has been used to compute it. In Table 4.1, the main numerical features are collected.

Mesh NACA 0012	
Triangles	55572
Boundary elements	650
DoFs	251049
Shape function	$\mathbb{P}^3$

TABLE 4.1. NACA 0012: numerical data.

4.3.1 Geometry and mesh

The computational domain is a rectangle 4 m high and 6 m long, as shown in Figure 4.1.
The mesh of the problem has been obtained using a specific command within the program itself (*Mesh*). The elements used are triangles having a minimum dimension of 0.003 m, while a criterion based both on the polynomial function chosen for the approximation and on *Shannon's sampling theorem* [9] was used for the maximum dimension in order not to lose any information about the shape and amplitude of solution. Since, in the neighborhood of the profile,

the computational domain is bent, an unstructured mesh discretization was chosen.

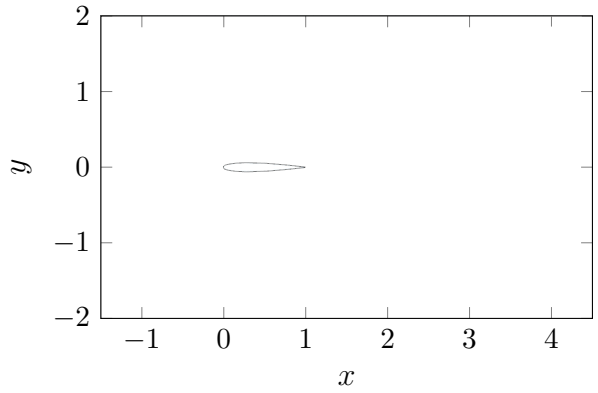

FIGURE 4.1. NACA 0012: problem geometry.

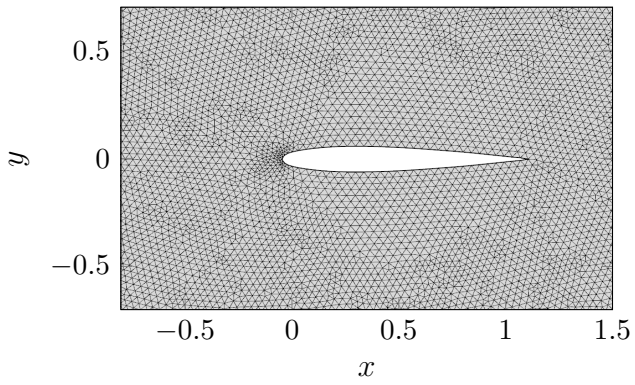

FIGURE 4.2. NACA 0012: particular of the mesh.

4.3.2 Results, representative visualizations and comparisons

The result obtained in the neighborhood of the profile is shown in figure 4.3. Graphically comparing the solution obtained with COMSOL and the solution obtained from O. Pironneau in [23], we remark that they are very similar. The simulations are shown in Figure 4.4. The accuracy of the solution is suggested, as well by visual comparison shown above, also by the type of solution obtained. In fact, (4.15) derives from a *Linear Time-Invariant* (LTI)

dynamic system [34], the solutions angular velocity ω will be equal to the source one. In order to demonstrate this property, the maximum peaks - denoted by a red colour in Figure 4.5 - are counted. Starting from the leading edge up to the trailing edge, the number is 5. Now, it is expected that the wavelength multiplied by the peaks number be equal to the profile length. The angular velocity is linked to the wavelength by the relation $\lambda = 2\pi/(\omega\sqrt{\mu\epsilon})$, thus:

$$\lambda \cdot 5 = \frac{2\pi}{\omega\sqrt{\mu\epsilon}} \cdot 5 = \frac{2\pi}{\sqrt{4\pi \cdot 10^{-7}\ 8.8541 \cdot 10^{-12}}} \cdot 5 \text{ m} = 1,0282 \text{ m} \approx 1 \text{ m}. \quad (4.20)$$

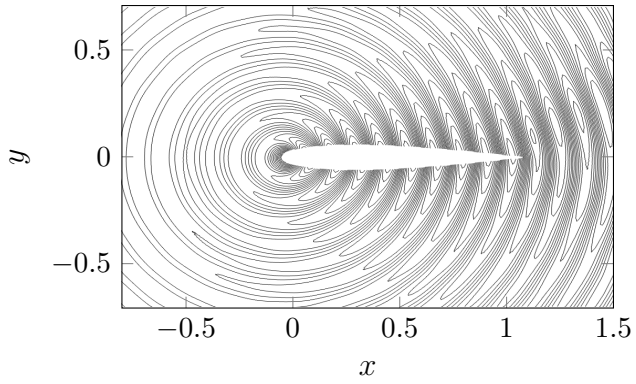

FIGURE 4.3. NACA 0012: isolines of the scattered wave in the neighbourhood of the airfoil.

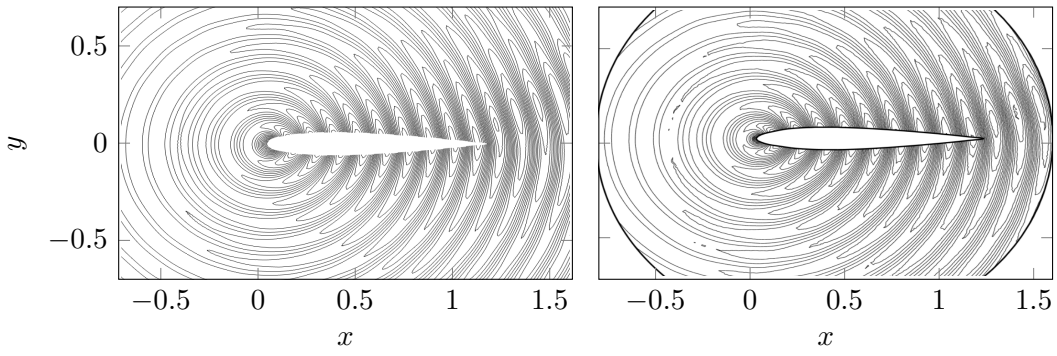

FIGURE 4.4. Comparison between the two simulations: on the left side the COMSOL solution and on the right side the solution by O. Pirroneau [23].

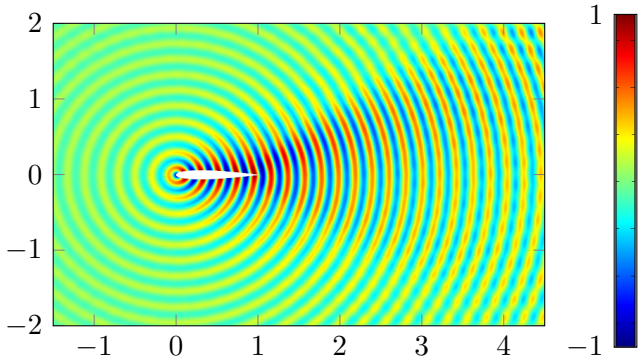

FIGURE 4.5. NACA 0012: COMSOL solution.

4.4 Parameterized case: a NACA 4312 airfoil

Let us consider a NACA 4312 profile with an angle of attack equal to $\alpha = 5°$ and length equal to $l = 1$ m, hit by an incident electromagnetic wave having direction $\gamma = 0°$, intensity $E_a = 60$ V/m. In this case, the parameter which will vary from one simulation to another will be the angular velocity, to which 51 values between a range of $\omega = \{6.283 : (9.425 - 6.283)/(51 - 1) : 9.425\} \cdot 10^9$ rad/s will be assigned. To make use of the formalism used in Chapter 3, we define $\mu = \omega \in \mathbb{P}$, con $M = 51$. Let us observe that, in this case, μ is not the magnetic permeability of the material, but the set of parameters. The incident wave travels in the air, then with values of $\epsilon_r = 1$ and $\mu_r = 1$. The finite element solver will be provided by COMSOL [7].
The solved analytical problem in strong form is (4.15), having chosen S_1^{fb} as the leading edge where the condition of perfect conductor has been applied, while the perimeter of the rectangle has been chosen as S_2^{∞}, where the Silver-Muller condition has been applied. The problem in weak formulation (4.16) has been solved using a class of approximating polynomials $\mathbb{P}_3$, while the associated algebraic linear problem has been characterized by 263004 DoFs, and a MUltifrontal Massively Parallel Sparse direct Solver (MUMPS) [4] solver has been used to compute it. In Table 4.2, the main numerical features are collected.

4.4.1 Geometry and mesh

The computational domain is a rectangle of 4 m height and 6 m width, as shown in Figure 4.6.
The mesh of the problem has been obtained using the tool (*Mesh*), which allows the creation of various types of mesh: in this case, a non-structured

Mesh NACA 4312	
Triangles	58223
Boundary elements	667
DoFs	263004
Shape function	$\mathbb{P}^3$

TABLE 4.2. NACA 4312: numerical data.

mesh - Figure 4.7 - has been chosen, having as a minimum dimension 0.003 m. Instead, a criterion based both on the degree of the approximating polynomial and *Shannon's sampling theorem* [9] has been used to determine the maximum size, in order not to miss any information given by the signal. In more detail, since the degree of the approximating polynomial is $N = 3$, and being the problem two-dimensional, it has been decided to *spatially sample* the signal with an interval of:

$$h = \frac{\lambda}{6}, \tag{4.21}$$

where h is the spatial dimension which characterizes the generic finite element, while λ is the wavelength of the incident signal, obtained from:

$$\begin{cases} f = \dfrac{1}{T} \\ T = \dfrac{2\pi}{\omega} \end{cases} \quad \Rightarrow \; f = \frac{\omega}{2\pi} \tag{4.22}$$

and

$$\begin{cases} k = \dfrac{2\pi}{\lambda} \\ k = \omega\sqrt{\epsilon\mu} \end{cases} \quad \Rightarrow \; \lambda = \frac{2\pi}{\omega\sqrt{\epsilon\mu}}, \tag{4.23}$$

whence

$$\lambda = \frac{2\pi}{2\pi f\sqrt{\epsilon\mu}} = \frac{1}{f\sqrt{\epsilon\mu}}. \tag{4.24}$$

Furthermore, it is evident how the Shannon theorem suggests that, to have an effective sampling of a signal with frequency f_s, it is needed a sampling frequency f_c such as to ensure the relation:

$$f_c > 2f_s \quad \text{whence} \quad \lambda_c < \frac{\lambda_s}{2}, \tag{4.25}$$

which confirms the accuracy of the choice made, for in the problems $\lambda_c = h$ and $\lambda_s = \lambda$.

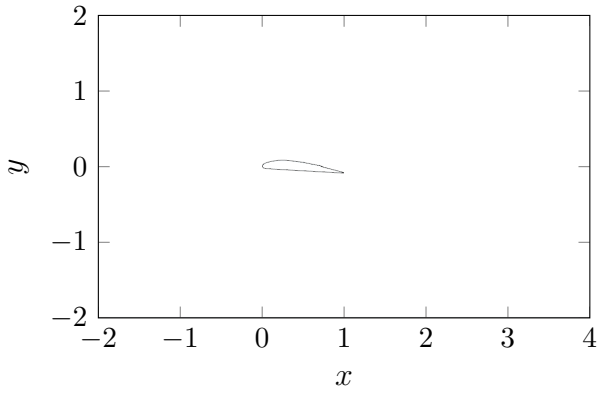

FIGURE 4.6. NACA 4312: problem geometry.

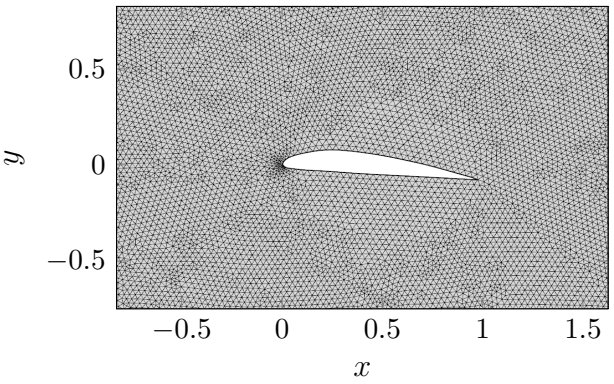

FIGURE 4.7. NACA 4312: particular of the mesh.

4.4.2 Results, representative visualizations and comparisons

Following what has been introduced in Chapter 3, the *offline* phase is the first step for the generation of a consistent POD method. As only one parameter has been changed to form the different configurations, the number of simulations generated was equal to the number of parameters. To prove the effectivity and feasibility of the method, three types of figures will be shown to provide an idea of how the method works in the *online* phase:

- all the obtained eigenvalues are shown in Figure 4.8, followed by those used on the construction of the basis. The selection criterion is based on a module control of each normalized eigenvalue, discarding those with $|\lambda_i / \sum_{k=1}^{51} \lambda_k| < 10^{-6}$ for $i = 1, ..., 51$, because of little energetic influence;

- the difference between the direct solution and the solution obtained through the POD method, given by $\Delta\tau = \tau_h - \tau_{POD}$, for a specific parameter. These were chosen in order to check and proof the uniformity in the reliability of the POD method. In fact, the chosen parameters are localized at the beginning, in the middle, and at the end of the set μ;

- the relative error given by $Err = \dfrac{|\tau_{h_i} - \tau_{POD_i}|}{|\tau_{h_i}|}$ for $i = 1..., 263004$ for the same chosen parameters above.

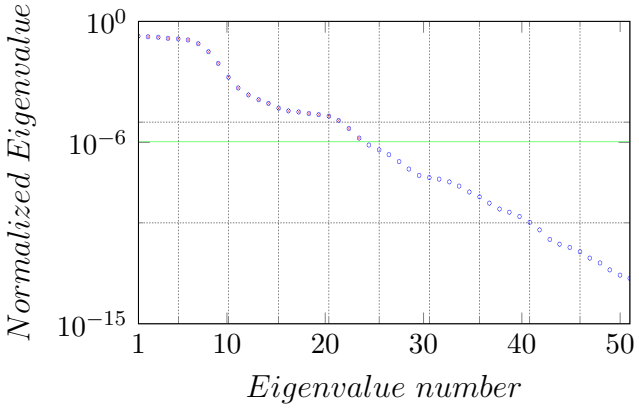

FIGURE 4.8. NACA 4312: eigenvalues of the correlation matrix. The blue circles are all the eigenvalues, while the red crosses indicate only the ones used to construct the reduce basis space. The green line is the tolerance limit. For the further example, the number of bases used to approximate the solution will be $N_{rb} = 23$.

The difference $\Delta\tau$ does not allow to give a clear interpretation of the accuracy of the solution τ_{POD}. For this reason, a finer indicator, the relative error Err, has been used. The error is certain when a reduced method is used, but in the engineering applications, if this turns out be limited below a specific value as $Err = 2.5\% \div 5\%$, the solution is considered accurate and usable. Thus, as shown in 4.9, 4.10 and 4.11, the POD reduced method works very well.

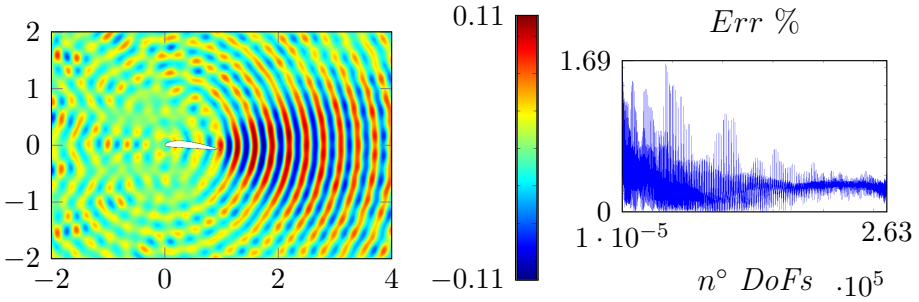

FIGURE 4.9. NACA 4312: the left figure shows $\Delta\tau$, while the right one shows Err; both are obtained by a parameter value of $\omega = 6.283 \cdot 10^9$ rad/s.

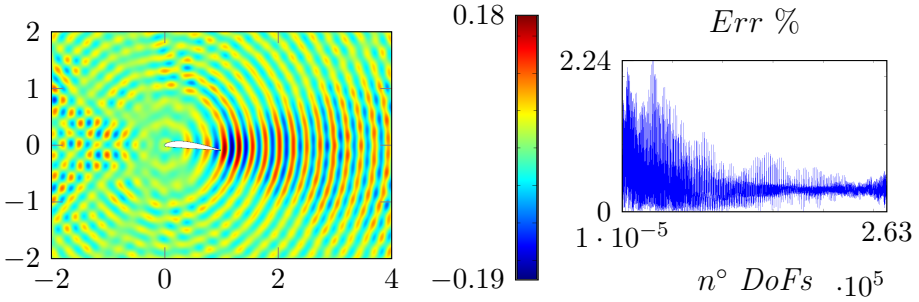

FIGURE 4.10. NACA 4312: the left figure shows $\Delta\tau$, while the right one shows Err; both are obtained by a parameter value of $\omega = 7.886 \cdot 10^9$ rad/s.

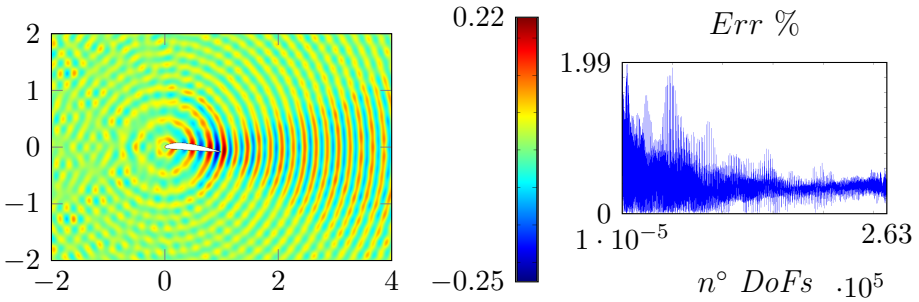

FIGURE 4.11. NACA 4312: the left figure shows $\Delta\tau$, while the right one shows Err; both are obtained by a parameter value of $\omega = 9.425 \cdot 10^9$ rad/s.

4.4.3 Computational saving

It will proceed now to show some significant data related to the computational savings guaranteed by the reduced method.
In order to compare the FEM approach and the RB one, the *avarege times* spent to compute the same number of simulations are computed. In particular,

the problems (3.24) and (3.25) were solved for every parametric configuration defined in the offline phase, hence for the set $\omega = \{6.283 : (9.425-6.283)/(51-1) : 9.425\} \cdot 10^9$ rad/s. The bases used for the reduced problem are $N_{rb} = 23$, which guarantees an accurate solution for the engineering standard. To solve the $M = 51$ algebraic systems of both approaches, Matlab [21] has been used and in particular the command *backslash*. Thanks to the command *mphmatrix* the stiffness matrix and the forcing term have been extracted.
As shown in Figure 4.12, the average time spent for the solution with FEM method is $T_{FEM} = 15.21$ s, while with RB method is $T_{RB} = 1.93$ s, which is approximately 8 times less, which means a saving of 87.37 % for the calculation time. Obviously, times for mesh generation and matrices assembly have not been counted as are common operations to both approaches. The commands *tic* and *toc* are used for time measurement. This is a practical demonstration of the efficiency of the online phase of the reduced method. While the offline part, done once, occupies most of the calculation time, in this case about 776 seconds, the online part can be considered, in comparison, a *real-time* phase. Thus, the reduced order approach is justified.

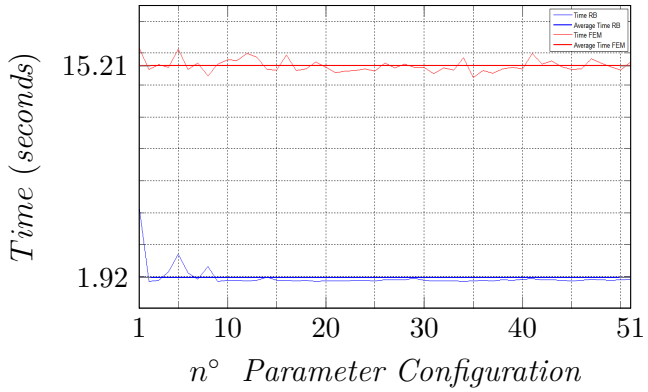

FIGURE 4.12. NACA 4312: relation among time spent to solved the FEM and RB systems (red line and blue line respectively) and the number of basis used to approximate the solution. The bold lines are the average times.

4.4.4 Convergence and consistency

As seen in Chapter 3, the reduced space $\mathbb{V}_{rb}$ is a subspace of the complete space $\mathbb{V}_h$ where the real solution τ_h is searched. It is expected that, as the reduced space tends to the complete space, also the solution τ_{rb} tends to the real one. In this way, also the relative error Err will tend to zero. Following

this argument, the convergence and consistency of the RB method will be demonstrated.
In Figure 4.13, the maximum and average relative error, obtained with the approximate RB solution and parameter fixed, $\omega = 7.886 \cdot 10^9$ rad/s, are shown. When the number of bases used to form the solution space is incremented, the behavior of both indicators tends to decrease, up to very low values. Obviously, there is no the need to choose a number of bases so high to obtain such values. In fact, the engineering standards of tolerability (about 2.5 % $\div$ 5 % of maximum relative error) is satisfied when the number of bases is in a range of 20/25 bases, that has already been confirmed by the previous comparison analysis among the FEM solutions and RB. In fact, as shown in Figure 4.8, the number of preserved eigenvalues is just equal to $N_{rb} = 23$.

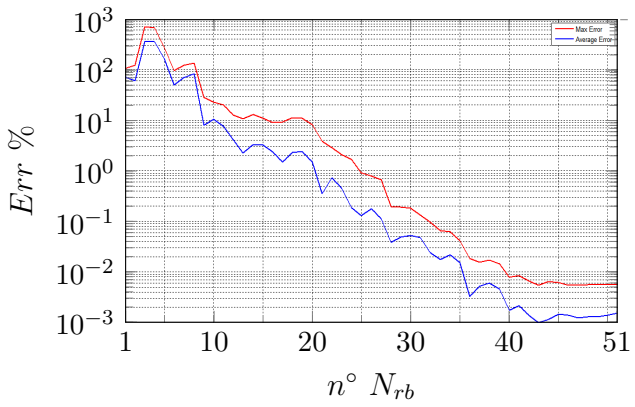

FIGURE 4.13. NACA 4312: relation among Maximum/Average relative error (red line and blue line respectively) and the number of basis used to approximate the solution with RB method.

5 | Model reduction of Maxwell's equations: the Optical Fiber problem

This Chapter deals with the so-called *open field problem*. In particular, a problem given by a dielectric, an optical fiber excited at the basis, immersed in the air will be considered. Initially, some brief history of fiber optics and their application will be introduced . Then, the mathematical theory of the problem will be developed, with the derivation of both the weak and the strong formulation. In the following section, the discussion of the geometry mapping - necessary to develop an appropriate POD model reduction method - and its parameterization will be explained. Then, the comparison of some direct solutions and the solutions obtained via model reduction will be shown.

5.1 Introduction

Fiber-optic communication is a technology solution that has had an interesting development in the twentieth century. The basic idea of this technology lies in the possibility of information/data transmission, through a dielectric medium, in the form of electromagnetic signal. The reasons for their massive development are the very low loss of information along distances with lengths of the order of kilometers, the immunity from electrical noise, weather conditions and temperature changes ([2], [18]).
The simplest and most widely used geometrical configuration is the coaxial cylinder made of different materials, the inner one called core and the more external called cladding. These ones are responsible for the transport of signals. The transmission of electromagnetic signals is enabled if and only if the two materials have different refractive indices, in particular if the refractive index of the core, n_1, is greater than the cladding one, n_2 (Figure 5.1). This type of optical fiber is called *step-index* because there is no gradual variation

between core and cladding (see Figure 5.2). In this thesis, only this type will be considered.

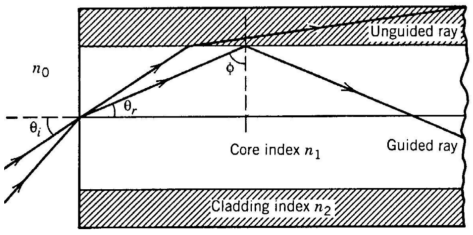

FIGURE 5.1. Section of a cylindrical optical fiber. Depending on the incidence angle of the electrical signals, some rays are reflect, remaining confined inside the core, while others pass through the cladding and are not transported.

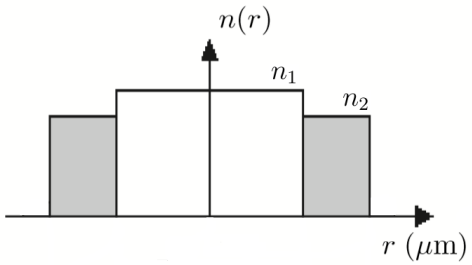

FIGURE 5.2. Pattern of a *step-index* refraction fiber with $n_1 > n_2$

A further classification depends on how many *modes* the fiber can carry: a single-mode fiber can carry only one mode; in other words, only signals having the shape of the selected mode are allowed to transit inside. A multi-mode fiber, instead, allows the transit of many modes([2], [18]).

In the last 40-50 years, optical fibers have grown rapidly in sectors like lighting and telecommunications, but it is barely known that the first scientific uses were made in the medical field. In fact, the first application, still rudimentary, was realized for medical purposes, namely to build a gastroscope. The first semi-flexible optical fiber gastroscope was patented by Basil Hirschowitz, C. Wilbur Peters, and Lawrence E. Curtiss in 1956. It is precisely in this same area that lies the reason for the qualitative and numerical study that will be described below. Some researchers of the SISSA neuroscience department,

led by Professor Vincent Torre, wanted to demonstrate the inhomogeneity of the photo-receptors within the cells used for the capture of light in vertebrate organisms. For this study, they have used optical fibers having cylindrical stems and conical tips. In particular, the tips have not been equipped with opening as in their classic configurations, but have been fully coated and for this reason called *tapered*. The thicknesses and characteristic sizes are on the order of micron, capable of producing very localized beams of electromagnetic waves. In this way, only the cell areas with dimensions comparable with the photo-receptors are excited. The core material is silica, SiO_2, while the cladding is gold, Au.

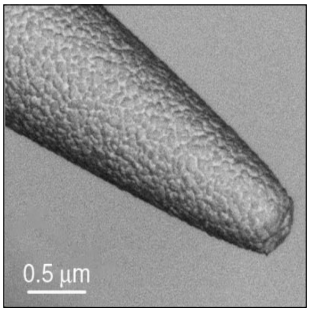

FIGURE 5.3. Microscope imagine of the Tapered Optical Fiber (TOP) tip.

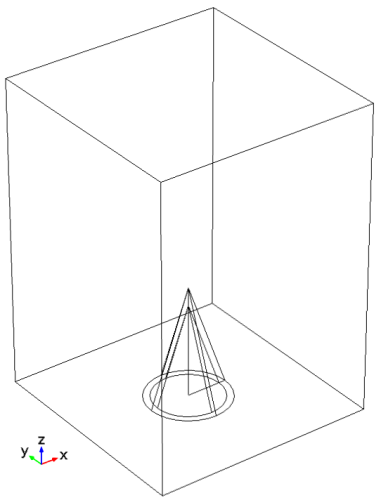

FIGURE 5.4. TOP geometric model used in this thesis.

5.2 Waveguides analytic theory: LP_{01} mode and its approximation

Let us consider a typical cylindrical geometry for the waveguide. This geometry allows us some analytical results, in particular, it leads to the analytical resolution of the proper modes of the system. Moreover, we consider a single-mode step-index fiber [2].

The equations for this type of analysis are (1.17) for $\mathbf{E}(\mathbf{r}, t)$ and (1.20) for $\mathbf{H}(\mathbf{r}, t)$. Let us consider the equations of the electromagnetic waves having the electric field as unknown, taking into account that the same derivation can be applied to the the magnetic field. To find the proper modes of the system, let us impose the source term $\mathbf{J}_N$ equal to zero.
The material of the core is a dielectric, which implies $\sigma = 0$. Furthermore, $\rho = 0$ in order to obtain the vanishing of the charge distribution.
The equation is:

$$\nabla^2 \mathbf{E} + \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = \mathbf{0}. \tag{5.1}$$

The cladding material is a metal. In this media, the characteristic time of the free charge distribution is smaller than the characteristic waveguide phenomena. For this reason, a metallic media can be modeled by imposing $\rho = 0$. The resulting equation is:

$$\nabla^2 \mathbf{E} + \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu\sigma \frac{\partial \mathbf{E}}{\partial t} = \mathbf{0}. \tag{5.2}$$

Equations (5.1) and (5.2) become, after applying (1.69):

$$\begin{cases} \nabla^2 \mathbf{E} + \omega^2 \mu\epsilon \mathbf{E} = \mathbf{0} \text{ for the core} \\ \nabla^2 \mathbf{E} + \omega^2 \mu\hat{\epsilon} \mathbf{E} = \mathbf{0} \text{ for the cludding,} \end{cases} \tag{5.3}$$

having introduced the complex electrical permittivity:

$$\hat{\epsilon} = \epsilon - j\frac{\sigma}{\omega}. \tag{5.4}$$

Thanks to (1.78) and the definition of refractive index $n = c/v_p$, with c the speed of light and v_p the wave propagation speed, we can derive the following

relationship:

$$\begin{cases} k^2 = \omega^2\mu\epsilon = \omega^2\mu_0\mu_r\epsilon_0\epsilon_r \\ k_0^2 = \omega^2\mu_0\epsilon_0 \\ n = \dfrac{1}{\sqrt{\mu_0\epsilon_0}}\sqrt{\mu\epsilon} = \dfrac{1}{\sqrt{\mu_0\epsilon_0}}\sqrt{\mu_0\epsilon_0}\sqrt{\mu_r\epsilon_r} = \sqrt{\mu_r\epsilon_r} \end{cases} \quad \Rightarrow \quad k^2 = k_0^2 n^2, \tag{5.5}$$

which, inserted in (5.3), allows the writing of the equations depending on the refractive index:

$$\begin{cases} \nabla^2\mathbf{E} + k_0^2 n^2(\omega)\mathbf{E} = \mathbf{0} \text{ for the core} \\ \nabla^2\mathbf{E} + k_0^2 \hat{n}^2(\omega)\mathbf{E} = \mathbf{0} \text{ for the cludding,} \end{cases} \tag{5.6}$$

with $\hat{n}$ complex refractive index.
There are two important considerations. The first one is that the refractive index depends on the frequency of the wave, as well on the material in which it travels. The frequency dependence is expressed according to empirical relationships. The second one is that, up to the complex nature of the index of refraction, the two equations are identical. This will allow us the development of the analytical theory independently of the material properties.

Let us consider a system of cylindrical coordinates (r, ϕ, z), with the z axis oriented in the direction of the longitudinal development of the waveguide. Either of (5.6) becomes:

$$\frac{\partial^2\mathbf{E}}{r^2} + \frac{1}{r}\frac{\partial\mathbf{E}}{\partial r} + \frac{1}{r^2}\frac{\partial^2\mathbf{E}}{\partial\phi^2} + \frac{\partial^2\mathbf{E}}{\partial z^2} + n^2 k_0^2\mathbf{E} = \mathbf{0}. \tag{5.7}$$

This system consists of three equations in three unknowns, namely the three components of the vector $\mathbf{E}$. Thanks to (1.71), we can express two of the three unknowns as a function of the remaining one, thus being able to solve only one of the three equations. In this case, we choose E_r as an independent unknown.
From the system (5.7), the radial equation is taken, then:

$$\frac{\partial^2 E_r}{r^2} + \frac{1}{r}\frac{\partial E_r}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_r}{\partial\phi^2} + \frac{\partial^2 E_r}{\partial z^2} + n^2 k_0^2 E_r = 0. \tag{5.8}$$

To find the closed form solution of (5.8), it is necessary to apply the decomposition of the unknown variable in separable form, that is:

$$E_r(r, \phi, z) = R(r)\Phi(\phi)A(z). \tag{5.9}$$

Inserting (5.9) in (5.8), we obtain:

$$\left(\frac{1}{R}\frac{d^2R}{dr^2}+\frac{1}{r}\frac{1}{R}\frac{dR}{dr}+\frac{1}{r^2}\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}+\frac{1}{A}\frac{d^2A}{dz^2}+n^2k_0^2\right)R\Phi A=0. \tag{5.10}$$

We can see that each term in (5.10) depends on a single variable. Therefore, grouping the addends depending on a variable only, we can deduce that the only way for the equation to be dimensionally homogeneous is that each group of functions is necessarily equal to a dimensionally correct constant. This allows us to extract three homogeneous differential equations such as:

$$\begin{cases}\dfrac{d^2A}{dz^2}\dfrac{1}{A}=-\beta^2\\ \dfrac{d^2\Phi}{d\phi^2}\dfrac{1}{\Phi}=-m^2\\ \dfrac{d^2R}{dr^2}+\dfrac{1}{r}\dfrac{dR}{dr}+(n^2k_0^2-\beta^2-\dfrac{m^2}{r^2})R=0.\end{cases} \tag{5.11}$$

The first two equations of (5.11), having chosen β and m positive and sign-changed in the equation, have a limited and harmonic solution as:

$$\begin{cases}A(z)=A_0e^{j\beta z}\\ \Phi(\phi)=\Phi_0e^{jm\phi}.\end{cases} \tag{5.12}$$

The third equation has as a particular solution, called *Bessel function*. Without going into the details of the mathematical steps, we can show that the function $R(r)$ has the following solution:

$$R(r)=\begin{cases}C_1J_m(\kappa r)+C_2Y_m(\kappa r) \text{ for r} \leqslant a\\ C_3K_m(\gamma r)+C_4I_m(\gamma r) \text{ for r} > a,\end{cases} \tag{5.13}$$

with a = *core radius*, $\kappa^2=n_1^2k_0^2-\beta^2$, $\gamma=\beta^2-n_2^2k_0^2$ and J_m, Y_m, K_m and I_m different types of Bessel functions. Of course, the first expression is related to the core, while the second one is related to the cladding. This notation will be maintained hereafter. Some assumptions, both physical and geometrical, are necessary to reach the goal.
The physical hypotheses are:

- to obtain physical solutions without singularities, it is necessary to impose that the electric field E_r - evaluated along the waveguide axis

- assumes a finite value, while for values going to the infinity of the waveguide radius, E_r must tend to zero, that is:

$$\begin{cases} E_r(r=0) \neq \infty \\ E_r(r \to \infty) = 0. \end{cases} \tag{5.14}$$

The geometric assumptions are:

- the waveguide is considered a mild guide; this means that the refractive index variation n along the radius of the fiber is very small, although the considered guide is a step-index fiber. In this way, the interface conditions between core and cladding are approximated with the continuity of $E(r)$ and $\dfrac{dE}{dr}$ at $r = a$, that is:

$$\begin{cases} E_r^{core}(r=a) = E_r^{cladding}(r=a) \\ \left.\dfrac{dE_r^{core}}{dr}\right|_{r=a} = \left.\dfrac{dE_r^{cladding}}{dr}\right|_{r=a}. \end{cases} \tag{5.15}$$

Using (5.14), we can simplify (5.13) and express the $E_r(r)$ component as:

$$E_r(r) = \begin{cases} C_1 J_m(\kappa r) e^{j\beta z} e^{jm\phi} \text{ for r} \leqslant a \\ C_2 K_m(\gamma r) e^{j\beta z} e^{jm\phi} \text{ for r} > a. \end{cases} \tag{5.16}$$

We can repeat the same argument to $H_r(r)$, thus obtaining:

$$H_r(r) = \begin{cases} C_3 J_m(\kappa r) e^{j\beta z} e^{jm\phi} \text{ for r} \leqslant a \\ C_4 K_m(\gamma r) e^{j\beta z} e^{jm\phi} \text{ for r} > a. \end{cases} \tag{5.17}$$

To determine the four constants, it is necessary to impose, in general, the interface conditions at $r = a$ for the tangential components of the two fields, E_z, E_ϕ, H_z, and H_ϕ. But in this case, the geometric hypothesis of mild guide allow to use (5.15) instead of the tangential continuity, obtaining a system of four equations in four unknowns like:

$$\begin{cases} E_r^{core}(r=a) = E_r^{cladding}(r=a) \\ \left.\dfrac{dE_r^{core}}{dr}\right|_{r=a} = \left.\dfrac{dE_r^{cladding}}{dr}\right|_{r=a} \\ H_r^{core}(r=a) = H_r^{cladding}(r=a) \\ \left.\dfrac{dH_r^{core}}{dr}\right|_{r=a} = \left.\dfrac{dH_r^{cladding}}{dr}\right|_{r=a}. \end{cases} \tag{5.18}$$

From this system we obtain an algebraic linear system of the form $\underline{\underline{A}}\ \underline{C} = \underline{0}$, where $\underline{\underline{A}}$ is a coefficient matrix and $\underline{C}$ is the vector containing the four unknown constants. This system has a solution only if the determinant of the coefficient matrix is equal to zero. Imposing this condition, the expression of the propagation constant $\beta = \beta(a, m, k_0, n_1, n_2)$ will be found, which depends on values set by the geometry - a - and by the type of signal - k_0 - as well as on the materials used, that is n_1 and n_2. The only parameter free is m, which indicates the azimuthal dependence of the signal profile. Then, $\beta = \beta(m) = \beta_{mn}$ with $n = 1, 2, ...\forall m$. Once β_{mn} is found, the mode is determined.

In the literature, the modes are written using the following notation:

- HE_{mn} in the case where either one of the magnetic field components prevails;
- EH_{mn} in the case where either one of the electric field components prevails.

Once the hypothesis (5.14) and (5.15) are satisfied, the found modes are not precisely the ones listed above, but their degenerate linear combination. As a linear combination of mode is still a mode, the latter ones are called *Linear Polarized Modes*, whose notation is LP_{mn}. For our future interests, only the LP_{01} mode will be considered , as it is the only one to always be supported by a single-mode guide. Let us note that $m = 0$ implies an invariance with respect to an azimuth angle ϕ.
Given the difficulty in dealing both with the Bessel functions and with using them to find the basic parameters of the modes, an approximate expression of the LP_{01} mode and of β is used often.
In particular, the chosen function for this mode is:

$$\begin{cases} E_r(r,z) = E_0 e^{-(W_0/r)^2} e^{j\beta z} \\ W_0 = a\ (0,65 + 1,619V^{-3/2} + 2.879V^{-6}) \\ V = \sqrt{X^2 + Y^2} \\ X = \kappa a \\ Y = \gamma a, \end{cases} \tag{5.19}$$

that is the product of a Gaussian function and a complex exponential function. For the eigenvalue β, the chosen function is:

$$\beta = \omega\sqrt{\mu_0\epsilon_0}\ (n_2 + (n_1 - n_2)(1.1428 - 0.996/V)^2). \tag{5.20}$$

These two approximations, (5.19) and (5.20), will be used within the numerical model to describe the source applied to the tapered fiber.

5.3 Problem formulation

In this section, the equations in strong and weak formulation will be obtained under simplifying assumptions. These assumptions are:

- tridimensional problem: the reference system is Cartesian, with the axis $\hat{\mathbf{z}}$ placed along the longitudinal direction of the fiber. The axis $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are orthogonal to each other as to satisfy $\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}$. The central system of reference is placed at the coordinate O(0,0,0) as shown in Figure 5.4;

- dielectric media: a dielectric media has the property of not being able to conduct charges, that is the absence of both currents and charge distribution. The consequence is that both the term σ and the term ρ in Maxwell's equations vanish. This assumption is valid only for the fiber core and the air. This implies that:

$$\begin{cases} \nabla \times \nabla \times \tilde{\mathbf{E}} = \nabla(\nabla \cdot \tilde{\mathbf{E}}) - \nabla^2 \tilde{\mathbf{E}} \\ \nabla \cdot \tilde{\mathbf{E}} = \tilde{\rho} = 0 \end{cases} \quad \Rightarrow \quad \nabla \times \nabla \times \tilde{\mathbf{E}} = -\nabla^2 \tilde{\mathbf{E}}; \tag{5.21}$$

- metallic media: in general, metallic materials allow the free charges to flow inside them, so a charge distribution ρ and an electric current are allowed in order to have $\sigma \neq 0$. The characteristic time necessary for the localized charges to be distributed is approximately 10^{-18} seconds, which is much smaller than the characteristic time of the phenomena discussed here. So this leads to considering $\rho = 0$, with the simplification in (5.21). Moreover, the term $j\omega\mu\sigma\tilde{\mathbf{E}}$ implies the possibility to define a complex value for the electric permittivity, which will be defined as:

$$\hat{\epsilon} = \epsilon - j\frac{\sigma}{\omega}, \tag{5.22}$$

from which, the complex relative dielectric permittivity can be defined:

$$\hat{\epsilon}_r = \epsilon_r - j\frac{\sigma}{\omega\epsilon_0} = \epsilon_r' + j\epsilon_r''. \tag{5.23}$$

Thanks to (5.23), it is possible to introduce the complex refractive index, $\hat{n}$, given by:

$$\hat{n} = \sqrt{\mu_r \hat{\epsilon}_r} = \sqrt{\mu_r \epsilon_r - j\frac{\mu_r \sigma}{\omega \epsilon_0}} = n + j\kappa. \tag{5.24}$$

We can remark that, if we set $\mu_r = 1$, an assumption we will consider in the continuation of the thesis, we have:

$$\hat{\epsilon} = \hat{n}^2 = n^2 - \kappa^2 + j2n\kappa. \tag{5.25}$$

Therefore the electric permittivity will be a function of n and κ.
The electromagnetic quantities, σ, ϵ_r, and μ_r are all functions of the frequency ω of the signal. For this reason, equation (5.24) is used to describe the behavior of *damping* materials - for a complex refractive index implies a damped trend of the electric field within these materials - and *dispersive* materials, because n and κ depend on ω so as to generate a phase shift between the forcing signal and the system response signal. This assumption only holds for the core and cladding.

- sources absence: since there is no current densities as sources, the term is canceled, then:

$$\tilde{\mathbf{J}}_N = \mathbf{0}; \tag{5.26}$$

- single wavelength: an exciting electric field having a unique wavelength in spectrum is taken into consideration. As a result, there will be a unique known angular frequency, ω, characterizing the problem. Moreover, the relation among the wavenumber and ω is obtained from (1.77). As for the impedance, the equation (1.81) becomes:

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}; \tag{5.27}$$

- imposed electrical field: it is applied to the base of the core and cladding, on the boundary of the type S_1^{fb}. In particular, only the component in the $\hat{\mathbf{y}}$ direction is fixed. It is assumed that the profile of $\tilde{E}_y$ corresponds to equation (5.19). This is suitable because, if the two reference systems have their origin and axis $\hat{\mathbf{z}}$ coincident, the $\tilde{E}_r$ component in cylindrical coordinates and the $\tilde{E}_y$ component in Cartesian coordinates are coplanar

and thus can be overlapped for a particular angle of rotation ϕ. The choice of the unique component $\tilde{E}_y$ implies that the signal travels only in the $\hat{\mathbf{z}}$ direction, thanks to the relation (1.71);

- *Silver-Muller* condition: it is applied on the border S_2^∞, since the problem is modeled as an open domain. The condition can be applied only to the outer domain, in this case the air.

Equation (1.67) is taken into account along with both the initial and the boundary conditions. For each element having different electromagnetic characteristics, a version will be written. Following the guidelines provided in Chapter 2, the problem will be written in the frequency domain and thanks to equations (5.21), (5.26), and (5.27), it will be simplified.
In the domain of integration, there are three different materials. We can assume that $V = V_f \cup V_g \cup V_a$, where V_f is the volume occupied by the fiber, V_g is the volume occupied by the gold coating and V_a is the volume occupied by the air. Moreover, there are two separation surfaces between the domains, $S_1^{f/g}$ relative to the interface between fiber and gold, and $S_1^{g/a}$ relative to the interface between gold and air. From now, for each physical quantity relating to one of the domains, the corresponding subscript will be associated.
For the fiber, the problem in strong form will be:

$$\begin{cases} -\nabla^2\tilde{\mathbf{E}}_f - \mu_f\epsilon_f\omega^2\tilde{\mathbf{E}}_f = \mathbf{0} \quad \text{in } V_f \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_f = \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_g \quad \text{on } S_1^{f/g} \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_f = \tilde{E}_y\hat{\mathbf{y}} \quad \text{on } S_1^f; \end{cases} \tag{5.28}$$

for the gold coating, the problem in strong form will be:

$$\begin{cases} -\nabla^2\tilde{\mathbf{E}}_g - \mu_g\hat{\epsilon}_g\omega^2\tilde{\mathbf{E}}_g = \mathbf{0} \quad \text{in } V_g \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_g = \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_f \quad \text{on } S_1^{f/g} \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_g = \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_a \quad \text{on } S_1^{g/a}; \end{cases} \tag{5.29}$$

and for the air, the problem in strong form will be:

$$\begin{cases} -\nabla^2\tilde{\mathbf{E}}_a - \mu_a\epsilon_a\omega^2\tilde{\mathbf{E}}_a = \mathbf{0} \quad \text{in } V_a \\ \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_a = \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_g \quad \text{on } S_1^{g/a} \\ \hat{\mathbf{n}} \times \nabla \times \tilde{\mathbf{E}}_a = -j\omega\dfrac{\mu_0}{Z_0}(\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_a) \quad \text{on } S_2^\infty. \end{cases} \tag{5.30}$$

Let us introduce the function space, H(curl,V), as defined in Chapter 1.

The weak formulation (2.5) is taken into account, applying the hypothesis of the problem. The weak formulation linked to the problem is:

$$\begin{aligned}
&\text{find } \tilde{\mathbf{E}}(\mathbf{r}) = (\tilde{\mathbf{E}}_f(\mathbf{r}), \tilde{\mathbf{E}}_g(\mathbf{r}), \tilde{\mathbf{E}}_a(\mathbf{r})) \in \text{H(curl,V) such that} \\
&\int_{V_f} (\nabla \times \mathbf{v}) \cdot (\nabla \times \tilde{\mathbf{E}}_f) - \mathbf{v} \cdot (\omega^2 \epsilon_f \mu_f \tilde{\mathbf{E}}_f) \\
&+ \int_{V_g} (\nabla \times \mathbf{v}) \cdot (\nabla \times \tilde{\mathbf{E}}_g) - \mathbf{v} \cdot (\omega^2 \hat{\epsilon}_g \mu_g \tilde{\mathbf{E}}_g) \\
&+ \int_{V_a} (\nabla \times \mathbf{v}) \cdot (\nabla \times \tilde{\mathbf{E}}_a) - \mathbf{v} \cdot (\omega^2 \epsilon_a \mu_a \tilde{\mathbf{E}}_a) \\
&+ \int_{S_{2a}^{\infty}} (-j\omega \frac{\mu_0}{Z_0} (\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_a)) \cdot \mathbf{v} \\
&= \mathbf{0} \quad \forall \mathbf{v} \in \text{ H(curl,V).}
\end{aligned} \tag{5.31}$$

The involved bilinear forms,

$$a_f(\tilde{\mathbf{E}_f}, \mathbf{v}) = \int_{V_f} (\nabla \times \mathbf{v}) \cdot (\nabla \times \tilde{\mathbf{E}}_f) - \mathbf{v} \cdot (\omega^2 \hat{\epsilon}_f \mu_f \tilde{\mathbf{E}}_f), \tag{5.32}$$

$$a_g(\tilde{\mathbf{E}_g}, \mathbf{v}) = \int_{V_g} (\nabla \times \mathbf{v}) \cdot (\nabla \times \tilde{\mathbf{E}}_g) - \mathbf{v} \cdot (\omega^2 \epsilon_g \mu_g \tilde{\mathbf{E}}_g), \tag{5.33}$$

$$\begin{aligned}
a_a(\tilde{\mathbf{E}}_a, \mathbf{v}) = &\int_{V_a} (\nabla \times \mathbf{v}) \cdot (\nabla \times \tilde{\mathbf{E}}_a) - \mathbf{v} \cdot (\omega^2 \epsilon_a \mu_a \tilde{\mathbf{E}}_a) \\
&+ \int_{S_{2a}^{\infty}} (-j\omega \frac{\mu_0}{Z_0} (\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \tilde{\mathbf{E}}_a)) \cdot \mathbf{v},
\end{aligned} \tag{5.34}$$

and the functional

$$f(\mathbf{v}) = \mathbf{0}, \tag{5.35}$$

are introduced.
Now, equation (5.31) can be rewritten as:

$$\begin{aligned}
&\text{find } \tilde{\mathbf{E}}(\mathbf{r}) = (\tilde{\mathbf{E}}_f(\mathbf{r}), \tilde{\mathbf{E}}_g(\mathbf{r}), \tilde{\mathbf{E}}_a(\mathbf{r})) \in \text{H(curl,V) such that} \\
&a_f(\tilde{\mathbf{E}}_f, \mathbf{v}) + a_g(\tilde{\mathbf{E}}_g, \mathbf{v}) + a_a(\tilde{\mathbf{E}}_f, \mathbf{v}) = f(\mathbf{v}) \quad \forall \mathbf{v} \in \text{ H(curl,V).}
\end{aligned} \tag{5.36}$$

The continuity and the coercivity of the three bilinear forms (5.32), (5.33), (5.34) and the continuity of the functional (5.35) can be proved easily. Thanks to this, the Lax-Milgram theorem is satisfied and the uniqueness and existence of the solution $\tilde{\mathbf{E}}(\mathbf{r}) = (\tilde{\mathbf{E}}_f(\mathbf{r}), \tilde{\mathbf{E}}_g(\mathbf{r}), \tilde{\mathbf{E}}_a(\mathbf{r}))$ is guaranteed. The unknown field $\tilde{\mathbf{E}}$ will be indicated as $\mathbf{E}$ to make the notation lighter.

5.4 Affine mapping

The affine mapping method is required as a part of the reduction method. In particular, when the variation of one or more parameters involves the modification of the starting geometry (real domain $\hat{V}$), the POD method requires that the DoFs number remains the same and that they are always associated with the same element of the mesh for each simulation. In order to implement the method, we need a supporting geometry (reference domain V) which will form the computational domain in which the governing equations will be solved, varying the parameters. The affine mapping allows us to find the geometric relationship - through coefficients depending on the parameters $\underline{\mu}$ - between the real geometry and the reference geometry. In this way, many different problems can be solved by simply changing these coefficients, while maintaining the same geometry.
With regard to the problem of the optical fiber, we know that the real domain $\hat{V}$ is the union of $R = 3$ different tridimensional domains - $d = 3$ - having different electromagnetic characteristics, so $\hat{V} = \hat{V}_f \cup \hat{V}_g \cup \hat{V}_a = \bigcup_{r=1}^{R} \hat{V}_r$.
Let us define:

$$\underline{x} = \mathcal{G}^r(\underline{\hat{x}}; \mu) = \underline{\underline{G}}^r(\mu)\underline{\hat{x}} + \underline{g}^r(\mu) \quad \text{with } 1 \leqslant r \leqslant R, \tag{5.37}$$

whence

$$\nabla(\underline{\hat{x}}) = \frac{\partial}{\partial \hat{x}_i} = \frac{\partial x_j}{\partial \hat{x}_i}\frac{\partial}{\partial x_j} = G^r_{ji}(\mu)\frac{\partial}{\partial x_j} = \underline{\underline{G}}^r(\mu)\nabla(\underline{x}), \tag{5.38}$$

where $\underline{x} \in V$, $\underline{\hat{x}} \in \hat{V}$, $\underline{\underline{G}}^r(\mu) \in \mathbb{R}^{dxd}$ is a piecewise constant matrix, $\underline{g}^r(\mu) \in \mathbb{R}^d$ is a piecewise constant vector, and $\mathcal{G}^r(\mu)$ is a piecewise affine geometric mapping.
Let us define the function space $H(\text{curl}, V) = \hat{H}(\text{curl}, \mathcal{G}^{-1}(\mu, V)) = \hat{H}(\text{curl}, \hat{V})$ and for every function $\hat{v} \in \hat{H}$, let us define $v \in H$ such that $v(\underline{x}) = \underline{\hat{v}}(\mathcal{G}^{-1}(\mu; \underline{x}))$.
Let us also define:

$$\frac{d\hat{V}}{dV} = |\underline{\underline{G}}(\mu)^{-1}| \quad \text{and} \quad \frac{d\hat{S}}{dS} = |\underline{\underline{G}}(\mu)^{-1}e^t|, \tag{5.39}$$

where e^t is the unit vector tangent to the generic boundary surface S. Let us consider equation (5.31). Every addend can be written as:

- $$\begin{aligned}&\int_{\hat{V}}(\nabla\times\hat{\mathbf{v}})\cdot(\nabla\times\hat{\mathbf{E}})-\hat{\mathbf{v}}\cdot(\mu\epsilon\omega^2\hat{\mathbf{E}})d\hat{V}=\\&=\sum_{r=1}^{R}\int_{V^r}(\nabla\times\hat{\mathbf{v}})\cdot(\nabla\times\hat{\mathbf{E}})-\hat{\mathbf{v}}\cdot(\mu^r\epsilon^r\omega^2\hat{\mathbf{E}})\;d\hat{V}^r;\end{aligned}\tag{5.40}$$

- $$\begin{aligned}&\int_{\hat{S}_2^\infty}(\hat{\mathbf{n}}^r\times\hat{\mathbf{v}})j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}^r\times\hat{\mathbf{E}})d\hat{S}=\\&=\sum_{r=1}^{R}\int_{\hat{S}_2^{\infty r}}(\hat{\mathbf{n}}^r\times\hat{\mathbf{v}})j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}^r\times\hat{\mathbf{E}})\;d\hat{S}^r.\end{aligned}\tag{5.41}$$

After these considerations, we can see that $a(\hat{\mathbf{E}},\hat{v})=a(\mathbf{E},v)$, so, applying (5.39) and (5.38), each addend can be written as:

- $$\begin{aligned}&\sum_{r=1}^{R}\int_{V^r}(\nabla\times\hat{\mathbf{v}})\cdot(\nabla\times\hat{\mathbf{E}})-\hat{\mathbf{v}}\cdot(\mu^r\epsilon^r\omega^2\hat{\mathbf{E}})\;d\hat{V}^r=\\&=\sum_{r=1}^{R}\int_{V^r}\Big\{(\underline{\underline{G}}^r(\mu)\nabla\times\mathbf{v})\cdot(\underline{\underline{G}}^r(\mu)\nabla\times\mathbf{E})\Big\}|\underline{\underline{G}}^r(\mu)^{-1}|\;dV\\&-\sum_{r=1}^{R}\int_{V^r}\Big\{\mathbf{v}\cdot(\mu^r\epsilon^r\omega^2\mathbf{E})\Big\}|\underline{\underline{G}}^r(\mu)^{-1}|\;dV;\end{aligned}\tag{5.42}$$

- $$\begin{aligned}&\sum_{r=1}^{R}\int_{\hat{S}_2^{\infty r}}(\hat{\mathbf{n}}^r\times\hat{\mathbf{v}})j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}^r\times\hat{\mathbf{E}})\;d\hat{S}^r=\\&=\sum_{r=1}^{R}\int_{\hat{S}_2^{\infty r}}\Big\{(\hat{\mathbf{n}}^r\times\mathbf{v})j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}^r\times\mathbf{E})\Big\}|\underline{\underline{G}}^r(\mu)^{-1}e^t|\;dS.\end{aligned}\tag{5.43}$$

Consequently, the weak form is written as:

$$\begin{aligned}&\text{find }\mathbf{E}(\mathbf{r})\in\mathrm{H(curl},V)\text{ such that}\\&\sum_{r=1}^{R}\int_{V^r}(\nabla\times\mathbf{v})^T\underline{\underline{G}}^r(\mu)^T|\underline{\underline{G}}^r(\mu)^{-1}|\underline{\underline{G}}^r(\mu)(\nabla\times\mathbf{E})\;dV\\&-\sum_{r=1}^{R}\int_{V^r}\Big\{\mathbf{v}\cdot(\mu^r\epsilon^r\omega^2\mathbf{E})\Big\}|\underline{\underline{G}}^r(\mu)^{-1}|\;dV\\&+\sum_{r=1}^{R}\int_{\hat{S}_2^{\infty r}}\Big\{(\hat{\mathbf{n}}^r\times\mathbf{v})j\omega\frac{\mu_0}{Z_0}(\hat{\mathbf{n}}^r\times\mathbf{E})\Big\}|\underline{\underline{G}}^r(\mu)^{-1}e^t|\;dS\\&\forall\mathbf{v}\in\mathrm{H(curl},V).\end{aligned}\tag{5.44}$$

Now, the 9 coefficients of $\underline{\underline{G}}(\mu)$ and the 3 coefficients of $\underline{g}(\mu)$ will have to be found for a practical use of (5.44). This implies that we will have to find 12 relations that will allow us to univocally find the coefficients.

5.4.1 Geometry mapping: cone and cladding

Given the shape of the optical fiber shown in Figure 5.5, the geometric element chosen for the modeling of the core was the cone, having base radius R_B, half-angle at tip θ and height h. Instead, the cladding was modeled with a hollow cone, having inner radius R_B thickness t and height h'.

Cone

Let us consider a geometric element that models the fiber, namely a cone. In this thesis work $\mu = \theta$ is taken into account as one of geometrical parameter that varies during simulations.

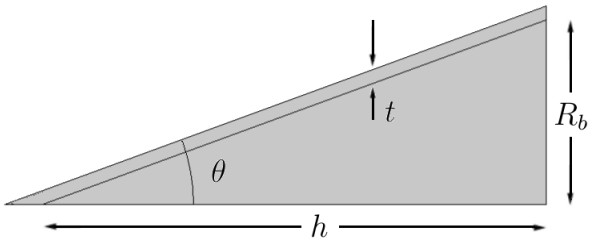

FIGURE 5.5. Section of the fiber tip and geometric quantities, with θ the only parameter used in this thesis.

From purely trigonometric considerations, it is possible to derive the relation between the height h as a function of the base radius R_b and the half angle at the tip θ:

$$h = \frac{R_b}{\tan(\theta)}. \tag{5.45}$$

Let us consider as reference geometry a cone having as a base radius R_b^{ref} and as height $h^{ref} = R_b^{ref} / \tan(\theta^{ref})$, while as real geometry a cone having a base radius R_b^{real} and as height $h^{real} = R_b^{real} / \tan(\theta^{real})$. Since it is a three-dimensional object, we need 4 points in order to find the 12 equations to solve the 12 unknown coefficients. These points are not chosen at random: in fact, three of them will be chosen on the same plane, to identify the base, while a fourth one will be chosen on the tip of the cone to set the height.

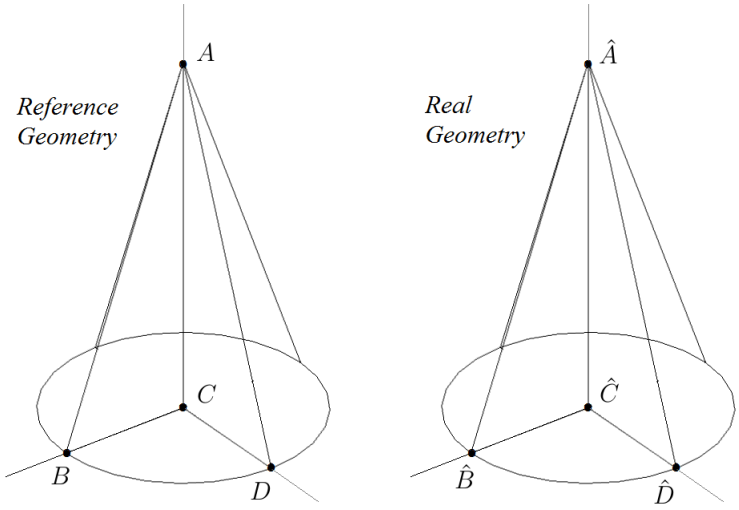

FIGURE 5.6. Core reference geometry (left) and the real one (right) with the points used for the affine mapping.

The chosen points are (Figure 5.6):

$$\begin{cases} A^{ref} = A = (0, 0, R_b^{ref} / \tan(\theta^{ref})) \\ B^{ref} = B = (R_b^{ref}, 0, 0) \\ C^{ref} = C = (0, 0, 0) \\ D^{ref} = D = (0, R_b^{ref}, 0) \end{cases}$$

and (5.46)

$$\begin{cases} A^{real} = \hat{A} = (0, 0, R_b^{real} / \tan(\theta^{real})) \\ B^{real} = \hat{B} = (R_b^{real}, 0, 0) \\ C^{real} = \hat{C} = (0, 0, 0) \\ D^{real} = \hat{D} = (0, R_b^{real}, 0), \end{cases}$$

for the reference geometry and the real geometry, respectively.
For the generic point P, writing out (5.37), we obtain:

$$\begin{cases} x_P = g_1 + G_{11}\hat{x}_P + G_{12}\hat{y}_P + G_{13}\hat{z}_P \\ y_P = g_2 + G_{21}\hat{x}_P + G_{22}\hat{y}_P + G_{23}\hat{z}_P \\ z_P = g_3 + G_{31}\hat{x}_P + G_{32}\hat{y}_P + G_{33}\hat{z}_P. \end{cases} \tag{5.47}$$

Applying (5.47) to every point of (5.46), we can obtain a linear system of the form:

$$\underline{c} = \underline{\underline{A}}^{-1}(\theta)\ \underline{b}, \tag{5.48}$$

where $\underline{c} \in \mathbb{R}^{12x1}$ is the vector of the unknown coefficients, $\underline{\underline{A}}^{-1}(\mu) \in \mathbb{R}^{12x12}$ is the inverse matrix of the known coefficients depending on the geometrical parameter and $\underline{b} \in \mathbb{R}^{12x1}$. The inverse matrix of $\underline{\underline{A}}(\theta)$ is not unique because the selected points are not all coplanar ([10], [29]). From this operation, we derive:

$$\underline{\underline{G}}_f(\theta) = \begin{bmatrix} \frac{R_b^{ref}}{R_b^{real}} & 0 & 0 \\ 0 & \frac{R_b^{ref}}{R_b^{real}} & 0 \\ 0 & 0 & \frac{R_b^{ref}}{R_b^{real}} \frac{\tan(\theta^{real})}{\tan(\theta^{ref})} \end{bmatrix} \quad \text{and} \quad \underline{g}_f(\theta) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (5.49)$$

Cladding

Now, let us consider the geometric element that represents the gold cladding. Even in this case, from purely trigonometric considerations, it is possible to derive the relation that binds the height h' as a function of the radius at the base R_b, of the thickness t of the cladding and of the half angle at the tip θ:

$$h' = \frac{R_b + t}{\tan(\theta)}. \quad (5.50)$$

Let us consider as a reference geometry a cladding having $R_b^{ref} + t^{ref}$ as a radius at the base and $h^{ref} = (R_b^{ref} + t^{ref})/\tan(\theta^{ref})$ as a height, while as a real geometry we will consider a cladding having $R_b^{real} + t^{real}$ as a radius at the base and $h^{real} = (R_b^{real} + t^{real})/\tan(\theta^{real})$ as a height. Even in this case, we will need 4 points in order to complete the mapping.

The chosen points are (Figure 5.7):

$$\begin{cases} A^{ref} = A = (0, 0, (R_b^{ref} + t^{ref})/\tan(\theta^{ref})) \\ B^{ref} = B = (R_b^{ref} + t^{ref}, 0, 0) \\ C^{ref} = C = (0, 0, 0) \\ D^{ref} = D = (0, R_b^{ref} + t^{ref}, 0) \end{cases}$$

and (5.51)

$$\begin{cases} A^{real} = \hat{A} = (0, 0, (R_b^{real} + t^{real})/\tan(\theta^{real})) \\ B^{real} = \hat{B} = (R_b^{real} + t^{real}, 0, 0) \\ C^{real} = \hat{C} = (0, 0, 0) \\ D^{real} = \hat{D} = (0, R_b^{real} + t^{real}, 0), \end{cases}$$

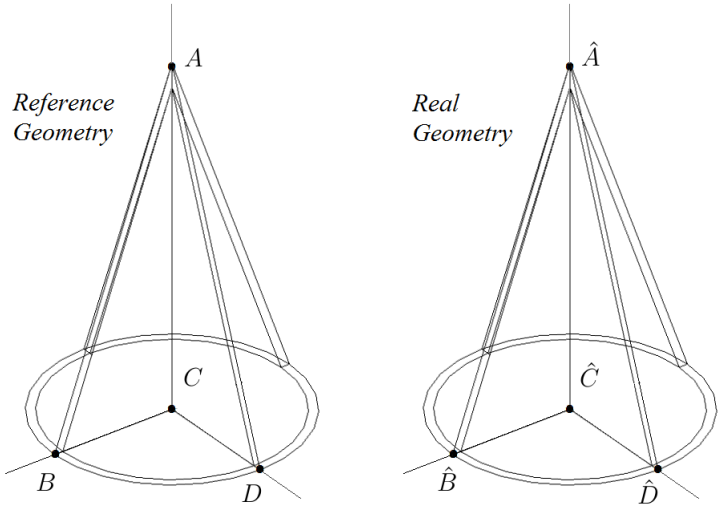

FIGURE 5.7. Cladding reference geometry (left) and the real one (right) with the points used to affine mapping.

for the reference geometry and the real geometry, respectively.
Proceeding in the same way as for the cone, the unknown coefficients will provide:

$$\underline{\underline{G}}_g(\theta) = \begin{bmatrix} \dfrac{R_b^{ref} + t^{ref}}{R_b^{real} + t^{real}} & 0 & 0 \\ 0 & \dfrac{R_b^{ref} + t^{ref}}{R_b^{real} + t^{real}} & 0 \\ 0 & 0 & \dfrac{R_b^{ref} + t^{ref}}{R_b^{real} + t^{real}} \dfrac{\tan(\theta^{real})}{\tan(\theta^{ref})} \end{bmatrix} \tag{5.52}$$

and

$$\underline{g}_g(\theta) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

5.5 Numerical results

Let us consider the final part of an optical fiber consisting of SiO_2 core and an Au cladding. The terminal part is cone-shaped, with a radius of $R_b = 0,85 \cdot 10^{-6}$ m, a semi-corner at the point $\theta = 20°$ and a height $h = R_b/\tan\theta$. The cladding, which is the coating that covers the whole cone, has a thickness of $t = 0.15 \cdot 10^{-6}$ m. Lastly, the volume of the air in which

the fiber and the coating are immersed, is parallelepiped-shaped, having as a basis a square of side $l = 2.5 \cdot 10^{-6}$ m, and a height equal to $h = 7.45 \cdot 10^{-6}$ m. The domain views are shown in Figure 5.8
The exciting signal to the fiber is expressed from approximation (5.19) and (5.20), with angular frequency $\omega = 1.79 \cdot 10^{15}$rad/s and electric field intensity $E_0 = 1$ V/m.
Since the domains consist of different materials, the electromagnetic characteristics will be different too. In Table 5.1 the values of n, k, μ_r and ϵ_r are shown for every material.

	Fiber(SiO_2)	**Cladding**(Au)	**Air**
n	1.4498	0.2724	1
κ	0	6.3390	0
μ_r	1	1	1
ϵ_r	2.1021	-40.1087+j3.4535	1

TABLE 5.1. Electromagnetic properties of materials for $\omega = 1.79 \cdot 10^{15}$rad/s. In particular, the air medium was modeled with the vacuum electromagnetic properties.

The problem solved in strong formulation is the union of (5.28), (5.29) and (5.30), having chosen as a boundary S_1^{fb} the basis of the fiber and the cladding, and as a boundary S_2^{∞} the outer walls of the parallelepiped. In the first case, the electric field has been imposed, in the second the Silver-Muller condition.
The problem in weak formulation (5.44) has been solved, having chosen approximating polynomial functions of degree $N = 1$, in order to obtain a reduced computational cost for the computer used. Despite the polynomial degree is the minimum possible, the solution will retain the main typical characteristics of the problem, useful to show the accuracy of the constructed model. The linear algebraic problem is characterized by 725547 DoFs and the Generalized Minimal RESidual method (GMRES) [33] with a Symmetric Successive Overrelaxation (SSOR) preconditioner [5] has been used to find the solution. In Table 5.2 the main numerical characteristics of the problem have been gathered.

As described earlier, we will use an *affine mapping* in order to develop a POD methodology. By following the notation introduced in Chapter 3, the geometric parameter in this numerical model is $\mu = \theta$, while the number of parametric configurations used is $M = 81$. In particular, the set of these parameters is given by $\theta = \{10 : (45 - 10)/(81 - 1) : 45\}$.
The finite element solver used is implemented in COMSOL [7].

Tetrahedron	543300
Surface elements	37531
Line elements	2255
DoFs	725547
Shape function	$\mathbb{P}^1$

TABLE 5.2. Numerical data

5.5.1 Geometry and mesh

The numeric domain is given by the union of three different domains. The mesh used has been automatically obtained by the solver itself (command *Mesh*) after setting some constraints on the type of mesh and on the size of all of the individual elementary volumes (Figure 5.9). An unstructured mesh was chosen for all three domains, while a different setup was used for the dimensions of the individual elements. In Table 5.3 the values used are shown.
As a criterion for selecting the maximum size, *Shannon sampling theorem* [9] was used. Using (4.23) e (4.25), we obtain that in every domain the following relationship must be satisfied $\lambda_c < \lambda_0/(2\sqrt{\mu_r \epsilon_r})$. The results are contained in Table 5.4.

	Fiber(SiO_2)	**Cladding**(Au)	**Air**
Min dimension	30nm	30nm	30nm
Max dimension	150nm	150nm	190.05nm

TABLE 5.3. Maximum and minimum dimensions used for the construction of the mesh elements.

	Fiber(SiO_2)	**Cladding**(Au)	**Air**
λ_c	362nm	296nm	525nm

TABLE 5.4. Maximum dimensions satisfying the *Shannon sampling theorem*.

All the values are abundantly higher than the values chosen for the discretization. We have deliberately chosen a spatial quality greater for the domain, both for the degree of the approximate polynomials chosen and for the three-dimensionality of the problem.

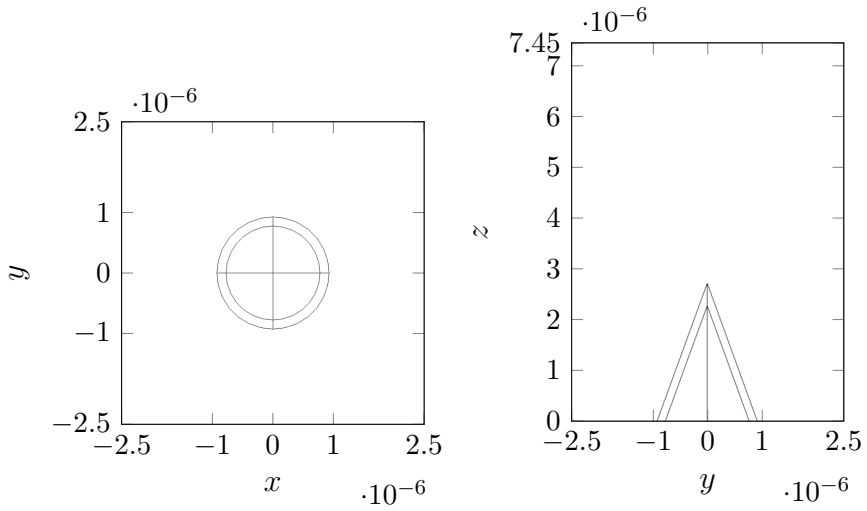

FIGURE 5.8. The top view (on the left) and the side view (on the rigth) of the domain. For the side view, the yz plane for $x = 0$ is shown.

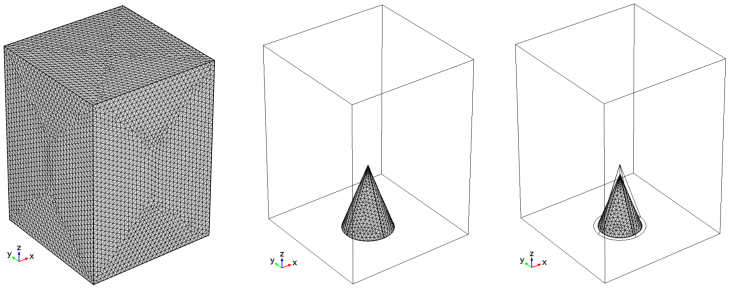

FIGURE 5.9. Meshes obtained for the different domains: from left to right, the air mesh, the cladding mesh and the fiber mesh.

5.5.2 Results, representative visualizations and comparisons

In order to understand whether the model adopted is in line with the literature [22], a simulation with the reference configuration as computational domain will be carried out. The results obtained are shown in Figures 5.10 and 5.11. As regards the view in planes $x - z$ e $y - z$ (Figure 5.10), there are three different scales in order to better appreciate the fields in the different domains. Inside the fiber there is a signal that gradually thickens as we approach the tip, with a consequent increase in the modulus of $\mathbf{E}$. This phenomenon is due to signal reflection every time it meets the interface between core and cladding. The absence of electrical field within the coating is due to the total reflection at the interface of the incident signal - Figure 5.10. In fact, it is

possible to show that the critical angle for n_{core} and $n_{cladding}$ is less than 70°, i.e. the angle between the signal direction and normal vector of the fiber oblique side. In this way it is not possible the passage of a refracted signal. Nevertheless, evanescent wave is generated, characterized by a propagation vector $\mathbf{s} = \mathbf{a} + j\mathbf{k}$. When the wave reaches the interface between the cladding and air, being the refractive index $n_{air} > n_{cladding}$, it generates a refracted wave which preserves both the wavenumber and the attenuation, being the air a dielectric medium. Thus, the damped behavior of the **E** module, caused by the attenuation vector **a**, is explained both in the longitudinal direction z and in transverse x and y directions, as shown in Figures 5.11. We can observe this effect even better thanks to the evaluations in picture 5.12, in which, once an altitude z is fixed, the trend of the electric field in the $x - z$ plane for $y = 0$ and $y - z$ plane for $x = 0$ is shown. These results are in agreement with what was found in the work [22].

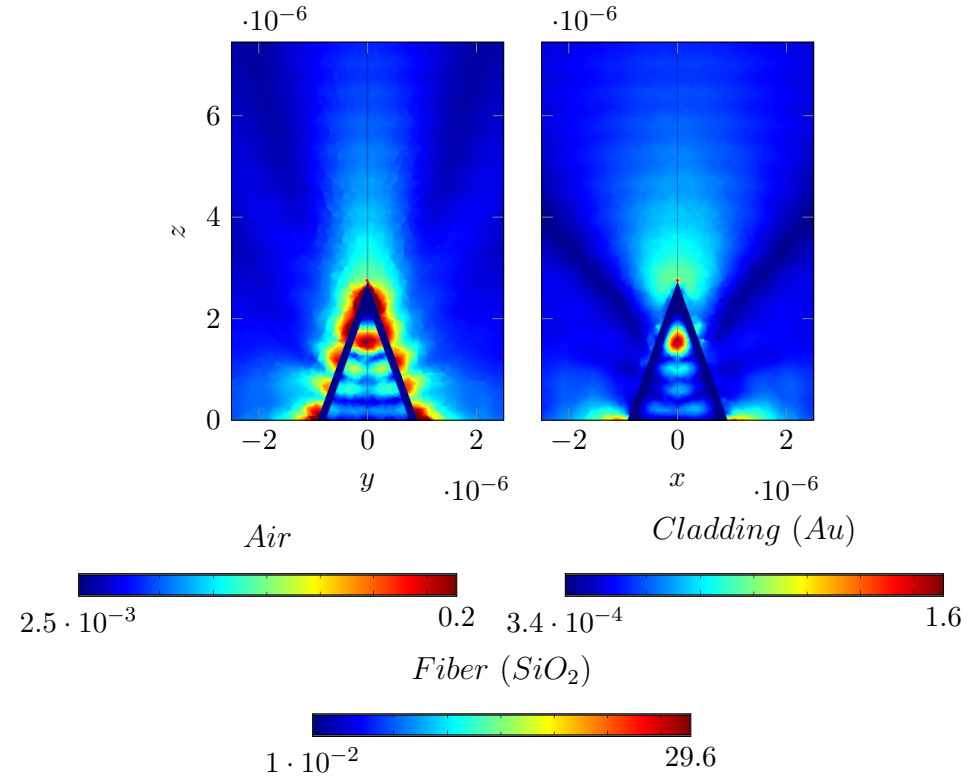

FIGURE 5.10. The **E** module in $y - z$ plane for $x = 0$ (left) and in $x - z$ plane for $y = 0$ (right).

Let us now examine the analysis of the method of POD reduced basis. As we described in Chapter 3, a relatively high number $M = 81$ of simulations have

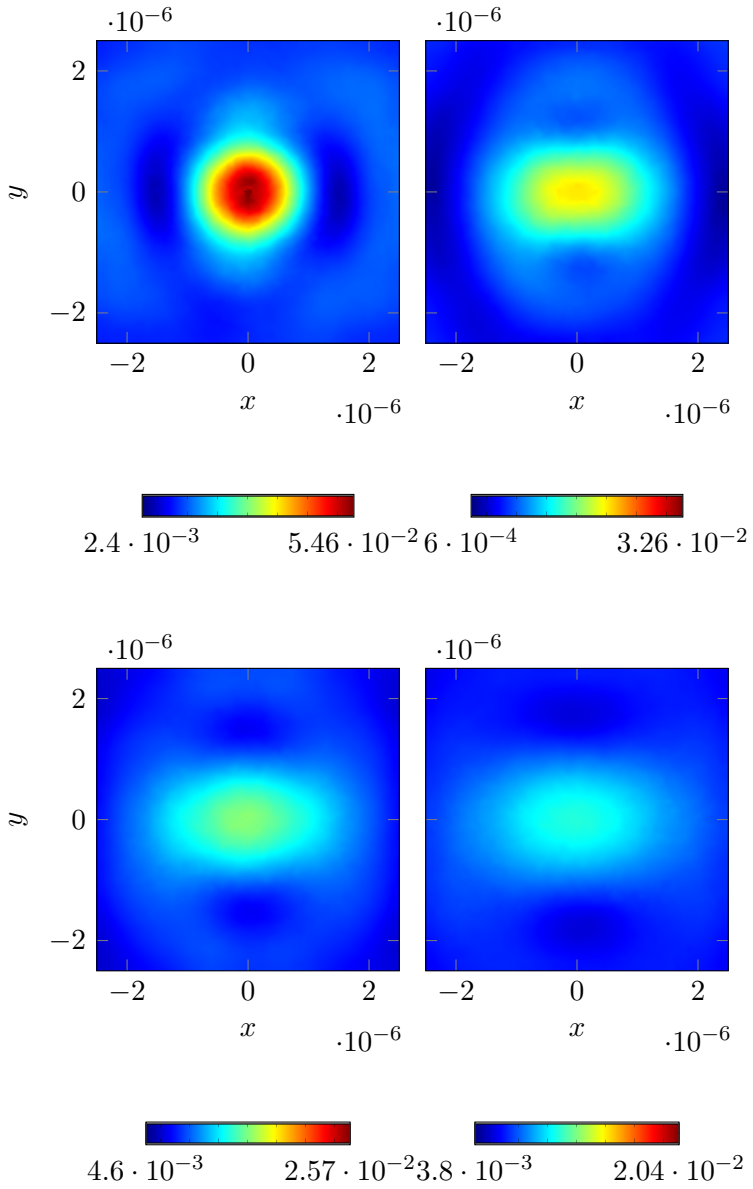

FIGURE 5.11. Section of the field $\mathbf{E}$ at different altitudes z: top left, $z = 3.2 \cdot 10^{-6}$ m; top right $z = 4.2 \cdot 10^{-6}$ m; lower left $z = 5.2 \cdot 10^{-6}$ m; lower right $z = 6.2 \cdot 10^{-6}$ m.

been solved, in order to find the solutions of the problem (5.44). The only parameter used was θ, thus obtaining a number of geometric configurations equal to the number of simulations. To demonstrate the quality and effectivity of the method, the most representative figures will be shown below. In particular:

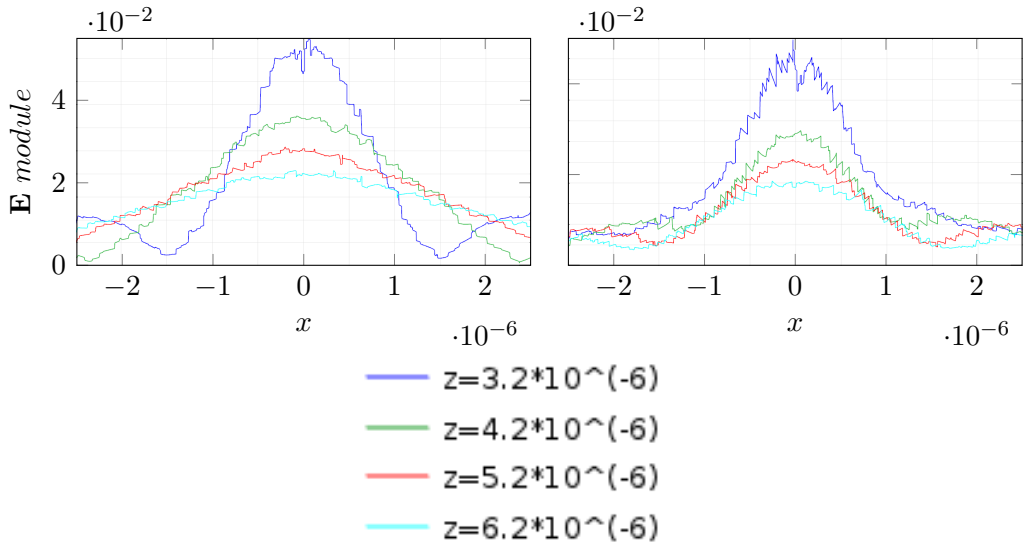

FIGURE 5.12. Field values $\mathbf{E}$ valued along lines passing the center with direction $\hat{\mathbf{x}}$ (left) and direction $\hat{\mathbf{y}}$ for the values of z considered in picture 5.11.

- all the eigenvalues of the correlation matrix C are shown in Figure 5.13. We have highlighted those that have allowed the construction of the new basis. The criterion to select those suitable is based on the relative weight of each eigenvalue. Namely, we adopted $|\lambda_i / \sum_{k=1}^{M} \lambda_k| < 10^{-13}$ for $i = 1, \ldots, M$ as a constraint to satisfy. To take the first N_{rb} eigenvalues means to consider only those having the greater associated energy. In this case, in fact, the elements of C are a form of electromagnetic energy. The diagonal elements are the energies related to each solution, while the extra-diagonal ones are the mutual energies. Consequently, C is real, symmetric and positive definite. Due to that, the spectral representation of the correlation matrix is $\underline{\underline{C}} = \underline{\underline{R}}^T \, \underline{\underline{\Lambda}} \, \underline{\underline{R}}$, with $\underline{\underline{R}} \in \mathbb{R}^{M \times M}$ the matrix of the right eigenvectors and $\underline{\underline{\Lambda}} \in \mathbb{R}^{M \times M}$ the diagonal matrix of the eigenvalues. It is easy to see now that, taking the first N_{rb} eigenvalues and related eigenvectors, the correlation matrix will be composed of the most energy-related solutions;

- the difference between the real solution and the one obtained by POD reconstruction, given by $\Delta\tau = \tau_h - \tau_{POD}$, for a specific configuration in the set μ. Those shown in Figure 5.14, 5.15, 5.16, 5.17 and 5.18 have been chosen in order to cover as equally as possible the set of parameter configurations;

- the relative error given by $Err = \dfrac{\tau_{h_i} - \tau_{POD_i}}{\tau_{h_i}}$ for $i = 1, \ldots, 725547$ for

the same parameter configuration, also shown in Figure 5.14, 5.15, 5.16, 5.17 and 5.18.

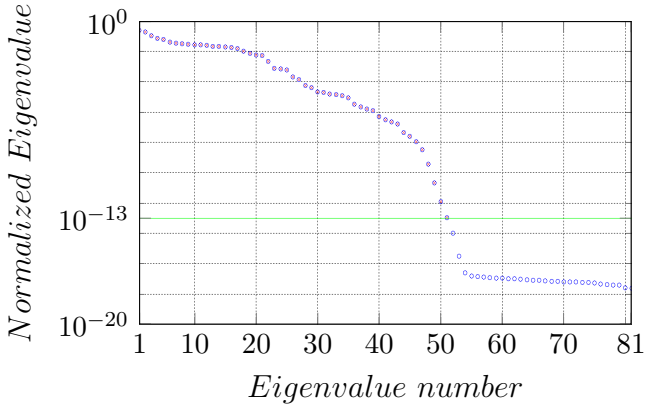

FIGURE 5.13. Eigenvalues of the correlation matrix. The blue circles are all the eigenvalues, while the red crosses indicate only the ones used to construct the reduce basis space. The green line is the tolerance limit. For the further example, the number of basis used to approximate the solution will be $N_{rb} = 51$

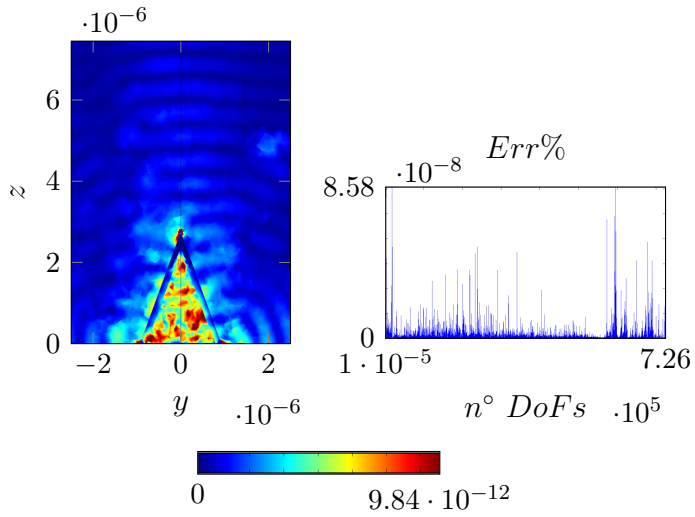

FIGURE 5.14. The left-side figure shows $\Delta\tau$, while the right-side one shows Err; both are obtained by a parameter value of $\theta = 10°$.

The point-wise difference among the real solution and the approximate one is not a clear indicator of the reliability of the POD method. For this

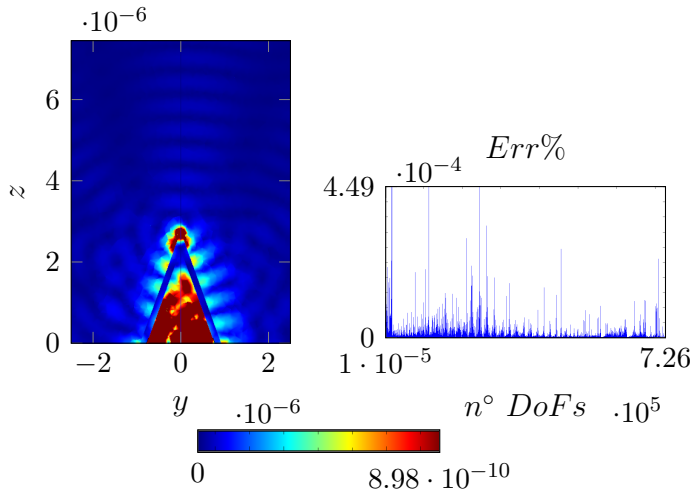

FIGURE 5.15. The left-side figure shows $\Delta\tau$, while the right-side one shows Err; both are obtained by a parameter value of $\theta = 18.75°$.

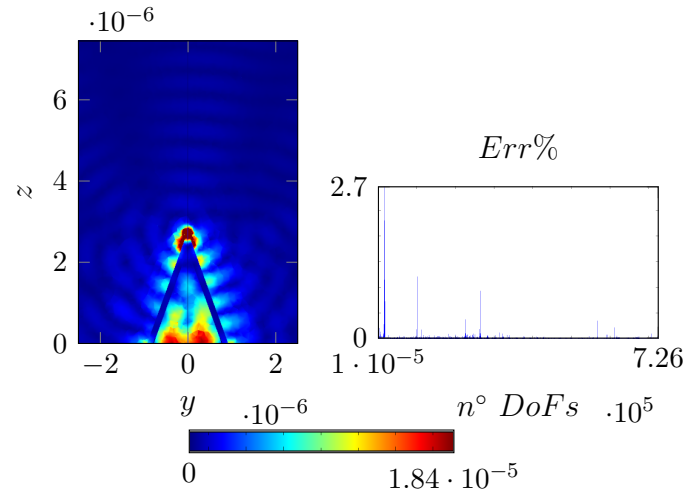

FIGURE 5.16. The left-side figure shows $\Delta\tau$, while the right-side one shows Err; both are obtained by a parameter value of $\theta = 27.5°$.

reason, a more practical and accurate indicator was used: the relative error. Thanks to its definition, all false positives are avoided, especially in those simulations where there are small field values. In fact, in these cases, a small difference value does not necessarily indicate a good approximation of the POD solution if not compared with the value of the real solution.
A solution with reduced basis is acceptable when it does not differ much from the real one. Since it constituted by a subspace of basis functions, it will definitely be affected by an error. In the engineering field, and in general in

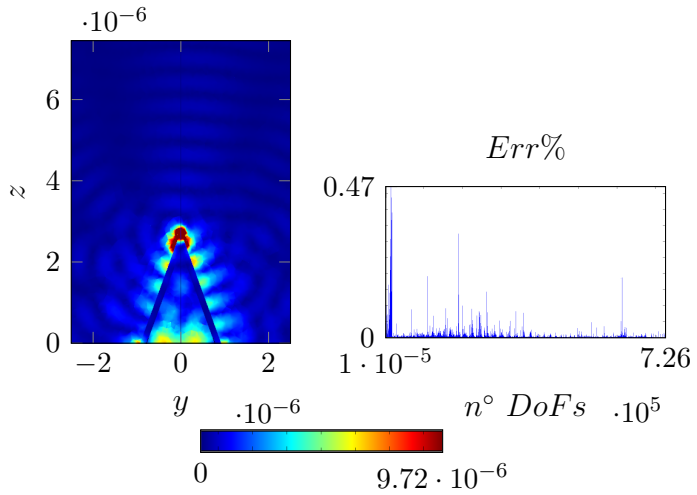

FIGURE 5.17. The left-side figure shows $\Delta\tau$, while the right-side one shows Err; both are obtained by a parameter value of $\theta = 36.69°$.

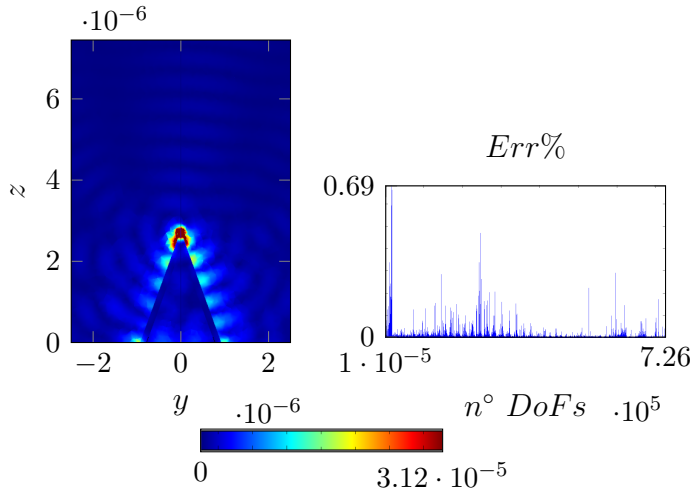

FIGURE 5.18. The left-side figure shows $\Delta\tau$, while the right-side one shows Err; both are obtained by a parameter value of $\theta = 45°$.

practical applications, however, if this error does not exceed a limit, which is normally set between $Err = 2.5\% \div 5\%$, the solution is still acceptable. As we can see from Figures 5.14, 5.15, 5.16, 5.17 and 5.18, the POD method works very well.

5.5.3 Other computational results

As already discussed in Chapter 4, some information about computational efficiency, performances and effectivity of the reduced basis method will be shown.
First of all, the time required to solved the FEM problem (3.24) and the RB one (3.25) for all the parameter of the configurations used in the offline phase will be compared in the left side of Figure 5.19. The average time for the FEM method is $T_{FEM} = 3180\ s$ and for RB one is $T_{RB} = 63\ s$. Even in this case, the ratio among the two times determines the advantage of the RB (POD-based) method, with a saving of 98% on the computation time. This is in agreement with several test performed in literature [14]. It is demonstrated once again the efficiency and the velocity of the online phase with respect to the offline one.
The convergence and the consistency of the reduced basis method is shown in the right side of Figure 5.19. Namely the need for a relatively low basis number in order to build a solution that satisfies the engineering limits previously cited is proved. As mentioned in Chapter 4, increasing the basis functions used, also the maximum and average relative error decrease, reaching a very low values.

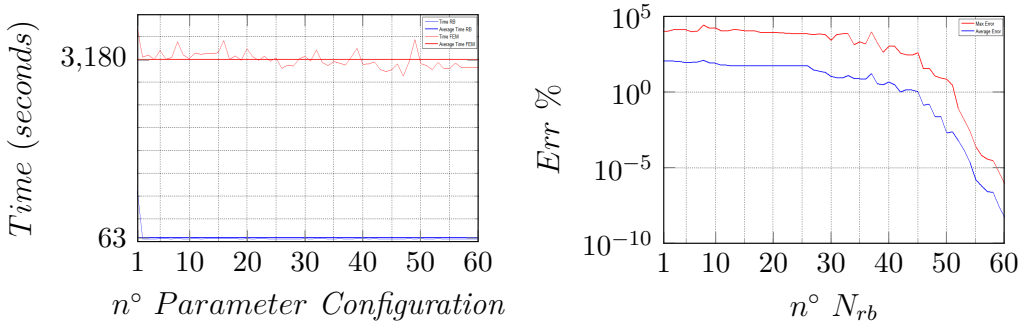

FIGURE 5.19. In the left figure is shown the relation among time spent to solved the FEM (red line) and the RB (blue line) systems. The bold lines are the average times. In the right figure is shown the relation among Maximum/Average relative error (red line and blue line respectively) and the number of basis used to approximate the solution with the RB method.

6 | Conclusions

This aim of the thesis is to develop a reduced order method for parameterized problems, held by partial differential equations, in particular Maxwell's equations.
The process of developing of a reduced model that could approximate in a satisfactory manner the true solution is divided into two parts:

- the *Offline* phase, very heavy from the computational point of view, where, for each parametric configuration chosen, the corresponding algebraic problem has to be solved;
- the *Online* phase, much more efficient and manageable, where, at each change of parametric configuration, the resolution of a new numerical solution is not required, but only the specialization of the stiffness matrix and the input term.

The undoubted advantage lies in computing the *Offline* part only once, while the *Online* one can be performed countless times, without requiring a so relevant waste in terms of time.

The first problem, treated in Chapter 4, concerns the reflection and diffusion of a two-dimensional airfoil hit by a wave, the so-called *scattering problem*. At first, we verified that the solution of a model problem (NACA 0012) was in agreement with the solution found in the literature [23]. Secondly, the problem shifted on the analysis of a more plausible profile from an applicative point of view (NACA 4312). A physical quantity (ω) has been used as parameter and consequently a reduced order method, namely POD has been applied with the comparison between the reconstructed solution and the real solution.
The problem treated in Chapter 5 concerns the study of the electric field generated by an optical fiber. As in the previous case, the first step has been the verification of the adopted model through comparison of the available experiments obtained in [22]. In a second step, the method to the parameterized problem in geometry has been applied thanks to the affine mapping of the

domain. The relative error among the truth and approximate solutions has been calculated in order to verify that its maximum was under a conventional limit (from 2,5 % to a maximum of 5 %).
In the light of the results obtained, we found out that the method of model reduction is in accordance with what has already been demonstrated in the literature ([12], [32], [11], [14]), and it also shows excellent efficiency and accuracy for these two applications specifically. In particular, for a fixed parametric configuration, the reduced order method compared with the finite element method guarantees a computational time saving by approximately 87% for the scattering problem and 98% for the optical fiber problem. This means that in a problem characterized by a considerable number of DoFs, like in the stealth one, the solution is quasi-instantaneous (the order is of seconds). But, when DoFs are many more, such in the optical fiber problem, the solution is computed in minutes instead of hours. In both cases, the use of the method is justified. Moreover, also the convergence was proven. In this sense, for both problems, the worst parametric configuration in term of the maximum relative error has been fixed, and then the approximate solution has been construct varying the number of bases used. The results show that the convergence is obtained with a relative few number of bases, which has a positively influence on computational time saving.

Once the numerical models have been developed and the results demonstrate the reliability and certainty in the use of the method, many more possible developments may arise.
For the problem of the *stealth* airfoil:

- since the parameterization occurred only for a physical variable, shifting focus to a geometric parameterization thanks to the use of more general affine mappings, such as the affine mapping with curved elements [31];

- construction of a 3D model for the wing configuration and addition of other parameters, such as swept angle and dihedral angle, thickness and curvature of the wing and thickness of the paint;

- simultaneous variation of all parameters, either geometrical or physical;

- setting an optimization problem for the wing geometry in order to minimize a single *output*, like the scattered electric field, with a constraint on aerodynamics [23].

For the problem of *tapered optical fiber*:

- simultaneous variation of the geometrical and physical parameters, including the basis radius of the fiber, the cladding thickness and wavelength of the exciting signal;
- given the aim of this specific optical fiber, finding the geometric shape that generates an electric field as localized as possible, by imposing more a constraint a particular *output*, like the profile of the electric field on *far field*.

A | Estratto in lingua italiana

I metodi di *riduzione di modello* ricoprono oggigiorno un ruolo di spicco durante le fasi di progettazione e di ottimizzazione di un prodotto ingegneristico. La loro diffusione, avvenuta intorno alla fine degli anni '70 del secolo scorso, fu dettata dalla necessità di ridurre i gradi di libertà (per esempio quelli di una risoluzione ad elementi finiti FEM) di problemi differenziali parametrizzati non lineari [25], i quali dovevano essere risolti innumerevoli volte per portare al raggiungimento di una soluzione soddisfacente. Quindi il loro primo impiego fu nell'ambito di ridurre i tempi computazionali di problemi differenziali. Con l'aumento della potenza computazionale dei moderni computer, la motivazione principe del loro utilizzo si è spostata dalla pura necessità di un metodo che riducesse i tempi di risoluzione ad una risorsa fondamentale per definire configurazioni di parametri che portassero ad una soluzione voluta.

I *metodi a riduzione di basi*, di cui fan parte anche il metodo POD, Proper Orthogonal Decomposition, possono essere applicati a tutti quei fenomeni fisici modellati e descritti da equazioni differenziali alle derivate parziali. Ed è proprio in questo ambito che si collocano le equazioni di Maxwell, le uniche a poter descrivere completamente un qualsiasi fenomeno elettromagnetico. Tra tutti, l'attenzione viene rivolta a due di questi:

- *problema di dispersione* o *scattering*, nel quale viene risolto il campo elettrico e magnetico generato dalla riflessione e diffusione di un'onda elettromagnetica che si infrange su un corpo. In particolare si prende in considerazione l'interazione tra un'onda e un profilo alare bidimensionale, caso più comunemente chiamato *invisibile*.

- *problema di campo lontano* relativo ad una fibra ottica, ossia una guida dielettrica che permette il transito digitale di informazioni al suo interno sotto forma di segnale elettromagnetico.

Per entrambi, non si procederà con un'analisi dettagliata della forma del campo elettrico risultante, ma si verificherà che il metodo di riduzione

di modello sia efficiente, consistente e convergente nella costruzione della soluzione con basi ridotte.

Modello matematico

Il comportamento dei fenomeni elettromagnetici è descritto dalle equazioni di Maxwell:

$$\begin{cases} \nabla \cdot \mathbf{D}(\mathbf{r},t) = \rho(\mathbf{r},t) \quad \text{Legge di Gauss elettrica,} \\ \dfrac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} = -\nabla \times \mathbf{E}(\mathbf{r},t) \quad \text{Legge di Faraday-Neumann-Lenz,} \\ \nabla \cdot \mathbf{B}(\mathbf{r},t) = 0 \quad \text{Legge di Gauss magnetica,} \\ \dfrac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} = \nabla \times \mathbf{H}(\mathbf{r},t) - \mathbf{J}(\mathbf{r},t) \quad \text{Legge di Ampére-Maxwell.} \end{cases} \tag{A.1}$$

Nella sua forma più famosa ed elegante, il sistema (A.1) presenta un conto delle incognite maggiore rispetto a quello delle equazioni (16 incognite contro 7 equazioni). Sono necessarie delle *relazioni costitutive* che permettano di pareggiare il conto. Per questo motivo, ipotizzando che i fenomeni presi in esame siano abbastanza deboli da non causare fenomeni non-lineari, che i materiali siano omogenei, isotropi, invarianti nel tempo e non dispersivi, valgono le approssimazioni $\mathbf{J} = \sigma\mathbf{E} + \mathbf{J}_N$, $\mathbf{D} = \epsilon\mathbf{E}$ e $\mathbf{B} = \mu\mathbf{H}$.
Un utile strumento per l'analisi e la risoluzione delle equazioni di Maxwell è la *trasformata di Fourier* che permette il passaggio dal dominio del tempo a quello delle frequenze. Una volta trasformato il sistema, si possono trovare alcune caratteristiche comuni al problema di scattering e a quello di fibra ottica guidante (assenza di distribuzione di cariche ρ e assenza di densità di corrente imposte $\mathbf{J}_N$). Grazie a queste ipotesi, il sistema da risolvere risulta essere:

$$\begin{cases} -\nabla^2\mathbf{E} - \omega^2\mu\epsilon\mathbf{E} + j\omega\mu\sigma\mathbf{E} = 0 \\ -\nabla^2\mathbf{H} - \omega^2\mu\epsilon\mathbf{H} + j\omega\mu\sigma\mathbf{H} = 0, \end{cases} \tag{A.2}$$

oltre all'imposizione di opportune condizioni al contorno.

Modello numerico

Per le simulazioni numeriche di entrambi i problemi in esame è stato utilizzato il software commerciale *COMSOL Multi-physics* il quale utilizza una discretizzazione spaziale secondo il metodo degli elementi finiti. In particolare, per

l'approssimazione del campo elettrico (unico effettivamente risolto tra i due campi incogniti) sono stati utilizzati i punti di integrazione di Gauss-Legendre. Grazie alla formalizzazione nel dominio delle frequenze non è stata necessaria nessuna approssimazione temporale. L'unica distinzione tra i due problemi riguarda il metodo di risoluzione dei relativi sistemi algebrici. Infatti, per il problema del profilo alare, è stato utilizzato il solutore diretto MUltifrontal Massively Parallel Sparse Solver (MUMPS), mentre per il problema della fibra ottica il solutore iterativo Generalized Minimal RESidual (GMRES) con un precondizionatore di tipo Symmetric Successive Overrelaxation (SSOR). La distinzione è dovuta al numero dei gradi di libertà da risolvere, notevolmente maggiore nel secondo caso.

Proper Orthogonal Decomposition (POD) per equazioni differenziali parametrizzate

Un problema differenziale si dice *parametrico* quando la risposta del sistema dinamico descritto dipende dal valore o dai valori assunti da parametri contenuti nelle equazioni di governo, nelle condizioni al contorno oppure nella geometria del problema. L'insieme di questi parametri compone una configurazione parametrica definita da $\mu \in \mathbb{P}$, il quale spazio ha una dimensione dipendete dal numero di parametri scelti.
Nel caso di un sistema dinamico stazionario, la forma astratta della formulazione debole approssimata con un metodo FEM è:

$$a(\tau_h(\mu), v_h; \mu) = f(v_h; \mu) \quad \forall v_h \in \mathbb{V}_h. \tag{A.3}$$

e, fissata una configurazione parametrica μ, la rispettiva soluzione è definita $\tau_h(\mu) \in \mathbb{V}_h$.

La peculiarità che contraddistingue i metodi di riduzione di modello, ed in particolare quello POD, è una netta differenza tra due principali fasi. Una, in cui vengono trovate un numero M di soluzioni aventi differenti configurazioni parametriche, è detta *Fase Offline*. Questa è la parte più onerosa dal punto di vista computazionale poiché si risolve (A.3) discretizzato, per esempio, tramite un metodo FEM. La seconda, in cui viene costruito lo spazio ridotto, è detta *Fase Online* che, al contrario della precedente, è molto più efficiente e rapida.
Nella metodologia POD, dato un insieme di M soluzioni $\{\tau_h(\mu_1), ..., \tau_h(\mu_M)\}$ aventi come basi $\{n_i\}_{i=1}^{N_h}$, l'obbiettivo è trovare delle basi ortogonali$\{\xi_i\}_{i=1}^{N_{rb}}$,

con dimensione N_{rb} molto inferiore rispetto N_h, minimizzando il *funzionale*

$$J = \sqrt{\frac{1}{M} \sum_{\mu \in \mathbb{P}} \inf_{v_{rb} \in \mathbb{V}_{rb}} ||\tau_h(\mu) - v_{rb}||^2_{\mathbb{V}}}. \tag{A.4}$$

La condizione di minimo è data dal problema agli autovalori $M \times M$

$$\underline{\underline{C}}\, \underline{\xi} = \lambda \underline{\xi}, \tag{A.5}$$

con λ_i i-esimo autovalore, ξ_i i-esimo autovettore e la *Matrice di Correlazione* definita come $\underline{\underline{C}} = 1/M(\tau_h(\mu_i), \tau_h(\mu_j))_{\mathbb{V}_h} \in \mathbb{R}^{M \times M}$. Disponendo in ordine decrescente gli autovalori e i rispettivi autovettori, le basi che comporranno il nuovo spazio approssimato $\mathbb{V}_{rb}$ sono i primi N_{rb} autovettori.

Applicazione del metodo POD: problema Stealth

Il problema del profilo alare colpito da un onda è una modellazione molto grossolana di una possibile situazione reale in cui un radar, mandando una *onda incidente*, cerca di acquisire informazioni su velocità, posizione e forma di un velivolo. Tutto ciò è possibile grazie all'*onda riflessa* o *onda scatterata* che si genera al momento del contatto tra profilo alare e onda incidente. Il profilo viene considerato come *conduttore perfetto*, mentre l'aria è considerata come un dielettrico ($\sigma = 0$), il che semplifica ulteriormente (A.2), ottenendo:

$$\begin{cases} -\nabla^2 \mathbf{E} - \omega^2 \mu \epsilon \mathbf{E} = 0 \\ -\nabla^2 \mathbf{H} - \omega^2 \mu \epsilon \mathbf{H} = 0. \end{cases} \tag{A.6}$$

I dati fondamentali dei due problemi studiati sono riassunti nella Tabella A.1.

NACA	α	γ	N	Triangoli	Elementi al contorno	GdL
0012	0°	15°	$\mathbb{P}^3$	55572	650	251049
4312	5°	0°	$\mathbb{P}^3$	58223	667	263004

TABELLA A.1. Dati della mesh e dati del problema, dove α è l'angolo d'incidenza del profilo, γ è la direzione dell'onda incidente e N è il grado dei polinomi approssimanti.

La soluzione per un profilo NACA 0012, utilizzata come validazione del modello definito in COMSOL, è in accordo con quanto ottenuto nel lavoro di

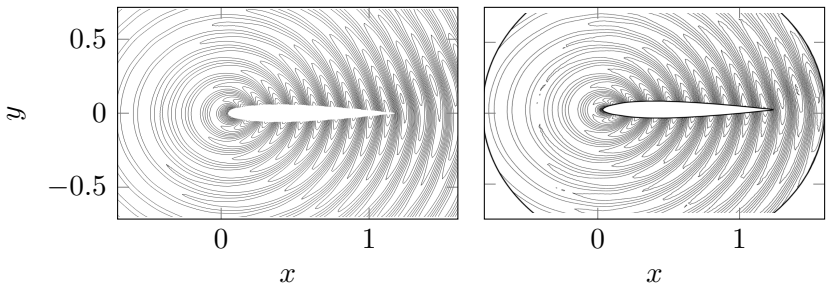

FIGURA A.1. Confrontto tra soluzioni: a sinistra, la soluzione ottenuta dal software COMSOL; a destra quella ottenuta da O. Pirroneau [23].

O. Pirroneau [23]. Entrambe sono mostrate nella Figura A.1.
La soluzione per un profilo NACA 4312 è stata utilizzata per dimostrare l'accuratezza e la validità della riduzione di modello con metodo POD.
Nella fase offline sono stati risolti $M = 51$ problemi (A.6). Il parametro scelto per la costruzione dell'insieme di soluzioni è la velocità angolare, la quale ha assunto i valori $\omega = \{6.283 : (9.425 - 6.283)/(51 - 1) : 9.425\} \cdot 10^9$ rad/s.
Nella fase online è stato creato uno script in ambiente *MATLAB* che permettesse l'implementazione della metodologia POD. Una volta costruita la matrice di correlazione $\underline{\underline{C}}$, sono stati selezionai i soli autovalori in grado di soddisfare la condizione $|\lambda_i / \sum_{k=1}^{51} \lambda_k| < 10^{-6}$ for $i = 1, ..., 51$, poichè gli unici energeticamente rilevanti - Figura A.2 -. L'insieme $\{\xi_i\}_{i=1}^{N_{rb}}$ degli autovettori corrispondenti sono stati usati per l'approssimazione della soluzione $\tau_{rb}(\mu)$ con basi ridotte.

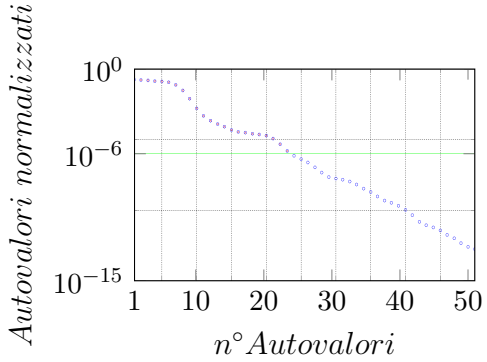

FIGURA A.2. NACA 4312: autovalori della matrice di correlazione. I cerchi blu indicano tutti gli autovalori, mentre quelli barrati dalla croce rossa indicano quelli utilizzati per la costruzione dello spazio delle basi ridotte. La linea verde indica la tolleranza scelta. In questo caso, il numero di basi utilizzate è $N_{rb} = 23$.

La differenza tra la soluzione ottenuta con il metodo FEM e quella con metodo di riduzione di base RB $\Delta\tau(\mu) = \tau_{FEM}(\mu) - \tau_{RB}(\mu)$, e l'errore relativo $Err = |\tau_{FEM_i} - \tau_{RB_i}|/|\tau_{FEM_i}|$ for $i = 1..., 263004$, sono stati usati per dimostrare la precisione della soluzione a basi ridotte, come mostrato in Figura A.3. Invece, per quanto riguarda la convergenza e la rapidità di approssimazione del metodo, sono stati confrontati sia i tempi medi di risoluzione dei metodi FEM e RB, sia l'errore relativo medio e massimo della soluzione RB al variare del numero di basi usate per la costruzione dello spazio ridotto - Figura A.4.

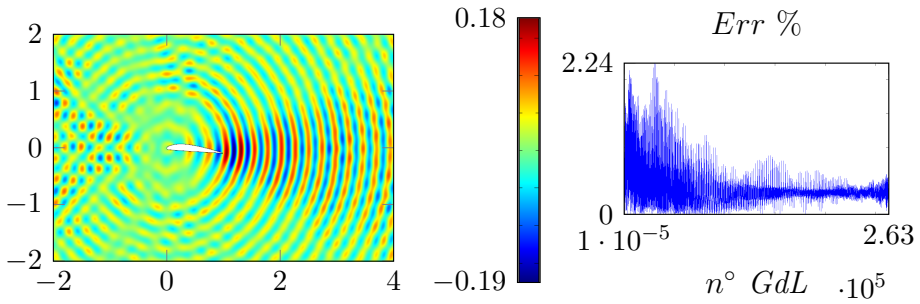

FIGURA A.3. NACA 4312: la figura a sinistra mostra $\Delta\tau$, mentre quella a destra mostra l'errore relativo Err, punto per punto, per ogni grado di libertà; entrambe le figure sono state ottenute con il valore del parametro $\omega = 7.886 \cdot 10^9$ rad/s.

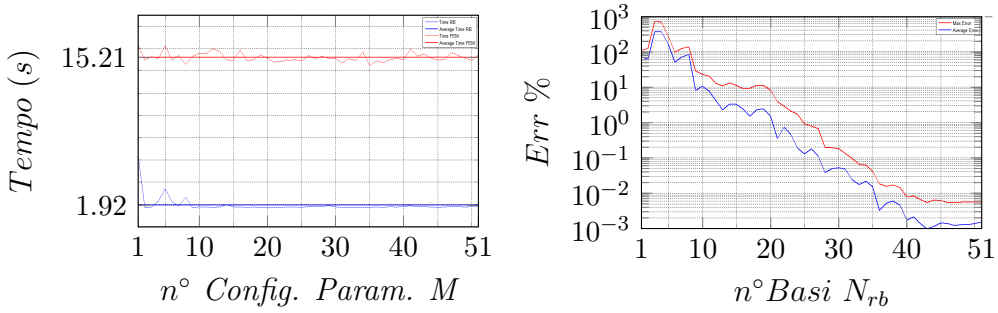

FIGURA A.4. NACA 4312: a sinistra, il grafico relativo ai tempi computazionali con metodo FEM (linee rosse) e con metodo RB (linee blu) in funzione del numero di basi utilizzate; a destra, relazione tra errore relativo massimo (linea rossa) ed errore relativo medio (linea blu) rispetto al numero di basi considerate.

Applicazione del metodo POD: problema della Fibra Ottica Tappata

Il problema della fibra ottica tappata (TOF) ha come scopo quello di replicare il comportamento del campo elettrico al di fuori dell'estremità della fibra. In particolare, per modellare in prima approssimazione la forma della fibra, si è utilizzato un cono. Come già detto in precedenza, non si è analizzato in dettaglio l'andamento del campo, bensì la bontà del metodo POD nel approssimare al meglio la soluzione ottenuta con il metodo FEM.
Il dominio è formato da un cono in Silice S_iO_2, da un rivestimento in oro Au immersi in aria. Viste le caratteristiche elettromagnetiche dei tre materiali, l'equazione da risolvere per ogni parte di dominio risulta una semplificazione di (A.2):

$$\begin{cases} -\nabla^2 \mathbf{E}_i - \omega^2 \mu_i \epsilon_i \mathbf{E}_i = 0 \\ -\nabla^2 \mathbf{H}_i - \omega^2 \mu_i \epsilon_i \mathbf{H}_i = 0, \end{cases} \tag{A.7}$$

dove con $i = 1, 2, 3$ si vuole indica che, avendo i tre materiali diverse caratteristiche elettromagnetiche, le equazioni risulteranno diverse tra loro. I dati numerici sono riassunti nella Tabella A.2.

Mesh	
Tetraedri	543300
Elementi di superficie	37531
Elementi di linea	2255
GdL	725547
N	$\mathbb{P}^1$

Tabella A.2. Dati numerici del problema, dove si indica con N il grado dei polinomi approssimanti.

Anche in questo caso, nella fase offline della metodologia POD si sono risolti $M = 81$ problemi (A.7). Il parametro scelto per la costruzione dell'insieme di soluzioni è il semi-angolo al vertice del cono, il quale assume i valori nell'insieme $\theta = \{10 : (45 - 10)/(81 - 1) : 45\}$. Poiché il parametro scelto non è un parametro fisico che compare nelle equazioni di governo, bensì geometrico, è necessaria una *mappatura affine* per poter completare la costruzione delle basi ridotte.
Nella fase online è stato creato, come nel caso precedente, uno script in ambiente *MATLAB* appositamente per questo problema in modo da implementare

la metodologia. Dopo aver costruito la matrice $\underline{\underline{C}}$, sono stati selezionati gli autovalori con la stessa tecnica di pesatura precedentemente esposta, ma con una tolleranza di $toll = 10^{-13}$. L'insieme degli autovalori è mostrato in Figura A.5.

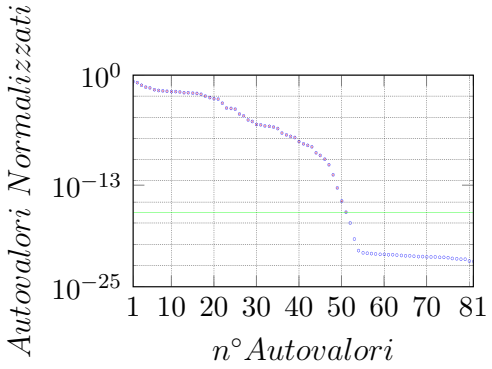

FIGURA A.5. Autovalori della matrice di correlazione. I cerchi blu indicano tutti gli autovalori, mentre quelli barrati dalla croce rossa indicano quelli utilizzati per la costruzione dello spazio delle basi ridotte. la linea verde indica la tolleranza scelta. In questo caso, il numero di basi utilizzate è $N_{rb} = 51$

Come nel caso precedente, alcuni grafici e figure sono mostrate per dimostrare prima l'accuratezza del metodo - Figura A.6 - e poi il risparmio computazionale tra metodo FEM e metodo RB e la velocità di approssimazione al variare del numero di basi utilizzate per la costruzione dello spazio ridotto approssimante - Figura A.7 -.

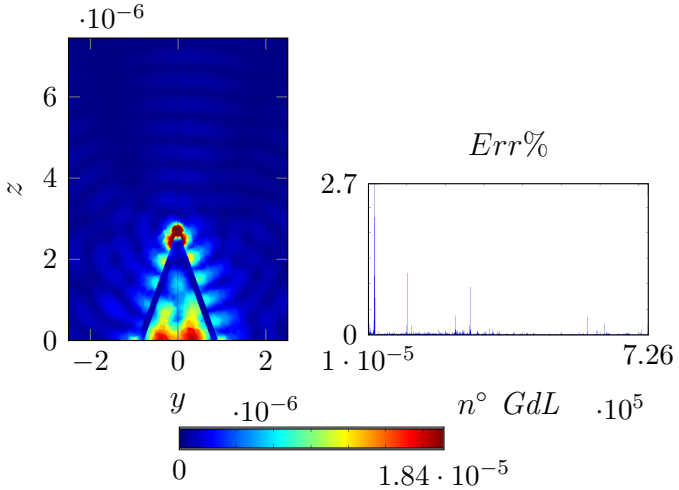

FIGURA A.6. La figura a sinistra mostra $\Delta\tau$ calcolata nel piano yz per $x = 0$, mentre quella a destra mostra l'errore relativo Err, punto per punto, per ogni grado di libertà; entrambe le figure sono state ottenute con il valore del parametro $\theta = 27.5°$.

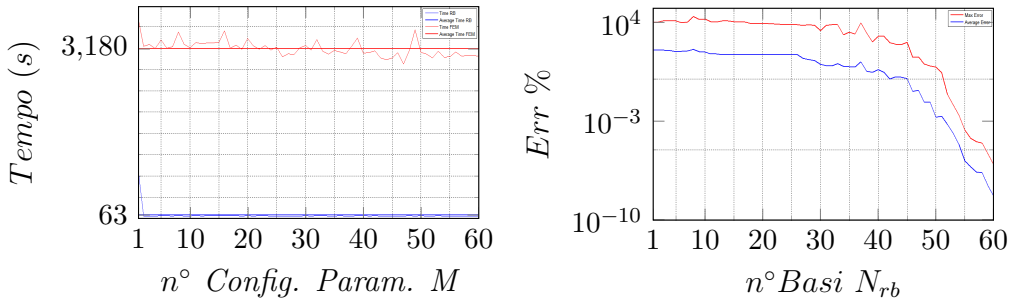

FIGURA A.7. A sinistra, il grafico relativo ai tempi computazionali con metodo FEM (linee rosse) e con metodo RB (linee blu) in funzione del numero di basi utilizzate; a destra, relazione tra errore relativo massimo (linea rossa) ed errore relativo medio (linea blu) rispetto al numero di basi considerate.

Bibliography

[1] Abbott, I. H., and Von Doenhoff, A. E. *Theory of wing sections, including a summary of airfoil data*. Courier Corporation, 1959.

[2] Adams, M. J. *An introduction to optical waveguides*. UMI Books on Demand, 1981.

[3] Adams, R. A. *A complete course calculus*. Addison-Wesley, Reading, Massachusetts,, 1995.

[4] Amestoy, P. R., Duff, I. S., and L'Excellent, J.-Y. Mumps multifrontal massively parallel solver version 2.0.

[5] Chen, X., Toh, K., and Phoon, K. A modified ssor preconditioner for sparse symmetric indefinite linear systems of equations. *International Journal for Numerical Methods in Engineering 65*, 6 (2006), 785–807.

[6] Chinesta, Huerta, Rozza, and Willcox. *Model Order Reduction, Enyclopedia of Computational Mechanics*. Elsevier, 2016.

[7] COMSOL, M. . User's guide. *Inc.: Burlington, MA, USA* (2014).

[8] Daniels, D. J. *Ground penetrating radar*, vol. 1. Iet, 2004.

[9] Dodson, M. Shannon's sampling theorem. *Current science 63*, 5c (1992).

[10] Gelsomino, F., and Rozza, G. Comparison and combination of reduced-order modelling techniques in 3d parametrized heat transfer problems. *Mathematical and Computer Modelling of Dynamical Systems 17*, 4 (2011), 371–394.

[11] Hess, M. W., and Benner, P. Fast evaluation of time–harmonic maxwell's equations using the reduced basis method. *Microwave Theory and Techniques, IEEE Transactions on 61*, 6 (2013), 2265–2274.

[12] Hess, M. W., Grundel, S., and Benner, P. Estimating the inf-sup constant in reduced basis methods for time-harmonic maxwell's equations. *Microwave Theory and Techniques, IEEE Transactions on 63*, 11 (2015), 3549–3557.

[13] Hesselmans, G., Calkoen, C., and Wensink, H. Mapping of seabed topography to and from synthetic aperture radar. *ESA SP* (1997), 1055–1058.

[14] Hesthaven, J. S., Rozza, G., and Stamm, B. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer, 2015.

[15] Hiptmair, R. Finite elements in computational electromagnetism. *Acta Numerica 11* (2002), 237–339.

[16] Jin, J.-M. *The finite element method in electromagnetics*. John Wiley & Sons, 2014.

[17] Joisel, A., Mallorqui, J., Broquetas, A., Geffrin, J., Joachimowicz, N., Lossera, M. V., Joire, L., and Bolomey, J.-C. Microwave imaging techniques for biomedical applications. In *Instrumentation and Measurement Technology Conference, 1999. IMTC/99. Proceedings of the 16th IEEE* (1999), vol. 3, IEEE, pp. 1591–1596.

[18] Kaminow, I., Li, T., and Willner, A. E. *Optical fiber telecommunications VB: systems and networks*. Academic press, 2010.

[19] Kepler, J. High order finite element methods for electromagnetic field computation.

[20] Marklein, R., Langenberg, K., Mayer, K., Miao, J., Shlivinski, A., Zimmer, A., Müller, W., Schmitz, V., Kohl, C., and Mletzko, U. Recent applications and advances of numerical modeling and wavefield inversion in nondestructive testing. *Advances in Radio Science 3*, B. 1 (2005), 167–174.

[21] MatLab, M. The language of technical computing. *The MathWorks, Inc. http://www. mathworks. com* (2012).

[22] Mazzolini, M., Facchetti, G., Andolfi, L., Zaccaria, R. P., Tuccio, S., Treu, J., Altafini, C., Di Fabrizio, E. M., Lazzarino, M., Rapp, G., et al. The phototransduction machinery in the rod outer segment has a strong efficacy gradient. *Proceedings of the National Academy of Sciences 112*, 20 (2015), E2715–E2724.

[23] Mohammadi, B., and Pironneau, O. *Applied shape optimization for fluids*, vol. 28. Oxford University Press Oxford, 2001.

[24] Monk, P. *Finite element methods for Maxwell's equations*. Oxford University Press, 2003.

[25] Noor, A. K., and Peters, J. M. Reduced basis technique for nonlinear analysis of structures. *Aiaa journal 18*, 4 (1980), 455–462.

[26] Quarteroni, A. *Numerical models for differential problems*, vol. 2. Springer Science & Business Media, 2010.

[27] Quarteroni, A., Sacco, R., and Saleri, F. *Numerical mathematics*, vol. 37. Springer Science & Business Media, 2010.

[28] Roif, H. I. Aircraft landing/taxiing system using lack of reflected radar signals to determine landing/taxiing area, Apr. 7 1998. US Patent 5,736,955.

[29] Rozza, G. Reduced-basis methods for elliptic equations in sub-domains with a posteriori error bounds and adaptivity. *Applied Numerical Mathematics 55*, 4 (2005), 403–424.

[30] Rozza, G., Huynh, D., Nguyen, N. C., and Patera, A. T. Real-time reliable simulation of heat transfer phenomena. In *proceeding of ASME 2009 Heat Transfer Summer Conference within the InterPACK09 and 3rd Energy Sustainability Conferences* (San Francisco, USA, 2009), American Society of Mechanical Engineers, pp. 851–860.

[31] Rozza, G., Huynh, D. B. P., and Patera, A. T. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering 15*, 3 (2008), 229–275.

[32] W. Hess, M., and Benner, P. A reduced basis method for microwave semiconductor devices with geometric variations. *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering 33*, 4 (2014), 1071–1081.

[33] Xu, J., and Cai, X.-C. A preconditioned gmres method for non-symmetric or indefinite problems. *Mathematics of computation 59*, 200 (1992), 311–319.

[34] Youcef-Toumi, K., and Reddy, S. Analysis of linear time invariant systems with time delay. *Journal of dynamic systems, measurement, and control 114*, 4 (1992), 544–555.