Contribution in topological optimization and application to nanophotonics

Nicolas Lebbe

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Mathématiques Appliquées**
Arrêté ministériel : 25 mai 2016

Présentée par

Nicolas LEBBE

Thèse dirigée par Édouard OUDET (UGA)
et codirigée par Alain GLIÈRE (CEA Leti)

préparée au sein du **Laboratoire Jean Kuntzmann**
dans **l'École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Contribution à l'optimisation de forme et application à la nanophotonique

Contribution in topological optimization and application to nanophotonics

Thèse soutenue publiquement le **26 novembre 2019**, devant le jury composé de :

Monsieur **Yannick Privat** *(Rapporteur)*
Professeur, Université de Strasbourg

Madame **Amélie Litman** *(Rapporteur)*
Maître de Conférences, Institut Fresnel

Monsieur **Stéphane Lanteri** *(Examinateur)*
Directeur de Recherche, Inria

Monsieur **François Jouve** *(Président du jury)*
Professeur, Université Paris Diderot

Monsieur **Édouard Oudet** *(Directeur de thèse)*
Professeur, Université Grenoble Alpes

Monsieur **Alain Glière** *(Co-directeur de thèse)*
Ingénieur chercheur, CEA Leti

Monsieur **Charles Dapogny** *(Encadrant)*
Chargé de recherche, Université Grenoble Alpes

Monsieur **Karim Hassan** *(Encadrant)*
Ingénieur chercheur, CEA Leti

This page is intentionally left blank.

Table of Contents

Introduction

Nanophotonic devices are components used to manipulate light, considered as an electromagnetic field, at the nanometric scale. They are tailored to accomplish specific tasks such as guiding an incident wave with negligible loss, splitting it into several output ports, converting a mode from an incoming waveguide into another mode of an outgoing waveguide, etc. Nanophotonic devices are the basic components of the photonic integrated circuits (PICs) used, for instance, in fiber optic communications, microscopy, biosensing and even in the prospective research about photonic computing.

The design of nanophotonic devices involves many variables, both in terms of physical properties of materials and geometry. As a result, more and more researchers are turning to numerical optimization, which in principle makes it possible for engineer to achieve the objectives described in the specifications in a shorter amount of time.

In the situation where the shapes to be determined are not intuitive, and therefore not simply parameterized a priori, specific geometric or topological optimization methods must be used. These numerical algorithms, initially developed in structural mechanics and aeronautics, are now applicable to the optimization of micro and nanometric scale components.

In this thesis, our primary goal was to set up a numerical framework for the systematic determination of the design of nanophotonic components with optimized performances based on geometrical methods of shape and topology optimization. Once implemented, our research led us to two different mathematical studies:

1. The first one concerns the modeling of uncertainties preventing the proper functioning of nanophotonic devices as well as the implementation of a method to take these uncertainties into account in the optimization process in order to obtain robust components. As a result, we have developed a method based on a gradient sampling algorithm which makes it possible to obtain robust components with respect to uncertainties over the incident wavelength or regarding the geometry of the produced component, which is sensitive to the lithography and etching manufacturing process variations.

2. The second one started as the optimization of the so-called “active” components in which we tried to optimize the shape of heaters which, when activated, allow to modify the materials properties by raising the temperature (Joule effect) and thus the behavior of the light inside a nanophotonic component beneath the heaters. This study allowed us to see that the underlying mathematical analysis was very rich and could be applied in a much more interesting way to the optimization of regions supporting different boundary conditions in mechanical devices.

Before moving to the summary of each chapter we would like to point out that an Index is supplied at the end of the document where the reader will find the definitions of the physical and mathematical concepts used throughout the text. The first occurrence of **important keywords** is also emphasized in **bold** in the text.

Chapter I: The physics behind photonics

The first chapter is an introduction to the fields of **electromagnetism** and nanophotonics, both from a physical and a mathematical point of views. We begin with the presentation of the amagnetic **time-harmonic vector wave equation** with no sources

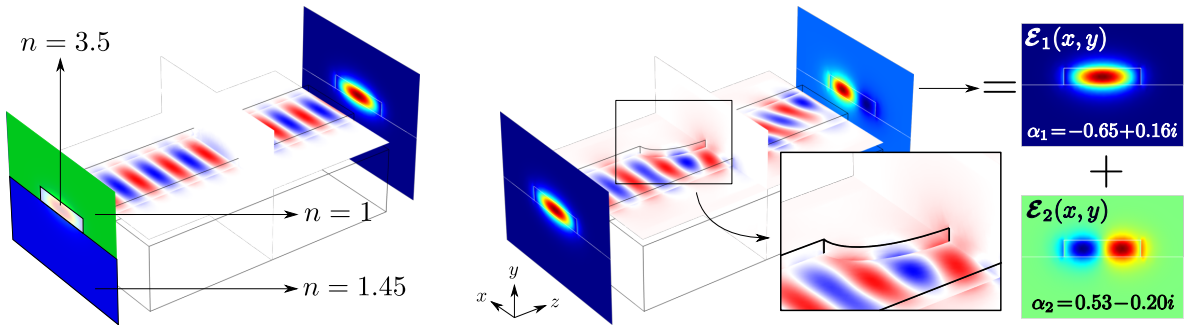
$$\nabla \times \nabla \times \mathbf{E} - k^2 n^2 \mathbf{E} = 0,$$

whose form is derived from the Maxwell equations and which describes the physical behavior of the **electric vector field E** at a given **wavelength** λ. This equation is fundamental in linear optics since it allows to study the behavior of light as an electromagnetic wave. The magnetic field **H** may be recovered from **E** using the Maxwell-Faraday equation

$$\mathbf{H} = \nabla \times \mathbf{E}/(i\omega\mu_0)$$

where $\omega = 2\pi c/\lambda$ with μ_0 and c constant physical values.

In the previous equation, n represents the **optical index** of the materials. By using materials with different optical indices, it is possible to produce **waveguides**, that is to say invariant structures in the z direction that allow for the confinement and propagation of light without loss; see Fig. A(a). The understanding of these waveguides is a key element in the study of nanophotonic components since they are the basic bricks connecting them together. In particular, the study of **guided modes** is of utter importance since they describes the state of the light which propagates within the waveguides.



(a) *Propagation of a guided mode inside a waveguide*

(b) *A perturbation of the waveguide's geometry results in a modification of the decomposition on the output modes*

Figure A: A guided mode injected inside a waveguide propagates without loss. If the optical indices are no longer invariant in the z direction then the decomposition of the electric field on the guided modes is modified.

More precisely we can show that any electric field propagating inside a z-invariant waveguide may be decomposed using $2N$ complex numbers α_j under the form:

$$\mathbf{E}(x,y,z) = \sum_{j=1}^{N} \underbrace{\alpha_j \boldsymbol{\mathcal{E}}_j(x,y)e^{i\beta_j z}}_{\text{forward}} + \underbrace{\alpha_{-j}\boldsymbol{\mathcal{E}}_{-j}(x,y)e^{-i\beta_j z}}_{\text{backward}},$$

where the guided modes $\boldsymbol{\mathcal{E}}_j$ are eigenvectors associated with the time-harmonic vector wave equation when looking for solutions of the form $\mathbf{E}(x,y,z) = \boldsymbol{\mathcal{E}}(x,y)e^{i\beta z}$ with an unknown **propagation constant** β. The sign in front of the modes number and coefficients is used to distinguish forward and backward propagating modes along the z

direction; see Fig. A(b) for an example of an electric field decomposed over two guided modes. Other concepts related to guided modes are presented such as the **TE and TM polarizations** used extensively by physicists and the **orthogonality relations** which allow to express the coefficients α_j in the previous decomposition as:

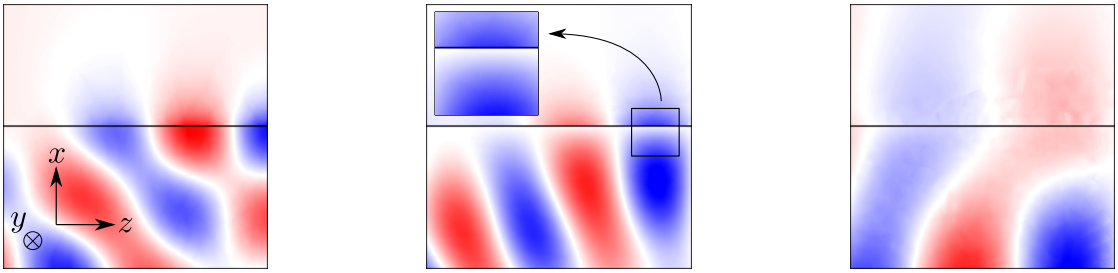
$$\alpha_j = \frac{1}{4}\int_\Gamma \left(\mathbf{E}\times\boldsymbol{\mathcal{H}}_j^* + \mathbf{H}\times\boldsymbol{\mathcal{E}}_j^*\right)\cdot\mathbf{n}\,\mathrm{d}s,$$

where Γ is the 2d cross section in the (x, y) plane of the waveguide.

A particular attention is devoted to the derivation of the appropriate **boundary conditions** which must be added to the time-harmonic vector wave equation in order to inject a particular guided mode inside a waveguide. The conclusion of this short study is the use of a **non-local Dirichlet-to-Neumann** condition which imposes the value of 1 to one coefficient α_j associated with a forward mode, while fixing the other ones to 0 except for the backward propagating modes whose coefficients are left free.

The **nanophotonic components** of interest in this thesis are then presented as silicon-based nanometric structures connected to each other through waveguides. Once stimulated by a guided mode of an input waveguide, the component redirects part of the light into the output waveguides so that the outgoing electric field contained inside each of them is controlled; the decomposition of $\mathbf{E}$ on the guided modes on each waveguide is the desired one. To characterize the performance of a nanophotonic component we mainly consider the **power carried by a mode $\boldsymbol{\mathcal{E}}_j$**, defined as $|\alpha_j|^2$ and introduce the **scattering matrix** of a component.

This first chapter also presents the mathematical tools underlying the study of electromagnetic fields. In particular, the Sobolev spaces $H(\boldsymbol{curl})$ and $H(\mathrm{div})$ are presented, thus allowing to obtain the **variational formulation** of the time-harmonic vector wave equation. The natural boundary conditions verified by the electric field at the interfaces between materials with different optical indices are expressed with the help of the trace theorems attached to the previous spaces: the tangential components $\mathbf{n}\times\mathbf{E}\times\mathbf{n}$ of the electric field is well-defined on such interfaces, as well as the product $n^2\mathbf{E}\cdot\mathbf{n}$ between the normal component of $\mathbf{E}$ and the squared optical index; see Fig. B for an illustration.



(a) *Real part of E_x (first tangential component)*

(b) *Real part of E_y (normal component)*

(c) *Real part of E_z (second tangential component)*

Figure B: Behavior of the electric field at the interface between two materials.

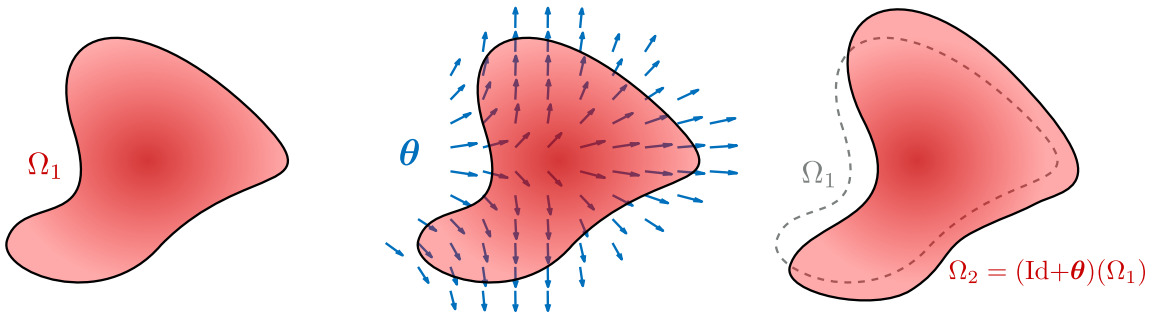
The chapter eventually present the two numerical methods which are extensively used throughout this thesis to solve the time-harmonic vector wave equation.

Chapter II: Geometric shape optimization

The second chapter focuses on geometrical **shape optimization** using the boundary variation method of Hadamard and the level-set framework. The goal of this chapter is to explain a way to find a shape $\Omega \subset \mathbb{R}^d$ which minimizes or maximizes a shape-dependent functional $\mathcal{J}(\Omega)$ using a gradient descent algorithm.

To define the gradient of such functional we will need to study the sensitivity of $\mathcal{J}(\Omega)$ when a small perturbation is applied on Ω. In this thesis we relies on **Hadamard's boundary variation method** which introduces the concept of **shape derivative**: a shape Ω_2 is said to be a small perturbation of another shape Ω_1 if we can transform one into to the other using a small vector field $\boldsymbol{\theta} : \mathbb{R}^d \to \mathbb{R}^d$. Mathematically it reads

$$\Omega_2 = (\mathrm{Id} + \boldsymbol{\theta})(\Omega_1) = \{\mathbf{x} + \boldsymbol{\theta}(\mathbf{x}), \mathbf{x} \in \Omega_1\}.$$



(a) *An initial shape* Ω_1 (b) *A vector field* $\boldsymbol{\theta}$ (c) *Modification of* Ω_1 *into* $(\mathrm{Id} + \boldsymbol{\theta})(\Omega_1) = \Omega_2$

Figure C: Example of deformation of a shape according to Hadamard's method.

The sensitivity of $\mathcal{J}(\Omega)$ with respect to the domain Ω is defined as the derivative of the mapping $\boldsymbol{\theta} \mapsto \mathcal{J}((\mathrm{Id} + \boldsymbol{\theta})(\Omega))$ at $\boldsymbol{\theta} = 0$. Namely, the following Taylor expansion holds:

$$\mathcal{J}((\mathrm{Id} + \boldsymbol{\theta})(\Omega)) = \mathcal{J}(\Omega) + \underbrace{\mathcal{J}'(\Omega)(\boldsymbol{\theta})}_{\text{shape derivative}} + o(\boldsymbol{\theta}).$$

Under mild assumptions on the considered shape functional $\mathcal{J}(\Omega)$, it turns out that most shape derivatives used in practice are of the form

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} V_\Omega \, \mathrm{d}s,$$

where $V_\Omega : \partial\Omega \to \mathbb{R}$ is a scalar field depending on the considered problem. In other words the sensitivity of $\mathcal{J}(\Omega)$ with respect to a deformation $\boldsymbol{\theta}$ only depends on the normal component of the vector field $\boldsymbol{\theta}$ on the border of the shape $\partial\Omega$. With this result in mind it follows that:

$$\mathcal{J}((\mathrm{Id} + \delta V_\Omega \mathbf{n})(\Omega)) = \mathcal{J}(\Omega) + \int_{\partial\Omega} |V_\Omega|^2 \, \mathrm{d}s + o(\delta),$$

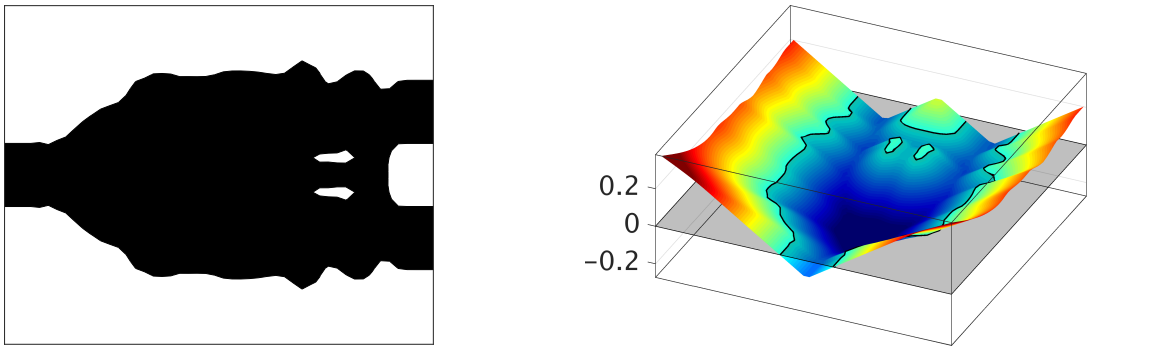
and so the maximization of $\mathcal{J}$ can be achieved by calculating iteratively V_Ω and performing the modification of Ω into $(\mathrm{Id} + \delta V_\Omega \mathbf{n})(\Omega)$ using a small scalar value δ.

Naturally, the mathematical determination of the scalar field V_Ω is not a trivial task. Even though most shape derivative may be found by means of **Céa's method**, we explain that

this technique must be applied with care if one does not want to end up with a wrong expression of the derivative $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$. This technical point can be understood with the help of the concepts of **Eulerian and Lagrangian derivatives**. In a nutshell these two derivatives are used when dealing with partial differential equations (PDE) constrained functionals such as the one considered in the next chapters. The potential difficulty with Céa's method is that it essentially relies on the existence of the Eulerian derivative while the latter may not exist if the solution to the considered PDE does not enjoy sufficient regularity on the border of the shape. Two examples of shape optimization problems where this lack of smoothness occurs are considered in the Chapters III and V.

The second part of this chapter concerns the numerical method used to represent and deform shapes. To achieve these goals, we consider the **level-set method** which represents a given shape $\Omega \subset \mathbb{R}^d$ by means of a function $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\partial\Omega$ is given by the zero level-set of ϕ, or more precisely that

$$\Omega = \{\mathbf{x}, \phi(\mathbf{x}) < 0\} \text{ and } \partial\Omega = \{\mathbf{x}, \phi(\mathbf{x}) = 0\}.$$



(a) *A two-dimensional shape* (b) *A level-set function representing the shape in* (a)

Figure D: A shape together with one of its level-set representations.

This representation has been proved to be useful in shape optimization since finding a level-set representation of the deformed shape $(\mathrm{Id} + V_\Omega \mathbf{n})(\Omega)$ is found as the solution at $t = 1$ to the following **Hamilton-Jacobi equation**:

$$\partial_t \psi(\mathbf{x}, t) + V_\Omega(\mathbf{x}) |\nabla \psi(\mathbf{x}, t)| = 0,$$

with the initial condition $\psi(\mathbf{x}, 0) = \phi(\mathbf{x})$.

We conclude this chapter with several important details concerning the numerical implementation of the gradient descent algorithm, such as the **velocity extension method**, and we provide a general pseudo-code to implement a shape optimization algorithm using Hadamard's method and a level-set framework.

Chapter III: Optimal design of photonic components

This chapter applies the information gathered in the two previous chapters to the optimization of some classical nanophotonic components. At first, a short review of the state of the art methods in shape and topology optimization of nanophotonic components is provided, explaining in particular binarization- and density-based methods.

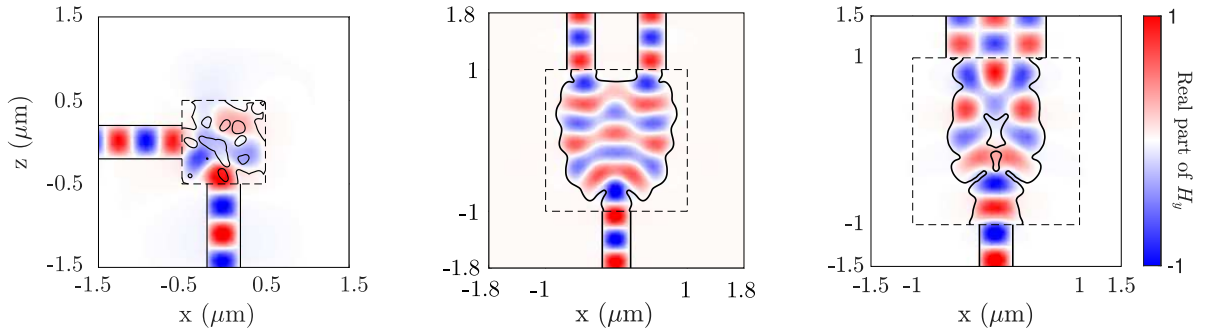
We then present the general optimization problem considered in most of this thesis: we are interested in the maximization of the power carried by a guided mode $(\boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{H}})$ on a cross section Γ_{out} of an outgoing waveguide when injecting light as a guided mode in an other (possibly the same) waveguide. This amounts to maximize the following **figure of merit**

$$\mathcal{J}(\Omega) = \left| \frac{1}{4} \int_{\Gamma_{\text{out}}} \mathbf{E}_\Omega \times \boldsymbol{\mathcal{H}}^* + \mathbf{H}_\Omega \times \boldsymbol{\mathcal{E}}^* \, \mathrm{d}s \right|^2 ,$$

where $\mathbf{E}_\Omega$ is the solution to the time-harmonic vector wave equation using the Dirichlet-to-Neumann boundary condition for the injection of the light and considering an optical index n_Ω which depends on the shape and the underlying repartition of the materials within the design domain. Recall that the magnetic field $\mathbf{H}_\Omega$ is derived from $\mathbf{E}_\Omega$ via the Maxwell-Faraday equation.

We then move on to the calculation of the shape derivative associated with $\mathcal{J}(\Omega)$ and we describe how the resulting formula may be adapted to obtain planar components, that is to say shapes that are invariant in the y-direction, which is a prerequisite for being able to produce the nanophotonic component through an etching process. As explained in the first chapter the electric field is not fully continuous at the interface between two materials. This notably means that $\mathbf{E}$ lacks regularity on $\partial\Omega$ resulting in the non-existence of the Eulerian derivative associated with $\mathbf{E}$ and therefore that we have to be careful in the application of the formal method of Céa or during the rigorous proof of the shape derivative.

This lack of continuity of the normal component of $\mathbf{E}$ also raises difficulties when computing the shape derivative numerically. To alleviate this issue we propose an **index smoothing method** where we replace the exact optical index with a smoothed counterpart obtained by convolution of n_Ω with a small gaussian, leading to a smooth optical index and thus to a fully-continuous electric field on $\partial\Omega$. This approximation is proved to be consistent, simplifies the simulation of the electric field using the Finite Element Method as it allows to have a constant mesh during the optimization process and justifies the possibility of evaluating $\mathbf{E}$ on $\partial\Omega$ as required by $\mathcal{J}'(\Omega)$ using a Finite Difference Method which is only defined on a Cartesian grid that does not coincide with $\partial\Omega$.



(a) *A ninety degrees bend which redirects the propagation direction of the light*

(b) *A power divider splitting equally the input power in the two top output waveguides*

(c) *A mode converter where the input guided mode is converted into an other mode in the top output waveguide*

Figure E: A collection of nanophotonic devices optimized in Chapter III.

The chapter continues with the presentation of many numerical results of our shape optimized nanophotonic devices. Some designs resulting from this study are shown in Fig. E.

The last section of this chapter finally presents an application of these methods to the optimization of **multi-level devices** and the application of this new methodology to obtain efficient **polarization rotators**. The idea here is to partially release the constraint that the optimized devices must be y-invariant and to authorize the shape to be composed of several layers stacked on top of each other.

This introduce a new constraint: for mechanical stability, each layer – represented by a two dimensional shape $\widehat{\Omega}_i$ for i from 1 to n_ℓ – must be located entirely above the lower layers, meaning that the shapes must verify $\widehat{\Omega}_1 \subset \ldots \subset \widehat{\Omega}_{n_\ell}$. To find shapes which respect this constraint we propose a projection-based algorithm and we discuss its limitations. We end up with a short explanation about another method to enforce this constraint which is not yet numerically implemented but which should, in theory, solve the problems encountered using the projection method.

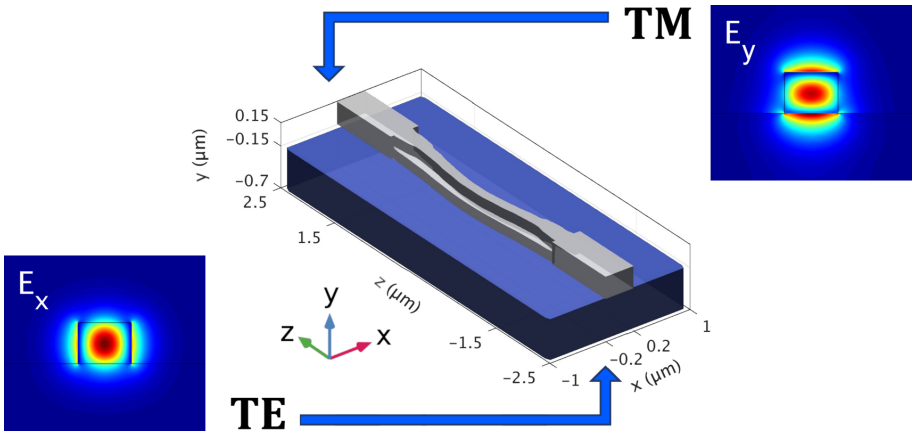


Figure F: A two-level polarization rotator converting an input guided mode with a TE polarization mode into a TM polarized one.

Chapter IV: Multi-objective problems and robustness to environmental uncertainties

In this chapter we present the numerical method developed during this thesis to find shapes which are robust to environmental uncertainties.

We first discuss the treatment of multi-objective shape optimization problems in which the goal is to simultaneously optimize several objectives functions $\mathcal{J}_i(\Omega)$ and for which it is desirable that each of the $\mathcal{J}_i$ perform equally well. Mathematically it means that we are interested in solving problems of the form:

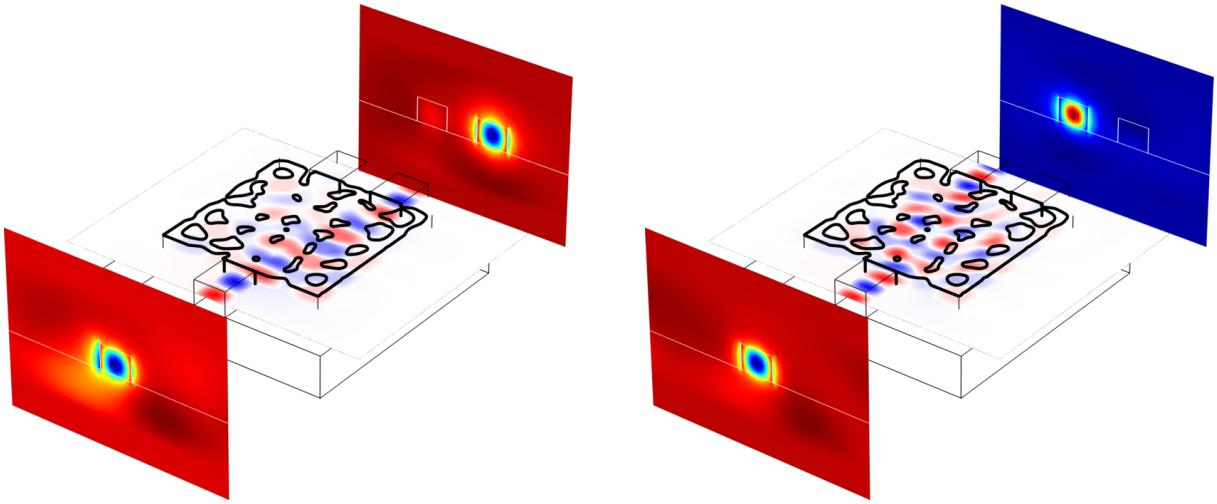
$$\max_{\Omega} \min_{i=1,\ldots,N} \mathcal{J}_i(\Omega).$$

Such problems are not easy to handle since it involves a min operator which is not differentiable. Moreover, numerically, when it can be computed, the gradient of the minimum will be equal to that of one of the $\mathcal{J}_i$ resulting in numerical oscillations since maximizing only one objective may degrade the values of the other ones. To deal with this problem, we then suggest to replace the gradient descent algorithm with a **gradient sampling**

one. This method cleverly considers all the "shape gradients" vector fields $\boldsymbol{\theta}_i$ that individually maximize each $\mathcal{J}_i(\Omega)$ and searches for a compromise vector field $\boldsymbol{\theta}$ as a convex combination (a "sampling") of the $\boldsymbol{\theta}_i$ which maximizes the minimum of the first-order Taylor expansion of all the $\mathcal{J}_i(\Omega)$. In other words we search for the solution of

$$\max_{\boldsymbol{\theta}\in\mathrm{conv}(\{\boldsymbol{\theta}_i\}_i)} \min_{i=1,\dots,N} \mathcal{J}_i(\Omega) + \mathcal{J}_i'(\Omega)(\boldsymbol{\theta}).$$

Finding such a vector field is actually fairly simple since it boils down to the resolution of a linear program. Using this method it is possible to consider the optimization of components such as diplexers, that is components which redirects the light into one of two output waveguides depending on the wavelength of the incident electric field; see Fig. G for an illustration.



(a) *Simulation at* $\lambda = 1.55\,\mu\mathrm{m}$ (b) *Simulation at* $\lambda = 1.31\,\mu\mathrm{m}$

Figure G: A nanophotonic diplexer in which the light is redirected to one of the two output waveguides depending on the wavelength of the incident guided mode.

The application of the gradient sampling algorithm to address **worst-case optimization** problems is then straightforward. Indeed, if the performances of a component are affected by an uncertain environmental parameter δ which could vary in the interval $[\delta_{\min}, \delta_{\max}]$ then we are interested in obtaining a component which retains good performances regardless of the value taken by δ in this interval. To obtain such devices, we could consider the following optimization problem in which we want to maximize the worst-case scenario:

$$\max_{\Omega} \min_{\delta\in[\delta_{\min},\delta_{\max}]} \mathcal{J}_\delta(\Omega).$$

In this formula, $\mathcal{J}_\delta(\Omega)$ is defined as the value of the objective function using the shape Ω when the uncertain parameter is equal to δ. Sampling the interval $[\delta_{\min}, \delta_{\max}]$ at a finite number of values this program drops down to a multi-objective problem such as those solved by the same gradient sampling algorithm. A direct application of this method is the design of components which are robust with respect to a small change of the wavelength induced for example by a lack of precision in the laser used to inject light.

We then move on to another type of uncertainties, this time caused by the **lithography and etching manufacturing process** used to produce nanophotonic components. After a presentation and a mathematical modeling of the main steps of this manufacturing

process, we propose a way to cope with the potential sources of uncertainties plaguing the proper production of the devices. We show in particular that it amounts to obtain a shape Ω which retains good performances even after an erosion or a dilation, namely that the border $\partial\Omega$ of the shape is perturbed into $(\mathrm{Id}+\delta\mathbf{n})(\partial\Omega)$ with δ a small scalar value. Since this random transformation of the shape involves the normal vector $\mathbf{n}$ which depends on the shape, a mathematical and numerical study is made to deal with the shape derivative of the altered figure of merit $\mathcal{J}_{\text{deformed}}(\Omega) = \mathcal{J}((\mathrm{Id}+\delta\mathbf{n})(\Omega))$. A variety of components are then optimized to obtain robust versions against the uncertainties inherent to the manufacturing processes. An example of such component is shown in Fig. H.

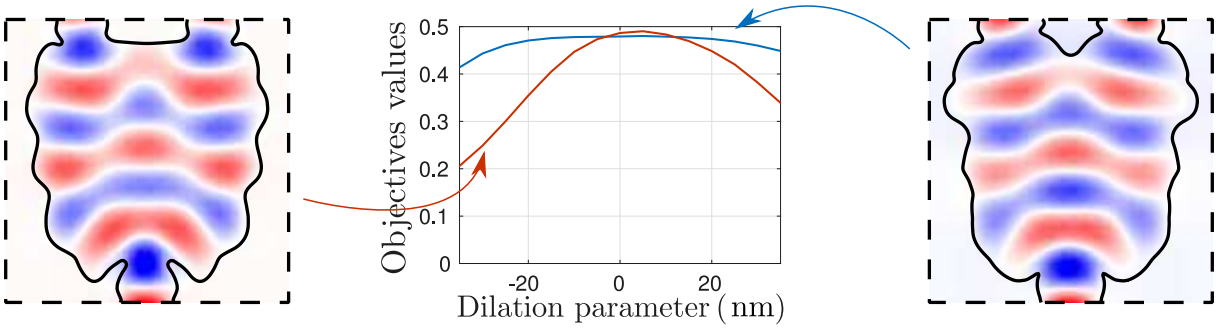


Figure H: A non robust (left) and a robust (right) power divider together with the real part of E_y propagating inside them. The performances of the components when a small dilation or erosion is applied on the shape is shown in the red and blue curves.

Chapter V: Boundary shape optimization

This last chapter is devoted to the application of geometrical shape optimization methods to the determination of the optimal repartition of regions supporting different boundary conditions in a partial differential equation. Considering at first the toy case consisting in a Laplacian PDE:

$$\begin{cases} -\Delta u_\Omega &= 0 \quad \text{in} \quad \Omega \\ \frac{\partial u_\Omega}{\partial \mathbf{n}} &= 0 \quad \text{on} \quad \Gamma \\ \frac{\partial u_\Omega}{\partial \mathbf{n}} &= g \quad \text{on} \quad \Gamma_N \\ u_\Omega &= 0 \quad \text{on} \quad \Gamma_D \end{cases},$$

we are interested in the shape optimization of the surface Γ_N and Γ_D representing the areas where respectively **inhomogeneous Neumann and homogeneous Dirichlet boundary conditions** are applied, the rest of the boundaries Γ bearing natural homogeneous Neumann conditions. The shape optimization of the interface between Γ and Γ_D, named Σ_D, happens to be more tricky than the case of the $\Gamma - \Gamma_N$ interface Σ_N. Indeed, as exemplified in Fig. I(c) the values of the solution u_Ω on Σ_D appear to be (weakly) **singular**. Precisely, in two dimensions, if Ω is locally flat around a point $\mathbf{x} \in \Sigma_D$ then using a local frame of polar coordinates (r, ν) centered at $\mathbf{x}$ we can express u_Ω as

$$u_\Omega(r,\nu) = u_{\text{reg}} + c r^{\frac{1}{2}} \cos(\nu/2),$$

where u_{reg} is an element in $H^2(\Omega)$ while the second term appear to be weakly singular in the sense that it only belongs to $H^s(\Omega)$ for $0 \leq s < 3/2$. As for the electric field, this lack of regularity at the interface of the optimized region implies that the Eulerian derivative of the considered PDE is not defined and therefore that a careful analysis must

(a) *Boundary conditions* (b) *Simulation of* u_Ω (c) *Plot of* $|\nabla u_\Omega|^2$

Figure I: Simulation of $-\Delta u_\Omega = 1$ using an arbitrary geometry and different boundary conditions; $\partial_{\mathbf{n}} u_\Omega = 0$ on Γ, $\partial_{\mathbf{n}} u_\Omega = 1$ on Γ_N and $u_\Omega = 0$ on Γ_D. We clearly see in (c) that there is less regularity near the interfaces between the homogeneous Neumann boundary Γ and the homogeneous Dirichlet boundary Γ_D than between Γ and the inhomogeneous Neumann boundary Γ_N.

be made in order to recover the correct shape derivative. In an even more surprising way, the incorrect application of Céa's method despite this lack of regularity yields a shape derivative equal to zero.

After some reminders concerning the Sobolev spaces adapted to the study of the solutions of the previous PDE in our context, we derive the correct shape derivatives corresponding to the optimization of Σ_N and Σ_D. Interestingly, in the case of Σ_D, the associated shape derivative only uses information coming from the singular part of u_Ω. From a numerical point of view however, this result is problematic since an accurate computation of the singular behavior of the solution is not easy to implement. We then propose a **regularization procedure** which replaces the boundary conditions on Γ and Γ_D by one of the form

$$\frac{\partial u_{\Omega,\varepsilon}}{\partial \mathbf{n}} + h_\varepsilon u_{\Omega,\varepsilon} = 0 \text{ on } \Gamma \cup \Gamma_D,$$

with h_ε a function chosen in such a way that a smooth transition between the Dirichlet and Neumann boundary condition is made.

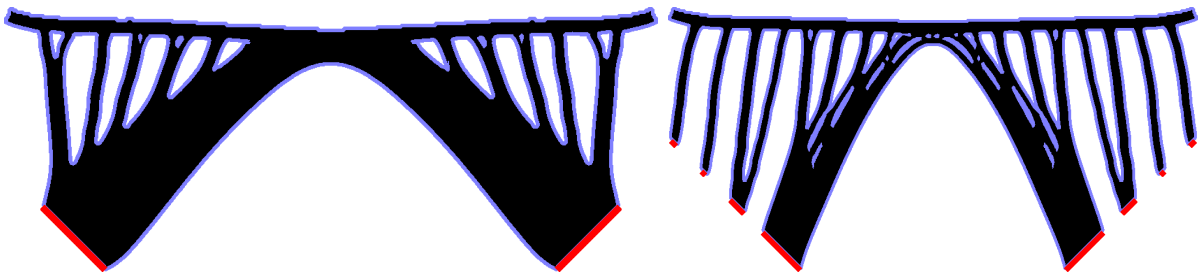


Figure J: Two optimized elastic bridges under their deformed configuration when a load is applied at the center of the bridge's deck. The red lines represents the fixed parts of the structures where a Dirichlet boundary condition is applied while the blue lines correspond to homogeneous Neumann boundary conditions.

The **consistency** of this approach is proved by showing that the approximate PDE solution $u_{\Omega,\varepsilon}$ using the regularization procedure is close to the exact solution u_Ω. It reads

$$\|u_{\Omega,\varepsilon} - u_\Omega\|_{H^1(\Omega)} \xrightarrow[\varepsilon \to 0]{} 0.$$

Moreover, since $u_{\Omega,\varepsilon}$ is more regular than u_Ω, we show that, for a large class of objective functions $\mathcal{J}(\Omega)$ depending on u_Ω, the shape derivative associated with the approximate objective function $\mathcal{J}_\varepsilon(\Omega)$ using $u_{\Omega,\varepsilon}$ instead of u_Ω may be calculated using Céa's method and is easier to compute numerically.

This chapter ends with applications of the regularization framework in the context of **linear elasticity** where Dirichlet boundary conditions account for the fixed part of a mechanical structure while the homogeneous Neumann boundary condition corresponds to the traction-free regions. An example of optimization result in the case of a bridge is given in Fig. J.

Scientific communications

Most of the works presented in this thesis have been published in the following papers:

[Leb19a] N. Lebbe, C. Dapogny, E. Oudet, K. Hassan, and A. Gliere. "Robust shape and topology optimization of nanophotonic devices using the level set method". In: *Journal of Computational Physics* (2019). DOI: 10.1016/j.jcp.2019.06.057

[Dap19] C. Dapogny, N. Lebbe, and E. Oudet. "Optimization of the shape of regions supporting boundary conditions". working paper or preprint. 2019. URL: https://hal.archives-ouvertes.fr/hal-02064477v1

[Leb19c] N. Lebbe, A. Gliere, K. Hassan, C. Dapogny, and E. Oudet. "Shape optimization for the design of passive mid-infrared photonic components". In: *Optical and Quantum Electronics* 51.5 (2019), pp. 166–179. DOI: 10.1007/s11082-019-1849-1

[Leb19b] N. Lebbe, A. Glière, and K. Hassan. "High-efficiency and broadband photonic polarization rotator based on multilevel shape optimization". In: *Optics Letters* 44.8 (2019), pp. 1960–1963. DOI: 10.1364/OL.44.001960

Oral presentations were also made at the following conferences:

[1] *(Poster)* Nicolas Lebbe, Edouard Oudet, Charles Dapogny, Karim Hassan, Alain Gliere. "Optimisation robuste de forme pour la nano-photonique", presented at SMAI-MODE, 2018, Autrans, France.

[2] *(Oral talk)* Nicolas Lebbe, Edouard Oudet, Charles Dapogny, Karim Hassan, Alain Gliere. "Optimisation robuste de forme pour la nano-photonique", presented at SMAI-CANUM, 2018, Cap d'Agde, France.

[3] *(Oral talk)* Nicolas Lebbe, Edouard Oudet, Charles Dapogny, Karim Hassan, Alain Gliere. "Dealing with uncertainties in shape optimizationof nano-photonic devices", presented at WCSMO13, 2019, Beijing, China.

[4] *(Oral talk)* Nicolas Lebbe, Edouard Oudet, Charles Dapogny, Karim Hassan, Alain Gliere. "Shape optimization for nanophotonic polarization rotator", presented at PIERS41, 2019, Rome, Italy.

Chapter I

The physics behind photonics

Summary — This first chapter gives a short presentation of nanophotonics and the devices of interest, as well as the physical equations governing light propagation inside them. We decided to give both a formal (Section I.1) and rigorous (Section I.4) description of the studied physics in order to point out where mathematical precision is really needed for our purpose.

In Section I.2 we define the so-called waveguide "modes" that underlie the study of nanophotonic devices and present their main properties. Two orthogonality relations for the waveguide modes are introduced and an emphasis is made on the link between these formulas and the power carried by a waveguide. The polarization of a mode is also defined by means of the two dimensional Maxwell equations.

Section I.3 then presents the full PDE allowing the study of nanophotonic devices (a particular attention is devoted to the boundary condition allowing the injection of a waveguide mode into the component) studied in the next chapters along with the mathematical formulations of the figure of merits that physicists seek to optimize.

The last section I.5 concludes this chapter by presenting the main numerical simulation methods used throughout this thesis to simulate the considered physics, that is the Finite Element Method (FEM) and the Finite Difference Time Domain (FDTD) method.

Most of the information presented in the following sections can be found in reference books such as [Jin14, Chapter 1] and [Orf02, Chapter 1] or [Mon03, Chapter 1, 3] for the mathematical aspects. We also highly suggest Manfred Hammer's lectures [Ham17] and the recent book of Westerveld & Urbach [Wes17] for a presentation specifically focused on photonics, as well as the book of Snyder & Love [Sny83] for a more exhaustive presentation of waveguides-related informations.

I.1 Waves as an electromagnetic field and the Maxwell equations

I.1.1 Time dependent Maxwell equations

At the length scales of interest in this thesis, light behaves as an electromagnetic wave and is fully described by the electric $\mathbf{E}(\mathbf{x}, t)$ (in V/m) and magnetic $\mathbf{H}(\mathbf{x}, t)$ (in A/m) fields. Even if the considered nanophotonic devices can reach sizes as small as ten nanometers, quantum effects, which are only at play when a corpuscular description of light is adopted

(as photons) have not been observed experimentally in this particular context and allows us to ignore them.

I.1.1.a General Maxwell equations

The theory of **Maxwell's equations** links the evolution of both $\mathbf{E} : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ and $\mathbf{H} : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ using a set of 4 coupled partial differential equations. In general this system of equations reads:

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} \qquad \textit{Maxwell-Faraday} \tag{I.1.1}$$

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{J} \qquad \textit{Maxwell-Ampère} \tag{I.1.2}$$

$$\nabla \cdot \mathbf{D} = \rho \qquad \textit{Maxwell-Gauss} \tag{I.1.3}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \textit{Maxwell-Thomson} \tag{I.1.4}$$

where ρ is the electric charge density and $\mathbf{B}, \mathbf{D}, \mathbf{J}$ are additional functions connected to the electromagnetic fields through constitutive equations. In the case of simple linear medium, which will always be the case in this thesis, these relations are simply

$$\mathbf{B} = \boldsymbol{\mu}\mathbf{H}, \quad \mathbf{D} = \boldsymbol{\varepsilon}\mathbf{E} \quad \text{and} \quad \mathbf{J} = \boldsymbol{\sigma}\mathbf{E} \tag{I.1.5}$$

with $\boldsymbol{\mu}$ the magnetic permeability, $\boldsymbol{\varepsilon}$ the dielectric permittivity and $\boldsymbol{\sigma}$ the conductivity. In general, the last three physical quantities are 3×3 symmetric positive definite matrices which may depend on the spatial position (when different materials are involved) and on time (memory effects).

I.1.1.b Simplifications in the nanophotonic context

In the context of nanophotonics we will mainly have to deal with amagnetic, dielectric and isotropic materials, meaning respectively that $\boldsymbol{\mu} = \mu_0 I_3$ (where μ_0 is the **permeability** of free space equal to $4\pi 10^{-7}$ T m/A), $\boldsymbol{\sigma} = \rho = 0$ and $\boldsymbol{\varepsilon} = \varepsilon I_3$ with $\varepsilon > 0$. The **permittivity** is more frequently given by the **optical index** n of the material with $n^2 = \varepsilon/\varepsilon_0$ and $\varepsilon_0 \simeq 8.85 \times 10^{-12}$ F/m the vacuum permittivity.

With these simplifications, a classical operation consists in injecting Eq. (I.1.2) into the curl of Eq. (I.1.1) in order to obtain the following equation involving only the electric field

$$\nabla \times \nabla \times \mathbf{E} = -c^{-2} n^2 \partial_{tt}^2 \mathbf{E} \tag{I.1.6}$$

and where we also used the relation $c^{-2} = \mu_0 \varepsilon_0$ with $c \simeq 3 \times 10^8$ m/s the **speed of light** in vacuum. The same type of equation is found for the magnetic field as

$$\nabla \times n^{-2} \nabla \times \mathbf{H} = -c^{-2} \partial_{tt}^2 \mathbf{H} \tag{I.1.7}$$

but will prove less convenient afterwards. It is important to note that, in addition to Eqs. (I.1.6) and (I.1.7), the vector fields $\mathbf{E}$ and $\mathbf{H}$ must also verify the conditions given by Eqs. (I.1.3) and (I.1.4), that is to say $\nabla \cdot (n^2 \mathbf{E}) = \nabla \cdot \mathbf{H} = 0$.

I.1.2 Time harmonic vector wave equation

Due to the linearity of Eq. (I.1.6), it is possible to study the frequency response of the electric field at a given frequency f (thereafter we will prefer to use the **wavelength** $\lambda = c/f$ in m) by using the Fourier transform. To do this, we consider a time-harmonic regime with

$$\mathbf{E}(\mathbf{x}, t) = \mathrm{Re}\left[\mathbf{E}(\mathbf{x}) \exp(-i\omega t)\right] \tag{I.1.8}$$

where $\omega = 2\pi c/\lambda$ is the pulsation and $\mathbf{E} : \mathbb{R}^3 \to \mathbb{C}^3$. Note here that we use the same notation for both time-dependent ($\mathbf{E}(\mathbf{x}, t) \in \mathbb{R}^3$) and time-harmonic ($\mathbf{E}(\mathbf{x}) \in \mathbb{C}^3$) fields. This should not be a problem afterwards since we will only rely on the second one. Injecting Eq. (I.1.8) in Eq. (I.1.6) we find that $\mathbf{E}(\mathbf{x})$ is solution to

$$\nabla \times \nabla \times \mathbf{E} - k^2 n^2 \mathbf{E} = 0 \tag{I.1.9}$$

where $k = \omega/c$ is the **wavenumber**. Note that if μ were not constant we would have to solve

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E} - \omega^2 \varepsilon \mathbf{E} = 0. \tag{I.1.10}$$

Respectively, we find that $\mathbf{H}$ is solution to

$$\nabla \times n^{-2} \nabla \times \mathbf{H} - k^2 \mathbf{H} = 0 \tag{I.1.11}$$

where an equivalent decomposition to Eq. (I.1.8) for $\mathbf{H}$ is considered. From Eqs. (I.1.1) and (I.1.2) we also have the relations

$$\nabla \times \mathbf{E} = i\omega\mu_0 \mathbf{H} \quad \text{and} \quad \nabla \times \mathbf{H} = -i\omega\varepsilon \mathbf{E}. \tag{I.1.12}$$

Equation (I.1.9) is most often refered as the **time-harmonic vector wave equation** in the literature and will be the basis for all our nanophotonics computations. Contrary to Eq. (I.1.6), a vector field satisfying Eq. (I.1.9) automatically verifies the **divergence condition** since applying the divergence operator to this equation leads to $\nabla \cdot (n^2 \mathbf{E}) = 0$ (reminding that of $\nabla \cdot \nabla \times = 0$).

Remark I.1.2.1: This observation explains why we prefer using Eq. (I.1.9) rather than the vectorial Helmholtz equation $\Delta\mathbf{E} + k^2 n^2 \mathbf{E}$ (which is more commonly found in physical textbooks) obtained using the identity $\nabla \times \nabla \times \mathbf{E} = \nabla\nabla \cdot \mathbf{E} - \Delta\mathbf{E}$ since a solution of the vectorial Helmholtz equation time the squared optical index does not necessarily verify the divergence-free condition.

I.1.3 Boundary conditions

I.1.3.a Limit at infinity

A large part of the phenomena studied in photonics involves the analysis of the scattered field $\mathbf{E}^{\mathrm{sc}}$ induced by the excitation of a device subject to an incident field $\mathbf{E}^{\mathrm{inc}}$. In this case the total field $\mathbf{E}$ is decomposed as

$$\mathbf{E} = \mathbf{E}^{\mathrm{sc}} + \mathbf{E}^{\mathrm{inc}}. \tag{I.1.13}$$

While $\mathbf{E}^{\mathrm{inc}}$ is known, the scattered field is obtained through Eq. (I.1.9) and should be an outgoing, exponentially decreasing wave at infinity. To ensure that this is verified, the Maxwell equations must be complemented with the **Silver-Müller radiation condition**.

Supposing that at infinity the medium is, in all direction, homogeneous (for instance air), then Silver-Müller condition is given by (see [Mon03, Section 1.4.3] or [Néd01, Section 5.2])

$$\lim_{\mathbf{x}\to\infty} |\mathbf{x}| \left(\nabla \times \mathbf{E}^{\mathrm{sc}} \times \frac{\mathbf{x}}{|\mathbf{x}|} - ik\mathbf{E}^{\mathrm{sc}} \right) = 0. \tag{I.1.14}$$

I.1.3.b Natural conditions at interfaces

Let us consider the situation depicted in Fig. I.1.1 for which we are going to look at the variations of the electromagnetic fields inside a domain $\mathcal{D} \subset \mathbb{R}^3$ and more precisely near the interface between two dielectric and amagnetic domains Ω and $\mathcal{D}\backslash\bar{\Omega}$ of respective optical indices n_1 and n_2.

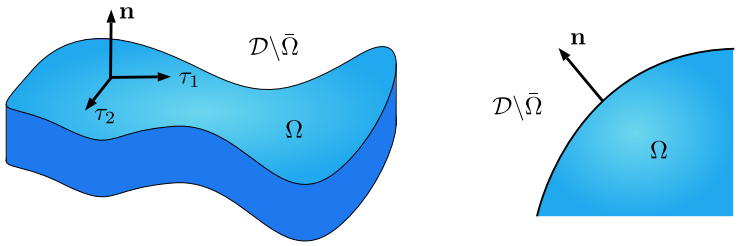


Figure I.1.1: Interface between two domain Ω and $\mathcal{D}\backslash\bar{\Omega}$ with different optical indices.

To study the precise behavior of the electric field at the interface we look at the restrictions of $\mathbf{E}$ in both domain, namely we define $\mathbf{E}_1 = \mathbf{E}_{|\mathcal{D}\backslash\bar{\Omega}}$ and $\mathbf{E}_2 = \mathbf{E}_{|\Omega}$. We will also need to use the normal vector exterior to Ω which is referred for any point $\mathbf{x} \in \partial\Omega$ as $\mathbf{n}_\Omega \in \mathbb{S}^2$. With this definition the **normal component** of any vector field $\mathbf{E}$ is defined as $\mathrm{E}_\perp = \mathbf{E} \cdot \mathbf{n}_\Omega$ and the tangential part as $\mathbf{E}_\parallel = \mathbf{E} - (\mathbf{E} \cdot \mathbf{n}_\Omega)\mathbf{n}_\Omega = \mathbf{n}_\Omega \times \mathbf{E} \times \mathbf{n}_\Omega$.

At the interfaces between two dielectric and amagnetic materials it can be derived that $\mathbf{E}$ and $\mathbf{H}$ must verify the following conditions (these conditions emerges naturally from the variational equations as we will see in Section I.4.1):

$$\mathbf{n}_\Omega \times \mathbf{E}_1 = \mathbf{n}_\Omega \times \mathbf{E}_2, \quad n_1^2 \mathbf{E}_1 \cdot \mathbf{n}_\Omega = n_2^2 \mathbf{E}_2 \cdot \mathbf{n}_\Omega \quad \text{and} \quad \mathbf{H}_1 = \mathbf{H}_2, \tag{I.1.15}$$

meaning that the electric field only has its tangential components continuous on interfaces whereas $\mathbf{H}$ is fully continuous. In the same way, using Eq. (I.1.12), we find that

$$\begin{gathered} \mathbf{n}_\Omega \times n_1^{-2} \nabla \times \mathbf{H}_1 = \mathbf{n}_\Omega \times n_2^{-2} \nabla \times \mathbf{H}_2, \\ \nabla \times \mathbf{H}_1 \cdot \mathbf{n}_\Omega = \nabla \times \mathbf{H}_2 \cdot \mathbf{n}_\Omega \quad \text{and} \quad \nabla \times \mathbf{E}_1 = \nabla \times \mathbf{E}_2, \end{gathered} \tag{I.1.16}$$

which this time gives the continuity of the electric field's curl. It is important to keep these **continuity relations** in mind when we will deal with interface motion in Chapter III because some quantities, for instance $\mathbf{E} \cdot \mathbf{n}_\Omega$, are not well-defined on the interface between different materials.

I.1.3.c Interface with perfect conductors

In the case of an interface between a dielectric medium and a perfectly conducting one, such as a metal, the boundary condition Eq. (I.1.15) is simplified into:

$$\mathbf{n} \times \mathbf{E} = 0 \tag{I.1.17}$$

since no electric field may exist inside the metal. Equation (I.1.17) is the same as a Dirichlet boundary condition on the tangential component of the electric field and is usually known as the **perfect electric conductor (PEC)** boundary condition. Let us also point out here that a similar relation exists at the interface between a dielectric medium and a **perfect magnetic conductor (PMC)**:

$$\mathbf{n} \times \nabla \times \mathbf{E} = 0, \tag{I.1.18}$$

where Eq. (I.1.18) acts as a Neumann boundary condition on $\mathbf{E}$ and is most often only used to impose symmetries into electromagnetic domains (see Remark I.4.3.3).

I.1.4 Two-dimensional approximation

We conclude this section by presenting the case of the two-dimensional Maxwell equations, which are used when propagation of the electromagnetic fields along one direction (say the y-axis) may be ignored due to invariances of the material properties, meaning that the derivative along this direction of any physical quantity is equal to zero

First of all, from Eqs. (I.1.9), (I.1.11) and (I.1.12) a direct calculation shows that

$$-\Delta_{x,z}\mathrm{E}_y - k^2 n^2 \mathrm{E}_y = 0 \quad \text{and} \quad -\nabla_{x,z} \cdot \left(n^{-2}\nabla_{x,z}\mathrm{H}_y\right) - k^2 n^2 \mathrm{H}_y = 0 \tag{I.1.19}$$

where the subscripts x, z indicates that only the partial derivatives with respect to x and z are considered in the operators. Equation (I.1.19) are usual scalar **Helmholtz equations**, which, once solved, give $\mathrm{E}_y, \mathrm{H}_y$ and subsequently all the other components using Eq. (I.1.12):

$$\mathrm{H}_x = -\partial_z \mathrm{E}_y/(i\omega\mu_0),\ \mathrm{H}_z = \partial_x \mathrm{E}_y/(i\omega\mu_0),\ \mathrm{E}_x = \partial_z \mathrm{H}_y/(i\omega\varepsilon),\ \text{and}\ \mathrm{E}_z = -\partial_x \mathrm{H}_y/(i\omega\varepsilon).$$

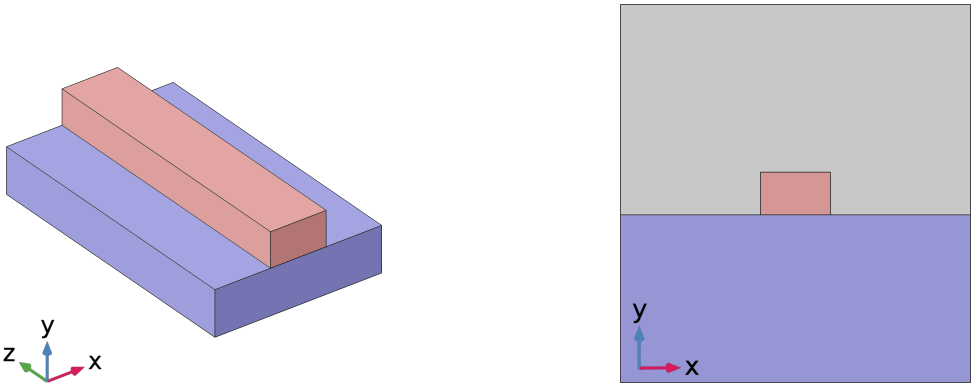
These equations notably imply that both triplets $(\mathrm{E}_x, \mathrm{H}_y, \mathrm{E}_z)$ and $(\mathrm{H}_x, \mathrm{E}_y, \mathrm{H}_z)$ may be found independently. This independence leads to defining two "**polarizations**" of the light; the components $(\mathrm{E}_x, \mathrm{H}_y, \mathrm{E}_z)$ are said to be in the **Transverse Electric (TE)** polarization (only the electric field in the transverse plane (x, z) are considered) whereas $(\mathrm{H}_x, \mathrm{E}_y, \mathrm{H}_z)$ belongs to the **Transverse Magnetic (TM)** polarization.

I.2 Waveguides modes and power

Propagation modes describe the fundamental structure of electromagnetic waves propagating in an infinite waveguide. As such, they play a central role in the boundary conditions accounting for injection of light at the entrance ports of nanophotonic devices (see Section I.3 later), as well as in the expression of the figure of merits considered for their optimization. A typical nanophotonic waveguide is depicted in Fig. I.2.1 and is usually divided into three main parts:

- A **core** (in red) containing a material with a high-value optical index $n = n_{\text{core}}$ allowing a good confinement of the light inside it. In practice this core is made of silicon (Si), germanium (Ge) or an alloy of both leading to $n_{\text{core}} > 3$.
- A **substrate** layer (in blue) with a lower optical index $n = n_{\text{subs}}$. In this thesis we will only consider silica (SiO_2) layers with $n_{\text{subs}} \simeq 1.4$ at wavelength in the near- or mid-infrared (i.e. λ from about 1 µm to 10 µm).

- A **cladding** (in gray) covering the rest of the core with index $n = n_{\text{clad}}$. Most of the time this cladding will simply be air ($n_{\text{clad}} = 1$); in the case of a different material the waveguide is said to be encapsulated.



(a) *A photonic waveguide* (b) *Two-dimensional view*

Figure I.2.1: Schematic representation of a photonic waveguide.

I.2.1 Modes in a waveguide, polarization

In this section we suppose that the waveguide is oriented (and thus propagates light) in the z direction meaning that the optical index inside the waveguide does not depend on z, that is $n(x, y, z) = n(x, y)$. Using some kind of separation of variables we search for electromagnetic fields $\mathbf{E}, \mathbf{H}$ which are equal to the product of a vector-valued function depending only on x, y (which is known as the field's profile) and a harmonic function on the z direction, meaning that

$$\mathbf{E}(x, y, z) = \mathbf{E}(x, y)e^{i\beta z} \quad \text{and} \quad \mathbf{H}(x, y, z) = \mathbf{H}(x, y)e^{i\beta z},$$

where $\beta \in \mathbb{R}$ is called the **propagation constant**. Thereafter it will be useful to distinguish between the component of $\mathbf{E}, \mathbf{H}$ in the propagation direction $\mathrm{E}_\perp = \mathrm{E} \cdot \hat{\mathbf{z}}$ and the other ones $\mathbf{E}_\parallel = \hat{\mathbf{z}} \times \mathbf{E} \times \hat{\mathbf{z}}$ which are perpendicular to it. Using these notations:

$$\mathbf{E} = (\mathbf{E}_\parallel(x, y) + \mathrm{E}_\perp(x, y)\hat{\mathbf{z}})e^{i\beta z} \quad \text{and} \quad \mathbf{H} = (\mathbf{H}_\parallel(x, y) + \mathrm{H}_\perp(x, y)\hat{\mathbf{z}})e^{i\beta z}. \tag{I.2.1}$$

I.2.1.a Eigenvalue problem

We are now searching for particular solutions of the time-harmonic vector wave equation (I.1.9) of the form given by Eq. (I.2.1). After some calculus we find that the components $\mathbf{E}_\parallel(x, y)$ and $\mathrm{E}_\perp(x, y)$ of the electric field are solution of the following 2D system on $\mathbb{R}^2$ (no z component is present)

$$\begin{cases} \nabla_\tau \times \nabla_\tau \times \mathbf{E}_\parallel - i\beta\nabla_\tau \mathrm{E}_\perp + (\beta^2 - k^2n^2)\mathbf{E}_\parallel &= 0 \\ -\Delta_\tau \mathrm{E}_\perp - i\beta\nabla_\tau \cdot \mathbf{E}_\parallel - k^2n^2\mathrm{E}_\perp &= 0 \end{cases}, \tag{I.2.2}$$

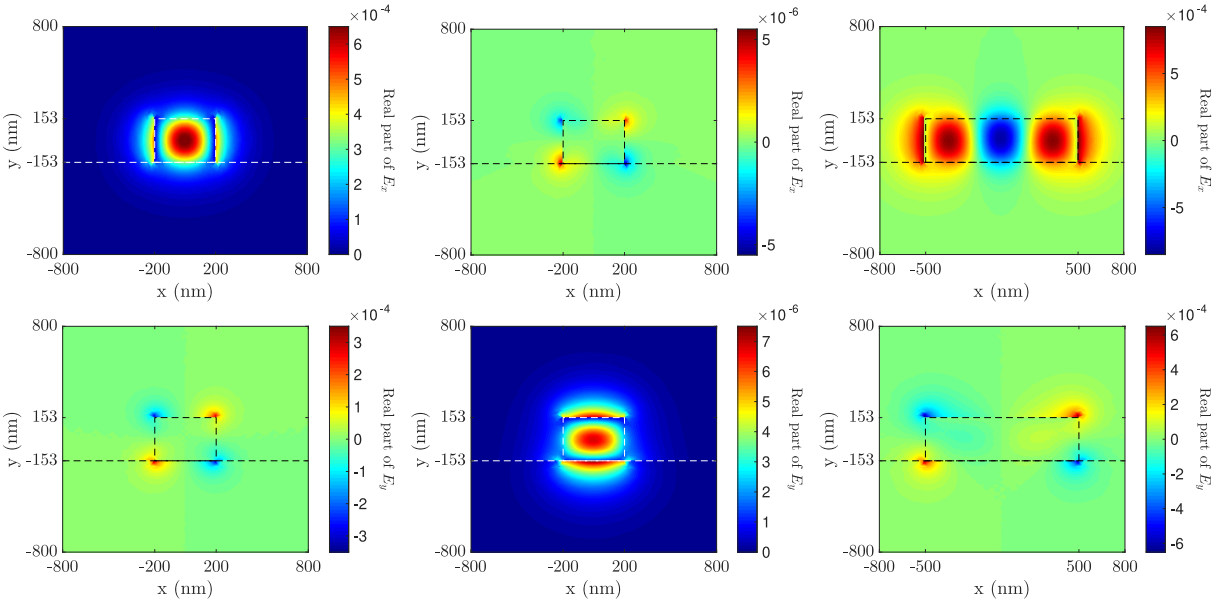
where $\nabla_\tau = (\partial_x, \partial_y, 0)^\top$. With the change of variables $\hat{\mathbf{E}}_\parallel = \beta\mathbf{E}_\parallel$ and $\hat{\mathrm{E}}_\perp = i\mathrm{E}_\perp$ this equation rewrites (assuming $\beta \neq 0$) into the following system involving only β^2

$$\begin{cases} \nabla_\tau \times \nabla_\tau \times \hat{\mathbf{E}}_\parallel - \beta^2\nabla_\tau \hat{\mathrm{E}}_\perp + (\beta^2 - k^2n^2)\hat{\mathbf{E}}_\parallel &= 0 \\ -\Delta_\tau \hat{\mathrm{E}}_\perp + \nabla_\tau \cdot \hat{\mathbf{E}}_\parallel - k^2n^2\hat{\mathrm{E}}_\perp &= 0 \end{cases}. \tag{I.2.3}$$

A solution $(\beta, \mathbf{E})$ to this (generalized) **eigenvalue problem** is said to be a propagation **mode** associated to the waveguide. Note that we now use a calligraphic $\boldsymbol{\mathcal{E}}$ as a means to indicate that this electric field belongs to a propagation mode and distinguish it from the electromagnetic field $\mathbf{E}$ solution of the time-harmonic vector wave equation (I.1.9). We also define the **effective index** of the mode as

$$n_{\text{eff}} = \beta / k. \tag{I.2.4}$$

This denomination is easily understood by looking at the behavior of a plane progressive wave. In a homogeneous medium with optical index n, there exist particular solutions of the time-harmonic vector-wave equation known as the progressive plane waves with electric fields of the form $\mathbf{E}_0 e^{inkz}$ where $\mathbf{E}_0$ is a vector and k the wavenumber. By comparison with Eq. (I.2.1) we see that the waveguide mode is oscillating in the z direction at the same speed than a progressive plane wave in an "effective" homogeneous material of optical index n_{eff} as defined in Eq. (I.2.4).



(a) *First TE mode (TE_0)* (b) *First TM mode (TM_0)* (c) *Third TE mode (TE_2)*

Figure I.2.2: Some guided mode inside two different waveguides with a silicon core of size 400×306 nm and 1000×306 nm, placed on a silica layer and surrounded by air at wavelength $\lambda = 1.55\,\mu$m.

Remark I.2.1.1: Many other formulations similar to Eq. (I.2.2) allow to obtain the same propagation modes. Here we consider a formulation involving the 3 components of the electric field but it is also possible to use only the components $(\mathrm{E}_z, \mathrm{H}_z)$ [Ham17, Part E] or the tangential parts $(\mathrm{E}_x, \mathrm{E}_y, \mathrm{H}_x, \mathrm{H}_y)$ [Wes17, Section 2.3]. As for Eq. (I.2.2), it is notably employed in the paper [Lee91].

I.2.1.b Polarization and denominations

Depending on both the value of β and the structure of $\boldsymbol{\mathcal{E}}$, many types of modes exist as solution to Eq. (I.2.2) with different mathematical nature and physical properties.

In this thesis we are mainly interested in the case of **guided modes**, meaning that β belongs to $\mathbb{R}$ with

$$k \max(n_{\text{clad}}, n_{\text{subs}}) < |\beta| < k n_{\text{core}}.$$

These eigenvalues exist in finite number and we will make a distinction between **forward** propagating ($\beta > 0$) and **backward** propagating ($\beta < 0$) modes. Moreover, for each guided mode $(\beta, \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{H}})$ there exist a mode $(\bar{\beta}, \bar{\boldsymbol{\mathcal{E}}}, \bar{\boldsymbol{\mathcal{H}}})$ propagating in the opposite direction with

$$\bar{\beta} = -\beta, \quad \bar{\boldsymbol{\mathcal{E}}} = \boldsymbol{\mathcal{E}}_{\parallel} - \mathcal{E}_{\perp}\hat{\mathbf{z}} \quad \text{and} \quad \bar{\boldsymbol{\mathcal{H}}} = -\boldsymbol{\mathcal{H}}_{\parallel} + \mathcal{H}_{\perp}\hat{\mathbf{z}}. \tag{I.2.5}$$

This is proved by injecting Eq. (I.2.5) into Eq. (I.2.2).

In addition to its direction of propagation, we also define the polarization of a guided mode using the same nomenclature as the one presented in Section I.1.4. More precisely if most of the mode's energy (see Section I.2.3 and Eq. (I.2.11)) comes from the $(\mathrm{E}_x, \mathrm{H}_y, \mathrm{E}_z)$ (resp. $(\mathrm{H}_x, \mathrm{E}_y, \mathrm{H}_z)$) components the mode is said to be in the TE (resp. TM) polarization or more properly quasi-TE (resp. quasi-TM). The difference between TE and TM is visible on Figs. I.2.2(a) and I.2.2(b).

Concerning the non-guided modes the reader is referred to [Wes17, Section 2.3.6] or [Gou10]. In the first reference we find that these other modes must verify either $\beta \in i\mathbb{R}$ or $|\beta| < k \max(n_{\text{clad}}, n_{\text{subs}})$, constitute a continuum of eigenvalues and are part of the essential spectrum of the operator associated to Eq. (I.2.2). In the aftermath, these particular modes will be gathered under the name of **radiative mode** and we will assume most of the time that they can be ignored since they are not physically realistic (and so not observed in practice).

I.2.2 Decomposition of the electric field

I.2.2.a Decomposition

In Section I.2.1 we defined the propagation modes of a waveguide as the eigenvectors associated to the operator of Eq. (I.2.2). A difficult result from spectral analysis (see [Gou10, Appendix A] for this result in the case of the scalar Helmholtz equation) allows to decompose any field solution of the time-harmonic vector wave equation in a waveguide as:

$$\mathbf{E} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\mathcal{E}}_i + \alpha_{-i} \boldsymbol{\mathcal{E}}_{-i} + \boldsymbol{\mathcal{E}}_{\text{rad}}, \tag{I.2.6}$$

$$\mathbf{H} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\mathcal{H}}_i + \alpha_{-i} \boldsymbol{\mathcal{H}}_{-i} + \boldsymbol{\mathcal{H}}_{\text{rad}} \tag{I.2.7}$$

where N is the number of forward propagating guided modes, $(\boldsymbol{\mathcal{E}}_i, \boldsymbol{\mathcal{H}}_i)$ (resp. $(\boldsymbol{\mathcal{E}}_{-i}, \boldsymbol{\mathcal{H}}_{-i})$) are the forward (resp. backward) modes and the vector fields $(\boldsymbol{\mathcal{E}}_{\text{rad}}, \boldsymbol{\mathcal{H}}_{\text{rad}})$ gather all the radiative modes (see Remark I.2.2.1). In other words the modes constitute a complete basis for the set of all the electromagnetic fields solution to the time-harmonic vector wave equation in a section of a waveguide. In the case of $N = 1$ the waveguide is said to be single-mode.

Remark I.2.2.1: Let us quickly comment about the structure of the radiative modes (see notably [Wes17, Section 2.3.6]). In general we can decompose $\boldsymbol{\mathcal{E}}_{\text{rad}}$ as

$$\boldsymbol{\mathcal{E}}_{\text{rad}} = \int_{-k\eta}^{k\eta} \alpha_s^{\text{rad}} \boldsymbol{\mathcal{E}}_s^{\text{rad}} \, \mathrm{d}s + \int_{-i\infty}^{i\infty} \alpha_s^{\text{evs}} \boldsymbol{\mathcal{E}}_s^{\text{evs}} \, \mathrm{d}s$$

where the presence of a continuum of real eigenvalues comes from the fact that the operator is unbounded (note that for closed waveguides these modes does not exists) and are sometime referred as radiation modes while the complex eigenvalues are inherent to the fact that the operator is not self-adjoint and are called evanescent modes. These radiative modes are part of the essential spectrum of the operator and are not physically relevant (radiation modes do not have finite energy while the evanescent ones are absorbed quickly after propagation in the waveguide).

I.2.2.b Orthogonality relations

In order to determine the coefficients α_i in Eq. (I.2.6) one crucial observation is that there exists an **orthogonality relation** between the propagation modes. For two propagating modes $(\boldsymbol{\mathcal{E}}_i, \boldsymbol{\mathcal{H}}_i), (\boldsymbol{\mathcal{E}}_j, \boldsymbol{\mathcal{H}}_j)$ with $|i| \neq |j|$ we have the relation (see the reciprocity theorem and its demonstration for guided modes in [Sny83, Section 31-3])

$$\int_{\Gamma_\infty} \left[\boldsymbol{\mathcal{E}}_i \times \boldsymbol{\mathcal{H}}_j^*\right] \cdot \hat{\mathbf{z}} \, \mathrm{d}s = 0 \tag{I.2.8}$$

where $\hat{\mathbf{z}}$ is the unit vector of the propagation direction (here the z axis) and Γ_∞ an infinite section of the waveguide, in our context $\Gamma_\infty = \{(x, y, 0), (x, y) \in \mathbb{R}^2\}$. To distinguish between forward and backward modes ($j = -i$) we need to consider a slightly different orthogonality condition valid for any $i \neq j$:

$$\int_{\Gamma_\infty} \left[\boldsymbol{\mathcal{E}}_i \times \boldsymbol{\mathcal{H}}_j^* + \boldsymbol{\mathcal{H}}_i \times \boldsymbol{\mathcal{E}}_j^*\right] \cdot \hat{\mathbf{z}} \, \mathrm{d}s = 0. \tag{I.2.9}$$

Remark I.2.2.2: These orthogonality relations are only true for non-absorbing waveguide modes (i.e. optical indices for the core, substrate and cladding are real-valued). An equivalent relationship may be found for absorbing ones by dropping the conjugate on $\boldsymbol{\mathcal{H}}_j$ in Eq. (I.2.8).

From these orthogonality relations we easily deduce that the coefficients α_i for $|i| \in [1, N]$ are given by

$$\alpha_i = \frac{\displaystyle\int_{\Gamma_\infty} [\mathbf{E} \times \boldsymbol{\mathcal{H}}_i^* + \mathbf{H} \times \boldsymbol{\mathcal{E}}_i^*] \cdot \hat{\mathbf{z}} \, \mathrm{d}s}{\displaystyle 2 \int_{\Gamma_\infty} \mathrm{Re}\, [\boldsymbol{\mathcal{E}}_i \times \boldsymbol{\mathcal{H}}_i^*] \cdot \hat{\mathbf{z}} \, \mathrm{d}s}. \tag{I.2.10}$$

For the rest of this manuscript we will impose the modes to have the following normalization (which corresponds to a unit power of 1 W as will be seen in Section I.2.3)

$$\frac{1}{2} \int_{\Gamma_\infty} \mathrm{Re}\, [\boldsymbol{\mathcal{E}}_i \times \boldsymbol{\mathcal{H}}_i^*] \cdot \hat{\mathbf{z}} \, \mathrm{d}s = \frac{1}{2} \int_{\Gamma_\infty} \mathrm{Re}\left[\mathcal{E}_{i,x} \mathcal{H}_{i,y}^* - \mathcal{E}_{i,y} \mathcal{H}_{i,x}^*\right] \mathrm{d}s = \mathrm{sign}(i) \tag{I.2.11}$$

where $\mathrm{sign}(i)$ equals 1 for $i > 0$ and -1 for $i < 0$. Using Eqs. (I.2.10) and (I.2.11) the coefficients α_i are given by

$$\alpha_i = \frac{\mathrm{sign}(i)}{4} \int_{\Gamma_\infty} [\mathbf{E} \times \boldsymbol{\mathcal{H}}_i^* + \mathbf{H} \times \boldsymbol{\mathcal{E}}_i^*] \cdot \hat{\mathbf{z}} \, \mathrm{d}s \tag{I.2.12}$$

which is known as the **overlap integral** between the electromagnetic field and the i-th mode.

I.2.3 Power carried by a mode

Physically the local (volumic) density of **electromagnetic energy** is defined by

$$\omega_{\text{em}} = \frac{\varepsilon}{2}|\mathbf{E}|^2 + \frac{\mu}{2}|\mathbf{H}|^2. \tag{I.2.13}$$

The power associated with this energy in a volume $V \subset \mathbb{R}^3$ is then, by definition, $P_V = \int_V \partial_t \omega_{\text{em}}\, \mathrm{d}\mathbf{x}$. Introducing the Poynting vector $\mathbf{\Pi} = \mathbf{E} \times \mathbf{H}$, a direct calculation using the Maxwell equations shows that for dielectric medium $\partial_t \omega_{\text{em}} = -\nabla \cdot \mathbf{\Pi}$ and thus

$$P_V = -\int_V \nabla \cdot \mathbf{\Pi}\, \mathrm{d}\mathbf{x} = -\int_{\partial V} \mathbf{\Pi} \cdot \mathbf{n}\, \mathrm{d}s. \tag{I.2.14}$$

Averaging this quantity over one period of time (between any t and $t + 2\pi/\omega$) leads to

$$\langle P_V \rangle = -\int_{\partial V} \langle \mathbf{\Pi} \rangle \cdot \mathbf{n}\, \mathrm{d}s = -\int_{\partial V} \frac{1}{2}\mathrm{Re}\left[\mathbf{E} \times \mathbf{H}^*\right] \cdot \mathbf{n}\, \mathrm{d}s.$$

We then define the **electromagnetic power** (also known as the **intensity**) in Watt crossing a surface Γ in $\mathbb{R}^3$ as

$$\mathcal{P}_\Gamma = \frac{1}{2}\int_\Gamma \mathrm{Re}\left[\mathbf{E} \times \mathbf{H}^*\right] \cdot \mathbf{n}\, \mathrm{d}s. \tag{I.2.15}$$

Remark I.2.3.1: In the two-dimensional situation of Section I.1.4, depending on the polarization, this equation may be simplified into either

$$\mathcal{P}_\Gamma^{\text{2D,TE}} = \frac{1}{2}\int_\Gamma \mathrm{Re}\left[\frac{1}{i\omega\varepsilon}\partial_z \mathrm{H}_y \mathrm{H}_y^*\right] \mathrm{d}s \quad \text{or} \quad \mathcal{P}_\Gamma^{\text{2D,TM}} = \frac{1}{2}\int_\Gamma \mathrm{Re}\left[\frac{1}{i\omega\mu_0}\mathrm{E}_y \partial_z \mathrm{E}_y^*\right] \mathrm{d}s.$$

Injecting the decomposition of Eq. (I.2.6) (again, we ignore the radiative modes) inside Eq. (I.2.15) with the orthogonality conditions (I.2.8) and (I.2.9) and the unit normalization (I.2.11) as well as the fact that $\mathrm{Re}\,[a] = 1/2(a + a^*)$ we find that:

$$\begin{aligned}\mathcal{P}_\Gamma &= \frac{1}{2}\int_\Gamma \mathrm{Re}\left[\left(\sum_{i=1}^N \alpha_i \boldsymbol{\mathcal{E}}_i + \alpha_{-i}\boldsymbol{\mathcal{E}}_{-i}\right) \times \left(\sum_{i=1}^N \alpha_i \boldsymbol{\mathcal{H}}_i + \alpha_{-i}\boldsymbol{\mathcal{H}}_{-i}\right)^*\right] \cdot \mathbf{n}\, \mathrm{d}s \\ &= \sum_{i=1}^N |\alpha_i|^2 - \sum_{i=1}^N |\alpha_{-i}|^2,\end{aligned} \tag{I.2.16}$$

meaning that the part of **power carried by the forward mode** $\boldsymbol{\mathcal{E}}_i$ (also known as the **transmission**) is given by $|\alpha_i|^2$ with α_i defined in Eq. (I.2.12). Physicists are also commonly using quantities in decibel, that is the value $\mathcal{P}_{\text{dB}} = 10\log_{10}(P)$ dB. The power carried by a mode in decibel is in turn defined as $20\log_{10}(|\alpha_i|)$.

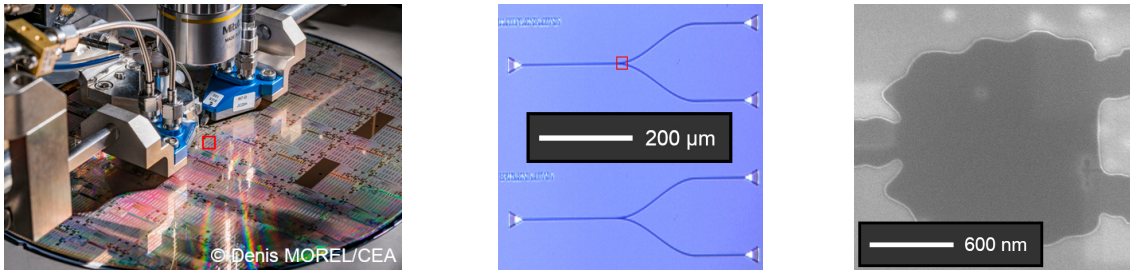
I.3 Nanophotonic components

I.3.1 General presentation

I.3.1.a Introduction

The nanophotonic devices of interest in this thesis pertain to the field of **silicon photonics**, which features photonic integrated circuits composed of a base wafer (that is to

say a thin slice of semiconductor as in Fig. I.3.1(a)), on which components are patterned by means of CMOS compatible microfabrication techniques (see Section IV.3.1 for more details on the manufacturing process). Silicon On Insulator (SOI) base wafers have recently aroused a tremendous enthusiasm among the integrated optics community for their relatively simple and cheap production, and for their high efficiency in terms of energy confinement.



(a) *Full wafer (*200 mm *diameter)* (b) *Zoom on two circuits* (c) *A single component*

Figure I.3.1: Successive magnifications of an engraved wafer.

Accordingly, the devices considered in the next chapters feature a core (remember that this denomination and the following ones were introduced in the preamble of Section I.2) which is made of either silicon (Si) or silicon-germanium (SiGe) whereas the substrate is typically composed of silica (SiO_2) or silicon-nitride (SiN). Moreover, in order to increase the contrast between optical indices, and thus the light guiding characteristics, the silicon pattern is surrounded by a cladding of air.

Multiple devices can then be tailored into the core material to accomplish specific tasks such as guiding an incident wave with negligible loss, splitting it into several output ports, converting a mode from an incoming waveguide into another mode of an outgoing waveguide, etc. These components constitute the unitary parts that, once chained and connected by means of waveguides, allow to build complex photonic integrated circuits (PICs) used, for instance, in fiber optic communications, microscopy, biosensing and even in the prospective research about photonic computing.

The peculiar, targeted properties of these nanophotonic devices are achieved by acting on the geometry of the repartition of core and cladding materials within a given design space $\mathcal{D}_{\text{opt}}$ (see Fig. I.3.1(c) and the following). The repartition of core material is characterized by the subset $\Omega \subset \mathcal{D}_{\text{opt}}$ and is typically composed of several simple geometric shapes with characteristic sizes varying from a few dozen to several hundreds of micrometers. Recent developments in the lithography-etching processes now even allow the inclusion of arbitrary patterns with sizes of about ten nanometers.

I.3.1.b Physical description

Let us now make more precise the geometrical domain's notations used for the full simulations of nanophotonic components. The ambient space is decomposed in several subdomains, with different optical indices defined as follows (and illustrated in Fig. I.3.2)

- The box $\mathcal{D} = [-w_x, w_x] \times [-w_y, w_y] \times [-w_z, w_z] \subset \mathbb{R}^3$ is the total computational domain, accounting for the whole three-dimensional space - at least the region where it is relevant to consider the electromagnetic fields surrounding the device.

- $\mathcal{D}_{\text{opt}} \subset \mathcal{D}$ is the fixed design domain; it is typically a box with small thickness h in the y direction, containing all possible shapes Ω. The optical index inside this domain is denoted by n_Ω since it will depend on the repartition of both the core (Ω) and cladding ($\mathcal{D}_{\text{opt}} \backslash \bar{\Omega}$) materials.

- $\mathcal{D}_{\text{wg}} \subset \mathcal{D}$ is the region occupied by the input and output waveguides.

- $\mathcal{D}_{\text{subs}} \subset \mathcal{D}$ is the layer supporting $\mathcal{D}_{\text{opt}}$, occupied by the substrate.

- $\mathcal{D}_{\text{PML}} \subset \mathcal{D}$ is a "Perfectly Matched Layer", a region of $\mathcal{D}$ filled with absorbing material aimed at imposing the correct behavior of the electromagnetic fields at infinity. It is generally shaped as a layer around $\mathcal{D}_{\text{opt}}$; see Section I.3.3.

- $\Gamma_{\text{in}} \subset \partial\mathcal{D}$ is a region of the boundary of $\mathcal{D}$ accounting for the entrance of a waveguide into the device (this area is also called a port); see Section I.3.2.

- $\Gamma_{\text{out}} \subset \partial\mathcal{D}$ is an internal surface in $\partial\mathcal{D}$ used for the computation of the optimization objective; see Section I.3.4.

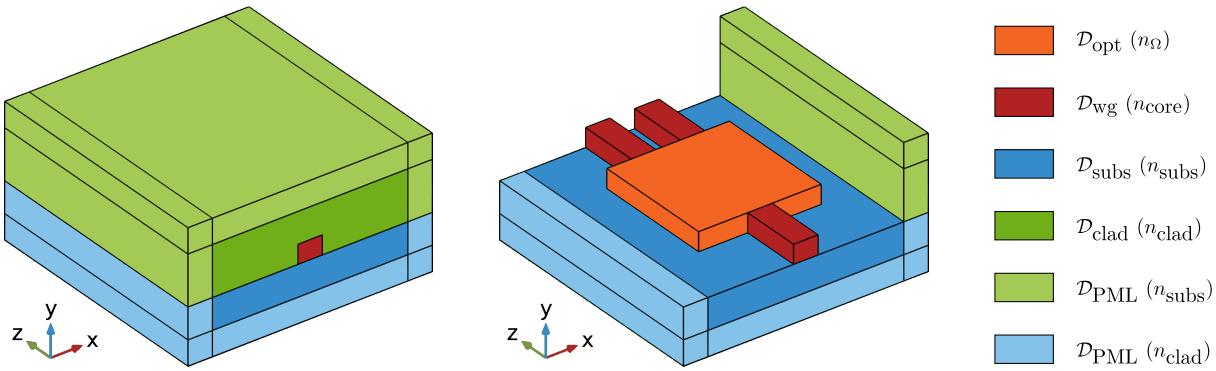
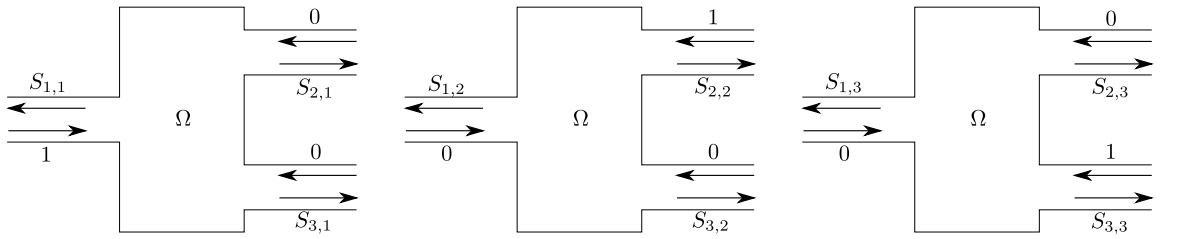


Figure I.3.2: Different subdomains associated to a nanophotonic component and their optical indices.

I.3.1.c Components as multi-ports

The behavior of a nanophotonic component is fully described by its **scattering matrix**. This type of representation exploits the linearity of Maxwell's equations, gathering all the required information to find, in a single product, the outgoing electromagnetic field inside each waveguide from the datum of the incoming electric field.



(a) *Injection from left* (b) *Injection from top-right* (c) *Injection from bottom-right*

Figure I.3.3: A multi-port nanophotonic component and its scattering parameters.

To define this matrix we first need to introduce the **S-parameters** featured in Fig. I.3.3. An S-parameter $S_{i,j}$ is defined using two indices i, j referring respectively to an ingoing and outgoing modes of two waveguides (possibly the same).

To better understand, consider a 2-port component where both waveguides are single-mode (meaning that only one forward/backward guided mode exists) but with different sizes. When injecting the fundamental mode $\boldsymbol{\mathcal{E}}_1^{\text{left}}$ into the left guide, the output electric field on the right waveguide is given by Eq. (I.2.6) as $\mathbf{E} = \alpha_{-1}^{\text{right}} \boldsymbol{\mathcal{E}}_{-1}^{\text{right}}$ (no forward mode and we ignore the radiative ones) whereas on the left waveguide it is $\mathbf{E} = \boldsymbol{\mathcal{E}}_1^{\text{left}} + \alpha_{-1}^{\text{left}} \boldsymbol{\mathcal{E}}_{-1}^{\text{left}}$. Each coefficients $\alpha_{-1}^{\text{right}}, \alpha_{-1}^{\text{left}}$ are given by the relation (I.2.12) and we will denote them respectively by $S_{2,1}$ and $S_{1,1}$. In the same way we can inject the fundamental mode of the right waveguide and look at the same coefficients α_i. In this case $\alpha_{-1}^{\text{right}}$ (resp. $\alpha_{-1}^{\text{left}}$) is referred as $S_{2,2}$ (resp. $S_{1,2}$). The output power transported by each modes when injecting any combination of the forward modes ($I^{\text{left}} \boldsymbol{\mathcal{E}}_1^{\text{left}}$ and $I^{\text{right}} \boldsymbol{\mathcal{E}}_1^{\text{right}}$) is then easily obtained performing

$$\begin{pmatrix} \alpha_{-1}^{\text{left}} \\ \alpha_{-1}^{\text{right}} \end{pmatrix} = \begin{pmatrix} S_{1,1} & S_{1,2} \\ S_{2,1} & S_{2,2} \end{pmatrix} \begin{pmatrix} I^{\text{left}} \\ I^{\text{right}} \end{pmatrix}. \tag{I.3.1}$$

For general components, $|S_{i,j}|^2$ is equal to the power transported by the backward mode number j when injecting the forward mode number i (note that the numbering is arbitrary and does not necessarily match the index of the modes).

$$S = \begin{pmatrix} S_{1,1} & \dots & S_{1,N} \\ \vdots & \ddots & \vdots \\ S_{N,1} & \dots & S_{N,N} \end{pmatrix} \in \mathrm{M}_{N,N}(\mathbb{C}) \tag{I.3.2}$$

Note also that since there must be no energy generation by the component we have for all i:

$$\sum_{j=1}^{N} |S_{i,j}|^2 \leq 1. \tag{I.3.3}$$

Moreover if we consider a lossless component (for all input vector $\mathbf{x}$ the output power $\|S\mathbf{x}\|_2^2$ is equal to the input power $\|\mathbf{x}\|_2^2$) then its associated scattering matrix should be unitary (that is $SS^{*,\top} = I_N$ where I_N is the identity matrix of size N).

Remark I.3.1.1: Most of the time we will only have to consider a submatrix of Eq. (I.3.2), the full scattering matrix being especially useful when looking at the global behavior of a circuit. Indeed, to evaluate the performance of a sequence of components, as some electric field may partly be converted into irrelevant output guided modes by each individual component, it is important to know how these wrong modes will be converted in each component to obtain the correct final output value of the whole circuit.

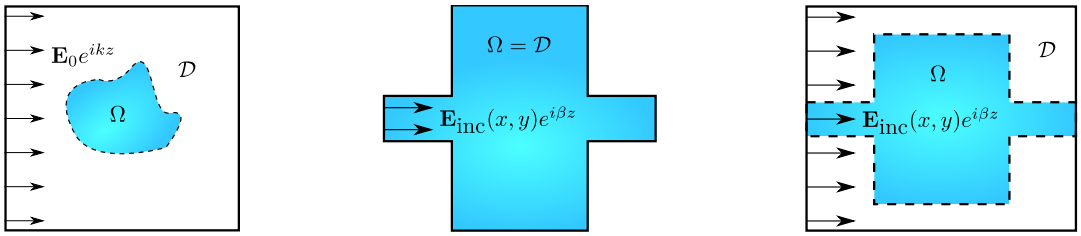
I.3.2 PDE with Dirichlet-to-Neumann boundary condition

In this subsection we deal with the derivation of a boundary condition which accounts for the injection of a wave inside the computational domain $\mathcal{D}$ from the border $\Gamma_{\text{in}} \subset \mathcal{D}$ where a waveguide is located. Generally speaking we are looking at the effect of an incoming incident wave $\mathbf{E}_{\text{inc}}$ into the physical domain $\mathcal{D}$ (where a component Ω is located). This

excitation will result in the emergence of a reflected field $\mathbf{E}_{\text{ref}}$ in such a way that the total electric field is then defined as

$$\mathbf{E} = \mathbf{E}_{\text{inc}} + \mathbf{E}_{\text{ref}}. \tag{I.3.4}$$

Many boundary conditions accounting for this effect are given in the literature and we propose here to give a short review of the most significant ones with an emphasis on their conditions of validity (with respect to the physical context). This subsection is limited to the case of flat borders with injection in the $\hat{z}$ direction (some references for curved borders may be found in [Jin14, Section 9.2]).



(a) *Plane wave injection* (b) *Mode in a closed waveguide* (c) *Mode in an open waveguide*

Figure I.3.4: Different configurations for the wave injection inside a component Ω (in blue) studied within a domain $\mathcal{D}$ (whose border $\partial\mathcal{D}$ is represented by solid black lines). Figures in (x, z)-plane view.

I.3.2.a Plane wave injection

Suppose that we are in the situation of Fig. I.3.4(a) and that we seek to inject an electromagnetic plane wave coming from a homogeneous area (say, air) into the physical domain $\mathcal{D}$. The incident field is therefore $\mathbf{E}_{\text{inc}}(x, y, z) = \mathbf{E}_0 e^{ikz}$ where k is the wavenumber and $\mathbf{E}_0 \in \mathbb{R}^3$ the injected field's amplitude. A direct calculation gives the relation

$$\hat{z} \times \nabla \times \mathbf{E}_{\text{inc}} = -ik\hat{z} \times \mathbf{E}_{\text{inc}} \times \hat{z} \quad \text{or equivalently} \quad \hat{z} \times \mathbf{H}_{\text{inc}} = -Z_0 \hat{z} \times \mathbf{E}_{\text{inc}} \times \hat{z} \tag{I.3.5}$$

where $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the impedance of free space. Supposing that the reflected field $\mathbf{E}_{\text{ref}}$ has the same profile as $\mathbf{E}_{\text{inc}}$ (but is an outgoing wave) we have $\mathbf{E}_{\text{ref}} = R\mathbf{E}_0 e^{-ikz}$ where $R \in \mathbb{R}$ is the reflection coefficient. From Eq. (I.3.4) applying $\hat{z} \times \nabla \times$ we then infer that on the border Γ_{in}

$$\hat{z} \times \nabla \times \mathbf{E} + ik\hat{z} \times \mathbf{E} \times \hat{z} = 2ik\hat{z} \times \mathbf{E}_{\text{inc}} \times \hat{z} \tag{I.3.6}$$

which is a Robin-like kind of equation known as the first order **Scattering Boundary Condition (SBC)**. The use of Eq. (I.3.6) makes it possible to inject a plane wave into $\mathcal{D}$ while allowing the plane wave to exit with arbitrary amplitude; the condition is said to be transparent for outgoing plane waves.

I.3.2.b General wave injection

When a more general form of the incoming wave is considered (such as the modes presented in Section I.2.1) the incident field is defined as $\mathbf{E}_{\text{inc}}(x, y, z) = \mathbf{E}_{\text{inc}}(x, y)e^{ikz}$ and Eq. (I.3.5) may be incorrect (and so does Eq. (I.3.6)).

Remark I.3.2.1: In two dimensions though, Eq. (I.3.6) is still working. Indeed, considering E_y (or H_y depending on the polarization) we have $\mathrm{E}_{\text{inc}}(x, z) = \mathrm{E}_{\text{inc}}(x)e^{ikz}$

and the relation $\partial_z \mathrm{E}_y = ik\mathrm{E}_y$ which lead to the following 2d-counterpart of Eq. (I.3.5)

$$\partial_z \mathrm{E}_y + ik\mathrm{E}_y = 2ik\mathrm{E}_y. \tag{I.3.7}$$

To address this problem, in [Tsi11] the authors proposed to modify Eq. (I.3.5) in such a way that the relation remain true but with a non-scalar impedance Z (that is, a 3×3 complex matrix) instead of Z_0. Indeed we can show that for any vector field such as a mode $(\boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{H}})$ we have

$$\hat{z} \times \boldsymbol{\mathcal{H}} = -Z\hat{z} \times \boldsymbol{\mathcal{E}} \times \hat{z} \quad \text{with} \quad Z = \begin{pmatrix} \mathcal{E}_x(x,y)/\mathcal{H}_y(x,y) & 0 & 0 \\ 0 & -\mathcal{E}_y(x,y)/\mathcal{H}_x(x,y) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{I.3.8}$$

And again we find that

$$\hat{z} \times \nabla \times \mathbf{E} + ikZ\hat{z} \times \mathbf{E} \times \hat{z} = 2ikZ\hat{z} \times \boldsymbol{\mathcal{E}} \times \hat{z}. \tag{I.3.9}$$

I.3.2.c Mode injection with scattering boundary condition

The problem with all the previous boundary conditions is that they only take into account reflected waves with profiles proportional to the one of the incident wave. For some devices this may be true (for instance in a single-mode waveguide with a sufficiently large domain of simulation) but fails to hold in general since different waves (or propagating modes) may be excited by a single incoming mode.

To allow for more general propagative modes we propose here to adapt the boundary condition of [Jin14, Section 11.1.2] to the case of open waveguides. Injecting a forward propagating mode inside the waveguide is equivalent to consider only one forward mode $\boldsymbol{\mathcal{E}}_n$ in the decomposition Eq. (I.2.6), that is to say (ignoring radiative modes) to impose that $\mathbf{E}$ is of the form

$$\mathbf{E} = \boldsymbol{\mathcal{E}}_n + \sum_{i=1}^{N} \alpha_{-i} \boldsymbol{\mathcal{E}}_{-i} \quad \text{on} \quad \Gamma_{\text{in}}. \tag{I.3.10}$$

By analogy this is equivalent as taking $\mathbf{E}_{\text{inc}} = \boldsymbol{\mathcal{E}}_n$ and $\mathbf{E}_{\text{ref}} = \sum_{i=1}^{N} \alpha_{-i} \boldsymbol{\mathcal{E}}_{-i}$ in Eq. (I.3.4). Unlike the previous examples, the scattered field $\mathbf{E}_{\text{ref}}$ considered here includes all propagation modes.

Invoking the orthogonality condition Eq. (I.2.8) and considering the normalization of Eq. (I.2.11) it comes

$$\alpha_{-i} = -\frac{1}{2} \int_{\Gamma_{\text{in}}} \left[(\mathbf{E} - \boldsymbol{\mathcal{E}}_n) \times \boldsymbol{\mathcal{H}}^*_{-i} \right] \cdot \hat{z} \, \mathrm{d}s \tag{I.3.11}$$

Applying $\hat{z} \times \nabla \times$ to Eq. (I.3.10) it follows that

$$\hat{z} \times \nabla \times \mathbf{E} = \hat{z} \times \nabla \times \boldsymbol{\mathcal{E}}_n - \frac{1}{2} \sum_{i=1}^{N} \int_{\Gamma_{\text{in}}} \left[(\mathbf{E} - \boldsymbol{\mathcal{E}}_n) \times \boldsymbol{\mathcal{H}}^*_{-i} \right] \cdot \hat{\mathbf{z}} \, \mathrm{d}s \, \hat{\mathbf{z}} \times \nabla \times \boldsymbol{\mathcal{E}}_{-i} \tag{I.3.12}$$

Using the relations $\hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_{-i} = -\hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_i$ between forward and backward propagating modes and the fact that $\nabla \times \mathbf{E} = i\omega\mu\mathbf{H}$ we finally find that

$$\hat{\mathbf{z}} \times \nabla \times \mathbf{E} + \frac{1}{2} i\omega\mu \sum_{i=1}^{N} \int_{\Gamma_{\text{in}}} [\mathbf{E} \times \boldsymbol{\mathcal{H}}^*_i] \cdot \hat{\mathbf{z}} \, \mathrm{d}s \, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_i =$$

$$i\omega\mu \, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_n + \frac{1}{2} i\omega\mu \sum_{i=1}^{N} \int_{\Gamma_{\text{in}}} [\boldsymbol{\mathcal{E}}_n \times \boldsymbol{\mathcal{H}}^*_i] \cdot \hat{\mathbf{z}} \, \mathrm{d}s \, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_i = 2i\omega\mu \, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_n.$$

The above equation accounts for a non-local Robin-like boundary condition of the form

$$\hat{\mathbf{z}} \times \nabla \times \mathbf{E} + \gamma(\mathbf{E}) = \mathbf{U} \quad \text{on} \quad \Gamma_{\text{in}} \tag{I.3.13}$$

where the operator γ and the field $\mathbf{U}$ are given by

$$\gamma(\mathbf{E}) = \frac{1}{2} i\omega\mu \sum_{i=1}^{N} \int_{\Gamma_{\text{in}}} [\mathbf{E} \times \boldsymbol{\mathcal{H}}_i^*] \cdot \hat{\mathbf{z}} \, \mathrm{d}s \, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_i \quad \text{and} \quad \mathbf{U} = 2i\omega\mu \, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_n. \tag{I.3.14}$$

Equation (I.3.13) maps the Dirichlet data $\boldsymbol{\mathcal{E}}_n$ into a Neumann boundary condition for $\mathbf{E}$ and will therefore be refered as a **Dirichlet-to-Neumann (DtN)** boundary condition.

I.3.3 Approximation of boundary conditions at infinity

The natural boundary condition accounting for the behavior of the electric field $\mathbf{E}$ at infinity was defined in Section I.1.3; see the Silver Müller radiation condition given by Eq. (I.1.14). Since numerical calculations takes place in a bounded computational domain $\mathcal{D} \subset \mathbb{R}^3$, there is the need to impose artificial boundary conditions on $\partial\mathcal{D}$ which mimic Eq. (I.1.14) without inducing too much reflection. There are several ways to achieve this goal.

I.3.3.a Absorbing boundary condition

The first choice to approximate Silver Müller's radiation condition is to no longer apply it to infinity but on the computational domain's edges. Applying $\mathbf{n}\times$ to Eq. (I.1.14) and dividing by $|\mathbf{x}|$ leads to

$$\mathbf{n} \times \nabla \times \mathbf{E} + ik \, \mathbf{n} \times \mathbf{E} \times \mathbf{n} = 0. \tag{I.3.15}$$

This boundary condition is called the first-order **Absorbing Boundary Condition (ABC)** and by comparison with Eq. (I.3.6) we see that it is equivalent to impose that the electric field on Γ_{in} is an outgoing plane wave escaping from the domain $\mathcal{D}$ along the direction of the boundary normal vector (for higher order ABC the reader may have a look at [Jin14, Section 9.2]).

I.3.3.b Perfectly Matched Layer

We now give a quick explanation about the **Perfectly Matched Layer (PML)** method. Contrary to the ABC, this method does not involve a new type of boundary condition on $\partial\mathcal{D}$ so to speak; it rather relies on a thin, "perfectly matched" layer $\mathcal{D}_{\text{PML}} \subset \mathcal{D}$ made of an artificial, absorbing material with the following properties:

1. Any electromagnetic wave, regardless of its angle of incidence, can penetrate inside $\mathcal{D}_{\text{PML}}$ without causing reflection into $\mathcal{D}$.

2. The amplitude of any electromagnetic wave propagating inside $\mathcal{D}_{\text{PML}}$ decreases exponentially fast to 0.

If both properties are fulfilled, imposing any type of homogeneous boundary conditions on $\partial\mathcal{D}$ for the electric field $\mathbf{E}$ – for instance the usual Dirichlet condition $\mathbf{n} \times \mathbf{E} = 0$ (known as a PEC condition in Section I.1.3.c) or the previously explained ABC – ensures a suitable approximation of the radiation condition Eq. (I.1.14).

Let us now briefly discuss the construction of such a perfectly matched layer, referring to [Jin14, Section 9.6] or [Mon03, Section 13.5.3.1] for details. Recalling that the computational domain $\mathcal{D}$ is a box with size $2w_x \times 2w_y \times 2w_z$ (see Section I.3.1), the perfectly matched layer $\mathcal{D}_{\text{PML}}$ is defined by:

$$\mathcal{D}_{\text{PML}} = \mathcal{D} \backslash [-w_x + \ell, w_x - \ell] \times [-w_y + \ell, w_y - \ell] \times [-w_z + \ell, w_z - \ell],$$

where ℓ is the thickness of the layer. Depending on where the input and output waveguides are located some part of $\mathcal{D}_{\text{PML}}$ may be removed (see Fig. I.3.5).

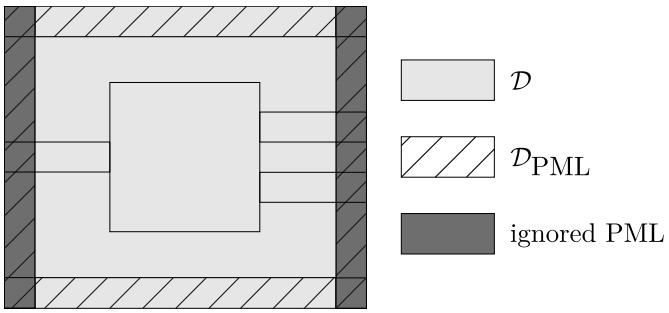


Figure I.3.5: A nanophotonic component together with a perfectly matched layer.

We then define for each component $\iota = x, y, z$:

$$\sigma_\iota(s) = \begin{cases} 1 & \text{if} \quad |s| \leq w_\iota - \ell, \\ 1 + \frac{i\lambda}{k} \frac{1}{w_\iota - |s|} & \text{if} \quad |s| > w_\iota - \ell, \end{cases} \tag{I.3.16}$$

and thence the anisotropic tensor Λ by:

$$\Lambda(\mathbf{x}) = \begin{pmatrix} \sigma_x^{-1}(x)\sigma_y(y)\sigma_z(z) & 0 & 0 \\ 0 & \sigma_x(x)\sigma_y(y)^{-1}\sigma_z(z) & 0 \\ 0 & 0 & \sigma_x(x)\sigma_y(y)\sigma_z(z)^{-1} \end{pmatrix}, \ \mathbf{x} \in \mathcal{D}. \tag{I.3.17}$$

In particular, $\Lambda(\mathbf{x})$ coincides with the identity when $\mathbf{x} \notin \mathcal{D}_{\text{PML}}$.

Finally, the electric field $\mathbf{E}$ is sought as the solution to Eq. (I.1.10), in which ε and μ are replaced by the tensors fields $\varepsilon\Lambda$ and $\mu\Lambda$, respectively; see Eq. (I.4.7) below.

I.3.3.c Absorbing boundary condition on waveguides

Even though the radiation condition accounts for the correct behavior of the electric field at infinity, this is only true if the medium is homogenenous in all directions after a finite distance from the component. In the presence of (infinitely long) waveguides on the borders of the domain (such as the ones in Fig. I.3.5), we can no longer consider the Silver-Müller radiation condition and therefore the boundary condition of Eq. (I.3.15) does not provide the correct behavior of the field on Γ_{out} (see [Ott17] for a more in-depth discussion and analysis on this subject). In a nutshell the problem with the radiation condition is that it suppose decaying fields at infinity while actually a lossless waveguide can propagate the fields indefinitely as guided modes.

To alleviate this problem we can add a perfectly matched layer after Γ_{out} or consider the same analysis as the one in Section I.3.2.c. Since no guided modes are entering the

component through Γ_{out} we can decompose the electric field by ignoring the backward propagating modes, which gives

$$\mathbf{E} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\mathcal{E}}_i \quad \text{on} \quad \Gamma_{\text{out}}.$$

We can then show that imposing this decomposition is equivalent to the boundary condition

$$\hat{z} \times \nabla \times \mathbf{E} + \gamma(\mathbf{E}) = 0 \quad \text{on} \quad \Gamma_{\text{out}} \tag{I.3.18}$$

with γ defined in Eq. (I.3.14).

Remark I.3.3.1: In practice placing a PML after the output side Γ_{out} has the advantage of not requiring to consider a finite number of modes but can cause spurious reflections. In Chapter III and particularly Remark III.2.1.3 we will see that it is interesting to consider the DtN boundary condition Eq. (I.3.18) instead of a PML since it allows for a simpler mathematical analysis.

I.3.4 Quantities of interest

In Section I.2.3 we showed that the outgoing power conveyed by the i-th mode is given by $|\alpha_i|^2$ with α_i defined in Eq. (I.2.12). In Section I.3.1 the scattering matrix was introduced as a means to represent the behavior of a nanophotonic component and for which we have seen that the scattering matrix coefficients are exactly the α_i. In many cases, the primary objective in nanophotonic will therefore be to find the shape Ω of the component such that its scattering matrix is equal to a target matrix. Returning to the two-ports example of Section I.3.1 (where we wanted to convert one mode into another) the target matrix would likely be

$$S_{\text{obj}} = \begin{pmatrix} 0 & - \\ 1 & - \end{pmatrix} \tag{I.3.19}$$

where the $-$ corresponds to values that do not interest us. In order to obtain this matrix we will have to find the optimal design $\Omega \subset \mathbb{R}^3$ of the core material such that, in this case, $|S_{2,1}|^2$ is as close as possible to 1. This amounts to maximize

$$\mathcal{J}(\Omega) = \left| \frac{1}{4} \int_{\Gamma} \left[\mathbf{E}_\Omega \times \boldsymbol{\mathcal{H}}_{-1}^{\text{out},*} + \mathbf{H}_\Omega \times \boldsymbol{\mathcal{E}}_{-1}^{\text{out},*} \right] \cdot \hat{z} \, \mathrm{d}s \right|^2. \tag{I.3.20}$$

Note that there is no intrinsic need to simultaneously minimize $|S_{1,1}|^2$ since by definition it should be equal to 0 if $|S_{2,1}|^2 = 1$. Since only outgoing guided modes may exists on Γ using the DtN boundary condition of Section I.3.2, Eq. (I.3.20) simplifies into

$$\mathcal{J}(\Omega) = \left| \frac{1}{2} \int_{\Gamma} \left[\mathbf{E}_\Omega \times \boldsymbol{\mathcal{H}}_{-1}^{\text{out},*} \right] \cdot \hat{z} \, \mathrm{d}s \right|^2. \tag{I.3.21}$$

Remark I.3.4.1: It is interesting to note that this objective does not depend on the phase $\phi \in \mathbb{R}$ of the injected mode. Indeed if $\boldsymbol{\mathcal{H}}_n$ is changed into $\boldsymbol{\mathcal{H}}_n e^{i\phi z}$ it is easy to see that a solution to Eqs. (I.1.9) and (I.3.13) is simply shifted in the same way (meaning that the solution is $\mathbf{E}_\Omega e^{i\phi z}$ with $\mathbf{E}_\Omega$ the unshifted solution) and therefore lead to the same objective (I.3.21) due to the absolute value.

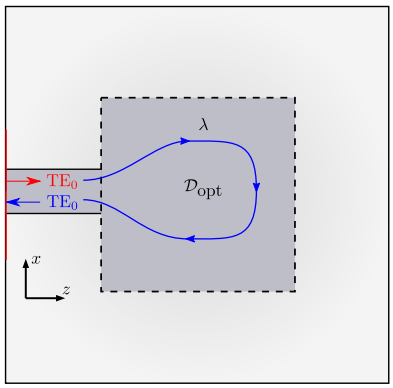
From this example we can easily infer why it is interesting to study the following mathematical program

$$\begin{cases} \max\limits_{\Omega\subset\mathbb{R}^3} & \left|\frac{1}{2}\int_\Gamma [\mathbf{E}_\Omega \times \boldsymbol{\mathcal{H}}^*]\cdot\hat{z}\,\mathrm{d}s\right|^2 \\ \textbf{s.t.} & \mathbf{E}_\Omega \text{ solution of Eqs. (I.1.9) and (I.3.13)} \end{cases} \tag{I.3.22}$$

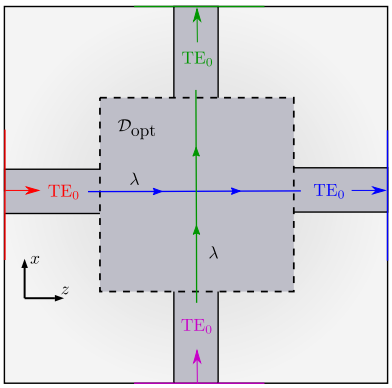
where Ω refer to the core material distribution inside the domain of interest $\mathcal{D}$. In Chapter III multiple components will be considered and most of them will be resembling Eq. (I.3.22).

I.3.5 Some examples of nanophotonic devices

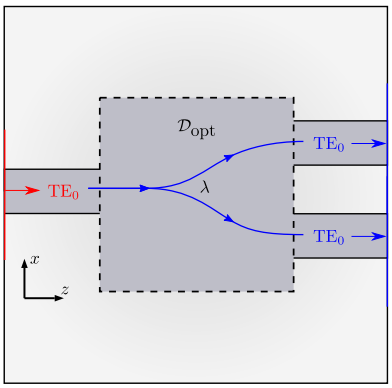
We conclude this section by presenting a number of conventional nanophotonic components that will be studied in Chapters III and IV.



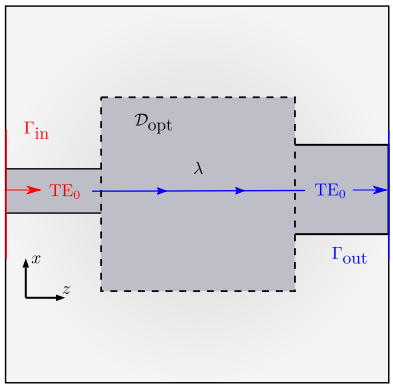
(a) *A mirror; this device seek to redirect the forward mode injected into the component into the same mode propagating in the opposite direction*



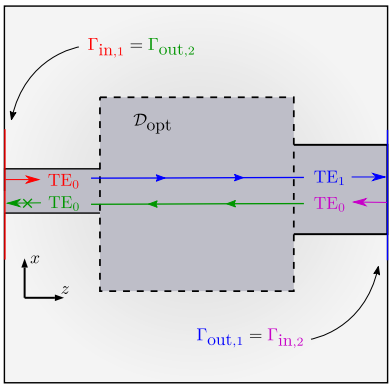
(b) *A crossing; this device allow the lossless transport of light in a straight manner from one waveguide to the one located on the other side*



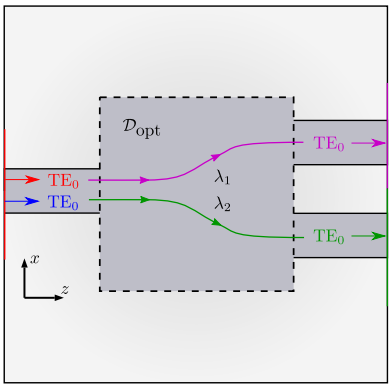
(c) *A power divider; the purpose of this device is to equally divide the input power into two output (or more) waveguides*



(d) *A mode converter; this device convert the input mode coming from the left into an other mode on the right waveguide*



(e) *A diode; the goal of this device is to allow the transmission of the TE_0 mode from the left to right waveguide while preventing the transfer of the TE_1 mode from the right into the TE_0 on the left*



(f) *A diplexer; the function of this device is to redirect the light coming from the left waveguide into one of the two waveguides on the right depending on the wavelength of the incident's light*

Figure I.3.6: Patchwork of several nanophotonic devices. The arrows shows where the light should be redirected depending on the input waveguide and wavelength.

For all the components in Fig. I.3.6, we are searching for a shape Ω inside the domain $\mathcal{D}_{\text{opt}}$ in order to optimize their respective objective functions which are all based on the optimization program Eq. (I.3.22). All these components and more will be studied in Section III.3.1 or Section IV.2.2.

I.4 Mathematical aspects

We now move on to the mathematical context. In particular we describes the variational framework that is necessary to study the time-harmonic vector wave equation as well as for the implementation of a finite element scheme to solve this PDE.

I.4.1 Functional spaces

In order to give the variational equation verified by the electric field we first need to review the functional spaces of interest in electromagnetism. We will not recall here the basics on variational formulations of PDE, Sobolev spaces and classical trace theorems (the reader is referred to [Mon03, Chapter 2], [Bre10, Chapter 5, 8, 9] or the very good introductory book [Man18, Section I.5 & Chapter II] for these details) except in the case of the $H(\text{div})$ and $H(\boldsymbol{curl})$ spaces which are extensively used in electromagnetism simulations and which are less classical from our viewpoint.

To understand why these spaces are important we will need the following integration by parts formulas:

- For any smooth vector fields $\boldsymbol{\phi}, \boldsymbol{\psi} \in (C^\infty(\Omega, \mathbb{R}))^3$ a direct calculation show that

$$\int_{\partial\Omega} (\mathbf{n} \times \boldsymbol{\phi}) \cdot (\mathbf{n} \times \boldsymbol{\psi} \times \mathbf{n}) \, \mathrm{d}s = \int_\Omega \nabla \times \boldsymbol{\phi} \cdot \boldsymbol{\psi} \, \mathrm{d}\mathbf{x} - \int_\Omega \boldsymbol{\phi} \cdot \nabla \times \boldsymbol{\psi} \, \mathrm{d}\mathbf{x}. \tag{I.4.1}$$

- For any smooth vector field $\boldsymbol{\phi} \in (C^\infty(\Omega, \mathbb{R}))^3$ and scalar field $\psi \in C^\infty(\Omega, \mathbb{R})$:

$$\int_{\partial\Omega} (\boldsymbol{\phi} \cdot \mathbf{n}) \psi \, \mathrm{d}s = \int_\Omega \boldsymbol{\phi} \cdot \nabla \psi \, \mathrm{d}\mathbf{x} + \int_\Omega (\nabla \cdot \boldsymbol{\phi}) \psi \, \mathrm{d}\mathbf{x}. \tag{I.4.2}$$

I.4.1.a $H(\boldsymbol{curl})$: the natural space for the electric field

As usual, to find the variational formulation associated to a PDE such as the time-harmonic vector wave equation $\nabla \times \nabla \times \mathbf{E} - k^2 n^2 \mathbf{E} = 0$ we proceed by multiplying this equation with a test function $\boldsymbol{\phi}$ and integrate over the whole simulation domain $\mathcal{D}$. Using Eq. (I.4.1) we then obtain that $\mathbf{E}$ must verify for all test function $\boldsymbol{\phi}$

$$\int_{\mathcal{D}} \nabla \times \mathbf{E} \cdot \nabla \times \boldsymbol{\phi}^* - k^2 n^2 \mathbf{E} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} + \int_{\partial\mathcal{D}} \mathbf{n} \times \nabla \times \mathbf{E} \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s = 0. \tag{I.4.3}$$

If we ignore for the moment the boundary integral on $\partial\Omega$ (see Section I.4.2.a below) we see that in order for the integrals in Eq. (I.4.3) to be well-defined, a sufficient condition is that both $\nabla \times \mathbf{E}, \nabla \times \boldsymbol{\phi}$ and $\mathbf{E}, \boldsymbol{\phi}$ are elements in $L^2(\mathcal{D}, \mathbb{C})$. This leads to the definition of the following fundamental (Sobolev) functional space to study electromagnetic fields, namely $H(\boldsymbol{curl})$.

Definition I.4.1.1 – The $H(\boldsymbol{curl})$ functional space.
For any smooth domain $\mathcal{D} \subset \mathbb{R}^3$ the space $H(\boldsymbol{curl}, \mathcal{D})$ is defined as

$$H(\boldsymbol{curl}, \mathcal{D}) = \left\{ \boldsymbol{\phi} \in (L^2(\mathcal{D}, \mathbb{C}))^3, \nabla \times \boldsymbol{\phi} \in (L^2(\mathcal{D}, \mathbb{C}))^3 \right\},$$

where the curl is taken in the sense of distribution meaning that $\nabla \times \boldsymbol{\phi}$ is defined as the only $\boldsymbol{g} \in (L^2(\mathcal{D}, \mathbb{C}))^3$ such that for all $\boldsymbol{\psi} \in (C_c^\infty(\mathcal{D}, \mathbb{C}))^3$ we have

$$\int_{\mathcal{D}} \boldsymbol{\phi} \cdot \nabla \times \boldsymbol{\psi}^* \, \mathrm{d}\mathbf{x} = \int_{\mathcal{D}} \boldsymbol{g} \cdot \boldsymbol{\psi}^* \, \mathrm{d}\mathbf{x}.$$

When equipped with the following inner product, the space $H(\boldsymbol{curl}, \mathcal{D})$ is a Hilbert space

$$\langle \boldsymbol{\phi}, \boldsymbol{\psi} \rangle_{H(\boldsymbol{curl}, \mathcal{D})} = \int_{\mathcal{D}} \boldsymbol{\phi} \cdot \boldsymbol{\psi}^* \, \mathrm{d}\mathbf{x} + \int_{\mathcal{D}} \nabla \times \boldsymbol{\phi} \cdot \nabla \times \boldsymbol{\psi}^* \, \mathrm{d}\mathbf{x}. \tag{I.4.4}$$

Physically $H(\boldsymbol{curl}, \mathcal{D})$ is "natural" in electromagnetism analyses as it is equal to the set of electromagnetic fields with finite energy; indeed in Section I.2.3 we defined the electromagnetic energy as $\omega_{\text{em}} = \varepsilon/2|\mathbf{E}|^2 + \mu/2|\mathbf{H}|^2$ whereas from Eq. (I.4.4) we see that

$$\|\mathbf{E}\|^2_{H(\boldsymbol{curl}, \mathcal{D})} = \int_{\mathcal{D}} |\mathbf{E}|^2 + |\nabla \times \mathbf{E}|^2 \, \mathrm{d}\mathbf{x} = \int_{\mathcal{D}} |\mathbf{E}|^2 + \omega^2 \mu_0^2 |\mathbf{H}|^2 \, \mathrm{d}\mathbf{x}, \tag{I.4.5}$$

which is equal (ignoring the finite factors ω, μ and ε) to the integral of ω_{em} on the whole domain $\mathcal{D}$.

I.4.1.b $H(\mathrm{div})$: natural space of $n^2\mathbf{E}$

If the electric field has more regularity than being a mere element of $H(\boldsymbol{curl})$ – for instance if it is solution of the strong formulation of the time-harmonic vector wave equation – then $\mathbf{E}$ verify Eq. (I.4.3) for all $\boldsymbol{\phi} \in (C^\infty(\mathcal{D}, \mathbb{R}))^3$. In particular, it is also true for $\boldsymbol{\phi} = \nabla\phi$ with $\phi \in C_c^\infty(\mathcal{D}, \mathbb{R})$. In this case, since $\nabla \times \nabla = 0$, Eq. (I.4.3) reduce to

$$\int_{\mathcal{D}} (n^2\mathbf{E}) \cdot (\nabla\phi^*) \, \mathrm{d}\mathbf{x} = 0.$$

Using now the integration by part formula Eq. (I.4.2) together with the fact that $\phi = 0$ on $\partial\mathcal{D}$ this results in

$$\int_{\mathcal{D}} (\nabla \cdot (n^2\mathbf{E}))\phi^* \, \mathrm{d}\mathbf{x} = 0.$$

In other words, if $\mathbf{E}$ is sufficiently regular then the weak divergence of $n^2\mathbf{E}$ is well-defined. This brings us to the consideration of the following space

Definition I.4.1.2 – The $H(\mathrm{div})$ functional space.
For any smooth domain $\mathcal{D} \subset \mathbb{R}^3$ the space $H(\mathrm{div}, \mathcal{D})$ is defined as

$$H(\mathrm{div}, \mathcal{D}) = \left\{ \boldsymbol{\phi} \in (L^2(\mathcal{D}, \mathbb{C}))^3, \nabla \cdot \boldsymbol{\phi} \in (L^2(\mathcal{D}, \mathbb{C}))^3 \right\}$$

where the divergence is taken in the sense of distribution meaning that $\nabla \cdot \boldsymbol{\phi}$ is defined as the only $g \in L^2(\mathcal{D}, \mathbb{C})$ such that for all $\psi \in C_c^\infty(\mathcal{D}, \mathbb{C})$ we have

$$\int_{\mathcal{D}} \boldsymbol{\phi} \cdot \nabla\psi^* \, \mathrm{d}\mathbf{x} = \int_{\mathcal{D}} g\psi^* \, \mathrm{d}\mathbf{x}.$$

Again, together with the inner product

$$\langle \boldsymbol{\phi}, \boldsymbol{\psi} \rangle_{H(\mathrm{div},\mathcal{D})} = \int_{\mathcal{D}} \boldsymbol{\phi} \cdot \boldsymbol{\psi}^* \, \mathrm{d}\mathbf{x} + \int_{\mathcal{D}} \nabla \cdot \boldsymbol{\phi} \nabla \cdot \boldsymbol{\psi}^* \, \mathrm{d}\mathbf{x},$$

the shape $H(\mathrm{div}, \mathcal{D})$ is a Hilbert space.

I.4.2 Traces theorems

We are now going to focus on the traces theorems associated to the spaces $H(\mathrm{div})$ and $H(\boldsymbol{curl})$. These theorems will allow us to manipulate the values of electromagnetic fields on the edges of a domain and to find the results stated in Section I.1.3.b.

I.4.2.a Tangential components in $H(\boldsymbol{curl})$

Let us start with the tangential component of the electric field on the border $\partial\Omega$ of a smooth domain Ω. As usual, trace theorems will be defined using an extension by continuity of an integration by part formula.

If the integral in the right hand side of Eq. (I.4.1) makes sense – that is for example the case if $\boldsymbol{\phi} \in H(\boldsymbol{curl}, \Omega)$ and $\boldsymbol{\psi} \in C^\infty(\Omega, \mathbb{R})$ – then we can define the action of the tangential components of $\boldsymbol{\phi}$ on $\mathbf{n} \times \boldsymbol{\psi}$ over the border $\partial\Omega$ through Eq. (I.4.1). This observation results in the following theorem which gives a precise definition of the tangential components of an element in $H(\boldsymbol{curl})$.

Theorem I.4.2.1 – Green theorem for $H(\boldsymbol{curl})$.
Let $\Omega \subset \mathbb{R}^3$ be a smooth domain. Then

(1) The mapping

$$\gamma_t : (C^\infty(\Omega, \mathbb{C}))^3 \to (C^\infty(\Omega, \mathbb{C}))^3, \ \gamma_t(\boldsymbol{\phi}) = (\mathbf{n} \times \boldsymbol{\phi})_{|\partial\Omega}$$

can be extended by continuity into a linear map from $H(\boldsymbol{curl}, \Omega)$ into the space $Y(\partial\Omega) \subset (H^{-1/2}(\partial\Omega))^3$ (the precise definition of $Y(\partial\Omega)$ is given in [Mon03, Remark 3.32]).

(2) In the same way the mapping

$$\gamma_T : (C^\infty(\Omega, \mathbb{C}))^3 \to (C^\infty(\Omega, \mathbb{C}))^3, \gamma_T(\boldsymbol{\phi}) = (\mathbf{n} \times \boldsymbol{\phi} \times \mathbf{n})_{|\partial\Omega}$$

can be extended into a linear map from $H(\boldsymbol{curl}, \Omega)$ into $(Y(\partial\Omega))'$.

(3) For any $\boldsymbol{\phi} \in H(\boldsymbol{curl}, \Omega)$ and $\boldsymbol{\psi} \in Y(\partial\Omega)$ there exists $\bar{\boldsymbol{\psi}} \in H(\boldsymbol{curl}, \Omega)$ such that $\boldsymbol{\psi} = \gamma_t(\bar{\boldsymbol{\psi}})$ and the following linear functional is well-defined

$$\langle \boldsymbol{\psi}, \gamma_T(\boldsymbol{\phi}) \rangle_{Y(\partial\Omega),(Y(\partial\Omega))'} = \int_\Omega \nabla \times \boldsymbol{\phi} \cdot \bar{\boldsymbol{\psi}}^* \, \mathrm{d}\mathbf{x} - \int_\Omega \boldsymbol{\phi} \cdot \nabla \times \bar{\boldsymbol{\psi}}^* \, \mathrm{d}\mathbf{x}. \tag{I.4.6}$$

A proof may be found in [Mon03, Theorem 3.31]. For simplicity we will refer to the tangential component of the electric field using the integral notation of Eq. (I.4.1) instead of the duality product in the left hand side of Eq. (I.4.6). $\langle \cdot, \cdot \rangle_{Y(\partial\Omega),(Y(\partial\Omega))'}$.

I.4.2.b Normal component in $H(\mathrm{div})$

Now we move on to the trace theorem associated with the $H(\mathrm{div})$ space. As for $H(\boldsymbol{curl})$ we can extend the integration by parts formula of Eq. (I.4.2) to give meaning of the normal component of an element $\boldsymbol{\phi}$ in $H(\mathrm{div})$ on the border of a domain as stated in the following theorem.

Theorem I.4.2.2 – Green theorem for $H(\mathrm{div})$.
Let $\Omega \subset \mathbb{R}^3$ be a smooth domain. Then

(1) The mapping

$$\gamma_n : (C^\infty(\bar{\Omega}, \mathbb{C}))^3 \to C^\infty(\bar{\Omega}, \mathbb{C}),\ \gamma_n(\boldsymbol{\phi}) = (\boldsymbol{\phi} \cdot \mathbf{n})_{|\partial\Omega}$$

can be extended by continuity into a linear map from $H(\mathrm{div}, \Omega)$ into $H^{-1/2}(\partial\Omega)$.

(2) The following Green's formula is valid for any $\boldsymbol{\phi} \in H(\mathrm{div}, \Omega)$ and $\psi \in H^1(\Omega)$:

$$\int_{\partial\Omega} \gamma_n(\boldsymbol{\phi})\psi \, \mathrm{d}s = \int_\Omega \boldsymbol{\phi} \cdot \nabla\psi \, \mathrm{d}\mathbf{x} + \int_\Omega (\nabla \cdot \boldsymbol{\phi})\psi \, \mathrm{d}\mathbf{x}.$$

Remembering that in Section I.4.1.b we showed that $n^2\mathbf{E}$ is an element of $H(\mathrm{div})$, Th. I.4.2.2 therefore imply that the normal component of $n^2\mathbf{E}$ is well-defined on the border of a shape.

We can summarize the information given by the two previous trace theorems as follows:

- Since any electric field is searched as an element of $H(\boldsymbol{curl})$ then its tangential components $\mathbf{n} \times \mathbf{E} \times \mathbf{n}$ are well-defined on a border $\partial\Omega$.
- If an electric field $\mathbf{E}$ is smooth inside a domain Ω then the normal component of $n^2\mathbf{E}$ is well-defined on $\partial\Omega$.

I.4.3 Variational formulation

We now have all the necessary elements to present the full variational formulation which we have to solve in order to simulate the electric field inside a nanophotonic component. Starting from the general PDE:

$$\begin{aligned}
\nabla \times \left(\Lambda^{-1}\nabla \times \mathbf{E}\right) - k^2 n^2 \Lambda \mathbf{E} &= 0 \quad \text{in} \quad \mathcal{D} \\
\mathbf{n} \times \mathbf{E} &= 0 \quad \text{on} \quad \partial\mathcal{D}\backslash(\Gamma_{\text{in}} \cup \Gamma_{\text{out}}) \\
\mathbf{n} \times \nabla \times \mathbf{E} + \frac{1}{2}\sum_{j=1}^{N} i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_j^{\text{in}} \int_{\Gamma_{\text{in}}} \left[\mathbf{E} \times \boldsymbol{\mathcal{H}}_j^{\text{in},*}\right] \cdot \mathbf{n} \, \mathrm{d}s &= \mathbf{U} \quad \text{on} \quad \Gamma_{\text{in}} \\
\mathbf{n} \times \nabla \times \mathbf{E} + \frac{1}{2}\sum_{j=1}^{N} i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_j^{\text{out}} \int_{\Gamma_{\text{out}}} \left[\mathbf{E} \times \boldsymbol{\mathcal{H}}_j^{\text{out},*}\right] \cdot \mathbf{n} \, \mathrm{d}s &= 0 \quad \text{on} \quad \Gamma_{\text{out}}
\end{aligned} \tag{I.4.7}$$

with $\mathbf{U} = 2i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_m^{\text{in}}$ (where m is the index of the injected mode) and Λ given by Eq. (I.3.17). Multiplying the first equation by $\boldsymbol{\phi}^*$, integrating other $\mathcal{D}$ and using Green's formula (I.4.1), we see that $\mathbf{E}$ is the solution of the following variational formulation for all $\boldsymbol{\phi} \in \mathcal{V}$ with $\mathcal{V} = \{\boldsymbol{\phi} \in H(\boldsymbol{curl}, \mathcal{D}), \mathbf{n} \times \boldsymbol{\phi} = 0 \text{ on } \partial\mathcal{D}\backslash(\Gamma_{\text{in}} \cup \Gamma_{\text{out}})\}$

$$a(\mathbf{E}, \boldsymbol{\phi}) = b(\boldsymbol{\phi}) \tag{I.4.8}$$

where the sesquilinear form $a : \mathcal{V} \times \mathcal{V} \to \mathbb{C}$ and antilinear map $b : \mathcal{V} \to \mathbb{C}$ are given by

$$\begin{aligned} a(\mathbf{E}, \boldsymbol{\phi}) = &\int_{\mathcal{D}} \Lambda^{-1} \nabla \times \mathbf{E} \cdot \nabla \times \boldsymbol{\phi}^* - k^2 n^2 \Lambda \mathbf{E} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} \\ &- \int_{\Gamma_{\text{in}}} \frac{1}{2} \sum_{j=1}^{N} i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_j^{\text{in}} \int_{\Gamma_{\text{in}}} \left[\mathbf{E} \times \boldsymbol{\mathcal{H}}_j^{\text{in},*} \right] \cdot \mathbf{n} \, \mathrm{d}s \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}t \\ &- \int_{\Gamma_{\text{out}}} \frac{1}{2} \sum_{j=1}^{N} i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_j^{\text{out}} \int_{\Gamma_{\text{in}}} \left[\mathbf{E} \times \boldsymbol{\mathcal{H}}_j^{\text{out},*} \right] \cdot \mathbf{n} \, \mathrm{d}s \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}t \quad \text{(I.4.9)} \end{aligned}$$

$$\text{and} \quad b(\boldsymbol{\phi}) = 2i\omega\mu_0 \int_{\Gamma_{\text{in}}} \mathbf{n} \times \boldsymbol{\mathcal{H}}_m^{\text{in}} \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}t. \quad \text{(I.4.10)}$$

Remark I.4.3.1: We do not provide a demonstration regarding the existence and uniqueness of a solution to Eq. (I.4.8) but we will have to assume that this is the case in Chapter III.

I.4.3.a Dirichlet-to-Neumann boundary condition

We present here a minor modification of the boundary conditions presented in Sections I.3.2.c and I.3.3.c which leads to great improvements of the finite element matrix sparsity (this method seems to be presented in [Zhu06, Section 6.3]). The idea is to impose the same boundary condition for the test function $\boldsymbol{\phi}$ on both Γ_{in} and Γ_{out} meaning that we have

$$\mathbf{E} = \boldsymbol{\mathcal{E}}_m^{\text{in}} + \sum_{j=1}^{N} \alpha_{-j}^{\text{in}} \boldsymbol{\mathcal{E}}_{-j}^{\text{in}} \quad \text{and} \quad \boldsymbol{\phi} = \sum_{j=1}^{N} \beta_{-j}^{\text{in}} \boldsymbol{\mathcal{E}}_{-j}^{\text{in}},$$

with no backward propagating modes. Putting this expression in Eqs. (I.4.9) and (I.4.10) we find that the integrals on Γ_{in} may be simplified into

$$I_{\text{in}} = \int_{\Gamma_{\text{in}}} \left(2i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_m^{\text{in}} - \frac{1}{2} \sum_{j=1}^{N} i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_j^{\text{in}} \alpha_{-j}^{\text{in}} \right) \cdot \mathbf{n} \times \sum_{j=1}^{N} \beta_{-j}^{\text{in},*} \boldsymbol{\mathcal{E}}_{-j}^{\text{in},*} \times \mathbf{n} \, \mathrm{d}t \quad \text{(I.4.11)}$$

where $\alpha_{-j}^{\text{in}}, \beta_{-j}^{\text{in}}$ are defined through Eq. (I.2.12). Remembering the orthogonality relations presented in Section I.2.2 lead to

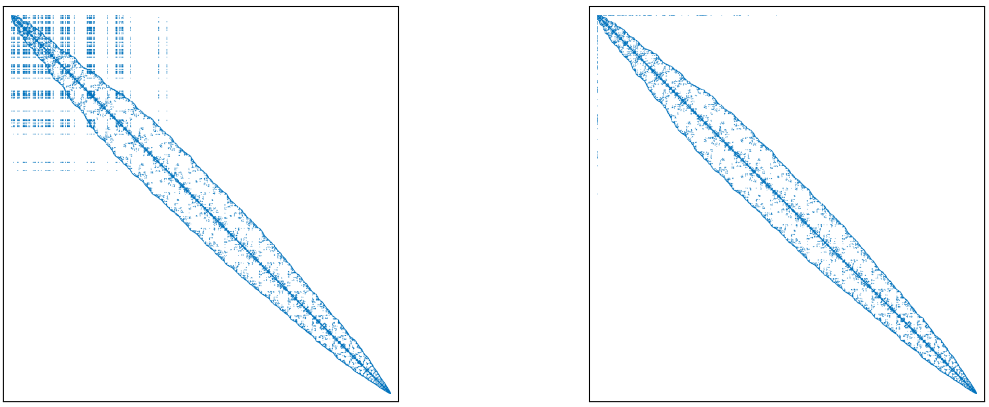
$$I_{\text{in}} = i\omega\mu_0 \sum_{j=1}^{N} (2\delta_{j,m} - \alpha_{-j}^{\text{in}}) \beta_{-j}^{\text{in},*}, \quad \text{(I.4.12)}$$

where $\delta_{j,m} = 1$ if $j = m$ and zero otherwise. The same simplification is applicable to the integral on Γ_{out} leading this time to

$$i\omega\mu_0 \sum_{j=1}^{N} \alpha_{-j}^{\text{out}} \beta_{-j}^{\text{out},*}.$$

Remark I.4.3.2: This representation is particularly useful once discretized using finite elements since it removes a lot of degrees of freedom and the associated matrix is sparse except for the (few) lines/columns linking α_i, β_i and $\mathbf{E}, \boldsymbol{\phi}$ together whereas using Eq. (I.4.8) results in a matrix where all the degree of freedom defined on Γ_{in} or

Γ_{out} are linked together. This difference is easily seen on Fig. I.4.1 where the matrices were constructed using only one mode ($N = 1$).



(a) *Using boundary condition Eq. (I.4.8), non-zeros elements* $\simeq 9 \times 10^6$

(b) *Using boundary condition Eq. (I.4.12), non-zeros elements:* $\simeq 8 \times 10^6$

Figure I.4.1: Finite element matrix sparse representation for the two representations of the DtN boundary condition.

All the examples presented in Chapter III use Eq. (I.4.12).

I.4.3.b Boundary conditions

To ensure that the boundary condition $\mathbf{n} \times \mathbf{E} = 0$ which is put at the external boundary of the PML is satisfied (we recall here that this boundary condition is known as a PEC and defined in Eq. (I.1.17)), we numerically add to the variational formulation the following term

$$10^{30} \int_{\partial \mathcal{D}_{\text{PML}}} \mathbf{n} \times \mathbf{E} \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s \tag{I.4.13}$$

which, once discretized using finite elements (see Section I.5.1), modifies the lines in the FEM system in such a way that the condition $\mathbf{n} \times \mathbf{E} = 0$ is necessarily verified numerically. Concerning the perfect magnetic conductor boundary condition (known as a PMC condition and defined by Eq. (I.1.18)), it is naturally verified by the electric field if no other integral on the corresponding edge is present in the variational formulation.

Remark I.4.3.3: If the physical configuration presents a symmetry (in terms of material properties and light excitation), say along an axis Γ, the numerical simulation can be realized on only one half of the computational domain by applying a PMC boundary condition on it. The same could be done for anti-symmetry around an axis using PEC instead of PMC, but is more rarely considered.

I.4.3.c Eigenvalue problem

The variational formulation associated to Eq. (I.2.2) is:

$$\int_{\Gamma} \beta^2 \left(\nabla_\tau \mathrm{E}_\perp \cdot \nabla_\tau \phi_\perp + \mathbf{E}_\parallel \cdot \boldsymbol{\phi}_\parallel - k^2 n^2 \mathrm{E}_\perp \cdot \phi_\perp \right) - k^2 n^2 \mathbf{E}_\parallel \cdot \boldsymbol{\phi}_\parallel + \nabla_\tau \mathbf{E}_\parallel \cdot \nabla_\tau \boldsymbol{\phi}_\parallel \, \mathrm{d}\mathbf{x}, \tag{I.4.14}$$

where Γ is a two-dimensional section of a waveguide. This generalized eigenvalue problem should be solved for $\beta \in \mathbb{C}$ and $\mathbf{E} \in H(\boldsymbol{curl}, \Gamma)$ with the additional Dirichlet boundary

condition that $\mathbf{n} \times \mathbf{E} = 0$ on $\partial\Gamma$. In practice we did not consider PML to compute the modes but rather a sufficiently large domain since the PMLs add complex values in the optical indices leading to complex-valued β even though a guided mode in a lossless medium should have a real propagation constant (see Section I.2.1.b). The reader is referred to [Lee91] for the full-presentation and implementation of a numerical solver of such eigenvalue problem.

I.4.4 Two dimensional approximation

I.4.4.a Functional space

In Section I.1.4, we have seen that in two dimensions, depending on the considered polarization, E_y and H_y are solutions of Eq. (I.1.19), that is to say a scalar Helmholtz equation. Interestingly since $(\mathrm{E}_x, \mathrm{H}_y, \mathrm{E}_z)$ and $(\mathrm{H}_x, \mathrm{E}_y, \mathrm{H}_z)$ are independent we easily see that the divergence-free conditions $\nabla \cdot (n^2\mathbf{E}) = \nabla \cdot \mathbf{H} = 0$ are always satisfied (in the TE case for instance we have $\partial_x \mathrm{H}_x = \partial_z \mathrm{H}_z = 0$ since only H_y is considered and $\partial_y \mathrm{H}_y = 0$ by definition of the two-dimensional approximation). With this in mind we derive that both E_y and H_y should simply be elements of

$$\mathcal{V}_{\mathrm{2D}} = H^1_{\partial\mathcal{D}_{\mathrm{PML}}}(\mathcal{D}, \mathbb{C}) = \{\phi \in L^2(\mathcal{D}, \mathbb{C}), \nabla\phi \in (L^2(\mathcal{D}, \mathbb{C}))^2, \phi_{|\partial\mathcal{D}_{\mathrm{PML}}} = 0\} \tag{I.4.15}$$

with, again, $\nabla\phi$ defined in the weak sense meaning that there exists $g_x, g_y \in L^2(\mathcal{D}, \mathbb{C})$ such that for all $\psi \in C_c^\infty(\mathcal{D}, \mathbb{C})$

$$\int_{\mathcal{D}} \phi \partial_x \psi^* \, \mathrm{d}\mathbf{x} = -\int_{\mathcal{D}} g_x \psi^* \, \mathrm{d}\mathbf{x} \quad \text{and} \quad \int_{\mathcal{D}} \phi \partial_y \psi^* \, \mathrm{d}\mathbf{x} = -\int_{\mathcal{D}} g_y \psi^* \, \mathrm{d}\mathbf{x}. \tag{I.4.16}$$

I.4.4.b Boundary conditions

In two dimensions, the approriate radiation condition is no longer given by Eq. (I.1.14) but rather the Sommerfeld condition

$$\lim_{\mathbf{x}\to\infty} \sqrt{|\mathbf{x}|} \left(\frac{\mathbf{x}}{|\mathbf{x}|} \cdot \nabla \mathrm{E}_y + ik\mathrm{E}_y \right) = 0. \tag{I.4.17}$$

This behavior at infinity of the electric field will once again be approximated by PML (see Section I.3.3). Concerning the injection of a mode we will this time simply consider Eq. (I.3.7) instead of the DtN boundary condition since, in 2D, it is sufficient to recover reasonable accuracy (see Remark I.3.2.1).

I.4.4.c Variational formulations

Using the previously formulated elements in this section, the full variational formulation in 2D for the E_y component of the electric field with PML and the boundary condition of Eq. (I.3.7) is given by

$$\int_{\mathcal{D}} \Lambda^{-1} \nabla \mathrm{E}_y \cdot \nabla \phi^* - k^2 n^2 \Lambda \mathrm{E}_y \phi^* \, \mathrm{d}\mathbf{x} + i\omega\mu_0 \int_{\Gamma_{\mathrm{in}} \cup \Gamma_{\mathrm{out}}} \mathrm{E}_y \phi^* \, \mathrm{d}\mathbf{x} = 2i\omega\mu_0 \int_{\Gamma_{\mathrm{in}}} \mathcal{E}_y^{\mathrm{in}} \phi^* \, \mathrm{d}s \tag{I.4.18}$$

where the PML matrix Λ is defined as

$$\Lambda(\mathbf{x}) = \begin{pmatrix} \sigma_x^{-1}(x)\sigma_y(y) & 0 \\ 0 & \sigma_x(x)\sigma_y(y)^{-1} \end{pmatrix}, \quad \mathbf{x} \in \mathcal{D}. \tag{I.4.19}$$

and σ_ι is given by Eq. (I.3.16).

I.5 Numerical aspects

Many numerical methods have been developed to solve either Eq. (I.1.9) or (I.1.6). In this thesis we only consider the two most widely used methods, namely the **Finite Element Method (FEM)** and the **Finite Difference Time Domain (FDTD)** method which are briefly presented in the next subsections.

I.5.1 Finite Element Method (FEM)

I.5.1.a Introduction

The FEM is a particular case of Galerkin approximation, which means that it searches for a solution of the variational formulation Eq. (I.4.8) in a finite-dimensional subset $H_h \subset H(\boldsymbol{curl}, \mathcal{D})$ (or $H^1(\mathcal{D})$ in 2D with Eq. (I.1.19)) which makes it possible to write the electric field as

$$\mathbf{E} = \sum_{i=1}^{d} \gamma_i \boldsymbol{\psi}_i \tag{I.5.1}$$

where γ_i are scalar values, d the dimension of H_h and $\boldsymbol{\psi}_i$ the **shape functions** (that is to say elements composing a basis of H_h). The index $h > 0$ is used to describe the approximation accuracy and it is assumed that H_h tends to $H(\boldsymbol{curl}, \mathcal{D})$ when $h \to 0$.

I.5.1.b Geometry meshing

To define the shape functions $\boldsymbol{\psi}_i$, a tetrahedral mesh of the domain $\mathcal{D}$ is used (see Fig. I.5.1) so that each $\boldsymbol{\psi}_i$ is different from zero only on a small number of vertices, faces or tetrahedron's volumes. This choice of support is important because it means that few functions will have a non-zero interaction with each other and it therefore leads to a sparse system of equations.

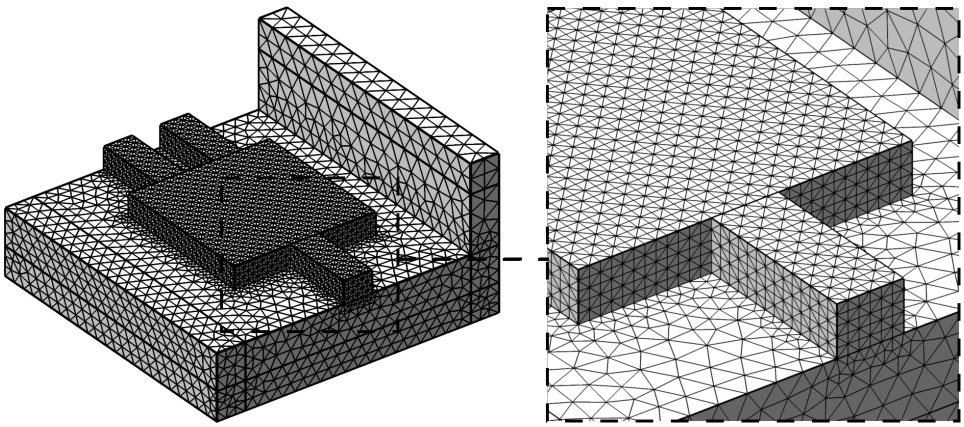


Figure I.5.1: Typical mesh of a nanophotonic component composed of approximately 100 000 tetrahedrons.

To simplify the meshing process and minimize the overall number of tetrahedrons we will most of the time rely on a mesh that does not exactly coincide everywhere with the edges of the shapes but which, using an index-smoothing method (see Section III.2.2), still gives consistent results. Regarding the elements size, it is commonly accepted that if a wave is propagating at the wavelength λ in a medium with optical index n then the tetrahedron's

edges must be of maximum size $\lambda/(5n)$ such that each oscillation of the wave may be properly discretized using at least 5 elements in the direction of propagation.

I.5.1.c Nédélec elements

Now that a mesh of the domain is defined let us present the (curl-conforming) edge elements of Nédélec and the corresponding shape functions. The full definition of these elements may be found in [Mon03, Section 5.5] or [Jin14, Section 8.3.2] but lets give here a glimpse of one of their most important properties.

As we have seen in both Sections I.1.3 and I.4.1, a solution $\mathbf{E}$ to the time-harmonic wave equation naturally have its tangential components $\mathbf{E}_\parallel$ continuous at the interfaces while $\mathrm{E}_\perp$ has the same property only if it is multiplied by n^2 (see Eq. (I.1.15)). The problem is that if we consider some subspace of $H(\boldsymbol{curl}, \mathcal{D})$ which only contains fully-continuous elements (such as $(P_k(\mathcal{D}))^3$ where P_k refer to the classic Lagrange elements of degree k) we numerically see the appearance of so-called "spurious" solutions. To solve this problem Nédélec notably introduced curl-conforming elements which are only imposing the continuity of tangential components of the electric field.

I.5.1.d Solving time-harmonic Maxwell equations

The solution $\mathbf{E} \in H_h$ of the variational formulation

$$a(\mathbf{E}, \boldsymbol{\phi}) = b(\boldsymbol{\phi}) \text{ for all } \boldsymbol{\phi} \in H_h$$

is, by definition of H_h, the same as the one obtained by solving

$$a(\mathbf{E}, \boldsymbol{\psi}_i) = b(\boldsymbol{\psi}_i) \text{ for all shape functions } \boldsymbol{\psi}_i.$$

By linearity, this is equivalent to finding the coefficients $(\gamma_i)_i = \boldsymbol{\gamma}$ such that:

$$\sum_{j=1}^{n} a(\boldsymbol{\psi}_j, \boldsymbol{\psi}_i)\gamma_j = b(\boldsymbol{\psi}_j) \text{ for all } i = 1, \dots, d. \tag{I.5.2}$$

That is to say the solution of the linear system $A\boldsymbol{\gamma} = b$ with $A = (a(\boldsymbol{\psi}_j, \boldsymbol{\psi}_i))_{i,j}$ and $b = (b(\boldsymbol{\psi}_i))_i$. To solve this system we rely on the direct sparse LU solver `MUMPS` (even though it requires a lot of RAM, it remains reasonable for the components described in Chapters III and IV) which was found to be faster than iterative methods based on GMRES or BiCGSTAB.

In practice the meshing and matrix assembly were done on either the commercial software `Comsol Multiphysics` (COMSOL AB, Stockholm, Sweden) for 3D simulations or `FreeFem++` [Hec12] for the 2D ones. For the records, nanophotonic components such as those in Chapter III were simulated using a mesh containing roughly 100 000 tetrahedrons and second-order Nédelec elements leading to approximately 500 000 degree of freedom. On the CEA 20-cores 3.3 Ghz cluster nodes with 128 GB of RAM this is solved in about 5 minutes.

I.5.2 Finite Difference Time Domain (FDTD)

Another widely popular method is the FDTD method (see [Sch10] for a good introduction and implementation details) which directly deals with the time-dependent Maxwell equations Eqs. (I.1.1) to (I.1.4). This method is useful when dispersion relations are present

(meaning that some parameters such as the optical indices depends on the wavelength) or when the user is looking at multiple wavelengths at the same time (the spectral response of the component). Even if most of our simulations uses the FEM, we resort to the FDTD on an ad hoc basis to confirm our results or to quickly obtain the full spectrum.

Since we only used FDTD through the commercial solvers `RSoft photonics` (RSoft Design Group, Inc., New York, U.S.A) and `Lumerical` (Lumerical Solutions, Inc., Vancouver, Canada) as a black-box method we will not dive into much details here (see [Sch10] for a presentation at introductory level). Nevertheless, it is important to be aware of two characteristics of this method:

- First, since finite differences are inherently defined on Cartesian grids, the core medium's shape Ω of the component may not be properly discretized and it raises the need to approximate the border $\partial\Omega$ using some kind of sub-pixel smoothing method (see Section III.2.2).
- The second point is about the injection of modes into the waveguides. Since FDTD considers time-dependent equations we now need to impose both the spatial profile and the temporal behavior of the mode. To do this, two approaches are possible; if we are only interested in a single frequency then this dependence could simply be $\exp(i\omega t)$ (a continuous wave) whereas if we try to study a whole interval of wavelengths then it is necessary to consider more than a single harmonic by injecting a time-dependence like $f(t)\exp(i\omega t)$ (a pulse). Once the wave has had enough time to propagate throughout the component, performing a Fourier transform then provides the full spectral response, such as the power carried by a mode at every considered wavelengths.

The consistency between FEM and FDTD is investigated in Section III.5.2 with the simulation of a polarization rotator.

This page is intentionally left blank.

Chapter II

Geometric shape optimization

Summary — One of the primary objectives of this thesis is to implement a numerical method that automatically finds the optimal design of devices such as those presented in Section I.3. To achieve this goal, most optimization algorithms are based on an iterative process which produces increasingly efficient shapes (according to a given figure of merit) and results in a locally optimized design.

For the past ten to twenty years, several shape optimization methods have been applied in the context of nanophotonics (some of them will be presented in Section III.1.3) based either on binary discretization and genetic algorithms, continuous approximations of material properties or geometric approaches. In this thesis we have decided to focus on the latter since it allows us to easily handle geometric quantities which will prove to be particularly useful to deal with manufacturing uncertainties as we will see in Section IV.3.

The full mathematical details being only very seldom documented in photonics papers which rely on geometric shape optimization, this chapter starts with a comprehensive overview of Hadamard's shape derivative concept in Section II.1.1, providing in particular the necessary ingredients for the implementation of a "shape" gradient descent algorithm.

Sections II.3 and II.4 are then devoted to the presentation of the level set-based numerical framework considered in this thesis with a particular emphasis on numerical details that are of great importance for the stability and efficiency of the algorithm but often overlooked in scientific papers (with the noticeable exceptions of the doctoral theses [De 05, Chapter 8] and [Vié16, Part IV]).

A direct application of the elements presented in this chapter to nanophotonics will be discussed in Chapter III and an extension of this material for the optimization of the geometry between regions of shapes supporting different types of PDE's boundary condition will be proposed Chapter V.

II.1 The sensitivity of a physical objective according to an infinitesimal variation of a shape

In this section we introduce the concept of shape derivative in the sense of Hadamard as well as the Eulerian and Lagrangian derivatives of PDE constraint functional used to define physical figure of merits.

Most of the material presented in this section may be found in the reference books [All07; Hen06; Sok09; Pir82] or doctoral theses [Mic14; Dap13; De 05; Vié16].

II.1.1 Shape derivatives in the sense of Hadamard

II.1.1.a Introduction

In Chapter I we saw that finding the design of a component which maximizes some figure of merit (also referred to as objective function) is a problem of interest to physicists. Since these figure of merits depends on the shape Ω of the design, we define in this section the "gradient" of a general (shape) functional $\mathcal{J}$ which depends on a shape $\Omega \subset \mathbb{R}^d$. This definition being the most important element in the implementation of a gradient descent method, it will allow us to solve the general shape optimization problem

$$\max_{\Omega \subset \mathbb{R}^d} \mathcal{J}(\Omega). \tag{II.1.1}$$

It is actually possible to define this "gradient" in many different ways depending on what is meant by a "small variation" of a shape. From an algorithmic point of view, this choice influences the allowed deformations of the shape between each iterations and therefore limits those that can be obtained numerically.

In his memoir [Had08] Jacques Hadamard proposed to study the sensitivity of a shape functional $\mathcal{J}(\Omega)$ when the shape's contour $\partial\Omega$ is shifted

> *" by applying on its normals infinitely small lengths δn "* (translated from french).

In modern language, he proposed that the contour $\partial\Omega$ be slightly modified into $(\mathrm{Id} + \delta\mathbf{n})(\partial\Omega) = \{\mathbf{x} + \delta\mathbf{n}(\mathbf{x}), \mathbf{x} \in \partial\Omega\}$ where $\delta \in \mathbb{R}$ is a small real number and $\mathbf{n}$ is the normal vector to $\partial\Omega$. Note that the choice to consider only modifications along the normal vector may seem surprising at first sight but this is due to the fact that at first order, a tangential movement simply corresponds to a reparameterization of the edges as can be seen in Fig. II.1.1 (see also Th. II.1.1.3).

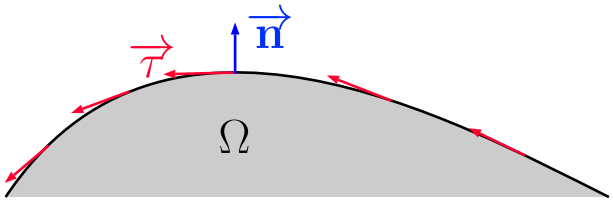


Figure II.1.1: A tangential displacement of $\partial\Omega$ is, at first order, negligible compared to a movement along the normal vector and is almost equivalent to a reparametrization.

This definition of a small variation of a shape was later considered and generalized in many works (see for instance the original paper of Murat & Simon [Mur75] or the books [Hen06] and [All07, Chapter 6]) as explained below.

For a given smooth shape $\Omega \subset \mathbb{R}^d$ we are going to continuously move its borders in the direction specified by a regular vector field of small amplitude. Defining $\boldsymbol{\theta} : \mathbb{R}^d \to \mathbb{R}^d$ as this vector field we obtain a variation $\Omega_{\boldsymbol{\theta}}$ of Ω by the operation (see Fig. II.1.2)

$$\Omega_{\boldsymbol{\theta}} = (\mathrm{Id} + \boldsymbol{\theta})(\Omega) = \{\mathbf{x} + \boldsymbol{\theta}(\mathbf{x}), \mathbf{x} \in \Omega\}. \tag{II.1.2}$$

This deformation is "small" insofar as $\boldsymbol{\theta}$ is "small" according to a given norm. The choice

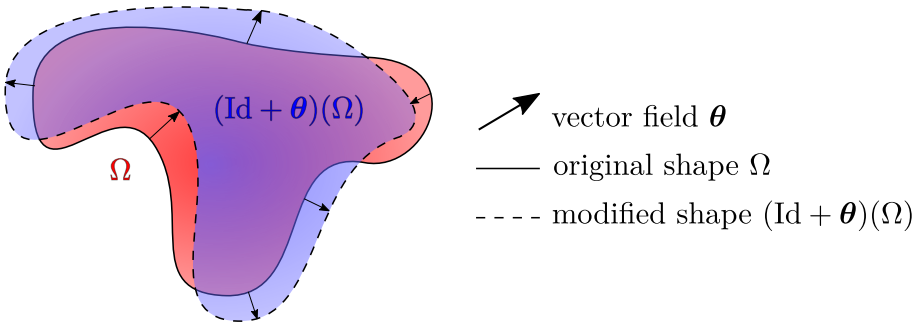


Figure II.1.2: Schematic representation of a variation of a shape Ω using Hadamard's method.

of this norm, and therefore of the associated functional space for $\boldsymbol{\theta}$, is not very important in practice (the fields will be chosen at least in $C^1(\mathbb{R}^d, \mathbb{R}^d)$ or even $C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ cf. Section II.4.2) but it is nevertheless important that the vector fields are at least continuous to ensure that the next theorems are true.

The most general space that can be chosen for $\boldsymbol{\theta}$ is the set of vector fields with bounded values and partial derivatives. More precisely we take $\boldsymbol{\theta} \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ the space of $L^\infty(\mathbb{R}^d)^d$ functions (essentially bounded functions) such that its weak partial derivatives exist and belong to $L^\infty(\mathbb{R}^d)^d$. This choice implies in particular that $\boldsymbol{\theta}$ is continuous. Endowing the space $W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ with the norm

$$\|\boldsymbol{\theta}\|_{W^{1,\infty}(\mathbb{R}^d,\mathbb{R}^d)} = \|\boldsymbol{\theta}\|_{L^\infty(\mathbb{R}^d)^d} + \|\nabla\boldsymbol{\theta}\|_{L^\infty(\mathbb{R}^d)^{d\times d}},$$

we obtain a Banach space. We will sometimes have to consider more regular deformations using the space $C^1_c(\mathbb{R}^d, \mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$.

Variations of shapes by a process of the form Eq. (II.1.2) turn out to be homeomorph to Ω, and in particular they share the same topology:

Theorem II.1.1.1 – Homeomorphism of $W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$.
Let $\boldsymbol{\theta} \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ with $\|\boldsymbol{\theta}\|_{W^{1,\infty}(\mathbb{R}^d,\mathbb{R}^d)} < 1$. The mapping $\mathrm{Id} + \boldsymbol{\theta}$ is an homeomorphism with $(\mathrm{Id} + \boldsymbol{\theta})^{-1} - \mathrm{Id} \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$.

The proof of this theorem is given for example in [All07, Lemma 6.13].

II.1.1.b Definition

With the previous choice of deformations the derivative of a shape-dependent functional is given in Def. II.1.1.1.

Definition II.1.1.1 – Differentiability of a shape functional.
Let $\mathcal{J} : \Omega \to \mathbb{R}$ be a given functional depending on a shape.

1. $\mathcal{J}$ is said to be shape differentiable at Ω if the functional
$$\mathcal{J}_\Omega : \boldsymbol{\theta} \mapsto \mathcal{J}((\mathrm{Id} + \boldsymbol{\theta})(\Omega))$$
is Fréchet differentiable in $W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ at 0.
2. The differential of $\mathcal{J}_\Omega$ at 0 is called the **shape derivative** of $\mathcal{J}$ at Ω and denoted by $\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \mathcal{J}'_\Omega(0)(\boldsymbol{\theta})$.

With these definitions the following first-order Taylor expansion of $\mathcal{J}$ holds for vector fields $\boldsymbol{\theta} \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$:

$$\mathcal{J}((\mathrm{Id} + \boldsymbol{\theta})(\Omega)) = \mathcal{J}(\Omega) + \mathcal{J}'(\Omega)(\boldsymbol{\theta}) + o(\boldsymbol{\theta}) \quad \text{where} \quad \lim_{\boldsymbol{\theta} \to 0} \frac{|o(\boldsymbol{\theta})|}{\|\boldsymbol{\theta}\|_{W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)}}.$$

Since the definition of the shape derivative is only valid for infinitely small perturbations of the reference shape, $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ is only defined using vector fields $\boldsymbol{\theta}$ with associated norm inferior to 1, meaning that this kind of sensitivity does not allow topology changes as explained in Th. II.1.1.1.

Remark II.1.1.1: The vector fields living only in a Banach space (and therefore no scalar product is available), there is, in general, no gradient associated to $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$. However, it is always possible to recover a gradient using the Hilbert projection method described in Section II.4.2.

II.1.1.c Application to integral functionals

By directly using the previous definitions we can determine the shape derivative of two types of functionals which are the basis for all the upcoming demonstrations.

Theorem II.1.1.2 – Shape derivatives of domain-dependent integrals.
Let Ω be an open, regular and bounded subset of $\mathbb{R}^d$, $f \in W^{1,1}(\mathbb{R}^d, \mathbb{R})$ and $g \in W^{2,1}(\mathbb{R}^d, \mathbb{R})$. Let

$$\mathcal{J}_1(\Omega) = \int_\Omega f(\mathbf{x}) \, \mathrm{d}\mathbf{x} \quad \text{and} \quad \mathcal{J}_2(\Omega) = \int_{\partial\Omega} g(s) \, \mathrm{d}s. \tag{II.1.3}$$

The functionals $\mathcal{J}_1$ and $\mathcal{J}_2$ are shape differentiable at Ω and we have the following shape derivatives for all $\boldsymbol{\theta}_1 \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ and $\boldsymbol{\theta}_2 \in C^1_c(\mathbb{R}^d, \mathbb{R}^d)$:

$$\mathcal{J}_1'(\Omega)(\boldsymbol{\theta}_1) = \int_{\partial\Omega} \boldsymbol{\theta}_1 \cdot \mathbf{n} \, f(s) \, \mathrm{d}s \quad \text{and} \quad \mathcal{J}_2'(\Omega)(\boldsymbol{\theta}_2) = \int_{\partial\Omega} \boldsymbol{\theta}_2 \cdot \mathbf{n} \, \left(\nabla g(s) \cdot \mathbf{n} + \kappa g(s)\right) \, \mathrm{d}s.$$

where $\kappa = \nabla \cdot \mathbf{n}$ is the mean curvature of $\partial\Omega$ and $\mathbf{n}$ the unitary, outer-pointing normal vector.

See [All07, Section 6.3.2] or [Hen06, Section 5.2] for the proof of Th. II.1.1.2 which is essentially based on a change of variables through the mapping $\mathbf{x} \mapsto (\mathrm{Id} + \boldsymbol{\theta})(\mathbf{x})$ that is only valid if it is a homeomorphism.

An interesting case is the one where $f = g = 1$. In this situation $\mathcal{J}_1$ corresponds to the volume of Ω while $\mathcal{J}_2$ represents its perimeter. Th. II.1.1.2 then gives the following derivatives

$$\mathcal{J}_1'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} \, \mathrm{d}s \quad \text{and} \quad \mathcal{J}_2'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \kappa \, \boldsymbol{\theta} \cdot \mathbf{n} \, \mathrm{d}s. \tag{II.1.4}$$

In other words, the most efficient way to minimize the volume of Ω (resp. the perimeter) consists in deforming its boundary using a vector field in the direction opposite to normal vectors (resp. opposite to normal vectors weighted by the value of the mean curvature).

II.1.1.d The structure theorem

In Th. II.1.1.2 we have seen that for two integrals functionals the shape derivative only considered the normal component $\boldsymbol{\theta} \cdot \mathbf{n}$ of the vector field on the border $\partial\Omega$ of the shape. The following theorem allows to extend this result for more general shape functionals.

Theorem II.1.1.3 – Structure theorem.
Let $\Omega \subset \mathbb{R}^d$ smooth and $\mathcal{J} : \Omega \to \mathbb{R}$ a shape differentiable functional with associate derivative $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$. Then there exists a linear form $l : \partial\Omega \to \mathbb{R}$ such that for all $\boldsymbol{\theta} \in C_c^1(\mathbb{R}^d, \mathbb{R}) \cap W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = l(\boldsymbol{\theta} \cdot \mathbf{n}). \tag{II.1.5}$$

In other words, the shape derivative only consider the normal component of the vector fields $\boldsymbol{\theta}$ on the border of the shape $\partial\Omega$. Another way to understand this result is that for any vector field $\boldsymbol{\theta}$ with compact support S such that $\partial\Omega \cap S = \emptyset$ we have $\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = 0$.

For the proof of this theorem see for instance [Hen06, Theorem 5.9.2].

II.1.1.e Summary of the shape optimization method: sketch of the algorithm

Let us put together everything we have presented so far by describing how we are going to be able, using Hadamard's method, to optimize an objective function $\mathcal{J}$ which depends on a shape Ω; namely to solve

$$\max_{\Omega \subset \mathbb{R}^d} \mathcal{J}(\Omega). \tag{II.1.6}$$

Considering a functional $\mathcal{J}$ as in Th. II.1.1.2 there exists a scalar field V_Ω such that:

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\mathcal{D}} \boldsymbol{\theta} \cdot \mathbf{n}\, V_\Omega(s)\, \mathrm{d}s.$$

Therefore, taking $\boldsymbol{\theta} = V_\Omega \mathbf{n}$ we find that

$$\mathcal{J}((\mathrm{Id} + \boldsymbol{\theta})(\Omega)) = \mathcal{J}(\Omega) + \int_{\partial\Omega} |V_\Omega(s)|^2\, \mathrm{d}s + o(\boldsymbol{\theta}).$$

Using a scalar τ small enough this implies that $\mathcal{J}((\mathrm{Id} + \tau V_\Omega \mathbf{n})(\Omega)) > \mathcal{J}(\Omega)$.

Starting from an initial shape Ω we can iteratively carry out the transformation $(\mathrm{Id} + \tau V_\Omega \mathbf{n})(\Omega)$ as described in Algorithm II.1.1 to obtain increasingly efficient shapes for the objective function $\mathcal{J}$.
Of course Algorithm II.1.1 does not present all the necessary ingredients to numerically implement this shape optimization algorithm, this is the purpose of Sections II.3 and II.4. The resulting fully-detailed algorithm is given at the end of the chapter in Section II.4.3.

II.1.1.f A note on the existence of global maximum in shape optimization problems

In Section II.1.1.e we proposed a general shape optimization algorithm to solve Eq. (II.1.6). As for any gradient-based method the convergence into a local maxima is ensured as soon as the step τ in Line 8 of Algorithm II.1.1 is chosen sufficiently small at each iteration.

Algorithm II.1.1: Sketch of the shape optimization algorithm to maximize an objective function.

```
begin (initialization)
    Ω := Ω_0 (initial shape);
    τ := 1 (the amplitude of the deformation);
repeat (optimization)
    θ := V_Ω the shape derivative scalar field (Eq. (II.2.9));
    τ := small amplitude for the gradient descent;
    Ω := (Id + τθn)(Ω);
until convergence;
return Ω;
```

Starting from different initial shapes Ω we may end up in different local maxima but an essential question then arises regarding the possibility to obtain a global maximum Ω^*;

$$\Omega^* = \underset{\Omega \subset \mathbb{R}^d}{\arg\max}\ \mathcal{J}(\Omega). \tag{II.1.7}$$

The problem with such a question is that it assumes the existence of such Ω^*. However, in general, the existence of a global maximum to this type of optimization problem is not guaranteed and it might be possible to build a sequence of shape with increasing efficiency for the objective function that does not converge to a shape $\Omega^* \subset \mathbb{R}^d$ (see on this topic [Hen06, Chapter 4] or [Pir82, Chapter 3]).

This also implies that the max in Eq. (II.1.6) does not really makes sense and should be replaced by a sup instead but we will keep this small abuse of notation later on.

II.1.2 Eulerian and Lagrangian derivatives

Even though Th. II.1.1.2 allows to find the shape derivative of integral functionals, its result does not apply to physically interesting figure of merits since the integrands f and g in Eq. (II.1.3) do not depend on the shape Ω.

This dependence is, however, of utmost importance to consider physical quantities (such as the electric field) that are inherently domain-dependent. Let us recall for instance that in Section I.3.4 we presented a general optimization program for photonic components which requires the resolution of Eq. (I.3.22), that is to maximize

$$\mathcal{J}(\Omega) = \left| \int_\Gamma \frac{1}{2} \left[\mathbf{E}_\Omega \cdot \boldsymbol{\mathcal{H}}^* \right] \cdot \mathbf{n} \, \mathrm{d}s \right|^2, \tag{II.1.8}$$

where $\mathbf{E}_\Omega$ is solution of the time-harmonic vector wave equation which depends on Ω through the values of the optical index.

In this section we will therefore describe the technique used to adapt Th. II.1.1.2 to integrals of the form

$$\mathcal{J}_1(\Omega) = \int_\Omega f(\mathbf{x}, \Omega) \, \mathrm{d}\mathbf{x} \quad \text{and} \quad \mathcal{J}_2(\Omega) = \int_{\partial\Omega} g(\mathbf{x}, \Omega) \, \mathrm{d}s. \tag{II.1.9}$$

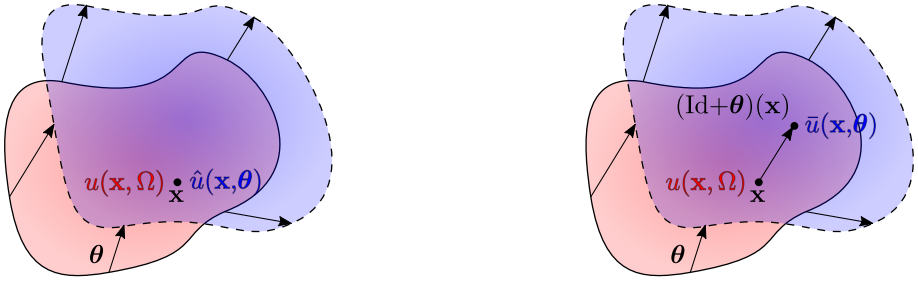
II.1.2.a Definition

To find the shape derivative of the functionals in Eq. (II.1.9), we need to give meaning to the derivative with respect to Ω of functions such as

$$u_\Omega : \mathcal{D} \to \mathbb{R},$$

which we assume to be an element of H, an Hilbert space defined on $\mathcal{D} \subset \mathbb{R}^d$ with $\Omega \subset \mathcal{D}$. This is the case of the electric field $\mathbf{E}_\Omega$ in Eq. (II.1.8) (as we have seen in Section I.4.3) which is defined as the solution to a PDE in $\mathcal{D} \subset \mathbb{R}^3$ using an optical index which depends on $\Omega \subset \mathcal{D}$.

To study these derivatives, two approaches are possible using either an Eulerian or Lagrangian framework as presented in Def. II.1.2.1, II.1.2.2 and Fig. II.1.3. These denominations will be explained in Remark II.1.2.1. Note that in general u_Ω may be defined only on Ω, thus requiring some modifications in the following definitions; see Remark II.1.2.2.



(a) *Eulerian point of view: we study the variation between the value of u before and after the deformation at the position $\mathbf{x}$*

(b) *Lagrangian point of view: we study the variation between the value of u at $\mathbf{x}$ before the deformation and after at $(\mathrm{Id}+\boldsymbol{\theta})(\mathbf{x})$*

Figure II.1.3: Difference between the Eulerian and Lagrangian points of view.

Definition II.1.2.1 – Eulerian derivative.

On the one hand we can study, for a given point $\mathbf{x}$ in space, the variation of the function $u_\Omega \in H$ at this fixed position $\mathbf{x}$ with respect to a modification of Ω (see Fig. II.1.3(a)). The new value of u_Ω at the same position is denoted by $\hat{u}_\Omega$ and defined as

$$\hat{u}_\Omega : W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d) \to H \quad \text{s.t.} \quad \hat{u}_\Omega(\boldsymbol{\theta}) = u_{(\mathrm{Id}+\boldsymbol{\theta})(\Omega)} \tag{II.1.10}$$

When it exists, the derivative of this mapping at 0, in the direction $\boldsymbol{\theta}$, is referred to as $u'_\Omega(\boldsymbol{\theta})$, the **Eulerian derivative** of u_Ω. In other words

$$u'_\Omega(\boldsymbol{\theta})(\mathbf{x}) = \mathrm{d}\hat{u}_\Omega(0)(\boldsymbol{\theta})(\mathbf{x}). \tag{II.1.11}$$

Notice that since we consider u_Ω as an element of Ω a subset of $\mathcal{D}$, then for any sufficiently small $\boldsymbol{\theta}$ with $\boldsymbol{\theta} = 0$ on $\partial\mathcal{D}$, the value of $\hat{u}_\Omega(\boldsymbol{\theta})$ is well defined and so does u'_Ω.

Definition II.1.2.2 – Lagrangian derivative.

Alternatively, one can follow the movement of $\mathbf{x} \in \mathcal{D}$ during the shape deformation and consider the variation of $u \in H$ along the path of $\mathbf{x}$ (see Fig. II.1.3(b)). The new

value of u_Ω at the modified position is denoted by $\bar{u}_\Omega$ and defined as

$$\bar{u}_\Omega : W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d) \to H \quad \text{s.t.} \quad \bar{u}_\Omega(\boldsymbol{\theta}) \mapsto u_{(\mathrm{Id}+\boldsymbol{\theta})(\Omega)} \circ (\mathrm{Id} + \boldsymbol{\theta}).$$

In this case, the derivative $\mathring{u}_\Omega(\boldsymbol{\theta})$ of $\bar{u}_\Omega$ at 0, in the direction $\boldsymbol{\theta}$, is called the **Lagrangian derivative** of u_Ω (also known as the material derivative of u_Ω), that is

$$\mathring{u}_\Omega(\boldsymbol{\theta})(\mathbf{x}) = \mathrm{d}\bar{u}_\Omega(0)(\boldsymbol{\theta})(\mathbf{x}). \tag{II.1.12}$$

Even though the Lagrangian derivative is more complex from a mathematical point of view, it seems more appropriate in shape optimization since it allows to study the precise behavior of the function u_Ω on interfaces such as $\partial\Omega$. Indeed, if $\boldsymbol{\theta} \neq 0$ and $\mathbf{x} \in \partial\Omega$ then $\mathbf{x} \notin (\mathrm{Id} + \boldsymbol{\theta})(\partial\Omega)$, hence the Eulerian derivative does not allow to track the variation of u_Ω on the border of Ω (it will be important in Chapter V when dealing with boundary conditions defined on $\partial\Omega$). Moreover, when u_Ω is only defined on Ω, the Lagrangian derivative is easier to handle; see Remark II.1.2.2.

Remark II.1.2.1: The strategies of Definitions II.1.2.1 and II.1.2.2 are named after two well-known viewpoints used, for instance, in fluid mechanics. When dealing with flows, depending on the applications, it is customary to observe either the velocity $v(\mathbf{x}, t)$ of the fluid at a given point $\mathbf{x} \in \mathbb{R}^d$ over the time $t \in \mathbb{R}$ or the evolution of a particle position $\chi(\mathbf{x}, t)$ (which begins at the location $\chi(\mathbf{x}, 0) = \mathbf{x}$ for $t = 0$) and therefore its individual speed $\partial_t \chi(\mathbf{x}, t) = v(\chi(\mathbf{x}, t), t)$.

When the quantity $v(\mathbf{x}, t)$ is considered, it is said that we are within the Eulerian frame of reference whereas considering $v(\chi(\mathbf{x}, t), t)$ falls down in the domain of Lagrangian coordinates.

A direct analogy between the time t in fluid mechanics and the transformation $\boldsymbol{\theta}$ in shape optimization is made if we let $\chi(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x} + \boldsymbol{\theta}(\mathbf{x})$ be the point $\mathbf{x}$ new position. Indeed if $w(\mathbf{x}, \boldsymbol{\theta}) = \hat{u}_\Omega(\boldsymbol{\theta})(\mathbf{x})$ is the Eulerian derivative at $\mathbf{x}$ then $w(\chi(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta}) = \bar{u}_\Omega(\boldsymbol{\theta})(\mathbf{x})$ is the Lagrangian one.

Remark II.1.2.2: If u_Ω is solution to a PDE which is only defined on Ω, then the Hilbert space H depends on Ω. This means that for any $\boldsymbol{\theta} \neq 0$, $\hat{u}_\Omega(\boldsymbol{\theta}) \notin H$.

Still, the definition of the Eulerian derivative may be adapted by modifying $\hat{u}_\Omega$ in Eq. (II.1.10) into $\hat{u}_{\Omega,\mathbf{x}}(\boldsymbol{\theta}) = u_{(\mathrm{Id}+\boldsymbol{\theta})(\Omega)}(\mathbf{x})$ (which is well defined for small enough vector fields $\boldsymbol{\theta}$) and $u'_\Omega(\boldsymbol{\theta})(\mathbf{x})$ as the associated derivative.

In this case, the Lagrangian derivative seems to give a more appropriate mathematical framework since it may still be defined globally as in Def. II.1.2.2 and does not require a definition at each position.

Nevertheless, whenever both $\hat{u}$ and $\bar{u}$ are well defined, they are linked through a simple formula. To show this we introduce $U_\Omega : \boldsymbol{\theta} \mapsto u_\Omega \circ (\mathrm{Id} + \boldsymbol{\theta})$ in such a way that a formal application of the chain rule gives:

$$\mathrm{d}\bar{u}_\Omega(0)(\boldsymbol{\theta}) = \mathrm{d}U_\Omega(0)(\boldsymbol{\theta}) + \mathrm{d}\hat{u}_\Omega(0)(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \nabla_{\mathbf{x}} u_\Omega + \mathrm{d}\hat{u}_\Omega(0)(\boldsymbol{\theta}).$$

Using Eqs. (II.1.11) and (II.1.12) this is equivalent to

$$\mathring{u}_\Omega(\boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \nabla_{\mathbf{x}} u_\Omega + u'_\Omega(\boldsymbol{\theta}). \tag{II.1.13}$$

Since the Lagrangian derivative is more rigorous from a mathematical point of view, it is common to find the Lagrangian derivative and then define the Eulerian one using Eq. (II.1.13) as

$$u'_\Omega(\boldsymbol{\theta}) = \mathring{u}_\Omega(\boldsymbol{\theta}) - \boldsymbol{\theta} \cdot \nabla_{\mathbf{x}} u_\Omega. \tag{II.1.14}$$

II.1.2.b Application to integral with shape-dependent integrands

Using the previous definitions we can now give the derivatives of Eq. (II.1.9) (see [All07, Prop 6.28]):

Theorem II.1.2.1 – Shape derivatives of domain-dependent integrals and integrands.
Let Ω be an open, regular and bounded subset of $\mathbb{R}^d$, $f_\Omega, g_\Omega \in L^1(\mathbb{R}^d, \mathbb{R})$ with $\mathring{g}_\Omega$ differentiable at 0 as a mapping from $C^1(\mathbb{R}^d, \mathbb{R}^d)$ into $L^1(\partial\Omega)$. Let

$$\mathcal{J}_1(\Omega) = \int_\Omega f_\Omega(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \text{and} \quad \mathcal{J}_2(\Omega) = \int_{\partial\Omega} g_\Omega(\mathbf{x})\,\mathrm{d}s. \tag{II.1.15}$$

The functionals $\mathcal{J}_1$ and $\mathcal{J}_2$ are shape differentiable at $\boldsymbol{\theta} = 0$ and we have the following shape derivatives for all $\boldsymbol{\theta}_1 \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$

$$\mathcal{J}'_1(\Omega)(\boldsymbol{\theta}_1) = \int_\Omega (\nabla \cdot \boldsymbol{\theta}_1) f_\Omega(\mathbf{x}) + \mathring{f}_\Omega(\boldsymbol{\theta}_1)\,\mathrm{d}\mathbf{x} = \int_{\partial\Omega} \boldsymbol{\theta}_1 \cdot \mathbf{n} f_\Omega(\mathbf{x})\,\mathrm{d}s + \int_\Omega f'_\Omega(\boldsymbol{\theta}_1)\,\mathrm{d}\mathbf{x}, \tag{II.1.16}$$

and for all $\boldsymbol{\theta}_2 \in C^1_c(\mathbb{R}^d, \mathbb{R}^d)$

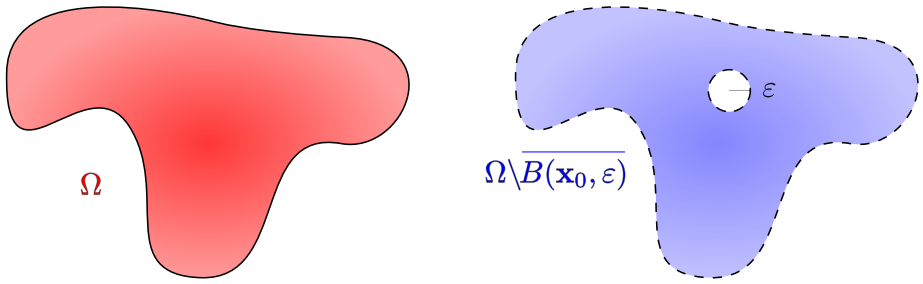
$$\begin{aligned}\mathcal{J}'_2(\Omega)(\boldsymbol{\theta}_2) &= \int_{\partial\Omega} g_\Omega(\mathbf{x})(\nabla \cdot \boldsymbol{\theta}_2 - \nabla\boldsymbol{\theta}_2 \mathbf{n} \cdot \mathbf{n}) + \mathring{g}_\Omega(\boldsymbol{\theta}_2)\,\mathrm{d}s \\ &= \int_{\partial\Omega} \boldsymbol{\theta}_2 \cdot \mathbf{n}\,(\nabla g_\Omega(\mathbf{x}) \cdot \mathbf{n} + \kappa g_\Omega(\mathbf{x}))\,\mathrm{d}s + \int_{\partial\Omega} g'_\Omega(\boldsymbol{\theta}_2)\,\mathrm{d}s. \end{aligned} \tag{II.1.17}$$

Note that using the Eulerian derivative the shape derivatives Eqs. (II.1.16) and (II.1.17) are exactly the ones expected using the chain rule applied formally and Th. II.1.1.2. As we will see in the next section, this sole theorem will allows us to find the shape derivative of all the functionals considered in this thesis. However, in practice, since both Eulerian and Lagrangian derivatives are difficult to obtain numerically we will resort to the adjoint method to obtain the shape derivatives without computing either u'_Ω or $\mathring{u}_\Omega$.

II.1.3 Other type of sensitivity: topological gradient

In this subsection we briefly discuss another type of shape sensitivity, namely the **topological gradient** which allows topological changes in the shape.

For more information about this method we refer to [Sok09, Chapter 1] or [Ams03, Chapter 1]. To better appraise the link between shape and topological derivatives we start this paragraph with a very crude attempt to reuse the previously defined shape derivative in the context of a topological modification which has the form of the removal of a sphere within the shape.



(a) *Unperturbed shape* Ω (b) *Modified shape* $\Omega_\varepsilon = \Omega\backslash\overline{B(\mathbf{x_0}, \varepsilon)}$

Figure II.1.4: Variation of the shape using the topological gradient.

To do this, first suppose that we are in the situation depicted in Fig. II.1.4(b). If we consider the shape $\Omega_\varepsilon = \Omega\backslash\overline{B(\mathbf{x_0}, \varepsilon)}$ then we can look at the shape derivative along the border of $B(\mathbf{x}_0, \varepsilon)$ and check if it should be better to dilate or erode this ball by looking at the sign of

$$\mathcal{J}'(\Omega_\varepsilon)(\mathbf{n}) = \int_{\partial B(\mathbf{x}_0,\varepsilon)} V_{\Omega_\varepsilon}(s)\,\mathrm{d}s$$

where V_Ω only depends on the values of u_Ω and v_Ω, the adjoint state as defined in Section II.2.1. From a numerical point of view we now make two approximations:

- For ε small enough the values of both u_{Ω_ε} and v_{Ω_ε} are constant on the border of the ball so that $\mathcal{J}'(\Omega_\varepsilon)(\mathbf{n}) \simeq 2\pi V_{\Omega_\varepsilon}(\mathbf{x}_0)$.
- If ε is small enough then the PDEs solutions (the state and the adjoint) using either Ω or Ω_ε should be equal so that $\mathcal{J}'(\Omega_\varepsilon)(\mathbf{n}) \simeq 2\pi V_\Omega(\mathbf{x}_0)$.

In other words considering only the simulation of the state and adjoint using Ω we can find if it would be better for Ω_ε to dilate or erode $B(\mathbf{x}_0, \varepsilon)$ by looking at the sign of $V_\Omega(\mathbf{x}_0)$ (negative values implying that a dilation would be better to increase the value of $\mathcal{J}(\Omega)$). Another way to interpret this result is the following if V_Ω is negative then it should be good to nucleate a hole in Ω at $\mathbf{x}_0$ since, if such a hole were to exist, then it would be better to expand it than to shrink it.

Of course, this conclusion does not result from a rigorous mathematical analysis and it proves wrong in many cases, even though it can be legitimate in some circumstances; see [Céa00]. In fact the most inaccurate assumption in the last paragraph was that we considered the shape derivative in the direction $\mathbf{n}$ using an amplitude of at most the hole's radii ε (remember that the shape derivative is not allowed to alter the shape's topology) which is also the magnitude of the perturbation caused by the removal of $\overline{B(\mathbf{x_0}, \varepsilon)}$. In other words, these two modifications (the edge of the hole in the normal direction and the drill of the hole itself) are both functions of ε and it is therefore not justified to disregard one of them without first checking that one of these perturbation is negligible compared to the other one (this is the case if they are not of the same order in ε).

A mathematically rigorous counterpart to the previous analysis relies on an asymptotic expansion of the objective function as

$$\mathcal{J}(\Omega_\varepsilon) = \mathcal{J}(\Omega) + f(\varepsilon)\mathcal{T}(\mathbf{x}_0) + R(f(\varepsilon)) \tag{II.1.18}$$

where $f(\varepsilon) \to 0$ and $R(f(\varepsilon)) = o(f(\varepsilon))$. In this case $\mathcal{T}(\mathbf{x}_0)$, which is called the topological gradient of the objective function, then gives information on where to remove balls in the

shape in order to optimize $\mathcal{J}$. Differentiating Eq. (II.1.18) by ε and dividing by $f'(\varepsilon)$ then yields

$$\mathcal{T}(\mathbf{x}_0) = \lim_{\varepsilon \to 0} \left(\frac{\varepsilon \mathcal{J}'(\Omega)(\mathbf{n})}{f'(\varepsilon)} - R'(f(\varepsilon)) \right). \tag{II.1.19}$$

And if we also have $R'(f(\varepsilon)) \to 0$ then $\mathcal{T}(\mathbf{x}_0) = V_\Omega(\mathbf{x}_0) \lim_{\varepsilon \to 0} \varepsilon f'(\varepsilon)^{-1}$. In the case of nanophotonics, proving a relation like Eq. (II.1.18) is very difficult; see Section III.3.2 where we give some optimization results using the value of the shape derivative as a means of knowing where to nucleate holes inside the shape.

II.2 How to find the shape derivative of a PDE constrainted functional

In Sections II.2.1 and II.2.2 we are interested in the shape derivatives of the following model physical objective in which Ω is a subset of the considered domain $\mathcal{D} \subset \mathbb{R}^d$ (this type of shape optimization problem where Ω is immersed into a larger domain is sometimes referred to as an interface optimization problem)

$$\mathcal{J}(\Omega) = \int_{\mathcal{D}} j(u_\Omega) \, \mathrm{d}\mathbf{x} \quad \text{with } u_\Omega \in H \text{ s.t.} \quad K_\Omega(u_\Omega) = 0 \tag{II.2.1}$$

where $u_\Omega \in H$ a given Hilbert space defined on the whole domain $\mathcal{D}$ and K_Ω a linear differential operator. Note here that to ensure the well-posedness of $\mathcal{J}$ we have to suppose the existence and uniqueness of a solution for all smooth Ω to the equation $K_\Omega(u_\Omega) = 0$ (additional regularity are also required in order to integrate j on $\mathcal{D}$ but we will ignore these details here). More precisely, we will consider the variational formulation associated to the PDE meaning that $u_\Omega \in H$ is solution of

$$a_\Omega(u_\Omega, v) = b_\Omega(v) \tag{II.2.2}$$

for all $v \in H$ where a (resp. b) is a bilinear (resp. linear) form.

According to Th. II.1.2.1 the shape derivative of Eq. (II.2.1) could be computed but it involves either the Lagrangian or Eulerian derivative of the mapping $\Omega \to u_\Omega$. Céa's method [Céa86] (also known as the adjoint or Lagrangian method) allows us to find the shape derivative of $\mathcal{J}$ without ever having to find this derivative but at the cost of a formal demonstration. The general structure of the complete proof is presented in Section II.2.2.

II.2.1 Céa's formal method

II.2.1.a General method

To find the derivative of Eq. (II.2.1) consider the following three steps:

1. Definition of the Lagrangian. Define a function $\mathcal{L}$, which will be referred as the Lagrangian, as the sum of the objective and the considered PDE's variational form (see Remark II.2.1.1 for some explanations about the link between this definition and the Lagrangian used in optimization)

$$\mathcal{L}(\Omega, u, v) = \mathcal{J}(\Omega) + a_\Omega(u, v) - b_\Omega(v). \tag{II.2.3}$$

Note here that $\mathcal{L}$ is well-defined for all $u, v \in H$ and that no dependency between u, v and Ω is required.

2. Expression of the shape derivative using the Lagrangian. Taking $u = u_\Omega$ the solution of the equation $a_\Omega(u, v) = b_\Omega(v)$ we have for all $v \in H$:

$$\mathcal{L}(\Omega, u_\Omega, v) = \mathcal{J}(\Omega). \tag{II.2.4}$$

And since this equality is verified for all shapes it also mean that for all $\boldsymbol{\theta}$:

$$\mathcal{L}((\mathrm{Id} + \boldsymbol{\theta})(\Omega), u_{(\mathrm{Id}+\boldsymbol{\theta})(\Omega)}, v) = \mathcal{J}((\mathrm{Id} + \boldsymbol{\theta})(\Omega)).$$

Now supposing that the mapping $\Omega \mapsto \mathcal{L}(\Omega, u_\Omega, v) \in \mathbb{R}$ is shape differentiable for every $v \in H$, we can use the chain rule and find that the shape derivative of $\mathcal{J}(\Omega)$ equals to

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \partial_\Omega \mathcal{L}(\Omega, u_\Omega, v)(\boldsymbol{\theta}) + \partial_u \mathcal{L}(\Omega, u_\Omega, v) \circ u'_\Omega(\boldsymbol{\theta}). \tag{II.2.5}$$

Remark here that finding $\partial_\Omega \mathcal{L}(\Omega, u, v)(\boldsymbol{\theta})$ is easier than $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ since the integrands in $\mathcal{L}$ does not depend on Ω so we can use the results of Th. II.1.1.2 instead of Th. II.1.2.1.

3. Introduction of the adjoint state. Since a_Ω is bilinear, Eq. (II.2.5) is equivalent to

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \partial_\Omega \mathcal{L}(\Omega, u_\Omega, v)(\boldsymbol{\theta}) + \int_{\mathcal{D}} j'(u_\Omega)(u'_\Omega(\boldsymbol{\theta}))\, \mathrm{d}\mathbf{x} + a_\Omega(u'_\Omega(\boldsymbol{\theta}), v_\Omega). \tag{II.2.6}$$

We will now see that it is possible to cancel the second and third term in Eq. (II.2.6) which has the advantage to greatly simplify the shape derivative's formula by not involving the Eulerian derivative. To do this it is sufficient to remark that taking $v = v_\Omega$ solution for all $\tilde{u} \in H$ of

$$\int_{\mathcal{D}} j'(u_\Omega)(\tilde{u})\, \mathrm{d}\mathbf{x} + a_\Omega(\tilde{u}, v) = 0. \tag{II.2.7}$$

From Eq. (II.2.6) the shape derivative of $\mathcal{J}$ is then found equal to

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \partial_\Omega \mathcal{L}(\Omega, u_\Omega, v_\Omega)(\boldsymbol{\theta}). \tag{II.2.8}$$

In summary to find $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ we will have to

1. Find u_Ω solution of $a_\Omega(u_\Omega, v) = b_\Omega(v)$ for all $v \in H$.
2. Find v_Ω solution of $a_\Omega(\tilde{u}, v_\Omega) = -\int_{\mathcal{D}} j'(u_\Omega)(\tilde{u})\, \mathrm{d}\mathbf{x}$ for all $\tilde{u} \in H$.
3. Compute the shape derivative as $\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \partial_\Omega \mathcal{L}(\Omega, u_\Omega, v_\Omega)(\boldsymbol{\theta})$ whose exact formula is obtained through Th. II.1.1.2. From this theorem it is also clear that the final shape derivative should be of the form

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} V_\Omega(s)\, \mathrm{d}s \tag{II.2.9}$$

where the scalar field $V_\Omega(s)$ may involve the values of both u_Ω and v_Ω (note that this expression is consistent with the structure theorem Th. II.1.1.3).

Remark II.2.1.1: The definition of the Lagrangian (Eq. (II.2.3)) comes from an analogy with the Lagrangian used in mathematical optimization. Indeed, we can first

see that maximizing Eq. (II.2.1) is equivalent to solving the following mathematical program

$$\begin{cases} \max\limits_{\substack{\Omega\subset\mathcal{D}\\ u\in H}} & \int_{\mathcal{D}} j(u)\,\mathrm{d}\mathbf{x} \\ \textbf{s.t.} & K_\Omega(u)(\mathbf{x}) = 0 \quad \text{for all } \mathbf{x}\in\mathcal{D} \end{cases}.$$

Introducing a Lagrangian multiplier $v(\mathbf{x})$ for each constraint on u leads us to the following classical definition for the Lagrangian

$$\mathcal{L}(\Omega, u, v) = \mathcal{J}(\Omega) + \int_{\mathcal{D}} v(\mathbf{x}) K_\Omega(u)\,\mathrm{d}\mathbf{x}$$

which, after integration by part, is equal to Eq. (II.2.3). See also [Del11, Chapter 10, Section 5.2] about this link between the Lagrangian in optimization and the one used in shape optimization.

II.2.1.b Limitations of Céa's method

Céa's method may be adapted to most physical problem and gives the correct shape derivative in many cases but its application still requires to carefully keep in mind the underlying assumptions that have been made to ensure that it is correctly adapted. Mainly two hypotheses are of great importance to find the correct shape derivative.

First, in Eq. (II.2.3), the definition of the Lagrangian supposed that u, v and Ω are independent. This may not be the case if the Hilbert space H in which u_Ω and v_Ω are defined depends on Ω. This occurs for instance when a Dirichlet boundary condition is applied on (a part of) the border of Ω which is also subject to optimization. It is still possible to adapt Céa's method by adding an additional Lagrange multiplier and we refer the reader to [All07, Section 6.4.3] for information about this particular case.

Secondly, Céa's method relies heavily on the formula (II.2.5) obtained by assuming the existence of the Eulerian derivative of u_Ω which, as we have seen in Section II.1.2.a, may not exist contrary to the Lagrangian derivative. We will see in Section III.2.1 that this exact problem occurs for objective functionals used in nanophotonics. This non differentiability may be readily seen from Eq. (II.1.14) since the Eulerian derivative is linked to the Lagrangian one by involving the gradient of the solution of the PDE (see Remark II.2.1.2) which may not be defined on the interface $\partial\Omega$ (in Section I.1.3.b we saw that only the tangential derivatives of the electric field are well-defined on an interface). Just like the first problem mentioned above, we can still adapt Céa's method to find the right shape derivative (see [Pan05] for more details).

Remark II.2.1.2: To be perfectly precise about the Eulerian derivative, it may be defined but will lack some regularity; u'_Ω may not be an element of H. In such a situation, even if v_Ω is defined by canceling the adjoint equation (II.2.7) for all $\tilde{u}\in H$, Eq. (II.2.8) is not verified since $u'_\Omega(\boldsymbol{\theta}) \notin H$. Even worse, it may not even be possible to evaluate a_Ω using an element which is not in H (for instance if the Eulerian derivative is only an element of L^2 and $a_\Omega(u, v)$ involve the gradient of u).

II.2.1.c Numerical considerations

In this subsection we discuss two numerical features about computation of the adjoint state. First, as we have seen u_Ω is solution of Eq. (II.2.2) and v_Ω the adjoint state is the solution of Eq. (II.2.7). Once discretized using the finite element method (see Section I.5.1) these equations drop down to:

$$A_\Omega \mathbf{u}_\Omega = \mathbf{l}_{1,\Omega} \quad \text{and} \quad A_\Omega^\top \mathbf{v}_\Omega = \mathbf{l}_{2,\Omega}$$

where A_Ω is the same stiffness matrix involved for both equations, $\mathbf{l}_{1,\Omega}, \mathbf{l}_{2,\Omega}$ are two vectors and $\mathbf{u}_\Omega$ (resp. $\mathbf{v}_\Omega$) is collecting the values of u_Ω(resp. v_Ω) at the degree of freedom of the finite element discretization. If during the resolution of $\mathbf{u}_\Omega$ the LU decomposition of A_Ω is found (this is most of the time the case when using a sparse direct solver) then $\mathbf{v}_\Omega$ is easily computed since it involves the same matrix.

It is worth noting that in Section II.2.1.a the Lagrangian involved u_Ω and v_Ω and not their discretized counterpart. The same analysis may be carried out using $\mathbf{u}_\Omega$ and $\mathbf{v}_\Omega$ in the Lagrangian and the question that can legitimately be asked is whether the shape derivative obtained by considering this new Lagrangian is the same as the previously obtained one once discretized (in other words, does considering a continuous model and then discretizing it give the same result as directly considering the discrete model?). For a lot of classical shape optimization problem these two approaches are indeed equivalent but unfortunately this is not the case of the problem studied in Section III.2.1; see also [All14b, Section 2.2].

II.2.2 Structure of the rigorous proof

Even when it gives the right result, Céa's method remains purely formal as we have seen at the end of Section II.2.1.a. This section is now dedicated to the explanation of the full proof required to obtain the shape derivative of Eq. (II.2.1), an application to nanophotonic will be studied in Section III.2.1.

This demonstration is divided into four main steps.

- First we look at the variational form of the transported field $u_{(\mathrm{Id}+\boldsymbol{\theta})(\Omega)} \circ (\mathrm{Id} + \boldsymbol{\theta})$.
- Later it is proved that the Lagrangian derivative of u_Ω is well defined by using the implicit function theorem.
- We differentiate both the objective function and the variational formulation of $u_{(\mathrm{Id}+\boldsymbol{\theta})(\Omega)} \circ (\mathrm{Id} + \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ to find that $\mathring{u_\Omega}(\boldsymbol{\theta})$ is obtained by solving a PDE where $\boldsymbol{\theta}$ acts as a parameter. Introducing an adjoint state v_Ω it is then possible to remove the dependence on the Lagrangian derivative and express the shape derivative using a volumic integral and v_Ω.
- The last step is then dedicated to the modification of the previously found expression into a surfacic one using integration by parts.

Unlike Section II.2.1.a we consider here a particular structure of the variational formulation (we could have once again remained with a general problem but we think that it would be too vague to adapt this demonstration to other physical problems). For this

introduction we consider that u_Ω is the solution to a fairly simple second order elliptic problem on $\mathcal{D}$ with Neumann boundary condition, precisely we take

$$a_\Omega(u_\Omega, v) = \int_{\mathcal{D}} \nabla u_\Omega \cdot \nabla v + c_\Omega u_\Omega v \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} f v \,\mathrm{d}\mathbf{x} = b_\Omega(v) \tag{II.2.10}$$

where $c_\Omega(\mathbf{x})$ equals $c_1 > 0$ inside Ω and $c_2 > 0$ outside Ω with $f \in C^1(\mathcal{D}, \mathbb{R})$. Here the Hilbert space H is $H^1(\mathcal{D})$. In details the proof is then as follows.

1. Mapping of the objective function and the variational formulation in the reference domain. Let $\boldsymbol{\theta} \in W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ a vector field such that its norm $\|\boldsymbol{\theta}\|_{W^{1,\infty}(\mathbb{R}^d,\mathbb{R}^d)}$ is strictly less than 1 (see Th. II.1.1.1). We start by transporting the objective functional for the shape $(\mathrm{Id} + \boldsymbol{\theta})(\Omega) = \Omega_{\boldsymbol{\theta}}$ on Ω:

$$\mathcal{J}(\Omega_{\boldsymbol{\theta}}) = \int_{\mathcal{D}} j(u_{\Omega_{\boldsymbol{\theta}}}) \,\mathrm{d}\mathbf{x} \quad \text{where} \quad \int_{\mathcal{D}} \nabla u_{\Omega_{\boldsymbol{\theta}}} \cdot \nabla v + c_{\Omega_{\boldsymbol{\theta}}} u_{\Omega_{\boldsymbol{\theta}}} v \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} f v \,\mathrm{d}\mathbf{x}$$

for all $v \in H$. Using a change of variables and defining the transported fields $\bar{u}_{\boldsymbol{\theta}} = u_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta})$ and $\bar{v}_{\boldsymbol{\theta}} = v \circ (\mathrm{Id} + \boldsymbol{\theta})$ we therefore have

$$\mathcal{J}(\Omega_{\boldsymbol{\theta}}) = \int_{\mathcal{D}} j(\bar{u}_{\boldsymbol{\theta}}) B(\boldsymbol{\theta}) \,\mathrm{d}\mathbf{x}, \tag{II.2.11}$$

and for a given $v \in H$

$$\int_{\mathcal{D}} A(\boldsymbol{\theta}) \nabla \bar{u}_{\boldsymbol{\theta}} \cdot \nabla \bar{v}_{\boldsymbol{\theta}} + c_\Omega B(\boldsymbol{\theta}) \bar{u}_{\boldsymbol{\theta}} \bar{v}_{\boldsymbol{\theta}} \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} B(\boldsymbol{\theta}) f \circ (\mathrm{Id} + \boldsymbol{\theta}) \bar{v}_{\boldsymbol{\theta}} \,\mathrm{d}\mathbf{x}$$

since by definition $c_\Omega = c_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta})$ with

$$A(\boldsymbol{\theta}) = |\det(\mathrm{Id} + \nabla\boldsymbol{\theta})|(\mathrm{Id} + \nabla\boldsymbol{\theta})^{-1}(\mathrm{Id} + \nabla\boldsymbol{\theta}^\top)^{-1} \quad \text{and} \quad B(\boldsymbol{\theta}) = |\det(\mathrm{Id} + \nabla\boldsymbol{\theta})|.$$

In the previous calculation we have used the fact that

$$\nabla(u_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta})) = (\mathrm{Id} + \nabla\boldsymbol{\theta}^\top)(\nabla u_{\Omega_{\boldsymbol{\theta}}}) \circ (\mathrm{Id} + \boldsymbol{\theta}).$$

Now from Th. II.1.1.1 we see that for $\boldsymbol{\theta}$ small enough we have $\{v \circ (\mathrm{Id} + \boldsymbol{\theta}), v \in H\} = H$ which means that $\bar{u}_{\boldsymbol{\theta}}$ is also solution for all $v \in H$ to

$$\int_{\mathcal{D}} A(\boldsymbol{\theta}) \nabla \bar{u}_{\boldsymbol{\theta}} \cdot \nabla v + c_\Omega B(\boldsymbol{\theta}) \bar{u}_{\boldsymbol{\theta}} v \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} B(\boldsymbol{\theta}) f \circ (\mathrm{Id} + \boldsymbol{\theta}) v \,\mathrm{d}\mathbf{x}. \tag{II.2.12}$$

Note that this first step could be achieved for any PDE and objective function since it amounts to express the objective function and variational formulation in terms of $\bar{u}_{\boldsymbol{\theta}}$.

2. Proof of the Lagrangian differentiability of u_Ω using the implicit function theorem. Now we consider the following functional from $W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d) \times H$ into H^* (the dual space of H)

$$F(\boldsymbol{\theta}, u) = \left(v \mapsto \int_{\mathcal{D}} A(\boldsymbol{\theta}) \nabla u \cdot \nabla v + c_\Omega B(\boldsymbol{\theta}) u v \,\mathrm{d}\mathbf{x} - \int_{\mathcal{D}} B(\boldsymbol{\theta}) f \circ (\mathrm{Id} + \boldsymbol{\theta}) v \,\mathrm{d}\mathbf{x} \right).$$

Using for instance the Lax-Milgram theorem we can prove the existence and uniqueness of the solution to

$$a_\Omega(u, \cdot) = \ell \tag{II.2.13}$$

for all Ω and $\ell \in H^*$. The function u_Ω is therefore the unique u such that $F(0,u) = 0$, that is $F(0,u_\Omega)(v) = 0$ for all $v \in H$.

Furthermore, F is an affine map with respect to u so $\partial_u F(0,u)(\tilde{u}) = a_\Omega(\tilde{u},\cdot)$. Since $a_\Omega(\tilde{u},\cdot) = \ell$ has a unique solution for all either fixed $\tilde{u} \in H$ or $\ell \in H^*$ then $\tilde{u} \to \partial_u F(0,u_\Omega)(\tilde{u})$ is an isomorphism from H into H^*.

Besides, $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ are polynomials in $\boldsymbol{\theta}$ which implies that F is a C^1 application with respect to $\boldsymbol{\theta}$ but also according to u since F is affine for a fixed value of $\boldsymbol{\theta}$.

In summary:

- The three sets $W^{1,\infty}(\mathbb{R}^d,\mathbb{R}^d), H$ and H^* are Banach spaces,
- $F : W^{1,\infty}(\mathbb{R}^d,\mathbb{R}^d) \times H \to H^*$ is a C^1 mapping,
- $F(0,u_\Omega) = 0$,
- $\partial_u F(0,u_\Omega)$ is an isomorphism from H into H^*.

Now owing to the implicit function theorem [Lan12, Chapter 1, Theorem 5.9] there exist two neighborhoods $U \subset W^{1,\infty}(\mathbb{R}^d,\mathbb{R}^d), V \subset H$ of $0 \in U$ and $u_\Omega \in V$ and a differentiable function $g : U \to V$ such that $F(\boldsymbol{\theta}, g(\boldsymbol{\theta})) = 0$ for all $\boldsymbol{\theta} \in U$.

Once again by uniqueness of a solution to the PDE we have $g(\boldsymbol{\theta}) = \bar{u}_{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta} \in U$ which allows to conclude that $\boldsymbol{\theta} \mapsto \bar{u}_{\boldsymbol{\theta}}$ is Fréchet differentiable or, equivalently, that u_Ω has a Lagrangian derivative $\mathring{u}_\Omega(\boldsymbol{\theta}) \in H$.

For more general PDEs the main difficulty of this step would be to prove the uniqueness of a solution to the equation $a_\Omega(u,\cdot) = \ell$ for every linear map ℓ, which is, in general, a fundamental assumption for the study of physical problems.

3. Volumetric shape derivative using an adjoint state. We can now differentiate the objective Eq. (II.2.11) to obtain

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\mathcal{D}} j'(u_\Omega)\mathring{u}_\Omega(\boldsymbol{\theta})\,\mathrm{d}\mathbf{x}, \tag{II.2.14}$$

as well as the variational formulation Eq. (II.2.12) verified by $\bar{u}_{\boldsymbol{\theta}}$

$$\begin{aligned}&\int_{\mathcal{D}} \nabla\mathring{u}_\Omega(\boldsymbol{\theta})\cdot\nabla v + c_\Omega\mathring{u}_\Omega(\boldsymbol{\theta})v\,\mathrm{d}\mathbf{x}\\ &= -\int_{\mathcal{D}} A'(0)(\boldsymbol{\theta})\nabla u_\Omega\cdot\nabla v + c_\Omega B'(0)(\boldsymbol{\theta})u_\Omega v\,\mathrm{d}\mathbf{x} + \int_{\mathcal{D}} B'(0)(\boldsymbol{\theta})fv + \boldsymbol{\theta}\cdot\nabla f v\,\mathrm{d}\mathbf{x},\end{aligned} \tag{II.2.15}$$

where simplifications were made using $A(0) = \mathrm{Id}, B(0) = 1$ with the derivatives of A and B given by

$$A'(0)(\boldsymbol{\theta}) = (\nabla\cdot\boldsymbol{\theta})\mathrm{Id} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top \quad \text{and} \quad B'(0)(\boldsymbol{\theta}) = \nabla\cdot\boldsymbol{\theta}.$$

Our goal is now to express Eq. (II.2.14) without resorting to $\mathring{u}_\Omega(\boldsymbol{\theta})$ since for the moment if we want to compute the value of $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ we have to solve one additional PDE (Eq. (II.2.15)) for each $\boldsymbol{\theta}$. Formula Eq. (II.2.14) is impractical from a numerical point of

view; in particular it does not lend itself to finding a descent direction $\boldsymbol{\theta}$ for $\mathcal{J}(\Omega)$ (i.e. $\boldsymbol{\theta}$ such that $\mathcal{J}'(\Omega)(\boldsymbol{\theta}) < 0$).

To understand how we will solve this problem, let us have a look at the variational formulations verified by both u_Ω and $\mathring{u}_\Omega(\boldsymbol{\theta})$. Taking $v = u_\Omega$ in Eq. (II.2.15) and $v = \mathring{u}_\Omega(\boldsymbol{\theta})$ in Eq. (II.2.10) we find that

$$\begin{aligned}&\int_{\mathcal{D}} \nabla \mathring{u}_\Omega(\boldsymbol{\theta}) \cdot \nabla u_\Omega + c_\Omega \mathring{u}_\Omega(\boldsymbol{\theta}) u_\Omega \,\mathrm{d}\mathbf{x} \\ &= -\int_{\mathcal{D}} A'(0)(\boldsymbol{\theta}) \nabla u_\Omega \cdot \nabla u_\Omega + c_\Omega B'(0)(\boldsymbol{\theta}) u_\Omega^2 \,\mathrm{d}\mathbf{x} + \int_{\mathcal{D}} B'(0)(\boldsymbol{\theta}) f u_\Omega + \boldsymbol{\theta} \cdot \nabla f u_\Omega \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{x},\end{aligned}$$

$$\int_{\mathcal{D}} \nabla u_\Omega \cdot \nabla \mathring{u}_\Omega(\boldsymbol{\theta}) + c_\Omega u_\Omega \mathring{u}_\Omega(\boldsymbol{\theta}) \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} f \mathring{u}_\Omega(\boldsymbol{\theta}) \,\mathrm{d}\mathbf{x},$$

and therefore

$$\int_{\mathcal{D}} f \mathring{u}_\Omega(\boldsymbol{\theta}) \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} B'(0)(\boldsymbol{\theta}) f u_\Omega + \boldsymbol{\theta} \cdot \nabla f u_\Omega - A'(0)(\boldsymbol{\theta}) \nabla u_\Omega \cdot \nabla u_\Omega - c_\Omega B'(0)(\boldsymbol{\theta}) u_\Omega^2 \,\mathrm{d}\mathbf{x}. \tag{II.2.16}$$

If $j'(u_\Omega)$ in Eq. (II.2.14) were both equal to f we could use the right-hand side of Eq. (II.2.16) to compute $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ without solving any additional PDE since it would not use $\mathring{u}_\Omega(\boldsymbol{\theta})$. Unfortunately this is not the case for a general function $\mathcal{J}(\Omega)$. The idea is then to introduce a new state, the adjoint, solution of Eq. (II.2.10) with this time a source term $j'(u_\Omega)$ instead of f. Indeed if we let v_Ω the solution for all $v \in H$ of

$$\int_{\mathcal{D}} \nabla v_\Omega \cdot \nabla v + c_\Omega v_\Omega v \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} j'(u_\Omega) v \,\mathrm{d}\mathbf{x} \tag{II.2.17}$$

then taking $v = L_u(\boldsymbol{\theta})$ in Eq. (II.2.17) and $v = v_\Omega$ in Eq. (II.2.15) we find that

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\mathcal{D}} B'(0)(\boldsymbol{\theta}) f u_\Omega + \boldsymbol{\theta} \cdot \nabla f u_\Omega - A'(0)(\boldsymbol{\theta}) \nabla u_\Omega \cdot \nabla v_\Omega - c_\Omega B'(0)(\boldsymbol{\theta}) u_\Omega v_\Omega \mathrm{d}\mathbf{x}, \tag{II.2.18}$$

which is known as the volumetric expression of the shape derivative.

4. Surfacic form of the shape derivative. To simplify the previous expression and obtain an integral on $\partial\Omega$ involving only the normal component of $\boldsymbol{\theta}$ (which should be possible according to the so-called structure theorem) we rely on the following integration by parts formulas valid for smooth enough $\boldsymbol{\theta}$, scalar fields u, v and vectorial ones $\mathbf{u}, \mathbf{v}$ (the strategy used here follow the development proposed in [Fep18, Section 3.4.3]):

$$\int_{\mathcal{D}} (\nabla \cdot \boldsymbol{\theta})(uv) \,\mathrm{d}\mathbf{x} = \int_{\partial\mathcal{D}} (\boldsymbol{\theta} \cdot \mathbf{n})(uv) \,\mathrm{d}s - \int_{\mathcal{D}} \nabla(uv) \cdot \boldsymbol{\theta} \,\mathrm{d}\mathbf{x} \tag{II.2.19}$$

$$\int_{\mathcal{D}} (\nabla \boldsymbol{\theta} \mathbf{u}) \cdot \mathbf{v} \,\mathrm{d}\mathbf{x} = \int_{\partial\mathcal{D}} (\boldsymbol{\theta} \cdot \mathbf{v})(\mathbf{u} \cdot \mathbf{n}) \,\mathrm{d}s - \int_{\mathcal{D}} ((\nabla \cdot \mathbf{u})\mathbf{v} + (\nabla \mathbf{v})\mathbf{u}) \cdot \boldsymbol{\theta} \,\mathrm{d}\mathbf{x}. \tag{II.2.20}$$

To apply these formulas note that the last term in the integrand of Eq. (II.2.18) is not smooth, more precisely since u_Ω, v_Ω are smooth (a classical result of elliptic PDE may be used to show that they are at least of class $H^k(\mathcal{D}, \mathbb{R})$ with $k > 1$, see for instance [Bre10, Section IX.6]) but c_Ω is not. We then rewrite the last integrand term as

$$\int_{\mathcal{D}} c_\Omega B'(0)(\boldsymbol{\theta}) u_\Omega v_\Omega \,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}\setminus\Omega} c_2 B'(0)(\boldsymbol{\theta}) u_\Omega v_\Omega \,\mathrm{d}\mathbf{x} + \int_{\Omega} c_1 B'(0)(\boldsymbol{\theta}) u_\Omega v_\Omega \,\mathrm{d}\mathbf{x}, \tag{II.2.21}$$

for which we can apply both Eqs. (II.2.19) and (II.2.20) to both integrals. Second remark, since values of $\boldsymbol{\theta}$ are only important inside $\bar{\Omega}$ and the shape Ω is an open subset of $\mathcal{D}$

we can choose any value for $\boldsymbol{\theta}$ on $\partial\mathcal{D}$ such as $\boldsymbol{\theta} = 0$. From these remarks we can now use Eqs. (II.2.19) and (II.2.20) on Eq. (II.2.18) to find that the first three integrands are of the form $\int_{\mathcal{D}} \boldsymbol{\theta} \cdot r_\Omega \, \mathrm{d}\mathbf{x}$ and that finally

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} (\boldsymbol{\theta} \cdot \mathbf{n})(c_1 - c_2) u_\Omega v_\Omega \, \mathrm{d}s + \int_{\mathcal{D}} \boldsymbol{\theta} \cdot r_\Omega \, \mathrm{d}\mathbf{x} \tag{II.2.22}$$

where r_Ω is some scalar field depending on both u_Ω and v_Ω. From Hadamard's structure theorem II.1.1.3 however, we know that $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ should vanish for compactly supported $\boldsymbol{\theta}$ in Ω or $\mathcal{D}\backslash\bar{\Omega}$ which is only possible if the last integral in Eq. (II.2.22) vanish. This conclude the four steps proof with the shape derivative given by Eq. (II.2.22) with $r_\Omega = 0$. □

Contrary to Céa's method we can see that this proof does not involve the Eulerian derivative of u_Ω and therefore only performs mathematically valid operations. The adaptation of this demonstration to nanophotonics is explained in Section III.2.1.

II.3 Numerical representation of shapes using level-set functions

This section is a short introduction to the level-set method used to numerically represent shapes and their deformations. Sections II.3.1 to II.3.3 gives theoretical results on level-set functions whereas Section II.3.4 deals with its numerical discretization. Most of the materials presented in this section can be found in the book of Sethian [Set99]. The application of level-set functions to shape optimization comes from the seminal paper [All04].

II.3.1 Introduction

Rather than representing a shape $\Omega \subset \mathbb{R}^d$ by its characteristic function $\mathbf{1}_\Omega : \mathbb{R}^d \to \{0, 1\}$, this method implicitly represents Ω using a smooth function $\phi : \mathbb{R}^d \to \mathbb{R}$ (a level-set function) such that $\partial\Omega$ corresponds to the zero level-set of ϕ (see Fig. II.3.1) and Ω to the negative subdomain of ϕ. More precisely ϕ is chosen such that

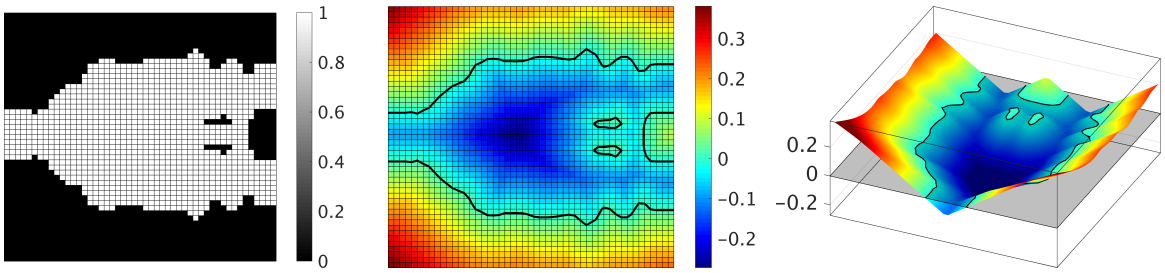
$$\Omega = \{\mathbf{x}, \phi(\mathbf{x}) < 0\}, \quad \partial\Omega = \{\mathbf{x}, \phi(\mathbf{x}) = 0\}, \quad (\mathbb{R}^d\backslash\bar{\Omega}) = \{\mathbf{x}, \phi(\mathbf{x}) > 0\}.$$

As we will see, this new degree of freedom (ϕ gives at each point a value in $\mathbb{R}$ instead of a binary result with $\mathbf{1}_\Omega$) is very convenient since a smooth representation means that small modifications of the shape Ω will corresponds to regular changes of ϕ allowing a finer control of the border of the shape and its properties.

Remark II.3.1.1: In the previous section we studied the sensitivity of an objective function when the shape Ω is deformed according to a vector field. In this section we have established that a level-set function allows to fully represent a shape. We could then study the sensitivity of the objective function when one of the level-set function ϕ associated with the shape Ω is slightly modified.

If we denote by Ω_δ the shape represented by a level-set function $\phi+\delta$ where $\delta : \mathbb{R}^d \to \mathbb{R}$, then we are looking for a first order Taylor expansion as

$$\mathcal{J}(\Omega_\delta) = \mathcal{J}(\Omega) + \mathcal{J}'(\Omega)(\delta) + o(\delta), \tag{II.3.1}$$



(a) *Characteristic function* $\mathbf{1}_\Omega$ (b) *A level-set representation* (c) *3D illustration of* (b)

Figure II.3.1: A shape Ω and some related discrete representations.

which allows to find the best modification δ to apply on ϕ in order to improve the objective function.

This method has been considered in several scientific articles but one criticism is often made about it; given that this method does not make any distinction between the shape and its numerical representation, level-set regularization processes such as redistanciation (see Section II.3.3.c) cannot be mathematically justified (although it is numerically useful). However, in the case of a derivation and modification of the shape with respect to a vector field, the level-set being there only to represent the shape, it is possible to modify it as we like as long as it always represents the same shape.

II.3.2 Geometric properties

The first interesting property of level-set functions is that they allow to conveniently calculate geometric quantities such as the normal vector or local curvature. Indeed, a result of differential geometry allows us to find that if a shape $\Omega \subset \mathcal{D}$ is represented by a smooth level-set function ϕ then

$$\mathbf{n} = \frac{\nabla\phi}{|\nabla\phi|} \quad \text{and} \quad \kappa = \nabla \cdot \mathbf{n} \tag{II.3.2}$$

where $\mathbf{n}$ is the normal vector to $\partial\Omega$ pointing outward Ω and κ the mean curvature.

Remark II.3.2.1: The geometrical definitions in Eq. (II.3.2) are only valid for a point $\mathbf{x}$ on the border of the shape Ω. Later on, however, we will sometimes refer to the value of the normal vector outside of $\partial\Omega$. For this purpose we define the extended normal vector on $\mathcal{D}$ through the level-set representation of Ω as $\mathbf{n}(\mathbf{x}) = \nabla\phi(\mathbf{x})/|\nabla\phi(\mathbf{x})|$ for all $\mathbf{x} \in \mathcal{D}$ whenever $\nabla\phi(\mathbf{x})$ is not equal to 0.

Its also worth noting that the perimeter and volume of Ω may be computed as

$$\text{Per}(\Omega) = \int_{\{\phi(\mathbf{x})=0\}} \mathrm{d}s = \int_{\mathbb{R}^d} \delta_0 \circ \phi(\mathbf{x}) |\nabla\phi(\mathbf{x})| \, \mathrm{d}\mathbf{x} \tag{II.3.3}$$

using the dirac function δ_0 and

$$\text{Volume}(\Omega) = \int_{\{\phi(\mathbf{x})<0\}} \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^d} \mathbf{1}_{\{\mathbf{x},\phi(\mathbf{x})<0\}} \, \mathrm{d}\mathbf{x}. \tag{II.3.4}$$

II.3.3 Movement along the normal vectors

II.3.3.a Hamilton-Jacobi equation

Let us get back to shape optimization. So far we have seen that in order to optimize a figure of merit $\mathcal{J}$ we need to move the borders of the shape along its normal vector; for a given initial shape Ω_0 we need to compute $\Omega_1 = (\mathrm{Id} + \boldsymbol{\theta})(\Omega_0)$ with a given vector field $\boldsymbol{\theta}$ (supplied by the shape derivative of $\mathcal{J}(\Omega)$) of small amplitude. Remember that a very crude sketch of the optimization algorithm was given in Section II.1.1.e.

We will now see that Line 8 of Algorithm II.1.1 where we performed the operation $\Omega := (\mathrm{Id} + \tau\theta\mathbf{n})(\Omega)$ is easily numerically performed using a level-set representation of the shapes. More precisely if Ω_0 is given by ϕ_0 then ϕ_1, a level-set representation of Ω_1, is found by solving an Hamilton-Jacobi equation.

To understand this, let us consider the general framework of the movement of a shape $\Omega_0 \subset \mathcal{D}$ through a time-dependent vector field. For all $t \in [0,1]$ we define $\Omega_t = (\mathrm{Id} + \boldsymbol{\theta}(\mathbf{x},t))(\Omega_0)$ where $\boldsymbol{\theta}(\mathbf{x},t)$ is a given vector field. The movement of any "particle" starting at a point $\mathbf{x}_0 \in \Omega_0$ is denoted by $\chi(\mathbf{x}_0,t)$ and defined as the solution

$$\partial_t \chi(\mathbf{x}_0,t) = \boldsymbol{\theta}(\chi(\mathbf{x}_0,t),t) \tag{II.3.5}$$

with the initial condition $\chi(\mathbf{x}_0,0) = \mathbf{x}_0$. Knowing $\phi(\cdot,0) = \phi_0$ a level-set representation of Ω_0, we are searching, for each $t \in]0,1]$, a function $\phi(\cdot,t)$ representing $(\mathrm{Id} + \boldsymbol{\theta}(\cdot,t))(\Omega_0)$. This implies in particular that a particle on $\partial\Omega_0$ has to stay on $\partial\Omega_t$ after its movement. In other word if $\mathbf{x}_0 \in \partial\Omega_0$ then for all t:

$$\phi(\chi(\mathbf{x}_0,t),t) = 0. \tag{II.3.6}$$

Differentiating with respect to t Eq. (II.3.6) using the chain rule and the definition of χ (Eq. (II.3.5)), we find that

$$\partial_t \phi(\mathbf{x}_0,t) + \boldsymbol{\theta}(\mathbf{x}_0,t) \cdot \nabla_{\mathbf{x}} \phi(\mathbf{x}_0,t) = 0. \tag{II.3.7}$$

If we add the more general constraint that for all $\mathbf{x} \in \mathcal{D}$ such that $\phi(\mathbf{x},t) = r \in \mathbb{R}$ the particle $\chi(\mathbf{x},t)$ associated with $\mathbf{x}$ must stay on the same level set value r, that is $\phi(\chi(\mathbf{x},t),t) = r$, then we find that Eq. (II.3.7) must be verified for all $\mathbf{x} \in \mathcal{D}$ and not only for $r = 0$ and $\mathbf{x}_0 \in \partial\Omega$. In the particular case of normal vector fields, $\boldsymbol{\theta}(\mathbf{x},t) = \theta(\mathbf{x},t)\mathbf{n}(\mathbf{x},t)$ and using Eq. (II.3.2) the Eq. (II.3.7) rewrites as the following **Hamilton-Jacobi equation**

$$\partial_t \phi(\mathbf{x},t) + \theta(\mathbf{x},t)|\nabla_{\mathbf{x}} \phi(\mathbf{x},t)| = 0. \tag{II.3.8}$$

Together with the initial condition given by $\phi(\mathbf{x},0) = \phi_0(\mathbf{x})$ this equation characterizes the motion of the domain Ω_t. We shall see in the next section that efficient finite difference schemes exist to solve the Hamilton-Jaocbi equation.

Remark II.3.3.1: As an example, it is interessting to note that solving Eq. (II.3.8) using $\theta(\mathbf{x},t) = \delta$ (resp. $-\delta$) should give the shape dilated (resp. eroded) by a length δ.

II.3.3.b Distance function and Eikonal equation

Among all the possible level-set functions ϕ to represent a shape $\Omega \subset \mathcal{D}$ there is one that is particularly useful in numerical practice. This is the **signed distance function** to Ω defined as the unique function d_Ω whose sign at $\mathbf{x}$ indicates whether or not $\mathbf{x}$ is inside Ω and whose absolute values gives the distance from $\mathbf{x}$ to $\partial\Omega$,

$$d_\Omega(\mathbf{x}) = \begin{cases} -d_{\partial\Omega}(\mathbf{x}) & \text{if} \quad \mathbf{x} \in \Omega \\ 0 & \text{if} \quad \mathbf{x} \in \partial\Omega \\ d_{\partial\Omega}(\mathbf{x}) & \text{if} \quad \mathbf{x} \in \mathcal{D}\backslash\bar{\Omega} \end{cases}, \tag{II.3.9}$$

where $d_{\partial\Omega}(\mathbf{x})$ is the distance from $\mathbf{x}$ to $\partial\Omega$ (the level-set representation in Fig. II.3.1(c) is a signed distance function). An interesting property of d_Ω is that it is solution of the following **Eikonal equation**:

$$|\nabla d_\Omega| = 1 \text{ in } \mathcal{D}, \quad d_\Omega = 0 \text{ on } \partial\Omega \quad \text{and} \quad d_\Omega < 0 \text{ in } \Omega. \tag{II.3.10}$$

The resolution of Eq. (II.3.10) provides a way to find a level-set function associated to a given shape Ω and is useful to initialize the optimization algorithm with an existing shape.

In addition of being useful numerically, the signed distance function is also used multiple times in Chapters III to V to simply express geometric properties or to regularize quantities close to the border of the shape; if $f : \mathcal{D} \to \mathbb{R}$ is a discontinuous function, taking a value f_1 in Ω and f_2 elsewhere then we can smooth the transition near $\partial\Omega$ using the signed distance function and a small parameter ε as $\tilde{f} = f_2 + (f_2 - f_1)H(d_\Omega)$ where H is a smooth heaviside function (see Remark III.2.2.2 or Section V.4.1 for examples using this regularization).

Handling the signed function as the level-set representation of a shape is convenient since, in practice, it has been reported many times that solving Eq. (II.3.8) will have the unfortunate tendency to flatten areas (resp. make very steeps areas) with near-zero values hindering the proper determination of the sign of ϕ (resp. preventing a precise calculation of the gradient of ϕ). However, it should be noted that we did not experience these kind of problems during the optimization of the nanophotonic components presented in Chapter III.

II.3.3.c Redistanciation

One last interesting equation concerning level-set functions that will be used later is the following:

$$\partial_t\phi(\mathbf{x}, t) + \text{sign}(\phi_0)(|\nabla_\mathbf{x}\phi| - 1) = 0 \quad \text{and} \quad \phi(\mathbf{x}, 0) = \phi_0, \tag{II.3.11}$$

where $\text{sign}(\phi_0)$ is defined as

$$\text{sign}(\phi_0(\mathbf{x})) = \begin{cases} 1 & \text{if} \quad \phi_0(\mathbf{x}) > 0 \\ 0 & \text{if} \quad \phi_0(\mathbf{x}) = 0 \\ -1 & \text{if} \quad \phi_0(\mathbf{x}) < 0 \end{cases}. \tag{II.3.12}$$

In order to grasp the behavior of such equation, suppose that ϕ_0 is a level-set function associated to a shape Ω. First of all, notice that Eq. (II.3.11) is the same as Eq. (II.3.8) using the vector field $\text{sign}(\phi_0)\mathbf{n}$ (where $\mathbf{n}$ is the extended normal vector as explained in Remark II.3.2.1) and an additional source term $\text{sign}(\phi_0)$. A formal analysis then reveals that all edge points $\mathbf{x}$ ($\phi_0(\mathbf{x}) = 0$) stay on $\partial\Omega$. Indeed:

1. The effect of the vector field $\text{sign}(\phi_0)\mathbf{n}$ is nearly the same as removing the value $\text{sign}(\phi_0)$ to all values of $\phi_0(\mathbf{x})$.

2. Meanwhile the same quantity $\text{sign}(\phi_0)$ is added by the presence of the source term.

Additionally, if a steady state is achieved, Eq. (II.3.11) is equivalent to Eq. (II.3.10), that is $|\nabla_{\mathbf{x}}\phi| = 1$. By combining all these remarks we can understand that Eq. (II.3.11) provides, when $t \to \infty$, the signed distance to the set associated with the level-set function ϕ_0. Solving Eq. (II.3.11) periodically allows to keep a well-defined level-set representation of the set.

II.3.4 Numerical discretization

II.3.4.a Introduction

In order to store and manipulate the level-set function numerically, we need to discretize the level-set function ϕ (i.e. parametrize ϕ by projecting it into a finite-dimensional space) representing a shape $\Omega \subset \mathcal{D}_{\text{opt}} \subset \mathbb{R}^d$. To do so, several options are feasible; see [Dij13] on this topic in the context of shape optimization. Most of them rely on a discretization of the domain $\mathcal{D}$ leading to ϕ being defined as

$$\phi(\mathbf{x}) = \sum_{i=1}^{n_{\text{dof}}} \alpha_i \phi_i(\mathbf{x}) \tag{II.3.13}$$

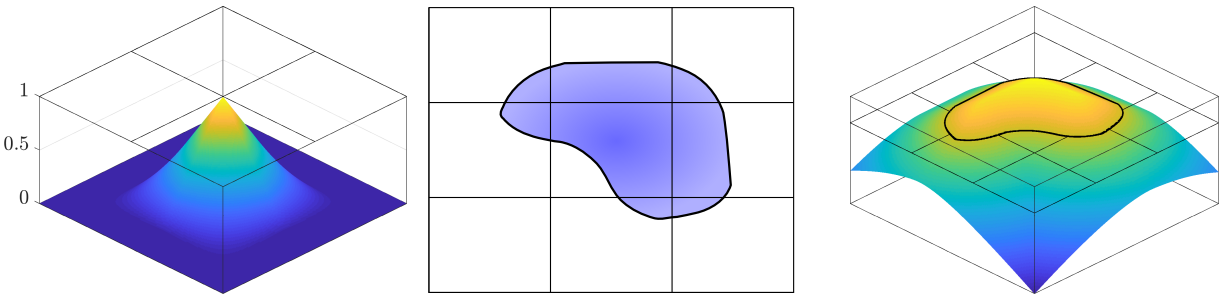
where the $(\alpha_i)_{i=1,\dots,n_{\text{dof}}}$ are scalar coefficients and $(\phi_i)_{i=1,\dots,n_{\text{dof}}}$ some basis functions. In this section, as in the rest of this thesis, we will limit ourselves to the two-dimensional case ($N = 2$) and a discretization of the domain through a Cartesian, regularly spaced grid so that Eq. (II.3.13) rewrites into

$$\phi(x, y) = \sum_{i=1}^{n_y} \sum_{j=1}^{n_x} \alpha_{i,j} \phi_{i,j}(x, y). \tag{II.3.14}$$

We also consider bilinear basis functions (see Fig. II.3.2) defined at each nodes of the grid:

$$\phi_{i,j}(x, y) = \left(1 - \frac{|x - x_{i,j}|}{\Delta x}\right)\left(1 - \frac{|y - y_{i,j}|}{\Delta y}\right) \mathbf{1}_{\{(x,y),|x-x_{i,j}|<\Delta x,|y-y_{i,j}|<\Delta y\}}. \tag{II.3.15}$$

See Fig. II.3.2(a) for a representation of this basis function.



(a) *One bilinear basis function defined in Eq. (II.3.15)*

(b) *A shape on a 4×4 Cartesian grid*

(c) *One level-set representation of the shape in* (b)

Figure II.3.2: Level-set function on a Cartesian grid.

Remark II.3.4.1: The bilinear interpolation scheme is mainly used for exportation purpose when generating the file containing the geometric information of the shape (`GDSII` format) and therefore the precise position of its edges. For most computations we are only interested in the nodal values of ϕ and the interpolation scheme is not important.

To enhance the precision of such parametrization it is possible to consider adaptative refinements of the grid using for example quad-trees or triangular meshes. Other basis functions include RBF (Radial Basis Functions) or spectral decomposition (again, see [Dij13, Section 2.2]) which both bring into play smoother level-set functions and therefore more regularity in numerical quantities such as $\mathbf{n}$ and κ.

II.3.4.b Geometric properties

For any matrix of scalar coefficients $\alpha = (\alpha_{i,j})_{i,j}$ we define the following operators approximating the first and second order partial derivatives considering either forward, centered or backward finite difference:

$$D^{+x}_{i,j}\alpha = \frac{\alpha_{i,j+1} - \alpha_{i,j}}{\Delta x}, \quad D^{-x}_{i,j}\alpha = \frac{\alpha_{i,j} - \alpha_{i,j-1}}{\Delta x}, \quad D^{=x}_{i,j}\phi = \frac{\alpha_{i,j+1} - \alpha_{i,j-1}}{2\Delta x}, \tag{II.3.16}$$

$$D^{++x}_{i,j}\alpha = \frac{\alpha_{i,j+2} - 2\alpha_{i,j+1} + \alpha_{i,j}}{(\Delta x)^2}, \quad D^{--x}_{i,j}\alpha = \frac{-\alpha_{i,j+2} + 2\alpha_{i,j+1} - 1\alpha_{i,j}}{(\Delta x)^2},$$

$$D^{+-x}_{i,j}\alpha = D^{-+x}_{i,j}\alpha = \frac{\alpha_{i,j+1} - 2\alpha_{i,j} + \alpha_{i,j-1}}{(\Delta x)^2},$$

and the same holds as far as the y-direction is concerned. With these finite differences operators the normal vector and curvature (defined in Eq. (II.3.2)) may be approximated by (see [Set99, Section II.6.7])

$$n_x = \frac{D^{=x}\alpha}{((D^{=x}\alpha)^2 + (D^{=y}\alpha)^2)^{\frac{1}{2}}}, \quad \frac{D^{=y}\alpha}{((D^{=x}\alpha)^2 + (D^{=y}\alpha)^2)^{\frac{1}{2}}},$$

$$\kappa = \frac{D^{+-x}\alpha(D^{=y}\alpha)^2 - 2D^{=x}\alpha D^{=y}\alpha(D^{=y}D^{=x}\alpha) + D^{+-y}\alpha(D^{=x}\alpha)^2}{((D^{=x}\alpha)^2 + (D^{=y}\alpha)^2)^{\frac{3}{2}}},$$

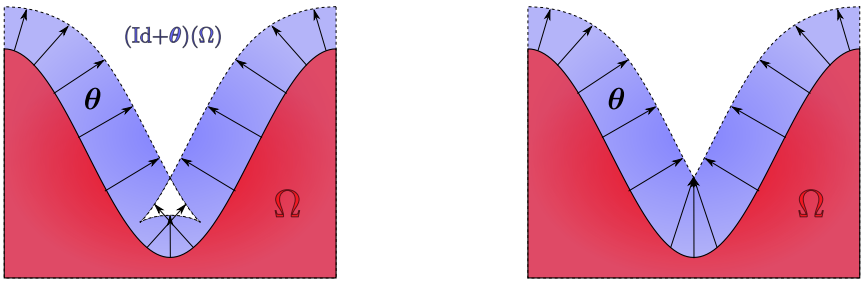
where we removed the i, j subscripts for simplicity. Note that for extremal values ($i = 1, n_y$ or $j = 1, n_y$) considering either Dirichlet or Neumann boundary conditions allows to define the previous operators and geometric values for all $i = 1, \dots, n_y$ and $j = 1, \dots, n_y$.

II.3.4.c Resolution of the PDEs using finite differences

Hamilton-Jacobi equation

The numerical resolution of either Eq. (II.3.8) or Eq. (II.3.11) first requires to choose one particular solution of these equations. Indeed, as it may be seen on Fig. II.3.3, several solutions can be considered mathematically but one in particular, the viscosity solution, seems to be the natural one. We do not give here the mathematical definition of such solution and the reader is referred to [Set99, Section II.2] for more details and to Fig. II.3.3 for an illustration.

The remainder of this section is devoted to the presentation of a second-order finite difference numerical scheme to obtain this viscosity solution. Information presented here



(a) *Moving each point of the border through normal vectors,* $(Id+V\mathbf{n})(\partial\Omega)$ *in dotted line*

(b) *The viscosity solution which border is different from the one in* (a)

Figure II.3.3: One shape Ω and two possible evolutions of its borders $\partial\Omega$ when moved along a vector field $V\mathbf{n}$ where V is a scalar field whose precise value is not important.

mostly comes from [Set99, Section II.6]. We consider an explicit and time-updwind finite different schemes of Eq. (II.3.8) of the form

$$\alpha_{i,j}^{n+1} = \alpha_{i,j}^{n} - \Delta t \mathcal{H}(i,j,\theta^n,\alpha^n) \tag{II.3.17}$$

where Δt is the time step and $\mathcal{H}(i,j,\theta^n,\alpha^n)$ is some numerical discretization of $\theta(\mathbf{x},t)|\nabla_{\mathbf{x}}\phi(\mathbf{x},t)|$ using the values $\alpha^n = (\alpha_{i,j}^n)_{i,j}$ (resp. $\theta^n = (\theta_{i,j}^n)_{i,j}$) of $\phi(\mathbf{x},n\Delta t)$ (resp. $\theta(\mathbf{x},n\Delta t)$) at each node of the Cartesian grid at the time n. From [Set99, Section II.6.4] we consider that the expression of $\mathcal{H}$ is given by the following formula in which different numerical schemes are considered depending on the sign of $\theta_{i,j}^n$

$$\mathcal{H}(i,j,\theta^n,\alpha^n) = \max(\theta_{i,j}^n,0)\mathcal{H}^+(i,j,\alpha^n) + \min(\theta_{i,j}^n,0)\mathcal{H}^-(i,j,\alpha^n). \tag{II.3.18}$$

In Eq. (II.3.18), $\mathcal{H}^{\pm}(i,j,\alpha^n)$ is chosen in order to provide a "good" discretization of $|\nabla\phi(\mathbf{x},t)|$. A second order strategy reads as:

$$\begin{aligned}\mathcal{H}^+(i,j,\alpha^n) &= \sqrt{\mathcal{H}_x^+(i,j,\alpha^n) + \mathcal{H}_y^+(i,j,\alpha^n)},\\ \mathcal{H}^-(i,j,\alpha^n) &= \sqrt{\mathcal{H}_x^-(i,j,\alpha^n) + \mathcal{H}_y^-(i,j,\alpha^n)},\end{aligned}$$

where, this time for $\nu = x, y$, $\mathcal{H}_\nu^{\pm}$ are approximating $(\partial_\nu\phi(\mathbf{x},t))^2$ by

$$\begin{aligned}\mathcal{H}_\nu^+(i,j,\alpha^n) &= \max(\mathrm{D}_+^\nu(i,j,\alpha^n),0)^2 + \min(\mathrm{D}_-^\nu(i,j,\alpha^n),0)^2,\\ \mathcal{H}_\nu^-(i,j,\alpha^n) &= \max(\mathrm{D}_-^\nu(i,j,\alpha^n),0)^2 + \min(\mathrm{D}_+^\nu(i,j,\alpha^n),0)^2,\end{aligned}$$

and $\mathrm{D}_+^\nu, \mathrm{D}_-^\iota$ are given by

$$\mathrm{D}_+^\nu(i,j,\alpha^n) = D_{i,j}^{-\nu}\alpha^n + \frac{\Delta\nu}{2} m(D_{i,j}^{--\nu}\alpha^n, D_{i,j}^{+-\nu}\alpha^n), \tag{II.3.19}$$

$$\mathrm{D}_-^\nu(i,j,\alpha^n) = D_{i,j}^{+\nu}\alpha^n + \frac{\Delta\iota}{2} m(D_{i,j}^{++\nu}\alpha^n, D_{i,j}^{+-\nu}\alpha^n). \tag{II.3.20}$$

The discrete derivatives are defined in Section II.3.4.b and m is the following function

$$m(a,b) = \begin{cases} \sqrt{a} & \text{if} \quad |a| \le |b| \quad \text{and} \quad ab \ge 0\\ b & \text{if} \quad |a| > |b| \quad \text{and} \quad ab \ge 0\\ 0 & \text{if} \quad ab < 0 \end{cases}. \tag{II.3.21}$$

Note that in Eqs. (II.3.19) and (II.3.20) D^{ι}_{+} and D^{ι}_{-} are either (depending on the value of m) first or second order approximation of the partial derivatives. The numerical implementation of Eq. (II.3.17) does not require further comments and since it only involve explicit finite difference its resolution is easily performed in parallel.

As is often the case for explicit finite difference schemes the time and spatial steps $\Delta t, \Delta x$ and Δy cannot be arbitrarily chosen in order to keep a stable numerical resolution. In the case where θ is time-independent (so that $\theta^n_{i,j} = \theta_{i,j}$ for all values of n), Δt is limited by both the spatial steps and the maximal value taken by the scalar field θ via the following relation

$$\Delta t \leq \frac{\min(\Delta x, \Delta y)}{\sup_{i,j} |\theta_{i,j}|} \tag{II.3.22}$$

which is known as a CFL condition, and we will refer to the right-hand side as the **CFL value**. From a geometrical point of view Eq. (II.3.22) simply means that at each iteration the boundary of the shape should not move by more than one grid node. Numerically we used a step Δt equal to 0.4 times the CFL value.

Redistanciation equation

Let us now turn to the numerical derivation of Eq. (II.3.11). The numerical scheme Eq. (II.3.17) is simply modified into

$$\alpha^{n+1}_{i,j} = \alpha^n_{i,j} - \Delta t \theta^n_{i,j} \mathcal{H}(i, j, \boldsymbol{\theta}^n, \boldsymbol{\alpha}^n) + \theta^n_{i,j} \tag{II.3.23}$$

with $\theta^n_{i,j} = \text{sign}(\alpha^n_{i,j})$. For stability reasons it is also advised to smooth the sign function near zero using for instance

$$\text{sign}(a) \simeq a / \sqrt{a^2 + \min(\Delta x, \Delta y)^2}\,.$$

Remark II.3.4.2: It is important to clarify that given the approximations made by the numerical resolution of Eq. (II.3.11), the level-set function $\phi(\cdot, n\Delta t)$ obtained after redistanciation of an other one $\phi(\cdot, 0)$ may not represent exactly the same shape near the borders, that is $\{\mathbf{x}, \phi(\mathbf{x}, n\Delta t)\} \neq \{\mathbf{x}, \phi(\mathbf{x}, 0)\}$. This fact is important since in the last iterations of the optimization, the successive shapes will be very similar and therefore the slight modification brought by the redistanciation process may cause too much numerical errors and prevent the optimization algorithm from properly converging.

Eikonal equation

Numerical schemes to solve the Eikonal equation $|\nabla \phi(\mathbf{x})| = 1$ (Eq. (II.3.10)) involve different methods than the one presented for the Hamilton-Jacobi equations since it is not a time-dependent equation. A good approximation may be obtained by solving the redistanciation PDE Eq. (II.3.11) starting from an initial level-set representing the same shape, but numerically it is more stable to use an algorithm such as the fast marching method for which good explanations may be found for instance in [Set99, Section III.8] or [Dap13, Section 1.3.1].

II.4 Shape optimization algorithm based on Hadamard's shape derivative

We now have all the required information to implement in details the numerical Algorithm II.1.1 for solving the shape optimization program:

$$\begin{cases} \max\limits_{\Omega} & \mathcal{J}(\Omega) = \int_{\Omega} j(u)\,\mathrm{d}\mathbf{x} \\ \textbf{s.t.} & u \in H, \text{ and } a_\Omega(u,v) = b_\Omega(v) \text{ for all } v \in H \end{cases} . \tag{II.4.1}$$

II.4.1 Gradient descent algorithm

II.4.1.a Optimization scheme

In Section II.1.2 we have seen how to get the shape derivative $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ of $\mathcal{J}(\Omega)$ and that in general this shape derivative may be expressed as

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n} V_\Omega(s)\,\mathrm{d}s = \langle \boldsymbol{\theta}, \mathbf{n}V_\Omega(s)\rangle_{L^2(\partial\Omega,\mathbb{R})}\,.$$

With this information a gradient descent algorithm (note here that we use the term of "descent" algorithm even though we are considering the maximization of a functional) consists in an iterative computation of $\boldsymbol{\theta} = \mathbf{n}f(s)$ and the modification of the shape into $(\mathrm{Id} + \tau\boldsymbol{\theta})(\Omega)$ where τ is the step size (see Algorithm II.1.1 for a sketch of the optimization algorithm), an operation which is achieved by solving the Hamilton-Jacobi equation presented in Section II.3.3.a.

Starting from an initial shape, the previous method should end up into one of the local maxima of the considered objective function where the gradient, that is $\boldsymbol{\theta} = \mathbf{n}f(s)$ vanishes. For this to happen in practice, high numerical accuracy is required at each step of the optimization process in order to guarantee that the computed gradient is indeed an ascent direction for $\mathcal{J}(\Omega)$. The precision of the overall computation being altered by many parameters such as the simulation software accuracy (mesh of the geometrical domain, iterative solver, ...), the computation required for the velocity extension process (explained later in Section II.4.2), the grid size of the level-set function or even the small modification of the shape induced by the redistanciation process. An easy way to asses the validity of the full numerical method is to simply compare the value of $\mathcal{J}((\mathrm{Id}+\tau\boldsymbol{\theta})(\Omega))$ for some small values of τ with its first order approximation obtained by $\mathcal{J}(\Omega) + \tau\mathcal{J}'(\Omega)(\boldsymbol{\theta})$.

In this thesis we only consider gradient descent algorithm and not more evolved scheme such as non-linear conjugate gradient or second-order methods like Newton's or BFGS. The interested reader is referred to [Vié16, Chapter 9] for more details on these numerical algorithms applied to shape optimization.

II.4.1.b Line search

In order to guarantee convergence towards a local maximum (as well as to improve the convergence speed of the gradient descent algorithm) it is advised to perform a line search at each iteration so that the step value τ of the gradient (used in the update formula $(\mathrm{Id} + \tau\boldsymbol{\theta})(\Omega)$) at each iteration is sought so that the objective function increased "sufficiently". Since the calculation of the objective function is usually very noisy due to

numerical errors, we have decided not to use classical indicators to find a satisfactory time step such as Wolfe conditions but rather rely on the following strategy. Starting from an initial step τ_0 and the iteration number $n_{\text{it}} = 0$ we perform a line search by looking at the value of $\mathcal{J}((\text{Id} + \tau_{n_{\text{it}}}\boldsymbol{\theta})(\Omega))$ and considering two cases:

Case 1: If $\mathcal{J}((\text{Id} + \tau_{n_{\text{it}}}\boldsymbol{\theta})(\Omega)) \geq \mathcal{J}(\Omega)$, then we update the shape Ω as $(\text{Id} + \tau_{n_{\text{it}}}\boldsymbol{\theta})(\Omega)$ and since the step for the gradient was sufficient for this iteration, we can consider that a slightly higher value would also work meaning that we can take $\tau_{n_{\text{it}}+1} = \gamma_{\text{accepted}}\tau_{n_{\text{it}}}$ where $\gamma_{\text{accepted}} > 1$ (in our numerical test we used $\gamma_{\text{accepted}} = 1.1$).

Case 2: If $\mathcal{J}((\text{Id} + \tau_{n_{\text{it}}}\boldsymbol{\theta})(\Omega)) < \mathcal{J}(\Omega)$, then the shape is not modified and we continue the line search using this time $\tau_{n_{\text{it}}+1} = \gamma_{\text{rejected}}\tau_{n_{\text{it}}}$ where $\gamma_{\text{rejected}} \in]0, 1[$ (in our numerical test we used $\gamma_{\text{rejected}} = 0.5$).

Even with a very small step τ it may happen that $\mathcal{J}((\text{Id} + \tau\boldsymbol{\theta})(\Omega)) < \mathcal{J}(\Omega)$ because of the lack of numerical accuracy. This remark led us to modify the condition of the two cases by comparing $\mathcal{J}((\text{Id} + \tau_{n_{\text{it}}}\boldsymbol{\theta})(\Omega))$ with $\mathcal{J}(\Omega) - \eta$ (where $\eta > 0$ is a small tolerance value) to ensure that the line search will end in the hope that subsequent iterations will be subject to fewer numerical errors. Let us conclude this subsection by pointing out [Pir82, Section 4.4.3] in which the author proposed a strategy to perform a line search using a local parabola approximation.

II.4.1.c Stopping criteria

Theoretically, a gradient descent algorithm should stop when the gradient of the objective functional equals zero. Obviously, it will be impossible to achieve such a precision numerically and a more common stopping criteria is to request that the norm of the gradient $\|f(s)\|^2_{L^2(\partial\Omega,\mathbb{R})}$ be less than a given threshold ε. Ending the algorithm in this manner means that at the end of the optimization, if the shape is moved in any direction with a maximum amplitude of τ then it should not increase the objective function by more than $\tau\|f(s)\|^2_{L^2(\partial\Omega,\mathbb{R})}$ (if τ is sufficiently small so that the first order approximation is valid).

However, in practice, we noticed in our numerical tests that it is really difficult to get precise values of the gradient's norm at the end of the optimization process. This observation lead us to two other stopping criteria.

- The first method is to look at the relative increase of the objective function in the last n_{it} steps. If it has not improved by more than a given threshold during this time then we can assume that allowing the optimization algorithm to continue even further will not allow to increase the objective function by more than several time the threshold value unless the algorithm is left for many more than n_{it} iterations.

- The other idea is to stop the algorithm if the line search (see Section II.4.1.b) cannot find a new shape with a better value for the objective function even with a very small step since this means that the numerical errors are too important (the next iterations will therefore mainly consist in some kind of noise-optimization).

II.4.2 Velocity extension

II.4.2.a Introduction

As we have seen in Th. II.1.1.3, Sections II.2.1 and II.2.2, in practice, the shape derivative of the considered objective functions are always in the form

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n}\, V_\Omega(s)\,\mathrm{d}s = \langle \boldsymbol{\theta}, \mathbf{n}V_\Omega\rangle_{L^2(\partial\Omega,\mathbb{R})}\,, \tag{II.4.2}$$

where $V_\Omega \in L^2(\partial\Omega,\mathbb{R})$. Considering only normal vector fields $\boldsymbol{\theta}(\mathbf{x}) = \theta(\mathbf{x})\mathbf{n}$ we have $\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \langle\theta, f\rangle_{L^2(\partial\Omega,\mathbb{R})}$. Taking $\theta = \eta f$ on $\partial\Omega$ with η sufficiently small will, in theory, increase the objective function. Unfortunately, there are several problems with the use of this vector field since it is only defined at the edges $\partial\Omega$ of the shape. Although this is in principle not a problem (any values can be taken outside $\partial\Omega$ according to Eq. (II.4.2)), the efficiency of the numerical schemes presented in Section II.3.4 for the resolution of the Hamilton-Jacobi equation depends very much on the regularity of $\boldsymbol{\theta}$ which $V_\Omega(s)$ may be lacking off.

As a matter of fact, in order to numerically compute the values of the scalar field $V_\Omega(s)$ on $\partial\Omega$ we need, in principle, a mesh of this border, which is not always the case in finite difference or finite element simulations (see Section III.2.2 in which this problem is described in more details). Secondly, from a theoretical point of view, as seen in Section I.4.1, the trace of a PDE's solution (like Maxwell's equations) on the edges may have limited regularity.

To alleviate all these problems at once, a common method (see for instance [De 05, Chapter 3]) is to project the normal component $V_\Omega(s)$ of the shape derivative Eq. (II.4.2) on a highly regular Hilbert space, for instance $H^1(\mathcal{D},\mathbb{R})$. In other words we are now looking for a scalar field $f_{\text{reg}} \in H^1(\mathcal{D},\mathbb{R})$ such that its associated scalar product with all regular θ in $H^1(\mathcal{D},\mathbb{R})$ is equal to

$$\langle\theta, V_{\Omega,\text{reg}}\rangle_{H^1(\mathcal{D},\mathbb{R})} = \langle\theta, V_\Omega\rangle_{L^2(\partial\Omega,\mathbb{R})}\,. \tag{II.4.3}$$

Once the regularized scalar field f_{reg} is found we see that taking $\boldsymbol{\theta} = \mathbf{n}f_{\text{reg}}$ gives $\langle V_{\Omega,\text{reg}}, V_\Omega\rangle_{L^2(\partial\Omega,\mathbb{R})} = \|V_{\Omega,\text{reg}}\|^2_{H^1(\mathcal{D},\mathbb{R})} \geq 0$ and so this new vector field is indeed an ascent direction. In theory, since we have restricted ourselves to a sub-space of $L^2(\mathcal{D},\mathbb{R})$, the gradient $V_{\Omega,\text{reg}}$ is not an as good ascent direction as V_Ω.

II.4.2.b Projection as the solution of a PDE

To properly define Eq. (II.4.3) we need to choose an inner product for the space $H^1(\mathcal{D},\mathbb{R})$. The most convenient option is to have recourse on

$$\langle u, v\rangle_{H^1_\varepsilon(\mathcal{D},\mathbb{R})} = \int_{\mathcal{D}} \varepsilon\nabla u\cdot\nabla v + uv\,\mathrm{d}\mathbf{x}, \tag{II.4.4}$$

where $\varepsilon > 0$ is a small parameter. With this choice of inner product Eq. (II.4.3) is equivalent to finding $V_{\Omega,\text{reg}} \in H^1(\mathcal{D},\mathbb{R})$ such that for all $\theta \in H^1(\mathcal{D},\mathbb{R})$:

$$\int_{\mathcal{D}} \varepsilon\nabla V_{\Omega,\text{reg}}\cdot\nabla\theta + V_{\Omega,\text{reg}}\theta\,\mathrm{d}\mathbf{x} = \int_{\partial\Omega} V_\Omega\theta\,\mathrm{d}s \tag{II.4.5}$$

which is the variational formulation associated to

$$-\varepsilon\Delta V_{\Omega,\text{reg}} + V_{\Omega,\text{reg}} = V_\Omega \delta_{\partial\Omega} \text{ in } \mathcal{D} \quad \text{and} \quad \partial_{\mathbf{n}} V_{\Omega,\text{reg}} = 0 \text{ on } \partial\mathcal{D} \tag{II.4.6}$$

where $\delta_{\partial\Omega}$ is the Dirac distribution on $\partial\Omega$. Physically, the solution Eq. (II.4.6) gives approximately the same values of V_Ω on $\partial\Omega$; if $\varepsilon = 0$ then Eq. (II.4.6) drop down to $V_{\Omega,\text{reg}} = V_\Omega$ on $\partial\Omega$). The presence of the Laplacian operator on his part causes the diffusion of the values of V_Ω over a width of about ε^d where d is the dimension in which the set $\mathcal{D}$ is immersed; this may be justified by looking at the Green function associated to Eq. (II.4.6). In other word, the solution $V_{\Omega,\text{reg}}$ drops down to V_Ω after a gaussian blur of radius ε. In practice we use $\varepsilon = \alpha(\Delta x)^d$ with $\alpha = 2$ and Δx the spatial step of the level-set function.

Remark II.4.2.1: A classical result on elliptic equations makes it possible to show that a solution to Eq. (II.4.6) is of higher regularity than f which provides an understanding of why this regularization process is of numerical interest.

Let us also note that we can limit modifications of the shape on boundaries. Indeed, to prevent changes on $\Gamma \subset \partial\Omega$ we can search for a regular field $V_{\Omega,\text{reg}} \in H^1_\Gamma(\mathcal{D},\mathbb{R})$ defined as

$$H^1_\Gamma(\mathcal{D},\mathbb{R}) = \{u \in H^1(\mathcal{D},\mathbb{R}), u_{|\Gamma} = 0\}.$$

This is useful in Chapter III to preserve the continuity between waveguides and the component.

II.4.2.c Numerical resolution

To solve equation Eq. (II.4.6) or (II.4.5) let us first remark that usually the Dirac distribution $\delta_{\partial\Omega}$ is approximated using the level-set function ϕ as $\delta_{\partial\Omega} \simeq 1/\eta\zeta(\phi/\eta)|\nabla\phi|$ where ζ is a mollifier like $\zeta(x) = 1/\beta \times \exp(-1/(1-x^2))\delta_{]-1,1[}$ where β is chosen such that $\int_{\mathbb{R}} \zeta(x)\,\mathrm{d}x = 1$ and $\eta > 0$ is taken as small as possible and generally equals to the level-set grid spacing (note that the unit normalization of ζ is important to keep consistent values of the scalar product).

Concerning the effective numerical resolution of Eq. (II.4.6) or (II.4.5) either finite element or finite differences may be considered, both of them leading to the resolution of a system $Ax = b$ where the matrix A is always the same no matter the value of f. If the size of the problem is not too large it is then possible to compute only once the LU decomposition of A subsequently making this velocity extension/regularization step at each iteration almost computationally-free.

II.4.2.d Other regularizations/post-processing of the scalar field

Even after regularization some part of the scalar field $V_{\Omega,\text{reg}}$ may still cause numerical issues. This is particularly the case when really small details are present in the shape (think of a thin bar or a small isolated ball) where the scalar field may take extremely high values. Because of their amplitudes these areas will be almost the only ones where the shape moves at each iteration. Moreover, since the values of $V_{\Omega,\text{reg}}$ in these areas could in turn take positive and negative values (a bar may want to shrink to a narrower width than the mesh resolution so that at each iteration it will require either to be larger or finer) this will cause oscillations of the shape update process in these areas and prevent other

locations from being modified as required. To prevent this problem, a simple method is to limit the extreme values of the scalar field by, for instance, modifying $V_{\Omega,\text{reg}}$ into

$$\tilde{V}_{\Omega,\text{reg}} = \text{sign}(V_{\Omega,\text{reg}}) \min(|V_{\Omega,\text{reg}}|, q_{0.95})$$

where $q_{0.95}$ is defined as the 95 % quantile of $|V_{\Omega,\text{reg}}|$ (the minimum is also numerically implemented using some smooth minimum function). It is worth noting here than even after all these modification on the scalar field it is still an element of $W^{1,\infty}(\mathbb{R}^d, \mathbb{R})$ meaning that $\mathcal{J}'(\Omega)(\tilde{V}_{\Omega,\text{reg}}\mathbf{n})$ is still able to inform us about the (first order) sensitivity of the objective function when following the vector field $\tilde{V}_{\Omega,\text{reg}}\mathbf{n}$.

II.4.3 General framework

Algorithm II.4.1: Numerical algorithm for topology optimization of Eq. (II.2.1) using a level set representation and a simple line search.

1 **begin** *(initialization)*
2 ϕ := Initial level set function ϕ^0 (see Section II.3.3.b);
3 $\Omega := \{\mathbf{x} \in \mathcal{D}, \phi(\mathbf{x}) < 0\}$ *(making the shape geometry for simulation software)*;
4 u_Ω := solution of the PDE (Eq. (II.2.2)) using Ω;
5 $\mathcal{J}$:= value of the objective (Eq. (II.2.1)) using u_Ω;
6 v_Ω := solution of the adjoint PDE (Eq. (II.2.7)) using Ω and u_Ω;
7 $\tau := 1$ *(step factor for the gradient descent)*;
8 $\gamma_{\text{CFL}} := 0.4$ *(factor for Hamilton-Jacobi equation)*;
9 $\gamma_{\text{accepted}} := 1.1, \gamma_{\text{rejected}} := 0.5$ *(factor if the step is accepted or rejected)*;
10 **repeat** *(optimization)*
11 $V := V_\Omega$ the shape derivative (Eq. (II.2.9));
12 V_{reg} := solution of the regularization PDE (Eq. (II.4.5)) using V;
13 $\tilde{V}_{\text{reg}}$:= eventual other regularizations/post-processing (see Section II.4.2.d);
14 $\delta_{\text{CFL}} := \gamma_{\text{CFL}} \times$ CFL value (Eq. (II.3.22)) using $\theta = \tilde{V}_{\text{reg}}$;
15 **begin** *(linear search)*
16 ψ := solve the Hamilton-Jacobi equation (Eq. (II.3.17)) for $t \in]0, \tau\delta_{\text{CFL}}/\gamma_{\text{CFL}}]$ using ϕ as initial condition and scalar field $\tilde{V}_{\text{reg}}$;
17 ψ := redistanciation of ψ solving Eq. (II.3.23);
18 $\Omega := \{\mathbf{x} \in \mathcal{D}, \psi(\mathbf{x}) < 0\}$ *(making the geometry for the software)*;
19 u_Ω := solution of the PDE (Eq. (II.2.2)) using Ω;
20 $\mathcal{J}_{\text{tmp}}$:= value of the objective (Eq. (II.2.1)) using u_Ω;
21 v_Ω := solution of the adjoint PDE (Eq. (II.2.7)) using Ω and u_Ω;
22 **if** $\mathcal{J}_{tmp}$ *is sufficiently larger than* $\mathcal{J}$ *(see Section II.4.1.b)* **then**
23 $\phi := \psi$, $\mathcal{J} := \mathcal{J}_{\text{tmp}}$;
24 $\tau := \gamma_{\text{accepted}}\tau$;
25 *goto Line 10*
26 **else**
27 $\tau := \gamma_{\text{rejected}}\tau$;
28 *goto Line 15*
29 **until** *convergence (see Section II.4.1.c)*;
30 **return** Ω;

We conclude this chapter with Algorithm II.4.1, the pseudo-code of a generic shape optimization algorithm using all the information presented throughout the previous sections.

Chapter III

Optimal design of photonic components

Summary — This chapter combines the results of Chapters I and II to achieve shape optimization of single-objective nanophotonic components.

Section III.1 starts with a short overview of the state of the art in topology optimization applied to nanophotonics devices; the bases of the most popular methods are presented, as well as some references where they have been used.

The second Section III.2 adapts the presentation of geometric shape optimization explained in Chapter II to the context of nanophotonics by giving the rigorous mathematical calculation of the shape derivative of the objective function presented in the first chapter. A strategy featuring a smoothing of the refractive index in order to stabilize the numerical implementation is discussed in details.

In Section III.3 we present several optimization results demonstrating that our numerical method manages to maximize the power carried by an outgoing waveguide mode when it comes to designing components such as crossings, mode converters, mirrors or power dividers. Some of the optimized devices presented in this section are coming from our published paper

> [Leb19a] N. Lebbe, C. Dapogny, E. Oudet, K. Hassan, and A. Gliere. "Robust shape and topology optimization of nanophotonic devices using the level set method". In: *Journal of Computational Physics* (2019). DOI: 10.1016/j.jcp.2019.06.057.

After presenting these results, some comments are added concerning the use of the topological gradient in the optimization process.

The short Section III.4 presents some of the technological constraints hindering the fabrication of arbitrary geometries and some recently published method to deal with such constraints in the context of geometric shape optimization.

Finally, Section III.5 is dedicated to the optimization of multi-layers components. This topic was, until very recently, ignored in shape optimization and has allowed us to achieve significant results concerning the optimization of nanophotonic polarization rotators. One part of the results presented in this section was published in the journal article

> [Leb19b] N. Lebbe, A. Glière, and K. Hassan. "High-efficiency and broadband photonic polarization rotator based on multilevel shape optimization". In: *Optics Letters* 44.8 (2019), pp. 1960–1963. DOI: 10.1364/OL.44.001960.

III.1 State of the art for the design of photonic components using topology optimization methods

In Section I.3.4, we presented a general problem whose resolution make it possible to optimize nanophotonic components. More precisely, in Section I.3.1.c, we have seen that it is desirable to control the way in which an incident light arriving in the form of a guided mode in one waveguide is transformed and redirected to other ones. To characterize a component, we defined the scattering parameters $S_{m,n}$ whose squared absolute values give the power carried by a guided mode on a waveguide. In details, to maximize the power carried by the n-th outgoing mode $(\boldsymbol{\mathcal{E}}_{-n}^{\text{out}}, \boldsymbol{\mathcal{H}}_{-n}^{\text{out}})$ on a waveguide's cross section Γ_{out} when the m-th input mode $(\boldsymbol{\mathcal{E}}_{m}^{\text{in}}, \boldsymbol{\mathcal{H}}_{m}^{\text{in}})$ is injected into the section Γ_{in} of a waveguide accounts to solve

$$\max_{\Omega \subset \mathcal{D}_{\text{opt}}} \mathcal{J}(\Omega) = |S_{m,n}(\mathbf{E}_\Omega)|^2 \quad \text{with} \quad S_{m,n}(\mathbf{E}_\Omega) = \frac{1}{2}\int_{\Gamma_{\text{out}}} \left[\mathbf{E}_\Omega \times \boldsymbol{\mathcal{H}}_{-n}^{\text{out},*}\right] \cdot \mathbf{n}\, \mathrm{d}s, \quad \text{(III.1.1)}$$

where $\mathcal{D}_{\text{opt}}$ is a subset of $\mathcal{D} \subset \mathbb{R}^3$. In Eq. (III.1.1), the field $\mathbf{E}_\Omega$ refer to the unique solution of the following time-harmonic vector wave equation

$$\left\{\begin{array}{rcll} \nabla \times \Lambda^{-1} \nabla \times \mathbf{E} - k^2 n_\Omega^2 \Lambda \mathbf{E} & = & 0 & \text{in } \mathcal{D} \\ \mathbf{n} \times \mathbf{E} & = & 0 & \text{on } \partial\mathcal{D} \backslash (\Gamma_{\text{out}} \cup \Gamma_{\text{in}}) \\ \mathbf{n} \times \nabla \times \mathbf{E} + \gamma_{\text{out}}(\mathbf{E}) & = & 0 & \text{on } \Gamma_{\text{out}} \\ \mathbf{n} \times \nabla \times \mathbf{E} + \gamma_{\text{in}}(\mathbf{E}) & = & 2i\omega\mu_0 \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_m^{\text{in}} & \text{on } \Gamma_{\text{in}} \end{array}\right., \quad \text{(III.1.2)}$$

in which n_Ω is the optical index equal, in $\mathcal{D}_{\text{opt}}$, to either n_{core} inside Ω and n_{clad} elsewhere. The matrix Λ is defined through the PML as Eq. (I.3.17) and the operators γ_{in} and γ_{out} are given by

$$\gamma_{\text{in}}(\mathbf{E}) = \frac{1}{2}\sum_{j=1}^{N} i\omega\mu_0\, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_j^{\text{in}} \int_{\Gamma_{\text{in}}} \left(\mathbf{E} \times \boldsymbol{\mathcal{H}}_j^{\text{in},*}\right) \cdot \hat{\mathbf{z}}\, \mathrm{d}s,$$

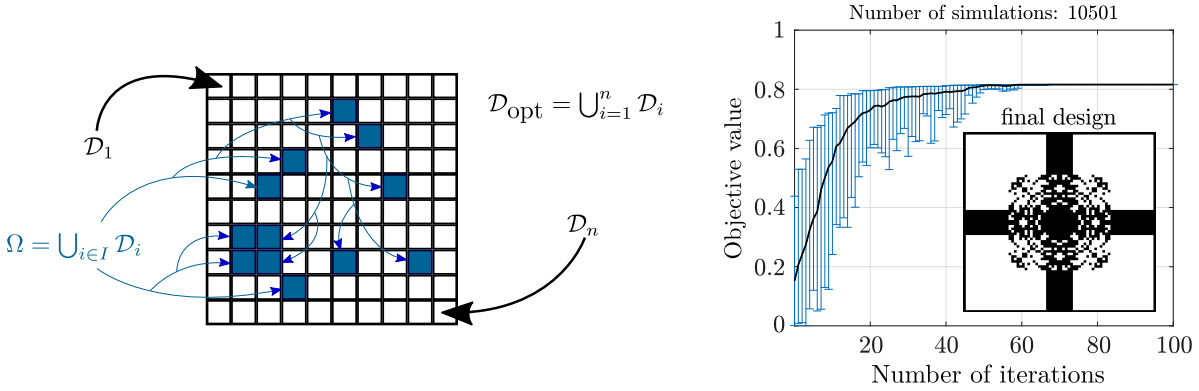
$$\gamma_{\text{out}}(\mathbf{E}) = \frac{1}{2}\sum_{j=1}^{M} i\omega\mu_0\, \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_{-j}^{\text{out}} \int_{\Gamma_{\text{out}}} \left(\mathbf{E} \times \boldsymbol{\mathcal{H}}_{-j}^{\text{out},*}\right) \cdot \hat{\mathbf{z}}\, \mathrm{d}s.$$

To design nanophotonic devices which maximize Eq. (III.1.1), physicists derived several methods based either on simplifications of the Maxwell equations and analytic calculations or on parameterized designs composed of geometric primitives and optimization of a few number of parameters. One limitation of these approaches is that, for each component, a new analysis must be made to adapt the existing tools to this new particular case; i.e. it requires to find a good approximation of the Maxwell equations and the interesting parameters to optimize. Moreover, the approximations of the considered PDE may be valid only for sufficiently large optimization domain $\mathcal{D}_{\text{opt}}$. This notably means that, with these methods, it is not possible to obtain designs with arbitrarily fixed dimensions.

In the past ten to twenty years, researchers have been working on methods to automatically find the design of photonic components without relying on an a priori parametrization of the desired design, with the hope to obtaining non-intuitive designs that may, among other advantages, be much more compact than the previously obtained components. In this section we briefly review the main non-parametric methods that have been applied to the shape optimization of nanophotonic components so far.

III.1.1 Binarization-based methods

We begin this review with the methods which involves a binary discretization of the desired shape. The general idea is to discretize the optimization domain $\mathcal{D}_{\text{opt}}$ into pixels (see Fig. III.1.1(a)) meaning that $\mathcal{D}_{\text{opt}} = \bigcup_{i=1}^{n} \mathcal{D}_i$ and search for a subset $I \subset 1, \dots, n$ of these pixels which allows to maximize the value of the objective function for the design $\Omega = \bigcup_{i \in I} \mathcal{D}_i$. These methods rely on optimization algorithms featuring a discrete (large in practice) set of binary variables.



(a) *Discretization of the domain $\mathcal{D}_{opt}$ into pixels $\mathcal{D}_i$; example of a shape Ω in blue consisting in some pixels*

(b) *Extreme values of the objective in the population at each iteration of the algorithm represented with the blue bar plots*

Figure III.1.1: Optimization of a nanophotonic crossing (see Section III.3.1.d for a presentation of this component) using genetic algorithm with the `ga` function of `Matlab` and its defaults parameters. About 10 000 simulations for the whole optimization. It should be noted that because of this high number of simulations (compared to the SIMP method presented below) the mesh used during this optimization was very coarse.

Among the multiple methods implemented to deal with this type of binary optimal design problems, let us mention

- Particle swarm methods; see [Mak16]. This metaheuristic considers several shapes (a population) and at each iteration of the algorithm, modifies all the shapes in different ways depending on the objective value obtained for each of the other shape. These modifications are carried out by considering a "distance" between the shapes and by transforming a shape in a "direction" given by its neighbors (that is, the shape which are the closest to this one) as in a swarm.

- Evolutionary algorithms; see [Gon08] or Fig. III.1.1. This metaheuristic also considers a population of shapes but, at each iteration of the algorithm, a new population of shapes is generated from some of the best shapes (according to the values of the objective function) by modifying the subset I of pixels constituting the shapes with operations inspired from biological evolution.

- Deep learning methods; see [Tah19]. Contrary to the two previous methods, this one is rather a way to reduce the overall computation time and thus to quickly test many shapes. By using a neural network, the algorithm learns how the electromagnetic field propagates into a component identified by the subset I used in the discretization. To this end, the results of many electromagnetic simulations are

supplied to the neural network, using for instance random sets I. Once these data have been processed, the neural network is then capable of predicting, at almost no additional computational cost, the output transmission for any shape Ω. Using this model, some optimization algorithms are then launched on the component in order to find the best shape to obtain the desired output transmissions.

Let us also mention [She15; Maj17] and [Xu17; Liu18] which respectively consider the Direct-binary-search and Fast-search methods but very few details concerning their implementations are given in these papers.

One of the main advantages of all these binary-based methods is that they do not require a precise analysis of the Maxwell equations and may be used as black-box optimization procedure. Another significant interest of these methods is that they produce a shape composed only of pixels. These "Manhattan" structures are well-known in microelectronic; several methods have been specifically developed to manufacture them correctly.

III.1.2 Density-based optimization methods

In order to use gradient based methods, it is necessary to manipulate continuous variables. The most logical idea in this direction judging from the presentation of the previous section is to relax the binary constraint on the pixels by allowing each of them to be composed of materials with optical indices taking intermediate values between the one of the cladding and core material. Mathematically, the problem is that of finding a density function $\rho : \mathcal{D}_{\text{opt}} \to [0, 1]$, associated with an optical index n_ρ defined for all $\mathbf{x} \in \mathcal{D}_{\text{opt}}$ by

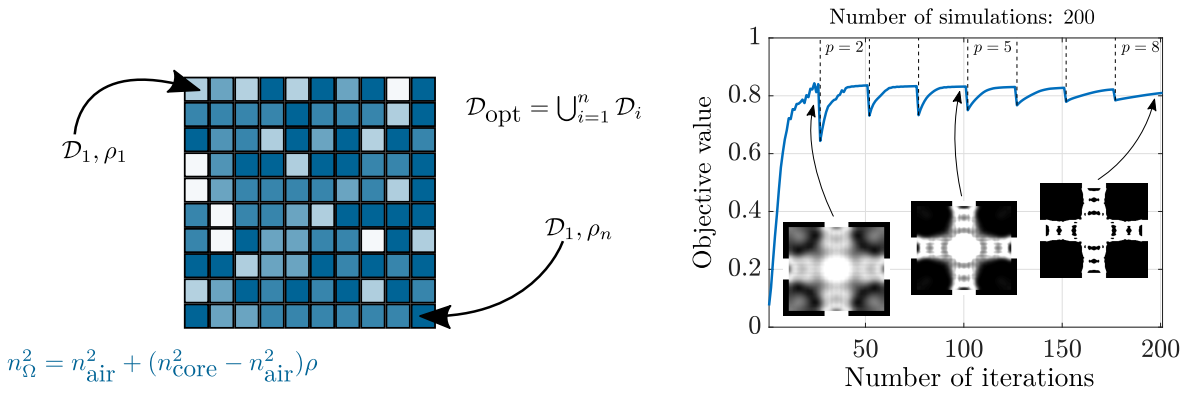
$$n_\rho(\mathbf{x})^2 = n_{\text{clad}}^2 + (n_{\text{core}}^2 - n_{\text{clad}}^2)\rho(\mathbf{x}),$$

which makes optimal a "relaxed" counterpart $\mathcal{J}(\rho)$ of the objective function $\mathcal{J}(\Omega)$. More precisely, the optimization program Eq. (III.1.1) which seeks to optimize $\mathcal{J}(\Omega)$ is rewritten as

$$\begin{array}{ll} \max\limits_{\rho,\mathbf{E}} & \mathcal{J}(\rho) = J(\mathbf{E}), \\ \textbf{s.t.} & \nabla \times \nabla \times \mathbf{E} - k^2 n_\rho^2 \mathbf{E} = 0 \end{array}, \tag{III.1.3}$$

where we simplified the time-harmonic vector wave equation for simplicity. Using for instance a gradient-based algorithm (the gradient may be found using a variation of the adjoint method presented in Section II.2.1) or a more advanced mathematical programming algorithm such as the MMA method, we can solve Eq. (III.1.3), resulting in a density ρ. This is unfortunately not a manufacturable shape since it is only possible to produce the cladding and core materials corresponding respectively to $\rho = 0$ or 1; see Fig. III.1.2. To alleviate this problem, one idea is to penalize the intermediate values of ρ. To achieve this, two popular methods known under the name of the SIMP method modifies the optical index n_ρ in Eq. (III.1.3) into n_{ρ^p} where $p > 1$ or into $H(n_\rho)$ where H is an "Heaviside filter" function, which makes it easier to obtain a black and white design. For more details see the review paper [Jen11] which presents this method in the context of nanophotonics.

Remark III.1.2.1: Another method, known as "objective-first", which grew popular after the publication of [Pig15] (for the implementation details see [Lu13, Appendix C]) is to exchange the objective and PDE-constraint of the optimization program in Eq. (III.1.3); we minimize the error (according to a given norm) in the fulfillment of the time-harmonic vector wave equation by the actual electric field $\mathbf{E}$ while constraining the "true" objective function of the optimization problem to be above a given



(a) *Discretization of the domain $\mathcal{D}_{opt}$ into pixels $\mathcal{D}_i$; example of a density ρ in blue*

(b) *Convergence history graph*

Figure III.1.2: Optimization of a nanophotonic crossing (see Section III.3.1.d for a presentation of this component) using a density-based gradient descent algorithm using the SIMP procedure. About 200 simulations for the whole optimization.

threshold τ:

$$\begin{array}{ll} \min\limits_{\rho,\mathbf{E}} & \|\nabla \times \nabla \times \mathbf{E} - k^2 n_\rho^2 \mathbf{E}\|, \\ \textbf{s.t.} & \mathcal{J}(\rho) > \tau \end{array},$$

the parameter τ being for instance taken equal to 0.9.

III.1.3 Geometrical shape optimization

Geometrical shape optimization methods are widely detailed in Chapter II. They have already been used in a few studies in the context of optimization of nanophotonic components.

To the best of our knowledge, the first papers dealing with geometric variations of a shape for the optimization of photonic devices is [Kao05], in the context of photonic crystals, and [Lal13] for nanophotonic components, such as the one presented hereafter. The analysis performed in this paper is fairly different from the one presented in Chapter II and the sections hereafter since it is based on the use of Green's functions to compute the sensitivity of the objective function in the case of a small geometric perturbation of the shape.

As denoted in Remark II.3.1.1, it is also possible to find a shape through the optimization of one of its level-set representation as it is done in the calculus of variation; see for instance the supplementary materials of [Pig17; Ver19a].

III.2 Hadamard's shape optimization method applied to photonic components

In Chapter II, a shape optimization algorithm based on Hadamard's boundary variation method was presented. In order to apply it to the optimization of nanophotonic components, the first necessary step is to find the shape derivative of the considered objective

function Eq. (III.1.1); this is the purpose of Section III.2.1. The second subsection III.2.2 then presents the index smoothing method which trades the sharp-interface (jump of the optical index value on $\partial\Omega$) into an approximate, smoothed interface, in order to reduce numerical errors when computing this shape derivative. Section III.2.3 concludes this section by presenting some variations of the first subsection's results in the case of other objective functions than the ones considered in Section III.2.1.

III.2.1 Shape derivative of the general model problem

III.2.1.a Rigorous calculation of the model objective

The following theorem gives the shape derivative of Eq. (III.1.1) and will be widely reused throughout this chapter and the next one to optimize nanophotonic components. Note that a similar result was already found in [Joh02] or [Mil13, Section 5.1], but without a full mathematical analysis.

Theorem III.2.1.1 – Shape derivative of the power carried by a mode.
The shape derivative associated to the problem Eq. (III.1.1) is given by:

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n}\, V_\Omega(s)\,\mathrm{d}s \tag{III.2.1}$$

with

$$V_\Omega(s) = k^2\mathrm{Re}\left[\frac{S_{m,n}(\mathbf{E}_\Omega)^*}{2i\omega\mu_0}\left(\left[\!\left[n_\Omega^2\right]\!\right]\mathbf{E}_{\Omega,\parallel}\cdot\mathbf{A}^*_{\Omega,\parallel} - \left[\!\left[\frac{1}{n_\Omega^2}\right]\!\right](n_\Omega^2\mathbf{E}_{\Omega,\perp})\,(n_\Omega^2\mathbf{A}^*_{\Omega,\perp})\right)\right] \tag{III.2.2}$$

where $[\![X]\!] = X_{|\bar{\mathcal{D}}\setminus\Omega} - X_{|\Omega}$ denotes the jump of the quantity X at a discontinuous interface, $\mathbf{X}_\parallel = \mathbf{n}\times\mathbf{X}\times\mathbf{n}$ the tangential components of $\mathbf{X}$, $\mathbf{X}_\perp = \mathbf{X}\cdot\mathbf{n}$ its normal component and $\mathbf{A}_\Omega$ the unique solution to the time-harmonic vector wave equation when injecting the output mode into the output waveguide, or equivalently stated that $\mathbf{A}_\Omega$ solves

$$\left\{\begin{array}{rcll} \nabla\times\Lambda^{-1}\nabla\times\mathbf{A} - k^2n_\Omega^2\Lambda\mathbf{A} &=& 0 & \text{in } \mathcal{D}\\ \mathbf{n}\times\mathbf{A} &=& 0 & \text{on } \partial\mathcal{D}\setminus(\Gamma_{\text{out}}\cup\Gamma_{\text{in}})\\ \mathbf{n}\times\nabla\times\mathbf{A} + \gamma_{\text{out}}(\mathbf{A}) &=& 2i\omega\mu_0\hat{\mathbf{z}}\times\boldsymbol{\mathcal{H}}_n^{\text{out}} & \text{on } \Gamma_{\text{out}}\\ \mathbf{n}\times\nabla\times\mathbf{A} + \gamma_{\text{in}}(\mathbf{A}) &=& 0 & \text{on } \Gamma_{\text{in}} \end{array}\right. . \tag{III.2.3}$$

Remark III.2.1.1: In order to prove this theorem we will need the following result for which elements of proof may be found in [Mon03, Remark 3.48] or [Dau12, Chapter IX]:

The restrictions $\mathbf{E}_{\Omega,1}$ and $\mathbf{E}_{\Omega,2}$ of $\mathbf{E}_\Omega$ on Ω and $\mathcal{D}\setminus\bar{\Omega}$ satisfy additional smoothness to that encoded in the spaces $H(\boldsymbol{curl},\Omega)$ and $H(\boldsymbol{curl},\mathcal{D}\setminus\bar{\Omega})$; this is a classical issue in the theory of elliptic partial differential equations, which follows from the smoothness of Ω; typically, in our context: $\mathbf{E}_{\Omega,1}\in H^1(\Omega)$ and $\mathbf{E}_{\Omega,2}\in H^1(\mathcal{D}\setminus\bar{\Omega})$. The same result holds for $\mathbf{A}_\Omega$.

Proof of Th. III.2.1.1. This proof follows the general four-parts method presented in Section II.2.2.

1. Variational form of the transported field

The variational form associated to $\mathbf{E}_{\boldsymbol{\theta}}$ solution of Eq. (III.1.2) for $\Omega_{\boldsymbol{\theta}} = (\mathrm{Id} + \boldsymbol{\theta})(\Omega)$ (and optical index $n_{\boldsymbol{\theta}}$) is given by:

$$\int_{\mathcal{D}} \Lambda^{-1} \nabla \times \mathbf{E}_{\boldsymbol{\theta}} \cdot \nabla \times \boldsymbol{\phi}^* - k^2 n_{\boldsymbol{\theta}}^2 \Lambda \mathbf{E}_{\boldsymbol{\theta}} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} - \int_{\Gamma_{\mathrm{in}}} \gamma_{\mathrm{in}}(\mathbf{E}_{\boldsymbol{\theta}}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s$$
$$- \int_{\Gamma_{\mathrm{out}}} \gamma_{\mathrm{out}}(\mathbf{E}_{\boldsymbol{\theta}}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s + \int_{\Gamma_{\mathrm{in}}} 2i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_m^{\mathrm{in}} \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s = 0, \quad \text{(III.2.4)}$$

where $\mathbf{E}_{\boldsymbol{\theta}}$ and $\boldsymbol{\phi}$ are elements of the Hilbert space $\mathcal{V}$ defined as

$$\mathcal{V} = \{\boldsymbol{\psi} \in H(\boldsymbol{curl}, \mathcal{D}), \boldsymbol{\psi} = 0 \text{ on } \partial\mathcal{D} \backslash (\Gamma_{\mathrm{out}} \cup \Gamma_{\mathrm{in}})\}.$$

Using a change of variables, the first integral in Eq. (III.2.4) may be expressed as

$$\int_{\mathcal{D}} [\Lambda^{-1}(\nabla \times \mathbf{E}_{\boldsymbol{\theta}}) \circ (\mathrm{Id} + \boldsymbol{\theta}) \cdot (\nabla \times \boldsymbol{\phi}^*) \circ (\mathrm{Id} + \boldsymbol{\theta})$$
$$- k^2 n_{\boldsymbol{\theta}}^2 \circ (\mathrm{Id} + \boldsymbol{\theta}) \Lambda \mathbf{E}_{\boldsymbol{\theta}} \circ (\mathrm{Id} + \boldsymbol{\theta}) \cdot \boldsymbol{\phi}^* \circ (\mathrm{Id} + \boldsymbol{\theta})] |\det(\mathrm{I} + \nabla\boldsymbol{\theta})| \, \mathrm{d}\mathbf{x}. \quad \text{(III.2.5)}$$

We define the transported field as $\bar{\mathbf{E}}_{\boldsymbol{\theta}} = (\mathrm{I} + \nabla\boldsymbol{\theta}^\top)\mathbf{E}_{\boldsymbol{\theta}} \circ (\mathrm{Id} + \boldsymbol{\theta})$ and its counterpart for $\bar{\boldsymbol{\phi}}_{\boldsymbol{\theta}}$. The following identity follows

$$(\nabla \times \mathbf{E}_{\boldsymbol{\theta}}) \circ (\mathrm{Id} + \boldsymbol{\theta}) = |\det(\mathrm{I} + \nabla\boldsymbol{\theta})|^{-1} (\mathrm{I} + \nabla\boldsymbol{\theta})(\nabla \times \bar{\mathbf{E}}_{\boldsymbol{\theta}}).$$

Together with the shortcuts

$$A(\boldsymbol{\theta}) = C(\boldsymbol{\theta})^{-1} = |\det(\mathrm{I} + \nabla\boldsymbol{\theta})|^{-1} (\mathrm{I} + \nabla\boldsymbol{\theta}^\top)(\mathrm{I} + \nabla\boldsymbol{\theta}),$$

Eq. (III.2.5) is equal to

$$\int_{\mathcal{D}} \Lambda^{-1} A(\boldsymbol{\theta}) \nabla \times \bar{\mathbf{E}}_{\boldsymbol{\theta}} \cdot \nabla \times \boldsymbol{\phi}^* - k^2 n^2 \Lambda C(\boldsymbol{\theta}) \bar{\mathbf{E}}_{\boldsymbol{\theta}} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x}.$$

Since $\boldsymbol{\theta} \mapsto (\mathrm{Id} + \boldsymbol{\theta})$ is an homeomorphism, $\boldsymbol{\theta} = 0$ on both $\Gamma_{\mathrm{in}}, \Gamma_{\mathrm{out}}$ and $n_{\boldsymbol{\theta}} \circ (\mathrm{Id} + \boldsymbol{\theta}) = n_\Omega$, $\bar{\mathbf{E}}_{\boldsymbol{\theta}}$, we get that $\bar{\mathbf{E}}_{\boldsymbol{\theta}}$ is solution for all $\boldsymbol{\phi} \in \mathcal{V}$ of

$$\int_{\mathcal{D}} \Lambda^{-1} A(\boldsymbol{\theta}) \nabla \times \bar{\mathbf{E}}_{\boldsymbol{\theta}} \cdot \nabla \times \boldsymbol{\phi}^* - k^2 n_\Omega^2 \Lambda C(\boldsymbol{\theta}) \bar{\mathbf{E}}_{\boldsymbol{\theta}} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} - \int_{\Gamma_{\mathrm{in}}} \gamma_{\mathrm{in}}(\bar{\mathbf{E}}_{\boldsymbol{\theta}}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s$$
$$- \int_{\Gamma_{\mathrm{out}}} \gamma_{\mathrm{out}}(\bar{\mathbf{E}}_{\boldsymbol{\theta}}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s + \int_{\Gamma_{\mathrm{in}}} 2i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_m^{\mathrm{in}} \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s = 0. \quad \text{(III.2.6)}$$

2. Differentiability of the transported field w.r.t. the vector field $\boldsymbol{\theta}$

To prove that $\mathbf{E}_\Omega$ has a Lagrangian derivative, we define a F map from $W^{1,\infty}(\mathbb{R}^3, \mathbb{R}^3) \times \mathcal{V}$ into the dual space $\mathcal{V}^*$:

$$F(\boldsymbol{\theta}, \boldsymbol{\psi}) : \boldsymbol{\phi} \mapsto a_{\boldsymbol{\theta}}(\boldsymbol{\psi}, \boldsymbol{\phi}) - b_{\boldsymbol{\theta}}(\boldsymbol{\phi}).$$

where $a_{\boldsymbol{\theta}}$ (resp. $b_{\boldsymbol{\theta}}$) is the sesquilinear (resp. antilinear) part of Eq. (III.2.6). Assuming the existence and uniqueness to a solution $\boldsymbol{\psi}$ of $a(\boldsymbol{\psi}, \boldsymbol{\phi}) = b(\boldsymbol{\phi})$ for all $\boldsymbol{\phi} \in \mathcal{V}$, the implicit function theorem implies that $\boldsymbol{\theta} \to \bar{\mathbf{E}}_{\boldsymbol{\theta}}$ is Fréchet differentiable (see Section II.2.2 for more details about this second step of the demonstration) meaning that it has a Lagrangian derivative $\mathring{\mathbf{E}}_\Omega$.

3. Volumetric shape derivative

Differentiation with respect to $\boldsymbol{\theta}$ of the variational form Eq. (III.2.6) gives:

$$\begin{aligned}\int_{\mathcal{D}} \Lambda^{-1} A'(0)(\boldsymbol{\theta}) \nabla \times \mathbf{E}_\Omega \cdot \nabla \times \boldsymbol{\phi}^* + \Lambda^{-1} \nabla \times \mathring{\mathbf{E}}(\boldsymbol{\theta}) \cdot \nabla \times \boldsymbol{\phi}^* \\ - k^2 n_\Omega^2 \Lambda C'(0)(\boldsymbol{\theta}) \mathbf{E}_\Omega \cdot \boldsymbol{\phi}^* - k^2 n_\Omega^2 \Lambda \mathring{\mathbf{E}}_\Omega(\boldsymbol{\theta}) \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} \\ - \int_{\Gamma_{\text{in}}} \gamma_{\text{in}}(\mathring{\mathbf{E}}_\Omega(\boldsymbol{\theta})) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s - \int_{\Gamma_{\text{out}}} \gamma_{\text{out}}(\mathring{\mathbf{E}}_\Omega(\boldsymbol{\theta})) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s = 0 \end{aligned} \quad \text{(III.2.7)}$$

where $A'(0)(\boldsymbol{\theta}) = -C'(0)(\boldsymbol{\theta}) = -(\nabla \cdot \boldsymbol{\theta})\mathrm{I} + \nabla\boldsymbol{\theta} + \nabla\boldsymbol{\theta}^\top$. The objective function Eq. (III.1.1) may also be expressed with $\bar{\mathbf{E}}_{\boldsymbol{\theta}}$ (remember again that $\boldsymbol{\theta} = 0$ on Γ_{out}):

$$\mathcal{J}(\Omega_{\boldsymbol{\theta}}) = \left| \frac{1}{2} \int_{\Gamma_{\text{out}}} \bar{\mathbf{E}}_{\boldsymbol{\theta}} \times \boldsymbol{\mathcal{H}}_n^{\text{out},*} \cdot \mathbf{n} \, \mathrm{d}s \right|^2 .$$

Its shape derivative is given by:

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \mathrm{Re}\left[\left(\int_{\Gamma_{\text{out}}} \mathring{\mathbf{E}}(\boldsymbol{\theta}) \times \boldsymbol{\mathcal{H}}_n^{\text{out},*} \cdot \mathbf{n} \, \mathrm{d}s \right) S_{m,n}(\Omega_{\boldsymbol{\theta}})^* \right]. \quad \text{(III.2.8)}$$

We consider an adjoint state $\mathbf{A}_\Omega$ solution of the following variational formulation:

$$\begin{aligned}\int_{\mathcal{D}} \Lambda^{-1} \nabla \times \mathbf{A} \cdot \nabla \times \boldsymbol{\phi}^* - k^2 n_\Omega^2 \Lambda \mathbf{A} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} - \int_{\Gamma_{\text{in}}} \gamma_{\text{in}}(\mathbf{A}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s \\ - \int_{\Gamma_{\text{out}}} \gamma_{\text{out}}(\mathbf{A}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s + 2i\omega\mu_0 \left(\int_{\Gamma_{\text{out}}} \boldsymbol{\phi} \times \boldsymbol{\mathcal{H}}_n^{\text{out},*} \cdot \mathbf{n} \, \mathrm{d}s \right)^* = 0. \end{aligned} \quad \text{(III.2.9)}$$

The antilinear part in the variational formulation Eq. (III.2.9) is of the same form as the one in Eq. (III.2.4), that is a mode injection:

$$2i\omega\mu_0 \left(\int_{\Gamma_{\text{out}}} \boldsymbol{\phi} \times \boldsymbol{\mathcal{H}}_n^{\text{out},*} \cdot \mathbf{n} \, \mathrm{d}s \right)^* = \int_{\Gamma_{\text{out}}} 2i\omega\mu_0 \mathbf{n} \times \boldsymbol{\mathcal{H}}_n^{\text{out}} \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s.$$

But since in the objective the mode is a backward one and the normal component on Γ_{out} point out in the opposite direction as the one on Γ_{in} ($\mathbf{n} = -\hat{\mathbf{z}}$ and $\mathbf{n} \times \boldsymbol{\mathcal{H}}_m = -\mathbf{n} \times \boldsymbol{\mathcal{H}}_{-m}$) we end up with the adjoint being the same as an injection of the forward mode in the output waveguide.

Note also that $(\mathbf{E}, \boldsymbol{\phi}) \to \langle \gamma(\mathbf{E}), \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \rangle_{L^2(\Gamma,\mathbb{C})}$ is sesquilinear, indeed:

$$\begin{aligned}\int_\Gamma \gamma(\mathbf{E}) \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}s &= \int_\Gamma \frac{1}{2} \sum_{j=1}^N i\omega\mu_0 \hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_j \int_\Gamma \left(\mathbf{E} \times \boldsymbol{\mathcal{H}}_j^* \right) \cdot \hat{\mathbf{z}} \, \mathrm{d}s \cdot \mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \, \mathrm{d}t \\ &= \int_\Gamma \frac{1}{2} \sum_{j=1}^N i\omega\mu_0 \left(\hat{\mathbf{z}} \times \boldsymbol{\mathcal{H}}_j^* \right) \int_\Gamma (\mathbf{n} \times \boldsymbol{\phi}^* \times \mathbf{n} \times \boldsymbol{\mathcal{H}}_j) \cdot \hat{\mathbf{z}} \, \mathrm{d}t \cdot \mathbf{E} \, \mathrm{d}s \\ &= \left(\int_\Gamma \gamma(\boldsymbol{\phi}) \cdot \mathbf{n} \times \mathbf{E}^* \times \mathbf{n} \, \mathrm{d}s \right)^* . \end{aligned}$$

Using $\boldsymbol{\phi} = \mathbf{A}_\Omega$ in (III.2.7) and $\boldsymbol{\phi} = \mathring{\mathbf{E}}(\boldsymbol{\theta})$ in (III.2.9) we find that

$$\begin{aligned}\int_{\mathcal{D}} \Lambda^{-1} \nabla \times \mathring{\mathbf{E}}(\boldsymbol{\theta}) \cdot \nabla \times \mathbf{A}_\Omega^* - k^2 n_\Omega^2 \Lambda \mathring{\mathbf{E}}(\boldsymbol{\theta}) \cdot \mathbf{A}_\Omega^* \, \mathrm{d}\mathbf{x} - \int_{\Gamma_{\text{in}}} \gamma_{\text{in}}(\mathring{\mathbf{E}}(\boldsymbol{\theta})) \cdot \mathbf{n} \times \mathbf{A}_\Omega^* \times \mathbf{n} \, \mathrm{d}s \\ - \int_{\Gamma_{\text{out}}} \gamma_{\text{out}}(\mathring{\mathbf{E}}(\boldsymbol{\theta})) \cdot \mathbf{n} \times \mathbf{A}_\Omega^* \times \mathbf{n} \, \mathrm{d}s \\ = - \int_{\mathcal{D}} \Lambda^{-1} A'(0)(\boldsymbol{\theta}) \nabla \times \mathbf{E}_\Omega \cdot \nabla \times \mathbf{A}_\Omega^* - k^2 n^2 \Lambda C'(0)(\boldsymbol{\theta}) \mathbf{E}_\Omega \cdot \mathbf{A}_\Omega^* \, \mathrm{d}\mathbf{x} \end{aligned}$$

and

$$\int_{\mathcal{D}} \Lambda^{-1}\nabla\times\mathbf{A}_\Omega\cdot\nabla\times\mathring{\mathbf{E}}(\boldsymbol{\theta})^* - k^2 n_\Omega^2\Lambda\mathbf{A}_\Omega\cdot\mathring{\mathbf{E}}(\boldsymbol{\theta})^*\,\mathrm{d}\mathbf{x} - \int_{\Gamma_{\text{in}}}\gamma_{\text{in}}(\mathbf{A}_\Omega)\cdot\mathbf{n}\times\mathring{\mathbf{E}}(\boldsymbol{\theta})^*\times\mathbf{n}\,\mathrm{d}s$$
$$-\int_{\Gamma_{\text{out}}}\gamma_{\text{out}}(\mathbf{A}_\Omega)\cdot\mathbf{n}\times\mathring{\mathbf{E}}(\boldsymbol{\theta})^*\times\mathbf{n}\,\mathrm{d}s = -2i\omega\mu_0\left(\int_{\Gamma_{\text{out}}}\mathring{\mathbf{E}}(\boldsymbol{\theta})\times\boldsymbol{\mathcal{H}}_n^{\text{out},*}\cdot\mathbf{n}\,\mathrm{d}s\right)^*.$$

Which gives the equality:

$$\int_{\Gamma_{\text{out}}}\mathring{\mathbf{E}}(\boldsymbol{\theta})\times\boldsymbol{\mathcal{H}}_n^{\text{out},*}\cdot\mathbf{n}\,\mathrm{d}s = \frac{-1}{2i\omega\mu_0}\int_{\mathcal{D}}\Lambda^{-1}A'(0)(\boldsymbol{\theta})\nabla\times\mathbf{E}_\Omega\cdot\nabla\times\mathbf{A}_\Omega^*$$
$$-k^2n_\Omega^2\Lambda C'(0)(\boldsymbol{\theta})\mathbf{E}_\Omega\cdot\mathbf{A}_\Omega^*\,\mathrm{d}\mathbf{x}.$$

The shape derivative (III.2.8) is therefore equal to:

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \mathrm{Re}\left[\frac{S_{m,n}(\Omega)^*}{2i\omega\mu_0}\int_{\mathcal{D}}\Lambda^{-1}A'(0)(\boldsymbol{\theta})\nabla\times\mathbf{E}_\Omega\cdot\nabla\times\mathbf{A}_\Omega^*\right.$$
$$\left.-k^2n_\Omega^2\Lambda C'(0)(\boldsymbol{\theta})\mathbf{E}_\Omega\cdot\mathbf{A}_\Omega^*\,\mathrm{d}\mathbf{x}\right]. \quad \text{(III.2.10)}$$

4. Surfacic shape derivative

In order to find the surfacic expression we use the following integration by parts formulas valid for smooth enough domain $\mathcal{D}$ and vector fields $\boldsymbol{\theta},\boldsymbol{\psi}$ and $\boldsymbol{\phi}$:

$$\int_{\mathcal{D}}(\nabla\cdot\boldsymbol{\theta})\,\boldsymbol{\psi}\cdot\boldsymbol{\phi}\,\mathrm{d}\mathbf{x} = \int_{\partial\mathcal{D}}\boldsymbol{\psi}\cdot\boldsymbol{\phi}\,(\boldsymbol{\theta}\cdot\mathbf{n})\,\mathrm{d}s - \int_{\mathcal{D}}\nabla(\boldsymbol{\psi}\cdot\boldsymbol{\phi})\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x},$$

$$\int_{\mathcal{D}}(\nabla\boldsymbol{\theta}\boldsymbol{\psi})\cdot\boldsymbol{\phi}\,\mathrm{d}\mathbf{x} = \int_{\partial\mathcal{D}}(\boldsymbol{\theta}\cdot\boldsymbol{\phi})\,(\boldsymbol{\psi}\cdot\mathbf{n})\,\mathrm{d}s - \int_{\mathcal{D}}(\nabla\cdot\boldsymbol{\psi})\,\boldsymbol{\theta}\cdot\boldsymbol{\phi}\,\mathrm{d}\mathbf{x} - \int_{\mathcal{D}}(\nabla\boldsymbol{\phi}\boldsymbol{\psi})\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x},$$

$$\int_{\mathcal{D}}(\nabla\boldsymbol{\theta}^\top\boldsymbol{\psi})\cdot\boldsymbol{\phi}\,\mathrm{d}\mathbf{x} = \int_{\partial\mathcal{D}}(\boldsymbol{\theta}\cdot\boldsymbol{\psi})\,(\boldsymbol{\phi}\cdot\mathbf{n})\,\mathrm{d}s - \int_{\mathcal{D}}(\nabla\cdot\boldsymbol{\phi})\,\boldsymbol{\theta}\cdot\boldsymbol{\psi}\,\mathrm{d}\mathbf{x} - \int_{\mathcal{D}}(\nabla\boldsymbol{\psi}\boldsymbol{\phi})\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}.$$

However, the second integrand in Eq. (III.2.10) is discontinuous across the border $\partial\Omega$ meaning that it is not smooth enough to apply one of the previous integration by part formula. We then decompose Eq. (III.2.10) into

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \mathrm{Re}\left[\frac{S_{m,n}(\Omega)^*}{2i\omega\mu_0}\,(I_1 - I_2 - I_3)\right],$$

so that

$$I_1 = \int_{\mathcal{D}}\Lambda^{-1}A'(0)(\boldsymbol{\theta})\nabla\times\mathbf{E}_\Omega\cdot\nabla\times\mathbf{A}_\Omega^*\,\mathrm{d}\mathbf{x},$$

$$I_2 = \int_\Omega k^2n_{\text{core}}^2\Lambda C'(0)(\boldsymbol{\theta})\mathbf{E}_{\Omega,1}\cdot\mathbf{A}_{\Omega,1}^*\,\mathrm{d}\mathbf{x} \quad\text{and}\quad I_3 = \int_{\mathcal{D}\setminus\bar{\Omega}}k^2n_{\text{clad}}^2\Lambda C'(0)(\boldsymbol{\theta})\mathbf{E}_{\Omega,2}\cdot\mathbf{A}_{\Omega,2}^*\,\mathrm{d}\mathbf{x},$$

in which $\mathbf{E}_{\Omega,1},\mathbf{A}_{\Omega,1}$ (resp. $\mathbf{E}_{\Omega,2},\mathbf{A}_{\Omega,2}$) are restrictions of $\mathbf{E}_\Omega,\mathbf{A}_\Omega$ on Ω (resp. $\mathcal{D}\setminus\bar{\Omega}$), that is areas of constant optical indices in which $\mathbf{E}_\Omega,\mathbf{A}_\Omega$ are smooth (at least in H^1 as explained in Remark III.2.1.1). Since $\boldsymbol{\theta}=0$ on $\partial\mathcal{D}$ and $\Lambda=\mathrm{Id}$ on $\partial\Omega$ we find that

$$I_1 = \int_{\mathcal{D}}\boldsymbol{\theta}\cdot r_{1,\Omega}\,\mathrm{d}\mathbf{x},$$

$$I_2 = k^2n_{\text{core}}^2\int_{\partial\Omega}(\mathbf{E}_{\Omega,1}\cdot\mathbf{A}_{\Omega,1}^*)\boldsymbol{\theta}\cdot\mathbf{n} + (\mathbf{E}_{\Omega,1}\cdot\mathbf{n})\mathbf{A}_{\Omega,1}^*\cdot\boldsymbol{\theta} + (\mathbf{A}_{\Omega,1}^*\cdot\mathbf{n})\mathbf{E}_{\Omega,1}\cdot\boldsymbol{\theta}\,\mathrm{d}s + \int_{\mathcal{D}}\boldsymbol{\theta}\cdot r_{2,\Omega}\,\mathrm{d}\mathbf{x},$$

$$I_3 = -k^2 n_{\text{clad}}^2 \int_{\partial\Omega} (\mathbf{E}_{\Omega,2}\cdot\mathbf{A}^*_{\Omega,2})\boldsymbol{\theta}\cdot\mathbf{n} + (\mathbf{E}_{\Omega,1}\cdot\mathbf{n})(\mathbf{A}^*_{\Omega,1}\cdot\boldsymbol{\theta}) + (\mathbf{A}^*_{\Omega,1}\cdot\mathbf{n})(\mathbf{E}_{\Omega,1}\cdot\boldsymbol{\theta})\,\mathrm{d}s + \int_{\mathcal{D}} \boldsymbol{\theta}\cdot r_{3,\Omega}\,\mathrm{d}\mathbf{x},$$

where $r_{i,\Omega}$ depends on $\mathbf{E}_\Omega, \mathbf{A}_\Omega$. Now using the continuity of the tangential components of $\mathbf{E}_\Omega, \mathbf{A}_\Omega$ and that of the normal components $n_\Omega^2\mathbf{E}_\Omega\cdot\mathbf{n}$, $n_\Omega^2\mathbf{A}_\Omega\cdot\mathbf{n}$ we can rewrite the last two integrals into

$$I_2 + I_3 = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n}\, k^2 \Big((n_{\text{core}}^2 - n_{\text{clad}}^2)(\mathbf{n}\times\mathbf{E}_\Omega\times\mathbf{n})\cdot(\mathbf{n}\times\mathbf{A}^*_\Omega\times\mathbf{n}) \\ -(n_{\text{core}}^{-2} - n_{\text{clad}}^{-2})(n_\Omega^2\mathbf{E}_\Omega\cdot\mathbf{n})(n_\Omega^2\mathbf{A}^*_\Omega\cdot\mathbf{n})\Big)\,\mathrm{d}s + \int_{\mathcal{D}} \boldsymbol{\theta}\cdot r_{4,\Omega}\,\mathrm{d}\mathbf{x}.$$

Canceling the terms of the form $\int_{\mathcal{D}} \boldsymbol{\theta}\cdot r_\Omega\,\mathrm{d}\mathbf{x}$ (see Section II.2.2 and Th. II.1.1.3) we find that the shape gradient is equal to

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n}\, k^2 \mathrm{Re}\left[\frac{S_{m,n}(\mathbf{E}_\Omega)^*}{2i\omega\mu_0} \Big((n_{\text{clad}}^2 - n_{\text{core}}^2)\mathbf{n}\times\mathbf{E}_\Omega\times\mathbf{n}\cdot\mathbf{n}\times\mathbf{A}^*_\Omega\times\mathbf{n} \right. \\ \left. -(n_{\text{clad}}^{-2} - n_{\text{core}}^{-2})(n_\Omega^2\mathbf{E}_\Omega)\cdot\mathbf{n}\,(n_\Omega^2\mathbf{A}^*_\Omega)\cdot\mathbf{n}\Big)\right]\mathrm{d}s, \quad \text{(III.2.11)}$$

concluding the proof. □

Remark III.2.1.2: One comment is about the units in Eq. (III.2.11). Since $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ gives the variation of $\mathcal{J}(\Omega)$ according to a small perturbation on $\boldsymbol{\theta}$, it must be expressed in W. This remark is useful since it gives a physical meaning to the previous quantity as well as giving a necessary condition on the obtained formula to ensure its correct calculation (the obtained formula for $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ must also be in W).

First, since the modes are normalized in Section I.2.2 we have $[S_{m,n}] = 1$ (we use the notation $[x]$ for the unit of x and 1 refer to an adimensional quantity). Secondly, by definition, $[k^2/(\omega\mu_0)] = \mathrm{S/m}$. Lastly, both $\mathbf{E}_\Omega$ and $\mathbf{A}_\Omega$ are electrical fields and so $[\mathbf{E}_\Omega] = [\mathbf{A}_\Omega] = \mathrm{V/m}$. Combining these results together we find that

$$[V_\Omega] = [\mathrm{S/m}][\mathrm{V/m}]^2 = \mathrm{W/m^3}.$$

In other word, the scalar field V_Ω can be interpreted as the local change of power transmission when the optical index is modified in a small volume. The integration on the surface $\partial\Omega$ and the multiplication by the perturbation $\boldsymbol{\theta}$ (which is a distance and therefore expressed in m) then leads to $[\mathcal{J}'(\Omega)(\boldsymbol{\theta})] = [\mathrm{m^2}][\mathrm{m}][\mathrm{W/m^3}] = \mathrm{W}$.

Remark III.2.1.3: In Th. III.2.1.1 the electric field $\mathbf{E}_\Omega$ is solution to a PDE using the Dirichlet-to-Neumann boundary condition defined in Section I.3.2.c for both the input and output waveguide surfaces. This means in particular that $\mathbf{E}_\Omega$ is only composed of backward propagating modes on Γ_{out}.

If, instead of using this boundary condition, we rather consider a PML after Γ_{out} as proposed in Remark I.3.3.1 we may have some spurious forward propagating modes in the decomposition of $\mathbf{E}_\Omega$ on Γ_{out}. This implies that the scattering parameter in Eq. (III.1.1) is now equal to Eq. (I.2.12) and that the adjoint state must be modified accordingly. This was our situation in [Leb19a] and we refer to our paper for details on the expression of the adjoint which, in this case, cannot be exactly written as the injection of

the output mode into Γ_{out}. The result proposed in this section is therefore possible to implement in commercial photonic simulation software since most of them only act as black-boxes which only provide a way to simulate the electric field when injecting a mode into a waveguide. A more general boundary condition like the one in our paper is impossible to implement in these software.

III.2.1.b Invariance of the shape in the etching direction

The result of Th. III.2.1.1 allows to optimize nanophotonic components but it lacks one important feature; even though the device's design $\Omega \subset \mathcal{D}_{\text{opt}}$ is three-dimensional, considering that it is manufactured through an etching process (explained in Section IV.3.1.b) it must be invariant in the etching direction (y-axis).

To ensure this constraint, the idea is to consider an initial design $\Omega_0 \subset \mathcal{D}_{\text{opt}}$ which is y-invariant, and deformation fields $\boldsymbol{\theta}$ in the method of Hadamard which are y-invariant. Indeed, let Ω be defined as

$$\Omega = \{(x, y, z), (x, z) \in \widehat{\Omega}, y \in [-h/2, h/2]\}$$

for some two dimensional shape $\widehat{\Omega} \subset \widehat{\mathcal{D}}_{\text{opt}}$ where $\widehat{\mathcal{D}}_{\text{opt}} \subset \mathbb{R}^2$ is a section of the (y-invariant) optimization domain $\mathcal{D}$. If $\boldsymbol{\theta}(x, y, z) = \boldsymbol{\theta}(x, z)$ is independent from the y-coordinate then Eq. (III.2.1) simplifies into

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\widehat{\Omega}} \boldsymbol{\theta}(x, z) \cdot \mathbf{n}(x, z) \left(\int_{-\frac{h}{2}}^{\frac{h}{2}} V_\Omega(x, y, z) \, \mathrm{d}y \right) \mathrm{ds}(x,z),$$

in which we see that an ascent direction supplied by the method of Hadamard which is y-invariant is given by

$$\boldsymbol{\theta}(x, z) = \mathbf{n}(x, z) \int_{-\frac{h}{2}}^{\frac{h}{2}} V_\Omega(x, y, z) \, \mathrm{d}y. \tag{III.2.12}$$

III.2.1.c Application of Céa's method

We propose to find here the result of Th. III.2.1.1 by using Céa's method. As pointed out in [Pan05] in the case of the Laplace equation, the 2-phases setting contains some difficulties which result in an erroneous shape derivative. In this subsection, we start with the formal application of Céa's method which ends up with a wrong result while the second part of this subsection shows how to modify Céa's method to recover, in theory, the correct shape derivative.

Formal application of Céa's method

If we apply Céa's method as presented in Section II.2.1.a without extra attention, we end up with the correct adjoint state but a wrong shape derivative. Indeed, let us consider the following Lagrangian

$$\mathcal{L}(\Omega, \mathbf{E}, \boldsymbol{\phi}) = \mathcal{J}(\Omega) + a_\Omega(\mathbf{E}, \boldsymbol{\phi}) - b(\boldsymbol{\phi}) \tag{III.2.13}$$

where $a(\mathbf{E}, \boldsymbol{\phi}) = b(\boldsymbol{\phi})$ is the variational formulation of Eq. (III.1.2) (defined in Eq. (III.2.4) with $\boldsymbol{\theta} = 0$). Using Th. II.1.1.2, we can infer the adjoint state $\mathbf{A}_\Omega$ equation by canceling

the derivative of Eq. (III.2.13) with respect to $\mathbf{E}$ and then find the shape derivative of $\mathcal{J}$ as the partial derivative of $\mathcal{L}$ with respect to the explicit dependence on Ω at $(\Omega, \mathbf{E}_\Omega, \mathbf{A}_\Omega)$. Since the Lagrangian may read as

$$\mathcal{L}(\Omega, \mathbf{E}, \boldsymbol{\phi}) = -k^2 \int_{\mathcal{D}} n_\Omega^2 \mathbf{E} \cdot \boldsymbol{\phi}^* \, \mathrm{d}\mathbf{x} + R(\mathbf{E}, \boldsymbol{\phi}),$$

where $R(\mathbf{E}, \boldsymbol{\phi})$ does not depends on Ω, we can separate the integral on both Ω and $\mathcal{D}\backslash\bar{\Omega}$ and end up with

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} \, k^2 (n_{\text{clad}}^2 \mathbf{E}_{\Omega,2} \cdot \mathbf{A}_{\Omega,2}^* - n_{\text{core}}^2 \mathbf{E}_{\Omega,1} \cdot \mathbf{A}_{\Omega,1}^*) \, \mathrm{d}s. \tag{III.2.14}$$

In Eq. (III.2.14), $\mathbf{E}_{\Omega,1}, \mathbf{A}_{\Omega,1}$ (resp. $\mathbf{E}_{\Omega,2}, \mathbf{A}_{\Omega,2}$) are the (smooth; see Remark III.2.1.1) restrictions of $\mathbf{E}_\Omega$ and $\mathbf{A}_\Omega$ on Ω (resp. $\mathcal{D}\backslash\bar{\Omega}$). Equation (III.2.14) rewrites as

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} \, k^2 (n_{\text{clad}}^2 (\mathbf{n} \times \mathbf{E}_{\Omega,2} \times \mathbf{n} \cdot \mathbf{n} \times \mathbf{A}_{\Omega,2}^* \times \mathbf{n} + \mathbf{E}_{\Omega,2} \cdot \mathbf{n} \mathbf{A}_{\Omega,2}^* \cdot \mathbf{n}) \\ - n_{\text{core}}^2 (\mathbf{n} \times \mathbf{E}_{\Omega,1} \times \mathbf{n} \cdot \mathbf{n} \times \mathbf{A}_{\Omega,1}^* \times \mathbf{n} + \mathbf{E}_{\Omega,1} \cdot \mathbf{n} \mathbf{A}_{\Omega,1}^* \cdot \mathbf{n})) \, \mathrm{d}s.\end{aligned}$$

Now using the continuity conditions of the electric field

$$\mathbf{n} \times \mathbf{E}_{\Omega,1} \times \mathbf{n} = \mathbf{n} \times \mathbf{E}_{\Omega,2} \times \mathbf{n} = \mathbf{n} \times \mathbf{E}_\Omega \times \mathbf{n} \quad \text{and} \quad n_{\text{core}}^2 \mathbf{E}_{\Omega,1} \cdot \mathbf{n} = n_{\text{clad}}^2 \mathbf{E}_{\Omega,2} \cdot \mathbf{n} = n_\Omega^2 \mathbf{E}_\Omega \cdot \mathbf{n},$$

the shape derivative is equal to

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} \, k^2 \left[(n_{\text{clad}}^2 - n_{\text{core}}^2) \mathbf{n} \times \mathbf{E}_\Omega \times \mathbf{n} \cdot \mathbf{n} \times \mathbf{A}_\Omega^* \times \mathbf{n} \right. \\ \left. + (n_{\text{clad}}^{-2} - n_{\text{core}}^{-2}) (n_\Omega^2 \mathbf{E}_\Omega \cdot \mathbf{n}) (n_\Omega^2 \mathbf{A}_\Omega^* \cdot \mathbf{n}) \right] \, \mathrm{d}s.\end{aligned} \tag{III.2.15}$$

The main difference between this expression and Eq. (III.2.11) is that the second term (the one involving the normal component of the fields) has opposite sign ! The error in the previous calculation comes from the fact that Céa's method assumes the existence of the Eulerian derivative of the electric field $\Omega \mapsto \mathbf{E}_\Omega$, which is not the case here due to the discontinuity of the normal component of $\mathbf{E}$; see Section II.2.1.a in which this problem was also explained.

It is also worth noting that the real part and the factor $S_{m,n}(\mathbf{E}_\Omega)^*/(2i\omega\mu_0)$ are absent from Eq. (III.2.15). This is due to the fact that all these information are contained here in the adjoint. Indeed, in Section II.2.1.a, we defined the adjoint $\mathbf{A}_\Omega$ as the solution of $\partial_{\mathbf{E}} \mathcal{L}(\Omega, \mathbf{E}_\Omega, \mathbf{A}_\Omega)(\tilde{\mathbf{E}}) = 0$ for all $\tilde{\mathbf{E}}$. Here, since $\mathcal{L}$ is not holomorphic with respect to $\mathbf{E}$ (due to the fact that $\mathcal{J}(\Omega)$ is real-valued), we need to consider separately both the real and imaginary part of $\mathbf{E}$. That is, if $\mathbf{E} = \mathbf{E}_{\text{re}} + i\mathbf{E}_{\text{im}}$, to define $\mathbf{A} = \mathbf{A}_{\text{re}} + i\mathbf{A}_{\text{im}}$ as the solution of both

$$\partial_{\mathbf{E}_{\text{re}}} \mathcal{L}(\Omega, \mathbf{E}_\Omega, \mathbf{A}_{\Omega,\text{re}}) = 0 \quad \text{and} \quad \partial_{\mathbf{E}_{\text{im}}} \mathcal{L}(\Omega, \mathbf{E}_\Omega, \mathbf{A}_{\Omega,\text{im}}) = 0,$$

using the fact that a_Ω is sesquilinear and $\mathcal{J}$ real-valued we find that for all $\tilde{\mathbf{E}}_{\text{re}}$ and $\tilde{\mathbf{E}}_{\text{im}}$:

$$a_\Omega(\tilde{\mathbf{E}}_{\text{re}}, \mathbf{A}_{\text{re}}) = -\partial_{\mathbf{E}_{\text{re}}} \mathcal{J}(\Omega)(\tilde{\mathbf{E}}_{\text{re}}), \quad a_\Omega(\tilde{\mathbf{E}}_{\text{re}}, \mathbf{A}_{\text{im}}) = 0 \tag{III.2.16}$$

$$a_\Omega(\tilde{\mathbf{E}}_{\text{im}}, \mathbf{A}_{\text{im}}) = -\partial_{\mathbf{E}_{\text{im}}} \mathcal{J}(\Omega)(\tilde{\mathbf{E}}_{\text{im}}), \quad a_\Omega(\tilde{\mathbf{E}}_{\text{im}}, \mathbf{A}_{\text{re}}) = 0. \tag{III.2.17}$$

For the objective function of Eq. (III.1.1), assuming without loss of generality that the objective mode $\boldsymbol{\mathcal{H}}^*_{-n} = \boldsymbol{\mathcal{H}}$ is real, we have

$$\partial_{\mathbf{E}_{\text{re}}}\mathcal{J}(\Omega)(\tilde{\mathbf{E}}_{\text{re}}) = \frac{1}{2}\text{Re}\left[S_{m,n}(\mathbf{E}_\Omega)\int_{\Gamma_{\text{out}}}\left[\tilde{\mathbf{E}}_{\text{re}}\times\boldsymbol{\mathcal{H}}\right]\cdot\mathbf{n}\,\mathrm{d}s\right],$$
$$\partial_{\mathbf{E}_{\text{im}}}\mathcal{J}(\Omega)(\tilde{\mathbf{E}}_{\text{im}}) = \frac{1}{2}\text{Im}\left[S_{m,n}(\mathbf{E}_\Omega)\int_{\Gamma_{\text{out}}}\left[\tilde{\mathbf{E}}_{\text{im}}\times\boldsymbol{\mathcal{H}}\right]\cdot\mathbf{n}\,\mathrm{d}s\right].$$

This allows us to summarize Eqs. (III.2.16) and (III.2.17) into

$$a_\Omega(\tilde{\mathbf{E}}_{\text{re}},\mathbf{A}_{\text{re}}) + a_\Omega(\tilde{\mathbf{E}}_{\text{im}},\mathbf{A}_{\text{im}}) = -\text{Re}\left[\frac{1}{2}S^*_{m,n}(\mathbf{E}_\Omega)\int_{\Gamma_{\text{out}}}\left[\tilde{\mathbf{E}}\times\boldsymbol{\mathcal{H}}\right]\cdot\mathbf{n}\,\mathrm{d}s\right],$$

where the missing real part and factor are present.

A (potential) correct calculation using Céa's method with two phases

In [Pan05] a method to find the correct shape derivative using Céa's method despite the non-existence of the Eulerian derivative of $\mathbf{E}_\Omega$ is explained. We have not been able to reproduce this method for our objective function and PDE but we still decided to provide some details in case someone is interested in doing this calculation.

In the previously mentioned paper, the author proposed to separately consider $\mathbf{E}_{\Omega,1}$ and $\mathbf{E}_{\Omega,2}$, the solutions of the time-harmonic vector wave equation inside Ω and $\mathcal{D}\backslash\bar{\Omega}$ with the addition of a Lagrange multiplier to link these two solutions to one another. In mathematical terms, the Lagrangian in Eq. (III.2.13) is modified into

$$\begin{aligned}\mathcal{L}(\Omega,\mathbf{E}_1,\boldsymbol{\phi}_1,\mathbf{E}_2,\boldsymbol{\phi}_2,\lambda,\mu) &= \mathcal{J}(\Omega) + a_{\Omega,1}(\mathbf{E}_1,\boldsymbol{\phi}_1) - b_1(\boldsymbol{\phi}_1) + a_{\Omega,2}(\mathbf{E}_2,\boldsymbol{\phi}_2) - b_2(\boldsymbol{\phi}_2)\\ &+ \int_{\partial\Omega}\lambda\cdot(\mathbf{n}\times\mathbf{E}_1\times\mathbf{n} - \mathbf{n}\times\mathbf{E}_2\times\mathbf{n})\,\mathrm{d}s + \int_{\partial\Omega}\mu\cdot(\mathbf{n}\times\nabla\times\mathbf{E}_1 - \mathbf{n}\times\nabla\times\mathbf{E}_2)\,\mathrm{d}s,\end{aligned} \tag{III.2.18}$$

with a_1, b_1, a_2, b_2 the sesquilinear and antilinear form of associated with the variational formulation of the time-harmonic vector wave equation in Ω and $\mathcal{D}\backslash\bar{\Omega}$. The two additional Lagrangian multiplier λ and μ account for the interface conditions of $\mathbf{E}$ on $\partial\Omega$. Since $\mathbf{E}_1$ and $\mathbf{E}_2$ are smooth on their respective domains of definitions (Remark III.2.1.1), the Eulerian derivatives of each of these functions are defined (through the relation Eq. (II.1.14)) and Céa's method can be carried out.

Canceling the partial derivatives of Eq. (III.2.18) with respect to $\mathbf{E}_1$ and $\mathbf{E}_2$ are equal to $\mathbf{A}$ defined in Eq. (III.2.3); $\mathbf{A}_1, \mathbf{A}_2$ are the restrictions of $\mathbf{A}$ into Ω and $\mathcal{D}\backslash\bar{\Omega}$. We also find the expressions of λ, μ as

$$\lambda = \mathbf{n}\times\nabla\times\boldsymbol{\phi}_1 = -\mathbf{n}\times\nabla\times\boldsymbol{\phi}_2 \text{ and } \mu = \mathbf{n}\times\mathbf{E}_1\times\mathbf{n} = -\mathbf{n}\times\mathbf{E}_2\times\mathbf{n}.$$

Using these values, Eq. (III.2.18) may be decomposed into

$$\begin{aligned}\mathcal{L}(\Omega,\mathbf{E}_1,\boldsymbol{\phi}_1,\mathbf{E}_2,\boldsymbol{\phi}_2) &= R(\mathbf{E},\boldsymbol{\phi}) - \int_{\mathcal{D}} k^2 n_\Omega^2\mathbf{E}\cdot\boldsymbol{\phi}^*\,\mathrm{d}\mathbf{x}\\ &- 2\int_{\partial\Omega}\mathbf{n}\times\nabla\times\mathbf{E}_1\cdot\mathbf{n}\times\boldsymbol{\phi}_1^*\times\mathbf{n} - \mathbf{n}\times\nabla\times\mathbf{E}_2\cdot\mathbf{n}\times\boldsymbol{\phi}_2^*\times\mathbf{n}\,\mathrm{d}s.\end{aligned} \tag{III.2.19}$$

A tedious differentiation of Eq. (III.2.19) with respect to $\boldsymbol{\theta}$ should give, this time, the correct shape derivative. Although we have not succeeded in achieving this calculation, we want to make two precision:

- Derivating Eq. (III.2.19) with respect to $\boldsymbol{\theta}$ brings into play terms of the form

$$\int_{\partial\Omega} (\boldsymbol{\theta}\cdot\mathbf{n})(\mathbf{n}\cdot\nabla)(\mathbf{n}\times\nabla\times\mathbf{E})\mathbf{n}\times\boldsymbol{\phi}^*\times\mathbf{n}\,\mathrm{d}s,$$

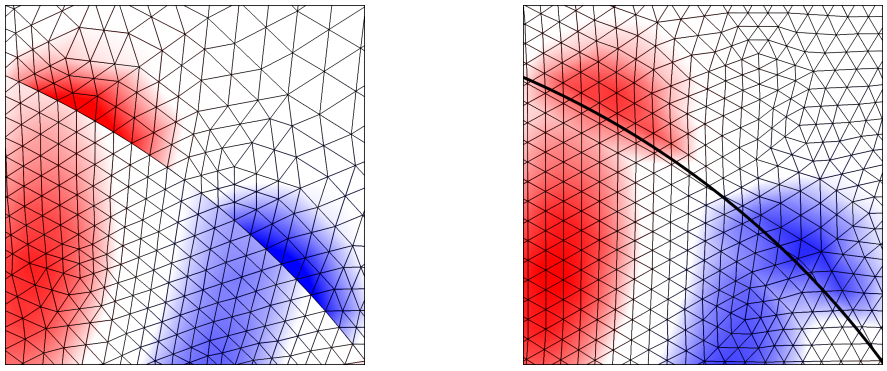
  where the term $(\mathbf{n}\cdot\nabla)(\mathbf{n}\times\nabla\times\mathbf{E})$ is close to $\mathbf{n}\times\nabla\times\nabla\times\mathbf{E}\times\mathbf{n} = k^2n^2\mathbf{n}\times\mathbf{E}\times\mathbf{n}$.

- In integrals of the form $\int_{\partial\Omega} f(\mathbf{n})$, the normal vector $\mathbf{n}$ depends on the shape Ω and therefore the derivative of this integral is found using Th. II.1.2.1 and the Eulerian derivative of the normal vector which is $\mathbf{n}'(\boldsymbol{\theta}) = -\nabla_{\partial\Omega}(\boldsymbol{\theta}\cdot\mathbf{n})$ (see [Hen06, Proposition 5.4.14])

III.2.2 Alleviating numerical instabilities using refractive index smoothing

III.2.2.a Isotropic smoothing

Since the shape derivative Eq. (III.2.1) depends on the evaluation of the electric field's tangential components and normal component $(n_\Omega^2\mathbf{E}_\Omega)\cdot\mathbf{n}$ at the interface $\partial\Omega$, we theoretically need a precise mesh of the shape at this discontinuity. Indeed, as can be seen in Fig. III.2.1, if the nodes of the mesh do not coincide exactly with the border of the shape, then, by definition of the finite element basis functions (for FDTD see Remark III.2.2.1), the electric field is continuous on these tetrahedrons crossing the interface. This means that, when evaluating the quantity $(n_\Omega^2\mathbf{E}_\Omega)\cdot\mathbf{n}$, we want to have a continuous value by computing the product between the two discontinuous ones n_Ω and $\mathbf{E}_\Omega\cdot\mathbf{n}$ while numerically $\mathbf{E}_\Omega\cdot\mathbf{n}$ is continuous.



(a) *Explicit mesh*

(b) *Implicit mesh. The black line indicates the interface between the two mediums*

Figure III.2.1: E_y component of the electric field using either an explicit or an implicit mesh of the interface. In the case of an implicit meshing, the numerical integration scheme on tetrahedrons crossing the interface leads to continuous values of the electric field's normal component.

Although it may be possible to use remeshing algorithms on the moving interface in the case of the finite element method to always have nodes of the mesh on this boundary, we prefer to rely on an **index-smoothing** (also known as sub-pixel smoothing) method which approximates the discontinuous optical index n_Ω by a smoothed version.

Remark III.2.2.1: This method also has the advantage of being useful in the case of finite difference simulations, like FDTD, which are inherently defined on Cartesian grids and therefore never give the possibility to accurately evaluate Eq. (III.2.1).

To explain this method, let us first define a sequence of mollifiers s_ε such that

$$s_\varepsilon(\mathbf{x}) = \varepsilon^{-2} s(\varepsilon^{-1}\mathbf{x}) \text{ with } s \in C_c^\infty(\mathbb{R}^2),\ \operatorname{supp}(s) \subset]-1,1[^2 \text{ and } \int_{\mathbb{R}^2} s(\mathbf{x})\mathrm{d}\mathbf{x} = 1. \quad \text{(III.2.20)}$$

We define the smoothed approximation $n_{\Omega,\varepsilon}$ of n_Ω by:

$$n^2_{\Omega,\varepsilon}(\cdot, y, \cdot) = s_\varepsilon * (n^2_\Omega(\cdot, y, \cdot)), \quad n^2_{\Omega,\varepsilon}(x, y, z) = \int_{\mathbb{R}}\int_{\mathbb{R}} s_\varepsilon(x-t, y-u) n^2_\Omega(t, y, u)\,\mathrm{d}t\,\mathrm{d}u. \quad \text{(III.2.21)}$$

The unique solution of the time-harmonic vector wave equation using the optical index $n_{\Omega,\varepsilon}$ is denoted by $\mathbf{E}_{\Omega,\varepsilon}$ and the smoothed objective function as $\mathcal{J}_\varepsilon(\Omega) = |S_{m,n}(\mathbf{E}_{\Omega,\varepsilon})|^2$. In this case, since $n^2_{\Omega,\varepsilon} \in C^\infty(\mathcal{D}, \mathbb{R}^*_+)$, the electric field $\mathbf{E}_{\Omega,\varepsilon}$ is at least of class $C^1(\mathcal{D})$ (see [Dau12, Chapter IX]) and the shape derivative associated to $\mathcal{J}_\varepsilon(\Omega)$ may be found using Céa's method:

$$\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n}\, k^2 \int_{-\frac{h}{2}}^{\frac{h}{2}} \operatorname{Re}\left[\frac{S_{m,n}(\mathbf{E}_{\Omega,\varepsilon})^*}{2i\omega\mu_0}(n^2_{\text{clad}} - n^2_{\text{core}}) s_\varepsilon * \left(\mathbf{E}_{\Omega,\varepsilon}\cdot\mathbf{A}^*_{\Omega,\varepsilon}\right)\right] \mathrm{d}y\,\mathrm{d}s, \quad \text{(III.2.22)}$$

where $\mathbf{A}_{\Omega,\varepsilon}$ is solution of Eq. (III.2.3) using the optical index $n_{\Omega,\varepsilon}$.

In summary, to obtain this shape derivative we need to convolve the optical index and solve for both electric fields ($\mathbf{E}_{\Omega,\varepsilon}$ and $\mathbf{A}^*_{\Omega,\varepsilon}$) and then convolve the scalar product between these two quantities. Equation (III.2.22) is much simpler to compute than Eq. (III.2.1) since it does not involve any discontinuous quantity across $\partial\Omega$.

Remark III.2.2.2: In [All14b] or [Ver19b] the authors also proposed to regularize the index (the conductivity of the Laplace equation in the original papers) into that $n_{\Omega,\varepsilon}$ defined by:

$$n^2_{\Omega,\varepsilon} = n^2_{\text{clad}} + (n^2_{\text{core}} - n^2_{\text{clad}})\varepsilon^{-1} h(\varepsilon^{-1} d_\Omega) \quad \text{(III.2.23)}$$

where h is a smooth, non-decreasing function such that $h(x) = 0$ if $x \leq -1$, $h(x) = 1$ if $x \geq 1$ and d_Ω the signed distance function defined in Section II.3.3.b. This additional dependency of $n^2_{\Omega,\varepsilon}$ on Ω in Eq. (III.2.23) through the signed distance function d_Ω leads to a different shape derivative which need to evaluate the electric field along rays; see [All14b, Section 3.2] for the definitions.

We conjecture that the following error estimate concerning the smoothed field is verified for all $p > 1$ if $\mathbf{E}_\Omega$ is sufficiently smooth:

$$\|\mathbf{E}_{\Omega,\varepsilon} - \mathbf{E}_\Omega\|_{(L^2(\mathcal{D},\mathbb{C}))^3} \underset{\varepsilon\to 0}{=} o(\varepsilon^{1-\frac{1}{p}}), \quad \text{(III.2.24)}$$

meaning that the convergence of the regularized field to the real one is at least sub-linear.

One difficulty in proving this result notably comes from the fact that the bilinear form associated with the time-harmonic vector wave equation is not coercive. Note that to alleviate this problem it is common to have recourse to the so-called Helmhotz decomposition (see [Mon03, Lemma 4.5]) of the electric field which breaks down $\mathbf{E}_\Omega$ (resp. $\mathbf{E}_{\Omega,\varepsilon}$)

into the sum of a divergence free field $\mathbf{u}_\Omega$ (resp. $\mathbf{u}_{\Omega,\varepsilon}$) and a curl free one ∇v_Ω (resp. $v_{\Omega,\varepsilon}$) such that v_Ω is solution to a Laplace equation. Proving that $v_{\Omega,\varepsilon}$ converges to v_Ω with the same order of convergence as in Eq. (III.2.24) is then achieved by classical energy estimations; see for instance [Dap13, Section 4.8.1].

Using the identity Eq. (III.2.24) and considering that the mapping $\mathbf{E} \mapsto S_{m,n}(\mathbf{E})$ defined in Eq. (III.1.1) is continuous (from $(L^2(\mathcal{D},\mathbb{C}))^3$ into $\mathbb{C}$) then we have $\mathcal{J}_\varepsilon(\Omega)$ which tends to $\mathcal{J}(\Omega)$ with at least the same order of convergence as the prior estimate. This validates the fact that optimizing $\mathcal{J}_\varepsilon(\Omega)$ will give a shape Ω with approximately the same value for $\mathcal{J}(\Omega)$.

With more calculations it may be possible to show that we also have an error estimate between $\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta})$ and $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ and therefore that a local optima found for $\mathcal{J}_\varepsilon$ is close to a local minima of $\mathcal{J}$ when ε tends to zero.

III.2.2.b Anisotropic smoothing

Apart from the smoothing method proposed in Section III.2.2.a, a larger order of convergence (larger value in the power of ε in Eq. (III.2.24)) may be obtained by using another type of regularization as proposed for instance in [Kot08, Section VI]. The method described here was initially found through the following formal observations:

- If we want to numerically compute a discontinuous quantity $y = ax$ equal to the product of a continuous value x and a discontinuous one a then it is desirable to smooth a in order to make y also continuous and thus enhancing the stability of its computation.
- If now y is continuous but with both a and x discontinuous then smoothing a will still lead to a discontinuous y. Instead we can re-arrange the equation as $a^{-1}y = x$ which brings us back to the previous case and lead to smooth a^{-1} instead of a.

In Eq. (III.2.1) we need to compute the continuous value $n_\Omega^2 \mathbf{E}_\Omega \cdot \mathbf{n}_\Omega$ which is the product of two discontinuous functions, thus we should smooth n_Ω^{-2} for the normal component of $\mathbf{E}_\Omega$. When it comes to the computation of the tangential part $n_\Omega^2 \mathbf{n}_\Omega \times \mathbf{E}_\Omega \times \mathbf{n}_\Omega$, we rather need to smooth n_Ω^2 on the tangential components. This leads to defining the following smooth, anisotropic, optical index $N^2_{\Omega,\varepsilon}$

$$N^2_{\Omega,\varepsilon} = \left(s_\varepsilon * (n_\Omega^2)\right) I_3 + \left(\left(s_\varepsilon * (n_\Omega^{-2})\right)^{-1} - \left(s_\varepsilon * (n_\Omega^2)\right)\right) N_\Omega \tag{III.2.25}$$

where the matrix N_Ω is equal to $N_\Omega = \mathbf{n}^\top \mathbf{n}$. In [Kot08, Section VI] the authors claim that this method is of second order in ε.

Let us conclude by saying that we have observed that the smoothing method presented in this section is very efficient in practice for shape optimization of photonic devices to get a stable algorithm. Many numerical software, such as `RSoft Synopsys` or `Lumerical`, both using FDTD, even implement this kind of smoothing procedure by default; see [Han14, Section 4.2] for the implementation of this idea in the FDTD method and [Mic18] which considers this kind of smoothing in the shape optimization of nanophotonic devices.

Although very practical for the numerical stability of the optimization algorithm, it is essential to take care of the fact that the electric field obtained numerically through the

smoothing method may differ significantly from the real solution even with the result of Eq. (III.2.24). Typically, if the width of the smoothing is of the order of a mesh tetrahedron, it seems very unreliable to have isolated shapes of the same size. In practice, it is therefore recommended to avoid as much as possible this kind of small shapes. A simple way to check the correct behavior of the electric field is to refine the mesh at the end of the optimization process or to ensure that other simulation software produces the same result (see Fig. III.5.6 in which we compare our results using three different software).

III.2.3 Results for other objectives

III.2.3.a Total power

In the previous sections, as far as the objective function is considered, we were only interested in the power carried by a given mode. In some situations however, it may be interesting to look at the total power going through a surface. By definition of $S_{m,i}$, ignoring the radiative modes, the total power $\mathcal{P}$ transmitted through an output waveguide Γ_{out} is given by (Eq. (I.2.16)):

$$\mathcal{P}(\Omega) = \frac{1}{2}\int_{\Gamma_{\text{out}}} \text{Re}\left[\mathbf{E}_\Omega \times \mathbf{H}_\Omega^*\right]\cdot\mathbf{n}\,\text{d}s \simeq \sum_{i=1}^{N}|S_{m,i}(\Omega)|^2, \tag{III.2.26}$$

where $S_{m,i}$ is given by Eq. (III.1.1). Summing N times the result of Th. III.2.1.1, the shape derivative is given by:

$$\mathcal{P}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega}\boldsymbol{\theta}\cdot\mathbf{n}\sum_{i=1}^{N}V_i(s)\,\text{d}s, \tag{III.2.27}$$

where V_i is defined in Eq. (III.2.2) and requires both the values of $\mathbf{E}_\Omega$ and $\mathbf{A}_{\Omega,i}$ the adjoint state (Eq. (III.2.3)) corresponding to the injection of the mode $(\boldsymbol{\mathcal{E}}_i, \boldsymbol{\mathcal{H}}_i)$ in the output waveguide. This shape derivative is not satisfying because it requires as many simulations as there are modes in the decomposition Eq. (III.2.26). If we look at Eq. (III.2.27) we can see that the sum of the scalar fields V_i may be written as

$$\sum_{i=1}^{N}V_i(s) = k^2\text{Re}\left[\frac{1}{2i\omega\mu_0}\left(\left[\!\left[n^2\right]\!\right]\mathbf{E}_{\Omega,\parallel}\cdot\left(\sum_{i=1}^{N}S_{m,i}(\mathbf{E}_\Omega)\mathbf{A}_{\Omega,i,\parallel}\right)^*\right.\right.$$
$$\left.\left. - \left[\!\left[\frac{1}{n^2}\right]\!\right]\mathbf{E}_{\Omega,\perp}\left(\sum_{i=1}^{N}S_{m,i}(\mathbf{E}_\Omega)\mathbf{A}_{\Omega,i,\perp}\right)^*\right)\right]. \tag{III.2.28}$$

Using the linearity of the time-harmonic vector-wave equation, we can define

$$\mathbf{A}_\Omega = \sum_{i=1}^{N}S_{m,i}(\mathbf{E}_\Omega)\mathbf{A}_{\Omega,i,\perp},$$

such that $\mathbf{A}_\Omega$ is solution of

$$\left\{\begin{array}{rcll} \nabla\times\Lambda^{-1}\nabla\times\mathbf{A} - k^2n^2\Lambda\mathbf{A} & = & 0 & \text{in } \mathcal{D} \\ \mathbf{n}\times\mathbf{A} & = & 0 & \text{on } \partial\mathcal{D}\backslash(\Gamma_{\text{out}}\cup\Gamma_{\text{in}}) \\ \mathbf{n}\times\nabla\times\mathbf{A} + \gamma_{\text{out}}(\mathbf{A}) & = & 2i\omega\mu_0\hat{\mathbf{z}}\times\boldsymbol{\mathcal{H}}_{\text{tot}}^{\text{out}} & \text{on } \Gamma_{\text{out}} \\ \mathbf{n}\times\nabla\times\mathbf{A} + \gamma_{\text{in}}(\mathbf{A}) & = & 0 & \text{on } \Gamma_{\text{in}} \end{array}\right., \tag{III.2.29}$$

where $\boldsymbol{\mathcal{H}}_{\text{tot}}^{\text{out}}$ is defined as

$$\boldsymbol{\mathcal{H}}_{\text{tot}}^{\text{out}} = \sum_{i=1}^{N} S_{m,i}(\mathbf{E}_\Omega)\boldsymbol{\mathcal{H}}_i^{\text{out}}.$$

This method effectively reduces the computational complexity of Eq. (III.2.27) since only one adjoint state is needed, which corresponds to the simultaneous injection of all the output modes multiplied by their associated S-parameters.

Note that instead of relying on Th. III.2.1.1, it is possible to directly compute the shape derivative of Eq. (III.2.26) using the same demonstration as in the theorem but the expression of the adjoint found in this case is more complex than Eq. (III.2.29) since it contains information on both the guided and radiative modes (the last equality in Eq. (III.2.26) is very accurate in practice but is still a mathematical approximation).

III.2.3.b Mode volume

An other figure of merit, which is often considered in, for instance, cavity (see Section III.3.1.e), is the so-called mode volume $V_{\text{mode}}(\Omega)$. This figure of merit characterizes the uniform distribution of the electric energy inside a volume $V \subset \mathbb{R}^3$ of core material (that is to say, the confinement of light inside V) and it is defined as

$$V_{\text{mode}}(\Omega) = \frac{\displaystyle\int_V n_{\text{core}}^2 |\mathbf{E}_\Omega|^2 \,\mathrm{d}\mathbf{x}}{\displaystyle\max_V n_{\text{core}}^2 |\mathbf{E}_\Omega|^2}. \tag{III.2.30}$$

Equation (III.2.30) is often simplified by dropping the denominator leading to

$$E_{\text{mode}}(\Omega) = \int_V n_{\text{core}}^2 |\mathbf{E}_\Omega(\mathbf{x})|^2 \,\mathrm{d}\mathbf{x}, \tag{III.2.31}$$

which is the total electric energy inside the volume V. The shape derivative associated to Eq. (III.2.31) is given by Eq. (III.2.1) with $\Gamma_{\text{out}} = \emptyset$ and an adjoint $\mathbf{A}_\Omega$ solution of

$$\left\{ \begin{array}{rcll} \nabla \times \Lambda^{-1} \nabla \times \mathbf{A} - k^2 n_\Omega^2 \Lambda \mathbf{A} & = & -2n_{\text{core}}^2 \mathbf{E}_\Omega \mathbf{1}_V & \text{in } \mathcal{D} \\ \mathbf{n} \times \mathbf{A} & = & 0 & \text{on } \partial\mathcal{D} \backslash \Gamma_{\text{in}} \\ \mathbf{n} \times \nabla \times \mathbf{A} + \gamma_{\text{in}}(\mathbf{A}) & = & 0 & \text{on } \Gamma_{\text{in}} \end{array} \right. .$$

From a physical point of view, this is the same as saying that $\mathbf{A}_\Omega$ is the solution of the time-harmonic vector wave equation with an imposed current density $2n_{\text{core}}^2 \mathbf{E}_\Omega$ inside the volume V. An application using this shape derivative is given in Section III.3.1.e.

Note that the mode volume, as well as many other figures of merit in photonics, is physically related to resonant phenomena and thus can be studied through the eigenvalues of the time-harmonic vector wave equation.

III.3 Numerical examples

III.3.1 Classical components

In this section, we evaluate the efficiency of our shape and topology optimization Algorithm II.4.1 on the design of various nanophotonic devices.

In all cases, the design domain $\mathcal{D}_{\text{opt}}$ is $[-d_x, d_x] \times [-h/2, h/2] \times [-d_z, d_z]$ where d_x and d_z vary between 1.5 µm to 3 µm depending on the situation, and $h = 306$ nm (see Fig. I.3.2 for a reminder of the considered 3d geometry). The PML width is 500 nm. The tetrahedral mesh $\mathcal{T}$ associated to the finite element resolution is composed of about 10^5 elements, with size at most $\lambda/(5n_\Omega)$ where λ is the wavelength and n_Ω the optical index, as commonly advised in practice for such simulations. The 2d Cartesian grid $\mathcal{G}$ of $\mathcal{D}_{\text{opt}}$ dedicated to practice of the level set method has uniform size $\Delta x = 10$ nm.

The values of the wavelength λ considered in this section lie within the typical range used in telecommunication applications, that is around 1.31 µm and 1.55 µm. The values of the refractive indices of the involved materials (core of silicon, cladding of air and substrate of silica) are:

$$\begin{aligned} & n_{\text{core}} = 3.476, \quad n_{\text{subs}} = 1.444, \quad n_{\text{clad}} = 1 \quad \text{at} \quad \lambda = 1.55\,\mu\text{m} \\ \text{and} \quad & n_{\text{core}} = 3.506, \quad n_{\text{subs}} = 1.447, \quad n_{\text{clad}} = 1 \quad \text{at} \quad \lambda = 1.31\,\mu\text{m} \end{aligned} \tag{III.3.1}$$

All the numerical computations are performed on a cluster node with 8 to 20 cores CPU clocked at 3.0 GHz with 128 GB of reserved memory. For some example, we provide a rough estimate of the needed CPU time; notice that in each case, more than 99 % of this time is devoted to the resolution of the state or adjoint time-harmonic vector wave equations; the effort related to the level set method is negligible by comparison.

III.3.1.a Mirror

Our first numerical example deals with the optimization of a nanophotonic mirror. This component is connected to the rest of the circuit through a single waveguide which acts as both an input and an output. The goal of this device is, for a given input forward mode, to send back this mode in the same waveguide without losing any power as represented in Fig. III.3.1.

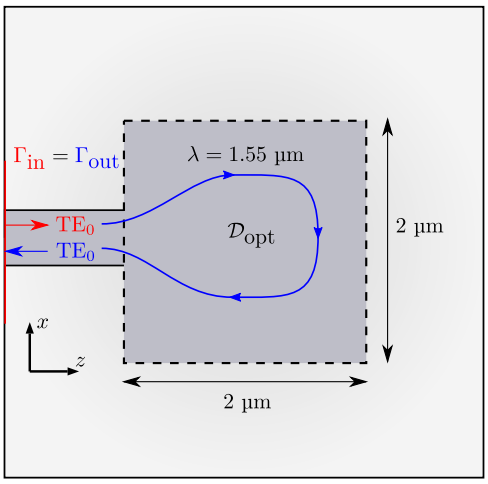


Figure III.3.1: Setting of the mirror test-case of Section III.3.1.a.

This situation falls into the framework of Th. III.2.1.1 with $\Gamma_{\text{in}} = \Gamma_{\text{out}}$ and $n = m = 1$. In particular, the adjoint state $\mathbf{A}_\Omega$ is equal to the direct solution of the PDE (the problem is said to be self-adjoint) and therefore the shape derivative only involves the electric field

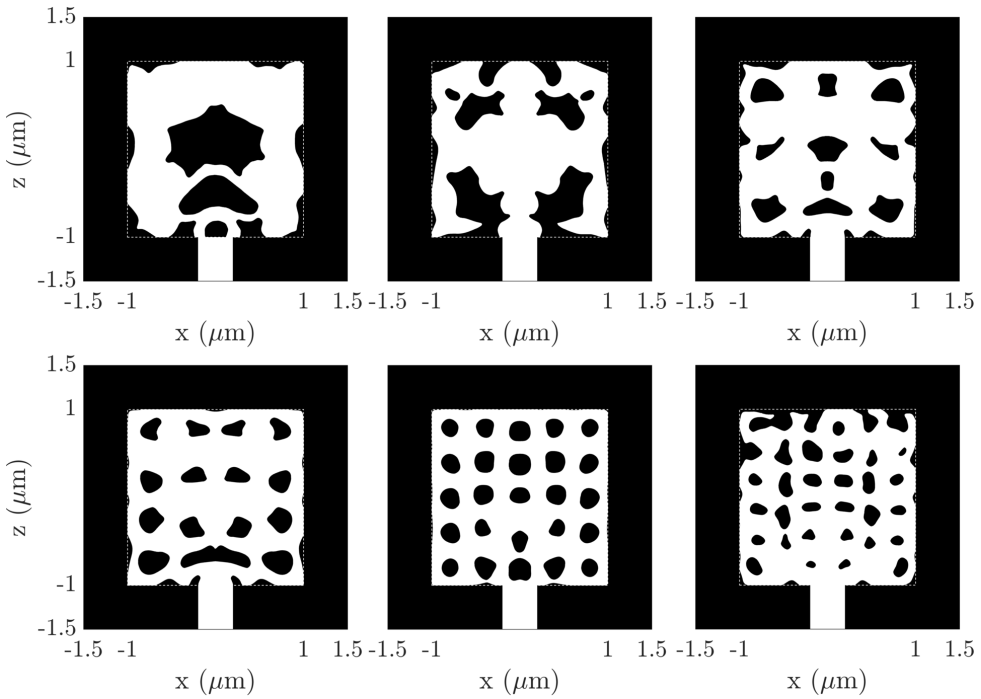


Figure III.3.2: Optimized shapes of the mirror device of Section III.3.1.a for different initializations. From left to right and top to bottom, each run was initialized with the full domain perforated by $n_h \times n_h$ holes with n_h ranging from 1 to 6.

$\mathbf{E}_\Omega$. Precisely, the shape derivative is given by

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n}\, k^2 \int_{-\frac{h}{2}}^{\frac{h}{2}} \mathrm{Re}\left[\frac{S_{m,n}(\mathbf{E}_\Omega)^*}{2i\omega\mu_0}\left((n_{\text{core}}^2 - n_{\text{cladding}}^2)|\mathbf{n}\times\mathbf{E}_\Omega\times\mathbf{n}|^2 - (n_{\text{core}}^{-2} - n_{\text{cladding}}^{-2})|n_\Omega^2\mathbf{E}_\Omega\cdot\mathbf{n}|^2\right)\right] \mathrm{d}y\,\mathrm{d}s. \quad \text{(III.3.2)}$$

Starting from an initial shape made of several holes we end up, after 200 iterations, with the shapes of Fig. III.3.2.

For each of these optimization examples, the convergence graphs are given in Fig. III.3.3 and as one would expect, the mirror's performance is almost always better when the process is started with a large number of holes. The optimized shape's successive transformations starting from a shape with one hole is shown in Fig. III.3.4 as well as a cross section of the electric field and energy density at $y = 0$.

Note that the resulting shapes are not perfectly symmetric while, in theory, since we started with a symmetric shape, the electric field should be symmetric as well and so does the vector field used to move the border of the shape. As can be seen on Figs. III.3.4(a) to III.3.4(d) this non-symmetry develops after a significant number of iterations, before it can be seen with our naked eye, and is due to the accumulation of small numerical errors. In the next examples the symmetry is enforced at each iteration.

III.3.1.b Power divider

Let us now turn to the optimization of a very useful device in nanophotonics, namely the power divider.

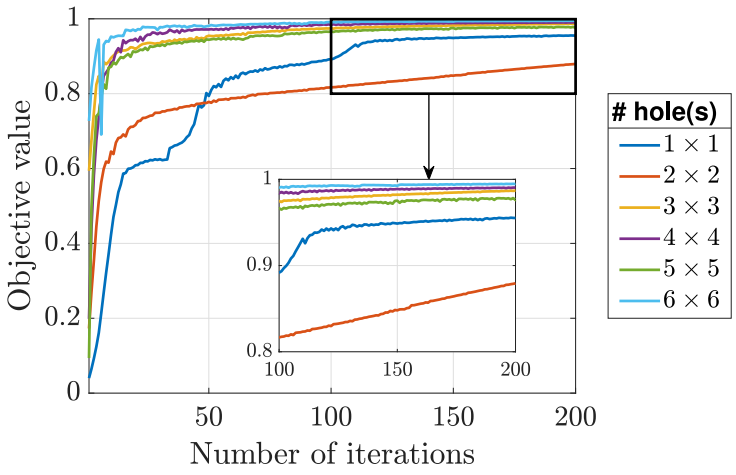
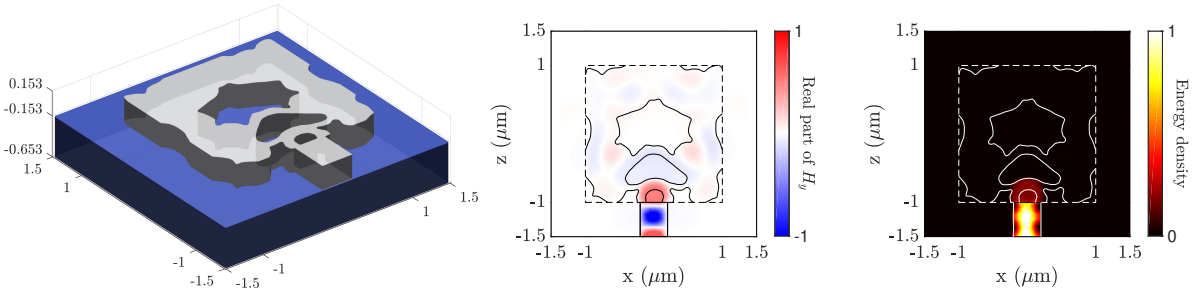


Figure III.3.3: Convergence history for the six optimization examples in Fig. III.3.2.

(a) *At* $n_{it} = 0$ (b) *At* $n_{it} = 25$ (c) *At* $n_{it} = 50$ (d) *At* $n_{it} = 100$



(e) *Optimized shape (axes in* µm*)* (f) H_y *in the section* $y = 0$ (g) ω_Ω *in the section* $y = 0$

Figure III.3.4: Evolution and result of the optimization process for the mirror test-case of Section III.3.1.a starting with an initial shape composed of one hole.

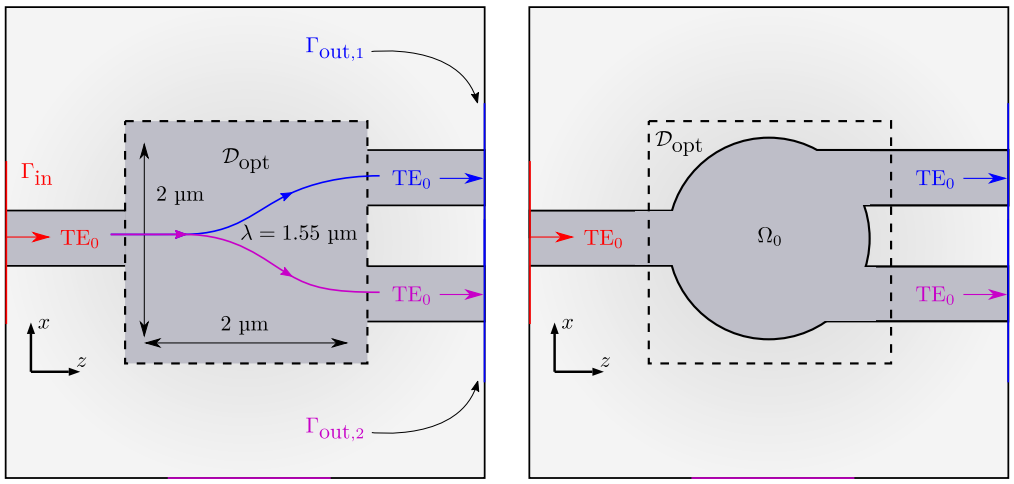
The physical setting is that of Fig. III.3.5(a): our aim is to divide the electromagnetic power conveyed by the incoming field through the waveguide figured by Γ_{in} in an equal way between both output waveguides $\Gamma_{\text{out},1}$ and $\Gamma_{\text{out},2}$. Considering the symmetry of the problem, we only discretize one half of the design domain $\mathcal{D}_{\text{opt}}$ and we restrict the set of the considered deformation fields $\boldsymbol{\theta}$ in the practice of Hadamard's method to vector fields of the form

$$\forall (x,z) \in \widehat{\mathcal{D}}_{\text{opt}}, \quad \tilde{\boldsymbol{\theta}}(x,z) = \frac{1}{2}\left(\boldsymbol{\theta}(x,z) + \boldsymbol{\theta}(-x,z)\right). \tag{III.3.3}$$

This allows to formulate our optimization problem as in Th. III.2.1.1, that is to consider the maximization of the single objective functional

$$\mathcal{J}_1(\Omega) = \frac{1}{2}\int_{\Gamma_{\text{out},1}} \left[\mathbf{E}_\Omega \times \boldsymbol{\mathcal{H}}_{-1}^{\text{out},*}\right] \cdot \mathbf{n}\, \mathrm{d}s. \tag{III.3.4}$$

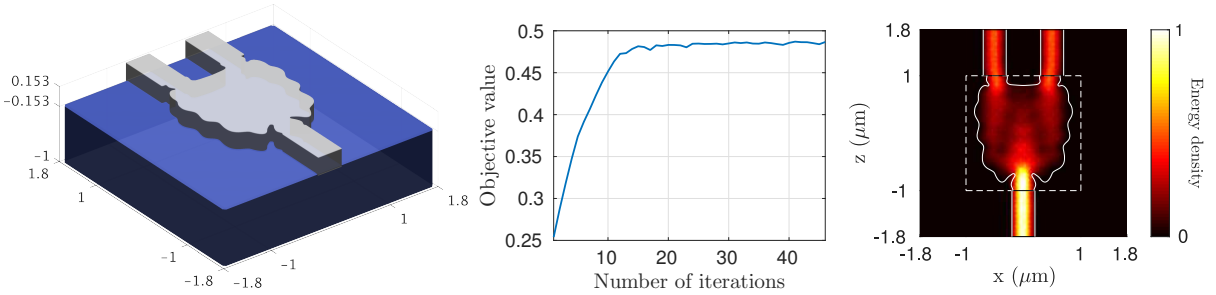
Note that considering the symmetrized vector field given by Eq. (III.3.3) is, if the shape Ω is also symmetric, the same as optimizing the objective function $\mathcal{J}_1(\Omega) + \mathcal{J}_2(\Omega)$ where $\mathcal{J}_i$ is given by Eq. (III.3.4) with $\Gamma_{\text{out},i}$ instead of $\Gamma_{\text{out},1}$.



(a) *Schematic representation of the power divider* (b) *Initial shape used by the algorithm*

Figure III.3.5: Setting of the power divider test-case of Section III.3.1.b.

Starting from the initial shape of Fig. III.3.5(b), 50 iterations of our optimization algorithm are performed, for a total computational time of roughly 3 hours. Details of the numerical computation are reported on Fig. III.3.6; the optimized device achieves approximately 49 % transmission into each output waveguides.



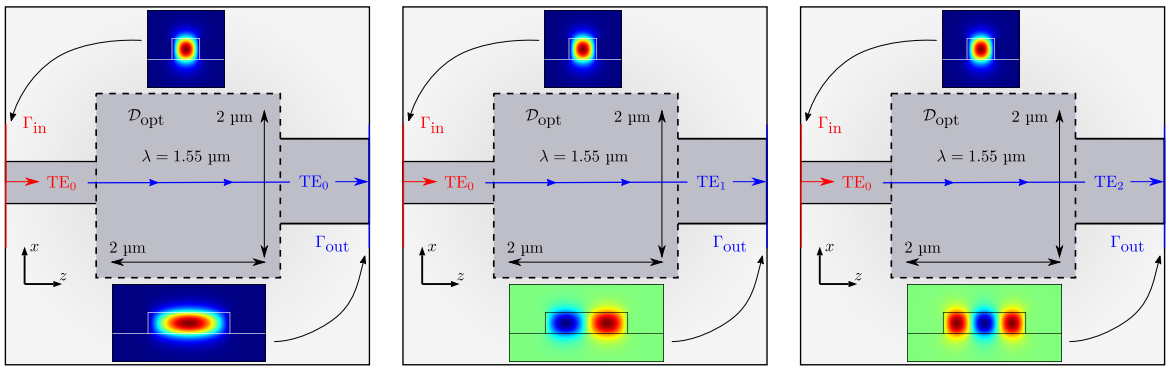
(a) *Optimized shape (axes in* µm*)* (b) *Convergence history* (c) ω_Ω *in the section* $y = 0$

Figure III.3.6: Optimized shape of the power divider of Section III.3.1.b and details of the numerical computation.

III.3.1.c Mode converters

Our third example deals with the optimization of the shape of devices whose common purpose is to transform the mode coming from an input waveguide into another mode of the output waveguide.
The physical settings of interest are depicted on Fig. III.3.7, where the output waveguide is wide enough to allow for the existence of multiple guided modes. In this context, the electromagnetic power is injected via the port Γ_{in}, using the fundamental mode TE_0, and we seek to transfer this power to the first, second or third TE mode of the output waveguide.

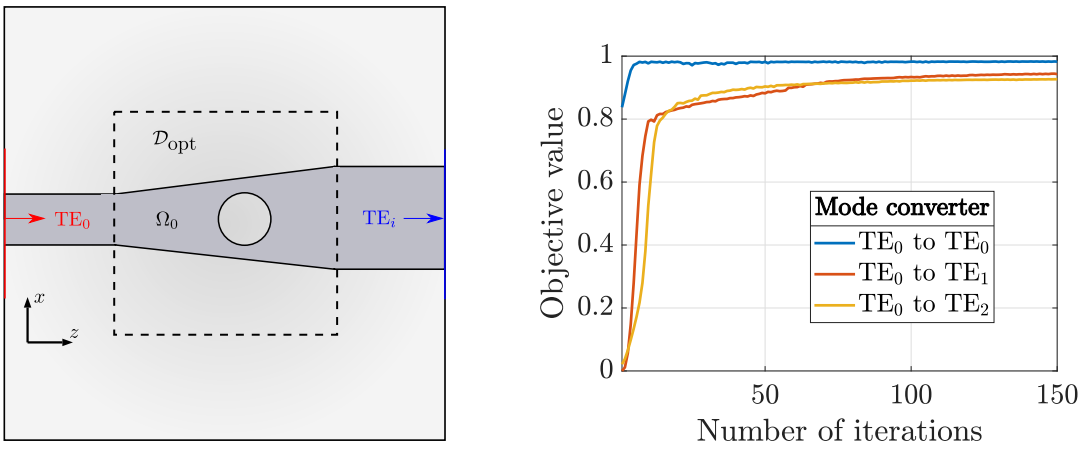
Due to the symmetry of the situation in the case of the TE_0 and TE_2 modes with respect to the x variable (see Fig. III.3.7; the TE_1 mode is, unlike the others, antisymmetric),



(a) TE_0 *to* TE_0 (b) TE_0 *to* TE_1 (c) TE_0 *to* TE_2

Figure III.3.7: Setting of the three mode converters test-cases of Section III.3.1.c.

only one half of the design domain $\mathcal{D}_{opt}$ is discretized in this case and we use symmetrized deformation fields of the form Eq. (III.3.3) in the practice of Hadamard's method.



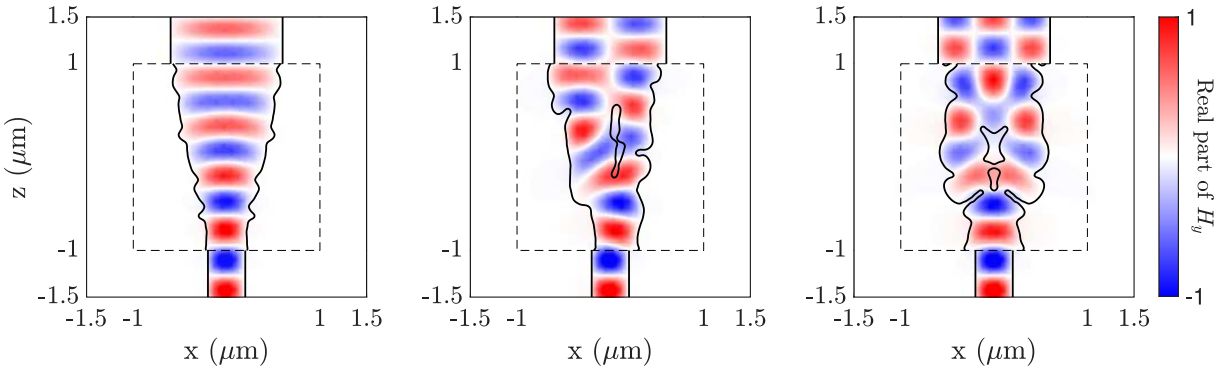
(a) *Initial shape used by the algorithm* (b) *Convergence history graphs*

Figure III.3.8: Application of our shape optimization algorithm to the mode converters of Section III.3.1.c.

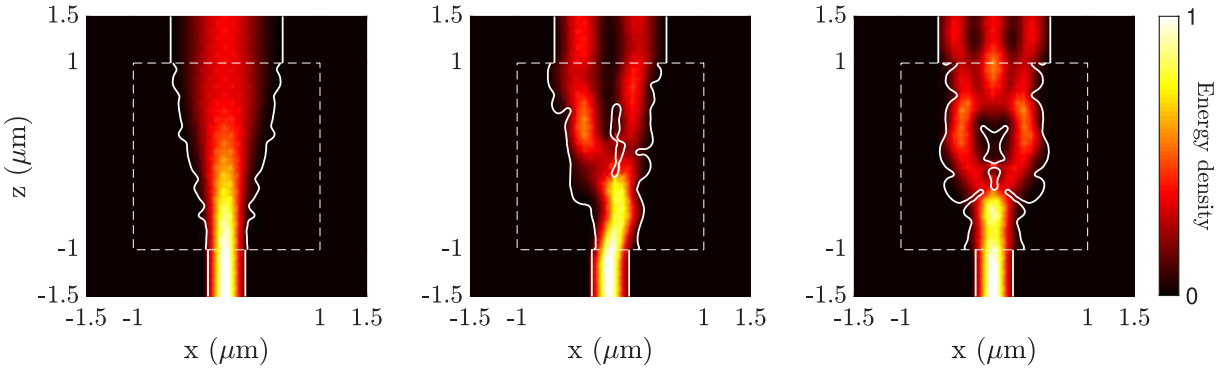
Starting from the initial shape of Fig. III.3.8(a) we perform 150 iterations of our optimization algorithm leading to the convergence graphs of Fig. III.3.8(b). We end up with the designs shown in Fig. III.3.9.
Note here that we only considered conversion from TE into TE modes. When the mode converter modifies also the polarization it is called a polarization rotator, which is a far more difficult effect to achieve; see Section III.5.

III.3.1.d Crossing

The purpose of this device is to facilitate circuit routing by limiting the crosstalk (undesired coupling power) between two intersecting waveguides (see Fig. III.3.10(a)). The design domain $\mathcal{D}_{opt}$ is connected to two input waveguides via the ports $\Gamma_{in,1}$ and $\Gamma_{in,2}$, and two outgoing waveguides via the surfaces $\Gamma_{out,1}$ and $\Gamma_{out,2}$. The fundamental mode TE_0 is injected at $\Gamma_{in,1}$ (resp. $\Gamma_{in,2}$) with a wavelength $\lambda = 1.55\,\mu m$ and our aim is to maximize the transmitted energy $\mathcal{J}_1(\Omega)$ to the fundamental mode in $\Gamma_{obj} = \Gamma_{out,1}$ (resp. $\Gamma_{out,2}$).

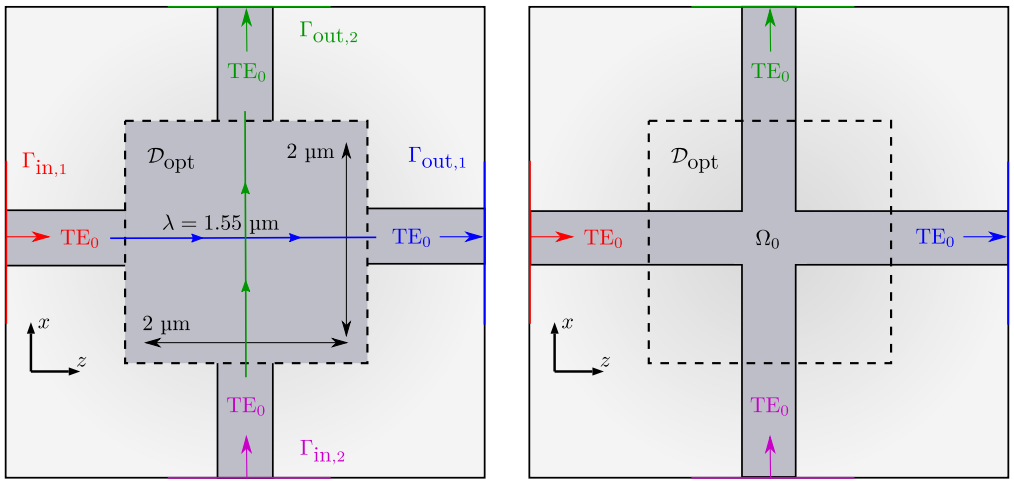


(a) *Real part of H_y in the section $y = 0$*



(b) *Energy density in the section $y = 0$*

Figure III.3.9: Simulation of the time-harmonic vector wave equation using the final designs of the mode converters obtained in Section III.3.1.c.



(a) *Schematic representation of the crossing*

(b) *Initial shape used by the algorithm*

Figure III.3.10: Setting of the crossing test-case of Section III.3.1.d.

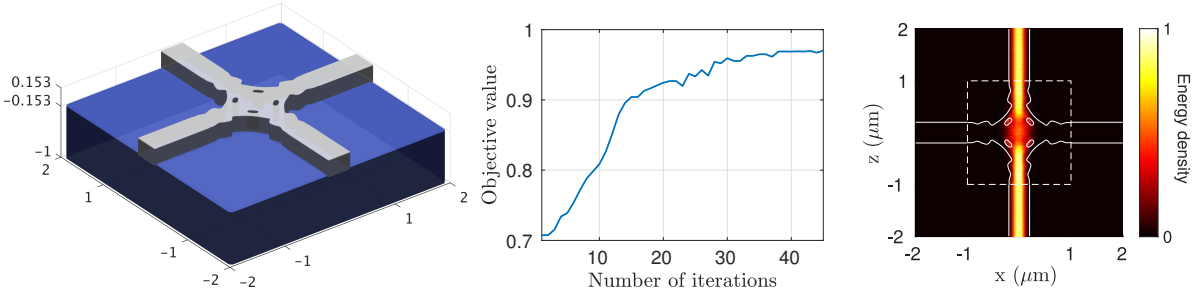
Taking advantage of the symmetry of the situation, we only consider shapes Ω which are symmetric with respect to the x and z axes; doing so ensures the symmetry between the electromagnetic fields $\mathbf{E}_\Omega$ and $\mathbf{H}_\Omega$ obtained in the situations where light is injected from $\Gamma_{in,1}$ and $\Gamma_{in,2}$. Hence, our shape optimization problem boils down to that of maximizing the single objective in Th. III.2.1.1, where the Maxwell equations describing the physics at play involve only injection through the port $\Gamma_{in,1}$ and where $\Gamma_{out} = \Gamma_{out,1}$. Accord-

ingly, in the practice of the boundary variation method of Hadamard, we only consider symmetrized vector fields $\boldsymbol{\theta}$ of the form

$$\forall (x,z) \in \widehat{\mathcal{D}}_{\text{opt}},\ \ \tilde{\boldsymbol{\theta}}(x,z) = \frac{1}{4}(\boldsymbol{\theta}(x,z) + \boldsymbol{\theta}(-x,z) + \boldsymbol{\theta}(x,-z) + \boldsymbol{\theta}(-x,-z)). \tag{III.3.5}$$

Starting from an initial shape made of the union of two orthogonal, straight waveguides connecting $\Gamma_{\text{in},1}$ to $\Gamma_{\text{out},1}$ and $\Gamma_{\text{in},2}$ to $\Gamma_{\text{out},2}$ (see Fig. III.3.10(b)), 50 iterations of our optimization algorithm are performed, for a total computational time of roughly 4 hours.

The optimized design, convergence history, as well as the normalized density of electromagnetic energy inside the computational domain are represented on Fig. III.3.11. We notice that more than 95 % of the electromagnetic energy contained in the incoming field is successfully conveyed to the output port, while this ratio equals only 70 % in the case of the initial device of Fig. III.3.10(b). Notice that the shape Ω has changed topology in the course of the optimization process. In this example, the crosstalk, that is the undesired coupling to the perpendicular waveguides is less than 1 %.



(a) *Optimized shape (axes in* µm*)* (b) *Convergence history* (c) ω_Ω *in the section* $y = 0$

Figure III.3.11: Optimized shape of the crossing device of Section III.3.1.d and details of the numerical computation.

III.3.1.e Cavities

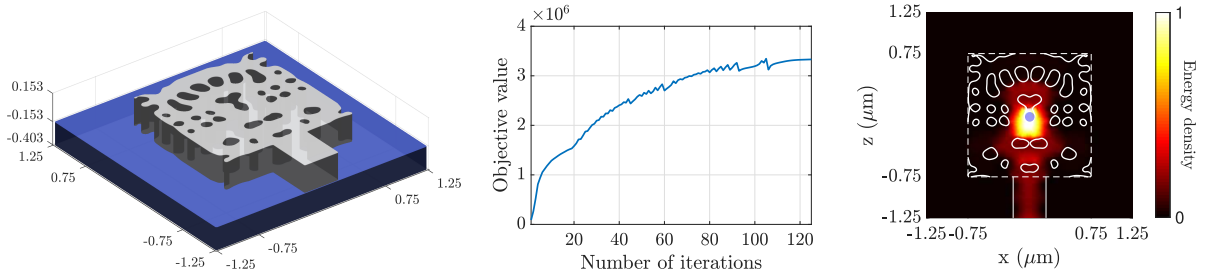
Here, we consider the mode volume objective of Section III.2.3.b. We consider the same geometry as in Fig. III.3.1 with an additional cylinder V at the center of $\mathcal{D}_{\text{opt}}$ composed only of cladding material and in which we want to maximize the mode volume defined by Eq. (III.2.31), that is

$$\mathcal{J}(\Omega) = \int_V n_{\text{core}}^2 |\mathbf{E}_\Omega(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x}.$$

Again, the symmetry of the problem allows to consider only 50% of the whole domain using vector fields of the form Eq. (III.3.3) in the practice of Hadamard's method. Applying 125 iterations of our numerical algorithm we end up with the result of Fig. III.3.12 and we can clearly see on Fig. III.3.12(c) that the energy is condensed in the small blue (cylindrical) volume.

III.3.2 Comments on the use of the topological gradient

In Section II.1.3 the topological gradient was presented as a means to optimize a functional $\mathcal{J}(\Omega)$ depending on a shape Ω by iteratively nucleating holes into Ω. We have seen that



(a) *Optimized shape (axes in* µm*)* (b) *Convergence history* (c) ω_Ω *in the section* $y = 0$

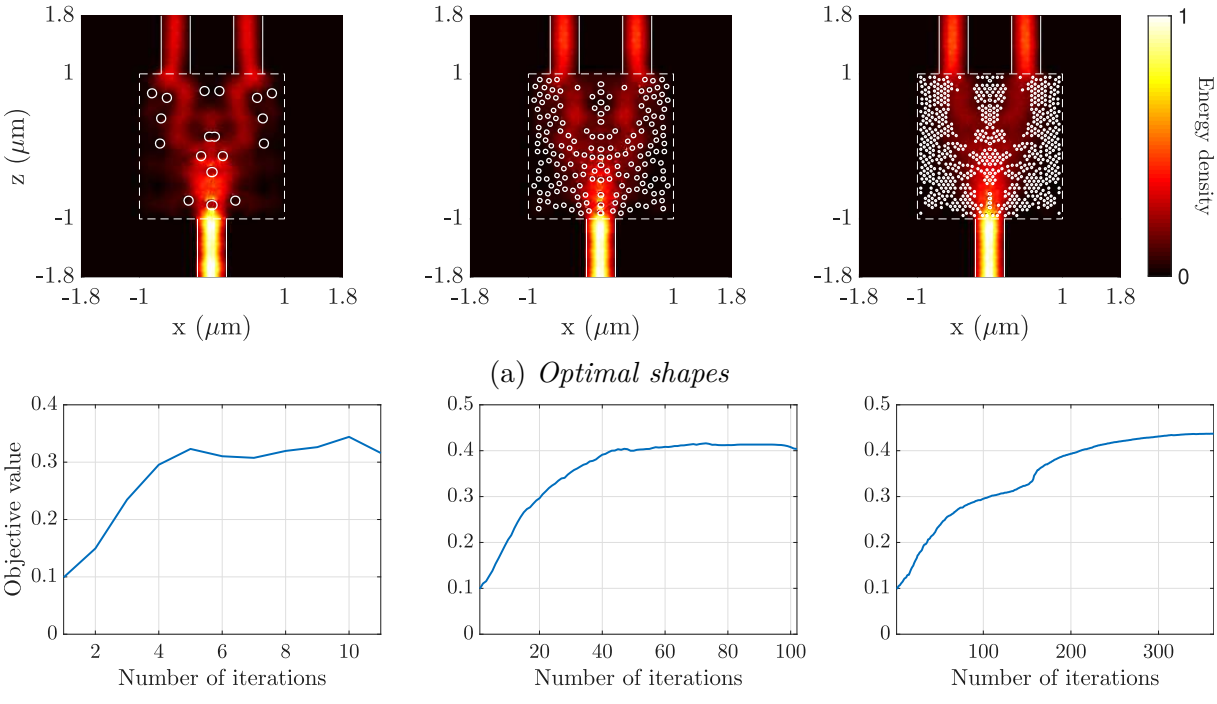
Figure III.3.12: Optimized shape of the cavity of Section III.3.1.e and details of the numerical computation. The initial shape was composed of 8×8, equally spaced holes.

the location of where a hole should be nucleated inside Ω relies on an asymptotic expansion of the objective functional as Eq. (II.1.18), that is

$$\mathcal{J}(\Omega_\varepsilon) = \mathcal{J}(\Omega) + f(\varepsilon)\mathcal{T}(\mathbf{x}_0) + o(f(\varepsilon)) \tag{III.3.6}$$

where $\Omega_\varepsilon = \Omega \backslash \overline{B(\mathbf{x_0}, \varepsilon)}$ and ε is the radius of the hole.

In our application, Ω is a 3d shape, which is y-invariant (see Section III.2.1.b), and where a 2d section $\widehat{\Omega}$ is optimized. Naturally, a small hole inside $\widehat{\Omega}$ amounts to performing a variation of the 3d shape of the form $\Omega \backslash \overline{C(\mathbf{x_0}, \varepsilon)}$ where $C(\mathbf{x}_0, \varepsilon)$ is a cylinder of radius ε, centered at $\mathbf{x}_0$ and with fixed height h (that of the component) in the y-axis.



(a) *Optimal shapes*

(b) *Convergence history graphs*

Figure III.3.13: Formal application of the "topological gradient" of Section III.3.2. From left to right: a diameter of the holes and a minimum distance between holes of 100, 50 and 25 nm is used. The optimization procedure was stopped when the topogical gradient was everywhere negative.

This type of asymptotic expansion is highly unusual and we did not try to find an equivalent to Eq. (III.3.6) in our context. Nevertheless, as suggested in Section II.1.3, a formal approach of this problem suggested that the sign of the shape gradient V_Ω in Eq. (III.2.2) (or more precisely its integrated values on $[-h/2, h/2]$ as in Eq. (III.2.12)), indicates where to remove matter into the shape. This method has been tested on the optimization of the power divider of Section III.3.1.b and some results are presented in Fig. III.3.13 using different sizes for the holes and a minimal distance between them. Interestingly, the convergence graphs of Fig. III.3.13(b) are almost always strictly increasing meaning that the approximation used here for the topological gradient works well in practice.

In reality this good result may be explained in this case by the optical index regularization process of Section III.2.2. Indeed, since the optical index n_Ω is smoothed by applying a convolution s_η on its values (defined in Eq. (III.2.20)), removing a cylinder $C(\mathbf{x}_0, \varepsilon)$ of radius ε into Ω amounts to changing $n^2_{\Omega,\eta}$ into $n^2_{\Omega,\eta} - n^2_{\varepsilon,\mathbf{x}_0}$ where $n^2_{\varepsilon,\mathbf{x}_0} \in C^\infty(\mathbb{R}^2, \mathbb{R})$ is defined by

$$n^2_{\varepsilon,\mathbf{x}_0}(x, y, z) = (n^2_{\text{core}} - n^2_{\text{clad}}) \int_{C(\mathbf{x}_0,\varepsilon)} s_\eta(x_0 - t, z_0 - u)\, \mathrm{d}t\, \mathrm{d}u,$$

implying in particular that $\lim_{\varepsilon\to 0} n^2_{\varepsilon,\mathbf{x}_0} = 0$. In this case, studying the sensitivity of the objective function with respect to the removal of an infinitely small cylinder then reduces to compute the derivative of the objective function $\mathcal{J}$ with respect to ε at 0. To do this, we first define for any fixed shape Ω and position $\mathbf{x}_0 \in \Omega$ the objective function $\mathcal{J}_h(\varepsilon) = \mathcal{J}(\Omega \backslash C(\mathbf{x}_0, \varepsilon))$. A direct calculation using Céa's method allows to find that the shape gradient with respect to ε of $\mathcal{J}_h(\varepsilon)$ is proportional to the scalar field in Eq. (III.2.22). This result means that using the shape derivative as a mean to know where to remove holes in the shape Ω is justified in the context of smoothed optical indices.

Once again we insist on the fact that this result is only true because of the index smoothing method. In order to obtain the real shape derivative without taking into account the smoothing it is still necessary to find an asymptotic expansion as in Eq. (II.1.18) and this analysis was not considered during this thesis.

III.4 Incorporating constraints into the optimization process

III.4.1 Introduction

For the moment, no constraints have been imposed on Ω during the resolution of the shape optimization problem. In practice, however, not all shapes are possible to manufacture; but defining precisely the set of shapes that may be produced is a complex task. Nowadays, engineers use programs known as Direct Rule Check (DRC) to verify if a design can be produced in reality. These programs mainly consists in the verification of three "rules":

- **Minimum gap space**: if two different regions of the design Ω are too close from one another then they are likely to end up accidentally merged in the concretely produced design (Fig. III.4.1, bottom left). Hence, such pattern are impossible to assemble without alterations.

- **Minimum feature size**: if a region of the shape is too thin then it will disappear in the final design (Fig. III.4.1, top left).

- **Maximum local curvature**: sharp corners (high local curvature value) are smoothed by the manufacturing process (Fig. III.4.1, right) and thus cannot be produced accurately.

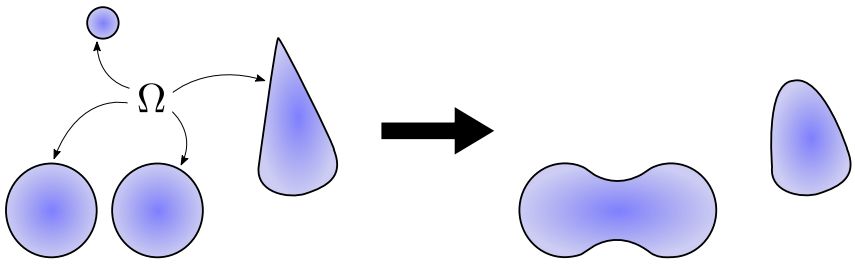


Figure III.4.1: (left) A shape Ω. (right) The associated produced shape.

These DRC programs provide criteria that should be respected for the desired design to be manufactured faithfully. In theory these rules are not "constraints" in the sense that we could still try to produce shapes which do not respect them but in practice engineers require physicists to respect these rules in order to guarantee the proper manufacturing of the designs.

Even though the modeling and incorporation of manufacturing constraints in the optimization process is not the primary target of this thesis, we hereafter provide some information about recently published paper on this topic concerning nanophotonic devices with a particular emphasis on strategies used in the geometric optimization framework. Let us also point out that this kind of constraints have already been studied extensively in the context of density-based shape optimization method; see for instance [Zho15].

III.4.2 Projection method

Let us suppose that it is possible to derive a function F which, to any shape $\Omega \subset \mathcal{D}_{\text{opt}}$, associates a "projection" of Ω onto the set of admissible shapes. In such a case, a projected gradient algorithm can be used to enforce these constraints. In [Pig17] the authors proposed to apply the two following projections after each iteration of the gradient descent in order to enforce the three previous rules of manufacturability:

- Remove the small isolated connected elements of Ω and the small isolated bubbles of air in Ω.

- Solve the Hamilton-Jacobi equation over a "short" time period using a scalar field equal to the opposite of the local curvature κ of the shape in order to reduce the values of κ.

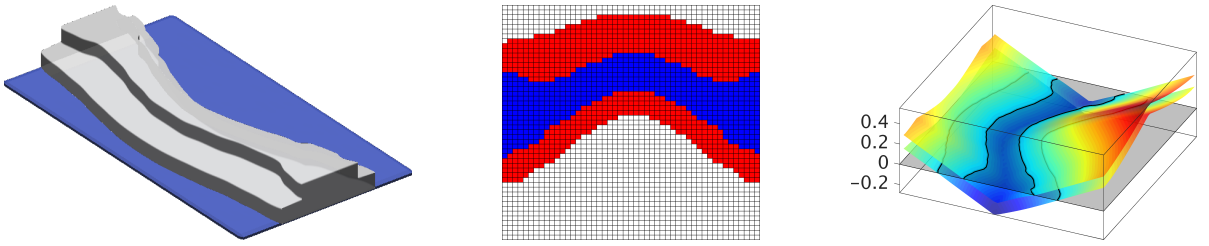
It is worth noting that in the second projection, using a scalar field equal to the opposite of κ in the Hamilton-Jacobi equation is mathematically equivalent to adding a penalization on the perimeter of the shape to the optimization problem (see Eq. (II.1.4)).

III.4.3 Penalization methods

Another possibility to enforce manufacturing constraints on the optimized shape is to rely on penalization. In this case, a term is added to the objective function in order to penalize the regions of the shape which violate the constraints. This penalization may be expressed by using the level-set numerical representation of the shape as in [Ver19a] or by using the signed distance function [Mic14, Section II.3].

III.5 Multi-levels: shapes that can be made through multiple levels of etching

This section is dedicated to the presentation of a method for the design of nanophotonic components exploiting the possibility of several etching levels (see Fig. III.5.1 for an example).



(a) *Three dimensional representation of a two-layer shape*

(b) *Views of the two 2d layers constituting the component, red: bottom layer, blue: top layer*

(c) *Level-set representations of the two layers*

Figure III.5.1: A two-layer component and its numerical representations.

III.5.1 Motivation: polarization rotator

In Sections I.1.4 and I.2.1.b the polarization of a mode was defined for a y-invariant material by looking at the behavior of the electromagnetic fields. In a nutshell, in the case of y-invariant optical indices, the time-harmonic vector wave equation is simplified into two uncoupled two-dimensional scalar Helmholtz equations (Eq. (I.1.19)) giving access to respectively $(\mathrm{E}_x, \mathrm{H}_y, \mathrm{E}_z)$ and $(\mathrm{H}_x, \mathrm{E}_y, \mathrm{H}_z)$. In the case of non-$y$-invariant optical indices, a mode is said to be in the TE polarization if the majority of its power is carried by the first set of three components and is said to be TM in the other case.

A polarization rotator is a component which converts an input TE (or TM) mode into a mode of the output waveguide with the other polarization as presented in Fig. III.5.2. Given the independence between TE and TM modes in the case of y-invariant optical indices, it is not possible to convert an input polarization into the other by using a device where the index is y-invariant. Even if general waveguides such as the one studied previously (Fig. I.2.1) are indeed not y-invariant due to the presence of a cladding and substrate with different optical indices than the one of the core, we can legitimately assume that having "more" variations of the optical indices in the y-direction may improve the performance of the polarization rotator. In order to enhance this variation, one idea is to

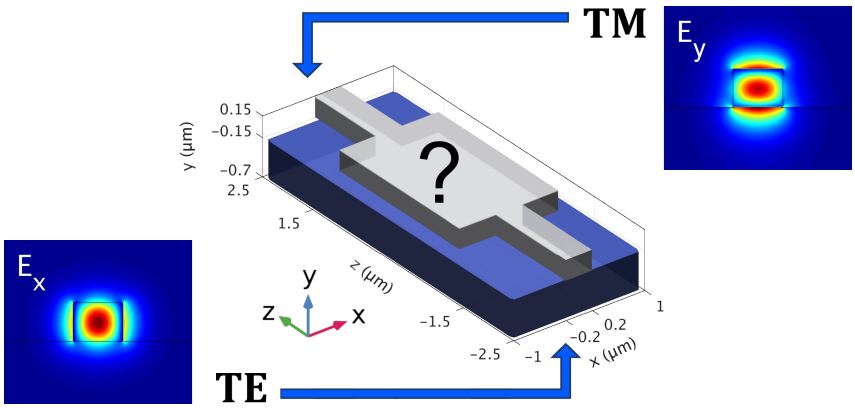
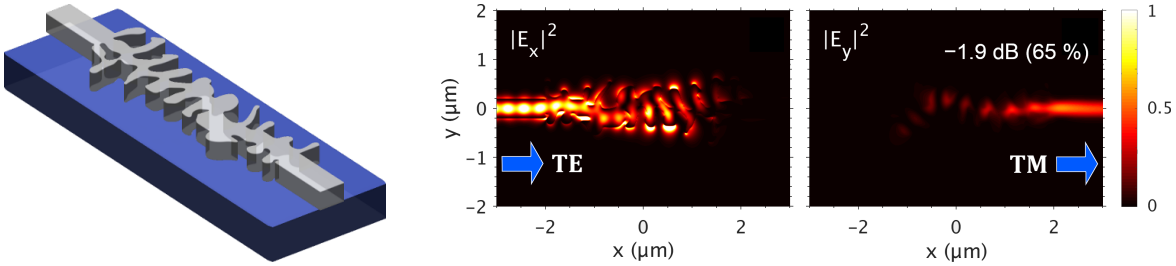


Figure III.5.2: Schematic representation of a polarization rotator.

consider components having several depths of etching (see Fig. III.5.1), that is allowing the shape to have different layers instead of being restricted to the one of the waveguides.

Still, if we try to optimize a polarization rotator using a single etching level though, starting from a straight waveguide we obtain, after more than a thousand iterations, the results shown in Fig. III.5.3.



(a) *Optimal shape* (b) *Cut of the energy density at* $y = 0$, $\lambda = 1.55\,\mu\text{m}$

Figure III.5.3: Classical shape optimization of a polarization rotator starting from a straight waveguide.

As we can see only 65 % of the input fundamental TE mode is well-converted into the fundamental TM mode, which still leaves a lot of room for improvement.

III.5.2 A first method using projections

In order to optimize the n_ℓ layers $\Omega_1, \ldots, \Omega_{n_\ell}$ constituting the shape $\Omega = \bigcup_{i=1,\ldots,n_\ell} \Omega_i$, we can still use the result of Th. III.2.1.1 by changing the height of integration in the shape derivative (III.2.1) for each layer (see below Eq. (III.5.2)). But we also need a way to ensure that each layer is always placed above the lower ones; this condition is necessary to produce the shape through etching. For each layer Ω_i, we denote by $\widehat{\Omega}_i \subset \mathbb{R}^2$ its two dimensional section such that

$$\Omega_i = \{(x, y, z), (x, z) \in \widehat{\Omega}_i, y \in [h_{i-1}, h_i]\}$$

with $-h/2 = h_0 < \ldots < h_{n_\ell} = h/2$. With these definitions, the manufacturing constraint reduce to enforcing that

$$\widehat{\Omega}_1 \supset \ldots \supset \widehat{\Omega}_{n_\ell},$$

The general shape optimization problem is now

$$\begin{cases} \max\limits_{\Omega_1,\dots,\Omega_n} & \mathcal{J}(\Omega) \\ \textbf{s.t.} & \widehat{\Omega}_1 \supset \dots \supset \widehat{\Omega}_n \end{cases} \tag{III.5.1}$$

where the value of $\mathcal{J}(\Omega)$ is given through the solution of the time-harmonic vector wave equation using the optical index defined through $\Omega_1, \dots, \Omega_n$. As in Eq. (III.2.12), the mathematical program Eq. (III.5.1) may be optimized by considering the following vector fields $\boldsymbol{\theta}_i$ to move each shape Ω_i in the practice of the boundary variation method of Hadamard:

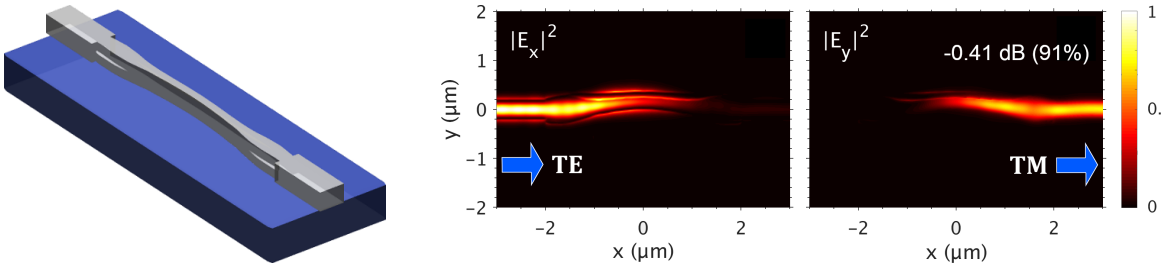
$$\boldsymbol{\theta}_i(x,z) = \mathbf{n}(x,z) \int_{h_{i-1}}^{h_i} V_\Omega(x,y,z)\,\mathrm{d}y. \tag{III.5.2}$$

where V_Ω was defined in Th. III.2.1.1.

To take into account the constraint, first note that it is sufficient to require that $\phi_1 < \dots < \phi_{n_\ell}$ where ϕ_i is a level-set representation of $\widehat{\Omega}_i$. With this new formulation of the constraint through level-set functions, a simple numerical algorithm to solve Eq. (III.5.1) is a projected gradient algorithm where, at each iteration, the constraint is ensured by using the following projection:

$$\text{for } i \text{ from } 2 \text{ to } n_\ell \text{ do:} \quad \phi_i = \max(\phi_i, \phi_{i-1}). \tag{III.5.3}$$

With this algorithm and two layers ($n_\ell = 2$) we end up this time with the results of Fig. III.5.4. As we can see, the final design is geometrically more simple than the one in Fig. III.5.3(a) and exhibit a drastic improvement of the transmission, greater than 90 % for the same compactnes (that is, the same size of the optimization domain). A comparison between the convergence graphs of both one- and two-layer designs is given in Fig. III.5.5 and, as can be expected, the geometrically simpler shape is obtained in far fewer iterations of the optimization algorithm.



(a) *Optimal shape* (b) *Cut of the energy density at* $y = 0$, $\lambda = 1.55\,\mu\text{m}$

Figure III.5.4: Shape optimization of a two-level polarization rotator starting from a straight waveguide.

Let us now optimize the same component with different lengths of the optimization domain, as represented in Fig. III.5.6. In order to increase the confidence in our numerical results, a comparison between the performances of the optimized design measured by three different numerical softwares was additionally performed (results in Fig. III.5.6(a)), namely `Lumerical` and `RSoft Photonics` using 3D-FDTD (Section I.5.2) as well as `Comsol Multiphysics` using 3D-FEM (Section I.5.1).
The transmission as a function of the wavelength is shown in Fig. III.5.6(b) (top) for polarization rotator length ranging from 3 µm to 6 µm. A clear improvement of the insertion losses (the power lost during the polarization rotation) can be seen from the blue

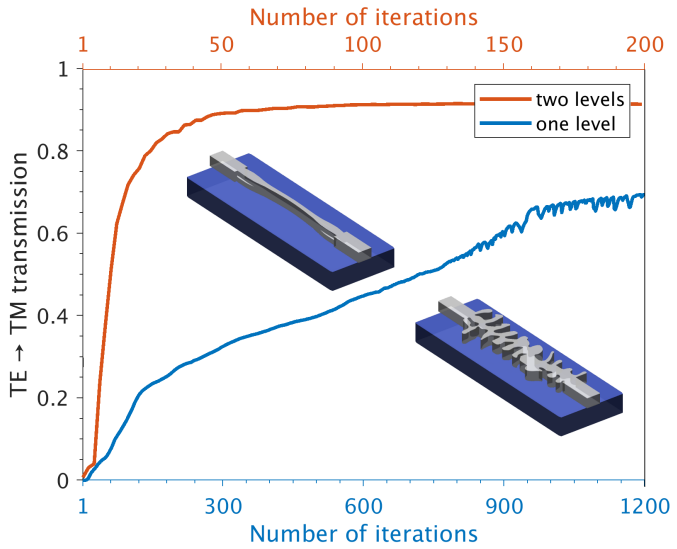
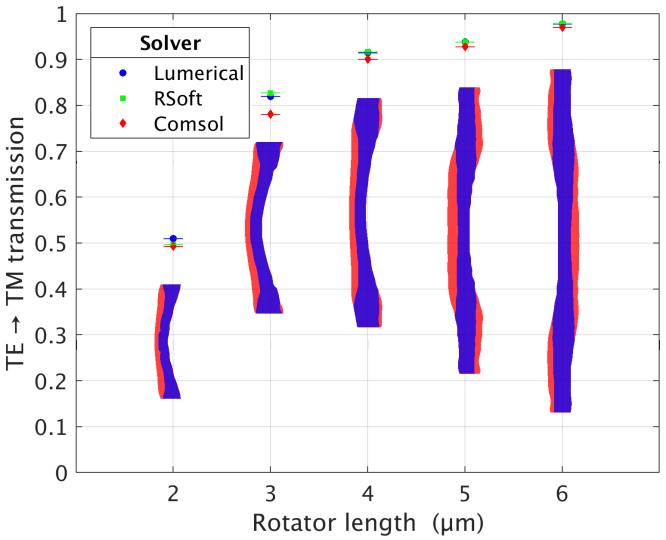
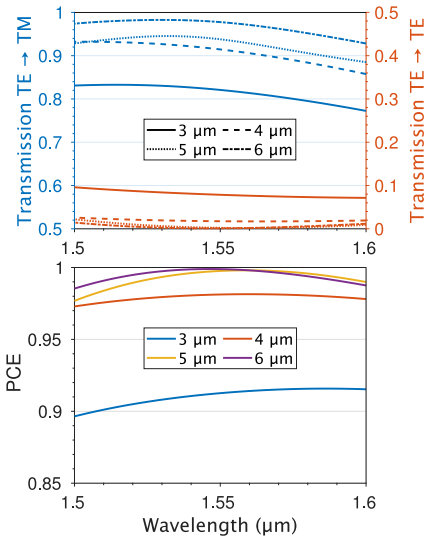


Figure III.5.5: Convergence graphs of both one and two layers designs. Note the different x-axis for both optimization.



(a) *Optimal shapes. Each device is represented in red (bottom layer) and blue (top layer). The multiple dots for each domain length corresponds to the TM transmission in the output waveguide computed using different commercial software.*



(b) *(top) Transmission into TM and TE polarizations inside the output waveguide for each two-levels designs, (bottom) corresponding polarization conversion efficiency defined in Eq. (III.5.4).*

Figure III.5.6: Shape optimization of several two-level polarization rotator starting from a straight waveguide and an optimization domain with length between 2 to 6 µm.

curves when increasing the length, down to −0.2 dB (95%) whereas the residual TE transmissions remains constant after 5 µm. This fact can be readily seen on the polarization conversion efficiency plot of Fig. III.5.6(b) (bottom), for which the broadband capability of our two-level designs is displayed. The **polarization conversion efficiency (PCE)** is defined as

$$\mathrm{PCE} = \frac{P_{\mathrm{TM}}}{P_{\mathrm{TE}} + P_{\mathrm{TM}}} \tag{III.5.4}$$

where P_{TE} (resp. P_{TM}) are given by Eq. (III.1.1) using the fundamental TE (resp. TM) mode in the overlap integral.

Let us conclude these results by having a few words about the projection method explained in Eq. (III.5.3). First, other projections methods may be used. Indeed by choosing any layer $j \in \{1, \dots, n_\ell\}$ the projections may be performed in a different order by doing

$$\begin{aligned} \text{for } i \text{ from } 1 \text{ to } j-1 \text{ do:} \quad & \phi_i = \min(\phi_i, \phi_{i+1}), \\ \text{for } i \text{ from } j+1 \text{ to } n_\ell \text{ do:} \quad & \phi_i = \max(\phi_i, \phi_{i-1}). \end{aligned}$$

One drawback of doing so is that, whatever the order of the projections, this method will always give priority to one level over the others. If we want to derive a general method which is not restricted by any order of projection it is necessary to find some sort of best common descent direction for all the layers, which satisfies the inclusion constraint.

III.5.3 Theoretical general frameworks

Before closing this chapter, we propose here two methods to derive an algorithm which does not give priority to any layer. This part is still an ongoing work and no numerical illustration is given for these methods.

III.5.3.a Penalization of the distance

Let us start with the most simple way to deal with this constraint. Again, we limit ourselves to only two layers for simplicity, meaning that Ω_1 (resp. Ω_2) represent the bottom (resp. top) layer. As we have seen in Section III.5.2, the inclusion constraint $\widehat{\Omega}_1 \supset \widehat{\Omega}_2$ may be equivalently reformulated as $\phi_1 < \phi_2$ where, for $i = 1, 2$, ϕ_i is one level-set representation of $\widehat{\Omega}_i$. In particular, this is true using the signed distance functions as level-sets (defined in Section II.3.3.b), that is $\phi_i = d_{\Omega_i}$. Since d_{Ω_i} is uniquely defined for every shape Ω_i, it is possible to study its variations when a deformation is applied on Ω_i. We can then define the penalized objective function to maximize as

$$\tilde{\mathcal{J}}_\delta(\Omega) = \mathcal{J}(\Omega) + \delta \int_{\mathcal{D}_{\text{opt}}} (d_{\Omega_2} - d_{\Omega_1}) \, \mathrm{d}\mathbf{x}, \tag{III.5.5}$$

for which we can prove that it is Gâteau-differentiable at $\boldsymbol{\theta} = 0$ (where $\boldsymbol{\theta} = \boldsymbol{\theta}^i$ in Ω_i) and that

$$\tilde{\mathcal{J}}_\delta(\Omega) = \mathcal{J}'(\Omega)(\boldsymbol{\theta}) + \delta t \int_{\mathcal{D}_{\text{opt}}} (\boldsymbol{\theta}^2 \cdot \mathbf{n}) \circ p_{\partial\Omega_2} - (\boldsymbol{\theta}^1 \cdot \mathbf{n}) \circ p_{\partial\Omega_1} \, \mathrm{d}\mathbf{x}, \tag{III.5.6}$$

where $p_{\partial\Omega}(\mathbf{x})$ is the projection of $\mathbf{x}$ on $\partial\Omega$; see [Dap13, Section 4.2.2]. Using a decomposition along rays, the second integral in Eq. (III.5.6) may be written as

$$\int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} \left(\int_{\text{ray}_{\partial\Omega_2}} (1 + d_{\Omega_2}\kappa_{\Omega_2}) \, \mathrm{dt} - \int_{\text{ray}_{\partial\Omega_1}} (1 + d_{\Omega_1}\kappa_{\Omega_1}) \, \mathrm{dt} \right) \mathrm{d}s, \tag{III.5.7}$$

where κ_{Ω_i} is the local curvature on $\partial\Omega_i$ and we refer again to [Dap13, Corollary 4.2] for the definition of $\text{ray}_{\partial\Omega_i}$. Equation (III.5.7) has the same structure as the previously obtained shape derivatives which allow to find an ascent direction. The drawback of this whole penalization scheme is that the optimized layers will tend to be as distant as possible one to the other. To alleviate this problem it is possible to change the integrand in Eq. (III.5.5) into $H_\varepsilon(d_{\Omega_2} - d_{\Omega_1})$ where H_ε is a smooth Heaviside function to only have a penalization where $\partial\Omega_1$ is really close to $\partial\Omega_2$ but this will still prevent these borders from being exactly on top of each other (that is, $\partial\widehat{\Omega}_1 \cap \partial\widehat{\Omega}_2 = \emptyset$).

III.5.3.b Modifying the Hamilton-Jacobi equation

We now move on to a different, more "algorithmic", approach of this problem. By "algorithmic" we mean that it will consider the numerical discretization of the shape instead of dealing with its continuous representation.

Remembering Section II.3.3.a, we saw that to find a level-set representation of $(\mathrm{Id}+\theta\mathbf{n})(\Omega)$ we have to solve the Hamilton-Jacobi equation $\partial_t\phi(\mathbf{x},t)+\theta(\mathbf{x})|\nabla_{\mathbf{x}}\phi(\mathbf{x},t)|=0$ on $t\in]0,1]$ with the initial condition $\phi(\mathbf{x},0)=\phi_0(\mathbf{x})$ where ϕ_0 is a level-set representation of Ω. Numerically, the resolution of the Hamilton-Jacobi equation was explained in Section II.3.4.c and we have seen that it relies on a finite difference scheme which, at each iteration, performs

$$\alpha_{i,j}^{n+1}=\alpha_{i,j}^{n}-\Delta t\mathcal{H}(i,j,\theta,\alpha^n) \tag{III.5.8}$$

where $\alpha_{i,j}^n$ are the coefficients of the level-set representation at each node of the grid and the Hamiltonian $\mathcal{H}$ is an approximation of $\theta|\nabla_{\mathbf{x}}\phi|$. When performing Eq. (III.5.8) on the level-sets ϕ_1,ϕ_2 corresponding to the lower (Ω_1) and upper (Ω_2) layer of the design, we can check at each node if the constraint $\phi_1<\phi_2$ is locally violated. If it is the case we need to modify the values of both level-sets at this node. Several choices are possible but a good heuristic is to consider the expected (first-order) improvement of $\mathcal{J}$ using only the vector field θ^k which is used to move Ω_k.

Let $\alpha_{i,j}^{k,n}$ be the coefficients of ϕ_k at the node position $\mathbf{x}_{i,j}$ and pseudo-time $n\Delta t$. We also define $\theta_{i,j}^k=\theta_k(\mathbf{x}_{i,j})$ where $\boldsymbol{\theta}_k=\mathbf{n}\theta_k$ is defined in Eq. (III.5.2). If we detect that $\alpha_{i,j}^{1,n+1}>\alpha_{i,j}^{2,n+1}$ then, depending on the value of $\theta_{i,j}^k$, we modify the coefficients as:

- if $|\theta_{i,j}^2|^2\geq|\theta_{i,j}^1|^2$ then $\alpha_{i,j}^{1,n+1}\leftarrow\alpha_{i,j}^{2,n+1}$,
- if $|\theta_{i,j}^2|^2<|\theta_{i,j}^1|^2$ then $\alpha_{i,j}^{2,n+1}\leftarrow\alpha_{i,j}^{1,n+1}$.

This algorithm works like a projection but this time it gives priority locally to the layer which is expected to increase most the objective function. Indeed, we can write the shape derivative $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ as

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta})=\sum_{k=1}^{2}\mathcal{J}_k'(\Omega_k)(\boldsymbol{\theta}_{|\Omega_k})\quad\text{where}\quad\mathcal{J}_k'(\Omega_k)(\boldsymbol{\theta}_{|\Omega_k})=\int_{\partial\widehat{\Omega}_k}\boldsymbol{\theta}_{|\Omega_k}\cdot\boldsymbol{\theta}_k\,\mathrm{d}s,$$

and $\boldsymbol{\theta}_k$ given by Eq. (III.5.2). If we now consider the following vector field:

$$\tilde{\boldsymbol{\theta}}_k^{i,j}(\mathbf{x})=\begin{cases}\delta_{\mathbf{x}_{i,j}}\boldsymbol{\theta}_k(\mathbf{x}) & \text{if } \mathbf{x}\in\Omega_k\\ 0 & \text{elsewhere}\end{cases} \tag{III.5.9}$$

with $\delta_{\mathbf{x}}$ the dirac distribution centered at $\mathbf{x}$, then $\mathcal{J}'(\Omega)(\tilde{\boldsymbol{\theta}}_k^{i,j})=|\theta_{i,j}^k|^2$. Note that this reasoning remains formal since the vector field in Eq. (III.5.9) is not an element of $W^{1,\infty}$ and therefore, in theory, $\mathcal{J}'(\Omega)(\tilde{\boldsymbol{\theta}}_k^{i,j})$ does not give the first order expected improvement of $\mathcal{J}(\Omega)$.

In the perspectives part of this manuscript we will also present a way to deal with the "limit case", that is when we consider an infinite number of layers ($n_\ell=\infty$). This is physically possible using a manufacturing process known as grayscale lithography.

Chapter IV

Multi-objective problems and robustness to environmental uncertainties

Summary — This chapter is devoted to the presentation of a method to deal with the simultaneous optimization of multiple objectives functions for the design of nanophotonic devices. An interesting application of this framework is related to the incorporation of a degree of robustness to the considered optimization problems. The method presented here has been published in the paper

> [Leb19a] N. Lebbe, C. Dapogny, E. Oudet, K. Hassan, and A. Gliere. "Robust shape and topology optimization of nanophotonic devices using the level set method". In: *Journal of Computational Physics* (2019). DOI: 10.1016/j.jcp.2019.06.057.

Section IV.1 opens the chapter with a presentation of the gradient sampling method to consider the optimization of multi-objective problems. Some of the numerical examples of the previous chapter are also formulated and tackled as multi-objectives problems and we show that using a "multi-objective point of view" allows to obtain results which are closer to what is physically expected from these components.

In Section IV.2 we adapt the previous gradient sampling method to take into account the uncertainties of some environmental parameters (such as the wavelength, the temperature, etc.) in the course of the optimization process. Our aim is to obtain nanophotonic devices which are robust to variations of these parameters.

The last section IV.3 concludes this chapter by showing how it is possible to take into consideration the geometrical uncertainties coming from the manufacturing process of nanophotonic devices. Several numerical examples are presented which reveal that our methodology indeed makes it possible to obtain components which are tolerant to these kind of geometric uncertainties.

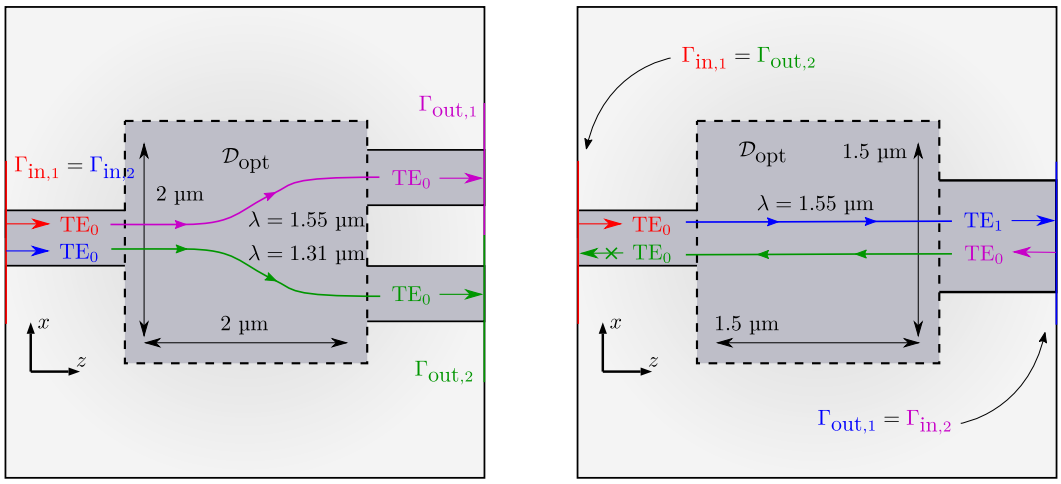
IV.1 The gradient sampling method for multi-objective problems

IV.1.1 Weighted sum of objective functions

In Section III.3.1 we have presented a variety of nanophotonic components but, still, we have not yet fully addressed the optimization of **multi-objective** components despite the fact that they are the most typical devices encountered in physical circuits. Among others, let us mention the cases of the diplexer and nanophotonic diode, as presented in Fig. IV.1.1 which both require the joint fulfillment of two objectives.

In the optimization of such devices we need to simultaneously consider at least two figures of merit. A first idea to solve a problem involving n_{obj} objective functions $\mathcal{J}_1(\Omega), \ldots, \mathcal{J}_{n_{\text{obj}}}(\Omega)$ is to gather all of them into a weighted sum $\mathcal{J}(\Omega)$:

$$\mathcal{J}(\Omega) = \sum_{i=1}^{n_{\text{obj}}} \alpha_i \mathcal{J}_i(\Omega). \tag{IV.1.1}$$



(a) *A diplexer in which the incident light, as a* TE_0 *mode, is redirected into one of the two output waveguides depending on the incoming wavelength*

(b) *A diode in which the incoming* TE_0 *mode may be converted into the* TE_1 *outgoing mode while preventing light coming from the output waveguide to come back*

Figure IV.1.1: Setting of two multi-objectives test-cases: a diplexer and a nanophotonic "diode".

One drawback of this formulation is that the coefficients α_i should be chosen appropriately so that the optimization algorithm does not end up in a local maxima which overly favors one of the objectives $\mathcal{J}_i(\Omega)$; see Fig. IV.1.2(a) in which $\alpha_1 = \alpha_2 = 1$ was chosen. In practice, the choice of adequate coefficients α_i, leading to a satisfactory optimization of all the involved objectives in the weighted sum is very much case dependent and requires fine tuning.

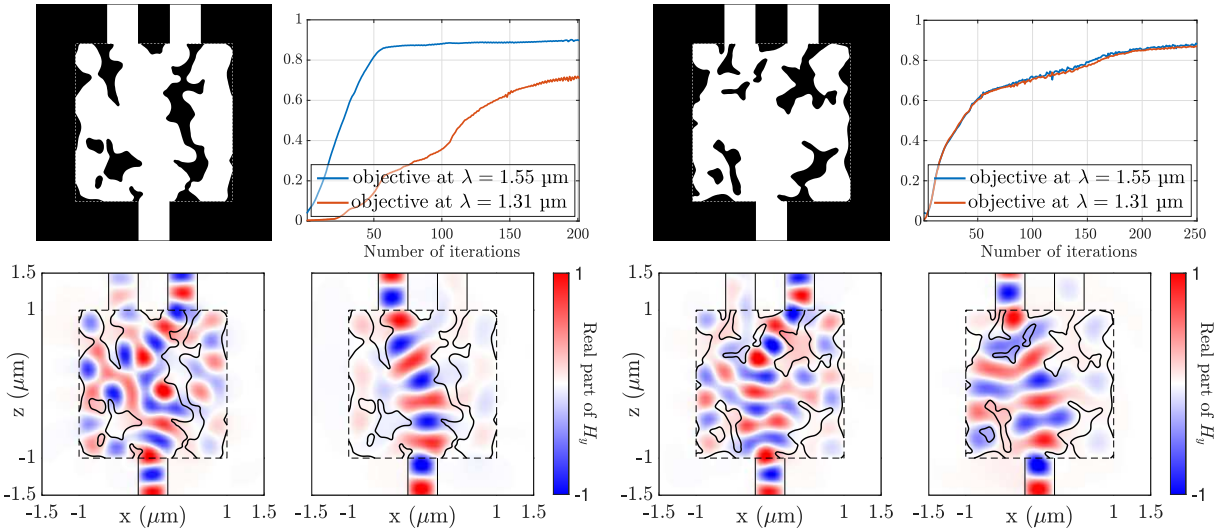
When dealing with the joint maximization of several objective functions $\mathcal{J}_i(\Omega)$, it is customary to consider the **Pareto front** $\mathcal{P}$ of the objectives. In our case this front is defined in the following way:

$$\mathcal{P} = \{\Omega^*, \forall \Omega, \exists i = 1, \ldots, n_{\text{obj}}, \mathcal{J}_i(\Omega^*) \geq \mathcal{J}_i(\Omega)\}.$$

In other words, $\Omega \subset \mathcal{P}$ if and only if there does not exists any shape which has strictly better performance for one of the objectives without being strictly worse for another one. Once the Pareto front is known, a shape Ω which establishes a compromise between each objective can be chosen. Of course, the precise determination of the set $\mathcal{P}$ is out of reach. In general, for nanophotonic components such as those depicted in Fig. IV.1.1 there is only one compromise of interest: we want all the objective functions to have approximately the same and largest possible values. That is to maximize

$$\mathcal{J}(\Omega) = \min_{i=1,\ldots,n_{\text{obj}}} \mathcal{J}_i(\Omega) \tag{IV.1.2}$$

instead of Eq. (IV.1.1). Note that Eq. (IV.1.2) makes sense since the $\mathcal{J}_i(\Omega)$ all represent the same physical quantity: a power transmission in the interval $[0, 1]$. If they do not correspond to the same quantities, it is necessary to normalize them in order to obtain comparable values.



(a) *Optimization of the sum of the two objectives ($\alpha_i = 1$ in Eq. (IV.1.1))*

(b) *Optimization of the minimum between the two objectives as in Eq. (IV.1.2)*

Figure IV.1.2: Optimization result of the diplexer test-case presented in Section IV.1.1 considering two different objective functions and starting from a shape made of 2×2 holes.

In Fig. IV.1.2 we compare the optimization of a diplexer device (see Fig. IV.1.1(a)) using the objective function Eq. (IV.1.1) with $\alpha_1 = \alpha_2 = 1$ (see Fig. IV.1.2(a)) or using the formulation of Eq. (IV.1.2). In the former case, the two objectives are not optimized at the same speed: after 200 iterations the first objective function $\mathcal{J}_1(\Omega)$ ends up with a transmission of almost 90 % while the second one $\mathcal{J}_2(\Omega)$ ended up at 70 %. On the contrary, using Eq. (IV.1.2), the optimized shape share similar performances for both figure of merits with $\mathcal{J}_1(\Omega) \simeq \mathcal{J}_2(\Omega) = 85$ %. This means that both objectives have indeed been optimized concurrently. Moreover, the final transmission of both $\mathcal{J}_i(\Omega)$ is an intermediate value between thus obtained using Eq. (IV.1.1). In the next section we explain how we manage to maximize the objective functional of Eq. (IV.1.2) even though it features a non-differentiable function due to the presence of the minimum operator.

IV.1.2 Algorithm for automatic coefficients adaptation

In the previous section, after the end of the optimization process in Fig. IV.1.2(a), we could increase the value of α_2 then start again the optimization in order to focus on the maximization of $\mathcal{J}_2(\Omega)$, at the expense of a decrease in $\mathcal{J}_1(\Omega)$. The modification of the coefficients α_i may be done by hand in a tedious trial and error procedure but it would be desirable to automate this process.

The method that we now discuss provides somehow, at each iteration, the "optimal" values of the coefficients α_i in order to maximize the minimum between all the objective functions $\mathcal{J}_i(\Omega)$, that is to maximize the program Eq. (IV.1.2). Our method is also very close to the Multi Gradient Descent Algorithm (MGDA) proposed in [Dés12].

Let us recall that our goal is to solve the following optimization problem

$$\max_{\Omega} \min_{i=1,\dots,n_{\text{obj}}} \mathcal{J}_i(\Omega). \tag{IV.1.3}$$

At a particular given shape Ω, we assume that the shape derivative $\mathcal{J}_i'(\Omega)(\boldsymbol{\theta})$ of each objective is known. In order to find an ascent direction $\boldsymbol{\theta}$ for Eq. (IV.1.3), we linearize each function $\mathcal{J}_i(\Omega)$ in the neighborhood of the actual shape Ω in Eq. (IV.1.3) so that it becomes:

$$\max_{\boldsymbol{\theta}} \min_{i=1,\dots,N} \mathcal{J}_i(\Omega) + \mathcal{J}_i'(\Omega)(\boldsymbol{\theta}), \tag{IV.1.4}$$

where $\boldsymbol{\theta}$ runs over the set of admissible perturbations in the practice of Hadamard's method. Note that, as it is common in the optimization field, Eq. (IV.1.4) may be reformulated with the addition of a dummy variable r into

$$\begin{cases} \max\limits_{\boldsymbol{\theta}, r \in \mathbb{R}} & r \\ \textbf{s.t.} & \min\limits_{i=1,\dots,n_{\text{obj}}} \mathcal{J}_i(\Omega) + \mathcal{J}_i'(\Omega)(\boldsymbol{\theta}) \geq r \end{cases}. \tag{IV.1.5}$$

The constraint is then equivalent to the following n_{obj} linear constraints:

$$\mathcal{J}_i(\Omega) + \mathcal{J}_i'(\Omega)(\boldsymbol{\theta}) > r, \text{ for } i = 1, \dots, n_{\text{obj}}.$$

Let us now introduce the shape gradients $V_{i,\Omega,\text{reg}}$ associated to the shape derivatives $\mathcal{J}_i'(\Omega)(\boldsymbol{\theta})$ via the extension and regularization process Eq. (II.4.5). We remind that $V_{i,\Omega,\text{reg}}$ is given as the only scalar field $V \in H^1(\mathcal{D}, \mathbb{R})$ solution of

$$a(V, \theta) = \int_{\mathcal{D}} \varepsilon \nabla V \cdot \nabla \theta + V\theta \, \mathrm{d}\mathbf{x} = \int_{\partial\Omega} V_{i,\Omega}\theta \, \mathrm{d}s \tag{IV.1.6}$$

for all $\theta \in H^1(\mathcal{D}, \mathbb{R})$ and $V_{i,\Omega}$ is defined through the shape derivatives as they are supposed to be of the form

$$\mathcal{J}_i'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta} \cdot \mathbf{n} V_{i,\Omega} \, \mathrm{d}s.$$

We then search for a solution $\boldsymbol{\theta} = \theta\mathbf{n}$ to Eq. (IV.1.4) whose amplitude θ is restricted to the convex hull $\text{conv}_{i=1,\dots,n_{\text{obj}}} \{V_{i,\Omega,\text{reg}}\}$ where

$$\text{conv}\{V_{i,\Omega,\text{reg}}\} := \left\{ \sum_{i=1}^{n_{\text{obj}}} \alpha_i V_{i,\Omega,\text{reg}}, \;\; 0 \leq \alpha_i \leq 1, \;\; \sum_{i=1}^{n_{\text{obj}}} \alpha_i = 1 \right\}. \tag{IV.1.7}$$

In other terms, the vector field $\boldsymbol{\theta}$ is sought under the form $\boldsymbol{\theta} = \theta\mathbf{n}$ with $\theta = \sum_{i=1}^{n_{\text{obj}}} \alpha_i V_{i,\Omega,\text{reg}}$, a convex combination of the scalar fields $V_{i,\Omega,\text{reg}}$. Assuming such a structure for $\boldsymbol{\theta}$, it comes:

$$\mathcal{J}_i'(\Omega)(\boldsymbol{\theta}) = a(\theta, V_{i,\Omega,\text{reg}}) = \sum_{j=1}^{n_{\text{obj}}} \alpha_j a(V_{i,\Omega,\text{reg}}, V_{j,\Omega,\text{reg}})$$

where $a(\cdot, \cdot)$ is the bilinear form defined in Eq. (IV.1.6). Using this choice of vector field $\boldsymbol{\theta}$, Eq. (IV.1.5) rewrites into

$$\begin{aligned} \max_{\boldsymbol{\alpha}, r} \quad & r \\ \textbf{s.t.} \quad & \boldsymbol{\alpha} \in [0,1]^{n_{\text{obj}}}, \; r \in \mathbb{R}, \\ & \sum_{i=1}^{n_{\text{obj}}} \alpha_i = 1, \\ & \mathcal{J}_i(\Omega) + \sum_{j=1}^{n_{\text{obj}}} \alpha_j a(V_{i,\Omega,\text{reg}}, V_{j,\Omega,\text{reg}}) \geq r, \text{ for } i = 1, \dots, n_{\text{obj}}. \end{aligned} \tag{IV.1.8}$$

Denoting by $P = (P_{i,j})_{i,j}$ the matrix with entries $P_{i,j} = a(V_{i,\Omega,\mathrm{reg}}, V_{j,\Omega,\mathrm{reg}})$ and by $v = (\mathcal{J}_i(\Omega))_i$ the vector of objective values, the program Eq. (IV.1.8) may be equivalently rewritten for $\mathtt{x} = (\boldsymbol{\alpha}, r)$ as

$$\begin{array}{rl} \max\limits_{\mathtt{x}\in\mathbb{R}^{n_{\mathrm{obj}}+1}} & \mathtt{c}^\top\mathtt{x} \\ \textbf{s.t.} & \mathtt{A}\mathtt{x} \leq \mathtt{b} \\ & \mathtt{A}_{\mathtt{eq}}\mathtt{x} = \mathtt{b}_{\mathtt{eq}} \\ & \mathtt{l} \leq \mathtt{x} \leq \mathtt{u} \end{array} \quad \text{where } \mathtt{c} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \ \mathtt{A} = \begin{pmatrix} & & 1 \\ & -P & \vdots \\ & & 1 \end{pmatrix}, \ \mathtt{A}_{\mathtt{eq}}^\top = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad \text{(IV.1.9)}$$

$\mathtt{b} = v$, $\mathtt{b}_{\mathtt{eq}} = 1$, $\mathtt{l}^\top = (0, \ldots, 0, -\infty)$ and $\mathtt{u}^\top = (1, \ldots, 1, +\infty)$. Equation (IV.1.9) is none other than a small linear program whose resolution may be carried out by using, for instance, the `linprog` function of `Matlab`.

Remark IV.1.2.1: Equation (IV.1.8) may be complemented with other linear constraints. For instance if $n_{\mathrm{obj}} = 2$ and the initial shape Ω_0 already provides good performances for one objective, for instance $\mathcal{J}_1 \simeq 1$ then it may be interesting to maximize $\mathcal{J}_2$ as much as possible without degrading too much the first objective by adding the linear constraint $\mathcal{J}_1(\Omega) + \sum_{j=1}^{2} \alpha_j a(V_{i,\Omega,\mathrm{reg}}, V_{j,\Omega,\mathrm{reg}}) \geq 0.9$ to Eq. (IV.1.8) in order to keep a first objective with transmission value above 0.9.

In other words, to solve Eq. (IV.1.4) we replace Line 11 to 13 in Algorithm II.4.1 into the following Algorithm IV.1.1. We refer to this process as a **gradient sampling** algorithm.

Algorithm IV.1.1: Gradient sampling algorithm to find a descent direction for the multi-objective topology optimization of Section IV.1.2: "gradient sampling".

1 $v := (\mathcal{J}_i(\Omega))_i$ the value of each objective (Eq. (II.2.1));
2 $V_i := V_{i,\Omega}$ the shape derivative of each objective (Eq. (II.2.9));
3 $V_{i,\Omega,\mathrm{reg}} :=$ solution of the regularization problem (Eq. (IV.1.6)) for each shape derivative;
4 $P := (P_{i,j})_{i,j}$ the matrix with entries the inner product between regularized gradients $V_{i,\Omega,\mathrm{reg}}$ (Eq. (IV.1.9));
5 $\alpha :=$ solution of the linear program (Eq. (IV.1.8)) involving v and P;
6 $V_{\mathrm{reg}} := \sum_{i=1}^{n_{\mathrm{obj}}} \alpha_i V_{i,\Omega,\mathrm{reg}}$;

Remark IV.1.2.2: If one wants to perform an additional post-processing on $V_{i,\mathrm{reg}}$ (see Section II.4.2.d) which is not encoded in the inner product (such as symmetries) then Line 4 in Algorithm IV.1.1 must be modified into

4a $\tilde{V}_{i,\Omega,\mathrm{reg}} :=$ optional other post-processing (see Section II.4.2.d);
4b $P := (P_{i,j})_{i,j}$ the scalar product (Eq. (II.4.5)) between $V_{i,\Omega,\mathrm{reg}}$ and $\tilde{V}_{j,\Omega,\mathrm{reg}}$;

and Line 6 becomes accordingly:

6 $V_{\mathrm{reg}} := \sum_{i=1}^{n_{\mathrm{obj}}} \alpha_i \tilde{V}_{i,\Omega,\mathrm{reg}}$;

It may happen that one of the shape gradients $V_{i,\Omega,\mathrm{reg}}$ has a much higher amplitude (L^∞ norm) than the others. In this case, considering a vector field whose amplitude is contained inside the convex hull of Eq. (IV.1.7) is not a wise choice.

Indeed, the shape gradient V_Ω associated with an objective function $\mathcal{J}(\Omega)$ is only optimal in the sense that there exists a step δ such that for any other scalar field W of the same amplitude, $\mathcal{J}((\mathrm{Id}+\delta W\mathbf{n})(\Omega)) \leq \mathcal{J}((\mathrm{Id}+\delta V_\Omega\mathbf{n})(\Omega))$. Since each $V_{i,\Omega}$ have different amplitudes, it may happen that for a fixed δ we have $\mathcal{J}_1((\mathrm{Id}+\delta V_2)(\Omega)) > \mathcal{J}_1((\mathrm{Id}+\delta V_1)(\Omega))$. Thus, solving Eq. (IV.1.8) could lead to a strange result where, to optimize $\mathcal{J}_1$, we should only consider $V_{2,\Omega}$.

To alleviate this issue we propose to normalize each shape gradient using the lowest L^∞ norm of the shape gradients by replacing Line 4 in Algorithm IV.1.1 by:

4a $V_{\mathrm{inf}} := \min_i(\|V_{i,\Omega,\mathrm{reg}}\|_\infty)$;
4b $V^*_{i,\Omega,\mathrm{reg}} := V_{i,\Omega,\mathrm{reg}} \times V_{\mathrm{inf}}/\|V_{i,\Omega,\mathrm{reg}}\|_\infty$;
4c $P := (P_{i,j})_{i,j}$ the scalar product (Eq. (II.4.5)) between each $V_{i,\Omega,\mathrm{reg}}$ and $V^*_{j,\Omega,\mathrm{reg}}$;

and Line 6 is replaced accordingly:

6 $V_{\mathrm{reg}} := \sum_{i=1}^{n_{\mathrm{obj}}} \alpha_i V^*_{i,\Omega,\mathrm{reg}}$;

Remark IV.1.2.3: Equation (IV.1.3) may be adapted to consider objective functions for which different values are requested. For instance, let us consider a non-symmetric version of the power divider depicted in Fig. III.3.5 where we aim that a fraction $\eta \in [0,1]$ of the total power be redirected in the $\Gamma_{\mathrm{out},1}$ port and $(1-\eta)$ in $\Gamma_{\mathrm{out},2}$. To obtain such a component we may consider the following objective function

$$\mathcal{J}(\Omega) = \min\{\mathcal{J}_1(\Omega)/\eta, \mathcal{J}_2(\Omega)/(1-\eta)\},$$

which, once maximized, should give the desired transmissions for both outputs (up to the same factor for each $\mathcal{J}_i$). More generally, it is also possible to prescribe a particular target value $\mathcal{J}_i^*$ for each objective by considering

$$\mathcal{J}(\Omega) = \min_{i=1,\dots,n_{\mathrm{obj}}} |\mathcal{J}_i(\Omega) - \mathcal{J}_i^*|^2,$$

which results in a modification of the considered shape derivatives in Eq. (IV.1.8).

IV.1.3 Numerical examples

In this section we show two numerical examples using the gradient sampling method introduced in the previous section. Note that in the optimization of the diplexer of Fig. IV.1.2(b) we already relied on this method to maximize both outputs at the same time.

IV.1.3.a Design of a power divider from 1 to n ports

This first example revisits the power divider example of Section III.3.1.b by considering a larger number of output waveguides. Even using symmetric vector fields as in Eq. (III.3.3), the design of such devices requires to consider multiple objectives at the same time. In Fig. IV.1.4 we optimize a 1 to 3 and a 1 to 4 power dividers starting with a shape as in Fig. III.3.5(b) and an optimization domain $\mathcal{D}_{\mathrm{opt}}$ of respectively 3×3 µm and 4×4 µm; see Fig. IV.1.3.

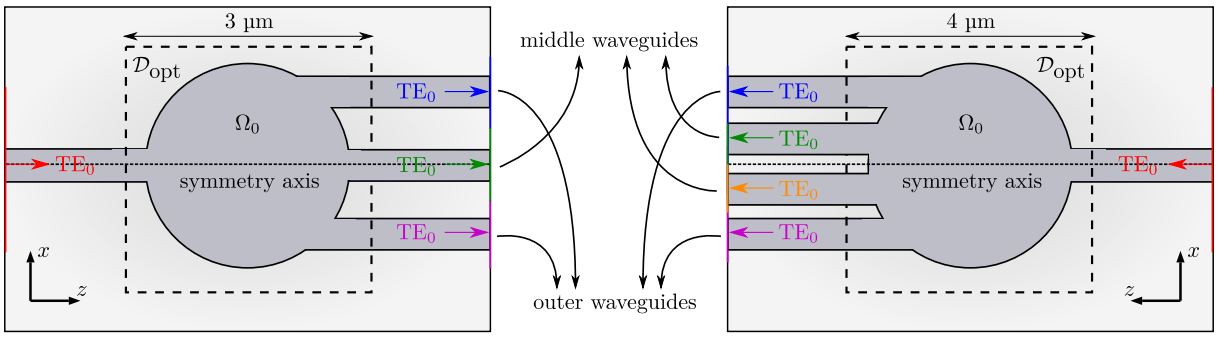
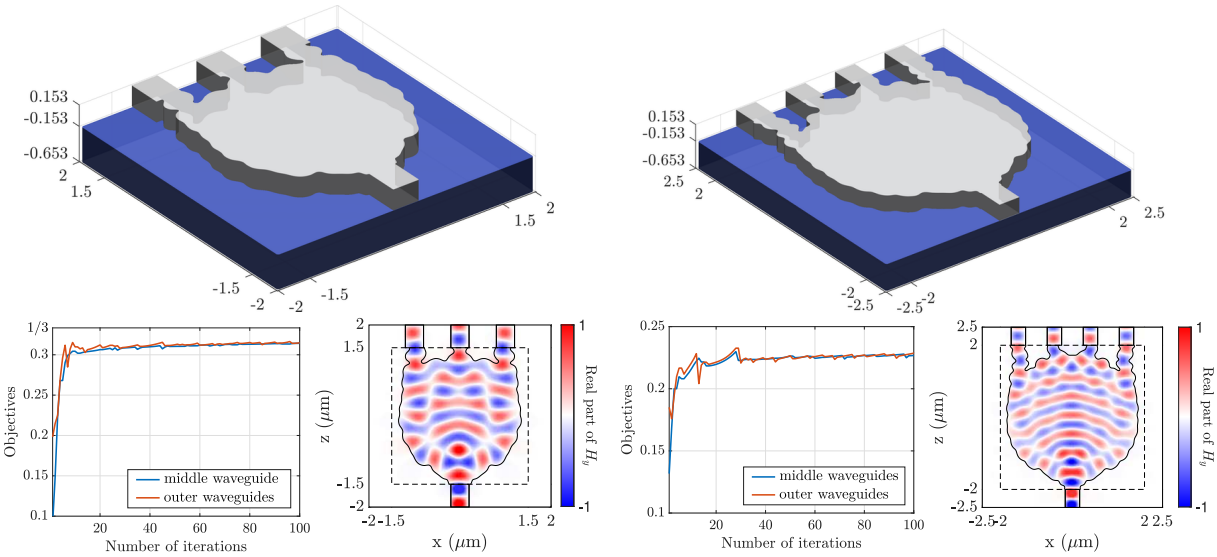


Figure IV.1.3: Settings of the 1 to 3 and 1 to 4 power divider test-cases of Section IV.1.3.a.

Due to the symmetry along z-axis we only have to consider two objective functions: the transmission in one of the central waveguides and one of the outer ones. As we can see on the convergence history graph, both objectives are maximized simultaneously. The algorithm is stopped after 100 iterations and results in a 1 to 3 power divider with 31 % transmission in each output waveguide (6 % of the total power is dissipated) and 22.5 % in the case of the 1 to 4 power divider (10 % of the total power is dissipated).



(a) *Optimization of a* 1 *to* 3 *power divider*

(b) *Optimization of a* 1 *to* 4 *power divider*

Figure IV.1.4: Optimization result of the power dividers test-cases presented in Section IV.1.3.a considering the initial shapes in Fig. IV.1.3.

Note also that non-symmetric versions of these power dividers may be obtained in the spirit of Remark IV.1.2.3.

IV.1.3.b Diode

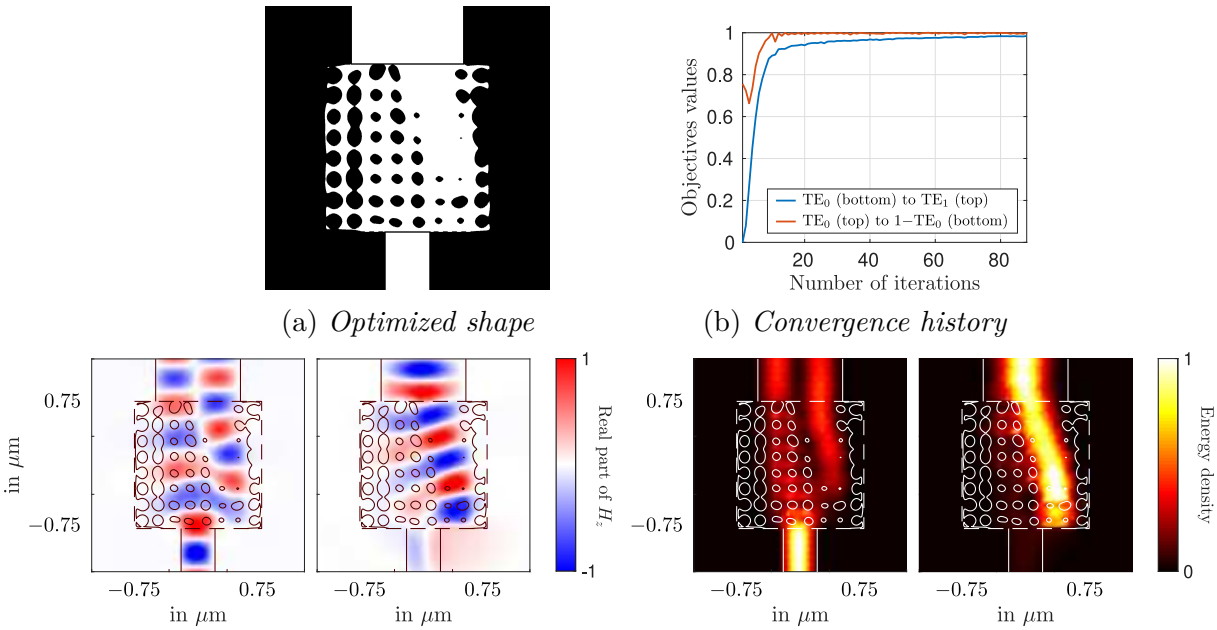
This second example is based on the test case of [Cal16] and it is illustrated in Fig. IV.1.1(b). This component is connected to two waveguides with different sizes: the bottom waveguide linked to the device via the port Γ_{bottom} is single-mode while the top one whose junction with the design domain is denoted by Γ_{top} allows the propagation of both a TE_0 and TE_1 mode. The two objective functions of this device is as follows:

- When the light is injected from Γ_{bottom} as the fundamental TE_0 mode we want to convert it into the TE_1 mode of the top waveguide.
- When injecting the TE_0 mode in Γ_{top} we want to prevent its transmission into the bottom waveguide.

To achieve these two objectives we consider here the maximization of

$$\mathcal{J}(\Omega) = \min\{\mathcal{J}_1(\Omega), 1 - \mathcal{J}_2(\Omega)\},$$

where $\mathcal{J}_1(\Omega)$ is the power carried by the TE_1 mode in Γ_{top} when the TE_0 mode is injected though Γ_{bottom} and $\mathcal{J}_2(\Omega)$ the power carried by the TE_0 mode in Γ_{bottom} when the TE_0 mode is injected from the top waveguide. Using the gradient sampling algorithm of Section IV.1.2 and starting with a shape made of 8×8 holes we end up with the result of Fig. IV.1.5 where we can see in the plot of the energy density that no power is transmitted into the fundamental mode of the bottom waveguide when injecting light from the top one.



(a) *Optimized shape* (b) *Convergence history*

(c) *Electric field and energy density in the section* $y = 0$

Figure IV.1.5: Optimized shape of the diode starting from 8×8 holes and details of the numerical computation.

IV.2 Dealing with robustness to the fluctuation of physical parameters

The physical and geometrical properties of the considered nanophotonic components and their environment are characterized by data (the incoming wavelength, the refractive indices of the media at play, or the morphology of shapes itself, to name a few) which are in practice known with some uncertainty. The electromagnetic fields around nanophotonic devices, and thereby their physical behaviors, being very sensitive to these data, it is of utmost importance to optimize their shapes in such a way that their performances are

robust with respect to such uncertainties, i.e. so that they retain an acceptable efficiency in a variety of fabrication or operating conditions.

Optimization of shapes in a way which is robust to uncertainties has been a burning issue in shape and topology optimization lately. In the recent review [Mol18], manufacturing uncertainties have been identified as the main obstacle preventing the use of nanophotonic components at the industrial level, and studies towards alleviating these problems have been initiated [Wan11; Ele12]. Beyond the field of nanophotonics, robustness issues in shape and topology optimization have been addressed from two fairly different viewpoints:

- When no information is available about the uncertain data but for a bound on their maximum amplitude, the worst value of the performance criterion under all possible uncertain data is optimized. These problems are generally ways too difficult to be dealt with in utter generality, since their treatment inherently involves a bilevel optimization program; yet several particular situations or approximations have provided quite satisfactory answers [All14a; Ams16]. The main drawback of such approaches is that they are generally too pessimistic: while the worst-case scenario is likely never to happen in practice, the specific optimization of this situation may conduct to shapes with poor nominal performance.

- When more information is available about the statistics of the uncertain data (e.g. about its first- and second-order moments), probabilistic approaches may be considered for the minimization of the average value or the standard deviation of the performance criterion; see [Mau14] for an overview. These approaches generally rely on very costly sampling strategies, such as Monte-Carlo, or collocation methods, involving a large number of evaluations of the considered cost function and its derivative; see for instance [Mar16] and the references therein. Linearized approximations of such problems have been proposed [Laz12; All15] which alleviate the dramatic computational cost of the aforementioned approaches.

In this section, we rely on a simple sampling strategy for the robust worst-case optimization when small uncertainties are expected; this method is particularly well-suited in situations where the uncertain data lie in a low-dimensional space. Our approach is guided by the large CPU cost of the numerical resolution of the time-harmonic vector wave equation, which makes methods involving a large number of evaluations of the objective function and its derivative totally impractical in our context. The general principle of the method is presented in Section IV.2.1 in an abstract and formal way. Its particular application to deal with robustness with respect to the incoming wavelength and to the geometry of shapes themselves are discussed in Sections IV.2.2 and IV.3, respectively.

IV.2.1 A general framework for dealing with robustness using a gradient sampling strategy

Our uncertain data are modelled by a (small) parameter δ lying in a set X. In practice we assume that

$$X \text{ is a ball with small radius } m > 0 \text{ in a low-dimensional vector space.} \tag{IV.2.1}$$

Denoting by $\mathcal{J}_\delta(\Omega)$ the value of the considered objective functional when the physical data δ are observed, our purpose is to maximize the worst value of $\mathcal{J}_\delta(\Omega)$ when δ runs

through X:

$$\max_{\Omega} \min_{\delta \in X} \mathcal{J}_\delta(\Omega). \tag{IV.2.2}$$

Taking advantage of the hypothesis Eq. (IV.2.1), the previous problem is consistently approximated by

$$\max_{\Omega} \min_{i=1,\dots,N} \mathcal{J}_i(\Omega), \quad \text{where } \mathcal{J}_i(\Omega) := \mathcal{J}_{\delta_i}(\Omega), \tag{IV.2.3}$$

and the δ_i, $i = 1, ..., N$ constitute a suitable sampling of X. Hence, the problem Eq. (IV.2.2) is reformulated as that Eq. (IV.2.3) of maximizing the minimum value between a finite number of objective functions, which falls exactly into the context of application of the gradient sampling method of Section IV.1.2.

In other words, to find a design Ω robust to some uncertain parameter δ, we only need at each iteration of our numerical Algorithm IV.1.1:

1. Perform the simulation of the time-harmonic vector wave equation on Ω by using different values $\delta_1, \dots, \delta_N$ of the uncertain parameters δ,
2. Compute the value of each objective function $\mathcal{J}_{\delta_i}(\Omega)$ as well as their respective shape derivatives (through the computation of the associated adjoint states),
3. Find a vector field $\boldsymbol{\theta}$ which maximizes the minimum between each first order approximation of the objective $\mathcal{J}_\delta(\Omega)$, i.e, which solves the program Eq. (IV.2.3).

Remark IV.2.1.1: The above strategy is solely based on a sampling $\mathcal{J}_i(\Omega) = \mathcal{J}_{\delta_i}(\Omega)$ of the perturbed functional $\mathcal{J}_\delta(\Omega)$ at particular values $\delta = \delta_i$, $i = 1, ..., N$, and on the derivative of the sampled functionals $\Omega \mapsto \mathcal{J}_i(\Omega)$. In particular, it does not involve the sensitivity of the objective function with respect to the perturbations, that is, the derivative of the mapping $\delta \mapsto \mathcal{J}_\delta(\Omega)$, which is a noticeable difference with the linearization method proposed in [All14a; All15].

IV.2.2 Robust design with respect to the incoming wavelength

Perhaps one of the most crucial aspects where robustness is desired in nanophotonics is related to the wavelength λ of the light injected into the component. Aiming at a performance which is little altered by small variations of the incoming wavelength is indeed a way to cope with the inaccuracy of the laser realizing the light injection, or simply to construct large bandwidth devices.

Using the notations of Section IV.2.1 the considered set X of perturbations is the interval $[\overline{\lambda} - m, \overline{\lambda} + m] \subset \mathbb{R}$, where $\overline{\lambda}$ is the ideal operating wavelength, and $m > 0$ is a user-defined tolerance for the range of wavelengths where the optimized design should retain good performances. Let us denote by $\mathcal{J}_\lambda(\Omega)$ the value of the considered objective function at a particular shape Ω when the operating wavelength equals λ; notice that λ influences three parameters of the physical model Eq. (III.1.1):

- the wavenumber $k = 2\pi/\lambda$,
- the optical indices of the silicon core and silica substrate depends on the wavelength as can be seen in Fig. IV.2.1,

- the optical modes $(\boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{H}})$, computed as the eigenvectors of Eq. (I.2.2), also depend on the values of the wavenumber and the optical indices; they are therefore modified by a wavelength shift (a dependence which is highly non-linear). From a mathematical viewpoint, this is reflected by the fact that the operator γ and the value $\mathbf{U}$ in the boundary condition Eq. (I.3.14) also depends on λ.

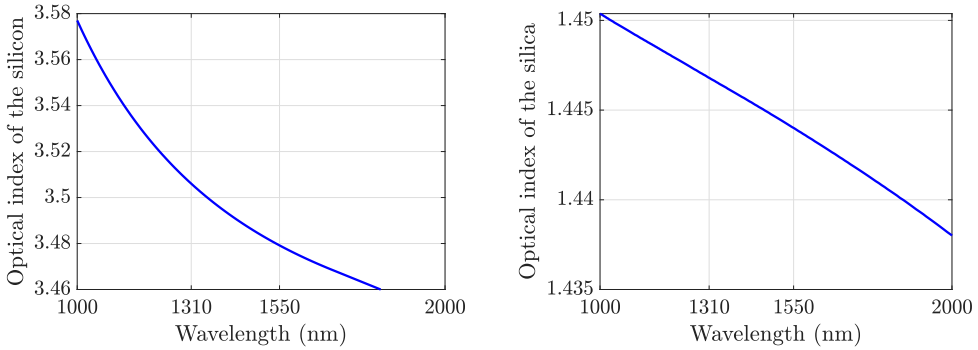


Figure IV.2.1: Dependence of the silicon and silica to the wavelength (ellipsometric measures made at the CEA).

The worst-case shape optimization problem reads, when uncertainties of amplitude m around the value $\overline{\lambda}$ are expected about the wavelength:

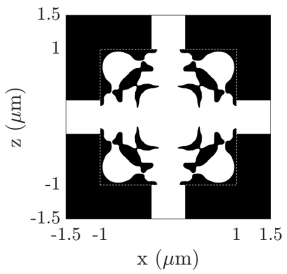
$$\max_{\Omega} \min_{\lambda \in X} \mathcal{J}_\lambda(\Omega). \tag{IV.2.4}$$

Following Section IV.2.1, the set X is sampled as $\{\lambda_i\}_{i=1,\dots,N}$, and Eq. (IV.2.4) readily boils down to a program of the type Eq. (IV.1.8), which is solved thanks to the methodology described in Section IV.2.1.

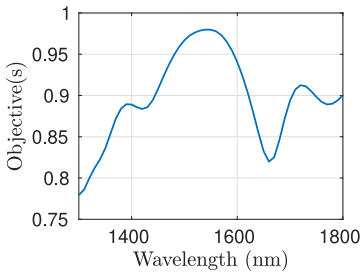
IV.2.2.a Robust crossing

Let us start our numerical analysis with the determination of a wavelength-robust crossing device (see Section III.3.1.d for the presentation of this component). We first computed an optimized crossing for the operating wavelength $\lambda = 1.55\,\mu\text{m}$ without taking into account robustness with respect to λ. Starting from a shape with 2×2 holes it produces the results in Figs. IV.2.2(a) to IV.2.2(c) where Fig. IV.2.2(b) shows that this component is not particularly robust in the (relatively large) wavelength band $[1.35, 1.75]$ µm: indeed, its performance drops from nearly 98 % at $\lambda = 1.55\,\mu\text{m}$ to 83 % at $\lambda = 1.65\,\mu\text{m}$ and even to less than 80 % at $\lambda = 1.35\,\mu\text{m}$.

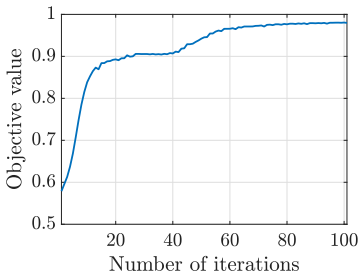
To enhance the robustness of this component with respect to uncertainties over the operating wavelength λ we rely on the multi-objective optimization presented in Section IV.2.1 by considering 3 wavelengths at respectively 1.45, 1.55 and 1.65 µm. Starting from the previously optimized crossing as an initial shape this leads to the results of Fig. IV.2.2. This new optimized component is noticeably more robust, exhibiting a transmission greater than 95 % in the considered bandwidth $[1.45, 1.65]$ µm.
We eventually turn our attention to the optimization of a robust crossing on an even broader bandwidth considering 5 different wavelengths at $1.35, 1.45, 1.55, 1.65$ and $1.75\,\mu\text{m}$ starting from the same non-robust shape and we end up after 100 iterations of our gradient sampling algorithm with the results of Fig. IV.2.3 where 90 % of transmission is expected in the whole bandwidth.



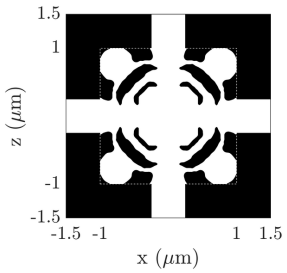
(a) *Non robust shape* Ω^*



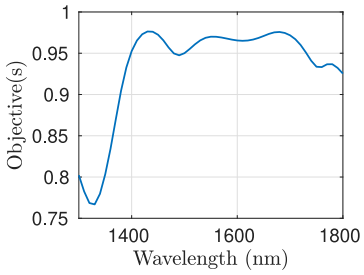
(b) *Sensitivity to wavelength*



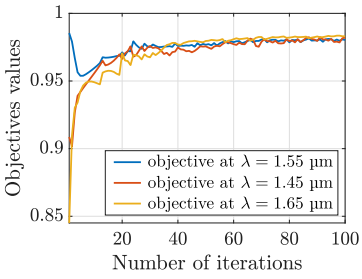
(c) *Convergence history*



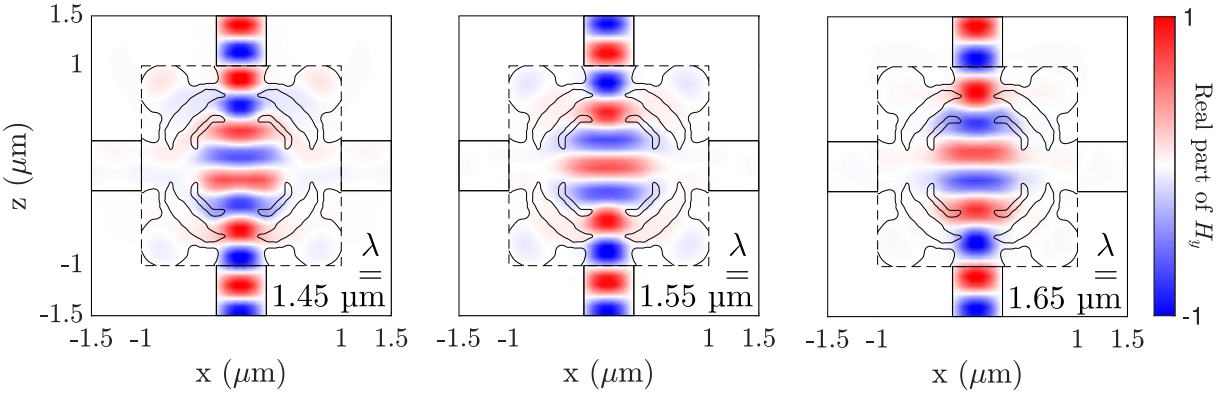
(d) *Robust optimized shape*



(e) *Sensitivity to wavelength*

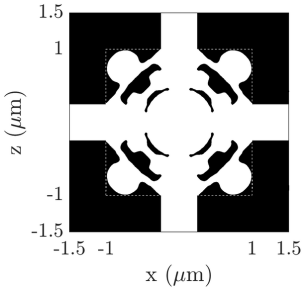


(f) *Convergence history*

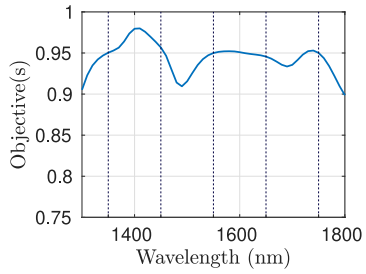


(g) *Electric field in the section* $y = 0$ *for different wavelengths*

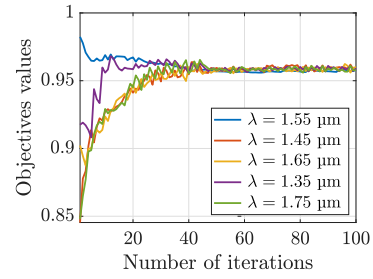
Figure IV.2.2: Results of the robust shape optimization of a crossing with respect to uncertainties over the wavelength.



(a) *Optimized shape*



(b) *Sensitivity to wavelength*



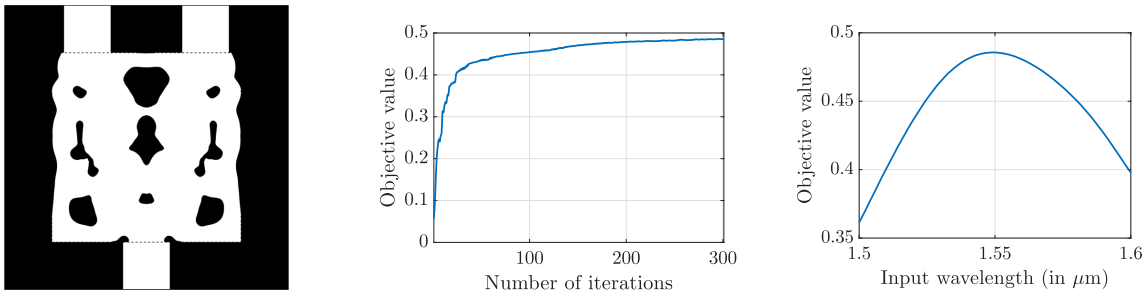
(c) *Convergence history*

Figure IV.2.3: Results of the robust shape optimization of a crossing with respect to uncertainties over the operating wavelength using a sampling of the interval $\lambda \in [1.35, 1.75]$ μm at five equally spaced wavelengths (dashed lines in Fig. IV.2.3(b)).

IV.2.2.b Robust power divider

In this subsection we revisit the power divider example of Section III.3.1.b with the aim to obtain a broadband component. The power divider obtained in Fig. III.3.6 being geo-

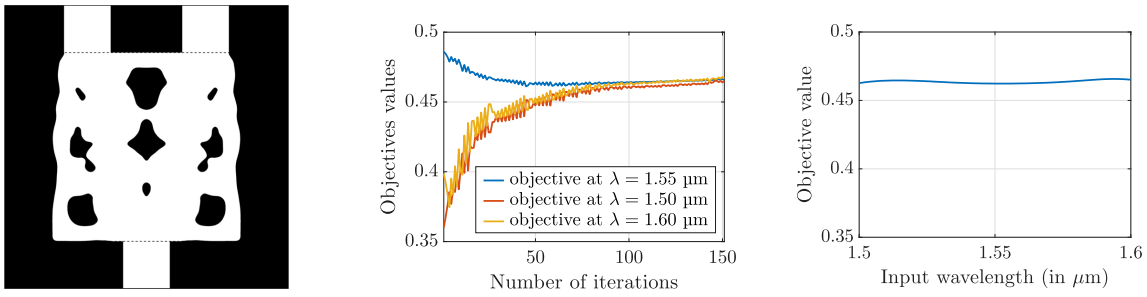
metrically simple, it is, understandably enough, also reasonably robust according to small variation of the input wavelength. To demonstrate the efficiency of our method we therefore consider another optimized power divider for the operating wavelength $\lambda = 1.55\,\mu m$ without considering robustness issues and starting from a shape composed of 3 by 3 holes. The result is represented in Fig. IV.2.4 and the graph in Fig. IV.2.4(c) reveals that this component is very sensitive to small variations of the wavelength.



(a) *Optimized shape* (b) *Convergence history* (c) *Spectrum*

Figure IV.2.4: Optimized shape of the power divider of Section III.3.1.b starting from 3×3 holes, details of the numerical computation and spectrum.

To obtain a robust component in the bandwidth $X = [1.5, 1.6]$ µm we now consider the problem Eq. (IV.2.4) with the uncertainty set X sampled at $\lambda = 1.5, 1.55$ and $1.6\,\mu m$. Starting from the optimized shape of Fig. IV.2.4, we perform 150 iterations of our gradient sampling algorithm which results with the component of Fig. IV.2.5.



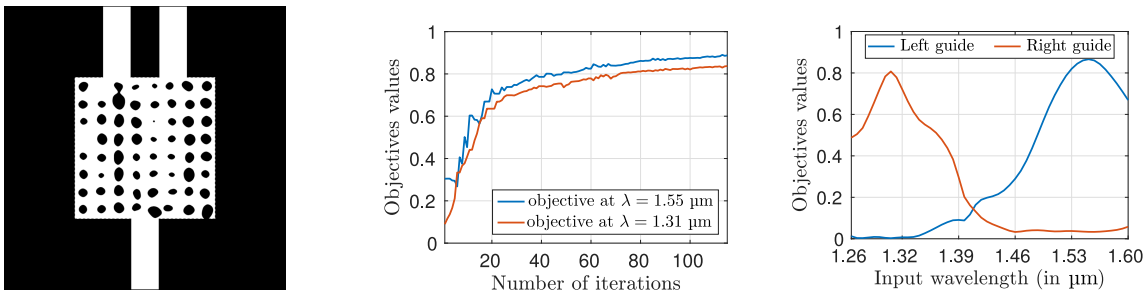
(a) *Optimized shape* (b) *Convergence history* (c) *Spectrum*

Figure IV.2.5: Optimized shape of the robust power divider starting from Fig. IV.2.4(a), details of the numerical computation and spectrum.

Comparing Figs. IV.2.4(c) and IV.2.5(c) shows that the new component is indeed far more robust than the previous one, exhibiting more than 46 % of transmission on the whole considered bandwidth whereas the non-robust device achieved less than 37 % of transmission at $\lambda = 1.5\,\mu m$.

IV.2.2.c Robust diplexer

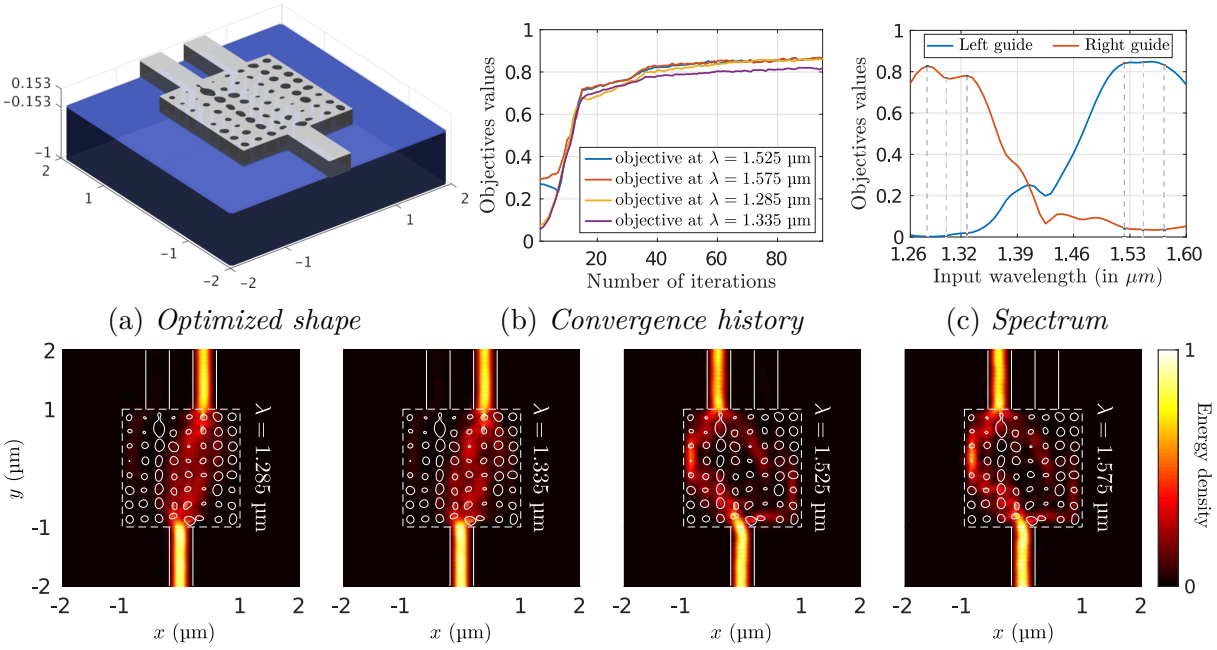
Our last example reconsiders the diplexer test-case of Fig. IV.1.1(a). A non-robust component is found by starting with an initial shape made of 8×8 holes and the result is depicted in Fig. IV.2.6.
As can be seen on its spectrum (Fig. IV.2.6(c)), more than 80 % of the input power is well transmitted into the fundamental TE mode of the top left waveguide at $\lambda = 1.31\,\mu m$ and on the top right waveguide at $\lambda = 1.55\,\mu m$. But as for the two previous examples we



(a) *Optimized shape* (b) *Convergence history* (c) *Spectrum*

Figure IV.2.6: Optimized shape of the diplexer of Fig. IV.1.1(a) starting from 8×8 holes, details of the numerical computation and spectrum.

see that a small perturbation of the wavelength dramatically jeopardize the performance of the component. For instance, considering the top right waveguide (in which we seek to redirect the light at $\lambda = 1.31\,\mu\text{m}$) an increase of 25 nm in the incident wavelength implies a decrease of the transmission of about 20 %.



(a) *Optimized shape* (b) *Convergence history* (c) *Spectrum*

(d) *Cut of the energy density at $y = 0$ for different wavelengths*

Figure IV.2.7: Optimized shape of the robust diplexer starting from 8×8 holes, details of the numerical computation and spectrum.

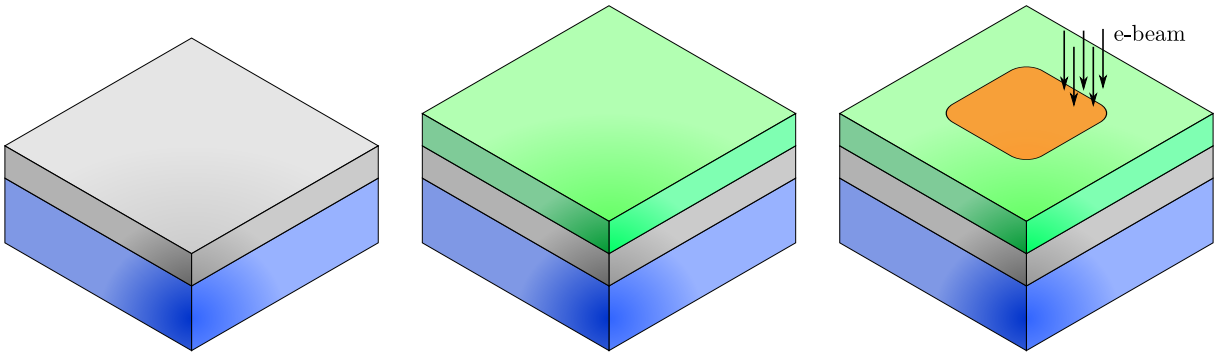
Starting from the same initial shape (8×8 holes) we perform 100 iterations of our multi-objective optimization algorithm. In this experiment we considered four objective functions to impose a 50 nm robustness: $\lambda = 1.285$ and $1.335\,\mu\text{m}$ for the top right waveguide while $\lambda = 1.525$ and $1.575\,\mu\text{m}$ was used for the top left one. The results are shown in Fig. IV.2.7. Comparing Figs. IV.2.6(c) and IV.2.7(c), we observe that the new component is way more robust, its spectrum exhibiting flat areas around the two wavelengths of interest.

IV.3 Geometrical uncertainties in lithography and etching processes

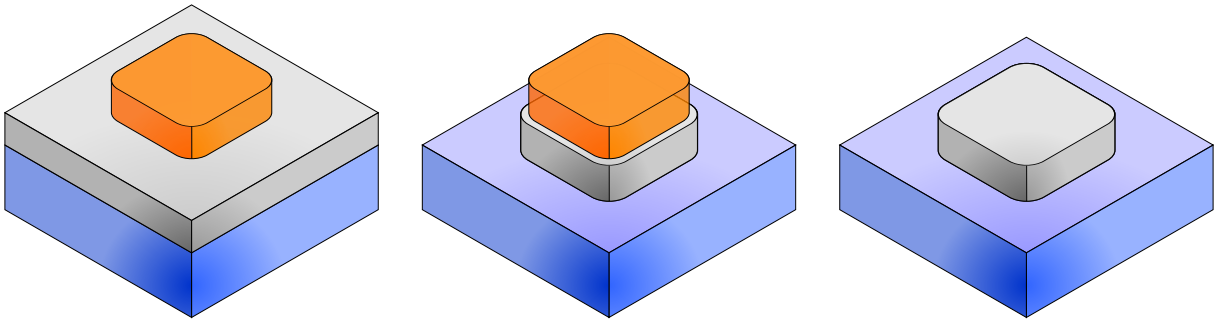
We now illustrate how the general framework of Section IV.2.1 may be adapted to impose the robustness of the optimized designs Ω with respect to uncertainties on their geometry. After a short presentation of the lithography and etching manufacturing process in Section IV.3.1 we present two ways to tackle the uncertainties coming from these production techniques. In Section IV.3.2 we first describe an a-posteriori method to find a shape which, once modified by the lithography process allows to recover the original shape that we wanted to produce. Section IV.3.3.a then propose an a-priori method to take into account the uncertainties related to the etching fabrication process during the topology optimization algorithm. An extension of this idea is used in Section IV.3.3.b to deal with uncertainties caused by lithography.

IV.3.1 Presentation of the main steps in silicon on insulator/wafer fabrication

The lithography-etching method to produce nanophotonic devices may be summarized into five main steps, as illustrated in Fig. IV.3.1. The production of such a component is achieved by etching a silicon wafer by means of some chemicals (reactive ion etching); see Fig. IV.3.1(e). Since this process does not allow for a precise engraving of the silicon plate locally, we start by making a **stencil** of the shape that we want to fabricate using a resist whose purpose is to prevent chemicals from engraving underneath it.



(a) *Silicon on insulator wafer* (b) *Resist thin film coating* (c) *e-beam* ***lithography***



(d) *Development* (e) *Reactive ion* ***etching*** (f) *Resist stripping*

Figure IV.3.1: Main steps of the lithography-etching manufacturing process. Silica is represented in blue, silicon is in gray, resist in green and exposed resist in orange.

This stencil is then produced by lithography; using deep UV or an **electron beam (e-beam)**, the resist is locally insolated by a high intensity field. After a sufficient exposition of the resist to the energy field, the resist will eventually change its physical states as can be seen in the orange part of Fig. IV.3.1(c). A development step (Fig. IV.3.1(d)) then makes it possible to remove the areas of the resist that have not been insolated and to recover the stencil. The final design is then obtained though reacting ion etching (Fig. IV.3.1(e)), i.e. by applying a chemical product on the plate that will etch the regions not covered by the stencil. The last step in Fig. IV.3.1(f) is dedicated to the removal of the resist.

In Section IV.3.1.a and Section IV.3.1.b we propose a mathematical model of the lithography and etching steps in the perspective of describing the type of geometrical uncertainties entailed by their action on the manufactured shape.

IV.3.1.a A mathematical model for lithography

Lithography involves many different physical processes. The model presented here results from [Fri89] and discussions with engineers from the CEA. Even though more precise model may be found in the literature (see for instance [Zho14]), the one proposed here catches the most important elements of the lithography process.

As it has been mentioned in the introduction, the lithography stage of the manufacturing process of nanophotonic devices relies on the alteration of the physical state of a resist caused by exposition to an e-beam. In order to obtain a stencil of the 2d section of a shape $\widehat{\Omega}$, the e-beam travels over the resist and we denote by (x_l, z_l) the position of the beam in the 2d plane corresponding to a section of the nanophotonic device. Whenever $(x_l, z_l) \in \widehat{\Omega}$, the e-beam is activated and it emits a Gaussian energy flux towards the position (x_l, z_l) on the resist. Mathematically, the **quantity of energy** E_{litho} received by the resist at each position $\boldsymbol{\xi} \in \widehat{\mathcal{D}}$ is therefore given by convolution product between $\mathbf{1}_{\widehat{\Omega}}$, the characteristic function of $\widehat{\Omega}$, and the energy flux $\mathcal{G}_\sigma$ which we assume to be a Gaussian kernel with zero mean and standard deviation σ:

$$\forall \boldsymbol{\xi} \in \widehat{\mathcal{D}},\ \ \mathcal{G}_\sigma(\boldsymbol{\xi}) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\boldsymbol{\xi}|^2}{2\sigma^2}}. \tag{IV.3.1}$$

The quantity of energy E_{litho} is defined by

$$E_{\text{litho}}(\boldsymbol{\xi}) = (\mathbf{1}_{\widehat{\Omega}} * \mathcal{G}_\sigma)(\boldsymbol{\xi}) = \int_{\widehat{\Omega}} \mathcal{G}_\sigma(\mathbf{x} - \boldsymbol{\xi})\, \mathrm{d}\mathbf{x}. \tag{IV.3.2}$$

To change the state of the resist at $\boldsymbol{\xi} \in \mathbb{R}^2$, a sufficient quantity of energy must be supplied at $\boldsymbol{\xi}$. Up to a dimensional constant, we shall assume that this quantity is equal to $1/2$, so that the shape of the stencil resulting from the lithography process is

$$\widehat{\Omega}_{\text{litho}} = \{\boldsymbol{\xi} \in \widehat{\mathcal{D}}, E_{\text{litho}}(\boldsymbol{\xi}) > 1/2\}. \tag{IV.3.3}$$

We will also denote by

$$L(\Omega) = \mathbf{1}_{[1/2,\infty[} \circ C(\Omega) \tag{IV.3.4}$$

the indicator function associated with the manufactured shape $\widehat{\Omega}_{\text{litho}}$ of Ω after the lithography process where $C : \Omega \mapsto \mathbf{1}_{\widehat{\Omega}} * \mathcal{G}_\sigma$. Thereafter we consider that the only possible source of uncertainties lying in the lithography process is the standard deviation parameter σ which is solely limited to belong in an interval of the form

$$[\bar{\sigma} - m, \bar{\sigma} + m], \tag{IV.3.5}$$

where $m > 0$.

IV.3.1.b Etching

Once a stencil is obtained via lithography, the etching step is used to produce the desired silicon shape Ω. In ideal conditions, the resulting shape from etching is exactly that of Eq. (IV.3.3) following the lithography stage. Unfortunately, in practice, the density of holes to be etched impacts the uniform distribution of the chemicals. This difficulty result in what is called **under- or over- etching** defects; the manufactured 2d section of the shape $\widehat{\Omega}_{\text{etch}}$ is a little bit dilated or eroded when compared to $\widehat{\Omega}_{\text{litho}}$:

$$\widehat{\Omega}_{\text{etch}} = (\text{Id} + \delta\mathbf{n})(\widehat{\Omega}_{\text{litho}}) \tag{IV.3.6}$$

with δ an uncertain scalar value; see [Sig09; Wan11] for the use of a similar model in the framework of density-based topology optimization and Fig. IV.3.3 for an illustration.

Remark IV.3.1.1: Depending on the particular etching technology, the manufactured shape $\widehat{\Omega}_{\text{etch}}$ may adopt a more complicated structure than that of Eq. (IV.3.6) featuring a constant parameter δ. For instance, in some situations, δ could be a function of the depth y; see the survey [Jan96].

Remark IV.3.1.2: It is sometime argued that if a nanophotonic component is wavelength-robust then it is naturally robust with respect to a dilation or erosion caused by the etching process. We propose here an attempt to justify this statement. The reasoning is as follows: let us consider that the shape Ω is modified into $F_+(\Omega)$ obtained from the mapping

$$F_+(\mathbf{x}) = (\text{Id} + m)(\mathbf{x}) = \mathbf{x} + m\mathbf{x} \tag{IV.3.7}$$

where $m \in \mathbb{R}$ and $\mathbf{x}_0 \in \mathbb{R}^2$. Let $\mathbf{E}$ be the solution to the time-harmonic vector wave equation $\nabla \times \nabla \times \mathbf{E} - k^2 n_\Omega^2 \mathbf{E} = 0$ using an optical index given by a shape Ω and a wavelength λ (we also remind here that the wavenumber is given by $k = 2\pi/\lambda$). Since $n_{F_+(\Omega)} = n_\Omega \circ F_+$, the solution $\mathbf{E}_+$ of the Maxwell equations associated to a shape $F_+(\Omega)$ and a wavenumber equal to $(1+m)k$ satisfies:

$$\nabla \times \nabla \times \mathbf{E}_+ - ((1+m)k)^2 n_\Omega^2 \circ F_+ \mathbf{E}_+ = 0. \tag{IV.3.8}$$

Now composing Eq. (IV.3.8) with F_+^{-1} we obtain

$$\nabla \times \nabla \times (\mathbf{E}_+ \circ F_+^{-1}) - k^2 n_\Omega^2 (\mathbf{E}_+ \circ F_+^{-1}) = 0,$$

and by uniqueness of a solution we get $\mathbf{E}_+ \circ F_+^{-1} = \mathbf{E}$. In other words, for a small value m, the performance of a nanophotonic component at the wavelength λ is exactly the same at $\lambda/(1+m)$ if the shape is modified into $(\text{Id} + m)(\Omega)$ (as long as n_Ω does not depend on λ). Equivalently, we could say that being robust to a small modification of the shape into $(\text{Id} + \delta)(\Omega)$ with $\delta \in [-m, m]$ is equivalent to robustness with respect to the wavelength in the interval $[\lambda/(1-m), \lambda/(1+m)]$.

Despite this result, the opening statement of this remark is not true (even as a first approximation). First of all, Eq. (IV.3.7) is not a dilation of the form $(\text{Id} + m\mathbf{n})(\Omega)$ but a **magnification**: $(\text{Id} + m)(\Omega) = \{\mathbf{x} + m\mathbf{x}, \mathbf{x} \in \Omega\}$. This notably differs by the fact that a dilation of Ω extends the outside boundary of the shape and shrinks the holes inside Ω while a magnification does not modify the size ratios. Moreover, even though the etching may dilate or erode the shape by a factor m, it does not modify the height of the silicon plate which is however the case in the previous analysis.

IV.3.2 Inverse lithography

In this section we consider an "optimal" shape Ω_{opt} found by our numerical Algorithm II.4.1 that is to be manufactured. We discuss an a posteriori optimization method which makes it possible to modify Ω_{opt} into another shape $\tilde{\Omega}_{\text{opt}}$ which, once manufactured, is "closer" to Ω_{opt} than $\widehat{\Omega}_{\text{litho}}$. This is in fact, another shape optimization problem: we seek for a shape Ω which, after the lithography-etching process, will match with Ω_{opt}. To simplify the presentation we only consider the lithography step meaning that $\widehat{\Omega}_{\text{etch}} = \widehat{\Omega}_{\text{litho}}$; we suppose that no defect is caused by etching. Moreover, we also suppose that the modification of the shape induced by the lithography process is known, that is to say that the value of the deviation parameter σ in Section IV.3.1.a is fixed equal to $\bar{\sigma}$ ($m = 0$ in Eq. (IV.3.5)).

Optimizing a shape in order to take into account the modifications induced by the lithography process has already been studied and is known in the literature as the **inverse lithography** problem. However, to our knowledge, no one has ever tried to solve this problem using the shape optimization framework presented in Chapter II (an attempt using the SIMP methodology presented in Section III.1.2 may be found in [Zho14; Jan13]).

The inverse lithography shape optimization problem may be modelized as the maximization of the following objective function (another model is given in [De 16] using the signed distance function)

$$\mathcal{J}_{\text{litho}}(\Omega) = \int_{\Omega_{\text{opt}}} L(\Omega)\,\mathrm{d}\mathbf{x} - \int_{\mathcal{D}\setminus\bar{\Omega}_{\text{opt}}} L(\Omega)\,\mathrm{d}\mathbf{x} = \int_{\mathcal{D}} \text{sign}(\Omega_{\text{opt}})L(\Omega)\,\mathrm{d}\mathbf{x}, \tag{IV.3.9}$$

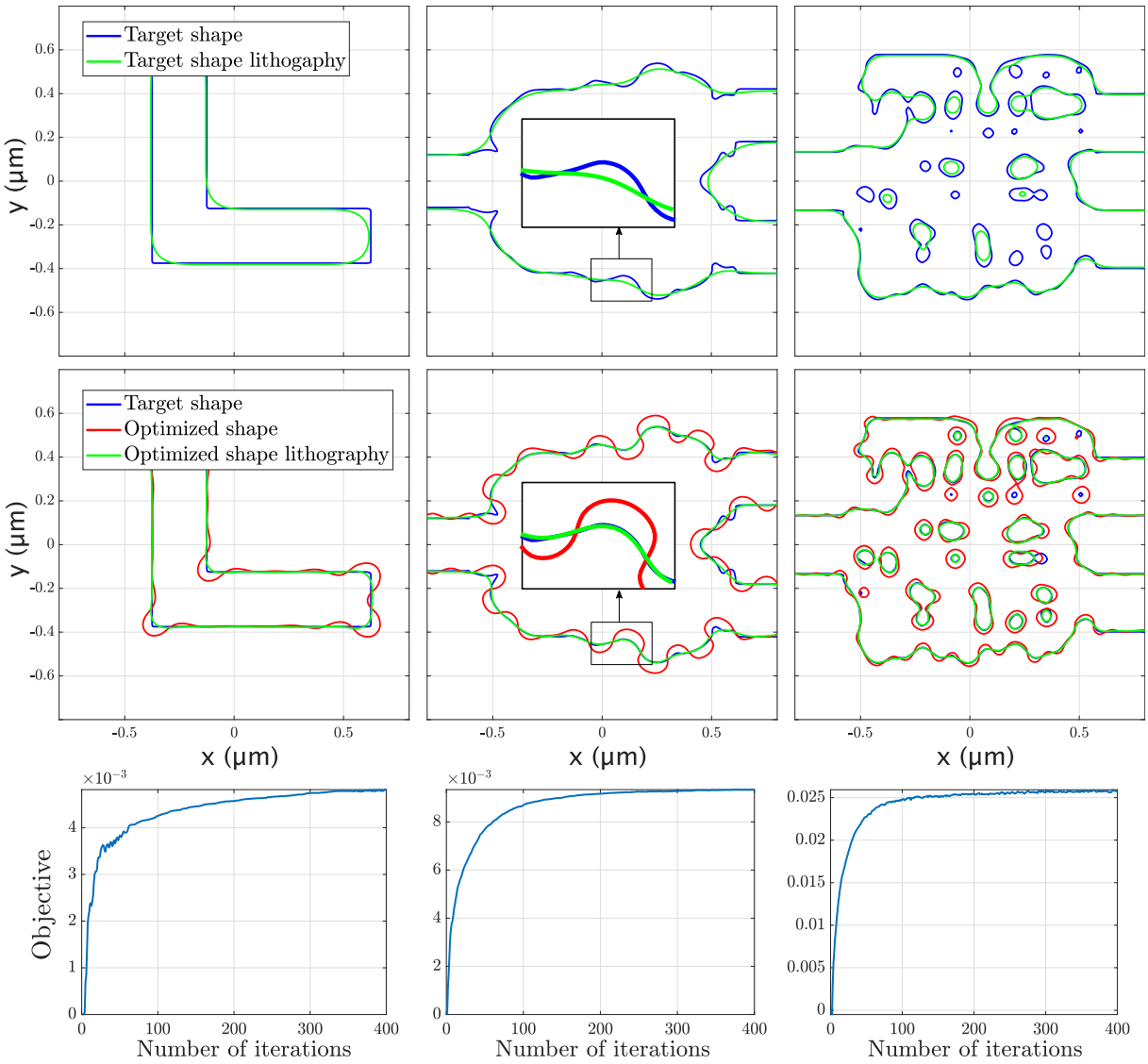
where $\text{sign}(\Omega)(\mathbf{x}) = 1$ if $\mathbf{x} \in \Omega$ and -1 otherwise. We also recall that $L(\Omega)$ is defined in Eq. (IV.3.4). Equation (IV.3.9) characterizes the fact that we want to maximize the value of $L(\Omega)$ inside Ω_{opt} while minimizing the value of $L(\Omega)$ elsewhere. The thresholding involved in $L(\Omega)$ causes Eq. (IV.3.9) to be non differentiable. We propose to modify it into

$$\mathcal{J}_{\text{litho},\varepsilon}(\Omega) = \int_{\mathcal{D}} \text{sign}(\Omega_{\text{opt}})H_\varepsilon \circ C(\Omega)\,\mathrm{d}\mathbf{x}, \tag{IV.3.10}$$

where $H_\varepsilon(x)$ is a smooth, non decreasing function equal to 0 if $x < 1/2 - \varepsilon$ and 1 if $x > 1/2 + \varepsilon$. Th. II.1.2.1 allows to find the shape derivative of Eq. (IV.3.10) as

$$\mathcal{J}'_{\text{litho},\varepsilon}(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} \boldsymbol{\theta}\cdot\mathbf{n} \int_{\mathcal{D}} \text{sign}(\Omega_{\text{opt}})(\mathbf{x})\mathcal{G}(\mathbf{x}-s)H'_\varepsilon \circ C(\Omega)(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}s. \tag{IV.3.11}$$

In Fig. IV.3.2 we present some optimization results associated to three different target shapes Ω_{opt} representing respectively an "L" shape, a power divider and a diplexer. As can be seen on Fig. IV.3.2 (middle), the optimized shape resulting from the above process (in red) exhibits large oscillations where $\partial\Omega_{\text{opt}}$ is sharp (high local curvature κ). This is clearly visible on the corners of the "L" shape where large oscillations of Ω_{opt} are observed. Interestingly, using the shape derivative $\mathcal{J}'(\Omega_{\text{opt}})$ we could modify the objective function of Eq. (IV.3.10) in order to specify where it is important to be close to the shape and whether it is less damaging to be inside or outside of Ω_{opt}; depending on the sign of $V_{\Omega_{\text{opt}}}$ in the shape derivative we know if the shape should be locally dilated or eroded. We did not pursue our research in this direction and therefore this addition has not been tested numerically.



(a) *A "L" shape* (b) *A power divider* (c) *A diplexer*

Figure IV.3.2: Optimization of the shape used as an input for the lithography process in order to obtain after fabrication a shape which is as close as possible to a target shape Ω_{opt}. In each run, 400 iterations of our shape optimization algorithm was performed to solve Eq. (IV.3.10) for a total duration of approximately 20 seconds. The standard deviation in the Gaussian kernel used in (a) and (b) equals $\sigma = 100\,\text{nm}$ and $\sigma = 50\,\text{nm}$ in (c).

IV.3.3 Geometrically robust shape optimization

We now illustrate how the general framework of Section IV.2 may be adapted to impose the robustness of the optimized designs Ω with respect to uncertainties on their geometry. Uncertainties related to the etching fabrication process are considered in Section IV.3.3.a, and an extension of these ideas is used in Section IV.3.3.b to deal with uncertainties caused by lithography. Note that using a worst-case approach for robustness to geometric uncertainties was also considered in [Zho14] for the optimization of micro-mechanical devices.

IV.3.3.a Robustness with respect to uncertainties caused by the etching process: an approach using dilation and erosion

As mentioned in Section IV.3.1.b, under- or over-etching of the design Ω is likely to occur in the course of the etching fabrication process. In other terms, the fabricated shape Ω_δ can be approximated as a uniform dilation or erosion of Ω,

$$\Omega_\delta := (\mathrm{Id} + \delta \mathbf{n}_\Omega)(\Omega), \tag{IV.3.12}$$

where δ is a real-valued parameter with small amplitude $|\delta| < m$.

In this context, the robust optimization problem Eq. (IV.2.3) of interest here brings into play the following perturbed functional $\mathcal{J}_\delta(\Omega)$ whose expression reads:

$$\mathcal{J}_\delta(\Omega) = \mathcal{J}(\Omega_\delta) = \mathcal{J}((\mathrm{Id} + \delta \mathbf{n}_\Omega)(\Omega)),$$

where $\mathcal{J}(\Omega)$ is given by Eq. (III.1.1). In particular, Eq. (IV.2.3) involves the optimized shape Ω via a modified version Ω_δ as for instance in the contribution [Che11]. In our way towards converting Eq. (IV.2.3) into a linear program of the form Eq. (IV.1.8), the shape derivative of the individual functions $\mathcal{J}_\delta(\Omega)$ is needed; the latter is the purpose of the next theorem.

Theorem IV.3.3.1 – Shape derivative considering dilated shapes.
Let Ω, and let $\delta > 0$ be small enough so that $(\mathrm{Id} + \delta \mathbf{n}_\Omega)$ is a diffeomorphism. The functional $\mathcal{J}_\delta(\Omega)$ is shape differentiable at Ω and its shape derivative reads:

$$\mathcal{J}'_\delta(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} g_{\Omega_\delta} \circ (\mathrm{Id} + \delta \mathbf{n}_\Omega) H(\boldsymbol{\theta}, \mathbf{n}_\Omega)\, \mathrm{d}s, \tag{IV.3.13}$$

where

$$H(\boldsymbol{\theta}, \mathbf{n}_\Omega) := |\det(\mathrm{Id} + \delta \nabla \mathbf{n}_\Omega)| \Big(((\mathrm{Id} + \delta \nabla \mathbf{n}_\Omega)^{-1} \mathbf{n}_\Omega \cdot \mathbf{n}_\Omega) \boldsymbol{\theta} \cdot \mathbf{n}_\Omega \Big).$$

Proof: Let $\delta > 0$ be sufficiently small for $(\mathrm{Id} + \delta \mathbf{n}_\Omega)$ to be a diffeomorphism from $\mathbb{R}^3$ into itself - where, as in Remark II.3.2.1, $\mathbf{n}_\Omega$ stands for a smooth extension of the unit normal vector on $\partial\Omega$ to $\mathbb{R}^3$. Our purpose is to calculate the shape derivative of the functional

$$\mathcal{J}_\delta(\Omega_{\boldsymbol{\theta}}) = \mathcal{J}((\mathrm{Id} + \delta \mathbf{n}_{\Omega_{\boldsymbol{\theta}}}) \circ (\mathrm{Id} + \boldsymbol{\theta})(\Omega)),$$

where we used the shortcut $\Omega_{\boldsymbol{\theta}} = (\mathrm{Id} + \boldsymbol{\theta})(\Omega)$. A straightforward calculation yields:

$$\begin{aligned}
\mathcal{J}_\delta(\Omega_{\boldsymbol{\theta}}) &= \mathcal{J}((\mathrm{Id} + \boldsymbol{\theta} + \delta \mathbf{n}_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta}))(\Omega)), \\
&= \mathcal{J}((\mathrm{Id} + \delta \mathbf{n}_\Omega + \boldsymbol{\theta} + \delta\, (\mathbf{n}_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta}) - \mathbf{n}_\Omega))(\Omega)), \\
&= \mathcal{J}((\mathrm{Id} + \boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta \boldsymbol{\xi}_2(\boldsymbol{\theta})) \circ (\mathrm{Id} + \delta \mathbf{n}_\Omega)(\Omega)), \\
&= \mathcal{J}((\mathrm{Id} + \boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta \boldsymbol{\xi}_2(\boldsymbol{\theta}))(\Omega_\delta)),
\end{aligned}$$

where we have defined

$$\boldsymbol{\xi}_1(\boldsymbol{\theta}) := \boldsymbol{\theta} \circ (\mathrm{Id} + \delta \mathbf{n}_\Omega)^{-1}, \text{ and } \boldsymbol{\xi}_2(\boldsymbol{\theta}) := (\mathbf{n}_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta}) - \mathbf{n}_\Omega) \circ (\mathrm{Id} + \delta \mathbf{n}_\Omega)^{-1},$$

considering that $\delta > 0$ is small enough so that $(\mathrm{Id} + \delta \mathbf{n}_\Omega)$ is a diffeomorphism. Using the following formula for the transformation of the normal vector where $\mathrm{com}(M)$ corresponds to the comatrix of M,

$$\mathbf{n}_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta}) = \frac{1}{|\mathrm{com}(\mathrm{I} + \nabla \boldsymbol{\theta}) \mathbf{n}_\Omega|} \mathrm{com}(\mathrm{I} + \nabla \boldsymbol{\theta}) \mathbf{n}_\Omega, \tag{IV.3.14}$$

whence the Lagrangian derivative of the normal vector field $\Omega \mapsto \mathbf{n}_\Omega$ is calculated, $\boldsymbol{\xi}_2$ expands on $\partial\Omega$ as (see [Dap13, Chapter 2]):

$$\boldsymbol{\xi}_2(\boldsymbol{\theta}) \circ (\mathrm{Id} + \delta\mathbf{n}_\Omega) = \Big((\nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega) \cdot \mathbf{n}_\Omega\Big)\, \mathbf{n}_\Omega - \nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega + o(\boldsymbol{\theta}).$$

Hence, using the Def. II.1.1.1 of the shape derivative of $\Omega \mapsto \mathcal{J}(\Omega)$ at Ω_δ, it follows that

$$\mathcal{J}_\delta(\Omega_{\boldsymbol{\theta}}) = \mathcal{J}_\delta(\Omega) + \mathcal{J}'(\Omega_\delta)(\boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta\boldsymbol{\xi}_2(\boldsymbol{\theta})) + o(\boldsymbol{\theta}), \tag{IV.3.15}$$

and we are now left with the calculation of the last quantity in the right-hand side of Eq. (IV.3.15); to this end, using Th. III.2.1.1 yields:

$$\mathcal{J}'(\Omega_\delta)(\boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta\boldsymbol{\xi}_2(\boldsymbol{\theta})) = \int_{\partial\Omega_\delta} g_{\Omega_\delta}\, (\boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta\boldsymbol{\xi}_2(\boldsymbol{\theta})) \cdot \mathbf{n}_{\Omega_\delta}\, \mathrm{d}s,$$

where $g_\Omega : \partial\Omega \to \mathbb{R}$ is given by Eq. (III.2.2). Changing variables in the integral in the above right-hand side, we finally obtain:

$$\begin{aligned}\mathcal{J}'(\Omega_\delta)(\boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta\boldsymbol{\xi}_2(\boldsymbol{\theta})) &= \int_{\partial\Omega} C_\Omega\, (g_{\Omega_\delta}\, (\boldsymbol{\xi}_1(\boldsymbol{\theta}) + \delta\boldsymbol{\xi}_2(\boldsymbol{\theta})) \cdot \mathbf{n}_{\Omega_\delta}) \circ (\mathrm{Id} + \delta\mathbf{n}_\Omega)\, \mathrm{d}s, \\ &= \int_{\partial\Omega} g_{\Omega_\delta} \circ (\mathrm{Id} + \delta\mathbf{n}_\Omega) \widetilde{H}(\boldsymbol{\theta}, \mathbf{n}_\Omega)\, \mathrm{d}s + o(\boldsymbol{\theta}). \end{aligned} \tag{IV.3.16}$$

where $C_\Omega = |\mathrm{com}(\mathrm{I} + \delta\nabla\mathbf{n}_\Omega)\mathbf{n}_\Omega|$ and

$$\widetilde{H}(\boldsymbol{\theta}, \mathbf{n}_\Omega) := \Big(\boldsymbol{\theta} + \delta\,\Big((\nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega \cdot \mathbf{n}_\Omega)\mathbf{n}_\Omega - \nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega\Big)\Big) \cdot C_\Omega.$$

where we have used again Eq. (IV.3.14). The derivative Eq. (IV.3.16) may now be given the convenient structure Eq. (III.2.1) owing to a little calculation. Indeed, let $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ be a local orthonormal basis of the tangent plane of $\partial\Omega$ around a fixed, arbitrary point $\mathbf{x}_0 \in \partial\Omega$, so that $(\boldsymbol{\tau}_1(\mathbf{x}), \boldsymbol{\tau}_2(\mathbf{x}), \mathbf{n}_\Omega(\mathbf{x}))$ is an orthonormal basis of $\mathbb{R}^3$ for any point $\mathbf{x} \in \partial\Omega$ close to $\mathbf{x}_0$. Then the Jacobian matrix of $\mathbf{n}_\Omega$ reads in this frame:

$$\nabla\mathbf{n}_\Omega = \begin{pmatrix} \kappa_1 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where κ_i is the principal curvature in direction $\boldsymbol{\tau}_i$. Hence,

$$\begin{aligned}\Big((\nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega) \cdot \mathbf{n}_\Omega\Big)\, \mathbf{n}_\Omega - \nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega &= -\sum_{i=1}^{2} \nabla\boldsymbol{\theta}^\top \mathbf{n}_\Omega \cdot \boldsymbol{\tau}_i, \\ &= -\sum_{i=1}^{2} \nabla(\boldsymbol{\theta} \cdot \mathbf{n}_\Omega) \cdot \boldsymbol{\tau}_i + \sum_{i=1}^{2} (\nabla\mathbf{n}_\Omega^\top \boldsymbol{\theta}) \cdot \boldsymbol{\tau}_i, \\ &= -\nabla_{\partial\Omega}(\boldsymbol{\theta} \cdot \mathbf{n}_\Omega) + \sum_{i=1}^{2} \kappa_i (\boldsymbol{\theta} \cdot \boldsymbol{\tau}_i)\boldsymbol{\tau}_i, \end{aligned} \tag{IV.3.17}$$

where $\nabla_{\partial\Omega} f := \nabla f - (\nabla f \cdot \mathbf{n}_\Omega)\mathbf{n}_\Omega$ is the tangential gradient of a smooth enough function $f : \partial\Omega \to \mathbb{R}$. It follows after a little more algebra that

$$\widetilde{H}(\boldsymbol{\theta}, \mathbf{n}_\Omega) = |\det(\mathrm{Id} + \delta\nabla\mathbf{n}_\Omega)|\Big(((\mathrm{Id} + \delta\nabla\mathbf{n}_\Omega)^{-1}\mathbf{n}_\Omega \cdot \mathbf{n}_\Omega)\boldsymbol{\theta} \cdot \mathbf{n}_\Omega\Big),$$

which is just the factor $H(\boldsymbol{\theta}, \mathbf{n}_\Omega)$ in Eq. (IV.3.13); this completes the proof. □

With this result at hand, the abstract framework of Section IV.2 can be readily used to tackle the robust optimization problem Eq. (IV.2.2), when small perturbations taking the form of a uniform dilation or erosion are expected on shapes.

As it is clear from the statement (and the proof) of Th. IV.3.3.1, these considerations are valid provided the maximum amplitude m of the expected dilations or erosions is small enough so that $(\mathrm{Id}+\delta\mathbf{n}_\Omega)$ is a diffeomorphism, i.e. Ω and Ω_δ share the same topology for all $|\delta|< m$. This restriction may impose unrealistically small values on m; so as to deal with more realistic situations, we rely on a heuristic adjustment of the above procedure when $(\mathrm{Id}+\delta\mathbf{n}_\Omega)$ is not a diffeomorphism. The starting point is the observation that $(\mathrm{Id}+\delta\mathbf{n}_\Omega)$ may fail to be a diffeomorphism because of the existence of points $\mathbf{x}\in\partial\Omega$ such that the segment with endpoints $\mathbf{x}$ and $\mathbf{x}+\delta\mathbf{n}_\Omega(\mathbf{x})$ crosses the **skeleton** (or sometimes also called **medial axis**) of the shape Ω, which is defined by

$$\mathrm{Sk}(\Omega) := \left\{\mathbf{x}\in\mathbb{R}^3,\ \exists \mathbf{y}_1,\mathbf{y}_2\in\partial\Omega,\ \mathbf{y}_1\neq\mathbf{y}_2 \text{ and } d(\mathbf{x},\partial\Omega)=|\mathbf{x}-\mathbf{y}_1|=|\mathbf{x}-\mathbf{y}_2|\right\}.$$

The skeleton $\mathrm{Sk}(\Omega)$ may alternatively be seen as the set of points $\mathbf{x}\in\mathbb{R}^3$ where the squared distance function d^2_Ω is not differentiable; see for instance [Del11, Section 6.3] about these points, and Fig. IV.3.3 for an illustration.

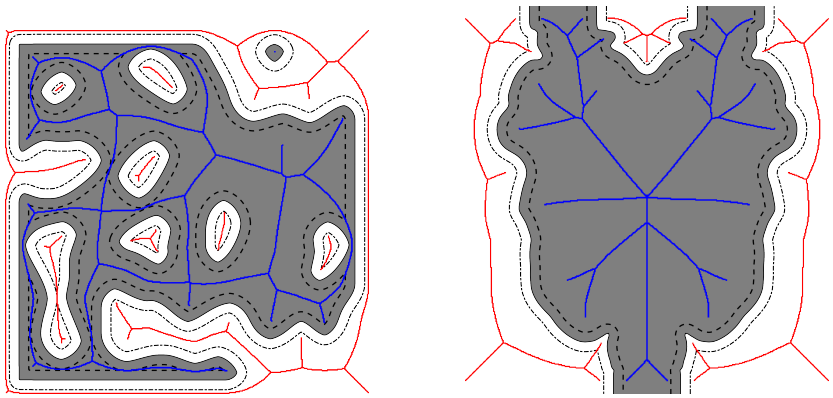


Figure IV.3.3: Two examples of shapes (in grey) with their skeleton. The contours of the dilated and eroded versions of both shapes are represented in dashed lines, and the interior (resp. exterior) part of the skeleton $\mathrm{Sk}(\Omega)\cap\Omega$ (resp. $\mathrm{Sk}(\Omega)\cap(\mathcal{D}\setminus\bar{\Omega})$) appears in blue (resp. red).

When this happens, we simply replace Ω_δ by the modified version

$$\Omega_s := (\mathrm{Id}+s(\mathbf{x})\mathbf{n}_\Omega)(\Omega),$$

where for $\mathbf{x}\in\partial\Omega$, $|s(\mathbf{x})|<\delta$ is calculated so that so that the segment joining $\mathbf{x}$ to $\mathbf{x}+s(\mathbf{x})\mathbf{n}_\Omega(\mathbf{x})$ does not intersect $\mathrm{Sk}(\Omega)$ - i.e. the considered dilation or erosion of Ω stops when $\mathrm{Sk}(\Omega)$ is encountered. To achieve this purpose, we proceed as in [Mic14, Section 3.6.2], relying on the knowledge of the signed distance function d_Ω: for every point $\mathbf{x}$, $s(\mathbf{x})$ is the first value $s<0$ (resp. $s>0$) such that the function $s\to d_\Omega(\mathbf{x}+s\nabla d_\Omega(\mathbf{x}))$ is no longer monotone in the case of an erosion (resp. dilation).

IV.3.3.b Robust shape optimization with respect to defects caused by the lithography process: a description using Gaussian kernels

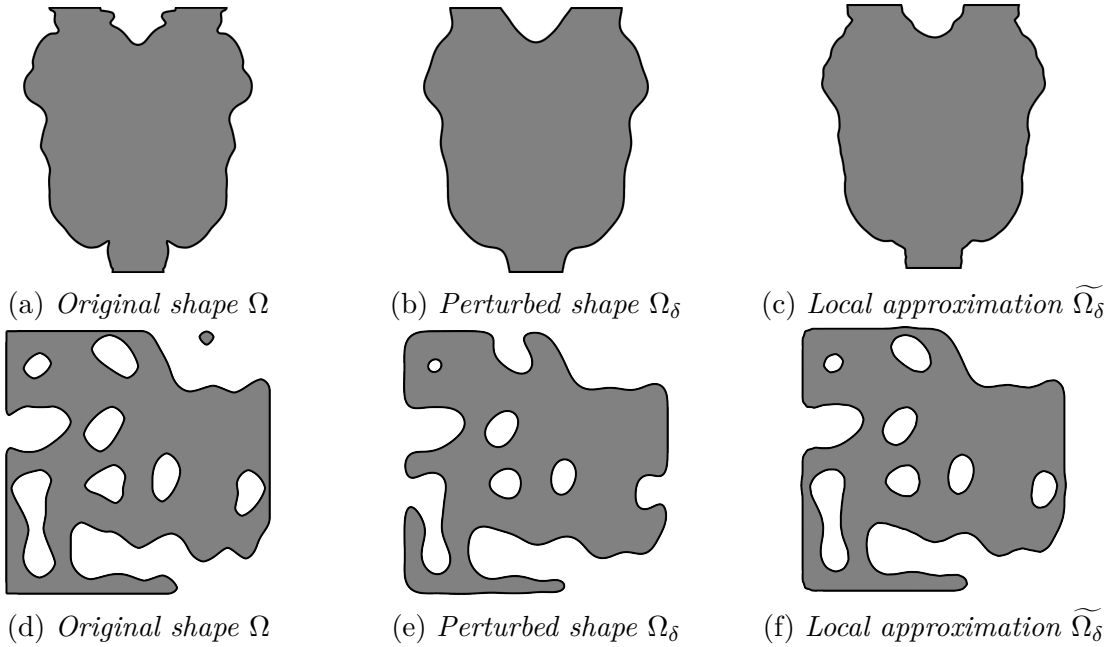
In Section IV.3.1.a we saw that when manufacturing a "blueprint", (y-invariant) shape

$$\Omega=\left\{(x,y,z)\in\mathbb{R}^3,\ (x,z)\in\widehat{\Omega},\ y\in(-h/2,h/2)\right\}$$

using the lithography process, the resulting shape is a smeared version $\Omega_\delta \subset \mathbb{R}^3$ given by:

$$\Omega_\delta = \left\{ (x,y,z) \in \mathbb{R}^3,\ (\mathbf{1}_{\widehat{\Omega}} * \mathcal{G}_\delta)(x,z) > \frac{1}{2},\ y \in (-h/2, h/2) \right\}, \tag{IV.3.18}$$

where $\mathcal{G}_\delta(\boldsymbol{\xi})$ is the Gaussian kernel with mean 0 and standard deviation δ defined in Eq. (IV.3.1).



(a) *Original shape* Ω (b) *Perturbed shape* Ω_δ (c) *Local approximation* $\widetilde{\Omega_\delta}$

(d) *Original shape* Ω (e) *Perturbed shape* Ω_δ (f) *Local approximation* $\widetilde{\Omega_\delta}$

Figure IV.3.4: Comparison between a shape Ω, its perturbed version Ω_δ given by Eq. (IV.3.18), and the approximation $\widetilde{\Omega_\delta}$ defined by Eq. (IV.3.19).

Let us recall that, intuitively, perturbations of the form Eq. (IV.3.18) imply that, if for instance the boundary $\partial\Omega$ is flat, $\partial\Omega_\delta$ coincides with $\partial\Omega$; however, if it has positive or negative curvature, the sharp feature of $\partial\Omega$ is smeared; see Fig. IV.3.4 for a two-dimensional illustration. In general, Ω_δ depends on global features of Ω, but it is mostly influenced by the curvature of $\partial\Omega$, in a rather non explicit fashion.

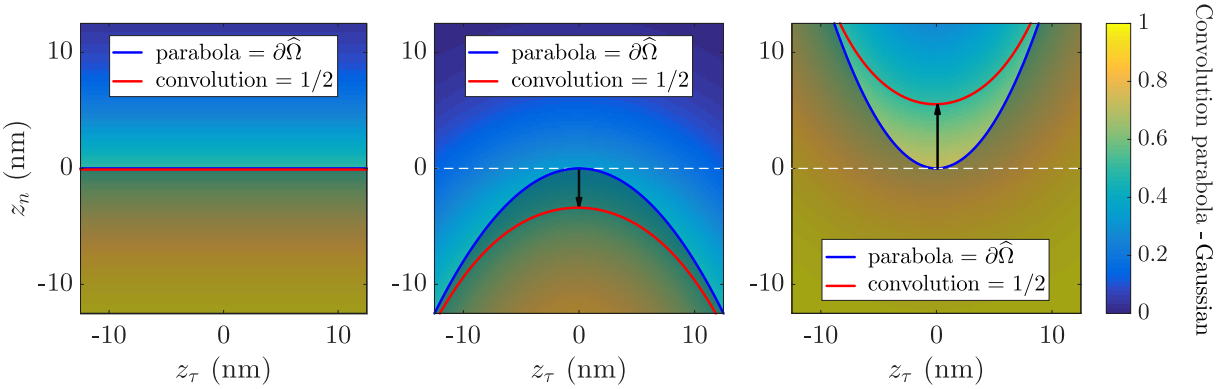


Figure IV.3.5: Schematic representation of the mapping $p_{\Omega,\delta}$ involved in the definition Eq. (IV.3.19) of the approximate perturbation $\widetilde{\Omega_\delta}$. Curvature $\kappa = 0$ (left), > 0 (middle) and < 0 (right).

The robust optimization problem Eq. (IV.2.3) with respect to uncertainties caused by lithography then involves the perturbed functional $\mathcal{J}_\delta(\Omega) := \mathcal{J}(\Omega_\delta)$. The study of $\mathcal{J}_\delta(\Omega)$,

and notably its shape derivative, is quite intricate because of the dependence of Ω_δ on global features of Ω. To alleviate this difficulty we trade Ω_δ for an approximate counterpart $\widetilde{\Omega_\delta}$ of the form:

$$\widetilde{\Omega_\delta} = (\mathrm{Id} + p_{\Omega,\delta}\mathbf{n}_\Omega)(\Omega), \tag{IV.3.19}$$

for a scalar field $p_{\Omega,\delta} : \partial\Omega \to \mathbb{R}$, which is defined below.

For an arbitrary, given point $\mathbf{x}_0 = (x_0, y_0, z_0) \in \partial\Omega$ with projection $\widehat{\mathbf{x}}_0 := (x_0, z_0)$ onto the 2d section $\widehat{\Omega}$, we consider the local, second-order approximation of the section $\widehat{\Omega}$ near $\widehat{\mathbf{x}}_0$, by means of the half-space $\mathcal{P}_{\Omega,\mathbf{x}_0}$ defined by (see Fig. IV.3.5):

$$\mathcal{P}_{\Omega,\mathbf{x}_0} = \left\{\widehat{\mathbf{x}}_0 + \hat{\mathbf{z}} \in \mathbb{R}^2,\ \ z_n < \kappa(\widehat{\mathbf{x}}_0) z_\tau^2\right\}.$$

In the latter formula, we have denoted by $z_n := \hat{\mathbf{z}} \cdot \mathbf{n}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0)$ and $z_\tau := \hat{\mathbf{z}} \cdot \boldsymbol{\tau}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0)$, the normal and tangential components of a vector $\hat{\mathbf{z}} \in \mathbb{R}^2$ in the local frame $(\boldsymbol{\tau}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0), \mathbf{n}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0))$ at $\widehat{\mathbf{x}}_0$ obtained by gathering the tangent $\boldsymbol{\tau}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0)$ and the normal vector $\mathbf{n}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0)$ to $\widehat{\Omega}$ at $\widehat{\mathbf{x}}_0$. Finally, $\kappa(\widehat{\mathbf{x}}_0)$ is the mean curvature of $\widehat{\Omega}$ at $\widehat{\mathbf{x}}_0$.

For $\widehat{\mathbf{x}} = (x, z) \in \mathbb{R}^2$ close to $\widehat{\mathbf{x}}_0$, taking advantage of the smallness of δ, we then have:

$$\left(\mathbf{1}_{\widehat{\Omega}} * \mathcal{G}_\delta\right)(x, z) = \int_{\mathbb{R}^2} \mathbf{1}_{\widehat{\Omega}}(\widehat{\mathbf{y}}) \mathcal{G}_\delta(\widehat{\mathbf{x}} - \widehat{\mathbf{y}})\, \mathrm{d}\widehat{\mathbf{y}} \approx F_{\Omega,\mathbf{x}_0}(\widehat{\mathbf{x}}),$$

where

$$F_{\Omega,\mathbf{x}_0}(\widehat{\mathbf{x}}) := \int_{\mathbb{R}^2} \mathbf{1}_{\mathcal{P}_{\Omega,\mathbf{x}_0}}(\widehat{\mathbf{y}}) \mathcal{G}_\delta(\widehat{\mathbf{x}} - \widehat{\mathbf{y}})\, \mathrm{d}\widehat{\mathbf{y}}$$

is the convolution between the characteristic function of the local second-order approximation of $\partial\widehat{\Omega}$ at $\widehat{\mathbf{x}}_0$ and the Gaussian kernel Eq. (IV.3.1). We then define $p_{\Omega,\delta}(\mathbf{x}_0)$ as the unique value $s \in \mathbb{R}$ such that

$$f(s) := F_{\Omega,\mathbf{x}_0}(\widehat{\mathbf{x}}_0 + s\mathbf{n}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0)) = \frac{1}{2},$$

which makes sense since

$$\begin{aligned} f(s) &= \int_{\mathbb{R}} \int_{-\infty}^{\kappa(\mathbf{x}_0) z_\tau^2 - s} \mathcal{G}_\delta(z_\tau \boldsymbol{\tau}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0) + t\mathbf{n}_{\widehat{\Omega}}(\widehat{\mathbf{x}}_0))\, \mathrm{d}t\, \mathrm{d}z_\tau \\ &= \frac{1}{2\sqrt{2\pi}\,\delta} \int_{\mathbb{R}} e^{-\frac{x^2}{2\delta^2}} \operatorname{erfc}\left(\frac{s - \kappa(\widehat{\mathbf{x}}_0)x^2}{\sqrt{2}\,\delta}\right) \mathrm{d}x \end{aligned}$$

(where $\operatorname{erfc}(x) = 2/\sqrt{\pi} \int_x^\infty \exp(-t^2)\, \mathrm{d}t$ refers to the so-called complementary error function) is a smooth, decreasing function with

$$\lim_{s \to -\infty} f(s) = 1 \text{ and } \lim_{s \to +\infty} f(s) = 0.$$

Notice that $p_{\Omega,\delta}(\mathbf{x}_0)$ only depends on $\widehat{\Omega}$ via its curvature at $\widehat{\mathbf{x}}_0$ however the dependence is not explicit. Nevertheless, $p_{\Omega,\delta}$ is easy to calculate numerically. As is exemplified on Fig. IV.3.4 this approximation performs reasonably well: the approximate perturbation $\widetilde{\Omega_\delta}$ is close to Ω_δ, except that it fails to capture the topological changes between Ω_δ and Ω.

Returning to our robust optimization problem, the implementation of Eq. (IV.2.3) relies on the shape derivative of the perturbated functional

$$\widetilde{\mathcal{J}}_\delta(\Omega) := \mathcal{J}((\mathrm{Id} + p_{\Omega,\delta}\mathbf{n}_\Omega)(\Omega)). \tag{IV.3.20}$$

The rigorous calculation and the practical use of this shape derivative are not simple since $p_{\Omega,\delta}$ brings into play the curvature of the interface $\partial\Omega$. To simplify this calculation, we neglect the dependence of $p_{\Omega,\delta}$ on Ω, so that the shape derivative $\mathcal{J}'_\delta(\Omega)$ is given by Eq. (IV.3.13) where δ is replaced by $p_{\Omega,\delta}$. Although quite rough, this approximation gives pretty good results as presented in Section IV.3.4.d and it has the advantage of being simple and fast to implement.

IV.3.4 Numerical examples

IV.3.4.a Geometrically robust power divider

In this section we consider a variant of the power divider test-case tackled in Section III.3.1.b in which robustness of the optimized design is desired with respect to uncertainties entailed by the etching manufacturing process, as discussed in Section IV.3.3.a.

The shape Ω^* resulting from the optimization study in Section III.3.1.b (that is, without taking robustness issues into account) is not robust with respect to uncertainties related to etching. Indeed, let us consider Fig. IV.3.6(b) where the variation of the performance criterion $\mathcal{J}$ is represented when a dilation or an erosion of at most $m = \pm\, 30\,\mathrm{nm}$ is performed on Ω^*: in particular, if Ω^* is eroded by 30 nm, the performance of the shape drops from 49 % down to only 20 %.

To remedy this, starting from Ω^* as initial shape, we solve the following robust problem which involves the dilated and eroded versions of the optimized shape:

$$\max_\Omega \min\left\{\mathcal{J}_{-m}(\Omega), \mathcal{J}_0(\Omega), \mathcal{J}_m(\Omega)\right\}, \tag{IV.3.21}$$

in which we have defined

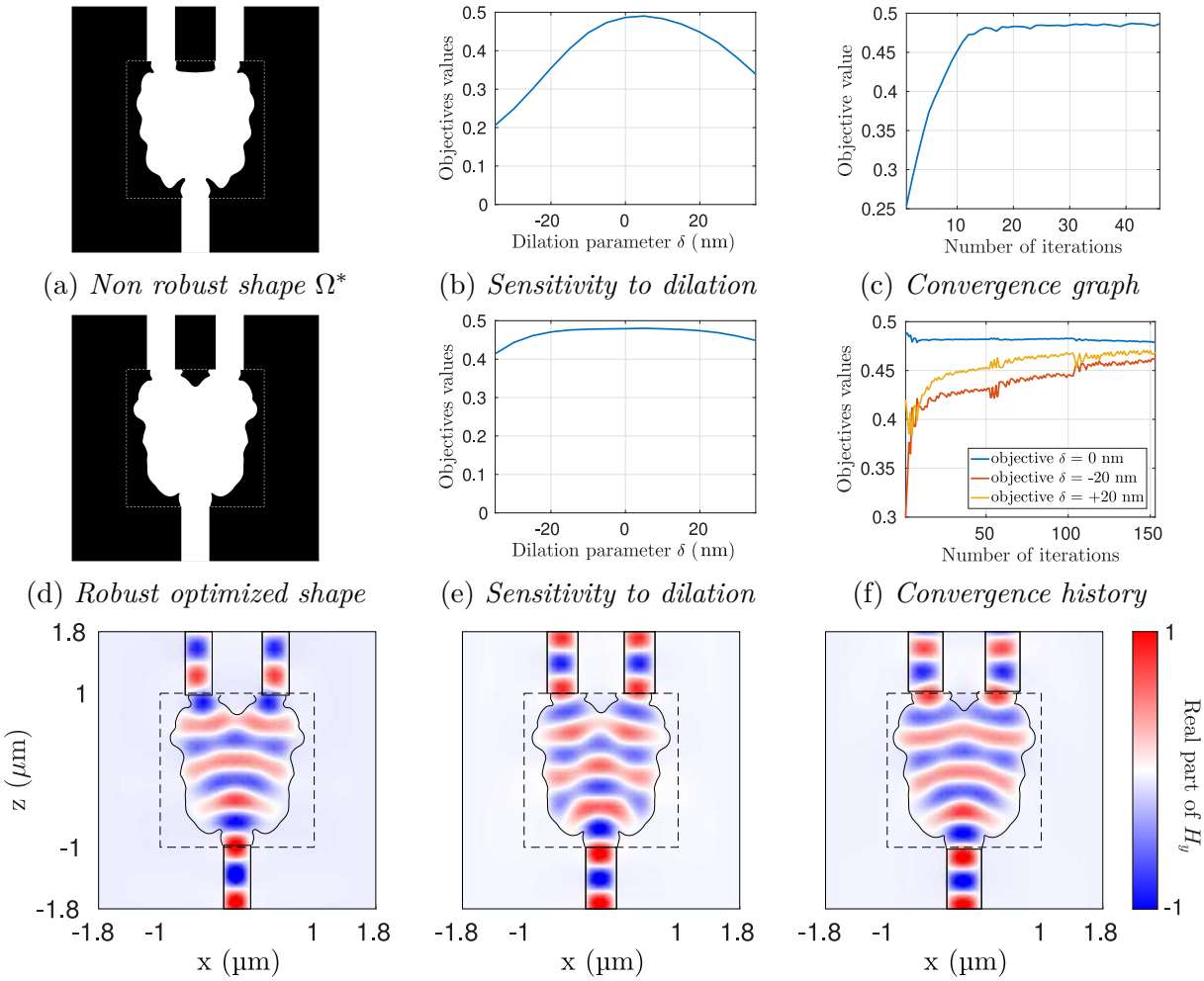
$$\mathcal{J}_\delta(\Omega) = \mathcal{J}(\Omega_\delta), \;\; \Omega_\delta := (\mathrm{Id} + \delta\mathbf{n}_\Omega)(\Omega),$$

where $\mathcal{J}(\Omega)$ is again given by Eq. (III.1.1). After 150 iterations of our optimization algorithm, we end up with the shape displayed in Fig. IV.3.6. The comparison of Fig. IV.3.6(b) and Fig. IV.3.6(e) suggests that the new power divider is much more robust to manufacturing uncertainties caused by etching. Moreover, the nominal performance of the device is not significantly degraded in the process, since it only suffers from a reduction by 1 % when compared to Ω^*.

IV.3.4.b Geometrically robust mirror

This second example is taken from our paper [Leb19c]: we consider the optimization of a mirror as in Section III.3.1.a. Note that materials and wavelength used here are different of the ones considered in the other experiments presented previously. Precisely,

- the operating wavelength lies within the mid-infrared region with $\lambda = 6.06\,\mu\mathrm{m}$,
- the core is made of an $\mathrm{Si}_{60\,\%}\mathrm{Ge}_{40\,\%}$ alloy whose optical index is taken equal to 3.54,



(a) *Non robust shape* Ω^* (b) *Sensitivity to dilation* (c) *Convergence graph*

(d) *Robust optimized shape* (e) *Sensitivity to dilation* (f) *Convergence history*

(g) *Electric field of the eroded by* 30 nm *(left), optimized (middle) and dilated by* 30 nm *(right) shape*

Figure IV.3.6: Results of the robust shape optimization of a power divider with respect to uncertainties linked to etching in Section IV.3.4.a; the upper row reproduces the features of the non robust optimized shape Ω^* of Section III.3.1.b.

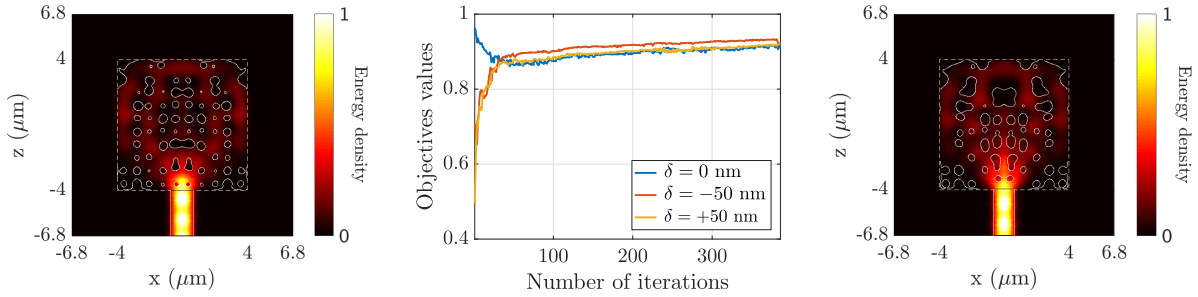
- a SiN substrate is used with an optical index of 1.7,
- the upper cladding is still air with $n_{\text{air}} = 1$,
- and finally waveguides with a 1.4 µm×1.4 µm cross-section are used, allowing the single mode operation of the fundamental TM mode.

The component design obtained after resolution of the mirror objective defined in Section III.3.1.a without taking uncertainties into account and using 400 iterations of the reference optimization algorithm has a reflectivity $\simeq 96\,\%$ (Fig. IV.3.7(a)).

It is then used as an initial condition for our robust design process as in the case of the previous power divider considering the objective function Eq. (IV.3.21) using the value of the figure of merit $\mathcal{J}$ evaluated at eroded and dilated versions of the shape. During the first iterations, as can be seen on the convergence curve presented on Fig. IV.3.7(b), the reflectivity of the $\pm$ 50 nm eroded and dilated shapes are significantly lower than that of the nominal shape. However, after a few dozen iterations, all three objectives reach

approximately the same value, and then continue to improve concurrently to reach an enhanced final solution, which is robust to fabrication uncertainties (Fig. IV.3.7(c)).

The reflectivity of several mirrors is plotted against the distance of erosion or dilation on Fig. IV.3.8. The reflectivity of the non robust design (black curve) is dramatically lowered as soon as the erosion or dilation exceeds a magnitude of ± 50 nm caused by etching. On the contrary, at the expense of a slight degradation of the maximum performance, the ± 25 nm (red curve) and ± 50 nm (green curve) robust designs retain a flat response on their respective variation interval.



(a) *Non-robust mirror used for initialization*

(b) *Convergence history*

(c) *Robust design to* ± 50 nm *fabrication uncertainty*

Figure IV.3.7: Optimization of a geometrically robust mirror to ± 50 nm of erosion/dilation starting from an initial non-robust optimized design.

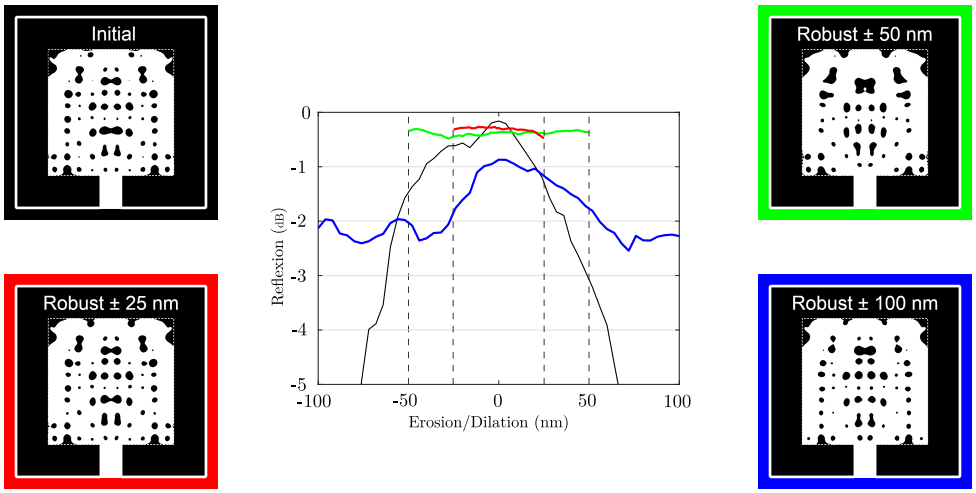


Figure IV.3.8: Robustness to fabrication uncertainty at 6.06 µm of several optimized mirror. The black curve represents the non robust design and the red, green and blue curves respectively represent the ± 25 nm, ± 50 nm and ± 100 nm robust designs.

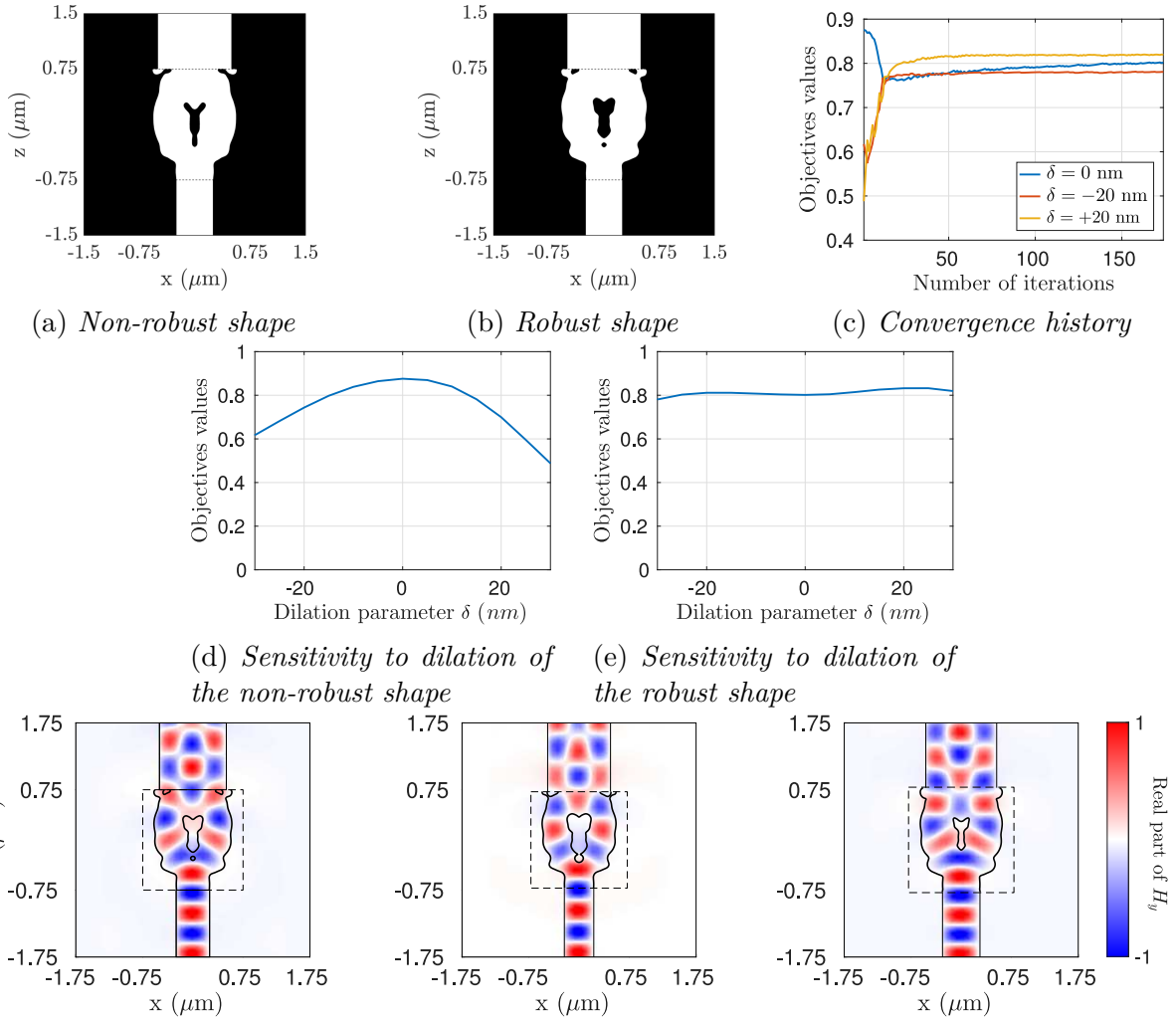
The loss is more important for the ± 100 nm (blue curve) robust design, which has been obtained by optimizing simultaneously five objective functions consisting on equally spaced eroded and dilated shapes. Let us note however that a ± 100 nm process uncertainty is excessively pessimistic in practice.

IV.3.4.c Geometrically robust mode converter

In quite the same spirit as in Section IV.2.2, we now revisit the mode converters test cases of Section III.3.1.c. When optimizing a 1.5 µm×1.5 µm TE_0 to TE_2 mode converter

(Fig. IV.3.9(a)) we see on Fig. IV.3.9(d) that the performance of the resulting shape proves to be very sensitive to small perturbations in the form of a uniform dilation or erosion.

Starting from this non-robust optimized shape as an initial shape, we now solve the robust counterpart optimization problem of a mode converter involving the worst-case between the values taken by the objective function $\mathcal{J}$ on the optimized shape Ω and its dilated and eroded perturbations by 20 nm. After 200 iterations of our shape and topology optimization algorithm, the results summarized in Fig. IV.3.9 are obtained. The new optimized shape is significantly more robust to the effect of dilation and erosion; visually, its main difference with that obtained without considering robustness effects lies in that the central hole has widened to make up for the effects of dilation.



(a) *Non-robust shape* (b) *Robust shape* (c) *Convergence history*

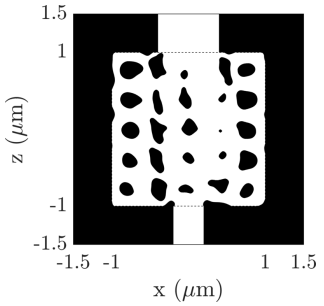
(d) *Sensitivity to dilation of the non-robust shape* (e) *Sensitivity to dilation of the robust shape*

(f) *Electric field at* $y = 0$ *of the robust shape* (g) *Electric field at* $y = 0$ *of the eroded robust shape by* 20 nm (h) *Electric field at* $y = 0$ *of the dilated robust shape by* 20 nm

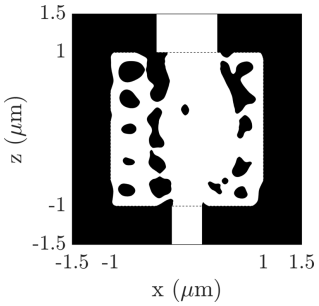
Figure IV.3.9: Optimization of a robust TE_0 to TE_2 mode converter as defined in Section III.3.1.c starting from an optimized non-robust 1.5 µm×1.5 µm wide TE_0 to TE_2 mode converter.

In the same way we move on to the optimization of a robust TE_0 to TE_1 mode converter. Compared to Fig. IV.3.9 which exhibits a "simple" topology with only one hole in the center of the design, we consider here our algorithm for geometrical robustness on a shape

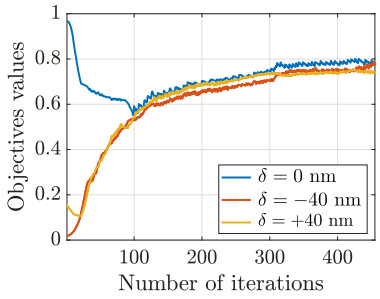
composed of multiple holes obtained after optimizing a non robust mode converter starting from an initial guess with 5×5 holes; see Fig. IV.3.10(a). As one would expect the resulting design is very sensitive to small dilations or erosions of the shape due to the presence of very small patterns as we can see on Fig. IV.3.10(d). The gradient sampling algorithm is used considering the normal, dilated by 40 nm and eroded by 40 nm shapes. On the convergence history of Fig. IV.3.10(c) we can see that the three objectives are optimized simultaneously. The final design, resulting from about 500 iterations of our algorithm is depicted on Fig. IV.3.10(b), in which several small holes has been removed from the non robust design and which is far less sensitive to dilation or erosion; see Fig. IV.3.10(e).



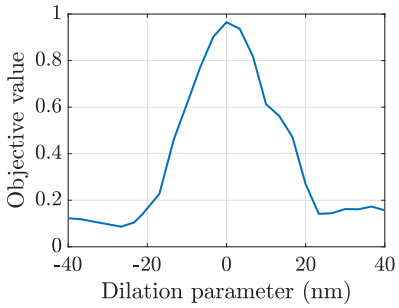
(a) *Non-robust shape*



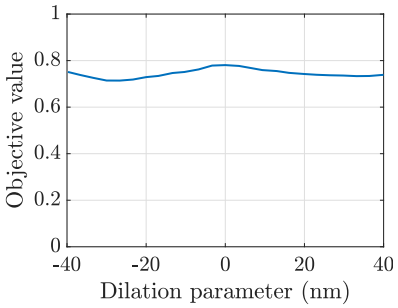
(b) *Robust shape*



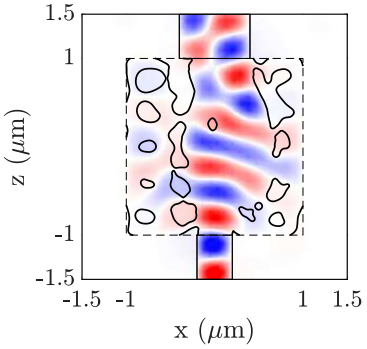
(c) *Convergence history*



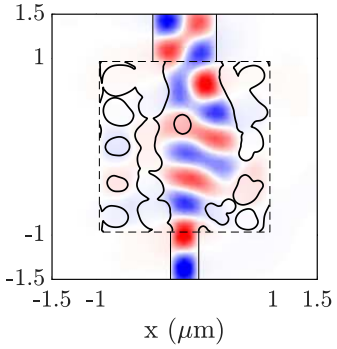
(d) *Sensitivity to dilation and erosion of the non-robust shape*



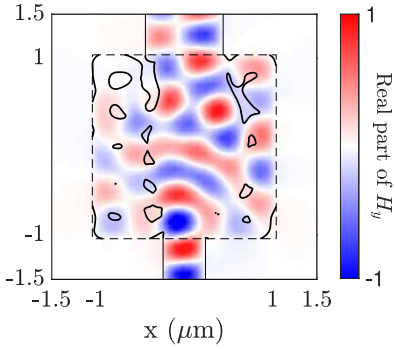
(e) *Sensitivity to dilation and erosion of the robust shape*



(f) *Electric field at $y = 0$ of the robust shape*



(g) *Electric field at $y = 0$ of the eroded robust shape by* 40 nm



(h) *Electric field at $y = 0$ of the dilated robust shape by* 40 nm

Figure IV.3.10: Optimization of a robust TE_0 to TE_1 mode converter as defined in Section III.3.1.c starting from a non-robust optimized mode converter.

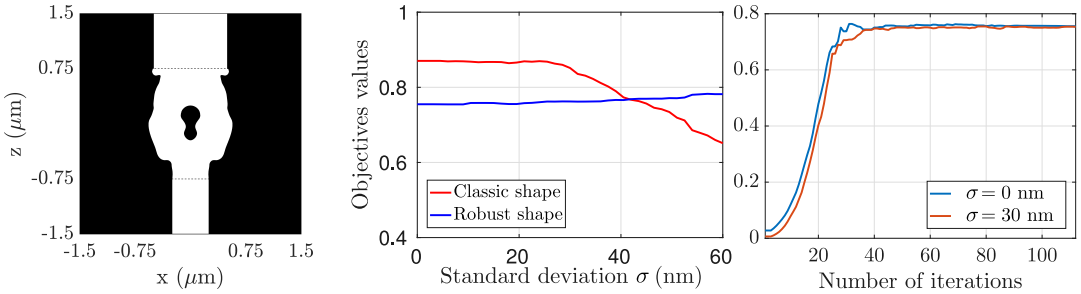
IV.3.4.d Robust components with respect to uncertainties caused by lithography

In this last subsection we finally turn to numerically appraising the procedure proposed in Section IV.3.3.b for imposing robustness of shapes with respect to the lithography

manufacturing process. The physical setting is that of the mode converter example, as discussed in Section III.3.1.c.

We consider the robust optimization program with respect to uncertainties caused by lithography as in Section IV.3.3.b. We recall that the approximate perturbation $\widetilde{\Omega_\sigma}$ of Ω is defined by Eq. (IV.3.19), and with a value $\sigma = 30\,\text{nm}$ for the parameter representing the standard deviation in the Gaussian kernel Eq. (IV.3.1).

Starting from the initial shape in Fig. III.3.8(a) and performing 120 iterations of our optimization algorithm yields the optimized design represented on Fig. IV.3.11.

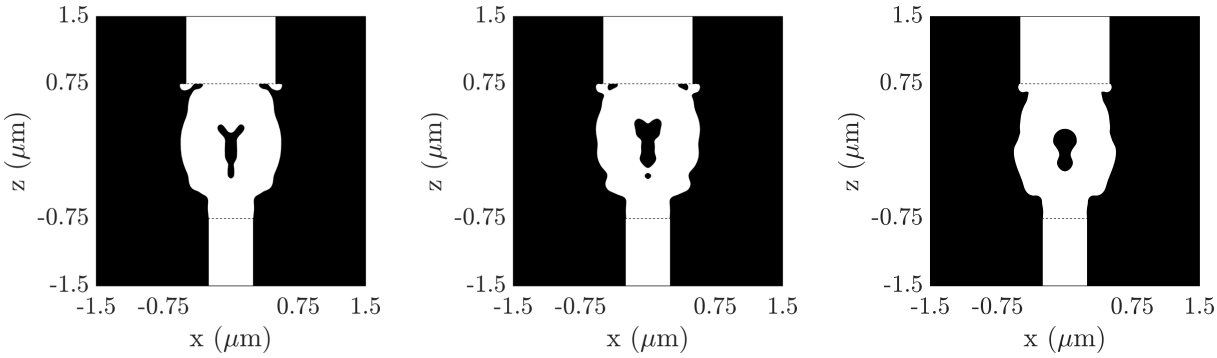


(a) *Robust optimized shape with respect to lithography uncertainties*

(b) *Sensitivity to uncertainties caused by lithography for the robust and non-robust (Fig. IV.3.12(a)) shapes.*

(c) *Convergence history*

Figure IV.3.11: Lithography robust optimization of a 1.5×1.5 µm mode converter.



(a) *Non robust (see Section III.3.1.c)*

(b) *Etching robust (see Section IV.3.4.c)*

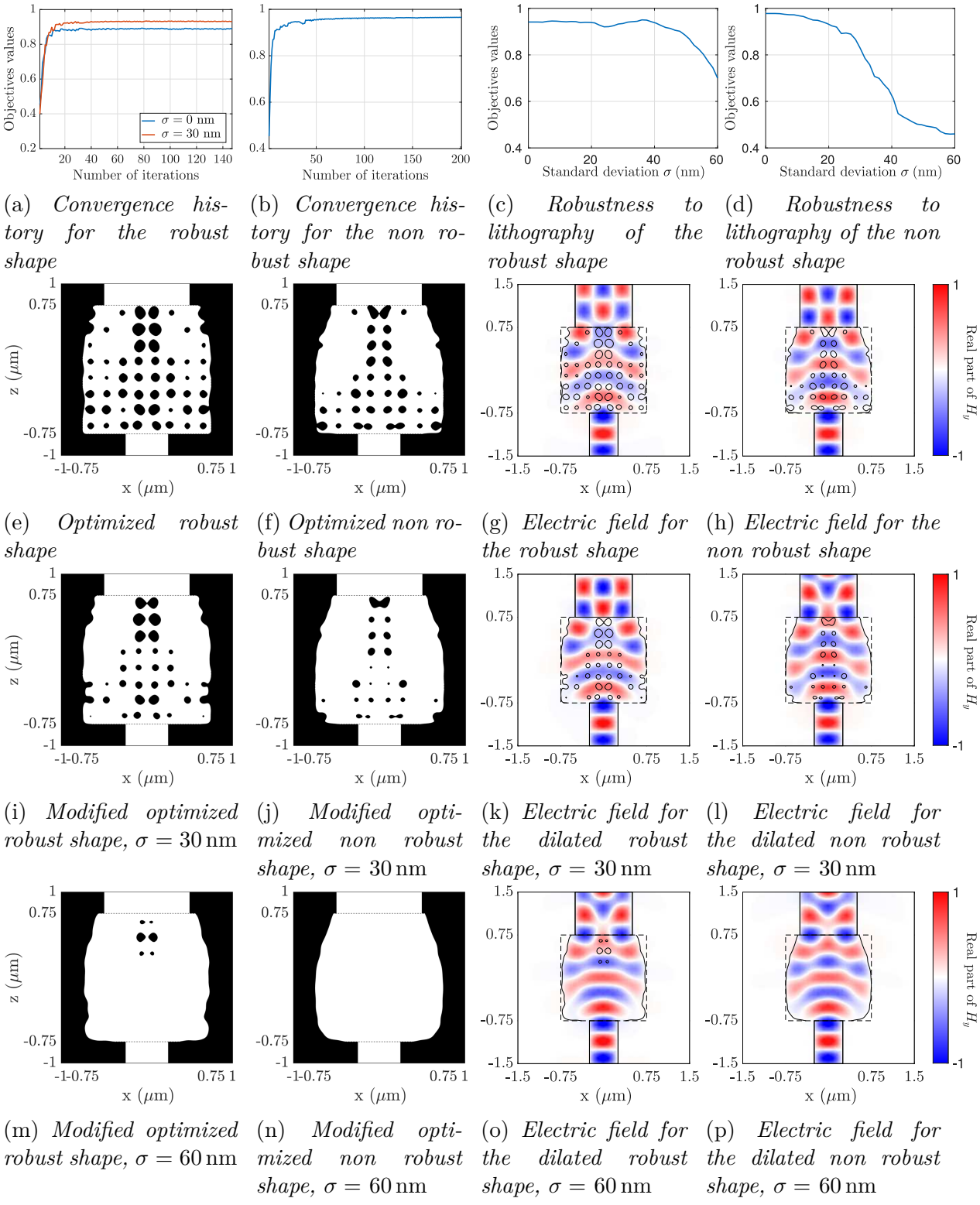
(c) *Lithography robust (see Section IV.3.4.d)*

Figure IV.3.12: Optimized shapes for the mode converter example without taking robustness effects into account, and considering robustness with respect to etching and lithography uncertainties.

For an easy to read visual comparison, three designs obtained for the mode converter test-case (non robust, robust with respect to etching uncertainties, and robust with respect to lithography uncertainties) are reproduced in Fig. IV.3.12.

We also performed another optimization, using this time an initial design perforated by 8×8 holes. The results obtained in the case of the non robust optimization problem and for its robust counterpart are summarized in Fig. IV.3.13. As expected, for small values

of σ the performances are better than the one in Fig. IV.3.11 but the shapes are also more sensitive to values of σ larger than $\sim 40\,\text{nm}$.



(a) *Convergence history for the robust shape*

(b) *Convergence history for the non robust shape*

(c) *Robustness to lithography of the robust shape*

(d) *Robustness to lithography of the non robust shape*

(e) *Optimized robust shape*

(f) *Optimized non robust shape*

(g) *Electric field for the robust shape*

(h) *Electric field for the non robust shape*

(i) *Modified optimized robust shape, $\sigma = 30\,\text{nm}$*

(j) *Modified optimized non robust shape, $\sigma = 30\,\text{nm}$*

(k) *Electric field for the dilated robust shape, $\sigma = 30\,\text{nm}$*

(l) *Electric field for the dilated non robust shape, $\sigma = 30\,\text{nm}$*

(m) *Modified optimized robust shape, $\sigma = 60\,\text{nm}$*

(n) *Modified optimized non robust shape, $\sigma = 60\,\text{nm}$*

(o) *Electric field for the dilated robust shape, $\sigma = 60\,\text{nm}$*

(p) *Electric field for the dilated non robust shape, $\sigma = 60\,\text{nm}$*

Figure IV.3.13: Optimization of a $1.5\,\mu\text{m} \times 1.5\,\mu\text{m}$ wide TE_0 to TE_2 mode-converter with an initial shape of 8×8 holes, with and without lithography robustness.

This page is intentionally left blank.

Chapter V

Boundary shape optimization

Summary — This chapter is quite different from the two previous ones in the sense that the numerical examples will not be applied to nanophotonic but rather mechanical problems whose behaviors are described using the linear elasticity equation. From a mathematical point of view this chapter is also different; although based on the geometric shape optimization method presented in Chapter II, it involve a more precise functional analysis of the regularity of PDEs solutions than the one used to derive the shape derivative in the case of the time-harmonic vector wave equation.

Our research on boundary shape optimization originally comes from the study of active components in nanophotonics. More precisely, we started our investigations with the optimization of the shape of electric contacts placed around a component and whose power supply allows the establishment of a current into the component. The electric current then causes an overheating through Joule effect resulting in the modification of the optical indices and therefore a change in the overall behavior of the device. During our mathematical research however, we noticed that given the power that can be injected into the electric contacts, the low sensitivity of the optical indices to temperature variation as well as the small size of the components that can be simulated, it is not possible to obtain sufficiently interesting results. Despite this, we also noticed that our mathematical study may largely be extended to other physical fields for which there are more concrete applications such as linear elasticity.

This chapter is presented as follows. In Section V.1 the difficulty behind boundary shape optimization in which we try to optimize the shape of the interface between two different boundary conditions is presented in the context of the Laplace PDE. In Sections V.2 and V.3 the cases of inhomogenous Neumann-homogeneous Neumann transition as well as homogeneous Neumann-homogeneous Dirichlet are studied extensively. In the case of transitions between Dirichlet and Neumann boundary conditions we find that the shape derivative involves quantities that are difficult to compute since they comes from the singular part of the solutions at the interface. We then propose in Section V.4 a regularization procedure to approximate this transition resulting in a problem that consistently approximates the initial PDE and for which its shape derivative may be efficiently computed. The chapter end with Section V.5 in which we apply the previously derived shape derivatives to the optimization of mechanical devices.

Most of the material of this chapter is coming from our preprint

[Dap19] C. Dapogny, N. Lebbe, and E. Oudet. “Optimization of the shape of regions supporting boundary conditions”. working paper or preprint. 2019. URL: https://hal.archives-ouvertes.fr/hal-02064477v1.

V.1 General problem

V.1.1 Introduction and motivations

In the previous chapters, we studied the optimization of a shape functional $\mathcal{J}(\Omega)$ which depends on the shape Ω via a state u_Ω arising as the solution to a boundary value problem for a certain partial differential equation. In several applications, it turns out that only one part of the boundary $\partial\Omega$ of the shape is subject to optimization, which is associated to one single type of boundary conditions for the state u_Ω in the underlying physical partial differential equation while the remaining regions are not meant to be modified. For example, in structural design, where u_Ω is the displacement of the structure and is the solution to the linearized elasticity system, it is customary to optimize only the traction-free part of $\partial\Omega$ (i.e. that bearing homogeneous Neumann boundary conditions). Likewise, in fluid applications (where u_Ω is the velocity of the fluid, solution to the Stokes or Navier-Stokes equations), one is often interested in optimizing only the region of $\partial\Omega$ supporting no-slip (that is, homogeneous Dirichlet) boundary conditions.

A little surprisingly, the dependence of a given performance criterion $\mathcal{J}(\Omega)$ with respect to the relative locations of regions accounting for different types of boundary conditions has been relatively seldom investigated in shape and topology optimization. Yet, situations where it is desirable to optimize not only the overall shape of Ω but also the repartition of the zones on $\partial\Omega$ bearing different types of boundary conditions are multiple in concrete applications. Let us mention a few of them:

- When the objective criterion involves thermal effects inside the optimized shape Ω, u_Ω is the temperature, solution to the stationary heat equation. The regions of $\partial\Omega$ associated to Dirichlet boundary conditions are those where a known temperature profile is imposed, while Neumann boundary conditions account for heat injection. It may be desirable to investigate the regions where heat should enter the medium Ω (or those which should be kept at fixed temperature) in order to minimize, for instance, the mean temperature inside Ω, or its variance; see for instance [Bar82] about such physical applications.

- In the context of linearly elastic structures, (homogeneous) Dirichlet boundary conditions account for the regions of $\partial\Omega$ where the structure is fixed, while inhomogeneous (resp. homogeneous) Neumann boundary conditions correspond to regions of $\partial\Omega$ where external loads are applied (resp. to traction-free regions). It may be of great interest to optimize the design of fixations, or the places where loads should be applied. One interesting application of this idea concerns the optimization of a clamping-locator fixture system; see for instance [Ma11] in the framework of density-based topology optimization method. More recently, in [Xia16; Xia14] the authors present an adapted level set method for the joint optimization of the shape of an elastic structure Ω and of the region of its boundary $\partial\Omega$ where it should be fixed.

- In acoustic applications, u_Ω is the solution to the time-harmonic Helmholtz equation. In this situation, the contribution [Des18] deals with the optimal repartition of an absorbing material (accounted for by Robin-like boundary conditions) on the walls of a room in order to minimize the sound pressure.

From the mathematical point of view, the above problems are of unequal difficulty: the calculation of the shape derivative $\mathcal{J}'(\Omega)$ of the optimized criterion $\mathcal{J}(\Omega)$ (or that of the

constraint functions) in the framework of Hadamard's method depends very much on the regularity of the physical state u_Ω of the problem in the neighborhood of the optimized transition region between zones bearing different types of boundary conditions.

In some of the above situations, u_Ω is "smooth enough" near these transitions, and the calculation of $\mathcal{J}'(\Omega)$ is achieved using Céa's method presented in Section II.2.1. On the contrary, the situation become much more difficult to analyze when u_Ω happens to be "weakly singular" near these transitions; indeed, Céa's method fail to give the correct result when u_Ω does not enjoy sufficient regularity (remember that the Eulerian derivative of u_Ω is not defined if u_Ω is not sufficiently regular as it was the case with the eletric field in Section III.2.1). Moreover, the calculation of the shape derivative of $\mathcal{J}(\Omega)$ requires a precise knowledge of this singular behavior of u_Ω. The resulting formula is also quite difficult to handle in algorithmic practice, since it brings into play quantities which somehow measure this singular behavior, that are difficult to evaluate from the numerical viewpoint; see Fig. V.1.1.

(a) *Boundary conditions* (b) *Simulation of* u_Ω (c) *Plot of* $|\nabla u_\Omega|^2$

Figure V.1.1: Simulation of $-\Delta u_\Omega = 1$ using an arbitrary geometry and different boundary conditions. We clearly see in (c) that there is less regularity near the interfaces between the homogeneous Neumann boundary Γ and homogeneous Dirichlet boundary Γ_D than between Γ and the inhomogeneous Neumann boundary Γ_N. Simulation results in arbitrary units. The mesh used here is extremely fine with more than 25 000 triangles to show that the non-regularity can be observed even with a high precision mesh.

To the best of our knowledge, the first theoretical calculation of shape derivatives in a context where regions bearing different types of boundary conditions are optimized – dealing with the difficulty of a "weakly singular" state u_Ω – dates back to [Fre01].

Our purpose in this chapter is to study such shape optimization problems in which the regions of the boundary of the optimized shape Ω bearing different types of boundary conditions are subject to optimization. Most of the theoretical analysis unfolds in a model two-dimensional situation where a functional $\mathcal{J}(\Omega)$ of the shape Ω is minimized; $\mathcal{J}(\Omega)$ depends on Ω via a state u_Ω which is the solution to a Laplace equation with mixed boundary conditions: the boundary $\partial\Omega$ is divided into three regions Γ_D, Γ_N and Γ, and u_Ω satisfies homogeneous Dirichlet boundary conditions on Γ_D, inhomogeneous Neumann boundary conditions on Γ_N, and homogeneous Neumann boundary conditions on Γ; see Section V.1.2 below.

At first, we rigorously calculate the derivative $\mathcal{J}'(\Omega)$ of $\mathcal{J}(\Omega)$ with respect to variations of the shape Ω in both situations where the transitions $\Sigma_N = \overline{\Gamma} \cap \overline{\Gamma_N}$ and $\Sigma_D = \overline{\Gamma} \cap \overline{\Gamma_D}$ are also subject to optimization. In the former case, the shape derivative turns out to have a classical structure, and it lends itself to a fairly simple treatment in numerical algorithms. On the contrary, in the latter context, the state u_Ω is weakly singular near Σ_D, which makes the formula for $\mathcal{J}'(\Omega)$ uneasy to handle in practice. To circumvent this drawback, our second contribution is to propose an approximation method for the considered state problem, and thereby for the resulting shape optimization problem: the considered "exact" Laplace equation with mixed boundary conditions is replaced with an approximate counterpart, parametrized by a "small" parameter ε, where Robin boundary conditions with ε-varying coefficients are imposed on the whole boundary $\partial\Omega$. The "sharp" transition Σ_D between regions equipped with homogeneous Dirichlet and Neumann boundary conditions is thus "smeared" into a zone with thickness ε (this regularization procedure is the same kind as the optical index smoothing method of Section III.2.2.a). We then turn to prove the consistency of this approach: namely, the approximate objective function $\mathcal{J}_\varepsilon(\Omega)$ and its shape derivative $\mathcal{J}'_\varepsilon(\Omega)$ converge to their exact counterparts $\mathcal{J}(\Omega)$ and $\mathcal{J}'(\Omega)$ when the smoothing parameter ε vanishes.

V.1.2 A model presenting a singularity at the Dirichlet-Neumann interface

In this section, we present the model problem under scrutiny in most of the chapter, which concentrates the main difficulties we plan to address in a simplified setting, and lends itself to a rather complete mathematical analysis.

V.1.2.a Presentation of the model physical problem and notations

Let $\Omega \subset \mathbb{R}^d$ be a smooth bounded domain ($d = 2$ or 3 in applications), whose boundary $\partial\Omega$ is divided into three disjoint, complementary open regions $\Gamma_D \neq \emptyset$, Γ_N and Γ:

$$\partial\Omega = \overline{\Gamma_D} \cup \overline{\Gamma_N} \cup \overline{\Gamma}. \tag{V.1.1}$$

We denote by u_Ω the unique solution in the space

$$H^1_{\Gamma_D}(\Omega) := \left\{ u \in H^1(\Omega),\ \ u = 0 \text{ on } \Gamma_D \right\}$$

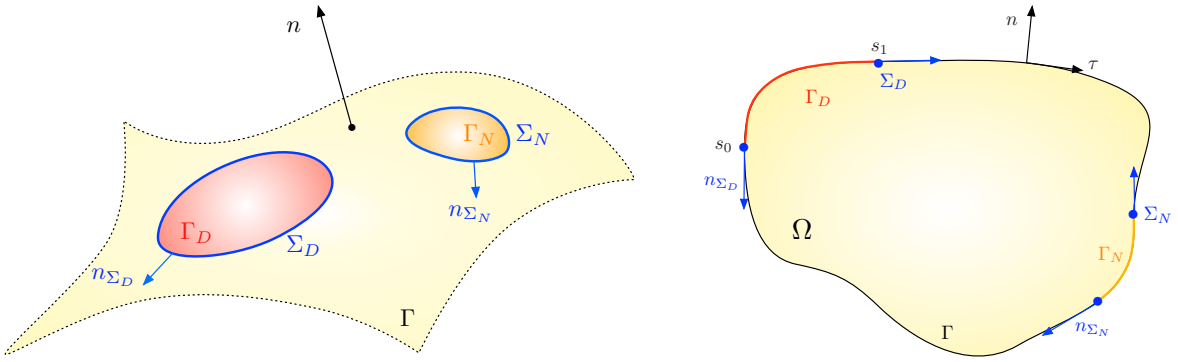
to the following mixed boundary value problem:

$$\left\{ \begin{array}{rcll} -\Delta u_\Omega & = & f & \text{in } \Omega, \\ u_\Omega & = & 0 & \text{on } \Gamma_D, \\ \frac{\partial u_\Omega}{\partial \mathbf{n}} & = & g & \text{on } \Gamma_N, \\ \frac{\partial u_\Omega}{\partial \mathbf{n}} & = & 0 & \text{on } \Gamma, \end{array} \right. \tag{V.1.2}$$

where the source term f and boundary flux g are supposed to be regular enough, say $f \in L^2(\mathbb{R}^d)$, $g \in H^1(\mathbb{R}^d)$. Note that g may vanish on some subset of Γ_N, so that the inclusion $\Gamma \subset \left\{ \mathbf{x} \in \partial\Omega,\ \frac{\partial u_\Omega}{\partial \mathbf{n}} = 0 \right\}$ may be strict. See Fig. V.1.1 for an example of solution u_Ω.

In this context, we denote by $\Sigma_D = \overline{\Gamma_D} \cap \overline{\Gamma} \subset \partial\Omega$ (resp. $\Sigma_N = \overline{\Gamma_N} \cap \overline{\Gamma} \subset \partial\Omega$) the boundary between the region $\Gamma_D \subset \partial\Omega$ bearing homogeneous Dirichlet boundary conditions (resp. the region $\Gamma_N \subset \partial\Omega$ bearing inhomogeneous Neumann boundary conditions) and that

Γ endowed with homogeneous Neumann boundary conditions; for simplicity, we assume that $\overline{\Gamma_D} \cap \overline{\Gamma_N} = \emptyset$. The sets Σ_D and Σ_N are both assumed to be smooth, codimension 1 submanifolds of $\partial\Omega$: in particular, they amount to collections of isolated points in the case $d = 2$, and to sets of smooth closed curves drawn on $\partial\Omega$ if $d = 3$. We denote by $\mathbf{n}_{\Sigma_D} : \Sigma_D \to \mathbb{R}^d$ (resp. $\mathbf{n}_{\Sigma_N} : \Sigma_N \to \mathbb{R}^d$) the unit normal vector to Σ_D (resp. Σ_N) pointing outward Γ_D (resp. Γ_N), inside the tangent plane to $\partial\Omega$; see Fig. V.1.2(a) for an illustration of these definitions.



(a) *Three dimensional setting* (b) *Two dimensional setting*

Figure V.1.2: Sketch of the considered setting in Section V.1.2.a and the simplified situation where Assumption Eq. (V.1.3) is fulfilled.

On several occurrences, the rigorous mathematical analysis of this model will be greatly simplified under further simplifying assumptions; in particular, in some duly specified situations, we shall proceed under the following hypotheses, in the situation where $d = 2$:

$$\text{The region } \overline{\Gamma_D} \cap \overline{\Gamma} \text{ consists of only two points } \overline{\Gamma_D} \cap \overline{\Gamma} = \{s_0, s_1\}\,, \quad \text{and} \quad \text{the domain } \Omega \subset \mathbb{R}^2 \text{ is locally flat around } s_0, s_1; \tag{V.1.3}$$

see Fig. V.1.2(b). In this last case, letting $\mathbf{n} = (n_1, n_2)$, we shall denote by $\boldsymbol{\tau} := (n_2, -n_1)$ the unit tangent vector to $\partial\Omega$, oriented so that $(\boldsymbol{\tau}, \mathbf{n})$ is a direct orthonormal frame of the plane.

V.1.2.b The shape optimization problem

In the setting of Section V.1.2.a, we consider the following shape optimization problem:

$$\inf_{\Omega \in \mathcal{U}_{\rm ad}} \mathcal{J}(\Omega), \tag{V.1.4}$$

(note here that in this chapter we consider minimization instead of maximization of the objective functional) featuring an objective criterion $\mathcal{J}(\Omega)$ of the form

$$\mathcal{J}(\Omega) = \int_\Omega j(u_\Omega)\, \mathrm{d}\mathbf{x} \tag{V.1.5}$$

where $j : \mathbb{R} \to \mathbb{R}$ is a smooth function satisfying appropriate growth conditions: there exists a constant $C > 0$ such that:

$$\forall t \in \mathbb{R},\ \ |j(t)| \leq C(1 + t^2),\ \ |j'(t)| \leq C(1 + |t|),\ \text{ and } |j''(t)| \leq C.$$

In Eq. (V.1.4), $\mathcal{U}_{\rm ad}$ is a set of smooth admissible shapes; in the following, two distinct shape optimization problems of this form are considered, implying different definitions of $\mathcal{U}_{\rm ad}$:

- On the one hand, the transition Σ_D between homogeneous Dirichlet and homogeneous Neumann boundary conditions is subject to optimization, and the region Γ_N bearing inhomogeneous Neumann boundary conditions is fixed. Then, $\mathcal{U}_{\text{ad}}$ corresponds to the set:

$$\mathcal{U}_{\text{DN}} := \left\{\Omega \subset \mathbb{R}^d \text{ is bounded and of class } \mathcal{C}^2,\ \Gamma_N \subset \partial\Omega\right\}.$$

- On the other hand, when the region Σ_N between Γ and Γ_N is optimized (while the region $\Gamma_D \subset \partial\Omega$ is fixed), $\mathcal{U}_{\text{ad}}$ reads:

$$\mathcal{U}_{\text{NN}} := \left\{\Omega \subset \mathbb{R}^d \text{ is bounded and of class } \mathcal{C}^2,\ \Gamma_D \subset \partial\Omega\right\}.$$

Notice that problem Eq. (V.1.4) is not guaranteed to have a solution; nevertheless, we assume in the following that it is the case or that, at least, local minima exist. Moreover, observe that the objective function $\mathcal{J}(\Omega)$ featured in Eq. (V.1.5) and the solution u_Ω to Eq. (V.1.2) depend on the particular subdivision Eq. (V.1.1) of the boundary $\partial\Omega$ into Γ_D, Γ_N and Γ; with some little abuse and so as to keep notations simple insofar as possible, this dependence is not made explicit in the formulation of our shape optimization problem.

In practice, so that variations of an admissible shape stay admissible, the considered deformations $\boldsymbol{\theta}$ are confined to a subset $\Theta_{\text{ad}} \subset W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d)$ of admissible deformations. In the present chapter, Θ_{ad} stands for one of the sets Θ_{DN} or Θ_{NN} defined by:

$$\Theta_{\text{DN}} := \left\{\boldsymbol{\theta} \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d),\ \boldsymbol{\theta} = 0 \text{ on } \Gamma_N\right\}, \tag{V.1.6}$$

and

$$\Theta_{\text{NN}} := \left\{\boldsymbol{\theta} \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d, \mathbb{R}^d),\ \boldsymbol{\theta} = 0 \text{ on } \Gamma_D\right\}, \tag{V.1.7}$$

both being equipped with the natural norm, when the considered set of admissible shapes is $\mathcal{U}_{\text{DN}}$ or $\mathcal{U}_{\text{NN}}$, respectively.

V.1.2.c A brief account of the regularity of u_Ω

As is well-known in the field of elliptic boundary value problems (see e.g. [Bre10, Section 9.6] for a comprehensive introduction), when the featured data f, g (and Ω itself) are smooth enough, the solution u_Ω to Eq. (V.1.2) is more regular than a mere element in $H^1(\Omega)$, as predicted by the classical Lax-Milgram variational theory. For instance, in the case where Eq. (V.1.2) only brings into play homogeneous Dirichlet boundary conditions (i.e. Γ and Γ_N are empty), a classical theorem states that, provided f belongs to $H^m(\Omega)$ for some $m \geq 0$ (and Ω is smooth), u_Ω belongs to $H^{m+2}(\Omega)$.

In the situations at stake in the present chapter, such as Eq. (V.1.2), things are a little more subtle. The assumptions that $f \in L^2(\mathbb{R}^d)$ and $g \in H^1(\mathbb{R}^d)$ still guarantee that u_Ω has H^2 regularity in some open neighborhood of an arbitrary point x_0 which is either interior to Ω, or which belongs to $\partial\Omega$ but is interior to one of the regions Γ_D, Γ or Γ_N. On the contrary, u_Ω has limited regularity around those points $x_0 \in \Sigma_D$ or $x_0 \in \Sigma_N$ marking the transition between regions subjected to different types of boundary conditions. A whole mathematical theory exists in the literature, which is devoted to the precise study of the "weakly singular" behavior of u_Ω in these regions; we refer for instance to the monographs [Dau06; Gri11; Koz97].

In this section, we first recall briefly some classical material about functional spaces in Section V.1.2.d, before summarizing in Section V.1.2.e the needed results about the regularity of the solution u_Ω to Eq. (V.1.2) for our purposes.

V.1.2.d Some functional spaces

Let Ω be a smooth bounded domain in $\mathbb{R}^d$. For $s > 0$ and $1 < p < \infty$, let us introduce:

- The usual Sobolev space $W^{s,p}(\Omega)$ (see [Di 12, Section 2]) is defined by, when $s = m$ is an integer:

$$W^{m,p}(\Omega) = \left\{ u \in L^p(\Omega),\ \ \partial^\alpha u \in L^p(\Omega) \text{ for } \alpha \in \mathbb{N}^d,\ \ |\alpha| \leq m \right\},$$

  and when $s = m + \sigma$ with $m \in \mathbb{N}$ and $\sigma \in (0,1)$,

$$W^{s,p}(\Omega) = \left\{ u \in W^{m,p}(\Omega),\ \ \int_\Omega \int_\Omega \frac{|\partial^\alpha u(\mathbf{x}) - \partial^\alpha u(\mathbf{y})|^p}{|\mathbf{x}-\mathbf{y}|^{d+p\sigma}}\, \mathrm{d}\mathbf{y}\, \mathrm{d}\mathbf{x} < \infty \text{ for all } |\alpha| = m \right\}. \tag{V.1.8}$$

  Both sets are equipped with the natural norms. Let us recall that the exact same definitions hold when the bounded domain Ω is replaced by the whole space $\mathbb{R}^d$.

- The subspace $\widetilde{W}^{s,p}(\Omega)$ of $W^{s,p}(\Omega)$ is that composed of functions whose extension $\widetilde{u}$ by 0 outside Ω belongs to $W^{s,p}(\mathbb{R}^d)$.

- The space $W_0^{s,p}(\Omega)$ is the closure of the set $\mathcal{C}_c^\infty(\Omega)$ of $\mathcal{C}^\infty$ functions with compact support in Ω in $W^{s,p}(\Omega)$.

As is customary, in the case $p = 2$ we use the notations $H^s(\Omega)$, $\widetilde{H}^s(\Omega)$ and $H_0^s(\Omega)$ for $W^{s,p}(\Omega)$, $\widetilde{W}^{s,p}(\Omega)$ and $W_0^{s,p}(\Omega)$ respectively.

In spite of their tight relation, the two spaces $\widetilde{W}^{s,p}(\Omega)$ and $W_0^{s,p}(\Omega)$ may not coincide depending on the values of s and p. On the one hand, for any $s > 0$ and $1 < p < \infty$, $W_0^{s,p}(\Omega) \subset \widetilde{W}^{s,p}(\Omega)$, but the converse inclusion may fail. In fact, the following characterization holds (see [Gri11, Lemma 1.3.2.6 and Corollary 1.4.4.10]):

$$\widetilde{W}^{s,p}(\Omega) = \left\{ u \in W_0^{m,p}(\Omega),\ \ \frac{1}{\rho^\sigma}\partial^\alpha u \in L^p(\Omega),\ \ |\alpha| = m \right\}, \tag{V.1.9}$$

where we have decomposed $s = m+\sigma$, with $m \in \mathbb{N}$ and $\sigma \in (0,1)$, and $\rho(\mathbf{x}) := d(\mathbf{x}, \partial\Omega)$ is the (unsigned) distance from $\mathbf{x}$ to the boundary of Ω. The space $\widetilde{W}^{s,p}(\Omega)$ is then endowed with the norm

$$\|u\|_{\widetilde{W}^{s,p}(\Omega)} = \left(\|u\|^p_{W^{m,p}(\Omega)} + \sum_{|\alpha|=m} \int_\Omega \frac{1}{\rho^{p\sigma}} |\partial^\alpha u|^p \, \mathrm{d}\mathbf{x} \right)^{\frac{1}{p}}, \tag{V.1.10}$$

which is equivalent to the natural norm $u \mapsto \|\widetilde{u}\|_{W^{s,p}(\mathbb{R}^d)}$. Let us also note that:

$$\widetilde{W}^{s,p}(\Omega) = W_0^{s,p}(\Omega) \text{ when } \left(s - \frac{1}{p}\right) \text{ is not an integer.}$$

We eventually mention that the above definitions and results hold in the more general context where Ω is replaced by a smooth submanifold of $\mathbb{R}^d$, e.g. (a region of) the boundary of a bounded smooth domain of $\mathbb{R}^d$. For instance, in the setting of Section V.1.2.a, we may consider the spaces $\widetilde{W}^{s,p}(\Gamma_D)$, $\widetilde{W}^{s,p}(\Gamma_N)$, etc.

V.1.2.e Local structure of u_Ω near the transition Σ_D

The classical variational theory for Eq. (V.1.2) (based on the Lax-Milgram theorem) features a solution u_Ω which naturally belongs to $H^1(\Omega)$. Moreover, assuming that the boundary $\partial\Omega$ is at least of class $\mathcal{C}^2$, and that $f \in L^2(\mathbb{R}^d)$, $g \in H^1(\mathbb{R}^d)$, the classical elliptic regularity theory ensures that u_Ω actually has H^2 regularity except perhaps near the transitions zones Σ_D and Σ_N; see [Bre10, Section 9.6]. On the contrary, u_Ω fails to enjoy H^2 regularity in the vicinity of Σ_D or Σ_N, where the boundary conditions it fulfills change types.

The precise behavior of u_Ω near Σ_D will be of utmost interest for our purpose; it is exemplified by the following theorem, which takes place under Assumption Eq. (V.1.3) (see [Gri11, Chapter 4 and notably Theorem 4.4.3.7]):

Theorem V.1.2.1 – Singular decomposition of u_Ω.
For any point $\mathbf{x}_0 \in \Omega$ or $\mathbf{x}_0 \in \partial\Omega \setminus (\Sigma_D \cup \Sigma_N)$, there exists an open neighborhood W of $\mathbf{x}_0$ in $\mathbb{R}^2$ such that u_Ω belongs to $H^2(\Omega \cap W)$. Furthermore, for either $i = 0$ or $i = 1$, there exists an open neighborhood V of s_i with the following property: introducing the polar coordinates (r, ν) at s_i, assuming without loss of generality that $s_i = 0$, $\Omega \cap V = \{\mathbf{x} \in V, \text{ s.t. } x_2 > 0\}$, and $\Gamma_D \cap V = \{\mathbf{x} \in V, \text{ s.t. } x_2 = 0,\ x_1 < 0\}$, there exist a function $u_r^i \in H^2(V)$, and a constant $c^i \in \mathbb{R}$ such that:

$$u_\Omega = u_r^i + c^i S^i \text{ on } V, \text{ where } S^i(r,\nu) = r^{\frac{1}{2}} \cos(\nu/2). \tag{V.1.11}$$

The function S^i is sometimes said to be weakly singular, in the sense that it belongs to $H^1(V)$, but not to $H^2(V)$. More precisely, invoking [Gri11, Theorem 1.4.5.3] to estimate the Sobolev regularity of functions of the form $r^\alpha \varphi(\nu)$, one proves that, for every $0 \leq s < \frac{3}{2}$, $u_\Omega \in H^s(V)$, with:

$$\|u_\Omega\|_{H^s(V)} \leq C_s \|f\|_{L^2(\mathbb{R}^d)}. \tag{V.1.12}$$

In the language of Section V.1.2.d, it follows in particular that $u_\Omega \in \widetilde{H}^{\frac{1}{2}}(\Gamma \cup \Gamma_N)$, while $\frac{\partial u_\Omega}{\partial \mathbf{n}} \in (\widetilde{H}^{\frac{1}{2}}(\Gamma \cup \Gamma_N))^*$.

Remark V.1.2.1: Higher-order versions of the expansion Eq. (V.1.11) are available. Actually, for any integer $m \geq 2$, if $f \in H^{m-2}(\mathbb{R}^d)$, the following decomposition holds in a neighborhood V of s_i (see [Gri11, Theorem 5.1.3.5]):

$$u_\Omega = u_{r,m}^i + \sum_{k=1}^{m-1} c_k^i S_k^i, \text{ where } u_{r,m}^i \in H^m(V) \text{ and } S_k^i(r,\nu) = r^{k-\frac{1}{2}} \cos((k-1/2)\nu).$$

Remark V.1.2.2: Still in the two-dimensional context, when the boundary $\partial\Omega$ is not flat in the vicinity of Σ_D, an expansion of the form Eq. (V.1.11) still holds: the weakly singular function S^i shows the same dependence $r^{\frac{1}{2}}$ with respect to r, but its dependence with respect to ν is no longer explicit. Nevertheless, there still holds that $u_\Omega \in H^s(V)$ with an estimate of the form Eq. (V.1.12), where V is an open neighborhood of Σ_D in Ω and $0 \leq s < \frac{3}{2}$ is arbitrary; see [Gri11, Chapter 5].

In Sections V.2 and V.3, we rigorously calculate the shape derivative of the objective function Eq. (V.1.4); we start in Section V.2 with the case where the transition region Σ_N is optimized and Σ_D is fixed (i.e. the sets of admissible shapes and admissible deformations

are $\mathcal{U}_{\mathrm{NN}}$ and Θ_{NN} respectively), before turning in Section V.3 to the more difficult case where Σ_D is optimized and Σ_N is not – i.e. $\mathcal{U}_{\mathrm{ad}} = \mathcal{U}_{\mathrm{DN}}$ and $\Theta_{\mathrm{ad}} = \Theta_{\mathrm{DN}}$ – under the simplifying assumption Eq. (V.1.3).

V.1.2.f Some facts from tangential calculus

In this section, we briefly review some facts from tangential calculus which come in handy in several parts of this article; see [Hen06, Section 5.4.3] for a more exhaustive presentation.

Let Ω be a smooth bounded domain in $\mathbb{R}^d$. There exists a tubular neighborhood U of its boundary $\partial\Omega$ such that the **projection mapping** $p_{\partial\Omega} : U \to \Gamma$ given by

$$p_{\partial\Omega}(\mathbf{y}) = \text{the unique } \mathbf{x} \in \Gamma \text{ s.t. } |\mathbf{x} - \mathbf{y}| = d(\mathbf{y}, \Gamma)$$

is well-defined and smooth; see [Amb94, Theorem 3.1]. This allows to define smooth extensions of the normal vector field $\mathbf{n}$ and of any tangential vector field $\boldsymbol{\tau} : \partial\Omega \to \mathbb{R}^d$ to U via the formulas:

$$\mathbf{n}(\mathbf{y}) \equiv \mathbf{n}(p_{\partial\Omega}(\mathbf{y})), \text{ and } \boldsymbol{\tau}(\mathbf{y}) \equiv \boldsymbol{\tau}(p_{\partial\Omega}(\mathbf{y})),$$

respectively. From these notions, we define the mean curvature κ of $\partial\Omega$ by $\kappa = \nabla \cdot \mathbf{n}$.

In this context, the **tangential gradient** $\nabla_{\partial\Omega} f$ of a smooth enough function $f : \partial\Omega \to \mathbb{R}$ is defined by $\nabla_{\partial\Omega} f = \nabla \widetilde{f} - (\nabla \widetilde{f} \cdot \mathbf{n})\mathbf{n}$, where $\widetilde{f}$ is any smooth extension of f to an open neighborhood of $\partial\Omega$.

In the same spirit, the **tangential divergence** $\nabla_{\partial\Omega} \cdot \mathbf{V}$ of a smooth vector field $\mathbf{V} : \partial\Omega \to \mathbb{R}^d$ is defined by $\nabla_{\partial\Omega} \cdot \mathbf{V} := \nabla \cdot \widetilde{\mathbf{V}} - (\nabla \widetilde{\mathbf{V}} \mathbf{n}) \cdot \mathbf{n}$, where $\widetilde{\mathbf{V}}$ is any extension of $\mathbf{V}$ to an open neighborhood of $\partial\Omega$.

Let us finally recall the following integration by parts formulas on the boundary of smooth domains; see [Hen06, Proposition 5.4.9], for the first point, and [Dap13, Section 5.5.4], for the second one.

Theorem V.1.2.2 – Integration by parts using tangential operators.
Let $\Omega \subset \mathbb{R}^d$ be a smooth bounded domain with boundary $\partial\Omega$;

1. Let $u \in H^1(\partial\Omega)$ and $\mathbf{V} \in H^1(\partial\Omega)^d$; then:
$$\int_{\partial\Omega} \nabla_{\partial\Omega} \cdot \mathbf{V} u \, \mathrm{d}s = \int_{\partial\Omega} (-\mathbf{V} \cdot \nabla_{\partial\Omega} u + \kappa u \mathbf{V} \cdot \mathbf{n}) \, \mathrm{d}s$$
2. Let G be a subset of $\partial\Omega$ with smooth boundary Σ, and denote by $\mathbf{n}_\Sigma$ its unit normal vector pointing outward G ($\mathbf{n}_\Sigma$ is a tangent vector field to $\partial\Omega$). Let $u \in H^1(\partial\Omega)$ and $\mathbf{V} \in H^1(\partial\Omega)^d$; then:
$$\int_G \nabla_{\partial\Omega} \cdot \mathbf{V} \, u \, \mathrm{d}s = \int_\Sigma u \mathbf{V} \cdot \mathbf{n}_\Sigma \, \mathrm{d}\ell + \int_G (-\nabla_{\partial\Omega} u \cdot \mathbf{V} + \kappa u \mathbf{V} \cdot \mathbf{n}) \, \mathrm{d}s,$$
where $\mathrm{d}\ell$ denotes integration over the codimension 2 submanifold Σ of $\mathbb{R}^d$.

V.2 Transition between Γ and Γ_N

V.2.1 Main result

Our main result in this section is the following.

Theorem V.2.1.1 – Shape derivative for Neumann-Neumann interfaces.
The functional $\mathcal{J}(\Omega)$ defined by Eq. (V.1.5) is shape differentiable over the admissible set $\mathcal{U}_{\mathrm{NN}}$; its shape derivative reads (volumetric form) for all $\boldsymbol{\theta} \in \Theta_{\mathrm{NN}}$:

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} (j(u_\Omega) - f p_\Omega)\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_\Omega j'(u_\Omega)\nabla u_\Omega\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} \\ + \int_\Omega ((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u_\Omega\cdot\nabla p_\Omega\,\mathrm{d}\mathbf{x} \\ + \int_\Omega f\nabla p_\Omega\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} - \int_{\Gamma_N}((\nabla_{\partial\Omega}\cdot\boldsymbol{\theta})g + \nabla g\cdot\boldsymbol{\theta})p_\Omega\,\mathrm{d}s, \quad \text{(V.2.1)}\end{aligned}$$

where $\nabla_{\partial\Omega}\cdot$ stands for the tangential divergence on $\partial\Omega$ (see Section V.1.2.f), and the adjoint state p_Ω is the unique solution in $H^1_{\Gamma_D}(\Omega)$ to the system:

$$\left\{\begin{array}{rcll} -\Delta p_\Omega &=& -j'(u_\Omega) & \text{in } \Omega, \\ p_\Omega &=& 0 & \text{on } \Gamma_D, \\ \frac{\partial p_\Omega}{\partial\mathbf{n}} &=& 0 & \text{on } \Gamma_N\cup\Gamma. \end{array}\right. \quad \text{(V.2.2)}$$

The above shape derivative has the alternative, surfacic form:

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\Gamma\cup\Gamma_N} j(u_\Omega)\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \int_{\Gamma\cup\Gamma_N}\nabla u_\Omega\cdot\nabla p_\Omega\,\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_{\Gamma\cup\Gamma_N} f p_\Omega\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s \\ - \int_{\Gamma_N}\left(\frac{\partial g}{\partial\mathbf{n}} + \kappa g\right)p_\Omega\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_{\Sigma_N} g p_\Omega\boldsymbol{\theta}\cdot\mathbf{n}_{\Sigma_N}\,\mathrm{d}\ell. \quad \text{(V.2.3)}\end{aligned}$$

Remark V.2.1.1: One comment is in order about the precise meaning of Eq. (V.2.3), and notably that of the term

$$\int_{\Gamma\cup\Gamma_N}\nabla u_\Omega\cdot\nabla p_\Omega\,\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s \quad \text{(V.2.4)}$$

featured in there. The function u_Ω belongs to the space

$$E(\Delta, L^2(\Omega)) := \left\{u\in H^1(\Omega),\ \Delta u\in L^2(\Omega)\right\}, \quad \text{(V.2.5)}$$

and as such, it has a normal trace $\frac{\partial u_\Omega}{\partial\mathbf{n}}\in H^{-1/2}(\partial\Omega)$, which is defined by the following Green's formula valid for all $w\in H^1(\Omega)$:

$$\int_{\partial\Omega}\frac{u_\Omega}{\partial\mathbf{n}}w\,\mathrm{d}s := \int_\Omega \Delta u_\Omega w\,\mathrm{d}\mathbf{x} + \int_\Omega\nabla u_\Omega\cdot\nabla w\,\mathrm{d}\mathbf{x}; \quad \text{(V.2.6)}$$

see [Gri11, Theorem 1.5.3.10], for more details about this point. Also, since $u_\Omega\in H^{1/2}(\partial\Omega)$, the tangential derivative $\frac{\partial u_\Omega}{\partial\boldsymbol{\tau}}$ naturally belongs to the dual space $H^{-1/2}(\partial\Omega)$. On the other hand, the function p_Ω enjoys H^2 regularity on account of elliptic regularity (see Section V.1.2.c), except perhaps near Σ_D where it has a weak singularity of the form Eq. (V.1.11). Since deformations $\boldsymbol{\theta}\in\Theta_{\mathrm{NN}}$ are smooth and vanish identically on Γ_D, the product $(\nabla p_\Omega)\boldsymbol{\theta}\cdot\mathbf{n}$ has a trace in $H^{1/2}(\partial\Omega)$, and so the integral Eq. (V.2.4) is well-defined as a duality product.

V.2.2 The proof

Again, the proof will be an application of the method described in Section II.2.2 with a particular attention devoted to the regularity of u_Ω and p_Ω. Since variations of this argument are used in the following, we present a sketch of it for the reader's convenience. The proof is decomposed into four steps.

Proof of the shape differentiability of $\mathcal{J}(\Omega)$ and derivation of Eq. (V.2.1)

This step amounts to the analysis of the differentiability of the mapping $\boldsymbol{\theta} \mapsto \mathcal{J}(\Omega_{\boldsymbol{\theta}})$ from Θ_{NN} into $\mathbb{R}$, which features the solution $u_{\Omega_{\boldsymbol{\theta}}}$ to the version of Eq. (V.1.2) posed on $\Omega_{\boldsymbol{\theta}}$. The main idea consists in recasting the latter problem as a boundary-value problem on Ω for the transported function $\overline{u_{\boldsymbol{\theta}}} := u_{\Omega_{\boldsymbol{\theta}}} \circ (\mathrm{Id} + \boldsymbol{\theta}) \in H^1(\Omega)$; thence, the implicit function theorem allows to calculate the derivative of the mapping $\boldsymbol{\theta} \mapsto \overline{u_{\boldsymbol{\theta}}}$.

For $\boldsymbol{\theta} \in \Theta_{\mathrm{NN}}$ with norm $\|\boldsymbol{\theta}\|_{\Theta_{\mathrm{NN}}} < 1$, the function $u_{\Omega_{\boldsymbol{\theta}}}$ is the unique solution in $H^1_{\Gamma_D}(\Omega_{\boldsymbol{\theta}})$ to the following variational problem:

$$\forall v \in H^1_{\Gamma_D}(\Omega_{\boldsymbol{\theta}}),\ \int_{\Omega_{\boldsymbol{\theta}}} \nabla u_{\Omega_{\boldsymbol{\theta}}} \cdot \nabla v \,\mathrm{d}\mathbf{x} = \int_{\Omega_{\boldsymbol{\theta}}} f v \,\mathrm{d}\mathbf{x} + \int_{(\Gamma_N)_{\boldsymbol{\theta}}} g v \,\mathrm{d}s.$$

Using test functions of the form $v \circ (\mathrm{Id} + \boldsymbol{\theta})^{-1}$, $v \in H^1_{\Gamma_D}(\Omega)$, a change of variables yields the following variational formulation for $\overline{u_{\boldsymbol{\theta}}} \in H^1_{\Gamma_D}(\Omega)$:

$$\begin{aligned}\forall v \in H^1_{\Gamma_D}(\Omega),\ \int_\Omega A(\boldsymbol{\theta}) \nabla \overline{u_{\boldsymbol{\theta}}} \cdot \nabla v \,\mathrm{d}\mathbf{x} = \int_\Omega |\det(\mathrm{I} + \nabla\boldsymbol{\theta})| f \circ (\mathrm{Id} + \boldsymbol{\theta}) v \,\mathrm{d}\mathbf{x} \\ + \int_{\Gamma_N} |\mathrm{com}(\mathrm{I} + \nabla\boldsymbol{\theta})\mathbf{n}| g \circ (\mathrm{Id} + \boldsymbol{\theta}) v \,\mathrm{d}s,\end{aligned}$$

where $A(\boldsymbol{\theta})$ is the $d \times d$ matrix $A(\boldsymbol{\theta}) = |\det(\mathrm{I} + \nabla\boldsymbol{\theta})|(\mathrm{I} + \nabla\boldsymbol{\theta})^{-1}(\mathrm{I} + \nabla\boldsymbol{\theta})^{-\top}$ and $\mathrm{com}(M)$ stands for the cofactor matrix of a $d \times d$ matrix M. Now introducing the mapping $\mathcal{F} : \Theta_{\mathrm{ad}} \times H^1_{\Gamma_D}(\Omega) \to (H^1_{\Gamma_D}(\Omega))^*$ defined by:

$$\begin{aligned}\forall v \in H^1_{\Gamma_D}(\Omega),\ \mathcal{F}(\boldsymbol{\theta}, u)(v) = \int_\Omega A(\boldsymbol{\theta}) \nabla u \cdot \nabla v \,\mathrm{d}\mathbf{x} - \int_\Omega |\det(\mathrm{I} + \nabla\boldsymbol{\theta})| f \circ (\mathrm{Id} + \boldsymbol{\theta}) v \,\mathrm{d}\mathbf{x} \\ - \int_{\Gamma_N} |\mathrm{com}(\mathrm{I} + \nabla\boldsymbol{\theta})\mathbf{n}| g \circ (\mathrm{Id} + \boldsymbol{\theta}) v \,\mathrm{d}s,\end{aligned}$$

it follows that for "small" $\boldsymbol{\theta} \in \Theta_{\mathrm{NN}}$, $\overline{u}_{\boldsymbol{\theta}}$ is the unique solution of the equation $\mathcal{F}(\boldsymbol{\theta}, \overline{u}_{\boldsymbol{\theta}}) = 0$. Then, the implicit function theorem together with classical calculations (again, see the proof in Section II.2.2) imply that the mapping $\boldsymbol{\theta} \mapsto \overline{u_{\boldsymbol{\theta}}}$ is Fréchet differentiable from a neighborhood of 0 in Θ_{NN} into $H^1_{\Gamma_D}(\Omega)$, and that its derivative $\boldsymbol{\theta} \mapsto \mathring{u}_\Omega(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = 0$ – the so-called Lagrangian derivative of the mapping $\Omega \mapsto u_\Omega$ – is the unique solution to the following variational problem:

$$\begin{aligned}\forall v \in H^1_{\Gamma_D}(\Omega), \int_\Omega \nabla \mathring{u}_\Omega(\boldsymbol{\theta}) \cdot \nabla v \,\mathrm{d}\mathbf{x} = -\int_\Omega ((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top) \nabla u_\Omega \cdot \nabla v \,\mathrm{d}\mathbf{x} \\ + \int_\Omega \nabla\cdot(f\boldsymbol{\theta}) v \,\mathrm{d}\mathbf{x} + \int_{\Gamma_N} ((\nabla_{\partial\Omega}\cdot\boldsymbol{\theta}) g + \nabla g \cdot \boldsymbol{\theta})\, v \,\mathrm{d}s. \quad \text{(V.2.7)}\end{aligned}$$

On the other hand, performing the same change of variables in the definition of $\mathcal{J}(\Omega)$ yields:

$$\mathcal{J}(\Omega_{\boldsymbol{\theta}}) = \int_\Omega |\det(\mathrm{I} + \nabla\boldsymbol{\theta})|\, j(\overline{u_{\boldsymbol{\theta}}}) \,\mathrm{d}\mathbf{x},$$

and so the mapping $\boldsymbol{\theta} \mapsto J(\Omega_{\boldsymbol{\theta}})$ from Θ_{NN} into $\mathbb{R}$ is Fréchet differentiable at $\boldsymbol{\theta} = 0$ with derivative:

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_\Omega (\nabla \cdot \boldsymbol{\theta}\, j(u_\Omega) + j'(u_\Omega)\mathring{u_\Omega}(\boldsymbol{\theta}))\, \mathrm{d}\mathbf{x}. \tag{V.2.8}$$

The material derivative $\mathring{u_\Omega}(\boldsymbol{\theta})$ can now be eliminated from Eq. (V.2.8) thanks to the introduction of the adjoint state p_Ω, solution to Eq. (V.2.2). Indeed, the variational formulation of p_Ω reads:

$$\forall v \in H^1_{\Gamma_D}(\Omega),\ \int_\Omega \nabla p_\Omega \cdot \nabla v\, \mathrm{d}\mathbf{x} = -\int_\Omega j'(u_\Omega) v\, \mathrm{d}\mathbf{x}. \tag{V.2.9}$$

Hence, combining Eqs. (V.2.7) to (V.2.9) yields:

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = &\int_\Omega \nabla \cdot (j(u_\Omega)\boldsymbol{\theta})\, \mathrm{d}\mathbf{x} - \int_\Omega j'(u_\Omega)\nabla u_\Omega \cdot \boldsymbol{\theta}\, \mathrm{d}\mathbf{x} \\ &+ \int_\Omega ((\nabla \cdot \boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u_\Omega \cdot \nabla p_\Omega\, \mathrm{d}\mathbf{x} - \int_\Omega \nabla \cdot (f\boldsymbol{\theta} p_\Omega)\, \mathrm{d}\mathbf{x} \\ &+ \int_\Omega f \nabla p_\Omega \cdot \boldsymbol{\theta}\, \mathrm{d}\mathbf{x} - \int_{\Gamma_N} ((\nabla_{\partial\Omega} \cdot \boldsymbol{\theta})g + \nabla g \cdot \boldsymbol{\theta})\, p_\Omega\, \mathrm{d}s, \end{aligned} \tag{V.2.10}$$

This results in the desired expression Eq. (V.3.1). Note that at this point, we have not used the fact that either u_Ω or p_Ω is more regular than $H^1(\Omega)$.

Derivation of the surface expression Eq. (V.2.3)

This expression is classically achieved from Eq. (V.2.1) using integration by parts; doing so requires a more careful attention to the regularity of u_Ω and p_Ω. Let us notice that the function u_Ω may not be much more regular than just H^1 in the neighborhood of the transition Σ_N. Actually, it belongs to the space $E(\Delta, L^2(\Omega))$, defined by Eq. (V.2.5). The key point is that p_Ω is locally H^2 in the neighborhood of Σ_N, on account of the material in Section V.1.2.c (note that $\frac{\partial p_\Omega}{\partial \mathbf{n}} = 0$ on $\Gamma \cup \Gamma_N$). Relying on the identity Eqs. (II.2.19) and (II.2.20) which holds for all smooth functions $v, w \in \mathcal{C}^\infty_c(\mathbb{R}^d)$,

$$\begin{aligned}\int_\Omega ((\nabla \cdot \boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla v \cdot \nabla w\, \mathrm{d}\mathbf{x} = &\int_\Omega (\Delta v \nabla w + \Delta w \nabla v) \cdot \boldsymbol{\theta}\, \mathrm{d}\mathbf{x} \\ &+ \int_{\Gamma \cup \Gamma_N} \left((\nabla v \cdot \nabla w)\, \boldsymbol{\theta} \cdot \mathbf{n} - \frac{\partial v}{\partial \mathbf{n}} \nabla w \cdot \boldsymbol{\theta} - \frac{\partial w}{\partial \mathbf{n}} \nabla v \cdot \boldsymbol{\theta} \right) \mathrm{d}s, \end{aligned} \tag{V.2.11}$$

and using the density of $\mathcal{C}^\infty_c(\mathbb{R}^d)$ in $E(\Delta, L^2(\Omega))$ and $H^2(\Omega)$ (see [Gri11, Lemma 1.5.3.9]), we obtain:

$$\begin{aligned}\int_\Omega ((\nabla \cdot \boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u_\Omega \cdot \nabla p_\Omega\, \mathrm{d}\mathbf{x} = &\int_\Omega (\Delta u_\Omega \nabla p_\Omega + \Delta p_\Omega \nabla u_\Omega) \cdot \boldsymbol{\theta}\, \mathrm{d}\mathbf{x} \\ &+ \int_{\Gamma \cup \Gamma_N} \left((\nabla u_\Omega \cdot \nabla p_\Omega)\, \boldsymbol{\theta} \cdot \mathbf{n} - \frac{\partial u_\Omega}{\partial \mathbf{n}} \nabla p_\Omega \cdot \boldsymbol{\theta} - \frac{\partial p_\Omega}{\partial \mathbf{n}} \nabla u_\Omega \cdot \boldsymbol{\theta} \right) \mathrm{d}s. \end{aligned} \tag{V.2.12}$$

Let us now work on the last integral in the right-hand side of Eq. (V.2.1); we obtain:

$$\begin{aligned}(\nabla_{\partial\Omega} \cdot \boldsymbol{\theta})g + \nabla g \cdot \boldsymbol{\theta} &= \nabla_{\partial\Omega} \cdot (g\boldsymbol{\theta}) + \tfrac{\partial g}{\partial \mathbf{n}}\, \boldsymbol{\theta} \cdot \mathbf{n}, \\ &= \nabla_{\partial\Omega} \cdot (g(\boldsymbol{\theta} - (\boldsymbol{\theta} \cdot \mathbf{n})\mathbf{n})) + \left(\tfrac{\partial g}{\partial \mathbf{n}} + \kappa g\right) \boldsymbol{\theta} \cdot \mathbf{n}. \end{aligned}$$

Hence, an integration by parts on the region Γ_N using Th. V.1.2.2 yields:

$$\int_{\Gamma_N} ((\nabla_{\partial\Omega}\boldsymbol{\theta})g + \nabla g \cdot \boldsymbol{\theta})\, p_\Omega\, \mathrm{d}s = \int_{\Sigma_N} g p_\Omega\, \boldsymbol{\theta} \cdot \mathbf{n}_{\Sigma_N} \mathrm{d}\ell + \int_{\Gamma_N} \left(\frac{\partial g}{\partial \mathbf{n}} + \kappa g\right) p_\Omega \boldsymbol{\theta} \cdot \mathbf{n}\, \mathrm{d}s. \tag{V.2.13}$$

Combining Eqs. (V.2.12) and (V.2.13) with the volumetric formula Eq. (V.2.1) and using the facts that $-\Delta u_\Omega = f$ and $-\Delta p_\Omega = -j'(u_\Omega)$ in Ω, we finally obtain the desired surface formula Eq. (V.2.3). □

V.3 Transition between Γ and Γ_D

V.3.1 Main result

In this section, we investigate the shape differentiability of the functional $\mathcal{J}(\Omega)$ defined in Eq. (V.1.5) in the particular case where the boundary Σ_D between the regions Γ_D and Γ of $\partial\Omega$ bearing homogeneous Dirichlet and homogeneous Neumann boundary conditions is also subject to optimization; in other terms, we suppose:

$$\mathcal{U}_{\text{ad}} = \mathcal{U}_{\text{DN}}, \text{ and } \Theta_{\text{ad}} = \Theta_{\text{DN}}.$$

The main difficulty of the present situation lies in the weakly singular behavior of the solution u_Ω to Eq. (V.1.2) near Σ_D. In particular, the use of the formal Céa's method, which implicitly relies on the smoothness of u_Ω since it have recourse to its Eulerian derivative (which is defined via Eq. (II.1.14)), gives rise to an erroneous shape derivative in the present context.

The conclusion of Th. V.3.1.1 was already observed in [Aze14; Fre01], but our proof is slightly different: we rely on a direct calculation based on integration by parts.

Theorem V.3.1.1 – Shape derivative for Dirichlet-Neumann interfaces.
The functional $\mathcal{J}(\Omega)$ is shape differentiable at any admissible shape $\Omega \in \mathcal{U}_{\text{DN}}$, and its shape derivative reads (volumetric form) for all $\boldsymbol{\theta} \in \Theta_{\text{DN}}$:

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = &\int_{\partial\Omega} (j(u_\Omega) - f p_\Omega)\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_\Omega j'(u_\Omega)\nabla u_\Omega\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}\\ &+ \int_\Omega ((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u_\Omega\cdot\nabla p_\Omega\,\mathrm{d}\mathbf{x} + \int_\Omega f\nabla p_\Omega\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}, \quad \text{(V.3.1)}\end{aligned}$$

where the adjoint state p_Ω is the unique solution in $H^1_{\Gamma_D}(\Omega)$ to the system:

$$\left\{\begin{array}{rcll} -\Delta p_\Omega &=& -j'(u_\Omega) & \text{in } \Omega,\\ p_\Omega &=& 0 & \text{on } \Gamma_D,\\ \frac{\partial p_\Omega}{\partial \mathbf{n}} &=& 0 & \text{on } \Gamma_N\cup\Gamma. \end{array}\right. \quad \text{(V.3.2)}$$

Moreover, under the assumption Eq. (V.1.3), let us write the local structure of u_Ω and p_Ω in an open neighborhood V^i of s_i, $i = 0, 1$ as follows:

$$u_\Omega = u_s^i + u_r^i \text{ and } p_\Omega = p_s^i + p_r^i; \quad \text{(V.3.3)}$$

in the above formula, $u_r^i, p_r^i \in H^2(V^i)$ and the weakly singular functions u_s^i and $p_s^i \in H^1(V^i)$ have the following expressions in local polar coordinates centered at s_i:

$$u_s^i(r,\nu) = c_u^i r^{\frac{1}{2}}\cos(\nu/2), \text{ and } p_s^i(r,\nu) = c_p^i r^{\frac{1}{2}}\cos(\nu/2), \text{ if } \mathbf{n}_{\Sigma_D}(s_i) = \mathbf{e}_1, \quad \text{(V.3.4)}$$

or

$$u_s^i(r,\nu) = c_u^i r^{\frac{1}{2}}\sin(\nu/2), \text{ and } p_s^i(r,\nu) = c_p^i r^{\frac{1}{2}}\sin(\nu/2), \text{ if } \mathbf{n}_{\Sigma_D}(s_i) = -\mathbf{e}_1, \quad \text{(V.3.5)}$$

where $(\mathbf{e}_1, \mathbf{e}_2)$ is the canonical basis of the plane. Then Eq. (V.3.1) rewrites (surface integral form):

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\Gamma_D \cup \Gamma} (j(u_\Omega) - f p_\Omega)\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_{\Gamma_D} \frac{\partial p_\Omega}{\partial \mathbf{n}}\frac{\partial u_\Omega}{\partial \mathbf{n}}\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \int_{\Gamma} \frac{\partial u_\Omega}{\partial \boldsymbol{\tau}}\frac{\partial p_\Omega}{\partial \boldsymbol{\tau}}\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \frac{\pi}{4}\sum_{i=0,1} c_u^i c_p^i (\boldsymbol{\theta}\cdot\mathbf{n}_{\Sigma_D})(s_i). \quad \text{(V.3.6)}$$

Remark V.3.1.1: In the surface formula Eq. (V.3.6), the integrals

$$-\int_{\Gamma_D} \frac{\partial p_\Omega}{\partial \mathbf{n}}\frac{\partial u_\Omega}{\partial \mathbf{n}}\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \int_{\Gamma} \frac{\partial u_\Omega}{\partial \boldsymbol{\tau}}\frac{\partial p_\Omega}{\partial \boldsymbol{\tau}}\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s$$

are not well-defined individually, since they may blow up around the points s_i, as is quite clear from the look of the structure Eqs. (V.3.4) and (V.3.5) of the singular parts of u_Ω and p_Ω. However, these integrals turn out to have compensating singularities at s_i, so that their sum is well-defined as a Cauchy principal value; see the proof below.

V.3.2 The proof

The calculation of the volumetric formula Eq. (V.3.1) unfolds almost as in the case of Th. V.2.1.1, and we focus on the derivation of the surface formula Eq. (V.3.6), assuming that Eq. (V.1.3) holds. Again, the main idea of the calculation is to perform integration by parts from Eq. (V.3.1), taking advantage of the smoothness of u_Ω and p_Ω far from the singularities at s_i, $i = 0, 1$, and of the local structure Eq. (V.3.3) of these functions in the vicinity of s_i.

Let $\boldsymbol{\theta} \in \Theta_{\mathrm{DN}}$ be fixed; for small $\delta > 0$, let $B^i(\delta) := B(s_i, \delta)$ be the ball centered at s_i with radius δ, and let $\Omega_\delta := \Omega \setminus (\overline{B^0(\delta)} \cup (\overline{B^1(\delta)})$. Since u_Ω and p_Ω belong to $H^1(\Omega)$, it holds from Eq. (V.3.1):

$$\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = \int_{\partial\Omega} (j(u_\Omega) - f p_\Omega)\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \lim_{\delta\to 0} I_\delta,$$

where:

$$I_\delta := -\int_{\Omega_\delta} j'(u_\Omega)\nabla u_\Omega\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} + \int_{\Omega_\delta} ((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u_\Omega\cdot\nabla p_\Omega\,\mathrm{d}\mathbf{x} + \int_{\Omega_\delta} f\nabla p_\Omega\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}.$$

Using the smoothness of u_Ω and p_Ω on Ω_δ, the fact that $\boldsymbol{\theta}$ vanishes on Γ_N, the definitions of u_Ω, p_Ω and an integration by parts on the second term in the above right-hand side based on the identity Eq. (V.2.11), we obtain:

$$I_\delta = \int_{\partial\Omega_\delta} \left(\nabla u_\Omega\cdot\nabla p_\Omega\,\boldsymbol{\theta}\cdot\mathbf{n} - \frac{\partial u_\Omega}{\partial \mathbf{n}}\nabla p_\Omega\cdot\boldsymbol{\theta} - \frac{\partial p_\Omega}{\partial \mathbf{n}}\nabla u_\Omega\cdot\boldsymbol{\theta}\right)\mathrm{d}s. \quad \text{(V.3.7)}$$

To proceed further, we decompose the boundary $\partial\Omega_\delta$ as the disjoint reunion:

$$\partial\Omega_\delta = \Gamma_{D,\delta} \cup \Gamma_\delta \cup \Gamma_N \cup \gamma_{0,\delta} \cup \gamma_{1,\delta},$$

where $\Gamma_{D,\delta} = \Gamma_D \cap \Omega_\delta$, $\Gamma_\delta = \Gamma \cap \Omega_\delta$, and $\gamma_{i,\delta} = \partial B^i(\delta) \cap \Omega$ is the half-circle with center s_i and radius δ. We also denote by $s_{i,\delta}$ the intersection point between $\partial B^i(\delta)$ and Γ; see Fig. V.3.1 about these notations.

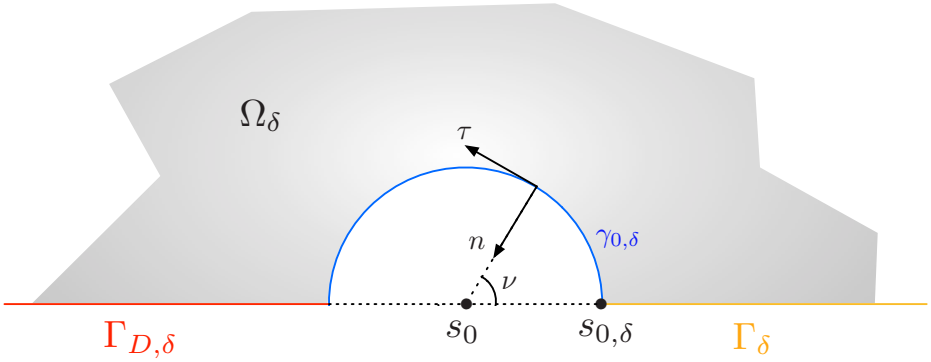


Figure V.3.1: The local situation around the point s_0 in the proof of Th. V.3.1.1.

Since $\boldsymbol{\theta} = 0$ on Γ_N, it follows that:

$$I_\delta = \int_{\Gamma_\delta} \frac{\partial u_\Omega}{\partial \boldsymbol{\tau}} \frac{\partial p_\Omega}{\partial \boldsymbol{\tau}} \boldsymbol{\theta} \cdot \mathbf{n} \, \mathrm{d}s - \int_{\Gamma_{D,\delta}} \frac{\partial u_\Omega}{\partial \mathbf{n}} \frac{\partial p_\Omega}{\partial \mathbf{n}} \boldsymbol{\theta} \cdot \mathbf{n} \, \mathrm{d}s + \sum_{i=0,1} \int_{\gamma_{i,\delta}} K(u_\Omega, p_\Omega) \, \mathrm{d}s, \qquad \text{(V.3.8)}$$

where we have introduced the shorthand:

$$K(v, w) = \nabla v \cdot \nabla w \, \boldsymbol{\theta} \cdot \mathbf{n} - \frac{\partial v}{\partial \mathbf{n}} \nabla w \cdot \boldsymbol{\theta} - \frac{\partial w}{\partial \mathbf{n}} \nabla v \cdot \boldsymbol{\theta}.$$

Let us now evaluate the contributions of I_δ on $\gamma_{i,\delta}$ for $i = 0, 1$ in the expression Eq. (V.3.8). Without loss of generality, we only deal with $i = 0$, and we assume that $s_0 = 0$; according to Eq. (V.1.3), $\partial\Omega$ is horizontal in the neighborhood of s_0 and we also assume that Γ_D (resp. Γ) lies on the left-hand side (resp. the right-hand side) of s_0; see again Fig. V.3.1. Introducing the polar coordinates (r, ν) with origin at $s_0 = 0$, taking into account our conventions for $\boldsymbol{\tau}$ and $\mathbf{n}$, we have:

$$\mathbf{n} = -\cos\nu \, \mathbf{e}_1 - \sin\nu \, \mathbf{e}_2, \quad \boldsymbol{\tau} = -\sin\nu \, \mathbf{e}_1 + \cos\nu \, \mathbf{e}_2,$$

and as far as derivatives are concerned $\frac{\partial}{\partial \mathbf{n}} = -\frac{\partial}{\partial r}$ and $\frac{\partial}{\partial \boldsymbol{\tau}} = \frac{1}{r}\frac{\partial}{\partial \nu}$. Recalling the local expressions Eq. (V.3.3) of u_Ω and p_Ω around s_0, we can see that the only possibly non vanishing contribution of $\int_{\gamma_{0,\delta}} K(u_\Omega, p_\Omega) \, \mathrm{d}s$ in the limit $\delta \to 0$ is given by the most singular part of its integrand:

$$\lim_{\delta \to 0} \int_{\gamma_{0,\delta}} K(u_\Omega, p_\Omega) \, \mathrm{d}s = \lim_{\delta \to 0} \widetilde{I_\delta^0}, \quad \text{where } \widetilde{I_\delta^0} := \int_{\gamma_{0,\delta}} K\left(u_s^0, p_s^0\right) \, \mathrm{d}s.$$

Let us then calculate the last integral. We have, on $\gamma_{0,\delta}$:

$$\begin{aligned} \nabla u_s^0 \cdot \nabla p_s^0 \boldsymbol{\theta} \cdot \mathbf{n} &= \left(\frac{\partial u_s^0}{\partial \boldsymbol{\tau}} \frac{\partial p_s^0}{\partial \boldsymbol{\tau}} + \frac{\partial u_s^0}{\partial \mathbf{n}} \frac{\partial p_s^0}{\partial \mathbf{n}} \right) \boldsymbol{\theta} \cdot \mathbf{n} \\ &= \left(\frac{1}{r^2} \frac{\partial u_s^0}{\partial \nu} \frac{\partial p_s^0}{\partial \nu} + \frac{\partial u_s^0}{\partial r} \frac{\partial p_s^0}{\partial r} \right) \boldsymbol{\theta} \cdot \mathbf{n} = \frac{c_u^0 c_p^0}{4r} \boldsymbol{\theta} \cdot \mathbf{n}. \qquad \text{(V.3.9)} \end{aligned}$$

Likewise,

$$\begin{aligned} -\frac{\partial u_s^0}{\partial \mathbf{n}} \nabla p_s^0 \cdot \boldsymbol{\theta} &= -\frac{\partial u_s^0}{\partial \mathbf{n}} \frac{\partial p_s^0}{\partial \boldsymbol{\tau}} \boldsymbol{\theta} \cdot \boldsymbol{\tau} - \frac{\partial u_s^0}{\partial \mathbf{n}} \frac{\partial p_s^0}{\partial \mathbf{n}} \boldsymbol{\theta} \cdot \mathbf{n}, \\ &= -\frac{c_u^0 c_p^0}{4r} \cos(\nu/2) \sin(\nu/2) \boldsymbol{\theta} \cdot \boldsymbol{\tau} - \frac{c_u^0 c_p^0}{4r} \cos^2(\nu/2) \boldsymbol{\theta} \cdot \mathbf{n}, \qquad \text{(V.3.10)} \end{aligned}$$

and

$$-\frac{\partial p_s^0}{\partial \mathbf{n}}\nabla u_s^0\cdot\boldsymbol{\theta} = -\frac{\partial u_s^0}{\partial \mathbf{n}}\nabla p_s^0\cdot\boldsymbol{\theta}$$
$$= -\frac{c_u^0 c_p^0}{4r}\cos(\nu/2)\sin(\nu/2)\boldsymbol{\theta}\cdot\boldsymbol{\tau} - \frac{c_u^0 c_p^0}{4r}\cos^2(\nu/2)\boldsymbol{\theta}\cdot\mathbf{n}. \quad \text{(V.3.11)}$$

Gathering Eqs. (V.3.9) to (V.3.11), we now obtain:

$$\begin{aligned} K(u_s^0,p_s^0) &= \frac{c_u^0 c_p^0}{4r}\left((1-2\cos^2(\nu/2))\boldsymbol{\theta}\cdot\mathbf{n} - 2\cos(\nu/2)\sin(\nu/2)\boldsymbol{\theta}\cdot\boldsymbol{\tau}\right), \\ &= \frac{c_u^0 c_p^0}{4r}\left(-(\sin\nu)\boldsymbol{\theta}\cdot\boldsymbol{\tau} - (\cos\nu)\boldsymbol{\theta}\cdot\mathbf{n}\right) = \frac{c_u^0 c_p^0}{4r}\theta_1, \end{aligned}$$

where θ_1 is the horizontal component of $\boldsymbol{\theta} = \theta_1\mathbf{e}_1 + \theta_2\mathbf{e}_2$. Therefore:

$$\widetilde{I_\delta^0} = \left(\int_0^\pi \frac{c_u^0 c_p^0}{4}\theta_1(\delta\cos\nu, \delta\sin\nu)\,\mathrm{d}\nu\right) \xrightarrow{\delta\to 0} \frac{\pi c_u^0 c_p^0}{4}\theta_1(0).$$

Combining all these results, we obtain the surface form Eq. (V.3.6) of the shape derivative $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$. □

Remark V.3.2.1: The result extends to the case where the boundary $\partial\Omega$ is not flat (but is still smooth) in the neighborhood of $\Sigma_D = \{s_0, s_1\}$. More precisely, let V be a small enough neighborhood of either of the s_i, and let us introduce a local description of Ω as a graph, assuming for simplicity that $s_i = 0$ and $\mathbf{n}(s_i) = -\mathbf{e}_2$: U is a neighborhood of 0 in $\mathbb{R}^2$ and $\psi(x_1,x_2) = (x_1, \varphi(_2))$ is a smooth diffeomorphism from U onto V such that:

$$\Omega\cap V = \left\{\mathbf{x} = (x_1,x_2)\in\mathbb{R}^2,\ \ x_2 > \varphi(x_1)\right\}\cap U.$$

Then, it follows from [Gri11, Section 5.2], that u_Ω reads in this case:

$$u_\Omega = c^i S^i\circ\psi^{-1} + u_r^i \text{ on } B_\delta(s_i),$$

where $u_r^i \in H^2(V)$. The proof extends to this latter context then.

Remark V.3.2.2: Interestingly, if u_Ω and p_Ω are assumed to be actually smooth (say H^2) in the neighborhood of the transition points s_0 and s_1, the shape derivative Eq. (V.3.6) no longer involves any term related to the geometry of the repartition of Γ_D and Γ. In other terms, all the information about the sensitivity of $\mathcal{J}(\Omega)$ with respect to the repartition of Γ_D and Γ is encoded in the (weak) singularities of u_Ω and p_Ω.

V.4 An approximate model to deal with the Dirichlet-Neumann transition

V.4.1 Regularization

We have calculated in Sections V.2 and V.3 the shape derivative of the functional $\mathcal{J}(\Omega)$ given by Eq. (V.1.5), in the situation where either the transition Σ_N or Σ_D is subject

to optimization. The resulting expression in the former case (see Th. V.2.1.1) may be readily used in a typical gradient-based shape optimization algorithm; see Section V.5.

On the other hand, in the case where Σ_D is optimized, the expression supplied by Th. V.3.1.1 is unfortunately awkward from both the theoretical and practical perspectives. Indeed,

- The calculation of the surface form Eq. (V.3.6) of $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ was enabled by the precise knowledge of the local behavior Eq. (V.3.3) of u_Ω and p_Ω near the singularities s_0 and s_1. In more involved situations, for instance in three space dimensions, or in more challenging physical contexts (such as those of the linearized elasticity system, or the Stokes equations), such precise information may not be available, or may be difficult to handle.

- The numerical evaluation of the shape derivative $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ requires the calculation of the coefficients c_u and c_p featured in Eq. (V.3.3); this is doable, but it demands adapted numerical techniques, for instance an enrichment of the finite element basis with the singular functions, or adapted p/hp finite mesh refinement methods; see [Bab90; Ell05; Li00] and the references therein. In our numerical setting, presented in Section V.5, such techniques are bound to be all the more difficult to carry out that the boundary Σ_D is not explicitly discretized in the computational mesh.

- Eventually, it is possible in principle to rely only on the volumetric form Eq. (V.3.1) of the shape derivative for algorithmic purposes, as is suggested for instance in [Hip15; Gia17] and the references therein; nevertheless, for many practical purposes, it is interesting to have a surface expression for this shape derivative - for instance when it comes to using advanced optimization algorithms such as that introduced in [Dun15].

We thenceforth focus our efforts on the instance of the problem Eq. (V.1.4) where this transition zone Σ_D is also subject to optimization (while Σ_N is not). To overcome the aforementioned difficulties, we introduce an approximation method which allows for the optimization of the boundary Σ_D between regions bearing homogeneous Dirichlet and Neumann boundary conditions, without requiring the knowledge of the weakly singular behavior of u_Ω (and that of the adjoint state p_Ω) in the neighborhood of Σ_D. As we shall see in Section V.5, this method lends itself to an easy generalization to more difficult situations: transitions between other types of boundary conditions involving a singular state, other physical contexts than that of the Laplace equation, etc.

Throughout this section, unless stated otherwise, we consider the shape optimization problem Eq. (V.1.4) in the physical setting of Section V.1.2, in the particular case where the transition Σ_D between the regions Γ_D and Γ of $\partial\Omega$ is subject to optimization: $\mathcal{U}_{\text{ad}} = \mathcal{U}_{\text{DN}}$ and $\Theta_{\text{ad}} = \Theta_{\text{DN}}$. After introducing a few notations and background material regarding the notion of geodesic distance function in Section V.4.1.a, in Section V.4.1.b, we present an approximate version of the physical problem Eq. (V.1.2), relying on a "small" parameter $\varepsilon > 0$, with the noticeable feature that its unique solution $u_{\Omega,\varepsilon}$ is smooth. This leads to the introduction of an approximate shape optimization problem of a smoothed functional $\mathcal{J}_\varepsilon(\Omega)$ in Section V.4.2.a; we calculate the shape derivative $\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta})$ by classical means, and the numerical evaluation of the resulting expression poses no particular difficulty. Finally, in Section V.4.2.b, we prove in the model context where Eq. (V.1.3) holds that the approximate shape optimization problem converges to its exact counterpart, in the

sense that $u_{\Omega,\varepsilon} \to u_\Omega$ as $\varepsilon \to 0$, and the values of $\mathcal{J}_\varepsilon(\Omega)$ and $\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta})$ converge to their exact counterparts $\mathcal{J}(\Omega)$ and $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$.

V.4.1.a About the signed distance function to a subset on a submanifold

This section is a concise summary of the result about the signed distance function on a submanifold presented in our paper [Dap19, Section 4.1].

Let $\mathcal{M}$ be an oriented, closed and smooth submanifold of $\mathbb{R}^d$ with codimension 1. $\mathcal{M}$ is equipped with a Riemannian structure by endowing its tangent bundle with the inner product induced by that of $\mathbb{R}^d$ and we denote by $\mathbf{n}$ its unit normal vector. In the context of Section V.1.2, $\mathcal{M}$ stands for the boundary $\partial\Omega$ of the considered shape Ω.

The length $\ell(\gamma)$ of a piecewise differentiable curve $\gamma : I \to \mathcal{M}$ defined on an interval $I \subset \mathbb{R}$ is

$$\ell(\gamma) = \int_I |\gamma'(t)| \, \mathrm{d}t.$$

The geodesic distance $d^{\mathcal{M}}(\mathbf{x}, \mathbf{y})$ between two points $\mathbf{x}, \mathbf{y} \in \mathcal{M}$ is then:

$$d^{\mathcal{M}}(\mathbf{x}, \mathbf{y}) = \inf \ell(\gamma),$$

where the infimum is taken over all piecewise differentiable curves $\gamma : (a, b) \to \mathcal{M}$ such that $\gamma(a) = \mathbf{x}$ and $\gamma(b) = \mathbf{y}$. Likewise, we denote by

$$d^{\mathcal{M}}(\mathbf{x}, K) = \inf_{\mathbf{y} \in K} d^{\mathcal{M}}(\mathbf{x}, \mathbf{y})$$

the distance between $\mathbf{x} \in \mathcal{M}$ and a subset $K \subset \mathcal{M}$.

We now turn to the notion of signed distance function on the submanifold $\mathcal{M}$: let $G \subset \mathcal{M}$ be an open subset which we assume to be smooth for simplicity; its boundary $\Sigma := \partial G$ is a closed, smooth submanifold of $\mathbb{R}^d$ with codimension 2, and we denote by $\mathbf{n}_\Sigma : \Sigma \to \mathbb{S}^1$ the unit normal vector to Σ pointing outward G. In particular, $\mathbf{n}_\Sigma$ is a vector field along Σ which is tangential to $\mathcal{M}$.

Definition V.4.1.1 – Signed distance function on a submanifold.
The signed distance function d_G to G is defined by:

$$\forall \mathbf{x} \in \mathcal{M}, \quad d_G(\mathbf{x}) = \begin{cases} -d^{\mathcal{M}}(\mathbf{x}, \Sigma) & \text{if } \mathbf{x} \in G, \\ 0 & \text{if } \mathbf{x} \in \Sigma, \\ d^{\mathcal{M}}(\mathbf{x}, \Sigma) & \text{if } \mathbf{x} \in \mathcal{M} \setminus \overline{G}. \end{cases}$$

For $\mathbf{y} \in \mathcal{M}$, we denote by $p_\Sigma(\mathbf{y})$ the projection of $\mathbf{y}$ onto Σ, that is, the unique point $\mathbf{x} \in \Sigma$ such that $d^{\mathcal{M}}(\mathbf{x}, \mathbf{y}) = d^{\mathcal{M}}(\mathbf{x}, \Sigma)$, when this makes sense (i.e. when there is indeed such a unique point).
We eventually consider the differentiation of the signed distance function d_G with respect to variations of the manifold $\mathcal{M}$ (and thus of Σ). The following result is new to the best of our knowledge (note here that for simplicity we only give the result concerning the Eulerian derivative of the signed geodesic distance function and we refer to our paper for details concerning the Lagrangian derivative):

Theorem V.4.1.1 – Eulerian derivative of the geodesic signed distance function.

When it is defined, the Eulerian derivative $d'_G(\boldsymbol{\theta})(\mathbf{y})$ of the geodesic signed distance function d_G at $\mathbf{y}$ is defined by the formula:

$$d'_G(\boldsymbol{\theta})(\mathbf{y}) = -\boldsymbol{\theta}(\mathbf{p}) \cdot \mathbf{n}_\Sigma(\mathbf{p}) + \int_0^{d_G(\mathbf{y})} \Pi^{\mathcal{M}}_{\sigma(t)}(\sigma'(t), \sigma'(t))\, (\boldsymbol{\theta} \cdot \mathbf{n})(\sigma(t))\, \mathrm{d}t,$$

where $\mathbf{p} = p_\Sigma \mathbf{y}$, $\sigma(t)$ is the geodesic curve joining $\mathbf{p}$ to $\mathbf{y}$ and $\Pi^{\mathcal{M}}_{\mathbf{p}}$ is the second fundamental form of $\mathcal{M}$ at $\mathbf{p}$, that is:

$$\forall \mathbf{v} \in T_{\mathbf{p}}\mathcal{M},\ \ \Pi^{\mathcal{M}}_{\mathbf{p}}\mathbf{v} \cdot \mathbf{v} = -\nabla \mathbf{n}(\mathbf{p})\mathbf{v} \cdot \mathbf{v},$$

where $\mathbf{n}$ is any extension of the normal vector $\mathbf{n} : \mathcal{M} \to \mathbb{R}^d$ to an open neighborhood of $\mathcal{M}$ in $\mathbb{R}^d$.

Remark V.4.1.1: The structure of Th. V.4.1.1 is quite intuitive: the first term is exactly the one featured in the formula for the Eulerian derivative of the signed distance function in the Euclidean case, i.e. without taking into account the curvature of the ambient space (see e.g. [Dap13, Section 4.2]), while the second one expresses the deformation with respect to $\boldsymbol{\theta}$ of the geodesic between $\mathbf{p}$ and $\mathbf{y}$ out of the (normal) variation of the manifold $\mathcal{M}$.

V.4.1.b The smoothed setting

In the setting of Section V.1.2 (see also Fig. V.1.2), and following the works [All14b; Des18], we trade the solution u_Ω to the "exact" problem Eq. (V.1.2) for that $u_{\Omega,\varepsilon}$ to the following approximate version, where the homogeneous Dirichlet and Neumann boundary conditions on Γ_D and Γ respectively are replaced by a Robin boundary condition on $\Gamma_D \cup \Gamma$:

$$\left\{ \begin{array}{rcll} -\Delta u_{\Omega,\varepsilon} & = & f & \text{in } \Omega, \\ \frac{\partial u_{\Omega,\varepsilon}}{\partial \mathbf{n}} + h_\varepsilon u_{\Omega,\varepsilon} & = & 0 & \text{on } \Gamma \cup \Gamma_D, \\ \frac{\partial u_{\Omega,\varepsilon}}{\partial \mathbf{n}} & = & g & \text{on } \Gamma_N. \end{array} \right. \tag{V.4.1}$$

Here, the coefficient h_ε is defined by:

$$h_\varepsilon(\mathbf{x}) = \frac{1}{\varepsilon} h\left(\frac{d_{\Gamma_D}(\mathbf{x})}{\varepsilon}\right), \tag{V.4.2}$$

where $h : \mathbb{R} \to \mathbb{R}$ is a smooth function with the properties:

$$0 \leq h \leq 1,\ \ h \equiv 1 \text{ on } (-\infty, -1],\ \ h(0) > 0,\ \ h \equiv 0 \text{ on } [1, \infty),$$

and $d_{\Gamma_D}(\mathbf{x})$ is the (geodesic) signed distance function to Γ_D on the surface $\partial\Omega$; see Def. V.4.1.1. In other words, h_ε equals $1/\varepsilon$ inside Γ_D, "far" from Σ_D, 0 on Γ "far" from Σ_D, and it presents a smooth transition between these two values in a tubular neighborhood of Σ_D with (geodesic) width ε, so that the boundary conditions in Eq. (V.1.2) are approximately satisfied; see Fig. V.4.1.

Remark V.4.1.2: Note here that the regularization function Eq. (V.4.2) is the same as the one explained in Remark III.2.2.2 and different than the one used to smooth

the optical index in Section III.2.2.a. We preferred to rely here on the signed distance function rather a convolution product since it seems to be more coherent in the case of curved structures.

In particular, h_ε vanishes on a neighborhood of Γ_N in $\partial\Omega$; notice also that our assumptions on h imply that there exists a real value $\alpha > 0$ which is independent of ε such that:

$$\forall \mathbf{x} \in \Gamma_D, \ \ \alpha \leq \varepsilon h_\varepsilon(\mathbf{x}). \tag{V.4.3}$$

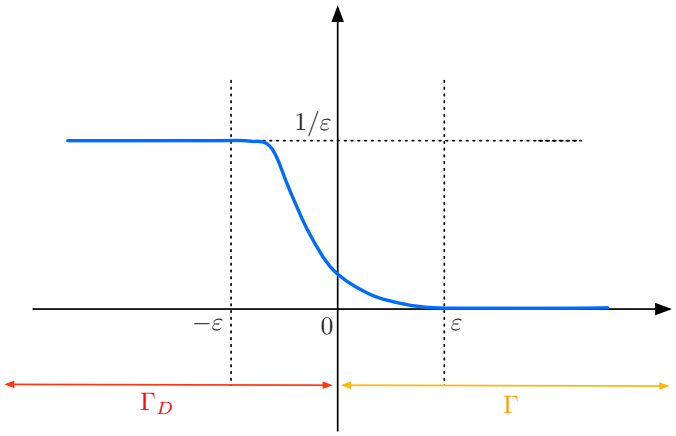


Figure V.4.1: Graph of the function h_ε defined by Eq. (V.4.2).

The variational formulation associated to Eq. (V.4.1) reads: $u_{\Omega,\varepsilon}$ is the unique function in $H^1(\Omega)$ such that

$$\forall v \in H^1(\Omega), \ \int_\Omega \nabla u_{\Omega,\varepsilon} \cdot \nabla v \, \mathrm{d}\mathbf{x} + \int_{\Gamma_D \cup \Gamma} h_\varepsilon u_{\Omega,\varepsilon} v \, \mathrm{d}s = \int_\Omega f v \, \mathrm{d}\mathbf{x} + \int_{\Gamma_N} g v \, \mathrm{d}s. \tag{V.4.4}$$

It follows from the standard Lax-Milgram theory that this problem is well-posed. In addition, for a fixed value of $\varepsilon > 0$, due to the smoothness of Ω and h_ε (see Section V.4.1.a about the smoothness of d_{Γ_D}), the solution $u_{\Omega,\varepsilon}$ to Eq. (V.4.1) actually enjoys H^2 regularity on a neighborhood of Σ_D, as a consequence of the standard regularity theory for elliptic equations; see e.g. [Bre10, Chapter 9].

Remark V.4.1.3: This approximation method can be straightforwardly adapted to different contexts, such as that of the linearized elasticity system; see Section V.5 for illustrations, and [Des18] for an adaptation in the context of the acoustic Helmholtz equation.

V.4.2 Regularized shape derivative

V.4.2.a The approximate shape optimization problem

We propose to replace the exact shape optimization problem Eq. (V.1.4) by its regularized counterpart:

$$\inf_{\Omega \in \mathcal{U}_{\mathrm{DN}}} \mathcal{J}_\varepsilon(\Omega), \ \text{where } \mathcal{J}_\varepsilon(\Omega) := \int_\Omega j(u_{\Omega,\varepsilon}) \, \mathrm{d}\mathbf{x}, \tag{V.4.5}$$

and $u_{\Omega,\varepsilon}$ is the solution to Eq. (V.4.1).

When it comes to the shape derivative of $\mathcal{J}_\varepsilon(\Omega)$, the result of interest is the following:

Theorem V.4.2.1 – Shape derivative of the regularized model.
The functional $\mathcal{J}_\varepsilon(\Omega)$ is shape differentiable at any admissible shape $\Omega \in \mathcal{U}_{\mathrm{DN}}$, and its shape derivative reads, for arbitrary $\boldsymbol{\theta} \in \Theta_{\mathrm{DN}}$ (volumetric form):

$$\begin{aligned}
\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta}) = &\int_{\partial\Omega} (j(u_{\Omega,\varepsilon}) - f p_{\Omega,\varepsilon})\, \boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_\Omega j'(u_{\Omega,\varepsilon})\nabla u_{\Omega,\varepsilon}\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}\\
&+\int_\Omega \left((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top\right)\nabla u_{\Omega,\varepsilon}\cdot\nabla p_{\Omega,\varepsilon}\,\mathrm{d}\mathbf{x}\\
&+\int_{\Gamma\cup\Gamma_D} \nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\, h_\varepsilon u_{\Omega,\varepsilon} p_{\Omega,\varepsilon}\,\mathrm{d}s + \int_\Omega f\nabla p_{\Omega,\varepsilon}\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}\\
+\frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right)&\left(-\boldsymbol{\theta}(\mathbf{x})\cdot\frac{\log_{\mathbf{x}}(p_{\Sigma_D}(\mathbf{x}))}{d_{\Gamma_D}(\mathbf{x})} - \boldsymbol{\theta}(p_{\Sigma_D}(\mathbf{x}))\cdot\mathbf{n}_{\Sigma_D}(p_{\Sigma_D}(\mathbf{x}))\right) u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s(\mathbf{x})\\
&\frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right)\left(\int_0^{d_{\Gamma_D}(\mathbf{x})} \Pi^{\partial\Omega}_{\sigma_\mathbf{x}(t)}(\sigma'_\mathbf{x}(t),\sigma'_\mathbf{x}(t))\,(\boldsymbol{\theta}\cdot\mathbf{n})(\sigma_\mathbf{x}(t))\,\mathrm{d}t\right) u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s(\mathbf{x}),
\end{aligned} \tag{V.4.6}$$

where $\sigma_\mathbf{x}(t)$ is the unique geodesic passing through $\mathbf{x}$ and $p_{\Sigma_D}(\mathbf{x})$, and the adjoint state $p_{\Omega,\varepsilon}$ is the unique solution in $H^1(\Omega)$ to the following system:

$$\left\{\begin{array}{rcll} -\Delta p_{\Omega,\varepsilon} &=& -j(u_{\Omega,\varepsilon}) & \text{in } \Omega,\\ \frac{\partial p_{\Omega,\varepsilon}}{\partial\mathbf{n}} + h_\varepsilon p_{\Omega,\varepsilon} &=& 0 & \text{on } \Gamma_D\cup\Gamma,\\ \frac{\partial p_{\Omega,\varepsilon}}{\partial\mathbf{n}} &=& 0 & \text{on } \Gamma_N. \end{array}\right. \tag{V.4.7}$$

Equivalently, this rewrites (surface form):

$$\begin{aligned}
\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta}) = &\int_{\Gamma\cup\Gamma_D}\Big[j(u_{\Omega,\varepsilon}) - f p_{\Omega,\varepsilon} + \nabla_{\partial\Omega}u_{\Omega,\varepsilon}\cdot\nabla_{\partial\Omega}p_{\Omega,\varepsilon}\\
&-\frac{\partial u_{\Omega,\varepsilon}}{\partial\mathbf{n}}\frac{\partial p_{\Omega,\varepsilon}}{\partial\mathbf{n}} - \kappa p_{\Omega,\varepsilon}\frac{\partial u_{\Omega,\varepsilon}}{\partial\mathbf{n}}\Big]\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right)\Big[-\boldsymbol{\theta}(p_{\Sigma_D}(\mathbf{x}))\cdot\mathbf{n}_{\Sigma_D}(p_{\Sigma_D}(\mathbf{x}))\\
&+\int_0^{d_{\Gamma_D}(\mathbf{x})}\Pi^{\partial\Omega}_{\sigma_\mathbf{x}(t)}(\sigma'_\mathbf{x}(t),\sigma'_\mathbf{x}(t))\,(\boldsymbol{\theta}\cdot\mathbf{n})(\sigma_\mathbf{x}(t))\,\mathrm{d}t\Big]u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s(\mathbf{x}).
\end{aligned} \tag{V.4.8}$$

Proof of Eq. (V.4.6): The proof is very similar to that of the volumetric formula Eq. (V.2.1) in Th. V.3.1.1, and we only sketch the main ingredients. At first, using the implicit function theorem, one sees that the solution $u_{\Omega,\varepsilon}$ to Eq. (V.4.1) has a Lagrangian derivative $\mathring{u}_{\Omega,\varepsilon}(\boldsymbol{\theta})$, which is the unique solution in $H^1(\Omega)$ to the following variational problem: for all $v \in H^1(\Omega)$,

$$\begin{aligned}
&\int_\Omega \nabla\mathring{u}_{\Omega,\varepsilon}(\boldsymbol{\theta})\cdot\nabla v\,\mathrm{d}\mathbf{x} + \int_{\Gamma\cup\Gamma_D} h_\varepsilon(d_{\Gamma_D})\mathring{u}_{\Omega,\varepsilon}(\boldsymbol{\theta})v\,\mathrm{d}s =\\
&\int_\Omega (\nabla\cdot(f\boldsymbol{\theta})v + (\nabla\boldsymbol{\theta} + \nabla\boldsymbol{\theta}^\top - (\nabla\cdot\boldsymbol{\theta})\mathrm{I})\nabla u_{\Omega,\varepsilon}\cdot\nabla v)\,\mathrm{d}\mathbf{x} - \int_{\Gamma\cup\Gamma_D}(\nabla_{\partial\Omega}\cdot\boldsymbol{\theta})\,h_\varepsilon(d_{\Gamma_D})u_{\Omega,\varepsilon}v\,\mathrm{d}s\\
&-\frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right)\left(-\boldsymbol{\theta}(\mathbf{x})\cdot\frac{\log_{\mathbf{x}}(p_{\Sigma_D}(\mathbf{x}))}{d_{\Gamma_D}(\mathbf{x})} - \boldsymbol{\theta}(p_{\Sigma_D}(\mathbf{x}))\cdot\mathbf{n}_{\Sigma_D}(p_{\Sigma_D}(\mathbf{x}))\right)u_{\Omega,\varepsilon}v\,\mathrm{d}s(\mathbf{x})\\
&-\frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right)\left(\int_0^{d_{\Sigma_D}(\mathbf{x})}\Pi^{\partial\Omega}_{\gamma(t)}(\gamma'(t),\gamma'(t))\,(\boldsymbol{\theta}\cdot\mathbf{n})(\gamma(t))\,\mathrm{d}t\right)u_{\Omega,\varepsilon}v\,\mathrm{d}s(\mathbf{x}),
\end{aligned} \tag{V.4.9}$$

where we have used Th. V.4.1.1 for the Lagrangian derivative of the geodesic distance.

On the other hand, using a change of variables yields:

$$\mathcal{J}_\varepsilon(\Omega_{\boldsymbol{\theta}}) = \int_\Omega |\det(\mathrm{Id} + \nabla\boldsymbol{\theta})| j(u_{\Omega_{\boldsymbol{\theta}},\varepsilon} \circ (\mathrm{Id} + \boldsymbol{\theta}))\, \mathrm{d}\mathbf{x},$$

whence, differentiating with respect to $\boldsymbol{\theta}$ and using the variational formulation for the adjoint system Eq. (V.4.7):

$$\begin{aligned} \mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta}) &= \int_\Omega \nabla\cdot\boldsymbol{\theta} j(u_{\Omega,\varepsilon})\,\mathrm{d}\mathbf{x} + \int_\Omega j'(u_{\Omega,\varepsilon})\mathring{u}_{\Omega,\varepsilon}(\boldsymbol{\theta})\,\mathrm{d}\mathbf{x}, \\ &= \int_\Omega \nabla\cdot\boldsymbol{\theta} j(u_{\Omega,\varepsilon})\,\mathrm{d}\mathbf{x} - \int_\Omega \nabla\mathring{u}_{\Omega,\varepsilon}(\boldsymbol{\theta})\cdot\nabla p_{\Omega,\varepsilon}\,\mathrm{d}\mathbf{x} \int_{\Gamma\cup\Gamma_D} h_\varepsilon(d_{\Gamma_D})\mathring{u}_{\Omega,\varepsilon}(\boldsymbol{\theta}) p_{\Omega,\varepsilon}\,\mathrm{d}s, \end{aligned} \tag{V.4.10}$$

Combining this with Eq. (V.4.9) eventually yields the desired result.

Proof of Eq. (V.4.8): To simplify the notations, until the end of the proof, we take the shortcuts $u \equiv u_{\Omega,\varepsilon}$ and $p \equiv p_{\Omega,\varepsilon}$. We decompose the volumetric expression Eq. (V.4.6) as:

$$\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta}) = I_1(\boldsymbol{\theta}) + I_2(\boldsymbol{\theta}),$$

where

$$\begin{aligned} I_1(\boldsymbol{\theta}) &= \int_{\partial\Omega} (j(u) - fp)\,\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_\Omega j'(u)\nabla u\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} \\ &\quad + \int_\Omega ((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u\cdot\nabla p\,\mathrm{d}\mathbf{x} + \int_{\Gamma\cup\Gamma_D} \nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\, h_{\Omega,\varepsilon} up\,\mathrm{d}s + \int_\Omega f\nabla p\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}, \end{aligned}$$

and

$$\begin{aligned} I_2(\boldsymbol{\theta}) = \frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right) \Bigg[&-\boldsymbol{\theta}(\mathbf{x})\cdot\frac{\log_{\mathbf{x}}(p_{\Sigma_D}(\mathbf{x}))}{d_{\Gamma_D}(\mathbf{x})} - \boldsymbol{\theta}(p_{\Sigma_D}(\mathbf{x}))\cdot\mathbf{n}_{\Sigma_D}(p_{\Sigma_D}(\mathbf{x})) \\ &+ \int_0^{d_{\Gamma_D}(\mathbf{x})} \Pi^{\partial\Omega}_{\sigma_{\mathbf{x}}(t)}(\sigma'_{\mathbf{x}}(t),\sigma'_{\mathbf{x}}(t))\,(\boldsymbol{\theta}\cdot\mathbf{n})(\sigma_{\mathbf{x}}(t))\,\mathrm{d}t \Bigg] up\,\mathrm{d}s(\mathbf{x}). \end{aligned}$$

Let us first rearrange the expression of $I_1(\boldsymbol{\theta})$. To this end, using the same type of calculations as in the proofs of Theorems V.2.1.1 and V.3.1.1, integration by parts together with the fact that u and p have at least H^2 regularity near $\Gamma_D \cup \Gamma$ yield straightforwardly:

$$\begin{aligned} I_1(\boldsymbol{\theta}) &= \int_{\partial\Omega} (j(u) - fp)\,\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_\Omega j'(u)\nabla u\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} + \int_\Omega f\nabla p\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} \\ &\quad + \int_{\Gamma\cup\Gamma_D} \nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\, h_{\Omega,\varepsilon} up\,\mathrm{d}s \\ &\quad + \int_{\Gamma\cup\Gamma_D} \left(\nabla u\cdot\nabla p\,\boldsymbol{\theta}\cdot\mathbf{n} - \frac{\partial u}{\partial\mathbf{n}}\nabla p\cdot\boldsymbol{\theta} - \frac{\partial p}{\partial\mathbf{n}}\nabla u\cdot\boldsymbol{\theta}\right)\mathrm{d}s \\ &\quad + \int_\Omega (-\nabla(\nabla u\cdot\nabla p) + \Delta u\nabla p + \Delta p\nabla u + \nabla^2 p\nabla u + \nabla^2 u\nabla p)\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x} \\ &= \int_{\partial\Omega} (j(u) - fp)\,\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s + \int_{\Gamma\cup\Gamma_D} \nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\, h_{\Omega,\varepsilon} up\,\mathrm{d}s \\ &\quad + \int_{\Gamma\cup\Gamma_D} \left(\nabla u\cdot\nabla p\,\boldsymbol{\theta}\cdot\mathbf{n} - \frac{\partial u}{\partial\mathbf{n}}\nabla p\cdot\boldsymbol{\theta} - \frac{\partial p}{\partial\mathbf{n}}\nabla u\cdot\boldsymbol{\theta}\right)\mathrm{d}s. \end{aligned} \tag{V.4.11}$$

Denoting by $\mathcal{D}$ the last integrand in the above right-hand side, we obtain:

$$\begin{aligned} \mathcal{D} &:= \nabla u\cdot\nabla p\,\boldsymbol{\theta}\cdot\mathbf{n} - \tfrac{\partial u}{\partial\mathbf{n}}\nabla p\cdot\boldsymbol{\theta} - \tfrac{\partial p}{\partial\mathbf{n}}\nabla u\cdot\boldsymbol{\theta}, \\ &= -\tfrac{\partial u}{\partial\mathbf{n}}\tfrac{\partial p}{\partial\mathbf{n}}\boldsymbol{\theta}\cdot\mathbf{n} - \left(\tfrac{\partial u}{\partial\mathbf{n}}\nabla_{\partial\Omega}p\cdot\boldsymbol{\theta} + \tfrac{\partial p}{\partial\mathbf{n}}\nabla_{\partial\Omega}u\cdot\boldsymbol{\theta}\right), \\ &= -\tfrac{\partial u}{\partial\mathbf{n}}\tfrac{\partial p}{\partial\mathbf{n}}\boldsymbol{\theta}\cdot\mathbf{n} + h_\varepsilon\left(u\nabla_{\partial\Omega}p\cdot\boldsymbol{\theta} + p\nabla_{\partial\Omega}u\cdot\boldsymbol{\theta}\right). \end{aligned} \tag{V.4.12}$$

On the other hand, integrating by parts on the surface $\partial\Omega$ (see again Th. V.1.2.2), we obtain:

$$\int_{\partial\Omega} \nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\, h_{\Omega,\varepsilon} up\,\mathrm{d}s = \int_{\Gamma_D\cup\Gamma} h_\varepsilon \kappa up\,\boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_{\Gamma\cup\Gamma_D} h_\varepsilon(p\nabla_{\partial\Omega}u\cdot\boldsymbol{\theta} + u\nabla_{\partial\Omega}p\cdot\boldsymbol{\theta})\,\mathrm{d}s - \frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{d_{\Gamma_D}}{\varepsilon}\right)(\nabla_{\partial\Omega}d_{\Gamma_D}\cdot\boldsymbol{\theta})up\,\mathrm{d}s. \quad \text{(V.4.13)}$$

Finally, combining Eqs. (V.4.11) to (V.4.13) with the definitions of $I_1(\boldsymbol{\theta})$ and $I_2(\boldsymbol{\theta})$, as well as Th. V.1.2.2 for the tangential gradient of the geodesic signed distance function, the desired result follows. □

V.4.2.b Study of the convergence of the approximate model to the exact problem

In this section, we are interested in evaluating in which capacity the exact shape optimization problem Eq. (V.1.4) is correctly approximated by its smoothed counterpart Eq. (V.4.5). More precisely, we investigate the convergence of the objective function $\mathcal{J}_\varepsilon(\Omega)$ and that of its shape derivative $\mathcal{J}'_\varepsilon(\Omega)$ to the exact versions $\mathcal{J}(\Omega)$ and $\mathcal{J}'(\Omega)$ respectively, for a fixed shape $\Omega\subset\mathbb{R}^2$. In order to keep the exposition as simple as possible, we proceed under the assumption Eq. (V.1.3), however we believe the result holds in greater generality, and notably in the 2d case where $\partial\Omega$ is not flat in the neighborhood of Σ_D; see Remark V.1.2.2. Let us mention that a quite similar problem is investigated from the theoretical viewpoint in [Cos96], with stronger conclusions. Our first result in this direction is the following:

Theorem V.4.2.2 – Convergence of the regularized model.
Under assumption Eq. (V.1.3), the function $u_{\Omega,\varepsilon}$ converges to u_Ω strongly in $H^1(\Omega)$, and the following estimate holds:

$$\|u_{\Omega,\varepsilon}-u_\Omega\|_{H^1(\Omega)} \le C_s\varepsilon^s\|f\|_{L^2(\Omega)}, \quad \text{(V.4.14)}$$

for any $0<s<\frac{1}{4}$, where the constant C_s depends on s.

Proof: The error $r_\varepsilon := u_{\Omega,\varepsilon}-u_\Omega$ is the unique solution in $H^1(\Omega)$ to the system:

$$\left\{\begin{array}{rcll} -\Delta r_\varepsilon &=& 0 & \text{in } \Omega,\\ \frac{\partial r_\varepsilon}{\partial\mathbf{n}} + h_\varepsilon r_\varepsilon &=& -\frac{\partial u_\Omega}{\partial\mathbf{n}} - h_\varepsilon u_\Omega & \text{on } \partial\Omega,\end{array}\right. \quad \text{(V.4.15)}$$

which rewrites, under variational form:

$$\forall v\in H^1(\Omega),\ \int_\Omega \nabla r_\varepsilon\cdot\nabla v\,\mathrm{d}\mathbf{x} + \int_{\partial\Omega} h_\varepsilon r_\varepsilon v\,\mathrm{d}s = -\int_{\partial\Omega}\frac{\partial u_\Omega}{\partial\mathbf{n}}v\,\mathrm{d}s - \int_{\Gamma_D\cup\Gamma} h_\varepsilon u_\Omega v\,\mathrm{d}s. \quad \text{(V.4.16)}$$

Note that the above variational problem is well-posed, as follows from the Lax-Milgram lemma and the following Poincaré-like inequality (which is proved by the standard contradiction argument):

$$\forall v\in H^1(\Omega),\ \|v\|^2_{H^1(\Omega)} \le C\left(\int_\Omega |\nabla v|^2\,\mathrm{d}\mathbf{x} + \int_{\Gamma_D} v^2\,\mathrm{d}s\right); \quad \text{(V.4.17)}$$

here and throughout the proof, C stands for a positive constant which is independent of ε. The estimate Eq. (V.4.14) is then obtained within two steps.

Step 1: We prove that r_ε is bounded in $H^1(\Omega)$, uniformly with respect to ε. To this end, we estimate the first term in the right-hand side of Eq. (V.4.16) as:

$$\left|\int_{\partial\Omega} \frac{\partial u_\Omega}{\partial \mathbf{n}} v \,\mathrm{d}s\right| \leq C \left\|\frac{\partial u_\Omega}{\partial \mathbf{n}}\right\|_{H^{-1/2}(\partial\Omega)} \|v\|_{H^1(\Omega)}, \tag{V.4.18}$$

where we have the control

$$\left\|\frac{\partial u_\Omega}{\partial \mathbf{n}}\right\|_{H^{-1/2}(\partial\Omega)} \leq C\|f\|_{L^2(\mathbb{R}^d)},$$

as follows from the Green's formula Eq. (V.2.6) applied to the function u_Ω in $E(\Delta, L^2(\Omega))$ (see Eq. (V.2.5)). We are thus left with the task of estimating the second term in the right-hand side of Eq. (V.4.16), that is, the integral:

$$\int_{\Gamma_D\cup\Gamma} h_\varepsilon u_\Omega v \,\mathrm{d}s = \int_\Gamma h_\varepsilon u_\Omega v \,\mathrm{d}s.$$

To achieve this goal, recall that, since $u_\Omega \in H^s(\Omega)$ for $\frac{1}{2} < s < \frac{3}{2}$, and owing to the continuity of the trace $u \mapsto u|_{\partial\Omega}$ from $H^s(\Omega)$ into $H^{s-\frac{1}{2}}(\partial\Omega)$, for $s > \frac{1}{2}$ (see e.g. [McL00, Theorem 3.37]), it comes that $u_\Omega \in H^{s-\frac{1}{2}}(\partial\Omega)$, and in fact, using Eq. (V.1.2), that $u_\Omega \in \widetilde{H}^{s-\frac{1}{2}}(\Gamma_N\cup\Gamma)$. Using now the characterization Eq. (V.1.9) of the space $\widetilde{H}^{s-\frac{1}{2}}(\Gamma_N\cup\Gamma)$, it follows that for all $0 < \sigma < 1$, the function u_Ω/ρ^σ belongs to $L^2(\Gamma)$, where we have introduced the weight $\rho(\mathbf{x}) := \min(|\mathbf{x}-s_0|, |\mathbf{x}-s_1|)$.

Using this fact in combination with the Sobolev embedding from $H^{1/2}(\Gamma)$ into $L^q(\Gamma)$ for any $1 \leq q < \infty$ (see e.g. [Di 12, Theorem 6.7] or [Man18, Theorem 32]), we get successively:

$$\begin{aligned}
\left|\int_\Gamma h_\varepsilon u_\Omega v \,\mathrm{d}s\right| &= \left|\int_\Gamma \rho^\sigma h_\varepsilon \frac{1}{\rho^\sigma} u_\Omega v \,\mathrm{d}s\right|, \\
&\leq \left(\int_\Gamma \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s\right)^{\frac{1}{p}} \left(\int_\Gamma \frac{1}{\rho^{2\sigma}} u_\Omega^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \left(\int_\Gamma v^q \,\mathrm{d}s\right)^{\frac{1}{q}}, \\
&\leq \left(\int_\Gamma \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s\right)^{\frac{1}{p}} \|u_\Omega\|_{\widetilde{H}^\sigma(\Gamma_N\cup\Gamma)} \|v\|_{L^q(\Gamma)}, \\
&\leq C\left(\int_\Gamma \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s\right)^{\frac{1}{p}} \|u_\Omega\|_{\widetilde{H}^\sigma(\Gamma_N\cup\Gamma)} \|v\|_{H^1(\Omega)},
\end{aligned} \tag{V.4.19}$$

for any $p > 2$ (the constant C depends on p), where we have used Hölder's inequality with $\frac{1}{2} + \frac{1}{p} + \frac{1}{q} = 1$ to pass from the first line to the second one.

To proceed further, let us decompose Γ as

$$\Gamma = \Gamma_\varepsilon \cup \overline{U_0} \cup \overline{U_1}, \text{ where } U_i := \{\mathbf{x} \in \Gamma,\ |\mathbf{x} - s_i| < \varepsilon\}, \text{ and } \Gamma_\varepsilon := \{\mathbf{x} \in \Gamma,\ \rho(\mathbf{x}) > \varepsilon\}.$$

Taking advantage of the structure Eq. (V.4.2) of h_ε, the first integral in the right-hand side of Eq. (V.4.19) is of the form

$$\begin{aligned}
\int_\Gamma \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s &= \int_{U_1} \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s + \int_{U_2} \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s + \int_{\Gamma_\varepsilon} \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s \\
&\leq \frac{C}{\varepsilon^p} \int_0^\varepsilon t^{p\sigma} h\left(\frac{t}{\varepsilon}\right)^p \mathrm{d}t, \\
&\leq \frac{C\varepsilon^{p\sigma+1}}{\varepsilon^p} \int_0^1 t^{p\sigma} h(t)^p \,\mathrm{d}t, \\
&\leq C\varepsilon^{p(\sigma-1)+1}.
\end{aligned}$$

Therefore,

$$\left(\int_\Gamma \rho^{p\sigma} h_\varepsilon^p \,\mathrm{d}s\right)^{\frac{1}{p}} \leq C\varepsilon^{\sigma-1+\frac{1}{p}};$$

now, choosing $p > 2$ and $\sigma < 1$ adequately and using Eqs. (V.1.10), (V.1.12) and (V.4.19), we have proved that, for all $s < \frac{1}{2}$, there exists a constant C_s:

$$\left|\int_\Gamma h_\varepsilon u_\Omega v \,\mathrm{d}s\right| \leq C_s \varepsilon^s \|f\|_{L^2(\mathbb{R}^d)} \|v\|_{H^1(\Omega)}. \tag{V.4.20}$$

Eventually, taking $v = r_\varepsilon$ as a test function in Eq. (V.4.16), we obtain the standard a priori estimate for r_ε:

$$\int_\Omega |\nabla r_\varepsilon|^2 \,\mathrm{d}\mathbf{x} + \int_{\partial\Omega} h_\varepsilon r_\varepsilon^2 \,\mathrm{d}s = -\int_{\partial\Omega} \frac{\partial u_\Omega}{\partial \mathbf{n}} r_\varepsilon \,\mathrm{d}s - \int_\Gamma h_\varepsilon u_\Omega r_\varepsilon \,\mathrm{d}s. \tag{V.4.21}$$

Combining Eq. (V.4.21) with the estimates Eqs. (V.4.18) and (V.4.20), the Poincaré inequality Eq. (V.4.17) and the fact that $h_\varepsilon \geq 1$ on Γ_D (see Eq. (V.4.3)), it follows that there exists a constant C, which does not depend on ε, such that:

$$\|r_\varepsilon\|_{H^1(\Omega)} \leq C\|f\|_{L^2(\mathbb{R}^d)}. \tag{V.4.22}$$

Step 2: We now turn to the proof of Eq. (V.4.14) so to speak.
Multiplying both sides of Eq. (V.4.21) by ε and using Eq. (V.4.3), we obtain:

$$\begin{aligned}
\|r_\varepsilon\|^2_{L^2(\Gamma_D)} &\leq C\varepsilon \int_{\Gamma_D} h_\varepsilon r_\varepsilon^2 \,\mathrm{d}s, \\
&\leq C\varepsilon \left(\int_\Omega |\nabla r_\varepsilon|^2 \,\mathrm{d}\mathbf{x} + \int_{\partial\Omega} h_\varepsilon r_\varepsilon^2 \,\mathrm{d}s\right) \\
&\leq C\varepsilon \left|\int_{\partial\Omega} \frac{\partial u_\Omega}{\partial \mathbf{n}} r_\varepsilon \,\mathrm{d}s\right| + C\varepsilon \left|\int_\Gamma h_\varepsilon u_\Omega r_\varepsilon \,\mathrm{d}s\right|, \\
&\leq C\varepsilon \|f\|^2_{L^2(\mathbb{R}^d)},
\end{aligned} \tag{V.4.23}$$

where we have used the estimate Eq. (V.4.20) with $v = r_\varepsilon$ and the bound Eq. (V.4.22) over r_ε. Interpolating between Eq. (V.4.22) and Eq. (V.4.23) (see for instance [Lio68, Proposition 2.3]), for all $0 \leq s \leq \frac{1}{2}$, $s = (1-t)0 + \frac{1}{2}t$, there exists a constant C_s such that:

$$\|r_\varepsilon\|_{H^s(\Gamma_D)} \leq C_s \|r_\varepsilon\|^{1-t}_{L^2(\Gamma_D)} \|r_\varepsilon\|^t_{H^{\frac{1}{2}}(\Gamma_D)} \leq C_s \varepsilon^{\frac{1}{2}-s} \|f\|_{L^2(\mathbb{R}^d)}.$$

Now, since $u_\Omega \in H^s(\Omega)$ for $\frac{1}{2} < s < \frac{3}{2}$, it comes that $\frac{\partial u_\Omega}{\partial \mathbf{n}} \in H^{s-\frac{3}{2}}(\partial\Omega)$ (see [Cos88, Lemma 4.3]), and so

$$\left|\int_{\Gamma_D} \frac{\partial u_\Omega}{\partial \mathbf{n}} r_\varepsilon \,\mathrm{d}s\right| \leq \left\|\frac{\partial u_\Omega}{\partial \mathbf{n}}\right\|_{H^{s-\frac{3}{2}}(\Gamma_D)} \|r_\varepsilon\|_{H^{\frac{3}{2}-s}(\Gamma_D)} \leq C_s \varepsilon^{s-1} \|f\|^2_{L^2(\mathbb{R}^d)}, \tag{V.4.24}$$

for all $1 < s < \frac{3}{2}$. Returning to Eq. (V.4.21) and using Eqs. (V.4.17), (V.4.20) and (V.4.24), we now see that, for any $s < \frac{1}{2}$ and $\sigma < \frac{1}{2}$, there exists a constant $C > 0$ (depending on s and σ) such that:

$$\begin{aligned}
\|r_\varepsilon\|^2_{H^1(\Omega)} &\leq C\left(\int_\Omega |\nabla r_\varepsilon|^2 \,\mathrm{d}s + \int_{\Gamma_D} r_\varepsilon^2 \,\mathrm{d}s\right) \\
&\leq C\left(\varepsilon^s \|f\|^2_{L^2(\mathbb{R}^d)} + \varepsilon^\sigma \|f\|_{L^2(\mathbb{R}^d)} \|r_\varepsilon\|_{H^1(\Omega)}\right),
\end{aligned}$$

Hence Eq. (V.4.14) holds, and this terminates the proof. □

As a straightforward consequence of Th. V.4.2.2, we obtain that under the assumption Eq. (V.1.3) then for any given admissible shape $\Omega \in \mathcal{U}_{\mathrm{DN}}$, the approximate shape functional $\mathcal{J}_\varepsilon(\Omega)$ converges to its exact counterpart $\mathcal{J}(\Omega)$. Let us now turn to the convergence of the derivative of $\mathcal{J}_\varepsilon(\Omega)$.

Theorem V.4.2.3 – Convergence of the regularized shape derivative.
Under Assumption Eq. (V.1.3), for a given admissible shape $\Omega \in \mathcal{U}_{\mathrm{DN}}$, the approximate shape derivative $\mathcal{J}'_\varepsilon(\Omega)$ converges to its exact counterpart $\mathcal{J}'(\Omega)$ in the sense that:

$$\sup_{\substack{\boldsymbol{\theta}\in\Theta_{\mathrm{DN}}\\ \|\boldsymbol{\theta}\|_{\Theta_{\mathrm{DN}}}\leq 1}} |\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta}) - \mathcal{J}'(\Omega)(\boldsymbol{\theta})| = 0.$$

Proof: We rely on the volumetric expressions Eq. (V.3.1) and Eq. (V.4.6) of the shape derivatives $\mathcal{J}'(\Omega)(\boldsymbol{\theta})$ and $\mathcal{J}'_\varepsilon(\Omega)(\boldsymbol{\theta})$. In our context where Eq. (V.1.3) is satisfied, the boundary $\partial\Omega$ is flat in the neighborhood of $\Sigma_D = \{s_0, s_1\}$; hence, for $\varepsilon > 0$ small enough, the second fundamental form of $\partial\Omega$ vanishes where $h_\varepsilon > 0$, and the normal vectors $\mathbf{n}_{\Sigma_D}(s_0)$, $\mathbf{n}_{\Sigma_D}(s_1)$ to Σ_D coincide with the tangent vectors $\pm\boldsymbol{\tau}(s_0)$ and $\pm\boldsymbol{\tau}(s_1)$ to $\partial\Omega$. Then, the approximate shape derivative $\mathcal{J}'_\varepsilon(\Omega)$ supplied by Th. V.4.2.1 simply boils down to:

$$\begin{aligned}\mathcal{J}'(\Omega)(\boldsymbol{\theta}) = &\int_{\partial\Omega} (j(u_{\Omega,\varepsilon}) - f p_{\Omega,\varepsilon})\, \boldsymbol{\theta}\cdot\mathbf{n}\,\mathrm{d}s - \int_\Omega j'(u_{\Omega,\varepsilon})\nabla u_{\Omega,\varepsilon}\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}\\ &+ \int_\Omega ((\nabla\cdot\boldsymbol{\theta})\mathrm{I} - \nabla\boldsymbol{\theta} - \nabla\boldsymbol{\theta}^\top)\nabla u_{\Omega,\varepsilon}\cdot\nabla p_{\Omega,\varepsilon}\,\mathrm{d}\mathbf{x}\\ &+ \int_{\Gamma\cup\Gamma_D} \nabla_{\partial\Omega}\cdot\boldsymbol{\theta} h_\varepsilon u_{\Omega,\varepsilon} p_{\Omega,\varepsilon}\,\mathrm{d}s + \int_\Omega f\nabla p_{\Omega,\varepsilon}\cdot\boldsymbol{\theta}\,\mathrm{d}\mathbf{x}\\ &+ \sum_{i=0}^{1}\frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{\mathbf{x}-s_i}{\varepsilon}\right)(\boldsymbol{\theta}(\mathbf{x})-\boldsymbol{\theta}(s_i))\cdot\mathbf{n}_{\Sigma_D}(s_i)\, u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s(\mathbf{x}). \quad \text{(V.4.25)}\end{aligned}$$

Given the expression Eq. (V.3.1) of the exact shape derivative $\mathcal{J}'(\Omega)$, and in light of Th. V.4.2.2, it is obviously enough to show that the three integrals

$$\begin{aligned} I_1(\boldsymbol{\theta}) &:= \int_{\Gamma\cup\Gamma_D} \nabla_\Gamma\cdot\boldsymbol{\theta}\, h_\varepsilon u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s,\\ I_2(\boldsymbol{\theta}) &:= \frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{\mathbf{x}-s_0}{\varepsilon}\right)(\boldsymbol{\theta}(\mathbf{x})-\boldsymbol{\theta}(s_0))\cdot\boldsymbol{\tau}(s_0)u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s(\mathbf{x}),\\ I_3(\boldsymbol{\theta}) &:= \frac{1}{\varepsilon^2}\int_{\Gamma\cup\Gamma_D} h'\left(\frac{\mathbf{x}-s_1}{\varepsilon}\right)(\boldsymbol{\theta}(\mathbf{x})-\boldsymbol{\theta}(s_1))\cdot\boldsymbol{\tau}(s_1)\, u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s(\mathbf{x}),\end{aligned}$$

converge to 0 as $\varepsilon\to 0$, uniformly with respect to $\boldsymbol{\theta}$ when $\|\boldsymbol{\theta}\|_{\Theta_{\mathrm{DN}}}\leq 1$.

As far as the integral $I_1(\boldsymbol{\theta})$ is concerned, Th. V.4.2.2 and the facts that $-\Delta u_\varepsilon = -\Delta u_\Omega = f$ imply that

$$\frac{\partial u_{\Omega,\varepsilon}}{\partial\mathbf{n}} \to \frac{\partial u_\Omega}{\partial\mathbf{n}} \text{ in } H^{-1/2}(\partial\Omega), \text{ and } u_{\Omega,\varepsilon}\to u_\Omega \text{ in } H^{1/2}(\partial\Omega) \text{ as } \varepsilon\to 0;$$

similar convergence results hold about $p_{\Omega,\varepsilon}$ and p_Ω. Therefore,

$$\int_{\Gamma\cup\Gamma_D} \nabla_\Gamma\cdot\boldsymbol{\theta}\, h_{\Omega,\varepsilon}u_{\Omega,\varepsilon}p_{\Omega,\varepsilon}\,\mathrm{d}s \xrightarrow{\varepsilon\to 0} -\int_{\Gamma\cup\Gamma_D}\nabla_\Gamma\cdot\boldsymbol{\theta}\,\frac{\partial u_\Omega}{\partial\mathbf{n}}p_\Omega\,\mathrm{d}s, \quad \text{(V.4.26)}$$

where the last integral may be decomposed as

$$\int_{\Gamma\cup\Gamma_D}\nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\,\frac{\partial u_\Omega}{\partial\mathbf{n}}p_\Omega\,\mathrm{d}s = \int_{\Gamma_D}\nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\,\frac{\partial u_\Omega}{\partial\mathbf{n}}p_\Omega\,\mathrm{d}s + \int_\Gamma\nabla_{\partial\Omega}\cdot\boldsymbol{\theta}\,\frac{\partial u_\Omega}{\partial\mathbf{n}}p_\Omega\,\mathrm{d}s = 0,$$

as follows from the boundary conditions satisfied by u_Ω and p_Ω. This convergence is easily seen to be uniform with respect to $\boldsymbol{\theta}\in\Theta_{\mathrm{DN}}$, $\|\boldsymbol{\theta}\|_{\Theta_{\mathrm{DN}}}\leq 1$.

Let us now turn to the treatment of $I_2(\boldsymbol{\theta})$, that of $I_3(\boldsymbol{\theta})$ being on all points identical. We assume for notation simplicity that $s_0 = 0$, and again, we identify the neighborhood of s_0 in $\partial\Omega$ (which is a horizontal line) with a subset of the real line $\mathbb{R}$. The key remark in the analysis of $I_2(\boldsymbol{\theta})$ is that there exists a vector field $\tilde{\boldsymbol{\theta}}(\mathbf{x})$ vanishing identically on Γ_N such that $(\boldsymbol{\theta}(\mathbf{x}) - \boldsymbol{\theta}(0)) \cdot \boldsymbol{\tau}(0) = \mathbf{x} \cdot \tilde{\boldsymbol{\theta}}(\mathbf{x})$, as is easily seen from a Taylor expansion at 0. This will allow to improve the available convergence rates of $u_{\Omega,\varepsilon}$ and $p_{\Omega,\varepsilon}$ in the integrand of $I_2(\boldsymbol{\theta})$. More precisely, using integration by parts on the boundary $\partial\Omega$, $I_2(\boldsymbol{\theta})$ rewrites:

$$\begin{aligned}
I_2(\boldsymbol{\theta}) &= \int_{\Gamma\cup\Gamma_D} \partial_{\boldsymbol{\tau}} h_\varepsilon \mathbf{x}\cdot\tilde{\boldsymbol{\theta}}(\mathbf{x}) u_{\Omega,\varepsilon} p_{\Omega,\varepsilon}\,\mathrm{d}s,\\
&= -\int_{\Gamma\cup\Gamma_D} h_\varepsilon \partial_{\boldsymbol{\tau}}\left(\mathbf{x}\cdot\tilde{\boldsymbol{\theta}}(\mathbf{x}) u_{\Omega,\varepsilon} p_{\Omega,\varepsilon}\right)\,\mathrm{d}s,\\
&= -\int_{\Gamma\cup\Gamma_D} h_\varepsilon \tilde{\boldsymbol{\theta}}(\mathbf{x})\cdot\left(\mathbf{x}\frac{\partial u_{\Omega,\varepsilon}}{\partial\boldsymbol{\tau}} p_{\Omega,\varepsilon} + \mathbf{x}\frac{\partial p_{\Omega,\varepsilon}}{\partial\boldsymbol{\tau}} u_{\Omega,\varepsilon}\right)\,\mathrm{d}s - h_\varepsilon u_{\Omega,\varepsilon} p_{\Omega,\varepsilon}\frac{\partial}{\partial\boldsymbol{\tau}}\left(\mathbf{x}\cdot\tilde{\boldsymbol{\theta}}(\mathbf{x})\right)\,\mathrm{d}s\\
&= -\int_{\Gamma\cup\Gamma_D} h_\varepsilon \tilde{\boldsymbol{\theta}}(\mathbf{x})\cdot\left(\mathbf{x}\frac{\partial u_{\Omega,\varepsilon}}{\partial\boldsymbol{\tau}} p_{\Omega,\varepsilon} + \mathbf{x}\frac{\partial p_{\Omega,\varepsilon}}{\partial\boldsymbol{\tau}} u_{\Omega,\varepsilon}\right)\,\mathrm{d}s + R_\varepsilon(\boldsymbol{\theta}),
\end{aligned}$$

where $R_\varepsilon(\boldsymbol{\theta})$ is a remainder (possibly changing from one line to the next) gathering several integrals which are proved to converge to 0 as $\varepsilon \to 0$, uniformly with respect to $\boldsymbol{\theta}$ when $\|\boldsymbol{\theta}\|_{\Theta_{\mathrm{DN}}} \le 1$ owing to similar calculations to those involved in the above proof of convergence of $I_1(\boldsymbol{\theta})$ (see Eq. (V.4.26)). Then, using the boundary conditions satisfied by $u_{\Omega,\varepsilon}$ and $p_{\Omega,\varepsilon}$,

$$\begin{aligned}
I_2(\boldsymbol{\theta}) &= \int_{\Gamma\cup\Gamma_D}\left(\rho(\mathbf{x})\frac{\partial u_{\Omega,\varepsilon}}{\partial\boldsymbol{\tau}}\frac{\partial p_{\Omega,\varepsilon}}{\partial\mathbf{n}} + \rho(\mathbf{x})\frac{\partial p_{\Omega,\varepsilon}}{\partial\boldsymbol{\tau}}\frac{\partial u_{\Omega,\varepsilon}}{\partial\mathbf{n}}\right)\left(\frac{\mathbf{x}}{\rho(\mathbf{x})}\cdot\tilde{\boldsymbol{\theta}}(\mathbf{x})\right)\,\mathrm{d}s + R_\varepsilon(\boldsymbol{\theta}),\\
&= \int_{\Gamma\cup\Gamma_D}\left(\frac{\partial(\rho u_{\Omega,\varepsilon})}{\partial\boldsymbol{\tau}}\frac{\partial p_{\Omega,\varepsilon}}{\partial\mathbf{n}} + \frac{\partial(\rho p_{\Omega,\varepsilon})}{\partial\boldsymbol{\tau}}\frac{\partial u_{\Omega,\varepsilon}}{\partial\mathbf{n}}\right)\left(\frac{\mathbf{x}}{\rho(\mathbf{x})}\cdot\tilde{\boldsymbol{\theta}}(\mathbf{x})\right)\,\mathrm{d}s + R_\varepsilon(\boldsymbol{\theta}),
\end{aligned}$$

where we have posed $\rho(\mathbf{x}) = |\mathbf{x}|$ and the same calculations as in Eq. (V.4.26) have been used.

At this point, we know from Th. V.4.2.2 that $\frac{\partial u_{\Omega,\varepsilon}}{\partial\mathbf{n}}$ (resp. $\frac{\partial p_{\Omega,\varepsilon}}{\partial\mathbf{n}}$) converges to $\frac{\partial u_\Omega}{\partial\mathbf{n}}$ (resp. $\frac{\partial p_\Omega}{\partial\mathbf{n}}$) in $H^{-1/2}(\partial\Omega)$. Hence, the proof of the convergence of $I_2(\boldsymbol{\theta})$, and thereby that of Th. V.4.2.3, follows from the following results:

$$\frac{\partial(\rho u_{\Omega,\varepsilon})}{\partial\boldsymbol{\tau}} \xrightarrow{\varepsilon\to 0} \frac{\partial(\rho u_\Omega)}{\partial\boldsymbol{\tau}} \text{ in } H^1(\Omega), \text{ and } \frac{\partial(\rho p_{\Omega,\varepsilon})}{\partial\boldsymbol{\tau}} \xrightarrow{\varepsilon\to 0} \frac{\partial(\rho p_\Omega)}{\partial\boldsymbol{\tau}} \text{ in } H^1(\Omega), \tag{V.4.27}$$

where $\boldsymbol{\tau}$ stands for any smooth extension to the whole Ω of the tangent vector $\boldsymbol{\tau}$ to $\partial\Omega$; see Section V.1.2.f. We now sketch the proof of this last statement focusing on the case of $u_{\Omega,\varepsilon}$; the counterpart result as regards $p_{\Omega,\varepsilon}$ being proved in a similar fashion.

The convergence Eq. (V.4.27) actually follows from exactly the same arguments as that in the proof of Eq. (V.4.14). At first, using the representation of Th. V.1.2.1 (or more exactly a higher-order avatar of it, see Remark V.1.2.1), observe that the function ρu_Ω belongs to $H^s(\Omega)$ for all $0 \le s < \frac{5}{2}$. Letting the notation $r_\varepsilon := u_{\Omega,\varepsilon} - u_\Omega$, and using test functions of the form $\rho(\mathbf{x})v \in H^1(\Omega)$ for $v \in H^1(\Omega)$ inside the variational formulation Eq. (V.4.16) of r_ε, we see that ρr_ε satisfies:

$$\begin{aligned}
\forall v \in H^1(\Omega),\ \int_\Omega \nabla(\rho r_\varepsilon)\cdot\nabla v\,\mathrm{d}\mathbf{x} + \int_{\Gamma\cup\Gamma_D} h_\varepsilon \rho r_\varepsilon v\,\mathrm{d}s =\\
-\int_\Gamma h_\varepsilon \rho u_\Omega v\,\mathrm{d}s - \int_{\Gamma_D}\frac{\partial u_\Omega}{\partial\mathbf{n}}\rho v\,\mathrm{d}s - \int_\Omega \nabla\rho\cdot(v\nabla r_\varepsilon - r_\varepsilon\nabla v)\,\mathrm{d}\mathbf{x}.
\end{aligned} \tag{V.4.28}$$

Now using test functions of the form $\frac{\partial v}{\partial \boldsymbol{\tau}}$, $v \in H^1(\Omega)$ in Eq. (V.4.28), then integrating by parts yields the following variational formulation for $q_\varepsilon := \frac{\partial(\rho r_\varepsilon)}{\partial \boldsymbol{\tau}}$:

$$\forall v \in H^1(\Omega), -\int_\Omega \nabla q_\varepsilon \cdot \nabla v \,\mathrm{d}\mathbf{x} - \int_{\Gamma \cup \Gamma_D} h_\varepsilon q_\varepsilon v \,\mathrm{d}s = \int_{\partial\Omega} \frac{\partial h_\varepsilon}{\partial \boldsymbol{\tau}} \rho r_\varepsilon v \,\mathrm{d}s + \int_\Gamma \frac{\partial}{\partial \boldsymbol{\tau}}(h_\varepsilon \rho(\mathbf{x}) u_\Omega) v \,\mathrm{d}s + \int_{\Gamma_D} \frac{\partial}{\partial \boldsymbol{\tau}} \left(\rho \frac{\partial u_\Omega}{\partial \mathbf{n}} \right) v \,\mathrm{d}s + \langle F_\varepsilon, v \rangle_{H^1(\Omega)^*, H^1(\Omega)}, \quad \text{(V.4.29)}$$

where the remainder F_ε is a sequence of linear forms in the dual $H^1(\Omega)^*$ of $H^1(\Omega)$ which converges to 0 in the strong dual topology.

Finally, using the result of Th. V.4.2.2, together with very similar calculations than those involved in its proof, the desired result Eq. (V.4.27) follows, which concludes the proof. □

V.5 Numerical applications

V.5.1 Optimization of the repartition of clamps and locators on the boundary of an elastic structure

We start with the application of the results of Sections V.3 and V.4 to the problem of optimal repartition of clamps and locators on an elastic structure; see [Baw04, Chapter 9] for a presentation of the physical context and [Kay06; Ma11; Sel13] for optimization studies conducted in this context.

V.5.1.a Description of the physical setting and of the optimization problem

In this example, Ω stands for a three-dimensional rectangular beam with size $4 \times 1 \times 1$, filled with a linearly elastic material, whose Hooke's law A is defined by, for any symmetric matrix e with size 3×3:

$$Ae = 2\mu e + \lambda \mathrm{tr}(e),$$

where λ, μ are the **Lamé parameters** of the material; in our context

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \quad \mu = \frac{E}{1+\nu} \quad \text{(V.5.1)}$$

with $E = 100, \nu = 0.3$. During its construction, Ω receives the vertical load $\mathbf{g}_{\text{tool}} = (0, 0, -1)$ from the manufacturing tool, which is applied on the upper side Γ_T of its boundary. So that the structure do not move under this effort, a **clamping-locator** system is used: Ω is attached on a subregion Γ_D of the left-hand side Λ_D of $\partial\Omega$ (locator), while it receives a prescribed load $\mathbf{g} = (0, -1, 0)$ on another region Γ_N of the right-hand side $\Lambda_N \subset \partial\Omega$ (clamping); the latter is exerted by an external mechanical device pressing against the structure; see Fig. V.5.1 for a sketch of the situation.
In this context, the displacement of Ω is the unique solution $\mathbf{u}_{\Gamma_D,\Gamma_N} \in H^1_{\Gamma_D}(\Omega)^3$ to the following linear elasticity system:

$$\begin{cases} -\nabla \cdot (Ae(\mathbf{u}_{\Gamma_D,\Gamma_N})) = 0 & \text{in } \Omega, \\ \mathbf{u}_{\Gamma_D,\Gamma_N} = 0 & \text{on } \Gamma_D, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N})\mathbf{n} = \mathbf{g}_{\text{tool}} & \text{on } \Gamma_T, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N})\mathbf{n} = \mathbf{g} & \text{on } \Gamma_N, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N})\mathbf{n} = 0 & \text{on } \Gamma, \end{cases} \quad \text{(V.5.2)}$$

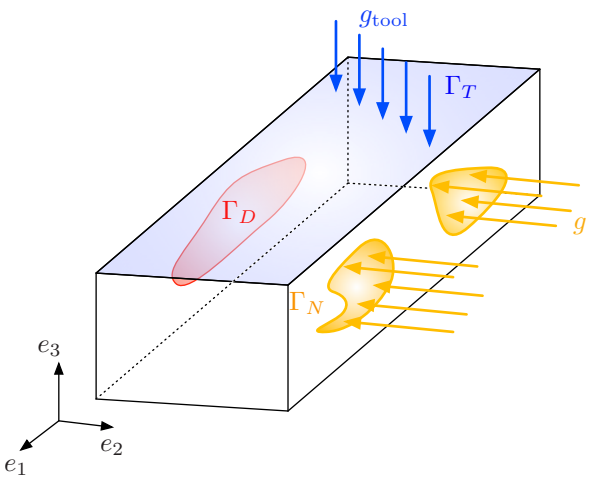


Figure V.5.1: Setting of the example of Section V.5.1 about the optimal repartition of clamps and locators on the boundary of an elastic structure.

where $e(\mathbf{u}) := \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^\top)$ is the strain tensor associated to a vector field $\mathbf{u} : \Omega \to \mathbb{R}^3$.

Our aim is to optimize the positions Γ_N and Γ_D of clamps and locators on the surface of the structure Ω, whose shape itself is not subject to optimization, so that the displacement of Ω under the action of $\mathbf{g}_{\text{tool}}$ be minimal. We also add constraints on the size of the regions Γ_N, Γ_D and on the perimeter of Γ_D via fixed penalizations of the objective function. More precisely, we consider the optimization problem:

$$\inf_{\substack{\Gamma_D \subset \Lambda_D \\ \Gamma_N \subset \Lambda_N}} \mathcal{J}(\Gamma_D, \Gamma_N), \text{ where } \mathcal{J}(\Gamma_D, \Gamma_N) = \int_\Omega |\mathbf{u}_{\Gamma_D,\Gamma_N}|^2 \, \mathrm{d}\mathbf{x} + \ell_D \int_{\Gamma_D} \mathrm{d}s + \ell_N \int_{\Gamma_N} \mathrm{d}s + \ell_{K_D} \int_{\Sigma_D} \mathrm{d}s, \quad \text{(V.5.3)}$$

where ℓ_D, ℓ_N and ℓ_{K_D} are fixed Lagrange multipliers: $\ell_D = 2.10^{-2}$, $\ell_N = 10^{-3}$, $\ell_{K_D} = 10^{-2}$. In the framework of Hadamard's method (see Section V.1.2.b), we consider deformations $\boldsymbol{\theta}$ such that:

$$\boldsymbol{\theta} \cdot \mathbf{n} = 0 \text{ on } \partial\Omega, \text{ and } \boldsymbol{\theta} = 0 \text{ on } \partial\Omega \setminus (\Lambda_D \cup \Lambda_N). \quad \text{(V.5.4)}$$

The numerical resolution of this problem relies on the knowledge of the shape derivatives of the partial mappings $\Gamma_D \mapsto \mathcal{J}(\Gamma_D, \Gamma_N)$ and $\Gamma_N \mapsto \mathcal{J}(\Gamma_D, \Gamma_N)$. In order to accomodate the presence of the transition $\Sigma_D := \overline{\Gamma_D} \cap \overline{\Gamma} \subset \Lambda_D$ between homogeneous Dirichlet and Neumann boundary conditions, we follow the lead of Section V.4 and consider the following approximate counterpart of Eq. (V.5.3):

$$\inf_{\substack{\Gamma_D \subset \Lambda_D \\ \Gamma_N \subset \Lambda_N}} \mathcal{J}_\varepsilon(\Gamma_D, \Gamma_N), \text{ where } \mathcal{J}_\varepsilon(\Gamma_D, \Gamma_N) := \int_\Omega |\mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon}|^2 \, \mathrm{d}\mathbf{x} + \ell_D \int_{\Gamma_D} \mathrm{d}s + \ell_N \int_{\Gamma_N} \mathrm{d}s + \ell_{K_D} \int_{\Sigma_D} \mathrm{d}s, \quad \text{(V.5.5)}$$

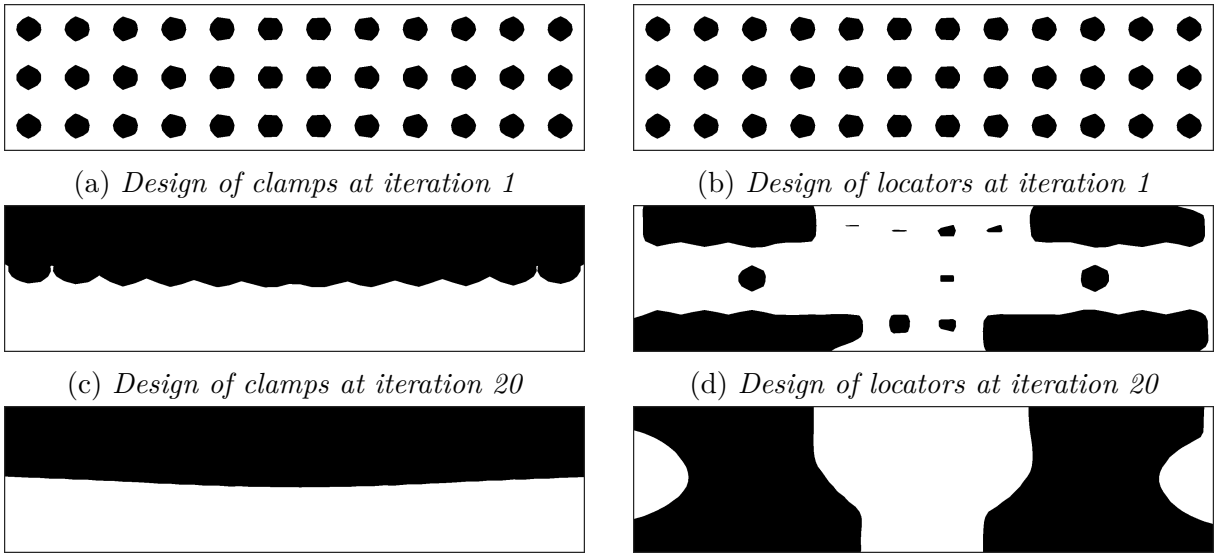
where $u_{\Gamma_D,\Gamma_N,\varepsilon}$ is the solution in $H^1(\Omega)^3$ to the system:

$$\begin{cases} -\nabla \cdot (Ae(\mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon}) &= 0 & \text{in } \Omega, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon}) + h_\varepsilon \mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon} &= 0 & \text{on } \Lambda_D, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon})\mathbf{n} &= \mathbf{g}_{\text{tool}} & \text{on } \Gamma_T, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon})\mathbf{n} &= \mathbf{g} & \text{on } \Gamma_N, \\ Ae(\mathbf{u}_{\Gamma_D,\Gamma_N,\varepsilon})\mathbf{n} &= 0 & \text{on } \Gamma \setminus \Lambda_D, \end{cases} \quad \text{(V.5.6)}$$

featuring the interpolation profile h_ε in Eq. (V.4.2). Thence, the calculation of the shape derivative of $\Gamma_N \mapsto \mathcal{J}_\varepsilon(\Gamma_D, \Gamma_N)$ is provided by Section V.2, or more exactly, the straightforward adaptation of its proof to the present linearized elasticity context. The shape derivative of the smoothed mapping $\Gamma_D \mapsto \mathcal{J}_\varepsilon(\Gamma_D, \Gamma_N)$ is calculated exactly as in the proof of Th. V.4.2.1 (or by using Céa's formal method), and we omit the formula for brevity.

V.5.1.b Numerical application

Let us now consider a concrete example in the previous context. A tetrahedral mesh of Ω composed of 45 000 vertices is used, and the optimization problem Eq. (V.5.5) is solved for the positions of Γ_D and Γ_N while, again, the shape of Ω itself is unchanged. Relying on the level set method for representing Γ_D and Γ_N, we use a standard gradient algorithm based on the knowledge of the shape derivatives of $\Gamma_D \mapsto \mathcal{J}_\varepsilon(\Gamma_D, \Gamma_N)$ and $\Gamma_N \mapsto \mathcal{J}_\varepsilon(\Gamma_D, \Gamma_N)$; the computation takes about 8 hours and the results are presented on Fig. V.5.2.



(a) *Design of clamps at iteration 1*

(b) *Design of locators at iteration 1*

(c) *Design of clamps at iteration 20*

(d) *Design of locators at iteration 20*

(e) *Design of clamps at iteration 100*

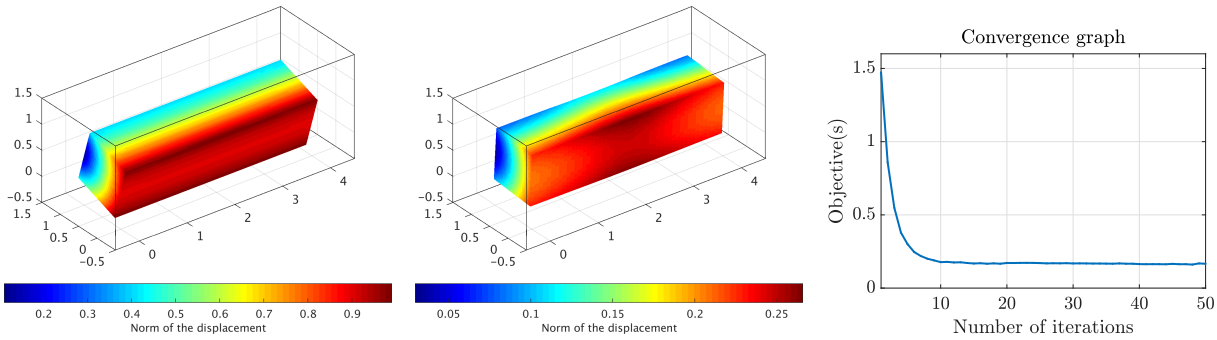
(f) *Design of locators at iteration 100*

Figure V.5.2: Initial, intermediate and optimized designs of clamps and locators in the test-case of Section V.5.1.

We notice in particular that the optimized design of the clamps is concentrated under Γ_T whereas the locators are symmetrically positioned at both ends of the beam. The deformed configurations of the initial and optimized shapes are displayed in Fig. V.5.3.

V.5.2 Joint optimization of the shape and the regions supporting different types of boundary conditions

We now turn to examples where the shape Ω of a 2d structure is optimized at the same time as the region Γ_D of its boundary supporting homogeneous Dirichlet boundary conditions. For simplicity, the region Γ_N supporting inhomogeneous Neumann boundary conditions is fixed, which means that we are exactly in the setting of Sections V.3 and V.4: in all the examples in this subsection, we consider the following shape and topology optimization



(a) *Deformed configuration of* Ω *with the initial configuration of clamps and locators*

(b) *Deformed configuration of* Ω *with the optimized configuration of clamps and locators*

(c) *Convergence history*

Figure V.5.3: Details of the optimization example of clamps and locators of Section V.5.1.

problem:

$$\inf_{\substack{\Omega\subset D\\ \Gamma_D\subset\Lambda_D\cap\partial\Omega}} \mathcal{J}(\Omega), \text{ where } \mathcal{J}(\Omega) = j(u_\Omega) + \ell_V \int_\Omega \mathrm{d}s + \ell_D \int_{\Gamma_D} \mathrm{d}s \tag{V.5.7}$$

is a weighted sum of a case-dependent objective defined from a smooth function j, involving the elastic displacement u_Ω of the shape, solution to Eq. (V.5.6), and of constraints on both the volume of shapes, and on the area of the Dirichlet boundary Γ_D (the latter constraints being enforced by means of fixed Lagrange multipliers ℓ_V, ℓ_D). Notice that in the statement Eq. (V.5.7) of the considered shape optimization problem, we have committed the same abuse of notations as in Section V.1.2.b: u_Ω and $\mathcal{J}(\Omega)$ depend on both the overall shape Ω of the structure and the position Γ_D of the region supporting homogeneous Dirichlet boundary conditions (the latter being constrained to belong to a fixed region Λ_D of the computational domain D), while only the first dependence is explicit.

As regards the numerical setting, the computational domain D is equipped with a fixed mesh. Each shape $\Omega \subset D$ is represented by the level set method, i.e. Ω is described via a level set function; see Section II.3.

Since the shape Ω is not discretized (it is only known via the datum of a level set function), no computational mesh is available to calculate the elastic displacement u_Ω by means of a standard finite element method. To alleviate this issue, the "ersatz material trick" (see e.g. [All01; All04; Ben13]) is used to approximate the considered linearized elasticity systems posed on Ω with systems posed on D as a whole: u_Ω is approximated by the solution u to:

$$\left\{\begin{array}{rcll} -\nabla\cdot(A_\eta e(\mathbf{u})) &=& 0 & \text{in } D,\\ \mathbf{u} &=& 0 & \text{on } \Gamma_D,\\ A_\eta e(\mathbf{u})\mathbf{n} &=& \mathbf{g} & \text{on } \Gamma_N,\\ A_\eta e(\mathbf{u}) &=& 0 & \text{on } \Gamma, \end{array}\right. \quad \text{where } A_\eta(\mathbf{x}) := \left\{\begin{array}{ll} A & \text{if } \mathbf{x}\in\Omega,\\ \eta A & \text{otherwise,}\end{array}\right.$$

and η is a small parameter so that the void region $D\setminus\overline{\Omega}$ is filled with a very soft material instead of void (typically, we take $\eta = 10^{-3}$). In this section the Lamé parameters are still given by Eq. (V.5.1) but using $E = 1, \nu = 0.3$.

As far as the representation of the optimized part Γ_D of $\partial\Omega$ is considered, it is constrained to belong to a planar subset Λ_D of the boundary ∂D in the examples of Sections V.5.2.a and V.5.2.b. In this case, it is represented by means of a level set function on a subset of the real line. In Section V.5.2.c, the set Λ_D is a whole region of D. Then, Γ_D is represented by means of a different level set function $\psi : \Lambda_D \to \mathbb{R}$ from that ϕ used to represent Ω. Both cases are simple adaptations from the general idea outlined in Section II.3; see also [Xia14; Xia16] about this type of representation.

The same process as before is applied to approximate the transition region Σ_D between homogeneous Dirichlet and Neumann boundary condition in the formulation of Problem Eq. (V.5.7), and we do not repeat the details for brevity.

V.5.2.a Optimization of the shape of a two-dimensional bridge and its supports

We first consider the joint optimization of the shape of a two-dimensional bridge Ω and of the location of its fixations. The situation is that depicted in Fig. V.5.4: Ω is enclosed inside a two-dimensional computational domain D meshed with 80 537 triangles; a unit vertical load is distributed along the upper deck Γ_N, a neighborhood of which is imposed to be part of Ω. We optimize Ω and the set of fixations Γ_D (which is restrained to a subset Λ_D of the lower part of ∂D) with respect to the elastic compliance of the configuration; more precisely, the optimization problem reads as Eq. (V.5.7) with the expressions:

$$j(u) = \int_{\Gamma_N} \mathbf{g} \cdot \mathbf{u} \, \mathrm{d}s, \quad \ell_V = 50, \quad \ell_D = 10.$$

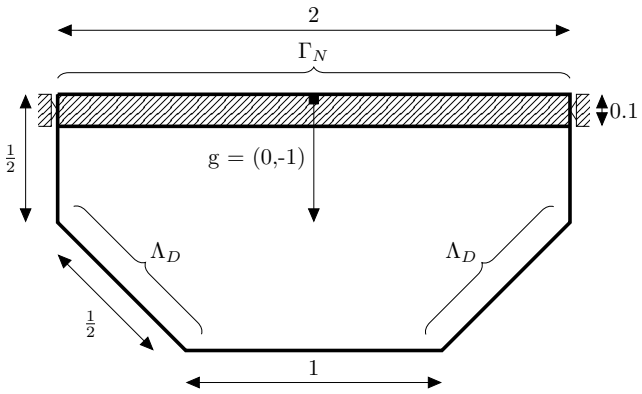
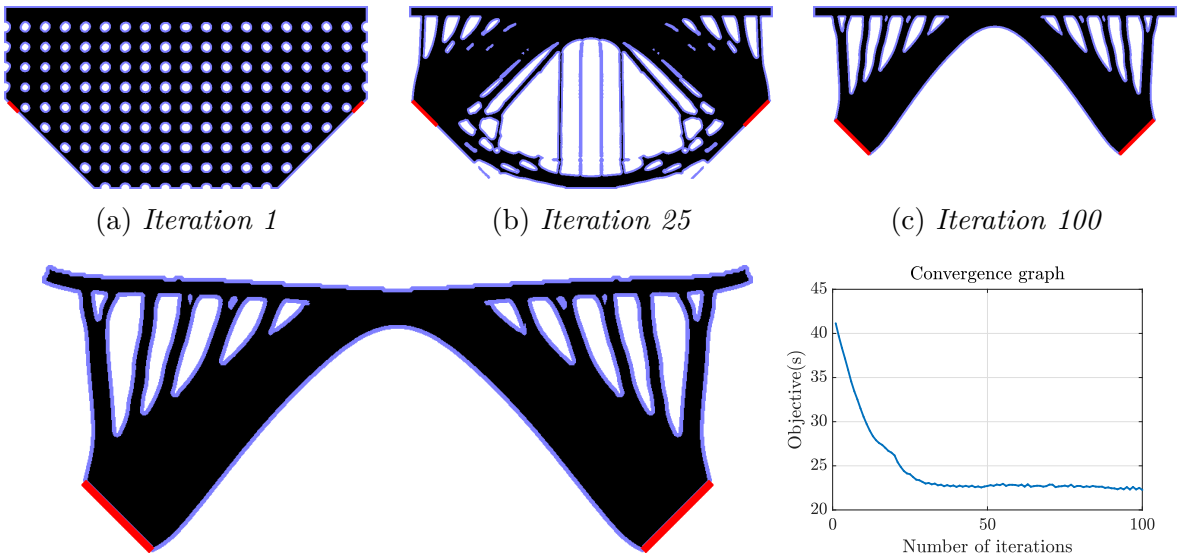


Figure V.5.4: Setting of the 2d bridge test-case of Section V.5.2.a; the dashed rectangle corresponds to the deck of the bridge, which is a non-optimizable area of Ω.

We perform two optimization experiments, corresponding to different initial states as for Ω and Γ_D; the results are reported in Fig. V.5.5 and Fig. V.5.6. In particular, we observe very different optimized topologies depending on the initial definition of the fixation region Γ_D.

V.5.2.b Optimization of the shape of a force inverter and of its fixations

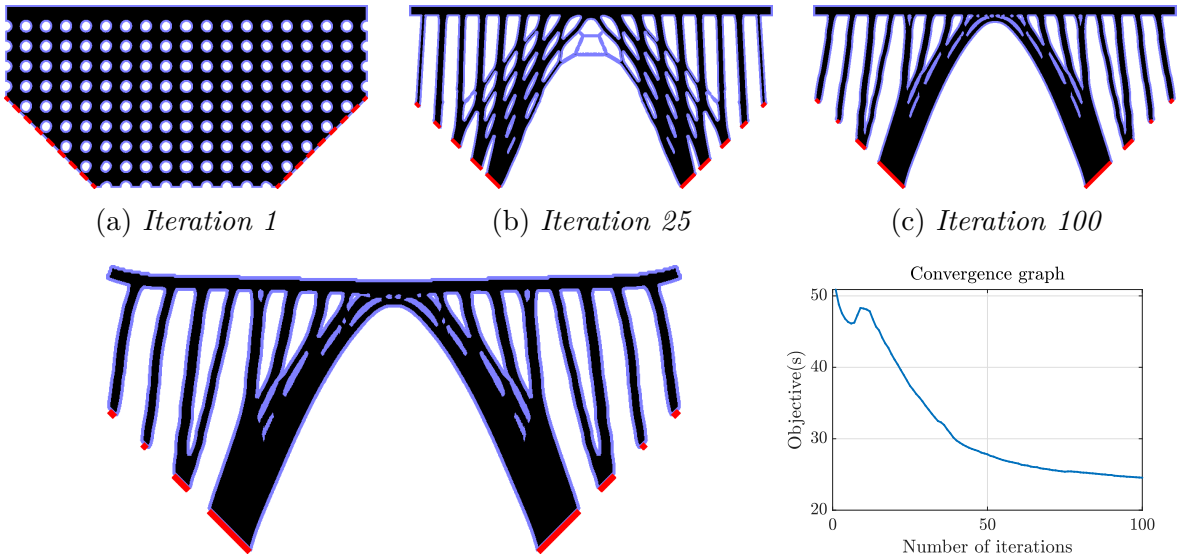
Our second example deals with the optimization of a force inverter mechanism, that is, a device which convert a pulling force into a pushing one. The details of the test-case



(a) *Iteration 1* (b) *Iteration 25* (c) *Iteration 100*

(d) *Deformed configuration of the optimized shape* (e) *Convergence history*

Figure V.5.5: Concurrent optimization of the shape and the fixation regions of the bridge of Section V.5.2.a, with an initial configuration for Γ_D composed of two line segments.



(a) *Iteration 1* (b) *Iteration 25* (c) *Iteration 100*

(d) *Deformed configuration of the optimized shape* (e) *Convergence history*

Figure V.5.6: Concurrent optimization of the shape and the fixation regions of the bridge of Section V.5.2.a, with an initial configuration for Γ_D composed of 18 line segments.

are presented on Fig. V.5.7: the considered shapes Ω are contained in a box D meshed with 78 408 triangles; they are subjected to a given load $\mathbf{g} = (-1, 0)$ applied on a non optimizable subset Γ_N of their left-hand side, and they are attached on another subset Γ_D of $\partial\Omega$, contained in the upper and lower sides of ∂D. In this context, the aim is to optimize the overall shape Ω and the location of the fixations Γ_D so that the elastic displacement of Ω on a non optimizable subset Γ_T is maximized. The symmetry of the optimized shapes with respect to the horizontal axis is enforced.

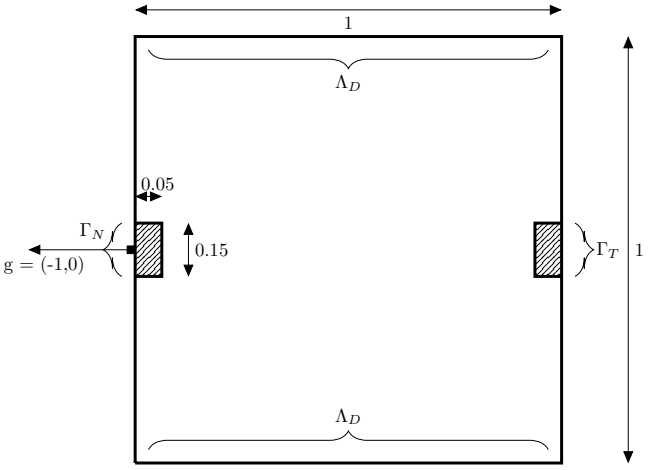
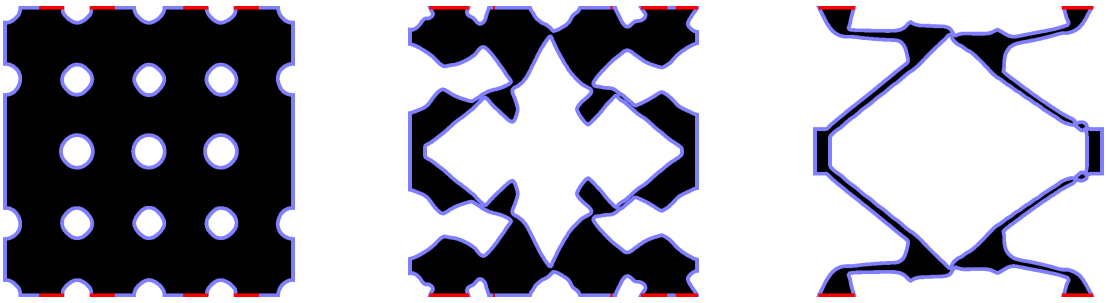


Figure V.5.7: Setting of the force inverter test-case of Section V.5.2.b. The two dashed rectangles represent non optimizable areas.

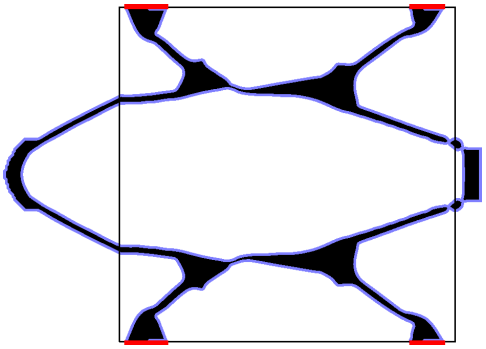
More precisely, in the general formulation of the problem Eq. (V.5.7), we set:

$$j(\mathbf{u}) = 10^{-1} \int_{\Gamma_T} |\mathbf{u} - (1,0)|^2 \, \mathrm{d}s - 10^{-3} \int_{\Gamma_N} \mathbf{u}_1 \, \mathrm{d}s, \quad \ell_V = 5 \times 10^{-3}, \quad \ell_D = 0$$

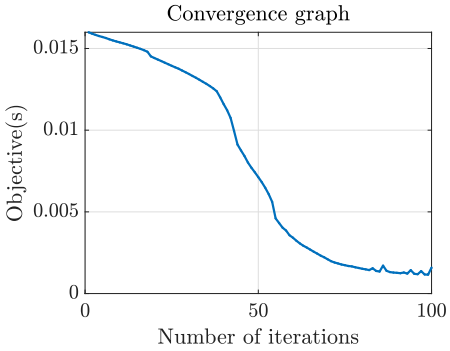
where the little penalization on the compliance was added to make it easier to obtain a connected structure and $\mathbf{u}_1$ corresponds to the first component of $\mathbf{u}$.



(a) *Iteration 1* (b) *Iteration 50* (c) *Iteration 100*

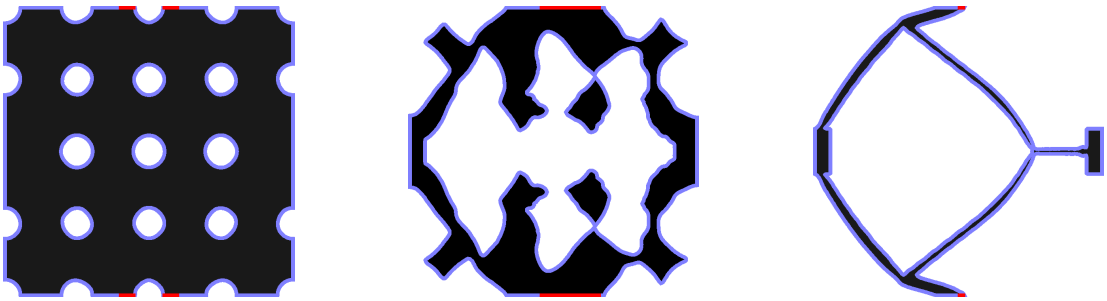


(d) *Deformed configuration of the optimized shape*

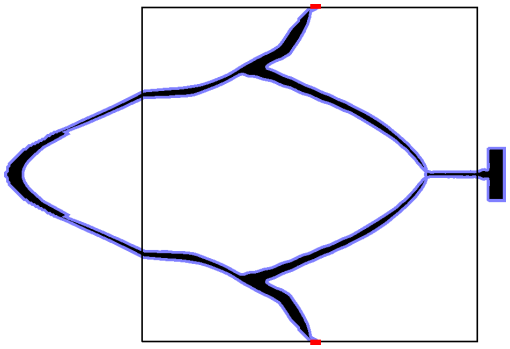


(e) *Convergence history*

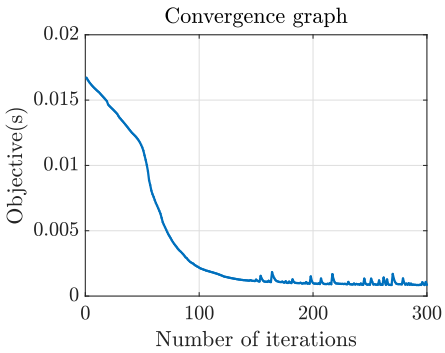
Figure V.5.8: Concurrent optimization of the shape and the fixation regions of the force inverter of Section V.5.2.b, with an initial configuration for Γ_D composed of 8 line segments.



(a) *Iteration 1* (b) *Iteration 50* (c) *Iteration 300*



(d) *Deformed configuration of the optimized shape*



(e) *Convergence history*

Figure V.5.9: Concurrent optimization of the shape and the fixation regions of the force inverter of Section V.5.2.b, with an initial configuration for Γ_D composed of 4 line segments.

V.5.2.c Optimization of the shape and the support regions of a two-dimensional cantilever beam

Our last example deals with the concurrent optimization of the shape of a classical 2d cantilever beam and its fixation zones. The considered shapes Ω are contained in a fixed computational domain D, meshed with 39 402 triangles. They are attached on the upper and lower left corners, as well as on a region Γ_D which is subjected to optimization, and which is constrained to be contained inside a given region $\mathcal{D}_D \subset D$. A vertical load $\mathbf{g} = (0, -1)$ is applied on a non optimizable subset Γ_N of the right-hand boundary; see Fig. V.5.10. Notice that, contrary to the previous two examples, the region Γ_D is not a subset of a region $\Lambda_D \subset \partial\mathcal{D}$ but it is allowed to evolve freely inside a region of $\mathcal{D}$. This demands a little adaptation of the framework described above (another level set function is used to identify the region Γ_D). Symmetry with respect to the horizontal axis is imposed on the optimized shape.

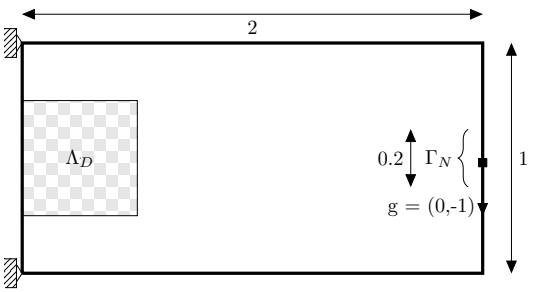
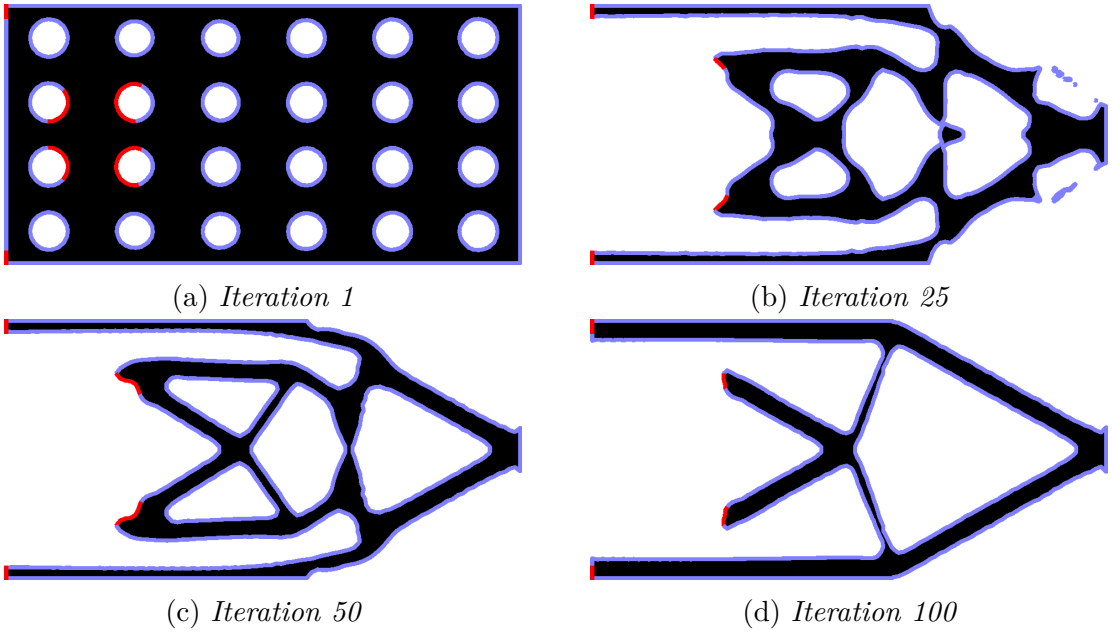


Figure V.5.10: Setting of the 2d cantilever test-case of Section V.5.2.c.

All things considered, we consider the optimization problem Eq. (V.5.7) with the expressions:

$$j(\mathbf{u}) = \int_{\Gamma_T} \mathbf{g} \cdot \mathbf{u} \, \mathrm{d}s, \quad \ell_V = 150, \quad \ell_D = 0.$$

Results are presented on Fig. V.5.11; obviously, the Dirichlet region aims to get as close as possible to the application region of the load $\mathbf{g}$. It also tends to concentrate on the top and bottom corners of the region $\mathcal{D}_D$, following insofar as possible the principal stress directions of the structure.

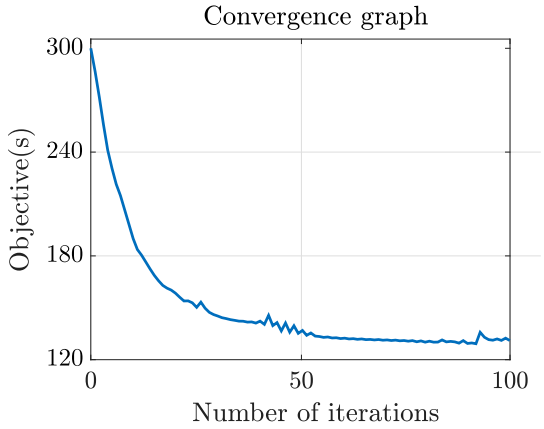


(a) *Iteration 1*

(b) *Iteration 25*

(c) *Iteration 50*

(d) *Iteration 100*



(e) *Convergence history*

Figure V.5.11: Concurrent optimization of the shape Ω and the fixation zones Γ_D of the two-dimensional cantilever of Section V.5.2.c (the latter are represented using red lines).

Conclusion & perspectives

General summary, and main research results

In Chapters I and II we introduced the concepts used throughout this thesis: the physical field of nanophotonics was presented as well as the mathematical tools for the implementation of a geometric shape optimization algorithm. From these reviews we have pursued research in several directions. Let us summarize here our main results.

- In Chapter III we have calculated the shape derivative of the perhaps most useful objective function when it comes to the optimal design of nanophotonic components. The numerical evaluation of this derivative being quite complex, we proposed an index smoothing method which provides an efficient way to obtain a good approximation of the derivative involving the solution of a regularized electric field. Several numerical examples were included to confirm the efficiency of our framework. The last section of the chapter also discussed an original work concerning the topology optimization of components featuring multiple levels of etching and the application of this method to the optimization of polarization rotators.

- In Chapter IV we described a method to maximize an objective functional defined as the worst value of a finite collection of figure of merits using a gradient sampling algorithm. We then showed how this methodology can be adapted in order to solve worst-case optimization problems when a small number of parameters are uncertain. The application of this algorithm to the design of robust components with respect to uncertainties over the wavelength or the geometry of the manufactured shape was studied on several test-cases and it was shown that this method does indeed provides satisfying results.

- In Chapter V the shape optimization of regions supporting different boundary conditions was studied in the case of the Laplace equation. We showed that the whole information about the sensitivity of the objective function with respect to the placement of the transition between Dirichlet and Neumann boundary conditions is encoded in the singular part of the PDE solution at the transition between these two boundary conditions. We then turned our attention to a regularized counterpart of this problem which makes the numerical implementation easier. This approximation is proved to be consistent with the original optimization problem. This chapter ended with numerical examples in the context of linear elasticity in both 2d and 3d situations.

Before outlining a series of research perspectives that were not fully studied during this thesis, we present several experimental results concerning shape optimized nanophotonic components that have been recently produced in the CEA clean room. We also describe the analyses that should be performed on these experimental results in order to recalibrate our simulations. This feedback between the practical results and the theory is essential if one wants to ensure the performance of future optimized devices.

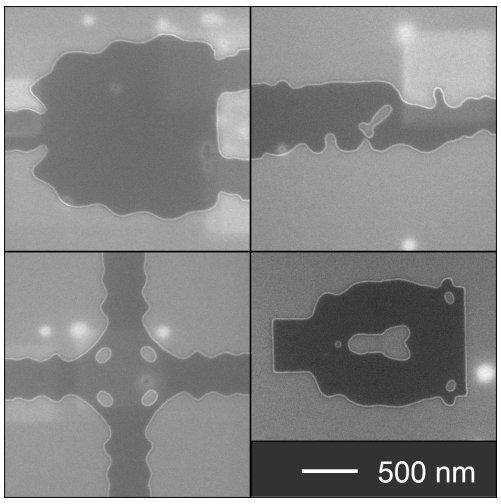
Experimental results of produced components

Since it took nearly two years between the delivery of the optimized components and the test of their performances, the designs presented here did not benefit from our most recent developments and, in particular, they did not use our method to take into account geometrical robustness. For the details of the manufacturing process conducted to obtain these results we refer to Section IV.3.1. The results are grouped thematically as follows:

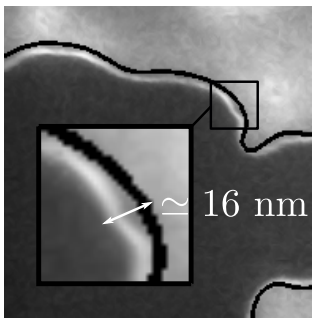
- First, we present some pictures of the realized components in order to asses the precision of the manufacturing process.
- We then move to the description of the experimental methodology used in order to test the performance of each component.
- The other subsections presents some experimental results obtained when testing the performances of our optimized nanophotonic components.

Geometrical precision of the manufactured devices

A selection of components optimized using the method discussed in this thesis which have been realized are represented in Fig. K(a).



(a) *From top to bottom, left to right: a power divider, a TE_0 to TE_1 mode converter, a crossing and a TE_0 to TE_2 mode converter*



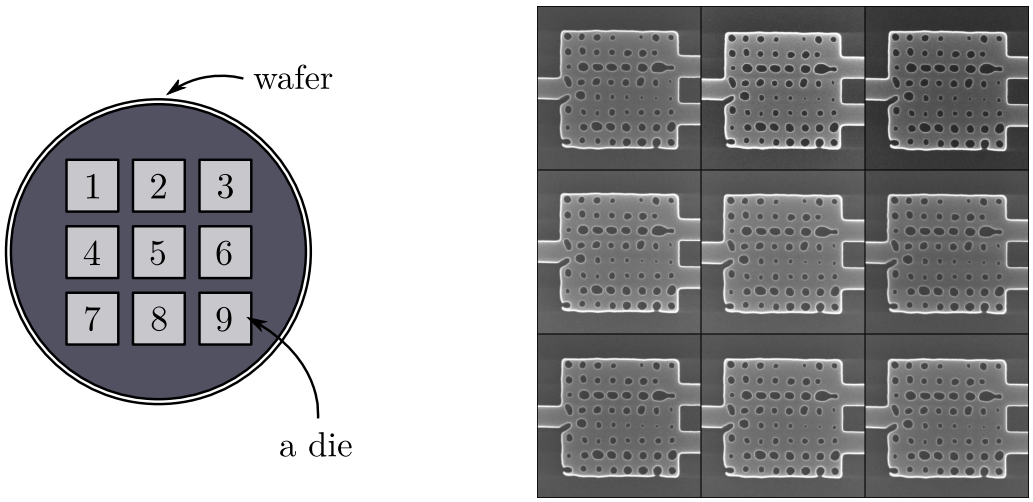
(b) *Zoom on a realized power divider and comparison with the numerical optimized shape (black lines)*

Figure K: Scanning Electron Microscope (SEM) images of the manufactured designs.

The quality of a manufactured component highly depends on its location on the wafer. We therefore decomposed the silicon plate into nine, uniformly distributed dies (see Fig. L(a)) in which the same components are manufactured.

As we now have the ability to produce shape optimized nanophotonic components, an image analysis study on Fig. L(b) would enable the precise determination of the uncertainty intervals regarding the dilation and erosion induced by the etching step. This calibration between the measurements and the numerical simulations is of utter importance for the practical efficiency of our shape optimization method. For an even greater accuracy it would be interesting to precisely analyze how the shape is etched on the edges of the

design as noted in Remark IV.3.1.1. We now present the method used in order to test the performance of each component on the wafer.



(a) *Repartition of the dies on the wafer*

(b) *Manufactured shapes of a diplexer depending on the die where it is produced*

Figure L: The 9 dies considered on our wafers.

Experimental methodology

In order to evaluate the actual performance of each individual component we consider a circuit where one shape optimized component is connected to two **grating couplers** (see Fig. M). Using a prober (see Fig. M, top left corner) light is injected using an optical fiber into one grating coupler which convert the light into the fundamental TE_0 mode of a waveguide linked to the input of the component.

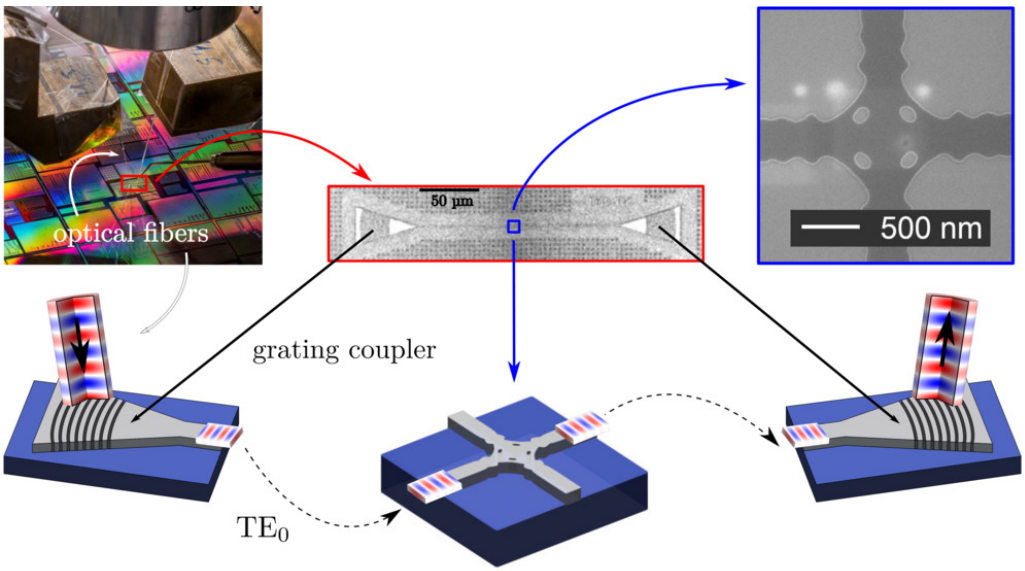


Figure M: The whole experimental setup with different levels of zoom.

Another waveguide connects the output of the component and a second grating coupler such that the electric field exiting of the component is received by another optical fiber

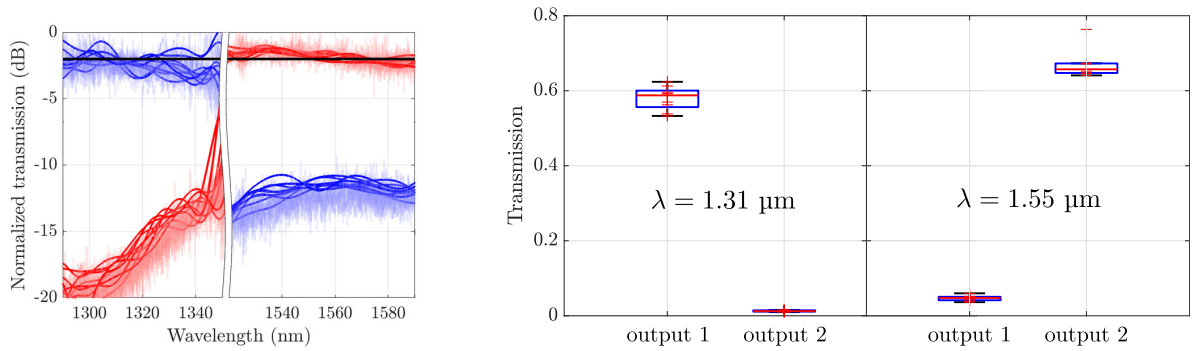
for which the prober can measure the transmitted power. By comparing the input and output powers we can then deduce the loss induced by the component. Adaptations of this setup are sometimes mandatory depending on the objective function (number of outputs, multiple wavelengths, different polarization etc.).

We now analyze experimental results concerning the performances of a selection of nanophotonic devices.

Experimental results

Performance of the wavelength robust diplexers

We begin with the robust wavelength diplexer presented in Section IV.2.2.c for which the associated fabricated components are depicted in Fig. L(b). Let us recall that the goal of this component is to redirect an incoming guided mode into either the top or bottom output waveguide depending on the wavelength. The graphs of Fig. N reveals that the routing of light depending on the incoming wavelength is achieved successfully with approximately 60 % to 65 % of transmission in each output. This is slightly below our theoretical predictions in Fig. IV.2.7 by about 20 %, but reproducible on nine dies.



(a) *Spectrum, the horizontal black line corresponds to* −2 dB

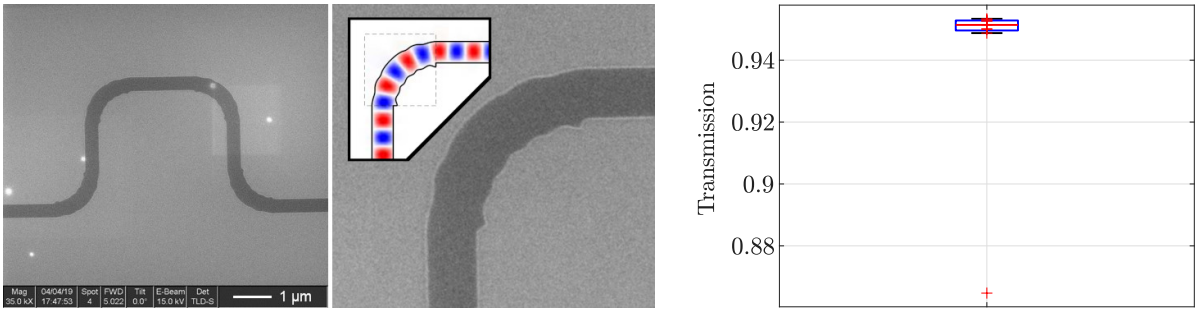
(b) *Performance at the nominal wavelengths of the transmissions in the top left and top right waveguide*

Figure N: Characterization of the nine diplexers on each die of the wafer around 1.31 and 1.55 µm. In (a) the quickly oscillating lines corresponds to the raw data for the 9 dies while the solid lines corresponds to a data interpolation with an "envelope". In (b) the average transmission for the top and bottom waveguides on the bandwidth 1.55 µm ± 1 nm and 1.31 µm ± 1 nm is averaged for each die and the resulting values are summarized using boxplots (the horizontal red line inside the boxes corresponds to the median value while the two extreme horizontal lines represents the 10 % and 90 % quantiles).

Performance of ninety degrees bends

We now move on to some results concerning bends, that is structures which redirect light coming from an input waveguide to an orthogonal output waveguide. This component has not been presented in Chapter III since it involves rather common geometries (see Fig. O(a)). The evaluation of the performance of such device is made using a chain structure as represented partially for four bends in Fig. O(a). Since this component has very low loss the chain allow to find the average transmission of one bend more easily.

The performance of one single bend is found here by using a chain composed of one hundred components. The results are summarized in Fig. O(b) for the 9 dies; a median transmission of 95 % is found as well as a variance inferior to 1 %.



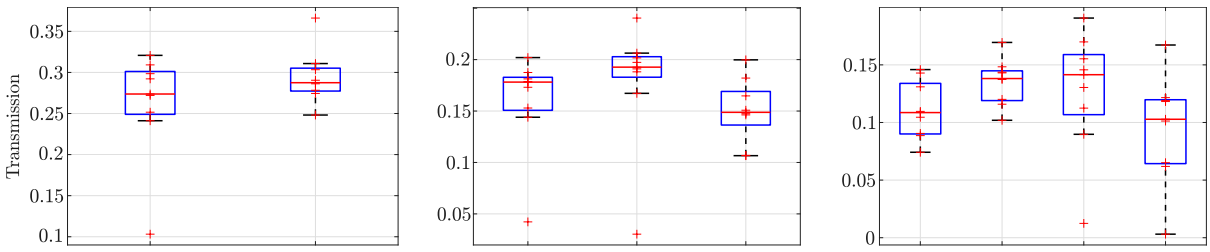
(a) *A chain composed of several bends* (b) *Performance of one bend*

Figure O: Evaluation of a ninety degrees nanophotonic bend.

Performance of power dividers

Concerning the power dividers, the results were not totally in line with the expectations despite the fact that the shapes seem to be well-manufactured (see discussion below). In Fig. P the performances of three power dividers are reported with respectively two, three and four outputs waveguides. One first remark is that although symmetric components are used, the transmissions measured on the output waveguides are not symmetric. This might be due to experimental measurements errors generated by misalignment of the input and output fibers.

Concerning the performances of the individual components, the 1 to 2 power dividers exhibit a median value of 28 % transmission in both outputs (44 % total loss), 17 % for the 1 to 3 (48 % total loss) and 12 % as for the 1 to 4 power divider (51 % total loss) making them useless in practice. We see here that the recalibration between experimental data and simulation results is very important since the theoretical performance of these components should be respectively of 49 % (2 % total loss), 31 % (6 % total loss) and 22.5 % (10 % total loss) as seen in Figs. III.3.6 and IV.1.4.



(a) *Two output waveguides* (b) *Three output waveguides* (c) *Four output waveguides*

Figure P: Performances on the nine dies of three 1 to n_{out} power dividers with n_{out} ranging from 2 to 4.

A partial explanation of such bad experimental behavior would be that an under- or over-etching of around 20 nm was made as suggested by Fig. K(b) and the robustness graph Fig. IV.3.6(b) on which a 35 % transmission is expected in the event of such a manufacturing error.

Research perspectives and other topics in integrated photonics that were not addressed in this thesis

We conclude this thesis with several research perspectives arising naturally from the previous works. We point out some related contributions from the literature, when available.

Other physical equations

In this thesis, we only considered the linear Maxwell equations in the time-harmonic regime. For the description of the physical behavior of the electric field however, very interesting components can be obtained by considering different equations or the coupling with other physical equation:

- The treatment of general electromagnetic phenomena does not rely on the time-harmonic vector wave equation but rather on its **time-dependent** counterpart defined in Eq. (I.1.6). The consideration of this more challenging equation requires some adaptations to the objective function involved in the shape optimization problem, to the definition of the Perfectly Matched Layers and to the Dirichlet-to-Neumann boundary condition for light injection. In exchange, once solved, the time-dependent field allows to recover the full frequency response of a component using Fourier transform. This could be interesting in the case of components such as the diplexer for which it is desired to optimize its whole spectrum. In such a time-dependent context, one could consider objection function such as

$$\mathcal{J}(\Omega) = \min_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |T(\lambda, \Omega) - T_{\text{obj}}(\lambda)|$$

  where $T(\lambda, \Omega)$ is the transmitted power into an output waveguide at the operating wavelength λ obtained via the Fourier transform of $\mathbf{E}$ and T_{obj} the desired value of the spectrum at λ.

- Another interesting variation of the time-harmonic vector wave equation concerns the consideration of **non-linear effects**. When the amplitude of the electric field is large, non-linearities coming from the dependence of the optical index n on the electric field can no longer be neglected. A first-order expansion of n at $|\mathbf{E}| = 0$ gives

$$n(|\mathbf{E}|) = n_0 + \alpha_1 |\mathbf{E}| + \frac{1}{2}\alpha_2 |\mathbf{E}|^2 + \dots$$

  where n_0 is the unperturbed optical index and the coefficients α_i are proportional to the "electro-optic coefficients" of the material. In some situations one coefficient is much more significant than the others. These phenomena are referred in the literature as Pockel effect, Kerr effect, etc. ... By exploiting one of these phenomena it is possible to obtain components which have a different behavior depending on the power of the injected light. Topology optimization of non-linear nanophotonic structures have recently started to be considered with the work of [Hug18].

- **Optomechanics** is the field studying the interaction between light and mechanical motion. When the electromagnetic fields are propagating inside a material, they exerts a force on it, called the Lorentz force F_{Lorentz}, equal to

$$F_{\text{Lorentz}} = \rho \mathbf{E} + \sigma \mathbf{E} \times \mathbf{H}$$

where ρ and σ are material-dependent values defined in Section I.1.1.a. Usually this force is expressed using the so-called Maxwell stress tensor (not shown here). As we have seen in Section V.5.1.a the mechanical movement of a structure under small deformations is given by the linear elasticity equation $-\nabla \cdot (Ae(\mathbf{u})) = F$ where A is the Hooke's law, $e(\mathbf{u})$ is the strain tensor associated to the displacement vector field $\mathbf{u}$, and F is the forces applied on the structure, here the Lorentz force. Since the motion of the material influences the value of the optical indices in space then it also modifies the flow of light. This mechanism translates mathematically into a coupled system between Maxwell equations and the linear elasticity one. To the best of our knowledge, no attempts were made to apply topology optimization methods to this kind of multiphysics problem.

Complete circuit

Since we now have the possibility to optimize each photonic component individually, it may be interesting to set up a method to efficiently design a complete **circuit**. In this direction, a subject has already been initiated at the CEA following this thesis with the aim to design a full photonic circuit with shape-optimized components which have the ability to perform mathematical operations. Without entering into the details about this particular circuit, let us give some generalities about the goal and difficulties in the optimization of a full circuit.

As presented in Section I.3.1.c and Remark I.3.1.1, a nanophotonic component acts as a $N \times N$ product where N is equal to the number of forward guided modes which may exist in all the adjacent waveguides. When N is large, i.e. when the number of inputs and/or outputs is important, the component is often divided into several smaller devices with a fewer number of waveguides connected to each of them and for which it is easier to obtain though optimization a design that achieves a desired functionality. One could for instance decompose any $N \times N$ S-matrix as the product between a diagonal matrix containing N phases shifts (coefficients of the form $\exp(i\phi)$) and $N(N-1)/2$ rotations matrices; see [Pai19].

Implementing a general optimization algorithm that, for a given S-matrix, determines how the circuit should be broken down into known or easily optimized components would speed-up drastically the creation process. Phase management in the circuit is a crucial point; depending on its length, each waveguide connecting two components introduce a phase shift and acts as a multiplication by $\exp(i\phi)$ of the transmission and must be taken into account in the design of the circuit. An important work concerning the treatment of losses and parasitic reflections should also be done in order to guarantee the overall performance of the circuit. Indeed if each component introduces some error of amplitude ε then the transmission error after propagation into d components will be of the order ε^d. A global optimization of the components location could compensate for these errors.

Optimization methods

We finish with a few numerical optimization methods that were only partially studied in this thesis and for which a more comprehensive analysis would be interesting:

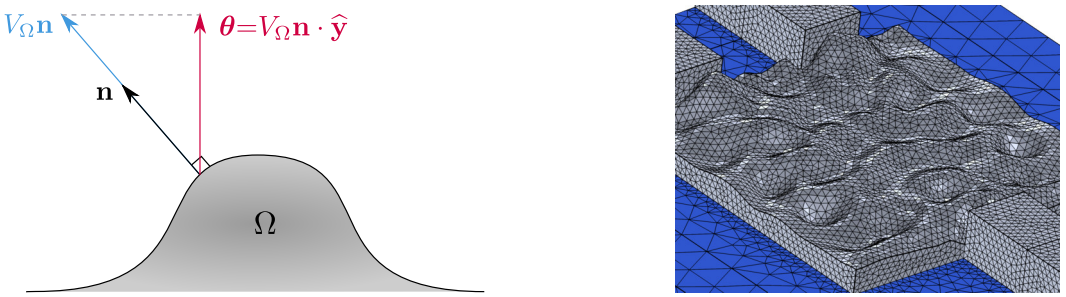
- First, as explained in both Sections II.1.3 and III.3.2 the full derivation of the **topological gradient** formula associated with the general optimization problem

of Th. III.2.1.1 has not been obtained in this thesis. The determination of this gradient would require a complete study of the limiting behavior of the electric field in the presence of cylindrical-shaped perturbation of the medium whose radius tends to zero. Note that the asymptotic in the case of a spherical perturbation has already been studied in [Mas05].

- At the end of Section III.5 we briefly mentioned a method called **grayscale lithography**. Let us enter here into more details about the opportunities offered by this manufacturing process. In most of this thesis, the considered components are y-invariant structures, which restricts the degrees of freedom of the shapes that can be obtained. We partially relaxed this constraint by considering several layers of etching in Section III.5 but we still are limited to a finite number of them. What about the limiting case where an infinite number of layers are considered? This is feasible using grayscale lithography, a manufacturing process which allows to produce components given by a heightmap; we trade the level-set representation of the shape with a height function $M : \mathbb{R}^2 \to [-h/2, h/2]$ such that

$$\Omega = \{(x, y, z) \in \mathcal{D}_{\text{opt}}, y < M(x, z)\}.$$

  This in fact considerably simplifies the numerical representation of shapes and a deformation field $\boldsymbol{\theta}$ preserving the fact that Ω must always be represented by a heightmap is simply found under the form $\boldsymbol{\theta} = \theta\widehat{\mathbf{y}}$ in the shape derivative given by Th. III.2.1.1. An example of power divider found using this methodology is displayed in Fig. Q. A more extensive study of this method would be necessary in order to know if this manufacturing process allows to significantly increase the performances of some nanophotonic components.



(a) *A one dimensional shape and the projection of the shape gradient $V_\Omega\mathbf{n}$ onto the vertical axis*

(b) *Shape optimized power divider with a shape defined as an heightmap*

Figure Q: Using the vector field depicted on the left it is possible to optimize components for the grayscale lithography manufacturing process and end up for instance with the power divider on the right.

- One last interesting research perspective concerns the shape optimization of regions bearing boundary conditions in the state PDE and more precisely the numerical **tracking of a shape on a submanifold** as was implicitly done in the test case of Section V.5.2.c. To consider the simultaneous optimization of a shape $\Omega \subset \mathbb{R}^d$ and a region $\Gamma \subset \mathbb{R}^{d-1}$ on its boundary $\partial\Omega$ a specific numerical representation of Γ must be implemented. One possibility would be to have recourse to the closest-point method [Mac08] which allows level-set representation of shapes living on a submanifold.

Bibliography

[All01] G. Allaire. *Shape Optimization by the Homogenization Method*. Vol. 146. Springer Science & Business Media, 2001. DOI: 10.1007/978-1-4684-9286-6 (cit. on p. 157).

[All04] G. Allaire, F. Jouve, and A.-M. Toader. "Structural optimization using sensitivity analysis and a level-set method". In: *Journal of computational physics* 194.1 (2004), pp. 363–393. DOI: 10.1016/j.jcp.2003.09.032 (cit. on pp. 48, 157).

[All07] G. Allaire. *Conception optimale de structures*. Vol. 58. Springer, 2007. DOI: 10.1007/978-3-540-36856-4 (cit. on pp. 31–34, 39, 43).

[All14a] G. Allaire and C. Dapogny. "A linearized approach to worst-case design in parametric and geometric shape optimization". In: *Mathematical Models and Methods in Applied Sciences* 24.11 (2014), pp. 2199–2257. DOI: 10.1142/S0218202514500195 (cit. on pp. 103, 104).

[All14b] G. Allaire, C. Dapogny, G. Delgado, and G. Michailidis. "Multi-phase structural optimization via a level set method". In: *ESAIM: control, optimisation and calculus of variations* 20.2 (2014), pp. 576–611. DOI: 10.1051/cocv/2013076 (cit. on pp. 44, 75, 145).

[All15] G. Allaire and C. Dapogny. "A deterministic approximation method in shape optimization under random uncertainties". In: *SMAI Journal of Computational Mathematics* 1 (2015), pp. 83–143. DOI: 10.5802/smai-jcm.5 (cit. on pp. 103, 104).

[Amb94] L. Ambrosio and H. M. Soner. "Level set approach to mean curvature flow in arbitrary codimension". In: *Journal of Differential Geometry* 43 (1994), pp. 693–737. DOI: 10.4310/jdg/1214458529 (cit. on p. 135).

[Ams03] S. Amstutz. "Aspects théoriques et numériques en optimisation de forme topologique". PhD thesis. Toulouse, INSA, 2003. URL: http://www.theses.fr/2003ISAT0025 (cit. on p. 39).

[Ams16] S. Amstutz and M. Ciligot-Travain. "A notion of compliance robustness in topology optimization". In: *ESAIM: Control, Optimisation and Calculus of Variations* 22.1 (2016), pp. 64–87. DOI: 10.1051/cocv/2014066 (cit. on p. 103).

[Aze14] H. Azegami, K. Ohtsuka, and M. Kimura. "Shape derivative of cost function for singular point: Evaluation by the generalized J integral". In: *JSIAM Letters* 6 (2014), pp. 29–32. DOI: 10.14495/jsiaml.6.29 (cit. on p. 139).

[Bab90] I. Babuška and M. Suri. "The p-and hp versions of the finite element method, an overview". In: *Computer Methods in Applied Mechanics and Engineering* 80.1-3 (1990), pp. 5–26. DOI: 10.1016/0045-7825(90)90011-A (cit. on p. 143).

[Bar82] M. Barone and D. Caulk. "Optimal arrangement of holes in a two-dimensional heat conductor by a special boundary integral method". In: *International Journal for Numerical Methods in Engineering* 18.5 (1982), pp. 675–685. DOI: 10.1002/nme.1620180505 (cit. on p. 128).

[Baw04] H. Bawa. *Manufacturing processes-II*. Vol. 2. Tata McGraw-Hill Education, 2004. ISBN: 978-0-07-058372-6 (cit. on p. 154).

[Ben13] M. P. Bendsoe and O. Sigmund. *Topology optimization: theory, methods, and applications*. Springer Science & Business Media, 2013. DOI: 10.1007/978-3-662-05086-6 (cit. on p. 157).

[Bre10] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010. DOI: 10.1007/978-0-387-70914-7 (cit. on pp. 20, 47, 132, 134, 146).

[Cal16] F. Callewaert, S. Butun, Z. Li, and K. Aydin. "Inverse design of an ultra-compact broadband optical diode based on asymmetric spatial mode conversion". In: *Scientific reports* 6 (2016), p. 32577. DOI: 10.1038/srep32577 (cit. on p. 101).

[Céa00] J. Céa, S. Garreau, P. Guillaume, and M. Masmoudi. "The shape and topological optimizations connection". In: *Computer methods in applied mechanics and engineering* 188.4 (2000), pp. 713–726. DOI: 10.1016/S0045-7825(99)00357-6 (cit. on p. 40).

[Céa86] J. Céa. "Conception optimale ou identification de formes, calcul rapide de la dérivée directionnelle de la fonction coût". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 20.3 (1986), pp. 371–402. URL: http://www.numdam.org/item/M2AN_1986__20_3_371_0 (cit. on p. 41).

[Che11] S. Chen and W. Chen. "A new level-set based approach to shape and topology optimization under geometric uncertainty". In: *Structural and Multidisciplinary Optimization* 44.1 (2011), pp. 1–18. DOI: 10.1007/s00158-011-0660-9 (cit. on p. 114).

[Cos88] M. Costabel. "Boundary integral operators on Lipschitz domains: elementary results". In: *SIAM Journal on Mathematical Analysis* 19.3 (1988), pp. 613–626. DOI: 10.1137/0519043 (cit. on p. 151).

[Cos96] M. Costabel and M. Dauge. "A Singularly mixed boundary value problem". In: *Communications in Partial Differential Equations* 21.11-12 (1996), pp. 1919–1949. DOI: 10.1080/03605309608821249 (cit. on p. 149).

[Dap13] C. Dapogny. "Shape optimization, level set methods on unstructured meshes and mesh evolution". PhD thesis. Université Pierre et Marie Curie-Paris 6, 2013. URL: https://tel.archives-ouvertes.fr/tel-00916224 (cit. on pp. 31, 55, 76, 93, 115, 135, 145).

[Dap19] C. Dapogny, N. Lebbe, and E. Oudet. "Optimization of the shape of regions supporting boundary conditions". working paper or preprint. 2019. URL: https://hal.archives-ouvertes.fr/hal-02064477v1 (cit. on pp. xvi, 127, 144).

[Dau06] M. Dauge. *Elliptic boundary value problems on corner domains: smoothness and asymptotics of solutions*. Vol. 1341. Springer, 2006. DOI: 10.1007/BFb0086682 (cit. on p. 132).

[Dau12] R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology: volume 3 Spectral Theory and Applications*. Springer Science & Business Media, 2012. DOI: 10.1007/978-3-642-61529-0 (cit. on pp. 66, 75).

[De 05] F. De Gournay. "Shape optimization by the level-set method". PhD thesis. Ecole Polytechnique X, 2005. URL: https://tel.archives-ouvertes.fr/tel-00446039/ (cit. on pp. 31, 58).

[De 16] M. De Buhan, C. Dapogny, P. Frey, and C. Nardoni. "An optimization method for elastic shape matching". In: *Comptes Rendus Mathématique* 354.8 (2016), pp. 783–787. DOI: 10.1016/j.crma.2016.05.007 (cit. on p. 112).

[Del11] M. C. Delfour and J.-P. Zolésio. *Shapes and geometries: metrics, analysis, differential calculus, and optimization*. Vol. 22. Siam, 2011. DOI: 10.1137/1.9780898719826 (cit. on pp. 43, 116).

[Dés12] J.-A. Désidéri. "Multiple-gradient descent algorithm (MGDA) for multiobjective optimization". In: *Comptes Rendus Mathematique* 350.5-6 (2012), pp. 313–318. DOI: 10.1016/j.crma.2012.03.014 (cit. on p. 97).

[Des18] J. Desai, A. Faure, G. Michailidis, G. Parry, and R. Estevez. "Topology optimization in acoustics and elasto-acoustics via a level-set method". In: *Journal of Sound and Vibration* 420 (2018), pp. 73–103. DOI: 10.1016/j.jsv.2018.01.032 (cit. on pp. 128, 145, 146).

[Di 12] E. Di Nezza, G. Palatucci, and E. Valdinoci. "Hitchhiker's guide to the fractional Sobolev spaces". In: *Bulletin des Sciences Mathématiques* 136.5 (2012), pp. 521–573. DOI: 10.1016/j.bulsci.2011.12.004 (cit. on pp. 133, 150).

[Dij13] N. P. van Dijk, K. Maute, M. Langelaar, and F. Van Keulen. "Level-set methods for structural topology optimization: a review". In: *Structural and Multidisciplinary Optimization* 48.3 (2013), pp. 437–472. DOI: 10.1007/s00158-013-0912-y (cit. on pp. 52, 53).

[Dun15] P. D. Dunning and H. A. Kim. "Introducing the sequential linear programming level-set method for topology optimization". In: *Structural and Multidisciplinary Optimization* 51.3 (2015), pp. 631–643. DOI: 10.1007/s00158-014-1174-z (cit. on p. 143).

[Ele12] Y. Elesin, B. S. Lazarov, J. S. Jensen, and O. Sigmund. "Design of robust and efficient photonic switches using topology optimization". In: *Photonics and nanostructures-Fundamentals and Applications* 10.1 (2012), pp. 153–165. DOI: 10.1016/j.photonics.2011.10.003 (cit. on p. 103).

[Ell05] M. Elliotis, G. Georgiou, and C. Xenophontos. "Solving Laplacian problems with boundary singularities: a comparison of a singular function boundary integral method with the p/hp version of the finite element method". In: *Applied Mathematics and Computation* 169.1 (2005), pp. 485–499. DOI: 10.1016/j.amc.2004.09.058 (cit. on p. 143).

[Fep18] F. Feppon, G. Allaire, F. Bordeu, J. Cortial, and C. Dapogny. "Shape optimization of a coupled thermal fluid-structure problem in a level set mesh evolution framework". In: *SeMA Journal* (2018), pp. 1–46. DOI: 10.1007/s40324-018-00185-4 (cit. on p. 47).

[Fre01] G. Fremiot and J. Sokolowski. "Shape sensitivity analysis of problems with singularities". In: *Lecture notes in pure and applied mathematics* (2001), pp. 255–276. DOI: 10.1201/9780203904169.ch9 (cit. on pp. 129, 139).

[Fri89] A. Friedman. *Mathematical problems in electron beam lithography*. Springer, 1989, pp. 79–87. DOI: 10.1007/978-1-4615-7402-6_9 (cit. on p. 110).

[Gia17] M. Giacomini, O. Pantz, and K. Trabelsi. "Volumetric expressions of the shape gradient of the compliance in structural shape optimization". In: *arXiv preprint arXiv:1701.05762* (2017). URL: https://arxiv.org/abs/1701.05762 (cit. on p. 143).

[Gon08] A. Gondarenko and M. Lipson. "Low modal volume dipole-like dielectric slab resonator". In: *Optics express* 16.22 (2008), pp. 17689–17694. DOI: 10.1364/OE.16.017689 (cit. on p. 63).

[Gou10] B. Goursaud. "Etude mathématique et numérique de guides d'ondes ouverts non uniformes, par approche modale". PhD thesis. Ecole Polytechnique X, 2010. URL: https://hal.inria.fr/pastel-00546093 (cit. on p. 8).

[Gri11] P. Grisvard. *Elliptic problems in nonsmooth domains*. SIAM, 2011. ISBN: 978-1-611972-02-3 (cit. on pp. 132–134, 136, 138, 142).

[Had08] J. Hadamard. *Mémoire sur le problème d'analyse relatif à l'équilibre des plaques élastiques encastrées*. Vol. 33. Imprimerie nationale, 1908. URL: https://gallica.bnf.fr/ark:/12148/bpt6k3338m/f611.item (cit. on p. 32).

[Ham17] M. Hammer. *Optical Waveguide Theory*. 2017. URL: https://www.computational-photonics.eu/theory.html (cit. on pp. 1, 7).

[Han14] P. Hansen. "Adjoint sensitivity analysis for nanophotonic structures". PhD thesis. Stanford University, 2014. URL: https://purl.stanford.edu/zm205vy5828 (cit. on p. 76).

[Hec12] F. Hecht. "New development in FreeFem++". In: *J. Numer. Math.* 20.3-4 (2012), pp. 251–265. ISSN: 1570-2820. URL: https://freefem.org/ (cit. on p. 28).

[Hen06] A. Henrot and M. Pierre. *Variation et optimisation de formes: une analyse géométrique*. Vol. 48. Springer Science & Business Media, 2006. DOI: 10.1007/3-540-37689-5 (cit. on pp. 31, 32, 34–36, 74, 135).

[Hip15] R. Hiptmair, A. Paganini, and S. Sargheini. "Comparison of approximate shape gradients". In: *BIT Numerical Mathematics* 55.2 (2015), pp. 459–485. DOI: 10.1007/s10543-014-0515-z (cit. on p. 143).

[Hug18] T. W. Hughes, M. Minkov, I. A. Williamson, and S. Fan. "Adjoint Method and Inverse Design for Nonlinear Nanophotonic Devices". In: *ACS Photonics* 5.12 (2018), pp. 4781–4787. DOI: 10.1021/acsphotonics.8b01522 (cit. on p. 168).

[Jan13] M. Jansen, B. S. Lazarov, M. Schevenels, and O. Sigmund. "On the similarities between micro/nano lithography and topology optimization projection methods". In: *Structural and Multidisciplinary Optimization* 48.4 (2013), pp. 717–730. DOI: 10.1007/s00158-013-0941-6 (cit. on p. 112).

[Jan96] H. Jansen, H. Gardeniers, M. de Boer, M. Elwenspoek, and J. Fluitman. "A survey on the reactive ion etching of silicon in microtechnology". In: *Journal of micromechanics and microengineering* 6.1 (1996), p. 14. DOI: 10.1088/0960-1317/6/1/002 (cit. on p. 111).

[Jen11] J. S. Jensen and O. Sigmund. "Topology optimization for nano-photonics". In: *Laser & Photonics Reviews* 5.2 (2011), pp. 308–321. DOI: 10.1002/lpor.201000014 (cit. on p. 64).

[Jin14] J.-M. Jin. *The finite element method in electromagnetics*. John Wiley & Sons, 2014. ISBN: 978-1-118-57136-1 (cit. on pp. 1, 14–17, 28).

[Joh02] S. G. Johnson et al. "Perturbation theory for Maxwell's equations with shifting material boundaries". In: *Physical review E* 65.6 (2002), p. 066611. DOI: 10.1103/PhysRevE.65.066611 (cit. on p. 66).

[Kao05] C. Y. Kao, S. Osher, and E. Yablonovitch. "Maximizing band gaps in two-dimensional photonic crystals by using level set methods". In: *Applied Physics B* 81.2-3 (2005), pp. 235–244. DOI: 10.1007/s00340-005-1877-3 (cit. on p. 65).

[Kay06] N. Kaya. "Machining fixture locating and clamping position optimization using genetic algorithms". In: *Computers in Industry* 57.2 (2006), pp. 112–120. DOI: 10.1016/j.compind.2005.05.001 (cit. on p. 154).

[Kot08] C. Kottke, A. Farjadpour, and S. G. Johnson. "Perturbation theory for anisotropic dielectric interfaces, and application to subpixel smoothing of discretized numerical methods". In: *Physical Review E* 77.3 (2008), p. 036611. DOI: 10.1103/PhysRevE.77.036611 (cit. on p. 76).

[Koz97] V. A. Kozlov, V. Mazia, and J. Rossmann. *Elliptic boundary value problems in domains with point singularities*. Vol. 52. American Mathematical Soc., 1997. DOI: 10.1090/surv/052 (cit. on p. 132).

[Lal13] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch. "Adjoint shape optimization applied to electromagnetic design". en. In: *Optics Express* 21.18 (2013), p. 21693. ISSN: 1094-4087. DOI: 10.1364/OE.21.021693 (cit. on p. 65).

[Lan12] S. Lang. *Fundamentals of differential geometry*. Vol. 191. Springer Science & Business Media, 2012. DOI: 10.1007/978-1-4612-0541-8 (cit. on p. 46).

[Laz12] B. S. Lazarov, M. Schevenels, and O. Sigmund. "Topology optimization with geometric uncertainties by perturbation techniques". In: *International Journal for Numerical Methods in Engineering* 90.11 (2012), pp. 1321–1336. DOI: 10.1002/nme.3361 (cit. on p. 103).

[Leb19a] N. Lebbe, C. Dapogny, E. Oudet, K. Hassan, and A. Gliere. "Robust shape and topology optimization of nanophotonic devices using the level set method". In: *Journal of Computational Physics* (2019). DOI: 10.1016/j.jcp.2019.06.057 (cit. on pp. xvi, 61, 70, 95).

[Leb19b] N. Lebbe, A. Glière, and K. Hassan. "High-efficiency and broadband photonic polarization rotator based on multilevel shape optimization". In: *Optics Letters* 44.8 (2019), pp. 1960–1963. DOI: 10.1364/OL.44.001960 (cit. on pp. xvi, 61).

[Leb19c] N. Lebbe, A. Gliere, K. Hassan, C. Dapogny, and E. Oudet. "Shape optimization for the design of passive mid-infrared photonic components". In: *Optical and Quantum Electronics* 51.5 (2019), pp. 166–179. DOI: 10.1007/s11082-019-1849-1 (cit. on pp. xvi, 119).

[Lee91] J.-F. Lee, D.-K. Sun, and Z. J. Cendes. "Full-wave analysis of dielectric waveguides using tangential vector finite elements". In: *IEEE Transactions on Microwave Theory and Techniques* 39.8 (1991), pp. 1262–1271. DOI: 10.1109/22.85399 (cit. on pp. 7, 26).

[Li00] Z.-C. Li and T.-T. Lu. "Singularities and treatments of elliptic boundary value problems". In: *Mathematical and Computer Modelling* 31.8-9 (2000), pp. 97–145. DOI: 10.1016/S0895-7177(00)00062-5 (cit. on p. 143).

[Lio68] J.-L. Lions and E. Magenes. "Problèmes aux limites non homogènes et applications. Volume I". In: (1968) (cit. on p. 151).

[Liu18] Y. Liu et al. "Very sharp adiabatic bends based on an inverse design". In: *Optics letters* 43.11 (2018), pp. 2482–2485. DOI: 10.1364/OL.43.002482 (cit. on p. 64).

[Lu13] J. Y.-s. Lu. "Nanophotonic Computational Design". PhD thesis. Stanford University, 2013. URL: http://purl.stanford.edu/sb598wt7448 (cit. on p. 64).

[Ma11] J. Ma, M. Y. Wang, and X. Zhu. "Compliant fixture layout design using topology optimization method". In: *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE. 2011, pp. 3757–3763. DOI: 10.1109/ICRA.2011.5979876 (cit. on pp. 128, 154).

[Mac08] C. B. Macdonald and S. J. Ruuth. "Level set equations on surfaces via the Closest Point Method". In: *Journal of Scientific Computing* 35.2-3 (2008), pp. 219–240 (cit. on p. 170).

[Maj17] A. Majumder, B. Shen, R. Polson, and R. Menon. "Ultra-compact polarization rotation in integrated silicon photonics using digital metamaterials". In: *Optics express* 25.17 (2017), pp. 19721–19731. DOI: 10.1364/OE.25.019721 (cit. on p. 64).

[Mak16] J. C. Mak, C. Sideris, J. Jeong, A. Hajimiri, and J. K. Poon. "Binary particle swarm optimized 2× 2 power splitters in a standard foundry silicon photonic platform". In: *Optics letters* 41.16 (2016), pp. 3868–3871. DOI: 10.1364/OL.41.003868 (cit. on p. 63).

[Man18] V. Manet. *Méthode des Éléments Finis: vulgarisation des aspects mathématiques et illustration de la méthode*. 2018. URL: https://cel.archives-ouvertes.fr/cel-00763690 (cit. on pp. 20, 150).

[Mar16] J. Martínez-Frutos, D. Herrero-Pérez, M. Kessler, and F. Periago. "Robust shape optimization of continuous structures via the level set method". In: *Computer Methods in Applied Mechanics and Engineering* 305 (2016), pp. 271–291. DOI: 10.1016/j.cma.2016.03.003 (cit. on p. 103).

[Mas05] M. Masmoudi, J. Pommier, and B. Samet. "The topological asymptotic expansion for the Maxwell equations and some applications". In: *Inverse Problems* 21.2 (2005), p. 547. DOI: 10.1088/0266-5611/21/2/008 (cit. on p. 170).

[Mau14] K. Maute. "Topology optimization under uncertainty". In: *Topology optimization in structural and continuum mechanics*. Springer, 2014, pp. 457–471. DOI: 10.1007/978-3-7091-1643-2_20 (cit. on p. 103).

[McL00] W. C. H. McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge university press, 2000. DOI: 978-0-52-166375-5 (cit. on p. 150).

[Mic14] G. Michailidis. "Manufacturing constraints and multi-phase shape and topology optimization via a level-set method". PhD thesis. Ecole Polytechnique X, 2014. URL: https://tel.archives-ouvertes.fr/pastel-00937306 (cit. on pp. 31, 89, 116).

[Mic18] A. Michaels and E. Yablonovitch. "Leveraging continuous material averaging for inverse electromagnetic design". In: *Optics express* 26.24 (2018), pp. 31717–31737. DOI: 10.1364/OE.26.031717 (cit. on p. 76).

[Mil13] O. D. Miller. "Photonic design: From fundamental solar cell physics to computational inverse design". PhD thesis. 2013. URL: https://arxiv.org/abs/1308.0212 (cit. on p. 66).

[Mol18] S. Molesky et al. "Inverse design in nanophotonics". In: *Nature Photonics* 12.11 (2018), p. 659. DOI: 10.1038/s41566-018-0246-9 (cit. on p. 103).

[Mon03] P. Monk. *Finite element methods for Maxwell's equations*. Oxford University Press, 2003. DOI: 10.1093/acprof:oso/9780198508885.001.0001 (cit. on pp. 1, 4, 17, 20, 22, 28, 66, 75).

[Mur75] F. Murat and J. Simon. "Etude de problèmes d'optimal design". In: (1975), pp. 54–62. DOI: 10.1007/3-540-07623-9_279 (cit. on p. 32).

[Néd01] J.-C. Nédélec. *Acoustic and electromagnetic equations: integral representations for harmonic problems*. Springer Science & Business Media, 2001. DOI: 10.1007/978-1-4757-4393-7 (cit. on p. 4).

[Orf02] S. J. Orfanidis. "Electromagnetic waves and antennas". In: (2002). URL: https://www.ece.rutgers.edu/~orfanidi/ewa/ (cit. on p. 1).

[Ott17] J. Ott. "Halfspace Matching: a Domain Decomposition Method for Scattering by 2D Open Waveguides". PhD thesis. Karlsruher Institut für Technologie (KIT), 2017. DOI: 10.5445/IR/1000070898 (cit. on p. 17).

[Pai19] S. Pai, B. Bartlett, O. Solgaard, and D. A. Miller. "Matrix optimization on universal unitary photonic devices". In: *Physical Review Applied* 11.6 (2019), p. 064044. DOI: 10.1103/PhysRevApplied.11.064044 (cit. on p. 169).

[Pan05] O. Pantz. "Sensibilité de l'équation de la chaleur aux sauts de conductivité". In: *Comptes Rendus Mathematique* 341.5 (2005), pp. 333–337. DOI: 10.1016/j.crma.2005.07.005 (cit. on pp. 43, 71, 73).

[Pig15] A. Y. Piggott et al. "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer". In: *Nature Photonics* 9.6 (2015), p. 374. DOI: 10.1038/nphoton.2015.69 (cit. on p. 64).

[Pig17] A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković. "Fabrication-constrained nanophotonic inverse design". In: *Scientific reports* 7.1 (2017), p. 1786. DOI: 10.1038/s41598-017-01939-2 (cit. on pp. 65, 88).

[Pir82] O. Pironneau. *Optimal shape design for elliptic systems*. Springer, 1982. DOI: 10.1007/978-3-642-87722-3 (cit. on pp. 31, 36, 57).

[Sch10] J. B. Schneider. "Understanding the finite-difference time-domain method". In: *School of electrical engineering and computer science Washington State University* (2010). URL: https://www.eecs.wsu.edu/~schneidj/ufdtd/ (cit. on pp. 28, 29).

[Sel13] S. Selvakumar, K. Arulshri, K. Padmanaban, and K. Sasikumar. "Design and optimization of machining fixture layout using ANN and DOE". In: *The International Journal of Advanced Manufacturing Technology* 65.9-12 (2013), pp. 1573–1586. DOI: 10.1007/s00170-012-4281-2 (cit. on p. 154).

[Set99] J. A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Vol. 3. Cambridge university press, 1999. ISBN: 978-0-52-164557-7 (cit. on pp. 48, 53–55).

[She15] B. Shen, P. Wang, R. Polson, and R. Menon. "An integrated-nanophotonics polarization beamsplitter with 2.4× 2.4 μm 2 footprint". In: *Nature Photonics* 9.6 (2015), p. 378. DOI: 10.1038/nphoton.2015.80 (cit. on p. 64).

[Sig09] O. Sigmund. "Manufacturing tolerant topology optimization". In: *Acta Mechanica Sinica* 25.2 (2009), pp. 227–239. DOI: 10.1007/s10409-009-0240-z (cit. on p. 111).

[Sny83] A. W. Snyder and J. D. Love. *Optical waveguide theory*. Springer Science & Business Media, 1983. DOI: 10.1007/978-1-4613-2813-1 (cit. on pp. 1, 9).

[Sok09] J. Sokolowski and A. Zochowski. "Topological derivative in shape optimization". In: *Encyclopedia of Optimization* (2009), pp. 3908–3918. DOI: 10.1007/978-0-387-74759-0_682 (cit. on pp. 31, 39).

[Tah19] M. H. Tahersima et al. "Deep Neural Network Inverse Design of Integrated Photonic Power Splitters". In: *Scientific reports* 9.1 (2019), p. 1368. DOI: 10.1038/s41598-018-37952-2 (cit. on p. 63).

[Tsi11] O. Tsilipakos, E. Kriezis, and T. Yioultsis. "Boundary condition for the efficient excitation and absorption of hybrid waveguide modes in finite element formulations". In: *Microwave and Optical Technology Letters* 53.11 (2011), pp. 2626–2631. DOI: 10.1002/mop.26364 (cit. on p. 15).

[Ver19a] D. Vercruysse, N. V. Sapra, L. Su, R. Trivedi, and J. Vučković. "Analytical level set fabrication constraints for inverse design". In: *Scientific reports* 9.1 (2019), p. 8999. DOI: 10.1038/s41598-019-45026-0 (cit. on pp. 65, 89).

[Ver19b] N. Vermaak et al. "Topological Optimization with Interfaces". In: *Architectured Materials in Nature and Engineering*. Springer, 2019, pp. 173–193. DOI: 10.1007/978-3-030-11942-3_6 (cit. on p. 75).

[Vié16] J.-L. Vié. "Second-order derivatives for shape optimization with a level-set method". PhD thesis. Paris Est, 2016. URL: https://hal.archives-ouvertes.fr/tel-01488770 (cit. on pp. 31, 56).

[Wan11] F. Wang, J. S. Jensen, and O. Sigmund. "Robust topology optimization of photonic crystal waveguides with tailored dispersion properties". In: *JOSA B* 28.3 (2011), pp. 387–397. DOI: 10.1364/JOSAB.28.000387 (cit. on pp. 103, 111).

[Wes17] W. J. Westerveld and H. P. Urbach. *Silicon Photonics*. 2053-2563. IOP Publishing, 2017. ISBN: 978-0-7503-1386-5. DOI: 10.1088/978-0-7503-1386-5 (cit. on pp. 1, 7–9).

[Xia14] Q. Xia, M. Y. Wang, and T. Shi. "A level set method for shape and topology optimization of both structure and support of continuum structures". In: *Computer*

Methods in Applied Mechanics and Engineering 272 (2014), pp. 340–353. DOI: 10.1016/j.cma.2014.01.014 (cit. on pp. 128, 158).

[Xia16] Q. Xia and T. Shi. "Topology optimization of compliant mechanism and its support through a level set method". In: *Computer Methods in Applied Mechanics and Engineering* 305 (2016), pp. 359–375. DOI: 10.1016/j.cma.2016.03.017 (cit. on pp. 128, 158).

[Xu17] K. Xu et al. "Integrated photonic power divider with arbitrary power ratios". In: *Optics letters* 42.4 (2017), pp. 855–858. DOI: 10.1364/OL.42.000855 (cit. on p. 64).

[Zho14] M. Zhou, B. S. Lazarov, and O. Sigmund. "Topology optimization for optical projection lithography with manufacturing uncertainties". In: *Applied optics* 53.12 (2014), pp. 2720–2729. DOI: 10.1364/AO.53.002720 (cit. on pp. 110, 112, 113).

[Zho15] M. Zhou, B. S. Lazarov, F. Wang, and O. Sigmund. "Minimum length scale in topology optimization by geometric constraints". In: *Computer Methods in Applied Mechanics and Engineering* 293 (2015), pp. 266–282. DOI: 10.1016/j.cma.2015.05.003 (cit. on p. 88).

[Zhu06] Y. Zhu and A. C. Cangellaris. *Multigrid finite element methods for electromagnetic field modeling*. Vol. 28. John Wiley & Sons, 2006. DOI: 10.1002/0471786381 (cit. on p. 24).

Remerciements

Je tiens à commencer mes remerciements en saluant individuellement chacun de mes quatre encadrants de thèse sans qui le contenu de ce manuscrit ne pourrait être aussi abouti aujourd'hui. Tout d'abord merci à Alain Glière pour ces trois années passées, je pense aujourd'hui que je n'aurais pas pu rêver meilleur encadrement au quotidien ; merci d'avoir toujours été à l'écoute, pour ta gentillesse et pour ta motivation indéfectible tout le long de ma thèse. Vient ensuite Charles Dapogny, que ce soit pour la modélisation mathématique et physique, les calculs numériques ou l'analyse des résultats, j'ai toujours pu compter sur toi et je t'en remercie chaleureusement. Merci aussi à mon directeur de thèse Édouard Oudet de m'avoir soutenu du début à la fin et ce toujours avec le sourire ! Et enfin merci à Karim Hassan d'avoir toujours su trouver des moments pour moi dans son emploi du temps pour toutes mes questions, mais aussi pour la fabrication et la caractérisation des composants qui n'auraient pas été envisageables sans un tel investissement de sa part.

Concernant la réalisation des composants justement, je me dois de remercier l'ensemble des ingénieurs et techniciens du CEA avec qui nous avons collaboré et qui, pendant plusieurs années, ont dû subir les demandes farfelues d'un jeune mathématicien qui souhaitait fabriquer des composants aux formes quelque peu extravagantes !

Je remercie bien sûr les deux rapporteurs de ma thèse Yannick Privat et Amélie Litman d'avoir bien voulu participer à l'évaluation de ce travail. Merci aussi à François Jouve et Stéphane Lantéri d'avoir accepté de faire le déplacement jusqu'à Grenoble pour ma présentation et pour le temps consacré à la lecture de mon manuscrit.

Arrivent à présent les remerciements de tous les membres du laboratoire des capteurs optiques du CEA avec qui j'ai passé trois années très enrichissantes, tant d'un point de vue humain que scientifique. Par ordre d'ancienneté pour les thésards : Cédric et Julien qui ont eu à supporter mes (trop) nombreux questionnements en début de thèse. L'affreux (copyright Cédric) Boris avec qui j'ai malheureusement dû passer ces trois années et qui m'a forcé à de nombreuses séances de blocs alors que j'aurais pu tranquillement me reposer chez moi ... Gabby (après décompte il y a 302 images dans ma thèse, t'as intérêt à faire mieux !), Thomas (bon courage avec Comsol !), et les petits derniers : Joris (normalement j'ai plus de trucs à te faire sous-traiter pour moi !), Joël (bon courage pour reprendre le flambeau du débogage de code !) et Jhouben (dans deux ans tu fais ta soutenance en français n'est-ce pas ?). Je remercie aussi tous les permanents et non-permanents du labo avec qui j'ai pris beaucoup de plaisir à discuter. Il m'est malheureusement impossible de citer tout le monde sans oublier de nom aussi j'espère que les oubliés me pardonneront : Mathieu (tu sais comment contacter la hotline Matlab !), Adrien, Salim, Laurent, Stone the bee, Pierre, Jean-Marc, Jean-Guillaume, Jules, Maryse, Olivier, Alexandre, Jade, Christophe, Gil (j'attends les résultats du postdoc !) ...

Enfin, je remercie aussi ma famille et mes amis pour leur soutien et encouragement durant ces trois dernières années. Une pensée particulière pour Lucie, Florian et Raphaël avec qui j'ai commencé ma thèse au même moment après notre école d'ingénieur et qui ont été de précieux soutiens pour moi.

Index

Abstract

This thesis focuses on the mathematical field of shape optimization and explores two topics: one concerns the systematic determination of the design of nanophotonic components and the other one the optimal shape and location of boundary conditions defining partial differential equations (PDE).

- In the mathematical setting of the three-dimensional, time-harmonic Maxwell equations, we propose a shape and topology optimization algorithm combining Hadamard's boundary variation method with a level set representation of shapes and their evolution. A particular attention is devoted to the robustness of the optimized devices with respect to small uncertainties over the physical or geometrical data of the problem. In this respect, we rely on a simple multi-objective formulation to deal with the two main sources of uncertainties plaguing nanophotonic devices, namely uncertainties over the incoming wavelength, and geometric uncertainties entailed by the lithography and etching fabrication process. Several numerical examples are presented and discussed to assess the efficiency of our methodology.
- The second application concern the optimization of the shape of the regions assigned to different types of boundary conditions in the definition of a "physical" PDE. This problem proves to be difficult in the case of a Dirichlet-Neumann transition since it requires a precise study of the singular nature of PDE solutions at the transition between two regions supporting these boundary conditions. On the one hand a full mathematical study is carried out on this theoretical problem and on the other hand a numerical method based on a regularization of the boundary conditions is proposed to optimize these regions. Various numerical examples are eventually presented in order to appraise the efficiency of the proposed process.

Keywords: *topology optimization · level-set method · nanophotonics · robustness · Maxwell equations · boundary conditions*

Résumé

Cette thèse contribue au domaine mathématique de l'optimisation de forme et explore deux sujets: l'une concerne la détermination automatique du design de composant nanophotonique et l'autre la répartion optimale des conditions aux limites permettant de définir une équation au dérivées partielles (EDP).

- Dans le cadre mathématique des équations de Maxwell tridimensionnelles et harmoniques dans le temps, nous proposons un algorithme d'optimisation de forme combinant la méthode d'Hadamard et une représentation des formes par la méthode level-set. Une attention particulière est accordée à la robustesse des dispositifs optimisés par rapport à des petites incertitudes sur les données physiques ou géométriques du problème. À cet égard, nous nous appuyons sur un algorithme provenant du domaine de l'optimisation multi-objectifs pour traiter les deux principales sources d'incertitudes affectant des dispositifs nanophotoniques, à savoir les incertitudes sur la longueur d'onde de la lumière injectée ainsi que les incertitudes géométriques liées au procédé de fabrication par lithographie-gravure. Plusieurs exemples numériques sont présentés permettant d'évaluer l'efficacité de la méthode proposé.
- La deuxième application concerne l'optimisation de la forme des régions affectées à différentes conditions limites dans la définition d'une EDP d'un problème "physique". L'étude de ce problème s'avère être délicate dans le cas d'une transition Dirichlet-Neumann car elle nécessite une analyse précise de la singularité des solutions d'une EDP à la transition entre deux régions supportant ces conditions limites. Nous proposons d'une part une étude mathématique complète de ce problème dans le cas de l'équation du Laplacien et d'autre part une méthode numérique basée sur une régularisation des conditions aux limites pour optimiser la forme de ces régions. Différents exemples numériques sont présentés afin d'évaluer l'efficacité de notre méthode.

Mots clés : *optimisation de forme · méthode level-set · nanophotonique · robustesse · équations de Maxwell · conditions aux limites*