UNIVERSITÀ DEGLI STUDI DI TRENTO

DOTTORATO DI RICERCA IN MATEMATICA

XIV CICLO

Tesi presentata per il conseguimento del titolo di Dottore di Ricerca

Luca Gerardo-Giorda

Domain Decomposition Algorithms for Transport and Wave Propagation Equations

Relatore

Prof. Alberto Valli

9 Dicembre 2002

Acknowledgements

Many people left a deep mark into my life during my four Ph.D. years, and I am afraid that there isn't enough room to thank them all in here.

First of all Alberto, my great advisor, who introduced me to the theory of Domain Decomposition Methods, for his many valuable suggestions and comments, but most of all for his friendship and his support throughout my work.

Patrick Le Tallec and Frédéric Nataf for having given me the opportunity to spend nine wonderful months at the École Polytechnique in Paris, and all the people there at CMAP (it has been a great pleasure being there!), for having offered me a very stimulating environment to work and for the many interesting and useful discussions. In particular I thank the team of Wednesday night dinners, who made me feel at home in Paris: Victorita, Paola, Hervé, Snorre (Takk!), and Olga.

Nicolas, Steve, Valérie, Patrice, Stephanie and all the people at the Maison des Étudiants Canadiens at the Cité Universitaire in Paris, in particular Fedo and Perico: I've found good friends among them and we've had some great time together!

All the people at the Department of Mathematics in Trento University, that made unique the period I spent there, from my roommate Velitchko, from whom, in 4 years, I've been able to learn nothing but 4 words in Bulgarian (it sounds like 1 word per year, I didn't apply that much, probably..) and the way to write down my name in Cyrillic, to Marcello, Erika, Michela, Marina, Claretta, Nadia, Vincenzo, Fabio, Gippo, and Beppe (in a rigorously randomized order).

My old companions outside the university, whose friendship is one of the keystones of my life, even if distance does not help us to meet frequently: thanks Alberto, Lollo, Marco, Aldo, and Manuele!

My Dad and my Mum, for their presence, their continuous encouragement and help (sometimes beyond my actual needs as any good parent does) and my little brother, who keeps me aware that mathematics is not everything, and whose discography is a wishing well for the soundtrack of my life.

Finally, the greatest thank goes to Francesca, because she exists, because of her encouragement, of her precious support, and because of her infinite patience to suffer my (sometimes) impossible character.

Contents

Chapter 1

Introduction

Domain decomposition methods for the numerical solution of partial differential equations is a relatively recent field of research. Though the earliest domain decomposition algorithm for elliptic problems is believed to be the alternating method discovered by Hermann A. Schwarz in 1869, who used it to establish the existence of harmonic functions on regions with complex geometries and non-smooth boundaries, the first key ideas emerged in the early eighties through the works of several scientists among which James Bramble, Toni Y. Chan, Roland Glowinski, Yuri Kuznetsov, Patrick Le Tallec, Pierre-Louis Lions, Alfio Quarteroni and Olof B. Widlund.

The numerical solution of differential problems of practical interest can be a difficult task to face: problems issued from Computational Mechanics are usually set on complex geometries and discretized on very fine grids, leading to large-scale algebraic systems. The widespread availability of parallel computers increased the need to design algorithms especially suited to better fit such architectures. The most promising answer to such need appeared to be the use of domain decomposition algorithms, which can be seen as a divide-and-conquer method, whose basic idea is the following: the given computational domain, denoted with Ω, is partitioned into subdomains Ω_i, $i = 1, \ldots, M$, which may or may not overlap. The original problem is then reformulated upon each subdomain Ω_i, yielding a family of subproblems of reduced size that are coupled one to another through the values of the unknown solution at subdomain interfaces. Very often the interface coupling is removed at the expense of introducing an iterative process among subdomains, yielding at each step independent subproblems (of lower complexity) upon subdomains, which can be efficiently faced by multiprocessor systems.

The earliest works on domain decomposition proposed, and analyzed, algorithms for linear, second order, self-adjoint, positive definite elliptic model problems in two (or very few) subregions. As the field matured, scientists were led to face more complex problems set on many subregions: many such problems issued from Computational Fluid Mechanics, where the need of improving the capability of calculation in regions with complex boundaries was accompanied by the willing to merge Euler's equation, Navier-Stokes equations, potential flows, and other models, each used in a suitable subregion of the computational domain, into a single computational model. The extension of domain decomposition methods to linear parabolic or hyperbolic problem was quite straightforward, whereas the task of dealing with problems that were both nonlinear and time dependent, appeared far more difficult.

If, on one hand, domain decomposition methods are nowadays well-understood, as the annual international symposium on the research on this area held in Cocoyoc (Mexico) in January this year was the fourteenth of the series, and some survey works have already been published (among them, we recall the book by P. Bjørstad, W. Gropp and B. Smith ([95] - 1996), the book by A. Quarteroni and A. Valli ([89] - 1999), and the recent book by A. Toselli and L. Pavarino, ([85] - 2002)), on the other hand the research on the subject is still very active.

In the recent years, many people worked on the Euler system of compressible gas dynamics, proposing algorithms for both sonic and transonic flows on three dimensional domains with unstructured meshes (among them, we can recall the works by X.-C. Cai and his colleagues). To our knowledge, only the work by V. Dolean, S. Lanteri and F. Nataf provides a theoretical convergence analysis (in both two and three dimensions) for a non-overlapping Schwarz algorithm applied to the Euler system (see [40]): the result is accomplished through a linearization of the flux via a *frozen coefficients* technique, which consists in linearizing in the neighborhood of a constant state. In this thesis, an attempt to provide a convergence result for an iteration-by-subdomain procedure is given. The result is obtained for one dimensional isentropic flows: the problem is advanced in time by means of a semi-implicit method, leading to a linearization without freezing the coefficients, and emphasis is put on the spatial decomposition. The algorithm is proved to converge at both time- and fully-discrete level.

Besides, stemming from the remark that they are intrinsically slow, domain decomposition algorithms are often used, in the case of linear problems, as preconditioners for Krylov subspace accelerator techniques such as the conjugate gradient (CG) or the generalized minimal residual (GMRES) method, or, in the case of nonlinear problems, as preconditioners for the solution of the linear system arising from the use of Newton's method (these methods are usually referred to as NKS - *Newton/Krylov/Schwarz* - methods) or as preconditioners for solvers such as the nonlinear conjugate gradient method. A wide part of the research in the field is thus now devoted to the optimization of the interface conditions: this amounts to replace at the numerical level the original interface conditions with new ones which either are equivalent or at least imply the original ones. The construction of a preconditioner of Robin/Robin type for the primal Schur solution of an heterogeneous advection-diffusion problem, presented in this thesis, can be seen as belonging to this direction of research.

A growing interest in the field, finally, led in the last years many scientists to study the equations describing the propagation of waves. B Després proposed, for both Maxwell's system and Helmholtz equation, a non-overlapping additive Schwarz algorithm. For Maxwell's system, A. Alonso and A. Valli proposed a substructuring method for the time-harmonic eddy-current problem, while F. Ben-Belgacem, A. Buffa, Y. Maday and F. Rapetti proposed a three-field mortar method. In this thesis, a convergence analysis of a Schwarz algorithm of Després' type for the time-harmonic Maxwell system is presented via a Fourier analysis of the interface operator.

The present thesis is thus twofold, as it does not focus neither on a single problem (or a single kind of problems) nor on a single algorithm. In the framework of non-overlapping partitions, different types of algorithms are used: in Chapter 2 the hyperbolic system of Euler equations is solved by means of an iteration-by-subdomains method; in Chapter 3 an heterogeneous model of advection-diffusion is solved by means of a primal Schur method, with the domain decomposition algorithm as a preconditioner; in Chapter 4, the Helmholtz equation of acoustics and the Maxwell system of electromagnetism are solved by an additive Schwarz method. In the following a brief survey of each chapter is given.

In Chapter 2 the inviscid Euler system of equations is considered. In the first part of the chapter, following the book by A. Chorin and J.E. Mardsen, the equations governing the dynamics of an inviscid compressible gas are derived. Then, an iteration-by-subdomains procedure of Dirichlet/Dirichlet type is proposed in the case of a two-domains decomposition. The underlying ideas stem from the analysis made by A. Quarteroni in [86] and L. Gastaldi in [56] for linear hyperbolic systems with constant coefficients: the natural requirement on the interface is the continuity of the normal inviscid flux, which splits into three conditions of Dirichlet type. The iteration-by-subdomains algorithm is studied, in the region of smooth flow, for one dimensional isentropic flows. When the flow is smooth, the Euler system can be put in quasi-linear form, emphasizing the hyperbolic character of the equations, and, if isentropic, it can be put in diagonal form, as the characteristic variables can actually be determined. The interface continuity required is thus the one of the characteristic variables of the system, since it entails the continuity of the normal inviscid flux. The original contribution of the thesis is the convergence analysis of the iteration-by-subdomains algorithm in characteristic variables. The flow is assumed to be subsonic in order to actually have an iteration-by subdomains procedure: in this case, in fact, one eigenvalue is positive on the interface while the other one is negative, thus there can be exchange of information from each subdomain to the other one. If the flow is supersonic, both eigenvalues have the same sign and the whole information travels form one subdomain to the other one: this trivially reduces the iterative procedure to a sequential solution firstly in a subdomain, then in the other one. The main interest is set on the spatial decomposition, and to this aim the original problem is integrated in time by a semi-implicit Euler scheme, which linearizes the system by evaluating at the *n-th* time step the space derivative and evaluating at the $(n-1)$-*th* step the matrix of the coefficients. On one hand, this kind of time marching scheme keeps trace of the information due to the nonlinearity of the problem (differently from what happens when a *frozen coefficients* type linearization is used), and, on the other hand, allows to decouple the linearized system in two scalar equations which are coupled only through the boundary conditions. For the continuous problem, it is not possible to have a system of scalar equations coupled only through the boundary conditions, since the non-zero elements of the diagonal matrix in the characteristic problem are linear combinations of the characteristic variables themselves, thus providing a stronger coupling. With this position, the iteration-by-subdomains procedure for the time-discretized problem is shown to converge for any choice of the time step Δt: the interface mapping is proved to be a contraction with constant of order $e^{-C/\Delta t}$, with $C > 0$. A fully discrete version of the iteration-by-subdomains algorithm is successively presented, where the problem is discretized in space with finite elements, and is stabilized via a Streamline-Diffusion method. Some inflow-outflow type estimate are then given for this scheme in the single domain case, and used to prove that the discrete mapping on the interface generated by the scheme is a contraction, provided the entries of the stabilizing matrix are sufficiently small, but independently of the mesh parameter h. In this sense the result is optimal. Then, some standard error estimates for the Streamline-Diffusion Method are presented, and used to give an energy estimate in the single domain case for the error between the exact solution in primitive variables at time $t = t^n$, and the fully discrete solution at time step n: under some not so restrictive assumptions, it can be shown that the L^2 error can be controlled by the L^2 error between the exact solution in characteristic variables and its time discretized one , plus some terms which depend on the approximation error at the previous time step: assuming an uniform convergence of the fully discrete solution to the time discrete one, as $h \to 0$, at the previous time steps, these terms vanish uniformly in h, and the approximation error depends only on the time marching scheme. The complete one dimensional Euler system is then considered: in this case the characteristic variables cannot effectively be determined, thus their continuity cannot be used as interface matching condition. An algorithm is then proposed which enforces the continuity of the characteristic variables of the isentropic case and of the entropy (the only known characteristic variable of the complete system). When the boundary conditions allow to evaluate the entropy at the left endpoint of the interval (*e.g.* if the density and the pressure of the gas are given), the time discretized system can be reduced, at time step n, to the isentropic one with an additional forcing term depending on the entropy at the same time step. When the boundary conditions do not allow to evaluate the entropy at the left endpoint of the interval the system can be put into a diagonal form with a time marching scheme which keeps space derivatives of the same order at both time step n and $n-1$. This may generate instabilities, and it implies that at the discrete level a CFL type condition must be taken into account in order to be able to treat the problem. The chapter is then concluded by the presentation of some domain decomposition algorithms proposed in literature for three dimensional flows.

In Chapter 3 advection-diffusion problems with strongly dominant convective part are considered. In the first part of the chapter a review of substructuring methods, previously proposed in literature for such problems, is addressed. In the second part of the chapter is located the original part of the thesis on this subject, where a preconditioner of Robin/Robin type for a primal Schur method is proposed and analyzed for strongly heterogeneous problems. This work has been done in collaboration with P. Le Tallec and F. Nataf at the CMAP of the École Polytechnique in Paris. An advection-diffusion equation with discontinuous viscosity coefficients has been considered, whose theoretical justification comes from the modeling of transport and diffusion of a species through heterogeneous media, where the jumps in the viscosity are due to the different materials present in the computational domain. A generalized Robin/Robin preconditioner is proposed for the solution of the Steklov-Poincaré equation on the interface. The idea is to extend both the generalized Neumann/Neumann preconditioner, introduced in [75], which deals with heterogeneity in the coefficients, and the Robin/Robin preconditioner, introduced in [2], which is especially suited for non-symmetric problems, in order to obtain a preconditioner whose performance is not affected by the amplitude of the jumps in the coefficients. The Fourier analysis is the main tool used to study the effectiveness of the preconditioner in the special case of the plane decomposed into the left $\{x < 0\}$ and right $\{x > 0\}$ half planes. The preconditioner is firstly defined for a constant convective field perpendicular to the interface, then its robustness with respect to the direction of the convective field is analyzed. Since the original problem is non-symmetric, one can estimate the reduction factor of a GMRES algorithm (to be used at the discrete level) for the preconditioned Schur complement system: the result obtained is optimal since this reduction factor is bounded from above by a constant which is independent of both the coefficient of the problem and the mesh parameter h. This is very important since it allows to deal with very large discontinuities in the viscosity. Moreover, the formula obtained for the upper bound implies that the reduction factor improves with the enlargement of the jump in the viscosity. The generalization to a decomposition into an arbitrary number of subdomains is then addressed, by means of the variational formulation of the problem. The chapter is then concluded by the presentation of some numerical tests in 3D, which have been carried out in collaboration with M. Vidrascu at the INRIA in Rocquencourt (France). The numerical evidence is in agreement with the theory, as the proposed preconditioner showed fair insensitivity to the jumps in the coefficients and to the variations of the convective field: it remained sensitive to the number of subdomains, but this seems unavoidable for advection-dominated problems without the introduction of a coarse grid correction. The original results of this chapter can be found also in [57] and, in shorter form, in [58].

Finally, Chapter 4 is devoted to additive Schwarz algorithms for the solution of time-harmonic acoustic and electromagnetic equations. In the first part of the chapter, following the book by J.-C. Nédélec [83], the Helmholtz equation of acoustics and the Maxwell's system of electromagnetism are introduced. Then, a survey of Schwarz algorithms previously proposed in literature for Helmholtz equation is presented, stemming from the early works of B. Després ([35], [38]), who proposed an interface condition of Robin type, linked to a radiation condition at finite distance, up to the recent contribution of M. Gander, F. Magoulès and F. Nataf ([51]): the convergence properties of these algorithms are studied, by means of a Fourier analysis, in order to enlighten their possible drawbacks, and the way these drawbacks have been overtaken. The Schwarz algorithm can be interpreted as an iteration operator acting on the interface, and its reduction factor is defined as being the modulus of the symbol in the Fourier space of this iteration operator: since the Helmholtz equation is scalar, the same holds for the symbol considered. In particular, Després' algorithm is shown to converge only for propagative modes (*i.e.* low frequencies in the Fourier space), where the reduction factor is strictly less than 1, whereas for evanescent modes (*i.e.* high frequencies in the Fourier space) the reduction factor is exactly 1, implying no convergence at all. Such drawback can be overtaken either with a slight modification of the Robin interface condition, as done by M. Gander *et al.* in [51], or with the addition of a second order space derivative in the direction tangential to the interface, as done by P. Chevalier in his thesis ([30]), and again by M. Gander *et al.* in [51]. The last part of the chapter is the original contribution of the thesis on this subject. Moving from the additive Schwarz algorithm with Robin type interface conditions proposed by B. Desprès for the Maxwell's system, we introduce a slightly more general interface condition, where the zero-*th* order term is multiplied by a complex number Z, with non-zero real part, instead of a purely imaginary one. The convergence properties of the algorithm are analyzed, in the case of $\mathbf{R}^3$ partitioned in two half-spaces, by means of a Fourier transform. Differently from Helmholtz equation, the interface problem is no longer scalar, but, in the Fourier space, its symbol is a 2×2 matrix. The reduction factor is thus introduced as being the spectral radius of the iteration matrix, as it equals the infimum of all compatible matrix norms. We showed that there is no possible choice of the the parameter Z ensuring convergence for both evanescent and propagative modes. If Z is is real, the reduction factor is 1 for propagative modes and greater than 1 for evanescent ones. If Z is purely imaginary (and this is the case of B. Després' algorithm), the spectral radius of the iteration matrix is strictly less than 1 for propagative modes, independently of the choice of the parameter Z, and convergence is ensured, whereas in the case of evanescent modes the spectral radius of the

matrix is exactly 1, and the iterative mapping does not converge. Finally, if $Z = p + iq$, the algorithm shows again a reduction factor greater than 1 for evanescent modes. Thus, on one hand, the algorithm with Z purely imaginary suffers of the same drawback as the algorithm for Helmholtz equation: convergence is ensured only for propagative modes. Moreover, in the case of Maxwell's system, it is not enough to multiply the zero-*th* order term in the Robin interface condition by a complex number with non-zero real part, in order to achieve convergence also for evanescent modes, as done in the case of Helmholtz equation. An opportunity to overcome this drawback is then proposed, by means of the addition, to B. Després' interface condition, of the vectorial tangential Laplacian, which is defined, for any vector field $\mathbf{u}$ tangential to the interface, as $\Delta_\Gamma \mathbf{u} := \nabla_\Gamma \mathrm{div}_\Gamma \mathbf{u} - \overrightarrow{\mathrm{rot}}_\Gamma \mathrm{rot}_\Gamma \mathbf{u}$. Since a similar approach, in the case of Helmholtz equation, ensured convergence for all modes, and since the Maxwell's system can be seen as a vectorial Helmholtz problem, we could expect (or, at least, hope) that the same could occur also in this case. However, a convergence analysis for this last algorithm has not yet been performed in this thesis.
A brief summary of the results obtained is addressed at the end of each chapter.

Chapter 2

Domain Decomposition Methods for Compressible Flows

In this chapter we deal with the motion of ideal compressible fluids. In the first part of the chapter we derive the equations governing the dynamics of an inviscid ideal compressible gas. In Section 2.2 we propose an iteration-by-subdomains algorithm to solve Euler system. In Section 2.3, stemming from some results obtained by A. Quarteroni in [86] and by L. Gastaldi in [56] for a domain decomposition approach to linear hyperbolic systems with constant coefficients, a convergence analysis is addressed for one dimensional flows, for both the problem continuous in space and discretized in time (the main attention is focused on the spatial decomposition) and the fully discrete problem. The result is obtained in the region of smooth flow, where the quasi-linear form of the system is valid, and for isentropic flows, since, in this latter case, the characteristic variables can actually be determined, and their continuity can be used as matching condition on the interface. In the last part of the chapter, in Section 2.4 we propose an algorithm for the complete one dimensional system, and in Section 2.5 we report some algorithms proposed in literature for three dimensional flows.

2.1 The Euler Equations for Compressible Flows

In this section, following mainly the book by A. Chorin and J. Mardsen, we develop the basic equation of the mechanics of ideal compressible fluids. These equations are derived from the conservation laws of mass, momentum, and energy.

2.1.1 Derivation of the Equations

Let Ω be a region in $\mathbf{R}^d$, $(d = 2, 3)$, the two- or three-dimensional space, filled with a fluid. Let $\mathbf{u}(\mathbf{x}, t) = (u_1(\mathbf{x}, t), \ldots, u_d(\mathbf{x}, t))$ be a vector, depending on the space-time variable $(\mathbf{x}, t) = (x_1, \ldots, x_d, t)$, representing the velocity of a particle of fluid moving through $\mathbf{x}$ at time t. We call $\mathbf{u}(\mathbf{x}, t)$ the *spatial velocity of the fluid.*

We assume that for each time t the fluid has a well-defined mass density $\rho(\mathbf{x}, t)$, so that if D is any subregion of Ω, the mass of fluid in D at time t is given by

$$m(D, t) = \int_D \rho(\mathbf{x}, t)\, d\mathbf{x},$$

where $d\mathbf{x}$ is the volume element in the plane or in space. The assumption that ρ exists is a *continuum assumption*, which does not hold if the molecular structure of the fluid is taken into account. However, for most macroscopic phenomena occurring in nature, this assumption is believed to be extremely accurate.

The derivation of the equations is based on three basic principles:

1) *mass is neither created nor destroyed*

2) *the rate of change of momentum of a portion of the fluid equals the force applied on it* (Newton's second law)

3) *energy is neither created nor destroyed*

Conservation of mass

Let D be a fixed (in time) subregion of Ω. The rate of change of mass in D is

$$\frac{d}{dt} m(D, t) = \frac{d}{dt} \int_D \rho(\mathbf{x}, t)\, d\mathbf{x} = \int_D \frac{\partial \rho}{\partial t}(\mathbf{x}, t)\, d\mathbf{x}.$$

Let $\mathbf{n}$ denote the unit outward normal defined at points of ∂D, the boundary of D, assumed to be smooth and with area element $d\sigma$: the volume flow rate across ∂D per unit area is $\mathbf{u} \cdot \mathbf{n}$, while the mass flow rate per unit area is $\rho \mathbf{u} \cdot \mathbf{n}$.

The principle of conservation of mass states that the rate of *increase* of mass in D equals the rate at which mass is crossing ∂D in the *inward* direction. Thus, the integral form of the law of conservation of mass reads

$$\frac{d}{dt} \int_D \rho(\mathbf{x}, t)\, d\mathbf{x} = -\int_{\partial D} \rho \mathbf{u} \cdot \mathbf{n}\, d\sigma.$$

By the divergence theorem, the previous statement is equivalent to

$$\int_D \left[\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{u}) \right] d\mathbf{x} = 0,$$

and, since this latter statement holds for all $D \subset \Omega$, it is equivalent to

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{u}) = 0,$$

which is the differential form of the law of conservation of mass, also called *continuity equation.*

Balance of momentum

For any continuum, forces acting on a piece of it are divided into two types: the *stress forces*, where a piece of material is acted on by forces across its surface by the rest of the continuum, and the *body forces* (also called *external forces*, such as gravity or magnetic fields), which exert a force per unit volume on the continuum.

Definition. *A continuum is called an ideal fluid if for any motion of the fluid there is a function $p(\mathbf{x}, t)$ called "pressure", such that if S is a surface in the fluid, with a chosen unit normal $\mathbf{n}$, the force of stresses exerted across the surface per unit area, at $\mathbf{x} \in S$ and at time t, is given by $p(\mathbf{x}, t)\mathbf{n}$.*

Without entering into details of the hypotheses underlying this definition, we only observe that the absence of tangential forces means that there is no way for rotations to start or stop. Equivalently, if $\operatorname{rot} \mathbf{u} = 0$ at time $t = 0$, then it must be identically zero for every time.
If D is a region in the fluid, the total force exerted, at time t, on the fluid inside D by the stresses on its boundary is

$$S_{\partial D} = \{\text{force on } D\} = -\int_{\partial D} p\mathbf{n}\, d\sigma,$$

with negative sign because $\mathbf{n}$ points outwards. The divergence theorem provides

$$S_{\partial D} = -\int_D \nabla p \, d\mathbf{x}.$$

Denoting with $\mathbf{b}(\mathbf{x}, t)$ the given body force per unit mass, the total body force is

$$\mathbf{B} = \int_D \rho \mathbf{u} \, d\mathbf{x}.$$

By Newton's second law (force =mass × acceleration) we are led to the differential form of the law of balance of momentum:

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p + \rho \mathbf{b},$$

where $\frac{D}{Dt} = \partial_t + \mathbf{u} \cdot \nabla$ is called the *material derivative*, since it takes into account the fact that the fluid is moving with velocity $\mathbf{u}$.

Conservation of energy

We have developed $d + 1$ equations for the $d + 2$ unknowns ρ, p, and $\mathbf{u}$, because the equation for $D\mathbf{u}/Dt$ is a vector equation consisting of d scalar equations. We therefore need one more equation to avoid an under-determined problem.
For a fluid moving in a domain Ω, with velocity field $\mathbf{u}$, the kinetic energy contained in a region $D \subset \Omega$ is

$$E_{\text{kinetic}} = \frac{1}{2} \int_D \rho \|\mathbf{u}\|^2, d\mathbf{x},$$

with $\|\mathbf{u}\|^2 = (u_1^2 + \ldots + u_d^2)$. We assume that the total energy of the fluid can be written as

$$E_{\text{total}} = E_{\text{internal}} + E_{\text{kinetic}},$$

namely the sum of the kinetic energy and the internal thermodynamic energy, which derives from sources such as intermolecular potentials and internal molecular vibrations. If energy is pulled into the system or we allow the fluid to do work, the amount of E_{total} will change.
For ideal gases, the internal energy is given by

$$\epsilon = \frac{p}{\rho}\left(\frac{1}{\gamma - 1}\right),$$

where the constant $\gamma > 1$ is the ratio between the specific heat at constant volume and the specific heat at constant pressure, thus the total energy per unit volume is

$$e = \frac{1}{2}\rho\|\mathbf{u}\|^2 + \rho\epsilon.$$

Assuming that no heat enters the fluid domain from its boundaries, the only variations in the total energy are induced when the fluid does work. The work done by a fluid volume D per unit time is given by $-\int_{\partial D} p\mathbf{u}\cdot\mathbf{n}\,d\sigma$, and it must equal the rate of change of the total energy in D. The divergence theorem provides the integral form of the conservation of energy:

$$\frac{\partial}{\partial t}\int_D e\,d\mathbf{x} = -\int_D \operatorname{div}(p\mathbf{u})\,d\mathbf{x}.$$

By the transport theorem (see [31], p. 9), this integral formulation is equivalent to the differential equation

$$\frac{\partial e}{\partial t} + \operatorname{div}\left[(e+p)\mathbf{u}\right] = 0,$$

which is called the *first law of thermodynamics.*
The Euler equations for an ideal compressible fluid are thus

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \operatorname{div}(\rho\mathbf{u}) = 0 \\[2ex] \rho\dfrac{D\mathbf{u}}{Dt} = -\nabla p + \rho\mathbf{b} \\[2ex] \dfrac{\partial e}{\partial t} + \operatorname{div}\left[(e+p)\mathbf{u}\right] = 0 \end{cases}$$

On the other hand, if we assume that the whole energy is kinetic, and that the rate of change of kinetic energy in a portion of fluid equals the rate at which the pressure and body forces do work, *i.e.*

$$\frac{d}{dt}E_{\text{kinetic}} = -\int_{\partial D_t} p\mathbf{n}\,d\sigma + \int_{D_t} \rho\mathbf{u}\cdot\mathbf{b}\,d\mathbf{x},$$

the application of divergence theorem and the previous formulas entail necessarily $\operatorname{div}\mathbf{u} = 0$. This means that if we assume $E_{\text{total}} = E_{\text{kinetic}}$, then the fluid must be incompressible. The Euler equations for incompressible flows are thus

$$\begin{cases} \dfrac{D\rho}{Dt} = 0 \\ \rho\dfrac{D\mathbf{u}}{Dt} = -\nabla p + \rho\mathbf{b} \\ \operatorname{div}\mathbf{u} = 0 \end{cases}$$

considered together with the boundary condition $\mathbf{u} \cdot \mathbf{n} = 0$.

2.1.2 Isentropic Fluids

A flow is called "*isentropic*" if there exists a function w called "*enthalpy*" such that

$$\nabla w = \frac{1}{\rho}\nabla p.$$

This terminology arises in thermodynamics, and without even entering here a detailed discussion of thermodynamic concepts, we make a few general comments.
The basic quantities appearing in thermodynamics, each of them a function of $\mathbf{x}$ and t depending on the given flow, are the pressure p, the density ρ, the temperature ϑ, the entropy s, the enthalpy (per unit mass) w and the internal energy per unit mass ϵ, which is given by

$$\epsilon = w - \frac{p}{\rho}.$$

These quantities are related by the *First Law of Thermodynamics*, commonly accepted as a basic principle

$$dw = \vartheta ds + \frac{1}{\rho}dp \tag{TD1}$$

which is a statement of conservation of energy: it can be equivalently expressed as

$$d\epsilon = \vartheta ds + \frac{p}{\rho^2}d\rho. \tag{TD2}$$

When the pressure is a function of the density ρ only, the flow is clearly isentropic, as

$$w = \int^{\rho} \frac{p'(\lambda)}{\lambda}\,d\lambda.$$

As a consequence the internal energy satisfies $d\epsilon = (p\,d\rho)/\rho^2$, or, equivalently

$$\epsilon = \int^{\rho} \frac{p(\lambda)}{\lambda^2}\,d\lambda.$$

For isentropic flows with $p = p(\rho)$, the integral form of conservation of energy states that the rate of change of energy in a portion of fluid equals the rate at which work is done in it:

$$\frac{d}{dt}E_{\text{total}} = \frac{d}{dt}\int_{W_t}\left(\frac{1}{2}\rho\|\mathbf{u}\|^2 + \rho\epsilon\right)dV = \int_{W_t}\rho\mathbf{u}\cdot\mathbf{b}\,dV - \int_{\partial W_t}p\mathbf{u}\cdot\mathbf{n}\,dA$$

Thus, Euler equations for isentropic flows, with $p = p(\rho)$, in a domain Ω are

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{u}) = 0 \\ \\ \rho \dfrac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla w + \rho \mathbf{b} \end{cases}$$

in Ω, and

$$\mathbf{u} \cdot \mathbf{n} = 0$$

on $\partial\Omega$ (or $\mathbf{u} \cdot \mathbf{n} = \mathbf{V} \cdot \mathbf{n}$ if $\partial\Omega$ is moving with velocity $\mathbf{V}$). In general, these equations lead to a well-posed initial value problem only if $p'(\rho) > 0$: this agrees with the common experience that the increase of the surrounding pressure on a volume of fluid causes a decrease in the occupied volume and thus an increase in density.
When dealing with ideal gas dynamics, the isentropic assumption is

$$p = K\rho^{\gamma},$$

where K and γ (the ratio of specific heats) are constants and $\gamma \geq 1$. Hence the enthalpy is given by

$$w = \int^{\rho} \frac{\gamma K s^{\gamma-1}}{s} \, ds = \frac{\gamma K \rho^{\gamma-1}}{\gamma - 1}$$

whereas the internal energy is

$$\epsilon = \frac{K\rho^{\gamma-1}}{\gamma - 1}.$$

2.1.3 The Navier-Stokes Equations for compressible fluids

In none of the cases of the previous section, the possibility of energy dissipation due to friction is taken into account: the viscous effects are neglected and we assume that the fluid is inviscid. When these effects are considered, we end up with the system of *Navier-Stokes* equations, which, in conservative form, reads (see [71]):

$$\frac{\partial \mathbf{W}}{\partial t} + \mathbf{div}\, \mathbf{F}(\mathbf{W}) = \mathbf{div}\, \mathbf{G}(\mathbf{W}) \tag{2.1.1}$$

in $\Omega \times (0, T)$, where $\mathbf{W}$ is the vector of conserved variables $\mathbf{W} = (\rho, \rho\mathbf{u}, \rho E)$, where we have simply indicated with E the total energy per unit mass, and where the convective and diffusive terms $\mathbf{F}(\mathbf{W})$ and $\mathbf{G}(\mathbf{W})$ are given by

$$\mathbf{F}(\mathbf{W}) = \begin{pmatrix} \rho\mathbf{u} \\ \rho\mathbf{u} \otimes \mathbf{u} + p\mathbf{I} \\ (\rho E + p)\mathbf{u} \end{pmatrix} \qquad \mathbf{G}(\mathbf{W}) = \begin{pmatrix} 0 \\ \tau \\ \tau \cdot \mathbf{u} - \mathbf{q} \end{pmatrix}.$$

Here $\mathbf{u} \otimes \mathbf{u}$ is the tensor whose components are $u_i u_j$, $\mathbf{I}$ is the unit tensor δ_{ij}, $\mathbf{q}$ is the heat flux, which is related to the absolute temperature by the standard Fourier law

$$\mathbf{q} = -\kappa \nabla \vartheta,$$

where $\kappa > 0$ is the heat conductivity coefficient, and finally $(\tau \cdot \mathbf{u})_i := \sum_j \tau_{ij} u_j$, where τ is the viscous stress tensor, whose components are defined as

$$\tau_{ij} := \mu(D_i u_j + D_j u_i) + \left(\zeta - \frac{2\mu}{d}\right) \operatorname{div} \mathbf{u}\, \delta_{ij},$$

with $\mu > 0$ and $\zeta \geq 0$ being the shear and bulk viscosity coefficients, respectively.
In system (2.1.1), the divergence of $\mathbf{F}(\mathbf{W})$ (and similarly for $\mathbf{G}(\mathbf{W})$) is the $(d+2)$-vector

$$\mathbf{div}\, \mathbf{F}(\mathbf{W}) = (\operatorname{div}(\rho \mathbf{u}), \operatorname{div}(\rho \mathbf{u} \otimes \mathbf{u} + p\mathbf{I}), \operatorname{div}\ [(\rho E + p)\mathbf{u}])$$

while the divergence of a tensor $\mathbf{T}$ is the vector with components

$$(\operatorname{div} \mathbf{T})_j := \sum_i D_i T_{ji}.$$

As the continuity equation for the density ρ is hyperbolic, equations (2.1.1) provide an incomplete parabolic system. It is not difficult to see that dropping all the diffusive terms (that is taking $\mu = \zeta = \kappa = 0$ or, equivalently $\tau = \mathbf{0}$ and $\mathbf{q} = \mathbf{0}$) leads back to the conservative formulation of Euler system.

2.2 Multidomain formulation of the Euler Equation

The formulation of the Euler system in terms of the conserved variables $\mathbf{W} = (\rho, \rho\mathbf{u}, \rho E)$ reads

$$\frac{\partial \mathbf{W}}{\partial t} + \mathbf{div}\, \mathbf{F}(\mathbf{W}) = \mathbf{0}. \quad \text{in } \Omega \times (0,T), \tag{2.2.1}$$

where $\Omega \subset \mathbf{R}^d$, $d = 2, 3$. Let the domain Ω be decomposed into two non-overlapping subdomains Ω_1 and Ω_2. If we denote with Γ the interface between Ω_1 and Ω_2, and with $\mathbf{W}_i$ the restriction of $\mathbf{W}$ on Ω_i, $= 1, 2$, equations (2.2.1) can be reformulated as

$$\begin{cases} \dfrac{\partial \mathbf{W}_i}{\partial t} + \mathbf{div}\, \mathbf{F}(\mathbf{W}_i) = \mathbf{0} & \text{in } \Omega_i \times (0,T) \\[2ex] \mathbf{F}(\mathbf{W}_1) \cdot \mathbf{n} = \mathbf{F}(\mathbf{W}_2) \cdot \mathbf{n} & \text{on } \Gamma \times (0,T). \end{cases} \tag{2.2.2}$$

for $i = 1, 2$, where $\mathbf{n}$ is the unit normal vector on Γ directed from Ω_1 to Ω_2, and where $\mathbf{F}(\mathbf{W}) \cdot \mathbf{n}$ is the $(d+2)$-vector

$$(\rho \mathbf{u} \cdot \mathbf{n}, \rho \mathbf{u}(\mathbf{u} \cdot \mathbf{n}) + p\mathbf{n}, (\rho E + p)\mathbf{u} \cdot \mathbf{n}).$$

In other words, the subdomain restrictions $\mathbf{W}_1$ and $\mathbf{W}_2$ satisfy the Euler equations in Ω_1 and Ω_2 separately, where they inherit the boundary and initial conditions prescribed for $\mathbf{W}$ on $\partial\Omega$ and at $t = 0$, together with suitable interface conditions. The interface equation $(2.2.2)_2$ prescribes the continuity across Γ of the normal inviscid flux, and this is a natural consequence of the fact that the variable $\mathbf{W}$ is a distributional solution of (2.2.1) in Ω. Condition $(2.2.2)_2$ can be rewritten as

$$\begin{aligned} \rho_1 \mathbf{u}_1 \cdot \mathbf{n} &= \rho_2 \mathbf{u}_2 \cdot \mathbf{n} \\ \rho_1 u_{1,j} \mathbf{u}_1 \cdot \mathbf{n} + p_1 n_j &= \rho_2 u_{2,j} \mathbf{u}_2 \cdot \mathbf{n} + p_2 n_j \quad j = 1, \ldots, d \\ (\rho_1 E_1 + p_1) \mathbf{u}_1 \cdot \mathbf{n} &= (\rho_2 E_2 + p_2) \mathbf{u}_2 \cdot \mathbf{n} \end{aligned} \tag{2.2.3}$$

Since $\mathbf{u}$ is continuous across Γ, condition $(2.2.3)_1$ is equivalent to

$$\rho_1 = \rho_2 \qquad \forall\, (\mathbf{x}, t) \in \Gamma \times (0, T) \text{ such that } \mathbf{u} \cdot \mathbf{n} \neq 0,$$

which is in agreement with the physics of compressible fluid flows, that allow two kind of discontinuities: shock waves and contact discontinuities. We won't enter here the details of this topic, but we refer the interested reader to [60], [66], [96] or [31] for an exhaustive treatment of this subject. We only recall that in case of contact discontinuities the normal velocity is zero, the pressure is continuous, but density, tangential velocity and temperature may have non-zero jumps.

Within the frame of iterative substructuring methods, the interface conditions (2.2.3) has to be split into Dirichlet conditions for $\mathbf{W}_1$ on $\Gamma \cap \{\mathbf{u} \cdot \mathbf{n} < 0\}$ and Dirichlet conditions for $\mathbf{W}_2$ on $\Gamma \cap \{\mathbf{u} \cdot \mathbf{n} > 0\}$. Thus, an iteration-by-subdomains approach would read:

Given $\mathbf{W}_1^0$ and $\mathbf{W}_2^0$, solve for $k \geq 1$

$$\begin{cases} \dfrac{\partial \mathbf{W}_1^{k+1}}{\partial t} + \mathbf{div}\, \mathbf{F}(\mathbf{W}_1^{k+1}) = \mathbf{0} & \text{in } \Omega_1 \times (0, T) \\ \mathbf{F}(\mathbf{W}_1^{k+1}) \cdot \mathbf{n} = \mathbf{F}(\mathbf{W}_2^{k}) \cdot \mathbf{n} & \text{on } (\Gamma \cap \{\mathbf{u} \cdot \mathbf{n} < 0\}) \times (0, T), \end{cases}$$

and

$$\begin{cases} \dfrac{\partial \mathbf{W}_2^{k+1}}{\partial t} + \mathbf{div}\, \mathbf{F}(\mathbf{W}_2^{k+1}) = \mathbf{0} & \text{in } \Omega_2 \times (0, T) \\ \mathbf{F}(\mathbf{W}_2^{k+1}) \cdot \mathbf{n} = \mathbf{F}(\mathbf{W}_1^{k}) \cdot \mathbf{n} & \text{on } (\Gamma \cap \{\mathbf{u} \cdot \mathbf{n} > 0\}) \times (0, T). \end{cases}$$

If the interface Γ moves with the time, $\Gamma = \Gamma(t)$, we denote with $\sigma(t)$ the velocity at time t along the normal direction $\mathbf{n} = \mathbf{n}(t)$, and the matching condition $(2.2.2)_2$ must be replaced by

$$[\mathbf{W}_1(t) - \mathbf{W}_2(t)]\, \sigma(t) = [\mathbf{F}(\mathbf{W}_1(t)) - \mathbf{F}(\mathbf{W}_2(t))] \cdot \mathbf{n}(t) \qquad \text{on } \Gamma(t). \tag{2.2.4}$$

In particular, if $\Gamma(t)$ coincides with (or simply intercepts) a shock front $\delta(t)$, then equation (2.2.4) can be easily recognized as the Rankine-Hugoniot jump condition across the shock front (again, see [31] or [60]).

In the following sections, we carry out a convergence analysis for an iteration-by-subdomains algorithm for one-dimensional flows, since they allow to enlighten the role of interface conditions in substructuring methods.

2.3 The 1-D Isentropic Euler Equation

We consider here an inviscid isentropic compressible fluid in one space dimension: the vector of conserved variables is $\mathbf{W} = (\rho, \rho u)$ and the flux vector $\mathbf{F}$ is given by

$$\mathbf{F}(\mathbf{W}) = \begin{pmatrix} \rho u \\ \rho u^2 + p \end{pmatrix}.$$

Thus the conservative form of the equation reads

$$\frac{\partial \mathbf{W}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{W})}{\partial x} = 0, \quad \text{in } Q_T := \Omega \times (0,T) \tag{2.3.1}$$

where $\Omega = (a,b) \subset \mathbf{R}$ is an interval.
In a region of smooth flow, using the Jacobian of the flux $\mathbf{F}(\mathbf{W})$, system (2.3.1) can be put in form of quasi-linear hyperbolic system

$$\frac{\partial \mathbf{U}}{\partial t} + A(\mathbf{U})\frac{\partial \mathbf{U}}{\partial x} = 0 \quad \text{in } Q_T := \Omega \times (0,T) \tag{2.3.2}$$

where $\mathbf{U} = (\rho, u) : \Omega \times (0,T) \to \mathbf{R}^2$ is the vector of physical unknowns, also called "*primitive variables*", whereas

$$A(\mathbf{U}) := \begin{pmatrix} u & \rho \\ c^2/\rho & u \end{pmatrix},$$

where $c = \sqrt{K\gamma\rho^{\gamma-1}}$ is the *speed of sound*, $K > 0$ being a suitable constant, and γ being the ratio of specific heats. The matrix A is diagonalizable with distinct real eigenvalues (thus system (2.3.2) is strictly hyperbolic), namely $A = L\Lambda L^{-1}$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$, with

$$\lambda_1 = u + c, \quad \lambda_2 = u - c,$$

while L is the matrix of left eigenvectors, given by

$$L := \begin{pmatrix} c/\rho & 1 \\ -c/\rho & 1 \end{pmatrix}.$$

From a mathematical point of view, equation (2.3.2) has to be considered together with an initial condition $\mathbf{U}_0(x) = \mathbf{U}(x,0)$ and with suitable boundary conditions in order to have a well-posed initial-boundary value problem. Without entering the details of well-posedeness, we simply recall that it is not admissible to assign values on the outgoing components, since they could contradict the effect of the initial condition making it impossible for a solution to exist (for an extensive discussion on boundary conditions for hyperbolic problems, see for instance [70] and [84]). Among the various set of boundary conditions that render this problem well posed, we consider the following ones

$$\begin{cases} \rho(a,t) = g_1(t) & t \in (0,T) \\ \rho(b,t) = g_2(t) & t \in (0,T), \end{cases} \tag{2.3.3}$$

namely, we assign the value of the density, or, equivalently, the value of the speed of sound. The same result we are going to present in the following could be obtained also with different choices of suitable boundary conditions, for instance assigning the velocity on the left endpoint of the interval, $u(a,t) = b_1(t)$, and the density on the right end one $\rho(b,t) = b_2(t)$.
We require the initial value $\mathbf{U}_0(x) = \mathbf{U}(x,0)$ to be a continuous vector function, with first component attaining the values $g_1(0)$ and $g_2(0)$ at the endpoints of the interval, hence the solution of our problem is continuous for the whole time of smooth flow. Moreover, we assume the solution $\mathbf{U}(t,x)$ to be bounded for the whole time of smooth flow. Concerning this assumption, we recall that the Euler system develops shocks in a finite time, even in the presence of regular initial data. So far, equation (2.3.2) fails, and one must use a weak formulation based on the conservative form of the equation (2.3.1).
With an iteration-by-subdomain approach in sight, we finally assume that the flow is subsonic, i.e. $0 < u < c$, so that $\lambda_1 > 0$ and $\lambda_2 < 0$ for each $(x,t) \in Q_T$, which amounts to have information traveling from each subdomain to the other one. In fact, if the flow is supersonic, both eigenvalues are positive, the whole information is a traveling wave from Ω_1 to Ω_2, and the domain decomposition approach is trivially reduced to the sequential solution firstly in Ω_1 and then in Ω_2.
The nonlinearity of the problem does not allow to define directly the characteristic variables $\mathbf{V}$: we therefore introduce them by means of the following differential form (see [62], as well as [89])

$$d\mathbf{V} := Ld\mathbf{U} = (\frac{c}{\rho}d\rho + du, -\frac{c}{\rho} + du). \tag{2.3.4}$$

Hence, a direct integration provides

$$\begin{aligned} \mathbf{V}_1 &= u + \int \frac{c}{\rho}\, d\rho = u + \int \sqrt{K\gamma}\rho^{\gamma/2-3/2}\, d\rho \\ &= u + \frac{2}{\gamma-1}c + \text{const}, \end{aligned}$$

and similarly for $\mathbf{V}_2$, so that we have in conclusion

$$\mathbf{V} = (R_+, R_-),\ R_\pm := u \pm \frac{2}{\gamma-1}c, \tag{2.3.5}$$

namely, $\mathbf{V}_1 = R_+$ and $\mathbf{V}_2 = R_-$ are the Riemann invariants which are constant along the *characteristic lines* $C_\pm = \{(x(t),t) \mid x'(t) = u \pm c\}$. Problem (2.3.2) can therefore be decoupled into its characteristic formulation

$$\begin{cases} \dfrac{\partial \mathbf{V}}{\partial t} + \Lambda(\mathbf{V})\dfrac{\partial \mathbf{V}}{\partial x} = 0 & \text{in } Q_T := \Omega \times (0,T) \\ \mathbf{V}_1(a,t) - \mathbf{V}_2(a,t) = \phi_1(t) & t \in (0,T) \\ \mathbf{V}_1(b,t) - \mathbf{V}_2(b,t) = \phi_2(t) & t \in (0,T) \end{cases} \tag{2.3.6}$$

where we have set

$$\phi_1(t) = \frac{4}{\gamma - 1}\sqrt{K\ (g_1(t))^\gamma} \qquad \text{and} \qquad \phi_2(t) = \frac{4}{\gamma - 1}\sqrt{K\ (g_2(t))^\gamma}.$$

More, we can observe from (2.3.5) that the eigenvalues λ_1 and λ_2 can be expressed as linear combinations of the characteristic variables

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \frac{1}{4}\begin{pmatrix} 1+\gamma & 3-\gamma \\ 3-\gamma & 1+\gamma \end{pmatrix}\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix}. \tag{2.3.7}$$

Denoting with C the matrix in (2.3.7), system (2.3.6) can therefore be rewritten as

$$\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix}_t + \begin{pmatrix} \sum_{j=1}^{2} C_{1j}\mathbf{V}_j & 0 \\ 0 & \sum_{j=1}^{2} C_{2j}\mathbf{V}_j \end{pmatrix} \cdot \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix}_x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{2.3.8}$$

Remark 2.3.1 From (2.3.7), we observe that, differently from the case of constant coefficients, system (2.3.8) is not constituted of two independent scalar equations coupled only through the boundary conditions, and this is a consequence of the nonlinearity of the original problem. □

2.3.1 An Iteration-by-subdomain algorithm for the time-discretized problem

Since the Riemann invariants are constant along the characteristics, we can introduce a domain decomposition of the spatial domain Ω, where we enforce the continuity of the characteristic variables on the interface. Notice that the continuity of these latter variables guarantees the continuity of the physical ones. In that order, let $\alpha \in (a, b)$ and set $\Omega_1 := (a, \alpha)$, as well as $\Omega_2 := (\alpha, b)$. So far, we can consider the decomposed problem

$$\begin{cases} \dfrac{\partial \mathbf{U}^1}{\partial t} + A(\mathbf{U}^1)\dfrac{\partial \mathbf{U}^1}{\partial x} = 0 & \text{in } \Omega_1 \times (0,T) \\ \dfrac{\partial \mathbf{U}^2}{\partial t} + A(\mathbf{U}^2)\dfrac{\partial \mathbf{U}^2}{\partial x} = 0 & \text{in } \Omega_2 \times (0,T) \\ \rho(a,t) = g_1(t) & \forall t \in (0,T) \\ \rho(b,t) = g_2(t) & \forall t \in (0,T) \\ \mathbf{V}_1^1(\alpha,t) = \mathbf{V}_1^2(\alpha,t) & \forall t \in (0,T) \\ \mathbf{V}_2^1(\alpha,t) = \mathbf{V}_2^2(\alpha,t) & \forall t \in (0,T), \end{cases} \tag{2.3.9}$$

that can be decoupled, owing to (2.3.4), into its characteristic form,

$$\begin{cases} \dfrac{\partial \mathbf{V}^1}{\partial t} + \Lambda(\mathbf{V}^1)\dfrac{\partial \mathbf{V}^1}{\partial x} = 0 & \text{in } \Omega_1 \times (0,T) \\[2ex] \dfrac{\partial \mathbf{V}^2}{\partial t} + \Lambda(\mathbf{V}^2)\dfrac{\partial \mathbf{V}^2}{\partial x} = 0 & \text{in } \Omega_2 \times (0,T) \\[2ex] \mathbf{V}_1(a,t) - \mathbf{V}_2(a,t) = \phi_1(t) & t \in (0,T) \\[2ex] \mathbf{V}_1(b,t) - \mathbf{V}_2(b,t) = \phi_2(t) & t \in (0,T) \\[2ex] \mathbf{V}_1^1(\alpha,t) = \mathbf{V}_1^2(\alpha,t) & \forall t \in (0,T) \\[2ex] \mathbf{V}_2^1(\alpha,t) = \mathbf{V}_2^2(\alpha,t) & \forall t \in (0,T). \end{cases} \tag{2.3.10}$$

We are mainly interested in a spatial decomposition, thus, owing to (2.3.8), we advance in time the decomposed problem (2.3.10) by means of a semi-implicit method: this can be interpreted as a linearisation of system (2.3.8), leading to the following two systems of ordinary differential equations

$$\begin{cases} \beta \mathbf{V}^{1,n+1} + \Lambda(\mathbf{V}^{1,n})\dfrac{d}{dx}\mathbf{V}^{1,n+1} = \beta \mathbf{V}^{1,n} & \text{in } \Omega_1 \\[2ex] \mathbf{V}_1^{1,n+1}(a) - \mathbf{V}_2^{1,n+1}(a) = \phi_1(t^{n+1}) \end{cases}$$

and

$$\begin{cases} \beta \mathbf{V}^{2,n+1} + \Lambda(\mathbf{V}^{2,n})\dfrac{d}{dx}\mathbf{V}^{2,n+1} = \beta \mathbf{V}^{2,n} & \text{in } \Omega_2 \\[2ex] \mathbf{V}_1^{2,n+1}(b) - \mathbf{V}_2^{2,n+1}(b) = \phi_2(t^{n+1}) \end{cases}$$

where $\beta = 1/\Delta t$ is the inverse of the time step, which are coupled only through the interface conditions

$$\begin{cases} \mathbf{V}_1^{2,n+1}(\alpha) = \mathbf{V}_1^{1,n+1}(\alpha) \\[2ex] \mathbf{V}_2^{2,n+1}(\alpha) = \mathbf{V}_2^{1,n+1}(\alpha). \end{cases}$$

At each time step an iterative procedure can be introduced to solve the coupled system. From now on, since we are not dealing with time, we drop any index referring to time discretisation, and we set $f^{(1)} := \beta \mathbf{V}^{1,n}$, $f^{(2)} := \beta \mathbf{V}^{2,n}$, $\phi_1 := \phi_1(t^{n+1})$, $\phi_2 := \phi_2(t^{n+1})$, as well as

$$\Lambda := \begin{cases} \Lambda(\mathbf{V}^{1,n}) & \text{in } \Omega_1 \\[2ex] \Lambda(\mathbf{V}^{2,n}) & \text{in } \Omega_2 \end{cases}$$

The iteration-by-subdomain procedure can therefore be written, for $k \geq 0$, as

$$\begin{cases} \beta \mathbf{V}^{1,k+1} + \Lambda \dfrac{d}{dx} \mathbf{V}^{1,k+1} = f^{(1)} \quad \text{in } \Omega_1 \\[2ex] \mathbf{V}_1^{1,k+1}(a) - \mathbf{V}_2^{1,k+1}(a) = \phi_1 \\[2ex] \mathbf{V}_2^{1,k+1}(\alpha) = \mathbf{V}_2^{2,k}(\alpha) \end{cases} \tag{2.3.11}$$

$$\begin{cases} \beta \mathbf{V}^{2,k+1} + \Lambda \dfrac{d}{dx} \mathbf{V}^{2,k+1} = f^{(2)} \quad \text{in } \Omega_2 \\[2ex] \mathbf{V}_1^{2,k+1}(\alpha) = \mathbf{V}_1^{1,k}(\alpha) \\[2ex] \mathbf{V}_1^{2,k+1}(b) - \mathbf{V}_2^{2,k+1}(b) = \phi_2, \end{cases} \tag{2.3.12}$$

having chosen any initial guess $\mathbf{V}_2^{1,0}(\alpha) \in \mathbf{R}$ and $\mathbf{V}_1^{2,0}(\alpha) \in \mathbf{R}$.

Convergence Analysis of the Iteration-by-Subdomain Method

In order to prove the convergence of the iterative algorithm, following what is done by A. Quarteroni in [86] for a spectral collocation method and by L. Gastaldi in [56], both in the case of constant coefficients, we define, for each subdomain, the error vector (in characteristic form) as

$$\mathbf{E}_1^{i,k+1} := \mathbf{V}_1^{i,k+1} - \mathbf{V}_1^i, \qquad \mathbf{E}_2^{i,k+1} := \mathbf{V}_2^{i,k+1} - \mathbf{V}_2^i, \tag{2.3.13}$$

for $i = 1, 2$.

It can be easily viewed that the vector functions $\mathbf{E}^{i,k+1} := (\mathbf{E}_1^{i,k+1}, \mathbf{E}_2^{i,k+1})$, $i = 1, 2$ satisfy the following error equations

$$\begin{cases} \beta \mathbf{E}_1^{1,k+1} + \lambda_1 \dfrac{d}{dx} \mathbf{E}_1^{1,k+1} = 0 \quad \text{in } \Omega_1 \\[2ex] \beta \mathbf{E}_2^{1,k+1} + \lambda_2 \dfrac{d}{dx} \mathbf{E}_2^{1,k+1} = 0 \quad \text{in } \Omega_1 \\[2ex] \mathbf{E}_1^{1,k+1}(a) = \mathbf{E}_2^{1,k+1}(a) \\[2ex] \mathbf{E}_2^{1,k+1}(\alpha) = \mathbf{E}_2^{2,k}(\alpha), \end{cases} \tag{2.3.14}$$

as well as

$$\begin{cases} \beta \mathbf{E}_1^{2,k+1} + \lambda_1 \dfrac{d}{dx} \mathbf{E}_1^{2,k+1} = 0 \quad \text{in } \Omega_2 \\[2ex] \beta \mathbf{E}_2^{2,k+1} + \lambda_2 \dfrac{d}{dx} \mathbf{E}_2^{2,k+1} = 0 \quad \text{in } \Omega_2 \\[2ex] \mathbf{E}_1^{2,k+1}(\alpha) = \mathbf{E}_1^{1,k}(\alpha) \\[2ex] \mathbf{E}_2^{2,k+1}(b) = \mathbf{E}_1^{2,k+1}(b). \end{cases} \tag{2.3.15}$$

Now, let us consider what happens within Ω_1. Since the system is completely decoupled, we can have an explicit representation of the solution of the error equation. We get from $(2.3.14)_2$-$(2.3.14)_4$:

$$\mathbf{E}_2^{1,k+1}(x) = \mathbf{E}_2^{2,k}(\alpha) \exp\{ -\beta \Psi^{(1)}(x)\}, \tag{2.3.16}$$

with

$$\Psi^{(1)}(x) := \int_\alpha^x \frac{dy}{\lambda_2(y)}.$$

Similarly, we get from $(2.3.14)_1$-$(2.3.14)_3$:

$$\mathbf{E}_1^{1,k+1}(x) = \mathbf{E}_2^{1,k+1}(a) \exp\{ -\beta \Phi^{(1)}(x)\}, \tag{2.3.17}$$

with

$$\Phi^{(1)}(x) := \int_a^x \frac{dy}{\lambda_1(y)}.$$

Then, we get from (2.3.16) and (2.3.17):

$$\mathbf{E}_1^{1,k+1}(\alpha) = \exp\Big\{ -\beta\Big(\Phi^{(1)}(\alpha) + \Psi^{(1)}(a)\Big)\Big\} \mathbf{E}_2^{2,k}(\alpha). \tag{2.3.18}$$

Notice that

$$\Psi^{(1)}(a) = \int_\alpha^a \frac{dy}{\lambda_2(y)} = -\int_a^\alpha \frac{dy}{\lambda_2(y)} > 0,$$

since $\lambda_2(x) < 0$ for all $x \in \Omega_1$. Analogously, $\lambda_1(x) > 0$ for all $x \in \Omega_1$, provides $\Phi^{(1)}(\alpha) > 0$. We can therefore state the following Lemma.

Lemma 2.3.1 *The solution of problem* (2.3.14) *satisfies:*

$$\mathbf{E}_1^{1,k+1}(\alpha) = \sigma_1 \mathbf{E}_2^{2,k}(\alpha) \tag{2.3.19}$$

with $\sigma_1 < 1$.

Proof. Setting

$$\sigma_1 := \exp\Big\{ -\beta\Big(\Phi^{(1)}(\alpha) + \Psi^{(1)}(a)\Big)\Big\}$$

(2.3.19) follows immediately from (2.3.18). □

A similar argument within Ω_2 provides

$$\mathbf{E}_2^{2,k+1}(\alpha) = \exp\Big\{ -\beta\Big(\Phi^{(2)}(b) + \Psi^{(2)}(\alpha)\Big)\Big\}\, \mathbf{E}_1^{1,k}(\alpha), \tag{2.3.20}$$

where we have

$$\Phi^{(2)}(b) := \int_\alpha^b \frac{dy}{\lambda_1(y)}, \quad \Psi^{(2)}(\alpha) := \int_b^\alpha \frac{dy}{\lambda_2(y)}.$$

Once again, since $\lambda_1(x) > 0$ and $\lambda_2(x) < 0$ for all $x \in \Omega_2$, we have $\Phi^{(2)}(b) > 0$, $\Psi^{(2)}(\alpha) > 0$, and we can state the counterpart in Ω_2 of Lemma 2.3.1.

Lemma 2.3.2 *The solution of problem* (2.3.15) *satisfies:*

$$\mathbf{E}_2^{2,k+1}(\alpha) = \sigma_2 \mathbf{E}_1^{1,k}(\alpha) \tag{2.3.21}$$

with $\sigma_2 < 1$.

Proof. Setting

$$\sigma_2 := \exp\Big\{ -\beta\Big(\Phi^{(2)}(b) + \Psi^{(2)}(\alpha)\Big)\Big\}$$

(2.3.21) follows immediately from (2.3.20). □

We are therefore in condition to prove the convergence of the iteration by subdomain procedure. In that order, let us introduce the following sequence of *interface errors*:

$$\mathbf{E}_\alpha^k := [\mathbf{E}_1^{1,k}(\alpha)]^2 + [\mathbf{E}_2^{2,k}(\alpha)]^2 \quad \text{for } k \geq 1. \tag{2.3.22}$$

From the previous Lemmas we can immediately deduce the following convergence result.

Theorem 2.3.1 *The interface error defined in* (2.3.22) *reduces at each iteration according to the law*

$$\mathbf{E}_\alpha^{k+1} \leq \sigma \mathbf{E}_\alpha^k \tag{2.3.23}$$

for each $k \geq 1$, *where the reduction factor is given by*

$$\sigma := \max\big[(\sigma_1)^2, (\sigma_2)^2\big] < 1.$$

Proof. Owing to (2.3.19) and (2.3.21) we get that the interface error is ruled by

$$\mathbf{E}_\alpha^{k+1} = [\sigma_2 \mathbf{E}_1^{1,k}(\alpha)]^2 + [\sigma_1 \mathbf{E}_2^{2,k}(\alpha)]^2$$

for each $k \geq 1$. □

From the previous theorem, we have

$$\lim_{k\to\infty} \mathbf{E}_\alpha^k = 0.$$

Next, we have to prove that the error $\mathbf{E}^k(x)$, defined in (2.3.13), can be controlled by the error on the interface $\mathbf{E}_\alpha^k$, for each $x \in \Omega$, so that the convergence for the iterations on the interface guarantees convergence in the whole Ω.

Let us focus on what happens within Ω_1:

$$\forall x \in (a,\alpha): \quad \left[\mathbf{E}^{1,k}(x)\right]^2 = \left[\mathbf{E}_1^{1,k}(x)\right]^2 + \left[\mathbf{E}_2^{1,k}(x)\right]^2 \tag{2.3.24}$$

Now, owing to (2.3.16)

$$\left[\mathbf{E}_2^{1,k}(x)\right]^2 < \left[\mathbf{E}_2^{1,k}(\alpha)\right]^2, \tag{2.3.25}$$

and, owing to (2.3.17)

$$\left[\mathbf{E}_1^{1,k}(x)\right]^2 < \left[\mathbf{E}_1^{1,k}(a)\right]^2 = \left[\mathbf{E}_2^{1,k}(a)\right]^2 < \left[\mathbf{E}_2^{1,k}(\alpha)\right]^2. \tag{2.3.26}$$

Therefore, (2.3.25)-(2.3.26)imply

$$\left[\mathbf{E}^{1,k}(x)\right]^2 < \left[\mathbf{E}_2^{1,k}(\alpha)\right]^2 + \left[\mathbf{E}_2^{1,k}(\alpha)\right]^2 < 2\left[\mathbf{E}_2^{1,k}(\alpha)\right]^2. \tag{2.3.27}$$

In a similar way, we get for each $x \in \Omega_2$:

$$\left[\mathbf{E}^{2,k}(x)\right]^2 < \left[\mathbf{E}_1^{2,k}(\alpha)\right]^2 + \left[\mathbf{E}_1^{2,k}(\alpha)\right]^2 < 2\left[\mathbf{E}_1^{2,k}(\alpha)\right]^2. \tag{2.3.28}$$

Consequently, (2.3.27)-(2.3.28) provide

$$\left[\mathbf{E}^{k}(x)\right]^2 < 2\left[\mathbf{E}_\alpha^{k}\right]^2, \tag{2.3.29}$$

for all $x \in \Omega$.

We can therefore state the following result.

Theorem 2.3.2 *The iteration-by-subdomain strategy in* (2.3.11)-(2.3.12) *converges as $k \to \infty$, for any choice of the time step Δt.*

Proof. In order to complete the proof we have to show that this convergence does not depend on the time step. This is an immediate consequence of the fact that each of the quantities $\Psi^{(1)}(a)$, $\Phi^{(1)}(\alpha)$, $\Psi^{(2)}(\alpha)$ and $\Phi^{(2)}(b)$ is greater than 0, so that the exponentials involved attain values less than 1, and we have a contraction on the interface for any choice of Δt. □

Notice that each one of the iterated solution of systems (2.3.14)-(2.3.15) can be viewed as an iteration on the interface, so that it can be reformulated in terms of a mapping $\mathcal{M} : \mathbf{R}^2 \to \mathbf{R}^2$, which is defined, for each $\xi = (\xi_1, \xi_2) \in \mathbf{R}^2$, as

$$\mathcal{M} : \begin{bmatrix} \xi_1 \\ \\ \xi_2 \end{bmatrix} \longmapsto \begin{bmatrix} \mathbf{V}_1^{1,k+1}(\alpha) \\ \\ \mathbf{V}_2^{2,k+1}(\alpha) \end{bmatrix} \tag{2.3.30}$$

where the values $\mathbf{V}_1^{1,k+1}(\alpha)$ and $\mathbf{V}_2^{2,k+1}(\alpha)$ are obtained from systems (2.3.11)-(2.3.12) with assigned incoming values on α given by $\mathbf{V}_2^{1,k+1}(\alpha) = \xi_2$ and $\mathbf{V}_1^{2,k+1}(\alpha) = \xi_1$. As an immediate consequence of Theorem 2.3.2, we can state the following result.

Lemma 2.3.3 *For any choice of the time step $\beta = 1/\Delta t$, the mapping $\mathcal{M}$ defined in* (2.3.30) *is a contraction. Moreover, there exists $C > 0$ such that the reduction factor K is given by*

$$K = e^{-\frac{C}{\Delta t}}.$$

Proof. Since all the problems involved are linear, it is enough to prove contractivity for the mapping $\mathcal{M}^0$, which is the one obtained from $\mathcal{M}$ when $f = \phi_1 = \phi_2 = 0$. Namely, it is enough to prove that there exists a constant $K < 1$ such that, for each $\xi \in \mathbf{R}^2$,

$$\left|\mathcal{M}^0\xi\right|^2 \leq K\left|\xi\right|^2 \tag{2.3.31}$$

From Lemmas 2.3.1 and 2.3.2, this latter mapping can be expressed in matrix form as

$$\mathcal{M}^0\xi = \begin{pmatrix} 0 & \sigma_1 \\ \\ \sigma_2 & 0 \end{pmatrix} \cdot \begin{bmatrix} \xi_1 \\ \\ \xi_2 \end{bmatrix},$$

with σ_1 and σ_2 as defined thereby. Thus, inequality (2.3.31), as well as its uniformity with respect to the time step $\beta = 1/\Delta t$, is an easy consequence of Theorem 2.3.1, with

$$K = \max\left[e^{-\frac{2}{\Delta t}\left[\Phi^{(1)}(\alpha)+\Psi^{(1)}(a)\right]},\ e^{-\frac{2}{\Delta t}\left[\Phi^{(2)}(b)+\Psi^{(2)}(\alpha)\right]}\right].$$

□

Remark 2.3.2 The expression for the contractive constant suggests to choose α in order to have K as small as possible: the constant K is optimal whenever the arguments in the maximum are equal, namely α must satisfy

$$\int_a^\alpha \frac{dy}{\lambda_1(y)} - \int_a^\alpha \frac{dy}{\lambda_2(y)} = \int_\alpha^b \frac{dy}{\lambda_1(y)} - \int_\alpha^b \frac{dy}{\lambda_2(y)},$$

which can be rewritten as

$$[\Lambda_1 - \Lambda_2](\alpha) = \frac{[\Lambda_1 - \Lambda_2](a) + [\Lambda_1 - \Lambda_2](b)}{2},$$

where we have denoted with Λ_k $(k = 1, 2)$ any primitive of $1/\lambda_k(x)$. Since in our framework $\lambda_k = \lambda_k(\mathbf{V}^n)$ $(k = 1, 2)$, $\mathbf{V}^n$ being the solution at the previous step in the time marching process, this could allow to possibly adapt the spatial decomposition at each time step. □

Remark 2.3.3 Notice that the fixed point of the mapping $\mathcal{M}^0$ is the solution of the Steklov-Poincaré interface equation, that reads

$$\begin{pmatrix} 1 & -\sigma_1 \\ \\ -\sigma_2 & 1 \end{pmatrix} \cdot \begin{bmatrix} \xi_1 \\ \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}$$

when $f = \phi_1 = \phi_2 = 0$. Therefore, each iteration of the mapping $\mathcal{M}^0$ can be viewed as one step in a non-preconditioned Richardson iterative method to solve the Steklov-Poincaré equation. □

An Equivalent Algorithm

As a matter of fact, the procedure in (2.3.11)-(2.3.12) could be advanced (at least in principle) in parallel, but this is somehow redundant. In fact, the iteration by subdomain algorithm can be efficiently exploited in the following sequential way:

STEP 1.

Given $\xi_1^0 \in \mathbf{R}$, solve for each $k \geq 0$:

$$\begin{cases} \beta \mathbf{V}_2^{1,k+1} + \lambda_1 \dfrac{d}{dx} \mathbf{V}_2^{1,k+1} = f_1^{(1)} & \text{in } \Omega_1 \\ \\ \mathbf{V}_2^{1,k+1}(\alpha) = \xi_1^k, & \end{cases}$$

then solve

$$\begin{cases} \beta \mathbf{V}_1^{1,k+1} + \lambda_2 \dfrac{d}{dx} \mathbf{V}_1^{1,k+1} = f_2^{(1)} \quad \text{in } \Omega_1 \\ \\ \mathbf{V}_1^{1,k+1}(a) = \mathbf{V}_2^{1,k+1}(a) + \phi_1. \end{cases}$$

Notice that the value of $\mathbf{V}_1^{1,k+1}(a)$ is completely determined by the value of $\mathbf{V}_2^{1,k+1}(a)$ and by the physical boundary condition at the left endpoint of the interval.

STEP 2.

Set $\xi_2^{k+1} = \mathbf{V}_1^{1,k+1}(\alpha)$ and solve

$$\begin{cases} \beta \mathbf{V}_1^{2,k+1} + \lambda_2 \dfrac{d}{dx} \mathbf{V}_1^{2,k+1} = f_2^{(2)} & \text{in } \Omega_2 \\ \\ \mathbf{V}_1^{2,k+1}(\alpha) = \xi_2^{k+1}, & \end{cases}$$

then solve

$$\begin{cases} \beta \mathbf{V}_2^{2,k+1} + \lambda_1 \dfrac{d}{dx} \mathbf{V}_2^{2,k+1} = f_1^{(2)} \quad \text{in } \Omega_2 \\ \\ \mathbf{V}_2^{2,k+1}(b) = \mathbf{V}_1^{2,k+1}(b) + \phi_2. \end{cases}$$

Once again, the value of $\mathbf{V}_2^{2,k+1}(b)$ is completely determined by the value of $\mathbf{V}_1^{2,k+1}(b)$ and by the physical boundary condition at the endpoint.

STEP 3.

Set $\xi_1^{k+1} = \mathbf{V}_2^{2,k+1}(\alpha)$, go to **STEP 1** and iterate.

We can easily prove convergence for this latter algorithm, stated in the following theorem.

Theorem 2.3.3 *The iteration by subdomain in* **STEP 1** - **STEP 3** *is equivalent to* (2.3.11)-(2.3.12)*, and it converges as $k \to \infty$, independently of the choice of the time step Δt.*

Proof. Consider the sequences

$$\left\{\mathbf{V}_2^{2,k}(\alpha)\right\}_k \quad \text{and} \quad \left\{\mathbf{V}_1^{1,k}(\alpha)\right\}_k,$$

stemming from the iteration-by-subdomain method in (2.3.11)-(2.3.12) as well as the sequences

$$\left\{\xi_1^k\right\}_k \quad \text{and} \quad \left\{\xi_2^k\right\}_k,$$

stemming from **STEP 1** - **STEP 3**. It can be easily viewed that the latter ones are subsequences of the previous ones, which are convergent. Convergence also for the sequences $\{\xi_1^k\}_k$ and $\{\xi_2^k\}$ is therefore straightforward. □

2.3.2 Fully discrete finite elements approximation for the single domain problem

In this section, following what is done in [56] for linear hyperbolic systems with constant coefficients, we focus our attention on the finite dimensional approximation for the system stemming from a semi-implicit time discretisation of system (2.3.10).

The scalar case and its finite elements approximation via the Streamline Diffusion Method

Since system (2.3.10) consists of two scalar transport equations coupled only through the boundary conditions, let us consider the following problem

$$\begin{cases} \dfrac{\beta}{\lambda(x)}\, u + u' = f(x) \qquad \text{in } \Omega = (a,\, b) \\[2ex] u(a) = \xi \end{cases} \tag{2.3.32}$$

where $\beta > 0$, $\lambda(x) \geq \lambda_* > 0$ $\forall x \in \overline{\Omega}$ and we have denoted with u' the space derivative of u with respect to x, *i.e.*

$$u' := \frac{du}{dx}.$$

The variational formulation of problem (2.3.32) reads

$$\text{Find } u \in V : \qquad a(u,v) = L(v), \qquad \forall v \in V \tag{2.3.33}$$

where

$$\begin{aligned} a(w,v) \;&= \int_\Omega \left\{ \frac{\beta}{\lambda(x)}\, wv + \frac{1}{2}[w'v - wv'] \right\} + \frac{1}{2}(wv)\,(b) \\ L_\xi(v) \;&= \int_\Omega fv + \frac{1}{2}\xi v(a) \end{aligned} \tag{2.3.34}$$

In order to approximate the solution of problem (2.3.32) with finite elements, let $\mathcal{T}_h$ be a subdivision of the interval Ω into a finite number of subintervals $[x_{j-1}, x_j]$ such that $|x_j - x_{j-1}| \leq h$ for $j = 1, \ldots, N$, where

$$a = x_0 < x_1 < \ldots < x_N = b.$$

We introduce the following finite element spaces (for further details on Finite Element Methods see the Appendix, whereas, for a more exhaustive treatment of the topic, we refer to the books by C. Johnson [67] or A. Quarteroni and A. Valli [88]):

$$V^h(\Omega) := \{v \in C^0(\overline{\Omega}) \mid v_{|K} \in \mathbb{P}_k, \ \forall K \in \mathcal{T}_h\} \qquad k \geq 1, \tag{2.3.35}$$

$$V_a^h(\Omega) := \left\{v \in V^h(\Omega) \mid v(a) = 0\right\}, \tag{2.3.36}$$

and a possible discrete version of (2.3.33) reads

$$\text{Find } u_h \in V^h: \qquad \tilde{a}_h(u_h, v_h) = \tilde{L}_h(v_h), \qquad \forall v_h \in V^h \tag{2.3.37}$$

with

$$\begin{aligned} \tilde{a}_h(w,v) &= \int_\Omega \left\{ \frac{\beta}{\lambda_h(x)} wv + \frac{1}{2}[w'v - wv'] \right\} + \frac{1}{2}(wv)\,(b) \\ \tilde{L}_h(v) &= \int_\Omega f_h v + \frac{1}{2}\xi_h v(a) \end{aligned} \tag{2.3.38}$$

where λ_h, f_h and ξ_h are suitable approximations of the data λ, f and ξ. This choice is unfortunately not satisfactory, since problem (2.3.37) must be stabilized. As it is quite natural in the case of transport problems, we will use a *Streamline Diffusion* technique, which consists in adding to the original variational formulation the element residual

$$\delta h \int_K \left[\frac{\beta}{\lambda_h(x)} w + w' - f_h \right] v',$$

where the value of $\delta > 0$ may depend on λ_h, and $meas(\Omega)$, but is independent of h. The discrete problem we deal with is therefore

$$\text{Find } u_h \in V^h: \qquad a_h(u_h, v_h) = L_h(v_h), \qquad \forall v_h \in V^h \tag{2.3.39}$$

where

$$\begin{aligned} a_h(w,v) &= \tilde{a}_h(w,v) + \sum_{K\in\mathcal{T}} \delta h \int_K \left[\frac{\beta}{\lambda_h(x)}\, w + w'\right] \cdot v' \\ &= \tilde{a}_h(w,v) + \int_\Omega \left[\frac{\beta}{\lambda_h(x)}\, w + w'\right] \cdot \delta h v' \\ L_h(v) &= \tilde{L}_h(v) + \sum_{K\in\mathcal{T}} \delta h \int_K f_h v' \\ &= \tilde{L}_h(v) + \int_\Omega f_h \delta h v' \end{aligned} \tag{2.3.40}$$

In order to have well-posedness for problem (2.3.39) we assume the bilinear form $a_h(.,.)$ to be positive, namely

$$\mu^* := \inf_{x\in\Omega} \left(\frac{\beta}{\lambda_h(x)} + \frac{1}{2} \frac{\beta \delta h \lambda_h'(x)}{\lambda_h^2(x)} \right) > 0, \tag{2.3.41}$$

which is fulfilled for each δ such that

$$0 < \delta < \delta^0 := 2\, \frac{\min_\Omega \lambda_h(x)}{\max\{-\min_\Omega \lambda_h'(x), 0\}}, \tag{2.3.42}$$

The above condition has to be interpreted in the following way: if $\min_\Omega \lambda_h'(x) > 0$, then no upper bound is needed for δ.

Inflow-outflow estimates for the finite elements approximation

In this section, adapting to our problem the approach of [53] and [56], we give some estimates of inflow-outflow type for the scalar problem (2.3.39) which will be used in the sequel. In that order, let $\Omega := (a,b)$ and let us consider the following problems:

(P1) *Find* $u_h \in V^h(\Omega)$ *such that*

$$\begin{aligned} \int_\Omega \left\{ \frac{\beta}{\lambda_h(x)}\, u_h + u_h' - f_h \right\} \cdot (v + \delta h v') &= 0 \qquad \forall v \in V_a^h(\Omega) \\ u_h(a) &= \chi \end{aligned} \tag{2.3.43}$$

and

(P2) *Find* $u_h^m \in V^h(\Omega)$ *such that*

$$\begin{aligned} \int_\Omega \left\{ \frac{\beta}{\lambda_h(x)}\, u_h^m + (u_h^m)' - f_h \right\} \cdot (v + \delta h v') &= 0 \qquad \forall v \in V_a^h(\Omega) \\ u_h^m(a) &= \chi_m \end{aligned} \tag{2.3.44}$$

We are in the position to prove the following result.

Lemma 2.3.4 *Assume $\lambda_h(x) \geq M_1 > 0$ for all $x \in \Omega$, $\lambda_h' \in L^\infty(\Omega)$ and (2.3.41), and let u_h and u_h^m be the solutions to problems (2.3.43) and (2.3.44) above. Then there exists a constant $H_\Omega < 1$ such that*

$$(u_h - u_h^m)^2\,(b) \leq H_\Omega\,(\chi - \chi_m)^2\,, \tag{2.3.45}$$

provided δ is sufficiently small.

Proof. The difference $e_m := u_h - u_h^m$ satisfies the following error equation

$$\begin{aligned} &\int_\Omega \left\{ \frac{\beta}{\lambda_h(x)}\, e_m + e_m' \right\} \cdot (v + \delta h v') = 0 \\ &e_m(a) = \chi - \chi_m \end{aligned} \tag{2.3.46}$$

If $\lambda_h(b) < \lambda_h(a)$, we take in (2.3.46) $v = e_m$ and we get

$$\begin{aligned} 0 &= \int_\Omega \frac{\beta}{\lambda_h} e_m^2 + \int_\Omega e_m e_m' + \delta h \int_\Omega \frac{\beta}{\lambda_h} e_m e_m' + \delta h \int_\Omega (e_m')^2 \\ &= \int_\Omega \left(\frac{\beta}{\lambda_h} + \frac{\beta\delta h \lambda_h'}{2\lambda_h^2} \right) e_m^2 + \frac{1}{2} \left[\left(1 + \frac{\beta\delta h}{\lambda_h} \right) e_m^2 \right]_a^b + \delta h \int_\Omega (e_m')^2. \end{aligned} \tag{2.3.47}$$

The third term on the right hand side is positive, while assumption (2.3.41) provides positivity also for the first one, so that we have

$$\left(1 + \frac{\beta\delta h}{\lambda_h(b)} \right) e_m^2(b) \;\leq\; \left(1 + \frac{\beta\delta h}{\lambda_h(a)} \right) e_m^2(a).$$

Inequality (2.3.45) follows with

$$H_\Omega \;:=\; \frac{1 + \frac{\beta\delta h}{\lambda_h(a)}}{1 + \frac{\beta\delta h}{\lambda_h(b)}} \;<\; 1.$$

If $\lambda_h(b) \geq \lambda_h(a)$, let $\varphi \in W^{1,\infty}(\Omega)$ be the linear function such that $\varphi(a) = 0$ and $\varphi(b) = 1$, *i.e.*

$$\varphi(x) = \frac{x - a}{b - a}.$$

We take $v = (1 + \eta\varphi)\, e_m$ in (2.3.46), with $\eta > 0$, and we get

$$
\begin{aligned}
0 &= \int_\Omega \left\{ \frac{\beta}{\lambda_h} e_m + e_m' \right\} \cdot \left\{ (1+\eta\varphi)e_m + \delta h \left[(1+\eta\varphi)e_m\right]' \right\} \\
&= \int_\Omega \frac{\beta}{\lambda_h} e_m^2 (1+\eta\varphi) + \underbrace{\int_\Omega (1+\eta\varphi)e_m e_m'}_{(1)} + \underbrace{\delta h \int_\Omega \frac{\beta}{\lambda_h}(1+\eta\varphi)e_m e_m'}_{(2)} + \delta h \int_\Omega \frac{\beta}{\lambda_h}\eta\varphi' e_m^2 \\
&\quad + \delta h \int_\Omega (1+\eta\varphi)(e_m')^2 + \underbrace{\delta h \int_\Omega \eta\varphi' e_m e_m'}_{(3)}
\end{aligned}
\tag{2.3.48}
$$

Since φ is linear,

$$
\begin{aligned}
(1) &= \frac{1}{2}\left[(1+\eta\varphi)e_m^2\right]_a^b - \frac{1}{2}\int_\Omega \eta\varphi' e_m^2 \\
(2) &= \frac{\delta h}{2}\left[\frac{\beta}{\lambda_h}(1+\eta\varphi)e_m^2\right]_a^b + \frac{1}{2}\int_\Omega \frac{\beta\delta h \lambda_h'}{2\lambda_h^2}(1+\eta\varphi)e_m^2 - \frac{1}{2}\int_\Omega \frac{\beta\delta h}{\lambda_h}\eta\varphi' e_m^2 \\
(3) &= \frac{\delta h}{2}\left[\eta\varphi' e_m^2\right]_a^b - \frac{\delta h}{2}\int_\Omega \eta\varphi'' e_m^2 \;=\; \frac{\delta h}{2}\left[\eta\varphi' e_m^2\right]_a^b,
\end{aligned}
$$

and this entails

$$
\begin{aligned}
0 = &\int_\Omega \left\{ \left(\frac{\beta}{\lambda_h} + \frac{\beta\delta h \lambda_h'}{2\lambda_h^2}\right)(1+\eta\varphi) + \frac{1}{2}\left(\frac{\beta\delta h}{\lambda_h} - 1\right)\eta\varphi' \right\} e_m^2 \\
&+ \frac{1}{2}\left[\left\{ \left(1 + \frac{\beta\delta h}{\lambda_h}\right)(1+\eta\varphi) + \delta h \eta\varphi' \right\} e_m^2 \right]_a^b + \delta h \int_\Omega (1+\eta\varphi)(e_m')^2
\end{aligned}
\tag{2.3.49}
$$

The third term of the sum is positive independently of η. Concerning the first one we have two opportunities: if $\beta\delta h - \lambda_h > 0$ in Ω, this term is positive without any further restriction on η, while if $\beta\delta h - \lambda_h < 0$ for some $x \in \Omega$, taking into account the definition of φ, its positivity is guaranteed if

$$
\eta \;\le\; \eta^* \;:=\; 2\,(b-a)\,\mu^*. \tag{2.3.50}
$$

where μ^* is the one defined in (2.3.41).
We obtain from (2.3.49)

$$
\left[\frac{\delta\eta h}{b-a} + \left(1 + \frac{\beta\delta h}{\lambda_h(b)}\right)(1+\eta)\right] e_m^2(b) \;\le\; \left[\frac{\delta\eta h}{b-a} + \left(1 + \frac{\beta\delta h}{\lambda_h(a)}\right)\right] e_m^2(a). \tag{2.3.51}
$$

Let

$$\delta_0 \ := \ \sup\left\{\delta > 0 \;\middle|\; \eta_* := \beta\delta h \frac{\lambda_h(b) - \lambda_h(a)}{\lambda_h(a)\left[\lambda_h(b) + \beta\delta h\right]} \ < \ \eta^*\right\} \tag{2.3.52}$$

and define

$$\delta^* := \min\left\{\delta^0 \ , \ \delta_0\right\}, \tag{2.3.53}$$

where δ^0 is the one introduced in (2.3.42). Thus, for any $\eta \in]\eta_*, \eta^*]$, inequality (2.3.45) follows with

$$H_\Omega \ := \ \frac{\frac{\delta\eta h}{b-a} + \left(1 + \frac{\beta\delta h}{\lambda_h(a)}\right)}{\frac{\delta\eta h}{b-a} + \left(1 + \frac{\beta\delta h}{\lambda_h(b)}\right)\ (1+\eta)} \ < \ 1, \tag{2.3.54}$$

provided $0 < \delta \ < \ \delta^*$.

□

We then introduce the finite element space

$$V_b^h(\Omega) := \left\{v \in V^h(\Omega) \mid v(b) = 0\right\} \tag{2.3.55}$$

and, in a similar way it is not difficult to prove the following lemma.

Lemma 2.3.5 *Assume there exists $M_2 > 0$ such that $\mu_h(x) \leq -M_2$ for all $x \in \Omega$, assume $\mu_h' \in L^\infty(\Omega)$ and*

$$\mu_* := \inf_\Omega \left(\frac{\beta}{|\mu_h|} + \frac{1}{2}\frac{\beta\delta h \mu_h'}{|\mu_h|^2}\right) > 0,$$

Let w_h and w_h^m be the solutions to problems

(P1) Find $w_h \in V^h(\Omega)$ such that

$$\int_\Omega \left\{\frac{\beta}{\mu_h(x)}\, w_h + w_h' - f_h\right\} \cdot (v + \delta h v') = \ 0 \qquad \forall v \in V_b^h(\Omega)$$

$$w_h(b) = \ \xi$$

and

(P2) Find $w_h^m \in V^h(\Omega)$ such that

$$\int_\Omega \left\{\frac{\beta}{\mu_h(x)}\, w_h^m + (w_h^m)' - f_h\right\} \cdot (v + \delta h v') = \ 0 \qquad \forall v \in V_b^h(\Omega)$$

$$w_h^m(b) = \ \xi_m$$

respectively. Then, there exists a constant $K_\Omega < 1$ *such that*

$$\left(w_h - w_h^m\right)^2(a) \le K_\Omega \, (\xi - \xi_m)^2, \tag{2.3.56}$$

provided δ *is sufficiently small.* □

The finite elements formulation in the vector case

In this section we go back to the complete system (2.3.10) and we introduce the finite element spaces:

$$\mathcal{W}^h(\Omega) := \left[V^h(\Omega)\right]^2 \qquad \text{and} \qquad \mathcal{W}_0^h(\Omega) := V_a^h(\Omega) \times V_b^h(\Omega). \tag{2.3.57}$$

At each time step, neglecting any index referring to the time step, the stabilized fully discrete formulation for system (2.3.6) reads:

Find $\mathbf{V}_h \in \mathcal{W}^h(\Omega)$ *such that*

$$\begin{cases} \displaystyle\int_\Omega \left(\beta\Lambda_h^{-1}\mathbf{V}_h + \frac{d\mathbf{V}_h}{dx} - \mathbf{f}_h, \varphi + hD\frac{d\varphi}{dx}\right) dx = 0 \qquad \forall\, \varphi \in \mathcal{W}_0^h(\Omega) \\ \left[\mathbf{V}_h\right]_1(a) - \left[\mathbf{V}_h\right]_2(a) = \phi_1 \\ \left[\mathbf{V}_h\right]_1(b) - \left[\mathbf{V}_h\right]_2(b) = \phi_2 \end{cases} \tag{2.3.58}$$

where clearly $\Lambda_h = \Lambda(\mathbf{V}_h^n)$ and $\mathbf{f}_h = \beta\Lambda_h^{-1}\mathbf{V}_h^n$, while $D = \operatorname{diag}(\delta_1, \delta_2)$ is so far a suitable diagonal matrix.

If, following what is done in the first partof this Section, we introduce the bilinear forms

$$a_h^+(u,v) := \int_\Omega \left\{ \left[\frac{\beta}{\lambda_{1h}} + \frac{1}{2}\frac{\beta\delta_1 h\lambda_{1h}'}{\lambda_{1h}^2}\right] uv + \frac{1}{2}\left(1 + \frac{\beta\delta_1 h}{\lambda_{1h}}\right)\left[uv' - u'v\right] + \delta_1 h u'v' \right\}$$
$$+\frac{1}{2}\left[1 + \frac{\beta\delta_1 h}{\lambda_{1h}(b)}\right] uv(b)$$

$$a_h^-(u,v) := \int_\Omega \left\{ \left[\frac{\beta}{|\lambda_{2h}|} + \frac{1}{2}\frac{\beta\delta_2 h\lambda_{2h}'}{\lambda_{2h}^2}\right] uv + \frac{1}{2}\left(1 - \frac{\beta\delta_2 h}{\lambda_{2h}}\right)\left[uv' - u'v\right] + \delta_2 h u'v' \right\}$$
$$+\frac{1}{2}\left[1 - \frac{\beta\delta_2 h}{\lambda_{2h}(a)}\right] uv(a)$$

as well as the linear forms

$$F_h^+(v) := \int_\Omega \mathbf{f}_{1h}\left(v + \delta h v'\right) + \frac{1}{2}\left(1 + \frac{\beta\delta h}{\lambda_{1h}(a)}\right) uv(a)$$

$$F_h^-(v) := \int_\Omega \mathbf{f}_{2h}\left(v + \delta h v'\right) + \frac{1}{2}\left(1 - \frac{\beta\delta h}{\lambda_{2h}(b)}\right) uv(b),$$

problem (2.3.58) can equivalently be written as

Find $\mathbf{V}_h \in \mathcal{W}^h(\Omega)$ *such that*

$$\begin{cases} \mathbf{a}_h(\mathbf{V}, \varphi) = \mathbf{F}_h(\varphi) \qquad \forall\, \varphi \in \mathcal{W}_0^h(\Omega) \\ \mathbf{V}_1(a) - \mathbf{V}_2(a) = \phi_1 \\ \mathbf{V}_1(b) - \mathbf{V}_2(b) = \phi_2 \end{cases} \tag{2.3.59}$$

where

$$\mathbf{a}_h(\mathbf{V}, \varphi) = \begin{pmatrix} a_h^+(\mathbf{V}_1, \varphi_1) \\ \\ a_h^-(\mathbf{V}_2, \varphi_2) \end{pmatrix} \quad \text{and} \quad \mathbf{F}_h(\varphi) = \begin{pmatrix} F_h^+(\varphi_1) \\ \\ F_h^-(\varphi_2) \end{pmatrix}$$

2.3.3 Fully Discrete Multidomain Formulation and Iterative Algorithm

In this section we go back to the multidomain formulation of Section 2.3.1 and we prove convergence for an iteration-by-subdomains procedure in the fully discrete case.
To this aim, let us denote once again with α the interface between the two subdomains (which, for sake of simplicity, is assumed to coincide with a node of the mesh), and consider the finite element spaces which are the restrictions to $\Omega_1 = (a, \alpha)$ and $\Omega_2 = (\alpha, b)$ of the spaces $\mathcal{W}^h(\Omega)$ and $\mathcal{W}_0^h(\Omega)$, namely

$$\mathcal{W}^h(\Omega_j) := \left[V^h(\Omega_j)\right]^2 \qquad j = 1, 2 \tag{2.3.60}$$

$$\mathcal{W}_0^h(\Omega_1) := V_a^h(\Omega_1) \times V_\alpha^h(\Omega_1) \qquad \text{and} \qquad \mathcal{W}_0^h(\Omega_2) := V_\alpha^h(\Omega_2) \times V_b^h(\Omega_2), \tag{2.3.61}$$

and consider the discretized version of the multidomain formulation (2.3.10)

$$\int_{\Omega_1} \left(\beta\Lambda_h^{-1}\mathbf{V}_h^1 + \frac{d}{dx}\mathbf{V}_h^1 - \mathbf{f}_h^{(1)}, \varphi + hD\frac{d\varphi}{dx}\right) dx = 0 \quad \forall\, \varphi \in \mathcal{W}_0^h(\Omega_1) \tag{2.3.62}$$

$$\int_{\Omega_2} \left(\beta\Lambda_h^{-1}\mathbf{V}_h^2 + \frac{d}{dx}\mathbf{V}_h^2 - \mathbf{f}_h^{(2)}, \psi\varphi + hD\frac{d\psi}{dx}\right) dx = 0 \quad \forall\, \psi \in \mathcal{W}_0^h(\Omega_2) \tag{2.3.63}$$

$$\left[\mathbf{V}_h^1\right]_1(a) = \left[\mathbf{V}_h^1\right]_2(a) + \phi_1 \tag{2.3.64}$$

$$\left[\mathbf{V}_h^2\right]_2(b) = \left[\mathbf{V}_h^2\right]_1(b) + \phi_2 \tag{2.3.65}$$

$$\left[\mathbf{V}_h^1\right]_2(\alpha) = \left[\mathbf{V}_h^2\right]_2(\alpha) \tag{2.3.66}$$

$$\left[\mathbf{V}_h^2\right]_1(\alpha) = \left[\mathbf{V}_h^1\right]_1(\alpha) \tag{2.3.67}$$

where, as usual, $\mathbf{f}_h^{(i)}$ denotes the restriction of $\mathbf{f}_h$ to Ω_i, $(i = 1, 2)$.
We introduce, as in the continuous case, an iterative procedure to solve system (2.3.62)-(2.3.67) above. At the $(m+1)$-*th* iteration, it reads as follows

$$\int_{\Omega_1} \left(\beta\Lambda_h^{-1}\mathbf{V}_h^{1,m+1} + \frac{d}{dx}\mathbf{V}_h^{1,m+1} - \mathbf{f}_h^{(1)}, \varphi + hD\frac{d\varphi}{dx}\right) dx = 0 \quad \forall\ \varphi \in \mathcal{W}_0^h(\Omega_1) \tag{2.3.68}$$

$$\int_{\Omega_2} \left(\beta\Lambda_h^{-1}\mathbf{V}_h^{2,m+1} + \frac{d}{dx}\mathbf{V}_h^{2,m+1} - \mathbf{f}_h^{(2)}, \psi + hD\frac{d\psi}{dx}\right) dx = 0 \quad \forall\ \psi \in \mathcal{W}_0^h(\Omega_2) \tag{2.3.69}$$

$$\left[\mathbf{V}_h^{1,m+1}\right]_1(a) = \left[\mathbf{V}_h^{1,m+1}\right]_2(a) + \phi_1 \tag{2.3.70}$$

$$\left[\mathbf{V}_h^{2,m+1}\right]_2(b) = \left[\mathbf{V}_h^{2,m+1}\right]_1(b) + \phi_2 \tag{2.3.71}$$

$$\left[\mathbf{V}_h^{1,m+1}\right]_2(\alpha) = \left[\mathbf{V}_h^{2,m}\right]_2(\alpha) \tag{2.3.72}$$

$$\left[\mathbf{V}_h^{2,m+1}\right]_1(\alpha) = \left[\mathbf{V}_h^{1,m}\right]_1(\alpha) \tag{2.3.73}$$

Convergence Analysis

Procedure (2.3.68)-(2.3.73) can be interpreted as a discrete iterative mapping $\mathcal{M}_h : \mathbf{R}^2 \to \mathbf{R}^2$, acting in the following way:

$$\mathcal{M}_h \ : \ \begin{pmatrix} \left[\mathbf{V}_h^{1,m}\right]_1(\alpha) \\ \\ \left[\mathbf{V}_h^{2,m}\right]_2(\alpha) \end{pmatrix} \longmapsto \begin{pmatrix} \left[\mathbf{V}_h^{1,m+1}\right]_1(\alpha) \\ \\ \left[\mathbf{V}_h^{2,m+1}\right]_2(\alpha) \end{pmatrix}. \tag{2.3.74}$$

where $\left[\mathbf{V}_h^{1,m+1}\right]_1(\alpha)$ and $\left[\mathbf{V}_h^{2,m+1}\right]_2(\alpha)$ stem from the solutions of systems (2.3.68)-(2.3.70)-(2.3.72) and (2.3.69)-(2.3.71)-(2.3.73), respectively.
The convergence properties of the mapping $\mathcal{M}_h$ are given in the following theorem.

Theorem 2.3.4 *The discrete mapping $\mathcal{M}_h$ is a contraction on the interface, provided the entries of the diagonal matrix D are sufficiently small.*

Proof. The mapping $\mathcal{M}_h$ is linear, thus it is enough to prove that it is contractive on the error, and to this aim it is immediate to see that the difference $\mathbf{E}_h^{j,m} := \mathbf{V}_h^j - \mathbf{V}_h^{j,m}$ $(j = 1, 2)$ satisfies the following error equations (as usual, subindices denote components)

$$\int_{\Omega_1} \left(\beta\Lambda_h^{-1}\mathbf{E}_h^{1,m+1} + \frac{d}{dx}\mathbf{E}_h^{1,m+1}, \varphi + hD\frac{d\varphi}{dx}\right) dx = 0 \quad \forall\ \varphi \in \mathcal{W}_0^h(\Omega_1) \tag{2.3.75}$$

$$\int_{\Omega_2}\left(\beta\Lambda_h^{-1}\mathbf{E}_h^{2,m+1}+\frac{d}{dx}\mathbf{E}_h^{2,m+1},\psi+hD\frac{d\psi}{dx}\right)dx=0\quad\forall\,\psi\,\in\,\mathcal{W}_0^h(\Omega_2)\tag{2.3.76}$$

$$\left[\mathbf{E}_h^{1,m+1}\right]_1(a)=\left[\mathbf{E}_h^{1,m+1}\right]_2(a)\tag{2.3.77}$$

$$\left[\mathbf{E}_h^{2,m+1}\right]_2(b)=\left[\mathbf{E}_h^{2,m+1}\right]_1(b)\tag{2.3.78}$$

$$\left[\mathbf{E}_h^{1,m+1}\right]_2(\alpha)=\left[\mathbf{E}_h^{2,m}\right]_2(\alpha)\tag{2.3.79}$$

$$\left[\mathbf{E}_h^{2,m+1}\right]_1(\alpha)=\left[\mathbf{E}_h^{1,m}\right]_1(\alpha)\tag{2.3.80}$$

Equations (2.3.75) and (2.3.76) consist of two scalar equations coupled only through the boundary conditions. The boundedness assumption on ρ and u entails boundedness also for $\mathbf{V}_h$ and $\mathbf{V}_h'$ at each time step. More, since

$$\left[\begin{array}{c}\lambda_{1h}\\ \\ \lambda_{2h}\end{array}\right]=\frac{1}{4}\left(\begin{array}{cc}1+\gamma & 3-\gamma\\ \\ 3-\gamma & 1+\gamma\end{array}\right)\left[\begin{array}{c}\mathbf{V}_{h,1}\\ \\ \mathbf{V}_{h,2}\end{array}\right],\tag{2.3.81}$$

where the vector $(\mathbf{V}_{h,1},\mathbf{V}_{h,2})$ is evaluated at the previous time step, we get $\lambda_{1h}',\lambda_{2h}'\in L^\infty(\Omega_j)$ (for $j=1,2$). As a consequence if the entries of D are small enough, we can apply Lemmas (2.3.4) and (2.3.5) in both Ω_1 and Ω_2.
Let us focus on Ω_1: from Lemma (2.3.5) there exists a constant $K_{\Omega_1}<1$ such that

$$\left[\mathbf{E}_h^{1,m+1}\right]_2^2(a)\le K_{\Omega_1}\left[\mathbf{E}_h^{1,m+1}\right]_2^2(\alpha).$$

Moreover, from (2.3.77) and Lemma (2.3.4) there exists a constant $H_{\Omega_1}<1$ such that

$$\left[\mathbf{E}_h^{1,m+1}\right]_1^2(\alpha)\le H_{\Omega_1}\left[\mathbf{E}_h^{1,m+1}\right]_2^2(a)\le H_{\Omega_1}\cdot K_{\Omega_1}\left[\mathbf{E}_h^{1,m+1}\right]_2^2(\alpha).\tag{2.3.82}$$

From a similar argument within Ω_2 there exist constants $H_{\Omega_2}<1$ and $K_{\Omega_2}<1$ such that

$$\left[\mathbf{E}_h^{2,m+1}\right]_2^2(\alpha)\le H_{\Omega_2}\cdot K_{\Omega_2}\left[\mathbf{E}_h^{2,m+1}\right]_1^2(\alpha)\tag{2.3.83}$$

Gathering together (2.3.79), (2.3.80), (2.3.82) and (2.3.83) we have

$$|\mathcal{M}_h\mathbf{E}_h^m|^2\le\mathcal{K}\,|\mathbf{E}_h^m|^2$$

where

$$\mathcal{K}:=\max\left\{H_{\Omega_1}\cdot K_{\Omega_1},\ H_{\Omega_2}\cdot K_{\Omega_2}\right\}<1,$$

and this concludes the proof. □

2.3.4 Error Estimates

In this section we study the approximation error we get from the characteristic approach to the Euler system. For that purpose, we firstly derive some standard approximation errors for the Streamline Diffusion Method in the single domain case, then we give an estimate for our approach in the same situation.

Error Estimates for the Streamline Diffusion Method

In this section we give some standard error estimates for the Streamline Diffusion finite elements discretisation of a transport problem. In that order, we consider the following problem

$$\begin{cases} \dfrac{\beta}{\lambda(x)}\, u + u' = f(x) \qquad \text{in } \Omega = (a,\, b) \\ \\ u(a) = \xi \end{cases} \tag{2.3.84}$$

where $\beta > 0$, $\lambda(x) \geq \lambda_* > 0$ $\forall x \in \overline{\Omega}$, and, as in equation (2.3.32), we have denoted with u' the space derivative of u with respect to x.
Problem (2.3.84) is well known to have a unique solution, which is given, for $x \in \Omega$, by

$$u(x) = \exp\left(-\beta \int_a^x \frac{dy}{\lambda(y)}\right) \times \left[\xi + \int_a^x f(t) \exp\left(\beta \int_a^t \frac{dy}{\lambda(y)}\right) dt\right] \tag{2.3.85}$$

For $f(x) \in L^2(\Omega)$, and $\beta/\lambda(x) \in L^\infty(\Omega)$, the solution u belongs to $H^1(\Omega)$ and satisfies the following a priori estimate

$$\|u\|_{\mathbf{H}^1} \leq C\left(\|f\|_0 + |\xi|\right) \tag{2.3.86}$$

Since in our framework λ and f depend on the solutions at the previous time step, the streamline diffusion method for problem (2.3.84) reads

Find $u_h \in V^h(\Omega)$ such that

$$\begin{cases} \displaystyle\int_\Omega \left[\frac{\beta}{\lambda_h(x)}\, u_h + u_h' - f_h\right] [\varphi + \delta h \varphi'] = 0 \qquad \forall \varphi \in V_a^h(\Omega) \\ \\ u_h(a) = \xi. \end{cases} \tag{2.3.87}$$

where λ_h and f_h are suitable approximations of λ and the right hand side f, respectively, and where the finite element spaces $V^h(\Omega)$ and $V_a^h(\Omega)$ are the ones introduced in (2.3.35) and (2.3.36). Problem (2.3.87) is well known to have a unique solution under the coerciveness assumption (2.3.41).
Let us consider the following auxiliary problem

$$\begin{cases} \dfrac{\beta}{\lambda_h(x)}\, \hat{u} + \hat{u}' = f_h(x) \qquad \text{in } \Omega = (a,\, b) \\ \\ \hat{u}(a) = \xi \end{cases} \tag{2.3.88}$$

whose exact solution is given by

$$\hat{u}(x) = \exp\left(-\beta \int_a^x \frac{dy}{\lambda_h(y)}\right) \times \left[\xi + \int_a^x f_h(t) \exp\left(\beta \int_a^t \frac{dy}{\lambda_h(y)}\right) dt\right], \tag{2.3.89}$$

and, if $f_h \in L^2(\Omega)$, it satisfies an a priori estimate analogous to (2.3.86).
Under the coerciveness assumption (2.3.41), it is not difficult, by means of standard arguments, to prove the following error estimates.

Lemma 2.3.6 *Let $\hat{u}$ and u_h be the solutions of problems* (2.3.88) *and* (2.3.87) *respectively. Assume that* (2.3.41) *is satisfied, that $\lambda_h(x) \in L^\infty(\Omega)$, and that $\lambda_h(x) \geq \lambda_* > 0$ for all $x \in \Omega$. Moreover, assume $f_h \in L^2(\Omega)$. Then there exist constants depending on β, δ, μ^* and k, but independent of h such that*

$$\begin{aligned} \mu^*\|\hat{u} - u_h\|_0^2 + (\hat{u} - u_h)^2(b) + \delta h\|\hat{u}' - u_h'\|_0^2 &\leq Ch\|\hat{u}\|_{\mathbf{H}^1}^2 \\ &\leq Ch\left(\|f_h\|_0^2 + |\xi|^2\right) \end{aligned} \tag{2.3.90}$$

where μ^* is the constant in the coerciveness assumption 2.3.41.

Proof. The difference $(\hat{u} - u_h)$ satisfies the following equation

$$\begin{cases} \displaystyle\int_\Omega \left[\frac{\beta}{\lambda_h(x)}(\hat{u} - u_h) + (\hat{u} - u_h)'\right][\varphi + \delta h \varphi'] = 0 \qquad \forall \varphi \in V_a^h(\Omega) \\ (\hat{u} - u_h)(a) = 0 \end{cases} \tag{2.3.91}$$

Let $\Pi_h^k \hat{u}$ be the interpolant of $\hat{u}$ in $V^h(\Omega)$ (notice that, since $\Omega \subset \mathbf{R}$, we have $H^1(\Omega) \subset C^0(\Omega)$, and the interpolant is well-defined for any $k \geq 1$); if we choose $\varphi = (\Pi_h^k \hat{u} - u_h)$, which belongs to $V_a^h(\Omega)$, we have

$$\begin{aligned} 0 = &\int_\Omega \left[\frac{\beta}{\lambda_h(x)}(\hat{u} - u_h) + (\hat{u} - u_h)'\right]\left[(\Pi_h^k\hat{u} - u_h) + \delta h(\Pi_h^k\hat{u} - u_h)'\right] = \\ = &\int_\Omega \frac{\beta}{\lambda_h(x)}(\hat{u} - \Pi_h^k\hat{u})(\Pi_h^k\hat{u} - u_h) + \int_\Omega \frac{\beta}{\lambda_h(x)}(\Pi_h^k\hat{u} - u_h)^2 + \int_\Omega (\hat{u} - \Pi_h^k\hat{u})'(\Pi_h^k\hat{u} - u_h) + \\ &+ \int_\Omega (\Pi_h^k\hat{u} - u_h)'(\Pi_h^k\hat{u} - u_h) + \delta h \int_\Omega \left[(\Pi_h^k\hat{u} - u_h)'\right]^2 + \delta h \int_\Omega (\hat{u} - \Pi_h^k\hat{u})'(\Pi_h^k\hat{u} - u_h)' \\ &+ \int_\Omega \frac{\beta\delta h}{\lambda_h(x)}(\Pi_h^k\hat{u} - u_h)(\Pi_h^k\hat{u} - u_h)' + \int_\Omega \frac{\beta\delta h}{\lambda_h(x)}(\hat{u} - \Pi_h^k\hat{u})(\Pi_h^k\hat{u} - u_h)', \end{aligned}$$

i.e.,

$$\begin{aligned} &\int_\Omega \frac{\beta}{\lambda_h(x)}(\Pi_h^k\hat{u} - u_h)^2 + \delta h \int_\Omega \left[(\Pi_h^k\hat{u} - u_h)'\right]^2 + \int_\Omega \left(1 + \frac{\beta\delta h}{\lambda_h(x)}\right)(\Pi_h^k\hat{u} - u_h)(\Pi_h^k\hat{u} - u_h)' = \\ &= \int_\Omega \left[\frac{\beta}{\lambda_h(x)}(\Pi_h^k\hat{u} - \hat{u}) + (\Pi_h^k\hat{u} - \hat{u})'\right]\left[(\Pi_h^k\hat{u} - u_h) + \delta h(\Pi_h^k\hat{u} - u_h)'\right]. \end{aligned} \tag{2.3.92}$$

Let us focus on the left hand side in (2.3.92): an integration by parts of the third term, together with the fact that $(\Pi_h^k \hat{u} - u_h)(a) = 0$, provides:

$$\begin{aligned}
&\int_\Omega \left(1 + \frac{\beta\delta h}{\lambda_h(x)}\right)(\Pi_h^k\hat{u} - u_h)(\Pi_h^k\hat{u} - u_h)' = \\
&\qquad = \left[\frac{1}{2}\left(1 + \frac{\beta\delta h}{\lambda_h(x)}\right)(\Pi_h^k\hat{u} - u_h)^2\right]_a^b + \frac{1}{2}\int_\Omega \frac{\beta\delta h}{\lambda_h^2(x)}\,\lambda_h'(x)\,(\Pi_h^k\hat{u} - u_h)^2 \\
&\qquad = \frac{1}{2}\left(1 + \frac{\beta\delta h}{\lambda_h(b)}\right)(\Pi_h^k\hat{u} - u_h)^2(b) + \frac{1}{2}\int_\Omega \frac{\beta\delta h}{\lambda_h^2(x)}\,\lambda_h'(x)\,(\Pi_h^k\hat{u} - u_h)^2.
\end{aligned}$$

Thus, we have the following estimate for the left hand side

$$\begin{aligned}
&\int_\Omega \frac{\beta}{\lambda_h(x)}(\Pi_h^k\hat{u} - u_h)^2 \;+\; \delta h\int_\Omega \left[(\Pi_h^k\hat{u} - u_h)'\right]^2 + \int_\Omega \left(1 + \frac{\beta\delta h}{\lambda_h(x)}\right)(\Pi_h^k\hat{u} - u_h)(\Pi_h^k\hat{u} - u_h)' \\
&\qquad = \int_\Omega \left(\frac{\beta}{\lambda_h(x)} + \frac{1}{2}\frac{\beta\delta h}{\lambda_h^2(x)}\,\lambda_h'(x)\right)(\Pi_h^k\hat{u} - u_h)^2 \;+\; \delta h\int_\Omega \left[(\Pi_h^k\hat{u} - u_h)'\right]^2 \\
&\qquad\quad + \frac{1}{2}\left(1 + \frac{\beta\delta h}{\lambda_h(b)}\right)(\Pi_h^k\hat{u} - u_h)^2(b) \qquad\qquad (2.3.93)\\
&\qquad \geq \mu^*\|\Pi_h^k\hat{u} - u_h\|_0^2 + \delta h\|(\Pi_h^k\hat{u} - u_h)'\|_0^2 + \frac{1}{2}(\Pi_h^k\hat{u} - u_h)^2(b).
\end{aligned}$$

where the inequality stems from the coerciveness assumption (2.3.41) and the positiveness of β, δ, h, and λ_h.
Now, let us consider the right hand side in (2.3.92): we have

$$\begin{aligned}
&\int_\Omega \left[\frac{\beta}{\lambda_h(x)}(\Pi_h^k\hat{u} - \hat{u}) + (\Pi_h^k\hat{u} - \hat{u})'\right]\left[(\Pi_h^k\hat{u} - u_h) + \delta h(\Pi_h^k\hat{u} - u_h)'\right] = \\
&\quad = \underbrace{\int_\Omega \frac{\beta}{\lambda_h(x)}\Big(\Pi_h^k\hat{u} - \hat{u}\Big)\Big(\Pi_h^k\hat{u} - u_h\Big)}_{(1)} + \underbrace{\int_\Omega (\Pi_h^k\hat{u} - \hat{u})'(\Pi_h^k\hat{u} - u_h)}_{(2)} + \underbrace{\int_\Omega \frac{\beta\delta h}{\lambda_h(x)}(\Pi_h^k\hat{u} - \hat{u})(\Pi_h^k\hat{u} - u_h)'}_{(3)} \\
&\qquad \underbrace{+\,\delta h\int_\Omega (\Pi_h^k\hat{u} - \hat{u})'(\Pi_h^k\hat{u} - u_h)'}_{(4)}.
\end{aligned}$$

So far, we focus on the terms in the above summation. By standard arguments, we have for the first one:

$$(1) \leq \frac{\beta}{\min_\Omega|\lambda_h|}\int_\Omega \left|(\Pi_h^k\hat{u} - \hat{u})\,(\Pi_h^k\hat{u} - u_h)\right| \leq \frac{\beta}{\lambda_*}\left(\alpha_1\left\|\Pi_h^k\hat{u} - \hat{u}\right\|_0^2 + \frac{1}{4\alpha_1}\left\|\Pi_h^k\hat{u} - u_h\right\|_0^2\right),$$

where $\alpha_1 > 0$ is a constant. Since the interpolant $\Pi_h^k \hat{u}$ coincides with $\hat{u}$ on the nodes of the mesh, an integration by parts of the second term, leads

$$\begin{aligned}(2) &= \Big[(\Pi_h^k \hat{u} - \hat{u})(\Pi_h^k \hat{u} - u_h)\Big]_a^b - \int_\Omega (\Pi_h^k \hat{u} - \hat{u})(\Pi_h^k \hat{u} - u_h)' \\ &= \int_\Omega (\hat{u} - \Pi_h^k \hat{u})(\Pi_h^k \hat{u} - u_h)' \leq \alpha_2 \delta h \left\|(\Pi_h^k \hat{u} - u_h)'\right\|_0^2 + \frac{1}{4\alpha_2 \delta h}\left\|\Pi_h^k \hat{u} - \hat{u}\right\|_0^2\end{aligned}$$

with $\alpha_2 > 0$ constant. We then have, for the third term:

$$(3) \leq \frac{\beta \delta h}{\min_\Omega \lambda_h}\int_\Omega \left|(\Pi_h^k \hat{u} - \hat{u})\,(\Pi_h^k \hat{u} - u_h)'\right| \leq \frac{\beta\delta h}{\lambda_*}\left(\alpha_3 \left\|\Pi_h^k \hat{u} - \hat{u}\right\|_0^2 + \frac{1}{4\alpha_3}\left\|(\Pi_h^k \hat{u} - u_h)'\right\|_0^2\right),$$

with $\alpha_3 > 0$ constant. Finally, we have for the last term:

$$(4) \leq \alpha_4 \delta h \left\|(\Pi_h^k \hat{u} - u_h)'\right\|_0^2 + \frac{\delta h}{4\alpha_4}\left\|(\Pi_h^k \hat{u} - \hat{u})'\right\|_0^2$$

with $\alpha_4 > 0$ constant.
We thus have

$$\begin{aligned}\mu^* \|\Pi_h^k \hat{u} - u_h\|_0^2 + \delta h \|(\Pi_h^k \hat{u} - u_h)'\|_0^2 + \frac{1}{2}\left(1 + \frac{\beta\delta h}{\lambda_h(b)}\right)(\Pi_h^k \hat{u} - u_h)^2(b) \leq \\ \leq \left(\frac{\beta}{\lambda_*}\alpha_1 + \frac{1}{4\alpha_2 \delta h} + \frac{\beta \delta h}{\lambda_*}\right)\|\Pi_h^k \hat{u} - \hat{u}\|_0^2 + \delta h \alpha_4 \|(\Pi_h^k \hat{u} - \hat{u})'\|_0^2 \\ + \frac{\beta}{\lambda_*}\frac{1}{4\alpha_1}\|\Pi_h^k \hat{u} - u_h\|_0^2 + \left(\alpha_2 + \frac{\beta}{\lambda_*}\frac{1}{4\alpha_3} + \frac{1}{4\alpha_4}\right)\delta h \|(\Pi_h^k \hat{u} - u_h)'\|_0^2\end{aligned}$$

For any α_j $(j = 1, .., 4)$ such that

$$\left(\alpha_2 + \frac{\beta}{\lambda_*}\frac{1}{4\alpha_3} + \frac{1}{4\alpha_4}\right) \leq \frac{1}{2}$$

and

$$\frac{\beta}{\lambda_*}\frac{1}{4\alpha_1} \leq \frac{\mu^*}{2},$$

we get

$$\begin{aligned}\frac{\mu^*}{2}\|\Pi_h^k \hat{u} - u_h\|_0^2 + \frac{\delta h}{2}\|(\Pi_h^k \hat{u} - u_h)'\|_0^2 + \frac{1}{2}\left(1 + \frac{\beta\delta h}{\lambda_h(b)}\right)(\Pi_h^k \hat{u} - u_h)^2(b) \leq \\ \leq \underbrace{\left(\frac{\beta}{\lambda_*}\alpha_1 + \frac{1}{4\alpha_2 \delta h} + \frac{\beta \delta h}{\lambda_*}\right)}_{\sim Ch^{-1}}\|\Pi_h^k \hat{u} - \hat{u}\|_0^2 + \delta h \alpha_4 \|(\Pi_h^k \hat{u} - \hat{u})'\|_0^2\end{aligned}$$

$$\leq Ch^{-1}\|\Pi_h^k\hat{u}-\hat{u}\|_0^2 + C\delta h\|(\Pi_h^k\hat{u}-\hat{u})'\|_0^2$$

$$\leq Ch\|\hat{u}\|_{H^1(\Omega)}^2,$$

where the last inequality follows from standard interpolation estimates for finite elements. Finally, using the fact that $\Pi_h^k\hat{u}(b)=\hat{u}(b)$, we can conclude

$$\mu^*\|\hat{u}-u_h\|_0^2 + \delta h\|(\hat{u}-u_h)'\|_0^2 + (\hat{u}-u_h)^2(b) \leq$$

$$\leq 2\mu^*\|\Pi_h^k\hat{u}-u_h\|_0^2 + 2\delta h\|(\Pi_h^k\hat{u}-u_h)'\|_0^2 + (\Pi_h^k\hat{u}-u_h)^2(b)$$

$$+2\mu^*\|\hat{u}-\Pi_h^k\hat{u}\|_0^2 + 2\delta h\|(\Pi_h^k\hat{u}-\hat{u})'\|_0^2$$

$$\leq 2\left(\mu^*+Ch^{-1}\right)\|\hat{u}-\Pi_h^k\hat{u}\|_0^2 + 2C\delta h\|(\Pi_h^k\hat{u}-\hat{u})'\|_0^2$$

$$\leq Ch\|\hat{u}\|_{H^1(\Omega)}^2,$$

and this concludes the proof. □

We can now consider the difference between the solution u of problem (2.3.84) and the solution u_h of problem (2.3.88). We can prove the following estimate.

Lemma 2.3.7 *Let u and u_h be the solutions of problems* (2.3.84) *and* (2.3.87), *respectively. Assume that* (2.3.41) *is satisfied, that f, $f_h \in L^2(\Omega)$, λ, $\lambda_h \in L^\infty(\Omega)$ and $\lambda(x), \lambda_h(x) \geq \lambda_* > 0$ for all $x \in \Omega$. Then, the following error estimate holds*

$$\mu^*\|u-u_h\|_0^2 + \delta h\|u'-u_h'\|_0^2 \leq$$

$$\leq Ch\Big(\|f_h\|_0 + \xi^2\Big) + C\Big(\|f-f_h\|_0^2 + \|\lambda-\lambda_h\|_0^2\Big) \tag{2.3.94}$$

where C is a constant depending on β, Ω, λ_*, δ and k, but independent of h.

Proof. We have

$$\mu^*\|u-u_h\|_0^2+\delta h\|u'-u_h'\|_0^2 \leq \mu^*\|u-\hat{u}\|_0^2+\mu^*\|\hat{u}-u_h\|_0^2+\delta h\|u'-\hat{u}'\|_0^2+\delta h\|\hat{u}'-u_h'\|_0^2 \tag{2.3.95}$$

where $\hat{u}$ is the solution of problem (2.3.88). The bound for the second and fourth term in the right hand side is given by Lemma 2.3.7 above. We therefore focus on the other two terms. For simplicity of notations, we set in the following

$$g(x) := \beta \int_a^x \frac{dy}{\lambda(y)} \quad \text{and} \quad g_h(x) := \beta \int_a^x \frac{dy}{\lambda_h(y)}. \tag{2.3.96}$$

Since $\lambda(x), \lambda_h(x) \geq \lambda_* > 0$ for all $x \in \Omega$, the functions $g(x)$ and $g_h(x)$ are positive and monotone increasing ($0 = g(a) < g(x) < g(b)$ and $0 = g_h(a) < g_h(x) < g_h(b)$, $\forall x \in \Omega$)
Up to a constant, we have by means of standard arguments

$$\begin{aligned}
\|u-\hat{u}\|_0^2 &= \int_\Omega \left| \xi(e^{-g(x)} - e^{-g_h(x)}) + e^{-g(x)} \int_a^x e^{g(s)} f(s)\, ds \;-\; e^{-g_h(x)} \int_a^x e^{g_h(s)} f_h(s)\, ds \right|^2 dx \\
&\leq \int_\Omega \left(\left| \xi(e^{-g(x)} - e^{-g_h(x)}) \right| + \left| e^{-g(x)} \int_a^x e^{g(s)} f(s)\, ds \;-\; e^{-g_h(x)} \int_a^x e^{g_h(s)} f_h(s)\, ds \right| \right)^2 dx \\
&\leq \int_\Omega \left(\left| \xi(e^{-g(x)} - e^{-g_h(x)}) \right| \right)^2 dx + \int_\Omega \left(\left| e^{-g(x)} \int_a^x e^{g(s)} f(s)\, ds \;-\; e^{-g_h(x)} \int_a^x e^{g_h(s)} f_h(s)\, ds \right| \right)^2 dx \\
&\leq \xi^2 \int_\Omega \left| e^{-g(x)} - e^{-g_h(x)} \right|^2 dx + \int_\Omega \Bigg(\left| e^{-g(x)} - e^{-g_h(x)} \right| \int_a^x \left| e^{g(s)} f(s) \right| ds \\
&\quad + \left| e^{-g_h(x)} \right| \int_a^x \left| (e^{g(s)} - e^{g_h(s)}) f(s) \right| ds + \left| e^{-g_h(x)} \right| \int_a^x \left| e^{g_h(s)} (f(s) - f_h(s)) \right| ds \Bigg)^2 dx \\
&\leq \xi^2 \int_\Omega \left| e^{-g(x)} - e^{-g_h(x)} \right|^2 dx + \int_\Omega \left(\left| e^{-g(x)} - e^{-g_h(x)} \right| \int_a^x \left| e^{g(s)} f(s) \right| ds \right)^2 dx \\
&\quad + \int_\Omega \left(e^{-g_h(x)} \int_a^x \left| e^{g(s)} - e^{g_h(s)} \right| |f(s)|\, ds \right)^2 dx + \int_\Omega \left(e^{-g_h(x)} \int_a^x e^{g_h(s)} \; |f(s) - f_h(s)|\, ds \right)^2 dx \\
&\leq \xi^2 \int_\Omega \left| e^{-g(x)} - e^{-g_h(x)} \right|^2 dx + \int_\Omega \left(\left| e^{-g(x)} - e^{-g_h(x)} \right| \int_\Omega \left| e^{g(s)} f(s) \right| ds \right)^2 dx \\
&\quad + \int_\Omega \left(e^{-g_h(x)} \int_\Omega \left| e^{g(s)} - e^{g_h(s)} \right| |f(s)|\, ds \right)^2 dx + \int_\Omega \left(e^{-g_h(x)} \int_\Omega e^{g_h(s)} \; |f(s) - f_h(s)|\, ds \right)^2 dx \\
&= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.
\end{aligned}$$

Focusing on the first term we immediately have

$$\mathrm{I} \leq \xi^2 \int_\Omega |g(x) - g_h(x)|^2 \, dx.$$

In order to estimate the terms II, III, and IV above, we observe that the bounds on $g(x)$ and $g_h(x)$ entail

$$1 \leq e^{g(x)} \leq e^{\frac{\beta|\Omega|}{\min_\Omega \lambda}}, \qquad 1 \leq e^{g_h(x)} \leq e^{\frac{\beta|\Omega|}{\min_\Omega \lambda_h}}.$$

Using the above estimate, as well as Jensen's inequality, we obtain for the second term:

$$\mathrm{II} \leq \int_\Omega \left| e^{-g(x)} - e^{-g_h(x)} \right| \, |\Omega| e^{\frac{2\beta|\Omega|}{\lambda_*}} \, \|f\|_0^2 \, dx \leq |\Omega|^2 \, e^{\frac{2\beta|\Omega|}{\lambda_*}} \, \|f\|_0^2 \int_\Omega |g(x) - g_h(x)|^2 \, dx.$$

A direct application of Hölder inequality provides for the third term

$$\mathrm{III} \leq \int_\Omega \left(\left\| e^{-g(x)} - e^{-g_h(x)} \right\|_0 \|f\|_0 \right)^2 \leq |\Omega| \, \|f\|_0^2 \int_\Omega |g(x) - g_h(x)|^2 \, dx.$$

Finally, again using Jensen's inequality in the last term, we get

$$\mathrm{IV} \leq \int_\Omega \left(\int_\Omega e^{g_h(s)} \, |f(s) - f_h(s)| \, ds \right)^2 dx \leq \int_\Omega |\Omega| \, e^{\frac{2\beta|\Omega|}{\lambda_*}} \, \|f - f_h\|_0^2 \, dx \leq |\Omega|^2 \, e^{\frac{2\beta|\Omega|}{\lambda_*}} \, \|f - f_h\|_0^2.$$

We therefore have

$$\|u - \hat{u}\|_0^2 \leq \left(\xi^2 + |\Omega|(1 + e^{\frac{2\beta|\Omega|}{\lambda_*}}) \, \|f\|_0^2 \right) \int_\Omega |g(x) - g_h(x)|^2 \, dx + |\Omega|^2 e^{\frac{2\beta|\Omega|}{\lambda_*}} \|f - f_h\|_0^2.$$

Observing that

$$\int_\Omega |g(x) - g_h(x)|^2 \, dx = \int_\Omega \left| \beta \int_a^x \left(\frac{1}{\lambda(s)} - \frac{1}{\lambda_h(s)} \, ds \right) \right|^2 dx \leq \int_\Omega \left(\beta \int_a^x \left| \frac{\lambda_h(s) - \lambda(s)}{\lambda(s)\lambda_h(s)} \right| ds \right)^2 dx$$
$$\leq \int_\Omega \left(\beta \int_\Omega \frac{|\lambda_h(s) - \lambda(s)|}{\lambda(s)\lambda_h(s)} \, ds \right)^2 \leq \int_\Omega \left(\frac{\beta^2}{\lambda_*^4} |\Omega| \int_\Omega |\lambda(s) - \lambda_h(s)|^2 \, ds \right) dx \leq \left[\frac{\beta \, |\Omega|}{\lambda_*^2} \right]^2 \|\lambda - \lambda_h\|_0^2 \, ,$$

we get, up to a constant

$$\|u - \hat{u}\|_0^2 \leq |\Omega|^2 e^{\frac{2\beta|\Omega|}{\lambda_*}} \left(\|f - f_h\|_0^2 + \|\lambda - \lambda_h\|_0^2 \right). \tag{2.3.97}$$

Since u and $\hat{u}$ are the solutions of equations (2.3.84) and (2.3.88), we have, up to a multiplicative factor

$$\begin{aligned}
\|u' - \hat{u}'\|_0^2 &= \int_\Omega \left| (f - f_h) + \frac{\beta}{\lambda\lambda_h} (\lambda\hat{u} - \lambda_h u) \right|^2 dx \\
&\leq \int_\Omega |f - f_h|^2 \, dx + \int_\Omega \left(\frac{\beta}{\lambda\lambda_h} \right)^2 |\lambda\hat{u} - \lambda_h u|^2 \, dx \\
&\leq \|f - f_h\|_0^2 + \left(\frac{\beta}{\min_\Omega \lambda \, \min_\Omega \lambda_h} \right)^2 \int_\Omega |\lambda\hat{u} - \lambda_h u|^2 \, dx \\
&\leq \|f - f_h\|_0^2 + \left[\frac{\beta}{\lambda_*^2} \right]^2 \left(\int_\Omega (|\lambda - \lambda_h| \, |\hat{u}|)^2 \, dx + \int_\Omega (|\lambda_h| \, |u - \hat{u}|^2 \, dx \right).
\end{aligned}$$

Since f_h is bounded in the L^2 norm, the same holds true for $\hat{u}$, and this entails, owing to the Hölder inequality,

$$\|u' - \hat{u}'\|_0^2 \leq \|f - f_h\|_0^2 + \frac{\beta^2}{\lambda_*^4}\left(\|\hat{u}\|_0^2 \, \|\lambda - \lambda_h\|_0^2 + \|\lambda_h\|_0^2 \, \|u - \hat{u}\|_0^2\right). \tag{2.3.98}$$

Gathering together estimates (2.3.90), (2.3.95), (2.3.97) and (2.3.98), the thesis follows. □

In a similar way it is not difficult to prove the corresponding result when the transport term is negative for each $x \in \Omega$, and the boundary condition is given in $x = b$.

Lemma 2.3.8 *Let v and v_h be the solutions to problems*

$$\begin{cases} \dfrac{\beta}{\zeta(x)}\, v + v' = f(x) & \text{in } \Omega = (a, b) \\ v(b) = \eta \end{cases} \tag{2.3.99}$$

and

Find $v_h \in V^h(\Omega)$ such that

$$\begin{cases} \displaystyle\int_\Omega \left[\frac{\beta}{\zeta_h(x)}\, v_h + v_h' - f_h\right] [\varphi + \delta h \varphi'] = 0 & \forall \varphi \in V_0^h(\Omega) \\ v_h(b) = \eta. \end{cases} \tag{2.3.100}$$

respectively, with $\beta > 0$, $\zeta(x)$, $\zeta_h(x) \leq \zeta_* < 0$ for each $x \in \Omega$. Assume that

$$\mu_* := \min_{x\in\Omega} \left| \frac{\beta}{\zeta_h(x)} - \frac{1}{2}\frac{\beta\delta h \zeta_h'(x)}{\zeta_h^2(x)} \right| > 0, \tag{2.3.101}$$

and that f, f_h, ζ, $\zeta_h \in L^\infty(\Omega)$. Then, there exist a constant C depending on β, δ, μ_*, Ω, ζ_* and k, but independent of h such that

$$\begin{aligned} &\mu_* \|v - v_h\|_0^2 + \delta h \|v' - v_h'\|_0^2 \leq \\ &\leq Ch\Big(\|f_h\|_0^2 + \eta^2\Big) + C\Big(\|f - f_h\|_0^2 + \|\zeta - \zeta_h\|_0^2\Big). \end{aligned} \tag{2.3.102}$$

□

Error estimates for the primitive variables

In this section we derive an energy estimate for the FEM approximation through the characteristic approach. Since our main attention focused on the spatial domain decomposition, we give in this section an estimate of the difference between the exact solution at time t^n, $\mathbf{U}(t^n, x)$, and the approximate one stemming from the characteristic approach. Since $\mathbf{V}_1 = u + \frac{2}{\gamma-1}c$ and $\mathbf{V}_2 = u - \frac{2}{\gamma-1}c$, the inverse change of variable is

$$\mathbf{U}(t,x) = \left\{ \begin{array}{l} \rho(t,x) = F_1(\mathbf{V}(t,x)) = \left\{ \dfrac{1}{K\gamma} \left[\dfrac{\gamma-1}{4}(\mathbf{V}_1 - \mathbf{V}_2)(t,x) \right]^2 \right\}^{1/\gamma-1} \\ \\ u(t,x) = F_2(\mathbf{V}(t,x)) = \dfrac{1}{2}(\mathbf{V}_1 + \mathbf{V}_2)(t,x). \end{array} \right. \tag{2.3.103}$$

Due to the nonlinearity of the change of variables, when we map the discretized (either in time or in both time and space) characteristic variables back to the primitive ones, we do not obtain the solution of a discretized version of the original problem in the primitive variables. However, we expect these resulting functions to be a good approximation of the primitive variables. Under these considerations, we denote with $\mathbf{V}^n$ the solution, at time step n, of the single domain problem discretized in time as in Section 2.3.1,

$$\beta\mathbf{V}^n + \Lambda^{n-1}\mathbf{V}^n_x = \beta\mathbf{V}^{n-1}, \tag{2.3.104}$$

where $\Lambda^{n-1} = \text{diag}(\lambda_1^{n-1}, \lambda_2^{n-1})$, as defined therein, and we define, with a little abuse of notation,

$$\mathbf{U}^n(x) := F(\mathbf{V}^n(x)) = \left\{ \begin{array}{l} \rho^n(x) = F_1(\mathbf{V}^n(x)) \\ \\ u^n(x) = F_2(\mathbf{V}^n(x)), \end{array} \right. \tag{2.3.105}$$

and

$$\mathbf{U}^n_h(x) := F(\mathbf{V}^n_h(x)) = \left\{ \begin{array}{l} \rho^n_h(x) = F_1(\mathbf{V}^n_h(x)) \\ \\ u^n_h(x) = F_2(\mathbf{V}^n_h(x)). \end{array} \right. \tag{2.3.106}$$

where $\mathbf{V}^n_h(x)$ is the fully discrete approximation of $\mathbf{V}(t^n, x)$ via the Streamline Diffusion FEM. We are in position to prove the following result.

Lemma 2.3.9 *Let* $\mathbf{U}(t^n, x)$ *and* $\mathbf{V}(t^n, x)$ *be the solutions of problems* (2.3.2) *and* (2.3.6) *respectively, at time* $t = t^n$, *and* $\mathbf{V}^n_h(x)$ *be the solution of problem* (2.3.58) *at time step* n. *Assume that,* $\mathbf{U}(t^n, x), \mathbf{V}(t^n, x) \in L^2(\Omega)$, *and that* $\mathbf{V}^{n-1}(x) \in L^\infty(\Omega)$. *Assume moreover that* $\lambda^* \geq |\lambda_j^{n-1}(x)|$, $|\lambda_{hj}^{n-1}(x)| \geq \lambda_* > 0$ *(*$j = 1, 2$*), for all* $x \in \Omega$, *that* $\gamma < 3$, *and that* (2.3.41) *and* (2.3.101) *are satisfied. Then, at time step* n, *the following error estimate holds*

$$\left\|\mathbf{U}(t^n, x) - \mathbf{U}^n_h(x)\right\|_0^2 \leq C \left\|\mathbf{V}(t^n, x) - \mathbf{V}^n(x)\right\|_0^2$$
$$+C\left\|\mathbf{V}^{n-1}(x) - \mathbf{V}^{n-1}_h(x)\right\|_0^2 + Ch\left(\left\|\mathbf{V}^{n-1}_h(x)\right\|_0^2 + |g_1(t^n)|^2 + |g_2(t^n)|^2\right),$$

where $g_1(t^n)$ *and* $g_2(t^n)$ *are the boundary conditions in* (2.3.3) *for* $t = t^n$, *and where the constant* C *may depend on* β, δ, Ω, *and* λ_*, *but is independent of* h.

Proof. First of all, notice that, under our assumptions, the function $\mathbf{V}^n$, solution of problem (2.3.104), belongs to $L^2(\Omega)$. Then, since

$$\left\|\mathbf{U}(t^n,x)-\mathbf{U}_h^n(x)\right\|_0^2=\left\|\mathbf{U}_1(t^n,x)-\mathbf{U}_{h,1}^n(x)\right\|_0^2+\left\|\mathbf{U}_2(t^n,x)-\mathbf{U}_{h,2}^n(x)\right\|_0^2,$$

we have to analyze both terms in the above summation. Concerning the first one, we observe that, since $\gamma<3$, the function $F_1(.)$ is Lipschitz continuous, with Lipschitz constant that we indicate with L_1. We set, for simplicity of notations, $\mathcal{C}:=\left[(\gamma-1)^2/(16K\gamma)\right]^{1/\gamma-1}$, and we obtain

$$\begin{aligned}
\int_\Omega\left|\mathbf{U}_1(t^n,x)-\mathbf{U}_{h,1}^n(x)\right|^2&=\int_\Omega\left|\mathcal{C}\,\left[\mathbf{V}_1(t^n,x)-\mathbf{V}_2(t^n,x)\right]^{\frac{2}{\gamma-1}}-\mathcal{C}\,\left[\mathbf{V}_{h,1}^n(x)-\mathbf{V}_{h,2}^n(x)\right]^{\frac{2}{\gamma-1}}\right|^2\\
&\le\mathcal{C}^2\int_\Omega\left|\left[\mathbf{V}_1(t^n,x)-\mathbf{V}_2(t^n,x)\right]^{\frac{2}{\gamma-1}}-\left[\mathbf{V}_{h,1}^n(x)-\mathbf{V}_{h,2}^n(x)\right]^{\frac{2}{\gamma-1}}\right|^2\\
&\le\mathcal{C}^2L_1^2\int_\Omega\left|\left[\mathbf{V}_1(t^n,x)-\mathbf{V}_2(t^n,x)\right]-\left[\mathbf{V}_{h,1}^n(x)-\mathbf{V}_{h,2}^n(x)\right]\right|^2\\
&\le 2\,\mathcal{C}^2L_1^2\left(\int_\Omega\left|\mathbf{V}_1(t^n,x)-\mathbf{V}_{h,1}^n(x)\right|^2+\int_\Omega\left|\mathbf{V}_2(t^n,x)-\mathbf{V}_{h,2}^n(x)\right|^2\right)\\
&\le K\left\|\mathbf{V}_1(t^n,x)-\mathbf{V}_1^n(x)\right\|_0^2+K\left\|\mathbf{V}_1^n(x)-\mathbf{V}_{h,1}^n(x)\right\|_0^2+K\left\|\mathbf{V}_2(t^n,x)-\mathbf{V}_2^n(x)\right\|_0^2\\
&\quad+K\left\|\mathbf{V}_2^n(x)-\mathbf{V}_{h,2}^n(x)\right\|_0^2,
\end{aligned}$$

where we have set $K:=4\,\mathcal{C}^2L_1^2$.
The linearity of $F_2(.)$ allows a simpler treatment of the second term:

$$\begin{aligned}
\int_\Omega\left|\mathbf{U}_2(t^n,x)-\mathbf{U}_{h,2}^n(x)\right|^2&=\int_\Omega\left|\frac{1}{2}\left[\mathbf{V}_1(t^n,x)+\mathbf{V}_2(t^n,x)\right]-\frac{1}{2}\left[\mathbf{V}_{h,1}^n(x)+\mathbf{V}_{h,2}^n(x)\right]\right|^2\\
&\le\frac{1}{2}\left(\int_\Omega\left|\mathbf{V}_1(t^n,x)-\mathbf{V}_{h,1}^n(x)\right|^2+\int_\Omega\left|\mathbf{V}_2(t^n,x)-\mathbf{V}_{h,2}^n(x)\right|^2\right)\\
&\le\frac{1}{2}\left\|\mathbf{V}_1(t^n,x)-\mathbf{V}_1^n(x)\right\|_0^2+\frac{1}{2}\left\|\mathbf{V}_1^n(x)-\mathbf{V}_{h,1}^n(x)\right\|_0^2+\frac{1}{2}\left\|\mathbf{V}_2(t^n,x)-\mathbf{V}_2^n(x)\right\|_0^2\\
&\quad+\frac{1}{2}\left\|\mathbf{V}_2^n(x)-\mathbf{V}_{h,2}^n(x)\right\|_0^2.
\end{aligned}$$

Therefore, there exists a constant $\mathcal{K}=\max\{1/2,4\mathcal{C}^2L^2\}$ such that

$$\left\|\mathbf{U}(t^n,x)-\mathbf{U}_h^n(x)\right\|_0^2\le\mathcal{K}\left(\left\|\mathbf{V}(t^n,x)-\mathbf{V}^n(x)\right\|_0^2+\left\|\mathbf{V}^n(x)-\mathbf{V}_h^n(x)\right\|_0^2\right)$$

Under our assumptions, we are in the position to use the estimates of the previous section, with $\bar{\mathbf{f}} = \beta\Lambda^{-1}\mathbf{V}^{n-1}$, and $\bar{\mathbf{f}}_h$ as in (2.3.58), and we get from Lemma 2.3.7 and 2.3.8

$$\begin{aligned}\left\|\mathbf{U}(t^n,x) - \mathbf{U}_h^n(x)\right\|_0^2 &\leq C\left\|\mathbf{V}(t^n,x) - \mathbf{V}^n(x)\right\|_0^2 \\ &+Ch\left(\left\|\bar{\mathbf{f}}_h\right\|_0^2 + |g_1(t^n)|^2 + |g_2(t^n)|^2\right) + C\left(\left\|\bar{\mathbf{f}} - \bar{\mathbf{f}}_h\right\|_0^2 + \left\|\bar{\lambda}^{n-1} - \bar{\lambda}_h^{n-1}\right\|_0^2\right)\end{aligned}$$

where $\bar{\lambda}^{n-1} = (\lambda_1^{n-1}, \lambda_2^{n-1})$, $\bar{\lambda}_h^{n-1} = (\lambda_{h1}^{n-1}, \lambda_{h2}^{n-1})$. Owing to (2.3.7) and (2.3.81), we easily have

$$\left\|\bar{\lambda}^{n-1} - \bar{\lambda}_h^{n-1}\right\|_0^2 \leq |||C|||^2\left\|\mathbf{V}^{n-1}(x) - \mathbf{V}_h^{n-1}(x)\right\|_0^2,$$

where C is the matrix in (2.3.7) and (2.3.81), whereas $|||.|||^2$ is any compatible matrix norm. We have

$$\|\bar{\mathbf{f}}_h\|_0^2 \leq \frac{\beta^2}{\lambda_*^2}\left\|\mathbf{V}^{n-1}\right\|_0^2,$$

as well as, for $j = 1, 2$

$$\begin{aligned}\left\|\bar{\mathbf{f}}_j - \bar{\mathbf{f}}_{h,j}\right\|_0^2 &= \beta^2\left\|(\lambda_j^{n-1})^{-1}\mathbf{V}_j^{n-1} - (\lambda_{h,j}^{n-1})^{-1}\mathbf{V}_{h,j}^{n-1}\right\|_0^2 \leq \beta^2\left\|\frac{1}{\lambda_j^{n-1}\lambda_{h,j}^{n-1}}\left(\lambda_{h,j}^{n-1}\mathbf{V}_j^{n-1} - \lambda_j^{n-1}\mathbf{V}_{h,j}^{n-1}\right)\right\|_0^2 \\ &\leq \frac{\beta^2}{\lambda_*^4}\left(\left\|\mathbf{V}_j^{n-1}(x)\right\|_\infty^2 \|\lambda_j^{n-1} - \lambda_{h,j}^{n-1}\|_0^2 + \|\lambda_j^{n-1}\|_\infty^2\left\|\mathbf{V}_j^{n-1} - \mathbf{V}_{h,j}^{n-1}\right\|_0^2\right).\end{aligned}$$

Thus,

$$\left\|\bar{\mathbf{f}} - \bar{\mathbf{f}}_h\right\|_0^2 \leq C\ \left\|\mathbf{V}^{n-1} - \mathbf{V}_h^{n-1}\right\|_0^2,$$

and this concludes the proof.

□

Remark 2.3.4 A few comments on Lemma 2.3.9 are in order. The assumption that the exact solution belongs to $L^2(\Omega)$, at time $t = t^n$, in both primitive ($\mathbf{U}$) and characteristic form ($\mathbf{V}$), is not that restrictive in the region of smooth flow. The bounds on the modulus of the time discrete λ_j^{n-1}, and fully discrete $\lambda_{h,j}^{n-1}$, $j = 1, 2$, approximations of the eigenvalues $u + c$ and $u - c$ are justified by the assumption we made on the flow to be subsonic. Finally, since for ideal gases the ratio $\gamma \sim 5/3$, the assumption on γ is not restrictive either. □

An immediate corollary of the above Lemma is the following.

Corollary 2.3.1 *Assume that*

$$\lim_{h\to 0}\left\|\mathbf{V}_h^{n-1}(x) - \mathbf{V}^{n-1}(x)\right\|_0^2 = 0$$

uniformly in h. Then, we have

$$\lim_{h\to 0}\left\|\mathbf{U}(t^n,x)-\mathbf{U}_h^n(x)\right\|_0^2 \leq \left\|\mathbf{V}(t^n,x)-\mathbf{V}^n(x)\right\|_0^2.$$

Remark 2.3.5 The above results are valid at time step n, and our attention was mainly paid to the convergence of the iteration-by-subdomain procedure. Further work needs however to be done in order to link the approximation error at the *n-th* time step to the initial condition $\mathbf{U}(0,x)$. □

2.4 The complete 1-D Euler System

The flow of an ideal inviscid compressible polytropic gas in one space dimension is governed by the complete one-dimensional Euler system,

$$\frac{\partial \mathbf{W}}{\partial t}+\frac{\partial \mathbf{F}(\mathbf{W})}{\partial x}=\mathbf{0} \quad \text{in } Q_T := \Omega\times(0,T), \tag{2.4.1}$$

$\Omega=(a,b)$ being an interval, where the vector of conserved variables is $\mathbf{V}=(\rho,\rho u,\rho E)$, as usual ρ being the density of the fluid, u its velocity, and E its energy per unit mass. The inviscid flux vector $\mathbf{F}(\mathbf{W})$ can therefore be written as

$$\mathbf{F}(\mathbf{W})=(\rho u,\rho u^2+p,(\rho E+p)u).$$

For ideal polytropic gases, the pressure p and the internal energy ϵ are related to the other thermodynamic quantities ρ and ϑ through the equations of state

$$p=R\rho\vartheta, \qquad \epsilon=c_V\vartheta,$$

where $R>0$ is the difference between the specific heat at constant pressure $c_P>0$ and the specific heat at constant volume $c_V>0$. These relations entail

$$p=(\gamma-1)\rho\epsilon,$$

with $\gamma>1$ being as usual the ratio of specific heats. Moreover, from (TD2) we obtain

$$p=k\rho^\gamma\exp(s/c_V), \tag{2.4.2}$$

for a suitable constant $K>0$.

In a region of smooth flow, using (TD2) to express the derivatives of ϵ in terms of s and ρ, the quasi-linear form of (2.4.1) in terms of the vector of primitive variables $\mathbf{U}=(\rho,u,s)$ reads

$$\frac{\partial \mathbf{U}}{\partial t}+A(\mathbf{U})\frac{\partial \mathbf{U}}{\partial x}=0 \quad \text{in } Q_T := \Omega\times(0,T) \tag{2.4.3}$$

where the matrix A is given by

$$A(\mathbf{U}) := \begin{pmatrix} u & \rho & 0 \\ c^2/\rho & u & p_s/\rho \\ 0 & 0 & u \end{pmatrix},$$

where $c = \sqrt{\frac{\partial p}{\partial \rho}}$ is the speed of sound, and $p_s := \frac{\partial p}{\partial s}$.
The matrix A is diagonalizable with distinct real eigenvalues (thus system (2.4.3) is strictly hyperbolic), namely $A = L^{-1}\Lambda L$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$, with

$$\lambda_1 = u + c, \quad \lambda_2 = u - c, \quad \lambda_3 = u$$

while L is the matrix of left eigenvectors, given by

$$L := \begin{pmatrix} c/\rho & 1 & p_s/(\rho c) \\ -c/\rho & 1 & -p_s/(\rho c) \\ 0 & 0 & 1 \end{pmatrix}.$$

For subsonic flows, in which the left endpoint of the interval is the upstream boundary, i.e. $0 < u < c$, the eigenvalues λ_1 and λ_3 are positive, while λ_2 is negative, whereas if the flow is supersonic $(u > c > 0)$, all the eigenvalues are positive. Since in this latter case the whole information travels from the left endpoint to the right one, this would trivially reduce an iteration-by-subdomains approach, based on Dirichlet transmission conditions, to a sequential solution of equation (2.4.3) from the left end subdomain to the right end one. Thus, we make again the assumption that the flow is subsonic.
In principle, since the matrix A is diagonalizable, system (2.4.3) can be transformed, using the left eigenvectors of A, into a fully decoupled problem. Similarly to the isentropic case, this can be accomplished introducing the characteristic variables, which are defined by means of the differential form (see [62])

$$d\mathbf{Y} := Ld\mathbf{U} = (\frac{c}{\rho}d\rho + du + \frac{p_s}{\rho c}ds, -\frac{c}{\rho}d\rho + du - \frac{p_s}{\rho c}ds, ds), \tag{2.4.4}$$

so that system (2.4.3) becomes

$$\frac{\partial \mathbf{Y}}{\partial t} + \Lambda\frac{\partial \mathbf{Y}}{\partial x} = 0 \quad \text{in } Q_T := \Omega \times (0,T)$$

and splits into three scalar equations

$$\frac{\partial \mathbf{Y}_j}{\partial t} + \lambda_j\frac{\partial \mathbf{Y}_j}{\partial x} = 0 \quad \text{in } Q_T := \Omega \times (0,T), \quad j = 1,2,3,$$

coupled only through the boundary conditions. This entails that, for $j = 1,2,3$, the component $\mathbf{Y}_j$ is constant along the characteristic curve

$$C_j = \left\{(x_j(t), t) \mid x_j'(t) = \lambda_j\right\}.$$

Unfortunately, though the matrix A is diagonalizable, the field (2.4.4) is not irrotational, and the characteristic variables cannot be explicitly determined, except for the entropy s, which is constant along the characteristic curve $C_0 = \{(x(t),t) \mid x'(t) = u\}$. As a clear consequence, we are not able to completely decouple system (2.4.3) into a system of scalar equations as in the isentropic case, and we have to compensate the lack of knowledge on the characteristic variables by choosing other variables to fulfill the interface continuity requirements in a domain decomposition setting.

A first opportunity, under the assumption that no eigenvalue of A is zero, is to enforce the continuity on the interface of the variables

$$\mathbf{Z} = L\mathbf{U} = (u + c + \frac{p_s}{\rho c}s, u - c - \frac{p_s}{\rho c}s, s), \tag{2.4.5}$$

and iterate accordingly, even if they are not characteristic variables (we can call them *pseudo-characteristic* variables). This amounts to consider the functions $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}_3$ associated with the eigenvalues $u + c$, $u - c$ and u respectively. In the subsonic case, the interface conditions in the iterative process become (subindices denote components, while superindices denote the subdomain and the iteration step)

$$\mathbf{Z}_2^{1,k+1} = \mathbf{Z}_2^{2,k} \quad \text{and} \quad \begin{cases} \mathbf{Z}_1^{2,k+1} = \mathbf{Z}_1^{1,k} \\ \\ \mathbf{Z}_3^{2,k+1} = \mathbf{Z}_3^{1,k} \end{cases}$$

since the characteristic curve C_2, associated with the eigenvalue $u - c$ is incoming in Ω_1, while the characteristic curves C_1 and C_3, associated with the eigenvalues $u + c$ and u respectively, are incoming in Ω_2. We are therefore prescribing for each subdomain a boundary condition at the interface for each variable associated to an incoming characteristic line.

An iteration by subdomain procedure of Dirichlet-Dirichlet type based on these transmission conditions for the problem continuous in space arising from a semi-implicit time discretisation can be shown to converge, provided the inverse of the time step $\beta = 1/\Delta t$ is sufficiently large.

Another opportunity relies on the Riemann invariants for the isentropic case: in the polytropic case the functions R_+ and R_-, defined in (2.3.5), are no longer constant along the characteristic lines $C_j = \{(x(t),t) \,|\, x'(t) = \lambda_j\}$, $j = 1, 2$, nevertheless we can enforce on the interface the continuity of a set of variables (we can call them *pseudo-Riemann invariants*), defined by the Riemann invariants of the isentropic case and the entropy s, which is the only characteristic variable that can be explicitly determined from (2.4.4), namely

$$\mathbf{V} := (R_+, R_-, s), \tag{2.4.6}$$

associated with the eigenvalues $u + c$, $u - c$ and u respectively.

In both cases, it is straightforward to see that the continuity across the interface of the variables $\mathbf{Z}$ or $\mathbf{V}$ entails the continuity of the inviscid flux, given in (2.2.3).

Remark 2.4.1 When considering discretisation, at the interface point x_Γ one has to enforce three additional conditions (besides the other three related to $\mathbf{Z}$ or $\mathbf{V}$), in order to recover all the six interface variables, and this can be accomplished by imposing to the variables $\mathbf{U}$ to satisfy equation (2.4.3) at the interface point x_Γ for any outgoing component. If the flow is subsonic, we have to impose two additional conditions for Ω_1 and one for Ω_2. An opportunity is to multiply equation (2.4.3) on the left by the matrix L and consider the components corresponding to the outgoing eigenvectors. Observing that, for $r = 1, 2, 3$,

$$\Big(L\left[\mathbf{U}_t + A(\mathbf{U})\mathbf{U}_x\right]\Big)_r = \Big(L\mathbf{U}_t + LA\mathbf{U}_x\Big)_r = \Big(L\mathbf{U}_t + \Lambda L\mathbf{U}_x\Big)_r = L^{(r)} \cdot \left[\mathbf{U}_t + \lambda_r \mathbf{U}_x\right],$$

where $L^{(r)}$ denotes the r-th column of the matrix L, this amounts to take, for any outgoing component, the scalar product at the interface point x_Γ between the left eigenvectors of A ($\mathbf{l}^r$, $r = 1, 2, 3$), and the equation. Thus, for sake of simplicity, we enforce the equations

$$\begin{aligned} &\left[\mathbf{l}^r \cdot \left(\frac{\partial \mathbf{U}_1}{\partial t} + \lambda_r \frac{\partial \mathbf{U}_1}{\partial x}\right)\right](x_\Gamma, t) = 0 \qquad \text{for } r = 1, 3 \\ &\left[\mathbf{l}^2 \cdot \left(\frac{\partial \mathbf{U}_2}{\partial t} + \lambda_2 \frac{\partial \mathbf{U}_2}{\partial x}\right)\right](x_\Gamma, t) = 0, \end{aligned} \tag{2.4.7}$$

with

$$\mathbf{l}^1 := \left(\frac{c}{\rho}, 1, \frac{p_s}{\rho c}\right), \quad \mathbf{l}^2 := \left(-\frac{c}{\rho}, 1, -\frac{p_s}{\rho c}\right), \quad \mathbf{l}^3 := (0, 0, 1).$$

Notice that, in the case of an hyperbolic system with constant coefficients, equations (2.4.7) above correspond to the natural choice of imposing the equation for the outgoing characteristic variable to be satisfied at the interface point x_Γ. Equations (2.4.7) can therefore be seen as a direct generalization of the constant coefficients case and are sometimes called the *compatibility* equations. □

Now, let us consider what happens in the framework of this latter opportunity. Owing to the diagonalisation of A, we can write (2.4.3) as

$$L\mathbf{U}_t + \Lambda L\mathbf{U}_x = 0. \tag{2.4.8}$$

We cannot actually decouple system (2.4.8) into a system of scalar equations, but we can evaluate the difference between the time derivative of the variables $\mathbf{V}$ and the quantity $L\mathbf{U}_t$ as well as the difference between the space derivative of $\mathbf{V}$ and $L\mathbf{U}_x$. Let us set

$$\begin{cases} \mathbf{R}^t := \mathbf{V}_t - L\mathbf{U}_t \\ \mathbf{R}^x := \mathbf{V}_x - L\mathbf{U}_x \end{cases} \tag{2.4.9}$$

So far, owing to (2.4.9), equation (2.4.8) becomes

$$\mathbf{V}_t + \Lambda \mathbf{V}_x = \mathbf{R}^t + \Lambda \mathbf{R}^x \tag{2.4.10}$$

Noticing that such a formulation for equation (2.4.3) can be achieved with any change of variables, let us see what happens with our choice. We have

$$L\mathbf{U}_t = \begin{bmatrix} \frac{c}{\rho}\rho_t + u_t + \frac{p_s}{\rho c}s_t \\ \\ -\frac{c}{\rho}\rho_t + u_t - \frac{p_s}{\rho c}s_t \\ \\ s_t \end{bmatrix}, \quad \text{and} \quad L\mathbf{U}_x = \begin{bmatrix} \frac{c}{\rho}\rho_x + u_x + \frac{p_s}{\rho c}s_x \\ \\ -\frac{c}{\rho}\rho_x + u_x - \frac{p_s}{\rho c}s_x \\ \\ s_x \end{bmatrix} \tag{2.4.11}$$

as well as

$$\mathbf{V}_t = \begin{bmatrix} u_t + \frac{2}{\gamma - 1}c_t \\ \\ u_t - \frac{2}{\gamma - 1}c_t \\ \\ s_t \end{bmatrix}, \quad \text{and} \quad \mathbf{V}_x = \begin{bmatrix} u_x + \frac{2}{\gamma - 1}c_x \\ \\ u_x - \frac{2}{\gamma - 1}c_x \\ \\ s_x \end{bmatrix}. \tag{2.4.12}$$

Therefore, taking into account the expression of c as well as (2.4.9) we can easily get

$$\mathbf{R}^t = \begin{bmatrix} \frac{1}{\gamma - 1}\frac{p_s}{\rho c}s_t \\ \\ -\frac{1}{\gamma - 1}\frac{p_s}{\rho c}s_t \\ \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{R}^x = \begin{bmatrix} \frac{1}{\gamma - 1}\frac{p_s}{\rho c}s_x \\ \\ -\frac{1}{\gamma - 1}\frac{p_s}{\rho c}s_x \\ \\ 0 \end{bmatrix} \tag{2.4.13}$$

Remark 2.4.2 As we could have expected since the entropy s is effectively a Riemann invariant, the third components of the "*rests*" $\mathbf{R}^t$ and $\mathbf{R}^x$ are zero. Moreover, since the only difference with respect to the isentropic case is that the entropy is no longer constant, the first two components of $\mathbf{R}^t$ and $\mathbf{R}^x$ depend only on s_t and s_x respectively. □

Setting $C_\gamma := 1/(\gamma - 1)$, system (2.4.10) reads

$$\begin{cases} (\mathbf{V}_1)_t + \lambda_1(\mathbf{V}_1)_x &= \quad C_\gamma \frac{p_s}{\rho c}[s_t + \lambda_1 s_x] \\ \\ (\mathbf{V}_2)_t + \lambda_2(\mathbf{V}_2)_x &= -C_\gamma \frac{p_s}{\rho c}[s_t + \lambda_2 s_x] \\ \\ (\mathbf{V}_3)_t + \lambda_3(\mathbf{V}_3)_x &= \quad 0 \end{cases} \tag{2.4.14}$$

In our framework $\mathbf{V}_3 = s$; using $(2.4.14)_3$ and taking into account the fact that $\lambda_1 - \lambda_3 = c$ as well as $\lambda_2 - \lambda_3 = -c$, we finally get the system

$$\left\{\begin{array}{rcl} (\mathbf{V}_1)_t + \lambda_1(\mathbf{V}_1)_x &=& K_\gamma(\mathbf{V}_3)_x \\ (\mathbf{V}_2)_t + \lambda_2(\mathbf{V}_2)_x &=& K_\gamma(\mathbf{V}_3)_x \\ (\mathbf{V}_3)_t + \lambda_3(\mathbf{V}_3)_x &=& 0 \end{array}\right. \tag{2.4.15}$$

where we have set $K_\gamma := C_\gamma(p_s/\rho)$, which is completely equivalent to (2.4.3).

2.4.1 Domain Decomposition

Letting

$$\mathbf{K} := \begin{bmatrix} 0 & 0 & K_\gamma \\ 0 & 0 & K_\gamma \\ 0 & 0 & 0 \end{bmatrix} \tag{2.4.16}$$

we can write system (2.4.15) into a matrix form:

$$\mathbf{V}_t + \mathbf{T}_\Lambda \mathbf{V}_x = 0. \tag{2.4.17}$$

where we have set $\mathbf{T}_\Lambda := \Lambda - \mathbf{K}$. Owing again to (2.3.7), system (2.4.17) reads

$$\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \end{bmatrix}_t + \begin{pmatrix} \sum_{j=1}^{2} C_{1j}\mathbf{V}_j & 0 & \alpha(\mathbf{V}_1 - \mathbf{V}_2)^2 \\ 0 & \sum_{j=1}^{2} C_{2j}\mathbf{V}_j & \alpha(\mathbf{V}_1 - \mathbf{V}_2)^2 \\ 0 & 0 & (\mathbf{V}_1 + \mathbf{V}_2)/2 \end{pmatrix} \cdot \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \end{bmatrix}_x = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.4.18}$$

where

$$\alpha = \frac{1-\gamma}{16 C_V \gamma}.$$

So far, letting as usual $\alpha \in (a, b)$ and setting $\Omega_1 := (a, \alpha)$ as well as $\Omega_2 := (\alpha, b)$, we introduce a domain decomposition method to solve system (2.4.17) where we enforce the continuity on the interface of the variables $\mathbf{V}$. It reads

$$\left\{\begin{array}{ll} \mathbf{V}_t^1 + \mathbf{T}_\Lambda \mathbf{V}_x^1 = 0 & \text{in } \Omega_1 \times (0, T) \\ \mathbf{V}_t^2 + \mathbf{T}_\Lambda \mathbf{V}_x^2 = 0 & \text{in } \Omega_2 \times (0, T) \\ \mathbf{V}^1(\alpha, t) = \mathbf{V}^2(\alpha, t) & \forall t \in (0, T), \end{array}\right. \tag{2.4.19}$$

considered with the occurring boundary conditions.
We can advance in time system (2.4.19) in order to obtain a problem continuous in space: so far, two rather different situations may occur, according to the boundary condition on the left endpoint if the interval $x = a$.

Case 1: $s(a)$ known

If the value of the entropy in $x = a$ can be effectively determined, which is the case of physical boundary conditions among the following ones

$$\begin{cases} u(a) \\ \\ s(a) \end{cases} \qquad \begin{cases} \rho(a) \\ \\ s(a), \end{cases} \tag{2.4.20}$$

(notice that the case when $\rho(a)$ and $p(a)$ are given, owing to (2.4.2), belongs to such a framework), we can advance in time system (2.4.19) by means of a semi-implicit method, obtaining

$$\begin{cases} \beta \mathbf{V}^{1,n+1} + \mathbf{T}^n_\Lambda \mathbf{V}^{1,n+1}_x = \beta \mathbf{V}^{1,n} = f^{1,n} & \text{in } \Omega_1 \\ \\ \beta \mathbf{V}^{2,n+1} + \mathbf{T}^n_\Lambda \mathbf{V}^{2,n+1}_x = \beta \mathbf{V}^{2,n} = f^{2,n} & \text{in } \Omega_2 \\ \\ \mathbf{V}^{1,n+1}(\alpha) = \mathbf{V}^{2,n+1}(\alpha), \end{cases} \tag{2.4.21}$$

where $\beta = 1/\Delta t$ is the inverse of the time step, and the boundary conditions must be satisfied at time step $n+1$..
At each time step, the third equation is completely decoupled from the others, so it can be solved once for all in the whole spatial domain Ω. This guarantees the continuity of $\mathbf{V}_3$ on the interface. Moreover, it can be used in the right hand side in the first two equations, and we get a reduced system. Indeed, defining in each subdomain the two dimensional vector functions

$$\begin{aligned} F^{1,n+1} &:= \begin{bmatrix} K_\gamma(\mathbf{V}^{1,n+1}_3)_x + f^{1,n}_1 \\ \\ K_\gamma(\mathbf{V}^{1,n+1}_3)_x + f^{1,n}_2 \end{bmatrix} \qquad \text{in } \Omega_1 \\ \\ F^{2,n+1} &:= \begin{bmatrix} K_\gamma(\mathbf{V}^{2,n+1}_3)_x + f^{2,n}_1 \\ \\ K_\gamma(\mathbf{V}^{2,n+1}_3)_x + f^{2,n}_2 \end{bmatrix} \qquad \text{in } \Omega_2, \end{aligned} \tag{2.4.22}$$

and dropping any index referring to the time discretisation, we get the system

$$\begin{cases} \beta \bar{\mathbf{V}}^1 + \bar{\Lambda} \bar{\mathbf{V}}^1_x = F^1 & \text{in } \Omega_1 \\ \\ \beta \bar{\mathbf{V}}^2_1 + \bar{\Lambda} \bar{\mathbf{V}}^2_x = F^2 & \text{in } \Omega_2 \\ \\ \bar{\mathbf{V}}^1(\alpha) = \bar{\mathbf{V}}^2(\alpha), \end{cases} \tag{2.4.23}$$

where we have set $\bar{\mathbf{V}}^i := (\mathbf{V}^i_1, \mathbf{V}^i_2)$, $i = 1, 2$, and $\bar{\Lambda} := \text{diag}\,(\lambda_1, \lambda_2)$, and where the remaining boundary conditions are only the ones for $\mathbf{V}_1$ and $\mathbf{V}_2$.
So far, at each time step, we reduce ourselves to the isentropic case with an additional forcing term, given by the derivative of the entropy at the same time step. An iteration-by-subdomains procedure can thus be used to solve the coupled problem, and its convergence analysis is exactly the same as in the case of isentropic flows.

Case 2: $s(a)$ **unknown**

If the value of the entropy in $x = a$ cannot be determined, the third equation in (2.4.15) is no longer decoupled from the others, and cannot be used as right hand side in the previous ones. We can therefore choose a different type of time discretization, which is almost explicit

$$\begin{cases} \beta \mathbf{V}^{1,n+1} + \Lambda^n \mathbf{V}_x^{1,n+1} = \beta \mathbf{V}^{1,n} + \mathbf{K} \cdot \mathbf{V}_x^{1,n} = f^{1,n} & \text{in } \Omega_1 \\ \beta \mathbf{V}^{2,n+1} + \Lambda^n \mathbf{V}_x^{2,n+1} = \beta \mathbf{V}^{2,n} + \mathbf{K} \cdot \mathbf{V}_x^{2,n} = f^{2,n} & \text{in } \Omega_2 \\ \mathbf{V}^{1,n+1}(\alpha) = \mathbf{V}^{2,n+1}(\alpha), \end{cases} \tag{2.4.24}$$

where $\mathbf{K}$ is the matrix in (2.4.16), while $\beta = 1/\Delta t$ is the inverse of the time step, and boundary conditions have to be satisfied at time step $n+1$. So far, at each time step we use in the right hand side of the first two equations the space derivative of the entropy at the previous one, but a few considerations are in order.

Since we have at both $(n+1)$-*th* and n-*th* step spatial derivatives of the same order, such a discretisation may suffer of serious instabilities. The stability of this scheme is at the moment an open problem and will not be discussed here. At the discrete level, however, this amounts to the necessity of taking into account a CFL-type condition, in order to overcome this stability drawback.

An iteration by subdomains procedure can then be introduced to solve, at each time step, system (2.4.24). Dropping any index referring to time discretization, at the $(k+1)$-*th* iteration step, the solutions $\mathbf{V}^{1,k+1}$ and $\mathbf{V}^{2,k+1}$ satisfy

$$\begin{cases} \beta \mathbf{V}^{i,k+1} + \Lambda \ \mathbf{V}_x^{i,k+1} = f^i & \text{in } \Omega_i, \ i = 1,2 \\ \mathbf{V}_1^{2,k+1}(\alpha) = \mathbf{V}_1^{1,k}(\alpha), \quad \mathbf{V}_j^{1,k+1}(\alpha) = \mathbf{V}_j^{2,k}(\alpha) & j = 1,3. \end{cases} \tag{2.4.25}$$

The resulting system is diagonal, and a convergence analysis for the problem discretized in time but continuous in space can be accomplished, following the lines of the isentropic case: the iteration-by-subdomains algorithm can be proved to converge under some restriction on the time step.

Theorem 2.4.1 *The iteration-by-subdomains method in* (2.4.25) *is convergent provided* β *is sufficiently large.*

Proof. We define, for each subdomain, the error vector

$$\mathbf{E}^{i,k} := \mathbf{V}^i - \mathbf{V}^{i,k}$$

for $i = 1,2$, which satisfies the following error equation

$$\begin{cases} \beta \mathbf{E}^{i,k+1} + \Lambda \mathbf{E}_x^{i,k+1} = 0 & \text{in } \Omega_i, \ i = 1,2 \\ \mathbf{E}_1^{2,k+1}(\alpha) = \mathbf{E}_1^{1,k}(\alpha), \quad \mathbf{E}_j^{1,k+1}(\alpha) = \mathbf{E}_j^{2,k}(\alpha) & j = 1,3, \end{cases} \tag{2.4.26}$$

Physical Boundary Conditions	Boundary Conditions for the Error Equation
$\rho(a),\ u(a),\ u(b)$	$\mathbf{E}_1^k(a) + \mathbf{E}_2^k(a) = 0$ $\mathbf{E}_2^k(a) = \mathbf{E}_1^k(a) - \frac{4}{\gamma - 1}\sqrt{K\gamma\rho^{\gamma-1}}(a)\left(e^{s^k(a)/2c_V} - e^{s(a)/2c_V}\right)$ $\mathbf{E}_1^k(b) + \mathbf{E}_2^k(b) = 0$
$\rho(a),\ u(a),\ \rho(b)$	$\mathbf{E}_1^k(a) + \mathbf{E}_2^k(a) = 0$ $\mathbf{E}_2^k(a) = \mathbf{E}_1^k(a) - \frac{4}{\gamma - 1}\sqrt{K\gamma\rho^{\gamma-1}}(a)\left(e^{s^k(a)/2c_V} - e^{s(a)/2c_V}\right)$ $\mathbf{E}_2^k(b) = \mathbf{E}_1^k(b) - \frac{4}{\gamma - 1}\sqrt{K\gamma\rho^{\gamma-1}}(b)\left(e^{s^k(b)/2c_V} - e^{s(b)/2c_V}\right)$

Table 2.1: Boundary Conditions for the Error Equation in Case 2.

with boundary conditions depending on the ones for the primitive variables and are reported in Table 2.1. Owing to (2.4.26), we define the interface error at the $(k+1)$-th iteration step as

$$\mathbf{E}_\alpha^{k+1} := \left(\mathbf{E}_1^{1,k+1}(\alpha),\ \mathbf{E}_2^{2,k+1}(\alpha),\ \mathbf{E}_3^{1,k+1}(\alpha)\right).$$

The solutions of system (2.4.26) are given by

$$\mathbf{E}^{1,k+1}(x) = \begin{bmatrix} \mathbf{E}_1^{1,k+1}(a)\, e^{-\beta\Phi^{(1)}(x)} \\ \mathbf{E}_2^{2,k}(\alpha)\, e^{-\beta\Psi^{(1)}(x)} \\ \mathbf{E}_3^{1,k+1}(a)\, e^{-\beta\Xi^{(1)}(x)} \end{bmatrix} \qquad \mathbf{E}^{2,k+1}(x) = \begin{bmatrix} \mathbf{E}_1^{1,k}(\alpha)\, e^{-\beta\Phi^{(2)}(x)} \\ \mathbf{E}_2^{2,k+1}(b)\, e^{-\beta\Psi^{(2)}(x)} \\ \mathbf{E}_3^{1,k}(\alpha)\, e^{-\beta\Xi^{(2)}(x)} \end{bmatrix}$$

where we have set

$$\left\{ \begin{array}{l} \Phi^{(1)}(x) := \displaystyle\int_a^x \frac{dy}{\lambda_1(y)} \\ \\ \Phi^{(2)}(x) := \displaystyle\int_\alpha^x \frac{dy}{\lambda_1(y)} \end{array} \right. , \quad \left\{ \begin{array}{l} \Psi^{(1)}(x) := \displaystyle\int_\alpha^x \frac{dy}{\lambda_2(y)} \\ \\ \Psi^{(2)}(x) := \displaystyle\int_b^x \frac{dy}{\lambda_2(y)} \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} \Xi^{(1)}(x) := \displaystyle\int_a^x \frac{dy}{\lambda_3(y)} \\ \\ \Xi^{(2)}(x) := \displaystyle\int_\alpha^x \frac{dy}{\lambda_3(y)} \end{array} \right. .$$

Recalling that, from the subsonic assumption, $\lambda_1(x) > 0$, $\lambda_2(x) < 0$ and $\lambda_3(x) > 0$ for all $x \in \Omega$, and proceeding as in the previous section, we immediately have

$$\left[\mathbf{E}_1^{1,k+1}(\alpha)\right]^2 = \left[\mathbf{E}_2^{2,k}(\alpha)\right]^2\ e^{-2\beta\left[\Phi^{(1)}(\alpha)+\Psi^{(1)}(a)\right]},$$

and a direct application of Lagrange's theorem provides

$$\mathbf{E}_1^{1,k+1}(a) - \mathbf{E}_2^{1,k+1}(a) = \frac{2\sqrt{K\gamma\rho^{\gamma-1}(a)}}{C_V(\gamma-1)}\, e^{\xi_0/2C_V}\,\mathbf{E}_3^{1,k+1}(a)$$

with $\xi_0 \in \left[\min\left\{s(a), s^{k+1}(a)\right\},\ \max\left\{s(a), s^{k+1}(a)\right\}\right]$, and this entails

$$\left[\mathbf{E}_3^{1,k+1}(\alpha)\right]^2 = [C_1(a)]^2\ \left[\mathbf{E}_2^k(\alpha)\right]^2\ e^{-2\beta\left[\Xi^{(1)}(\alpha)+\Psi^{(1)}(a)\right]}$$

where we have set $C_1(a) := -\frac{C_V(\gamma-1)}{\sqrt{K\gamma\rho^{\gamma-1}(a)}}\ e^{-\xi_0/2C_V}$.

If $u(b)$ is given, we have

$$\left[\mathbf{E}_2^{2,k+1}(\alpha)\right]^2 = \left[\mathbf{E}_1^{1,k}(\alpha)\right]^2\ e^{-2\beta\left[\Phi^{(2)}(b)+\Psi^{(2)}(\alpha)\right]},$$

whereas, if $\rho(b)$ is given, we have from Lagrange's theorem

$$\mathbf{E}_1^{k+1}(b) - \mathbf{E}_2^{k+1}(b) = \frac{2\sqrt{K\gamma\rho^{\gamma-1}(b)}}{C_V\,(\gamma-1)}\, e^{\xi_1/2C_V}\ \mathbf{E}_3^{k+1}(b)$$

where $\xi_1 \in \left[\min\left\{s(b), s^{k+1}(b)\right\},\ \max\left\{s(b), s^{k+1}(b)\right\}\right]$, which implies

$$\left[\mathbf{E}_2^{2,k+1}(\alpha)\right]^2 \leq \left[\mathbf{E}_1^{1,k}(\alpha)\right]^2\ e^{-2\beta\left[\Phi^{(2)}(b)+\Psi^{(2)}(\alpha)\right]} +\ [C_2(b)]^2\ \left[\mathbf{E}_3^{1,k}(\alpha)\right]^2\ e^{-2\beta\left[\Xi^{(2)}(b)+\Psi^{(2)}(\alpha)\right]}$$

where we have set $C_2(b) := -\frac{2\sqrt{K\gamma\rho^{\gamma-1}(b)}}{C_V\,(\gamma-1)}\, e^{\xi_1/2C_V}$. At each iteration step, we thus have, if $u(b)$ is given,

$$\left|\mathbf{E}_\alpha^{k+1}\right|^2 = \mathcal{K}_1(\beta)\ \left[\mathbf{E}_1^{1,k}(\alpha)\right]^2 + \mathcal{K}_2(\beta)\ \left[\mathbf{E}_2^{2,k}(\alpha)\right]^2,$$

where we have set

$$\mathcal{K}_1(\beta) := e^{-2\beta\left[\Phi^{(2)}(b)+\Psi^{(2)}(\alpha)\right]},\quad \mathcal{K}_2(\beta) := \left[e^{-2\beta\left[\Phi^{(1)}(\alpha)+\Psi^{(1)}(a)\right]} + [C_1(a)]^2\ e^{-2\beta\left[\Xi^{(1)}(\alpha)+\Psi^{(1)}(a)\right]}\right],$$

whereas, if $\rho(b)$ is given,

$$\left|\mathbf{E}_\alpha^{k+1}\right|^2 \leq \mathcal{K}_1(\beta)\ \left[\mathbf{E}_1^{1,k}(\alpha)\right]^2 + \mathcal{K}_2(\beta)\ \left[\mathbf{E}_2^{2,k}(\alpha)\right]^2 + \mathcal{K}_3(\beta)\ \left[\mathbf{E}_3^{1,k}(\alpha)\right]^2$$

where

$$\mathcal{K}_3(\beta) := [C_2(b)]^2\ e^{-2\beta\left[\Xi^{(2)}(b)+\Psi^{(2)}(\alpha)\right]}.$$

Noticing that $\mathcal{K}_1(\beta) < 1$, the iterative mapping on the interface is a contraction provided β is large enough to have

$$\max\left\{\mathcal{K}_2(\beta),\ \mathcal{K}_3(\beta)\right\}\ <\ 1,$$

i.e., provided

$$\beta > \max\left\{ -\frac{\log\left[\frac{1}{1+[C_1(a)]^2}\right]}{2\left[\Phi^{(1)}(\alpha)+\Psi^{(1)}(a)\right]}, -\frac{\log\left[\frac{1}{[C_2(b)]^2}\right]}{2\left[\Psi^{(2)}(\alpha)+\Xi^{(2)}(b)\right]} \right\}$$

With an argument similar to the one in the previous section, we can show that, for each $x \in \Omega$, the error $\mathbf{E}^k(x)$ is controlled, in the Euclidean norm, by the error on the interface, ensuring convergence in the whole spatial domain and concluding the proof. □

Remark 2.4.3 The above results suffers evidently of a couple of severe drawbacks. First of all, no stability analysis has been made for the time integration scheme, and the procedure could blow up. Secondly, the resulting algorithm is proved to converge if β satisfies a lower bound: this can actually be a quite restrictive upper bound on the time step Δt, which might be forced to be very close to zero, causing an unsustenable increase of the computational cost at the discrete level. □

2.5 Three dimensional flows

In this last section we present some algorithm proposed in literature for compressible flows in the three dimensional space. The three-dimensional Euler equations in the region of smooth flow, can be written in quasi-linear form,with respect to the primitive variables $\mathbf{U} = (\rho, u_1, u_2, u_3, s)$, as

$$\frac{\partial \mathbf{U}}{\partial t} + \sum_{k=1}^{3} A_k D_k \mathbf{U} = \mathbf{0} \qquad \text{in } \Omega \times (0,T), \tag{2.5.1}$$

where

$$A_1 := \begin{pmatrix} u_1 & \rho & 0 & 0 & 0 \\ c^2/\rho & u_1 & 0 & 0 & p_s/\rho \\ 0 & 0 & u_1 & 0 & 0 \\ 0 & 0 & 0 & u_1 & 0 \\ 0 & 0 & 0 & 0 & u_1 \end{pmatrix}, \quad A_2 := \begin{pmatrix} u_2 & 0 & \rho & 0 & 0 \\ 0 & u_2 & 0 & 0 & 0 \\ c^2/\rho & 0 & u_2 & 0 & p_s/\rho \\ 0 & 0 & 0 & u_2 & 0 \\ 0 & 0 & 0 & 0 & u_2 \end{pmatrix},$$

and

$$A_3 := \begin{pmatrix} u_3 & 0 & 0 & \rho & 0 \\ 0 & u_3 & 0 & 0 & 0 \\ 0 & 0 & u_3 & 0 & 0 \\ c^2/\rho & 0 & 0 & u_3 & p_s/\rho \\ 0 & 0 & 0 & 0 & u_3 \end{pmatrix},$$

where we have set $c := \sqrt{\frac{\partial p}{\partial \rho}}$ and $p_s := \frac{\partial p}{\partial s}$.

As usual, we partition the domain Ω into two non-overlapping subdomains Ω_1 and Ω_2, and we denote with Γ the interface. For any point $\mathbf{x} \in \Gamma$ and any time $t \in (0,T)$, we denote by $C = C(\mathbf{n})$ the characteristic matrix $C = \sum_k \mathbf{n}_k A_k$, $\mathbf{n}$ being the unit vector normal to Γ directed from Ω_1 to Ω_2. The eigenvalues of C are given by

$$\lambda_1 = \mathbf{u}\cdot\mathbf{n} + c, \quad \lambda_2 = \mathbf{u}\cdot\mathbf{n} - c, \quad \lambda_{3,4,5} = \mathbf{u}\cdot\mathbf{n}, \tag{2.5.2}$$

and, since they are real but not distinct, system (2.5.1) is not strictly hyperbolic. We finally, denote by L the matrix of the left eigenvectors of C, which is given by

$$L := \begin{pmatrix} \frac{c}{\rho} & n_1 & n_2 & n_3 & \frac{p_s}{\rho c} \\ -\frac{c}{\rho} & n_1 & n_2 & n_3 & -\frac{p_s}{\rho c} \\ 0 & \tau_1^{(1)} & \tau_2^{(1)} & \tau_3^{(1)} & 1 \\ 0 & \tau_1^{(2)} & \tau_2^{(2)} & \tau_3^{(2)} & 1 \\ 0 & -\tau_1^{(1)} - \tau_1^{(2)} & -\tau_2^{(1)} - \tau_2^{(2)} & -\tau_3^{(1)} - \tau_3^{(2)} & 1 \end{pmatrix}, \qquad (2.5.3)$$

where $\tau^{(1)}$ and $\tau^{(2)}$ are two unit orthogonal vectors, spanning the plane orthogonal to $\mathbf{n}$.
Assuming that at time t the interface Γ is not characteristic at point $\mathbf{x}$, namely, that no eigenvalue is zero at $\mathbf{x}$, A. Quarteroni and A. Valli proposed in [89] matching conditions which are naturally extrapolated from the one-dimensional case. The idea is thus to enforce the continuity of the '*characteristic*' variables $L\mathbf{U}$, and reads as follows

$$\sum_{q=1}^{5} L_{rq}\mathbf{U}_{1,q} = \sum_{q=1}^{5} L_{rq}\mathbf{U}_{2,q} \qquad \text{at } \mathbf{x} \in \Gamma, \quad r = 1, \dots, 5. \qquad (2.5.4)$$

It can be easily verified that, as a consequence of (2.5.4), the interface conditions (2.2.3) are satisfied.
In order to have well-posed problems, the iteration-by-subdomain algorithm used for solving the multi-domain problem alternates the solution of the Euler equations (2.5.1) in Ω_1 and in Ω_2, with the Dirichlet boundary condition (2.5.4) imposed at a point $\mathbf{x}$ on Γ for all indices r corresponding to incoming characteristic lines. For instance, if we assume that at time t the interface point $\mathbf{x}$ is an outflow point for Ω_1 and that the flow is subsonic (namely, $0 < \mathbf{u} \cdot \mathbf{n} < c$, with $\mathbf{n}$ directed from Ω_1 to Ω_2), one has to impose at the $(m+1)$-th iteration

$$\sum_{q=1}^{5} L_{2q}\mathbf{U}_{1,q}^{m+1} = \sum_{q=1}^{5} L_{2q}\mathbf{U}_{2,q}^{m} \qquad \text{at } \mathbf{x} \in \Gamma,$$

and

$$\sum_{q=1}^{5} L_{kq}\mathbf{U}_{2,q}^{m+1} = \sum_{q=1}^{5} L_{kq}\mathbf{U}_{1,q}^{m} \qquad \text{at } \mathbf{x} \in \Gamma, \quad k = 1, 3, 4, 5.$$

When a numerical discretisation is applied, the compatibility equations, described in Remark 2.4.1 have to be imposed at $\mathbf{x}$. Proceeding as in (2.4.7), in each subdomain Ω_i, $i = 1, 2$, they are obtained by taking the scalar product of (2.5.1) (stated for $\mathbf{U}_i^{m+1}$ instead of $\mathbf{U}$) with the k-th left eigenvector $\mathbf{l}^k$, $k = 1, \dots, 5$, but only for those values of k for which the eigenvalue λ_k is associated with a characteristic line that is directed outward from Ω_i at $\mathbf{x}$.
The same kind of approach has been proposed by V. Dolean *et al.* in [40], where they consider the quasi-linear form of the Euler system (2.5.1) in the unknowns $\tilde{\mathbf{U}} = (\rho, u_1, u_2, u_3, p)$, where the Jacobian matrices $\mathcal{A}_k$ $(k = 1, .., 3)$ are given by

$$\mathcal{A}_1 := \begin{pmatrix} u_1 & \rho & 0 & 0 & 0 \\ 0 & u_1 & 0 & 0 & 1/\rho \\ 0 & 0 & u_1 & 0 & 0 \\ 0 & 0 & 0 & u_1 & 0 \\ 0 & \rho c^2 & 0 & 0 & u_1 \end{pmatrix}, \quad \mathcal{A}_2 := \begin{pmatrix} u_2 & 0 & \rho & 0 & 0 \\ 0 & u_2 & 0 & 0 & 0 \\ 0 & 0 & u_2 & 0 & 1/\rho \\ 0 & 0 & 0 & u_2 & 0 \\ 0 & 0 & \rho c^2 & 0 & u_2 \end{pmatrix},$$

and

$$\mathcal{A}_3 := \begin{pmatrix} u_3 & 0 & 0 & \rho & 0 \\ 0 & u_3 & 0 & 0 & 0 \\ 0 & 0 & u_3 & 0 & 0 \\ 0 & 0 & 0 & u_3 & 1/\rho \\ 0 & 0 & 0 & \rho c^2 & u_3 \end{pmatrix}.$$

Considering a general partitioning of the domain $\Omega = \bigcup_k \Omega_k$ $(k = 1,..,N)$ with interfaces $\Gamma_{kj} = \partial\Omega_k \cap \partial\Omega_j$, the idea is the following: integrate in time system (2.5.1) by a backward Euler implicit scheme, involving a linearisation of the flux functions in the neighborhood of a constant state (*i.e. freeze* the coefficients), then decompose the operator $\mathcal{A}_{kj} = \sum_{m=1}^{3}(\mathbf{n}_{kj})_m \mathcal{A}_m$, where $\mathbf{n}_{kj}$ is the unit vector normal to Γ_{kj} directed from Ω_k to Ω_j, into its positive and negative part,

$$\mathcal{A}_{kj} = \mathcal{A}_{kj}^{+} + \mathcal{A}_{kj}^{-}.$$

Owing to the diagonalisation of $\mathcal{A}_{kj}$, one has $\mathcal{A}_{kj}^{\pm} = T\Lambda_{kj}^{\pm}T^{-1}$, where $\Lambda_{kj}^{\pm} = \operatorname{diag}\left([\lambda_n^{kj}]^{\pm}\right)_{1\leq n\leq 5}$ with $[\lambda_n^{kj}]^{\pm} = \frac{1}{2}(\lambda_n^{kj} \pm |\lambda_n^{kj}|)$. We recall that, on the interface Γ_{kj}, the eigenvalues of $\mathcal{A}_{kj}$ are given by $\lambda_1^{kj} = \mathbf{u}\cdot\mathbf{n}_{kj} + c$, $\lambda_2^{kj} = \mathbf{u}\cdot\mathbf{n}_{kj} - c$, $\lambda_{3,4,5}^{kj} = \mathbf{u}\cdot\mathbf{n}_{kj}$.
So far, V. Dolean *et al.* propose an additive Schwarz algorithm without overlap that reads as follows.

Given $\tilde{\mathbf{U}}_k^0$ $(k = 1,..,N)$, find $\tilde{\mathbf{U}}_k$ such that

$$\begin{cases} \mathcal{L}\tilde{\mathbf{U}}_k^{m+1} = f & \text{in } \Omega_k \\ \mathcal{A}_{kj}^{+}\tilde{\mathbf{U}}_k^{m+1} = \mathcal{A}_{kj}^{+}\tilde{\mathbf{U}}_j^{m} & \text{on } \Gamma_{kj} \\ \mathcal{A}_{kj}^{-}\tilde{\mathbf{U}}_k^{m+1} = \mathcal{A}_{kj}^{-}\tilde{\mathbf{U}}_j^{m} & \text{on } \Gamma_{kj} \end{cases}$$

where the operator $\mathcal{L}$ is defined as

$$\mathcal{L}\tilde{\mathbf{U}} := \frac{1}{\Delta t}\tilde{\mathbf{U}} + \mathcal{A}_1\partial_x\tilde{\mathbf{U}} + \mathcal{A}_2\partial_y\tilde{\mathbf{U}} + \mathcal{A}_3\partial_z\tilde{\mathbf{U}}.$$

In [40] a convergence analysis of the above algorithm is performed via a Fourier analysis for a two-domain decomposition in both the two- and three-dimensional cases, and some numerical results are given (see also [39]).

2.6 Conclusions

We proposed an iteration-by-subdomain algorithm with interface matching conditions of Dirichlet/Dirichlet type, and we proved its convergence, for both the time discrete and the fully discrete problem, in the case of one dimensional isentropic flows. Convergence is achieved for any choice of the time step Δt in the time marching scheme, and independently of the mesh parameter h. Aside the fact that ideal fluids are not real ones, showing possibly a dramatically different behavior, a few more comments are in order.

Firstly, we considered isentropic flows, which is not such a restrictive assumption, since this is a good approximation of several phenomena occurring in nature. Then, the result has been obtained for the quasi-linear form of Euler system, thus the convergence of the iterative algorithm is ensured only in the region of smooth flow. Since it is well known that the Euler system develops shocks in a finite time, further work needs to be done to extend this result also in the presence of shocks or rarefaction waves. Finally, a convergence result for the quasi-linear system in higher dimensions without freezing the coefficients is not yet available, and it appears rather complicated.

The result obtained is clearly not optimal, nevertheless it is an attempt to give a theoretical convergence analysis for a domain decomposition approach to the Euler system, a task that, to our knowledge, hasn't been faced yet.

Chapter 3

Substructuring Methods for Advection-Diffusion Equations

This chapter deals with a class of elliptic equations that are not symmetric, due to the presence of first order terms in the differential operator. These equations describe advection-diffusion processes which typically arise in fluid mechanics or in the modeling of a wide range of physical phenomena. The problem is important in itself in both engineering and environmental sciences, and this is for instance the case of the diffusion and transport of polluant in air and water, or the transport of electrons in semiconductor devices, and it is a key ingredient in the Navier-Stokes equations. We will focus our attention here on the case in which the transport phenomena, governed by the advective terms, are dominant with respect to the diffusive ones, driven by the principal second-order part of the operator, which is the case of the majority of situations of practical interest.

We investigate the solution of advection-diffusion equations in the framework of iterative substructuring methods with non-ovelapping multi-domain partitions (for overlapping partitions see for instance the works by X.-C. Cai and O. B. Widlund - [25], [26], [27], and [98]- and M. Garbey [52]). Such methods based on transmission conditions at the interface are very effective when the diffusive part of the operator is more relevant, whereas when the problem is convection-dominated the natural interface conditions may generate instabilities.

Many scientists worked in the past years on this subject, and, in the first part of the chapter, we address a review of the principal substructuring methods appeared in literature: we describe adaptive methods, originally introduced by C. Carlenzoli and A. Quarteroni in [28], and futherly developed by L. Trotta [97] and F. Gastaldi, L. Gastaldi and A. Quarteroni [55], coercive methods, which have been studied by A. Alonso, L. Trotta and A. Valli in [6], by F. Nataf and F. Rogier in [81], and by A. Auge, G. Lube and F.Otto in [13], and the Robin/Robin method, firstly introduced, by Y. Achdou and F. Nataf in [3], as an extension to non-symmetric problems of the Neumann/Neumann preconditioner for the Steklov-Poincarè interface equation. In the second part of the chapter, we present a work done in collaboration with P. Le Tallec and F. Nataf at CMAP of the École Polytechnique in Paris: we consider an advection-diffusion problem with discontinuous viscosity coefficient, a kind of problem which may arise, for instance, from the modeling of transport and diffusion of a species through an heterogeneous medium, where the different viscosity coefficients depend on the physical properties of different materials present in each subregion of the computational domain. We propose and analyze a preconditioner of Robin/Robin type for the solution of the associated Steklov-Poincarè interface equation. In the last part of the chapter, some numerical results in three dimensions are presented. The original results of this section can also be found in [57] and, in shorter form, in [58].

3.1 Advection-diffusion problems and their multi-domain formulation

Let Ω be a bounded domain in $\mathbf{R}^d$ (with $d = 2, 3$). We consider in Ω the boundary value problem:

$$\begin{cases} L_\nu u := -\nu\Delta u + \operatorname{div}(\mathbf{b}u) + a\,u & = f \quad \text{in } \Omega \\ \\ u & = 0 \quad \text{on } \partial\Omega, \end{cases} \tag{3.1.1}$$

where $\nu > 0$ is a diffusion coefficient, $\mathbf{b} = \mathbf{b}(\mathbf{x})$ is a given flow field, and $a = a(\mathbf{x})$ is a reaction term: when a is constant, it may arise from an implicit time discretization of the evolution problem and represent the inverse of the time step, namely $a = 1/\Delta t$. Finally, $f = f(\mathbf{x})$ represents a given body force.

We consider the domain Ω partitioned into two non-overlapping open subdomains Ω_1 and Ω_2, we define the interface as

$$\Gamma = \partial\Omega_1 \cap \partial\Omega_2,$$

we denote with $\mathbf{n}_1(\mathbf{x})$ and $\mathbf{n}_2(\mathbf{x})$ the unit vectors normal to $\partial\Omega_1$ and $\partial\Omega_2$ respectively, pointing outwards, and we set $\mathbf{n}(\mathbf{x}) := \mathbf{n}_1(\mathbf{x})$ for $\mathbf{x} \in \Gamma$. We finally denote with u_i, $i = 1, 2$, the restrictions of the solution u of problem (3.1.1) to each subdomain Ω_i.

With these positions, problem (3.1.1) can be equivalently reformulated in multidomain form, with different suitable choices of matching conditions at the interface.

An immediate choice consists in enforcing the continuity across Γ of the solution u and of its normal derivative $\frac{\partial u}{\partial n}$. This is called the *Dirichlet/Neumann* (DN) formulation, which reads as follows.

For $i = 1, 2$, find $u_i = u_{|\Omega_i}$ such that

$$\begin{cases} L_\nu u_i = f & \text{in } \Omega_i, \quad i = 1, 2 \\ \\ u_i = 0 & \text{on } \partial\Omega_i \cap \partial\Omega, \quad i = 1, 2 \\ \\ u_1 = u_2 & \text{on } \Gamma \\ \\ \nu\dfrac{\partial u_1}{\partial n} = \nu\dfrac{\partial u_2}{\partial n} & \text{on } \Gamma. \end{cases} \tag{3.1.2}$$

Assuming that the interface Γ is Lipschitz, taking into account the local direction of the flow field $\mathbf{b}(\mathbf{x})$, we partition it as $\Gamma = \Gamma^{\text{in}} \cup \Gamma^{\text{out}} \cup \Gamma^0$, with

$$\begin{aligned} \Gamma^{\text{in}} &:= \{\mathbf{x} \in \Gamma \,|\, \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\} \\ \Gamma^{\text{out}} &:= \{\mathbf{x} \in \Gamma \,|\, \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) > 0\} \\ \Gamma^{0} &:= \{\mathbf{x} \in \Gamma \,|\, \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = 0\}\,. \end{aligned} \tag{3.1.3}$$

When the subset Γ^0 has zero surface measure, we can replace the Dirichlet matching condition $(3.1.2)_3$ with a Robin type one, which arises from an alternative weak formulation of problem (3.1.1), where also the convective term is integrated by parts. We obtain the following equivalent *Robin/Neumann* (RN) formulation.

For $i = 1, 2$, find $u_i = u_{|\Omega_i}$ such that

$$\begin{cases} L_\nu u_i = f & \text{in } \Omega_i, \quad i = 1,2 \\ u_i = 0 & \text{on } \partial\Omega_i \cap \partial\Omega, \quad i = 1,2 \\ \nu \dfrac{\partial u_1}{\partial n} - \mathbf{b}\cdot\mathbf{n}\, u_1 = \nu \dfrac{\partial u_2}{\partial n} - \mathbf{b}\cdot\mathbf{n}\, u_2 & \text{on } \Gamma \\ \nu \dfrac{\partial u_1}{\partial n} = \nu \dfrac{\partial u_2}{\partial n} & \text{on } \Gamma. \end{cases} \tag{3.1.4}$$

Another set of interface conditions that can be used also when Γ^0 has positive surface measure is obtained by simply replacing the term $\mathbf{b}\cdot\mathbf{n}u_i$ in $(3.1.4)_3$ with βu_i $(i = 1, 2)$, where β is a given function in $L^\infty(\Gamma)$ such that $\beta \neq 0$ almost everywhere in Γ. The associated multi-domain formulation is the so-called *β-Robin/Neumann* (R$_\beta$N) formulation, which reads as follows.

For $i = 1, 2$, find $u_i = u_{|\Omega_i}$ such that

$$\begin{cases} L_\nu u_i = f & \text{in } \Omega_i, \quad i = 1,2 \\ u_i = 0 & \text{on } \partial\Omega_i \cap \partial\Omega, \quad i = 1,2 \\ \nu \dfrac{\partial u_1}{\partial n} - \beta\, u_1 = \nu \dfrac{\partial u_2}{\partial n} - \beta\, u_2 & \text{on } \Gamma \\ \nu \dfrac{\partial u_1}{\partial n} = \nu \dfrac{\partial u_2}{\partial n} & \text{on } \Gamma. \end{cases} \tag{3.1.5}$$

Let us notice that any of the choices above for the matching conditions on the interface can be rigorously justified. Moreover, although their equivalence can be proved (see F. Gastaldi *et al.* [55]), each set of conditions DN, RN, and R$_\beta$N generates different iterative substructuring schemes between Ω_1 and Ω_2. Clearly, the choice of the matching conditions depends heavily on the data and the structure of the problem one is focusing on.

3.1.1 Variational formulation

Problem (3.1.1) can be put in variational form, by introducing the bilinear form associated with the operator L_ν, which is defined, for each $w, v \in H^1(\Omega)$, as

$$a^0(w,v) := \int_\Omega [\nu \nabla w \nabla v + \mathrm{div}\,(\mathbf{b}w)v + awv]\,. \tag{3.1.6}$$

which can be easily verified to be continuous in $H^1(\Omega)$. The weak formulation of problem (3.1.1) is thus

$$\text{Find } u \in H^1_0(\Omega) \ : \ a^0(u,v) = (f,v)_\Omega, \quad \forall v \in H^1(\Omega) \tag{3.1.7}$$

where $(.,.)_\Omega$ denotes the inner product in $L^2(\Omega)$.
When dealing with existence and uniqueness analysis of problem (3.1.7) it is usual to assume that there exists $\mu > 0$ such that

$$\frac{1}{2}\mathrm{div}\ \mathbf{b}(\mathbf{x}) + a(\mathbf{x}) \geq \mu > 0 \tag{3.1.8}$$

for almost every $\mathbf{x} \in \Omega$. Under this assumption, the bilinear form $a^0(.,.)$ is coercive, and the well-known Lax-Milgram Lemma ensures that there exists a unique solution of problem (3.1.7) (for more details see for instance [88]).
The multidomain formulations presented in differential form in the previous section can be easily rewritten in a weak form, and to this aim we define, for $i = 1, 2$, the spaces

$$V_i := \{v_i \in H^1(\Omega_i) \mid v_{i|\partial\Omega\cap\partial\Omega_i} = 0\}. \tag{3.1.9}$$

If we introduce the restricted bilinear forms

$$a_i^0(w_i, v_i) := \int_{\Omega_i} [\nu \nabla w_i \nabla v_i + \mathrm{div}\,(\mathbf{b}w_i)v_i + aw_iv_i]\,. \tag{3.1.10}$$

it can be shown that problem (3.1.7) is equivalent to the following multi-domain problem:

Find $u_1 \in V_1$, $u_2 \in V_2$ such that

$$\begin{cases} a_1^0(u_1, v_1) = (f, v_1)_{\Omega_1} & \forall v_1 \in H^1_0(\Omega_1) \\ u_1 = u_2 & \text{on } \Gamma \\ a_2^0(u_2, v_2) = (f, v_2)_{\Omega_2} & \forall v_2 \in H^1_0(\Omega_2) \\ \displaystyle\sum_{i=1}^2 a_i^0(u_i, \mathcal{R}_i\mu) = \sum_{i=1}^2 (f, \mathcal{R}_i\mu)_{\Omega_i} & \forall \mu \in \Lambda, \end{cases} \tag{3.1.11}$$

where $\Lambda = \mathrm{Tr}_\Gamma(H^1_0(\Omega))$, is the space of traces on Γ of functions belonging to $H^1_0(\Omega)$, and $\mathcal{R}_i\mu$ denotes any possible extension of μ to Ω_i. This is the weak form of the Dirichlet/Neumann formulation.

If we integrate by parts also the convective term in $a^0(.,.)$, we derive an alternative weak formulation which uses the following bilinear form:

$$a^R(w,v) := \int_\Omega [\nu \nabla w \nabla v - w\mathbf{b} \cdot \nabla v + awv] . \tag{3.1.12}$$

When Γ^0 has zero surface measure, we consider the restrictions of $a^R(.,.)$ to Ω_i, for $i = 1, 2$,

$$a_i^R(w_i, v_i) := \int_{\Omega_i} [\nu \nabla w_i \nabla v_i - w_i\mathbf{b} \cdot \nabla v_i + aw_iv_i] . \tag{3.1.13}$$

Notice that, for $i = 1, 2$, the bilinear forms $a_i^0(.,.)$ and $a_i^R(.,.)$ coincide on the space $H_0^1(\Omega_i)$, since we have

$$a_i^R(w_i, v_i) = a_i^0(w_i, v_i) - \int_\Gamma \mathbf{b} \cdot \mathbf{n}_i w_i v_i ,$$

for all $w_i, v_i \in V_i$, and this entails well-posedeness for the Robin/Neumann formulation (3.1.4), whose weak form reads

Find $u_1 \in V_1$, $u_2 \in V_2$ such that

$$\begin{cases} a_1^R(u_1, v_1) = (f, v_1)_{\Omega_1} & \forall v_1 \in H_0^1(\Omega_1) \\ \sum_{i=1}^2 a_i^R(u_i, \mathcal{R}_i\mu) = \sum_{i=1}^2 (f, \mathcal{R}_i\mu)_{\Omega_i} & \forall \mu \in \Lambda, \\ a_2^R(u_2, v_2) = (f, v_2)_{\Omega_2} & \forall v_2 \in H_0^1(\Omega_2) \\ \sum_{i=1}^2 a_i^0(u_i, \mathcal{R}_i\mu) = \sum_{i=1}^2 (f, \mathcal{R}_i\mu)_{\Omega_i} & \forall \mu \in \Lambda. \end{cases} \tag{3.1.14}$$

Finally, for any type of Γ^0, the variational form of the β-Robin/Neumann formulation reads:

Find $u_1 \in V_1$, $u_2 \in V_2$ such that

$$\begin{cases} a_1^R(u_1, v_1) = (f, v_1)_{\Omega_1} & \forall v_1 \in H_0^1(\Omega_1) \\ \sum_{i=1}^2 a_i^\beta(u_i, \mathcal{R}_i\mu) = \sum_{i=1}^2 (f, \mathcal{R}_i\mu)_{\Omega_i} & \forall \mu \in \Lambda, \\ a_2^R(u_2, v_2) = (f, v_2)_{\Omega_2} & \forall v_2 \in H_0^1(\Omega_2) \\ \sum_{i=1}^2 a_i^0(u_i, \mathcal{R}_i\mu) = \sum_{i=1}^2 (f, \mathcal{R}_i\mu)_{\Omega_i} & \forall \mu \in \Lambda. \end{cases} \tag{3.1.15}$$

where the bilinear form $a_i^\beta(.,.)$ $(i = 1, 2)$ is defined, for each $w_i, v_i \in V_i$, as

$$a_i^\beta(w_i, v_i) := a_i^R(w_i, v_i) + \int_\Gamma (\mathbf{b} \cdot \mathbf{n} - \beta)\mathbf{n} \cdot \mathbf{n}_i w_i v_i. \tag{3.1.16}$$

3.2 Iterative Substructuring Methods

We face here the task of solving the multi-domain problems, described in the previous section, by iterative methods. The key idea is the following: given an initial guess u_1^0 and u_2^0 for the restriction of the solution u of (3.1.1) to Ω_1 and Ω_2 respectively, we build a sequence of subproblems in Ω_1 and Ω_2 with boundary conditions on the interface Γ that are either of Dirichlet, Neumann or Robin type, according to the formulation we choose, generating two sequences of functions $\{u_1^k\}$ and $\{u_2^k\}$, which will converge to u_1 and u_2. This approach based on transmission conditions at the interface is usually referred to as *Iterative Substructuring.*

We present in the following three types of iterative methods: we firstly deal with methods that, considering the fact that for strongly dominant convection the problem is almost hyperbolic, take into account the direction of the flow field across the interface (we refer to these methods as *adaptive*). The second family of methods are based on bilinear forms which are coercive in each subdomain (we refer to these methods as *coercive*). Finally, the third family of methods we consider here seeks for a direct parallel preconditioning of the Steklov-Poincarè interface equation. Among them, the first two family of procedures are sequential, whereas the third one is naturally parallel. Since this is the most desirable feature when seeking for a decomposition into a great number of subdomains, also the first two family of methods can be written in parallel. However, their sequential nature turns out to generate a redundant sequence of iterate solutions, which contains as a subsequence the solutions stemming from the sequential procedure.

3.2.1 Adaptive Methods: ADN, ARN, AR$_\beta$N

The difficulties arising when the viscosity coefficient ν is small can be heuristically explained in the limiting situation as $\nu \to 0^+$. When considering a subproblem in either Ω_1 or Ω_2, with a boundary condition at the interface, this should be consistent with the hyperbolic limit. This amounts to take into account the local direction of the characteristic curves at the interface, and to impose a Neumann boundary condition on the *outflow* part of the interface and a Dirichlet boundary condition on the *inflow* part.

The idea is to define a sequence $\{u_1^k, u_2^k\}$, where u_i^k satisfies $L_\nu u_i^k = f$ in Ω_i, together with suitable boundary conditions at the interface Γ that depend on the local direction of the flow field $\mathbf{b}(\mathbf{x})$. These algorithms are thus called *adaptive*, because the role played by the conditions on Γ varies according to the flow conditions, and we have Adaptive Dirichlet/Neumann (ADN), Adaptive Robin/Neumann, or Adaptive β-Robin/Neumann algorithms, depending on the choice of interface conditions. Such methods have been proposed, and motivated by the fact that, in the one-dimensional case with $a = 0$, the convergence rate of the algorithm is of order $\exp(-b/\nu)$, by Carlenzoli and Quarteroni in [28], then further developed by Trotta [97] and F. Gastaldi *et al.* [55]

Recalling that the interface can be partitioned as $\Gamma = \Gamma^{\text{in}} \cup \Gamma^{\text{out}} \cup \Gamma^0$, with Γ^{in}, Γ^{out}, and Γ^0 defined as in (3.1.3), in the rest of the section we present these adaptive algorithms.

The Adaptive Dirichlet/Neumann method

The Adaptive Dirichlet/Neumann algorithm enforces on each subdomain a Dirichlet condition on the corresponding inflow part of the interface Γ and a Neumann condition on the outflow part. On Γ^0, where the flux is parallel to the interface, one can either enforce the Dirichlet or the Neumann condition. However, once this choice is made for one domain, one has to impose on the complementary domain the condition which has not been enforced on the other one. Its differential form reads as follows.

Given u_i^0 in Ω_i, solve for each $k \geq 0$

$$\begin{cases} L_\nu u_1^{k+1} = f & \text{in } \Omega_1 \\ u_1^{k+1} = 0 & \text{on } \partial\Omega_1 \cap \partial\Omega \\ u_1^{k+1} = \lambda^k & \text{on } \Gamma^{\text{in}} \cup \Gamma^0 \\ \nu\dfrac{\partial u_1^{k+1}}{\partial n} = \nu\dfrac{\partial u_2^k}{\partial n} & \text{on } \Gamma^{\text{out}} \end{cases} \tag{3.2.1}$$

and

$$\begin{cases} L_\nu u_2^{k+1} = f & \text{in } \Omega_2 \\ u_2^{k+1} = 0 & \text{on } \partial\Omega_2 \cap \partial\Omega \\ u_2^{k+1} = \mu^{k+1} & \text{on } \Gamma^{\text{out}} \\ \nu\dfrac{\partial u_2^{k+1}}{\partial n} = \nu\dfrac{\partial u_1^{k+1}}{\partial n} & \text{on } \Gamma^{\text{in}} \cup \Gamma^0 \end{cases} \tag{3.2.2}$$

with

$$\lambda^k := \vartheta' u^k_{2|\Gamma^{\text{in}}\cup\Gamma^0} + (1-\vartheta')u^k_{1|\Gamma^{\text{in}}\cup\Gamma^0} \qquad \text{on } \Gamma^{\text{in}} \cup \Gamma^0, \tag{3.2.3}$$

and

$$\mu^{k+1} := \vartheta'' u^{k+1}_{1|\Gamma^{\text{out}}} + (1-\vartheta'')u^k_{2|\Gamma^{\text{out}}} \qquad \text{on } \Gamma^{\text{out}}, \tag{3.2.4}$$

where ϑ' and ϑ'' are two positive parameters that allow, if needed, *under-relaxation* to guarantee convergence. Typically a single parameter ϑ is enough (and sometimes $\vartheta' = \vartheta'' = 1$, that is no relaxation at all, is a suitable choice), whereas the presence of two parameters provides more flexibility in achieving optimal convergence.

As the sequence $\{u_1^k, u_2^k\}$ converges in a suitable sense, its limit $\{u_1, u_2\}$ satisfies (at least formally) the differential problem (3.2.25), and would be the desired solution. Convergence for the ADN scheme has been proved by F. Gastaldi *et al.* in [55], for a problem set on the unit square $(0,1)^2$, with a constant advective field $\mathbf{b} = (b, 0)$ and a constant reaction term $a = a$.

The solution of (3.2.1)-(3.2.4) is sequential as it yields the solution of (3.2.2) in Ω_2 only after having solved problem (3.2.1) in Ω_1. This sequence guarantees the quickest convergence, but it might not be the best option when seeking for parallelism, especially when dealing with many subdomains. In that order, the algorithm can be parallelized with an obvious modification which consists of solving problem (3.2.1) in Ω_1, and in parallel the following modified problem in Ω_2:

$$\begin{cases} L_\nu u_2^{k+1} = f & \text{in } \Omega_2 \\ u_2^{k+1} = 0 & \text{on } \partial\Omega_2 \cap \partial\Omega \\ u_2^{k+1} = \eta^k & \text{on } \Gamma^{\text{out}} \\ \nu \dfrac{\partial u_2^{k+1}}{\partial n} = \nu \dfrac{\partial u_1^k}{\partial n} & \text{on } \Gamma^{\text{in}} \cup \Gamma^0 \end{cases} \tag{3.2.5}$$

with

$$\eta^k := \vartheta'' u^k_{1|\Gamma^{\text{out}}} + (1-\vartheta'') u^k_{2|\Gamma^{\text{out}}} \qquad \text{on } \Gamma^{\text{out}}. \tag{3.2.6}$$

It is straightforward to see that the iterative algorithm (3.2.1)-(3.2.5)-(3.2.3)-(3.2.6) generalizes naturally to the case in which the domain Ω is decomposed into many subdomains.

The Adaptive Robin/Neumann and β-Robin/Neumann methods

These methods enforce on each subdomain a Robin condition on the corresponding inflow part of the interface Γ and a Neumann condition on the outflow part. If Γ^0 has a zero surface measure, the *Adaptive Robin/Neumann algorithm* (ARN) reads as follows.

Given u_i^0 in Ω_i, solve for each $k \geq 0$

$$\begin{cases} L_\nu u_1^{k+1} = f & \text{in } \Omega_1 \\ u_1^{k+1} = 0 & \text{on } \partial\Omega_1 \cap \partial\Omega \\ \psi(u_1^{k+1}) = \lambda^k & \text{on } \Gamma^{\text{in}} \\ \nu \dfrac{\partial u_1^{k+1}}{\partial n} = \nu \dfrac{\partial u_2^k}{\partial n} & \text{on } \Gamma^{\text{out}} \end{cases} \tag{3.2.7}$$

and

$$\begin{cases} L_\nu u_2^{k+1} = f & \text{in } \Omega_2 \\ u_2^{k+1} = 0 & \text{on } \partial\Omega_2 \cap \partial\Omega \\ \psi(u_2^{k+1}) = \mu^{k+1} & \text{on } \Gamma^{\text{out}} \\ \nu \dfrac{\partial u_2^{k+1}}{\partial n} = \nu \dfrac{\partial u_1^{k+1}}{\partial n} & \text{on } \Gamma^{\text{in}} \end{cases} \tag{3.2.8}$$

with

$$\psi(v) := \nu \frac{\partial v}{\partial n} - \mathbf{b}\cdot\mathbf{n}\, v, \tag{3.2.9}$$

$$\lambda^k := \vartheta' \psi(u_2^k)_{|\Gamma^{\rm in}} + (1-\vartheta')\psi(u_1^k)_{|\Gamma^{\rm in}} \qquad \text{on } \Gamma^{\rm in}, \tag{3.2.10}$$

and

$$\mu^{k+1} := \vartheta'' \psi(u_1^{k+1})_{|\Gamma^{\rm out}} + (1-\vartheta'')\psi(u_2^k)_{|\Gamma^{\rm out}} \qquad \text{on } \Gamma^{\rm out}. \tag{3.2.11}$$

The convergence of the ARN algorithm is a consequence of Theorem 3.2.2, as a particular case of a family of iteration-by-subdomain methods.

The *Adaptive β-Robin/Neumann algorithm* is obtained from the previous one by simply replacing the flux $\psi(v)$ in (3.2.9) with the modified flux

$$\psi(v) := \nu \frac{\partial v}{\partial n} - \beta\, v, \tag{3.2.12}$$

which allows to omit the assumption on Γ^0. In fact, with this choice of the flux $\psi(v)$, one may impose on Γ^0 either the same condition of $\Gamma^{\rm in}$ or the same condition of $\Gamma^{\rm out}$. In any case, once this choice is made for one domain, one has to impose the other condition on the complementary domain. However, in order to guarantee solvability for (3.2.7) and (3.2.8) in the AR$_\beta$N framework, we have to make the following assumptions on β:

$$\beta \le \frac{1}{2}\mathbf{b}\cdot\mathbf{n} \text{ on } \Gamma^{\rm in}, \quad \beta \ge \frac{1}{2}\mathbf{b}\cdot\mathbf{n} \text{ on } \Gamma^{\rm out},$$

and $\beta \gtreqless \frac{1}{2}\mathbf{b}\cdot\mathbf{n}$ on Γ^0 according to choice made for the boundary condition on this part of the interface.

In the same way as the ADN method, the ARN and the AR$_\beta$N algorithms are sequential, but they can be easily parallelized by simply replacing $(3.2.8)_3$, $(3.2.8)_4$, and (3.2.11) by

$$\psi(u_2^{k+1}) = \mu^k \qquad \text{on } \Gamma^{\rm out}, \qquad\qquad \nu\frac{\partial u_2^{k+1}}{\partial n} = \nu\frac{\partial u_1^k}{\partial n} \qquad \text{on } \Gamma^{\rm in},$$

and

$$\mu^k := \vartheta'' \psi(u_1^k)_{|\Gamma^{\rm out}} + (1-\vartheta'')\psi(u_2^k)_{|\Gamma^{\rm out}} \qquad \text{on } \Gamma^{\rm out},$$

so that the resulting algorithm can be easily extended to a decomposition of the domain Ω into many subdomains.

3.2.2 Coercive Methods: γ-DR and γ-RR

Other kind of iterative procedures do not pay a significant attention to the local direction of the advective field $\mathbf{b}$ on Γ, but they require that the bilinear forms associated with the boundary value subproblem in each subdomain Ω_1 and Ω_2 are coercive in $H^1(\Omega_1)$ and $H^1(\Omega_2)$ respectively, under the only assumption that $\frac{1}{2}\text{div}\,\mathbf{b} + a \ge \mu > 0$ in Ω. These methods are in fact a family of schemes depending on a real parameter $\gamma = \gamma(\mathbf{x})$ which is in general a given non-negative function of $L^\infty(\Gamma)$, which influences the rate of convergence of the algorithm. No requirement on the boundary value to vanish on any part of $\partial\Omega_j$, $j = 1, 2$ is made, though the bilinear forms $a_1^0(.,.)$ and $a_2^0(.,.)$ introduced in (3.1.10) are coercive in $V_1 \cap H^1_{\Gamma^{\rm in}}(\Omega_1)$ and $V_2 \cap H^1_{\Gamma^{\rm out}}(\Omega_2)$,

respectively, and the bilinear forms $a_1^R(.,.)$ and $a_2^R(.,.)$ introduced in (3.1.13) are coercive in $V_1 \cap H^1_{\Gamma_{\text{out}}}(\Omega_1)$ and $V_2 \cap H^1_{\Gamma_{\text{in}}}(\Omega_2)$ respectively, but not in $H^1(\Omega_1)$ nor in $H^1(\Omega_2)$.
The iterative methods we present in this section have, with respect to the ones introduced in the previous sections, the drawback to use bilinear forms that are somehow a little more complicated, and an additional parameter γ which has to be considered besides the relaxation parameter ϑ. However, the flow direction does not have to be taken into account and they can easily extend to systems of equations.

The γ-Dirichlet/Robin method

This scheme has been proposed by Alonso *et al.* in [6], and reads as follows: at each step we have a boundary value problem in Ω_1 with a Dirichlet condition on Γ and a boundary value problem in Ω_2 with a Robin condition on Γ. Given $\gamma \geq 0$, the differential form of the scheme is the following:

Given λ^0, solve for each $k \geq 0$

$$\begin{cases} L_\nu u_1^{k+1} = f & \text{in } \Omega_1 \\ u_1^{k+1} = 0 & \text{on } \partial\Omega_1 \cap \partial\Omega \\ u_1^{k+1} = \lambda^k & \text{on } \Gamma \end{cases} \tag{3.2.13}$$

and

$$\begin{cases} L_\nu u_2^{k+1} = f & \text{in } \Omega_2 \\ u_2^{k+1} = 0 & \text{on } \partial\Omega_2 \cap \partial\Omega \\ \nu \dfrac{\partial u_2^{k+1}}{\partial n} - \left(\dfrac{1}{2}\mathbf{b}\cdot\mathbf{n} + \gamma\right) u_2^{k+1} = \nu \dfrac{\partial u_1^{k+1}}{\partial n} - \left(\dfrac{1}{2}\mathbf{b}\cdot\mathbf{n} + \gamma\right) u_1^{k+1} & \text{on } \Gamma, \end{cases} \tag{3.2.14}$$

then set

$$\lambda^{k+1} := \vartheta u_{2|\Gamma}^{k+1} + (1-\vartheta)\lambda^k \qquad \text{on } \Gamma. \tag{3.2.15}$$

In order to have the variational formulation of this scheme, we introduce the local bilinear forms

$$a_i^b(w_i, v_i) := \int_{\Omega_i} \left\{ \nu \nabla w_i \cdot \nabla v_i + \left(\frac{1}{2}\operatorname{div}\mathbf{u} + a\right) w_i v_i \right\} + \frac{1}{2}\int_{\Omega_i} (v_i \mathbf{b}\cdot\nabla w_i - w_i \mathbf{b}\cdot\nabla v_i), \tag{3.2.16}$$

which are continuous and coercive in $H^1(\Omega_i)$, $i = 1, 2$, with continuity and coercivity constants β_i^b and α_i^b respectively, and we define the spaces V_1, V_2 and Λ as in Section 3.1.1. The variational formulation of the γ-Dirichlet/Robin method is therefore

$$\begin{cases} \text{find } u_1^{k+1} \in V_1 \; : \\ a_1^b(u_1^{k+1}, v_1) = (f, v_1)_{\Omega_1} \qquad \forall v_1 \in V_1^0 \\ u_{1|\Gamma}^{k+1} = \lambda^k \end{cases} \tag{3.2.17}$$

and

$$\begin{cases} \text{find } u_2^{k+1} \in V_2 \; : \\ a_2^b(u_2^{k+1}, v_2) + \displaystyle\int_\Gamma \gamma u_{2|\Gamma}^{k+1} v_{2|\Gamma} = (f, v_2)_{\Omega_2} + (f, \mathcal{R}_1 v_{2|\Gamma})_{\Omega_1} \\ \qquad -a_1^b(u_1^{k+1}, \mathcal{R}_1 v_{2|\Gamma}) + \displaystyle\int_\Gamma \gamma u_{1|\Gamma}^{k+1} v_{2|\Gamma} \qquad \forall v_2 \in V_2, \end{cases} \tag{3.2.18}$$

where $\mathcal{R}_i$ denotes any extension operator from Λ to V_i, and finally setting λ^{k+1} as in (3.2.15). Notice that problem (3.2.18) is coercive in V_2, for any $\gamma \geq 0$. The variational iterative scheme up above is thus well defined.

Remark 3.2.1 This scheme is different from the ADN scheme, since the Dirichlet boundary condition is imposed on the whole interface, disregarding the fact whether is an inflow or an outflow boundary. However, if the flow has the same direction on the whole interface Γ, we recover the ADN scheme. □

Let us introduce, for $i = 1, 2$ and for $\lambda \in \Lambda$, the a_i^b-harmonic extension of λ, that we denote with $E_i^b \lambda$, as being the solution of the Dirichlet boundary value problem

$$\begin{cases} a_i^b(E_i^b \lambda, v_i) = 0 \qquad \forall\, v_i \in V_i^0 \\ (E_i^b \lambda)_{|\Gamma} = \lambda. \end{cases}$$

For each $\lambda, \mu \in \Lambda$ and $i = 1, 2$ we define the Steklov-Poincaré operators $S_i : \Lambda \to \Lambda'$, as

$$\langle S_i \lambda, \mu \rangle := a_i^b(E_i^b \lambda, E_i^b \mu),$$

and we set

$$\langle S_1^{(\gamma)} \lambda, \mu \rangle := a_1^b(E_1^b \lambda, E_1^b \mu) - \gamma(\lambda, \mu)_\Lambda$$
$$\langle S_2^{(\gamma)} \lambda, \mu \rangle := a_2^b(E_2^b \lambda, E_2^b \mu) + \gamma(\lambda, \mu)_\Lambda.$$

We therefore have

$$S = S_1^{(\gamma)} + S_2^{(\gamma)} = S_1 + S_2$$

and the iterative scheme (3.2.17)-(3.2.18)-(3.2.15) is equivalent to a preconditioned Richardson method for the Steklov-Poincaré operator S, with $S_2^{(\gamma)}$ as a preconditioner:

$$\lambda^{k+1} = \lambda^k + \vartheta (S_2^{(\gamma)})^{-1} (\Upsilon - S \lambda^k).$$

The convergence of the γ-Dirichlet/Robin method is ensured by the following result, which is proved in [6] and [89].

Theorem 3.2.1 *There exists $\gamma^* \geq 0$ such that for each $\gamma \geq \gamma^*$ and for each $\lambda^0 \in \Lambda$ the iterative scheme* (3.2.17)-(3.2.18)-(3.2.15) *is convergent in Λ, provided the relaxation parameter ϑ is chosen in a suitable interval* $(0, \vartheta_\gamma)$. □

The γ-Robin/Robin method

The γ-Robin/Robin method is another iteration-by-subdomain procedure proposed by Alonso *et al.* in [6]. It extends to the non-symmetric case the Robin method proposed by P.-L. Lions in [79] for symmetric elliptic operators. Given a function $\gamma = \gamma(\mathbf{x})$ in $L^\infty(\Gamma)$ stisfying $\gamma(\mathbf{x}) \geq \hat{\gamma} > 0$, the scheme reads as follows.

Given $\lambda^0 \in L^2(\Gamma)$, for each $k \geq 0$ solve

$$\begin{cases} L_\nu u_1^{k+1} = f & \text{in } \Omega_1 \\ u_1^{k+1} = 0 & \text{on } \partial\Omega_1 \cap \partial\Omega \\ \nu \dfrac{\partial u_1^{k+1}}{\partial n} - \left(\dfrac{1}{2}\mathbf{b}\cdot\mathbf{n} - \gamma\right) u_1^{k+1} = \lambda^k & \text{on } \Gamma \end{cases} \tag{3.2.19}$$

and

$$\begin{cases} L_\nu u_2^{k+1} = f & \text{in } \Omega_2 \\ u_2^{k+1} = 0 & \text{on } \partial\Omega_2 \cap \partial\Omega \\ \nu \dfrac{\partial u_2^{k+1}}{\partial n} - \left(\dfrac{1}{2}\mathbf{b}\cdot\mathbf{n} + \gamma\right) u_2^{k+1} = \nu \dfrac{\partial u_1^{k+1}}{\partial n} - \left(\dfrac{1}{2}\mathbf{b}\cdot\mathbf{n} + \gamma\right) u_1^{k+1} & \text{on } \Gamma, \end{cases} \tag{3.2.20}$$

then set

$$\lambda^{k+1} := \nu \frac{\partial u_2^{k+1}}{\partial n} - \left(\frac{1}{2}\mathbf{b}\cdot\mathbf{n} - \gamma\right) u_2^{k+1} \qquad \text{on } \Gamma. \tag{3.2.21}$$

It is worthwhile to note that

$$\lambda^{k+1} = \nu \frac{\partial u_1^{k+1}}{\partial n} - \left(\frac{1}{2}\mathbf{b}\cdot\mathbf{n} + \gamma\right) u_1^{k+1} + 2\gamma u_2^{k+1} = \lambda^k + 2\gamma(u_2^{k+1} - u_1^{k+1}),$$

so that, since $\lambda^0 \in L^2(\Gamma)$ and $\gamma \in L^\infty(\Gamma)$, we have $\lambda^k \in L^2(\Gamma)$ for each $k \geq 0$. We can therefore introduce the variational form of the scheme up above as

Given $\lambda^0 \in L^2(\Gamma)$, for each $k \geq 0$

$$\begin{cases} \text{find } u_1^{k+1} \in V_1 \; : \\ a_1^b(u_1^{k+1}, v_1) + \displaystyle\int_\Gamma \gamma u_{1|\Gamma}^{k+1} v_{1|\Gamma} = (f, v_1)_{\Omega_1} + \int_\Gamma \gamma \lambda^k v_{1|\Gamma} \qquad \forall v_1 \in V_1 \end{cases} \tag{3.2.22}$$

and

$$\begin{cases} \text{find } u_2^{k+1} \in V_2 \ : \\ a_2^b(u_2^{k+1}, v_2) + \displaystyle\int_\Gamma \gamma u_{2|\Gamma}^{k+1} v_{2|\Gamma} = (f, v_2)_{\Omega_2} + (f, \mathcal{R}_1 v_{2|\Gamma})_{\Omega_1} \\ \qquad\qquad -a_1^b(u_1^{k+1}, \mathcal{R}_1 v_{2|\Gamma}) + \displaystyle\int_\Gamma \gamma u_{1|\Gamma}^{k+1} v_{2|\Gamma} \qquad \forall v_2 \in V_2, \end{cases} \tag{3.2.23}$$

then set

$$\lambda^{k+1} = \lambda^k + 2\gamma(u_2^{k+1} - u_1^{k+1}) \qquad \text{on } \Gamma, \tag{3.2.24}$$

where again $\mathcal{R}_i$ denotes any extension operator from Λ to V_i, and $a_i^b(.,.)$, $i = 1, 2$ are the bilinear forms introduced in (3.2.16). It is thus straightforward to see that the bilinear forms

$$a_i^b(u_i, v_i) + \int_\Gamma \gamma u_i v_i, \quad i = 1, 2$$

used in (3.2.22) and (3.2.23) are coercive in V_i, for each $\gamma \geq 0$.

The γ-Robin/Robin methods generalizes some other methods appeared in literature, for instance by choosing $\gamma = \frac{1}{2}|\mathbf{b} \cdot \mathbf{n}|$ we recover the ARN method without relaxation, by choosing $\gamma = \frac{1}{2}\sqrt{|\mathbf{b} \cdot \mathbf{n}|^2 + 4a\nu}$ we recover the method proposed by Nataf and Rogier in [81], and by choosing $\gamma = \frac{1}{2}\sqrt{|\mathbf{b} \cdot \mathbf{n}|^2 + 4\kappa\nu}$, with $\kappa > 0$, the method proposed by Auge *et al.* in [13].

The convergence of the γ-Robin/Robin method is provided by the following result which is proved in [6] (see also [89]) and is inspired by the results of P.-L. Lions and Nataf and Rogier.

Theorem 3.2.2 *Assume that either Ω is a Lipschitz polygonal domain or that $\partial\Omega$ is regular enough, say $\partial\Omega \in C^2$. Moreover, suppose that $\mathbf{b}_{|\Gamma} \in (L^\infty(\Gamma))^d$. Then, for each $\lambda^0 \in L^2(\Gamma)$ and for each $i = 1, 2$, the sequences u_i^k converge in $H^1(\Omega_i)$ to the restriction $u_{|\Omega_i}$ of the solution u of* (3.1.1). □

3.2.3 Primal Schur methods: the Robin/Robin algorithm

In this section we present another method based on the solution of problem (3.1.1) by a primal Schur method, which amounts to the reduction of the problem in Ω to an interface problem on Γ, and the parallel direct preconditioning of the Steklov-Poincaré equation. This kind of approach was originally introduced by J.-F. Bourgat *et al.* in [19], and is well suited for parallelism in a multidomain formulation. We assume throughout this section that Ω is a rectangular domain $\mathbf{R}^2$, say $\Omega =]0, L[\times]0, \eta[$, and is partitioned into N non-overlapping vertical strips $\Omega_i = (l_i, l_{i+1})\times]0, \eta[$, $1 \leq i \leq N - 1$, with interfaces denoted by $\Gamma_{i,i+1} = \{l_{i+1}\}\times]0, \eta[$. We consider the global interface $\Gamma = \bigcup_{i=1}^{N-1} \Gamma_{i,i+1}$, we consider the restriction on the interface of the solution of (3.1.1), $U_\Gamma = (u_{|\Gamma_{i,i+1}})_{1\leq i\leq N-1}$, and we define the Steklov-Poincaré operator on Γ as

$$\Sigma : \left(H_{00}^{1/2}(]0,\eta[)\right)^{N-1} \times L^2(\Omega) \to \left[\left(H_{00}^{1/2}(]0,\eta[)\right)'\right]^{N-1}$$

$$\Sigma : ((u_i)_{1\le i\le N-1}, f) \longmapsto \left(\frac{\nu}{2}\left(\frac{\partial w_i}{\partial n_i} + \frac{\partial w_{i+1}}{\partial n_{i+1}}\right)_{\Gamma_{i,i+1}}\right)_{1\le i\le N-1}$$

where w_i, for $i = 1, 2$, are the solutions of

$$\begin{cases} L_\nu w_i &= f & \text{in } \Omega_i \\ w_i &= 0 & \text{on } \partial\Omega_i \cap \partial\Omega \\ w_i &= u_{i-1} & \text{on } \Gamma_{i-1,i} \text{ for } i = 2,\dots,N \\ w_i &= u_i & \text{on } \Gamma_{i,i+1} \text{ for } i = 1,\dots,N-1. \end{cases} \tag{3.2.25}$$

where we have set $u_i := u_{|\Gamma_{i,i+1}}$.
Problems (3.2.25) are linear, thus, setting $\mathcal{S}U_\Gamma = \Sigma(U_\Gamma, 0)$ and $\chi = -\Sigma(0, f)$, the interface problem can be written as

$$\mathcal{S}U_\Gamma = \chi. \tag{3.2.26}$$

The idea is the following: we split the Steklov-Poincaré operator into a sum of local operators in each subdomain,

$$\mathcal{S} = \mathcal{S}_1 + \dots + \mathcal{S}_N, \tag{3.2.27}$$

which solve problems with Dirichlet boundary condition on Γ, and we precondition the interface equation (3.2.26) with a weighted sum of the local inverses, which approximates the inverse of the operator $\mathcal{S} = \Sigma(.,0)$. This procedure can be interpreted as a preconditioned Richardson algorithm where, given an initial guess λ^0 for the trace on the interface of the solution u of problem (3.1.1), the we seek the fixed point of the sequence $\{\lambda^k\}$, where

$$\lambda^{k+1} = \lambda^k + \vartheta\left(\sigma_1\mathcal{S}_1^{-1} + \dots + \sigma_N\mathcal{S}_N^{-1}\right)(\chi - \mathcal{S}\lambda^k)$$

for each $k \ge 0$, where $\sigma_i > 0$ are averaging parameters, whereas ϑ is possibly a relaxation parameter.
The original method in [19] proposed, for the Poisson problem in a two domain decomposition setting, to split $\mathcal{S}$ into a sum of Dirichlet to Neumann operators, $\mathcal{S}_i$, for $i = 1, 2$, with

$$\mathcal{S}_i : u_\Gamma \longmapsto \nu\frac{\partial u_i}{\partial n_i}_{|\Gamma},$$

and to use as a preconditioner a system of problems in each subdomain with Neumann conditions on the interface. This algorithm was given the name of *Neumann/Neumann* method. When applied to advection-diffusion equations, such method showed a poor behavior, since the symmetry of the local operators reflected into a lack of capability to handle the non-symmetry of the global differential operator.

The Robin/Robin Algorithm

This method, proposed by Y. Achdou *et al.* in [3] is a generalization of the Neumann/Neumann algorithm to non-symmetric problems: the key idea relies in replacing the Dirichlet to Neumann local operators with Dirichlet to Robin ones, which are able to take into account the action of the convective field. The Steklov-Poincarè operator $\mathcal{S}$ is therefore split as in (3.2.27), where, for $i = 1, \dots, N$, we have

$$\mathcal{S}_i : u_\Gamma \mapsto \left(\nu \frac{\partial u_i}{\partial n_i} - \frac{1}{2} \mathbf{b} \cdot \mathbf{n}_i \, u_i \right)_\Gamma .$$

Notice that, since $\mathbf{n}_i = -\mathbf{n}_{i+1}$ for $i = 1, \dots, N-1$, the terms $\frac{1}{2}\mathbf{b} \cdot \mathbf{n}_i \, u_i$ vanish in the sum and we recover the operator $\mathcal{S}$.
The approximate inverse of $\mathcal{S}$ proposed in [3] is, at the continuous level, the operator $\mathcal{T}$ defined as:

$$\begin{aligned} &\mathcal{T} : \left[\left(H_{00}^{1/2}(]0,\eta[) \right)' \right]^{N-1} \to \left(H_{00}^{1/2}(]0,\eta[) \right)^{N-1} \\ &\mathcal{T} : (g_i)_{1 \le i \le N-1} \longmapsto \left(\frac{1}{2} \left(v_i + v_{i+1} \right)_{\Gamma_{i,i+1}} \right)_{1 \le i \le N-1} . \end{aligned} \tag{3.2.28}$$

where v_i (for $i = 1, \dots, N$) is the solution of

$$\left\{ \begin{array}{rll} L_\nu v_i = & 0 & \text{in } \Omega_i \\ v_i = & 0 & \text{on } \partial\Omega_i \cap \partial\Omega \\ \nu \dfrac{\partial v_i}{\partial n_i} - \dfrac{\mathbf{b} \cdot \mathbf{n}_i}{2} v_i = & g_i & \text{on } \Gamma_{i,i+1} \text{ for } i = 1, \dots, N-1 \\ \nu \dfrac{\partial v_i}{\partial n_i} - \dfrac{\mathbf{b} \cdot \mathbf{n}_i}{2} v_i = & g_{i-1} & \text{on } \Gamma_{i-1,i} \text{ for } i = 2, \dots, N. \end{array} \right. \tag{3.2.29}$$

Although the Robin boundary conditions in (3.2.29) are not standard ones, they nevertheless stem from an integration by parts of the advective term $\frac{1}{2}(\mathbf{b} \cdot \nabla u)v$ in (3.1.1), and they lead to a well-posed problem in each subdomain, as stated in the following Proposition (see [3]).

Proposition 3.2.1 *Let Ω be an open set of $\mathbf{R}^2$, $f \in L^2(\Omega)$, $\lambda \in H^{-1/2}(\partial\Omega)$, $\mathbf{b} \in (C^1(\overline{\Omega}))^2$, $a \in \mathbf{R}$ such that $a - \frac{1}{2}\text{div } \mathbf{b} \ge \mu > 0$ for some $\mu \in \mathbf{R}$. Then there exists an unique $u \in H^1(\Omega)$ such that*

$$\int_\Omega \nu \nabla u \cdot \nabla v + (\mathbf{b} \cdot \nabla u)v + auv - \int_{\partial\Omega} \frac{1}{2} \mathbf{b} \cdot \mathbf{n} \, uv = \langle \lambda, v \rangle + \int_\Omega fv \quad \forall v \in H^1(\Omega),$$

where $\langle .,. \rangle$ denotes the duality pairing between $H^{-1/2}(\partial\Omega)$ and $H^{1/2}(\partial\Omega)$.

Proof. It relies on the Lax-Milgram Lemma, where the only thing that is not obvious is the coercivity of the bilinear form

$$(u,v) \longmapsto \int_\Omega \nu\nabla u \cdot \nabla v + (\mathbf{b}\cdot\nabla u)v + auv - \int_{\partial\Omega} \frac{1}{2}\mathbf{b}\cdot\mathbf{n}\, uv.$$

An integration by parts provide

$$\int_\Omega \nu|\nabla u|^2 + (\mathbf{b}\cdot\nabla u)u + au^2 - \int_{\partial\Omega}\frac{1}{2}\mathbf{b}\cdot\mathbf{n}\, u^2 = \int_\Omega \nu|\nabla u|^2 + (a - \frac{1}{2}\text{div } \mathbf{b})u^2 \geq \min(\mu,\nu)\|u\|^2_{H^1(\Omega)}.$$

and this concludes the proof. □

When the operator is symmetric, or when the flow field is parallel to the interface (*i.e.*, $\mathbf{b}\cdot\mathbf{n} = 0$), one recovers the Neumann/Neumann preconditioner, which can be showed to be exact on a two-domain decomposition with uniform velocity. A Fourier analysis shows that the Robin/Robin preconditioner shares this feature with the Neumann/Neumann one. This is assessed in the following proposition (again, see [3]).

Proposition 3.2.2 (Y. Achdou, F. Nataf) *In the case where the plane $\mathbf{R}^2$ is decomposed into the left ($\Omega_1 =]-\infty, 0[\times\mathbf{R}$) and right ($\Omega_2 =]0,+\infty[\times\mathbf{R}$) half planes and where the velocity field is uniform, we have*

$$\mathcal{T}\circ\mathcal{S} = Id.$$

Proof. For $i = 1,2$, we denote with τ_i the tangential vector to Ω_i, and we set $\mathbf{n} := \mathbf{n}_1$. We express the action of the operator $\mathcal{S}u_0$, for $u_0 \in H^{1/2}(\mathbf{R})$, by means of the Fourier transform with respect to y, and to this aim we denote with ξ the Fourier variable and with $\mathcal{F}^{-1}$ the inverse Fourier transform. The Fourier transform w.r.t. y of $(3.2.25)_1$ yields

$$\left(a + \mathbf{b}\cdot\mathbf{n}\partial_x - \nu\partial_{xx} + \mathbf{i}\mathbf{b}\cdot\tau_i\xi + \nu\xi^2\right)\hat{w}_i(x,\xi) = 0,$$

where $\mathbf{i}^2 = -1$. For a given ξ these are ODEs in x whose solutions must be bounded at infinity and satisfy the Dirichlet condition $\hat{w}(0,\xi) = \hat{u}_0(\xi)$. Thus, the solutions are $w_1 = \mathcal{F}^{-1}(\hat{u}_0(\xi)\, e^{\lambda_1^+(\xi)x})$ and $w_2 = \mathcal{F}^{-1}(\hat{u}_0(\xi)\, e^{\lambda_2^-(\xi)x})$, where, for $i = 1,2$,

$$\lambda_i^\pm = \frac{-\mathbf{b}\cdot\mathbf{n} \pm \sqrt{4a\nu + (\mathbf{b}\cdot\mathbf{n})^2 + 4\mathbf{i}\,\mathbf{b}\cdot\tau_i\,\xi\nu + 4\xi^2\nu^2}}{2\nu}.$$

Computing $\widehat{S}\hat{u}_0$, since $\partial_{n_1} = \partial_x$ and $\partial_{n_2} = -\partial_x$, we have

$$\widehat{S}\hat{u}_0 = \frac{\nu}{2}\left(\partial_x\hat{w}_1 - \partial_x\hat{w}_2\right)_{|x=0} = \frac{\nu}{2}\left(\lambda_1^+(\xi) - \lambda_2^-(\xi)\right)\hat{u}_0(\xi),$$

hence

$$\mathcal{S}(u_0) = \frac{1}{2}\mathcal{F}^{-1}\left(\sqrt{4a\nu + (\mathbf{b}\cdot\mathbf{n})^2 + 4\mathbf{i}\,\mathbf{b}\cdot\tau_i\,\xi\nu + 4\xi^2\nu^2}\;\hat{u}_0(\xi)\right).$$

In the same way it is possible to compute $\mathcal{T}(g)$, for $g \in H^{-1/2}(\mathbf{R})$, and the Robin condition at $x = 0$ entails that the solutions v_1 and v_2 of problem (3.2.29) may be expressed as $v_1 =$

$\mathcal{F}^{-1}\left(\frac{2\hat{g}(\xi)}{\sqrt{4a\nu+(\mathbf{b}\cdot\mathbf{n})^2+4\mathbf{i}\,\mathbf{b}\cdot\tau_i\,\xi\nu+4\xi^2\nu^2}}\,e^{\lambda_1^+(\xi)x}\right)$ and $v_2 = \mathcal{F}^{-1}\left(\frac{2\hat{g}(\xi)}{\sqrt{4a\nu+(\mathbf{b}\cdot\mathbf{n})^2+4\mathbf{i}\,\mathbf{b}\cdot\tau_i\,\xi\nu+4\xi^2\nu^2}}\,e^{\lambda_2^-(\xi)x}\right)$. Finally, since $\widehat{\mathcal{T}}(\hat{g}) = \frac{1}{2}(\hat{v}_1(0,\xi) + \hat{v}_2(0,\xi))$, we have

$$\mathcal{T}(g) = 2\mathcal{F}^{-1}\left(\frac{1}{\sqrt{4a\nu + (\mathbf{b}\cdot\mathbf{n})^2 + 4\mathbf{i}\,\mathbf{b}\cdot\tau_i\,\xi\nu + 4\xi^2\nu^2}}\hat{g}(\xi)\right),$$

and the thesis follows. □

As long as the domain Ω is subdivided into strips, the Robin/Robin preconditioner is very close to a nilpotent operator, whose nilpotency is the number of subdomains. We set $\mathbf{n} := (1,0)$, we introduce the space $\mathcal{H}_N^s = \prod_{i=1}^{N-1} H^s(\Gamma_{i,i+1})$, $s \in \mathbf{R}$, consisting of H^s functions on the $N-1$ interfaces: as, for $U \in \mathcal{H}_N^s$ we have $U = (u_i)_{1\le i\le N-1}$, the space $\mathcal{H}_N^s$ is endowed with the norm $\|U\|_{\mathcal{H}_N^s} = \sup_{1\le i\le N-1}\|u_i\|_{H^s(\Gamma_{i,i+1})}$ and we can state the following result, which is proved in [2].

Theorem 3.2.3 (Y. Achdou, P. Le Tallec, F. Nataf, M. Vidrascu) *Let $\Omega = (0,H_\Omega)\times\mathbf{R}$ be decomposed into N non-overlapping vertical strips $\Omega_k = (l_i, l_{i+1})\times\mathbf{R}$, $0 \le i \le N-1$, with interfaces $\Gamma_{i,i+1} = \{l_{i+1}\}\times\mathbf{R}$, and let $H = \min_k(l_{i+1}-l_i)$ be the size of the smallest subdomain. Assume that the velocity field $\mathbf{b}$ is uniform, that*

$$\epsilon \equiv \exp\{-(\mathbf{b}\cdot\mathbf{n} + \sqrt{(\mathbf{b}\cdot\mathbf{n})^2 + 4a\nu})H/\nu\} < 1, \quad \textit{and} \quad \rho \equiv \frac{3N\epsilon}{(1-\epsilon)^{N+1}} < 1.$$

Then, for $n \ge [N/2]$, we have

$$\|(\mathcal{T}\circ\mathcal{S} - Id)^n\|_{L(\mathcal{H}_N^s)} \le \frac{N}{2(1-\rho)}\,\frac{1}{(1-\epsilon)^{N-2}}\,\rho^{[n/[N/2]]-1},$$

where $[x]$ denotes the integer part of a real number x. □

3.3 Heterogeneous Advection-Diffusion Problems

This section contains the original contribution of the thesis to the substructuring approach for advection-diffusion problems, which has been developed in collaboration with P. Le Tallec and F. Nataf at the CMAP of the École Polytechnique in Paris.

We consider an advection-diffusion problem with discontinuous viscosity coefficients. Such problems arise from the modeling of transport and diffusion of a species through heterogeneous media, where different materials with different physical properties are present in the computational domain. These differences may be rather significant, and this would reflect into large discontinuities for the coefficients of the problem. For instance, the project "Couplex" of the French National Agency ANDRA deals with the modeling of the far field simulation of a nuclear waste disposal, constituted by a central layer of clay (with a viscosity coefficient of order 10^{-7} m^2/year) which contains the repository, and is surrounded by layers of dogger, marble and limestone (where the viscosity is of order 10^{-4} m^2/year), but there are some problems in physics and engineering with even larger jumps in the coefficients.

The idea is to extend both the generalized Neumann/Neumann preconditioner, introduced in [75], which deals with heterogeneity in the coefficients, and the Robin/Robin preconditioner, described in the previous section, which is especially suited for non-symmetric problems, in order to obtain a preconditioner whose performance is not affected by the amplitude of the jumps in the coefficients. Although each subregion can be furtherly subdivided into smaller subdomains, the basic decomposition is driven by the physics of the problem, and we focus in this section on the treatment of the interfaces of discontinuity for the viscosity coefficients. The results of this section can be found, in shorter form, in [57] and [58].

3.3.1 The Domain Decomposition Algorithm

Let Ω be a bounded domain in $\mathbf{R}^d$, $(d = 2, 3)$, which we assume to be partitioned into two non-overlapping subdomains Ω_1 and Ω_2, with interface Γ. We consider the following general advection-diffusion problem

$$\begin{aligned} -\mathrm{div}\,(\nu(x)\nabla u) + \mathbf{b}\cdot\nabla(u) + au = &\ f && \text{in } \Omega \\ u = &\ 0 && \text{on } \partial\Omega_D \\ \nu\frac{\partial u}{\partial n} = &\ \varphi && \text{on } \partial\Omega_N \end{aligned} \tag{3.3.1}$$

where the function $\nu(x)$ is piecewise constant

$$\nu(x) = \begin{cases} \nu_1 & \text{if } x \in \Omega_1 \\ \nu_2 & \text{if } x \in \Omega_2 \end{cases}$$

and where the reaction term may arise from an implicit time discretization of the evolution problem, and represent the inverse of the time step (*i.e.* $a = 1/\Delta t$). Throughout the rest of the chapter we assume, without loss of generality, that $\nu_1 < \nu_2$.

We propose a domain decomposition algorithm of Robin/Robin type, and we introduce, at the continuous level, the global interface operator

$$\begin{aligned} \Sigma : H^{1/2}_{00}(\Gamma) \times L^2(\Omega) \times L^2(\partial\Omega) &\longrightarrow H^{-1/2}(\Gamma) \\ (u_\Gamma, f, \varphi) &\longmapsto \left(\nu_1\frac{\partial w_1}{\partial n_1} + \nu_2\frac{\partial w_2}{\partial n_2}\right)_\Gamma, \end{aligned} \tag{3.3.2}$$

w_j $(j = 1, 2)$ being the solution of problem

$$\begin{aligned} L_j w_j &= f && \text{in } \Omega_j \\ w_j &= 0 && \text{on } \partial\Omega_D \cap \partial\Omega_j \\ \nu_j\frac{\partial w_j}{\partial n} &= \varphi && \text{on } \partial\Omega_N \cap \partial\Omega_j \\ w_j &= u_\Gamma && \text{on } \Gamma, \end{aligned} \tag{3.3.3}$$

where we have denoted with L_j $(j = 1, 2)$ the operators

$$L_j w := -\nu_j \Delta w + \mathbf{b} \cdot \nabla w + aw. \tag{3.3.4}$$

Since the operator Σ is linear with respect to each variable, we can easily reduce (3.3.1) to the Steklov-Poincaré formulation of a coupled problem on the interface

$$\mathcal{S}(u_\Gamma) = \chi, \tag{3.3.5}$$

where we have set $\mathcal{S}(.) := \Sigma(., 0, 0)$ as well as $\chi := -\Sigma(0, f, \varphi)$. We split the operator $\mathcal{S}$ into the sum of two Dirichlet to Robin local operators, $\mathcal{S} = \mathcal{S}_1 + \mathcal{S}_2$, with

$$\mathcal{S}_j : u_\Gamma \mapsto \left(\nu_j \frac{\partial w_j}{\partial n_j} - \frac{\mathbf{b} \cdot \mathbf{n}_j}{2} w_j \right)_\Gamma \quad \text{for } j = 1, 2. \tag{3.3.6}$$

Since $\mathbf{n}_1 = -\mathbf{n}_2$, the terms $\frac{1}{2}\mathbf{b} \cdot \mathbf{n}_j \, u_j$ vanish in the sum and we recover the operator $\mathcal{S}$.
Following [2], we propose as a preconditioner for the Steklov-Poincaré equation at the continuous level an approximate inverse of $\mathcal{S}$, which is the weighted sum of the inverses of the operators $\mathcal{S}_1$ and $\mathcal{S}_2$, namely

$$\mathcal{T} = D_1 \mathcal{S}_1^{-1} D_1 + D_2 \mathcal{S}_2^{-1} D_2 \tag{3.3.7}$$

where D_1 and D_2 are two suitable operators on the interface satisfying $D_1 + D_2 = Id$, and must be able to handle large jumps in the viscosity coefficients. The operator $\mathcal{T}$ is therefore defined as follows

$$\begin{aligned} \mathcal{T} : H^{-1/2}(\Gamma) &\longrightarrow H^{1/2}_{00}(\Gamma) \\ g &\longmapsto D_1 \, (v_1)_\Gamma + D_2 \, (v_2)_\Gamma \end{aligned} \tag{3.3.8}$$

where v_j $(j = 1, 2)$ is the solution of

$$\begin{aligned} \mathcal{L}_j(v_j) &= 0 && \text{in } \Omega_j \\ v_j &= 0 && \text{on } \partial\Omega_D \cap \partial\Omega_j \\ \nu_j \frac{\partial v_j}{\partial n} &= 0 && \text{on } \partial\Omega_N \cap \partial\Omega_j \\ \left(\nu_j \frac{\partial v_j}{\partial n_j} - \frac{\mathbf{b} \cdot \mathbf{n}_j}{2} v_j \right)_\Gamma &= D_j \, (g) && \text{on } \Gamma \end{aligned} \tag{3.3.9}$$

In the following section, by means of Fourier techniques, we study the choice of the weighting interface operators.

3.3.2 Construction of the preconditioner

Recalling that the Robin/Robin preconditioner is exact in a two half-plane decomposition of $\mathbf{R}^2$, (see Proposition 3.2.2), in this section we consider the case where $\Omega = \mathbf{R}^2$ is decomposed into the left ($\Omega_1 =]-\infty, 0[\times \mathbf{R}$) and right ($\Omega_2 =]0, +\infty[\times \mathbf{R}$) half-planes, we assume the convective field to be uniform and directed from Ω_1 to Ω_2 perpendicularly to the interface, *i.e.* $\mathbf{b} = (b_x, 0)$, and we assume that the solution u of problem (3.1.1) is bounded as $|x| \to +\infty$.
We can express the action of the operator $\mathcal{S}$ in terms of its Fourier transform in the y direction as

$$\mathcal{S}u_\Gamma = \mathcal{F}^{-1}\left(\hat{\mathcal{S}}(\xi)\hat{u}_\Gamma(\xi)\right), \quad u_\Gamma \in H_{00}^{1/2}(\Gamma)$$

where we have denoted with ξ the Fourier variable and with $\mathcal{F}^{-1}$ the inverse Fourier transform. We consider, for $j = 1, 2$, the problem

$$\begin{aligned} L_j(w_j) &= 0 \qquad \text{in } \Omega_j \\ w_j &= u_\Gamma \qquad \text{on } \Gamma, \end{aligned} \tag{3.3.10}$$

and we have to compute the Fourier transform of $(\nu_1(\partial w_1/\partial n_1) + \nu_2(\partial w_2/\partial n_2))_\Gamma$. Performing a Fourier transform in the y direction on the operators L_j, we get

$$\left(a + b_x\partial_x - \nu_j\partial_{xx} + \nu_j\xi^2\right)\hat{w}_j(x,\xi) = 0, \tag{3.3.11}$$

for $j = 1, 2$. For a given ξ, equation (3.3.11) is an ordinary differential equation in x whose solutions have the form $\alpha_j(\xi)\exp\{\lambda_j^-(\xi)x\} + \beta_j(\xi)\exp\{\lambda_j^+(\xi)x\}$, where

$$\lambda_j^{\pm}(\xi) = \frac{b_x \pm \sqrt{b_x^2 + 4a\nu_j + 4\nu_j^2\xi^2}}{2\nu_j} \tag{3.3.12}$$

The boundedness assumption on the solutions w_j ($j = 1, 2$) for $x \to \pm\infty$, implies $\alpha_1(\xi) = \beta_2(\xi) = 0$, while the Dirichlet condition on the interface provides $\beta_1(\xi) = \alpha_2(\xi) = \hat{u}_\Gamma$. Hence

$$\begin{aligned} \nu_1\left(\frac{\partial\hat{w}_1}{\partial n_1}\right)_\Gamma &= \nu_1\left(\frac{\partial\hat{w}_1}{\partial x}\right)_{|x=0} \\ &= \frac{1}{2}\hat{u}_\Gamma\left(b_x + \sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}\right) \end{aligned}$$

In a similar way we get

$$\begin{aligned} \nu_2\left(\frac{\partial\hat{w}_2}{\partial n_2}\right)_\Gamma &= \nu_2\left(-\frac{\partial\hat{w}_2}{\partial x}\right)_{|x=0} \\ &= -\frac{1}{2}\hat{u}_\Gamma\left(b_x - \sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}\right) \end{aligned}$$

Therefore we have the following expression for $\hat{\mathcal{S}}$:

$$\hat{\mathcal{S}}\hat{u}_\Gamma = \frac{1}{2}\left(\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2} + \sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}\right)\hat{u}_\Gamma \tag{3.3.13}$$

In order to find the operators D_1 and D_2 in (3.3.7), we look for an operator $\mathcal{T}$ that, in the Fourier space, can be represented as

$$\hat{\mathcal{T}} := \hat{\mathcal{T}}_{d_1,d_2} = d_1\hat{\mathcal{S}}_1^{-1}d_1 + d_2\hat{\mathcal{S}}_2^{-1}d_2, \tag{3.3.14}$$

and whose symbol is therefore given by

$$\hat{\mathcal{T}}\hat{g} = 2\left(\frac{d_1^2}{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}} + \frac{d_2^2}{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}}\right)\hat{g}. \tag{3.3.15}$$

Hence, we immediately have that

$$Symb(\mathcal{T}\circ\mathcal{S}) = d_1^2\left(1 + \frac{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}}{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}}\right) + d_2^2\left(1 + \frac{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}}{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}}\right) \tag{3.3.16}$$

We recall that the condition number of a matrix A is given by

$$\kappa_2(A) := ||A||_2\,||A^{-1}||_2 = \frac{\sqrt{\lambda_{\max}(A^TA)}}{\sqrt{\lambda_{\min}(A^TA)}}.$$

Thus, the condition number of the discrete version of the operator $(\mathcal{T}\circ\mathcal{S})$ can be estimated by means of its Fourier transform, as

$$\mathrm{cond}(\mathcal{T}\circ\mathcal{S}) \sim \frac{\sqrt{\max_\xi |Symb(\mathcal{T}\circ\mathcal{S})|^2}}{\sqrt{\min_\xi |Symb(\mathcal{T}\circ\mathcal{S})|^2}}.$$

Since the operator $(\mathcal{T}\circ\mathcal{S})$ is symmetric with respect to y, its symbol is real and positive, thus the monotonicity of the square root allows us to estimate the condition number of the discrete version of the operator $(\mathcal{T}\circ\mathcal{S})$ as

$$\mathrm{cond}(\mathcal{T}\circ\mathcal{S}) \sim \frac{\max_\xi Symb(\mathcal{T}\circ\mathcal{S})}{\min_\xi Symb(\mathcal{T}\circ\mathcal{S})}.$$

In the following, therefore, we focus on the choice of the weights d_1 and d_2 in order to achieve a good conditioning of the operator $\mathcal{T}\circ\mathcal{S}$, *i.e.* $\mathrm{cond}(\mathcal{T}\circ\mathcal{S}) \leq K$, with K a constant independent of the coefficients of the problem, as well as good parallelization properties for the algorithm.

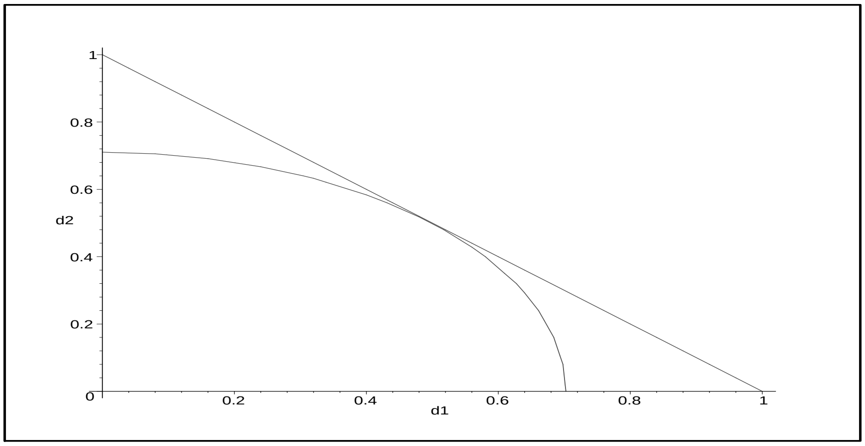

Figure 3.1: $a = b_x = 1, \nu_1 = 10^{-6}, \nu_2 = 10^{-2}, \xi = 3$

Exactness on a two domains decomposition

Since the original Robin/Robin algorithm provides exact preconditioning on a two domain decomposition, an immediate choice is to look for d_1 and d_2 such that $\mathcal{T}$ shares the same feature. This is equivalent to solve the following system

$$\left\{\begin{aligned} d_1^2\left(1+\frac{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}}{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}}\right)+d_2^2\left(1+\frac{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}}{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}}\right) &= 1 \\ d_1+d_2 &= 1 \end{aligned}\right. \tag{3.3.17}$$

Notice that, for a given ξ, the first equation in (3.3.17) describes, in the (d_1, d_2) plane, an ellipse, to which the straight line $d_1 + d_2 = 1$ is tangent. As an example, see Figure 3.1, where we have plotted the ellipse and the line for positive values of d_1 and d_2 (which are the ones we are interested in) with the choices $a = b = 1, \nu_1 = 10^{-6}, \nu_2 = 10^{-2}, \xi = 3$.
Therefore the solution of system (3.3.17) is unique and is given, for each ξ, by

$$d_1(\xi) = \frac{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}}{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}+\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}}$$

$$d_2(\xi) = \frac{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}}{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}+\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}}$$

It is not difficult to observe that the functions $d_1(\xi)$ and $d_2(\xi)$ are symmetric on the (d, ξ) plane with respect to the line $d = 1/2$, are bounded, monotone ($d_1(\xi)$ decreasing, and $d_2(\xi)$ increasing, respectively) and tend to the values $\nu_1/(\nu_1 + \nu_2)$ and $\nu_2/(\nu_1 + \nu_2)$ respectively, as ξ tends to infinity. In Figure 3.2 we have plotted the behavior of $d_2(\xi)$ with the same coefficients used in Figure 3.1.

Exactness for the two domain decomposition setting can therefore be achieved in a unique way as a function of the Fourier variable ξ. Unfortunately this is not satisfactory, since if we consider the operators in the physical space associated to the symbols $d_1(\xi)$ and $d_2(\xi)$, we have

$$\begin{aligned} D_1 &= \mathcal{S}_1 \circ (\mathcal{S}_1 + \mathcal{S}_2)^{-1} \\ D_2 &= \mathcal{S}_2 \circ (\mathcal{S}_1 + \mathcal{S}_2)^{-1} \end{aligned}$$

It is evident that the two operators above depend on both subdomains, hence a parallel algorithm based upon D_1 and D_2 cannot be carried on, since we cannot express the operator $(\mathcal{S}_1 + \mathcal{S}_2)^{-1}$ in terms of a boundary value PDE in each subdomain. A different choice is therefore in order.

Approximation of $d_1(\xi)$ and $d_2(\xi)$

For sake of simplicity at the computational level, we look for constant approximations of the functions $d_1(\xi)$ and $d_2(\xi)$, which amounts to the multiplication by a constant in the physical space.

We define

$$\mathcal{T}_d = \mathcal{F}^{-1}\left(\hat{\mathcal{T}}_d\right) \tag{3.3.18}$$

where, taking into account the normalization condition $d_1 + d_2 = 1$ and the assumption $\nu_1 < \nu_2$, we have set $d_2 = d$, $d_1 = 1 - d$, and where, for sake of simplicity, we have set $\hat{\mathcal{T}}_d := \hat{\mathcal{T}}_{1-d,d}$, the latter being defined in (3.3.14).

Remark 3.3.1 If $\nu_1 > \nu_2$, a symmetric argument stemming from the choice $d_1 = d$, $d_2 = 1 - d$ would lead the same results we present in the following. □

From (3.3.18) we reduce $Symb(\mathcal{T}_d \circ \mathcal{S})$ in (3.3.16) to a bivariate function $F(d, \xi)$

$$F(d,\xi) = (1-d)^2\left(1 + \frac{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}}{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}}\right) + d^2\left(1 + \frac{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}}{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}}\right) \tag{3.3.19}$$

which is defined on the subset of the (d, ξ) plane given by $[1/2, \nu] \times [0, \xi_{max}]$, where we have set

$$\nu := \frac{\nu_2}{\nu_1 + \nu_2}. \tag{3.3.20}$$

Notice that the bounds on the variable d stem from the boundedness of the function $d_2(\xi)$ that we want to approximate. On the other hand, the function $F(d, \xi)$ is symmetric in ξ and this allows

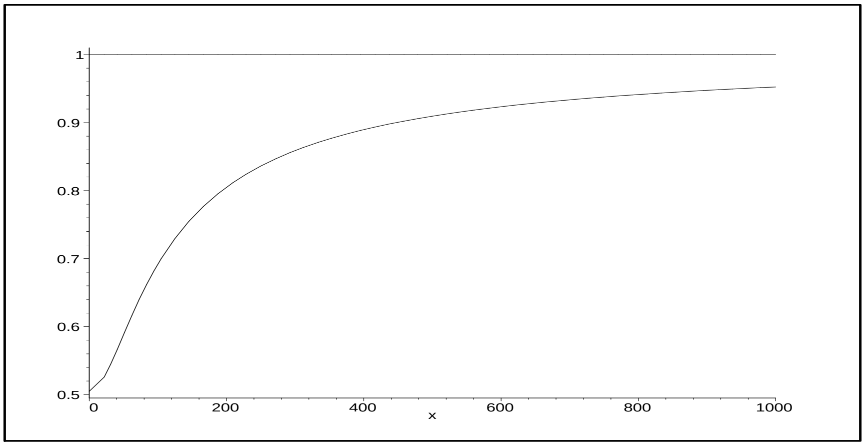

Figure 3.2: $d_2(\xi)$, with $a = b_x = 1, \nu_1 = 10^{-6}, \nu_2 = 10^{-2}$

us to consider only positive values of ξ, whereas ξ_{max} denotes the largest frequency supported by the numerical grid: in that order, we recall that the minimal representable wavelength is $\lambda_{\min} = 2h$, where h is the mesh size, thus $\xi_{\max}$ is of order π/h. Since $d_2(0) \le d_2(\xi) \le d_2(\xi_{max})$, and is monotonically increasing, we have to face three different cases.

$$\textbf{Case 1}: \qquad d \in \left[\frac{1}{2}, d_2(0)\right]$$

Let the constant d be fixed in the interval $[1/2, d_2(0)[$: the function $F(d,\xi)$ is shown to be strictly increasing in ξ. Infact, for a given d, a tedious rather than difficult computation provides

$$\begin{aligned}\partial_\xi F(d,\xi) = 4\,\xi\, &\frac{\left[b_x^2(\nu_1+\nu_2) + 4a\nu_1\nu_2\right]\,(\nu_2-\nu_1)}{\left(\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}\,\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}\right)^3}\times \\ &\times\left(\frac{1-d}{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}} + \frac{d}{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2}}\right)\times \\ &\times\left[(1-d)\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi^2} - d\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi^2}\right],\end{aligned} \tag{3.3.21}$$

whose sign is ruled by the last factor in the product which, since $\nu_2 > \nu_1$, is positive.
Thus, we have

$$\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S}) \sim \frac{F(d,\xi_{max})}{F(d,0)} = \mathcal{G}(d).$$

Setting

$$A := \frac{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi_{max}^2}+\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi_{max}^2}}{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi_{max}^2}}\;\frac{\sqrt{b_x^2+4a\nu_1}}{\sqrt{b_x^2+4a\nu_1}+\sqrt{b_x^2+4a\nu_2}}$$

and noting that $A>0$, we can expand $\mathcal{G}(d)$ as

$$\mathcal{G}(d) = A\,\frac{(1-d)^2+\dfrac{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi_{max}^2}}{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi_{max}^2}}\,d^2}{(1-d)^2+\dfrac{\sqrt{b_x^2+4a\nu_1}}{\sqrt{b_x^2+4a\nu_2}}\,d^2}$$

We have, for each $d\in[1/2,d_2(0)]$:

$$\partial_d\mathcal{G}(d) = \frac{2Ad(1-d)}{\left[(1-d)^2+\frac{\sqrt{b_x^2+4a\nu_1}}{\sqrt{b_x^2+4a\nu_2}}\,d^2\right]^2}\left(\frac{\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi_{max}^2}}{\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi_{max}^2}}-\frac{\sqrt{b_x^2+4a\nu_1}}{\sqrt{b_x^2+4a\nu_2}}\right)<0$$

since the sign is ruled by the last factor in the product which (under our assumption $\nu_1<\nu_2$) is negative. Therefore, $\mathrm{cond}(\mathcal{T}_d\circ\mathcal{S})$ is decreasing as a function of d in $[1/2,d_2(0)]$. If, following what is done in [2] when the viscosity coefficients coincide ($\nu_1=\nu_2$), we set $d=1/2$, we get

$$F\left(\frac{1}{2},\xi\right)=\frac{1}{2}+\frac{1}{4}\left(\sqrt{\frac{b_x^2+4a\nu_1+4\nu_1^2\xi^2}{b_x^2+4a\nu_2+4\nu_2^2\xi^2}}+\sqrt{\frac{b_x^2+4a\nu_2+4\nu_2^2\xi^2}{b_x^2+4a\nu_1+4\nu_1^2\xi^2}}\right).$$

In order to evaluate the performance of the preconditioner $\mathcal{T}_{1/2}$, let us consider what happens if $b_x=0$. Assuming $a\neq0$, we have:

$$\mathrm{cond}(\mathcal{T}_{1/2}\circ\mathcal{S})\sim\frac{2+\sqrt{\frac{\nu_1(1+\eta\nu_1)}{\nu_2(1+\eta\nu_2)}}+\sqrt{\frac{\nu_2(1+\eta\nu_2)}{\nu_1(1+\eta\nu_1)}}}{2+\sqrt{\frac{\nu_1}{\nu_2}}+\sqrt{\frac{\nu_2}{\nu_1}}} \tag{3.3.22}$$

where we have set $\eta:=\frac{\xi_{max}^2}{a}=\Delta t\,\xi_{max}^2$. Developing the ratio in (3.3.22), we can easily estimate $\mathrm{cond}(\mathcal{T}_{1/2}\circ\mathcal{S})$ with

$$\left[\frac{\sqrt{\nu_1}}{\sqrt{\nu_1}+\sqrt{\nu_2}}\right]^2\sqrt{\frac{1+\eta\nu_1}{1+\eta\nu_2}}+2\frac{\sqrt{\nu_1}\sqrt{\nu_2}}{\left[\sqrt{\nu_1}+\sqrt{\nu_2}\right]^2}+\left[\frac{\sqrt{\nu_2}}{\sqrt{\nu_1}+\sqrt{\nu_2}}\right]^2\sqrt{\frac{1+\eta\nu_2}{1+\eta\nu_1}} \tag{3.3.23}$$

where the first two terms in the sum are bounded between 0 and 1, while for the last one we have:

$$\left[\frac{\sqrt{\nu_2}}{\sqrt{\nu_1}+\sqrt{\nu_2}}\right]^2\sqrt{\frac{1+\eta\nu_2}{1+\eta\nu_1}}>\frac{1}{4}\sqrt{\frac{1+\eta\nu_2}{1+\eta\nu_1}}.$$

If $\eta\nu_1 \gg 1$, we have

$$\mathrm{cond}(\mathcal{T}_{1/2} \circ \mathcal{S}) \ > \ \frac{1}{4}\sqrt{\frac{\nu_2}{\nu_1}},$$

which turns out to be very large when $\nu_1 \ll \nu_2$.

Case 2: $\quad d \in [d_2(0), d_2(\xi_{\max})]$

Let the constant d be fixed in the interval $[d_2(0), d_2(\xi_{max})]$. We can prove the following result.

Lemma 3.3.1 *Let $\mathcal{T}_d$ be the operator whose Fourier transform is defined in* (3.3.14) *with $d_1 = 1 - d$ and $d_2 = d$. Then, for each $d \in [d_2(0), d_2(\xi_{max})]$, the condition number of the discrete version of the operator $(\mathcal{T}_d \circ \mathcal{S})$ satisfies the following estimate*

$$1 < \mathrm{cond}(\mathcal{T}_d \circ \mathcal{S}) \lesssim \max\left\{F(d_2(0), \xi_{max}), F\,(d_2(\xi_{max}), 0)\right\}. \tag{3.3.24}$$

and there exists $d_0 \in]d_2(0), d_2(\xi_{max})[$ which minimizes $\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S})$.

Proof. Developing the square $(1-d)^2$, and gathering together all the terms involving d and ξ, $F(d, \xi)$ can be written as

$$F(d,\xi) = 1 + \frac{\left[(1-d)\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2} - d\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}\right]^2}{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2}\,\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2}}. \tag{3.3.25}$$

As an immediate consequence of (3.3.25) we have

$$F(d, \xi) \geq 1$$

$\forall (d, \xi) \in [d_2(0), d_2(\xi_{max})]$, and the value 1 is attained along the curve $d = d_2(\xi)$.

Owing to (3.3.21), when d is fixed in the interval $[d_2(0), d_2(\xi_{max})]$, $\partial_\xi F(d, \xi)$ is positive for $\xi > d_2^{-1}(d)$, negative for $\xi < d_2^{-1}(d)$ and vanishes in $\xi_0 := d_2^{-1}(d)$, *i.e.* on the curve $d = d_2(\xi)$, hence the function $F(d, \xi)$ is strictly decreasing in $[0, \xi_0[$, attains its minimum in ξ_0 and is increasing in $[\xi_0, \xi_{max}]$. Therefore, for a given $d \in [d_2(0), d_2(\xi_{max})]$, the condition number of the operator $\mathcal{T}_d \circ \mathcal{S}$ is easily estimated as

$$\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S}) \sim \max\left\{F(d, 0), F(d, \xi_{max})\right\}, \tag{3.3.26}$$

and the lower estimate in (3.3.17) follows.
If we focus on the functions $F(d, 0)$ and $F(d, \xi_{max})$, it is not difficult to see that in $[d_2(0), d_2(\xi_{max})]$ the first one is increasing , while the latter one is decreasing. Since $F(d_2(0), 0) = F(d_2(\xi_{max}), \xi_{max}) = 1$, there exists d_0 such that

$$F(d_0, 0) = F(d_0, \xi_{max}). \tag{3.3.27}$$

Hence,

$$\max_\xi F(d, \xi) = \begin{cases} F(d, \xi_{max}) & \text{for } d \in [d_2(0), d_0[\\ \\ F(d, 0) & \text{for } d \in [d_0, d_2(\xi_{max})[\end{cases} \tag{3.3.28}$$

The upper estimate in (3.3.24) follows. In order to evaluate d_0, let us go back to equation (3.3.27): it is easy to see that the only solution occurring in the interval $]d_2(0), d_2(\xi_{max})[$ is

$$d_0 = \frac{\left[\left(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi_{max}^2\right)\left(b_x^2 + 4a\nu_2\right)\right]^{1/4}}{\left[\left(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi_{max}^2\right)\left(b_x^2 + 4a\nu_2\right)\right]^{1/4} + \left[\left(b_x^2 + 4a\nu_1 + 4\nu_1^2\xi_{max}^2\right)\left(b_x^2 + 4a\nu_1\right)\right]^{1/4}},$$

while the other one

$$d^* = \frac{\left[\left(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi_{max}^2\right)\left(b_x^2 + 4a\nu_2\right)\right]^{1/4}}{\left[\left(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi_{max}^2\right)\left(b_x^2 + 4a\nu_2\right)\right]^{1/4} - \left[\left(b_x^2 + 4a\nu_1 + 4\nu_1^2\xi_{max}^2\right)\left(b_x^2 + 4a\nu_1\right)\right]^{1/4}}$$

is greater than 1.
We have in conclusion that the condition number of the operator $\mathcal{T}_d \circ \mathcal{S}$ is minimum when $d = d_0$, and

$$\operatorname{cond}(\mathcal{T}_{d_0} \circ \mathcal{S}) \sim \frac{\left(\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi_{max}^2}+\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi_{max}^2}\right)\left(\sqrt{b_x^2+4a\nu_1}+\sqrt{b_x^2+4a\nu_2}\right)}{\left(\left[\sqrt{b_x^2+4a\nu_2+4\nu_2^2\xi_{max}^2}\,\sqrt{b_x^2+4a\nu_2}\right]^{1/2}+\left[\sqrt{b_x^2+4a\nu_1+4\nu_1^2\xi_{max}^2}\,\sqrt{b_x^2+4a\nu_1}\right]^{1/2}\right)^2}. \qquad (3.3.29)$$

□

Case 3: $\quad d \in [d_2(\xi_{\max}), \nu]$

When d is fixed in the interval $[d_2(\xi_{\max}), \nu]$, owing again to (3.3.21), the function $F(d,\xi)$ is strictly decreasing in ξ, therefore

$$\operatorname{cond}(\mathcal{T}_d \circ \mathcal{S}) \sim \frac{F(d,0)}{F(d,\xi_{max})} = \mathcal{H}(d).$$

Since $\mathcal{H}(d) = [\mathcal{G}(d)]^{-1}$, we have, for each $d \in \left[d_2(\xi_{\max}), \frac{\nu_2}{\nu_1+\nu_2}\right]$

$$\partial_d \mathcal{H}(d) = -\partial_d \mathcal{G}(d)\, \frac{[F(d,0)]^2}{[F(d,\xi_{max})]^2} > 0$$

where the inequality is an immediate consequence of the previous section. Therefore, the condition number of $(\mathcal{T}_d \circ \mathcal{S})$ is increasing in the interval $[d_2(\xi_{\max}), \nu]$ as a function of d.
More, since $F(\nu, \xi_{max}) > 1$, we have

$$\operatorname{cond}(\mathcal{T}_\nu \circ \mathcal{S}) < F(\nu, 0)$$

If $b_x \neq 0$, we define $\eta := \frac{4a}{b_x^2}$, and we have

$$\operatorname{cond}(\mathcal{T}_\nu \circ \mathcal{S}) < \left[\frac{\nu_1}{\nu_1+\nu_2}\right]^2 (1+\varphi(\eta)) + \left[\frac{\nu_2}{\nu_1+\nu_2}\right]^2 \left(1 + \frac{1}{\varphi(\eta)}\right) \qquad (3.3.30)$$

where

$$\varphi(\eta) := \sqrt{\frac{1+\eta\nu_2}{1+\eta\nu_1}}. \tag{3.3.31}$$

The right hand side in (3.3.30) is decreasing as a function of η, since

$$\varphi'(\eta) = \frac{1}{2}\sqrt{\frac{1+\eta\nu_1}{1+\eta\nu_2}}\,\frac{\nu_2-\nu_1}{(1+\eta\nu_1)^2} > 0$$

and it is not difficult to see that

$$\nu_1^2\,[\varphi(\eta)]^2 - \nu_2^2 < 0.$$

Hence

$$\mathrm{cond}(\mathcal{T}_\nu \circ \mathcal{S}) < \left[\frac{\nu_1}{\nu_1+\nu_2}\right]^2 (1+\varphi(0)) + \left[\frac{\nu_2}{\nu_1+\nu_2}\right]^2 \left(1+\frac{1}{\varphi(0)}\right). \tag{3.3.32}$$

We therefore have from (3.3.31) and (3.3.32)

$$\begin{aligned}\mathrm{cond}(\mathcal{T}_\nu \circ \mathcal{S}) &< 2\left[\frac{\nu_1}{\nu_1+\nu_2}\right]^2 + 2\left[\frac{\nu_2}{\nu_1+\nu_2}\right]^2 \\ &= 2\,\frac{\nu_1^2+\nu_2^2}{(\nu_1+\nu_2)^2} \\ &< 2.\end{aligned} \tag{3.3.33}$$

On the other hand, if $b_x = 0$ (*i.e.* if there is no convective term) we simply have

$$\begin{aligned}F(\nu,0) &= \left[\frac{\nu_1}{\nu_1+\nu_2}\right]^2\left(1+\sqrt{\frac{\nu_2}{\nu_1}}\right) + \left[\frac{\nu_2}{\nu_1+\nu_2}\right]^2\left(1+\sqrt{\frac{\nu_1}{\nu_2}}\right) \\ &< \frac{1}{(\nu_1+\nu_2)^2}\left[\nu_1^2+\nu_2^2+(\nu_1+\nu_2)\sqrt{\nu_1\nu_2}\right] \\ &\le \frac{\nu_1^2+\nu_2^2}{(\nu_1+\nu_2)^2} + \frac{1}{2} \\ &< 1+\frac{1}{2} < 2.\end{aligned} \tag{3.3.34}$$

Since $F(d_2(0),0) = F(d_2(\xi_{max}),\xi_{max}) = 1$, the continuity of $F(d,\xi)$ guarantees the continuity of $\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S})$, as a function of d, in the whole interval $[1/2,\nu]$. We have therefore proved the following result.

Interval	$[1/2, d_2(0)]$	$[d_2(0), d_0]$	$[d_0, d_2(\xi_{\max})]$	$[d_2(\xi_{\max}), \nu]$
$\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S})$	$\dfrac{F(d,\xi_{max})}{F(d,0)}$	$F(d,\xi_{max})$	$F(d,0)$	$\dfrac{F(d,0)}{F(d,\xi_{max})}$
Behavior	↓	↓	↑	↑

Table 3.1: Definition and behavior of $\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S})$ with respect to d

Theorem 3.3.1 *Let* $\mathcal{T}_d = \mathcal{F}^{-1}\left(\hat{\mathcal{T}}_d\right)$. *When the plane* $\mathbf{R}^2$ *is decomposed into the left and right half-planes, and the convective field* $\mathbf{b}$ *is uniform and perpendicular to the interface, the condition number* $\mathrm{cond}(\mathcal{T}_d \circ \mathcal{S})$*, as a function of* d*, is optimal in*

$$d_0 = \frac{\left[(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi_{max}^2)\,(b_x^2 + 4a\nu_2)\,\right]^{1/4}}{\left[(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi_{max}^2)\,(b_x^2 + 4a\nu_2)\,\right]^{1/4} + \left[(b_x^2 + 4a\nu_1 + 4\nu_1^2\xi_{max}^2)\,(b_x^2 + 4a\nu_1)\,\right]^{1/4}},$$

where the condition number is given in (3.3.29)*, is decreasing in the interval* $[1/2, d_0)$*, and increasing in the interval* $(d_0, \nu]$.
Moreover, we have

$$\mathrm{cond}(\mathcal{T}_\nu \circ \mathcal{S}) < 2, \tag{3.3.35}$$

independently of the coefficients of the problem.

□

We have resumed the definition and the behavior of the estimate condition number of $\mathcal{T}_d \circ \mathcal{S}$ in Table 3.1, but a few comments on the above result are in order. We have shown that the choice $d = 1/2$, which is natural in the case of constant viscosity may lead to problems whose conditioning is very bad. The optimal choice of the weight, $d = d_0$, depends on all the coefficients of the problem; moreover, d_0 depends also on $\xi_{\max}$, thus it depends on the mesh parameter h. On the other hand, the weight ν (3.3.20) depends only on the viscosity coefficients, where the discontinuity is located. For sake of generality and simplicity of implementation, since also the conditioning of $\mathcal{T}_\nu \circ \mathcal{S}$ is very good, we choose as a preconditioner the operator $\mathcal{T}_\nu$. Notice that such weights are exactly the ones used by P. Le Tallec *et al.* in [75] when introducing the *Generalized Neumann/Neumann* method for elasticity problems. Such choice pays a major attention to the information stemming from the subdomain in which the viscosity is larger: this is not so evident when the viscosity jump is small, but for very large jumps (say, $10^4 - 10^6$) this amounts to almost neglect the contribution of one of the two subdomains. This is not so strange, however, since for advection-dominated flows, the problem in the less viscous region in the presence of large jumps of the viscosity is very close to a pure transport problem.

b_x	a	d_0	ν	cond$(\mathcal{T}_{1/2} \circ \mathcal{S})$	cond$(\mathcal{T}_{d_0} \circ \mathcal{S})$	cond$(\mathcal{T}_{\nu} \circ \mathcal{S})$
1	1	.999715601	.99999900	$\sim$ 3004.798211	$\sim$ 1.015268461	$\sim$ 1.015835611
1	0.1	.999495544	.99999900	$\sim$ 8884.104495	$\sim$ 1.048893915	$\sim$ 1.049942042
1	10	.999837253	.99999900	$\sim$ 971.3739450	$\sim$ 1.004777051	$\sim$ 1.005093540
0	1	.999967615	.99999900	$\sim$ 951.5605638	$\sim$ 1.000936218	$\sim$ 1.000997997
0	0.1	.999968299	.99999900	$\sim$ 993.0520980	$\sim$ 1.000937543	$\sim$ 1.000997999

Table 3.2: Condition Numbers for $\nu_1 = 10^{-3}$, $\nu_2 = 10^{3}$, $\xi_{max} = 100$

b_x	a	d_0	ν	cond$(\mathcal{T}_{1/2} \circ \mathcal{S})$	cond$(\mathcal{T}_{d_0} \circ \mathcal{S})$	cond$(\mathcal{T}_{\nu} \circ \mathcal{S})$
1	1	.602083675	.99990001	$\sim$ 1.172493844	$\sim$ 1.037780797	$\sim$ 1.370231679
1	0.1	.602083675	.99990001	$\sim$ 1.170997997	$\sim$ 1.037780797	$\sim$ 1.370231679
1	10	.623797395	.99990001	$\sim$ 1.180152596	$\sim$ 1.026982600	$\sim$ 1.290022659
0	1	.996847690	.99990001	$\sim$ 9.822576218	$\sim$ 1.004646014	$\sim$ 1.008981928
0	0.1	.998182603	.99990001	$\sim$ 29.59142199	$\sim$ 1.006665791	$\sim$ 1.009635852

Table 3.3: Condition Numbers for $\nu_1 = 10^{-6}$, $\nu_2 = 10^{-2}$, $\xi_{max} = 100$

To conclude our analysis in the case of a convective field perpendicular to the interface, we have reported in Tables 3.2-3.5 the values of d_0 and ν as well as the condition number for the operators $(\mathcal{T}_{1/2} \circ \mathcal{S})$, $(\mathcal{T}_{d_0} \circ \mathcal{S})$ and $(\mathcal{T}_{\nu} \circ \mathcal{S})$, with $\xi_{max} = 100$ and different choices for the parameters involved (b_x, a, ν_1 and ν_2). It turns out that, as we expected, the preconditioner $\mathcal{T}_{1/2}$ provides large condition numbers when $\nu_2 \, \Delta t \, \xi_{max}^2 \gg 1$ and $\nu_1 \ll \nu_2$. More, as we expected from the theoretical analysis, the performance of the preconditioner $\mathcal{T}_{\nu}$ is not affected by the growth of the ratio ν_2/ν_1, as d_0 and ν get very close each other. We observe that both preconditioner $\mathcal{T}_{d_0}$ and $\mathcal{T}_{\nu}$ perform very well and are very close to be exact.

3.3.3 Robustness with respect to the convective field

In this section we focus our attention on the effectiveness of the Robin/Robin type preconditioner $\mathcal{T}_{\nu} = \mathcal{F}^{-1}\left(\hat{\mathcal{T}_{\nu}}\right)$ (which, from now on, will be simply denoted by $\mathcal{T}$) when the convective filed is uniform but not orthogonal to the interface, namely $\mathbf{b} = (b_x, b_y)$, always with the additional

b_x	a	d_0	ν	cond$(\mathcal{T}_{1/2} \circ \mathcal{S})$	cond$(\mathcal{T}_{d_0} \circ \mathcal{S})$	cond$(\mathcal{T}_{\nu} \circ \mathcal{S})$
1	1	.829608824	.99999900	$\sim$ 5.482335873	$\sim$ 1.333314647	$\sim$ 1.757435981
1	0.1	.818812164	.99999900	$\sim$ 5.518448891	$\sim$ 1.394195189	$\sim$ 1.886383877
1	10	.870267866	.99999900	$\sim$ 4.734773820	$\sim$ 1.150535609	$\sim$ 1.378705905
0	1	.999822203	.99999900	$\sim$ 31.56162183	$\sim$ 1.000675726	$\sim$ 1.000968316
0	0.1	.999899763	.99999900	$\sim$ 99.31197087	$\sim$ 1.000809394	$\sim$ 1.000989842

Table 3.4: Condition Numbers for $\nu_1 = 10^{-7}$, $\nu_2 = 10^{-1}$, $\xi_{max} = 100$

b_x	a	d_0	ν	$\mathrm{cond}(\mathcal{T}_{1/2} \circ \mathcal{S})$	$\mathrm{cond}(\mathcal{T}_{d_0} \circ \mathcal{S})$	$\mathrm{cond}(\mathcal{T}_{\nu} \circ \mathcal{S})$
1	1	.991240729	.99999990	$\sim$ 233.8988119	$\sim$ 1.136576301	$\sim$ 1.155596278
1	0.1	.985266826	.99999990	$\sim$ 427.4783464	$\sim$ 1.405586166	$\sim$ 1.446490440
1	10	.995027989	.99999990	$\sim$ 90.69553856	$\sim$ 1.040043706	$\sim$ 1.049413925
0	1	.999982173	.99999990	$\sim$ 314.4616791	$\sim$ 1.000281568	$\sim$ 1.000315213
0	0.1	.999989758	.99999990	$\sim$ 952.8609609	$\sim$ 1.000296071	$\sim$ 1.000315866

Table 3.5: Condition Numbers for $\nu_1 = 10^{-6}$, $\nu_2 = 10$, $\xi_{max} = 100$

requirement for the solutions u_j to be bounded as $|x| \to +\infty$. A Fourier transform in the y direction on the operator L_j yields

$$\left(a + b_x\partial_x - \nu_j\partial_{xx} + ib_y\xi + \nu_j\xi^2\right)\hat{w}_j(x,\xi) = 0, \tag{3.3.36}$$

for $j = 1, 2$, where $i^2 = -1$. For a given ξ, equation (3.3.36) is again an ordinary differential equation in x whose solutions have the form $\alpha_j(\xi)\exp\{\lambda_j^-(\xi)x\} + \beta_j(\xi)\exp\{\lambda_j^+(\xi)x\}$, where

$$\lambda_j^{\pm}(\xi) = \frac{b_x \pm \sqrt{b_x^2 + 4a\nu_j + 4\nu_j^2\xi^2 + 4ib_y\nu_j\xi}}{2\nu_j}, \tag{3.3.37}$$

with $\mathrm{Re}(\lambda_j^{\pm}) \gtreqless 0$, as $\mathrm{Re}(z)$ indicates the real part of a complex number z. The boundedness assumption on the solutions w_j $(j = 1, 2)$ for $x \to \pm\infty$, still implies $\alpha_1(\xi) = \beta_2(\xi) = 0$, while the Dirichlet condition on the interface provides $\beta_1(\xi) = \alpha_2(\xi) = \hat{u}_\Gamma$. Hence, once again

$$\begin{aligned}\nu_1\left(\frac{\partial\hat{w}_1}{\partial n_1}\right)_\Gamma &= \nu_1\left(\frac{\partial\hat{w}_1}{\partial x}\right)_{|x=0} \\ &= \frac{1}{2}\hat{u}_\Gamma\left(b_x + \sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2 + 4ib_y\nu_j\xi}\right)\end{aligned}$$

as well as

$$\begin{aligned}\nu_2\left(\frac{\partial\hat{w}_2}{\partial n_2}\right)_\Gamma &= \nu_2\left(-\frac{\partial\hat{w}_2}{\partial x}\right)_{|x=0} \\ &= -\frac{1}{2}\hat{u}_\Gamma\left(b_x - \sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2 + 4ib_y\nu_j\xi}\right)\end{aligned}$$

and we have the following expression for $\hat{\mathcal{S}}$:

$$\hat{\mathcal{S}}\hat{u}_\Gamma = \frac{1}{2}\left(\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2 + 4ib_y\nu_1\xi} + \sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2 + 4ib_y\nu_2\xi}\right)\hat{u}_\Gamma \tag{3.3.38}$$

Using as approximate inverse the operator $\mathcal{T}$, we define

$$N_1 = \left[\frac{\nu_1}{\nu_1+\nu_2}\right]^2, \qquad N_2 = \left[\frac{\nu_2}{\nu_1+\nu_2}\right]^2, \tag{3.3.39}$$

and the symbol of the preconditioned operator is easily determined as

$$\begin{aligned}\Phi(\xi) = N_1 & \left(1 + \frac{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2 + 4ib_y\nu_2\xi}}{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2 + 4ib_y\nu_1\xi}}\right) \\ + N_2 & \left(1 + \frac{\sqrt{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2 + 4ib_y\nu_1\xi}}{\sqrt{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2 + 4ib_y\nu_2\xi}}\right).\end{aligned} \tag{3.3.40}$$

Remark 3.3.2 Notice that in the case of purely elliptic problems, *i.e.* when $a = b_x = b_y = 0$, we have $\Phi(\xi) \equiv 1$, which implies exact preconditioning in this simple case. The proposed preconditioner is thus an extension of the generalized Neumann/Neumann one introduced in [75]. □

Due to the presence of the first order term, the resulting linear system at the discrete level is non-symmetric, and we use an iterative method of Krylov type, such as GMRES. In that order, we recall that the reduction factor ρ in a GMRES iteration is given, for a positive real matrix A with symmetric part M, (see [94]) by

$$\rho = 1 - \frac{(\lambda_{\min}(M))^2}{\lambda_{\max}(A^TA)},$$

We can prove the following result.

Theorem 3.3.2 *Let* $\mathcal{T} = \mathcal{F}^{-1}\left(\hat{\mathcal{T}}_\nu\right)$. *In the case where the plane* $\mathbf{R}^2$ *is decomposed into the left and right half-planes and the convective field is uniform, the reduction factor for the associated* GMRES *preconditioned by* $\mathcal{T}$ *can be bounded from above by a constant independent of the time step* Δt*, the convective field* $\mathbf{b}$ *and the viscosity coefficients* ν_1 *and* ν_2*. Moreover, under the assumption* $\nu_1 < \nu_2$ *we have*

$$\rho\,(\mathcal{T}\circ\mathcal{S}) < 1 - \frac{1}{5 + 6\,(\nu_1/\nu_2)^2 + 5\,(\nu_1/\nu_2)^4}, \tag{3.3.41}$$

and, if the convective field is parallel to the interface

$$\rho\,(\mathcal{T}\circ\mathcal{S}) < 1 - \frac{1}{1 + 2\sum_{k=1}^{7}(\nu_1/\nu_2)^{k/2} + (\nu_1/\nu_2)^4} \tag{3.3.42}$$

Proof. The reduction factor for the associated GMRES algorithm preconditioned by $\mathcal{T}$ can be estimated, in the Fourier space, by the quantity

$$1 - \frac{(\min_\xi \operatorname{Re}\Phi(\xi))^2}{\max_\xi |\Phi(\xi)|^2}.$$

where $\Phi(\xi)$ is the function defined in (3.3.40). Thus, it is enough to show that

$$\frac{\max_\xi |\Phi(\xi)|^2}{(\min_\xi \operatorname{Re}\Phi(\xi))^2} \le C \tag{3.3.43}$$

with the constant C independent of a, b_x, b_y, ν_1 and ν_2. Recalling that, for a complex number $z \in \mathbb{C}$, $z^{-1} = \bar{z}/|z|^2$, the function $\Phi(\xi)$ can be written as

$$\Phi(\xi) = N_1\,[1 + z(\xi)] + N_2\left[1 + \frac{\bar{z}(\xi)}{|z(\xi)|^2}\right], \tag{3.3.44}$$

where we have set

$$z(\xi) := \sqrt{\frac{b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2 + 4ib_y\nu_2\xi}{b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2 + 4ib_y\nu_1\xi}}.$$

Hence

$$\operatorname{Re}\Phi(\xi) = N_1 + N_2 + \left[N_1 + \frac{N_2}{|z(\xi)|^2}\right]\operatorname{Re} z(\xi), \tag{3.3.45}$$

as well as

$$\operatorname{Im}\Phi(\xi) = \left[N_1 - \frac{N_2}{|z(\xi)|^2}\right]\operatorname{Im} z(\xi). \tag{3.3.46}$$

Since $\operatorname{Re} z(\xi) > 0$, we have from (3.3.45)

$$\operatorname{Re}\Phi(\xi) \ge N_1 + N_2 = \frac{\nu_1^2}{(\nu_1+\nu_2)^2} + \frac{\nu_2^2}{(\nu_1+\nu_2)^2} > \frac{\nu_2^2}{(\nu_1+\nu_2)^2} \tag{3.3.47}$$

In order to have an estimate for $\max_\xi\ |\Phi(\xi)|^2$, let us consider $|z(\xi)|$. We have

$$\begin{aligned} |z(\xi)| &= \left[\frac{(b_x^2 + 4a\nu_2 + 4\nu_2^2\xi^2)^2 + (4b_y\nu_2\xi)^2}{(b_x^2 + 4a\nu_1 + 4\nu_1^2\xi^2)^2 + (4b_y\nu_1\xi)^2}\right]^{1/4} \\ &= \left(\frac{[b_x^2 + 4a\nu_2]^2 + 8\,[b_x^2 + 2b_y^2 + 4a\nu_2]\,\nu_2^2\xi^2 + 16\,\nu_2^4\xi^4}{[b_x^2 + 4a\nu_1]^2 + 8\,[b_x^2 + 2b_y^2 + 4a\nu_1]\,\nu_1^2\xi^2 + 16\,\nu_1^4\xi^4}\right)^{1/4}, \end{aligned} \tag{3.3.48}$$

which is bounded, as

$$1 \,\le\, |z(\xi)| \,\le \frac{\nu_2}{\nu_1}, \tag{3.3.49}$$

and it is not difficult (altought rather tedious) to see that its first derivative is given by

$$\frac{d\,|z(\xi)|}{d\xi} = \frac{1}{2}\,|z(\xi)|^{-3/4}\,\frac{F + 2G\xi^2 + H\xi^4}{[Q(\xi)]^2}\,(\nu_2 - \nu_1)\,\xi,$$

where

$$F = \left[b_x^2(b_x^2 + 2b_y^2) + 32a^2\nu_1\nu_2\right]\,\left[b_x^2(\nu_1 + \nu_2) + 2a\nu_1\nu_2\right] + 2ab_x^2\,\left[b_x^2(2\nu_2^2 + 5\nu_1\nu_2 + 2\nu_1^2) + 6b_y^2\nu_1\nu_2\right]$$

$$G = \left[b_x^4(\nu_2^2 + \nu_1^2) + 16a^2\nu_1^2\nu_2^2\right]\,(\nu_1 + \nu_2) + 8ab_x^2\nu_1\nu_2\,\left(\nu_2^2 + \nu_1\nu_2 + \nu_1^2\right)$$

$$H = \left[b_x^2 + 2b_y^2\right]\,\nu_1^2\nu_2^2\,(\nu_2 + \nu_1) + 4a\nu_1^3\nu_2^3$$

$$Q(\xi) = [b_x^2 + 4a\nu_1]^2 + 8\,[b_x^2 + 2b_y^2 + 4a\nu_1]\,\nu_1^2\xi^2 + 16\,\nu_1^4\xi^4.$$

As the coefficients F, G and H are positive, the function $|z(\xi)|$ is decreasing in $(-\infty, 0)$, increasing in $(0, +\infty)$, and we have

$$\min_{\xi} |z(\xi)| \;=\; |z(0)| = \sqrt{\frac{b_x^2 + 4a\nu_2}{b_x^2 + 4a\nu_1}}, \quad \sup_{\xi} |z(\xi)| \;=\; \lim_{\xi\to\pm\infty} |z(\xi)| = \frac{\nu_2}{\nu_1}.$$

So far, we focus on $|\Phi(\xi)|^2$ to prove its boundedness from above. We have from (3.3.45) and (3.3.46)

$$|\Phi(\xi)|^2 = [N_1 + N_2 + \psi_1(\xi)\,\cos\vartheta]^2 + [\psi_2(\xi)\,\sin\vartheta]^2, \tag{3.3.50}$$

where $\vartheta = \vartheta(\xi)$ is the argument of $z(\xi)$, and $\psi_1(\xi)$ and $\psi_2(\xi)$ are defined as

$$\psi_1(\xi) = N_1\,|z(\xi)| + \frac{N_2}{|z(\xi)|} \tag{3.3.51}$$

and

$$\psi_2(\xi) = N_1\,|z(\xi)| - \frac{N_2}{|z(\xi)|}. \tag{3.3.52}$$

Hence we have, for all ξ

$$|\Phi(\xi)|^2 \leq [N_1 + N_2 + \psi_1(\xi)]^2 + [\psi_2(\xi)]^2 = \Psi(\xi). \tag{3.3.53}$$

The left inequality in (3.3.49) entails $\psi_1(\xi) > 0$, as well as the right one entails $\psi_2(\xi) < 0$, for all $\xi \in \mathbf{R}$. Moreover, since

$$\psi_1'(\xi) = \left[\frac{N_1|z(\xi)|^2 - N_2}{|z(\xi)|^2}\right]\,\frac{d|z(\xi)|}{d\xi}$$

and

$$\psi_2'(\xi) = \left[\frac{N_1|z(\xi)|^2 + N_2}{|z(\xi)|^2}\right]\,\frac{d|z(\xi)|}{d\xi},$$

the same argument shows that $\psi_1(\xi)$ is increasing in $(-\infty, 0)$ and decreasing in $(0, +\infty)$, while $\psi_2(\xi)$ behaves in the opposite way.

The function $\Psi(\xi)$ is therefore increasing in $(-\infty, 0)$ and decreasing in $(0, +\infty)$, as

$$\Psi'(\xi) = 2\,[N_1 + N_2 + \psi_1(\xi)]\,\psi_1'(\xi) + 2\,[\psi_2(\xi)]\,\psi_2'(\xi),$$

where the two terms on the right hand side have the same sign. This entails

$$\max_{\xi} |\Phi(\xi)|^2 \le \Psi(0),$$

and we have to focus on the calculation of $\Psi(0)$, considering two different cases.

i) If $b_x \neq 0$, let us define $\eta := 4a/b_x^2$. We have

$$\Psi(0) = \left[N_1\left(1+\sqrt{\frac{1+\eta\nu_2}{1+\eta\nu_1}}\right) + N_2\left(1+\sqrt{\frac{1+\eta\nu_1}{1+\eta\nu_2}}\right)\right]^2 + \left[N_1\sqrt{\frac{1+\eta\nu_2}{1+\eta\nu_1}} - N_2\sqrt{\frac{1+\eta\nu_1}{1+\eta\nu_2}}\right]^2 \tag{3.3.54}$$

It can be easily verified that the right hand term is decreasing as a function of η: since η is positive, it attains its maximum when $\eta = 0$. This provides

$$\begin{aligned} \max_{\xi} |\Phi(\xi)|^2 &\le (2N_1 + 2N_2)^2 + (N_1 - N_2)^2 \\ &= 5N_1^2 + 6N_1N_2 + 5N_2^2 \\ &= 5\,\frac{\nu_1^4}{(\nu_1+\nu_2)^4} + 6\,\frac{\nu_1^2\nu_2^2}{(\nu_1+\nu_2)^4} + 5\,\frac{\nu_2^4}{(\nu_1+\nu_2)^4}. \end{aligned} \tag{3.3.55}$$

So far, gathering together (3.3.47) and (3.3.55), we can conclude

$$\begin{aligned} \frac{\max_{\xi} |\Phi(\xi)|^2}{(\min_{\xi} \operatorname{Re}\Phi(\xi))^2} &\le \frac{(\nu_1+\nu_2)^4}{\nu_2^4}\left[5\,\frac{\nu_1^4}{(\nu_1+\nu_2)^4} + 6\,\frac{\nu_1^2\nu_2^2}{(\nu_1+\nu_2)^4} + 5\,\frac{\nu_2^4}{(\nu_1+\nu_2)^4}\right] \\ &= 5 + 6\left(\frac{\nu_1}{\nu_2}\right)^2 + 5\left(\frac{\nu_1}{\nu_2}\right)^4 < 16, \end{aligned} \tag{3.3.56}$$

where the last inequality follows from the assumption $\nu_1 < \nu_2$.

ii) If $b_x = 0$, namely the flux term is parallel to the interface, $|z(0)| = \sqrt{\nu_2/\nu_1}$, and we have

$$\begin{aligned} \max_{\xi} |\Phi(\xi)|^2 &\le \left[N_1\left(1+\sqrt{\frac{\nu_2}{\nu_1}}\right) + N_2\left(1+\sqrt{\frac{\nu_1}{\nu_2}}\right)\right]^2 + \left[N_1\sqrt{\frac{\nu_2}{\nu_1}} - N_2\sqrt{\frac{\nu_1}{\nu_2}}\right]^2 \\ &= \frac{1}{(\nu_1+\nu_2)^4}\,[\nu_1^4 + \nu_2^4 + 2\,\nu_1^3\nu_2 + 2\,\nu_1\nu_2^3 + 2\,\nu_1^{7/2}\nu_2^{1/2} + 2\,\nu_1^2\nu_2^2 \\ &\quad + 2\,\nu_1^{3/2}\nu_2^{5/2} + 2\,\nu_1^{3/2}\nu_2^{5/2} + 2\,\nu_1^{1/2}\nu_2^{7/2}]. \end{aligned} \tag{3.3.57}$$

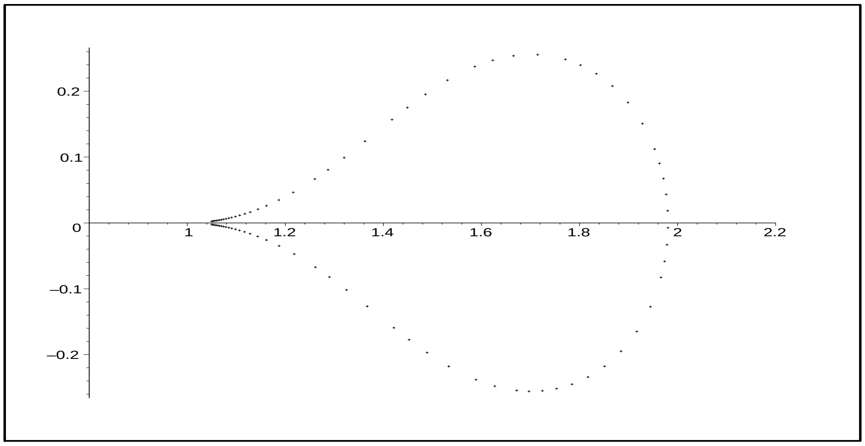

Figure 3.3: $\mathbf{b} = (1,1)$, $a = 0.1$, $\nu_1 = 10^{-7}, \nu_2 = 10^{-1}$

Gathering together (3.3.47) and (3.3.57), we get

$$\frac{\max_\xi |\Phi(\xi)|^2}{(\min_\xi \operatorname{Re}\Phi(\xi))^2} \leq 1 + 2\sum_{k=1}^{7}\left(\frac{\nu_1}{\nu_2}\right)^{k/2} + \left(\frac{\nu_1}{\nu_2}\right)^4 < 16 \tag{3.3.58}$$

where, once again, the last inequality follows from the assumption $\nu_1 < \nu_2$.
From (3.3.56) and (3.3.58), on one hand, estimates (3.3.41) and (3.3.42) are straightforward, while, on the other hand,

$$\rho\,(\mathcal{T} \circ \mathcal{S}) < \frac{15}{16},$$

independently of the coefficients of the problem, and this concludes the proof. □

Remark 3.3.3 The assumption $\nu_1 < \nu_2$ is not restrictive, since it can be easily seen that a symmetric argument would give the same result as long as $\nu_2 < \nu_1$. □

Remark 3.3.4 It appears from estimates (3.3.41) and (3.3.42) that the reduction factor of the GMRES algorithm for the preconditioned system improves with the growth of the ratio ν_2/ν_1: this allows the treatment of large discontinuities. Moreover, since the bound does not depend on the Fourier variable ξ, its independence from the mesh easily follows. □

Figures 3.3 through 3.8 show the distribution of the eigenvalues of the preconditioned operator $\mathcal{T} \circ \mathcal{S}$ for several different choices of the coefficients involved, considering cases in which $\nu_1 \ll \nu_2$ and a maximal frequency $\xi_{max} = 100$. It turns out that, as we could expect from (3.3.41) and (3.3.42), the preconditioner performs very well for large values of the ratio ν_2/ν_1. Moreover, for advection dominated problems $\mathcal{T}$ is almost exact.

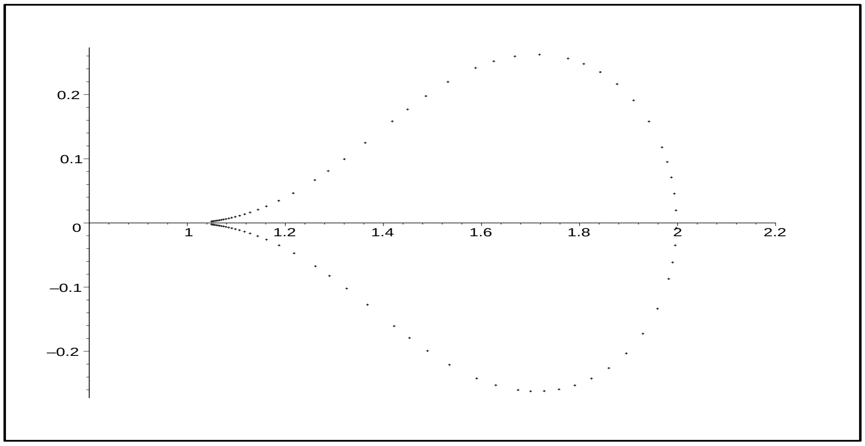

Figure 3.4: $\mathbf{b} = (1,1)$, $a = 0.1$, $\nu_1 = 10^{-6}, \nu_2 = 10^{-2}$

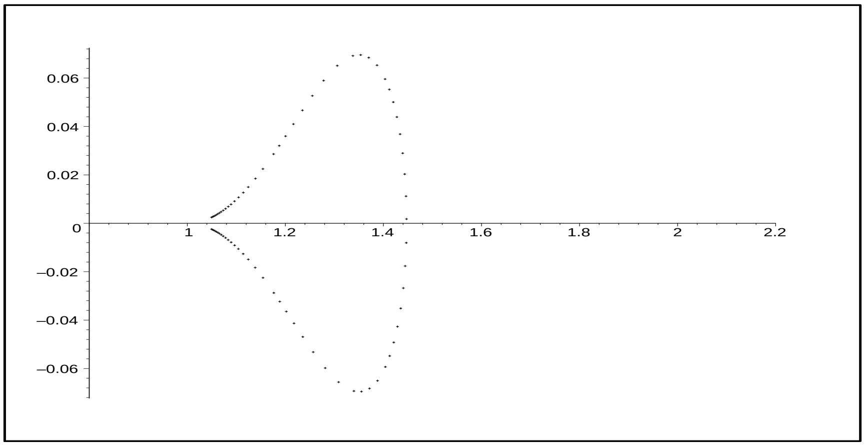

Figure 3.5: $\mathbf{b} = (1,1)$, $a = 10$, $\nu_1 = 10^{-7}, \nu_2 = 10^{-1}$

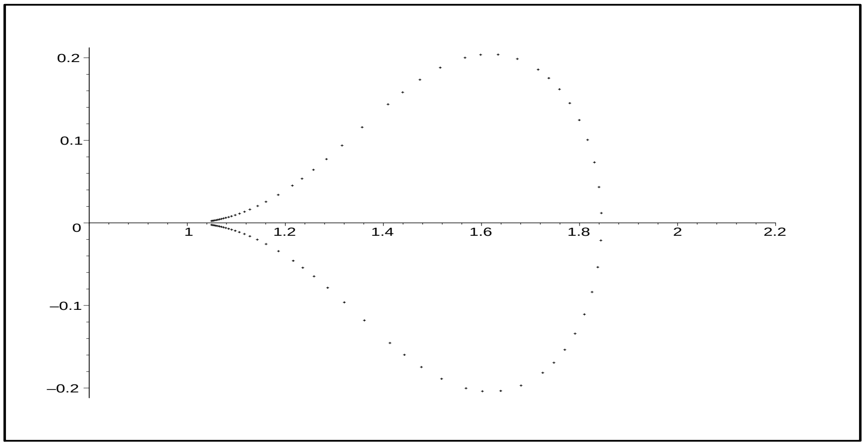

Figure 3.6: $\mathbf{b} = (1, 1)$, $a = 10$, $\nu_1 = 10^{-6}, \nu_2 = 10^{-2}$

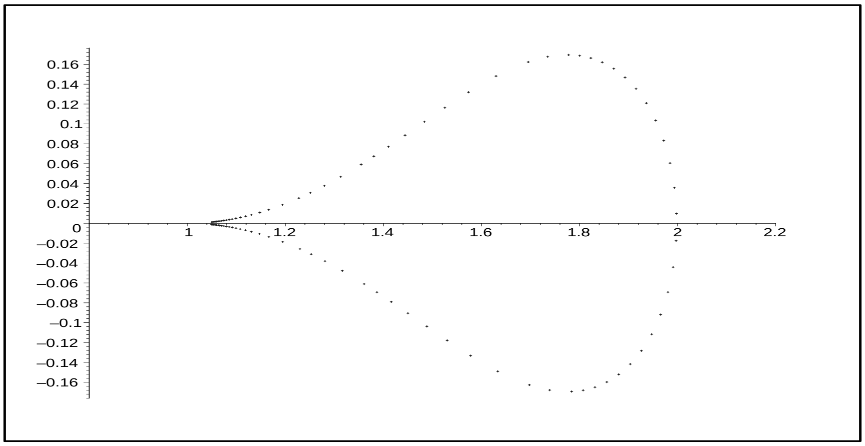

Figure 3.7: $\mathbf{b} = (10, 5)$, $a = 1$, $\nu_1 = 10^{-7}, \nu_2 = 10^{-1}$

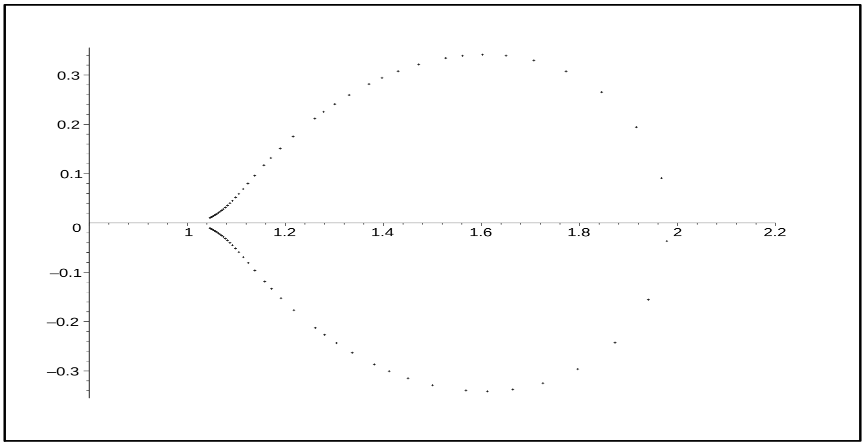

Figure 3.8: $\mathbf{b} = (10, 5)$, $a = 1$, $\nu_1 = 10^{-6}, \nu_2 = 10^{-2}$

3.3.4 The case of many subdomains

In this section we focus on the extension of the proposed preconditioner to the case of a decomposition into an arbitrary number of subdomains. In order to be able to treat also more general partitions, we firstly introduce the variational form of the algorithm, then we extend to this case the substructuring technique discussed in the previous sections.

Variational formulation of the continuous problem

Let us consider in $\mathbf{R}^d$ (with $d = 2, 3$) the domain partition

$$\Omega = \bigcup_{k=1}^{N} \Omega_k,$$

with $\Omega_j \cap \Omega_k = \emptyset$ for $j \neq k$, on which we are solving the general advection-diffusion problem

$$\begin{array}{rcll} -\mathrm{div}\,(\nu(x)\nabla u) + \mathbf{b}(x) \cdot \nabla(u) + a(x)u & = & f & \text{in } \Omega \\ u & = & 0 & \text{on } \partial\Omega_D \\ \nu(x)\dfrac{\partial u}{\partial n} & = & \varphi & \text{on } \partial\Omega_N \end{array} \tag{3.3.59}$$

with piecewise constant viscosity

$$\nu(x) := \sum_{k=1}^{N} \nu_k \, \mathbf{1}_{\Omega_k}(x)$$

where $\mathbf{1}_{\Omega_k}$ is the characteristic function of the domain Ω_k.
In order to restrict ourselves to well-posed problems, we assume that the velocity field $\mathbf{b} \in W^{1,\infty}(\Omega)$ is of bounded divergence, that

$$a - \frac{1}{2}\mathrm{div}(\mathbf{b}) \geq \mu > 0,$$

for some $\mu \in \mathbf{R}$, and that the Neumann boundary conditions are given only on a subset $\partial\Omega_N$ of the domain boundary where we have outflow conditions,

$$\mathbf{b} \cdot \mathbf{n} \geq 0 \qquad \forall x \in \partial\Omega_N,$$

where as usual $\mathbf{n}$ denotes the unit vector normal to $\partial\Omega$ pointing outwards. The variational formulation of (3.3.59) reads

$$\text{Find } u \in \mathbb{H}(\Omega): \qquad a(u,v) = L(v) \quad \forall v \in \mathbb{H}(\Omega), \tag{3.3.60}$$

where

$$\mathbb{H}(\Omega) = \left\{ v \in H^1(\Omega) : v_{|\partial\Omega_D} = 0 \right\},$$

and

$$\begin{aligned} a(u,v) &= \int_\Omega \nu \nabla u \cdot \nabla v + (\mathbf{b} \cdot \nabla u) v + auv, \\ L(v) &= \int_\Omega fv + \int_{\partial\Omega_N} \varphi v. \end{aligned}$$

In order to extend the substructuring technique discussed in the previous section to this general partitioning, we define the interfaces

$$\Gamma_k := \partial\Omega_k \setminus \partial\Omega, \quad \Gamma = \cup_k \Gamma_k,$$

and we have to describe the action of the advection-diffusion operator on each subdomain Ω_k. The simple restriction of the bilinear form $a(u,v)$ to Ω_k

$$\hat{a}_k(u,v) = \int_{\Omega_k} \nu_k \nabla u \cdot \nabla v + (\mathbf{b} \cdot \nabla u) v + auv$$

is not satisfactory because of its lack of positiveness. To overcome this problem, an integration by parts of the advective term $1/2(\mathbf{b}(x) \cdot \nabla u)v$ leads to the local symmetrized form

$$\begin{aligned} a_k(u,v) &:= \int_{\Omega_k} \left\{ \nu_k \nabla u \cdot \nabla v + \frac{1}{2}\left[(\mathbf{b} \cdot \nabla u)v - (\mathbf{b} \cdot \nabla v)u\right] + (a - \frac{1}{2}\mathrm{div}\,\mathbf{b})uv \right\} + \int_{\partial\Omega_N \cap \partial\Omega_k} \frac{1}{2}\mathbf{b} \cdot \mathbf{n}_k uv \\ &= \hat{a}_k(u,v) - \int_{\Gamma_k} \frac{1}{2}\mathbf{b} \cdot \mathbf{n}_k uv. \end{aligned}$$

Summing up on k, and letting

$$L_k(v) := \int_{\Omega_k} fv + \int_{\partial\Omega_N \cap \partial\Omega_k} \varphi v,$$

the variational problem (3.3.60) is equivalent to

$$\text{Find } u \in \mathbb{H}(\Omega) : \qquad \sum_{k=1}^{n} \{a_k(u,v) - L_k(v)\} = 0 \quad \forall v \in \mathbb{H}(\Omega), \tag{3.3.61}$$

since the interface terms $-\int_{\Gamma_k} 1/2\, \mathbf{b}\cdot\mathbf{n}_k uv$ added locally to each form $\hat{a}_k$ cancel each other by summation. However, since we have by construction

$$a_k(u,u) = \int_{\Omega_k} \left\{ \nu_k |\nabla u|^2 + (a - \frac{1}{2}\mathrm{div}\,\mathbf{b})u^2 \right\} + \int_{\partial\Omega_N \cap \partial\Omega_k} \frac{1}{2}\mathbf{b}\cdot\mathbf{n}_k u^2 \qquad \forall u \in \mathbb{H}(\Omega_k),$$

where we have denoted with $\mathbb{H}(\Omega_k) = \left\{ v_k = v_{|\Omega_k},\ v \in \mathbb{H}(\Omega) \right\}$ the space of restrictions, their presence is very important since it guarantees that the local bilinear form $a_k(u,v)$ is positive on $\mathbb{H}(\Omega_k)$.

Finite Element Approximation

In order to approximate numerically the variational problem (3.3.61) above with finite elements, we replace the space $\mathbb{H}(\Omega)$ with a suitable finite element space $\mathbb{H}_h(\Omega)$ (for a brief outline of finite element methods for advection-diffusion problems see the Appendix). In the numerical tests reported at the end of this Chapter, we use second order isoparametric finite elements defined on regular triangulations of Ω, as they are a good compromise between accuracy and cost-efficiency. Other choices are of course possible, but in any case the triangulations respect the geometry of subdomain decomposition: the interfaces Γ_k will coincide with interelement boundaries, which means that each subdomain can be obtained as the union of a given subset of elements in the original triangulation.

When problem (3.3.59) is advection-dominated, these finite elements techniques must be stabilized. In the following we will use *Galerkin Least-Squares* techniques (*GALS*), but different choices (such as the *Streamline Diffusion* introduced in Chapter 1) can be made. The *GALS* technique consists in adding to the original variational formulation the element residuals

$$\int_{T_i} \delta_i(h)\,(-\mathrm{div}\,(\nu(x)\nabla u) + \mathbf{b}(x)\cdot\nabla u + a(x)u - f)\,(-\mathrm{div}\,(\nu(x)\nabla v) + \mathbf{b}(x)\cdot\nabla v + a(x)v)$$

where T_i is an element of the triangulation, with a suitable choice of the local positive stabilization parameter $\delta_i(h)$. The stabilized finite elements formulation then reads

$$\text{Find } u_h \in \mathbb{H}_h(\Omega) : \qquad \sum_{k=1}^{n} \{a_{kh}(u_h,v_h) - L_{kh}(v_h)\} = 0 \quad \forall v_h \in \mathbb{H}_h(\Omega), \tag{3.3.62}$$

where

$$\begin{aligned} a_{kh}(u,v) &= a_k(u,v) + \sum_{T_i \subset \Omega_k} \int_{T_i} \delta_i \left(-\text{div}\,(\nu_k \nabla u) + \mathbf{b}\cdot\nabla u + au\right)\left(-\text{div}\,(\nu_k \nabla v) + \mathbf{b}\cdot\nabla v + av\right), \\ L_{kh}(v) &= L_k(v) + \sum_{T_i \subset \Omega_k} \int_{T_i} \delta_i(h) f \left(-\text{div}\,(\nu_k \nabla v) + \mathbf{b}\cdot\nabla v + av\right). \end{aligned}$$

Substructuring

The variational structure of problems (3.3.61) and (3.3.62) allows to reduce them to an interface problem by means of standard substructuring techniques. Notice that, since the variational structure of the original problem and of its finite elements discretization are very similar, and this will be true also for the numerical domain decomposition we introduce in this section, we will use the same notation for both the continuous and the discrete problem and omit all the subscripts h in all finite elements formulations: just remember that, when dealing with finite elements, the bilinear and linear forms $a_k(.,.)$ and $L_k(.)$ should be replaced by their discrete counterparts $a_{kh}(.,.)$ and $L_{kh}(.)$ as defined in the previous section.
Following [2], we consider the local space of restrictions $\mathbb{H}(\Omega_k)$ defined in the previous section and we introduce the space

$$\mathbb{H}^0(\Omega_k) = \left\{ v_k \in \mathbb{H}(\Omega), v_k = 0 \text{ in } \overline{\Omega \setminus \Omega_k} \right\}$$

consisting of functions of $\mathbb{H}(\Omega_k)$ with zero continuous extension in $\overline{\Omega \setminus \Omega_k}$, the global trace space $\mathbb{V} = \text{Tr}\mathbb{H}(\Omega)_{|\Gamma}$, the local trace spaces

$$\mathbb{V}_k = \left\{ \bar{v}_k = \text{Tr}\, v_{k|\Gamma_k},\ v_k \in \mathbb{H}(\Omega_k) \right\} = \left\{ \bar{v}_k = \text{Tr}\, v_{|\Gamma_k},\ v \in \mathbb{H}(\Omega) \right\},$$

the restriction operators

$$R_k : \mathbb{H}(\Omega) \to \mathbb{H}(\Omega_k), \quad \bar{R}_k : \mathbb{V} \to \mathbb{V}_k,$$

the a_k-harmonic extension $\text{Tr}_k^{-1} : \mathbb{V}_k \to \mathbb{H}(\Omega_k)$, defined as

$$a_k(\text{Tr}_k^{-1}\bar{u}_k, v_k) = 0 \quad \forall v_k \in \mathbb{H}^0(\Omega_k), \quad \text{Tr}(\text{Tr}_k^{-1}\bar{u}_k)_{|\Gamma_k} = \bar{u}_k, \quad \text{Tr}_k^{-1}\bar{u}_k \in \mathbb{H}(\Omega_k) \tag{3.3.63}$$

as well as its adjoint Tr_k^{-*}, defined by

$$a_k(v_k, \text{Tr}_k^{-*}\bar{u}_k) = 0 \quad \forall v_k \in \mathbb{H}^0(\Omega_k), \quad \text{Tr}(\text{Tr}_k^{-*}\bar{u}_k)_{|\Gamma_k} = \bar{u}_k, \quad \text{Tr}_k^{-*}\bar{u}_k \in \mathbb{H}(\Omega_k). \tag{3.3.64}$$

Since the bilinear form a_k is elliptic on $\mathbb{H}^0(\Omega_k)$ by construction, problems (3.3.63) and (3.3.64) are well-posed, and we can define the local Schur complement operator $S_k : \mathbb{V}_k \to \mathbb{V}'_k$ as

$$\langle S_k \bar{u}_k, \bar{v}_k \rangle = a_k(\text{Tr}_k^{-1}\bar{u}_k, \text{Tr}_k^{-*}\bar{v}_k) \qquad \forall \bar{u}_k, \bar{v}_k \in \mathbb{V}_k.$$

If we decompose the local degrees of freedom U_k of $u_k = R_k u$ into internal (U_k^0) and interface $(\bar{U}_k)$ degrees of freedom, the matrix A_k associated to the bilinear form a_k can be decomposed into

$$A_k = \left[\begin{array}{cc} A_k^0 & B_k \\ \tilde{B}_k^T & \bar{A}_k \end{array} \right],$$

and we have

$$Tr_k^{-1} = \left(\begin{array}{c} -(A_k^0)^{-1} B_k \\ \\ Id \end{array} \right),$$

as well as

$$S_k \bar{U}_k = \left(\bar{A}_k - \tilde{B}_k^T (A_k^0)^{-1} B_k \right) \bar{U}_k.$$

We can therefore decompose each restriction of the solution u and test function v into $R_k u = u_k^0 + Tr_k^{-1}(\bar{R}_k u)$ and $R_k v = v_k^0 + Tr_k^{-*}(\bar{R}_k v)$, and eliminate the local internal component u_k^0 since it is solution of the local well-posed problem

$$a_k(u_k^0, v_k) = L_k(v_k) \qquad \forall v_k \in \mathbb{H}^0(\Omega_k), \; u_k^0 \in \mathbb{H}^0(\Omega_k).$$

Thus, we can introduce the global Schur complement operator

$$S = \sum_{k=1}^{N} \bar{R}_k^T S_k \bar{R}_k$$

and we reduce problems (3.3.61) and (3.3.62) to the interface problem

$$S\bar{u} = F \qquad \text{in } \mathbb{V}, \tag{3.3.65}$$

where the right-hand side is defined as

$$\begin{aligned} \langle F, \bar{v} \rangle &= \sum_k L_k(Tr_k^{-*}(\bar{R}_k \bar{v})) \\ &= \sum_k \left[L_k(v_k) - L_k(v_k - Tr_k^{-*}(\bar{R}_k \bar{v})) \right] \\ &= \sum_k \left[L_k(v_k) - a_k(u_k^0, v_k - Tr_k^{-*}(\bar{R}_k \bar{v})) \right] \quad \text{(construction of } u_k^0\text{)} \\ &= \sum_k \left[L_k(v_k) - a_k(u_k^0, v_k) \right] \quad \text{(definition of } Tr_k^{-*}\text{)}, \end{aligned}$$

where v_k is any function in $\mathbb{H}(\Omega_k)$ such that $v_k = \bar{v}$ on Γ_k.

Definition of the preconditioner

The preconditioner we propose here for the solution of (3.3.65) is an extension of the ones proposed in [2] and [75] and a generalization of the one discussed in the previous section to an arbitrary number of subdomains. We precondition the interface operator $S = \sum_{k=1}^{N} \bar{R}_k^T S_k \bar{R}_k$ with a weighted sum of inverses:

$$T = \sum_{k=1}^{N} D_k^T (S_k)^{-1} D_k, \tag{3.3.66}$$

with

$$\sum_{k=1}^{N} D_k \bar{R}_k = Id_\Gamma. \tag{3.3.67}$$

Notice that, as well known in domain decomposition literature, for any $F_k \in \mathbb{V}_k'$ the action of the operator $(S_k)^{-1} F_k$ is simply equal to the trace on Γ_k of the solution w_k of the local variational problem

$$a_k(w_k, v_k) = \langle F_k, Tr_k v_k \rangle \quad \forall v_k \in \mathbb{H}(\Omega_k), w_k \in \mathbb{H}(\Omega_k),$$

which, by construction of the bilinear form a_k, is associated to the operator

$$-\text{div}(\nu_k \nabla w) + \mathbf{b} \cdot \nabla w + aw$$

with Robin boundary condition on the interface

$$\nu_k \frac{\partial w}{\partial n_k} - \frac{1}{2} \mathbf{b} \cdot \mathbf{n}_k w = F_k \qquad \text{on } \Gamma_k.$$

In order to achieve good parallelization properties for the preconditioned algorithm, as the bilinear form changes with the subdomains, the weights D_k should be chosen as local as possible. The following section is dedicated to their construction.

Construction of the weights D_k

Since we have to take into account what happens in the neighborhood of each interface point, the map D_k is defined on each degree of freedom of the interface Γ_k. For $P \in \Gamma_k$ we define the set

$$N_P := \{ j \in \{1, \ldots, N\} \mid P \in \Gamma_j \},$$

consisting of all indexes corresponding to the subdomains Ω_j whose interface boundary contains P. We define the weight D_k on the degree of freedom $\bar{u}(P)$ by

$$D_k \, \bar{u}(P) \; = \; C_P \frac{\nu_k}{\sum_{j \in N_P} \nu_j} \, \bar{u}(P),$$

where the constant C_P is chosen in a suitable way to satisfy (3.3.67), and it depends only on the number of subdomains to which the point P belongs. As an example, consider a domain

$\Omega \in \mathbf{R}^3$ decomposed into N parallelepipedal subdomains: if the point P is a vertex that belongs to 8 subdomains, the set N_P will consists of 8 indexes and we choose $C_P = 1/3$, if P lies on a side which separates 4 subdomains, N_P consists of 4 indexes and we choose $C_P = 1/2$, and finally if P belongs to a face and separates only two subdomains we choose $C_P = 1$.

Remark 3.3.5 The preconditioner introduced in this section transforms naturally into the Robin/Robin one as long as the viscosity is continuous (and into the generalized Neumann/Neumann as long as the operator is symmetric). Thus, it is well-suited for a subdivision of each region physically homogeneous into smaller subdomains. In this latter case, however, if the number of subdomains is large, the introduction of a coarse space may help to reduce the high number of initial iterations due to the presence of constant-like functions which scale badly in energy norm when expanded from a local subdomain to the full domain. The traditional remedy consists in the introduction of a small coarse global space which includes these constant like functions and to solve the interface problem by a direct solver on the coarse space and by a Robin/Robin preconditioned iterative solver on its orthogonal. □

3.4 Numerical results in three-dimensions

This section contains the results of some numerical tests carried out at the INRIA in Rocquencourt (France) with the group of M. Vidrascu. The advection-diffusion problem (3.3.1) is discretized by means of the stabilized Galerkin Least-Squares technique described in the previous section using second order elements on an hexaedral decomposition. The interface problem (3.3.65) is solved by a GMRES algorithm preconditioned by the operator $\mathcal{T}$. The algorithm stops when the ℓ^2 norm on the interface of the initial residual is reduced by a factor of 10^{-10}.

3.4.1 A two-domains model problem

The first experiment deals with a partition of the unit cube $[0,1] \times [0,1] \times [0,1]$ into two subdomains $\Omega_1 = [0,0.5] \times [0,1] \times [0,1]$ and $\Omega_2 = [0.5,1] \times [0,1] \times [0,1]$. We choose different convective fields

i) $\vec{b} = \vec{e}_1$: the velocity is perpendicular to the interface,

ii) $\vec{b} = \vec{e}_2 + \vec{e}_3$: the velocity is parallel to the interface,

iii) $\vec{b} = \vec{e}_1 + 3\,\vec{e}_2 + 5\,\vec{e}_3$: we refer to this velocity as "*oblique*",

as well as $a = 1$. We consider large jumps between the viscosity coefficients, we choose $f \equiv 0$ in the whole Ω and we impose $u = 1$ on the bottom face of the cube as well as homogeneous Dirichlet conditions on the rest of the boundary $\partial\Omega$.

The total number of finite elements is 1728, the total number of degrees of freedom is 14023 and the number of degrees of freedom on the interface is 625. The number of iterations is reported in Table 3.6: when two results are present, the first one refers to a convective field directed from the more viscous region to the less viscous one, while the second refers to the opposite case. The results show that the preconditioner is almost insensitive to the choice of the convective fields, although it performs slightly better when the flux is directed towards the less viscous region. Nevertheless, a strong improvement in the number of iterations is observed when one

ν_1, ν_2	ν_1/ν_2	$\vec{b} = (\pm 1, 0, 0)$		$\vec{b} = (0, 1, 1)$	$\vec{b} = (\pm 1, 3, 5)$	
$10^{-1}, 10^{-5}$	10^4	10	11	17	15	17
$10^{-2}, 10^{-6}$		12	16	13	7	8
$10^{-1}, 10^{-6}$	10^5	10	11	17	15	17
$10^{-6}, 10^{-11}$		5	5	2	7	7
$10^{-1}, 10^{-7}$	10^6	10	11	17	15	17
$10^3, 10^{-3}$		3	3	3	3	3
$1, 10^{-7}$	10^7	6	7	9	11	11

Table 3.6: Number of iterations for the two-domain 3D model problem: res $< 10^{-10}$

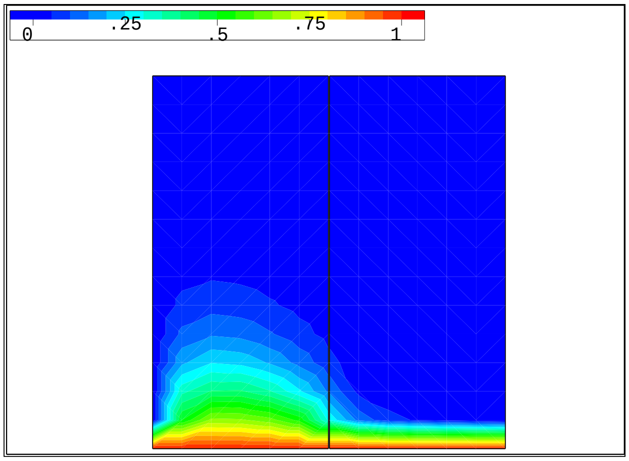

Figure 3.9: $\vec{b} = (-1, 0, 0)$, $\nu_1 = 10^{-1}, \nu_2 = 10^{-6}$. Section: $y = 0.5$.

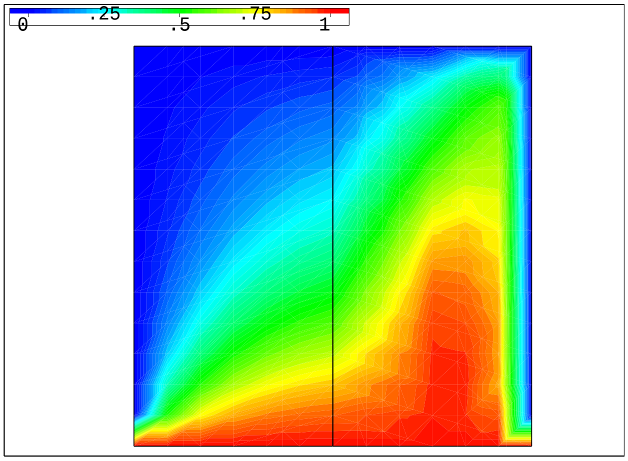

Figure 3.10: $\vec{b} = (1, 3, 5)$, $\nu_1 = 1, \nu_2 = 10^{-7}$. Section: $3x - y = 0.5$.

ν_1, ν_2	ν_1/ν_2	Test 1	Test 2	Test 3
$10^{-1}, 10^{-5}$	10^4	33	33	34
$10^{-1}, 10^{-6}$	10^5	32	33	34
$10^{-1}, 10^{-7}$	10^6	32	33	34
$10^3, 10^{-3}$	10^6	29	28	21
$1, 10^{-7}$	10^7	29	31	29

Table 3.7: Number of iterations for the multidomain model problem: res $< 10^{-10}$

of the two subproblems is not advection-dominated as well as when both subdomains have very little viscosity. However, the number of iterations is reasonable in all cases and it appears to be, as we expected from the theory, fairly insensitive to the viscosity jumps. Finally, we have represented in Figure 3.9 and 3.10 two cross-sections (which take into account the direction of the convective field) of the results for $\vec{b} = (-1, 0, 0)$, $\nu_1 = 10^{-1}, \nu_2 = 10^{-6}$ and for $\vec{b} = (1, 3, 5)$, $\nu_1 = 1, \nu_2 = 10^{-7}$ respectively; in both cases, Ω_1 is on the left side of the figure.

3.4.2 Influence of the number of subdomains

We investigate here the robustness of the preconditioner with respect to the number of subdomains and to their mutual position. We consider the cube $\Omega = [-0.5, 0.5] \times [-0.5, 0.5] \times [0, 1]$ partitioned into 8 subdomains, numbered in a clockwise helicoidal way from $\Omega_1 = [-0.5, 0] \times [-0.5, 0] \times [0, 0.5]$ to $\Omega_8 = [0, 0.5] \times [-0.5, 0] \times [0.5, 1]$. We consider the velocity field

$$\vec{b} = -2\pi y\, \vec{e}_1 + 2\pi x\, \vec{e}_2 + \sin(2\pi x)\, \vec{e}_3.$$

and we consider the cube as constituted of different materials disposed in the following ways:

Test 1: $\nu_1 = \nu_4 = \nu_5 = \nu_8$, and $\nu_2 = \nu_3 = \nu_6 = \nu_7$: this is the configuration considered in the previous section, but each physical domain here is decomposed into four smaller subdomains.

Test 2: $\nu_1 = \nu_5 = \nu_6 = \nu_8$, and $\nu_2 = \nu_3 = \nu_4 = \nu_7$: the homogeneous subdomains Ω_1 and Ω_2 are shown in Figure 3.11.

Test 3: $\nu_1 = \nu_3 = \nu_6 = \nu_8$, and $\nu_2 = \nu_4 = \nu_5 = \nu_7$: this case is a black and white decomposition where each subdomain of one kind is surrounded by subdomains of the other one. Figure 3.12 shows Ω_2.

Test 4: We choose $\nu_1 = 10^{-1}$, $\nu_3 = 10^{-2}$, $\nu_6 = 10^{-3}$, $\nu_8 = 10^{-4}$ and $\nu_2 = \nu_4 = \nu_5 = \nu_7 = 10^{-6}$.

We choose again $f \equiv 0$ in the whole Ω, and Dirichlet conditions $u = 1$ on the bottom face and $u = 0$ on the rest of $\partial\Omega$. The total number of finite elements is still 1728, the total number of degrees of freedom 14023, but the number of interface degrees of freedom has raised to 1801. We report in Table 3.7 the number of iterations, and we observe that the preconditioner is sensitive to the number of subdomains (33 against 20), but it appears once again insensitive to the jumps in the viscosity coefficients. Even more interesting, the preconditioned system is not affected by the larger number of different viscosity coefficients (see the results of Test 4 in Table 3.8). Finally, we have reported in Figure 3.13 a section of a result for Test 3.

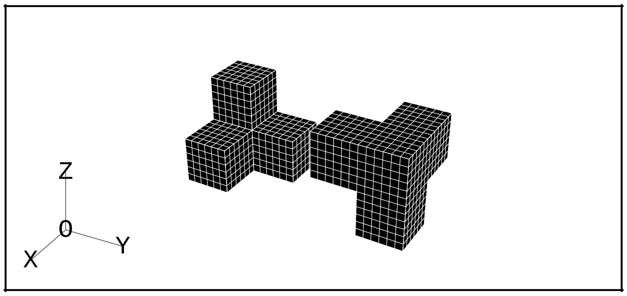

Figure 3.11: The subdomains Ω_1 (left) and Ω_2 (right) in Test 2.

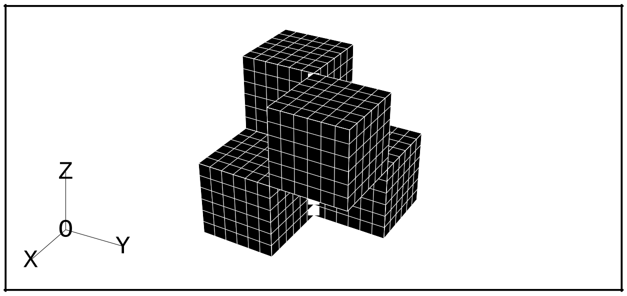

Figure 3.12: The domain Ω_2 in Test 3.

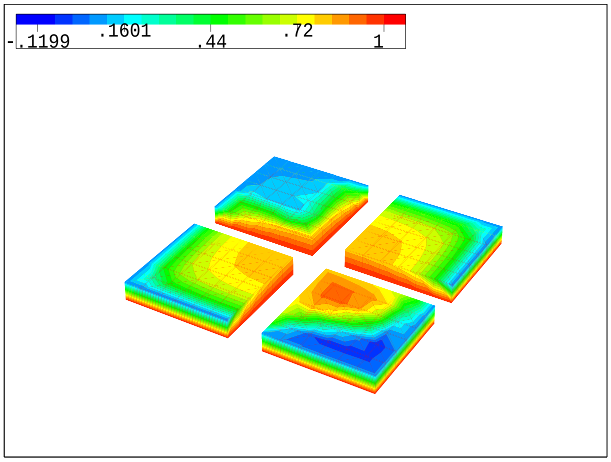

Figure 3.13: Test 3, $\nu_1 = 10^{-1}, \nu_2 = 10^{-7}$. Section: $z = 0.25$.

ν_1	$\nu_2, \nu_4, \nu_5, \nu_7$	ν_3	ν_6	ν_8	ITER
10^{-1}	10^{-6}	10^{-2}	10^{-3}	10^{-4}	34

Table 3.8: Number of iterations in Test 4. Residual $< 10^{-10}$

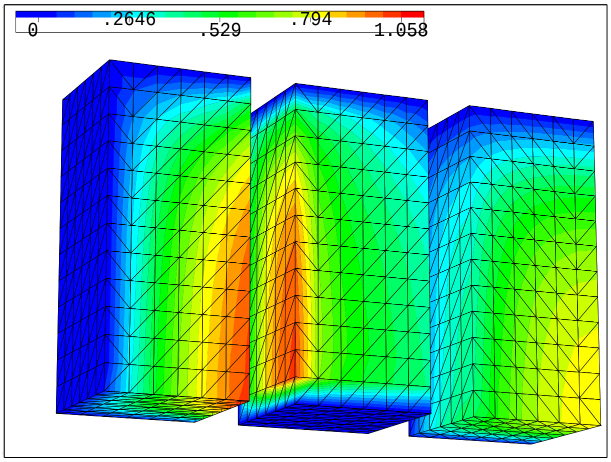

Figure 3.14: The three layers model. Section: $y = 0.5$

3.4.3 A three layers model problem

The third experiment deals with a parallelpipedal domain $\Omega = [0, 1.5] \times [0, 1] \times [0, 1]$ which is partitioned into three layers $\Omega_1 = [0, 0.5] \times [0, 1] \times [0, 1]$, $\Omega_2 = [0.5, 1] \times [0, 1] \times [0, 1]$, $\Omega_3 = [1, 1.5] \times [0, 1] \times [0, 1]$. It is a very simplified model of transport and diffusion of a species through different layers of materials, and is inspired by the project "Couplex" of the French national agency ANDRA. We choose $\nu_1 = \nu_3 = 0.1$, $\nu_2 = 10^{-4}$, a discontinuous convective field given by

$$\begin{cases} \vec{b} \;= 3\,\vec{e}_2 - 2\,\vec{e}_3 & \text{in } \Omega_1\text{and } \Omega_3 \\ \vec{b} \;= -\,\vec{e}_1 & \text{in } \Omega_2, \end{cases}$$

and a discontinuous reaction term given by

$$\begin{cases} a = .001 & \text{in } \Omega_1 \text{ and } \Omega_3 \\ a = .1 & \text{in } \Omega_2. \end{cases}$$

We choose $f \equiv 1$ and we impose the following boundary conditions:

$$\begin{cases} \dfrac{\partial u}{\partial n} = 0 & \text{on } [0, 0.5] \times [0, 1] \times \{0\},\ [1, 1.5] \times [0, 1] \times \{0\} \text{ and } \{0\} \times [0, 1] \times [0, 1] \\ u = 0 & \text{elsewhere} \end{cases}$$

We report in Table 3.9 the total number of finite elements (NE), the total number of degrees of freedom (NDF), the number of interface degrees of freedom (NIDF) and the number of iterations (ITER). Once again the result is quite satisfactory: discontinuity in all coefficients appears not to affect the performance of the preconditioner. We finally report in Figure 3.14 the cross-section $y = 0.5$ of the solution.

Partition	NE	NDF	NIDF	ITER
$3 \times 1 \times 1$	2808	14275	838	13

Table 3.9: A three layers model problem

3.5 Conclusions

We proposed a preconditioner which is a generalization of the Robin/Robin preconditioner to advection-diffusion problems with discontinuous coefficients. We have shown its robustness, assessed theorically by a Fourier analysis in the special case of the two half-planes, where the preconditioner provides a reduction factor for the associate GMRES which is bounded from above independently of the coefficients of the problem. Moreover, a more accurate estimate entails that the reduction factor improves with the growth of the jump in the viscosity, and this allows to handle very large viscosity jumps. The robustness of the preconditioner has then been confirmed by some numerical tests in three dimensions, which are reported in Section 3.4, where the preconditioner has shown fair insensitivity to the jumps in the viscosity coefficients as well as to the convective field. Unfortunately, the preconditioner remains sensitive to the number of subdomains, but this seems unavoidable in the case of advection-dominated problems without coarse grid correction. Consequently, if, on one hand, the extension of the preconditioner to a partition into many subdomains is quite straightforward stemming from the variational formulation of the problem, on the other hand, when the number of subdomains is large, the introduction of a coarse grid correction space could become mandatory, in order to avoid (or at least reduce) the initial stagnation of the algorithm.

The preconditioner transforms naturally into the Robin/Robin one when the viscosity is continuous, and, probably being its most interesting feature, it has the same algebraic structure as this latter one. Therefore it can be easily implemented into a software which contains the Robin/Robin or the Neumann/Neumann preconditioner.

However, if the extension to systems of advection-diffusion equations appears to be quite straightforward, further work needs to be done: a convergence analysis in a more general setting is not yet available (and it appears to be quite difficult), the extension to the case of discontinuous convective fields should be addressed, the introduction of a coarse space to reduce the sensitivity to the number of subdomains should be analyzed, and, finally, the algorithm should be tested on more complex situations.

Chapter 4

Schwarz Algorithms for Wave Equations

This chapter is devoted to a domain decomposition approach to the solution of equations describing the propagation of waves. Wave equations are used to model phenomena such as the sound emitted by a loud speaker, or the electromagnetic field generated by an antenna, or the diffraction of such field by a building.

In the last years, a growing interest in the field led many scientists to work on wave equations, especially in the numerical approximation framework, and this involved also people working on domain decomposition methods. Several contributions appeared in this directions, on both the Helmholtz equation, following the early works by B. Després ([35], [36], [16]), and the Maxwell system, among which we recall the works by A. Alonso and A. Valli ([9], [12], [11] [7], [10]), F. Ben-Belgacem ([15]), and Y. Maday, F. Rapetti and B. Wohlmuth ([80]). The most recent developments on this subject seek for an optimization of the interface conditions in order to improve the numerical efficiency of the algorithms proposed. Among them, we recall the thesis by P. Chevalier ([30]), and the work by M. Gander and F. Nataf ([51]) on Helmholtz equations.

In the first part of the chapter we briefly introduce the equations of acoustics and electromagnetics. Then, in Section 4.2 we analyze the convergence properties of some Schwarz algorithm, previously appeared in literature, for Helmholtz equations. Finally, in Section 4.3, we analyze the convergence properties of a slightly more general version of the Schwarz algorithm proposed by B. Després for the Maxwell system.

4.1 Acoustic and Electromagnetic Waves

The most common wave equations that can be encountered in physics or in engineering are the acoustic wave equation, also known as Helmholtz equation, describing the propagation of a sound through a medium, and the Maxwell system, which describes the propagation of the electromagnetic fields. We won't present here a detailed mathematical analysis of such equations, but in the rest of this section, following mainly the book by J.-C. Nédélec [83], we describe their main features.

4.1.1 The Helmholtz Equation

The propagation of the sound in a medium is described by the acoustic wave equation, which can be derived from the inviscid Navier-Stokes equations for compressible gases. In the case of small displacements of the gas, in fact, a linearization leads to an equation for the displacements and the small pressure variation in the gas. If the medium is homogeneous with mean density ρ_0, this results in the system of equations

$$\begin{cases} \rho_0 \dfrac{\partial \mathbf{u}}{\partial t} + \nabla p = 0, \\[2ex] \dfrac{1}{c^2}\dfrac{\partial p}{\partial t} + \rho_0 \mathrm{div}\, \mathbf{u} = 0, \end{cases} \tag{4.1.1}$$

where c is the speed of sound in the medium, $\mathbf{u}$ is the velocity of displacement in the medium and p is the pressure, which is supposed to be isotropic. After eliminating the displacement $\mathbf{u}$, we are left with a scalar wave equation for the pressure p:

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} - \Delta p = 0. \tag{4.1.2}$$

To be able to solve equation (4.1.2), we have to specify the domain in which we look for a solution, and we must prescribe initial data as well as suitable boundary conditions to have a well-posed problem.

In that order, let Ω be a region in $\mathbf{R}^d$, $(d = 2, 3)$, with bounded regular boundary $\partial\Omega$ (a curve or a surface). The domain in which the above equation has to be solved is either an interior domain, denoted by Ω_i, which coincides with Ω, or its complement $\Omega_e = \mathbf{R}^d \backslash \Omega_i$. The unit normal $\mathbf{n}$ is defined as pointing outwards with respect to Ω_i.

A complete set of data is obtained by prescribing initial values of p and its first time derivative $\partial p/\partial t$, while the classical boundary conditions are the Dirichlet and Neumann ones. The problem to be solved, up to a multiplicative constant depending on the choice of physical units, is therefore

$$\begin{cases} \dfrac{\partial^2 p}{\partial t^2} - \Delta p = 0 & x \in \Omega,\ t > 0 \\[2ex] p(0,x) = p_0(x) & x \in \Omega \\[2ex] \dfrac{\partial p}{\partial t}(0,x) = p_1(x) & x \in \Omega \\[2ex] \mathcal{B}(p)(t,x) = 0 & x \in \partial\Omega,\ t > 0 \end{cases} \tag{4.1.3}$$

where the trace operator $\mathcal{B}(p)$ is either a Dirichlet or a Neumann one.

Harmonic solutions and plane waves

It is usual to look for time harmonic solutions of the equation (4.1.3): it amounts to seek for solutions in the form

$$p(t,x) = \mathrm{Re}\left[u(x)\, e^{i\omega t}\right],$$

where i is the imaginary unit and ω is called the pulsation of the wave. The function $u(x)$ is complex valued and equation (4.1.3) reduces to the well-known Helmholtz equation

$$-\Delta u - k^2 u = 0 \qquad \text{in } \Omega, \tag{4.1.4}$$

with Dirichlet or Neumann boundary conditions. The quantity $k = \omega/c$ is called the wave number, whereas the quantity $f = \omega/2\pi$ is called the frequency.

The Helmholtz equation has some very special family of solutions. The first one are the so-called "*plane waves*": up to a multiplicative factor, these solutions are complex-valued functions of the form

$$u(x) = e^{i(\vec{k}\cdot\vec{x})}, \quad |\vec{k}| = k.$$

If the vector $\vec{k}$ is real, these solutions are of modulus 1, while if $\vec{k}$ is complex and such that $(\vec{k}\cdot\vec{k}) = k^2$, these solutions are exponentially decreasing in an half-space determined by the imaginary part of the vector $\vec{k}$ and exponentially increasing in the other half-space.

The second special family of solutions consists of the "*spherical waves*", which are functions that depend only on the radial variable r (the distance from the origin): an example in $\mathbf{R}^3$ is given by the function

$$u(x) = \frac{\sin kr}{r}, \qquad r = \sqrt{x_1^2 + x_2^2 + x_3^2},$$

which can be easily seen to satisfy equation (4.1.4). Notice that when r is large, r times a spherical wave is asymptotically a plane wave: this latter wave can therefore be used as a model describing a remote punctual acoustic source.

When we consider equation (4.1.4) in the framework of an interior problem we speak of stationary wave solutions. In that case, it is well-known that the operator $-\Delta$, with its boundary conditions on $\partial\Omega$, is self-adjoint and has a compact resolvent in $L^2(\Omega_i)$, admitting a spectral decomposition with positive eigenvalues of finite order. As a consequence, the Fredholm alternative argument tells us that either k^2 is not an eigenvalue and there is a unique solution of equation (4.1.4), or k^2 is an eigenvalue and the corresponding eigenfunctions are solutions of equation (4.1.4) with zero right-hand side.

When we deal with an exterior problem, we speak of progressive wave solutions, a situation extremely different from the previous one. The operator $-\Delta$, with its boundary conditions, neither is self-adjoint, nor it has a compact resolvent in $L^2(\Omega_e)$. Since plane waves are solutions to the homogeneous system in the whole space, it is thus natural to impose additional boundary conditions at infinity to equation (4.1.4) in order to guarantee uniqueness. We can eliminate the plane waves by simply look for solutions u that decrease at infinity as $1/r$. However, this is not enough for uniqueness (for instance the spherical wave $(\sin kr)/r$ is a non-zero solution of (4.1.4) in the free space), and we add therefore at infinity the extra "*Sommerfeld radiation condition*", (also called "*outgoing wave condition*")

$$\left|\frac{\partial u}{\partial r}+iku\right| \leq \frac{c}{r^2} \qquad \text{at infinity} \tag{4.1.5}$$

The formulation of the exterior problem is therefore

$$\left\{\begin{array}{ll} -\Delta u - k^2 u = 0 & \text{in } \Omega, \\ \\ u = 0 \quad (\text{or } \dfrac{\partial u}{\partial n} = 0) & \text{on } \partial\Omega \end{array}\right. \tag{4.1.6}$$

together with the radiation condition (4.1.5), which can be written in a weaker form as

$$\int_{\Omega_e} \left|\frac{\partial u}{\partial r}+iku\right|^2 dx \leq c.$$

We denote by H the Hilbert space

$$H = \left\{u,\ \frac{u}{(1+r^2)^{1/2}},\ \frac{\nabla u}{(1+r^2)^{1/2}}, \frac{\partial u}{\partial r}+iku \in L^2(\Omega_e)\right\}$$

endowed with the norm

$$\|u\|_H^2 := \|u\|^2_{L^2(\Omega_e)} + \left\|\frac{u}{(1+r^2)^{1/2}}\right\|^2_{L^2(\Omega_e)} + \left\|\frac{\nabla u}{(1+r^2)^{1/2}}\right\|^2_{L^2(\Omega_e)} + \left\|\frac{\partial u}{\partial r}+iku\right\|^2_{L^2(\Omega_e)}$$

and we state the following result (for proof see [83]).

Theorem 4.1.1 *The exterior Dirichlet and Neumann problem* (4.1.6)-(4.1.5) *admit at most one solution in the Hilbert space H.* □

4.1.2 The Maxwell Equations

Maxwell equations describe the propagation of electromagnetic waves, which are defined by the electric field, denoted with E, and the magnetic field, denoted with H. We start by describing their laws in a dielectric isotropic medium, which is characterized by the *electric permittivity* ε and by the *magnetic permeability* μ.
The speed of waves in the dielectric medium is given by $1/\sqrt{\varepsilon\mu}$. We denote by ε_0 and μ_0 respectively the permittivity and the permeability of the vacuum, and by c the speed of light in the vacuum, which is

$$c = \frac{1}{\sqrt{\varepsilon_0\mu_0}}.$$

The relative permittivity and the permeability of the medium are defined as

$$\left\{\begin{array}{ll} \varepsilon = \varepsilon_r\varepsilon_0, & \varepsilon_r \geq 1, \\ \\ \mu = \mu_r\mu_0. & \mu_r \geq 1. \end{array}\right.$$

In the absence of electric and magnetic charges and currents, the electric and magnetic fields are governed by the system of equations

$$\begin{cases} -\varepsilon \dfrac{\partial E}{\partial t} + \mathrm{rot} H = 0 \\ \\ \mu \dfrac{\partial H}{\partial t} + \mathrm{rot} E = 0 \end{cases} \tag{4.1.7}$$

The system above has to be completed by transmission conditions at the interfaces separating different dielectric media. Along a surface Γ of discontinuity of ε or μ the tangential components of the fields E and H are continuous and, denoting with $\mathbf{n}$ the unit normal to Γ, these jump conditions take the form:

$$\begin{cases} [E \times \mathbf{n}]_\Gamma = 0, \\ \\ [H \times \mathbf{n}]_\Gamma = 0. \end{cases} \tag{4.1.8}$$

Isotropic conducting media are characterized, aside of ε and μ, by the *conductivity* σ, which is a real positive number. In such medium Maxwell system reads

$$\begin{cases} -\varepsilon \dfrac{\partial E}{\partial t} + \mathrm{rot} H - \sigma E = 0 \\ \\ \mu \dfrac{\partial H}{\partial t} + \mathrm{rot} E = 0, \end{cases}$$

while the interface conditions (4.1.8) holding for dielectric media remain unchanged.

Time-harmonic solutions

Also in the case of Maxwell system, it is quite usual to consider harmonic solutions which are the complex-valued fields $\mathbf{E}$ and $\mathbf{H}$ such that the fields

$$\begin{cases} E(t,x) = \mathrm{Re}\left(\mathbf{E}(x) e^{i\omega t}\right), \\ \\ H(t,x) = \mathrm{Re}\left(\mathbf{H}(x) e^{i\omega t}\right), \end{cases}$$

where i is the imaginary unit and ω is the pulsation, satisfy the Maxwell system. Thus, the fields $\mathbf{E}$ and $\mathbf{H}$ must satisfy the harmonic Maxwell equations:

$$\begin{cases} -i\omega\varepsilon \mathbf{E} + \mathrm{rot}\mathbf{H} - \sigma \mathbf{E} = \mathbf{0} \\ \\ i\omega\mu \mathbf{H} + \mathrm{rot}\mathbf{E} = \mathbf{0}. \end{cases} \tag{4.1.9}$$

As in the case of acoustic waves, to be able to solve system (4.1.9) one must specify the domain in which is looking for a solution, as well as suitable boundary conditions. Thus, let Ω be a region in $\mathbf{R}^d$ ($d = 2, 3$), with bounded regular boundary $\partial\Omega$ (a curve or a surface). Once again, we denote with Ω_i the interior domain, which coincides with Ω, and with Ω_e the exterior domain,

which coincides with its complement $\Omega_e = \mathbf{R}^d \backslash \Omega_i$ (the unit normal $\mathbf{n}$ being defined as pointing outwards with respect to Ω_i).
Waves usually propagate in unbounded domains, and, similarly to the case of the Helmholtz equation, one must be able to describe precisely the behavior of the solutions at infinity, which is equivalent to define the radiation conditions. They typically have the form

$$\begin{cases} |\mathbf{E}(x)| \leq \dfrac{C}{r}, & \text{for large } r, \\[2ex] |\mathbf{H}(x)| \leq \dfrac{C}{r}, & \text{for large } r, \\[2ex] \left|\sqrt{\varepsilon}\mathbf{E} - \sqrt{\mu}\mathbf{H} \times \dfrac{\vec{r}}{r}\right| \leq \dfrac{C}{r^2}, & \text{for large } r, \end{cases} \tag{4.1.10}$$

where C is a constant.
Thus, a possible formulation of the *time harmonic exterior Maxwell problem* is the following:

Find $\mathbf{E}$ and $\mathbf{H}$, such that

$$\begin{cases} -i\omega\varepsilon\mathbf{E} + \mathrm{rot}\mathbf{H} - \sigma\mathbf{E} = \mathbf{0} & \text{in } \Omega_e \\[1ex] i\omega\mu\mathbf{H} + \mathrm{rot}\mathbf{E} = \mathbf{0} & \text{in } \Omega_e \\[1ex] \mathbf{E} \times \mathbf{n} = \boldsymbol{\Upsilon} & \text{on } \partial\Omega \\[1ex] \mathbf{E} \text{ and } \mathbf{H} \text{ satisfy (4.1.10)} \end{cases}$$

where the tangential vector field Υ is a given datum defined on $\partial\Omega$. In the case of an object lit by an incident plane wave, with electric field $\mathbf{E}^{\mathrm{inc}}$, the boundary condition takes the form $\boldsymbol{\Upsilon} = \mathbf{E}^{\mathrm{inc}} \times \mathbf{n}$.
Similarly, a possible formulation of the *time harmonic interior Maxwell problem* is the following:

Find $\mathbf{E}$ and $\mathbf{H}$, such that

$$\begin{cases} -i\omega\varepsilon\mathbf{E} + \mathrm{rot}\mathbf{H} - \sigma\mathbf{E} = \mathbf{0} & \text{in } \Omega_i \\[1ex] i\omega\mu\mathbf{H} + \mathrm{rot}\mathbf{E} = \mathbf{0} & \text{in } \Omega_i \\[1ex] \mathbf{E} \times \mathbf{n} = \boldsymbol{\Upsilon} & \text{on } \partial\Omega \end{cases}$$

$\boldsymbol{\Upsilon}$ again being given, tangent to $\partial\Omega$. Another boundary condition quite common deals with the so-called impedance conditions, also called *Leontovich conditions*:

$$[(\mathbf{E} \times \mathbf{n}) + \beta\mathbf{n} \times (\mathbf{H} \times \mathbf{n})]_{|\partial\Omega} = g,$$

where β and g are given.
In any case, we can rewrite system (4.1.9) in terms of the unknown $\mathbf{E}$: for the interior problem we have

$$\begin{cases} \operatorname{rot}\ \left(\mu^{-1}\operatorname{rot}\mathbf{E}\right) - \omega^2\varepsilon\mathbf{E} + i\omega\sigma\mathbf{E} = \mathbf{0} & \text{in } \Omega \\ \mathbf{E}\times\mathbf{n} = \boldsymbol{\Upsilon} & \text{on } \partial\Omega \end{cases} \tag{4.1.11}$$

with the additional requirement that $\operatorname{div}\mathbf{E} = 0$ in Ω.
Setting

$$\mathbf{u} := \mathbf{E} - \mathbf{E}_\Upsilon, \tag{4.1.12}$$

where $\mathbf{E}_\Upsilon$ is defined in Ω and satisfies $\mathbf{E}_\Upsilon\times\mathbf{n} = \Upsilon$ on $\partial\Omega$, we can finally rewrite system (4.1.11) as

$$\begin{cases} \operatorname{rot}\ \left(\mu^{-1}\operatorname{rot}\mathbf{u}\right) - \omega^2\varepsilon\mathbf{u} + i\omega\sigma\mathbf{u} = \mathbf{F} & \text{in } \Omega \\ \mathbf{u}\times\mathbf{n} = \mathbf{0} & \text{on } \partial\Omega \end{cases} \tag{4.1.13}$$

where we have set $\mathbf{F} := -\operatorname{rot}(\mu^{-1}\operatorname{rot}\mathbf{E}_\Upsilon) + \omega^2\varepsilon\mathbf{E}_\Upsilon - i\omega\sigma\mathbf{E}_\Upsilon$.

Weak formulation of the problem

We introduce the Hilbert spaces

$$\begin{aligned} H(\operatorname{rot};\Omega) &:= \left\{\mathbf{v}\in\left[L^2(\Omega)\right]^3 \mid \operatorname{rot}\mathbf{v}\in\left[L^2(\Omega)\right]^3\right\} \\ H_0(\operatorname{rot};\Omega) &:= \left\{\mathbf{v}\in H(\operatorname{rot};\Omega) \mid (\mathbf{n}\times\mathbf{v})_{|\partial\Omega} = \mathbf{0}\right\}, \end{aligned}$$

endowed with the graph norm

$$||\mathbf{v}||_{H(\operatorname{rot};\Omega)} := (||\mathbf{v}||^2_{0,\Omega} + ||\operatorname{rot}\mathbf{v}||^2_{0,\Omega})^{1/2}.$$

as well as the bilinear form

$$m(\mathbf{w},\mathbf{v}) := \int_\Omega \left(\mu^{-1}\operatorname{rot}\mathbf{w}\cdot\operatorname{rot}\overline{\mathbf{v}} - \omega^2\varepsilon\mathbf{w}\cdot\overline{\mathbf{v}} + i\omega\sigma\mathbf{w}\cdot\overline{\mathbf{v}}\right),$$

defined on $H(\operatorname{rot};\Omega)$, where $\overline{\mathbf{v}}$ denotes the conjugated of a complex number $\mathbf{v}$. The weak formulation of (4.1.13) reads

$$\text{find}\quad \mathbf{u}\in H_0(\operatorname{rot};\Omega)\ :\ m(\mathbf{u},\mathbf{v}) = (\mathbf{F},\mathbf{v})\qquad \forall\mathbf{v}\in H_0(\operatorname{rot};\Omega). \tag{4.1.14}$$

In the case of a conductor, ε, μ and σ are assumed to be symmetric matrices, uniformly positive definite in Ω, and with these positions the bilinear form $m(.,.)$ is continuous and coercive in $H(\operatorname{rot};\Omega)$; thus, existence and uniqueness of the solution follow from the Lax-Milgram lemma. Notice that in the so-called *low-frequency* case, which corresponds to omitting the term

$$-\int_\Omega \omega^2\varepsilon\mathbf{w}\cdot\overline{\mathbf{v}}$$

in the definition of $m(.,.)$, the coerciveness result still holds with an even easier proof. Finally, as long as $\sigma = 0$, the bilinear form $m(.,.)$ fails to be coercive, and existence and uniqueness of the solution stem from the Fredholm alternative theorem.

4.2 Domain Decomposition for Helmholtz Equations

In this section we present a domain decomposition approach to solve Helmholtz equation (4.1.4) and we introduce the algorithms appeared in literature within the framework of non-overlapping Schwarz methods.

Let Ω be a bounded domain in $\mathbf{R}^2$, with boundary $\partial\Omega$ regular enough (say, Lipschitz continuous), and let u be the unique weak solution of problem

$$\begin{cases} -\Delta u - \omega^2 u = f & \text{in } \Omega \\ (\partial_n + i\omega)u = 0 & \text{on } \partial\Omega, \end{cases}$$

that can be rewritten as

Find $u \in H^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v - \omega^2 \int_\Omega uv + i\omega \int_{\partial\Omega} uv = \int_\Omega fv, \tag{4.2.1}$$

for each $v \in H^1(\Omega)$. Notice that the boundary condition on $\partial\Omega$ is a simplification of the Sommerfeld radiation at infinity.

We partition the domain Ω into two non-overlapping subdomains Ω_1 and Ω_2, with interface Γ, and we can show that the equation (4.2.1) is equivalent to a transmission problem.

Proposition 4.2.1 *Problem* (4.2.1) *is equivalent to the following multidomain formulation. Find u_1 and u_2 weak solutions of*

$$\begin{cases} -\Delta u_1 - \omega^2 u_1 = f_1 & \text{in } \Omega_1 \\ -\Delta u_2 - \omega^2 u_2 = f_1 & \text{in } \Omega_2 \\ (\partial_n + i\omega)u_1 = 0 & \text{on } \partial\Omega_1 \cap \partial\Omega, \\ (\partial_n + i\omega)u_2 = 0 & \text{on } \partial\Omega_2 \cap \partial\Omega, \\ u_1 = u_2 & \text{on } \Gamma, \\ \dfrac{\partial u_1}{\partial n_1} = -\dfrac{\partial u_2}{\partial n_2} & \text{on } \Gamma. \end{cases} \tag{4.2.2}$$

with $f_i = f_{|\Omega_i}$, for $i = 1, 2$.

Proof. We denote with Λ the usual space of traces on the interface,

$$\Lambda := \{\eta \in H^{1/2}(\Gamma) \mid \eta = v_{|\Gamma}, \text{ for a suitable } v \in H^1(\Omega)\}.$$

Then, the weak formulation of (4.2.2) is the following

Find $u_1 \in H^1_\Gamma(\Omega_1)$, $u_2 \in H^1_\Gamma(\Omega_2)$ such that

$$\begin{cases} a_1(u_1, v_1) = (f, v_1)_{\Omega_1} & \forall v_1 \in H^1_\Gamma(\Omega_1) \\ u_1 = u_2 & \text{on } \Gamma \\ a_2(u_2, v_2) = (f, v_2)_{\Omega_2} & \forall v_2 \in H^1_\Gamma(\Omega_2) \\ a_2(u_2, \mathcal{R}_2\mu) = (f, \mathcal{R}_1\mu)_{\Omega_1} + (f, \mathcal{R}_2\mu)_{\Omega_2} - a_1(u_1, \mathcal{R}_1\mu) & \forall \mu \in \Lambda, \end{cases} \qquad (4.2.3)$$

where $\mathcal{R}_i\mu$ denotes any possible extension of μ to Ω_i, and where the bilinear forms $a_i(.,.)$ $(i = 1, 2)$ are defined as

$$a_i(u_i, v_i) := \int_{\Omega_i} \nabla u_i \cdot \nabla v_i - \omega^2 \int_{\Omega_i} u_i v_i + i\omega \int_{\partial\Omega_i \cap \partial\Omega} u_i v_i \qquad \forall u_i,\ v_i \in H^1(\Omega_1).$$

We start by considering the solution u to (4.2.1). If we set $u_i := u_{|\Omega_i}$, $i = 1, 2$, we have that $u_i \in H^1_\Gamma(\Omega_i)$, and that $(4.2.3)_1$, $(4.2.3)_2$, and $(4.2.3)_3$ are trivially satisfied. Moreover, for each $\mu \in \Lambda$, the function $\mathcal{R}\mu$ defined as

$$\mathcal{R}\mu := \begin{cases} \mathcal{R}_1\mu & \text{in } \Omega_1 \\ \mathcal{R}_2\mu & \text{in } \Omega_2 \end{cases}$$

belongs to $H^1(\Omega)$, thus we have

$$a(u, \mathcal{R}\mu) = (f, \mathcal{R}\mu),$$

which is equivalent to $(4.2.3)_4$, the weak form of the Neumann condition $(4.2.2)_6$
On the other hand, let u_i, $i = 1, 2$, be the solutions of (4.2.3). Setting

$$u := \begin{cases} u_1 & \text{in } \Omega_1 \\ u_2 & \text{in } \Omega_2, \end{cases}$$

we immediately have from $(4.2.3)_2$ that $u \in H^1(\Omega)$. Then, taking $v \in H^1(\Omega)$, we have that $\mu := v_{|\Gamma} \in \Lambda$. Define $\mathcal{R}\mu$ as before: clearly, $(v|_{\Omega_i} - \mathcal{R}_i\mu) \in H^1_\Gamma(\Omega_i)$, thus from $(4.2.3)_1$, $(4.2.3)_3$, and $(4.2.3)_4$ it follows that

$$\begin{aligned} a(u, v) &= \sum_{i=1}^{2} [a_i(u_i, v|\Omega_i - \mathcal{R}_i\mu) + a_i(u_i, \mathcal{R}_i\mu)] \\ &= \sum_{i=1}^{2} [(f, v|\Omega_i - \mathcal{R}_i\mu) + (f, \mathcal{R}_i\mu)_{\Omega_i}] \\ &= (f, v) \end{aligned}$$

and this concludes the proof. □

4.2.1 A Schwarz algorithm for Helmholtz equation

In this section we introduce an iterative algorithm of Schwarz type with non overlapping subdomains. This kind of approach has been used in literature by several people, among which B. Després in [35] and [16], P. Chevalier in [30], and M. Gander *et al.* in [51], who proposed different suitable interface conditions, that we are going to outline in the following.
Let Ω_j $(j = 1,..,N)$ be open sets such that $\Omega = \cup_j \Omega_j$. We define the global external boundary $\Sigma := \partial\Omega$, the local external boundaries $\Sigma_j := \partial\Omega_j \cap \Sigma$, and the interfaces $\Gamma_{jp} := \partial\Omega_j \cap \partial\Omega_p$. We set $\Gamma_j := \cup_j \Gamma_{jp}$, so that $\partial\Omega_j = \Gamma_j \cup \Sigma_j$. We denote with n_j the outward normal to $\partial\Omega_j$. Finally, let $\mathcal{B}_j$ be an operator acting on the interface Γ_j. We introduce the following Schwarz iterative procedure.

Given u_j^0 $(j = 1,..,N)$ in Ω_j, solve for $m \geq 1$

$$\begin{cases} -\Delta u_j^{m+1} - \omega^2 u_j^{m+1} = f_j & \text{in } \Omega_j \\ \left(\dfrac{\partial}{\partial n_j} + i\omega\right) u_j^{m+1} = 0 & \text{on } \Sigma_j, \\ \mathcal{B}_j(u_j^{m+1}) = \mathcal{B}_j(u_p^m) & \text{on } \Gamma_{jp},\ \forall p \ \text{ s.t. } \Gamma_{jp} \neq \emptyset. \end{cases} \tag{4.2.4}$$

The key point in the above algorithm is the shape of the interface operators $\mathcal{B}_j$, that must be chosen in a suitable way to recover, at convergence, the continuity of both the solution and of its normal derivative across the interfaces Γ_{jp}. Such algorithm has been firstly proposed by B. Després in [35], with the choice

$$\mathcal{B}_j := \partial_{n_j} + i\omega$$

for the interface transmission operator. In the rest of this section, we will outline a convergence analysis of the above algorithm (4.2.4) in a two-domains decomposition, following mainly the analysis made by P. Chevalier in his thesis [30], and M. Gander *et al.* in [51], who generalized Després' approach.
For that purpose, it is useful to consider the local Steklov-Poincaré operators. In a two-domain decomposition, owing to $(4.2.2)_5$-$(4.2.2)_6$, these local operators are defined as

$$\begin{aligned} \mathcal{S}_j : H^{1/2}(\Gamma) \times L^2(\Omega) &\longrightarrow H^{-1/2}(\Gamma) \\ (u_\Gamma, f) &\longmapsto \frac{\partial w_j}{\partial n_j} \end{aligned} \tag{4.2.5}$$

w_j $(j = 1, 2)$ being the solution of problem

$$\begin{cases} -\Delta w_j - \omega^2 w_j = f_j & \text{in } \Omega_j \\ \left(\dfrac{\partial}{\partial n_j} + i\omega\right) w_j = 0 & \text{on } \Sigma_j, \\ w_j = u_\Gamma & \text{on } \Gamma \end{cases}$$

It is immediate to see that $\mathcal{S}_2(u_{2|\Gamma}, f_2) = \frac{\partial u_2}{\partial n_2}_{|\Gamma}$, thus the transmission problem (4.2.2) entails that u_1 is solution of the problem

$$\begin{cases} -\Delta u_1 - \omega^2 u_1 = f_1 & \text{in } \Omega_1 \\ \left(\dfrac{\partial}{\partial n_1} + i\omega\right) u_1 = 0 & \text{on } \Sigma_1, \\ \dfrac{\partial u_1}{\partial n_1} + \mathcal{S}_2(u_{1|\Gamma}, 0) = -\mathcal{S}_2(0, f_2) & \text{on } \Gamma \end{cases} \tag{4.2.6}$$

A transparent condition for the interface Γ is thus given by the operator $(\partial_{n_1} + \mathcal{S}_2)$. An optimal choice for the operators $\mathcal{B}_j$ in (4.2.4) would thus be

$$\mathcal{B}_j = \partial_{n_j} + \mathcal{S}_{j+1}.$$

Unfortunately, as it is well known, the operators $\mathcal{S}_j$ are nonlocal, thus they are not differential operators along the interface.
We can clarify this with a Fourier analysis of the operator $\mathcal{S}_2$. For that purpose, we consider the plane $\mathbf{R}^2$ decomposed into the left $\Omega_1 = (-\infty, 0) \times \mathbf{R}$ and right half plane $\Omega_2 = (0, +\infty) \times \mathbf{R}$, with interface $\Gamma = \{0\} \times \mathbf{R}$. Let v be the solution of

$$\begin{cases} -\Delta v - \omega^2 v = 0 & \text{in } \Omega_2 \\ v = u & \text{on } \Gamma \\ \lim\limits_{r\to\infty} r\left(\dfrac{\partial v}{\partial r} + i\omega v\right) = 0 \end{cases} \tag{4.2.7}$$

where the Sommerfeld outgoing condition stems from the unboundedness of the domain. We perform a Fourier transform with respect to y, which is defined as

$$\mathcal{F} : v(x,y) \longmapsto \widehat{v}(x,k) = \int_{\mathbf{R}} v(x,y)\, e^{-iky}\, dy,$$

where k is the Fourier variable, with inverse denoted by $\mathcal{F}^{-1}$, and we can show that if v is solution of (4.2.7), then $\hat{v}$ is solution of

$$\begin{cases} \dfrac{\partial^2 \hat{v}}{\partial x^2} + (|k|^2 - \omega^2)\hat{v} = 0 & \text{in } \mathbf{R}^+ \\ \hat{v}(0,k) = \hat{u}(k). \end{cases} \tag{4.2.8}$$

Among the two solutions of this system, one must choose the one physically admissible (the one not exponentially growing), obtaining

$$\hat{v}(x,k) = \hat{u}(k)\, e^{-\lambda(k)\, x},$$

where $\lambda(k) := \sqrt{|k|^2 - \omega^2}$ is the root of the characteristic equation $\lambda^2 + (\omega^2 - k^2) = 0$ which is either positive real or purely imaginary:

$$\lambda(k) = \sqrt{k^2 - \omega^2} \text{ for } |k| > \omega, \quad \lambda(k) = i\sqrt{\omega^2 - k^2} \text{ for } |k| < \omega.$$

So far, we can easily evaluate the symbol of the Steklov-Poincaré operator, which is given by

$$\widehat{\mathcal{S}}(k) = \sqrt{|k|^2 - \omega^2}.$$

Since its symbol is not a polynomial in k, the Steklov-Poincaré is not a differential operator, thus, one must choose the interface operators $\mathcal{B}_j$ by looking for a suitable approximation of the Steklov-Poincaré operator by means of a differential one.

Remark 4.2.1 The peculiarity of the Steklov-Poincaré operator stems from its difference of behavior according to the frequency k. In fact, when $|k| > \omega$, the symbol $\widehat{\mathcal{S}}$ is real and problem (4.2.8) is elliptic, whereas as long as $|k| < \omega$, $\widehat{\mathcal{S}}$ is purely imaginary and problem (4.2.8) is not coercive. □

Convergence Analysis

We denote with Π_j the approximate Steklov-Poincaré operator on the interface Γ_j, we express the interface operator $\mathcal{B}_j$ as

$$\mathcal{B}_j = \frac{\partial}{\partial n_j} + \Pi_j, \quad j = 1, 2,$$

and we analyze the convergence to the zero solution of the coupled problem (4.2.4) when $f(x, y) = 0$ (it suffices, since the problem is linear). After a Fourier transform with respect to y, we obtain

$$\begin{aligned} \frac{\partial^2 \hat{u}_1^{n+1}}{\partial x^2} + (|k|^2 - \omega^2)\hat{u}_1^{n+1} &= 0 \qquad\qquad x < 0,\ k \in \mathbf{R} \\ (\partial_x + \pi_1(k))\hat{u}_1^{n+1}(0) &= (\partial_x + \pi_1(k))\hat{u}_2^n(0) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 \hat{u}_2^{n+1}}{\partial x^2} + (|k|^2 - \omega^2)\hat{u}_2^{n+1} &= 0 \qquad\qquad x > 0,\ k \in \mathbf{R} \\ (\partial_x + \pi_2(k))\hat{u}_2^{n+1}(0) &= (\partial_x + \pi_2(k))\hat{u}_1^n(0) \end{aligned}$$

where we have denoted with k the Fourier variable and with $\pi_j(k)$ the symbol of the operator Π_j. Since the Sommerfeld condition excludes both growing solutions and incoming modes at infinity, the solutions of these ordinary differential equations are

$$\hat{u}_1^{n+1}(x, k) = \hat{u}_2^n(0, k)\, e^{\lambda(k)x}, \qquad \hat{u}_2^{n+1}(x, k) = \hat{u}_1^n(0, k)\, e^{-\lambda(k)x},$$

where, again, $\lambda(k) := \sqrt{|k|^2 - \omega^2}$ is either real (and positive) or purely imaginary. Using the transmission conditions and the fact that

$$\frac{\partial \hat{u}_1^{n+1}}{\partial x}(x,k) = \lambda(k)\,\hat{u}_1^{n+1}, \qquad \frac{\partial \hat{u}_2^{n+1}}{\partial x}(x,k) = -\lambda(k)\,\hat{u}_2^{n+1}$$

we get one step of the Schwarz iteration as

$$\hat{u}_1^{n+1}(x,k) = \frac{-\lambda(k)+\pi_1(k)}{\lambda(k)+\pi_1(k)} e^{\lambda(k)x} \hat{u}_2^{n}(0,k),$$

$$\hat{u}_2^{n+1}(x,k) = \frac{\lambda(k)+\pi_2(k)}{-\lambda(k)+\pi_2(k)} e^{\lambda(k)x} \hat{u}_1^{n}(0,k).$$

If we evaluate the second equation at $x=0$ for the n-th iteration step, we insert it into the first one, and we evaluate this latter in $x=0$, we get

$$\hat{u}_1^{n+1}(0,k) = \frac{-\lambda(k)+\pi_1(k)}{\lambda(k)+\pi_1(k)}\,\frac{\lambda(k)+\pi_2(k)}{-\lambda(k)+\pi_2(k)}\,\hat{u}_1^{n-1}(0,k).$$

Then, introducing the convergence rate $\varrho(k)$ defined as

$$\varrho(k) \;=\; \frac{-\lambda(k)+\pi_1(k)}{\lambda(k)+\pi_1(k)}\,\frac{\lambda(k)+\pi_2(k)}{-\lambda(k)+\pi_2(k)}$$

we obtain by induction

$$\hat{u}_1^{2n}(0,k) = \varrho^n(k)\,\hat{u}_1^{0}(0,k).$$

It is not difficult to see that a similar argument in Ω_2 gives

$$\hat{u}_2^{2n}(0,k) = \varrho^n(k)\,\hat{u}_2^{0}(0,k).$$

Remark 4.2.2 It is immediate to observe, from the above calculations, that if we could choose $\Pi_j = \mathcal{S}_j$, the Steklov-Poincaré operator, we would have $\pi_1(k)=\lambda(k)$ and $\pi_2(k) = -\lambda(k)$, and consequently $\varrho(k)=0$, ensuring convergence in two iterations for the algorithm, independently of the initial guess $u_j^0(0,k)$, $j=1,2$. Unfortunately, as we observed in the previous section, due to the presence of the square root in the symbol, the Steklov-Poincaré is a non-local operator in the real domain, and the choice $\pi_j(k) = \pm\lambda(k)$ is not admissible. □

In Després' framework, we can easily evaluate the convergence rate of the Schwarz algorithm (4.2.4) as

$$|\varrho(k)| = \left|\frac{-\lambda(k)+i\omega}{\lambda(k)+i\omega}\,\frac{\lambda(k)-i\omega}{-\lambda(k)-i\omega}\right| = \left|\frac{(\lambda(k)-i\omega)^2}{(\lambda(k)+i\omega)^2}\right| = \left|\frac{\lambda(k)-i\omega}{\lambda(k)+i\omega}\right|^2.$$

The behavior of the symbol $\lambda(k)$ varies according to the frequency k, thus we have the following situations.

For $k<\omega$ (*i.e.* for propagative modes), $\lambda(k)$ is purely imaginary and

$$|\varrho(k)| = \left|\frac{i\sqrt{\omega^2-k^2}-i\omega}{i\sqrt{\omega^2-k^2}+i\omega}\right|^2 = \left|\frac{\sqrt{\omega^2-k^2}-\omega}{\sqrt{\omega^2-k^2}+\omega}\right|^2 < 1.$$

For $k > \omega$ (*i.e.* for evanescent modes), $\lambda(k)$ is real and

$$|\varrho(k)| = \left|\frac{\sqrt{k^2-\omega^2} - i\omega}{\sqrt{k^2-\omega^2} + i\omega}\right|^2 = 1,$$

since it is the ratio of two conjugated complex numbers.

Remark 4.2.3 The case $k = \omega$ represents the resonance frequency, and it corresponds to the resolution of an *ill-posed* problem, since the global problem could be rewritten as

$$\begin{cases} \partial_{xx} u = 0 & \text{in } \Omega, \\ \partial_n u = 0 & \text{on } \Gamma, \end{cases}$$

which admits infinitely many solutions. □

What we observe from the above argument, is that the Schwarz algorithm (4.2.4) does not converge for evanescent modes, whereas the reduction factor $\varrho(k)$ is very small for low frequencies ($k \ll \omega$). This is not in contradiction with Després' work, since he considered a problem with radiation conditions at finite distance, which allows only to propagative modes to "*enter*" the system.

4.2.2 Modified Transmission Conditions on the Interface

As we showed in the previous section, the Schwarz algorithm (4.2.4) with the choice $\mathcal{B}_j = \partial_{n_j} + i\omega$ does not converge for evanescent modes. To overcome this problem, several propositions have been made in literature, seeking for an optimization of the interface conditions in order to improve the convergence rate of the algorithm for the propagating modes and to achieve convergence also for the evanescent ones. In this section, we present the approaches introduced by P. Chevalier in [30], which consists in adding to Després interface conditions a space derivative of second order in the direction tangential to the interface, and M. Gander *et al.* in [51], who seek for an optimization of the convergence rate of the algorithm based on a slight modification of the transmission conditions proposed so far.

If we consider the Taylor expansion of the symbol $\widehat{\mathcal{S}}$, in the neighborhood of $k = 0$,

$$\widehat{\mathcal{S}}(k) = i\omega\sqrt{1 - \frac{k^2}{\omega^2}} = i\omega - i\frac{k^2}{2\omega} + i\frac{k^4}{8\omega^3} + O(k^6),$$

and we denote with $\hat{\mathcal{S}}^n$ its truncated at the n-th order, its inverse Fourier transform $\mathcal{F}^{-1}\hat{\mathcal{S}}^n$ is a differential operator along the interface in the physical space. The choice made by B. Després in [35] consists in an approximation of order zero of the Steklov-Poincaré operator, *i.e.*

$$\mathcal{B}_j = \frac{\partial}{\partial n_j} + i\omega,$$

which is the exact condition for a plane wave.

In his thesis P. Chevalier observes that choosing a second order approximation of the symbol of the Steklov-Poincaré operator does not help at all the convergence of evanescent modes, since the convergence rate in this case is given by (see [30], p. 79):

$$|\varrho_{(2)}(k)| = \begin{cases} \left| \dfrac{i\sqrt{\omega^2 - k^2} - i\omega + i\frac{k^2}{2\omega}}{i\sqrt{\omega^2 - k^2} + i\omega - i\frac{k^2}{2\omega}} \right| & \text{if } k < \omega \\[2ex] \left| \dfrac{\sqrt{k^2 - \omega^2} - i\omega + i\frac{k^2}{2\omega}}{\sqrt{k^2 - \omega^2} + i\omega - i\frac{k^2}{2\omega}} \right| & \text{if } k > \omega \end{cases}$$

and, again, $|\varrho_{(2)}(k)| = 1$ for $k > \omega$. This is a consequence of the fact that the zero-*th* order Taylor expansion is a good approximation of the symbol $\hat{\mathcal{S}}$ as long as it is purely imaginary (*i.e.* for $k < \omega$), thus one must improve the approximation along the real axis.
Observing that $\hat{\mathcal{S}}$ behaves like $|k|$ for large k, and it is symmetric in k, P. Chevalier proposes to approach $\hat{\mathcal{S}}$ with

$$\hat{\mathcal{S}}^1(k) = i\omega + \eta k^2, \qquad \eta \in \mathbb{C}.$$

Within this framework, since a low frequency approximation, given by $i(\omega + \mathrm{Im}(\eta)k^2)$, is mixed with an high frequency approximation, given by $\mathrm{Re}(\eta)k^2$, the real part of η will allow to approach $\hat{\mathcal{S}}$ for large values of k. Moreover, since $\hat{\mathcal{S}}^1(k)$ coincides with $\hat{\mathcal{S}}(k)$ when $k = 0$, also this condition is exact for a plane wave.
An inverse Fourier transform on the operator $\hat{\mathcal{S}}^1(k)$ leads, in the physical space, to the differential operator $\mathcal{S}^1 = i\omega - \eta\partial_{yy}$, which can be written in the more general form

$$\mathcal{S}^1 = i\omega - \eta\partial_\tau^2$$

where ∂_τ^2 denotes the second order derivative along the direction tangential to the interface. The convergence of the Schwarz algorithm with this interface conditions is given by the following result, which is proved in [30].

Theorem 4.2.1 (P. Chevalier) *If Re*$(\eta) > 0$ *and Im*$(\eta) > -\frac{1}{\omega}$*, the Schwarz algorithm* (4.2.4) *for a two domain decomposition with interface conditions given by*

$$\mathcal{B}_j = \partial_{n_j} + i\omega - \eta\partial_\tau^2, \qquad j = 1, 2$$

converges for all modes. □

Optimized Interface Conditions

Another way to overcome the problem of convergence for the evanescent modes is the optimization of interface conditions. Stemming from an analogous approach used for advection-diffusion problems by C. Japhet *et al.* (see [63], [64], and [65]), M. Gander *et al.* proposed in [51] an optimization procedure for the interface conditions, that we present in this section. They approximate the symbol of the Steklov-Poincaré operator with polynomials of degree at most 2 (in order to avoid an increase in bandwidth for the local subproblems), the choice of which relies on the optimization of the convergence rate, and it can be either a constant $\Pi_j^{\mathrm{R}} = \pm\alpha$, $\alpha \in \mathbb{C}$, or a constant plus a second order derivative along the interface, $\Pi_j^{\mathrm{O2}} = \pm(\zeta_1 + \zeta_2\partial_\tau^2)$, with

$\zeta, \zeta_2 \in \mathbb{C}$. According to the choice of the approximate operator Π_j, we thus will speak of "*Optimized Robin Transmission Conditions*" and "*Optimized Second Order Transmission Conditions*" (**OO2**), respectively.

Optimized Robin Transmission Conditions
The idea is to approximate the Steklov-Poincaré operator with

$$\Pi_j^{\mathrm{R}} = \pm(p + iq), \qquad p, q \in \mathbf{R},$$

and to optimize the convergence rate over p and q. It can be easily seen that, with this position, the convergence rate (so far a function of p, q, and the frequency k) becomes

$$\varrho(p,q,k) = \begin{cases} \dfrac{p^2 + (q - \sqrt{\omega^2 - k^2})^2}{p^2 + (q + \sqrt{\omega^2 - k^2})^2} & k \leq \omega^2 \\[2ex] \dfrac{q^2 + (p - \sqrt{\omega^2 - k^2})^2}{q^2 + (p + \sqrt{\omega^2 - k^2})^2} & k > \omega^2 \end{cases} \tag{4.2.9}$$

Remark 4.2.4 Notice that the case $k = \omega$ is still the resonance frequency, with $\varrho(p,q,k) = 1$ independently of the choice of the parameters p and q. However, the point $k = \omega$ represents one single mode in the spectrum and a Krylov method can easily take care of it as long as the Schwarz algorithm is used as a preconditioner. Moreover, if the point $k = \omega$ does not lie on a node of the discretization grid, this is not necessary. □

Differently from the case of positive definite problems (see, for instance, [64] or [65]), one cannot minimize ϱ over all the relevant frequencies, and one must face the following optimization problem

$$\min_{p,q\in\mathbf{R}} \left(\max_{k\in(k_{\min},\,\omega^-)\cup(\omega^+,\,k_{\max})} |\varrho(p,q,k)| \right), \tag{4.2.10}$$

where ω^- and ω^+ are parameter to be chosen in a suitable way to exclude the inflexion point $k = \omega$, while $k_{\min}$ is the smallest frequency relevant to the subdomain whereas $k_{\max}$ denotes the largest frequency supported by the numerical grid (of order π/h, h being the mesh size).
The solution of the optimization problem (4.2.10) is given by the following result.

Theorem 4.2.2 (M. Gander, F. Magoulès, F. Nataf) *Under the three assumption*

$$\begin{array}{rcll} 2\omega^2 & \leq & (\omega^-)^2 + (\omega^+)^2, & \omega^- < \omega \\ 2\omega^2 & > & k_{\min}^2 + (\omega^+)^2, & \\ 2\omega^2 & < & k_{\min}^2 + k_{\max}^2, & \end{array} \tag{4.2.11}$$

the min-max problem (4.2.10) *has a unique solution and the optimal parameters are given by*

$$p^* = q^* = \sqrt{\frac{\sqrt{\omega^2 - (\omega^-)^2}\sqrt{k_{\max}^2 - \omega^2}}{2}}.$$

The optimal convergence rate is then given by

$$\max_{k\in(k_{\min},\omega^-)\cup(\omega^+,k_{\max})} \varrho(p^*,q^*,k) = \frac{1-\sqrt{2}\left(\frac{\omega^2-(\omega^-)^2}{k_{\max}^2-\omega^2}\right)^{\frac{1}{4}}+\sqrt{\frac{\omega^2-(\omega^-)^2}{k_{\max}^2-\omega^2}}}{1+\sqrt{2}\left(\frac{\omega^2-(\omega^-)^2}{k_{\max}^2-\omega^2}\right)^{\frac{1}{4}}+\sqrt{\frac{\omega^2-(\omega^-)^2}{k_{\max}^2-\omega^2}}}.$$

□

The proof of the above theorem is rather tricky and can be found in [51]. We only observe that assumptions (4.2.11) are not restrictive. In fact, letting $\omega^{\pm}=\omega\pm\delta_{\pm}$, with $\delta_{\pm}>0$, assumption $(4.2.11)_1$ amounts to $2(\delta_+ +\delta_-)\omega+\delta_+^2+\delta_-^2\geq 0$, which is satisfied, for instance, for $\delta_-=\delta_+>0$. Since in practice $k_{\min}$ is small and ω_+ is close to ω, also assumption $(4.2.11)_2$ is not restrictive either. Finally, the thumb rule to take at least ten points per wavelength leads typically to a mesh size $h\leq\frac{\pi}{5\omega}$: if $k_{\max}=\pi/h$, this gives $k_{\max}>5\omega$ and also assumption $(4.2.11)_3$ is easily satisfied.

Optimized Second Order Transmission Conditions
The idea is to approximate the Steklov-Poincaré operator with

$$\Pi_j^{O2}=\pm(\zeta_1+\zeta_2\,\partial_\tau^2),\qquad \zeta_1,\,\zeta_2\in\mathbb{C}.$$

and to optimize the convergence rate over ζ_1 and ζ_2. The following Lemma allows to simplify the above interface condition

Lemma 4.2.1 *Let u_1 and u_2 be the solutions in Ω_j $(j=1,2)$ of*

$$-\Delta u_j-\omega^2u_j=f$$

with interface conditions

$$\left(\frac{\partial}{\partial n_1}+\alpha\right)\left(\frac{\partial}{\partial n_1}+\beta\right)(u_1)=\left(-\frac{\partial}{\partial n_2}+\alpha\right)\left(-\frac{\partial}{\partial n_2}+\beta\right)(u_2)$$

with $\alpha,\beta\in\mathbb{C}$, $\alpha+\beta\neq 0$, and n_j denoting the outward normal to the domain Ω_j. Then, the following second order interface condition is satisfied as well

$$\left(\frac{\partial}{\partial n_1}+\frac{\alpha\beta-\omega^2}{\alpha+\beta}-\frac{1}{\alpha+\beta}\frac{\partial^2}{\partial\tau_1^2}\right)(u_1)=\left(-\frac{\partial}{\partial n_2}+\frac{\alpha\beta-\omega^2}{\alpha+\beta}-\frac{1}{\alpha+\beta}\frac{\partial^2}{\partial\tau_2^2}\right)(u_2).$$

□

Owing to Lemma 4.2.1, M. Gander *et al.* propose to approximate the symbol of the Steklov-Poincarè operator with

$$\sigma_j^{\text{app}}:=\pm\left(\frac{\alpha\beta-\omega^2}{\alpha+\beta}+\frac{1}{\alpha+\beta}k^2\right)$$

which leads to a very simple formula for the convergence rate, given by

$$\varrho(\alpha,\beta,k) = \left(\frac{\lambda(k)-\alpha}{\lambda(k)+\alpha}\right)^2 \left(\frac{\lambda(k)-\beta}{\lambda(k)+\beta}\right)^2$$

where, as usual, $\lambda(k) = \sqrt{k^2-\omega^2}$, and one can play with the role of the parameters α, $\beta \in \mathbb{C}$ to optimize ϱ. By symmetry, it is enough to consider only positive values of the frequency k, and to optimize the convergence rate one has to solve the min-max problem

$$\min_{\alpha,\beta\in\mathbb{C}} \left(\max_{k\in(k_{\min},\omega^-)\cup(\omega^+,k_{\max})} |\varrho(\alpha,\beta,k)| \right), \tag{4.2.12}$$

where, again the parameters ω^- and ω^+ are chosen to exclude the resonance frequency $k=\omega$. With these positions, the convergence rate $\varrho(\alpha,\beta,k)$ consists of two factors, whereas $\lambda(k)$ is real for evanescent modes and purely imaginary for propagative ones, so if one chooses α purely imaginary ($\alpha \in i\mathbf{R}$), and β real ($\beta \in \mathbf{R}$), one of the two factors in the convergence rate is of modulus one, the first factor for vanishing modes, and the second one for propagative ones. This allows to optimize the convergence rate on a single parameter according to the region in the frequency domain. With this choice of α and β the min-max problem decouples, and one can consider the simpler problem

$$\min_{\alpha\in i\mathbf{R},\beta\in\mathbf{R}} \left(\max_{k\in(k_{\min},\omega^-)\cup(\omega^+,k_{\max})} |\varrho(\alpha,\beta,k)| \right), \tag{4.2.13}$$

which is solved by the following result, proved in [51], that we state here without proof.

Theorem 4.2.3 (M. Gander, F. Magoulès, F. Nataf) *The solution of the min-max problem* (4.2.13) *is unique and the optimal parameters are are given by*

$$\alpha^* := i[(\omega^2 - k_{\min}^2)(\omega^2 - (\omega^-)^2)]^{1/4} \quad \in i\mathbf{R}$$

and

$$\beta^* := [(k_{\max}^2 - \omega^2)((\omega^+)^2 - \omega^2)]^{1/4} \quad \in \mathbf{R}.$$

The convergence rate is then, for propagating modes, given by

$$\max_{k\in(k_{\min},\omega^-)} |\varrho(\alpha^*,\beta^*,k)| = \left(\frac{(\omega^2-(\omega^-)^2)^{1/4} - (\omega^2-k_{\min}^2)^{1/4}}{(\omega^2-(\omega^-)^2)^{1/4} + (\omega^2-k_{\min}^2)^{1/4}} \right)^2$$

whereas, for evanescent modes is given by

$$\max_{k\in(\omega^+,k_{\max})} |\varrho(\alpha^*,\beta^*,k)| = \left(\frac{(k_{\max}^2-\omega^2)^{1/4} - ((\omega^+)^2-\omega^2)^{1/4}}{(k_{\max}^2-\omega^2)^{1/4} + ((\omega^+)^2-\omega^2)^{1/4}} \right)^2 .$$

□

4.3 Domain Decomposition for Maxwell Equations

Let Ω be a bounded domain in $\mathbf{R}^3$. We consider the Maxwell problem in the auxiliary unknown $\mathbf{u}$, defined in (4.1.12)

$$\begin{cases} \text{rot }(\text{rot}\,\mathbf{u}) - \omega^2\mathbf{u} = \mathbf{F} & \text{in } \Omega \\ (\text{rot}\,\mathbf{u} \times \mathbf{n}) \times \mathbf{n} + i\,\omega\mathbf{u} \times \mathbf{n} = 0 & \text{on } \partial\Omega, \end{cases} \tag{4.3.1}$$

where we have set $\mu = \varepsilon = 1$ for simplicity of notations (notice that since $\sigma = 0$ the original wave problem is not positive). The boundary condition is a simplification of the Silver-Muller condition at infinity, and, recalling that the unknown $\mathbf{u}$ must be divergence free since it equals a **curl**, it is immediate to see that the solution $\mathbf{u}$ of (4.3.1) satisfies in Ω the vector Helmholtz problem

$$-\Delta\mathbf{u} - \omega^2\mathbf{u} = \mathbf{F}. \tag{4.3.2}$$

4.3.1 Multidomain formulation

We split Ω into two non-overlapping subdomains Ω_1 and Ω_2, we denote as usual the interface with Γ. According to the physics of the problem, we have to enforce the continuity across the interface of the tangential components of $\mathbf{u}$ and $\text{rot}\,\mathbf{u}$. Since these are the natural conditions, we end up with well-posed problems. The coupled problem reads
Find $\mathbf{u}_j \in H(\text{rot}\,,\Omega_j)$ such that

$$\begin{cases} \text{rot}\,(\text{rot}\,\mathbf{u}_j) - \omega^2\mathbf{u}_j = \mathbf{F} & \text{in } \Omega_j \quad j = 1,2 \\ (\text{rot}\,\mathbf{u}_j \times \mathbf{n}) \times \mathbf{n} + i\,\omega\mathbf{u}_j \times \mathbf{n} = 0 & \text{on } \partial\Omega \cap \partial\Omega_j \quad j = 1,2 \\ \mathbf{u}_1 \times \mathbf{n} = \mathbf{u}_2 \times \mathbf{n} & \text{on } \Gamma \\ \text{rot}\,\mathbf{u}_1 \times \mathbf{n} = \text{rot}\,\mathbf{u}_2 \times \mathbf{n} & \text{on } \Gamma \end{cases} \tag{4.3.3}$$

which can be proved to be equivalent to the single domain problem (4.3.1)

4.3.2 A Non-overlapping Schwarz Algorithm

During the last years, several proposition have been made in order to solve the coupled problem above. We consider here a Schwarz algorithm without overlap which reads

$$\begin{cases} \text{rot }\left(\text{rot}\,\mathbf{u}_1^{n+1}\right) - \omega^2\mathbf{u}_1^{n+1} = \mathbf{F} & \text{in } \Omega_1 \\ \text{rot }\left(\text{rot}\,\mathbf{u}_2^{n+1}\right) - \omega^2\mathbf{u}_2^{n+1} = \mathbf{F} & \text{in } \Omega_2 \\ \text{rot}\,\mathbf{u}_1^{n+1} \times \mathbf{n} + Z\,\mathbf{n} \times \mathbf{u}_1^{n+1} \times \mathbf{n} = \text{rot}\,\mathbf{u}_2^{n} \times \mathbf{n} + Z\,\mathbf{n} \times \mathbf{u}_2^{n} \times \mathbf{n} & \text{on } \Gamma, \\ \text{rot}\,\mathbf{u}_2^{n+1} \times \mathbf{n} - Z\,\mathbf{n} \times \mathbf{u}_2^{n+1} \times \mathbf{n} = \text{rot}\,\mathbf{u}_1^{n} \times \mathbf{n} - Z\,\mathbf{n} \times \mathbf{u}_1^{n} \times \mathbf{n} & \text{on } \Gamma \end{cases} \tag{4.3.4}$$

where $Z = p + iq \in \mathbb{C}$ is a suitable parameter, and n indicates the iteration step. This approach with the choice $Z = i\omega$ has been firstly proposed by Després *et al.* in [38], where they show that such algorithm converges weakly in $H(\text{rot}\,,\Omega_j)$, $(j = 1, 2)$ to the solution of the single domain problem, and that $\mathbf{u}_j^n$ and $\text{rot}\,\mathbf{u}_j^n \times \mathbf{n}$ converge to $\mathbf{u}_j$ and $\text{rot}\,\mathbf{u}_j \times \mathbf{n}$ respectively. We will show in the following that the convergence properties of this algorithm depend heavily on the radiation boundary condition. To be more specific, we will show that the algorithm (4.3.4) converges only for propagative modes, while for the evanescent ones the convergence rate is exactly 1. However, since the radiation condition allows only the propagative modes to come into the system, this is not in contradiction with Després' work.

Convergence Analysis

Since the problems involved are linear, it is enough to analyze the convergence to the zero solution for the homogeneous system. The tool for performing a convergence analysis is the Fourier transform. In that order, we denote by (x, y, z) a point in $\mathbf{R}^3$, and we consider the domain $\Omega = \mathbf{R}^3$ partitioned in the left $(\Omega_1 =]-\infty, 0[\times\mathbf{R}^2)$, and right $(\Omega_1 =]0, +\infty[\times\mathbf{R}^2)$ half spaces, with interface $\Gamma = \{0\} \times \mathbf{R}^2$. With this position, the unit outward normal directed from Ω_1 to Ω_2 is $\mathbf{n} = (1, 0, 0)$, and, for sake of simplicity in notations, throughout this section, we will denote $\mathbf{u}$ the solution in Ω_1 and $\mathbf{v}$ the one in Ω_2.

Since the solutions $\mathbf{u}$ and $\mathbf{v}$ represent the rotational of the magnetic fields in Ω_1 and Ω_2, , they must be divergence free, and using this fact in $(4.3.4)_1$-$(4.3.4)_2$, each iteration step in the Schwarz algorithm solves the following coupled problem of vectorial Helmholtz equations.

$$\begin{cases} -\Delta\mathbf{u}^{n+1} - \omega^2\mathbf{u}^{n+1} = \mathbf{0} & \text{in } \Omega_1 \\ \text{div}\,\mathbf{u}^{n+1} = 0 & \text{in } \Omega_1 \\ \text{rot}\,\mathbf{u}^{n+1} \times \mathbf{n} + Z\,\mathbf{n} \times \mathbf{u}^{n+1} \times \mathbf{n} = \text{rot}\,\mathbf{v}^n \times \mathbf{n} + Z\,\mathbf{n} \times \mathbf{v}^n \times \mathbf{n} & \text{on } \Gamma \end{cases} \tag{4.3.5}$$

and

$$\begin{cases} -\Delta\mathbf{v}^{n+1} - \omega^2\mathbf{v}^{n+1} = \mathbf{0} & \text{in } \Omega_2 \\ \text{div}\,\mathbf{v}^{n+1} = 0 & \text{in } \Omega_2 \\ \text{rot}\,\mathbf{v}^{n+1} \times \mathbf{n} - Z\,\mathbf{n} \times \mathbf{v}^{n+1} \times \mathbf{n} = \text{rot}\,\mathbf{u}^n \times \mathbf{n} - Z\,\mathbf{n} \times \mathbf{u}^n \times \mathbf{n} & \text{on } \Gamma, \end{cases} \tag{4.3.6}$$

We perform a partial Fourier transform in the y and z directions, that we denote with $\mathcal{F}$, and we call k_1 and k_2 the corresponding dual variables. The transform $\mathcal{F}$ is defined as

$$\mathcal{F} : \mathbf{E}(x, y, z) \longmapsto \widehat{\mathbf{E}}(x, k_1, k_2) = \int\int_{\mathbf{R}^2} \mathbf{E}(x, y, z)\, e^{-i(k_1 y + k_2 z)}\, dydz.$$

We can show that, if $\mathbf{u}$ and $\mathbf{v}$ are solutions of (4.3.5)-(4.3.6), then $\widehat{\mathbf{u}}$ and $\widehat{\mathbf{v}}$ are solutions of

$$\left\{\begin{array}{ll} -\dfrac{\partial^2 \widehat{\mathbf{u}}^{n+1}}{\partial x^2} + (k_1^2 + k_2^2 - \omega^2)\widehat{\mathbf{u}}^{n+1} = \mathbf{0} & \text{in } \Omega_1 \\[2ex] -\dfrac{\partial^2 \widehat{\mathbf{v}}^{n+1}}{\partial x^2} + (k_1^2 + k_2^2 - \omega^2)\widehat{\mathbf{v}}^{n+1} = \mathbf{0} & \text{in } \Omega_2 \end{array}\right. \tag{4.3.7}$$

which must satisfy the transformed divergence free conditions

$$\begin{array}{l} \partial_x \widehat{\mathbf{u}}_1^{n+1} - ik_1 \widehat{\mathbf{u}}_2^{n+1} - ik_2 \widehat{\mathbf{u}}_3^{n+1} = 0 \quad \text{in } \Omega_1 \\[1ex] \partial_x \widehat{\mathbf{v}}_1^{n+1} - ik_1 \widehat{\mathbf{v}}_2^{n+1} - ik_2 \widehat{\mathbf{v}}_3^{n+1} = 0 \quad \text{in } \Omega_2, \end{array} \tag{4.3.8}$$

as well as the transformed interface conditions on Γ. For any fixed k_1 and k_2, (4.3.7) are systems of ordinary differential equations whose solutions are

$$\hat{\mathbf{u}} = \vec{\alpha}_1(k_1, k_2)\ e^{\lambda x} + \vec{\beta}_1(k_1, k_2)\ e^{-\lambda x}$$

$$\hat{\mathbf{v}} = \vec{\alpha}_2(k_1, k_2)\ e^{\lambda x} + \vec{\beta}_2(k_1, k_2)\ e^{-\lambda x}$$

where we have set $\lambda = \lambda(k_1, k_2) := \sqrt{k_1^2 + k_2^2 - \omega^2}$, which is either real (and positive) or purely imaginary. Among these solutions, we must choose the ones physically admissible, that is the ones which are not exponentially increasing at infinity: this implies $\vec{\alpha}_2 = \vec{\beta}_1 = (0, 0, 0)$, and the solutions are completely determined by the boundary conditions at the interface Γ and the divergence free conditions (4.3.8).

Since, for each $\mathbf{w} \in H(\text{rot}\,, \Omega)$, we have

$$\widehat{\text{rot}\,\mathbf{w}} = \Big(- ik_1 \widehat{\mathbf{w}}_3 + ik_2 \widehat{\mathbf{w}}_2,\ -\partial_x \widehat{\mathbf{w}}_3 - ik_2 \widehat{\mathbf{w}}_1,\ \partial_x \widehat{\mathbf{w}}_2 + ik_1 \widehat{\mathbf{w}}_1 \Big), \tag{4.3.9}$$

and

$$\text{rot}\,\mathbf{w} \times \mathbf{n} = (0, [\text{rot}\,\mathbf{w}]_3, -[\text{rot}\,\mathbf{w}]_2), \qquad \mathbf{n} \times \mathbf{w} \times \mathbf{n} = (0, \mathbf{w}_2, \mathbf{w}_3), \tag{4.3.10}$$

where the subindeces denote components, it can be easily shown that the interface conditions $(4.3.5)_3$ and $(4.3.6)_3$ transform naturally into

$$\left\{\begin{array}{l} \partial_x \widehat{\mathbf{u}}_2^{n+1} + ik_1 \widehat{\mathbf{u}}_1^{n+1} + Z\widehat{\mathbf{u}}_2^{n+1} = \partial_x \widehat{\mathbf{v}}_2^n + ik_1 \widehat{\mathbf{v}}_1^n + Z\widehat{\mathbf{v}}_2^{n+1} \\[1ex] \partial_x \widehat{\mathbf{u}}_3^{n+1} + ik_2 \widehat{\mathbf{u}}_1^{n+1} + Z\widehat{\mathbf{u}}_3^{n+1} = \partial_x \widehat{\mathbf{v}}_2^n + ik_2 \widehat{\mathbf{v}}_1^n + Z\widehat{\mathbf{v}}_3^n \end{array}\right. \tag{4.3.11}$$

and

$$\left\{\begin{array}{l} \partial_x \widehat{\mathbf{v}}_2^{n+1} + ik_1 \widehat{\mathbf{v}}_1^{n+1} - Z\widehat{\mathbf{v}}_2^{n+1} = \partial_x \widehat{\mathbf{u}}_2^n + ik_1 \widehat{\mathbf{u}}_1^n - Z\widehat{\mathbf{u}}_2^n \\[1ex] \partial_x \widehat{\mathbf{v}}_2^{n+1} + ik_2 \widehat{\mathbf{v}}_1^{n+1} - Z\widehat{\mathbf{v}}_3^{n+1} = \partial_x \widehat{\mathbf{u}}_3^n + ik_2 \widehat{\mathbf{u}}_1^n - Z\widehat{\mathbf{u}}_3^n \end{array}\right. \tag{4.3.12}$$

all evaluated at $x = 0$. The divergence free conditions (4.3.8) provide, for all (x, k_1, k_2),

$$\widehat{\mathbf{u}}_1(x, k_1, k_2) = i\ \frac{k_1 \widehat{\mathbf{u}}_2(x, k_1, k_2) + k_2 \widehat{\mathbf{u}}_3(x, k_1, k_2)}{\lambda}$$

and

$$\widehat{\mathbf{v}}_1(x,k_1,k_2) = -i\ \frac{k_1\widehat{\mathbf{v}}_2(x,k_1,k_2)+k_2\widehat{\mathbf{v}}_3(x,k_1,k_2)}{\lambda},$$

that can be used in (4.3.11) and (4.3.12) obtaining

$$\left\{\begin{array}{l} \lambda\widehat{\mathbf{u}}_2^{n+1} - \dfrac{k_1^2}{\lambda}\widehat{\mathbf{u}}_2^{n+1} - \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{u}}_3^{n+1} + Z\widehat{\mathbf{u}}_2^{n+1} = -\lambda\widehat{\mathbf{v}}_2^n + \dfrac{k_1^2}{\lambda}\widehat{\mathbf{v}}_2^n + \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{v}}_3^n + Z\widehat{\mathbf{v}}_2^n \\ \\ \lambda\widehat{\mathbf{u}}_3^{n+1} - \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{u}}_2^{n+1} - \dfrac{k_2^2}{\lambda}\widehat{\mathbf{u}}_3^{n+1} + Z\widehat{\mathbf{u}}_3^{n+1} = -\lambda\widehat{\mathbf{v}}_2^n + \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{v}}_2^n + \dfrac{k_2^2}{\lambda}\widehat{\mathbf{v}}_3^n + Z\widehat{\mathbf{v}}_3^n, \end{array}\right. \qquad (4.3.13)$$

where $\hat{\mathbf{u}}_i^{n+1} := \hat{\mathbf{u}}_i^{n+1}(0,k_1,k_2)$ and $\hat{\mathbf{v}}_i^n := \hat{\mathbf{v}}_i^n(0,k_1,k_2)$, for $i=2,3$, and

$$\left\{\begin{array}{l} -\lambda\widehat{\mathbf{v}}_2^{n+1} + \dfrac{k_1^2}{\lambda}\widehat{\mathbf{v}}_2^{n+1} + \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{v}}_3^{n+1} + Z\widehat{\mathbf{v}}_2^{n+1} = \lambda\widehat{\mathbf{u}}_2^n - \dfrac{k_1^2}{\lambda}\widehat{\mathbf{u}}_2^n - \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{u}}_3^n + Z\widehat{\mathbf{u}}_2^n \\ \\ -\lambda\widehat{\mathbf{v}}_2^{n+1} + \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{v}}_2^{n+1} + \dfrac{k_2^2}{\lambda}\widehat{\mathbf{v}}_3^{n+1} + Z\widehat{\mathbf{v}}_3^{n+1} = \lambda\widehat{\mathbf{u}}_3^n - \dfrac{k_1k_2}{\lambda}\widehat{\mathbf{u}}_2^n - \dfrac{k_2^2}{\lambda}\widehat{\mathbf{u}}_3^n + Z\widehat{\mathbf{u}}_3^n. \end{array}\right. \qquad (4.3.14)$$

We set $\mathbf{u} := (\mathbf{u}_2,\mathbf{u}_3)$, $\mathbf{v} := (\mathbf{v}_2,\mathbf{v}_3)$, and we express the action of one iteration of the Schwarz algorithm as

$$B_1\widehat{\mathbf{u}}^{n+1} = \widetilde{B}_1\widehat{\mathbf{v}}^n, \qquad B_2\widehat{\mathbf{v}}^{n+1} = \widetilde{B}_2\widehat{\mathbf{u}}^n, \qquad (4.3.15)$$

where the matrices $B_1, \widetilde{B}_1, B_2$ and $\widetilde{B}_2$ are the following ones:

$$B_1 = \begin{pmatrix} \lambda - \dfrac{k_1^2}{\lambda} + Z & -\dfrac{k_1k_2}{\lambda} \\ \\ -\dfrac{k_1k_2}{\lambda} & \lambda - \dfrac{k_2^2}{\lambda} + Z \end{pmatrix}, \qquad \widetilde{B}_1 = \begin{pmatrix} -\lambda + \dfrac{k_1^2}{\lambda} + Z & \dfrac{k_1k_2}{\lambda} \\ \\ \dfrac{k_1k_2}{\lambda} & -\lambda + \dfrac{k_2^2}{\lambda} + Z \end{pmatrix}$$

as well as

$$B_2 = \begin{pmatrix} -\lambda + \dfrac{k_1^2}{\lambda} - Z & \dfrac{k_1k_2}{\lambda} \\ \\ \dfrac{k_1k_2}{\lambda} & -\lambda + \dfrac{k_2^2}{\lambda} - Z \end{pmatrix}, \qquad \widetilde{B}_2 = \begin{pmatrix} \lambda - \dfrac{k_1^2}{\lambda} - Z & -\dfrac{k_1k_2}{\lambda} \\ \\ -\dfrac{k_1k_2}{\lambda} & \lambda - \dfrac{k_2^2}{\lambda} - Z \end{pmatrix}$$

It is immediate to see that $B_1 = -B_2$ and $\widetilde{B}_1 = -\widetilde{B}_2$, providing

$$\widehat{\mathbf{u}}^{n+1} = B_1^{-1}\widetilde{B}_1\widehat{\mathbf{v}}^n = A\widehat{\mathbf{v}}^n, \qquad \widehat{\mathbf{v}}^{n+1} = B_2^{-1}\widetilde{B}_2\widehat{\mathbf{u}}^n = A\widehat{\mathbf{u}}^n.$$

Thus, given an initial guess $\widehat{\mathbf{u}}^0$ and $\widehat{\ }\ \mathbf{v}^0$, we have, for each $n \geq 1$

$$\widehat{\mathbf{u}}^{n+1} = A^{2n}\widehat{\mathbf{u}}^0, \qquad \widehat{\mathbf{v}}^{n+1} = A^{2n}\widehat{\mathbf{v}}^0. \qquad (4.3.16)$$

Since

$$\|\widehat{\mathbf{u}}^{n+1}\| \leq |||A|||^{2n}\,\|\widehat{\mathbf{u}}^{0}\|,$$

where $|||.|||$ is any matrix norm compatible with the vector norm $\|.\|$, we define the convergence rate (or, equivalently, the reduction factor) of the algorithm as being the spectral radius of the matrix A. Owing to the definition of the spectral radius of a matrix A, we therefore have

$$\rho(A) := \max_{\gamma\in\sigma(A)} |\gamma| \tag{4.3.17}$$

where we have denoted with $\sigma(A)$ the spectrum of the matrix A,

Remark 4.3.1 The choice of defining the reduction factor as the spectral radius of the interface mapping matrix is quite natural, since it equals the infimum of all compatible norms:

$$\rho(A) := \inf_{|||\cdot|||} |||A|||.$$

Thus, if $\rho(A) < 1$, there exists a norm such that the interface mapping is a contraction with respect to this norm, and the algorithm converges for all compatible norms. □

A simple algebraic manipulation provides

$$A = \begin{pmatrix} \dfrac{\omega^2 + Z^2 + \frac{k_1^2-k_2^2}{\lambda}\,Z}{-\omega^2 + Z^2 + (2\lambda - \frac{k_1^2+k_2^2}{\lambda})\,Z} & \dfrac{2\frac{k_1k_2}{\lambda}\,Z}{-\omega^2 + Z^2 + (2\lambda - \frac{k_1^2+k_2^2}{\lambda})\,Z} \\ & \\ \dfrac{2\frac{k_1k_2}{\lambda}\,Z}{-\omega^2 + Z^2 + (2\lambda - \frac{k_1^2+k_2^2}{\lambda})\,Z} & \dfrac{\omega^2 + Z^2 - \frac{k_1^2-k_2^2}{\lambda}\,Z}{-\omega^2 + Z^2 + (2\lambda - \frac{k_1^2+k_2^2}{\lambda})\,Z} \end{pmatrix}. \tag{4.3.18}$$

Since $\lambda = \sqrt{|k|^2 - \omega^2}$ (with $|k|^2 := k_1^2 + k_2^2$ for sake of simplicity in notations) is either real or purely imaginary according to the sign of $|k|^2 - \omega^2$, the matrix A will have a different shape in the two different regions of the Fourier space separated by the resonance frequency $|k| = \omega$. The eigenvalues of A are thus

$$\begin{cases} \gamma_L^{\pm} = \dfrac{\omega^2 + Z^2 \pm i|k|^2\frac{Z}{\sqrt{\omega^2-|k|^2}}}{-\omega^2 + Z^2 + i\frac{Z}{\sqrt{\omega^2-|k|^2}}\,(2\omega^2 - |k|^2)} & \text{if } |k|^2 < \omega^2 \\ & \\ \gamma_H^{\pm} = \dfrac{\omega^2 + Z^2 \pm |k|^2\frac{Z}{\sqrt{|k|^2-\omega^2}}}{-\omega^2 + Z^2 + \frac{Z}{\sqrt{|k|^2-\omega^2}}\,(|k|^2 - 2\omega^2)} & \text{if } |k|^2 > \omega^2, \end{cases} \tag{4.3.19}$$

and we thus have two different spectral radii to consider , $\rho_L(A)$ and $\rho_H(A)$ (where the subscripts L and H stay, in a certain sense, for *Low* and *High* frequency), depending on the value of $|k|^2 - \omega^2$. In the following lemma we prove that the non-overlapping Schwarz algorithm (4.3.4) cannot converge for all modes.

Lemma 4.3.1 *Let $Z = p + iq$, $p, q \in \mathbf{R}$, be a complex number. There is no possible choice of the parameters p and q in $\mathbf{R}$ such that the Schwarz algorithm* (4.3.4) *is convergent for all modes.*

Proof. If $q = 0$ and $Z = p$ is real, the eigenvalues of the iteration matrix are

$$\begin{cases} \gamma_L^\pm = \dfrac{\omega^2 + p^2 \pm i|k|^2 \frac{p}{\sqrt{\omega^2 - |k|^2}}}{-\omega^2 + p^2 + i\frac{p}{\sqrt{\omega^2-|k|^2}}(2\omega^2 - |k|^2)} & \text{if } |k|^2 < \omega^2 \\ \gamma_H^\pm = \dfrac{\omega^2 + p^2 \pm |k|^2 \frac{p}{\sqrt{|k|^2 - \omega^2}}}{-\omega^2 + p^2 + \frac{p}{\sqrt{|k|^2-\omega^2}}(|k|^2 - 2\omega^2)} & \text{if } |k|^2 > \omega^2 \end{cases}$$

If $|k|^2 < \omega^2$, the eigenvalues of A are complex: the numerators are conjugated, while the denominator remains unchanged. Thus we have $|\gamma_L^+| = |\gamma_L^-| = \rho_L$, and it is quite immediate to see that

$$\rho_L^2(A) = \frac{(\omega^2 + p^2)^2 + |k|^4 \frac{p^2}{\omega^2 - |k|^2}}{(-\omega^2 + p^2)^2 + \frac{p^2}{\omega^2-|k|^2}(2\omega^2 - |k|^2)^2} = 1,$$

since, evaluating the difference between the numerator (N) and the denominator (D) in the above formula, we easily have

$$\mathrm{N} - \mathrm{D} = (\omega^2 + p^2)^2 - (-\omega^2 + p^2)^2 + \frac{p^2}{\omega^2 - |k|^2}\left[|k|^4 - (2\omega^2 - |k|^2)^2\right] = 0.$$

For evanescent modes, the situation is even worse. Both eigenvalues are real and if $p > 0$ we have $|\gamma_H^-| \leq |\gamma_H^+|$. Thus,

$$\rho_H^2(A) = \frac{\left(\omega^2 + p^2 + |k|^2 \frac{p}{\sqrt{|k|^2-\omega^2}}\right)^2}{\left(-\omega^2 + p^2 + \frac{p}{\sqrt{|k|^2-\omega^2}}(|k|^2 - 2\omega^2)\right)^2} > 1.$$

Indeed, the difference between the numerator (N) and the denominator (D) in the above formula, after some algebraic manipulations, reads

$$\mathrm{N} - \mathrm{D} = 4\omega^2 \left(p^2 + \frac{p}{\sqrt{|k|^2 - \omega^2}}(|k|^2 - \omega^2 + p^2) + \frac{p^2}{|k|^2 - \omega^2}(|k|^2 - \omega^2)\right) > 0,$$

since $|k|^2 > \omega^2$. On the other hand, if $p < 0$, we have $|\gamma_H^+| \leq |\gamma_H^-|$ and

$$\rho_H^2(A) = \frac{\left(\omega^2 + p^2 - |k|^2 \frac{p}{\sqrt{|k|^2-\omega^2}}\right)^2}{\left(-\omega^2 + p^2 + \frac{p}{\sqrt{|k|^2-\omega^2}}(|k|^2 - 2\omega^2)\right)^2} > 1,$$

once again, since in this case

$$\mathrm{N}-\mathrm{D} = 4\omega^2 \left(p^2 + \frac{p^2}{|k|^2-\omega^2}(|k|^2-\omega^2) - 4p\frac{p^2}{\sqrt{|k|^2-\omega^2}} \right) > 0.$$

If $p = 0$ and $Z = iq$ is purely imaginary, the eigenvalues are

$$\begin{cases} \gamma_L^{\pm} = \dfrac{\omega^2 - q^2 \pm |k|^2 \frac{q}{\sqrt{\omega^2-|k|^2}}}{-\omega^2 - q^2 - \frac{q}{\sqrt{\omega^2-|k|^2}}(2\omega^2-|k|^2)} & \text{if } |k|^2 < \omega^2 \\ \\ \gamma_H^{\pm} = \dfrac{\omega^2 - q^2 \pm i|k|^2 \frac{q}{\sqrt{|k|^2-\omega^2}}}{-\omega^2 - q^2 + i\frac{q}{\sqrt{|k|^2-\omega^2}}(|k|^2-2\omega^2)} & \text{if } |k|^2 > \omega^2 \end{cases}$$

If $|k|^2 < \omega^2$, the eigenvalues of A are real and it is not difficult to see that, if $\omega^2 - q^2 \geq 0$, then $|\gamma_L^-| \leq |\lambda_L^+|$, whereas, if $\omega^2 - q^2 \leq 0$, $|\gamma_L^-| \geq |\gamma_L^+|$. Thus, if $\omega^2 - q^2 \geq 0$, we have

$$\rho_L^2(A) = \frac{\left(\omega^2 - q^2 + |k|^2 \frac{q}{\sqrt{\omega^2-|k|^2}}\right)^2}{\left(\omega^2 + q^2 + \frac{q}{\sqrt{\omega^2-|k|^2}}(2\omega^2-|k|^2)\right)^2} < 1, \tag{4.3.20}$$

since $|k|^2 < 2\omega^2 - |k|^2$. On the other hand, if $\omega^2 - q^2 \leq 0$, it is still more evident that

$$\rho_L^2(A) = \frac{\left(\omega^2 - q^2 - |k|^2 \frac{q}{\sqrt{\omega^2-|k|^2}}\right)^2}{\left(\omega^2 + q^2 + \frac{q}{\sqrt{\omega^2-|k|^2}}(2\omega^2-|k|^2)\right)^2} < 1. \tag{4.3.21}$$

for any choice of q.

If $|k|^2 > \omega^2$, the eigenvalues of A are complex: the numerators are conjugated, while the denominator remains unchanged. Thus $|\gamma_H^+| = |\gamma_H^-| = \rho_H$, and we have

$$\rho_H^2(A) = \frac{(\omega^2-q^2)^2 + |k|^4\frac{q^2}{|k|^2-\omega^2}}{(\omega^2+q^2)^2 + \frac{q^2}{|k|^2-\omega^2}(|k|^2-2\omega^2)^2} = 1, \tag{4.3.22}$$

since, evaluating the difference between the numerator (N) and the denominator (D) in the above formula, we easily have

$$\mathrm{N}-\mathrm{D} = (\omega^2-q^2)^2 - (\omega^2+q^2)^2 + \frac{q^2}{|k|^2-\omega^2}\left[|k|^4 - (|k|^2+2\omega^2)^2\right] = 0.$$

Finally, let us consider the case $Z = p + iq$, with $p, q \neq 0$. In this case the eigenvalues of A are

$$\begin{cases} \gamma_L^{\pm} = \dfrac{\left[\omega^2 + p^2 - q^2 \pm q\,\frac{|k|^2}{\sqrt{\omega^2-|k|^2}}\right] + i\left[2pq \mp p\,\frac{|k|^2}{\sqrt{\omega^2-|k|^2}}\right]}{\left[-\omega^2 + p^2 - q^2 - q\,\frac{2\omega^2-|k|^2}{\sqrt{\omega^2-|k|^2}}\right] + i\left[2pq + p\,\frac{2\omega^2-|k|^2}{\sqrt{\omega^2-|k|^2}}\right]} & \text{if } |k|^2 < \omega^2 \\[2ex] \gamma_H^{\pm} = \dfrac{\left[\omega^2 + p^2 - q^2 \pm p\,\frac{|k|^2}{\sqrt{|k|^2-\omega^2}}\right] + i\left[2pq \pm q\,\frac{|k|^2}{\sqrt{|k|^2-\omega^2}}\right]}{\left[-\omega^2 + p^2 - q^2 + p\,\frac{|k|^2-2\omega^2}{\sqrt{|k|^2-\omega^2}}\right] + i\left[2pq + q\,\frac{|k|^2-2\omega^2}{\sqrt{|k|^2-\omega^2}}\right]} & \text{if } |k|^2 > \omega^2 \end{cases}$$

Since most of the problems occurred when $|k|^2 > \omega^2$, we start our analysis from this case: both eigenvalues have the same denominator, thus, in order to evaluate the spectral radius of A, it is enough to compare the moduli of the numerators, which we denote with $\mathrm{N}(\gamma_H^+)$ and $\mathrm{N}(\gamma_H^-)$, respectively. After some calculations, we have

$$|\mathrm{N}(\gamma_H^+)|^2 - |\mathrm{N}(\gamma_H^-)|^2 = 4p\,\frac{|k|^2}{\sqrt{|k|^2-\omega^2}}(\omega^2 + p^2 + q^2),$$

thus, if $p > 0$, then $|\gamma_H^+| \geq |\gamma_H^-|$, and the opposite if $p < 0$. The spectral radius is thus

$$\rho_H^2(A) = \begin{cases} |\gamma_H^+|^2 = \dfrac{\left[\omega^2+p^2-q^2+q\,\frac{|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2 + \left[2pq-p\,\frac{|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2}{\left[-\omega^2+p^2-q^2-q\,\frac{2\omega^2-|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2 + \left[2pq+p\,\frac{2\omega^2-|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2} & \text{if } p > 0, \\[2ex] |\gamma_H^-|^2 = \dfrac{\left[\omega^2+p^2-q^2-q\,\frac{|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2 + \left[2pq+p\,\frac{|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2}{\left[-\omega^2+p^2-q^2-q\,\frac{2\omega^2-|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2 + \left[2pq+p\,\frac{2\omega^2-|k|^2}{\sqrt{\omega^2-|k|^2}}\right]^2} & \text{if } p < 0. \end{cases} \tag{4.3.23}$$

We then calculate the difference between the numerator N and the denominator D in both the occurrences in (4.3.23). If $p > 0$, we get after some calculations

$$\mathrm{N} - \mathrm{D} = 4\omega^2\left(2p^2 + p\,\frac{p^2 + |k|^2}{\sqrt{|k|^2-\omega^2}}\right) > 0,$$

independently of q, while, if $p < 0$,

$$\begin{aligned} \mathrm{N} - \mathrm{D} &= 8\omega^2 p^2 - 4\,\frac{p}{\sqrt{|k|^2-\omega^2}}\Big\{(|k|^2-\omega^2)\,(p^2-q^2) + 2q^2|k|^2 + \omega^2\Big\} \\ &= 8\omega^2 p^2 - 4\,\frac{p}{\sqrt{|k|^2-\omega^2}}\Big\{p^2(|k|^2-\omega^2) + \omega^2 + q^2(|k|^2+\omega^2)\Big\} > 0, \end{aligned}$$

again independently of q. So far, it is not necessary to analyze what happens in the case $|k|^2 < \omega^2$. □

Remark 4.3.2 We can resume the above analysis in the following way: if the coefficient Z in the Robin transmission condition on the interface is real, the Schwarz algorithm without overlap (4.3.4) does not converge neither for propagative modes, where the reduction factor is exactly 1, nor for evanescent modes, where the reduction factor is greater than 1. On the other hand, if Z is purely imaginary, the Schwarz algorithm without overlap (4.3.4) converges for propagating modes, where it shows a reduction factor that can be very small for the low frequencies, but it does not for evanescent ones, namely

$$\begin{cases} \rho_L(A) < 1 & \text{for propagative modes} \\ \\ \rho_H(A) = 1 & \text{for evanescent modes} \end{cases}$$

This is the same situation occurring in the case of Helmholtz equation (see Section 4.2.1) for a non-overlapping Schwarz algorithm with Robin transmission conditions at the interface. Finally, if the coefficient Z is complex with non zero real part the reduction factor for evanescent modes is greater than 1. This is a situation completely different from the case of Helmholtz equation, where the reduction factor associated to the Schwarz algorithm with a complex interface coefficient was smaller than 1 for both propagative and evanescent modes: in that occurrence the real part of Z allowed to control the interface operator a long as $|k|$ increased. Unfortunately, for Maxwell's system, this is not the case. An hint in that direction could have come from the following simple remark. In the case of Helmholtz equation, there is a sort of duality between a purely imaginary and a real coefficient, which does not hold in the case of Maxwell's system: in each region of the frequency space, one of the twos guarantees convergence, while the other one does not. See (4.2.9): the choice $Z = iq$ provides a reduction factor strictly less than 1 for propagative modes and exactly 1 for the evanescent ones, and the vice versa happens with $Z = p$. □

Remark 4.3.3 Notice that the choice made by B. Després, $Z = i\omega$, simplifies a little bit the expression of the spectral radius, but shows the same drawback. Infact, with Després choice, we have

$$\rho^2(A) = \begin{cases} \rho_L^2(A) = \dfrac{\frac{|k|^4}{\omega^2-|k|^2}}{\left(2\omega + \frac{2\omega-|k|^2}{\sqrt{\omega^2-|k|^2}}\right)^2} < 1 & \text{if } |k|^2 < \omega^2 \\ \\ \rho_H^2(A) = \dfrac{\frac{|k|^4}{|k|^2-\omega^2}}{4\omega^2 + \frac{(|k|^2-2\omega)^2}{|k|^2-\omega^2}} = 1 & \text{if } |k|^2 > \omega^2 \end{cases}$$

and, again the algorithm does not converge for evanescent modes. However, B. Després, on one hand, considers a radiation condition at finite distance, which allows only the propagative modes to be significant in the system, and, on the other hand, states a weak convergence result, which is not in contradiction with the fact that the interface iteration map is non-expansive. □

Second Order Interface Conditions

Another way to overcome the lack of convergence for evanescent modes, in the case of Helmholtz equation, relied on the addition of a Laplacian in the direction tangential to the interface. The use of a second order derivative, stemmed from the need to preserve the symmetry of the original problem. A similar approach can be used in the case of the Maxwell's system by means of the vectorial tangential Laplacian operator Δ_Γ, which is a second order differential operator acting along the direction tangential to the interface. The vectorial tangential Laplacian (also known as Hodge operator) is defined, for any vector field $\mathbf{u}$ tangent to the surface, as

$$\Delta_\Gamma \mathbf{u} := \nabla_\Gamma \mathrm{div}_\Gamma \mathbf{u} - \overrightarrow{\mathrm{rot}}_\Gamma \mathrm{rot}_\Gamma \mathbf{u}. \tag{4.3.24}$$

where the operators ∇_Γ (tangential gradient), $\overrightarrow{\mathrm{rot}}_\Gamma$ (tangential rotational), div_Γ (surfacic divergence), and rot_Γ (surfacic rotational), are defined as

$$\nabla_\Gamma \varphi = \mathbf{n} \times \nabla\varphi \times \mathbf{n} \qquad \overrightarrow{\mathrm{rot}}_\Gamma \varphi = \nabla_\Gamma \varphi \times \mathbf{n}$$

$$\mathrm{div}_\Gamma \mathbf{u} = \mathrm{div}\,(\mathbf{u} \times \mathbf{n}) \qquad \mathrm{rot}_\Gamma \mathbf{u} = \mathrm{div}_\Gamma(\mathbf{u} \times \mathbf{n})$$

for any scalar field φ defined on Γ, and for any vector field $\mathbf{u}$ tangential to Γ.
The idea is to consider again a non-overlapping Schwarz algorithm as in (4.3.4), with modified interface conditions: we thus propose the following procedure

Given $\mathbf{u}_1^0$ and $\mathbf{u}_2^0$ in Ω_1 and Ω_2 respectively, solve for $n \geq 1$

$$\begin{cases} \mathrm{rot}\,\left(\mathrm{rot}\,\mathbf{u}_1^{n+1}\right) - \omega^2 \mathbf{u}_1^{n+1} = \mathbf{F} & \text{in } \Omega_1 \\ (\mathrm{rot}\,\mathbf{u}_1 \times \mathbf{n}) \times \mathbf{n} + i\,\omega \mathbf{u}_1 \times \mathbf{n} = 0 & \text{on } \partial\Omega \cap \partial\Omega_1 \\ \mathrm{rot}\,\mathbf{u}_1^{n+1} \times \mathbf{n} + i\omega\,\mathbf{n} \times \mathbf{u}_1^{n+1} \times \mathbf{n} + \eta\Delta_\Gamma \mathbf{u}_1^{n+1} = & \\ \qquad = \mathrm{rot}\,\mathbf{u}_2^{n} \times \mathbf{n} + i\omega\,\mathbf{n} \times \mathbf{u}_2^{n} \times \mathbf{n} + \eta\Delta_\Gamma \mathbf{u}_2^{n} & \text{on } \Gamma, \end{cases} \tag{4.3.25}$$

and

$$\begin{cases} \mathrm{rot}\,\left(\mathrm{rot}\,\mathbf{u}_2^{n+1}\right) - \omega^2 \mathbf{u}_2^{n+1} = \mathbf{F} & \text{in } \Omega_2 \\ (\mathrm{rot}\,\mathbf{u}_2 \times \mathbf{n}) \times \mathbf{n} + i\,\omega \mathbf{u}_2 \times \mathbf{n} = 0 & \text{on } \partial\Omega \cap \partial\Omega_2 \\ \mathrm{rot}\,\mathbf{u}_2^{n+1} \times \mathbf{n} - i\omega\,\mathbf{n} \times \mathbf{u}_2^{n+1} \times \mathbf{n} + \eta\Delta_\Gamma \mathbf{u}_2^{n+1} = & \\ \qquad = \mathrm{rot}\,\mathbf{u}_1^{n} \times \mathbf{n} - i\omega\,\mathbf{n} \times \mathbf{u}_1^{n} \times \mathbf{n} + \eta\Delta_\Gamma \mathbf{u}_1^{n} & \text{on } \Gamma \end{cases} \tag{4.3.26}$$

where $\eta = \alpha + i\beta$, $\alpha, \beta \in \mathbf{R}$, is a complex number, while the reason of the choice $Z = i\omega$ relies in the convergence analysis of the previous section: the case $Z = iq$ is the only one such that the iterative map is a contraction for *"low"* frequencies, and it is not expansive for *"high"* ones. Finally, we set $q = \omega$ linking the interface condition to the radiation condition on the boundary. A convergence analysis for the above algorithm is not yet available, but a few comments are in order. On one hand, the use of a tangential second order operator in the case of Helmholtz equation led to an algorithm convergent for all modes: since we have reduced the original Maxwell's system to a vectorial Helmholtz problem, an analogous approach with the addition of a surfacic second order operator of rotational type could be promising in this latter case. On the other hand, the use of a complex coefficient in the Robin transmission condition in the case of Helmholtz equation ensured convergence for all modes, but the same approach does not help in the case of Maxwell's system. A convergence analysis of this latter algorithm must therefore be faced.

4.4 Conclusions

We have analyzed the non-overlapping additive Schwarz algorithm proposed by B. Després for the Maxwell's system, with a Robin interface condition depending upon a parameter. We have performed a convergence analysis in the special case of the space $\mathbf{R}^3$ partitioned in two half spaces by means of the Fourier transform: we have reduced the iteration step to a mapping on the interface, whose symbol, in the Fourier space, is a 2×2 matrix, and we have defined its convergence rate as being the spectral radius of this matrix. The algorithm is showed to converge for propagative modes (*i.e.* for low frequencies in the Fourier space), independently of the choice of the parameter. Unfortunately, for evanescent modes (*i.e.* for high frequencies in the Fourier space) the convergence rate is exactly 1, no matter how one chooses the parameter, and the algorithm does not converge. This is the same drawback occurring when such algorithm is applied to the Helmholtz equation. Differently from this latter case, it is not enough to replace the purely imaginary coefficient, in front of the zero-*th* order term in the Robin interface condition, with a complex one with non zero real part, to achieve convergence also for evanescent modes: the convergence rate becomes greater than 1. We then proposed, as an opportunity to overcome this drawback, the addition to the interface condition of the vectorial tangential Laplace operator (multiplied by some constant $\eta \in \mathbb{C}$): since a similar approach, consisting in the addition of a second order derivative in the direction tangential to the interface, was successfully used for Helmholtz equation, this could be promising in order to achieve convergence also for evanescent modes, but the convergence analysis is not yet available and further work needs to be done in this direction.

Appendix A

In this appendix, we give a brief review of some instances which are well-known in literature. We recall some basic definitions and properties of some function spaces, and we give a brief presentation of the finite elements method for the approximation of elliptic partial differential equation.

A.1 Function Spaces

In this section we recall the definitions of some function spaces which have been often used in the book. A complete presentation of this subject can be found for instance in Yosida [100], Brezis [22], J.-L. Lions and Magenes [76], and Adams [4].

1. Hilbert and Banach spaces

Let V be a complex linear space. A *scalar product* on V is a map $(\cdot,\cdot) : V \times V \to \mathbf{C}$, linear in the first argument and such that $(w,v) = \overline{(v,w)}$ for each $w, v \in V$ (symmetry; the bar denotes complex conjugation); $(v,v) \geq 0$ for each $v \in V$ (positivity); and $(v,v) = 0$ if, and only if, $v = 0$. In the case of a real linear space, the map $(\cdot,\cdot)$ takes values in $\mathbf{R}$.
A *seminorm* is a map $||\cdot|| : V \to \mathbf{R}$ such that $||v|| \geq 0$ for each $v \in V$; $||cv|| = |c|\,||v||$ for each $c \in \mathbf{C}$ and $v \in V$; and $||w+v|| \leq ||w|| + ||v||$ for each $w, v \in V$ (triangular inequality).
A *norm* on V is a seminorm satisfying the additional property that $||v|| = 0$ if, and only if, $v = 0$. Two norms $||\cdot||$ and $|||\cdot|||$ on V are equivalent if there exist two positive constants M_1 and M_2 such that

$$M_1||v|| \leq |||v||| \leq M_2||v||$$

for each $v \in V$.
It is readily verified that any scalar product defines a norm by setting: $||v|| := (v,v)^{1/2}$. Moreover, any norm defines a distance: $d(w,v) := ||w-v||$.
A linear space V endowed with a scalar product (respectively, a norm) is called *pre-hilbertian* (respectively, *normed*) space. A sequence v_n is a Cauchy sequence in a normed space V if it is a Cauchy sequence with respect to the distance $d(w,v) = ||w-v||$. If any Cauchy sequence in a pre-hilbertian (respectively, normed) space V is convergent, the space V is called a *Hilbert space* (respectively, *Banach space*).

In a Hilbert space the *Schwarz inequality* holds:

$$|(w, v)| \leq ||w|| \, ||v|| \ \text{ for each } w, v \in V.$$

2. Dual spaces

If $(V, ||\cdot||_V)$ and $(W, ||\cdot||_W)$ are normed spaces, we denote by $\mathcal{L}(V; W)$ the set of linear, continuous functionals from V into W, and for $\mathcal{F} \in \mathcal{L}(V; W)$ we define the norm

$$||\mathcal{F}||_{\mathcal{L}(V;W)} := \sup_{\substack{v \in V \\ v \neq 0}} \frac{||\mathcal{F}(v)||_W}{||v||_V}.$$

Thus $\mathcal{L}(V; W)$ is a normed space; if W is a Banach space, then $\mathcal{L}(V; W)$ is a Banach space, too. If $W = \mathbf{C}$ (respectively, $W = \mathbf{R}$, if V is a real normed space), the space $\mathcal{L}(V; \mathbf{C})$ (respectively, $\mathcal{L}(V; \mathbf{R})$) is called the *dual space* of V and is denoted by V'. The norm in V' is indicated by $|| \cdot ||_{V'}$.

The bilinear form $\langle \cdot, \cdot \rangle$ from $V' \times V$ into $\mathbf{C}$ (respectively, $\mathbf{R}$) defined by $\langle \mathcal{F}, v \rangle := \mathcal{F}(v)$ is called the *duality pairing* between V' and V.

3. Weak convergence

In a normed space V it is possible to introduce another type of convergence, which is called *weak convergence.* It is defined as follows: a sequence v_n is called weakly convergent to $v \in V$ if $\mathcal{F}(v_n)$ converges to $\mathcal{F}(v)$ for each $\mathcal{F} \in V'$. Clearly, if the sequence v_n converges to v in V, it is also weakly convergent. The converse is not true, unless V is finite dimensional.

It can be proved that the weak limit v, if it exists, is unique. Moreover, if v_n is weakly convergent to $v \in V$, one has

$$||v|| \leq \liminf_{n \to \infty} ||v_n||.$$

4. The Riesz theorem and the Lax-Milgram lemma

An important result which holds in Hilbert spaces is the following one.

Theorem A.1.1 (Riesz representation theorem) *Let V be a (real or complex) Hilbert space, endowed with the scalar product $(\cdot, \cdot)_V$. If $\mathcal{F} \in V'$, then there exists a unique $w_{\mathcal{F}} \in V$ such that*

$$\mathcal{F}(v) = (w_{\mathcal{F}}, v)_V \qquad \forall \ v \in V.$$

As a consequence of the Riesz representation theorem, if V is a Hilbert space, then the dual V' is a Hilbert space which can be canonically identified to V.

Another consequence of the Riesz theorem is the following result.

Theorem A.1.2 (Lax–Milgram lemma) *Assume that V is a real Hilbert space, endowed with the scalar product $(\cdot,\cdot)_V$ and the norm $||\cdot||_V$, that $\mathcal{A} : V \times V \to \mathbf{R}$ is a bilinear form, and that $\mathcal{F} : V \to \mathbf{R}$ is a linear continuous functional, namely, $\mathcal{F} \in V'$. Assume, moreover, that $\mathcal{A}$ is continuous, namely,*

$$\exists\ \gamma > 0 : |\mathcal{A}(w,v)| \leq \gamma ||w||_V \, ||v||_V \qquad \forall\ w, v \in V,$$

and coercive, namely,

$$\exists\ \alpha > 0 : \mathcal{A}(v,v) \geq \alpha ||v||_V^2 \qquad \forall\ v \in V.$$

Then there exists a unique $u \in V$ solution to

$$\mathcal{A}(u,v) = \mathcal{F}(v) \qquad \forall\ v \in V,$$

and, moreover,

$$||u||_V \leq \frac{1}{\alpha} ||\mathcal{F}||_{V'}.$$

If V is a complex Hilbert space, the Lax–Milgram lemma holds provided that (9.1.5) is substituted by the assumption

$$\exists\ \alpha > 0 : |\mathcal{A}(v,v)| \geq \alpha ||v||_V^2 \qquad \forall\ v \in V.$$

5. L^p spaces

Let Ω be an open set contained in $\mathbf{R}^d$, $d \geq 1$, and consider in Ω the Lebesgue measure. A very important family of Banach spaces is the following one. Let $1 \leq p \leq \infty$, and consider the set of measurable functions v such that

$$\int_\Omega |v(\mathbf{x})|^p d\mathbf{x} < \infty, \quad 1 \leq p < \infty, \tag{A.1.1}$$

or, when $p = \infty$,

$$\sup\{|v(\mathbf{x})| \,|\, \mathbf{x} \in \Omega\} < \infty. \tag{A.1.2}$$

These spaces are usually denoted by $L^p(\Omega)$ and the associated norm is

$$||v||_{L^p(\Omega)} := \left(\int_\Omega |v(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}, \quad 1 \leq p < \infty, \tag{A.1.3}$$

or, when $p = \infty$,

$$||v||_{L^\infty(\Omega)} := \sup\{|v(\mathbf{x})| \,|\, \mathbf{x} \in \Omega\}. \tag{A.1.4}$$

More precisely, $L^p(\Omega)$ is indeed the space of classes of equivalence of measurable functions, satisfying (A.1.1) or (A.1.2), with respect to the equivalence relation: $w \equiv v$ if w and v are different on a subset having zero-measure. In other words, in the space $L^p(\Omega)$ two functions, different on a subset which has zero-measure, are identified to each other. Thus the definition of the space $L^\infty(\Omega)$ in (A.1.2) and of its norm in (A.1.4) should be modified in the following way: $v \in L^\infty(\Omega)$ if

$$\inf\{M \geq 0 \,|\, |v(\mathbf{x})| \leq M \text{ almost everywhere in } \Omega\} < \infty,$$

and

$$||v||_{L^\infty(\Omega)} := \inf\{M \geq 0 \,|\, |v(\mathbf{x})| \leq M \text{ almost everywhere in } \Omega\},$$

where 'almost everywhere in Ω' means 'everywhere except on a subset of Ω having zero-measure'. The space $L^2(\Omega)$ is a Hilbert space, endowed with the scalar product

$$(w, v)_{L^2(\Omega)} := \int_\Omega w(\mathbf{x})\, v(\mathbf{x})\, d\mathbf{x},$$

often indicated by $(w, v)_{0,\Omega}$ or simply (w, v).
If $1 \leq p < \infty$, the dual space of $L^p(\Omega)$ is given by $L^{p'}(\Omega)$, where $(1/p) + (1/p') = 1$ (and $p' = \infty$ if $p = 1$).

6. Distributions

Let us recall that $C_0^\infty(\Omega)$ (or $\mathcal{D}(\Omega)$) denotes the space of infinitely differentiable functions having compact support; that is, vanishing outside a bounded open set $\Omega' \subset \Omega$ which has a positive distance from the boundary $\partial\Omega$ of Ω.
It is useful to define the concept of convergence for sequences of $\mathcal{D}(\Omega)$. We say that $v_n \in \mathcal{D}(\Omega)$ converges to $v \in \mathcal{D}(\Omega)$ if there exists a closed bounded subset $K \in \Omega$ such that v_n vanishes outside K for each n, and for every non-negative multi-index α the derivative $D^\alpha v_n$ converges to $D^\alpha v$ uniformly in Ω. We recall that if $\alpha = (\alpha_1, ..., \alpha_d)$, α_i non-negative integers, then

$$D^\alpha v := \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} ... \partial x_d^{\alpha_d}},$$

where $|\alpha| := \alpha_1 + ... + \alpha_d$ is the length of α.
The space of linear functionals on $\mathcal{D}(\Omega)$ which are continuous with respect to the convergence introduced above is denoted by $\mathcal{D}'(\Omega)$ and its elements are called *distributions*. If $L \in \mathcal{D}'(\Omega)$ and $v \in \mathcal{D}(\Omega)$, the action of the functional L on v is usually denoted by the duality pairing $\langle L, v\rangle$.
It is easily seen that each function $w \in L^p(\Omega)$, $1 \leq p \leq \infty$, can be associated with the following distribution:

$$v \to \int_\Omega w(\mathbf{x})\, v(\mathbf{x})\, d\mathbf{x}, \quad v \in \mathcal{D}(\Omega).$$

However, setting for instance $\Omega = (-1, 1)$, the Dirac functional

$$v \to \langle \delta, v\rangle := v(0), \quad v \in \mathcal{D}(\Omega),$$

is a distribution which cannot be represented through any function belonging to $L^p(\Omega)$, $1 \leq p \leq \infty$.

We introduce now the *derivative* of a distribution. Let α be a non-negative multi-index and L a distribution. Then $D^\alpha L$ is the distribution defined as follows:

$$\langle D^\alpha L, v\rangle := (-1)^{|\alpha|}\langle L, D^\alpha v\rangle \qquad \forall\, v \in \mathcal{D}(\Omega).$$

Note that, from this definition, a distribution turns out to be infinitely differentiable. On the other hand, when L is a smooth function, it is easily verified by integrating by parts that its derivative in the sense of distributions coincides with the usual derivative.

Let us also recall that the Dirac distribution δ is the distributional derivative of the Heaviside function:

$$H(x) := \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} .$$

Finally, we say that the α-derivative of a distribution L is a function belonging to $L^p(\Omega)$, $1 \leq p \leq \infty$, if there exists a function $g_\alpha \in L^p(\Omega)$ such that

$$\langle D^\alpha L, v\rangle = \int_\Omega g_\alpha(\mathbf{x})\, v(\mathbf{x})\, d\mathbf{x} \qquad \forall\, v \in \mathcal{D}(\Omega).$$

7. Sobolev spaces

We finally introduce another class of functions, which furnish the natural environment for the variational theory of partial differential equations.

The *Sobolev space* $W^{k,p}(\Omega)$, k a non-negative integer and $1 \leq p \leq \infty$, is the space of functions $v \in L^p(\Omega)$ such that all the distributional derivatives of v of order up to k belong to $L^p(\Omega)$. In short

$$W^{k,p}(\Omega) := \{v \in L^p(\Omega) \mid \quad D^\alpha v \in L^p(\Omega) \text{ for each non-negative multi-index } \alpha \text{ such that } |\alpha| \leq k\}.$$

Clearly, for each p, $1 \leq p \leq \infty$, $W^{0,p}(\Omega) = L^p(\Omega)$ and $W^{k_2,p}(\Omega) \subset W^{k_1,p}(\Omega)$ when $k_1 \leq k_2$. For $1 \leq p < \infty$, $W^{k,p}(\Omega)$ is a Banach space with respect to the norm

$$||v||_{k,p,\Omega} := \left(\sum_{|\alpha| \leq k} ||D^\alpha v||^p_{L^p(\Omega)} \right)^{1/p} .$$

Moreover, its seminorm is defined as follows:

$$|v|_{k,p,\Omega} := \left(\sum_{|\alpha| = k} ||D^\alpha v||^p_{L^p(\Omega)} \right)^{1/p} .$$

On the other hand, $W^{k,\infty}(\Omega)$ is a Banach space with respect to the norm

$$||v||_{k,\infty,\Omega} := \max_{|\alpha| \leq k} ||D^\alpha v||_{L^\infty(\Omega)},$$

while the corresponding seminorm is denoted by

$$|v|_{k,\infty,\Omega} := \max_{|\alpha|=k} ||D^\alpha v||_{L^\infty(\Omega)}.$$

In particular, when $p = 2$ we write $H^k(\Omega)$ instead of $W^{k,2}(\Omega)$, $||\cdot||_{k,\Omega}$ and $|\cdot|_{k,\Omega}$ instead of $||\cdot||_{k,2,\Omega}$ and $|\cdot|_{k,2,\Omega}$, respectively.
Note that $H^k(\Omega)$ is a Hilbert space with respect to the scalar product

$$(w,v)_{k,\Omega} := \sum_{|\alpha|\le k} (D^\alpha w, D^\alpha v)_{0,\Omega}.$$

Finally, for $1 \le p < \infty$ we denote by $W_0^{k,p}(\Omega)$ the closure of $C_0^\infty(\Omega)$ with respect to the norm $||\cdot||_{k,p,\Omega}$, and with $W^{-k,p'}(\Omega)$ the dual space of $W_0^{k,p}(\Omega)$. As before, when $p = 2$ we write $H_0^k(\Omega)$ and $H^{-k}(\Omega)$ instead of $W_0^{k,2}(\Omega)$ and $W^{-k,2}(\Omega)$, respectively.
It can also be proved that $W_0^{0,p}(\Omega) = L^p(\Omega)$ and that, if Ω has a Lipschitz continuous boundary and $1 \le p < \infty$, $W^{k,p}(\Omega)$ is indeed the closure of $C^\infty(\overline{\Omega})$ with respect to the norm $||\cdot||_{k,p,\Omega}$. In other words, $C^\infty(\overline{\Omega})$ is *dense* in $W^{k,p}(\Omega)$ for $1 \le p < \infty$.
It is sometimes useful to consider the Sobolev space $W^{s,p}(\Omega)$, where $s \in \mathbf{R}$ and $1 \le p \le \infty$. For the general definition, we refer to Adams ([4]). In particular, we only recall that, if $\Omega = \mathbf{R}^d$ and $p = 2$, $W^{s,2}(\mathbf{R}^d) = H^s(\mathbf{R}^d)$ can be characterized as follows by means of the Fourier transform $\hat{v}(\xi)$:

$$H^s(\mathbf{R}^d) = \{v \in L^2(\mathbf{R}^d) \,|\, (1+|\xi|^2)^{s/2}\hat{v}(\xi) \in L^2(\mathbf{R}^d)\}.$$

When considering vector-valued functions $\mathbf{v} : \Omega \to \mathbf{R}^d$, the Hilbert space

$$H(\mathrm{div}\,;\Omega) := \{\mathbf{v} \in (L^2(\Omega))^d \,|\, \mathrm{div}\,\mathbf{v} \in L^2(\Omega)\}$$

is also often used. It is endowed with the graph norm

$$||\mathbf{v}||_{H(\mathrm{div}\,;\Omega)} := (||\mathbf{v}||_{0,\Omega}^2 + ||\mathrm{div}\,\mathbf{v}||_{0,\Omega}^2)^{1/2}.$$

Similarly to the preceding cases, if Ω has a Lipschitz continuous boundary, it can be proved that $H(\mathrm{div}\,;\Omega)$ is the closure of $(C^\infty(\overline{\Omega}))^d$ with respect to the norm $||\cdot||_{H(\mathrm{div}\,;\Omega)}$.
For three-dimensional vector-valued functions we also introduce the Hilbert space

$$H(\mathrm{rot}\,;\Omega) := \{\mathbf{v} \in (L^2(\Omega))^3 \,|\, \mathrm{rot}\,\mathbf{v} \in (L^2(\Omega))^3\},$$

endowed with the graph norm

$$||\mathbf{v}||_{H(\mathrm{rot}\,;\Omega)} := (||\mathbf{v}||_{0,\Omega}^2 + ||\mathrm{rot}\,\mathbf{v}||_{0,\Omega}^2)^{1/2}.$$

Again, if Ω has a Lipschitz continuous boundary, then $H(\mathrm{rot}\,;\Omega)$ is the closure of $(C^\infty(\overline{\Omega}))^3$ with respect to the norm $||\cdot||_{H(\mathrm{rot}\,;\Omega)}$.
Another important class of Sobolev spaces is given by $W^{s,p}(\Sigma)$, where $s \ge 0$, $1 \le p \le \infty$ and Σ is a suitable subset of the boundary $\partial\Omega$ (again, we write $H^s(\Sigma)$ instead of $W^{s,2}(\Sigma)$). Their definition needs the introduction of some technical tools, especially if Σ is a non-smooth hypersurface (for instance, the boundary of a polygonal domain). For this we refer to Adams ([4]) or Brezzi and Gilardi ([23]); however, we return on a characterization of these spaces in the following section. When $\Sigma = \partial\Omega$, the dual space of $H^s(\partial\Omega)$ is denoted by $H^{-s}(\partial\Omega)$.

A.1.1 Some results about Sobolev spaces

In this section we present without proofs some relevant properties enjoyed by functions belonging to Sobolev spaces. We mainly limit ourselves to the Hilbert spaces $H^s(\Omega)$, referring the reader to J.-L. Lions and Magenes ([76]) or Adams ([4]) for the general case and all the proofs.
Let us start with the so-called *trace* theorems. The trace on the boundary $\partial\Omega$ of a function $v \in H^s(\Omega)$ is, in a sense to make precise, the value of v restricted to $\partial\Omega$. Indeed, the latter statement has not even a meaning, as a function in $H^s(\Omega)$ is not univocally defined on subsets having measure equal to zero. If we denote by $C^0(\overline{\Omega})$ the space of continuous functions on $\overline{\Omega}$, the precise result reads as follows.

Theorem A.1.3 (Trace theorem) *Let Ω be a bounded open set of $\mathbf{R}^d$. Assume that the boundary $\partial\Omega$ is smooth and that $s > 1/2$.*

(a) There exists a unique linear continuous map $\gamma_0 : H^s(\Omega) \to H^{s-1/2}(\partial\Omega)$ such that $\gamma_0 v = v_{|\partial\Omega}$ for each $v \in H^s(\Omega) \cap C^0(\overline{\Omega})$.

(b) There exists a linear continuous map $\mathcal{R}_0 : H^{s-1/2}(\partial\Omega) \to H^s(\Omega)$ such that $\gamma_0\mathcal{R}_0\varphi = \varphi$ for each $\varphi \in H^{s-1/2}(\partial\Omega)$.

Analogous results hold true if we consider the trace $\gamma_\Sigma v$ over a smooth subset Σ of the boundary $\partial\Omega$. In particular, for $1/2 < s \leq 1$, it is enough to assume that the boundary $\partial\Omega$ or the set Σ are Lipschitz continuous.

Thus, we have seen that any function belonging to $H^{s-1/2}(\Sigma)$, $s > 1/2$ and Σ smooth, is the trace on Σ of a function in $H^s(\Omega)$. This provides a useful characterization of the space $H^{s-1/2}(\Sigma)$.
For vector functions belonging to $H(\mathrm{div}\,;\Omega)$ the following trace result can be proved.

Theorem A.1.4 (Normal trace theorem) *Let Ω be a bounded open set of $\mathbf{R}^d$ with a Lipschitz continuous boundary $\partial\Omega$.*

(a) There exists a unique linear continuous map $\gamma_n : H(\mathrm{div}\,;\Omega) \to H^{-1/2}(\partial\Omega)$ such that $\gamma_n\mathbf{v} = (\mathbf{v}\cdot\mathbf{n}^*)_{|\partial\Omega}$ for each $\mathbf{v} \in H(\mathrm{div}\,;\Omega) \cap (C^0(\overline{\Omega}))^d$.

(b) There exists a linear continuous map $\mathcal{R}_n : H^{-1/2}(\partial\Omega) \to H(\mathrm{div}\,;\Omega)$ such that $\gamma_n\mathcal{R}_n\varphi = \varphi$ for each $\varphi \in H^{-1/2}(\partial\Omega)$.

Here we have denoted by $\mathbf{n}^*$ the unit outward normal vector on $\partial\Omega$. Let us note, moreover, that the normal trace of a vector function $\mathbf{v} \in H(\mathrm{div}\,;\Omega)$ over a Lipschitz continuous subset Σ of $\partial\Omega$ different from the whole boundary $\partial\Omega$ does not belong in general to $H^{-1/2}(\Sigma)$, but to a larger space, which is usually denoted by $H_{00}^{-1/2}(\Sigma)$ (see, for instance, J.-L. Lions and Magenes, [76]).
Now, let us introduce the space

$$\mathcal{X}_{\partial\Omega} := \{\psi \in (H^{-1/2}(\partial\Omega))^3 \mid \psi\cdot\mathbf{n}^* = 0,\ \mathrm{div}_\tau\psi \in H^{-1/2}(\partial\Omega)\},$$

where $\mathrm{div}_\tau\psi$ denotes the tangential divergence of ψ (see, for example, Bègue *et al.*, [14]).
For three-dimensional vector functions belonging to $H(\mathrm{rot}\,;\Omega)$ the following trace result can be proved (see Alonso and Valli, [8]).

Theorem A.1.5 (Tangential trace theorem) *Let Ω be a bounded open set of $\mathbf{R}^3$ with a Lipschitz continuous boundary $\partial\Omega$.*

(a) There exists a unique linear continuous map $\gamma_\tau : H(\mathrm{rot}\,;\Omega) \to \mathcal{X}_{\partial\Omega}$ such that $\gamma_\tau \mathbf{v} = (\mathbf{n}^* \times \mathbf{v})_{|\partial\Omega}$ for each $\mathbf{v} \in H(\mathrm{rot}\,;\Omega) \cap (C^0(\overline{\Omega}))^3$.

(b) If either the boundary $\partial\Omega$ is smooth or Ω is a convex polyhedron, then there exists a linear continuous map $\mathcal{R}_\tau : \mathcal{X}_{\partial\Omega} \to H(\mathrm{rot}\,;\Omega)$ such that $\gamma_\tau \mathcal{R}_\tau \psi = \psi$ for each $\psi \in \mathcal{X}_{\partial\Omega}$.

By means of these trace operators it is possible to characterize the spaces $H^1_0(\Omega)$, $H_0(\mathrm{div}\,;\Omega) := \overline{(C_0^\infty(\Omega))^d}$ and $H_0(\mathrm{rot}\,;\Omega) := \overline{(C_0^\infty(\Omega))^3}$ (here the closure has to be intended with respect to the norms $||\cdot||_{H(\mathrm{div}\,;\Omega)}$ and $||\cdot||_{H(\mathrm{rot}\,;\Omega)}$, respectively). As a matter of fact, if the boundary $\partial\Omega$ is Lipschitz continuous, we have:

$$\begin{aligned} H^1_0(\Omega) &= \{v \in H^1(\Omega) \,|\, \gamma_0 v = 0\} \\ H_0(\mathrm{div}\,;\Omega) &= \{\mathbf{v} \in H(\mathrm{div}\,;\Omega) \,|\, \gamma_n \mathbf{v} = 0\} \\ H_0(\mathrm{rot}\,;\Omega) &= \{\mathbf{v} \in H(\mathrm{rot}\,;\Omega) \,|\, \gamma_\tau \mathbf{v} = \mathbf{0}\}. \end{aligned}$$

A similar characterization holds for the space

$$H^1_\Sigma(\Omega) := \{v \in H^1(\Omega) \,|\, \gamma_\Sigma v = 0\}.$$

An important result is the so-called Poincaré inequality.

Theorem A.1.6 (Poincaré inequality) *Assume that Ω is a bounded connected open set of $\mathbf{R}^d$ and that Σ is a (non-empty) Lipschitz continuous subset of the boundary $\partial\Omega$. Then there exists a constant $C_\Omega > 0$ such that*

$$\int_\Omega v^2(\mathbf{x})\, d\mathbf{x} \le C_\Omega \int_\Omega |\nabla v(\mathbf{x})|^2\, d\mathbf{x}$$

for each $v \in H^1_\Sigma(\Omega)$.

As a consequence of the density of $C^\infty(\overline{\Omega})$ in $H^1(\Omega)$ (under the assumption that $\partial\Omega$ is Lipschitz continuous), it is easily proved that for each $w, v \in H^1(\Omega)$ the following *Green formula* holds:

$$\int_\Omega (D_j w)\, v\, d\mathbf{x} = -\int_\Omega w\, D_j v\, d\mathbf{x} + \int_{\partial\Omega} (\gamma_0 w)\, (\gamma_0 v)\, n_j^*\, d\gamma, \quad j = 1, ..., d,$$

where we have denoted by D_j the partial derivative $\frac{\partial}{\partial x_j}$ and by $d\gamma$ the surface measure on $\partial\Omega$. Similarly, if $\mathbf{w} \in H(\mathrm{div}\,;\Omega)$ and $v \in H^1(\Omega)$, we find that

$$\int_\Omega (\mathrm{div}\,\mathbf{w})\, v\, d\mathbf{x} = -\int_\Omega \mathbf{w} \cdot \nabla v\, d\mathbf{x} + \int_{\partial\Omega} (\gamma_n \mathbf{w})\, (\gamma_0 v)\, d\gamma.$$

Finally, if $\mathbf{w} \in H(\mathrm{rot}\,;\Omega)$ and $\mathbf{v} \in (H^1(\Omega))^3$, we find that

$$\int_\Omega (\mathrm{rot}\,\mathbf{w}) \cdot \mathbf{v}\, d\mathbf{x} = \int_\Omega \mathbf{w} \cdot \mathrm{rot}\,\mathbf{v}\, d\mathbf{x} + \int_{\partial\Omega} (\gamma_\tau \mathbf{w}) \cdot (\gamma_0 \mathbf{v})\, d\gamma.$$

As we have already noted, the functions belonging to the Sobolev spaces $W^{s,p}(\Omega)$ are not univocally defined over subsets having measure equal to zero. However, if suitable restrictions on the indices s and p are assumed, these functions indeed turn out to be regular functions. This is made clear by the following theorem.

Theorem A.1.7 (Sobolev embedding theorem) *Assume that Ω is a (bounded or unbounded) open set of $\mathbf{R}^d$ with a Lipschitz continuous boundary, and that $1 \le p < \infty$. Then the following embeddings are continuous.*

(a) If $0 \le sp < d$, then $W^{s,p}(\Omega) \subset L^{p^*}(\Omega)$ for $p^* = dp/(d - sp)$.

(b) If $sp = d$, then $W^{s,p}(\Omega) \subset L^q(\Omega)$ for any q such that $p \le q < \infty$.

(c) If $sp > d$, then $W^{s,p}(\Omega) \subset C^0(\overline{\Omega})$.

In the one-dimensional case, we have in particular that $H^1(\Omega) \subset C^0(\overline{\Omega})$, with continuous embedding.

A.2 Finite elements approximation of elliptic problems

In this section we give a brief presentation of the finite element approximation theory. For more details, we refer the interested reader, for example, to Ciarlet ([32]) and Quarteroni and Valli ([88]).

To start with, assume that the set $\Omega \subset \mathbf{R}^d$, $d = 2, 3$, is a *polygonal* domain, that is, Ω is a bounded open connected subset such that $\overline{\Omega}$ is the union of a finite number of polygons (for $d = 2$) or polyhedra (for $d = 3$).

The finite element approximation is based on a finite decomposition

$$\overline{\Omega} = \bigcup_{K \in \mathcal{T}_h} K,$$

where

- each K is a polygon or a polyhedron with a non-empty internal part $\overset{\circ}{K}$
- $\overset{\circ}{K}_1 \cap \overset{\circ}{K}_2 = \emptyset$ for each distinct $K_1, K_2 \in \mathcal{T}_h$
- if $F = K_1 \cap K_2 \neq \emptyset$ (K_1 and K_2 being distinct elements of $\mathcal{T}_h$) then F is a common face, side, or vertex of K_1 and K_2
- diam $(K) \le h$ for each $K \in \mathcal{T}_h$.

$\mathcal{T}_h$ is called a *triangulation* of $\overline{\Omega}$ (see Fig. A.1).

In what follows we assume further that each element K of $\mathcal{T}_h$ can be obtained as $K = T_K(\hat{K})$, where $\hat{K}$ is a reference polygon or polyhedron and T_K is a suitable invertible affine map, i.e. $T_K(\hat{\mathbf{x}}) = B_K\hat{\mathbf{x}} + \mathbf{b}_K$, B_K being a non-singular matrix.

Moreover, we will confine ourselves to considering two different cases:

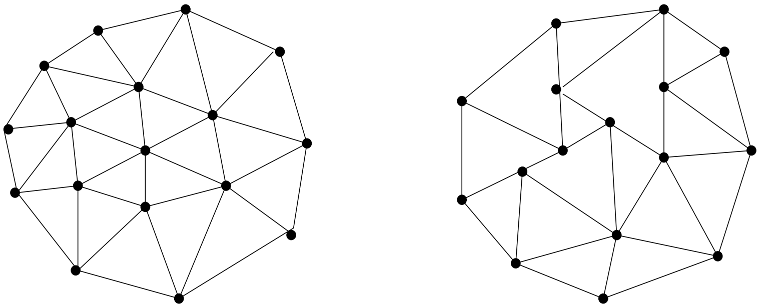

Figure A.1: Triangulation of Ω: admissible (left), non-admissible (right).

- The reference element $\hat{K}$ is the unit d-simplex; that is, the triangle with vertices (0,0), (1,0), (0,1) (when $d = 2$), or the tetrahedron with vertices (0,0,0), (1,0,0), (0,1,0), (0,0,1) (when $d = 3$). As a consequence, each $K = T_K(\hat{K})$ is a triangle or a tetrahedron.
- The reference element $\hat{K}$ is the unit d-cube $[0,1]^d$. As a consequence, each $K = T_K(\hat{K})$ is a parallelogram (when $d = 2$) or a parallelepiped (when $d = 3$).

Let V_h denote a finite dimensional subspace of $H_0^1(\Omega)$. A Galerkin finite element approximation to (3.1.7) is defined as follows:

$$\text{find } u_h \in V_h \ : \ a(u_h, v_h) = (f, v_h) \qquad \forall\, v_h \in V_h, \tag{A.2.1}$$

where we have set, for simplicity, $a(.,.) := a^0(.,.)$, the latter one being defined in (3.1.6).
The most frequent example is when V_h is given by piecewise polynomials. They can be introduced as follows. When the reference element $\hat{K}$ is the unit d-simplex, let us define

$$X_h^r := \{v_h \in C^0(\overline{\Omega}) \mid v_{h|K} \in \mathbf{P}_r(K)\ \forall\, K \in \mathcal{T}_h\}, \quad r \geq 1,$$

where $\mathbf{P}_r(K)$ denotes the set of polynomials defined in K and of degree less than or equal to r globally with respect to all space coordinates. Then we set

$$\begin{aligned} V_h : \ &= \{v_h \in X_h^r \mid v_{h|\partial\Omega} = 0\} \\ &= X_h^r \cap H_0^1(\Omega). \end{aligned}$$

When the reference element $\hat{K}$ is the unit d-cube, the space V_h is defined in the same way, but in this case the space X_h^r is given by

$$X_h^r := \{v_h \in C^0(\overline{\Omega}) \mid v_{h|K} \circ T_K \in \mathbf{Q}_r(K)\ \forall\, K \in \mathcal{T}_h\},$$

where $\mathbf{Q}_r(K)$ denotes the set of polynomials defined in K and of degree less than or equal to r with respect to each variable $x_1, \ldots, x_d$.
The family of triangulations $\mathcal{T}_h$ is said to be regular if there exists a constant $\sigma \geq 1$ such that

$$\frac{h_K}{\rho_K} \leq \sigma \qquad \forall\, K \in \mathcal{T}_h, \ \ \forall\, h > 0,$$

where h_K denotes the diameter of K and ρ_K the maximum diameter of a ball contained in K. Under this assumption, denoting by $\pi_h v \in V_h$ the interpolant of a continuous function v at the nodes of $\mathcal{T}_h$, the following interpolation error estimate holds:

$$||u - \pi_h u||_{0,\Omega} + h||u - \pi_h u||_{1,\Omega} \leq Ch^{r+1}|u|_{r+1,\Omega}. \tag{A.2.2}$$

It is well known from the Lax–Milgram lemma that problem (A.2.1) has a unique solution under the assumption that the bilinear form $a(\cdot,\cdot)$ is continuous and coercive in $H_0^1(\Omega)$ (see assumption (3.1.8)).

Besides, from the Céa lemma it follows that

$$||u - u_h||_{1,\Omega} \leq \frac{\gamma}{\alpha} \inf_{v_h \in V_h} ||u - v_h||_{1,\Omega},$$

where γ and α are the continuity and coerciveness constants of $a(\cdot,\cdot)$, respectively. By the interpolation error estimate (A.2.2), we finally find that

$$||u - u_h||_{1,\Omega} \leq Ch^r|u|_{r+1,\Omega},$$

provided that $u \in H^{r+1}(\Omega)$.

A.2.1 Algebraic formulation of the discrete problem

The unknowns of the finite dimensional problem (A.2.1) are given by the pointvalues of u_h at the finite element nodes $\mathbf{a}_j$. In fact, denoting by N_h the total number of the nodes and by φ_j the basis functions of V_h; that is, the unique functions in V_h satisfying $\varphi_j(\mathbf{a}_i) = \delta_{ij}$ for each $i, j = 1, \ldots, N_h$, each element $u_h \in V_h$ can be represented through

$$u_h(\mathbf{x}) = \sum_{j=1}^{N_h} u_h(\mathbf{a}_j)\varphi_j(\mathbf{x}).$$

Introducing the notation

$$\mathbf{u} := \{u_h(\mathbf{a}_j)\}_{j=1,\ldots,N_h}$$

and

$$\mathbf{f} := \{(f, \varphi_j)\}_{j=1,\ldots,N_h},$$

problem (A.2.1) can be rewritten as

$$A\mathbf{u} = \mathbf{f}.$$

The matrix A is called the finite element *stiffness* matrix and is given by

$$A_{lj} := a(\varphi_j, \varphi_l), \quad l, j = 1, \ldots, N_h.$$

The stiffness matrix A is positive definite; that is, for any $\mathbf{v} \in \mathbf{R}^{N_h}$, $\mathbf{v} \neq \mathbf{0}$, $(A\mathbf{v}, \mathbf{v}) > 0$, where $(\cdot,\cdot)$ denotes the Euclidean scalar product. Indeed, let $v_h \in V_h$ be the function defined as

$$v_h(\mathbf{x}) = \sum_{j=1}^{N_h} v_j \, \varphi_j(\mathbf{x}).$$

Then

$$\begin{aligned}(A\mathbf{v},\mathbf{v}) &= \sum_{l,j=1}^{N_h} v_i\, a(\varphi_j,\varphi_l)\, v_j \\ &= a(v_h,v_h) \geq 0,\end{aligned}$$

and $(A\mathbf{v},\mathbf{v}) = 0$ if and only if $v_h = 0$, or, equivalently, $\mathbf{v} = \mathbf{0}$. In particular, any eigenvalue of A has a positive real part. Besides, when the bilinear form $a(\cdot,\cdot)$ is symmetric, it follows immediately that A is also symmetric.
Another important remark is concerned with the condition number

$$\kappa_2(A) := ||A||_2\, ||A^{-1}||_2 = \frac{\sqrt{\lambda_{\max}(A^TA)}}{\sqrt{\lambda_{\min}(A^TA)}}.$$

In the symmetric case, we have the simplified relation

$$\kappa_2(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)},$$

and it can be proved that

$$\kappa_2(A) = O(h^{-2}).$$

A.2.2 The multi-domain formulation for finite elements

For sake of simplicity inpresentation, we split Ω into two subdomains Ω_1 and Ω_2, such that $\overline{\Omega_1} \cup \overline{\Omega_2} = \overline{\Omega}$, $\Omega_1 \cap \Omega_2 = \emptyset$, and we set $\Gamma := \overline{\Omega_1} \cap \overline{\Omega_2}$. We also suppose that the interface Γ does not cut any finite element T. This implies that the global triangulation $\mathcal{T}_h$ of $\overline{\Omega}$ induces two triangulations $\mathcal{T}_{h,1}$ of $\overline{\Omega_1}$ and $\mathcal{T}_{h,2}$ of $\overline{\Omega_2}$ that are compatible on Γ; that is, they share the same edges on Γ (see Fig. A.2).
Let us start with the variational formulation of our problem (an alternative characterisation based on a purely algebraic argument is given in Section 2.3). To this purpose let us define

$$\Lambda_h := \{v_{h|\Gamma} \mid v_h \in V_h\}, \quad V_{i,h} := \{v_{h|\Omega_i} \mid v_h \in V_h\},$$

and set

$$V_{i,h}^0 := \{v_h \in V_{i,h} \mid v_{h|\Gamma} = 0\}.$$

It can be shown that the finite element problem (A.2.1), after identifying $u_{1,h}$ with $u_{h|\Omega_1}$ and $u_{2,h}$ with $u_{h|\Omega_2}$, is equivalent to the multi-domain problem

$$\begin{cases} a_1(u_{1,h},v_{1,h}) = (f,v_{1,h})_{\Omega_1} \qquad \forall\, v_{1,h} \in V_{1,h}^0 \\ u_{1,h} = u_{2,h} \qquad \text{on } \Gamma \\ a_2(u_{2,h},v_{2,h}) = (f,v_{2,h})_{\Omega_2} \qquad \forall\, v_{2,h} \in V_{2,h}^0 \\ \displaystyle\sum_{i=1}^{2} a_i(u_{i,h},\mathcal{R}_{i,h}\mu_h) = \sum_{i=1}^{2} (f,\mathcal{R}_{i,h}\mu_h)_{\Omega_i} \qquad \forall\, \mu_h \in \Lambda_h, \end{cases} \tag{A.2.3}$$

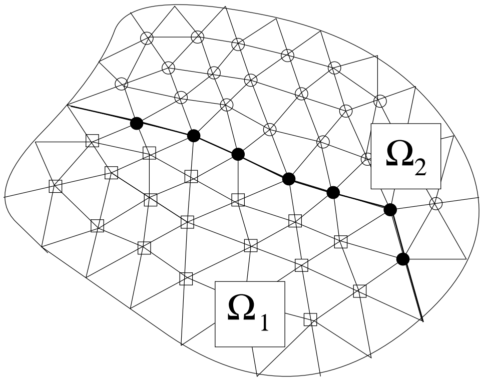

Figure A.2: Splitting of Ω and finite element triangulation

where the bilinear forms $a_i(.,.)$ are any of the ones defined in Section 3.1.1, while $\mathcal{R}_{i,h}$, $i = 1, 2$, is any extension operator from Λ_h into $V_{i,h}$. In practical implementation, these extension operators will be taken equal to the finite element interpolant $\pi_{i,h}\mu_h$, which belongs to $V_{i,h}$, equals μ_h at the nodes on the interface Γ, and vanishes at the internal nodes in Ω_i.
The formulation (A.2.3) may be generalised to many subdomains, possibly including cross-points. Note, in particular, that if $\mathcal{R}_{i,h}\mu_h$ is the restriction to Ω_i of the finite element shape function associated with a cross-point P, then (A.2.3)$_4$ enforces the continuity of the 'normal' derivative in P in a natural form.

A.2.3 Algebraic formulation of the discrete Steklov-Poincaré operator

In order to give the algebraic interpretation of the operator Σ at the finite dimensional level, we distinguish between the finite element nodes belonging to Γ and to each subdomain Ω_1 and Ω_2. We denote the corresponding vectors of finite element unknowns with $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_\Gamma$ respectively, and their lenghts by N_1, N_2 and N_Γ. The stiffness matrix A of the finite element problem is therefore a $N_h \times N_h$ matrix with $N_h = N_1 + N_2 + N_\Gamma$, while the vector corresponding to the datum f is denoted by $\mathbf{f} \in \mathbf{R}^{N_h}$. The resulting system can therefore be written in block form as

$$\begin{pmatrix} A_{11} & 0 & A_{1\Gamma} \\ 0 & A_{22} & A_{2\Gamma} \\ A_{\Gamma 1} & A_{\Gamma 2} & A_{\Gamma\Gamma} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_\Gamma \end{pmatrix} = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_\Gamma \end{pmatrix}, \tag{A.2.4}$$

where we have used the following notations: for $i = 1, 2$

$$(A_{ii})_{jk} := a_i(\varphi_k^{(i)}, \varphi_j^{(i)}), \quad j, k = 1, \dots, N_i,$$

where $\varphi_j^{(i)}$ are the basis functions associated to the nodes lying inside Ω_i, while $a_i(.,.)$ denotes the restriction to Ω_i of $a(.,.)$, which is either the bilinear form $a^0(.,.)$ defined in (3.1.6), or the bilinear form $a^R(.,.)$ defined in (3.1.12), according to the choice of interface conditions. Moreover, we have

$$(A_{\Gamma\Gamma})_{pq} := a_1(\varphi_q^{(\Gamma)}, \varphi_p^{(\Gamma)}) + a_2(\varphi_q^{(\Gamma)}, \varphi_p^{(\Gamma)}), \quad p, q = 1, \dots, N_\Gamma,$$

where $\varphi_j^{(i)}$ are the basis functions associated to the nodes lying on Γ, and finally, for $i = 1, 2$

$$(A_{i\Gamma})_{jp} := a_i(\varphi_p^{(\Gamma)}, \varphi_j^{(i)}), \quad j = 1, \dots, N_i, \ p = 1, \dots, N_\Gamma,$$

while $A_{\Gamma i}$ denotes the transpose of $A_{i\Gamma}$, for $i = 1, 2$.
After eliminating $\mathbf{u}_1$ and $\mathbf{u}_2$ we get the reduced system:

$$\Sigma_h \mathbf{u}_\Gamma = \chi_\Gamma, \tag{A.2.5}$$

with

$$\chi_\Gamma = \mathbf{f}_\Gamma - A_{\Gamma 1} A_{11}^{-1} \mathbf{f}_1 - A_{\Gamma 2} A_{22}^{-1} \mathbf{f}_2$$

and

$$\Sigma_h = A_{\Gamma\Gamma} - A_{\Gamma 1} A_{11}^{-1} A_{1\Gamma} - A_{\Gamma 2} A_{22}^{-1} A_{2\Gamma},$$

Once the solution $\mathbf{u}_\Gamma$ of (A.2.5) is available, the subdomain components $\mathbf{u}_1$ and $\mathbf{u}_2$ can be easily recovered from (A.2.4) at the cost of two independent solves A_{11}^{-1} and A_{22}^{-1}.
We can split the matrix of interface contributions as

$$A_{\Gamma\Gamma} = A_{\Gamma\Gamma}^{(1)} + A_{\Gamma\Gamma}^{(2)},$$

where $A_{\Gamma\Gamma}^{(i)}$ denotes the contribution from the subdomain Ω_i, $i = 1, 2$. We can therefore write

$$\Sigma_h = \Sigma_{1,h} + \Sigma_{2,h}$$

with

$$\Sigma_{i,h} := A_{\Gamma\Gamma}^{(i)} - A_{\Gamma i} A_{ii}^{-1} A_{i\Gamma}, \quad i = 1, 2.$$

The matrix Σ_h is the *Schur complement matrix*, the algebraic counterpart of the discrete Steklov-Poincaré operator

Bibliography

[1] Y. Achdou - Y. A. Kuznetsov, *Substructuring preconditioners for finite element methods on nonmatching grids*, East-West J. Numer. Math., 3 (1995), pp. 1–28.

[2] Y. Achdou - P. Le Tallec - F. Nataf - M. Vidrascu, *A domain decoposition preconditioner for an advection-diffusion problem*, Comp. Meth. Appl. Mech. Engng, 184 (2000), pp. 145–170.

[3] Y. Achdou - F. Nataf, *A Robin-Robin preconditioner for an advection-diffusion problem*, C. R. Acad. Sci. Paris, 325, Série I (1997), pp. 1211–1216.

[4] R. Adams, *Sobolev spaces*, Academic Press, New York, 1975.

[5] V. I. Agoshkov, *Poincaré-Steklov operators and domain decomposition methods in finite dimensional spaces*, in First International Symposium on Domain Decomposition Methods for Partial Differential Equations, R. Glowinski - G. H. Golub - G. A. Meurant - J. Périaux, eds., Philadelphia, PA, 1988, SIAM.

[6] A. Alonso - R. L. Trotta - A. Valli, *Coercive domain decomposition algorithms for advection-diffusion equations and systems*, J. Comput. Appl. Math., 96 (1) (1998), pp. 51–76.

[7] A. Alonso - A. Valli, *A new approach to the coupling of viscous and inviscid stokes equations.*, East-West J. Numer. Math., 3 (1995), pp. 29–41.

[8] A. Alonso - A. Valli, *Some remarks on the characterization of the space of tangential traces of* $H(\mathrm{rot};\Omega)$ *and the construction of an extension operator*, Manuscripta Math., 89 (1996), pp. 159–178.

[9] A. Alonso - A. Valli, *A domain decomposition approach for heterogeneous time-harmonic Maxwell equations*, Comput. Meth. Appl. Mech. Engrg., 143 (1997), pp. 97–112.

[10] A. Alonso - A. Valli, *Finite element approximation of heterogeneous time-harmonic Maxwell equations via a domain decomposition approach*, in International Conference on Differential Equations (Lisboa, 1995), River Edge, NJ,, 1998, World Sci. Publishing, pp. 227–232.

[11] A. Alonso - A. Valli, *Unique solvability for high-frequency heterogeneous time-harmonic Maxwell equations via the fredholm alternative theory*, Math. Methods Appl. Sci., 21 (1998), pp. 463–477.

[12] A. Alonso - A. Valli, *An optimal domain decomposition preconditioner for low-frequency time-harmonic Maxwell equations*, Math. Comp., 68 (1999), pp. 607–631.

[13] A. Auge - A. Kapurkin - G. Lube - F. C. Otto, *A note on domain decomposition of singularly perturbed elliptic problems*, in Domain Decomposition Methods in Sciences and Engineering, P. E. Bjørstad - M. Espedal - D. Keyes, eds., John Wiley & Sons, 1997. Proceedings from the Ninth International Conference, June 1996, Bergen, Norway.

[14] C. Bègue - C. Conca - F. Murat - O. Pironneau, *Les equations de Stokes et de Navier-Stokes avec des conditions aux limites sur la pression*, in Nonlinear partial differential equations and their applications. Collège de France seminar, Vol. IX, H. Brezis - J.-L. Lions, eds., Longman, Harlow, 1988, pp. 179–264.

[15] F. Ben Belgacem - A. Buffa - Y. Maday, *The mortar method for the Maxwell's equations in 3d*, C. R. Acad. Sci. Paris Sr. I Math., 329 (1999), pp. 903–908.

[16] J.-D. Benamou - B. Després, *A domain decomposition method for the Helmholtz equation and related optimal control problems*, J. of Comp. Physics, 136 (1997), pp. 68–82.

[17] L. Bers - F. John - M. Schechter, *Partial Differential Equations*, AMS Lectures in Applied Mathematics, Vol. 3a, Providence, 1964.

[18] L. C. Berselli - F. Saleri, *New substructuring domain decomposition methods for advection-diffusion equations*, J. Comput. Appl. Math., 116 (2000), pp. 201–220.

[19] J.-F. Bourgat - R. Glowinski - P. Le Tallec - M. Vidrascu, *Variational formulation and algorithm for trace operator in domain decomposition calculations*, in Domain Decomposition Methods, T. Chan - R. Glowinski - J. Périaux - O. Widlund, eds., Philadelphia, PA, 1989, SIAM, pp. 3–16.

[20] J. H. Bramble - J. E. Pasciak - A. H. Schatz, *The construction of preconditioners for elliptic problems by substructuring, I*, Math. Comp., 47 (1986), pp. 103–134.

[21] S. C. Brenner - L. R. Scott, *The Mathematical Theory of Finite Element Method*, Springer-Verlag, New York, 1994.

[22] H. Brezis, *Analyse fonctionnelle*, Masson, Paris, 1983.

[23] F. Brezzi - G. Gilardi, *Functional Analysis and Functional Spaces*, in Finite Element Handbook, Chapters 1-2, H. Kardestuncer, ed., McGraw-Hill, New York, 1987.

[24] A. Buffa - Y. Maday - F. Rapetti, *A sliding mesh-mortar method for a two dimensional eddy currents model of electric engines*, M2AN Math. Model. Numer. Anal., 35 (2001), pp. 191–228.

[25] X.-C. Cai, *An additive Schwarz algorithm for nonselfadjoint elliptic equations*, in Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, T. Chan - R. Glowinski - J. Périaux - O. Widlund, eds., SIAM, Philadelphia, PA, 1990, pp. 232–244.

[26] X.-C. Cai - O. B. Widlund, *Domain decomposition algorithms for indefinite elliptic problems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 243–258.

[27] X.-C. Cai - O. B. Widlund, *Multiplicative Schwartz algorithms for some nonsymmetric and indefinite problems*, SIAM J. Num. Anal., 30(4) (1993), pp. 936–952.

[28] C. Carlenzoli - A. Quarteroni, *Adaptive domain decomposition methods for advection-diffusion problems*, The IMA Volumes in Mathematics and its Applications, Springer Verlag, 75 (1995), pp. 165–186.

[29] L. M. Carvalho - L. Giraud - P. Le Tallec, *Algebraic two-level preconditioners for the schur complement method*, SIAM J. Scientific Computing, 22 (2001), pp. 1987–2005.

[30] P. Chevalier, *Méthodes numériques pour les tubes hyperfréquences. Résolution par décomposition de domaine*, PhD thesis, Université Paris VI, 1998.

[31] A. J. Chorin - J. E. Mardsen, *A Mathematical Introduction to Fluid Mechanics*, Springer-Verlag, New York, 1993.

[32] P.-G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[33] P. Collino - G. Delbue - P. Joly - A. Piacentini, *A new interface condition in the non-overlapping domain decomposition for the Maxwell equations*, Comput. Methods Appl. Mech. Engrg., 148 (1997), pp. 195–207.

[34] L. C. Cowsar - J. Mandel - M. F. Wheeler, *Balancing domain decomposition for mixed finite elements*, Math. Comp., 64 (1995), pp. 989–1015.

[35] B. Desprès, *Décomposition de domaine et problème de Helmholtz*, C.R. Acad. Sci. Paris, 1 (1990), pp. 313–316.

[36] B. Després, *Domain decomposition method and the Helmholtz problem*, in First International Conference on Mathematical and Numerical Aspects of Wave Propagation (Strasbourg, 1991), Philadelphia, PA, 1991, SIAM, pp. 44–52.

[37] B. Després, *Domain decomposition method and the Helmholtz problem.II*, in Second International Conference on Mathematical and Numerical Aspects of Wave Propagation (Newark, DE, 1993), Philadelphia, PA, 1993, SIAM, pp. 197–206.

[38] B. Després - P. Joly - J. E. Roberts, *A domain decomposition method for the harmonic Maxwell equations*, in Iterative methods in linear algebra (Brussels, 1991), Amsterdam, 1992, North-Holland, pp. 475–484.

[39] V. Dolean, *Algorithmes par décomposition de domaine et accélération multigrille pour le calcul d'écoulements compressibles*, PhD thesis, Université de Nice-Sophia Antipolis, 2001.

[40] V. Dolean - S. Lanteri - F. Nataf, *Convergence Analysis of a Schwarz Type Domain Decomposition Method for the Solution of the Euler Equations*, Rapp. Rech. INRIA, RR-3916 (2000).

[41] M. Dryja - O. B. Widlund, *An additive variant of the Schwarz alternating method for the case of many subregions*, Tech. Rep. 339, also Ultracomputer Note 131, Department of Computer Science, Courant Institute, 1987.

[42] M. Dryja - O. B. Widlund, *Additive Schwarz methods for elliptic finite element problems in three dimensions*, in Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations, D. E. Keyes - T. F. Chan - G. A. Meurant - J. S. Scroggs - R. G. Voigt, eds., Philadelphia, PA, 1992, SIAM, pp. 3–18.

[43] L. C. Evans, *Partial Differential Equations*, AMS, Providence, 1998.

[44] R. S. Falk - G. R. Richter, *Local Error Estimates for a Finite Element Method for Hyperbolic and Convection-Diffusion Equations*, SIAM J. Num. Anal., 29 (1992), pp. 730–754.

[45] R. S. Falk - G. R. Richter, *Explicit Finite Element Methods for Symmetric Hyperbolic Equations*, SIAM J. Num. Anal., 36 (1999), pp. 935–952.

[46] C. Farhat - J. Mandel - F.-X. Roux, *Optimal convergence properties of the FETI domain decomposition method*, Comput. Methods Appl. Mech. Engrg., 115 (1994), pp. 367–388.

[47] C. Farhat - F.-X. Roux, *A Method of Finite Element Tearing and Interconnecting and its Parallel Solution Algorithm*, Int. J. Numer. Meth. Engng., 32 (1991), pp. 1205–1227.

[48] M. J. Gander, *Optimized Schwarz methods for Helmholtz problems*, in Proceedings of the 13th international conference on domain decomposition, 2001. submitted.

[49] M. J. Gander - L. Halpern - F. Nataf, *Optimal convergence for overlapping and non-overlapping Schwarz waveform relaxation*, in Eleventh international Conference of Domain Decomposition Methods, C.-H. Lai - P. Bjørstad - M. Cross - O. Widlund, eds., ddm.org, 1999.

[50] M. J. Gander - L. Halpern - F. Nataf, *Optimized Schwarz methods*, in 12th international conference on domain decomposition methods, 2000.

[51] M. J. Gander - F. Magoulès - F. Nataf, *Optimized schwarz methods without overlap for the helmholtz equation*, SIAM J. Numer. Anal., (2001). to appear.

[52] M. Garbey, *A schwarz alternating procedure for singular perturbation problems*, SIAM J. Sci. Comput., 17 (1996), pp. 1175–1201.

[53] F. Gastaldi - L. Gastaldi, *On a domain decomposition for the transport equation: theory and finite element approximation*, IMA J. Num. Anal., 14 (1993), pp. 111–135.

[54] F. Gastaldi - L. Gastaldi, *Convergence of Subdomain Iteration for the Transport Equation*, Boll. U.M.I., 7 (1995), pp. 175–202.

[55] F. Gastaldi - L. Gastaldi - A. Quarteroni, *Adaptive domain decomposition methods for advection dominated equations*, East-West J. Numer. Math., 4 (1996), pp. 165–206.

[56] L. Gastaldi, *A domain decomposition method associated with streamline diffusion FEM for linear hyperbolic systems*, App. Num. Math., 10 (1992), pp. 357–380.

[57] L. Gerardo Giorda - P. Le Tallec - F. Nataf, *A Robin-Robin preconditioner for advection-diffusion equations with discontinuous coefficients*, Tech. Rep. 492, CMAP (Ecole Polytechnique), 2002. submitted.

[58] L. Gerardo Giorda - P. Le Tallec - F. Nataf, *A Robin-Robin preconditioner for strongly heterogenous advection-diffusion equations*, in 14th International Conference on Domain Decomposition Methods, DDM.org, 2002.

[59] V. Girault - P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, Berlin, 1995.

[60] E. Godlewski - P. A. Raviart, *Hyperbolic Systems of Conservation Laws*, Springer-Verlag, New York, 1994.

[61] E. Godlewski - P. A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer-Verlag, New York, 1996.

[62] C. Hirsch, *Numerical Computation of Internal and External Flows*, J. Wiley & Sons, Chichester, 1990.

[63] C. Japhet, *Conditions aux limites artificielles et décomposition de domaine: Méthode oo2 (optimisé d'ordre 2). application à la résolution de problèmes en mécanique des fluides*, Tech. Rep. 373, CMAP (Ecole Polytechnique), 1997.

[64] C. Japhet - F. Nataf - F. Rogier, *The optimized order 2 method. application to convection-diffusion problems*, Future Generation Computer Systems FUTURE, 18 (2001).

[65] C. Japhet - F. Nataf - F.-X. Roux, *The Optimized Order 2 Method with a coarse grid preconditioner. application to convection-diffusion problems*, in Ninth International Conference on Domain Decompositon Mehods in Science and Engineering, P. Bjorstad - M. Espedal - D. Keyes, eds., John Wiley & Sons, 1998, pp. 382–389.

[66] A. Jeffrey, *Quasilinear Hyperbolic Systems and Waves*, Pitman Publ., London, 1976.

[67] C. Johnson, *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Cambridge University Press, Cambridge, 1987.

[68] C. Johnson - U. Nävert - J. Pitkäranta, *Finite Element Methods for Linear Hyperbolic Problems*, Comp. Met. App. Mech. Eng., 45 (1984), pp. 285–312.

[69] C. Johnson - J. Pitkäranta, *An Analysis of the Discontinuous Galerkin Method for a Scalar Hyperbolic Equation*, Math. Comp., 46 (1986), pp. 1–26.

[70] H. O. Kreiss, *Initial Boundary Value Problems for Hyperbolic Systems*, Comm. Pure Appl. Math., 23 (1970), pp. 277–298.

[71] L. D. Landau - E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, Oxford, 1959.

[72] P. D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, SIAM, Philadelphia, 1973.

[73] P. Le Tallec, *Domain decomposition methods in computational mechanics*, in Computational Mechanics Advances, J. T. Oden, ed., vol. 1 (2), North-Holland, 1994, pp. 121–220.

[74] P. Le Tallec - Y.-H. De Roeck - M. Vidrascu, *Domain-decomposition methods for large linearly elliptic three dimensional problems*, J. of Computational and Applied Mathematics, 34 (1991).

[75] P. Le Tallec - M. Vidrascu, *Generalized Neumann-Neumann preconditioners for iterative substructuring*, in Domain Decomposition Methods in Sciences and Engineering, P. E. Bjørstad - M. Espedal - D. Keyes, eds., John Wiley & Sons, 1997. Proceedings from the Ninth International Conference, June 1996, Bergen, Norway.

[76] J.-L. Lions - E. Magenes, *Non-homogeneous Boundaryu Value Problems and Applications, 1*, Springer-Verlag, Berlin, 1972.

[77] P.-L. Lions, *On the Schwarz alternating method. I.*, in First International Symposium on Domain Decomposition Methods for Partial Differential Equations, R. Glowinski - G. H. Golub - G. A. Meurant - J. Périaux, eds., Philadelphia, PA, 1988, SIAM, pp. 1–42.

[78] P.-L. Lions, *On the Schwarz alternating method. II.*, in Domain Decomposition Methods, T. Chan - R. Glowinski - J. Périaux - O. Widlund, eds., Philadelphia, PA, 1989, SIAM, pp. 47–70.

[79] P.-L. Lions, *On the Schwarz alternating method. III: a variant for nonoverlapping subdomains*, in Third International Symposium on Domain Decomposition Methods for Partial Differential Equations , held in Houston, Texas, March 20-22, 1989, T. F. Chan - R. Glowinski - J. Périaux - O. Widlund, eds., Philadelphia, PA, 1990, SIAM.

[80] Y. Maday - F. Rapetti - B. Wohlmuth, *Coupling between scalar and vector potentials by the mortar element method.*, C. R. Math. Acad. Sci. Paris, 334 (2002), pp. 933–938.

[81] F. Nataf - F. Rogier, *Factorization of the convection-diffusion operator and the Schwarz algorithm*, M^3AS, 5 (1995), pp. 67–93.

[82] J.-C. Nédélec, *Mixed finite elements in* $\mathbb{R}^3$, Numer. Math., 35 (1980), pp. 315–341.

[83] J.-C. Nédélec, *Acoustic and Electromagnetic Equations*, Springer, New York, 2001.

[84] J. Oliger - A. Sundström, *Theoretical and Practical Aspects of some Initial Boundary value Problems in Fluid Dynamics*, SIAM J. Appl. Math., 35 (1978), pp. 419–446.

[85] L. F. Pavarino - A. Toselli, *Recent Developments in Domain Decomposition Methods*, vol. 23 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, 2002.

[86] A. Quarteroni, *Domain decomposition methods for systems of conservation laws: spectral collocation approximation.*, SIAM J. Sci. Stat. Comput., 11 (1990), pp. 1029–1052.

[87] A. Quarteroni - F. Gastaldi - G. S. Landriani, *On the Coupling of Two Dimensional Hyperbolic and Elliptic Equations: Analytical and Numerical Approach*, Proc. of the 3rd Int. Symp. on Domain Decompositon Methods for Partial Differential Equations, (1990), pp. 22–63.

[88] A. Quarteroni - A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer-Verlag, Berlin, 1994.

[89] A. Quarteroni - A. Valli, *Domain Decompostion Methods for Partial Differential Equations*, Oxford University Press, 1999.

[90] F. Rapetti - A. Toselli, *A FETI preconditioner for two-dimensional edge element approximations of Maxwell's equations on nonmatching grids*, SIAM J. Sci. Comput., 23 (2001), pp. 92–108.

[91] B. L. Roždestvenskiĭ - N. N. Janenko, *Systems of Quasilinear Equations and Their Application to Gas Dynamics*, AMS Translation of Mathematical Monographs, 1983.

[92] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14(2) (1993), pp. 461–469.

[93] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS, Boston, 1996.

[94] Y. Saad - H. Schultz, *GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7(3) (1986), pp. 856–869.

[95] B. F. Smith - P. E. Bjørstad - W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, 1996.

[96] J. Smoller, *Shock Waves and Reaction-Diffusion Equations*, Springer-Verlag, New York, 1983.

[97] R. Trotta, *Multidomain finite elements for advection-diffusion equations*, Appl. Numer. Math., 21 (1996), pp. 91–118.

[98] O. B. Widlund, *Some Schwarz methods for symmetric and nonsymmetric elliptic problems*, in Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations, D. E. Keyes - T. F. Chan - G. A. Meurant - J. S. Scroggs - R. G. Voigt, eds., Philadelphia, PA, 1992, SIAM, pp. 19–36.

[99] F. Willien - I. Faille - F. Nataf - F. Schneider, *Domain decomposition methods for fluid flow in porous medium*, in 6th European Conference on the Mathematics of Oil Recovery, September 1998.

[100] K. Yosida, *Functional Analysis*, Springer-Verlag, Berlin, 1973.