Subgrid scale stabilized finite elements for low speed flows

Javier Principe

Advisor: **Ramon Codina**

A Laura, a mamá y a Martin

Acknowledgments

You enter his office and, once again, make a silly question. He patiently picks a piece of paper and starts, once again, explaining you the thing. This disposition goes in hand with the freedom you need to try things, to get wrong (many times) and to finally do it by yourself. There is nothing to say but...*merci Ramon!*

After this guide, the next important thing I had to develop this work is the environment, meaning the RMEE department of the Universitat Politècnica de Catalunuya and CIMNE. I would like to thank all the people that are part of this, starting from Eugenio Oñate, for giving me the chance of working here. Although it is impossible to give a complete list, I would like to particularly thank the every day work with Guillaume Houzeaux, Santi Badia, Oriol Guasch, Herbert Coppola and Noel Hernandez (lately also with Matias Avila, Christian Muñoz and Joan Bages) and the collaboration with Florian Henke; the conversations with the people of second floor, Romain Aubry, Facundo del Pin and Monica de Mier, Ricardo Rossi and Pooyan Dadvan and also with Carlos Labra and with Mariano Vazquez. Thanks to the CIMNE Terrassa people, Jordi and Dani. Thanks to my present and past office mates Roberto, Jeovan, Pablo and Maritzabel. A special thank to Gerardo Valdez for his sense of humor and for his help with the format of this thesis.

My previous work at the Center for Industrial Research (CINI) gave me the chance of being in contact with real applications of numerical methods, what motivated me to begin doctoral studies. I would like to thank the people I worked with during these years, in particular Marcela Goldschmit and Eduardo Dvorkin.

I would like to particularly thank prof. Bernard Schrefler and Francesco Pesavento for involving me in the joint project "Numerical modelling of smoke dynamics generated by fires within tunnels", which is the frame in which this thesis has been developed. Likewise, the financial support to this project provided by International Centre For Mechanical Sciences (CISM) and Autostrada del Brennero S.p.A. is also acknowledged. This project has been (and still is) a broad source of motivating challenges.

The financial support received from the Agència de Gestió d'Ajuts Universitaris i de Recerca of the *Generalitat de Catalunya* (Catalan Government) and the *European Social Fund* through a doctoral grant is acknowledged.

Finally, I would like to thank my friends for these shared years here in Barcelona,

specially Ana, el chute, Mateo, Mano, Pablo, Dolo and Maria. My old friends, those I made when I was a kid, Alequis, Chapa, Diego, Facu, Flor, Salva and Xime, make me a lucky guy. I cannot finish without specially thank the support given by my family, my mother and my brother. And I couldn't have done it without my love, Laura.

Abstract

A general description of a fluid flow is given by the compressible Navier-Stokes equations, a very complex problem whose mathematical structure is not well understood. Therefore, simplified models are derived by asymptotic analysis, under some assumptions made in terms of dimensionless parameters that measure the relative importance of different physical processes. Low speed flows can be described by the incompressible Navier Stokes equations whose mathematical structure is much better understood. However, many important flows cannot be considered as incompressible, even at low speed, due to the presence of thermal effects. In these cases another class of simplified equations can be derived: the Boussinesq equations and the Low Mach number equations.

The complexity of these problems makes their numerical solution very difficult as the standard finite element method is unstable. In the incompressible Navier Stokes equations, two well known sources of numerical instabilities are the incompressibility constraint and the presence of the convective term. Many stabilization techniques used nowadays are based on scale separation, splitting the unknown into a coarse part induced by the discretization of the domain and a fine subgrid part. The modelling of the subgrid scale and its influence leads to a modified coarse scale problem providing stability.

Although stabilization techniques are nowadays widely used, important problems remain open. Contributing to their understanding, several aspects of the subgrid scale modelling are analyzed in this work. For second order scalar problems, the dependence of the subgrid scale on the mesh size, in the general anisotropic case, is clarified. These ideas are extended to systems of equations to consider the Oseen problem. The modelling of the subgrid scales in transient problems is also analyzed, leading to an improved time discretization scheme for the coarse scale problem. To consider low speed flow models, the extension of these techniques to nonlinear and coupled problems is presented, something that is intimately related to the problem of turbulence modelling, which a entire subject on its own right.

Thermally coupled flow problems are important from an engineering point of view. An accurate solution of a flow problem is needed to define thermal loads on structures which, in many cases have a strong response, making the problem coupled. This kind of problems, that motivated this work, include the problem of a structural response in the case of fires.

Contents

Chapter 1

Introduction

The general description of a fluid flow involves the solution of the compressible Navier Stokes equations. It is widely accepted that these equations provide an accurate description of any problem in fluid mechanics. This set of equations, the mathematical formulation of the physical principles of mass, momentum and energy conservation coupled with a state equation, is very complex and very little is known about its mathematical structure. Results on the boundary conditions that make the problem well posed, on the existence of a solution and on uniqueness can be found in [104]. The mathematical complexity of the problem is the manifestation of the also complex physical behavior of these flows. Many different non linear physical mechanisms are coupled in fluid mechanics problems. For these reasons, depending on the physics of the problem under consideration, different models can be derived from the compressible Navier Stokes equations [104, 148]. The derivation of these reduced sets of equations is based on some assumptions on the problem, usually made in terms of some dimensionless parameters that measure the relative importance of different physical processes, like the Mach or Reynolds numbers. The most important of these models is described by the incompressible Navier Stokes equations. This set of equations is smaller than the compressible one and its mathematical structure is much better understood. Furthermore, two physical effects that are difficult to predict, shock waves and sound waves, are not found in incompressible problems. However many important flows cannot be considered as incompressible due to the presence of thermal effects. In such kind of problems another class of simplified equations can be derived: the Boussinesq equations and the Low Mach number equations.

The complexity of the mathematical problems found in fluid mechanics makes their numerical solution very difficult. Special techniques are needed because when the standard Galerkin method is used, numerical instabilities appear. The nature of these instabilities depends on the problem under consideration but the manifestation is usually a solution that presents node to node oscillations of numerical (non physical) nature. In the incompressible case, two well known sources of numerical instabilities are the incompressibility constraint and the presence of the convective terms. The convective

instability is also present in the convection diffusion reaction problem (CDR) and was early understood as a lack of diffusion of the discrete problem. The first attempts to remedy the situation consisted in the addition of an extra stabilizing term of diffusive type and were called artificial viscosity methods. These methods are not consistent, i.e. the exact solution of the continuous problem does not satisfy the perturbed equation, what results in a loss of accuracy. The first consistent method, the streamline upwind Petrov Galerkin method (SUPG), was developed in the late seventies [76, 93, 18]. This method and many of its successors consist then in the addition of a stabilizing term to the original Galerkin formulation which is proportional to the residual and we refer to [24] for a comparison of different methods of this type.

The incompressibility constraint gives rise to a instability of the pressure and can be also found in the Stokes problem. The standard Galerkin method applied to solve this problem is stable provided the Ladyzhenskaya-Babuska-Brezzi (LBB) condition is satisfied, which requires a compatibility of the spaces where the velocity and pressure belong. It is satisfied in the case of the continuous problem but it may not hold in the discrete case depending on the interpolation used. In particular, equal order interpolations do not satisfy this condition. It is important to mention that the compressible Navier Stokes equations, as well as the simplified equations derived from them , can be written as a system of second order convection diffusion reaction (CDR) equations and that the pressure gradient and incompressibility appear in the first order convective term. This observation was exploited in [79] to apply a technique similar to SUPG to obtain a stabilized formulation allowing the use of equal order interpolations.

The way of understanding these methods has changed since the introduction of the variational multiscale method (VMM) in [75, 78]. This method is based on the split of the unknown into a coarse scale resolvable part and a fine scale subgrid part. This split corresponds to a decomposition of the space in which the solution of the problem is sought as a direct sum of a coarse scale space and a fine scale one. The coarse scale space is the one induced by the discretization of the domain and the fine scale space is any complement to yield the continuous space. In this way, the problem is decomposed into a resolvable coarse scale problem induced by the discretization and a small scale problem that cannot be exactly solved because it is as complex as the original continuous problem. The subgrid scale problem is approximately solved and the influence of the subgrid scale on the coarse scale problem is approximately taken into account. The final result is a modified discrete problem that now can be shown to be stable. This technique has been extended to incompressible Navier Stokes equations (see for example [28]) and has been used to solve many different kind of problems. Its extension to general CDR systems has been analyzed in [25] where it is shown that the natural extension cannot be performed in general. In particular a general expression for the stabilization parameters is still unknown. This fact implies that a stabilized finite element formulation needs to

be developed for each set of equations separately.

Although stabilization techniques are nowadays widely used, there are important questions that have not been answered. In the first place we have the definition of the stabilization parameters. We know how these parameters depend on the equation coefficients but they also depend on some measure of the mesh size, whose precise definition is open, and on some constants, whose values are known from numerical experiments only. Then, we have the question of how these parameters depend on the size of the time discretization what, in fact, gives rise to the question of how to extend the stabilization techniques to consider transient problems. When this formulation is applied to the incompressible Navier Stokes equations, apart from the definition of the stabilization parameters for this system we also face the problem of extending stabilization techniques to nonlinear problems. The answer to these questions is implicit in the subgrid scale model finally used. In particular, the subgrid scale modelling in the case of nonlinear problems is intimately related to the problem of turbulence modelling which is an entire subject on its own right.

After a discrete formulation of the problem considered has been defined, a discrete algebraic problem needs to be solved. Apart from the potential numerical instabilities, another manifestation of the complexity of the problems considered is the highly nonlinear nature of the associated discrete system. Therefore the numerical solution also requires a proper linearization strategy which can be written, in general, as a fixed point scheme. Several possibilities can be considered, from fully coupled Newton type to segregated Picard type linearization schemes.

Thermally coupled flow problems, despite the intrinsic interest they deserve, are important from an engineering point of view. Many structural problems, for example, involve the solution of a flow problem to define the loads. It is also common to have a strong response from the structure, what makes the problem coupled. This is the kind of problems we have in mind for the application of the developed model. We are specially concerned with the problem of a thermal load on a structure due to a fire which is an example of strongly thermally coupled flow that will be described in the following section.

1.1 An application problem: fire in tunnels

The results of these work will be applied to the problem of a fire in a tunnel. This problem is of great interest, particularly in the European Union, due to the recent fires occurred in European tunnels in the last years [131]. As fires cause loss of human beings and have a strong impact in economy due to the high reparation costs, big efforts are carried out to understand this problem.

The problem is geometrically and physically very complex and its numerical solution challenging. An important aspect that needs to be determined is the structural response

of the tunnel construction. As mentioned in [131], structural damage can be attributed to two main factors: spalling of concrete and excessive temperatures attained in the concrete and steel components. In turn the temperature field inside the concrete depends on the temperature on the walls of the tunnel which is a result of a fluid mechanics problem. Actually both problems are coupled trough the boundary conditions on the walls. A complete model to predict the concrete behavior of the structure is presented in [131].

In order to model the flow problem, several physical phenomena should be taken into account. The problem of simulating flow dynamics due to a fire can be stated as the dynamics of several fluids into a domain Ω with a chemical interaction between them. That chemical reaction, which transforms elements, defines the type of fire. In the case of fire in tunnels, the detailed mechanism is unknown except in some experiments specially designed. Therefore, the model should include the appropriate balances of mass, momentum and energy and a combustion model that define the species to be considered and the relation between them. It is to be noted that the mass balance should be performed for all the species, what is usually done through the inclusion of transport equations for the species concentration. Due to the high temperatures attained, the model for the heat flux on the flow should contain a radiative mechanism apart from the usual convective one. The radiation properties of the medium may depend, of course, on the species concentration. Different approaches to fire modelling can be found in [108] and the references therein.

The simplest combustion model that can be considered is the volumetric heat source (VHS) model where it is considered that the combustion is a source of heat that does not depend on any species concentration. However, the concentration of the smoke, which is a product of the combustion, needs to be determined as it greatly affects the radiation problem. In [144] a comparison of different combustion models, including the VHS, is presented as well as their performance on the simulation of a room fire, a shopping mall fire and a tunnel fire.

This application problem is a motivation for the objectives posed in this work. The problem of a fire is that of a fluid with strong thermal coupling and the usual Boussinesq approximation cannot be used as the temperature variation can be higher than the mean temperature. The low Mach number model discussed in chapter 2 is much more appropriate. To accurately solve this problem we need to define a robust discrete approximation in order to avoid numerical instabilities. Finally, a good strategy for the solution of the whole thermally coupled fluid-solid problems is needed.

1.2 Objectives and organization

Let us close this introduction describing the organization of the work according to the objectives defined.

The first objective of this work is to understand the derivation of the simplified models that describe low speed flows as well as the relation between them. This will be done in chapter 2, where a unified asymptotic approach is proposed and all these models, whose justification was separately known, are recovered. This approach enables us to go further and, in particular, to predict the range of applicability of each model in terms of the dimensionless parameters already mentioned.

The second and most important objective of this work is to develop a subgrid scale stabilized finite element formulation for the kind of problems we are considering. To achieve this goal we follow a natural way, starting from the scalar convection diffusion equation in chapter 3, where a new definition of the stabilization parameters is presented. Then we extend these results to the incompressible Navier Stokes problem. This extension involves two main aspects, the definition of the stabilization parameters, which is treated in chapter 4, and the extension of the stabilization techniques to transient nonlinear problems, which is teated in chapter 5. Finally we extend these results to thermally coupled flows in chapter 6, where the final discrete formulation is presented.

The third objective is to develop a finite element code to solve this problems. Apart from the discrete formulation of the problems, the final ingredient that we need is an algorithm for the solution of the discrete problem. In chapter 7 different linearization strategies are compared and the final algorithm is presented.

The fourth and last objective of this work is to apply the developed code to the problem of thermal coupling of fluids and solids. To achieve this goal, a coupling strategy based on a domain decomposition approach has been developed. This strategy implies the development of a small code to manage the coupling between the solid and the fluid. This development was applied to the problem of a fire in a tunnel described above. Both, the strategy and the application are described in chapter 8.

We close the work with chapter 9, where conclusions and further possible research lines are summarized. Let us finally mention that chapters are quite self contained even if this implies the need of repeating some information. This is due to the fact that each chapter is based on the following publications:

- Chapter2: "On the low Mach number and the Boussinesq approximations for low speed flows", J. Principe and R. Codina, Submitted.

- Chapter3: "The modelling of subgrid scales in the finite element approximation of convection diffusion reaction problems on anisotropic meshes", J. Principe and R. Codina, In preparation.

- Chapter 4:"The modelling of subgrid scales in the finite element approximation of incompressible flows", J. Principe and R. Codina, In preparation.

- Chapter 5: "Time dependent subscales in the stabilized finite element approximation

of incompressible flow problems", R. Codina, J. Principe, O. Guasch and S. Badia, *Computer Methods in Applied Mechanics and Engineering*, 196 (2007), 2413-2430.

- Chapter 6: "Dynamic subscales in the finite element approximation of thermally coupled incompressible flows", R. Codina and J. Principe, *International Journal for Numerical Methods in Fluids*, 54 (2007), 707-730.

- Chapter 7: "A stabilized finite element approximation of low speed thermally coupled flows", J. Principe and R. Codina, *International Journal of Numerical Methods for Heat & Fluid Flow*, Accepted.

- Chapter 8: "A numerical approximation of the thermal coupling of fluids and solids", J. Principe and R. Codina, Submitted.

Chapter 2

Physical problems

In this chapter we present an asymptotic analysis of the compressible Navier Stokes equations at low speeds. Compressible flows at low speeds behave as incompressible in a sense that we make precise. In the absence of heat exchange (the isentropic regime) this limit is well understood and rigorous results are available. When heat exchange is considered, different simplified models can be obtained. These models have been used during the years for different applications (usually on different academic environments) the most famous being the Boussinesq approximation. Here a unified formal justification of these models, based on an asymptotic analysis, is presented. Special attention is paid to the relation between the low Mach number and the Boussinesq approximations.

2.1 Introduction

Many flows of interest can be considered as incompressible. This assumption is useful as it makes the problem much simpler than if a full compressible flow is considered. The compressible flow equations have different structure depending on the Mach number. If the Mach number is of the order of or greater than one, shock waves may be present. A number of issues have to be considered when numerically solving compressible flows, such as the set of variables to be used and the prediction of such shock waves. In the incompressible case, the system of equations is smaller and shocks as well as sound waves are absent. Furthermore, the mathematical structure of incompressible equations is much better understood than the general one. For ideal fluids, in the absence of heat sources (the isentropic case), solutions of the incompressible Navier Stokes equations can be found as the limit of solutions of the compressible ones as the Mach number tends to zero under certain assumptions on the initial data. Rigorous mathematical results were established in [96] (see also [105]).

When heat exchange is taken into account, the limit is quite different, since the energy equation is not uncoupled and one needs to keep the state equation to close the system.

The zero Mach number limit gives rise to a splitting of the pressure into a constant-in-space thermodynamic pressure p^{th} and a mechanical pressure p that has to be used in the momentum equation. This leads to a removal of the acoustic modes and the flow behaves as incompressible, in the sense that the mechanical pressure is determined by the mass conservation equation and not by the state equation. However, large variations of density due to temperature variations are allowed. This limit has been studied first in [126] in the inviscid case, and generalized to the viscous case in [119]. A rigorous derivation including combustion was presented in [106]. This zero Mach number model has also been presented in [47] and [145]. The numerical implications of this limit have been studied in [97] and [113], for example.

However the most widely used model in the context of thermally coupled flows is the so called Boussinesq approximation. In 1903, based on his observations on the behavior of thermal flows, J. Boussinesq [14] proposed to ignore the variations of density except where they multiply the gravity acceleration (historical issues can be found in [149]). Since that moment, many authors have looked for a formal justification of the Boussinesq approximation. In [135] the Boussinesq approximation is found expressing the thermodynamic variables as a constant and an static part plus a fluctuating part resulting from the motion. It is showed that for a thin layer of fluid (compared to the scale of variation of the static fields), the Boussinesq approximation follows. However density variations are retained in the momentum equation even when they are of higher order based on physical arguments and not on a limiting process. The first attempt to present a rigorous derivation of the Boussinesq approximation was performed in [110],where an expansion in two parameters, ε_1 and ε_2, of the full compressible equations is proposed. The Boussinesq approximation is found to the lowest order in both ε_1 and ε_2. Several problems of this approach are described in [120]. On the one hand, the two parameters introduced in [110] are of order $\varepsilon_1 \sim 10^{-4}$ and $\varepsilon_2 \sim 10^{-11}$ for typical fluids in a standard Rayleigh Bénard experiment, indicating a second order approximation for ε_1 to have the same order as a first order approximation for ε_2. On the other hand, the starting point of Mihaljan's approach is the compressible equations but using an equation of state that relates temperature to density only. According to [120] this assumption and the selection of the parameters "destroyed the self-consistency of the scheme" making second order approximations meaningless. It is interesting to note some thermodynamic consequences of an state equation of the form $\rho = \rho(\vartheta)$, where ρ is the density and ϑ is the temperature. Although they result from classical thermodynamics, they were only noted in two articles. In [12] it is mentioned that the constant volume specific heat diverges whereas in [7] it is shown that convexity inequalities are violated. The Mihaljan's approach was improved first by Malkus (in an unpublished work mentioned in [120]) and in [120]. The new ingredient was the selection of an appropriate reference state. In [63] a derivation of the Boussinesq equations was presented taking a reference state into account

and allowing temperature and pressure dependent properties. We note that all these works are concerned with natural convection as the velocity is made dimensionless using another variable as scale (viscosity or gravity, for example). An asymptotic justification of the Boussinesq approximation was developed in the works of Zeytounian [146, 148, 149] and Bois [12, 13]. These developments dealt first with polytropic gases (in [146] and [12]) and the main conclusion was that the Mach number is a small parameter in the Boussinesq approximation. An asymptotic derivation of the Boussinesq approximation for liquids was then presented in [147]. Finally a unified approach for liquids and gases was presented in [13]. Another widely used model, the anelastic approximation, was proposed in [5] and [115] and has been used for a long time in the context of atmospheric flows (see also [44] and [61]). This approximation removes the height limitation present in the Boussinesq's one.

The formal justification of these models has been developed but the connection between them has not been fully analyzed. Although the Boussinesq approximation was considered in [126] and [119] it was not found following the same asymptotic procedure used to derive the low Mach number model. In this work we present the zero Mach number model, the anelastic and the Boussinesq approximations, the density dependent incompressible Navier Stokes equations and the usual incompressible equations in a unified asymptotic setting. As a consequence we show that the Boussinesq and the anelastic approximations are found in the limit of small Mach and small Froude numbers with some restrictions. Particular attention is paid to way in which the asymptotic justification of the Boussinesq approximation is related to that of the other models.

2.2 Equations of motion

The flow of a compressible fluid in a domain Ω is described in terms of the velocity ($\boldsymbol{u}$), pressure (p), density (ρ), and temperature (ϑ) fields (bold characters are used to denote vectors and tensors). These fields are solutions of the equations that describe the dynamics of the system and that are statements of conservation of mass, momentum and energy and a state equation relating the thermodynamic variables. They can be found, for example, in [6] and [104] and can be written as

$$\begin{aligned} \frac{D\rho}{Dt} + \rho \boldsymbol{\nabla} \cdot \boldsymbol{u} &= 0 \\ \rho \frac{D\boldsymbol{u}}{Dt} + \boldsymbol{\nabla} p &= \boldsymbol{\nabla} \cdot \left(2\mu \boldsymbol{\varepsilon}\prime(\boldsymbol{u})\right) + \rho \boldsymbol{g} \\ \rho c_p \frac{D\vartheta}{Dt} - \beta \vartheta \frac{Dp}{Dt} &= \boldsymbol{\nabla} \cdot (k \boldsymbol{\nabla} \vartheta) + \Phi + Q \\ \rho &= F\left(p, \vartheta\right) \end{aligned} \tag{2.1}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla}$ is the material derivative, $\boldsymbol{g}$ the external source of momentum, Q the external source of energy, $\boldsymbol{\varepsilon}\prime(\boldsymbol{u}) = \boldsymbol{\varepsilon} - \frac{1}{3}(\boldsymbol{\nabla} \cdot \boldsymbol{u})\boldsymbol{I}$ the deviatoric part of the rate of deformation tensor ($\boldsymbol{\varepsilon}$ is the symmetric part of the velocity gradient, $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}\left(\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}\boldsymbol{u}^t\right)$), μ the viscosity, c_p the constant pressure specific heat, k the thermal conductivity, β the thermal expansion coefficient and Φ (the Rayleigh dissipation function) is a non-negative contribution due to mechanical dissipation of energy in sheared motion defined as

$$\Phi = 2\mu\varepsilon\prime(\boldsymbol{u}) : \varepsilon\prime(\boldsymbol{u})$$

When an isentropic flow is considered a state equation of the form $\rho = F(p)$ can be assumed (where F could depend on the initial distribution of entropy) and the equations to be solved simplify to

$$\begin{aligned} \frac{D\rho}{Dt} + \rho\boldsymbol{\nabla} \cdot \boldsymbol{u} &= 0 \\ \rho\frac{D\boldsymbol{u}}{Dt} + \boldsymbol{\nabla}p &= \boldsymbol{\nabla} \cdot (2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u})) + \rho\boldsymbol{g} \end{aligned}$$

These equations of motion can be written in dimensionless form in different ways. The process depends on the choice of reference values in a way stated by the π theorem proved in [19]. Having in the system r different units and taking n reference values for the adimensionalization process, the system will have $n - r$ dimensionless numbers defining classes of similar solutions. The system to be solved in a compressible flow is given by (2.1). In this system we have $r = 4$ different units (length, time, mass and temperature). Different choices for reference values have been found in the literature and different non-dimensional numbers result. Our approach is based on taking different scales for each field and for dependent properties. To this end, we introduce the Strouhal, Mach, Reynolds, Péclet, Froude and a heat release rate number, defined as

$$\begin{aligned} \mathrm{S} &= \frac{l_0}{u_0 t_0}, \quad \mathrm{M} = \frac{u_0}{\sqrt{p_0/\rho_0}}, \quad \mathrm{R} = \frac{\rho_0 u_0 l_0}{\mu_0} \\ \mathrm{P} &= \frac{\rho_0 c_{p_0} u_0 l_0}{k_0}, \quad \mathrm{F} = \frac{u_0}{\sqrt{g_0 l_0}}, \quad \mathrm{H} = \frac{t_0 Q_0}{\rho_0 c_{p_0} \vartheta_0}, \quad \varepsilon = \frac{\Delta\vartheta}{\vartheta_0} \end{aligned}$$

where l_0, t_0, ρ_0, p_0, ϑ_0, u_0, μ_0, k_0, c_{p_0}, g_0, Q_0 and $\Delta\vartheta$ are the scales of length, time, density, pressure, temperature, velocity, viscosity, conductivity, specific heat, external acceleration and external heat and temperature variation respectively. The choice of viscosity and conductivity reference values is needed to allow variable physical properties (temperature dependent, for example) whereas the choice of a temperature variation scale is needed to define dimensionless boundary conditions. The dimensionless numbers are defined in terms of 12 parameters but, thanks to the state equation $\rho_0 = F(p_0, \vartheta_0)$, we have 11 reference values, giving rise to the seven dimensionless numbers already defined. We would like to stress that we do not assume the existence of 11 reference scales because,

as it will be shown in the particular cases considered, if a reference scale is not available, its value can be defined eliminating a dimensionless number. For example, if an independent time scale is not available for a particular problem, we can define it from the velocity scale taking $\mathrm{S}=1$.

The dimensionless variables we take (denoted by $\widetilde{\ }$) are

$$\begin{aligned}&\boldsymbol{x}=l_0\widetilde{\boldsymbol{x}},\quad t=t_0\widetilde{t},\quad \rho=\rho_0\widetilde{\rho},\quad p=p_0\widetilde{p},\quad \vartheta=\vartheta_0\widetilde{\vartheta}\\&\boldsymbol{u}=u_0\widetilde{\boldsymbol{u}},\quad \boldsymbol{g}=g_0\widetilde{\boldsymbol{g}},\quad Q=Q_0\widetilde{Q},\quad \mu=\mu_0\widetilde{\mu},\quad k=k_0\widetilde{k},\quad c_p=c_{p_0}\widetilde{c_p}\end{aligned}$$

and the thermal expansion coefficient can be written in dimensionless form using its definition

$$\beta=-\frac{1}{\rho}\left.\frac{\partial\rho}{\partial\vartheta}\right|_p=-\frac{1}{\rho_0\tilde{\rho}}\left.\frac{\partial\rho_0\tilde{\rho}}{\partial\vartheta_0\tilde{\vartheta}}\right|_{\tilde{p}}=-\frac{1}{\vartheta_0}\frac{1}{\tilde{\rho}}\left.\frac{\partial\tilde{\rho}}{\partial\tilde{\vartheta}}\right|_{\tilde{p}}=\frac{1}{\vartheta_0}\tilde{\beta}$$

The dimensionless equations are (omitting $\widetilde{\ }$)

$$\mathrm{S}\frac{\partial\rho}{\partial t}+\boldsymbol{\nabla}\cdot(\rho\boldsymbol{u})=0 \tag{2.2}$$

$$\rho\left(\mathrm{S}\frac{\partial\boldsymbol{u}}{\partial t}+\boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}\right)+\frac{1}{\mathrm{M}^2}\boldsymbol{\nabla}p=\frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot(2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u}))+\frac{1}{\mathrm{F}^2}\rho\boldsymbol{g} \tag{2.3}$$

$$\rho c_p\left(\mathrm{S}\frac{\partial\vartheta}{\partial t}+\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta\right)-\mathrm{S_t}\beta\vartheta\left(\mathrm{S}\frac{\partial p}{\partial t}+\boldsymbol{u}\cdot\boldsymbol{\nabla}p\right)=\frac{\mathrm{M}^2}{\mathrm{R}}\Phi+\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot(k\boldsymbol{\nabla}\vartheta)+\mathrm{HS}Q \tag{2.4}$$

The state equation is made dimensionless using that $\rho_0=F\left(p_0,\vartheta_0\right)$. In the case of an ideal gas it reads

$$p=\rho\vartheta$$

The parameter $\mathrm{S_t}$ depends on the state equation and is defined by

$$\mathrm{S_t}=\frac{p_0}{\rho_0 c_{p_0}\vartheta_0}=\frac{p_0}{F\left(p_0,\vartheta_0\right)c_{p_0}\vartheta_0}$$

which for an ideal gas becomes

$$\mathrm{S_t}=\frac{\gamma-1}{\gamma}$$

Finally, for each particular problem, the boundary conditions have to be written in dimensionless form. For example, the boundary conditions for the momentum equations are of the form

$$\begin{aligned}\boldsymbol{u}&=\boldsymbol{u}_D\quad\text{on}\quad\Gamma_D^{\boldsymbol{u}}\\(-p\boldsymbol{I}+2\mu\varepsilon\prime(\boldsymbol{u}))\cdot\boldsymbol{n}&=\boldsymbol{t}\quad\text{on}\quad\Gamma_N^{\boldsymbol{u}}\end{aligned}$$

where $\Gamma_D^{\boldsymbol{u}}$ ($\Gamma_N^{\boldsymbol{u}}$) is the part of the domain boundary where Dirichlet (Neumann) boundary conditions for the velocity are given, $\Gamma=\partial\Omega=\overline{\Gamma_D^{\boldsymbol{u}}\cup\Gamma_N^{\boldsymbol{u}}}$ is the boundary of the domain

and $\boldsymbol{n}$ its exterior normal. We can introduce the dimensionless form of the data as

$$\boldsymbol{u}_D = u_0 \widetilde{\boldsymbol{u}}_D$$
$$\boldsymbol{t} = p_0 \widetilde{\boldsymbol{t}}$$

and the dimensionless form of the boundary conditions is

$$\widetilde{\boldsymbol{u}} = \widetilde{\boldsymbol{u}}_D \quad \text{on} \quad \Gamma_D^{\boldsymbol{u}} \tag{2.5}$$
$$\left(-\frac{1}{\mathrm{M}^2}\tilde{p}\boldsymbol{I} + \frac{1}{\mathrm{R}} 2\widetilde{\mu}\varepsilon\prime(\widetilde{\boldsymbol{u}})\right) \cdot \boldsymbol{n} = \frac{1}{\mathrm{M}^2}\widetilde{\boldsymbol{t}} \quad \text{on} \quad \Gamma_N^{\boldsymbol{u}} \tag{2.6}$$

In the same way the boundary conditions for the energy equation are of the form

$$\widetilde{\vartheta} = 1 + \varepsilon \widetilde{\vartheta}_D \quad \text{on} \quad \Gamma_D^{\vartheta} \tag{2.7}$$
$$\widetilde{k}\boldsymbol{n} \cdot \nabla \widetilde{\vartheta} = \widetilde{q} \quad \text{on} \quad \Gamma_N^{\vartheta} \tag{2.8}$$

where Γ_D^{ϑ} (Γ_N^{ϑ}) is the part of the domain boundary where Dirichlet (Neumann) boundary conditions for the temperature are given, and $\Gamma = \partial\Omega = \overline{\Gamma_D^{\vartheta} \cup \Gamma_N^{\vartheta}}$. The parameter ε appears when the given function ϑ_D is rescaled to satisfy $0 \leq \widetilde{\vartheta}_D \leq 1$. In order to close the definition of the problem, initial conditions need also to be specified.

Having defined the equations of the motion and rather general boundary conditions, let us consider some particular problems we are interested in that will also help us to illustrate the application of the asymptotic scheme. We are interested in natural convection problems and we consider two examples. The first one is the differentially heated cavity studied in [23] and [102] that consists of a rectangular cavity whose left (hot) wall has a fixed temperature ϑ_h and whose right (cold) wall has a fixed temperature ϑ_c. Upper and lower walls are adiabatic and initially the gas is at rest with a temperature ϑ_0 and density ρ_0. The second one is the well known Rayleigh-Bénard problem (see [99]) which consists in a layer of fluid between two infinite horizontal walls. On the lower wall a higher temperature (ϑ_h) is imposed whereas on the upper one a lower temperature is imposed (ϑ_c) and again, initially the gas is at rest with temperature ϑ_0 and density ρ_0 depending linearly on the vertical coordinate.

However, we want also to consider the case in which a velocity field is prescribed on the boundary and therefore the Poiseuille-Rayleigh-Bénard (PRB) problem (see [114]) is also taken into account. Although several boundary conditions can be applied, we assume a prescribed Poiseuille velocity profile on the inlet and prescribed temperatures on the upper (ϑ_c) and lower walls (ϑ_h). We assume initially a Poiseuille velocity distribution in the whole channel and, as in the Rayleigh Bénard problem, an initial temperature ϑ_0 and density ρ_0 depending linearly on the vertical coordinate.

2.3 The small Mach number limit

The limit when the Mach number tends to zero can be found using standard procedures of asymptotic analysis described for example in [95]. The first step is to expand all flow variables in power series of the small parameter considered

$$\xi\left(\boldsymbol{x},t,\mathrm{M}\right)=\xi^{(0)}\left(\boldsymbol{x},t\right)+\mathrm{M}^{2}\xi^{(2)}\left(\boldsymbol{x},t\right)+\mathcal{O}(\mathrm{M}^{4}) \tag{2.9}$$

for $\xi=\boldsymbol{u}$, $\xi=p$, $\xi=\rho$, $\xi=\vartheta$. The second step is to substitute this expansion into equations 2.2 to 2.4 and to require that all terms in the expanded equations that are multiplied by the same power of M^2 vanish to obtain a hierarchy of equations. The limit is carried out considering that the remaining parameters that appear in the equations are fixed.

This asymptotic setting cannot be used in any situation and in particular we have to mention the problem of the behavior near the initial time. In this case it is necessary to introduce a fast time scale $\tau=t/\mathrm{M}$ and assume an expansion of the form

$$\xi\left(\boldsymbol{x},t,\mathrm{M}\right)=\xi^{(0)}\left(\boldsymbol{x},t,\tau\right)+\mathrm{M}\xi^{(1)}\left(\boldsymbol{x},t,\tau\right)+\mathcal{O}(\mathrm{M}^{2})$$

This is done in [148] and [113], for example. We also have to mention the problem of the behavior of the flow in the far field when unbounded domains are considered. In this case it is necessary to introduce a long space variable $\boldsymbol{\eta}=\mathrm{M}\boldsymbol{x}$ that "looks" to the infinity and to assume an expansion of the form

$$\xi\left(\boldsymbol{x},t,\mathrm{M}\right)=\xi^{(0)}\left(\boldsymbol{x},\boldsymbol{\eta},t\right)+\mathrm{M}\xi^{(1)}\left(\boldsymbol{x},\boldsymbol{\eta},t\right)+\mathcal{O}(\mathrm{M}^{2})$$

This is done in [97] and [109], for example. The objective of these variables is to separate scales and to perform a multiple scale analysis of the problem. The multiple scale analysis of the compressible Navier Stokes equations is of crucial importance to analyze acoustic phenomena. Since we are not interested in the acoustic problem we restrict ourselves to a single scale analysis assuming an asymptotic expansion of the form 2.9. The selection of M^2 as the expansion parameter is due to the fact that this is the parameter that appears in the system of equations, as well as in the boundary conditions (a single scale expansion in terms of M gives the same result).

Any physical property χ (where χ can be μ, k, c_p, or β) can be considered to depend on the temperature and pressure. Using the expansion for the temperature and pressure defined above it follows that

$$\begin{aligned}\chi=\chi\left(\vartheta,p\right)=\chi\left(\vartheta^{(0)},p^{(0)}\right)+\left.\frac{\partial\chi}{\partial\vartheta}\right|_{\left(\vartheta^{(0)},p^{(0)}\right)}\left(\vartheta-\vartheta^{(0)}\right)+\left.\frac{\partial\chi}{\partial p}\right|_{\left(\vartheta^{(0)},p^{(0)}\right)}\left(p-p^{(0)}\right)\\+\mathcal{O}\left(\left(\vartheta-\vartheta^{(0)}\right)^{2},\left(p-p^{(0)}\right)^{2}\right)\end{aligned}$$

Considering that the derivatives of the physical properties are bounded we have

$$\chi\left(\vartheta, p\right) = \chi\left(\vartheta^{(0)}, p^{(0)}\right) + \mathcal{O}(\mathrm{M}^2)$$

The following notation will be used

$$\chi^{(0)} \equiv \chi\left(\vartheta^{(0)}, p^{(0)}\right)$$

To order zero in M^2, the mass conservation equation gives

$$\mathrm{S}\frac{\partial \rho^{(0)}}{\partial t} + \boldsymbol{\nabla}\cdot\left(\rho^{(0)}\boldsymbol{u}^{(0)}\right) = 0$$

The momentum conservation equation gives

$$\begin{aligned}
\mathcal{O}\left(\mathrm{M}^{-2}\right): &\qquad \boldsymbol{\nabla} p^{(0)} = 0 \\
\mathcal{O}\left(1\right): &\qquad \rho^{(0)}\left(\mathrm{S}\frac{\partial u^{(0)}}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}^{(0)}\right) + \boldsymbol{\nabla} p^{(2)} = \frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u}^{(0)})\right) + \frac{1}{\mathrm{F}^2}\rho^{(0)}\boldsymbol{g}
\end{aligned}$$

The first equation implies $p^{(0)} = p^{(0)}(t)$. This is a very important result: the pressure splits into $p^{(0)}$, a reference thermodynamic pressure and $p^{(2)}$ a mechanical pressure. The first one, constant over the whole domain, changes its value only by global heating or mass adding, as will be shown below. The mechanical pressure component $p^{(2)}$ is determined from a velocity constraint playing the same role as in incompressible equations.

The zero order energy equation is

$$\rho^{(0)}c_p^{(0)}\left(\mathrm{S}\frac{\partial \vartheta^{(0)}}{\partial t} + \boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}\right) - \mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\mathrm{S}\frac{dp^{(0)}}{dt} = \frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(0)}\right) + \mathrm{HS}Q$$

In the zero Mach number limit a system of equations for $\rho^{(0)}$, $\vartheta^{(0)}$, $p^{(2)}$ and $\boldsymbol{u}^{(0)}$ has to be solved. The reference pressure $p^{(0)}$, also called thermodynamic pressure, depends on the boundary conditions of the problem. If $\Gamma_N^{\boldsymbol{u}} \neq \emptyset$ the thermodynamic pressure is determined by the boundary condition. This can be seen introducing the asymptotic expansion 2.9 in the dimensionless boundary condition 2.6, from where

$$\begin{aligned}
\mathcal{O}\left(\mathrm{M}^{-2}\right): &\qquad p^{(0)} = \boldsymbol{t}^{(0)}\cdot\boldsymbol{n} \\
\mathcal{O}\left(1\right): &\qquad \left(-p^{(2)}\boldsymbol{I} + \frac{1}{\mathrm{R}}2\mu\varepsilon\prime(\boldsymbol{u}^{(0)})\right)\cdot\boldsymbol{n} = \boldsymbol{t}^{(2)}
\end{aligned}$$

This justifies what was noted in [126]: if the domain is "open" to the atmosphere, the reference pressure is determined by the external pressure. In a "closed" domain ($\Gamma_N^{\boldsymbol{u}} = \emptyset$) the thermodynamic pressure is determined by a global balance. Using the zero order mass and energy conservation equations and the state equation an equation relating the velocity divergence and the thermodynamic pressure can be found. In the case of an ideal gas, this constraint is

$$p^{(0)}\boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)} = -\frac{1}{\gamma}\mathrm{S}\frac{dp^{(0)}}{dt} + \frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(0)}\right) + \mathrm{HS}Q \tag{2.10}$$

This equation, integrated over the domain, gives an ordinary differential equation for the reference pressure. In the case of an ideal gas this equation is explicit and given by

$$p^{(0)} \int_{\partial\Omega} \boldsymbol{u}^{(0)} \cdot \boldsymbol{n} = -\frac{V_\Omega}{\gamma} \mathrm{S} \frac{dp^{(0)}}{dt} + \frac{1}{\mathrm{P}} \int_{\partial\Omega} \boldsymbol{q}^{(0)} \cdot \boldsymbol{n} + \mathrm{HS} \int_\Omega Q \tag{2.11}$$

where $V_\Omega = \mathrm{meas}\,(\Omega)$ is the volume of the domain and $\boldsymbol{q}^{(0)}$ is the zero order term of the heat flux on the boundary (either prescribed as boundary condition or computed from the temperature). In general this equation will be an implicit equation for the reference pressure. In the case of an ideal gas, a physical interpretation is possible. The constant-in-space thermodynamic pressure changes in time due to the addition or subtraction of mass (left hand side term) or to heat addition or subtraction either by the boundary (second right hand side term) or by volumetric sources (last right hand side term).

2.3.1 The incompressible Navier Stokes equations

Let us consider a non-conducting fluid in absence of heat sources. In the case of open flows the thermodynamic pressure is constant. In the case of closed flows, if there is no addition of mass, we have

$$\int_{\partial\Omega} \boldsymbol{u}^{(0)} \cdot \boldsymbol{n} = 0$$

For closed flows this depends on the boundary conditions of the problem. In this case equation 2.11 gives

$$\frac{dp^{(0)}}{dt} = 0$$

Therefore, for open flows or closed flows without addition of mass, in absence of heating effects, we have a constant thermodynamic pressure. In such a case equation 2.10 gives

$$\boldsymbol{\nabla} \cdot \boldsymbol{u}^{(0)} = 0$$

and the system to be solved, called the non-homogeneous Navier Stokes equations in [104], is given by

$$\begin{aligned}
\boldsymbol{\nabla} \cdot \boldsymbol{u}^{(0)} &= 0 \\
\mathrm{S} \frac{\partial \rho^{(0)}}{\partial t} + \boldsymbol{u}^{(0)} \cdot \boldsymbol{\nabla} \rho^{(0)} &= 0 \\
\rho^{(0)} \left(\mathrm{S} \frac{\partial \boldsymbol{u}^{(0)}}{\partial t} + \boldsymbol{u}^{(0)} \cdot \boldsymbol{\nabla} \boldsymbol{u}^{(0)} \right) + \boldsymbol{\nabla} p^{(2)} &= \frac{1}{\mathrm{R}} \boldsymbol{\nabla} \cdot \left(2\mu^{(0)} \boldsymbol{\varepsilon}(\boldsymbol{u}^{(0)}) \right) + \frac{1}{\mathrm{F}^2} \rho^{(0)} \boldsymbol{g}
\end{aligned}$$

The temperature is recovered from the state equation

$$\rho = F\left(p^{(0)}, \vartheta\right)$$

If the temperature (or density) distribution is initially constant, it remains constant for all times and we have the homogeneous incompressible Navier Stokes equations, given by

$$\boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)}=0$$
$$\rho^{(0)}\left(\mathrm{S}\frac{\partial\boldsymbol{u}^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\boldsymbol{u}^{(0)}\right)+\boldsymbol{\nabla}p^{(2)}=\frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu^{(0)}\boldsymbol{\varepsilon}(\boldsymbol{u}^{(0)})\right)+\frac{1}{\mathrm{F}^2}\rho^{(0)}\boldsymbol{g}$$

2.3.2 The zero Mach number equations

The system to be solved in this case is given by

$$\mathrm{S}\frac{\partial\rho^{(0)}}{\partial t}+\boldsymbol{\nabla}\cdot\left(\rho^{(0)}\boldsymbol{u}^{(0)}\right)=0$$
$$\rho^{(0)}\left(\mathrm{S}\frac{\partial\boldsymbol{u}^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\boldsymbol{u}^{(0)}\right)+\boldsymbol{\nabla}p^{(2)}=\frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu^{(0)}\boldsymbol{\varepsilon\prime}(\boldsymbol{u}^{(0)})\right)+\frac{1}{\mathrm{F}^2}\rho^{(0)}\boldsymbol{g}$$
$$\rho^{(0)}c_p^{(0)}\left(\mathrm{S}\frac{\partial\vartheta^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}\right)-\mathrm{S}_t\beta\vartheta^{(0)}\mathrm{S}\frac{dp^{(0)}}{dt}=\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(0)}\right)+\mathrm{HS}Q$$

which has to be completed with a state equation of the form

$$\rho=F\left(p^{(0)},\vartheta\right)$$

where the thermodynamic pressure $p^{(0)}$ is either given by 2.11 or determined by the boundary conditions.

This system of equations does not present acoustic phenomena that are present in a compressible flow as shown in [126], [119] and [113]. Acoustic phenomena are pressure and density waves of small amplitude and fast propagation velocity (the sound speed c) that satisfy the system of equations. It is easy to see that a wave equation for the pressure can be deduced from the full compressible equations 2.2 to 2.4. When the Mach number is small the hyperbolic wave equation for the pressure becomes an elliptic equation for the first order pressure $p^{(2)}$, thus showing the implicit ("incompressible" or "mechanical") character of this pressure component. It is not an evolving variable but can be understood as an implicit Lagrange multiplier determined by the mass conservation.

2.4 The small Mach number and small Froude number limit

The low Mach number approximation developed in the previous section was carried out considering the rest of the dimensionless numbers fixed. In this section the possibility of a low Froude number is taken into account and the Boussinesq and anelastic approximations are presented. As previously mentioned, successive improvements of the derivation of the Boussinesq approximation have been made in [120] and [63] introducing a reference

state about which a perturbative scheme is developed. An asymptotic derivation of the Boussinesq approximation was considered in [12, 13] and [146, 148, 149]. In order to study this limit it is useful to introduce the Boussinesq number, defined as

$$\mathrm{B} = \frac{\rho_0 g l_0}{p_0} = \frac{\mathrm{M}^2}{\mathrm{F}^2}$$

This number was defined first in [146] but its importance in vertically stratified flows was already noted in [5]. When $\mathrm{M} \to 0$ and $\mathrm{F} \to 0$, the Boussinesq number can be finite, tend to zero or tend to infinity depending on the relation between F and M. The external force will be considered due to gravity and supposed in the $(-\hat{\boldsymbol{z}})$ direction where $\hat{\boldsymbol{z}} = (0, 0, 1)^t$. Depending on the Boussinesq number, the external force will be different in the hierarchy of equations obtained after the introduction of the low Mach number expansion. Two different cases can be considered.

1. If $\mathrm{B} \to 0$ as $\mathrm{B} = \mathcal{O}(\mathrm{M})$ when $\mathrm{M} \to 0$, we have

   $$\frac{\mathrm{M}}{\mathrm{F}^2} = \mathcal{O}(1)$$

   and (under some conditions to be given below) the Boussinesq approximation is found.

2. If $\mathrm{B} \not\to 0$, $\mathrm{B} = \mathcal{O}(1)$ when $\mathrm{M} \to 0$, we have

   $$\frac{\mathrm{M}}{\mathrm{F}} = \mathcal{O}(1)$$

   and the anelastic or the quasistatic approximations are found.

Let us mention that the condition of small Boussinesq number is a restriction on the height of the flow analyzed, as the quantity $\rho_0 g_0 / p_0$ is the height scale of the thermodynamic field. This is the reason why it is usually mentioned that the anelastic approximation removes the height limitation of the Boussinesq one (see [115], [44] and [61].

2.4.1 The Boussinesq approximation

In the case of $\mathrm{M} \simeq \mathrm{F}^2$ (here and below the symbol $\simeq$ denotes "of the same order") we have that when $\mathrm{M} \to 0$, $\mathrm{M} = g\mathrm{F}^2$ where g is a constant (when dimensions are restored it will be the gravity modulus) and the equations of motion become

$$\mathrm{S}\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho \boldsymbol{u}) = 0$$

$$\rho\left(\mathrm{S}\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}\right) + \frac{1}{\mathrm{M}^2}\boldsymbol{\nabla} p = \frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot(2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u})) - \frac{1}{\mathrm{M}}\rho g\hat{\boldsymbol{z}}$$

$$\rho c_p\left(\mathrm{S}\frac{\partial \vartheta}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta\right) - \mathrm{S}_t\beta\vartheta\left(\mathrm{S}\frac{\partial p}{\partial t} + \boldsymbol{u}\cdot\boldsymbol{\nabla} p\right) = \frac{\mathrm{M}^2}{\mathrm{R}}\Phi + \frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot(k\boldsymbol{\nabla}\vartheta) + \mathrm{HS}Q$$

The analysis is carried out assuming an asymptotic expansion of the form

$$\xi\left(\boldsymbol{x},t,\mathrm{M}\right)=\xi^{(0)}\left(\boldsymbol{x},t\right)+\mathrm{M}\xi^{(1)}\left(\boldsymbol{x},t\right)+\mathrm{M}^{2}\xi^{(2)}\left(\boldsymbol{x},t\right)+\mathcal{O}(\mathrm{M}^{4}) \tag{2.12}$$

for $\xi=\boldsymbol{u}$, $\xi=p$, $\xi=\rho$ and $\xi=\vartheta$. The choice of M as the parameter expansion is due to the fact that it is the parameter that appears in the system of equations. We also consider that the heat sources are small that is $\mathrm{H}\simeq\mathrm{M}$ when $\mathrm{M}\rightarrow 0$ or, more precisely,

$$\mathrm{H}=c\mathrm{M}$$

where c is a constant that is absorbed redefining Q. Introducing the expansion into the equations and considering $\mathrm{M}\rightarrow 0$ the following hierarchy of equations is obtained:

$$\mathcal{O}(\mathrm{M}^{0}):\quad \mathrm{S}\frac{\partial\rho^{(0)}}{\partial t}+\boldsymbol{\nabla}\cdot\left(\rho^{(0)}\boldsymbol{u}^{(0)}\right)=0 \tag{2.13}$$

$$\mathcal{O}(\mathrm{M}^{-2}):\qquad \boldsymbol{\nabla}p^{(0)}=0 \tag{2.14}$$

$$\mathcal{O}(\mathrm{M}^{-1}):\qquad \boldsymbol{\nabla}p^{(1)}=-\rho^{(0)}g\hat{\boldsymbol{z}} \tag{2.15}$$

$$\mathcal{O}(\mathrm{M}^{0}):\quad \rho^{(0)}\left(\mathrm{S}\frac{\partial\boldsymbol{u}^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\boldsymbol{u}^{(0)}\right)+\boldsymbol{\nabla}p^{(2)}=\frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu\boldsymbol{\varepsilon\imath}(\boldsymbol{u}^{(0)})\right)-\rho^{(1)}g\hat{\boldsymbol{z}} \tag{2.16}$$

$$\begin{aligned}\mathcal{O}(\mathrm{M}^{0}):\quad &\rho^{(0)}c_{p}^{(0)}\left(\mathrm{S}\frac{\partial\vartheta^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}\right)-\mathrm{S}_{t}\beta^{(0)}\vartheta^{(0)}\left(\mathrm{S}\frac{\partial p^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}p^{(0)}\right)\\ &=\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k\boldsymbol{\nabla}\vartheta^{(0)}\right)\end{aligned} \tag{2.17}$$

$$\begin{aligned}\mathcal{O}(\mathrm{M}^{1}):\quad &\rho^{(0)}c_{p}^{(0)}\left(\mathrm{S}\frac{\partial\vartheta^{(1)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(1)}\right)-\mathrm{S}_{t}\beta^{(0)}\vartheta^{(0)}\left(\mathrm{S}\frac{\partial p^{(1)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}p^{(1)}\right)\\ &+\left(\rho^{(1)}c_{p}^{(0)}+\rho^{(0)}c_{p}^{(1)}\right)\left(\mathrm{S}\frac{\partial\vartheta^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}\right)\\ &-\mathrm{S}_{t}\left(\beta^{(0)}\vartheta^{(1)}+\beta^{(1)}\vartheta^{(0)}\right)\left(\mathrm{S}\frac{\partial p^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}p^{(0)}\right)\\ &+\rho^{(0)}c_{p}^{(0)}\boldsymbol{u}^{(1)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}-\mathrm{S}_{t}\beta^{(0)}\vartheta^{(0)}\boldsymbol{u}^{(1)}\cdot\boldsymbol{\nabla}p^{(0)}\\ &=\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(1)}+k^{(1)}\boldsymbol{\nabla}\vartheta^{(0)}\right)+\mathrm{S}Q\end{aligned} \tag{2.18}$$

Under the assumptions considered, the pressure evolution equation 2.10 is written as

$$p^{(0)}\boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)}=-\frac{1}{\gamma}\mathrm{S}\frac{dp^{(0)}}{dt}+\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(0)}\right) \tag{2.19}$$

From equation 2.14 it follows that $p^{(0)}=p^{(0)}(t)$ and from equation 2.15 that $p^{(1)}=p^{(1)}(z,t)$ and $\rho^{(0)}=\rho^{(0)}(z,t)$. Then, from the state equation we have that $\vartheta^{(0)}=\vartheta^{(0)}(z,t)$. The form of the zero order thermodynamic variables depends on the

(boundary conditions of the) particular problem under consideration. Introducing the asymptotic expansion 2.12 in the boundary condition 2.7 and 2.8 we obtain

$$\mathcal{O}\left(\mathrm{M}^0\right): \qquad \vartheta^{(0)} = 1 \tag{2.20}$$

$$\mathcal{O}\left(\mathrm{M}^1\right): \qquad \vartheta^{(1)} = \vartheta_D^{(0)} \tag{2.21}$$

and

$$\mathcal{O}\left(\mathrm{M}^0\right): \qquad \frac{1}{\mathrm{P}} k\boldsymbol{n}\cdot\nabla\vartheta^{(0)} = 0 \tag{2.22}$$

$$\mathcal{O}\left(\mathrm{M}^1\right): \qquad \frac{1}{\mathrm{P}} k\boldsymbol{n}\cdot\nabla\vartheta^{(1)} = \mathrm{S}q^{(0)} \tag{2.23}$$

The solution $\vartheta^{(0)} = 1$ satisfies 2.17 and its boundary conditions 2.20 and 2.22. Therefore if the initial temperature perturbations are small (meaning order ε or higher), $\vartheta^{(0)} = 1$. Otherwise, the evolution problem of the zero order fields must be solved. In the first case, thanks to the state equation $\rho^{(0)} = \rho^{(0)}(t)$ and in the case of open flows or closed flows without addition of mass, the thermodynamic pressure equation 2.19 implies that $p^{(0)}$ and therefore $\rho^{(0)}$ are constants. Finally, equation 2.15 implies $p^{(1)} = p^{(1)}(z)$ and the system to be solved (given by 2.13,2.16 and 2.18) reads

$$\boldsymbol{\nabla}\cdot\boldsymbol{u}^{(0)} = 0$$

$$\rho^{(0)}\left(\mathrm{S}\frac{\partial\boldsymbol{u}^{(0)}}{\partial t} + \boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\boldsymbol{u}^{(0)}\right) + \boldsymbol{\nabla}p^{(2)} = \frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu\boldsymbol{\varepsilon\prime}(\boldsymbol{u}^{(0)})\right) - \rho^{(1)}g\hat{\boldsymbol{z}}$$

$$\rho^{(0)}c_p^{(0)}\left(\frac{\partial\vartheta^{(1)}}{\partial t} + \boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(1)}\right) + \mathrm{S}_t\beta^{(0)}\vartheta^{(0)}w^{(0)}\frac{dp^{(1)}}{dz} = \frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(1)}\right) + \mathrm{S}Q$$

where w is the component of $\boldsymbol{u}$ in the $\hat{\boldsymbol{z}}$ direction.

This system has to be completed with a state equation. For an ideal gas we have that

$$\mathcal{O}(1): \qquad p^{(0)} = \rho^{(0)}\vartheta^{(0)}$$

$$\mathcal{O}(\mathrm{M}): \qquad p^{(1)} = \rho^{(0)}\vartheta^{(1)} + \vartheta^{(0)}\rho^{(1)}$$

and the first order pressure is determined from 2.15 as

$$p^{(1)} = -\rho^{(0)}gz$$

to obtain

$$\rho^{(1)} = -\frac{\rho^{(0)}}{\vartheta^{(0)}}z - \frac{\rho^{(0)}}{\vartheta^{(0)}}\vartheta^{(1)}$$

The first term can be absorbed by the pressure gradient through a redefinition of the second order pressure and *the Boussinesq equations are obtained*.

Let us mention that the derivation of the Boussinesq approximation given in [12, 13] and [146, 148, 149] is somewhat different. First a fixed Boussinesq number is considered

and the expansion on powers of the Mach number is performed, to obtain, to the first order in the momentum equation

$$\boldsymbol{\nabla} p^{(0)} = -\mathrm{B}\rho^{(0)}\hat{\boldsymbol{z}} \tag{2.24}$$

from where it follows that

$$p^{(0)} = p^{(0)}(z,t)$$
$$\rho^{(0)} = \rho^{(0)}(z,t)$$

and using a state equation

$$\vartheta^{(0)} = \vartheta^{(0)}(z,t)$$

The *first hypothesis* made in [12, 13] and [146, 148, 149] is that *this reference state is independent of time*. The *second hypothesis*, motivated by equation 2.24, is that *the zero order thermodynamic fields depend on z only trough the variable $\zeta = \mathrm{B}z$* . Under these assumptions

$$p^{(0)} = p^{(0)}(z) = p^{(0)}(\zeta)$$
$$\rho^{(0)} = \rho^{(0)}(z) = \rho^{(0)}(\zeta)$$

from where

$$\frac{dp^{(0)}}{dz} = \frac{d\zeta}{dz}\frac{dp^{(0)}}{d\zeta} = \mathrm{B}\frac{dp^{(0)}}{d\zeta}$$

and equation 2.24 becomes

$$\frac{dp^{(0)}}{d\zeta} = -\rho^{(0)}.$$

Next the limit of small Boussinesq number is considered taking $\mathrm{B} \to 0$ as $\mathrm{B} \simeq \mathrm{M}$ and the Boussinesq approximation is recovered. For example, as $\rho^{(0)} = \rho^{(0)}(\zeta)$ the continuity equation 2.13 will give

$$0 = \boldsymbol{u}^{(0)} \cdot \boldsymbol{\nabla}\rho^{(0)} + \rho^{(0)}\boldsymbol{\nabla} \cdot \boldsymbol{u}^{(0)} = \boldsymbol{u}^{(0)} \cdot \hat{\boldsymbol{z}}\,\frac{d\rho^{(0)}}{dz} + \rho^{(0)}\boldsymbol{\nabla} \cdot \boldsymbol{u}^{(0)} = \boldsymbol{u}^{(0)} \cdot \hat{\boldsymbol{z}}\,\mathrm{B}\frac{d\rho^{(0)}}{d\zeta} + \rho^{(0)}\boldsymbol{\nabla} \cdot \boldsymbol{u}^{(0)}$$

that gives

$$\boldsymbol{\nabla} \cdot \boldsymbol{u}^{(0)} = 0$$

when $\mathrm{B} \to 0$.

In this way, taking the limits consecutively, a reference state that depends weakly on z is obtained instead of a constant one. Let us stress that, through this reasoning, the gravity term, that naturally should appear modifying the first order pressure (as it was presented above), appears modifying the zero order pressure and this is what makes possible to deal with the dependence of the reference state on z. What is actually shown in [12, 13] and [146, 148, 149] is that the Boussinesq approximation is found in the limit

of small Mach number and small Boussinesq number *assuming* an asymptotic expansion of the form

$$\xi\left(\boldsymbol{x},t,\mathrm{M}\right)=\xi^{(0)}\left(\zeta\right)+\mathrm{M}\xi^{(1)}\left(\boldsymbol{x},t\right)+\mathrm{M}^{2}\xi^{(2)}\left(\boldsymbol{x},t\right)+\mathcal{O}(\mathrm{M}^{3})$$

that is to say, assuming that the reference state depends weakly on z through ζ. Note that as $\mathrm{B}\rightarrow 0$ this variable represents a very small scale compared to the scale given by z. As it has been mentioned in the previous section, this type of variable is introduced to separate scales when a multiple scale analysis is performed, something that only makes sense when an unbounded domain is considered.

Finally, let us also stress that in any case *some hypothesis on the zero order thermodynamic fields are needed*. Although the derivation given by [12, 13] and [146, 148, 149] can explain the dependence of the reference state with respect to z, it is still necessary to assume that they do not depend on time.

2.4.2 The anelastic and quasistatic approximations

In the case of $\mathrm{M}\simeq\mathrm{F}$ the equations of motion are

$$\begin{aligned}\mathrm{S}\frac{\partial\rho}{\partial t}+\boldsymbol{\nabla}\boldsymbol{\cdot}(\rho\boldsymbol{u})&=0\\ \rho\left(\mathrm{S}\frac{\partial\boldsymbol{u}}{\partial t}+\boldsymbol{u}\cdot\boldsymbol{\nabla}\boldsymbol{u}\right)+\frac{1}{\mathrm{M}^{2}}\boldsymbol{\nabla}p&=\frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u})\right)-\frac{\mathrm{B}}{\mathrm{M}^{2}}\rho\hat{\boldsymbol{z}}\\ \rho c_{p}\left(\mathrm{S}\frac{\partial\vartheta}{\partial t}+\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta\right)-\mathrm{S}_{t}\beta\vartheta\left(\mathrm{S}\frac{\partial p}{\partial t}+\boldsymbol{u}\cdot\boldsymbol{\nabla}p\right)&=\frac{\mathrm{M}^{2}}{\mathrm{R}}\Phi+\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k\boldsymbol{\nabla}\vartheta\right)+\mathrm{HS}Q\end{aligned}$$

and the analysis is carried out assuming an asymptotic expansion of the form

$$\xi\left(\boldsymbol{x},t,\mathrm{M}\right)=\xi^{(0)}\left(\boldsymbol{x},t\right)+\mathrm{M}^{2}\xi^{(2)}\left(\boldsymbol{x},t\right)+\mathcal{O}(\mathrm{M}^{4}) \tag{2.25}$$

for $\xi=\boldsymbol{u}$, $\xi=p$, $\xi=\rho$, $\xi=\vartheta$. The hierarchy of equations obtained is similar to the one obtained for the case $\mathrm{M}\simeq\mathrm{F}^{2}$ in the previous section except for the momentum equation, and reads:

$$\mathcal{O}(\mathrm{M}^{0}):\qquad \mathrm{S}\frac{\partial\rho^{(0)}}{\partial t}+\boldsymbol{\nabla}\cdot\left(\rho^{(0)}\boldsymbol{u}^{(0)}\right)=0 \tag{2.26}$$

$$\mathcal{O}(\mathrm{M}^{-2}):\qquad \boldsymbol{\nabla}p^{(0)}=-\mathrm{B}\rho^{(0)}\hat{\boldsymbol{z}} \tag{2.27}$$

$$\mathcal{O}(1):\quad \mathrm{S}\rho^{(0)}\frac{\partial\boldsymbol{u}^{(0)}}{\partial t}+\rho^{(0)}\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\boldsymbol{u}^{(0)}+\boldsymbol{\nabla}p^{(2)}=\frac{1}{\mathrm{R}}\boldsymbol{\nabla}\cdot\left(2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u}^{(0)})\right)-\mathrm{B}\rho^{(2)}\hat{\boldsymbol{z}} \tag{2.28}$$

$$\mathcal{O}(\mathrm{M}^0): \qquad \rho^{(0)}c_p^{(0)}\left(\mathrm{S}\frac{\partial\vartheta^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}\right)-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\left(\mathrm{S}\frac{\partial p^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}p^{(0)}\right)$$
$$=\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(0)}\right)+\mathrm{S}Q \tag{2.29}$$

$$\mathcal{O}(\mathrm{M}^2): \qquad \rho^{(0)}c_p^{(0)}\left(\frac{\partial\vartheta^{(2)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(2)}\right)-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\left(\frac{\partial p^{(2)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}p^{(2)}\right)$$
$$+\left(\rho^{(2)}c_p^{(0)}+\rho^{(0)}c_p^{(2)}\right)\left(\frac{\partial\vartheta^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}\right)$$
$$-\mathrm{S}_t\left(\beta^{(0)}\vartheta^{(2)}+\beta^{(2)}\vartheta^{(0)}\right)\left(\frac{\partial p^{(0)}}{\partial t}+\boldsymbol{u}^{(0)}\cdot\boldsymbol{\nabla}p^{(0)}\right)$$
$$+\rho^{(0)}c_p^{(0)}\boldsymbol{u}^{(2)}\cdot\boldsymbol{\nabla}\vartheta^{(0)}-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\boldsymbol{u}^{(2)}\cdot\boldsymbol{\nabla}p^{(0)}$$
$$=\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(2)}+k^{(2)}\boldsymbol{\nabla}\vartheta^{(0)}\right)+\frac{1}{\mathrm{R}}\Phi^{(0)} \tag{2.30}$$

From equation 2.27 it follows that $p^{(0)}=p^{(0)}(z,t)$ and that $\rho^{(0)}=\rho^{(0)}(z,t)$. Then, from the state equation $\vartheta^{(0)}=\vartheta^{(0)}(z,t)$. Now *If* $\rho^{(0)}$, $\vartheta^{(0)}$ *and* $p^{(0)}$ *are independent of time* we have that 2.29 gives

$$w^{(0)}\left[\rho^{(0)}c_p^{(0)}\frac{d\vartheta^{(0)}}{dz}-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\frac{dp^{(0)}}{dz}\right]=\frac{1}{\mathrm{P}}\frac{d}{dz}\left(k^{(0)}\frac{d\vartheta^{(0)}}{dz}\right)+\mathrm{S}Q \tag{2.31}$$

Two different cases are found:

- If the reference state is such that

$$\rho^{(0)}c_p^{(0)}\frac{d\vartheta^{(0)}}{dz}-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\frac{dp^{(0)}}{dz}\neq 0 \tag{2.32}$$

  then

$$w^{(0)}=\frac{\frac{1}{\mathrm{P}}\boldsymbol{\nabla}\cdot\left(k^{(0)}\boldsymbol{\nabla}\vartheta^{(0)}\right)+\mathrm{S}Q}{\rho^{(0)}\frac{d\vartheta^{(0)}}{dz}-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\frac{dp^{(0)}}{dz}}$$

  or, for an ideal fluid in absence of external heating

$$w^{(0)}=0$$

  This case is called in [12, 13] and [146, 148, 149] the *quasi-static approximation*. The vertical velocity is constrained by an hydrostatic equilibrium in the vertical direction and only plane motions can occur. Further details can be found in the references already mentioned.

- If the reference state is such that

$$\rho^{(0)}c_p^{(0)}\frac{d\vartheta^{(0)}}{dz}-\mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\frac{dp^{(0)}}{dz}\approx 0 \tag{2.33}$$

the *anelastic approximation* follows. This condition, together with the zero order momentum and energy equations define the reference (zero order) state

$$\rho^{(0)} c_p^{(0)} \frac{d\vartheta^{(0)}}{dz} - \mathrm{S}_t \beta^{(0)} \vartheta^{(0)} \frac{dp^{(0)}}{dz} = 0 \tag{2.34}$$

$$\frac{1}{\mathrm{P}} \frac{d}{dz} \left(k^{(0)} \frac{d\vartheta^{(0)}}{dz} \right) + \mathrm{S}Q = 0 \tag{2.35}$$

$$\frac{dp^{(0)}}{dz} = \mathrm{B}\rho^{(0)} \tag{2.36}$$

where also the zero order state equation needs to be considered. For an ideal gas $p^{(0)} = \rho^{(0)}\vartheta^{(0)}$. The final set of equations to be solved, using this reference state, is given by

$$\boldsymbol{\nabla} \cdot \left(\rho^{(0)} \boldsymbol{u}^{(0)} \right) = 0$$

$$\mathrm{S}\rho^{(0)} \frac{\partial \boldsymbol{u}^{(0)}}{\partial t} + \rho^{(0)} \boldsymbol{u}^{(0)} \cdot \boldsymbol{\nabla} \boldsymbol{u}^{(0)} + \boldsymbol{\nabla} p^{(2)} = \frac{1}{\mathrm{R}} \boldsymbol{\nabla} \cdot \left(2\mu \boldsymbol{\varepsilon}\prime(\boldsymbol{u}^{(0)}) \right) - \mathrm{B}\rho^{(2)} \hat{\boldsymbol{z}}$$

$$\rho^{(0)} c_p^{(0)} \left(\mathrm{S} \frac{\partial \vartheta^{(2)}}{\partial t} + \boldsymbol{u}^{(0)} \cdot \boldsymbol{\nabla} \vartheta^{(2)} \right) + \left(\rho^{(2)} c_p^{(0)} + \rho^{(0)} c_p^{(2)} \right) w^{(0)} \frac{d\vartheta^{(0)}}{dz} - \mathrm{S}_t \left(\beta^{(2)} \vartheta^{(0)} + \beta^{(0)} \vartheta^{(2)} \right) \frac{dp^{(0)}}{dz}$$

$$-\mathrm{S}_t \beta^{(0)} \vartheta^{(0)} \left(\mathrm{S} \frac{\partial p^{(2)}}{\partial t} + \boldsymbol{u}^{(0)} \cdot \boldsymbol{\nabla} p^{(2)} \right) = \frac{1}{\mathrm{P}} \frac{d}{dz} \left(k^{(0)} \frac{d\vartheta^{(2)}}{dz} + k^{(2)} \frac{d\vartheta^{(0)}}{dz} \right) + \frac{1}{\mathrm{R}} \Phi^{(0)}$$

which also need to be closed by the state equation that in the case of an ideal gas is

$$p^{(2)} = \rho^{(0)}\vartheta^{(2)} + \vartheta^{(0)}\rho^{(2)}$$

This set of equations was presented in [119], where it is mentioned that they were written in this form in [100] and that they are a generalization of those obtained in [44] and [61]. As noted in [119], in the case of an ideal gas with constant c_p equations 2.34-2.35-2.36 can be solved and the reference state can be written as

$$\vartheta^{(0)} = \left(1 - \frac{\gamma - 1}{\gamma} \mathrm{B}z \right)$$

$$\rho^{(0)} = \left(1 - \frac{\gamma - 1}{\gamma} \mathrm{B}z \right)^{\frac{1}{\gamma - 1}}$$

$$p^{(0)} = \left(1 - \frac{\gamma - 1}{\gamma} \mathrm{B}z \right)^{\frac{\gamma}{\gamma - 1}}$$

If now the limit of $\mathrm{B} \to 0$ is taken, the Boussinesq approximation is recovered following the same steps as [12, 13] and [146, 148, 149] that is, taking the limits consecutively. In a bounded domain this also gives a constant reference state. The reference state defined by equations 2.34, 2.35 and 2.36 is the one introduced in [120] to improve the derivation of the Boussinesq approximation proposed in [110].

Let us close this section noting that condition 2.33 can be written as

$$\rho^{(0)}c_p^{(0)}\frac{d\vartheta^{(0)}}{dz} - \mathrm{S}_t\beta^{(0)}\vartheta^{(0)}\frac{dp^{(0)}}{dz} = \rho^{(0)}c_p^{(0)}\frac{ds^{(0)}}{dz} \approx 0$$

As shown in [101] the condition for the thermomechanical equilibrium of a fluid is

$$\frac{ds}{dz} > 0$$

A medium having $\frac{ds}{dz} = 0$ is neutrally stratified and a medium having $\frac{ds}{dz} < 0$ is unstably stratified. Then, the two cases to be considered when the Boussinesq number is not small, defined in 2.32 and 2.33, correspond to a neutral reference state or a stratified one (stable or unstable). If the reference state is stratified, the vertical velocity is constrained by the hydrostatic equilibrium. If the reference state is neutral, the anelastic approximation can be used. This is the first condition required in [115] to derive the anelastic approximation.

2.4.3 Applications

In this subsection we apply the developed framework to the problems defined at the end of section 2.2. Let us start considering the case of natural convection problems. We consider $\mathrm{S} = 1$, that is to say that we take l_0/u_0 as a reference time and we have the scales l_0, ρ_0, ϑ_0, μ_0, k_0, c_{p_0}, g_0 and $\Delta\vartheta$. Therefore these problems are described in terms of five parameters: M, F, R, P and ε. In the natural convection context, the Rayleigh-Bénard problem and the differentially heated cavity, it is common to consider the Rayleigh number Ra and the Prandtl number Pr, defined as

$$\mathrm{Ra} = \frac{gl_0^3}{\nu_0\alpha_0}\frac{\Delta\vartheta}{\vartheta_0}, \quad \mathrm{Pr} = \frac{\nu_0}{\alpha_0}$$

where ν_0 is the kinematic viscosity ($\nu_0 = \mu_{\prime 0}/\rho_0$), which satisfy

$$\mathrm{Ra} = \frac{\varepsilon}{\mathrm{F}^2}\mathrm{R}^2\mathrm{Pr}, \quad \mathrm{R} = \mathrm{PPr}$$

and to describe the problems in terms of M, F, Ra, Pr, and ε. A definition of the velocity scale eliminates one of these numbers. The problem of the differentially heated cavity has been analyzed by [23] and the Rayleigh Bénard problem has been analyzed in [110] in both cases defining *the velocity scale of the problem as the diffusive speed*, given by

$$u_0 = \frac{k_0}{\rho_0 c_{p_0} l_0} = \frac{\alpha_0}{l_0}$$

where α_0 is the thermal diffusivity scale ($\alpha_0 = k_0/\rho_0 c_{p_0} l_0$), what corresponds to make $\mathrm{P} = 1$. The differentially heated cavity problem has also been analyzed by [136] assuming that *the velocity scale problem is the viscous speed*, given by $u_0 = \nu_0/l_0$, what corresponds

to make R = 1. However, in our view, the most appropriate scaling for the velocity is the one used by [120] in the context of the Rayleigh Bénard problem and by [63], given by

$$u_0 = (\beta_0 \Delta\vartheta g l_0)^{1/2}$$

that we may call "¨buoyancy speed", that is obtained from

$$\mathrm{F}^2 = \varepsilon$$

The low Mach number approximation is valid when the Mach number is small, what physically means the velocity scale of the problem smaller than the sound speed. With our choice of the velocity scale we have $\mathrm{M}^2 = \mathrm{B}\varepsilon$ and this happens for thin layers *or* small temperature differences. The Boussinesq approximation requires also F and ε small with the restrictions $\mathrm{F}^2 = \mathcal{O}(\mathrm{M})$ and $\varepsilon = \mathcal{O}(\mathrm{M})$. The first one is equivalent to $\mathrm{B} = \mathcal{O}(\mathrm{M})$ and, with our choice of the velocity scale, the second condition is automatically satisfied. Therefore, it will be valid for thin layers *and* small temperature difference.

Another important aspect of the proposed approach is the possibility of analyzing mixed convection problems. In the case of the Poiseuille-Rayleigh-Bénard problem we have a velocity u_0 defined by the boundary condition. The low Mach number approximation is valid when the velocity u_0 is small compared to the sound speed. The Boussinesq approximation also requires $\mathrm{F}^2 = \mathcal{O}(\mathrm{M})$, what is equivalent to $\mathrm{B} = \mathcal{O}(\mathrm{M})$, and implies a restriction on the height, and $\varepsilon = \mathcal{O}(\mathrm{F}^2)$, what implies a velocity u_0 of the order of the buoyancy speed or, equivalently, small temperature differences (i.e. $\varepsilon = \mathcal{O}(\mathrm{M})$). However, if the velocity prescribed on the boundary is much smaller than the buoyancy speed, the problem will be similar to the Rayleigh-Bénard problem and the fluid motion will be driven by temperature differences. Therefore, when $\mathrm{F}^2 \ll \varepsilon$ we redefine the velocity scale by $\mathrm{F}^2 = \varepsilon$ and the case of natural convection is recovered. If $\mathrm{F}^2 \gg \varepsilon$ we keep the velocity scale given by the boundary conditions and the validity of the low Mach number approximation will depend directly on the Mach number. The Boussinesq approximation will be valid if the Froude number is also small, as the asymptotic analysis shows. Note that, in this case, the temperature difference is small because $\mathrm{F}^2 \gg \varepsilon$.

2.5 Summary and conclusions

The zero Mach number limit of the compressible flow equations yields different sets of equations depending on the type of flow analyzed. If an isentropic flow is considered, the incompressible Navier Stokes equations are recovered. When heat exchange is taken into account, different sets of equations are found. This limit was obtained using an expansion of the unknowns in series of the Mach number, which according to [134] is valid (i.e. yields a convergent solution) in the near field (see also [141]). When the Froude number is also small several situations can be found. On the one hand, the anelastic and the quasistatic

approximations are found when $\mathrm{M}, \mathrm{F} \to 0$ and $\mathrm{M} \simeq \mathrm{F}$, if a reference state depending on z is assumed. On the other hand, the Boussinesq approximation is found when $\mathrm{M}, \mathrm{F} \to 0$ and $\mathrm{M} \simeq \mathrm{F}^2$, $\mathrm{H} \simeq \mathrm{F}^2$ and $\varepsilon \simeq \mathrm{F}^2$ assuming appropriate initial and boundary conditions. This approximation is also valid in an unbounded domain if the reference state depends weakly on z through $\zeta = \mathrm{B}z$ as shown by [12, 13] and [146, 148, 149]. All these limits have been obtained under the same asymptotic setting proposed here. The physical meaning of the similarity rules introduced in the asymptotic analysis has been made precise in the case of bounded domains for both natural and mixed convection problems.

The three approximations considered when heat exchange is taken into account (the zero Mach number model, the anelastic approximation and the Boussinesq approximation) describe the basic mechanism of thermal coupling which is due to the dependence of the density on the temperature. When a fluid element is heated, it expands and moves up. None of the three approximations describe acoustic phenomena, what is certainly desirable from a numerical point of view. The main difference between them is how they take into account the compressibility of the medium. While in the Boussinesq approximation the flow is incompressible, in the zero Mach number model the density distribution is predicted and the velocity field is affected by expansions or contractions due to heating. Between them, the anelastic approximation (mainly used in atmospheric sciences) takes into account the density of the medium in the mass balance.

Chapter 3

The convection diffusion reaction problem

In this chapter we revisit the definition of the stabilization parameters for the convection diffusion reaction equation. We restrict ourselves to scalar problems and we focus our attention on the extension of the well known one dimensional case to the multidimensional one, considering also an anisotropic diffusion coefficient. The new definition of the parameter also takes into account anisotropy of the mesh used, what is possible thanks to a precise definition of the element size. The proposal is based on an approximation of the subgrid scale equation in the context of the variational multiscale method. The constants involved in the definition of the parameters arise naturally from the approximations performed. Some numerical experiments illustrating the contributions are also presented.

3.1 Introduction

The convection diffusion reaction (CDR) equation is a simple equation that describes several physical phenomena like, for example, heat transfer. In the development of numerical methods this simplicity is important because the problems found when solving more complex transport equations can be reproduced using this simplified model.

When attempting the numerical solution of the CDR equation, the first problem identified is the lack of stability of the Galerkin formulation when the convective term is important which manifest itself as numerical oscillations that pollute the solution in the whole domain and specially near boundary layers. After understanding this problem as a lack of diffusion in the discrete problem, the first solution was to add numerical dissipation developing upwind techniques in the context of the finite difference method. The inconsistent extra terms implied a loss of accuracy and the situation was fixed with the introduction of the SUPG method in [76, 93, 18], which was analyzed in [91]. This method depends on a parameter called the stabilization parameter and denoted usually by

τ. This parameter is also present in the Galerkin least squares method (GLS), introduced in [80] and analyzed in [50] as well as the Douglas-Wang method introduced in [43] in the context of the Stokes problem. These methods were related to the introduction of bubble functions in [15, 4, 17, 49], where it was shown that a choice of the bubble implies a choice of the stabilization parameter. The optimal bubble is given by the solution of a local subproblem driven by the residual [54], and is therefore named residual free. A general approach to the development of stabilized formulations is the variational multiscale method (VMM) introduced in [75, 78], based on a decomposition of the space into a coarse scale resolvable part and a fine scale subgrid part that, after some approximations, is found as the solution a local problems driven by the residual through the Green function approach. The equivalence between the residual free bubble and the variational multiscale method was established in [16]. Other methods introduced to solve this problem are the Characteristic Galerkin method [42] and the Taylor Galerkin method [41]. A comparison of all these methods was performed in [24]. A recent review of stabilization techniques for the CDR equation can be found in [51].

Another problem identified is the lack of stability when the reaction term is important which manifest itself as numerical oscillations localized near boundary layers. The methods mentioned lead to a stable discrete formulation but some of them (VMM) are much more accurate than others (GLS). The expression of the stabilization parameter needs to be modified to take reaction into account. An expression based on the satisfaction of the discrete maximum principle was proposed in [24]. If we denote the diffusion coefficient by ε the norm of the advective velocity by a and the reaction by s, this expression reads

$$\tau = \left(\frac{c_1 \varepsilon}{h^2} + \frac{c_2 a}{h} + \frac{1}{s} \right)^{-1} \tag{3.1}$$

where h is a characteristic element length and c_1 and c_2 are constants whose values, determined by numerical experiments, are 4 and 2 respectively. The expression proposed in [50] for the convection diffusion case, based on the error analysis, was extended to the reactive case in [55], obtaining an expression that behaves asymptotically as 3.1, what means that the limits of the expression with respect to any of the coefficients and with respect to mesh size are the same.

The dependence of the stabilization parameters with respect to the equation coefficients and the mesh size is determined by the error analysis. However, as pointed out in [67], convergence proofs are performed using functional analysis inequalities which depend on unknown constants what is sufficient as the error bounds are obtained up to a constant. Therefore constants appearing in 3.1 cannot be determined by error analysis except in particular problems. At the same time, the analysis is performed under strong assumptions on the mesh such as regularity of the elements or quasi-uniformity and general definitions of the mesh size parameter h are used (like the maximum or minimum

element length for example). On the other hand, precise definitions of the constants and the mesh parameter h are implemented in finite element codes, which are then used to solve application problems in meshes that are far from satisfying these constrains. The performance of the stabilized method presented in [140] (which is similar to the variational multiscale method in the context of the Navier Stokes equations) when high aspect ratio elements are used was analyzed in [111] and the need of incorporating the stretching of the grid in the definition of the stabilization parameter was emphasized.

An important effort in this direction is reported in [48] and the references therein, where anisotropic error estimates are developed for the convection diffusion equation using linear elements. Still some assumptions on the mesh are needed and the final error bound depends on a stretching factor that diverges when only one side of the element is reduced. In particular, the definition of the stabilization parameter using the minimum element length, as the analysis of [48] suggest, is not the most convenient as will be shown here. Another way in which the element length has been incorporated into the definition of the stabilization parameter is through the Jacobian of the isoparametric transformation, as in [81, 133]. A completely different approach, based on the calculation of norm of the element matrices and vectors, is presented in [139]. We can finally mention the finite calculus (FIC) method, based on expressing the equation of balance of fluxes in a domain of finite size, originally proposed in [117, 116] and modified in [118] by the introduction of a nonlinear stabilization parameter.

Although stabilization techniques have been extended to consider many different kinds of problems, a general definition of the stabilization parameters is still an open problem. In this work the definition of the stabilization parameters for scalar convection diffusion equations is revisited. The purpose of this chapter is to present a new definition of the stabilization parameters that can be directly implemented in a finite element code and that contains a precise definition of the element length and the values of the constants. The chapter is organized as follows. In section 3.2 we state the problem to be solved including the discrete formulation which is based on the variational multiscale method of [75, 78]. In section 3.3, the method to find an approximate solution of the fine scale problem is presented and the functional form of the stabilization parameter is defined. In section 3.4 the choice of the constants and the definition of the element length is discussed to arrive to the proposed definition of the stabilization parameters. In section 3.5 we will present a quite *standard error analysis* of the method *valid in the anisotropic case* in which we obtain an error estimate that depends on the interpolation error *without considering anisotropic interpolation estimates*. The analysis will pose a condition on the stabilization parameters due to the use of the inverse estimate. Numerical experiments illustrating the benefits of the method are presented in section 3.6 and final conclusions are drawn in section 3.7.

3.2 Problem statement

3.2.1 Continuous problem

We consider a convection diffusion reaction problem consisting of finding u such that

$$\begin{aligned} \mathcal{L}u := -\partial_i \left(\varepsilon_{ij}\partial_j u\right) + a_i \partial_i u + su &= f \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega \end{aligned}$$

Here $\Omega \subset \mathbb{R}^{\mathrm{d}}$ is an open domain in ($\mathrm{d} = 2, 3$ is the number of space dimensions) and $\partial\Omega$ its boundary, ε_{ij} is the constant (positive definite) diffusion tensor, a_i the solenoidal advection velocity, $s \geq 0$ the constant reaction coefficient and f a given internal force (the index summation convention is used here and in what follows). We restrict ourselves to the case of positive reaction, which corresponds to the exponential regime, and we refer to [71, 68] for the case of negative reaction, which corresponds to the propagation regime.

As usual, the space of functions whose p power ($1 \leq p < \infty$) is integrable in a domain ω, denoted by $L^p(\omega)$ and when $p = 2$ the inner product is denoted by $(\cdot, \cdot)_\omega$. The space of functions whose distributional derivatives of order up to $m \geq 0$ (integer) belong to $L^2(\omega)$ is denoted by $H^m(\omega)$. The space $H_0^1(\omega)$ consists of functions in $H^1(\omega)$ vanishing on $\partial\omega$. The topological dual of $H_0^1(\omega)$ is denoted by $H^{-1}(\omega)$ and $\langle \cdot, \cdot \rangle_\omega$ is used to denote the duality pairing between them.

The problem can be written in a weak form as follows: given $f \in H^{-1}(\Omega)$ and $a \in L^\infty(\Omega)$,find $u \in V := H_0^1(\Omega)$ such that

$$B(u, v) = L(v) \quad \forall v \in V$$

where

$$\begin{aligned} B(u, v) &= \left(\partial_i v, \varepsilon_{ij}\partial_j u\right)_\Omega + \left(v, a_i \partial_i u\right)_\Omega + \left(v, su\right)_\Omega \\ L(v) &= \langle v, f \rangle_\Omega \end{aligned}$$

The discretization of the problem is based on a finite element partition of the domain, $\mathcal{P}_h = \{K\}$, of size $h > 0$, which is a set of n_{el} elements K such that they cover the domain and their are either disjoint or share a complete edge (face). Based on this partition, the space V is approximated by a finite dimensional space V_h defined as

$$V_h = \left\{ w \in V : w \circ F^{-1}\big|_K \in P_p\left(\widehat{K}\right), 1 \leq p \leq \infty \right\}$$

where $P_p\left(\widehat{K}\right)$ denotes the set of polynomials of degree at most p (on each space variable if tetrahedral/hexahedral elements are used) and F the affine mapping from the reference element $\widehat{K}$ to the physical element K. Then, the Galerkin discrete problem consists in finding $u_h \in V_h$ such that

$$B(u_h, v_h) = L(v_h) \quad \forall v_h \in V_{0,h} \tag{3.2}$$

This formulation is not stable if diffusive terms are small compared either to convective or reactive ones.

3.2.2 Multiscale decomposition

Different stabilization techniques are used depending on the instability of the problem under consideration. A rather general method (that can be used in many cases) is the variational multiscale method. It is based on a decomposition of the unknown u into a resolvable part u_h and a subgrid scale part $\widetilde{u}$ which cannot be captured by the finite element mesh, what corresponds to a decomposition of the space V as

$$V = V_h \oplus \widetilde{V}.$$

The above decomposition, applied to the weak form of the problem, leads to

$$B(u_h, v_h) + B(\widetilde{u}, v_h) = L(v_h) \quad \forall v_h \in V_h \tag{3.3}$$

$$B(u_h, \widetilde{v}) + B(\widetilde{u}, \widetilde{v}) = L(\widetilde{v}) \quad \forall \widetilde{v} \in \widetilde{V} \tag{3.4}$$

The first equation is the equation for the resolvable scale u_h and has two terms: the first one is the Galerkin contribution and the second one takes into account the influence of the subgrid scale on u_h. The second one is an equation for the subgrid scale contribution. Let us introduce the following notation

$$\Omega^h = \bigcup_{K \in \mathcal{P}_h} K \quad \text{and} \quad \Gamma^h = \bigcup_{K \in \mathcal{P}_h} \partial K$$

and

$$(\cdot,\cdot)_h = \sum_{K \in \mathcal{P}_h} (\cdot,\cdot)_K, \quad (\cdot,\cdot)_{\partial h} = \sum_{K \in \mathcal{P}_h} (\cdot,\cdot)_{\partial K} \quad \text{and} \quad \|\cdot\|_h^2 = \sum_{K \in \mathcal{P}_h} \|\cdot\|_K^2$$

Integrating by parts within each element, equations 3.3 and 3.4 can be written as

$$\begin{aligned} B(u_h, v_h) + (\mathcal{L}^* v_h, \widetilde{u})_h + (n_i \varepsilon_{ij} \partial_j v_h, \widetilde{u})_{\partial h} &= L(v_h) \quad \forall v_h \in V_h \\ (\widetilde{v}, \mathcal{L}\widetilde{u})_h + (\widetilde{v}, n_i \varepsilon_{ij} \partial_j \widetilde{u})_{\partial h} &= (\widetilde{v}, (f - \mathcal{L}u_h))_h - (\widetilde{v}, n_i \varepsilon_{ij} \partial_j u_h)_{\partial h} \quad \forall \widetilde{v} \in \widetilde{V} \end{aligned}$$

where $\mathcal{L}^*$ is the adjoint of the operator $\mathcal{L}$ (with Dirichlet boundary conditions) given by

$$-\mathcal{L}^*(v) = \partial_i(\varepsilon_{ij} \partial_j u) + \partial_i(a_i u) - su$$

As the normal fluxes of the exact solution are continuous across any surface, it follows that

$$(\widetilde{v}, n_i \varepsilon_{ij} \partial_j u)_h = (\widetilde{v}, n_i \varepsilon_{ij} \partial_j \widetilde{u})_h + (\widetilde{v}, n_i \varepsilon_{ij} \partial_j u_h)_h = 0$$

Then, the second equation is equivalent to: find $\widetilde{u} \in \widetilde{V}$ such that

$$\begin{aligned} \mathcal{L}\widetilde{u} &= f - \mathcal{L}u_h + \widetilde{v}^{\perp} \quad \text{in} \quad \Omega^h \\ \widetilde{u} &= u_{\text{ske}} \quad \text{on} \quad \Gamma^h \end{aligned} \tag{3.5}$$

where u_{ske} is a function defined on the element boundaries and $\widetilde{v}^{\perp}$ is any function in $\widetilde{V}^{\perp}$ (the orthogonal complement of $\widetilde{V}$ in the $L^2\left(\Omega^h\right)$ sense). The function u_{ske} must be such that the normal fluxes of u are continuous across element boundaries. In turn, the function $\widetilde{v}^{\perp}$ is responsible for guaranteeing that $\widetilde{u} \in \widetilde{V}$. A modelling step is necessary to solve the system what means a choice of u_{ske}, $\widetilde{v}^{\perp}$ and an approximate solution of 3.5.

Note that 3.5 is posed in Ω^h which consists of the union of the elements of the mesh. Therefore, any choice of u_{ske} leads to n_{el} uncoupled problems posed on each element K. As a discrete approximation that gives exact nodal values would be optimal, one may ask the subscales to vanish at the nodes. In one dimensional problems, this gives homogeneous boundary conditions for problems 3.5 which are now decoupled and can be solved on each element. This has been done for the convection diffusion and Helmholtz equations (see [78] and the references therein). In more than one space dimension the choice $u_{\text{ske}} = 0$ is an approximation.

The approximated solution that will be constructed in the following section can be written as

$$\widetilde{u}|_K = \mathcal{L}^{-1}\left[(f - \mathcal{L}u_h) + \widetilde{v}^{\perp}\right]\big|_K \simeq \tau_K\left[(f - \mathcal{L}u_h) + \widetilde{v}^{\perp}\right]$$

This equation emphasizes that τ_K is an approximation to the (formal) inverse of the differential operator on each element K, a fact that will be used to construct an expression for it.

Finally we have to impose $\widetilde{u} \in \widetilde{V}$ which is equivalent to

$$0 = \left(\widetilde{u}, \widetilde{w}^{\perp}\right) \quad \forall \widetilde{w}^{\perp} \in \widetilde{V}^{\perp}$$

To this end let us consider the inner product

$$(\cdot,\cdot)_{\tau} = \sum_{K \in \mathcal{P}_h} (\tau_K \cdot, \cdot)_K$$

and let us consider the projection $\widetilde{P}_{\tau}^{\perp}$ onto $\widetilde{V}^{\perp}$ associated to the product $(\cdot,\cdot)_{\tau}$. We have

$$0 = \left(\widetilde{u}, \widetilde{w}^{\perp}\right) = \left(f - \mathcal{L}u_h, \widetilde{w}^{\perp}\right)_{\tau} + \left(\widetilde{v}^{\perp}, \widetilde{w}^{\perp}\right)_{\tau} \quad \forall \widetilde{w}^{\perp} \in \widetilde{V}^{\perp}$$

what implies

$$\widetilde{v}^{\perp} = -\widetilde{P}_{\tau}^{\perp}\left(f - \mathcal{L}u_h\right)$$

The projection $\widetilde{P}_{\tau}^{\perp}$ differs from the $L^2\left(\Omega^h\right)$ projection $\widetilde{P}^{\perp}$ in the element-by-element weights τ_K. If the stabilization parameter is the same for all elements we have $\widetilde{P}_{\tau}^{\perp} = \widetilde{P}^{\perp}$. In this simple case $\widetilde{u} \in \widetilde{V}$ if

$$0 = \widetilde{P}^{\perp}\widetilde{u} = \widetilde{P}^{\perp}\left[(f - \mathcal{L}u_h) + \widetilde{v}^{\perp}\right]$$

from where $\widetilde{v}^{\perp} = -\widetilde{P}^{\perp}\left(f - \mathcal{L}u_h\right)$. The final approximation is

$$\widetilde{u} = \tau\widetilde{P}_{\tau}\left(f - \mathcal{L}u_h\right)$$

where $\widetilde{P}_\tau = I - \widetilde{P}_\tau^\perp$ is the projection onto the subscale space $\widetilde{V}$ (I is the identity in V). A typical choice of the subscales space is given by $\widetilde{P}_\tau = I$ which is called in [29] the Algebraic Subgrid-Scale formulation (ASGS) and consists simply in taking $\widetilde{v}^\perp = 0$ to obtain

$$\widetilde{u}|_K = \tau_K \left(f - \mathcal{L}u_h\right)$$

In that reference the choice $\widetilde{P}_\tau = I - P_h := P_h^\perp$ is advocated, P_h being the $L^2\left(\Omega^h\right)$ projection onto the finite element space. The resulting formulation is called Orthogonal Subscales Stabilization (OSS) because when τ is the same for all elements this choice corresponds to take $\widetilde{V}$ as the orthogonal complement of V_h. If the element-by-element variation of the stabilization parameter is to be considered, in order to have $\widetilde{V} = V_h^\perp$ we need to take $\widetilde{P}_\tau = I - P_{h\tau}$ where $P_{h\tau}$ is the projection onto the finite element space in the sense of $(\cdot,\cdot)_\tau$. However, as the $L^2\left(\Omega^h\right)$ projection is very convenient from a computational point of view, the first choice is always considered and in this case we have

$$\widetilde{u}|_K = \tau_K P_h^\perp \left(f - \mathcal{L}u_h\right)$$

Neglecting boundary terms, the final stabilized discrete problem is: find $u_h \in V_h$ such that

$$B_\tau\left(u_h, v_h\right) = L_\tau\left(v_h\right) \quad \forall v_h \in V_h \tag{3.6}$$

where the stabilized forms are

$$B_\tau\left(u_h, v_h\right) = B\left(u_h, v_h\right) - \left(\mathcal{L}^* v_h, \tau \mathcal{L}u_h\right)_h$$
$$L_\tau\left(v_h\right) = L\left(v_h\right) - \left(\mathcal{L}^* v_h, \tau f\right)_h$$

3.3 Approximate solution of the subscale equation

In this section an approximate solution of equation 3.5 is presented. This equation for the subscale can be thought as an equation for the error and it is the equation used for the derivation of a posteriori error estimators[1], a fact already noted in [78]. In fact, it is used as error estimator in [45, 69, 70]. Two approaches are typical in a posteriori error estimation: an explicit expression for the error based on the residuals (derived from this equation) or the numerical solution of this equation (the so called implicit methods) [1]. In this case the first approach is followed because the problem is actually solved a priori and this relation between the subscale (the error) and the residual is used to stabilize the finite element problem.

The approximate solution is based on two properties that will be presented in the following subsections. The first is how the subgrid scale depends on the element size and will be determined by transforming the fine scale equation to the reference domain. The isoparametric transformation to the reference domain as a tool to define the stabilization

parameters was used first in [81, 133] but only for implementation purposes and it has not been related to the fine scale equation in the variational multiscale context that was developed later on in [78]. The second property is how the subgrid scale depends on the coefficients of the equation and will be determined by a heuristic argument already presented in [29] that will be revisited and extended. In this section we consider $\widetilde{v}^{\perp} = 0$ as it does not affect the discussion.

3.3.1 Transformation to the reference domain

Instead of directly solving

$$\begin{aligned} \mathcal{L}\widetilde{u} &= f - \mathcal{L}u_h := r \quad \text{in} \quad K \\ \widetilde{u} &= 0 \quad \text{on} \quad \partial K \end{aligned}$$

on each element K, we will transform this equation to the reference domain. The isoparametric transformation is defined by a mapping $\boldsymbol{x} = F(\boldsymbol{\xi})$, relating the element K (with coordinates $\boldsymbol{x}$) to the reference element $\hat{K}$ (with coordinates $\boldsymbol{\xi}$) whose Jacobian (J) verifies

$$J_{kl} = \frac{\partial x_l}{\partial \xi_k}, \quad J^{-t}_{kl} = \frac{\partial \xi_k}{\partial x_l}.$$

Therefore, we can write the fine scale problem as

$$-\frac{\partial}{\partial \xi_i}\left(\varepsilon^r_{ij}\frac{\partial \widetilde{u}}{\partial \xi_j}\right) + a^r_i \frac{\partial \widetilde{u}}{\partial \xi_i} + s\widetilde{u} = r \quad \text{in} \quad \hat{K} \tag{3.7}$$

where the modified velocity and diffusion coefficients are defined by

$$\begin{aligned} \varepsilon^r_{kl} &= J^{-t}_{ki} J^{-t}_{lj} \varepsilon_{ij} \\ a^r_k &= \left(\frac{\partial J^{-t}_{li}}{\partial \xi_l}\varepsilon_{ij} + a_j\right) J^{-t}_{kj} \end{aligned}$$

Note that the term in a^r_k that depends on the spatial derivatives of the Jacobian would not be present if the weak form of the problem is considered. Therefore another possibility that could be considered is to take

$$\varepsilon^r_{kl} = J^{-t}_{li} \varepsilon_{ij} J^{-t}_{kj} \tag{3.8}$$

$$a^r_k = a_j J^{-t}_{kj} \tag{3.9}$$

3.3.2 A Fourier analysis of the subscale problem

As in [29], let us consider the Fourier transform of a function v defined in $\hat{K}$ as

$$\widehat{v}(\boldsymbol{k}) = \int_{\hat{K}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\xi}}\, v(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

where $\mathrm{i} = \sqrt{-1}$ and $\boldsymbol{k}$ is the vector wave number. If $\boldsymbol{n}$ denotes the normal to the element $\hat{K}$ we have that

$$\widehat{\frac{\partial v}{\partial \xi_j}}(\boldsymbol{k}) = \mathrm{i}k_j \widehat{v}(\boldsymbol{k}) + \int_{\partial \hat{K}} n_j e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\xi}}\, v \mathrm{d}\xi$$

When this transform is applied to functions that vanish on the element boundary, the second term on the right hand side vanishes and we have

$$\widehat{\frac{\partial v}{\partial \xi_j}}(\boldsymbol{k}) = \mathrm{i}k_j \widehat{v}(\boldsymbol{k})$$

Transforming equation 3.7 we arrive to

$$\mathcal{T}^{-1}(\boldsymbol{k})\, \widehat{\widetilde{u}} = \widehat{r}$$

where

$$\mathcal{T}^{-1}(\boldsymbol{k}) := \left(k_i k_j \varepsilon_{ij}^r + s + \mathrm{i}k_j a_j^r\right)$$

Using the inverse Fourier transform the subgrid scale can be written as

$$\widetilde{u}(\boldsymbol{\eta}) = \int_{\mathbb{R}^{\mathrm{d}}} e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\eta}}\, \mathcal{T}(\boldsymbol{k})\, \widehat{r}(\boldsymbol{k})\, \mathrm{d}k$$

It is to be noted that the exact solution to the problem will depend on the element domain and the integration on the wave number space will be replaced by a sum over the values of $\boldsymbol{k}$ that make boundary conditions to be satisfied. In the above expression we can identify the Fourier representation of the Green function of the subscale problem [75] given by

$$\widetilde{u}(\boldsymbol{\eta}) = \int_{\hat{K}} G(\boldsymbol{\xi}, \boldsymbol{\eta})\, r(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi}$$

where

$$G(\boldsymbol{\xi}, \boldsymbol{\eta}) = \int_{\mathbb{R}^{\mathrm{d}}} \left(k_i k_j \varepsilon_{ij}^r + s + \mathrm{i}k_j a_j^r\right)^{-1} e^{-\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{\xi}-\boldsymbol{\eta})} \mathrm{d}\boldsymbol{k} \tag{3.10}$$

Up to this point no approximation has been performed except for the use of Fourier transforms in a bounded domain (and the assumption of $\widetilde{u} = 0$ on the element boundary). This expression, with the appropriate replacement of the integral on the wave number space by a sum, can be used to exactly calculate the subscale. However, this sum contains an infinite number of terms and must be truncated at some point. Doing this is equivalent to solving the fine scale problem with a discrete formulation, what has already been done in [53, 52] using a finite element or finite difference formulation instead of a spectral one. Apart from efficiency considerations such approach has a conceptual problem: the fine scale problem will suffer the same numerical instability as the problem defined in V_h. Although in this problem the mentioned instability will not manifest when the submesh is fine enough, this is not the case when other problems (i.e. Stokes) are

solved and the extension of the method will be difficult. Our main concern here is to find an approximation of 3.10.

It is well known [24, 75] that the use of a stabilization parameter τ corresponds to the approximation

$$G(\boldsymbol{\xi},\boldsymbol{\eta}) = \tau\delta(\boldsymbol{\xi}-\boldsymbol{\eta})$$

where δ denotes the Dirac distribution. From expression 3.10 it is quite clear that this corresponds to the approximation

$$\begin{aligned} G(\boldsymbol{\xi},\boldsymbol{\eta}) &\approx \left|\left(k_i k_j \varepsilon_{ij}^r + s + \mathrm{i}k_j a_j^r\right)^{-1}\right| \int_{\mathbb{R}^{\mathrm{d}}} e^{-\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{\xi}-\boldsymbol{\eta})} d\boldsymbol{k} \\ &= \left(\left(k_i^0 k_j^0 \varepsilon_{ij}^r + s\right)^2 + \left(k_j^0 a_j^r\right)^2\right)^{-1/2} \delta(\boldsymbol{\xi}-\boldsymbol{\eta}) \end{aligned}$$

for some $\boldsymbol{k}^0$ to be defined, and then

$$\boxed{\tau = \left(\left(k_i^0 k_j^0 \varepsilon_{ij}^r + s\right)^2 + \left(k_j^0 a_j^r\right)^2\right)^{-1/2}} \tag{3.11}$$

A justification for this approximation was presented in [29] and is briefly recalled here. Thanks to Plancharel's formula, the subgrid scale norm is given by

$$\begin{aligned} \|\widetilde{u}\|^2_{L^2(\hat{K})} &= \frac{1}{(2\pi)^{2\mathrm{d}}} \left\|\widehat{\widetilde{u}}\right\|^2_{L^2(\mathbb{R}^{\mathrm{d}})} = \frac{1}{(2\pi)^{2\mathrm{d}}} \|\mathcal{T}(\boldsymbol{k})\,\widehat{r}\|^2_{L^2(\mathbb{R}^{\mathrm{d}})} \\ &= \frac{1}{(2\pi)^{2\mathrm{d}}} \int_{\mathbb{R}^{\mathrm{d}}} |\mathcal{T}(\boldsymbol{k})\,\widehat{r}|^2 d\boldsymbol{k} = \frac{1}{(2\pi)^{2\mathrm{d}}} \int_{\mathbb{R}^{\mathrm{d}}} |\mathcal{T}(\boldsymbol{k})|^2 |\widehat{r}|^2 d\boldsymbol{k} \end{aligned}$$

and thanks to the mean value theorem, there exists $\boldsymbol{k}^0$ for which

$$\int_{\mathbb{R}^{\mathrm{d}}} |\mathcal{T}(\boldsymbol{k})|^2 |\widehat{r}|^2 d\boldsymbol{k} = \left|\mathcal{T}\left(\boldsymbol{k}^0\right)\right|^2 \int_{\mathbb{R}^{\mathrm{d}}} |\widehat{r}|^2 d\boldsymbol{k} = \left|\mathcal{T}\left(\boldsymbol{k}^0\right)\right|^2 \|\widehat{r}\|^2_{L^2(\mathbb{R}^{\mathrm{d}})}$$

Therefore, using again the Plancharel's formula

$$\frac{1}{(2\pi)^{2\mathrm{d}}} \|\widehat{r}\|^2_{L^2(\mathbb{R}^{\mathrm{d}})} = \|r\|^2_{L^2(\hat{K})}$$

from where

$$\|\widetilde{u}\|^2_{L^2(\hat{K})} = \left|\mathcal{T}\left(\boldsymbol{k}^0\right)\right|^2 \|r\|^2_{L^2(\hat{K})}$$

It follows that if we approximate the subscale as

$$\widetilde{u}^{\mathrm{ap}} = \tau R$$

and τ is defined as

$$\tau = \left|\mathcal{T}\left(\boldsymbol{k}^0\right)\right|$$

then

$$\|\widetilde{u}\|^2_{L^2(\hat{K})} = \|\widetilde{u}^{\mathrm{ap}}\|^2_{L^2(\hat{K})}$$

3.4 Definition of the stabilization parameter

Having established the functional form of the stabilization parameters let us finally consider the definition of $\boldsymbol{k}^0$, whose superscript will be omitted in this section. It will be shown that its magnitude is related to the constant factors involved in the definition of the parameter whereas its direction is related to the definition of the element length, thus answering the question posed in [67]: what are c and h?

3.4.1 The one dimensional problem

Let us first consider the problem in one space dimension without reaction. In this case the stabilization parameter presented above (see 3.11) is given by

$$\tau = \left[\left(\frac{h_{\text{nat}}^2 k^2 \varepsilon}{h^2} \right)^2 + \left(\frac{h_{\text{nat}} k a}{h} \right)^2 \right]^{-1/2}$$

where h_{nat} is the size of the reference domain and using the Péclet number defined as

$$\mathrm{P} = \frac{ah}{2\varepsilon}$$

it can be written in dimensionless form as

$$\alpha = \frac{2a\tau}{h} = \left(\left(\frac{h_{\text{nat}}^2 k^2}{4P^2} \right)^2 + \left(\frac{h_{\text{nat}} k}{2} \right)^2 \right)^{-1/2}$$

The advective limit of this expression is

$$\lim_{\mathrm{p}\to\infty} \alpha = 2 h_{\text{nat}}^{-1} k^{-1}$$

The analytic solution to the problem can be used to obtain the function α^{opt} that guarantees exact nodal values [24] which is given by

$$\alpha^{\text{opt}} = \coth(\mathrm{P}) - \frac{1}{\mathrm{P}}$$

The advective limit is of α^{opt} is 1 and therefore we conclude that

$$k = 2 h_{\text{nat}}^{-1}$$

must be taken. Both expressions are compared in figure 3.1 for this choice of k.

Note that the final expression for the stabilization parameter does not depend on the reference domain, as expected. Just to simplify the notation, we consider $h_{\text{nat}} = 2$ in what follows.

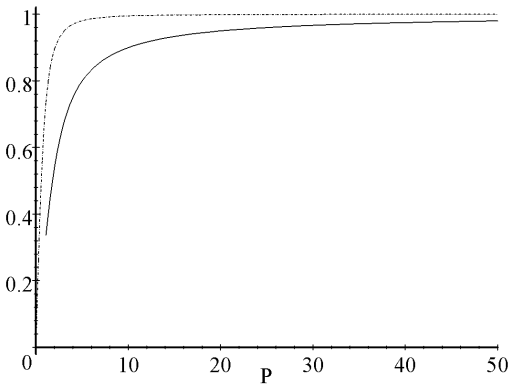

Figure 3.1: Upwind funtions

3.4.2 Extension to several dimensions: an isotropic approximation

The definition of the stabilization parameter given in 3.11 depends on the constant vector $\boldsymbol{k}$. It is therefore invariant under transformations of the reference system. In order to preserve this invariance, an invariant approximation needs to be performed. This is possible if it is assumed that the products $k_j a_j^r$ and $k_i k_j \varepsilon_{ij}^r$ depend on the invariants of $\boldsymbol{a}^r$ and $\boldsymbol{\varepsilon}^r$. In the first case the only invariant available is $\|\boldsymbol{a}^r\|$ whereas in the second we have three possible invariants. Considering the first invariant we can perform the approximations

$$k_j a_j^r \simeq \|\boldsymbol{k}\| \, \|\boldsymbol{a}^r\| \tag{3.12}$$

and

$$k_i k_j \varepsilon_{ij}^r \simeq \|\boldsymbol{k}\|^2 \, \varepsilon_{ii}^r \tag{3.13}$$

From the previous subsection we know that to obtain exact results in a one dimensional problem we need $\|\boldsymbol{k}\| = 1$ and therefore we arrive to the expression

$$\tau = \left((\varepsilon_{ii}^r + s)^2 + \|\boldsymbol{a}^r\|^2 \right)^{-1/2} \tag{3.14}$$

When $s = 0$ and the diffusion coefficient ε_{ij} is isotropic (given by $\varepsilon_{ij} = \varepsilon \delta_{ij}$) we have

$$\varepsilon_{ii}^r = \varepsilon J_{ij}^{-t} J_{ij}^{-t} = \varepsilon \frac{\partial \xi_i}{\partial x_j} \frac{\partial \xi_i}{\partial x_j}$$

and

$$\|\boldsymbol{a}^r\|^2 = a_i J_{ki}^{-t} a_j J_{kj}^{-t} = a_i \frac{\partial \xi_k}{\partial x_i} \frac{\partial \xi_k}{\partial x_j} a_j$$

and we arrive to

$$\tau = \left(\varepsilon^2 g_{ij} g_{ij} + a_i g_{ij} a_j \right)^{-1/2} \tag{3.15}$$

where

$$g_{ij} = \frac{\partial \xi_k}{\partial x_i}\frac{\partial \xi_k}{\partial x_j}$$

is the metric tensor related to the isoparametric mapping.

The stabilization parameter defined by 3.15 was proposed first in [133] in the context of compressible flow equations and has been used in several applications [137, 143, 21]. It is clear that the approximations 3.12 and 3.13 do not take into account the angle between the equation coefficients and the vector $\boldsymbol{k}$. It is due to these approximations, based on invariant quantities only, that the information on the anisotropy of the mesh is lost. Another possibility is analyzed in the next subsection.

3.4.3 Extension to several dimensions: an anisotropic approximation.

In order to extend the one dimensional definition to several space dimensions an intrinsic definition of the vector $\boldsymbol{k}$ is needed.

General considerations

Let us first consider a pure convection diffusion problem ($s = 0$). If we move from one to several space dimensions, the same argument that was used to pass from the artificial diffusion method to the streamline diffusion method [76, 93] can be used here. If a constant velocity is considered and the problem is written in a reference system such that one direction coincides with the streamlines, we actually obtain a one dimensional problem in this direction and a pure diffusion problem in the orthogonal ones. This implies that the diffusion that needs to be considered to define the Péclet number is the diffusion along the streamlines, what immediately suggest to take $\boldsymbol{k} = \frac{\boldsymbol{a}^r}{\|\boldsymbol{a}^r\|}$, arriving to

$$\tau = \left[\left(\frac{1}{\|\boldsymbol{a}^r\|^2} a_i^r a_j^r \varepsilon_{ij}^r \right)^2 + \|\boldsymbol{a}^r\|^2 \right]^{-1/2} \tag{3.16}$$

Remark 1 ***The subgrid problem solved in the reference domain in many cases will present an anisotropic diffusion.*** *As an example we may consider a two dimensional convection diffusion problem defined on the unit square with an isotropic diffusion ε and a velocity of the form $\boldsymbol{a} = (a, 0)^t$. If the discretization is performed using rectangular elements of sizes $h_1 = 1/n_1$ and $h_2 = 1/n_2$, n_1 and n_2 being the number of elements along each side of the domain, we have that*

$$\boldsymbol{\varepsilon}^r = \begin{bmatrix} \frac{2}{h_1} & 0 \\ 0 & \frac{2}{h_2} \end{bmatrix} \begin{bmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{bmatrix} \begin{bmatrix} \frac{2}{h_1} & 0 \\ 0 & \frac{2}{h_2} \end{bmatrix} = \begin{bmatrix} \frac{4}{h_1^2}\varepsilon & 0 \\ 0 & \frac{4}{h_2^2}\varepsilon \end{bmatrix}$$

and

$$\boldsymbol{a}^r = \begin{bmatrix} \frac{2}{h_1} & 0 \\ 0 & \frac{2}{h_2} \end{bmatrix} \boldsymbol{a} = \left(\frac{2a}{h_1}, 0 \right)^t$$

Then 3.16 gives

$$\tau = \left[\left(\frac{4\varepsilon}{h_1^2} \right)^2 + \left(\frac{2a}{h_1} \right)^2 \right]^{-1/2}$$

This simple example shows that expression 3.16 takes into account the fact that refining the mesh in the direction orthogonal to $\boldsymbol{a}$ *will not have any effect on the solution and only refining along the direction of* $\boldsymbol{a}$ *makes sense. We could say, roughly speaking, that refining the mesh is like adding diffusion (what is clearly seen in the factor* $4\varepsilon/h_1^2$*) and as this example shows that if the mesh is refined only in one direction the added diffusion is anisotropic.*

Remark 2 ***The selection of the vector k implies a definition of the element length used in the definition of the stabilization parameter.*** *As an example we may consider the problem of the previous remark but with a general velocity* $\boldsymbol{a} = (a_1, a_2)^t$*. In the case*

$$\boldsymbol{a}^r = \begin{bmatrix} \frac{2}{h_1} & 0 \\ 0 & \frac{2}{h_2} \end{bmatrix} \boldsymbol{a} = \left(\frac{2a_1}{h_1}, \frac{2a_2}{h_2} \right)^t$$

from where

$$\|\boldsymbol{a}^r\|^2 = 4 \left[\frac{a_1^2}{h_1^2} + \frac{a_2^2}{h_2^2} \right]$$

and expression 3.16 gives

$$\begin{aligned} \tau &= \left[\left(\frac{1}{\|\boldsymbol{a}^r\|^2} \left(a_1^r a_1^r \varepsilon_{11}^r + a_2^r a_2^r \varepsilon_{22}^r \right) \right)^2 + \|\boldsymbol{a}^r\|^2 \right]^{-1/2} \\ &= \left[\left(\frac{16\varepsilon}{\|\boldsymbol{a}^r\|^2} \left(\frac{a_1^2}{h_1^4} + \frac{a_2^2}{h_2^4} \right) \right)^2 + \|\boldsymbol{a}^r\|^2 \right]^{-1/2} \end{aligned}$$

This expression can be written as

$$\tau = \left[\left(\frac{4\varepsilon}{h_\varepsilon^2} \right)^2 + \left(\frac{2\|\boldsymbol{a}\|}{h_a} \right)^2 \right]^{-1/2} \tag{3.17}$$

where

$$h_\varepsilon^2 = \frac{1}{4} \|\boldsymbol{a}^r\|^2 \left(\frac{a_1^2}{h_1^4} + \frac{a_2^2}{h_2^4} \right)^{-1} = \left[\frac{a_1^2}{h_1^2} + \frac{a_2^2}{h_2^2} \right] \left(\frac{a_1^2}{h_1^4} + \frac{a_2^2}{h_2^4} \right)^{-1} \tag{3.18}$$

and

$$h_a = 2 \frac{\|\boldsymbol{a}\|}{\|\boldsymbol{a}^r\|} \tag{3.19}$$

are length scales that depend on the velocity direction. As mentioned before, natural candidates for the definition of the element length are the maximum element length ($h_{\max}$*),*

the minimum element length ($h_{\min}$) and the streamline element length (h_a). The definition of the element length in the direction of the flow 3.19 was considered in [34] and in [111]. Let us emphasize that neither the minimum nor the maximum element length can be used. If the minimum element length is used refining the mesh in the direction orthogonal to the velocity would make $h \to 0$ and $\tau \to 0$ without eliminating the instability. If the maximum element length is used and the mesh in the direction orthogonal to the velocity is too coarse would give a non zero value of τ even if the instability has been eliminated. In any case, the convergence analysis presented below imposes a condition on the choice of the element length that needs to be satisfied (see 3.23).

Expression 3.16 presents an important conceptual problem: the definition of the stabilization parameter will depend on the velocity direction even when $\|\boldsymbol{a}\| \to 0$, a fact that is seen in the definition of the element length 3.18. This problem is not shared by the isotropic expression 3.14 of the previous section, which can also be written in the form 3.17 using the streamline length for h_a but taking

$$h_\varepsilon^2 = \left(\frac{1}{h_1^2} + \frac{1}{h_2^2} \right)^{-1}$$

which is similar to $h_{\min}$.

In any case, neither definition 3.16 nor definition 3.14 can be used when reaction is present and the mesh anisotropic. Consider a convection diffusion reaction problem defined on the unit square with an isotropic diffusion ε and a velocity of the form $\boldsymbol{a} = (a, 0)^t$ solved using a mesh of rectangular elements whose sizes are such that

$$\frac{ah_1}{2\varepsilon} \ll 1, \quad \frac{sh_1^2}{4\varepsilon} \ll 1 \quad \text{and} \quad \frac{sh_2^2}{2\varepsilon} \gg 1 \tag{3.20}$$

that is, a mesh that is fine in direction 1 and coarse in direction 2. In this case the stabilization parameter using either 3.16 or 3.14 would be

$$\begin{aligned} \tau^{-2} &\simeq \left(\frac{4\varepsilon}{h_1^2} + s \right)^2 + \left(\frac{2a}{h_1} \right)^2 \\ &= \frac{16\varepsilon^2}{h_1^4} \left[\left(1 + \frac{sh_1^2}{4\varepsilon} \right)^2 + \left(\frac{ah_1}{2\varepsilon} \right)^2 \right] \end{aligned}$$

giving in the limit

$$\lim_{h_1 \to 0} \tau = \lim_{h_1 \to 0} \frac{h_1^2}{4\varepsilon} = 0$$

and will not take into account the reactive instability of the problem.

The choice of the direction

In order to get an insight of how the vector $\boldsymbol{k}$ should be taken let us consider two directions $\boldsymbol{k}_1 = (1, 0)$ and $\boldsymbol{k}_2 = (0, 1)$, and compare the stabilization parameter obtained using each of them in the following two examples.

Example 1 *Let us consider a convection diffusion problem defined on the unit square with an isotropic diffusion ε and a velocity of the form $\boldsymbol{a} = (a, 0)^t$. We have*

$$\tau(\boldsymbol{k}_1) = \left[\left(\frac{4\varepsilon}{h_1^2}\right)^2 + \left(\frac{2a}{h_1}\right)^2\right]^{-1/2}, \quad \tau(\boldsymbol{k}_2) = \left(\frac{4\varepsilon}{h_2^2}\right)^{-1}$$

Let us consider a convection dominated problem in a uniform mesh of size $h_1 = h_2 = h$. We have that

$$\frac{ah}{2\varepsilon} \gg 1$$

and

$$\tau(\boldsymbol{k}_1) \sim \frac{h}{2a}, \quad \tau(\boldsymbol{k}_2) = \frac{h^2}{4\varepsilon}$$

In this case

$$\tau(\boldsymbol{k}_1) \ll \tau(\boldsymbol{k}_2)$$

and the stabilization parameter should be given by $\tau(\boldsymbol{k}_1)$, what suggest to take the minimum of $\tau(\boldsymbol{k}_1)$ and $\tau(\boldsymbol{k}_2)$. As shown before, this is equivalent to take $\boldsymbol{k}$ in the direction of the velocity.□

Example 2 *Let us consider a diffusion reaction problem ($\boldsymbol{a} = \boldsymbol{0}$) defined on the unit square with an isotropic diffusion ε. We have*

$$\tau(\boldsymbol{k}_1) = \left(\frac{4\varepsilon}{h_1^2} + s\right)^{-1}, \quad \tau(\boldsymbol{k}_2) = \left(\frac{4\varepsilon}{h_2^2} + s\right)^{-1}$$

Let us consider a reaction dominated problem (small diffusion) and the two cases of anisotropic refinement. First if the mesh in direction 1 is very fine but is coarse in direction 2 then

$$\frac{sh_1^2}{4\varepsilon} \ll 1 \quad \textit{and} \quad \frac{sh_2^2}{2\varepsilon} \gg 1$$

and we have that

$$\tau(\boldsymbol{k}_1) \sim \frac{h_1^2}{4\varepsilon} \quad \tau(\boldsymbol{k}_2) \sim \frac{1}{s}$$

what implies

$$\tau(\boldsymbol{k}_1) \ll \tau(\boldsymbol{k}_2)$$

In this case the stabilization parameter should be given by $\tau(\boldsymbol{k}_2)$. Second, if the mesh in the direction 1 is fine enough but coarse in the direction 2 then

$$\frac{sh_1^2}{4\varepsilon} \gg 1 \quad \textit{and} \quad \frac{sh_2^2}{2\varepsilon} \ll 1$$

and we have that

$$\tau(\boldsymbol{k}_1) \sim \frac{1}{s}, \quad \tau(\boldsymbol{k}_2) \sim \frac{h_2^2}{4\varepsilon}$$

what implies

$$\tau(\boldsymbol{k}_2) \ll \tau(\boldsymbol{k}_1)$$

In this case the stabilization parameter should be given by $\tau(\boldsymbol{k}_1)$*. Therefore, in the case of a pure reactive problem we could consider the maximum of* $\tau(\boldsymbol{k}_1)$ *and* $\tau(\boldsymbol{k}_2)$ *or simply their sum. This is equivalent to consider the direction* $\boldsymbol{k}$ *of maximum element length or, in other words, the direction of minimum diffusivity, i.e. direction* $\boldsymbol{k}$ *that makes* $k_i k_j \varepsilon_{ij}^r$ *minimum.*□

The situation is similar in the two examples in a sense that gives rise to an important conclusion: $\boldsymbol{k}$ *depends on the direction in which the instability of the problem appears.* It is clear that in these examples the way to determine which is the correct definition of the stabilization parameter was by determining the direction in which the instability appears and this was done by comparing the dimensionless numbers defined on each direction. These numbers are

$$\frac{ah_1}{2\varepsilon}, \quad \frac{ah_2}{2\varepsilon}, \quad \frac{sh_1^2}{4\varepsilon}, \quad \frac{sh_2^2}{2\varepsilon}$$

and the one that is dominant defines the direction that needs to be considered.

In a general case, these numbers naturally appear if we consider the dimensionless parameter

$$k_i k_j \varepsilon_{ij}^r \tau(\boldsymbol{k}) = \left[\left(1 + \frac{s}{k_i k_j \varepsilon_{ij}^r} \right)^2 + \left(\frac{k_l a_l^r}{k_i k_j \varepsilon_{ij}^r} \right)^2 \right]^{-1/2}$$

which immediately suggests the definitions

$$\mathrm{P}_{\boldsymbol{k}} = \frac{\left| k_j a_j^r \right|}{k_i k_j \varepsilon_{ij}^r} \quad \mathrm{D}_{\boldsymbol{k}} = \frac{s}{k_i k_j \varepsilon_{ij}^r}$$

Then, the direction of maximum instability, that of the maximum $\mathrm{P}_{\boldsymbol{k}}$ and $\mathrm{D}_{\boldsymbol{k}}$, will be given by the minimum of $k_i k_j \varepsilon_{ij}^r \tau(\boldsymbol{k})$. Equivalently, we can define the direction of maximum instability ($\boldsymbol{k}^{\mathrm{I}}$) as

$$\boxed{\boldsymbol{k}^{\mathrm{I}} = \arg\max_{\|\boldsymbol{k}\|=1} \frac{\tau^{-1}(\boldsymbol{k})}{k_i k_j \varepsilon_{ij}^r}} \tag{3.21}$$

and the stabilization parameter we propose is given by $\tau\left(\boldsymbol{k}^{\mathrm{I}}\right)$, that is,

$$\boxed{\tau = \tau\left(\boldsymbol{k}^{\mathrm{I}}\right) = \left(\left(k_i^{\mathrm{I}} k_j^{\mathrm{I}} \varepsilon_{ij}^r + s \right)^2 + \left(k_j^{\mathrm{I}} a_j^r \right)^2 \right)^{-1/2}} \tag{3.22}$$

The computation of the direction

Definition 3.21 implies the maximization of the function

$$H(\boldsymbol{k}) = \frac{\tau^{-1}(\boldsymbol{k})}{k_i k_j \varepsilon_{ij}^r} = \left[\left(1 + \frac{s}{k_i k_j \varepsilon_{ij}^r} \right)^2 + \left(\frac{k_j a_j^r}{k_i k_j \varepsilon_{ij}^r} \right)^2 \right]^{1/2}$$

but, as the square root is a monotone function, we may solve the equivalent problem of maximizing $H^2(\boldsymbol{k})$. This optimization problem will be approximately solved. After

multiplying its gradient by $\left(k_i k_j \varepsilon^r_{ij}\right)^3$ we arrive to the equation

$$-\left[\left(k_j a^r_j\right)^2 + \left(k_i k_j \varepsilon^r_{ij} + s\right) s\right] \nabla\left(k_i k_j \varepsilon^r_{ij}\right) + \left(k_i k_j \varepsilon^r_{ij}\right)\left(k_j a^r_j\right) \nabla\left(k_j a^r_j\right) = 0$$

As the minimization is performed under the restriction $\|\boldsymbol{k}\| = 1$, in 2D we can take $\boldsymbol{k} = (\cos\theta, \sin\theta)$ and after a change of variables of the form $x = \tan\theta$ we arrive to a fourth order polynomial equation whose solution can be explicitly found. Let us consider some particular cases in two dimensions

- When $s = 0$ the problem simplifies to

$$-\left(k_j a^r_j\right) \nabla\left(k_i k_j \varepsilon^r_{ij}\right) + \left(k_i k_j \varepsilon^r_{ij}\right) \nabla\left(k_j a^r_j\right) = 0$$

and after taking $\boldsymbol{k} = (\cos\theta, \sin\theta)$ we arrive to a third order polynomial equation

$$-a^r_1\left(\varepsilon^r_{12} + \varepsilon^r_{21}\right) + a^r_2 \varepsilon^r_{11} + \left[a^r_1 \varepsilon^r_{11} - 2a^r_1 \varepsilon^r_{22} - 2a^r_2\left(\varepsilon^r_{12} + \varepsilon^r_{21}\right)\right] x + \left[2a^r_2 \varepsilon^r_{11} - a^r_2 \varepsilon^r_{22}\right] x^2 + \left[a^r_2\left(\varepsilon^r_{12} + \varepsilon^r_{21}\right) - a^r_1 \varepsilon^r_{22}\right] x^3 = 0$$

where $x = \tan\theta$. If we further assume the situation of remark 1 this equation simplifies to

$$\left[\varepsilon^r_{11} - 2\varepsilon^r_{22}\right] x - \varepsilon^r_{22} x^3 = 0$$

and we have two possible solutions that can be found as illustrated in figure 3.2

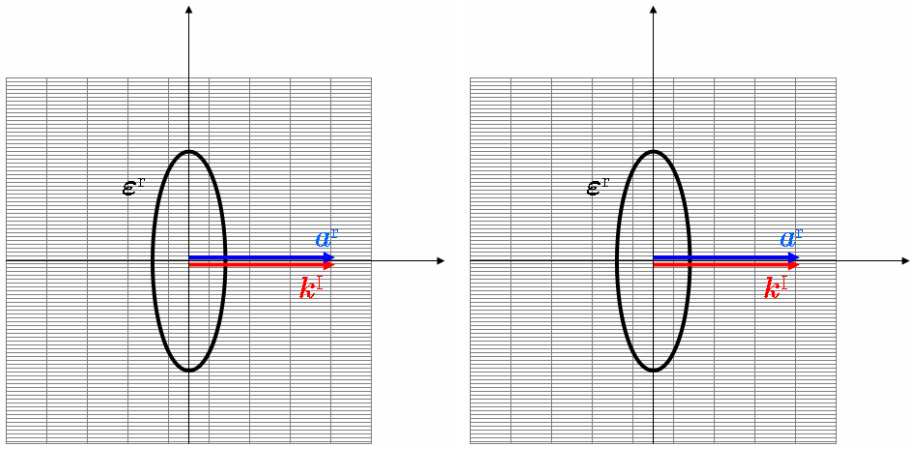

Figure 3.2: The definitiion of the instability direction in cases a (left) and b (right)

a When the mesh is such that

$$\varepsilon^r_{11} < 2\varepsilon^r_{22} \Leftrightarrow \frac{4\varepsilon}{h_1^2} < 2\frac{4\varepsilon}{h_2^2} \Leftrightarrow h_2^2 < 2h_1^2$$

or, equivalently

$$\mathrm{A} := \frac{h_2}{h_1} < \sqrt{2}$$

we have the solution $\theta = 0$.

b When the mesh is such that

$$\varepsilon_{11}^r > 2\varepsilon_{22}^r \Leftrightarrow \frac{4\varepsilon}{h_1^2} > 2\frac{4\varepsilon}{h_2^2} \Leftrightarrow h_2^2 > 2h_1^2$$

or, equivalently

$$\mathrm{A} = \frac{h_2}{h_1} > \sqrt{2}$$

we also have the solution

$$\tan^2\theta = \frac{\varepsilon_{11}^r - 2\varepsilon_{22}^r}{\varepsilon_{22}^r} = a^2 - 2$$

and the stabilization parameter is given by

$$\tau = \left(\left(\frac{4\varepsilon}{h_1^2}\cos\theta + \frac{4\varepsilon}{h_2^2}\sin^2\theta \right)^2 + \left(\frac{2a_1}{h_1}\cos\theta \right)^2 \right)^{-1/2}$$

- When $\boldsymbol{a} = \mathbf{0}$ the problem simplifies to

$$\nabla\left(k_i k_j \varepsilon_{ij}^r\right) = 0$$

that corresponds to find the direction of minimum diffusion. We can follow the same procedure used before to show that when $\boldsymbol{\varepsilon}^r$ is diagonal the solutions are $\theta = 0$ or $\theta = \pi/2$ and if it is not we have

$$\tan\theta = -\frac{(\varepsilon_{11}^r - \varepsilon_{22}^r)}{(\varepsilon_{12}^r + \varepsilon_{21}^r)} \pm \left(\frac{(\varepsilon_{11}^r - \varepsilon_{22}^r)^2}{(\varepsilon_{12}^r + \varepsilon_{21}^r)^2} + 1 \right)^{1/2}$$

As these particular cases illustrate the maximum of $H(\boldsymbol{k})$ will occur somewhere between the direction of minimum diffusion and the direction of $\boldsymbol{a}^r$. Therefore, in practice, we find this maximum simply evaluating the function on a given number of points between these two directions.

3.5 Error analysis

In this section we present the error analysis of the method in the case of $\varepsilon_{ij} = \varepsilon\delta_{ij}$, following a standard approach. We start by proving stability in a discrete norm to be defined and then we obtain a bound in terms of the interpolation error. A key ingredient is the anisotropic inverse estimate [2] which can be derived from a scaling argument

$$\left\|\nabla^2 u_h\right\|_K^2 \leq \frac{C_{\mathrm{inv}}^2}{h_{\min}^2} \left\|\boldsymbol{\nabla} u_h\right\|_K^2 \quad \forall u_h \in V_h \tag{3.23}$$

The important result of this section is that, due to the need of using the inverse estimate 3.23, the stabilization parameter must satisfy the following condition

$$\tau^{-1} > 4\frac{\varepsilon C_{\mathrm{inv}}^2}{h_{\min}^2} + s \tag{3.24}$$

In the case of linear elements $C_{\text{inv}} = 0$ and the condition is automatically satisfied by 3.22 and 3.1. If higher order elements are used, only taking $h_{\min}$ in 3.1 will satisfy it. In this case, the direction of instability used in 3.22 should take this condition into account. An estimation of the constant C_{inv} can be found in [67]. Defining $\widetilde{s} = s\,(1 - \tau s)$ and the discrete norm

$$\|u_h\|_\tau^2 = \varepsilon \|\boldsymbol{\nabla} u_h\|_h^2 + \left\|\widetilde{s}^{1/2} u_h\right\|_h^2 + \left\|\tau^{1/2} \boldsymbol{a} \cdot \nabla u_h\right\|_h^2$$

we have the following

Lemma 1 *(stability) Assume that the stabilization parameter satisfies condition 3.24. Then, there exists a constant $C > 0$ such that*

$$B_\tau\,(u_h, u_h) \geq C\,\|u_h\|_\tau^2$$

Proof. Taking $v_h = u_h$ in 3.6 and taking into account the skewsymmetry of the convective term we have

$$B_\tau\,(u_h, u_h) \geq \varepsilon\,\|\boldsymbol{\nabla} u_h\|_\Omega^2 + s\,\|u_h\|_\Omega^2 + \left\|\tau^{1/2} \boldsymbol{a} \cdot \nabla u_h\right\|_h^2 - \left\|\tau^{1/2}\left(-\varepsilon \nabla^2 u_h + s u_h\right)\right\|_h^2$$

As

$$\left\|-\varepsilon \nabla^2 u_h + s u_h\right\|_K^2 \leq \left\|\varepsilon \nabla^2 u_h\right\|_K^2 + \|s u_h\|_K^2 + 2\left\|\varepsilon \nabla^2 u_h\right\|_K \|s u_h\|_K$$

using the inverse estimate 3.23 and that for any $\alpha > 0$ we have $-2xy \geq -\frac{1}{\alpha} x^2 - \alpha y^2$, we arrive to

$$\begin{aligned} B_\tau\,(u_h, u_h) \;\geq\; & \varepsilon \left\| \left(1 - \tau \frac{\varepsilon C_{\text{inv}}^2}{h_{\min}^2} - \tau s \frac{1}{\alpha}\right)^{1/2} \boldsymbol{\nabla} u_h \right\|_h^2 \\ & + s \left\| \left(1 - \tau s - \alpha \tau \frac{\varepsilon C_{\text{inv}}^2}{h_{\min}^2}\right)^{1/2} u_h \right\|_h^2 + \left\|\tau^{1/2} \boldsymbol{a} \cdot \nabla u_h\right\|_K^2 \end{aligned}$$

Note that

$$1 - \tau \frac{\varepsilon C_{\text{inv}}^2}{h_{\min}^2} - \tau s \frac{1}{\alpha} \geq C_1$$

iff

$$(1 - C_1)\,\tau^{-1} \geq \frac{\varepsilon C_{\text{inv}}^2}{h_{\min}^2} + \frac{1}{\alpha} s$$

what is implied by 3.24 when $C_1 = 1 - 1/\alpha$ and $\alpha = 2$. In the same way

$$1 - \tau s - \alpha \tau \frac{\varepsilon C_{\text{inv}}^2}{h_{\min}^2} \geq C_2\,(1 - \tau s)$$

iff

$$(1 - C_2)\left(\tau^{-1} - s\right) \geq \alpha \frac{\varepsilon C_{\text{inv}}^2}{h_{\min}^2}$$

what is again implied by 3.24 when $C_2 < 1/2$. Therefore, the result holds for $C = min\,(C_1, C_2, 1)$. ■

Let us now consider $\widehat{u}_h$ an interpolant of the solution of the continuous problem u and define the interpolation error $\eta = u - \widehat{u}_h$. We will present now a bound of $B_\tau(\eta, v_h)$ in terms of a function of the interpolation error $E(\eta)$ defined by

$$\begin{aligned} E(\eta) &= \varepsilon^{1/2} \|\boldsymbol{\nabla}\eta\|_h + 2\left\|\tau^{1/2}\boldsymbol{a}\cdot\nabla\eta\right\|_h + 2\left\|\widetilde{s}^{1/2}\eta\right\|_h \\ &\quad +2\varepsilon\left\|\tau^{1/2}\nabla^2\eta\right\|_h + \left\|\left(\tau^{-1}-s\right)^{1/2}\eta\right\|_h + \left\|\frac{\tau s}{\widetilde{s}^{1/2}}\boldsymbol{a}\cdot\boldsymbol{\nabla}\eta\right\|_h \end{aligned}$$

In turn, this function can be bounded relying on some result from interpolation theory, although we will not consider this type of bound here.

Lemma 2 *Assume that the stabilization parameter satisfies condition 3.24. Then*

$$\|\eta\|_\tau \leq E(\eta)$$

and

$$B_\tau(\eta, v_h) \leq E(\eta)\|v_h\|_\tau$$

Proof. The first inequality is evident. To prove the second one we start from the definition

$$\begin{aligned} B_\tau(\eta, v_h) &= (\varepsilon\boldsymbol{\nabla}\eta, \boldsymbol{\nabla}v_h)_h && (3.25)\\ &\quad + (\boldsymbol{a}\cdot\nabla\eta, v_h)_h && (3.26)\\ &\quad + (s\eta, v_h)_h && (3.27)\\ &\quad + \left(-\varepsilon\nabla^2\eta, \tau\varepsilon\nabla^2 v_h\right)_h && (3.28)\\ &\quad + \left(-\varepsilon\nabla^2\eta, \tau\boldsymbol{a}\cdot\nabla v_h\right)_h && (3.29)\\ &\quad - \left(-\varepsilon\nabla^2\eta, \tau s v_h\right)_h && (3.30)\\ &\quad + \left(\boldsymbol{a}\cdot\nabla\eta, \tau\varepsilon\nabla^2 v_h\right)_h && (3.31)\\ &\quad + (\boldsymbol{a}\cdot\nabla\eta, \tau\boldsymbol{a}\cdot\nabla v_h)_h && (3.32)\\ &\quad - (\boldsymbol{a}\cdot\nabla\eta, \tau s v_h)_h && (3.33)\\ &\quad + \left(s\eta, \tau\varepsilon\nabla^2 v_h\right)_h && (3.34)\\ &\quad + (s\eta, \tau\boldsymbol{a}\cdot\nabla v_h)_h && (3.35)\\ &\quad - (s\eta, \tau s v_h)_h && (3.36) \end{aligned}$$

In order to bound these 12 terms we will use that

$$\frac{\tau^{1/2}\varepsilon^{1/2}C_{\text{inv}}}{h_{\min}} < 1 \tag{3.37}$$

and that

$$\tau^{1/2}s^{1/2}\frac{\varepsilon^{1/2}C_{\text{inv}}}{h_{\min}} \leq s^{1/2}(1-\tau s)^{1/2} \tag{3.38}$$

which are implied by 3.24. Also as $0 < 1 - \tau s < 1$ we have

$$1 - \tau s \leq (1 - \tau s)^{1/2} = \tau^{1/2} \left(\tau^{-1} - s\right)^{1/2} \tag{3.39}$$

Term 3.25 can be bounded as

$$(\varepsilon \boldsymbol{\nabla} \eta, \boldsymbol{\nabla} v_h)_h \leq \varepsilon \left\| \boldsymbol{\nabla} \eta \right\|_h \left\| \boldsymbol{\nabla} v_h \right\|_h \leq \varepsilon^{1/2} \left\| \boldsymbol{\nabla} \eta \right\|_h \left\| v_h \right\|_\tau$$

Term 3.32 can be bounded as

$$(\boldsymbol{a} \cdot \nabla \eta, \tau \boldsymbol{a} \cdot \nabla v_h)_h \leq \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla \eta \right\|_h \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla v_h \right\|_h \leq \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla \eta \right\|_h \left\| v_h \right\|_\tau$$

Terms 3.27 and 3.36 can be bounded as

$$(s\eta, v_h)_h - \tau \left(s\eta, s v_h\right)_h = (\eta, \widetilde{s} v_h)_h \leq \left\| \widetilde{s}^{1/2} \eta \right\|_h \left\| \widetilde{s}^{1/2} v_h \right\|_h \leq \left\| \widetilde{s}^{1/2} \eta \right\|_h \left\| v_h \right\|_\tau$$

Term 3.28 can be bounded using 3.37 as

$$\begin{aligned} \left(-\varepsilon \nabla^2 \eta, \tau \varepsilon \nabla^2 v_h\right)_h &\leq \left\| \tau^{1/2} \varepsilon \nabla^2 \eta \right\|_h \left\| \tau^{1/2} \varepsilon \nabla^2 v_h \right\|_h \\ &\leq \left\| \tau^{1/2} \varepsilon \nabla^2 \eta \right\|_h \left\| \frac{\tau^{1/2} \varepsilon C_{\text{inv}}}{h} \boldsymbol{\nabla} v_h \right\|_h \\ &\leq \left\| \tau^{1/2} \varepsilon \nabla^2 \eta \right\|_h \left\| \varepsilon^{1/2} \boldsymbol{\nabla} v_h \right\|_h \leq \left\| \tau^{1/2} \varepsilon \nabla^2 \eta \right\|_h \left\| v_h \right\|_\tau \end{aligned}$$

Term 3.29 can be bounded as

$$\left(-\varepsilon \nabla^2 \eta, \tau \boldsymbol{a} \cdot \nabla v_h\right)_h \leq \left\| \tau^{1/2} \varepsilon \nabla^2 \eta \right\|_h \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla v_h \right\|_h \leq \left\| \tau^{1/2} \varepsilon \nabla^2 \eta \right\|_h \left\| v_h \right\|_\tau$$

Term 3.31 can be bounded using 3.37 as

$$\begin{aligned} \left(\boldsymbol{a} \cdot \boldsymbol{\nabla} \eta, \tau \varepsilon \nabla^2 v_h\right)_h &\leq \left\| \tau^{1/2} \boldsymbol{a} \cdot \boldsymbol{\nabla} \eta \right\|_h \left\| \tau^{1/2} \varepsilon \nabla^2 v_h \right\|_h \leq \left\| \tau^{1/2} \boldsymbol{a} \cdot \boldsymbol{\nabla} \eta \right\|_K \left\| \frac{\tau^{1/2} \varepsilon C_{\text{inv}}}{h} \boldsymbol{\nabla} v_h \right\|_K \\ &\leq \left\| \tau^{1/2} \boldsymbol{a} \cdot \boldsymbol{\nabla} \eta \right\|_h \left\| \varepsilon^{1/2} \boldsymbol{\nabla} v_h \right\|_h \leq \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla \eta \right\|_h \left\| v_h \right\|_\tau \end{aligned}$$

Term 3.34 can be bounded using 3.38 as

$$\begin{aligned} \left(s\eta, \tau \varepsilon \nabla^2 v_h\right)_h &\leq \sum_{K \in \mathcal{P}_h} \tau s \varepsilon \left\| \eta \right\|_K \left\| \nabla^2 v_h \right\|_K \\ &\leq \sum_{K \in \mathcal{P}_h} (\tau s)^{1/2} \frac{(\tau s)^{1/2} \varepsilon^{1/2} C_{\text{inv}}}{h} \left\| \eta \right\|_K \left\| \varepsilon^{1/2} \boldsymbol{\nabla} v_h \right\|_K \\ &\leq \sum_{K \in \mathcal{P}_h} s^{1/2} (1 - \tau s)^{1/2} \left\| \eta \right\|_K \left\| \varepsilon^{1/2} \boldsymbol{\nabla} v_h \right\|_K \leq \left\| \widetilde{s}^{1/2} \eta \right\|_h \left\| v_h \right\|_\tau \end{aligned}$$

Terms 3.26 and 3.35 are bounded integrating by parts the convective term 3.26 and using 3.39 as

$$\begin{aligned} (\boldsymbol{a} \cdot \nabla \eta, v_h)_\Omega + (s\eta, \tau \boldsymbol{a} \cdot \nabla v_h)_h &= -\left(\eta, \boldsymbol{a} \cdot \nabla v_h\right)_\Omega + \left(s\eta, \tau \boldsymbol{a} \cdot \nabla v_h\right)_h \\ &= -\left(\tau^{-1/2} (1 - \tau s) \eta, \tau^{1/2} \boldsymbol{a} \cdot \nabla v_h\right)_h \\ &\leq \left\| \tau^{-1/2} (1 - \tau s) \eta \right\|_h \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla v_h \right\|_h \\ &\leq \left\| \left(\tau^{-1} - s\right)^{1/2} \eta \right\|_h \left\| \tau^{1/2} \boldsymbol{a} \cdot \nabla v_h \right\|_h \\ &\leq \left\| \left(\tau^{-1} - s\right)^{1/2} \eta \right\|_h \left\| v_h \right\|_\tau \end{aligned}$$

Term 3.33 can be bounded as

$$\begin{aligned} -\left(\boldsymbol{a}\cdot\nabla\eta,\tau s v_h\right)_h &= -\left(\frac{\tau s}{\widetilde{s}^{1/2}}\boldsymbol{a}\cdot\nabla\eta,\widetilde{s}^{1/2}v_h\right)_h \leq \left\|\frac{\tau s}{\widetilde{s}^{1/2}}\boldsymbol{a}\cdot\nabla\eta\right\|_h \left\|\widetilde{s}^{1/2}v_h\right\|_h \\ &\leq \left\|\frac{\tau s}{\widetilde{s}^{1/2}}\boldsymbol{a}\cdot\nabla\eta\right\|_h \|v_h\|_\tau \end{aligned}$$

The result is obtained grouping terms. ■

Using these results we can prove convergence using Céa's lemma. The result is the following

Theorem 1 *Assume that the stabilization parameter satisfies condition 3.24. Then*

$$\|u-u_h\|_\tau \leq E(\eta) = E(u-\widehat{u}_h)$$

where E is the function defined above.

Let us close this section with the following

Remark 3 *The only condition needed to prove convergence in the anisotropic case is 3.24. After satisfying this condition, there is still some freedom for the selection of the stabilization parameter (and in the case of linear elements this condition is satisfied for any definition). Therefore, the difference between the definition 3.1 or 3.22 is the norm in which this convergence is proved and the form of the function $E(\eta)$ (which depends on τ). In the isotropic case this estimate is optimal (see the discussion of [28] about the norm $\|\cdot\|_\tau$). In the anisotropic case we would need appropriate interpolation estimates to decide about this optimality. However, numerical experiments will show the convenience of choosing 3.22.*

3.6 Numerical examples

In this section we present numerical examples illustrating the behavior of the method proposed. The first two of them illustrate the behavior of the method on anisotropic meshes, the third one shows the importance of satisfying the restriction imposed by the error analysis when elements of order higher than one are used and the last one how the method behaves on isotropic but unstructured meshes.

3.6.1 Convection diffusion under anisotropic refinement

In this subsection we consider a convection diffusion problem ($s = 0$) on the domain $\Omega = [0,1]\times[0,1]$ with zero Dirichlet boundary conditions on $\partial\Omega$ and a force $f = 1$. We consider a diffusion coefficient $\varepsilon_{ij} = \varepsilon\delta_{ij}$ where $\varepsilon = 10^{-4}$ and three different velocities

1. $\boldsymbol{a} = (1,0)$

2. $\boldsymbol{a} = (0, 1)$

3. $\boldsymbol{a} = \left(\sqrt{2}/2, \sqrt{2}/2\right)$

A reference solution of this problem was found using a mesh of 100×100 elements refined according to the velocity. In case 1 it was refined in the direction x near the right wall and was uniform in the direction y; in the second case it was uniform in the direction x and refined in the direction y near the upper wall and in the third case it was refined in both directions near the right and upper walls. The smallest element size is about 2.5×10^{-5}. The results for the three cases are shown in figure 3.3. Note the presence of a strong boundary layer on the right wall in the case 1, on the upper wall in the case 2 and on both the upper and right walls in the case 3.

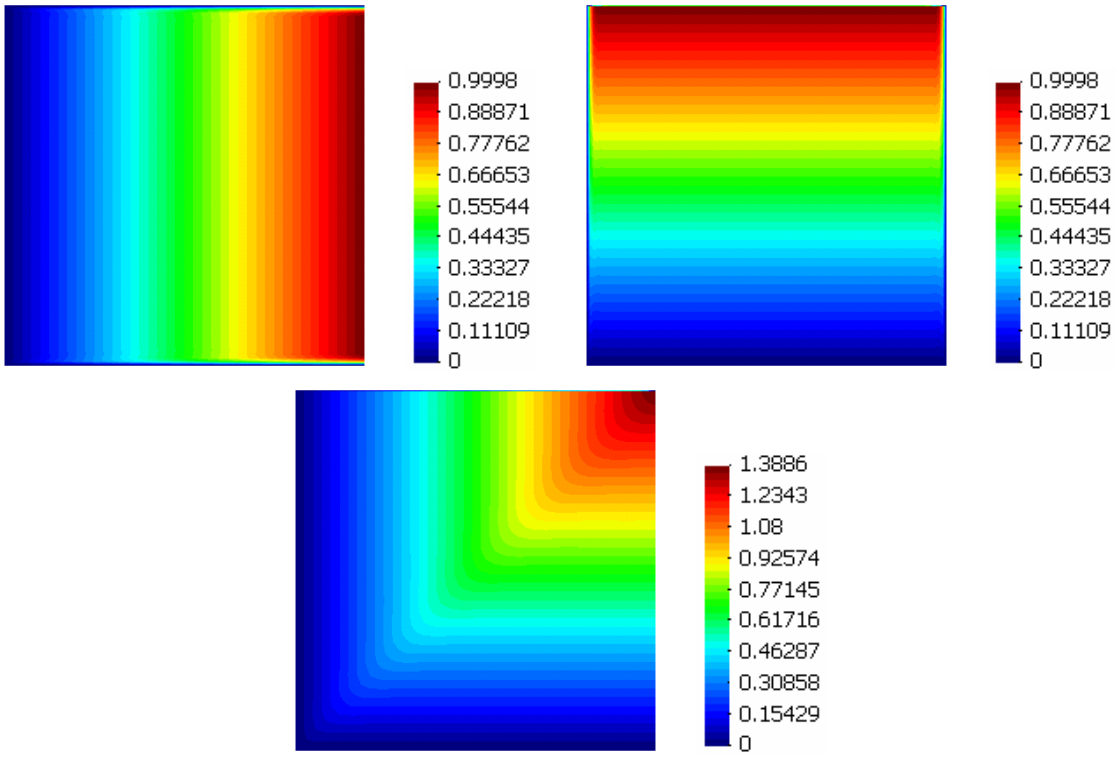

Figure 3.3: Reference solutions

For each case, the problem was also solved using a uniform mesh of 10×10 elements using the definition of the stabilization parameter given by 3.1 taking h as $h_{\min}$ (the minimum element length), as $h_{\max}$ (the maximum element length) and as h_a (the streamline element length) and also using expression 3.22 which is what we propose here. When the elements of the mesh are squares the definition given by 3.1 yields the same result taking h as $h_{\min}$ or $h_{\max}$ or h_a (when $h_1 = h_2 = h$ expressions 3.18 and 3.19 give $h_\varepsilon = h_a = h$). In these cases also expression 3.22 gives a similar result. These results are shown in figure 3.4

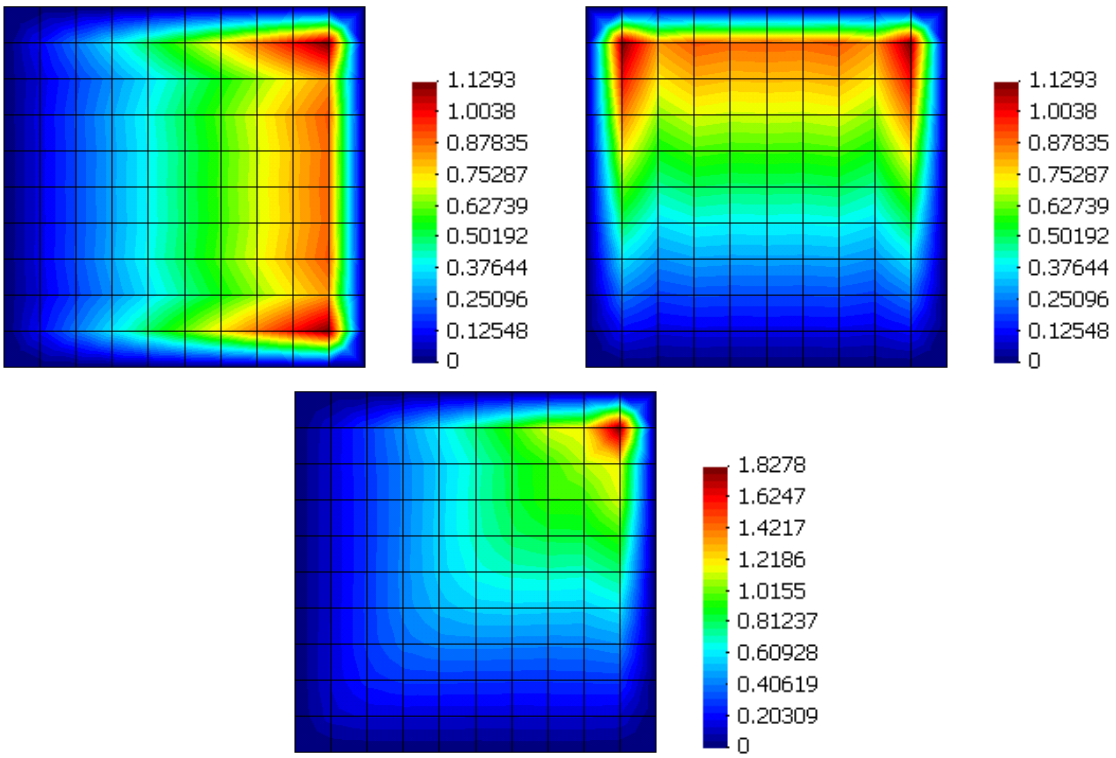

Figure 3.4: Solutions obtained using a mesh of 10×10 elements

Then the behavior of the method with respect to the mesh aspect ratio was analyzed. To this end, the problem was solved using meshes of 10×10 and also 100×10, 1000×10, 10000×10 giving aspect ratios A $= 10^0, 10^1, 10^2, 10^3$. To analyze the results we plot the unknown along the line $y = 0.5$ in the case 1, along the line $x = 0.5$ in the case 2 and along the lines $x = 0.9$ and $y = 0.9$ in the case 3. The results using the stabilization parameter defined by 3.1 taking h as $h_{\min}$ are shown in figure 3.5, those obtained taking h as $h_{\max}$ are shown in figure 3.6, those obtained using h_a are shown in figure 3.7 and those obtained using the stabilization parameter defined by 3.22 are shown in figure 3.8.

The minimum requirement we should pose to evaluate the behavior of a method is that the solution obtained using any anisotropic grid cannot be worse than the solution obtained using the 10×10 grid, or in other words, we should require that the solution can not get worse when the dimension of the finite element space is increased in a nested way. This is what happens if we use the stabilization parameter defined by 3.1 taking h as $h_{\min}$ or as h_a. In the first case, the solution obtained in case 1 is improved but in cases 2 and 3 numerical oscillations appear when the stretching factor increases. In the second case the solution obtained in cases 1 and 2 is improved but in case 3 numerical oscillations appear when the mesh is anisotropically refined. On the other hand, the solution obtained using the stabilization parameter defined by 3.1 taking h as $h_{\max}$ or the solution obtained using the expression 3.22 satisfy this requirement. Both methods give similar results in cases

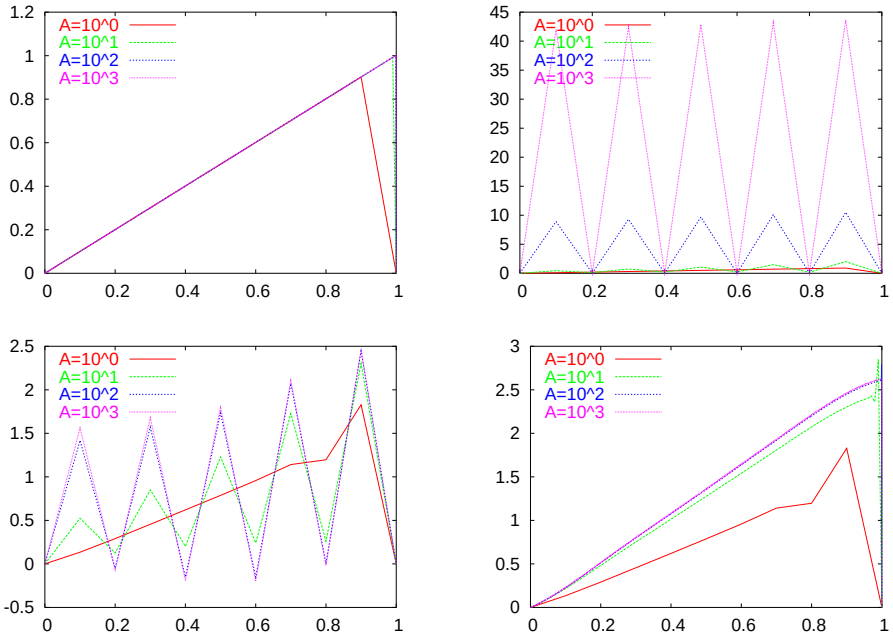

Figure 3.5: Solutions obtained using 3.1 with $h_{\min}$ in the case 1 (top left), in the case 2 (top right) and in the case 3 (bottom).

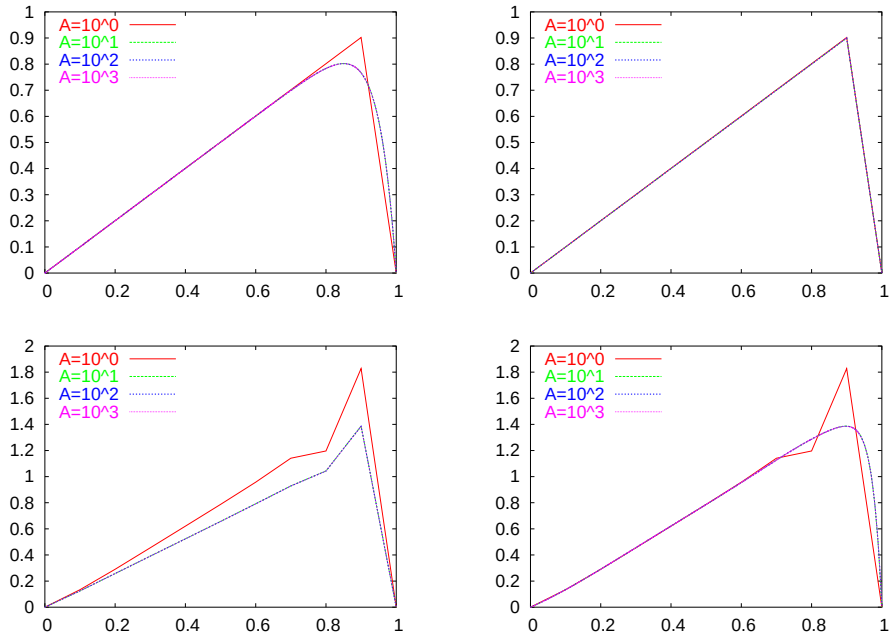

Figure 3.6: Solutions obtained using 3.1 with $h_{\max}$ in the case 1 (top left), in the case 2 (top right) and in the case 3 (bottom).

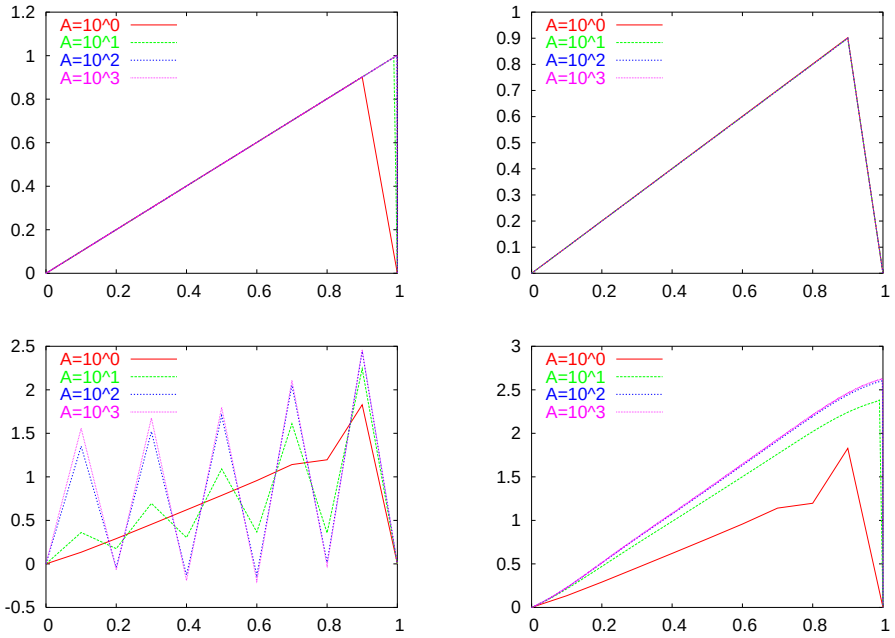

Figure 3.7: Solutions obtained using 3.1 with h_a in the case 1 (top left), in the case 2 (top right) and in the case 3 (bottom).

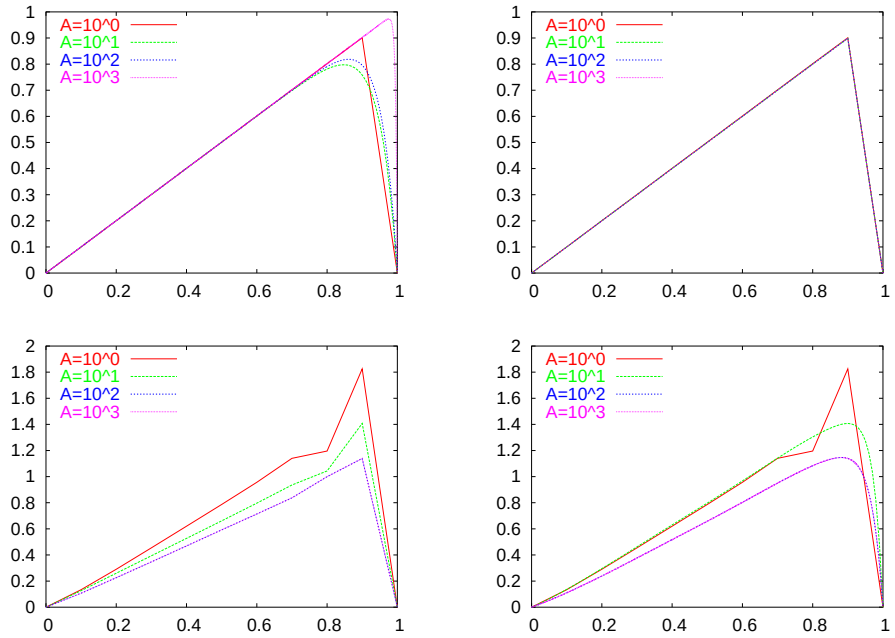

Figure 3.8: Solutions obtained using 3.22 in the case 1 (top left), in the case 2 (top right) and in the case 3 (bottom).

2 and 3 but only when the stabilization parameter proposed here is used the solution in case 1 is improved when the grid is refined. The method defined by 3.1 with h as $h_{\max}$ does take advantage of the new points added in the direction x even if the solution has a boundary layer on the right wall. Let us finally remark that in some cases the solution obtained using 3.1 taking h as $h_{\min}$ or as h_a can give a better solution than the method proposed here, as occurs in case 1, even if they present numerical oscillations in other cases.

3.6.2 Diffusion reaction under anisotropic refinement

In this subsection we consider a diffusion reaction problem on the domain $\Omega = [0,1]\times[0,1]$ with zero Dirichlet boundary conditions on $\partial\Omega$ and a force $f = 40$. We consider again $\varepsilon_{ij} = \varepsilon\delta_{ij}$ where $\varepsilon = 10^{-4}$ and a reaction $s = 40$. The problem is solved using meshes of 10×10 elements and also 100×10, 1000×10, 10000×10 elements, giving aspect ratios $\mathrm{A} = 10^0, 10^1, 10^2, 10^3$. To analyze the results, we plot the unknown along the line $y = 0.5$ and along the line $x = 0.5$. In this case, the results obtained using the stabilization parameter defined by 3.1 taking h as $h_{\max}$ and those obtained using 3.22 are the same. Therefore, we compare the results obtained using 3.1 taking h as $h_{\min}$ shown in figure 3.9 to those obtained taking h as $h_{\max}$ shown in figure 3.10.

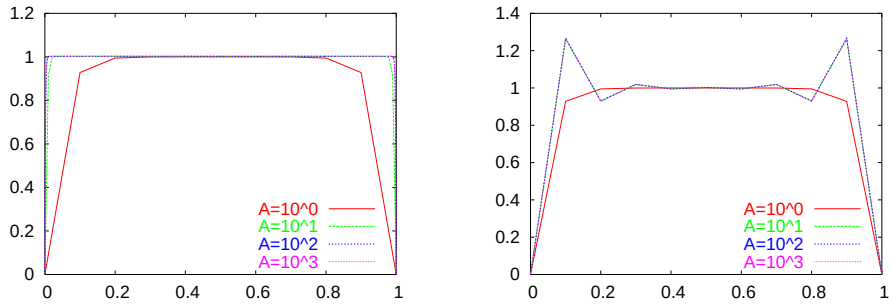

Figure 3.9: Solution obtained using 3.1 with $h_{\min}$ along the lines $y = 0.5$ and $x = 0.5$.

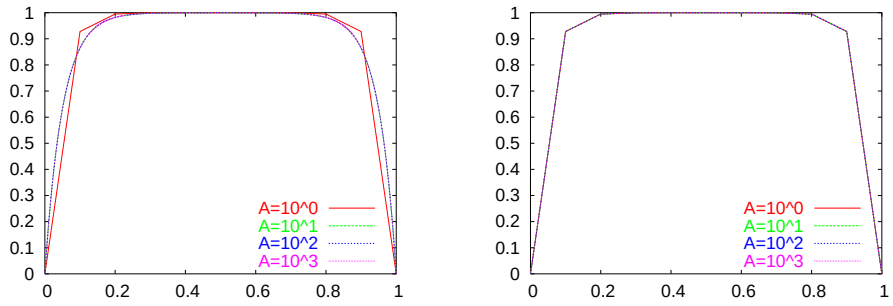

Figure 3.10: Solution obtained using 3.1 with $h_{\max}$ along the lines $x = 0.5$ and $y = 0.5$.

As in the convection diffusion problem shown in the previous subsection, the result obtained using 3.1 taking h as $h_{\min}$ shows numerical oscillations when the mesh is anisotropically refined whereas the results obtained using 3.1 taking h as $h_{\max}$ do not change. Both results are compared in figure 3.11 where the solutions obtained using the mesh of 100×10 elements are shown.

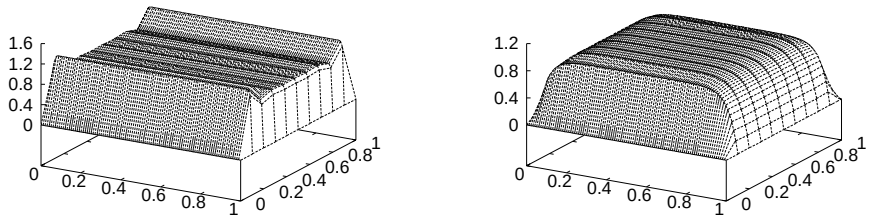

Figure 3.11: Solution obtained using 3.1 with $h_{\min}$ (left) and using 3.1 with $h_{\max}$ (right)

3.6.3 The Poisson problem using quadratic elements

In this subsection we consider a pure diffusive Poisson problem which corresponds to the CDR problem in the limit of vanishing convection and reaction. The domain considered is $\Omega = [0,1] \times [0,1]$ and zero Dirichlet boundary conditions on $\partial\Omega$ are prescribed. In order to activate instabilities we introduce a forcing term that gives a solution of the form

$$u(x,y) = \left(1 + e^{-\alpha} - e^{-\alpha x} - e^{\alpha(x-1)}\right)\left(1 + e^{-\alpha} - e^{-\alpha y} - e^{\alpha(y-1)}\right)$$

which presents boundary layers on the domain boundary whose width can be controlled using the parameter α. We solve the problem using a uniform mesh of 10×10 biquadratic elements and expression 3.1 for different values of c_1. The results are shown in figure 3.12.

For biquadratic elements $C_{\text{inv}} = 24$ [67] and it can be observed that when condition 3.24 is not satisfied numerical oscillations appear. Note that the Galerkin method is recovered when $c_1 \to \infty$

3.6.4 A convection diffusion reaction problem on isotropic meshes.

In this subsection we consider a convection diffusion problem on the domain $\Omega = [0,1] \times [0,1]$ with zero Dirichlet boundary conditions on $\partial\Omega$ and a force $f = 20$. The equation coefficients are $\varepsilon_{ij} = \varepsilon\delta_{ij}$ where $\varepsilon = 10^{-2}$, $s = 20$ and $\boldsymbol{a} = (3,2)$. We solve the problem on different meshes:

- Case 2 structured triangular elements of size 0.1 titled to the right /

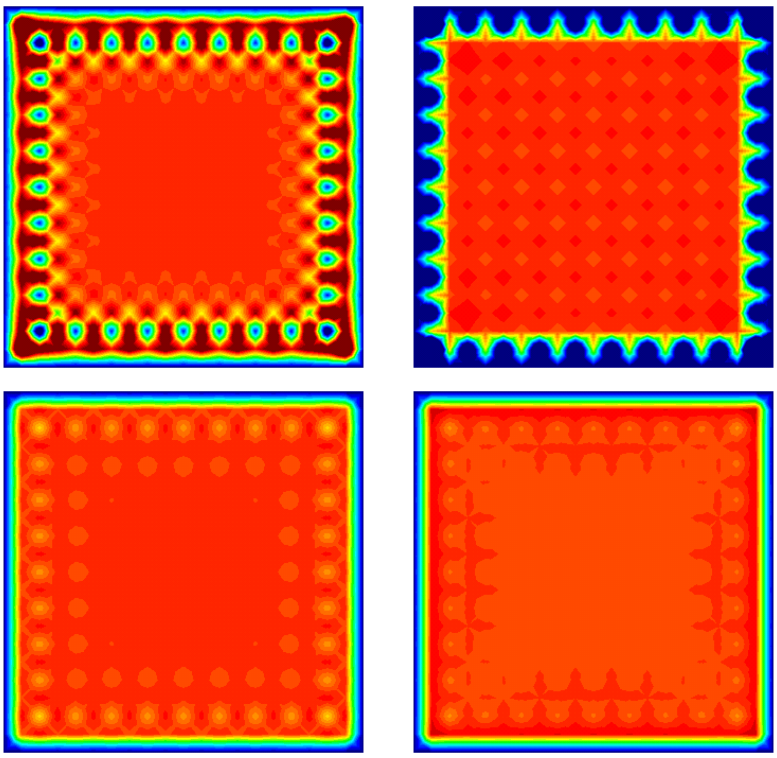

Figure 3.12: Solution to the Poisson problem obtained using 3.1 with $c_1 = 16$ (top left), with $c_1 = 24$ (top right), with $c_1 = 48$ (bottom left) and with $c_1 = 96$ (bottom right).

- Case 3: structured triangular elements of size 0.1 titled to the left \
- Case 4: unstructured triangular elements of size 0.1
- Case 5 structured 10×10 square elements
- Case 6: unstructured square elements of size 0.1

In any of these cases the mesh size is around 0.1 but it varies slightly according to the mesh design. In the case of triangular elements the element lengths are calculated as $h_1 = J_{1k}^{-t} J_{1k}^{-t}$ and $h_2 = J_{2k}^{-t} J_{2k}^{-t}$. The dimensionless numbers of the problem are given by

$$\mathrm{D} = \frac{sh^2}{4\varepsilon} = 5$$

$$\mathrm{P} = \frac{ah}{2\varepsilon} \sim 18$$

A reference solution (case 1) was computed on a 200×200 mesh for which these numbers are

$$\mathrm{D} = 0.0125$$

$$\mathrm{P} = 0.9$$

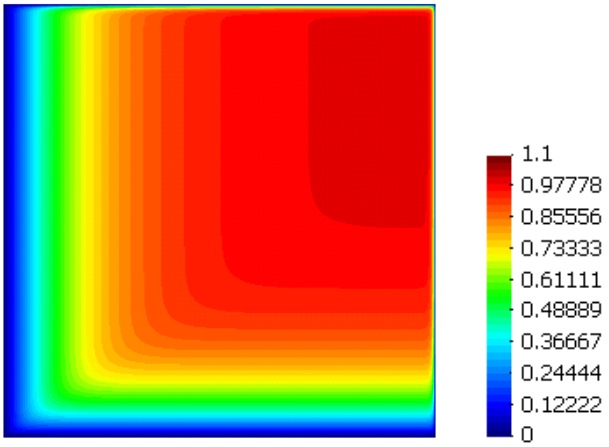

Figure 3.13: Reference solution

The result is shown in figure 3.13. The maximum value obtained is $u_{\max} = 0.99807$.

For each of these cases we compare the results obtained using the multiscale formulation using four possible definitions of the stabilization parameter. The first three are given by 3.1 taking the length h as the minimum, the maximum and the length in the velocity direction. The fourth is the definition given by 3.22. Table 3.1 shows the maximum values obtained for each case and method.

	Case 2 (ST /)	Case 3 (ST \)	Case 4 (UT)	Case 5 (SQ)	Case 6 (UQ)
3.1 using $h_{\min}$	1.3841	1.2712	1.2052	1.2973	1.3543
3.1 using $h_{\max}$	1.1915	1.1486	1.1318	1.2973	1.2332
3.1 using h_a	1.0281	1.3154	1.2437	1.2973	1.4191
3.22	0.9639	0.9762	0.9773	1.0828	1.0248

Table 3.1: Maximum values obtained

Figures 3.14 to 3.18 show contours for each case and method, all given in the same scale as figure 3.13.

3.7 Conclusions

The definition of the stabilization parameters in the case of the scalar convection diffusion reaction problem has been revisited. The variational multiscale method provides a natural framework to understand the problem. Starting from this point and introducing a transformation of the fine scale problem to the reference domain the dependence of the stabilization parameters on the equation coefficients and element length (through the Jacobian of such transformation) has been identified. A deeper inspection of the Fourier argument presented in [29] permitted to obtain an exact representation of the

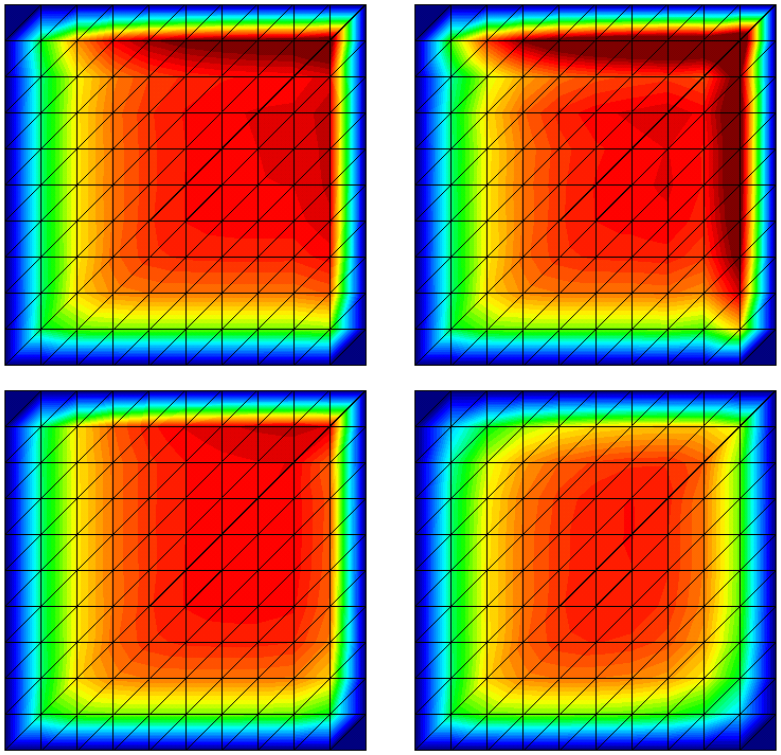

Figure 3.14: Results obtained in case 2 using 3.1 with $h_{\max}$ (top left), with $h_{\min}$ (top right), with h_a (bottom left) and using 3.22 (bottom right).

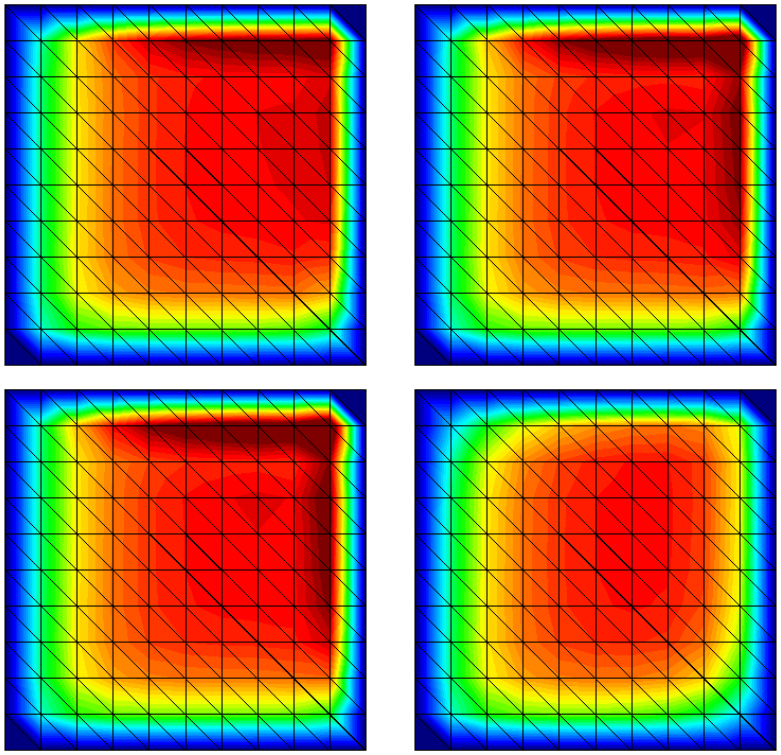

Figure 3.15: Results obtained in case 3 using 3.1 with $h_{\max}$ (top left), with $h_{\min}$ (top right), with h_a (bottom left) and using 3.22 (bottom right).

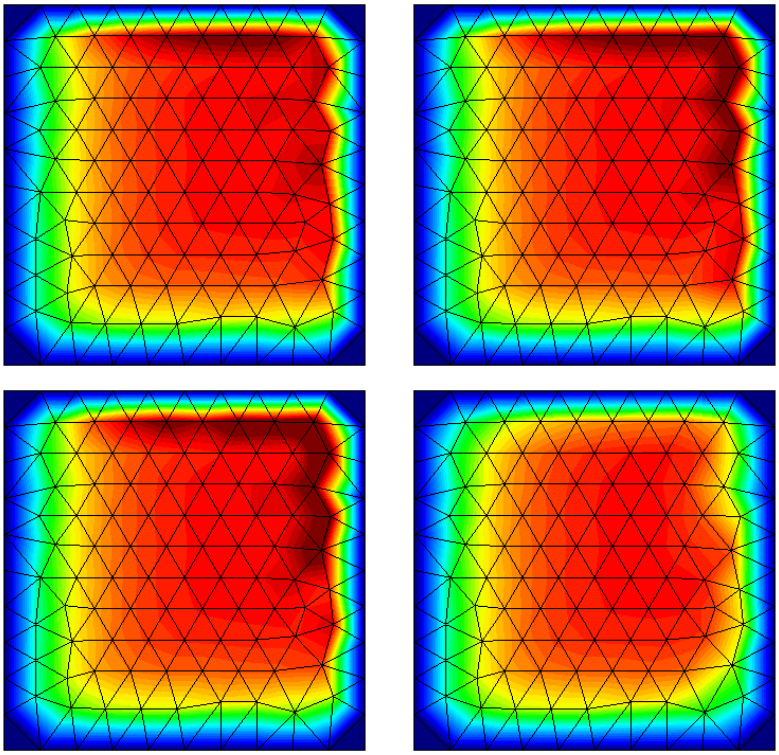

Figure 3.16: Results obtained in case 4 using 3.1 with $h_{\max}$ (top left), with $h_{\min}$ (top right), with h_a (bottom left) and using 3.22 (bottom right).

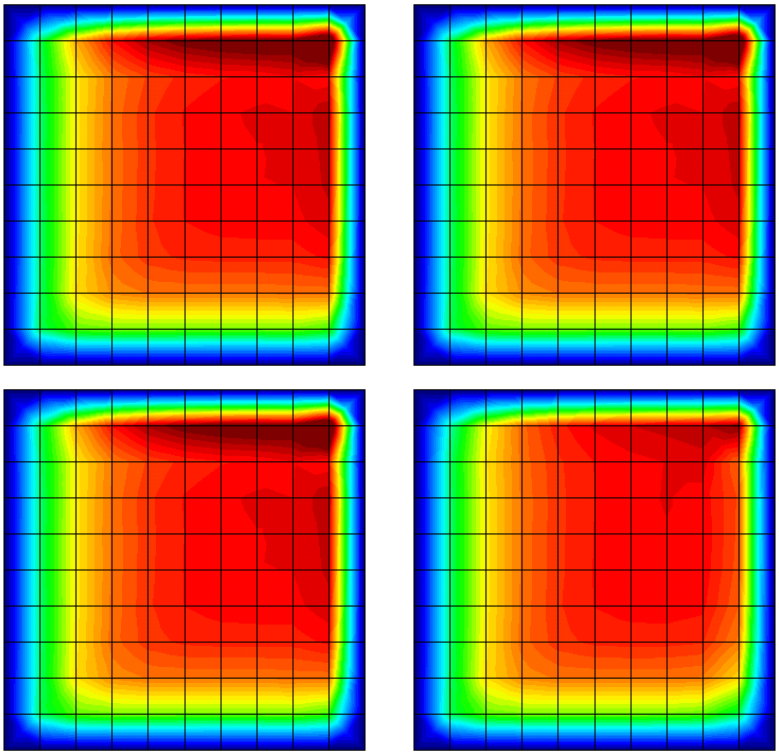

Figure 3.17: Results obtained in case 5 using 3.1 with $h_{\max}$ (top left), with $h_{\min}$ (top right), with h_a (bottom left) and using 3.22 (bottom right).

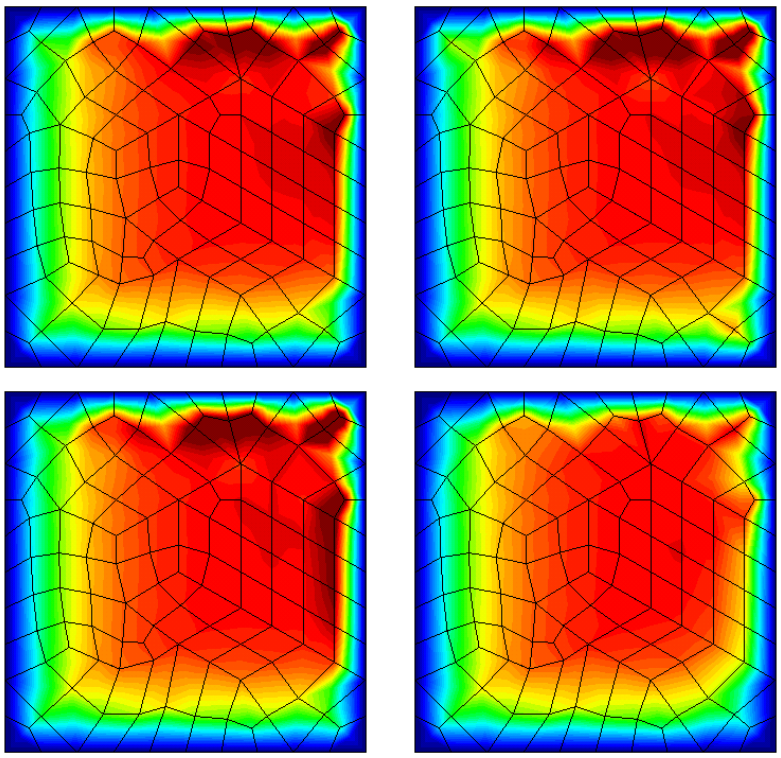

Figure 3.18: Results obtained in case 6 using 3.1 with $h_{\max}$ (top left), with $h_{\min}$ (top right), with h_a (bottom left) and using 3.22 (bottom right).

Green function 3.10 and a first approximation to it. The well known exact solution to the one dimensional problem has been used to find the constants of the parameter in a very natural way. Finally, the direction of the wave vector $\boldsymbol{k}$ has been identified as the direction of the maximum instability of the problem as the one of maximum

The expression proposed gives excellent results as numerical experiments have shown. In the case of linear elements, these experiments have also shown that when the usual expression 3.1 with $h = h_{\max}$ is used numerical oscillations do not appear. This conclusion is important for more complex problems where the direction of the instability could be difficult to find. When higher order elements are used condition 3.24 must be satisfied for the error analysis to be valid. This is confirmed by the results obtained in the case of the Poisson problem that present numerical oscillations if the condition is not satisfied. This also implies that it is not possible to stabilize reaction when higher order elements are used, at least in the anisotropic case. Although desirable, this stability is not essential as the oscillations are local thanks to the L^2 stability provided by the reactive term.

Chapter 4

The Oseen problem

In this chapter we present a new subgrid scale model for the Oseen equations in the context of the variational multiscale method. We extend the method of the previous chapter to systems of second order equations and two possible approximations of the solution of the fine scale problem are presented. Following the line of the previous chapter we evaluate the proposed model when anisotropic meshes are used. The stability of the linearized problem is proved and numerical examples illustrating the behavior of the method are provided.

4.1 Introduction

Although the incompressible Navier Stokes equations have been extensively studied, several points remain unclear. At the continuous level these problems involve uniqueness of weak solutions or global existence of strong solutions and are summarized in the Clay Institute Prize Problem. At the discrete level many different approximations have been proposed and several results have been established. All of them involve, in a way or another, the solution of the linearized problem. It is well known that these equations present different types of numerical instabilities.

In the first place we have the instability due to the dominance of the convective term over the viscous one in the high Reynolds number regime. This instability is also present in the scalar convection diffusion problem and it is well understood. A stable and accurate approximation to this problem has been presented in the previous chapter, where a new definition of the stabilization parameters has proven to give excellent results.

As a second problem, we have the pressure instability that may appear if the compatibility of the velocity and pressure spaces posed by the inf-sup condition is not satisfied. This instability is also present in the Stokes problem and is also well understood. It is not related to the dominance of a term in the equations but rather to the vectorial structure of the problem. When the Navier Stokes equations are written as a system

of second order equations, the pressure appears in the first order term and the diffusion matrix is not positive definite, but only semidefinite.

In this work we present a stabilized finite element formulation based on the subgrid-scale approach introduced in [75, 78] for the scalar convection–diffusion equation. The idea is to split the solution of the continuous problem φ into a finite element component φ_h and the difference $\widetilde{\varphi} = \varphi - \varphi_h$, called subscale, which cannot be reproduced by the finite element mesh. This splitting corresponds to a decomposition of the continuous space V as a direct sum of the finite element space V_h and a subgrid space $\widetilde{V}$ to be defined. The approximation of the problem projected onto $\widetilde{V}$, which is driven by the strong residual of the finite element problem, will give an approximated subscale $\widetilde{\varphi}^{\text{ap}}$ whose effect on the discrete problem for φ_h will be taken into account. Hopefully, this approximation will enhance the stability properties of the discrete problem projected on V_h, allowing the use of equal order velocity-pressure interpolations and the solution of convection dominated problems. This approach is a general framework in which it is possible to design different stabilized formulations depending on the approximation performed for solving the fine scale problem and on the selection of the space of subscales. After stating the problem in section 4.2, the approximation of the fine scale problem in presented in section 4.3. The whole process can be divided in three steps.

The first one consists in approximating the boundary conditions of the small scale problem on the edges of the finite element mesh in order to obtain uncoupled local problems posed on each element. It is common to assume that the subscales vanish on the element boundaries, but other possibilities could be considered and are currently under investigation.

The second step consists of some approximation of the inverse differential operator to write the subscales in terms of the residual of the finite element component. In the case of the scalar convection diffusion problem the inverse of the differential operator $\mathcal{L}$ is replaced by an algebraic operator τ that depends on the equation coefficient and on the finite element mesh, including its stretching, as shown in the previous chapter. In the case of the Stokes or Oseen equations the differential operator is of vectorial character and therefore so is its inverse. Nevertheless a diagonal matrix of stabilization parameters is commonly employed, although some efforts to understand the vectorial structure of the equations have been recently made in [121], where a stabilization matrix has been derived by dimensional analysis and the stability of the final method has been proved. Non diagonal approximations have also been proposed to consider anisotropic finite element approximations in [9, 8, 10] but they do not result from the approximation of the fine scale problem in the variational multiscale context and are rather ad hoc.

In this chapter we present two approximations to the inverse of the Oseen operator using ideas that can be applied to any second order system of equations. Both are extensions of the ideas presented in the previous chapter applied in different ways. The

first one deals with the whole operator and results in a condition for the design of the stabilization matrix. This permits to define a stabilization matrix taking into account the simplicity of the final method. Using this approach we recover the standard matrix diag $(\tau_m \boldsymbol{I}, \tau_c)$ (τ_m and τ_c are defined in section 4.3). The second approximation deals with each equation separately and naturally takes into account the coupling between variables. The result is not a stabilization matrix but a stabilization operator that consists in the usual algebraic term diag $(\tau_m \boldsymbol{I}, \tau_c)$ and some extra differential terms that couple equations. In both cases the anisotropy of the grid is incorporated in the definition of the scalar parameters as in chapter 3 and not in the coupling between equations as done before in [9, 8, 10].

The third step consists in imposing that the approximated subscale belongs to the selected subspace $\widetilde{V}$, what is done by a projection. The approach followed originally in [75, 78] and described also in [28], consists in taking the subscales directly proportional to the residual of the finite element component. In this case the space of subscales is the space of the residuals $\mathcal{L}V_h$ (when the force is a finite element function) and no projection is needed. Another approach, described in [29], is to take only the component of these residuals L^2 orthogonal to the finite element space. This idea was first introduced in [26] as an extension of a stabilization method originally introduced for the Stokes problem in [30].

In section 4.4 we also prove stability of the formulation and we will show that the extra differential terms in the second approximation, which are of high order, do not provide any extra stability (except for some control of $\nabla^2 \boldsymbol{\nabla} \cdot \boldsymbol{u}_h$) and must be controlled by the usual diagonal terms. Therefore, they can only be considered if a high order polynomial approximation is employed. How important these terms are is something that needs further research and that we leave for a future work. The point we want to evaluate here is how the anisotropy of the grid has to be taken into account and to do that we need to clarify how the relation between variables should be.

Finally, numerical experiments are presented in section 4.5 and conclusions are drawn in section 4.6.

4.2 Problem Statement

Let us start writing the Oseen equations with zero Dirichlet boundary conditions. To this end, let us consider the space of functions whose p power ($1 \leq p < \infty$) is integrable in a domain ω, denoted by $L^p(\omega)$, and the space of bounded functions in ω, denoted by $L^\infty(\omega)$. The space of functions whose distributional derivatives of order up to $m \geq 0$ (integer) belong to $L^2(\omega)$ is denoted by $H^m(\omega)$. The space $H_0^1(\omega)$ consists of functions in $H^1(\omega)$ vanishing on $\partial\omega$. The topological dual of $H_0^1(\omega)$ is denoted by $H^{-1}(\omega)$. A bold character is used to denote the vector counterpart of all these spaces. If f and g are functions (or

distributions) such that $f\,g$ is integrable in the domain ω under consideration, we denote

$$\langle f, g\rangle_\omega = \int_\omega f\, g \,\mathrm{d}\omega,$$

so that, in particular, $\langle\cdot,\cdot\rangle_\omega$ is the duality pairing between $H^{-1}(\omega)$ and $H^1_0(\omega)$. When $f, g \in L^2(\omega)$, we write the inner product as $\langle f, g\rangle_\omega \equiv (f, g)_\omega$ and the norm $(g, g)_\omega^{1/2}$ is denoted by $\|g\|_\omega$. Using this notation, the Oseen problem consists in finding the velocity field $\boldsymbol{u} \in \boldsymbol{V} := \boldsymbol{H}^1_0(\Omega)$, and the pressure field $p \in Q := L^2(\Omega)/\mathbb{R}$ such that

$$\begin{aligned} -\nu\nabla^2\boldsymbol{u} + \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}p &= \boldsymbol{f} \quad \text{in} \quad \Omega && (4.1)\\ \boldsymbol{\nabla}\cdot\boldsymbol{u} &= 0 \quad \text{in} \quad \Omega && (4.2)\\ \boldsymbol{u} &= \boldsymbol{0} \quad \text{on} \quad \Gamma && (4.3) \end{aligned}$$

where $\Gamma = \partial\Omega$ is the boundary of the domain $\Omega \subset \mathbb{R}^{\mathrm{d}}$, $(\mathrm{d} = 2, 3)$, $\boldsymbol{f} \in \boldsymbol{H}^{-1}(\Omega)$ is the external force, ν is the kinematic viscosity and $\boldsymbol{a} \in L^\infty(\Omega)$ is the given solenoidal advection velocity. This problem can be written in a weak form as follows: find $(\boldsymbol{u}, p) \in \boldsymbol{V} \times Q$ such that

$$B(\boldsymbol{u}, p; \boldsymbol{v}, q) = L(\boldsymbol{v}, q) \quad \forall\,(\boldsymbol{v}, q) \in \boldsymbol{V} \times Q \tag{4.4}$$

where

$$\begin{aligned} B(\boldsymbol{u}, p; \boldsymbol{v}, q) &= (\boldsymbol{\nabla}\boldsymbol{v}, \nu\boldsymbol{\nabla}\boldsymbol{u})_\Omega + (\boldsymbol{v}, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u})_\Omega - (\boldsymbol{\nabla}\cdot\boldsymbol{v}, p)_\Omega + (\boldsymbol{\nabla}\cdot\boldsymbol{u}, q)_\Omega\\ L(\boldsymbol{v}, q) &= \langle\boldsymbol{v}, \boldsymbol{f}\rangle_\Omega \end{aligned}$$

Let us consider the multiscale decomposition of the space $\boldsymbol{V}$ and Q

$$\boldsymbol{V} = \boldsymbol{V}_h \oplus \widetilde{\boldsymbol{V}}, \quad Q = Q_h \oplus \widetilde{Q}$$

where h is used to indicate spaces (and functions) constructed using a finite element partition of the domain $\mathcal{P}_h = \{K\}$ as

$$\begin{aligned} \boldsymbol{V}_h &= \left\{\boldsymbol{v}_h \in \boldsymbol{V} : \boldsymbol{v}_h \circ F^{-1}\big|_K \in \boldsymbol{P}_p\left(\widehat{K}\right)\right\}\\ Q_h &= \left\{w_h \in Q : w \circ F^{-1}\big|_K \in P_p\left(\widehat{K}\right)\right\} \end{aligned}$$

where $P_p\left(\widehat{K}\right)$ denotes the set of polynomials of degree at most p (on each space variable if quadrilateral/hexahedral elements are used) and F the affine mapping from the reference element $\widehat{K}$ to the physical element K. In the same way $\widetilde{}$ is used to indicate subgrid spaces (and functions) which are any completion of the finite element spaces to the continuous spaces. Applied to the weak form of the problem, this decomposition leads to

$$\begin{aligned} B(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h) + B(\widetilde{\boldsymbol{u}}, \widetilde{p}; \boldsymbol{v}_h, q_h) &= L(\boldsymbol{v}_h, q_h) \quad \forall\,(\boldsymbol{v}_h, q_h) \in \boldsymbol{V}_h \times Q_h && (4.5)\\ B(\boldsymbol{u}_h, p_h; \widetilde{\boldsymbol{v}}, \widetilde{q}) + B(\widetilde{\boldsymbol{u}}, \widetilde{p}; \widetilde{\boldsymbol{v}}, \widetilde{q}) &= L(\widetilde{\boldsymbol{v}}, \widetilde{q}) \quad \forall\,(\widetilde{\boldsymbol{v}}, \widetilde{q}) \in \widetilde{\boldsymbol{V}} \times \widetilde{Q} && (4.6) \end{aligned}$$

The first equation is the equation for the resolvable scales (the functions of the spaces $\boldsymbol{V}_h$ and Q_h) and has two terms: the first one is the Galerkin contribution and the second one takes into account the influence of the subgrid scale on the finite element components.

Let us introduce the following notation

$$\Omega^h = \bigcup_{K\in\mathcal{P}_h} K \quad \text{and} \quad \Gamma^h = \bigcup_{K\in\mathcal{P}_h} \partial K$$

and

$$(\cdot,\cdot)_h = \sum_{K\in\mathcal{P}_h} (\cdot,\cdot)_K\,, \quad (\cdot,\cdot)_{\partial h} = \sum_{K\in\mathcal{P}_h} (\cdot,\cdot)_{\partial K} \quad \text{and} \quad \|\cdot\|_h^2 = \sum_{K\in\mathcal{P}_h} \|\cdot\|_K^2$$

Integrating by parts within each element we have

$$(\boldsymbol{v}_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}))_\Omega = (\boldsymbol{v}_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h)_\Omega + (\boldsymbol{v}_h, \boldsymbol{a}\cdot\boldsymbol{n}\widetilde{\boldsymbol{u}})_{\partial h} - (\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{v}_h, \widetilde{\boldsymbol{u}})_h - (\boldsymbol{v}_h, \boldsymbol{\nabla}\cdot\boldsymbol{a}\widetilde{\boldsymbol{u}})_h$$

and

$$\begin{aligned}(\boldsymbol{\nabla}\boldsymbol{v}_h, \nu\boldsymbol{\nabla}\widetilde{\boldsymbol{u}})_\Omega &= (\nu\boldsymbol{n}\cdot\boldsymbol{\nabla}\boldsymbol{v}_h, \widetilde{\boldsymbol{u}})_{\partial h} - \left(\nu\nabla^2\boldsymbol{v}_h, \widetilde{\boldsymbol{u}}\right)_h \\ (q_h, \boldsymbol{\nabla}\cdot\widetilde{\boldsymbol{u}})_\Omega &= (q_h, \boldsymbol{n}\cdot\widetilde{\boldsymbol{u}})_{\partial h} - (\boldsymbol{\nabla}q_h, \widetilde{\boldsymbol{u}})_h\end{aligned}$$

Then, we can write the first equation 4.5 as

$$\begin{aligned}(\boldsymbol{\nabla}\boldsymbol{v}_h, \nu\boldsymbol{\nabla}\boldsymbol{u}_h)_\Omega + (\boldsymbol{v}_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h)_\Omega - (\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, p_h)_\Omega + (q_h, \boldsymbol{\nabla}\cdot\boldsymbol{u}_h)_\Omega \\ + (\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \widetilde{\boldsymbol{u}})_h - (\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, \widetilde{p})_h \underbrace{- (\boldsymbol{v}_h, \boldsymbol{\nabla}\cdot\boldsymbol{a}\widetilde{\boldsymbol{u}})_h}_{1} \\ \underbrace{+ (\boldsymbol{v}_h, \nu\boldsymbol{n}\cdot\boldsymbol{\nabla}\widetilde{\boldsymbol{u}})_{\partial h}}_{2} \underbrace{+ (\boldsymbol{v}_h, \boldsymbol{n}\cdot\boldsymbol{a}\widetilde{\boldsymbol{u}})_{\partial h}}_{3} \underbrace{+ (q_h, \boldsymbol{n}\cdot\widetilde{\boldsymbol{u}})_{\partial h}}_{4} &= \langle\boldsymbol{v}_h, f\rangle_\Omega\end{aligned} \tag{4.7}$$

for any $(\boldsymbol{v}_h, q_h) \in \boldsymbol{V}_h \times Q_h$, where $\mathcal{L}^*$ is the adjoint of the convection diffusion operator $\mathcal{L}$, defined as

$$\begin{aligned}\mathcal{L} &= -\nu\nabla^2 + \boldsymbol{a}\cdot\boldsymbol{\nabla} \\ \mathcal{L}^* &= -\nu\nabla^2 - \boldsymbol{a}\cdot\boldsymbol{\nabla}\end{aligned}$$

Let us remark that, up to this point, no approximation has been performed. Term 1 in 4.7 vanishes because $\boldsymbol{a}$ is assumed to be solenoidal. Terms 3 and 4 in 4.7 also vanishes thanks to the continuity of the subscales and test functions across interelement boundaries. Finally term 2 does not vanish and could be accounted for but will be neglected in this work.

Integrating again by parts within each element we have

$$\begin{aligned}(\boldsymbol{\nabla}\widetilde{\boldsymbol{v}}, \nu\boldsymbol{\nabla}\boldsymbol{u})_\Omega &= (\widetilde{\boldsymbol{v}}, \nu\boldsymbol{n}\cdot\boldsymbol{\nabla}\boldsymbol{u})_{\partial h} + \left(\widetilde{\boldsymbol{v}}, -\nu\nabla^2\boldsymbol{u}\right)_h \\ (\boldsymbol{\nabla}\cdot\widetilde{\boldsymbol{v}}, p)_\Omega &= (\widetilde{\boldsymbol{v}}, p\boldsymbol{n})_{\partial h} - (\widetilde{\boldsymbol{v}}, \boldsymbol{\nabla}p)_h\end{aligned}$$

Then, we can write the second equation 4.6 as

$$(\widetilde{\boldsymbol{v}}, \mathcal{L}\widetilde{\boldsymbol{u}} + \boldsymbol{\nabla}\widetilde{p})_h + (\boldsymbol{\nabla}\cdot\widetilde{\boldsymbol{u}}, \widetilde{q})_h + (\widetilde{\boldsymbol{v}}, -p\boldsymbol{n}+\nu\boldsymbol{n}\cdot\boldsymbol{\nabla}\boldsymbol{u})_{\partial h} = (\widetilde{\boldsymbol{v}}, \boldsymbol{R}_m)_h + (\widetilde{q}, R_c)_h$$

for any $(\widetilde{\boldsymbol{v}}, \widetilde{q}) \in \widetilde{\boldsymbol{V}} \times \widetilde{Q}$ where

$$\begin{aligned} \boldsymbol{R}_m &= \boldsymbol{f} - \mathcal{L}\boldsymbol{u}_h - \boldsymbol{\nabla}p_h \\ R_c &= -\boldsymbol{\nabla}\cdot\boldsymbol{u}_h \end{aligned}$$

are the residuals of the momentum and continuity equations. As the continuous tractions $\boldsymbol{\sigma} = -p\boldsymbol{n}+\nu\boldsymbol{n}\cdot\boldsymbol{\nabla}\boldsymbol{u}$ are continuous across any surface, the last term on the left hand side vanishes and the problem is equivalent to find $(\widetilde{\boldsymbol{u}}, \widetilde{p}) \in \widetilde{\boldsymbol{V}} \times \widetilde{Q}$ such that

$$\mathcal{L}\widetilde{\boldsymbol{u}} + \boldsymbol{\nabla}\widetilde{p} = \boldsymbol{R}_m + \widetilde{\boldsymbol{v}}^{\perp} \quad \text{in} \quad \Omega^h \tag{4.8}$$

$$\boldsymbol{\nabla}\cdot\widetilde{\boldsymbol{u}} = R_c + \widetilde{q}^{\perp} \quad \text{in} \quad \Omega^h \tag{4.9}$$

$$\widetilde{\boldsymbol{u}} = \boldsymbol{u}_{\text{ske}} \quad \text{on} \quad \Gamma^h \tag{4.10}$$

where $\boldsymbol{u}_{\text{ske}}$ is a function defined on the element boundaries and $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$ are any functions on the orthogonal complement of $\widetilde{\boldsymbol{V}}$ and $\widetilde{Q}$ respectively (in the $L^2\left(\Omega^h\right)$ sense). The function $\boldsymbol{u}_{\text{ske}}$ must be such that the exact tractions are continuous across element boundaries. In turn functions $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$ are responsible for guaranteeing that $\widetilde{\boldsymbol{u}} \in \widetilde{\boldsymbol{V}}$ and $\widetilde{p} \in \widetilde{Q}$. A modelling step is necessary to solve the system, what means a choice of $\boldsymbol{u}_{\text{ske}}$ and of $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$ and an approximate solution of 4.8-4.9. Note that 4.8-4.9 is posed in Ω^h, which consists in the union of the elements of the mesh. Therefore, any choice of $\boldsymbol{u}_{\text{ske}}$ leads to uncoupled problems posed on each element K. In turn, a choice of $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$ is a choice of the spaces where the subscales belong.

4.3 Approximate solution of the subscale equation

In this section we present two approximated solutions to the fine scale problem. The first one is based on an extension of the ideas of chapter 3 to systems of second order equations. This rather general approach, that permits to motivate the standard use of a diagonal stabilization matrix, is presented the next subsection. However, in the case of the Oseen problem, a better argument can be developed based on the same ideas but treating the coupling exactly as shown in subsection 4.3.2. The choice of $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$ is discussed in subsection 4.3.3 (we consider the simple case $\widetilde{\boldsymbol{v}}^{\perp} = 0$ and $\widetilde{q}^{\perp} = 0$ in the first two subsections) and we finally summarize the possibilities for discrete problem in subsection 4.3.4.

4.3.1 Approximating the Oseen equations as a system

Let us consider the generic problem of n equations for n unknowns $\boldsymbol{U} \in \boldsymbol{\mathcal{V}}$ of the form

$$\boldsymbol{\mathcal{L}U} := -\partial_i \left(\boldsymbol{K}_{ij} \partial_j \boldsymbol{U} \right) + \boldsymbol{A}_i \partial_i \boldsymbol{U} + \boldsymbol{S}\boldsymbol{U} = \boldsymbol{F} \quad \text{in} \quad \Omega$$

where $\boldsymbol{K}_{ij}, \boldsymbol{A}_i$ and $\boldsymbol{S}$ (for$1 \leq i, j \leq \mathrm{d}$) are square coefficient matrices of $n \times n$ components and $\boldsymbol{F} \in \boldsymbol{\mathcal{W}} = \mathcal{L}\left(\boldsymbol{\mathcal{V}}\right)$ is the vector of external forces. Several systems can be written in this form with an appropriate definition of the matrices $\boldsymbol{K}_{ij}, \boldsymbol{A}_i$ and $\boldsymbol{S}$. In particular, when $\mathrm{d} = 2$, the Oseen problem is obtained taking

$$\boldsymbol{K}_{ij} = \nu \delta_{ij} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{A}_i = \begin{bmatrix} a_i & 0 & \delta_{i1} \\ 0 & a_i & \delta_{i2} \\ \delta_{i1} & \delta_{i2} & 0 \end{bmatrix}, \quad \boldsymbol{S} = \boldsymbol{0}$$

and $\boldsymbol{U} = (u_1, u_2, p)^t$. The fine scale 4.8-4.9 problem is now written as

$$\boldsymbol{\mathcal{L}}\widetilde{\boldsymbol{U}} = \boldsymbol{F} - \boldsymbol{\mathcal{L}U}_h := \boldsymbol{R} \tag{4.11}$$

where $\widetilde{\boldsymbol{U}} = (\widetilde{u}_1, \widetilde{u}_2, \widetilde{p})^t$ and $\boldsymbol{U}_h = (u_{h1}, u_{h2}, p_h)^t$. As in the scalar case of chapter 3, we first determine the dependence of the solution with respect to the mesh size by transforming to a unitary reference domain. To this end let us define an isoparametric mapping F relating the element K (with coordinates $\boldsymbol{x}$) to a reference element $\hat{K}$ (with coordinates $\boldsymbol{\xi}$)

$$\boldsymbol{x} = F\left(\boldsymbol{\xi}\right)$$

The Jacobian of the mapping F, J, verifies

$$J_{kl} = \frac{\partial x_l}{\partial \xi_k}, \quad J^{-t}_{kl} = \frac{\partial \xi_k}{\partial x_l}.$$

and transforming 4.11 we obtain

$$\boldsymbol{\mathcal{L}}\widetilde{\boldsymbol{U}} := -\partial_i \left(\boldsymbol{K}^r_{ij} \partial_j \widetilde{\boldsymbol{U}} \right) + \boldsymbol{A}^r_i \partial_i \widetilde{\boldsymbol{U}} + \boldsymbol{S}\widetilde{\boldsymbol{U}} = \boldsymbol{R} \tag{4.12}$$

where now ∂_i stands for $\partial / \partial \xi_i$ and

$$\begin{aligned} \boldsymbol{K}^r_{ij} &= J^{-t}_{ip} J^{-t}_{jq} \boldsymbol{K}_{pq} \\ \boldsymbol{A}^r_i &= J^{-t}_{pi} \boldsymbol{A}_p \end{aligned}$$

The next step is to Fourier transform equation 4.12 and to this end we consider [29] the Fourier transform of a function v defined in $\hat{K}$ as

$$\widehat{v}\left(\boldsymbol{k}\right) = \int_{\hat{K}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\xi}} \, v\left(\boldsymbol{\xi}\right) d\boldsymbol{\xi}$$

where $\mathrm{i} = \sqrt{-1}$ and $\boldsymbol{k}$ is the wave number. If $\boldsymbol{n}$ denotes the normal to the element $\hat{K}$ we have that

$$\widehat{\frac{\partial v}{\partial \xi_j}}(\boldsymbol{k}) = \mathrm{i}k_j \widehat{v}(\boldsymbol{k}) + \int_{\partial \hat{K}} n_j e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\xi}}\, v d\Gamma_\xi$$

Applying this transform is applied to functions that vanish on the element boundary, the second term on the right hand side vanishes and we have the classical Fourier derivation formula. Transforming 4.12 we obtain

$$\mathcal{T}^{-1}(\boldsymbol{k})\,\widehat{\widetilde{\boldsymbol{U}}} = \widehat{\boldsymbol{R}} \tag{4.13}$$

where

$$\mathcal{T}(\boldsymbol{k}) := \left(k_i k_j \boldsymbol{K}^r_{ij} + \boldsymbol{S}^r + \mathrm{i}k_j \boldsymbol{A}^r_j\right)^{-1}$$

if $\mathcal{T}^{-1}(\boldsymbol{k})$ is assumed to be invertible. Using the inverse Fourier transform the subgrid scale can be written as

$$\widetilde{\boldsymbol{U}}(\boldsymbol{\eta}) = \int_{\mathbb{R}^{\mathrm{d}}} e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{\eta}}\, \mathcal{T}(\boldsymbol{k})\, \widehat{\boldsymbol{R}}(\boldsymbol{k})\, d\boldsymbol{k}$$

As in the scalar case, in the above expression we can identify the Fourier representation of the Green function of the subscale problem [75] given by

$$\widetilde{\boldsymbol{U}}(\boldsymbol{\eta}) = \int_{\hat{K}} \boldsymbol{G}(\boldsymbol{\xi}, \boldsymbol{\eta})\, \boldsymbol{R}(\boldsymbol{\xi})\, d\boldsymbol{\xi}$$

where

$$\boldsymbol{G}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \int_{\mathbb{R}^{\mathrm{d}}} \mathcal{T}(\boldsymbol{k})\, e^{-\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{\xi}-\boldsymbol{\eta})} d\boldsymbol{k} \tag{4.14}$$

If we approximate [24, 75]

$$\boldsymbol{G}(\boldsymbol{\xi}, \boldsymbol{\eta}) = \boldsymbol{\tau}\delta(\boldsymbol{\xi} - \boldsymbol{\eta}) = \boldsymbol{\tau} \int_{\mathbb{R}^{\mathrm{d}}} e^{-\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{\xi}-\boldsymbol{\eta})} d\boldsymbol{k}$$

expression 4.14 permits us to *identify the stabilization matrix as some norm* of $\mathcal{T}(\boldsymbol{k}_0)$ for certain $\boldsymbol{k}_0$. In the scalar case it is possible to show that if we consider the approximated subscale as $\widetilde{u}^{\mathrm{ap}} = \tau R$ then

$$\|\widetilde{u}\|^2_{L^2(\hat{K})} = \|\widetilde{u}^{\mathrm{ap}}\|^2_{L^2(\hat{K})}$$

provided τ is defined as $\tau = |\mathcal{T}(\boldsymbol{k}_0)|$ and the existence of $\boldsymbol{k}_0$ of is guaranteed by the mean value theorem [29].

In order to extend this argument to systems of equations, we need to define appropriate norms of $\widetilde{\boldsymbol{U}}$ and $\boldsymbol{R}$. In general neither $\boldsymbol{U}^t\boldsymbol{U}$ nor $\boldsymbol{F}^t\boldsymbol{F}$ are dimensionally meaningful. Only the product $\boldsymbol{U}^t\boldsymbol{F}$, that represents the work done by $\boldsymbol{U}$ against $\boldsymbol{F}$ is defined, because we assume the duality paring $\langle , \rangle : \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{W}} \longrightarrow \mathbb{R}$ to be defined. Therefore, we introduce a positive definite scaling matrix $\boldsymbol{M}$ such that the product $(\boldsymbol{F}_1, \boldsymbol{F}_2)_M := \boldsymbol{F}_1^t \boldsymbol{M} \boldsymbol{F}_2$ is

pointwise well defined and we define the corresponding norm $|\cdot|_M$ and $\|\cdot\|_M$ the $L^2(\Omega)$-norm of $|\cdot|_M$. We will also write $(\boldsymbol{U}_1, \boldsymbol{U}_2)_{M^{-1}} := \boldsymbol{U}_1^t \boldsymbol{M}^{-1} \boldsymbol{U}_2$ and the corresponding norms $|\cdot|_{M^{-1}}$ and $\|\cdot\|_{M^{-1}}$. Using these norms we can define the scaling of an operator as

$$\left|\mathcal{T}^{-1}\right|_M = \sup_{\boldsymbol{U} \in \boldsymbol{\mathcal{V}}} \frac{|\mathcal{T}^{-1}\boldsymbol{U}|_M}{|\boldsymbol{U}|_{M^{-1}}} = \sup_{\boldsymbol{U} \in \boldsymbol{\mathcal{V}}} \frac{\boldsymbol{U}^t \mathcal{T}^{-*} \boldsymbol{M} \mathcal{T}^{-1} \boldsymbol{U}}{\boldsymbol{U}^t \boldsymbol{M}^{-1} \boldsymbol{U}}$$

for any $\boldsymbol{U} \in \boldsymbol{\mathcal{V}}$. The choice of the scaling $\boldsymbol{M}$ is equivalent to choose the way the equations are written in dimensionless form, if this is the option adopted.

Taking the M-norm of 4.13 we have that

$$\left\|\widehat{\boldsymbol{R}}\right\|_M = \left\|\mathcal{T}^{-1}(\boldsymbol{k})\,\widehat{\widetilde{\boldsymbol{U}}}\right\|_M = \int_{\mathbb{R}^{\mathrm{d}}} \left|\mathcal{T}^{-1}(\boldsymbol{k})\,\widehat{\widetilde{\boldsymbol{U}}}\right|_M \leq \int_{\mathbb{R}^{\mathrm{d}}} \left|\mathcal{T}^{-1}(\boldsymbol{k})\right|_M \left|\widehat{\widetilde{\boldsymbol{U}}}\right|_{M^{-1}}$$

and by the mean value theorem

$$\left\|\widehat{\boldsymbol{R}}\right\|_M \leq \left|\mathcal{T}^{-1}(\boldsymbol{k}_0)\right|_M \left\|\widehat{\widetilde{\boldsymbol{U}}}\right\|_{M^{-1}}$$

Now if we approximate the subscale as $\widetilde{\boldsymbol{U}}_{\mathrm{ap}} = \boldsymbol{\tau}\boldsymbol{R}$, and we perform the same steps we arrive to

$$\left\|\widehat{\boldsymbol{R}}\right\|_M \leq \left|\boldsymbol{\tau}^{-1}\right|_M \left\|\widehat{\widetilde{\boldsymbol{U}}}\right\|_{M^{-1}}$$

Therefore we impose the condition

$$\left|\boldsymbol{\tau}^{-1}\right| = \left|\mathcal{T}^{-1}(\boldsymbol{k}_0)\right| \tag{4.15}$$

which means that the approximated subscale bounds the residual in the same way the exact subscale does. In the scalar case this means that the approximated and exact subscales have the same norm. In practice we impose condition 4.15 by computing the spectrum with respect to $\boldsymbol{M}^{-1}$ of $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$ and of $\mathcal{T}^{-*}(\boldsymbol{k}_0)\,\boldsymbol{M}\mathcal{T}^{-1}(\boldsymbol{k}_0)$ and imposing the equality of the largest eigenvalues. Actually, the ideal situation is found when both matrices have the same spectrum so $\boldsymbol{\tau}^{-1}$ is a better approximation of $\mathcal{T}^{-1}(\boldsymbol{k}_0)$. We omit the subscript in $\boldsymbol{k}_0$ in what follows.

In the case of the Oseen problem we have

$$\mathcal{T}(\boldsymbol{k})^{-1} = k_i k_j \boldsymbol{K}_{ij}^r + \boldsymbol{S}^r + \mathrm{i} k_j \boldsymbol{A}_j^r = \begin{bmatrix} \nu\kappa^2 + \mathrm{i}\kappa_j a_j & 0 & \mathrm{i}\kappa_1 \\ 0 & \nu\kappa^2 + \mathrm{i}\kappa_j a_j & \mathrm{i}\kappa_2 \\ \mathrm{i}\kappa_1 & \mathrm{i}\kappa_2 & 0 \end{bmatrix} \tag{4.16}$$

where $\kappa_i = k_j J_{ji}^{-t}$. Taking a scaling matrix $\boldsymbol{M} = diag(\mu_u, \mu_u, \mu_p)$ the eigenvalues of $\mathcal{T}^{-*}\boldsymbol{M}\mathcal{T}^{-1}$ with respect to $\boldsymbol{M}^{-1}$ are

$$\lambda_1 = A, \quad \lambda_2 = \frac{1}{2}A + B + C, \quad \lambda_3 = \frac{1}{2}A + B - C$$

$$A := \mu_u^2 \left(\nu^2\kappa^4 + (\boldsymbol{\kappa}\cdot\boldsymbol{a})^2\right)$$

$$B := \kappa^2 \mu_p \mu_u$$

$$C := \frac{1}{2}\mu_u \sqrt{\left(4\kappa^2\mu_p\mu_u\left(\nu^2\kappa^4 + (\boldsymbol{\kappa}\cdot\boldsymbol{a})^2\right) + \mu_u^2\left(\nu^2\kappa^4 + (\boldsymbol{\kappa}\cdot\boldsymbol{a})^2\right)^2\right)}$$

and *assuming* $\boldsymbol{\tau} = \mathrm{diag}\,(\tau_m, \tau_m, \tau_c)$ those of $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$ are μ_u^2/τ_m^2, μ_u^2/τ_m^2, and μ_p^2/τ_c^2. It can be easily shown that taking the scaling $\mu_u = \left(\nu^2\kappa^4 + (\boldsymbol{\kappa}\cdot\boldsymbol{a})^2\right)^{-1}$ and $\mu_p = 2\kappa^{-2}$ the spectrum of both matrices is identical (a condition stronger than 4.15) and in this case we have

$$\tau_m\,(\boldsymbol{k}) = \left(\nu^2\kappa^4 + (\boldsymbol{\kappa}\cdot\boldsymbol{a})^2\right)^{-1/2} \tag{4.17}$$

and

$$\tau_c\,(\boldsymbol{k}) = \frac{2\kappa^{-2}}{\tau_m} \tag{4.18}$$

This argument determines the functional form of the stabilization parameters. Let us finally consider the definition of $\boldsymbol{k}$. As discussed in chapter 3, its magnitude is related to the constant factors involved in the definition of the parameter whereas its direction is related to the definition of the element length. In order to reproduce the exact solution of the one dimensional convection diffusion equation we need $\|\boldsymbol{k}\| = 2h_{\mathrm{nat}}^{-1}$. On the other hand, the optimal choice of the direction is that of the instability presented by the problem, defined as

$$\boldsymbol{k}^{\mathrm{I}} = \arg\max_{\|\boldsymbol{k}\|=1} \frac{\tau_m^{-1}\,(\boldsymbol{k})}{\nu\kappa^2} \tag{4.19}$$

Note that

$$\frac{\tau_m^{-1}}{\nu\kappa^2} = \left(1 + \left(\frac{\boldsymbol{\kappa}\cdot\boldsymbol{a}}{\nu\kappa^2}\right)^2\right)^{-1/2} \sim \frac{|\boldsymbol{\kappa}\cdot\boldsymbol{a}|}{\nu\kappa^2} = \mathrm{P}_{\boldsymbol{k}}$$

where $\mathrm{P}_{\boldsymbol{k}}$ is the Péclet number in the direction of $\boldsymbol{k}$. Therefore, this definition of the instability direction is meaningless when $\boldsymbol{a} = \boldsymbol{0}$. In the case of the scalar CDR equation this is not a problem because in this case it reduces to the Poisson equation that needs not to be stabilized and any direction can be taken provided the stability condition implied by the use of the inverse estimate is satisfied. In particular, for linear elements, any direction can be considered. In the case of the Stokes problem, however, numerical experiments presented in the following section show that when the minimum element length is used numerical instabilities may appear. The direction of the maximum element length corresponds to the direction of minimum diffusion in the reference domain, what is an intuitive way to understand the problem. Therefore, when $\mathrm{P}_{\boldsymbol{k}} < 1$ we consider $\boldsymbol{k}^{\mathrm{I}}$ in the direction of the maximum element length.

An isotropic approximation to the parameters in 4.17 and 4.18, as shown in chapter 3, is given by

$$\tau_m = \left(\left(\frac{c_1\nu}{h^2}\right)^2 + \left(\frac{c_2 a}{h}\right)^2\right)^{-1/2} \tag{4.20}$$

$$\tau_c = c_1\tau_m^{-1}h^{-2} \tag{4.21}$$

4.3.2 Approximating each equation

After transforming the problem to the reference domain and applying the Fourier transform we arrive to 4.16, which exactly inverted gives

$$\mathcal{T}(\boldsymbol{k}) = \begin{bmatrix} (\nu\kappa^2 + \mathrm{i}\kappa_j a_j)^{-1} P_{11} & (\nu\kappa^2 + \mathrm{i}\kappa_j a_j)^{-1} P_{12} & -\frac{\mathrm{i}\kappa_1}{\kappa^2} \\ (\nu\kappa^2 + \mathrm{i}\kappa_j a_j)^{-1} P_{21} & (\nu\kappa^2 + \mathrm{i}\kappa_j a_j)^{-1} P_{22} & -\frac{\mathrm{i}\kappa_2}{\kappa^2} \\ -\frac{\mathrm{i}\kappa_1}{\kappa^2} & -\frac{\mathrm{i}\kappa_2}{\kappa^2} & \frac{\left(\nu\kappa^2 + \mathrm{i}\kappa_j a_j\right)}{\kappa^2} \end{bmatrix} \tag{4.22}$$

where

$$P_{ij}(\boldsymbol{k}) = \delta_{ij} - \frac{\kappa_i \kappa_j}{\kappa^2}$$

In 4.22 we identify the Fourier transform of the convection diffusion reaction operator

$$\widehat{\mathcal{L}}(\boldsymbol{\kappa}) = \left(\nu\kappa^2 + \mathrm{i}\kappa_j a_j\right)$$

and that of the Laplace operator, κ^2. The solution of the fine scale problem in the Fourier space is given by

$$\widehat{\widetilde{u}}_i = \widehat{\mathcal{L}}^{-1}(\boldsymbol{\kappa}) P_{ij} \widehat{R_{m,j}} - \frac{\mathrm{i}\kappa_i}{\kappa^2} \widehat{R_c} \tag{4.23}$$

$$\widehat{\widetilde{p}} = \frac{\widehat{\mathcal{L}}(\boldsymbol{\kappa})}{\kappa^2} \widehat{R_c} - \frac{\mathrm{i}\kappa_i}{\kappa^2} \widehat{R_{m,i}} \tag{4.24}$$

The residual of the momentum equation is multiplied by P_{ij}, the projector onto the direction orthogonal to $\boldsymbol{\kappa}$. This projection implies the satisfaction of the continuity equation as

$$\mathrm{i}\kappa_i \left[\delta_{ij} - \frac{\kappa_i \kappa_j}{\kappa^2}\right] \widehat{d}_j = \left[\mathrm{i}\kappa_j - \mathrm{i}\kappa_j\right] \widehat{d}_j = 0 \quad \forall \boldsymbol{d} \tag{4.25}$$

Therefore, if we multiply 4.23 by $\mathrm{i}\kappa_i$ and we use 4.25 continuity is exactly recovered.

The main idea presented in this section is to approximate the scalar operators $\widehat{\mathcal{L}}(\boldsymbol{\kappa})$ and κ^2 and to exactly account for the coupling between equations. To do that we use the inverse Fourier transform to obtain

$$\begin{aligned} \widetilde{u}_i(\boldsymbol{x}) &= \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \widehat{\widetilde{u}}_i(\boldsymbol{k}) \, \mathrm{d}\boldsymbol{k} \\ &= \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \widehat{\mathcal{L}}^{-1} \widehat{R_{m,i}} \mathrm{d}\boldsymbol{k} - \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \frac{\widehat{\mathcal{L}}^{-1}}{\kappa^2} \kappa_i \kappa_j \widehat{R_{m,j}} \mathrm{d}\boldsymbol{k} - \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \frac{i\kappa_i}{\kappa^2} \widehat{R_c} \mathrm{d}\boldsymbol{k} \end{aligned}$$

where the integrals extend over the wave number space. Next we approximate the operator $\widehat{\mathcal{L}}^{-1}$ by τ_m (defined below) and the Fourier transformed Laplace operator (κ^{-2}) by τ_p (also defined below). We have

$$\begin{aligned} \widetilde{u}_i(\boldsymbol{x}) &\simeq \tau_m \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \widehat{R_{m,i}} \mathrm{d}\boldsymbol{k} - \tau_m \tau_p \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} \kappa_i \kappa_j \widehat{R_{m,j}} \mathrm{d}\boldsymbol{k} - \tau_p \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} i\kappa_i \widehat{R_c} \mathrm{d}\boldsymbol{k} \\ &= \tau_m R_{m,i} + \tau_p \tau_m \partial_i \partial_j R_{m,j} - \tau_p \partial_i R_c \end{aligned}$$

In the same way

$$\begin{aligned}\widetilde{p}(\boldsymbol{x}) &= \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,\widehat{\widetilde{p}}(\boldsymbol{k})\,\mathrm{d}\boldsymbol{k} \\ &= \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,\frac{\widehat{\mathcal{L}}(\boldsymbol{\kappa})}{\kappa^2}\widehat{R_c}\mathrm{d}\boldsymbol{k} - \int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,\frac{i\kappa_i}{\kappa^2}\widehat{R_{m,i}}\mathrm{d}\boldsymbol{k}\end{aligned}$$

and performing the same approximations

$$\begin{aligned}\widetilde{p}(\boldsymbol{x}) &\simeq \tau_m^{-1}\tau_p\int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,\widehat{R_c}\mathrm{d}\boldsymbol{k} - \tau_p\int e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\,i\kappa_i\widehat{R_{m,i}}\mathrm{d}\boldsymbol{k} \\ &= \tau_p\tau_m^{-1}R_c - \tau_p\partial_i R_{m,i}\end{aligned}$$

Finally, the same argument of the previous section applied to the scalar case permits to define τ_m and τ_p as

$$\tau_m = \left(\left(\nu\kappa_0^2\right)^2 + \left(\boldsymbol{\kappa}_0\cdot\boldsymbol{a}\right)^2\right)^{-1/2}$$

and

$$\tau_p = \kappa_0^{-2}$$

Let us conclude this section with a different view of the approximation performed. Assuming $\boldsymbol{u}$ and p regular enough, taking the divergence of the momentum equation 4.8 and using the continuity equation 4.9 we find a Poisson equation for the pressure subscale

$$\nabla^2\widetilde{p} = \boldsymbol{\nabla}\cdot\boldsymbol{R}_m - \mathcal{L}R_c$$

We can formally solve this equation to obtain

$$\widetilde{p} = \nabla^{-2}\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m\right) - \nabla^{-2}\mathcal{L}R_c$$

where ∇^{-2} must satisfy appropriate boundary conditions. Introducing this solution into the first equation we have

$$\mathcal{L}\widetilde{\boldsymbol{u}} = \boldsymbol{R}_m - \boldsymbol{\nabla}\nabla^{-2}\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m\right) + \boldsymbol{\nabla}\nabla^{-2}\mathcal{L}R_c$$

We can formally solve this equation as

$$\widetilde{\boldsymbol{u}} = \mathcal{L}^{-1}\boldsymbol{R}_m - \mathcal{L}^{-1}\boldsymbol{\nabla}\nabla^{-2}\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m\right) + \mathcal{L}^{-1}\boldsymbol{\nabla}\nabla^{-2}\mathcal{L}R_c$$

Finally approximating $\mathcal{L}^{-1}$ by τ_m and $-\nabla^{-2}$ by τ_p we arrive to

$$\widetilde{\boldsymbol{u}} = \tau_m\boldsymbol{R}_m + \tau_m\tau_p\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m\right) - \tau_p\boldsymbol{\nabla}R_c \tag{4.26}$$

$$\widetilde{p} = -\tau_p\boldsymbol{\nabla}\cdot\boldsymbol{R}_m + \tau_p\tau_m^{-1}R_c \tag{4.27}$$

The approximated solution obtained can be written as

$$\begin{bmatrix}\widetilde{\boldsymbol{u}}\\ \widetilde{p}\end{bmatrix} = \boldsymbol{\tau}\begin{bmatrix}\boldsymbol{R}_m\\ R_c\end{bmatrix}$$

but the usual matrix of stabilization parameters $\boldsymbol{\tau}$ defined in the previous section as

$$\begin{bmatrix} \tau_m \boldsymbol{I} & 0 \\ 0 & \tau_c \end{bmatrix}$$

has to be replaced by a stabilization operator of the form

$$\begin{bmatrix} \tau_m \boldsymbol{I} + \tau_m \tau_p \boldsymbol{\nabla}\boldsymbol{\nabla}\cdot & -\tau_p \boldsymbol{\nabla} \\ -\tau_p \boldsymbol{\nabla}\cdot & \tau_p \tau_m^{-1} \end{bmatrix}$$

For appropriate definitions of τ_m and τ_p this operator is positive and gives rise to a stable scheme as will be shown in the following section. Neglecting the differential terms we recover the standard approach.

Remark 4 *If we define on each element the spaces of the residuals as*

$$\begin{aligned} \boldsymbol{R} &= \left\{ \boldsymbol{v} : \Omega^h \to \mathbb{R} : \boldsymbol{v} = \alpha \left(\boldsymbol{f} - \mathcal{L}\boldsymbol{u}_h - \boldsymbol{\nabla} q_h \right)|_K , \, \boldsymbol{u}_h \in \boldsymbol{V}_h, \, q_h \in Q_h, \, \alpha \in \mathbb{R} \right\} \\ \mathcal{R} &= \left\{ q : \Omega^h \to \mathbb{R} : q = \alpha \, \boldsymbol{\nabla} \cdot \boldsymbol{u}_h|_K , \, \boldsymbol{u}_h \in \boldsymbol{V}_h, \, \alpha \in \mathbb{R} \right\} \end{aligned}$$

the solution 4.26-4.27 implies

$$\begin{aligned} \widetilde{\boldsymbol{V}} &= \boldsymbol{R} + \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \boldsymbol{R} \right) + \boldsymbol{\nabla} \mathcal{R} \\ \widetilde{Q} &= \boldsymbol{\nabla} \cdot \boldsymbol{R} + \mathcal{R} \end{aligned}$$

which can be written as

$$\widetilde{\boldsymbol{V}} = \boldsymbol{R} + \boldsymbol{\nabla} \widetilde{Q}$$

It is immediately clear that theses spaces satisfy the inf-sup condition

$$\inf_{\widetilde{q} \in \widetilde{Q}} \sup_{\widetilde{\boldsymbol{v}} \in \widetilde{\boldsymbol{V}}} \frac{(\widetilde{q}, \boldsymbol{\nabla} \cdot \widetilde{\boldsymbol{v}})}{\|\widetilde{q}\| \, \|\boldsymbol{\nabla} \widetilde{\boldsymbol{v}}\|} \geq \beta > 0$$

because $\forall \widetilde{q} \in \widetilde{Q}, \boldsymbol{\nabla} \widetilde{q} \in \widetilde{\boldsymbol{V}}$.

4.3.3 The choice of the space of subscales

Let us finally consider the choice of the space of subscales or, equivalently, the definition of $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$. The diagonal approximation developed in subsection 4.3.1 can be obtained neglecting differential terms in 4.26-4.27 and we restrict the discussion to this case. Considering the general case of $\widetilde{\boldsymbol{v}}^{\perp} \neq \boldsymbol{0}$ and $\widetilde{q}^{\perp} \neq 0$ instead of 4.26-4.27 we have

$$\widetilde{\boldsymbol{u}} = \tau_m \left(\boldsymbol{R}_m + \widetilde{\boldsymbol{v}}^{\perp} \right) + \tau_m \tau_p \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \left(\boldsymbol{R}_m + \widetilde{\boldsymbol{v}}^{\perp} \right) \right) - \tau_p \boldsymbol{\nabla} \left(R_c + \widetilde{q}^{\perp} \right) \tag{4.28}$$

$$\widetilde{p} = -\tau_p \boldsymbol{\nabla} \cdot \left(\boldsymbol{R}_m + \widetilde{\boldsymbol{v}}^{\perp} \right) + \tau_p \tau_m^{-1} \left(R_c + \widetilde{q}^{\perp} \right) \tag{4.29}$$

To simplify the discussion that follows let us consider (only in the remaining part of this subsection) that the stabilization parameters are the same for all elements. This point is not essential and the reader is referred to the previous chapter for the general case. If we denote by P the projection onto the orthogonal complement of $\widetilde{Q}$ we have $P\widetilde{p} = 0$ and $P\widetilde{q}^{\perp} = \widetilde{q}^{\perp}$ and therefore from 4.29 we obtain

$$\widetilde{q}^{\perp} = \tau_m P\boldsymbol{\nabla}\cdot\left(\boldsymbol{R}_m + \widetilde{\boldsymbol{v}}^{\perp}\right) - PR_c \tag{4.30}$$

In the same way, if we denote by $\boldsymbol{P}$ the projection onto the orthogonal complement of $\widetilde{\boldsymbol{V}}$ we have $\boldsymbol{P}\widetilde{\boldsymbol{u}} = \boldsymbol{0}$ and $\boldsymbol{P}\widetilde{\boldsymbol{v}}^{\perp} = \widetilde{\boldsymbol{v}}^{\perp}$ and therefore from 4.28 we obtain

$$\widetilde{\boldsymbol{v}}^{\perp} = -\tau_m \boldsymbol{P}\boldsymbol{R}_m - \tau_m\tau_p \boldsymbol{P}\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m + \boldsymbol{\nabla}\cdot\widetilde{\boldsymbol{v}}^{\perp}\right) + \tau_p \boldsymbol{P}\boldsymbol{\nabla}\left(R_c + \widetilde{q}^{\perp}\right)$$

After some manipulation we arrive to

$$\widetilde{\boldsymbol{v}}^{\perp} = -\boldsymbol{P}\boldsymbol{R}_m - \tau_p \boldsymbol{P}\boldsymbol{\nabla}\left(I-P\right)\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m + \boldsymbol{\nabla}\cdot\widetilde{\boldsymbol{v}}^{\perp}\right) + \tau_m^{-1}\tau_p \boldsymbol{P}\boldsymbol{\nabla}\left(I-P\right)R_c \tag{4.31}$$

where I is the identity operator in Q. This gives an implicit definition of $\widetilde{\boldsymbol{v}}^{\perp}$ (as a differential equation) in terms of projections of the residuals. However, we do not need to *explicitly* compute $\widetilde{\boldsymbol{v}}^{\perp}$ and $\widetilde{q}^{\perp}$ to impose $\widetilde{\boldsymbol{u}} \in \widetilde{\boldsymbol{V}}$ and $\widetilde{p} \in \widetilde{Q}$ but only to write

$$\widetilde{\boldsymbol{u}} = \widetilde{\boldsymbol{P}}\left[\tau_m \boldsymbol{R}_m + \tau_m\tau_p \boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\boldsymbol{R}_m\right) - \tau_p\boldsymbol{\nabla}R_c\right] \tag{4.32}$$

$$\widetilde{p} = \widetilde{P}\left[-\tau_p\boldsymbol{\nabla}\cdot\boldsymbol{R}_m + \tau_p\tau_m^{-1}R_c\right] \tag{4.33}$$

where $\widetilde{P} = I - P$ and $\widetilde{\boldsymbol{P}} = \boldsymbol{I} - \boldsymbol{P}$ are the projections onto the spaces of subscales $\widetilde{\boldsymbol{V}}$ and $\widetilde{Q}$. Equating 4.28 with 4.32 and 4.29 with 4.33 we obtain conditions 4.31 and 4.30.

Two possibilities have been considered for the choice of the spaces of subscales. The easier approach is to take $P = 0$ and $\boldsymbol{P} = \boldsymbol{0}$ which is equivalent to take $\widetilde{\boldsymbol{v}}^{\perp} = 0$ and $\widetilde{q}^{\perp} = 0$ and is called in [29] the Algebraic Subgrid-Scale formulation (ASGS). In that reference the choice $P = P_h$ (and $\boldsymbol{P} = \boldsymbol{P}_h$) is advocated, P_h being the $L^2\left(\Omega^h\right)$ projection onto the finite element space Q_h (and $\boldsymbol{V}_h$). The resulting formulation is called Orthogonal Subscales Stabilization (OSS) because it corresponds to take $\widetilde{Q}$ ($\widetilde{\boldsymbol{V}}$) as the orthogonal complement of Q_h ($\boldsymbol{V}_h$) (in the $L^2\left(\Omega^h\right)$ sense).

4.3.4 The final discrete problem

Let us summarize the possibilities for the discrete problem. It consists in finding $(\boldsymbol{u}_h, p_h) \in \boldsymbol{V}_h \times Q_h$ such that

$$B_\tau\left(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h\right) = L_\tau\left(\boldsymbol{v}_h, q_h\right) \quad \forall\left(\boldsymbol{v}_h, q_h\right) \in \boldsymbol{V}_h \times Q_h \tag{4.34}$$

where the bilinear form B_τ and the linear form L_τ are given by

- Diagonal approximation:

$$\begin{aligned} B_\tau\left(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h\right) &= \left(\boldsymbol{\nabla}\boldsymbol{v}_h, \nu\boldsymbol{\nabla}\boldsymbol{u}_h\right)_\Omega + \left(\boldsymbol{v}_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right)_\Omega \\ &\quad - \left(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, p_h\right) + \left(q_h, \boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right) \\ &\quad - \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_m\widetilde{\boldsymbol{P}}\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right)\right)_h \\ &\quad + \left(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, \tau_p\tau_m^{-1}\widetilde{P}\left(\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)\right)_h \end{aligned} \tag{4.35}$$

and

$$L_\tau\left(\boldsymbol{u}_h, p_h\right) = \left\langle\boldsymbol{v}_h, \boldsymbol{f}\right\rangle_\Omega - \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_m\widetilde{\boldsymbol{P}}\boldsymbol{f}\right)_h \tag{4.36}$$

- Coupled approximation

$$\begin{aligned} B_\tau\left(\boldsymbol{u}_h, p_h; \boldsymbol{v}_h, q_h\right) &= \left(\boldsymbol{\nabla}\boldsymbol{v}_h, \nu\boldsymbol{\nabla}\boldsymbol{u}_h\right)_\Omega + \left(\boldsymbol{v}_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right)_\Omega \\ &\quad - \left(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, p_h\right) + \left(q_h, \boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right) \\ &\quad - \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_m\widetilde{\boldsymbol{P}}\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right)\right)_h \\ &\quad - \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_m\tau_p\widetilde{\boldsymbol{P}}\left[\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right)\right)\right]\right)_h \\ &\quad + \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_p\widetilde{\boldsymbol{P}}\left[\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)\right]\right)_h \\ &\quad - \left(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, \tau_p\widetilde{P}\left[\boldsymbol{\nabla}\cdot\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right)\right]\right)_h \\ &\quad + \left(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, \tau_p\tau_m^{-1}\widetilde{P}\left(\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)\right)_h \end{aligned} \tag{4.37}$$

and

$$\begin{aligned} L_\tau\left(\boldsymbol{u}_h, p_h\right) &= \left\langle\boldsymbol{v}_h, \boldsymbol{f}\right\rangle_\Omega - \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_m\widetilde{\boldsymbol{P}}\boldsymbol{f}\right)_h \\ &\quad - \left(\mathcal{L}^*\boldsymbol{v}_h - \boldsymbol{\nabla}q_h, \tau_p\tau_m^{-1}\widetilde{\boldsymbol{P}}\left[\boldsymbol{\nabla}\left(\boldsymbol{\nabla}\cdot\boldsymbol{f}\right)\right]\right)_h \\ &\quad - \left(\boldsymbol{\nabla}\cdot\boldsymbol{v}_h, \tau_p\widetilde{\boldsymbol{P}}\left(\boldsymbol{\nabla}\cdot\boldsymbol{f}\right)\right)_h \end{aligned} \tag{4.38}$$

4.4 Stability analysis

In this section we present a stability analysis of the final discrete problem in the case of $\widetilde{\boldsymbol{P}} = \boldsymbol{I}$ and $\widetilde{P} = I$ for the coupled approximation to the subscales. The stability of the diagonal approximation has already been shown in [28]. The stability in the cases of $\widetilde{\boldsymbol{P}} = \boldsymbol{P}_h^\perp$ and $\widetilde{P} = P_h^\perp$ require the bound of the finite element component of $\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h$ and this should be done using the techniques of [30]. We will make use of the following anisotropic inverse estimates [2], which can derived from a scaling argument

$$\left\|\nabla^2 v_h\right\|_K^2 \leq \frac{C_{\mathrm{K}}^2}{h_{\min}^2}\left\|\boldsymbol{\nabla}v_h\right\|_K^2, \quad \left\|\boldsymbol{\nabla}\cdot\boldsymbol{v}_h\right\|_K^2 \leq \frac{C_{\mathrm{K}}^2}{h_{\min}^2}\left\|\boldsymbol{v}_h\right\|_K^2 \tag{4.39}$$

We will also need an inverse trace estimate [142] of the form

$$\|v_h\|^2_{\partial K} \leq \frac{C_{\mathrm{E}}}{h_{\min}} \|v_h\|^2_K \tag{4.40}$$

In both cases the constants depend on the order of the polynomial approximation and estimates of their values are given in [67, 142] and references therein. In this section we will omit the min subscript and will denote the minimum element length by h unless otherwise specified and, just to simplify the notation, we will consider the parameters τ_m and τ_p constant on each element and satisfying

$$\tau_m^{-1} \geq \frac{8\nu}{h^2} \max\left(C_{\mathrm{K}}^2, C_{\mathrm{E}}^2\right) \tag{4.41}$$

and

$$\tau_p^{-1} \geq \frac{8}{h^2} \max\left(C_{\mathrm{K}}^2, C_{\mathrm{E}}^2\right) \tag{4.42}$$

Note that when the isotropic definitions 4.20 and 4.21 are used, these conditions are satisfied when the constant c_1 is such that

$$c_1 \geq 8 \max\left(C_{\mathrm{K}}^2, C_{\mathrm{E}}^2\right)$$

We will also make use of the following algebraic inequality

$$xy \leq \frac{1}{2\alpha} x^2 + \frac{\alpha}{2} y^2$$

with $\alpha > 0$. Defining the discrete norm

$$\begin{aligned} \|\boldsymbol{u}_h\|^2_\tau \;\; : \;\; &= \left\|\nu^{1/2} \boldsymbol{\nabla} \boldsymbol{u}_h\right\|^2_h + \left\|\tau_m^{1/2} \left(\boldsymbol{a} \cdot \boldsymbol{\nabla} \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right\|^2_h \\ &+ \left\|\nu \tau_m^{1/2} \tau_p^{1/2} \nabla^2 \boldsymbol{\nabla} \cdot \boldsymbol{u}_h\right\|^2_h + \left\|\tau_p^{1/2} \tau_m^{-1/2} \boldsymbol{\nabla} \cdot \boldsymbol{u}_h\right\|^2_h \end{aligned}$$

we have the following

Theorem 2 *(stability) Assume that conditions 4.41 and 4.42 are valid. Then, there exists a constant $C > 0$ such that*

$$B_\tau\left(\boldsymbol{u}_h, p_h; \boldsymbol{u}_h, p_h\right) \geq C \|\boldsymbol{u}_h\|^2_\tau$$

Proof. Taking $\boldsymbol{v}_h = \boldsymbol{u}_h$ and $q_h = p_h$ in 4.37 we have

$$\begin{aligned} B_\tau\left(\boldsymbol{u}_h, p_h; \boldsymbol{u}_h, p_h\right) = \nu \|\boldsymbol{\nabla} \boldsymbol{u}_h\|^2_\Omega &+ \left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h, \tau_m \left(\mathcal{L} \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_h \\ &+ \left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h, \tau_m \tau_p \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \left(\mathcal{L} \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)\right)_h \\ &- \left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h, \tau_p \boldsymbol{\nabla} \left(\boldsymbol{\nabla} \cdot \boldsymbol{u}_h\right)\right)_h \\ &- \left(\boldsymbol{\nabla} \cdot \boldsymbol{u}_h, \tau_p \boldsymbol{\nabla} \cdot \left(\mathcal{L} \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_h \\ &+ \left(\boldsymbol{\nabla} \cdot \boldsymbol{u}_h, \tau_p \tau_m^{-1} \boldsymbol{\nabla} \cdot \boldsymbol{u}_h\right)_h \end{aligned} \tag{4.43}$$

As usual, the second term on the left hand side of 4.43 provides stability of the convective term and the pressure gradient. Integrating by parts the third and fourth terms in 4.43 we have

$$\begin{aligned}
B_\tau\left(\boldsymbol{u}_h, p_h; \boldsymbol{u}_h, p_h\right) \;\geq\; & \nu \left\|\boldsymbol{\nabla} \boldsymbol{u}_h\right\|_\Omega^2 && (4.44)\\
& + \left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h, \tau_m \left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_h && (4.45)\\
& - \left(\boldsymbol{\nabla}\cdot\left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right), \tau_m\tau_p \boldsymbol{\nabla}\cdot\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_h && (4.46)\\
& + \left(\boldsymbol{n}\cdot\left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right), \tau_m\tau_p \boldsymbol{\nabla}\cdot\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_{\partial h} && (4.47)\\
& + \left(\boldsymbol{\nabla}\cdot\left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right), \tau_p \boldsymbol{\nabla}\cdot \boldsymbol{u}_h\right)_h && (4.48)\\
& - \left(\boldsymbol{n}\cdot\left(-\mathcal{L}^* \boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right), \tau_p \boldsymbol{\nabla}\cdot \boldsymbol{u}_h\right)_{\partial h} && (4.49)\\
& - \left(\boldsymbol{\nabla}\cdot \boldsymbol{u}_h, \tau_p \boldsymbol{\nabla}\cdot\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_h && (4.50)\\
& + \left\|\tau_m^{-1/2}\tau_p^{1/2}\boldsymbol{\nabla}\cdot \boldsymbol{u}_h\right\|_h^2 && (4.51)
\end{aligned}$$

Using the inverse estimate in 4.45 we have

$$\begin{aligned}
\tau_m\left(-\mathcal{L}^*\boldsymbol{u}_h, \mathcal{L}\boldsymbol{u}_h\right)_K &= \tau_m\left(\left(\nu\nabla^2 + \boldsymbol{a}\cdot\boldsymbol{\nabla}\right)\boldsymbol{u}_h, \left(-\nu\nabla^2 + \boldsymbol{a}\cdot\boldsymbol{\nabla}\right)\boldsymbol{u}_h\right)_K\\
&= -\tau_m\left\|\nu\nabla^2\boldsymbol{u}_h\right\|_K^2 + \tau_m\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2\\
&\geq -\tau_m\nu\frac{C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2 + \tau_m\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2
\end{aligned}$$

We also have

$$\tau_m\left(\boldsymbol{\nabla} p_h, \left(\mathcal{L} - \mathcal{L}^*\right)\boldsymbol{u}_h\right)_K = 2\tau_m\left(\boldsymbol{\nabla} p_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right)_K$$

and therefore

$$\begin{aligned}
\left(-\mathcal{L}^*\boldsymbol{u}_h + \boldsymbol{\nabla} p_h, \mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)_K \;\geq\; & -\tau_m\nu\frac{C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2 + \tau_m\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2\\
& + 2\tau_m\left(\boldsymbol{\nabla} p_h, \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right)_K + \tau_m\left\|\boldsymbol{\nabla} p_h\right\|_K^2\\
=\; & -\tau_m\nu\frac{C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2 + \tau_m\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right\|_K^2
\end{aligned}$$

In the same way 4.46 is bounded as

$$\begin{aligned}
-\tau_m\tau_p\left(\boldsymbol{\nabla}\cdot\left(-\mathcal{L}^*\boldsymbol{u}_h\right), \boldsymbol{\nabla}\cdot\mathcal{L}\boldsymbol{u}_h\right)_K &= -\tau_m\tau_p\left(\boldsymbol{\nabla}\cdot\left(\nu\nabla^2 + \boldsymbol{a}\cdot\boldsymbol{\nabla}\right)\boldsymbol{u}_h, \boldsymbol{\nabla}\cdot\left(-\nu\nabla^2 + \boldsymbol{a}\cdot\boldsymbol{\nabla}\right)\boldsymbol{u}_h\right)_K\\
&= \tau_m\tau_p\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2 - \tau_m\tau_p\left\|\boldsymbol{\nabla}\cdot\left(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right)\right\|_K^2
\end{aligned}$$

and

$$-\tau_m\tau_p\left(\nabla^2 p_h, \boldsymbol{\nabla}\cdot\left(\mathcal{L} - \mathcal{L}^*\right)\boldsymbol{u}_h\right)_K = -2\tau_m\tau_p\left(\nabla^2 p_h, \boldsymbol{\nabla}\cdot\left(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h\right)\right)_K$$

so using the inverse estimate

$$\begin{aligned}
& -\left(\boldsymbol{\nabla}\cdot\left(-\mathcal{L}^*\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right), \tau_m\tau_p\boldsymbol{\nabla}\cdot\left(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right)_K\\
&\qquad = \tau_m\tau_p\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2 - \tau_m\tau_p\left\|\boldsymbol{\nabla}\cdot\left(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right)\right\|_K^2\\
&\qquad \geq \tau_m\tau_p\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2 - \tau_m\tau_p\frac{C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla} p_h\right\|_K^2
\end{aligned}$$

The product of the pressure gradient and velocity divergence in 4.48 cancels with the same product in 4.50. The remaining parts of 4.48 and 4.50 are bounded as

$$\begin{aligned}\tau_p \left(\boldsymbol{\nabla}\cdot(-\mathcal{L}^*-\mathcal{L})\,\boldsymbol{u}_h, \boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_K &= \tau_p\left(2\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h, \boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_K \\ &\geq -\tau_p \left\|2\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K \left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K \\ &\geq -2\tau_p\frac{\nu C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\end{aligned}$$

It remains to bound the boundary terms. To this end, if y, z are finite element functions or derivatives of finite element functions we have

$$(y,z)_{\partial K} \geq -\left\|y\right\|_{\partial K}\left\|z\right\|_{\partial K} \geq -\frac{1}{2\alpha}\left\|y\right\|_{\partial K}^2 - \frac{\alpha}{2}\left\|z\right\|_{\partial K}^2$$

and using the inverse estimate 4.40 we arrive to

$$(y,z)_{\partial K} \geq \frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|y\right\|_K^2 - \frac{\alpha}{2}\left\|z\right\|_K^2\right) \tag{4.52}$$

for any $\alpha > 0$. For each contribution to 4.47 we have

$$\begin{aligned}&\left(\boldsymbol{n}\cdot(-\mathcal{L}^*\boldsymbol{u}_h + \boldsymbol{\nabla}p_h), \boldsymbol{\nabla}\cdot(\mathcal{L}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h)\right)_{\partial K} \\ &\quad= \left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h + \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right), \boldsymbol{\nabla}\cdot\left(-\nu\nabla^2\boldsymbol{u}_h + \boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right)\right)_{\partial K} \\ &\quad= -\left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right), \boldsymbol{\nabla}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right)\right)_{\partial K} && (4.53) \\ &\quad+ \left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right), \boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h)\right)_{\partial K} && (4.54) \\ &\quad- \left(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h), \boldsymbol{\nabla}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right)\right)_{\partial K} && (4.55) \\ &\quad+ \left(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h), \boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h)\right)_{\partial K} && (4.56)\end{aligned}$$

Using 4.52 and the inverse estimate 4.39 we have

$$\begin{aligned}-\left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right), \boldsymbol{\nabla}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right)\right)_{\partial K} &\geq \frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\nu\nabla^2\boldsymbol{u}_h\right\|_K^2 - \frac{\alpha}{2}\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\right) \\ &\geq \frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\frac{C_{\mathrm{K}}^2}{h^2}\left\|\nu\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2 - \frac{\alpha}{2}\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\right)\end{aligned}$$

and taking $\alpha = h/\left(2C_{\mathrm{E}}\right)$ we have a bound for 4.53

$$-\left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right), \boldsymbol{\nabla}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right)\right)_{\partial K} \geq -\frac{C_{\mathrm{K}}^2 C_{\mathrm{E}}^2}{h^4}\left\|\nu\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2 - \frac{1}{4}\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2$$

In the same way

$$\begin{aligned}&\left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right), \boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h)\right)_{\partial K} \\ &\qquad\geq \frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\nu\nabla^2\boldsymbol{u}_h\right\|_K^2 - \frac{\alpha}{2}\left\|\boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h)\right\|_K^2\right) \\ &\qquad\geq \frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\frac{C_{\mathrm{K}}^2}{h^2}\left\|\nu\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2 - \frac{\alpha}{2}\frac{C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h + \boldsymbol{\nabla}p_h\right\|_K^2\right)\end{aligned}$$

and taking $\alpha = h/(2C_{\mathrm{E}})$ we have a bound for 4.54

$$\left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right),\boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h)\right)_{\partial K}\geq-\frac{C_{\mathrm{E}}^2C_{\mathrm{K}}^2}{h^4}\left\|\nu\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2-\frac{C_{\mathrm{K}}^2}{4h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2$$

Using again 4.52 and the inverse estimate we have

$$\begin{aligned}-\left(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h),\boldsymbol{\nabla}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right)\right)_{\partial K}&\\ \geq\frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{\alpha}{2}\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\right)&\end{aligned}$$

and taking $\alpha = h/(2C_{\mathrm{E}})$ we have a bound for 4.55

$$-\left(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h),\boldsymbol{\nabla}\cdot\left(\nu\nabla^2\boldsymbol{u}_h\right)\right)_{\partial K}\geq-\frac{C_{\mathrm{E}}^2}{h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{1}{4}\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2$$

In the case of 4.56 we have

$$\begin{aligned}(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h),\boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h))_{\partial K}&\\ \geq\frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{\alpha}{2}\left\|\boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h)\right\|_K^2\right)&\\ \geq\frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{\alpha}{2}\frac{C_{\mathrm{K}}^2}{h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2\right)&\end{aligned}$$

and taking again $\alpha = h/(2C_{\mathrm{E}})$ we have a bound for 4.56

$$\begin{aligned}(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h),\boldsymbol{\nabla}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h))_{\partial K}&\\ \geq-\frac{C_{\mathrm{E}}}{h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{C_{\mathrm{K}}^2}{4h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2&\end{aligned}$$

Finally for each contribution to 4.49 we have

$$\begin{aligned}-\left(\boldsymbol{n}\cdot\left(\nu\nabla^2\boldsymbol{u}_h+\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right),\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}&=-\left(\boldsymbol{n}\cdot\nu\nabla^2\boldsymbol{u}_h,\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}\\ &\quad-\left(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h),\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}\end{aligned}$$

and we have

$$\begin{aligned}-\left(\boldsymbol{n}\cdot\nu\nabla^2\boldsymbol{u}_h,\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}&\geq\frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\nu\nabla^2\boldsymbol{u}_h\right\|_K^2-\frac{\alpha}{2}\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\right)\\ &\geq\frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\frac{C_{\mathrm{K}}^2}{h^2}\left\|\nu\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2-\frac{\alpha}{2}\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\right)\end{aligned}$$

so taking $\alpha = \nu C_{\mathrm{E}}/h$ we have

$$-\left(\boldsymbol{n}\cdot\nu\nabla^2\boldsymbol{u}_h,\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}\geq-\frac{\nu C_{\mathrm{K}}^2}{2h^2}\left\|\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2-\frac{\nu C_{\mathrm{E}}^2}{2h^2}\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2$$

Also

$$-\left(\boldsymbol{n}\cdot(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h),\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}\geq\frac{C_{\mathrm{E}}}{h}\left(-\frac{1}{2\alpha}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{\alpha}{2}\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\right)$$

and taking $\alpha = h/\left(\tau_m C_{\mathrm{E}}\right)$ we have

$$-\left(\boldsymbol{n}\cdot\left(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right),\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right)_{\partial K}\geq-\frac{\tau_m C_{\mathrm{E}}^2}{2h^2}\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2-\frac{\tau_m^{-1}}{2}\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2$$

Grouping terms we arrive to

$$\begin{aligned}B_\tau\left(\boldsymbol{u}_h,p_h;\boldsymbol{u}_h,p_h\right)\geq&\sum_{K\in\mathcal{P}_h}\nu\left(1-\tau_m\frac{\nu C_{\mathrm{K}}^2}{h^2}-\tau_m\tau_p\frac{2\nu C_{\mathrm{K}}^2C_{\mathrm{E}}^2}{h^4}-\tau_p\frac{C_{\mathrm{K}}^2}{2h^2}\right)\left\|\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_K^2\\&+\sum_{K\in\mathcal{P}_h}\tau_m\left(1-\tau_p\frac{C_{\mathrm{K}}^2}{h^2}-\tau_p\frac{C_{\mathrm{K}}^2}{2h^2}-\tau_p\frac{2C_{\mathrm{E}}^2}{h^2}-\tau_p\frac{C_{\mathrm{E}}^2}{2h^2}\right)\left\|\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right\|_K^2\\&+\sum_{K\in\mathcal{P}_h}\tau_p\tau_m^{-1}\left(1-\tau_m\frac{2\nu C_{\mathrm{K}}^2}{h^2}-\tau_m\frac{\nu C_{\mathrm{E}}^2}{2h^2}-\frac{1}{2}\right)\left\|\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\\&+\sum_{K\in\mathcal{P}_h}\frac{1}{2}\tau_m\tau_p\left\|\nu\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_K^2\end{aligned}$$

and using conditions 4.41 and 4.42 we have

$$\begin{aligned}B_\tau\left(\boldsymbol{u}_h,p_h;\boldsymbol{u}_h,p_h\right)\quad\geq\quad&\frac{1}{2}\left\|\nu^{1/2}\boldsymbol{\nabla}\boldsymbol{u}_h\right\|_h^2+\frac{1}{2}\left\|\tau_m^{1/2}\left(\boldsymbol{a}\cdot\boldsymbol{\nabla}\boldsymbol{u}_h+\boldsymbol{\nabla}p_h\right)\right\|_h^2\\&+\frac{1}{2}\left\|\nu\tau_m^{1/2}\tau_p^{1/2}\nabla^2\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_h^2+\frac{3}{16}\left\|\tau_p^{1/2}\tau_m^{-1/2}\boldsymbol{\nabla}\cdot\boldsymbol{u}_h\right\|_h^2\end{aligned}$$

that immediately implies the result. ■

4.5 Numerical examples

In this section we present two numerical examples both of them using the diagonal approximation to the subscales, with the objective of studying the influence of the anisotropic mesh refinement and the influence of the choice of the space of subscales. The performance of the coupled approximation needs further research. The first example is the simple problem of a 2D Stokes flow in a channel and the superior performance of the OSS method will be clearly demonstrated. The second example is an anisotropic refinement study using an analytic solution. Again, the OSS method gives better results. Numerical instabilities are found when the classical expression for the stabilization parameter using the minimum element length is used.

4.5.1 Stokes flow in a channel

In this subsection we consider the Stokes problem on the domain $\Omega = [0, 10] \times [0, 1]$ with Dirichlet boundary conditions. The boundary of the domain can divided into an inflow part at $x = 0$, an outflow part at $x = 10$ and two walls at $y = 0$ and $y = 1$. On the inflow and outflow part a Poiseuille (quadratic) velocity profile is imposed and on the walls the

non slip condition is prescribed. The flow is driven by an external force $\boldsymbol{f} = (2, 0)$ (which is equivalent to an imposed pressure gradient). The exact solution of the problem is

$$\boldsymbol{u} = (y\,(1-y)\,, 0)\,, \quad p = 0$$

The problem was solved using a uniform mesh of 10×10 bilinear elements whose aspect ratio is only 10. The problem was solved using the ASGS and OSS formulations in which the differential terms of the stabilization operator are neglected (they differ in the projection of the residual, see section 3). For each formulation the results obtained defining the stabilization parameter with $h = h_{\min}$ and $h = h_{\max}$ are compared.

In the case of the ASGS formulation, the impact of the choice of the element length in the definition of the stabilization parameters is very important, as can be seen in figure 4.1, where the velocity field obtained is shown. This difference can also be seen in figure 4.2, where the y component of the velocity, that should vanish, is more than one order of magnitude bigger when $h = h_{\max}$ than when $h = h_{\min}$. The same consideration can be made about the pressure field, shown in figure 4.3, where the maximum pressure difference of 1.05 in the case of $h = h_{\min}$ and of 15.52 in the case of $h = h_{\max}$ can be seen. Let us note that when quadratic elements are used the exact solution is found for any mesh regardless of the stabilization parameter taken because this solution belongs to the finite element space.

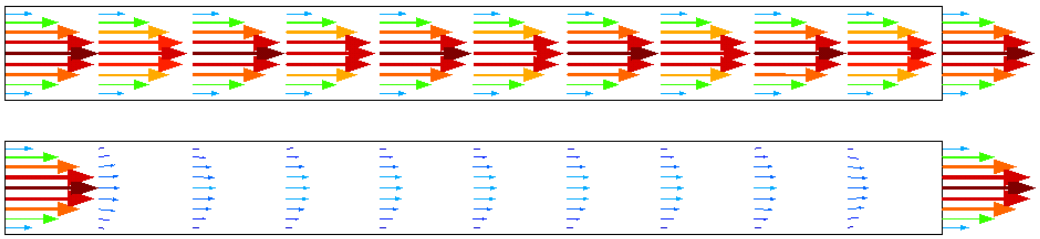

Figure 4.1: Velocity fields obtained using the ASGS formulation with $h = h_{\min}$ (top) and $h = h_{\max}$ (bottom).

On the contrary, in the case of the OSS formulation the impact of the choice of the element length is smaller. In fact, the velocity fields, shown in figure 4.4, are indistinguishable. However, some differences between the solution with $h = h_{\min}$ and $h = h_{\max}$ can be found, as can be seen in figure 4.5, where the y component of the velocities are shown, and in figure 4.6, where the pressure fields are shown. When $h = h_{\min}$ is used the exact result is obtained and when $h = h_{\max}$ is used still a non zero y component of the velocity and a non zero pressure are obtained. Note, however, that the result obtained using the OSS formulation using $h = h_{\max}$ is better than the result obtained using the ASGS formulation and $h = h_{\min}$. The explanation we give for these results is that when the projection is included, the stabilizing term added to the Galerkin one

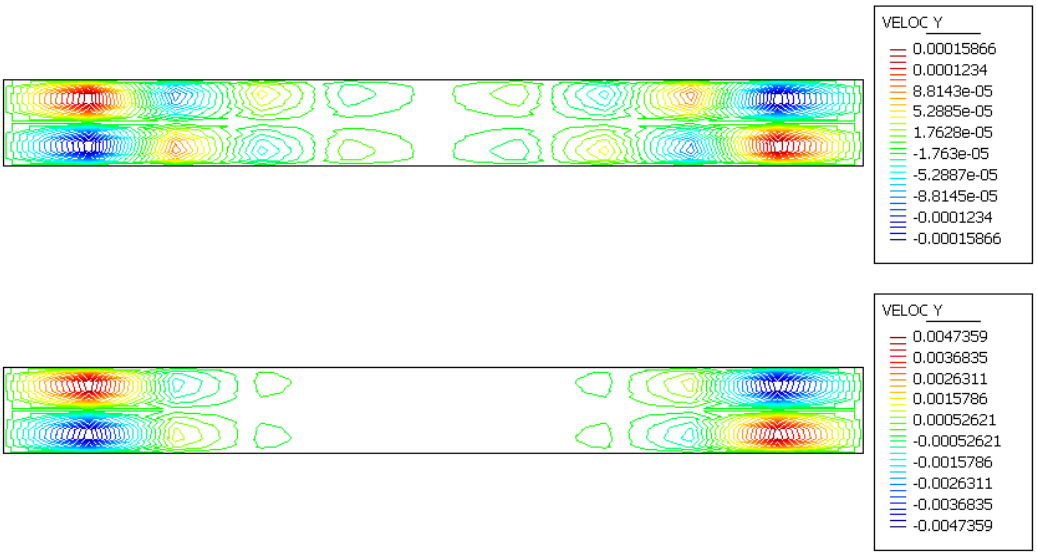

Figure 4.2: y component of the velocities obtained using the ASGS formulation with $h = h_{\min}$ (top) and $h = h_{\max}$ (bottom).

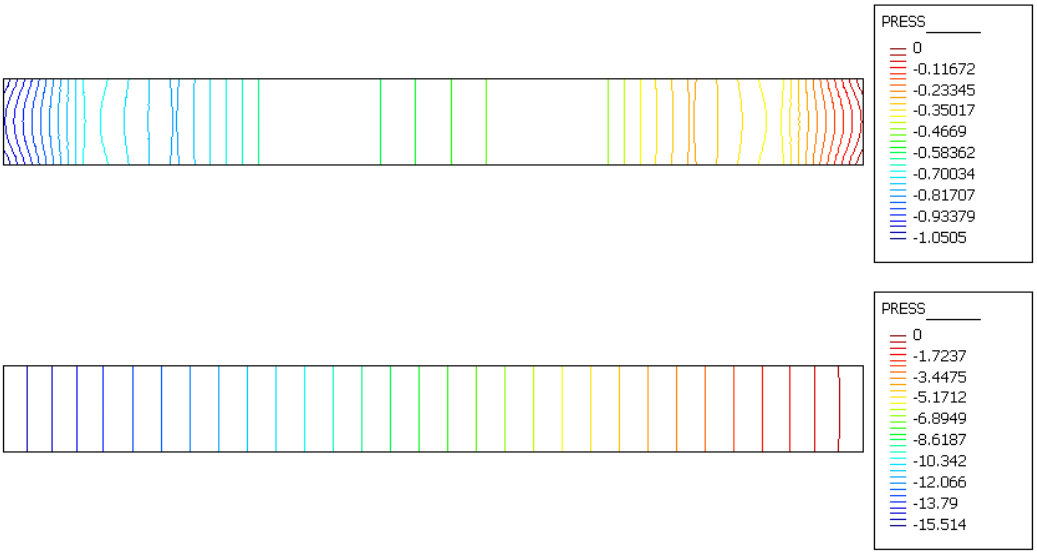

Figure 4.3: Pressure fields obtained using the ASGS formulation with $h = h_{\min}$ (top) and $h = h_{\max}$ (bottom).

is smaller (because the projection vanishes when $h \to 0$) although it is enough to have optimal convergence. This term vanishes when $h \to 0$ because it is proportional to the residual of the finite element component. In the case of linear elements the approximation of the residual is quite poor (the Laplacians of finite element functions vanish) even if it is of the correct order. The inclusion of the projection remedies the situation.

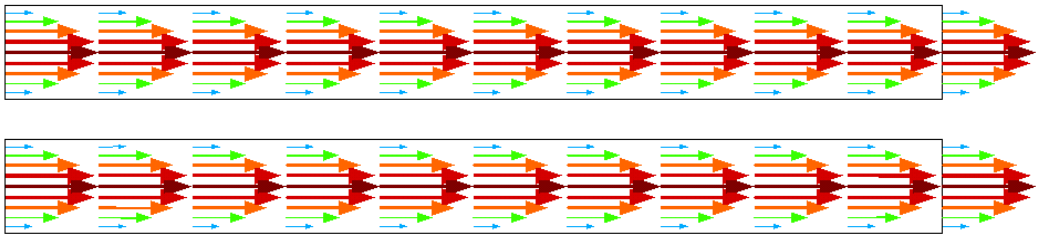

Figure 4.4: Velocity fields obtained using the OSS formulation with $h = h_{\min}$ (top) and $h = h_{\max}$ (bottom).

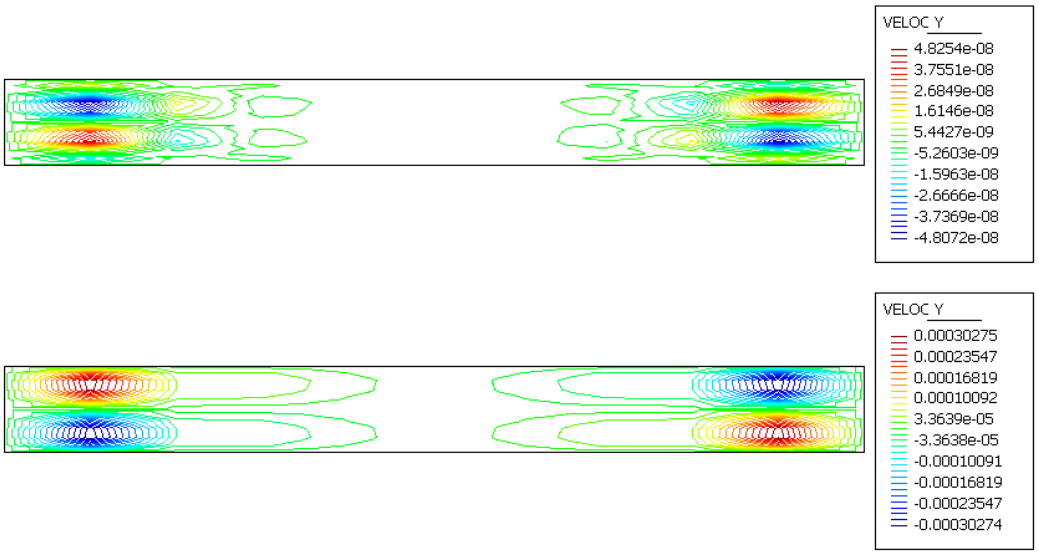

Figure 4.5: y component of the velocities obtained using the OSS formulation with $h = h_{\min}$ (top) and $h = h_{\max}$ (bottom).

Two main conclusions can be drawn from this example: the choice $h = h_{\min}$ gives better results than the choice $h = h_{\max}$ and the OSS formulation gives better results than the ASGS one, in particular being much less sensitive to the definition of the stabilization parameter. However, the choice $h = h_{\min}$ can give rise to numerical oscillations as shown in the next example.

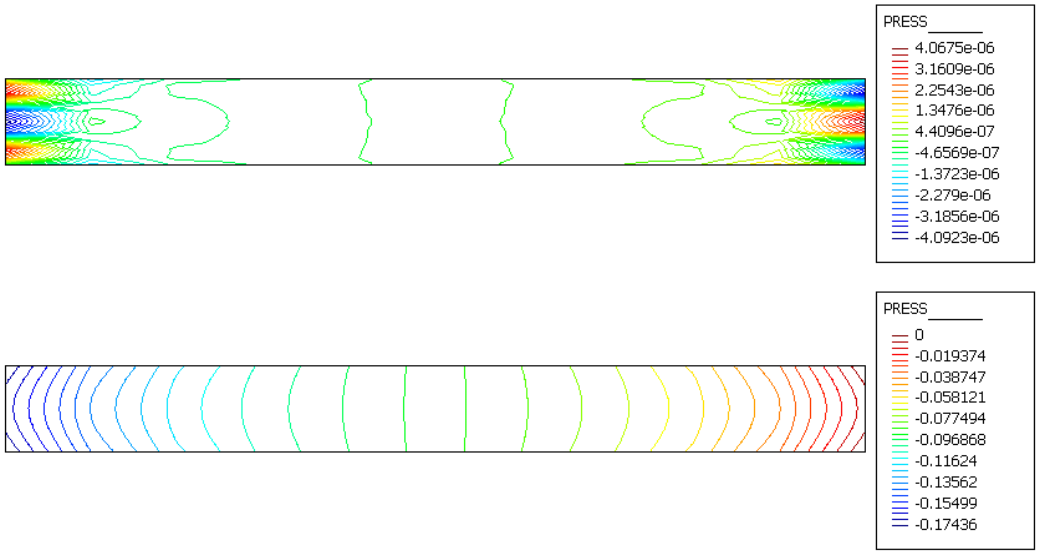

Figure 4.6: Pressure fields obtained using the ASGS formulation with $h = h_{\min}$ (top) and $h = h_{\max}$ (bottom).

4.5.2 An anisotropic convergence test

In this subsection we consider the Oseen problem in the domain $\Omega = [0,1]\times[0,1]$ with zero Dirichlet boundary conditions on $\partial\Omega$. A forcing term is prescribed to have the solution given by

$$\boldsymbol{u} = (u_x(y), u_y(x)) = \left(1 + e^{-\alpha} - e^{-\alpha y} - e^{\alpha(y-1)}, 1 + e^{-\alpha} - e^{-\alpha x} - e^{\alpha(x-1)}\right)$$

and

$$p = 1 + x + x^2$$

which presents boundary layers on the domain boundary whose width can be controlled using the parameter α. We consider $\alpha = 100$ and the advection velocity is given by the exact solution for Reynolds numbers of 0 (Stokes) and 10^2. We solve the problem using meshes of 10×10 and also 100×10, 1000×10, 10000×10 giving aspect ratios $\mathrm{A} := h_1/h_2 = 10^0, 10^1, 10^2, 10^3$. As in the case of the scalar convection diffusion problem, the impact of the choice of the element length is very important. We plot the pressure along the line $y = 0.9$ and the y component of the velocity along the line $y = 0.5$. The results of the Stokes problem are shown in figures 4.7, 4.8, 4.9 and 4.10. When $h = h_{\min}$ is used to define the stabilization parameter numerical oscillations show up, specially in the pressure but also in the velocity. When $h = h_{\max}$ is used to define the stabilization parameter the solution is free of such oscillations and in this case the results are much better when the OSS method is used.

The results of the Oseen problem for $Re = 10^2$ are shown in figures 4.11, 4.12, 4.13 and 4.14. Similar conclusions can be obtained in this case. The only important point

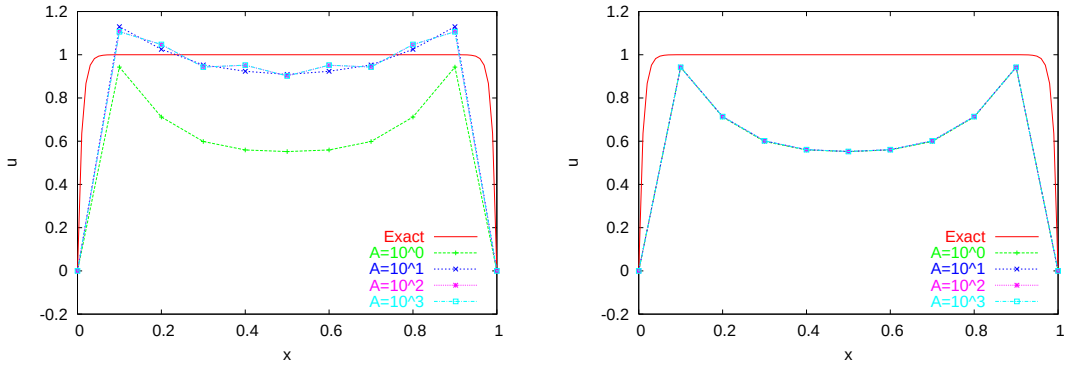

Figure 4.7: y component of the velocities obtained using the ASGS formulation with $h = h_{\min}$ (left) and $h = h_{\max}$ (right) at Re= 0.

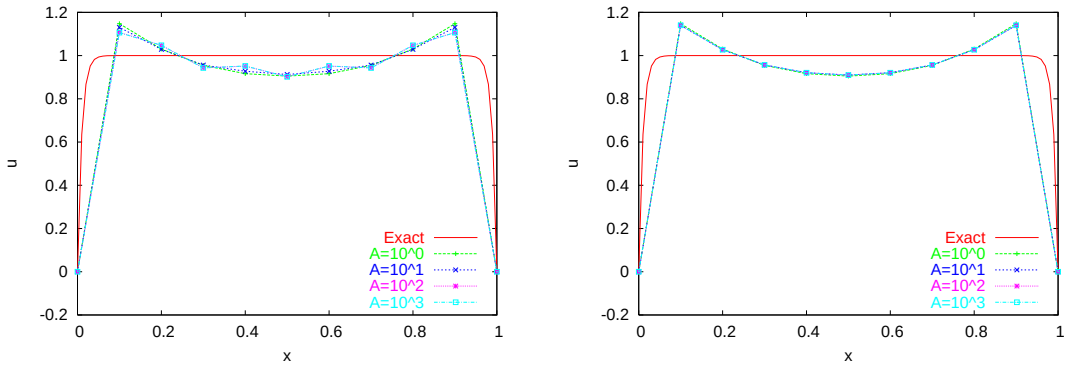

Figure 4.8: y component of the velocities obtained using the OSS formulation with $h = h_{\min}$ (left) and $h = h_{\max}$ (right) at Re= 0.

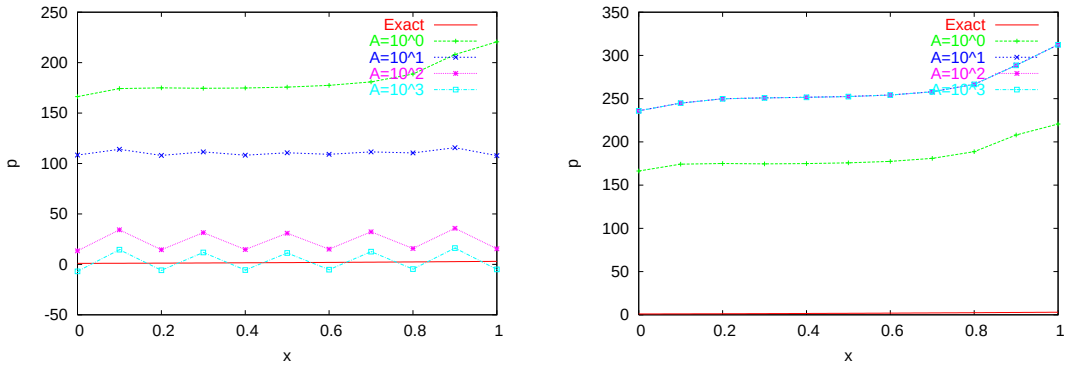

Figure 4.9: Pressures obtained using the ASGS formulation with $h = h_{\min}$ (left) and $h = h_{\max}$ (right) at Re= 0.

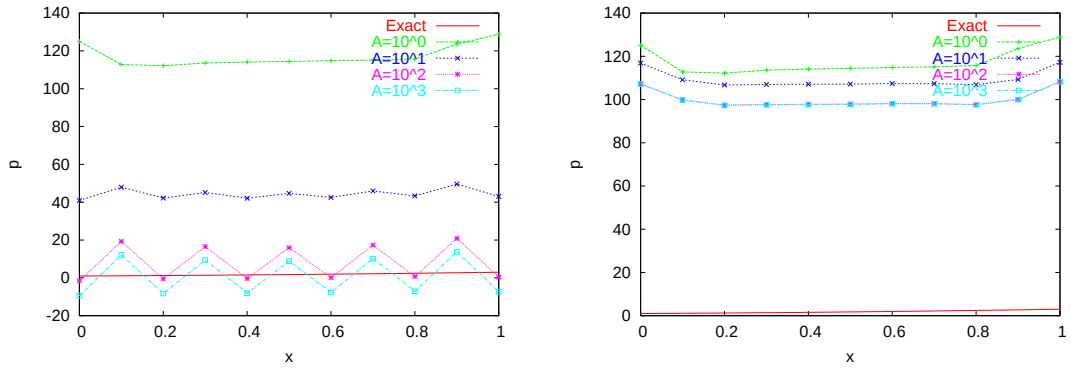

Figure 4.10: Pressures obtained using the OSS formulation with $h = h_{\min}$ (left) and $h = h_{\max}$ (right) at Re= 0.

to mention is that the new definition of the stabilization parameter does not improve the results with respect to those obtained using $h = h_{\max}$ as it does in the case of the CDR equation, although more numerical experiments are needed to clarify the point. The difference in the behavior could be due to the terms present in the coupled approximation and neglected in the diagonal one. Again, this is a point that needs further research.

4.6 Conclusions

We have presented a procedure to derive stabilized formulations for systems of equations. In the general case the steps to be performed are

- Fourier transform the system of equation.
- Compute the spectrum of $\mathcal{T}^{-*}(\boldsymbol{k}_0)\,\boldsymbol{M}\mathcal{T}^{-1}(\boldsymbol{k}_0)$ and of $\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{\tau}^{-1}$ with respect to the scaling matrix $\boldsymbol{M}$ and impose (at least) the equality of the largest eigenvalue. The optimal situation is when the spectrum of both matrices is identical.

In the particular case of the Oseen problem (but hopefully in other systems as well) the following alternative procedure can be followed

- Fourier transform the system of equation
- Invert the algebraic matrix.
- Apply the inverse Fourier transform approximating the scalar operators but keeping the differential operators for the coupling.
- Perform a stability analysis to show that the resulting formulation provides stability.

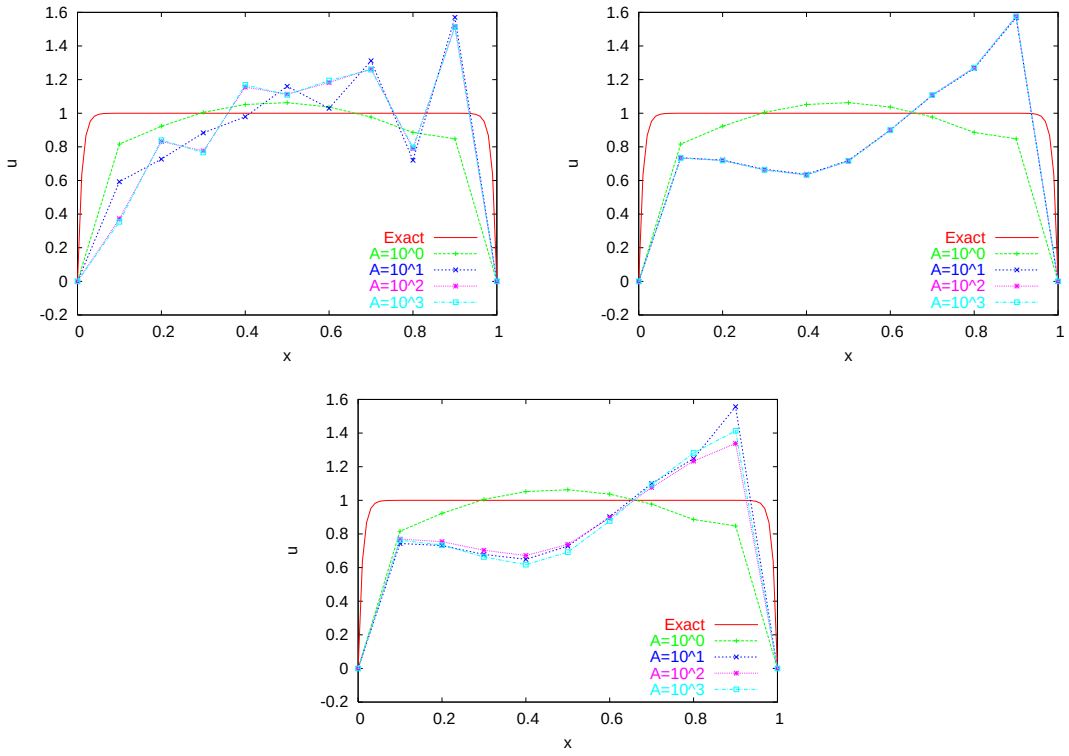

Figure 4.11: y component of the velocities obtained using the ASGS formulation with $h = h_{\min}$ (top left), $h = h_{\max}$ (top right) and 4.19 (bottom) at Re= 10^2.

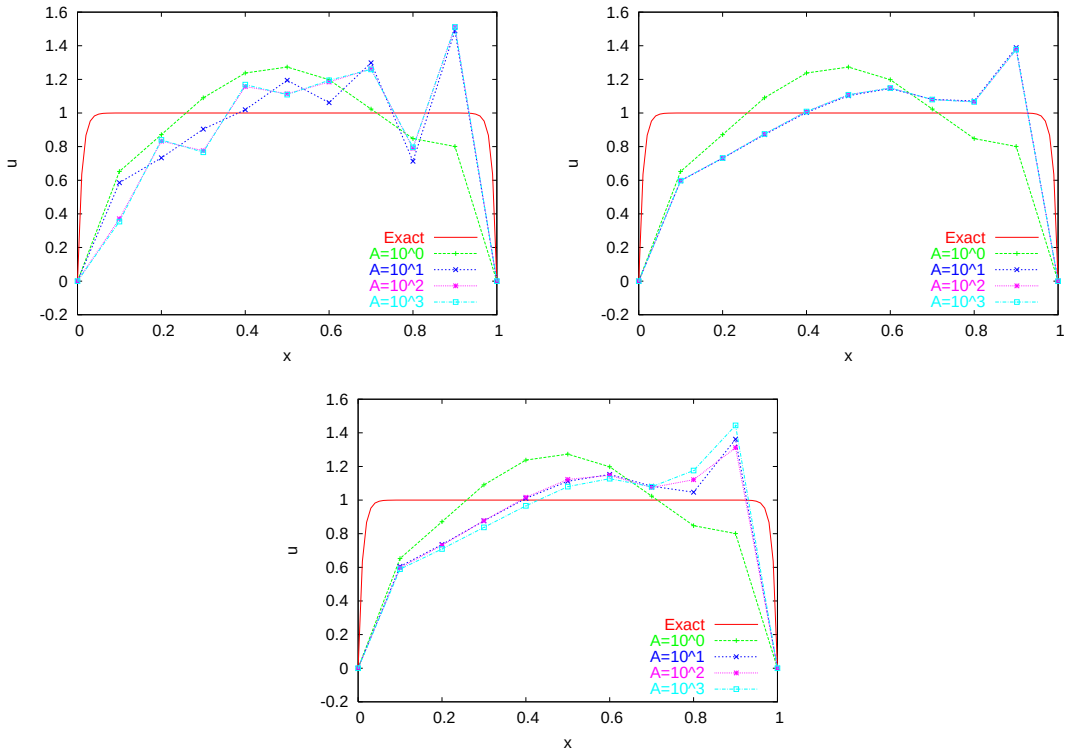

Figure 4.12: y component of the velocities obtained using the OSS formulation with $h = h_{\min}$ (top left), $h = h_{\max}$ (top right) and 4.19 (bottom) at Re= 10^2.

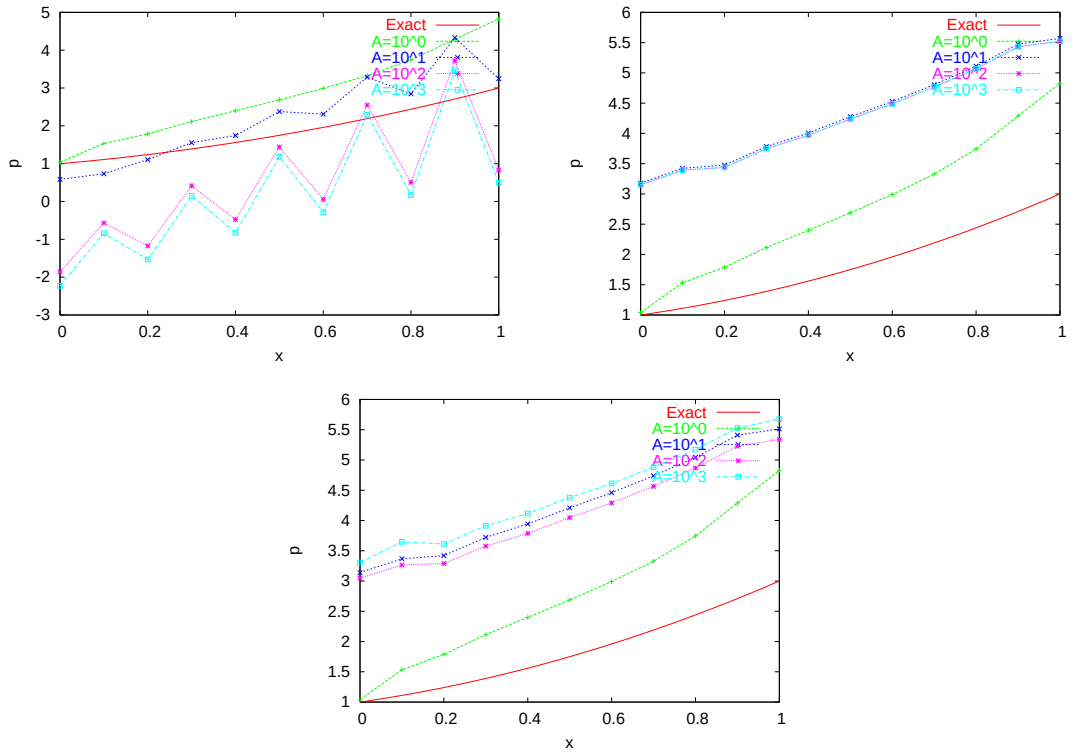

Figure 4.13: Pressures obtained using the ASGS formulation with $h = h_{\min}$ (top left), $h = h_{\max}$ (top right) and 4.19 (bottom) at Re= 10^2.

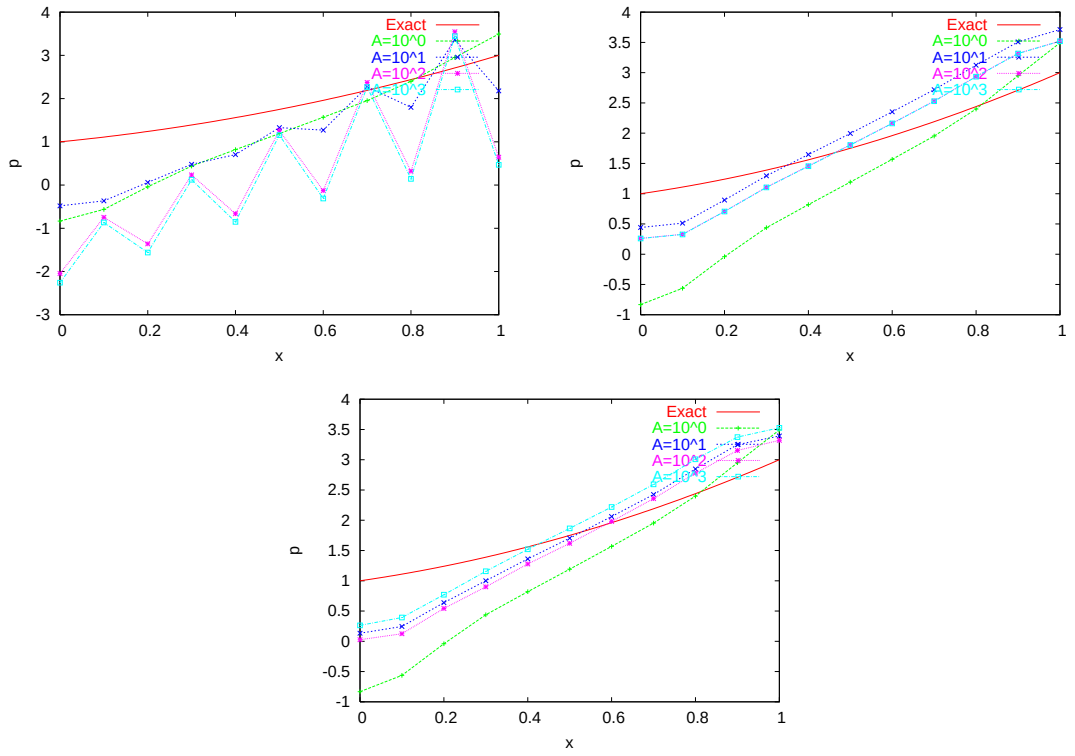

Figure 4.14: Pressures obtained using the OSS formulation with $h = h_{\min}$ (top left), $h = h_{\max}$ (top right) and 4.19 (bottom) at Re= 10^2.

The second procedure highlights the vectorial structure of the Navier Stokes equations and does not give rise to a matrix of stabilization parameters, but to a stabilization operator of the form

$$\begin{bmatrix} \tau_m \boldsymbol{I} + \tau_m \tau_p \boldsymbol{\nabla}\boldsymbol{\nabla}\cdot & -\tau_p \boldsymbol{\nabla} \\ -\tau_p \boldsymbol{\nabla}\cdot & \tau_p \tau_m^{-1} \end{bmatrix}$$

Although the extra terms could be important when anisotropic finite element meshes are used, the mesh information is taken into account only in the definition of the stabilization parameters τ_m and τ_p. This is an important point since formulations of anisotropic stabilized approximations of the Navier Stokes equations in which the stabilization parameters where replaced by matrices that depend on the stretching of the grid have been used in the past[9, 8, 10]. As shown here, a formulation of this type is rather ad hoc and does not naturally follow from the multiscale concept.

It has been also proved that under some conditions on the parameters τ_m and τ_p the resulting scheme is stable. In principle the proof of the stability bound is valid for any mesh (isotropic or anisotropic) provided the conditions on the stabilization parameters are satisfied. However, in the limit of vanishing advection that would imply the need of using the minimum element length in the definition of the stabilization parameters. In the case of the diagonal approximation this choice gives rise to spurious oscillations as shown in the numerical examples.

The results of the numerical examples presented show that the OSS method performs much better than the ASGS method as it is much less sensitive to the choice of the element length. As mentioned, the standard definition taking $h = h_{\min}$ results in an unstable scheme even for the Stokes problem. The results obtained using the standard definition taking $h = h_{\max}$ and those obtained using 4.17 and 4.18 are almost identical. Both deteriorate when the mesh is anisotropically refined. This odd behavior might be due to the neglect of the differential terms of the stabilization operator but it might be also due to the definition of the direction of instability. This definition has been made in the previous chapter for the convection diffusion reaction equation but it might be not the correct one for the Oseen problem. In fact, for the Stokes problem the arguments of the previous chapter do not permit to select any direction. Further research is needed to clarify the situation.

Chapter 5

The incompressible Navier Stokes problem

In this chapter we extend the stabilized finite element approximation developed in the previous chapters for the CDR and Oseen problems to the incompressible Navier-Stokes equations. Two aspects of the problem make it different from the ones considered in previous chapters: it is a time dependent problem and it is non linear. We explore the properties of the discrete formulation that results allowing the subgrid-scales to depend on time. This apparently "natural" idea avoids several inconsistencies of previous formulations. Likewise, we consider the complete multiscale decomposition of the nonlinear term, following the variation of the subscale along the iterative process. This also "natural" idea gives rise to a discrete formulation with enhanced properties.

5.1 Introduction

Let us start by writing the incompressible Navier-Stokes equations. Consider a domain Ω in $\mathbb{R}^{\mathrm{d}}$, where $\mathrm{d} = 2, 3$ is the number of space dimensions, with boundary $\Gamma = \partial\Omega$, in which we want to solve an incompressible flow problem in the time interval $[0, T]$. If $\boldsymbol{u}$ is the velocity of the fluid and p the pressure, the incompressible Navier-Stokes equations are

$$\partial_t \boldsymbol{u} - \nu\Delta\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad \text{in } \Omega,\ t \in (0,T) \tag{5.1}$$

$$\nabla\cdot\boldsymbol{u} = 0 \qquad \text{in } \Omega,\ t \in (0,T) \tag{5.2}$$

where ν is the kinematic viscosity and $\boldsymbol{f}$ is the force vector. These equations must be supplied with an initial condition of the form $\boldsymbol{u} = \boldsymbol{u}^0$ in $\Omega,\ t = 0$, and a boundary condition which, for simplicity, will be taken as $\boldsymbol{u} = \boldsymbol{0}$ on $\Gamma,\ t \in (0,T)$.

Let us introduce some standard notation. The space of functions whose p power $(1 \le p < \infty)$ is integrable in a domain ω is denoted by $L^p(\omega)$, $L^\infty(\omega)$ being the space of

bounded functions in ω. The space of functions whose distributional derivatives of order up to $m \geq 0$ (integer) belong to $L^2(\omega)$ is denoted by $H^m(\omega)$. The space $H_0^1(\omega)$ consists of functions in $H^1(\omega)$ vanishing on $\partial\omega$. The topological dual of $H_0^1(\omega)$ is denoted by $H^{-1}(\omega)$. A bold character is used to denote the vector counterpart of all these spaces. If f and g are functions (or distributions) such that $f\,g$ is integrable in the domain ω under consideration, we denote

$$\langle f, g \rangle_\omega = \int_\omega f\, g \, \mathrm{d}\omega$$

so that, in particular, $\langle \cdot, \cdot \rangle_\omega$ is the duality pairing between $H^{-1}(\omega)$ and $H_0^1(\omega)$. When $f, g \in L^2(\omega)$, we write the inner product as $\langle f, g \rangle_\omega \equiv (f, g)_\omega$. The norm in a Banach space X is denoted by $\|\cdot\|_X$, and $L^p(0,T;X)$ is the space of time dependent functions such that their X-norm is $L^p(0,T)$. This notation is simplified in some cases as follows: $(\cdot,\cdot)_\Omega \equiv (\cdot,\cdot)$, $\langle \cdot, \cdot \rangle_\Omega \equiv \langle \cdot, \cdot \rangle$ and $\|\cdot\|_{L^2(\Omega)} \equiv \|\cdot\|$.

Using this notation, the velocity and pressure finite element spaces for the continuous problem are $\boldsymbol{L}^2(0,T;\boldsymbol{V}^{\mathrm{st}})$ and $L^1(0,T;Q^{\mathrm{st}})$, respectively, where $\boldsymbol{V}^{\mathrm{st}} := \boldsymbol{H}_0^1(\Omega)$, $Q^{\mathrm{st}} := L^2(\Omega)/\mathbb{R}$. The weak form of the problem consists in finding $[\boldsymbol{u}, p] \in \boldsymbol{L}^2(0,T;\boldsymbol{V}^{\mathrm{st}}) \times L^1(0,T;Q^{\mathrm{st}})$ such that

$$(\partial_t \boldsymbol{u}, \boldsymbol{v}) + \nu(\nabla \boldsymbol{u}, \nabla \boldsymbol{v}) + \langle \boldsymbol{u} \cdot \nabla \boldsymbol{u}, \boldsymbol{v} \rangle - (p, \nabla \cdot \boldsymbol{v}) = \langle \boldsymbol{f}, \boldsymbol{v} \rangle \tag{5.3}$$

$$(q, \nabla \cdot \boldsymbol{u}) = 0 \tag{5.4}$$

for all $[\boldsymbol{v}, q] \in \boldsymbol{V}^{\mathrm{st}} \times Q^{\mathrm{st}}$, and satisfying the initial condition in a weak sense.

The Galerkin finite element approximation of problem 5.3-5.4 consists in seeking the unknowns in finite dimensional spaces $\boldsymbol{V}_h \subset \boldsymbol{V}^{\mathrm{st}}$ and $Q_h \subset Q^{\mathrm{st}}$ and taking the test functions also in these spaces. Using the method of lines, the problem discretized in space, but still continuous in time, consists in finding $[\boldsymbol{u}_h(t), p_h(t)] \in \boldsymbol{L}^2(0,T;\boldsymbol{V}_h) \times L^1(0,T;Q_h)$ such that

$$(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) + \langle \boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle - (p_h, \nabla \cdot \boldsymbol{v}_h) = \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle \tag{5.5}$$

$$(q_h, \nabla \cdot \boldsymbol{u}_h) = 0 \tag{5.6}$$

for all $[\boldsymbol{v}_h, q_h] \in \boldsymbol{V}_h \times Q_h$.

Once discretized in time (using for example a finite difference scheme), it is well known that problem 5.5-5.6 suffers from different types of numerical instabilities. Two of them are inherited from the stationary problem, namely, the dominance of the (nonlinear) convective term over the viscous one when ν is small and the compatibility required for the velocity and pressure finite element spaces posed by the inf-sup condition. There are also numerical instabilities encountered when the time step size of the time discretization is small, particularly in early stages of the time integration.

A vast literature exists dealing with the instabilities due to the dominance of convection and to the velocity-pressure compatibility condition. In this work we adopt a

stabilized finite element formulation based on the subgrid-scale concept and, in particular, in the approach introduced by Hughes in [75, 78] for the scalar convection–diffusion equation. The basic idea is to approximate the *effect* of the component of the continuous solution which can not be resolved by the finite element mesh, which we will call *subscale*, on the discrete finite element solution. This approach is a general framework in which it is possible to design different stabilized formulations. We will restrict our attention to two approaches, described in [28] and [29]. In the first case, the velocity and pressure subscales are taken proportional to the residual of the finite element component in the momentum and in the continuity equations, respectively. The bottom line of the second approach is to take only the component of these residuals L^2 orthogonal to the finite element space. This idea was first introduced in [28] as an extension of a stabilization method originally introduced for the Stokes problem in [30] and fully analyzed for the stationary Navier-Stokes equations in [31].

However, the main interest of this chapter is *not* how to stabilize convection-dominated flows or how to be able to use equal velocity-pressure interpolation, thus avoiding the need to satisfy the inf-sup condition that problem 5.5-5.6 demands. Our objective in this chapter is *to analyze the formulation that stems from considering time dependent subscales*. In fact, the idea we will follow is not new, and was already introduced in [29]. In this sense, the present work can be considered as a continuation of this reference.

The chapter is organized as follows. The numerical formulation is described in Section 5.2, and its main features are presented in Sections 5.3 where we detail the benefits of considering the subscales time dependent, and how some of the misbehaviors of classical stabilized finite element methods are overcome. We also end Section 5.3 with a speculative subsection considering the tracking of subscales along the nonlinear process as a way to model turbulence. This idea was also pointed out in [29]. In Section 5.4 we present the results of three simple numerical examples that show the benefits of our approach and we conclude with some final remarks in Section 5.5.

5.2 Stabilized finite element problem

Let us consider a finite element partition $\mathcal{P}_h = \{K\}$ of the domain Ω with n_{el} elements. We will assume that *all the finite element spaces constructed are continuous* and of the same order for the velocity and the pressure. The starting idea of the formulation we propose is the variational multiscale formulation proposed in [75, 78]. Let $\boldsymbol{V}^{\text{st}} = \boldsymbol{V}_h \oplus \widetilde{\boldsymbol{V}}$, where $\boldsymbol{V}_h$ is the velocity finite element space and $\widetilde{\boldsymbol{V}}$ any space to complete $\boldsymbol{V}_h$ in $\boldsymbol{V}^{\text{st}}$. Similarly, let $Q^{\text{st}} = Q_h \oplus \widetilde{Q}$. The original continuous problem 5.3-5.4 is equivalent to find $[\boldsymbol{u}_h(t), p_h(t)] \in$

$\boldsymbol{L}^2(0,T;\boldsymbol{V}_h)\times L^1(0,T;Q_h)$, as well as $[\widetilde{\boldsymbol{u}}(t),\widetilde{p}(t)]\in\ \boldsymbol{L}^2(0,T;\widetilde{\boldsymbol{V}})\times L^1(0,T;\widetilde{Q})$, such that

$$(\partial_t(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}),\boldsymbol{v})+\nu(\nabla(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}),\nabla\boldsymbol{v})$$
$$+\langle(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}})\cdot\nabla(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}),\boldsymbol{v}\rangle-(p_h+\widetilde{p},\nabla\cdot\boldsymbol{v})=\langle\boldsymbol{f},\boldsymbol{v}\rangle \tag{5.7}$$
$$(q,\nabla\cdot(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}))=0 \tag{5.8}$$

for all $[\boldsymbol{v},q]\in\boldsymbol{V}^{\rm st}\times Q^{\rm st}$. These equations can be split into two systems by taking first $[\boldsymbol{v},q]=[\boldsymbol{v}_h,q_h]\in\boldsymbol{V}_h\times Q_h$ and then $[\boldsymbol{v},q]=[\widetilde{\boldsymbol{v}},\widetilde{q}]\in\widetilde{\boldsymbol{V}}\times\widetilde{Q}$. Denoting by $\boldsymbol{n}$ the exterior unit normal to an integration domain, after integrating some terms by parts the first choice leads to

$$(\partial_t(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}),\boldsymbol{v}_h)+\nu(\nabla\boldsymbol{u}_h,\nabla\boldsymbol{v}_h)$$
$$+\nu\sum_K[-(\widetilde{\boldsymbol{u}},\Delta\boldsymbol{v}_h)_K+\langle\widetilde{\boldsymbol{u}},\ \boldsymbol{n}\cdot\nabla\boldsymbol{v}_h\rangle_{\partial K}]$$
$$+\langle(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}})\cdot\nabla\boldsymbol{u}_h,\boldsymbol{v}_h\rangle-\langle\widetilde{\boldsymbol{u}},(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}})\cdot\nabla\boldsymbol{v}_h\rangle-(p_h+\widetilde{p},\nabla\cdot\boldsymbol{v}_h)=\langle\boldsymbol{f},\boldsymbol{v}_h\rangle \tag{5.9}$$
$$(q_h,\nabla\cdot\boldsymbol{u}_h)-(\nabla q_h,\widetilde{\boldsymbol{u}})=0 \tag{5.10}$$

where we have used the fact that $\nabla\cdot(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}})=0$, that the sum of the integral of $\boldsymbol{n}\cdot(\boldsymbol{u}_h+\widetilde{\boldsymbol{u}})$ on the boundaries of two adjacent elements (and thus with opposite normal $\boldsymbol{n}$) must be zero and that $\boldsymbol{u}_h=\widetilde{\boldsymbol{u}}=\boldsymbol{0}$ on Γ.

The second system is obtained by taking $[\boldsymbol{v},q]=[\widetilde{\boldsymbol{v}},\widetilde{q}]\in\widetilde{\boldsymbol{V}}\times\widetilde{Q}$ in 5.7-5.8. Of course, the resulting system, together with 5.9-5.10, is exactly equivalent to 5.3-5.4. A *stabilized finite element method* is obtained if $\widetilde{\boldsymbol{u}}$ and $\widetilde{p}$ *are approximated* and their expression inserted into 5.9-5.10. However it is not our purpose in this chapter to emphasize how to obtain the approximations for $\widetilde{\boldsymbol{u}}$ and $\widetilde{p}$ because this problem has been considered in previous chapters (and still needs further research). Our purpose here is

- To allow $\widetilde{\boldsymbol{u}}$ to be time dependent, and therefore to keep its time dependency in 5.9.
- To note that the advection velocity in 5.9 is $\boldsymbol{u}_h+\widetilde{\boldsymbol{u}}$, and not only $\boldsymbol{u}_h$.

In fact, we will not explore in detail the second item. Some comments about this point will be made later on. Our main concern will be to study the properties of the numerical formulation that emanates from considering $\widetilde{\boldsymbol{u}}$ time dependent. For this purpose, it is enough to make some simplifying assumptions:

- The term involving integrals over interelement boundaries will be neglected. This can be understood as considering the velocity subscales as bubble functions, vanishing on the boundaries of the elements (see, e.g., [4, 16]). Even though its consideration can bring important stabilization properties, it is not essential for what follows.

- The approximation of the subgrid-scales is performed as follows. The system for the subscales $[\widetilde{\boldsymbol{u}}(t), \widetilde{p}(t)]$, obtained taking $[\boldsymbol{v}, q] = [\widetilde{\boldsymbol{v}}, \widetilde{q}] \in \widetilde{\boldsymbol{V}} \times \widetilde{Q}$, can be understood as

$$\begin{aligned} \partial_t \widetilde{\boldsymbol{u}} + (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \widetilde{\boldsymbol{u}} - \nu \Delta \widetilde{\boldsymbol{u}} + \nabla \tilde{p} &= \boldsymbol{R}_m \\ \nabla \cdot \widetilde{\boldsymbol{u}} &= R_c \end{aligned}$$

  where $\boldsymbol{R}_m$ and R_c are appropriate residuals of the finite element components $\boldsymbol{u}_h$ and p_h adequately projected onto the space of subscales ($\widetilde{\boldsymbol{V}}$ for the first equation and $\widetilde{Q}$ for the second). Using the arguments of chapter 4, the following approximation to the previous equations can be motivated:

$$\partial_t \widetilde{\boldsymbol{u}} + \frac{1}{\tau_m} \widetilde{\boldsymbol{u}} = \boldsymbol{R}_m \tag{5.11}$$

$$\frac{1}{\tau_c} \widetilde{p} = R_c + \tau_m \partial_t R_c \tag{5.12}$$

  where

$$\tau_m = \left[c_1 \frac{\nu}{h^2} + c_2 \frac{|\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}|}{h} \right]^{-1} \tag{5.13}$$

$$\tau_c = \frac{h^2}{c_1 \tau_m} \tag{5.14}$$

$$\boldsymbol{R}_m = -\mathcal{P} \left[\partial_t \boldsymbol{u}_h + (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \boldsymbol{u}_h - \nu \Delta \boldsymbol{u}_h + \nabla p_h - \boldsymbol{f} \right] \tag{5.15}$$

$$R_c = -\mathcal{P} \left[\nabla \cdot \boldsymbol{u}_h \right] \tag{5.16}$$

  This formulation is obtained if the differential terms of the stabilization operator presented in chapter 4 are neglected and the isotropic approximation to the stabilization parameters, in which $c_1 = 4$ and $c_2 = 2$, is considered. As in chapter 4, the projection $\mathcal{P}$ can be either the identity for "classical" stabilized finite element methods (which can be traced back to [18], for example) or the projection orthogonal to the finite element space (we have used the same symbol for the scalar and vector counterparts of this operator). As in previous chapters, we will refer to the choice $\mathcal{P} = I$ (identity) as the Algebraic Subgrid Scale formulation (ASGS), whereas $\mathcal{P} = \Pi_h^{\perp}$, Π_h being the L^2 projection onto the appropriate finite element space (of velocities or of pressures), will lead to the so called Orthogonal Subscales Stabilization (OSS).

  As is shown in [83], the fine-scale component of the solution is related to the residual of the coarse scales through so-called small-scale Green's function. It was also shown in [83] that the small-scale Green's function is highly localized for the right choice of the projector, rendering local algebraic approximations 5.11-5.12 a viable model for the fine scales.

Again, let us stress that the two assumptions described are not essential for our discussion and could be modified. The important point is that $\partial_t \widetilde{\boldsymbol{u}}$ appears in the

approximate equation for the velocity subscale. In our case, this approximation turns out to be the differential equation in time 5.11.

Remark 5 *Observe that equation 5.11 must hold at each point, and therefore it is in fact an* ordinary differential equation *rather than a partial differential equation.*

Remark 6 *Neglecting the time derivative in 5.11 could be understood as considering that the subscales adapt automatically to the finite element residual. The subscales obtained from this assumption were defined in [29] as* quasi-static*.*

Remark 7 *Observe that 5.11 is a nonlinear equation, due to the dependence of τ_m and $\boldsymbol{R}_m$ on $\widetilde{\boldsymbol{u}}$. Obviously, this does not depend on whether the subscales vary in time or not, and was also noticed in [21] for what we have called quasi-static subscales. In this case, it is possible to tackle directly the resulting nonlinear* algebraic *equation and solve for $\widetilde{\boldsymbol{u}}$ in terms of $\boldsymbol{R}_m$ accounting for this nonlinearity. However, in our case this is not possible, and we will have to linearize 5.11 to integrate it in time.*

The formulation we want to analyze is now complete. It consists of solving 5.9-5.10 *together with* 5.11-5.12 for $\boldsymbol{u}_h$, $\widetilde{\boldsymbol{u}}$, p_h and $\widetilde{p}$, neglecting the integrals over interelements boundaries, as it has been mentioned. Although it does not introduce any particular complication, as it can be observed from the analysis in [28, 29], we will take $\widetilde{p} = 0$ for the sake of simplicity (in fact, we have used expression 5.12 with τ_c given by 5.14 in the numerical examples of Section 5) . Therefore, the final problem we have to solve can be written as a single variational equation as follows: find $[\boldsymbol{u}_h(t), p_h(t)] \in \boldsymbol{L}^2(0,T;\boldsymbol{V}_h) \times L^1(0,T;Q_h)$ such that

$$\begin{aligned}(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) + \langle \boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h\rangle - (p_h, \nabla \cdot \boldsymbol{v}_h) \\ +(q_h, \nabla \cdot \boldsymbol{u}_h) - \sum_K \langle \widetilde{\boldsymbol{u}}, \nu\Delta \boldsymbol{v}_h + \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h + \nabla q_h\rangle_K \\ + (\partial_t \widetilde{\boldsymbol{u}}, \boldsymbol{v}_h) + \langle \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h\rangle - \langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h\rangle = \langle \boldsymbol{v}_h,\ \boldsymbol{f}\rangle \end{aligned} \tag{5.17}$$

for all $[\boldsymbol{v}_h, q_h] \in \boldsymbol{V}_h \times Q_h$, where $\widetilde{\boldsymbol{u}}$ is solution of the nonlinear differential equation 5.11, with τ_m given by 5.13 and $\boldsymbol{R}_m$ by 5.15. In what follows, we will rename $\tau_m \equiv \tau$.

Remark 8 *From the point of view of the implementation of the method, it is clear from 5.17 that $\widetilde{\boldsymbol{u}}$ is needed at the numerical integration points within each element. Therefore, 5.11 has to be integrated in time at each integration point. In this sense, $\widetilde{\boldsymbol{u}}$ acts as what would be called* internal variable *in solid mechanics.*

Remark 9 *If the subscales are assumed to be orthogonal to the finite element space, the term $(\partial_t \widetilde{\boldsymbol{u}}, \boldsymbol{v}_h)$ vanishes and, as explained in [29], the term $\sum_K \langle \widetilde{\boldsymbol{u}}, \nu\Delta \boldsymbol{v}_h + \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h + \nabla q_h\rangle_K$ can be replaced by $\sum_K \langle \widetilde{\boldsymbol{u}}, \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h + \nabla q_h\rangle_K$ and still keep the same accuracy of the method.*

Remark 10 *Problem 5.9-5.10 and 5.11-5.12 needs to be completed with initial conditions* $\boldsymbol{u}_h = \boldsymbol{u}_h^0$ *and* $\widetilde{\boldsymbol{u}} = \widetilde{\boldsymbol{u}}^0$ *at* $t = 0$*, where the functions* $\boldsymbol{u}_h^0$ *and* $\widetilde{\boldsymbol{u}}^0$ *depend on the way to choose the space of subscales. We assume that the projections onto the finite element space and the space of subscales are* L^2 *continuous (this is obvious if* $\mathcal{P} = \Pi_h^\perp$ *in 5.15), and therefore* $\|\boldsymbol{u}_h^0\| \leq C\|\boldsymbol{u}^0\|$, $\|\widetilde{\boldsymbol{u}}^0\| \leq C\|\boldsymbol{u}^0\|$ *for a certain constant* C*.*

5.3 Main features of the formulation

The left-hand-side of the discrete variational form of the problem given by 5.17 consists of the following terms:

$$
\begin{aligned}
&(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) + \langle \boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle \\
&- (p_h, \nabla \cdot \boldsymbol{v}_h) + (q_h, \nabla \cdot \boldsymbol{u}_h) - \langle \boldsymbol{v}_h, \ \boldsymbol{f} \rangle \qquad \text{Galerkin terms} \qquad (5.18) \\
&- \sum_K \langle \widetilde{\boldsymbol{u}}, \nu \Delta \boldsymbol{v}_h + \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h + \nabla q_h \rangle_K \qquad \text{Stabilization terms} \qquad (5.19) \\
&(\partial_t \widetilde{\boldsymbol{u}}, \boldsymbol{v}_h) + \langle \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle - \langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h \rangle \qquad \text{Effect of } \widetilde{u}(t) \qquad (5.20)
\end{aligned}
$$

The stabilization terms appear also in the stationary and linearized problem, and it is now well known that they allow to overcome the instability problems of the classical Galerkin formulation, which in this case are the instabilities found in convection dominated flows and the need to satisfy an inf-sup condition for the velocity and pressure interpolations.

The terms associated to the effect of $\widetilde{\boldsymbol{u}}$ in the material derivative are precisely those that come from accepting the decomposition $\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}$ in the expression of

$$
\begin{aligned}
\frac{\mathrm{D}}{\mathrm{D}t}\boldsymbol{u} &= \frac{\mathrm{D}}{\mathrm{D}t}(\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \\
&= \partial_t \boldsymbol{u}_h + \partial_t \widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h + \boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h + \widetilde{\boldsymbol{u}} \cdot \nabla \widetilde{\boldsymbol{u}} + \boldsymbol{u}_h \cdot \nabla \widetilde{\boldsymbol{u}}. \qquad (5.21)
\end{aligned}
$$

Only the last of these terms where $\widetilde{\boldsymbol{u}}$ appears contributes to the stabilization terms. Our objective is to discuss precisely the effect of the other terms contributed by $\widetilde{\boldsymbol{u}}$.

5.3.1 Commutation of space and time discretization

Let us start our discussion on the properties of the method just presented by noting that we have been able to formulate a *stabilized finite element method* without any reference to the time discretization. Usually, the problem of formulating stabilized methods for time dependent problems has been tackled using two main approaches:

- By using space-time finite element formulations, and considering the temporal derivative in the same way as the first order spatial derivatives of the convective term. This is the approach adopted for example in the early papers on this subject [91, 132].

- By discretizing first in time, and then using a stabilized finite element method for the resulting spatially-continuous problem. This is perhaps the most popular approach in the literature. The design of the time integration scheme is in principle independent of the stabilization formulation used, but can be adapted to improve the behavior in time of the solution (see, e.g., [88]).

Space time formulations of order higher than one require predictor-corrector strategies to avoid an unacceptable increase in the number of unknowns treated at once (see, e.g. [132]). On the other hand, first order methods, with piecewise constant interpolations in time, lead to very poor schemes, that need to be modified *a posteriori* to improve their accuracy [87]. In particular, it turns out to be essential to include an approximation of the time derivative in the residual given by 5.15. This comes out naturally if the equations are first discretized in time using a finite difference scheme.

Nevertheless, in the subgrid scale formulation we are analyzing, the fact of considering the subscales time dependent allows us either to start from the time discrete problem, as in [29], or to use a method of lines, discretizing first in space and then in time, which is the approach we are following here. Both methods *will lead exactly to the same fully discrete scheme, that is to say, space and time discretization commute*, even when using finite difference schemes in time. In general, this property is trivial only for stabilized methods that do not involve the residual of the equations to be solved, as the method proposed in [20] or even the stabilization with quasi-static orthogonal subscales [29].

Let us consider now which would be a finite difference time discretization of problem 5.17, with $\tilde{\boldsymbol{u}}$ solution of 5.11. To fix ideas, let us apply the generalized trapezoidal rule. Consider a uniform finite element partition of $[0,T]$ of size δt, and for a time dependent function f let f^n denote an approximation to it at $t^n = n\delta t$, $\delta f^n := f^{n+1} - f^n$, $\delta_t f^n := \delta f^n/\delta t$ and $f^{n+\theta} = \theta f^{n+1} + (1-\theta) f^n$, with $1/2 \leq \theta \leq 1$. The generalized trapezoidal rule applied to 5.17 leads to the following fully discrete variational problem: given $\boldsymbol{u}_h^n$ and $\widetilde{\boldsymbol{u}}^n$, find $\boldsymbol{u}_h^{n+1}$, p_h^{n+1} and $\widetilde{\boldsymbol{u}}^{n+1}$ by solving

$$
\begin{aligned}
(\delta_t \boldsymbol{u}_h^n, \boldsymbol{v}_h) + \nu(\nabla \boldsymbol{u}_h^{n+\theta}, \nabla \boldsymbol{v}_h) + \langle \boldsymbol{u}_h^{n+\theta} \cdot \nabla \boldsymbol{u}_h^{n+\theta}, \boldsymbol{v}_h \rangle - (p_h^{n+1}, \nabla \cdot \boldsymbol{v}_h) \\
+ (q_h, \nabla \cdot \boldsymbol{u}_h^{n+\theta}) - \sum_K \langle \widetilde{\boldsymbol{u}}^{n+\theta}, \nu \Delta \boldsymbol{v}_h + \boldsymbol{u}_h^{n+\theta} \cdot \nabla \boldsymbol{v}_h + \nabla q_h \rangle_K \\
+ (\delta_t \widetilde{\boldsymbol{u}}^n, \boldsymbol{v}_h) + \langle \widetilde{\boldsymbol{u}}^{n+\theta} \cdot \nabla \boldsymbol{u}_h^{n+\theta}, \boldsymbol{v}_h \rangle - \langle \widetilde{\boldsymbol{u}}^{n+\theta}, \widetilde{\boldsymbol{u}}^{n+\theta} \cdot \nabla \boldsymbol{v}_h \rangle = \langle \boldsymbol{v}_h, \ \boldsymbol{f}^{n+\theta} \rangle \quad (5.22) \\
\delta_t \widetilde{\boldsymbol{u}}^n + \frac{1}{\tau^{n+\theta}} \widetilde{\boldsymbol{u}}^{n+\theta} = \boldsymbol{R}_m^{n+\theta} \quad (5.23)
\end{aligned}
$$

for all $[\boldsymbol{v}_h, q_h] \in \boldsymbol{V}_h \times Q_h$ (we have assumed $\boldsymbol{f}$ continuous in time, otherwise $\boldsymbol{f}^{n+\theta}$ has to be understood as a time average between t^n and t^{n+1}). In 5.23 it is understood that the time derivative in $\boldsymbol{R}_m^{n+\theta}$ is already discretized. From this equation we can obtain $\widetilde{\boldsymbol{u}}^{n+\theta}$ and insert it into 5.22. Obviously, the result will depend on $\widetilde{\boldsymbol{u}}^n$, and thus the subscales *need to be tracked in time.*

Equation 5.23 can be considered the "natural" choice for the time integration of the equation for the subscales, in the sense that they are integrated using the same scheme as the finite element component of the velocity. Likewise, if we had first discretized the continuous Navier-Stokes equations in time and then applied the splitting $\boldsymbol{u}^n = \boldsymbol{u}_h^n + \widetilde{\boldsymbol{u}}^n$ we would have arrived also to 5.22-5.23 (with the adequate modeling of the subscales). However, there is also the possibility of using a different time integration for $\boldsymbol{u}_h$ and $\widetilde{\boldsymbol{u}}$. For example, assuming given a guess for $\widetilde{\boldsymbol{u}}^{n+1}$ to evaluate τ^{n+1} and $\boldsymbol{R}_m^{n+1}$, within the time interval $[t^n, t^{n+1}]$ we could consider the *time continuous* equation for $\widetilde{\boldsymbol{u}}$

$$\partial_t \tilde{\boldsymbol{u}} + \frac{1}{\tau^{n+\alpha}} \tilde{\boldsymbol{u}} = \boldsymbol{R}_m^{n+\alpha}$$

with $0 \leq \alpha \leq 1$, which can be integrated to yield

$$\widetilde{\boldsymbol{u}}^{n+1} = \left(\widetilde{\boldsymbol{u}}^n - \tau^{n+\alpha}\, \boldsymbol{R}_m^{n+\alpha}\right) \exp\left(-\frac{\delta t}{\tau^{n+\alpha}}\right) + \tau^{n+\alpha}\, \boldsymbol{R}_m^{n+\alpha}. \tag{5.24}$$

Remember that both $\tau^{n+\alpha}$ and $\boldsymbol{R}_m^{n+\alpha}$ depend on $\widetilde{\boldsymbol{u}}^{n+1}$, and therefore 5.24 is a nonlinear algebraic equation for this subscale (except if $\alpha = 0$, of course), which can be solved for example using the strategy proposed in [21], or simply linearized and solved iteratively.

Remark 11 *Even though we are considering $\frac{1}{2} \leq \theta \leq 1$, 5.23 makes sense also for $\theta = 0$ (explicit integration of the subscales), case in which it yields $\widetilde{\boldsymbol{u}}^{n+1} = (1 - \delta t/\tau^n)\widetilde{\boldsymbol{u}}^n + \delta t\, \boldsymbol{R}_m^{n+\alpha}$. This expression corresponds also to 5.24 with $\alpha = 0$ and expanding the exponential to first order in $\delta t/\tau^n$.*

5.3.2 Why τ must depend on δt (but this is not enough)

Let us consider equation 5.23 and re-write it as

$$\widetilde{\boldsymbol{u}}^{n+\theta} = \left(\frac{1}{\theta\delta t} + \frac{1}{\tau^{n+\theta}}\right)^{-1} \left(\boldsymbol{R}_m^{n+\theta} + \frac{1}{\theta\delta t}\widetilde{\boldsymbol{u}}^n\right) \tag{5.25}$$

From this expression we see that the residual of the momentum equation is multiplied by

$$\tau_t := \left(\frac{1}{\theta\delta t} + \frac{1}{\tau^{n+\theta}}\right)^{-1} \tag{5.26}$$

This is what can be considered the stabilization parameter for the transient incompressible Navier-Stokes equations. Expressions with asymptotic behavior similar to 5.26 in terms of h, ν, $|\boldsymbol{u}_h|$ *and* δt can be often found in the literature (see, e.g. [132, 138]). The way to motivate it can be explained in a simplified way by saying that the temporal derivative of the velocity is considered as a reaction-like term (with a zero order derivative) with factor $1/(\theta\delta t)$, after considering for a given time step the equations discretized in time. This explanation can be found for example in [58], or in [85], where it motivates a careful design of the stabilization parameters for reaction dominated problems.

In reference [11] there is a study of the *instability* encountered when the ASGS method is used and 5.25 is replaced by the simplified equation

$$\widetilde{\boldsymbol{u}}^{n+\theta} = \tau^{n+\theta}\ \boldsymbol{R}_m^{n+\theta} \tag{5.27}$$

that corresponds to what we have called quasi-static subscales. It is shown in the reference mentioned that for the Stokes time continuous problem the Schur complement matrix for the pressure is not uniformly invertible, and this property is inherited as $\delta t \to 0$ if h, and therefore $\tau^{n+\theta}$, remains fixed (the case $\theta = 1$ is considered in [11]).

It is easily shown that the instability described disappears if

$$\delta t \geq C\tau^{n+\theta} \tag{5.28}$$

where C is a positive constant. This is a condition that appears very often and about which there are several remarks to be made:

- As it has been mentioned, under condition 5.28 the instability problems described in [11] for the ASGS method do not appear. This condition prevents the possibility of letting $\delta t \to 0$ while keeping h fixed.
- In fact, if 5.28 holds it is irrelevant from the analysis point of view if the residual in 5.25 is multiplied by τ_t defined in 5.26 or simply by $\tau^{n+\theta}$, since this parameter and τ_t have the same asymptotic behavior in terms of h, ν and $|\boldsymbol{u}_h|$.
- Condition 5.28 was needed in the analysis of the stabilization with orthogonal subscales for the convection-diffusion equation analyzed in [32], also considering time dependent subscales.

From this discussion it seems clear that the stabilization parameter and the time step size *must be related in classical stabilized finite element methods*. This is clear from the heuristic arguments presented in the references mentioned above, the instability described in [11] for the ASGS method and the reasons found to comply with condition 5.28 just mentioned. However, we have not mentioned yet the fact that *in 5.25 we are tracking the subscales in time*. This has two major benefits, which justifies why taking the stabilization parameter as indicated by 5.26 *is not enough:*

- If, as it is done in [58, 132, 138], among other references, the stabilization parameter adopted has an expression similar to 5.26 but the subscales are not considered time dependent, *the steady-state solution depends on the time step size*. This is clearly not an optimal situation. The amount of stabilization will depend on the way the equations are integrated to the steady-state. This does not happen if expression 5.25 is used. It can be easily checked that, when the steady-state is reached, 5.27 (now without any superscript) is recovered.

- Stability *for all δt and h, without any need to satisfy 5.28* can be obtained for the linearized Navier Stokes equations[36]. Further, a complete convergence analysis of the transient approximation to the Stokes problem can be found in [3]. This is particularly relevant, since it allows us to use *arbitrary combinations of h and δt*. In other words, we may use what could be called *anisotropic* space-time discretizations. Of course, it is possible to use directly 5.26 without considering time-dependent subscales, and in that case 5.28 is automatically verified. However, that would lead to stability estimates that become meaningless *in space* when $\delta t \to 0$.

5.3.3 Tracking of subscales along the nonlinear process

Up to now we have considered the effect of the term $(\partial_t \widetilde{\boldsymbol{u}}, \boldsymbol{v}_h)$ in 5.18 and of $\partial_t \widetilde{\boldsymbol{u}}$ in 5.11. In this subsection we describe the effect of the other two terms in 5.18. Summarizing, $\langle \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle$ allows us to guarantee global conservation of momentum, whereas $-\langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h \rangle$ may be understood as the term coming from the subgrid scale tensor in a LES approach.

Conservation of momentum

Let us start by analyzing the effect of $\langle \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle$. The purpose of what follows is to present a version of the results in [84], simplified and adapted to the present setting.

Let $\boldsymbol{V}_h^d$ the velocity finite element space without imposing the Dirichlet boundary conditions, that is, with degrees of freedom also associated to the boundary nodes. Let $\boldsymbol{t}$ be the stress vector (traction) on the boundary Γ and consider the following augmented problem instead of 5.17:

$$
\begin{aligned}
(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) + \langle \boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle - (p_h, \nabla \cdot \boldsymbol{v}_h) \\
+(q_h, \nabla \cdot \boldsymbol{u}_h) - \sum_K \langle \widetilde{\boldsymbol{u}}, \nu \Delta \boldsymbol{v}_h + \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h + \nabla q_h \rangle_K \\
+ (\partial_t \widetilde{\boldsymbol{u}}, \boldsymbol{v}_h) + \langle \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle - \langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h \rangle = \langle \boldsymbol{v}_h, \ \boldsymbol{f} \rangle + \langle \boldsymbol{v}_h, \ \boldsymbol{t} \rangle_\Gamma
\end{aligned}
$$

where now $\boldsymbol{v}_h \in \boldsymbol{V}_h^d$ (not just $\boldsymbol{V}_h$). Considering $d = 3$ and taking for example $\boldsymbol{v}_h = (1, 0, 0)$ and $q_h = 0$, this equation yields

$$
\begin{aligned}
\int_\Omega \partial_t (u_{h,1} + \tilde{u}_1) \mathrm{d}\Omega + \int_\Gamma u_{h,1} \boldsymbol{u}_n \cdot \boldsymbol{n} \mathrm{d}\Gamma - \int_\Omega u_{h,1} \nabla \cdot \boldsymbol{u}_h \mathrm{d}\Omega \\
+ \int_\Omega \widetilde{\boldsymbol{u}} \cdot \nabla u_{h,1} \mathrm{d}\Omega \ = \ \int_\Omega f_1 \mathrm{d}\Omega + \int_\Gamma t_1 \mathrm{d}\Gamma
\end{aligned}
$$

where now the zero Dirichlet condition for the velocity is not explicitly required. This statement provides *global* momentum conservation if

$$
- \int_\Omega u_{h,1} \nabla \cdot \boldsymbol{u}_h \mathrm{d}\Omega + \int_\Omega \widetilde{\boldsymbol{u}} \cdot \nabla u_{h,1} \mathrm{d}\Omega = 0. \tag{5.29}
$$

This is implied by the continuity equation obtained by taking $\boldsymbol{v}_h = \boldsymbol{0}$

$$(q_h, \nabla \cdot \boldsymbol{u}_h) - \sum_K \langle \widetilde{\boldsymbol{u}}, \nabla q_h \rangle_K = 0, \tag{5.30}$$

provided $V_h/\mathbb{R} \subseteq Q_h$, that is to say, the velocity component $u_{h,1}$ belongs to the pressure space ($u_{h,1}$ can be considered modulo constants, since they do not affect neither the first nor the second terms in 5.29). This holds, in particular, for the "natural" choice $V_h/\mathbb{R} = Q_h$, that is to say, equal velocity-pressure interpolations. For the standard Galerkin method, *this condition is impossible to be satisfied*, since equal interpolation does not satisfy the inf-sup condition. As a conclusion, *the term* $\langle \tilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle$ *provides global momentum conservation*, since without it in the discrete momentum equation, we would have obtained $-\int_\Omega u_{h,1} \nabla \cdot \boldsymbol{u}_h \mathrm{d}\Omega = 0$ instead of 5.29, which is not implied by 5.30.

A door to turbulence

Let us conclude this section with some speculative comments on the contribution of the term $-\langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h \rangle$. In the standard large eddy simulation (LES) approach to solve turbulent flows (see e.g., [122], [130]) an equation is obtained for the large, filtered scales of the flow, which we will denote with an overbar. This equation includes an extra term when compared with the incompressible Navier-Stokes equations 5.1-5.2: the divergence of the so-called *residual stress tensor* or *subgrid scale tensor* $\boldsymbol{R} := \overline{\boldsymbol{u} \otimes \boldsymbol{u}} - \overline{\boldsymbol{u}} \otimes \overline{\boldsymbol{u}}$. Tensor $\boldsymbol{R}$ has to be modeled in terms of $\overline{\boldsymbol{u}}$ to obtain a self-contained equation, a problem known as the *closure problem*, and, once this is done, the resulting LES equation can be solved numerically.

The residual stress tensor, $\boldsymbol{R}$, is often decomposed into the so-called Reynolds, Cross and Leonard stresses to keep the Galilean invariance of the original Navier-Stokes equation in the LES equation. This invariance is automatically inherited by the formulation presented in this work and we observe that analogous terms to the various stress types are recovered in a "natural" way from our pure numerical approach (this was also the case in [82]). Let us have a look at this point. We first consider the last four terms in the material derivative 5.21 as they appear in the variational equation 5.17. The term $-\langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h \rangle$ can be rewritten as

$$-\langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h \rangle = -\langle \widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}}, \nabla \boldsymbol{v}_h \rangle$$

and can be identified with the Reynolds stress. The addition of the other three terms becomes, after integration by parts,

$$\langle \boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle - \langle \widetilde{\boldsymbol{u}}, \boldsymbol{u}_h \cdot \nabla \boldsymbol{v}_h \rangle + \langle \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle = -\langle \boldsymbol{u}_h \otimes \boldsymbol{u}_h, \nabla \boldsymbol{v}_h \rangle - \langle \boldsymbol{u}_h \otimes \widetilde{\boldsymbol{u}} + \widetilde{\boldsymbol{u}} \otimes \boldsymbol{u}_h, \nabla \boldsymbol{v}_h \rangle$$

and we can identify the second term on the right hand side with the cross stress. If we now pay attention to the convective term of the residual in the subscale equation 5.11

and take, for simplicity, $\mathcal{P} = I$, we observe that

$$\langle (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \boldsymbol{u}_h, \widetilde{\boldsymbol{v}} \rangle = - (\boldsymbol{u}_h \otimes \boldsymbol{u}_h, \nabla \widetilde{\boldsymbol{v}}) - (\boldsymbol{u}_h \otimes \widetilde{\boldsymbol{u}}, \nabla \widetilde{\boldsymbol{v}})$$

and the first term on the right hand side can be identified with the Leonard stress. Hence, we can effectively conclude that the modifications introduced by the presence of the divergence of $\boldsymbol{R}$ in the LES equations are somehow automatically included in our subgrid scale stabilized finite element approach.

How good our formulation will work as a turbulent model will mainly depend on the validity of the approximation made to derive the evolution equation for the subscales 5.11, being the ASGS or the OSS methods two available possibilities. In order to check this performance, benchmark problems for turbulent flows should be used. A widely used benchmark problem is the decay of isotropic turbulence. Our model should be able to reproduce the Kolmogorov energy cascade in the wavenumber Fourier space that displays an inertial range, where $E(k,t) \sim C_K \varepsilon^{2/3} k^{-5/3}$ (ε being the energy dissipation rate, k the wavenumber modulus, C_K the Kolmogorov constant in energy space and E the kinetic energy). The model should be also able to capture the appropriate decay in time of energy, enstrophy and other related statistical variables. Other more intricate questions such as if the model allows for backscatter or if the dimension of the global attractor is properly reproduced could be also addressed. We remind that the heuristic estimate for this dimension is $\mathcal{N} \sim (L/\lambda_K)^3 \sim \mathrm{Re}^{\frac{9}{4}}$ (where λ_K is the Kolmogorov length scale) and that the closest estimate analytically proved is (roughly) $(L/\lambda_K)^{4.8}$ (see [59]). Another standard test for turbulence is the turbulent channel flow. In this case the model should be able to approximate the turbulent boundary layer that, according to Prandtl theory, exhibits a log behavior after the laminar sublayer. Finally, we should mention that in an attempt to find a more mathematical foundation for the LES approach to turbulence, the concept of *suitable approximations* to the Navier-Stokes equations has been introduced in [65, 66]. It is expected that approximate solutions converge (in a weak sense) to *suitable solutions*. This seems to be the case for low order finite elements and the *standard Galerkin method* [64]. Hopefully, our *enhanced formulations* will have this property.

The original idea of using the multiscale formulation with local approximation to the fine scales to compute turbulent flows was already pointed out in [29] and elaborated in [77, 21]. Very good results were obtained for fully developed and transitional turbulent flows. In fact, some promising results of numerical simulation of turbulent flows *only* with stabilization can be found in [73, 39](see also [62] for a review).

5.4 Numerical examples

In this section we present three simple numerical examples that illustrate the performance of the method. The first is a convergence test that shows that for solutions with a smooth

behavior in time both quasi-static and transient subscales lead to the same optimal convergence rate. In the second example we demonstrate the improvement obtained when the subscales are tracked in time in the example introduced in [11]. Finally, the last example is the classical flow over a cylinder, for which considering transient subscales leads to better results, both in terms of accuracy (with higher amplitudes and frequencies, that is, less numerical dissipation) and of stability, eliminating some pressure oscillations in time encountered when the subscales are considered quasi-static. In all the cases we have used the ASGS method, that is, $\mathcal{P} = I$ (identity) in 5.15-5.16.

5.4.1 A convergence test

In this example, already presented in [27], we consider the time dependent Navier-Stokes equations in the unit square with homogeneous Dirichlet boundary conditions and taking the force $\boldsymbol{f}$ and boundary and initial values to have the exact solution defined by

$$\boldsymbol{u} = 100h(t)\left(f(x)f'(y), -f'(x)f(y)\right), \qquad p = 100x^2,$$

where

$$h(t) = \cos(\pi t)\,\mathrm{e}^{-t}, \qquad f(x) = x^2(1-x)^2.$$

Uniform meshes of 10×10, 20×20, 40×40 and 80×80 bilinear elements have been used to discretize the computational domain. The time interval of the analysis is $[0, 1]$ and the viscosity is 0.1.

The objective of this test is to check the convergence of the time approximation to the exact solution using the method proposed here. To this end we compare the results obtained using transient subscales (TRS) to those obtained using quasi-static subscales (QSS) (see Remark 2). We compute the error as the discrete approximation to the L^2 norm of the difference between the exact and the approximated solution at time $t = 1$ and we normalize it using the discrete approximation to the L^2 norm of the exact solution. Numerical experiments have been performed using a first and a second order temporal discretization (Crank Nicolson scheme) and several time step sizes. In the case of the second order approximation we have also considered a first and second order time integration of equation 5.23. The convergence of the velocity approximation is shown in figure 5.1, from where it is seen that stabilized approximation converges to the exact solution at the expected rate either using the time dependent or the quasi-static subscales (see Remark 9). We also note that the integration of the subgrid scale equation 5.23 using a first or a second order method has little influence on the results.

5.4.2 Stability in the small time step limit

The second example, presented in [11], shows the instability of the approximation to the Stokes problem when quasi-static subscales are considered (recall that we are using the

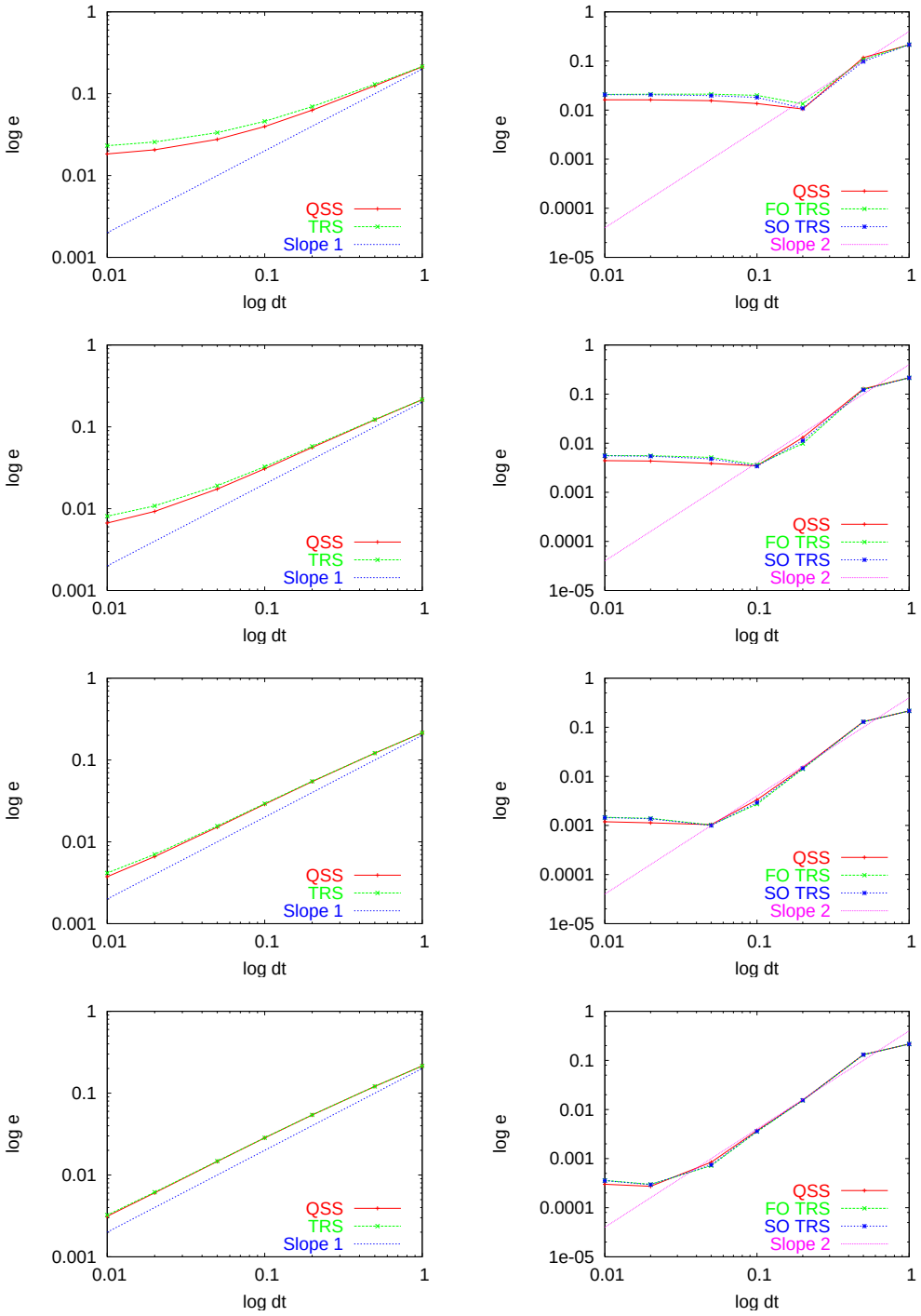

Figure 5.1: Convergence of the time approximation using quasi-static subscales (QSS) and transient subscales (TRS). First order approximation on the left and second order approximation on the right. In the second order approximation first order (FO) or second order (SO) subscales are considered. From top to bottom meshes of sizes $h = 1/20$, $h = 1/40$ and $h = 1/80$. Note that the convergence curves loose the optimal slope in time (1 or 2) when the error becomes dominated by the spatial component.

ASGS method in all the examples). It consists again of an exact solution problem in which the time dependent Navier-Stokes equations are solved in the unit square with Dirichlet boundary conditions taking the force $\boldsymbol{f}$ and boundary and initial values to have the exact (steady state) solution defined by

$$\boldsymbol{u} = (\sin(\pi x - 0.7)\sin(\pi y + 0.2)\,,\, \cos(\pi x - 0.7)\cos(\pi y + 0.2))\,,$$
$$p = \sin(\pi x)\cos(\pi y) + (\cos(1) - 1)\sin(1).$$

Numerical examples presented in [11] show that spurious oscillations in the pressure are found when the time step is small enough and that this effect is more dramatic when the order of the polynomial approximation is increased. We have solved this problem using different meshes for time step sizes $\delta t^n = 10^{-n}$ using a first order time approximation.

Figure 5.2 shows the convergence of the approximation using bilinear elements at the first time step, while figure 5.3 shows the same results corresponding to the second time step. The instability mentioned can be seen in figure 5.2, as for a given mesh size the error increases when the time step is decreased. As a first order approximation is being used and the solution of the problem is steady, the error should decrease linearly with the time step size. This is not the case in the first step, neither using the quasi-static subscales as shown in [11], nor using transient subscales. However, as shown in figure 5.3, when the transient subscales are considered the instability is eliminated at the second time step. This behavior leads to consider the practical problem of the initial conditions for the subgrid scale (we have taken them to be zero), which has not been considered here. It has to be noted that, in any case, the instability observed disappears as time advances and, obviously, the stationary solution is equally approximated using quasi-static and transient subscales.

The situation is different when higher order elements are used. Figure 5.4 shows the convergence of the approximation using biquadratic elements while figure 5.5 shows the convergence of the approximation using bicubic elements, both at the first time step. Similar results are found for the second time step. From figures 5.4 and 5.5 it is seen that when quasi-static subscales are considered the method could not converge as the mesh is refined for small time steps. This is even more dramatic than the result presented in [11], where only a fixed mesh of 10×10 elements was considered. In the case of transient subscales, although some dependence of the error on the time step size is still observed, convergence under mesh refinement is always achieved. This effect is seen in figure 5.6, where pressure contours for different mesh sizes obtained using quasi-static and transient subscales are compared.

5.4.3 Flow past a cylinder

The last example is the flow past a cylinder at $Re = 100$, a well known benchmark. The domain is $[0, 16] \times [0, 8] \setminus D$, where the cylinder D has a diameter 1 and is located at $(4, 4)$.

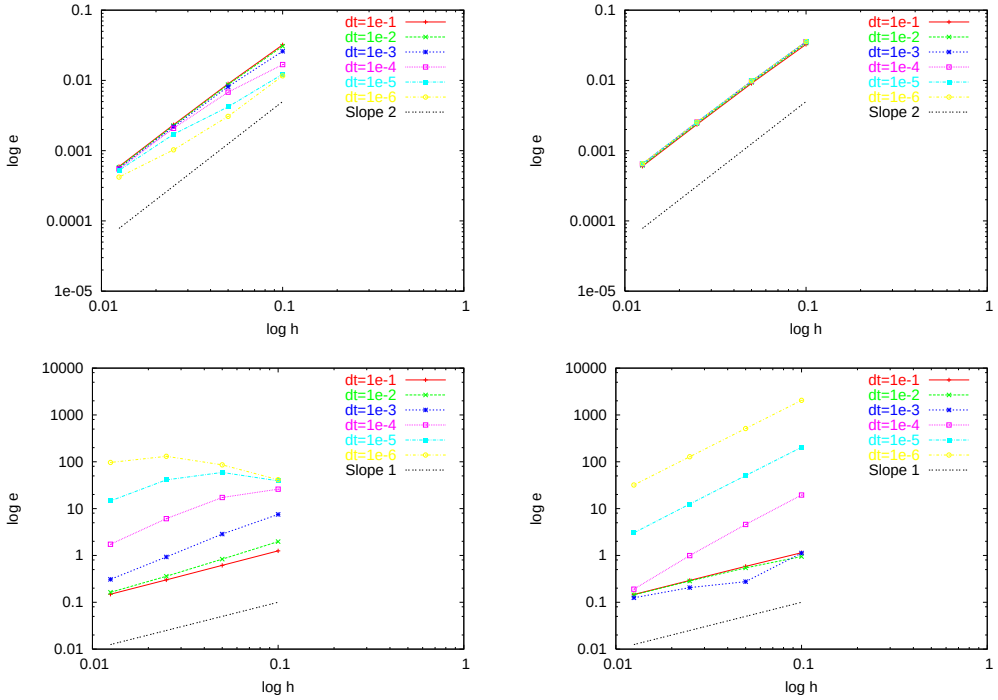

Figure 5.2: Convergence of the approximation using bilinear elements at the first time step. Quasi-static subscales on the left and transient subscales on the right. Velocity error at the top and pressure error at the bottom.

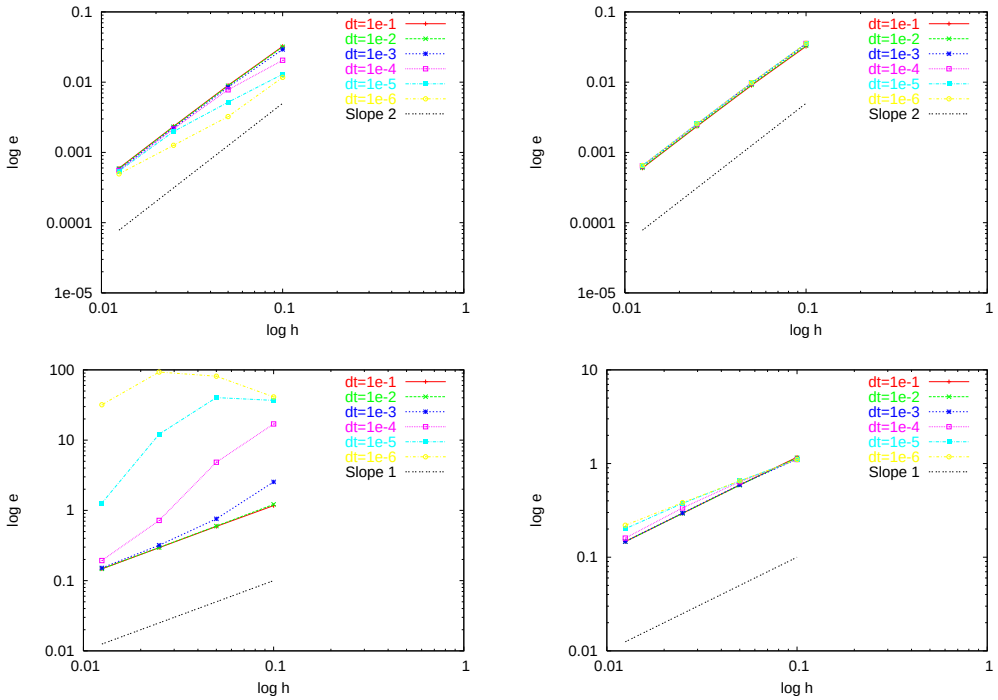

Figure 5.3: Convergence of the approximation using bilinear elements at the second time step. Quasi-static subscales on the left and transient subscales on the right. Velocity error at the top and pressure error at the bottom.

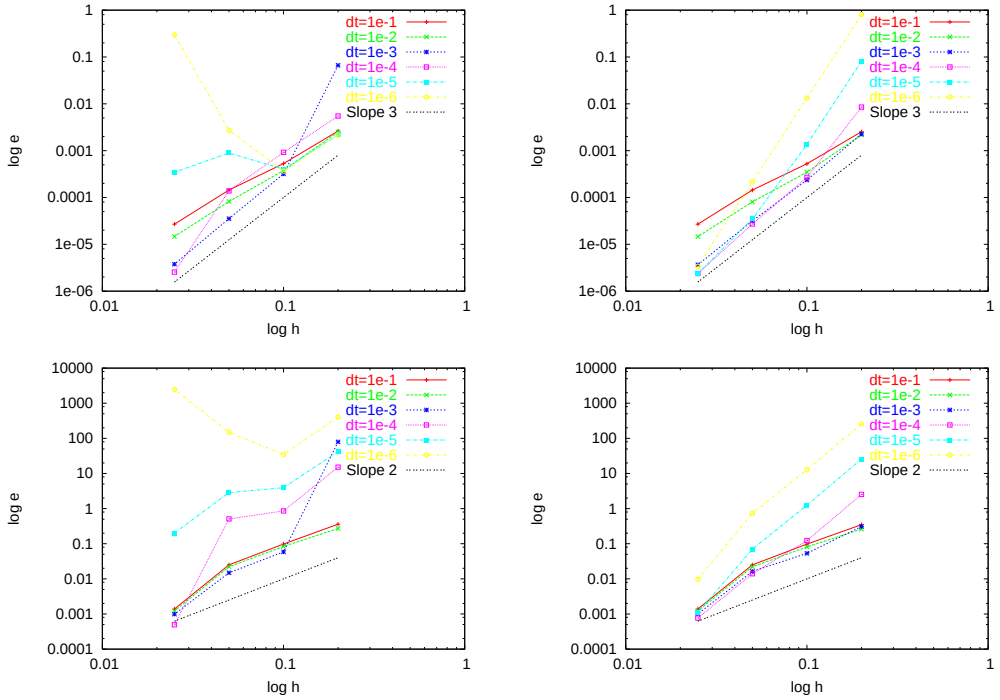

Figure 5.4: Convergence of the approximation using biquadratic elements at the first time step. Quasi-static subscales on the left and transient subscales on the right. Velocity error at the top and pressure error at the bottom.

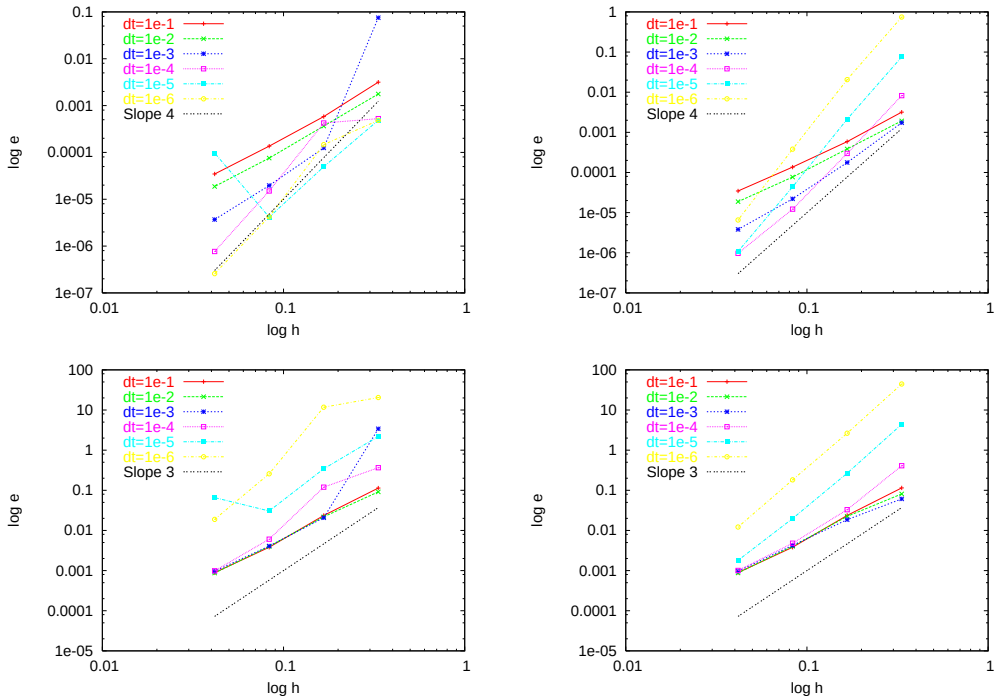

Figure 5.5: Convergence of the approximation using bicubic elements at the first time step. Quasi-static subscales on the left and transient subscales on the right. Velocity error at the top and pressure error at the bottom.

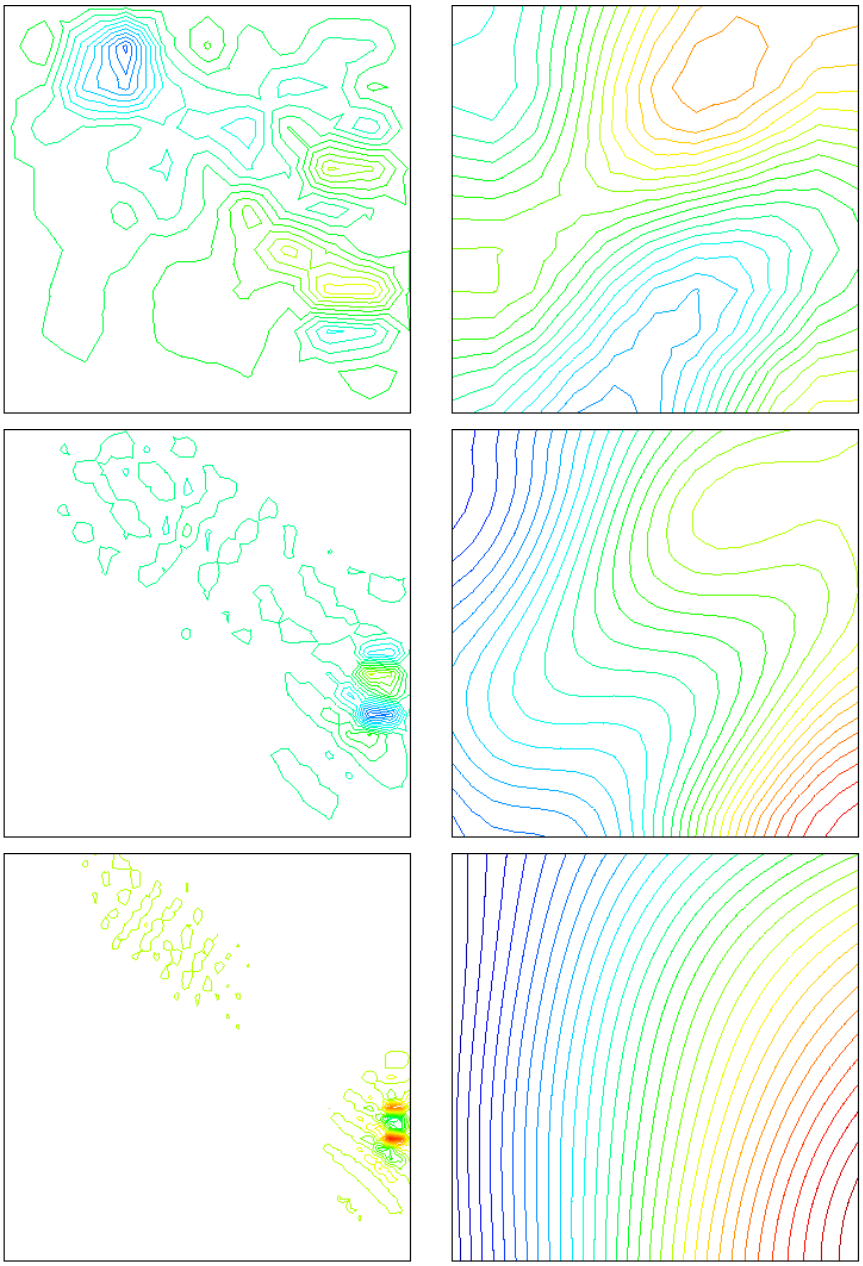

Figure 5.6: Pressure contours for $\delta t = 10^{-6}$ and (from top to bottom) $h = 1/20$, $h = 1/40$ and $h = 1/80$ using biquadratic elements. Quasi-static subscales on the left and transient subscales on the right.

A uniform velocity is prescribed at the inlet, zero y component is prescribed at $y = 0$ and $y = 8$ and zero traction is prescribed at the outlet. Two meshes have been used to test the behavior of the method, a coarse one of 1360 nodes and a fine one of 5280. The results will be compared to those obtained using a reference mesh of 20800 nodes.

The initial condition is $\boldsymbol{u} = (1, 0)$ except at the cylinder surface. From this initial condition the flow evolves to a symmetric solution that becomes unstable around $t = 100$ and the characteristic vortex shedding appears. To visualize the problem setting, a pressure distribution snapshot in the fully developed regime is shown in figure 5.7. A second order method has been used with time step size $\delta t = 0.2$ and 10 Euler time steps have been performed at the beginning of the calculations for all the meshes. A convergence tolerance of 10^{-8} was required at each step, which was achieved typically after 8 to 10 Picard iterations.

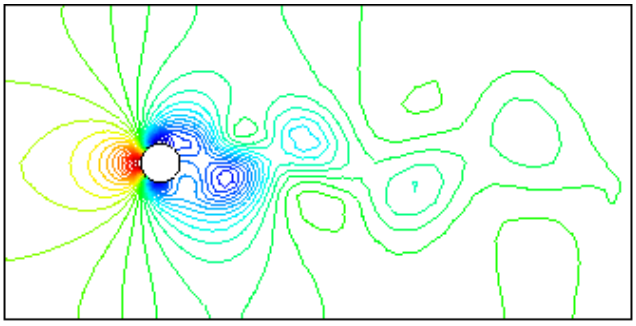

Figure 5.7: Pressure distribution at $t = 160$

Figures 5.8 and 5.9 show the evolution of the x-velocity at point $(6.15, 4)$, figures 5.10 and 5.11 that of the y-velocity and figures 5.12 and 5.13 that of the pressure, always at the same point and for the two meshes considered, comparing the results obtained using quasi-static subscales and transient subscales to those obtained using the reference mesh. It can be seen from figures 5.8 and 5.9 how the use of the transient subscales gives a better mean value of the x-velocity when the flow is fully developed, specially in the coarse mesh. From 5.10 and 5.11 it can be observed how the use of the transient subscales gives a higher amplitude and a higher frequency of the oscillation, that is to say, less numerical dissipation. Finally, in figure 5.12 some time step-to-time step oscillations can be observed when the quasi-static subscales are used and how these oscillations do not appear when transient subscales are considered. These oscillations, already reported in [29], depend on the length used in the definition of the stabilization parameters. They appear when there is a variation of the element size from one element to another and they disappear if a fixed mesh size is used to define the stabilization parameter. From figure 5.13 it is seen that they also disappear in the fine mesh. In this case there is almost no gain in the pressure using transient subscales (but there is in the velocity, as shown in figures 5.9 and 5.11).

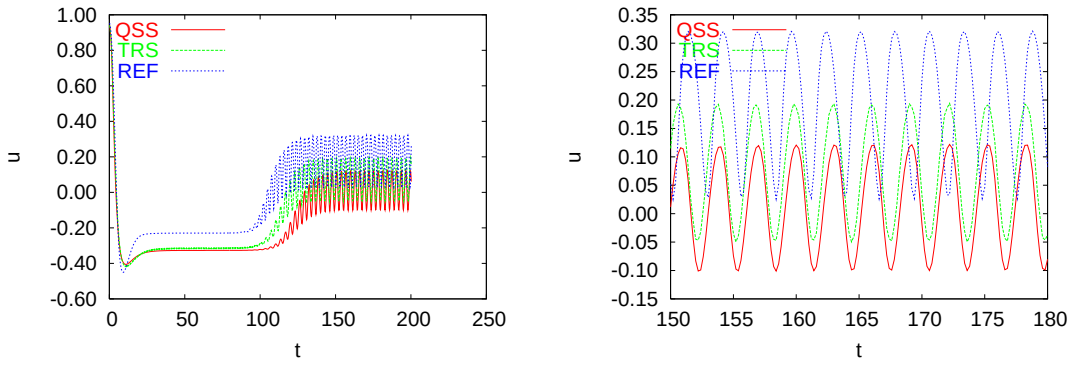

Figure 5.8: Horizontal velocity evolution at $(6.15, 4.0)$ using the coarse mesh (left) and its detail (right).

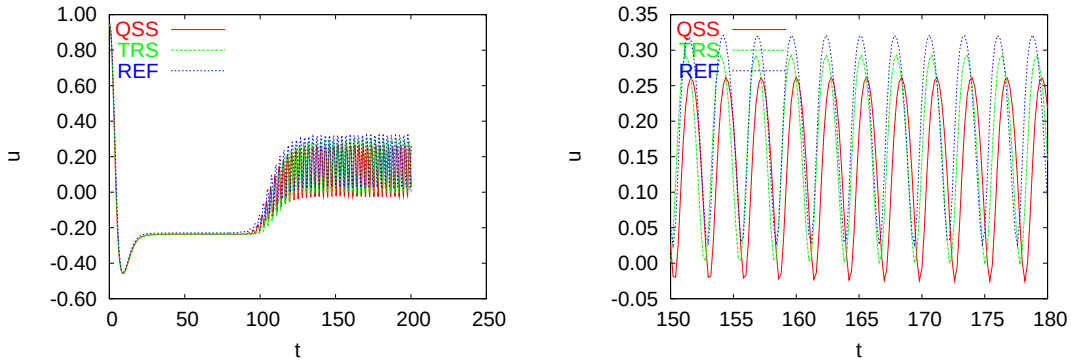

Figure 5.9: Horizontal velocity evolution at $(6.15, 4.0)$ using the fine mesh (top) and its detail (bottom).

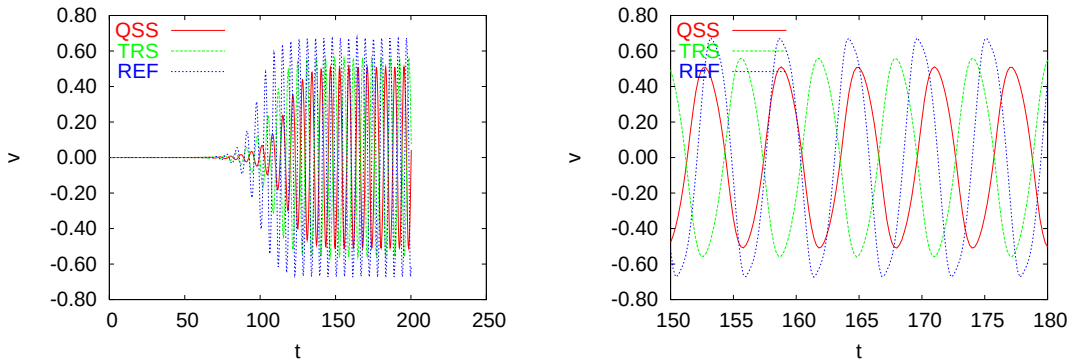

Figure 5.10: Vertical velocity evolution at $(6.15, 4.0)$ using the coarse mesh (top) and its detail (bottom).

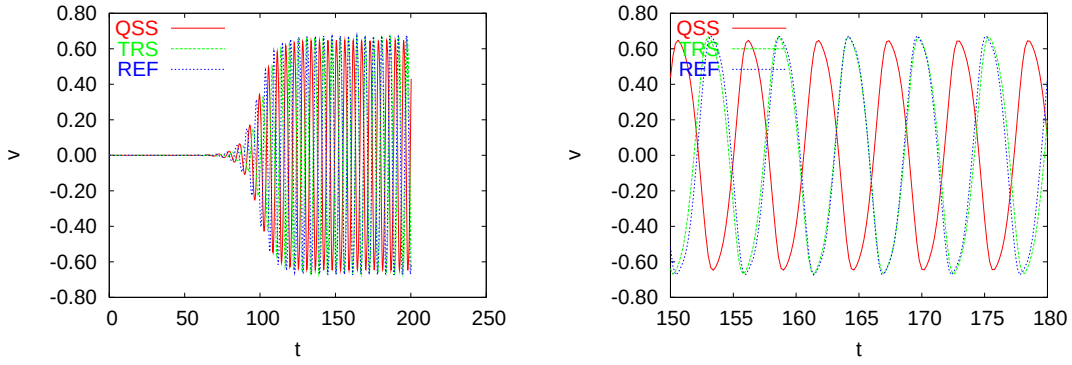

Figure 5.11: Vertical velocity evolution at $(6.15, 4.0)$ using the fine mesh (top) and its detail (bottom).

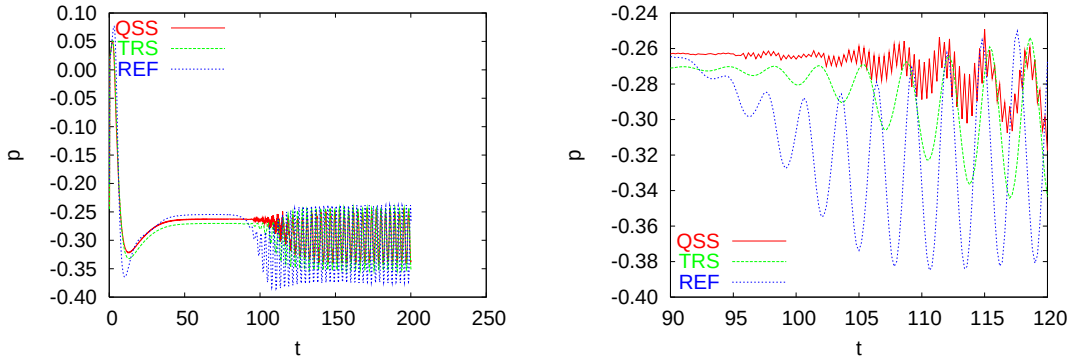

Figure 5.12: Pressure evolution at $(6.15, 4.0)$ using the coarse mesh (top) and its detail (bottom).

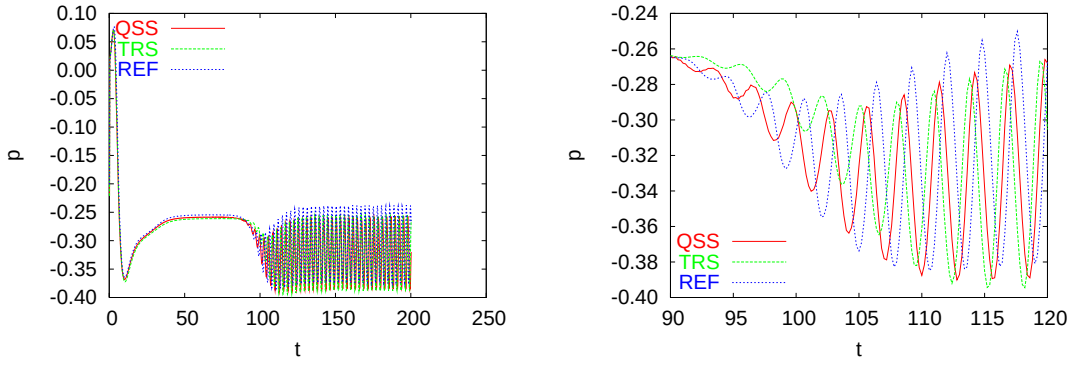

Figure 5.13: Pressure evolution at $(6.15, 4.0)$ using the fine mesh (top) and its detail (bottom).

5.5 Conclusions

The main conclusion of this chapter is simple: we believe it is worth to track the subscales in time in a variational multiscale approach to the transient incompressible Navier-Stokes equations and to take into account all their contributions in the convective term.

The first and very simple reason is that it leads to global momentum conservation, a rare property. A second reason can be the door opened to turbulence modeling, although we have touched this point only marginally. What has been the main focus of this chapter is the study of the advantages of tracking the subscales from the point of view of the time integration scheme. First, we have remarked that the resulting formulation leads in a natural way to the correct behavior of the stabilization parameters with the time step while steady-state solutions do not depend on it. Moreover, the conflict about the design of the stabilization terms for time dependent problems (either at the semi-discrete or the fully discrete level) disappears, since space and time discretization can be commuted. The numerical experiments show that the gain with respect to quasi-static subscales is notorious.

Chapter 6

Thermally coupled flow problems

In this chapter we propose a variational multiscale finite element approximation of thermally coupled flows. We consider the thermal coupling in the context of the Boussinesq approximation but the same formulation is used to solve the low Mach number equations with minor modifications (we refer the reader to the next chapter, in which some implementation details are discussed). The main feature of the formulation in contrast to other stabilized methods is that we consider the subscales as time dependent. They are solution of a differential equation in time that needs to be integrated. Likewise, we keep the effect of the subscales both in the nonlinear convective terms of the momentum and temperature equations and, if required, the coupling between them.

6.1 Introduction

Thermally coupled incompressible flows are of particular interest from the numerical point of view for different reasons. Apart from their obvious practical interest, very often these flows exhibit instabilities and even transition to turbulence in situations simpler than for isothermal flows. The numerical modeling of these instabilities that take place in rather simple cases is an excellent test for numerical formulations.

In this chapter we propose a finite element formulation for thermally coupled flows based on the variational multiscale formalism [78]. The basic idea is to split the unknowns, velocity, pressure and temperature, into their finite element component and a subgrid scale component, hereafter referred to as subscale. The particular approximation used for these subscales defines the numerical model. The main feature of the model we propose is that we consider the subscales time dependent and that we keep their effect in all the terms of the equations to be solved, both the nonlinear convective terms of the momentum and the heat equation and in the coupling term due to the Boussinesq model.

The basic formulation for isothermal incompressible flows was described in [36] and chapter 5. As it is explained there, considering the subscales time dependent and tracking

them along the iterative process to deal with the nonlinear terms has several benefits, such as a better performance in time of the final formulation, the conservation of momentum or the possibility to model turbulence. In this chapter we extend the formulation to thermally coupled flows using the Boussinesq approximation.

The need to stabilize the standard Galerkin finite element approximation comes from two main sources, namely, the wish to use equal velocity-pressure interpolations and to deal with convection dominated flows. As it is now well known, both sources of instability can be overcome by using stabilized formulations. However, the main interest of this chapter is not to explain how the stabilized formulation employed here allows to use equal interpolations or is able to avoid convection instabilities. Our main concern is to explain how to consider *dynamic* subscales, how to integrate them in time and how to track them along the iterative process, accounting in particular for the coupling of the heat and the momentum equations.

The chapter is organized as follows. In the following section we define the problem and in section 6.3 we consider the multiscale formulation extended to thermally coupled flows and we present the time integration scheme in section 6.4, summarizing its main properties in section 6.5. Two numerical examples are presented in Section 6.6, both of them two-dimensional. They involve two situations of thermally coupled flows that display a bifurcation of the solution due to the instability of the basic flow. One of them is the classical Rayleigh-Bénard instability coupled with a Poiseuille flow, which leads to a transient flow even if the bifurcation is of stationary type. The second example is the classical flow in a cavity with differentially heated vertical walls. When the Prandtl number is small, the flow exhibits a Hopf bifurcation that leads to an oscillating flow pattern. The chapter concludes in Section 6.7 with some final remarks and comments.

6.2 Physical problem

Let $\Omega \subset \mathbb{R}^{\mathrm{d}}$, with $\mathrm{d} = 2, 3$, be the computational domain in which the flow takes place during the time interval $[0, T]$, and let Γ be its boundary. The initial and boundary value problem to be considered consists in finding a velocity field $\boldsymbol{u}$, a pressure p and a

temperature ϑ such that

$$\begin{aligned}
\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} - \nu\Delta\boldsymbol{u} + \nabla p + \beta\boldsymbol{g}\vartheta &= \boldsymbol{f} + \beta\boldsymbol{g}\vartheta_0 && \text{in } \Omega,\ t\in(0,T)\\
\nabla\cdot\boldsymbol{u} &= 0 && \text{in } \Omega,\ t\in(0,T)\\
\partial_t\vartheta + \boldsymbol{u}\cdot\nabla\vartheta - \alpha\Delta\vartheta &= Q && \text{in } \Omega,\ t\in(0,T)\\
\boldsymbol{u} &= \boldsymbol{0} && \text{on } \Gamma,\ t\in(0,T)\\
\boldsymbol{u} &= \boldsymbol{u}^0 && \text{in } \Omega,\ t=0\\
\vartheta &= 0 && \text{on } \Gamma,\ t\in(0,T)\\
\vartheta &= \vartheta^0 && \text{in } \Omega,\ t=0
\end{aligned}$$

In these equations, ν is the kinematic viscosity, α the thermal diffusivity, β the thermal expansion coefficient, $\boldsymbol{f}$ the external body forces, ϑ_0 the reference temperature, $\boldsymbol{g}$ the gravity acceleration vector, Q the heat source and $\boldsymbol{u}^0$ and ϑ^0 the initial conditions for velocity and temperature, respectively. For simplicity in the exposition, we have assumed homogeneous Dirichlet boundary conditions for both velocity and temperature.

To define the functional setting, let $H^1(\Omega)$ be the space of functions such that they and their first derivatives belong to $L^2(\Omega)$ (that is, they are square integrable), and let $H^1_0(\Omega)$ be the subspace of functions in $H^1(\Omega)$ vanishing on the boundary. Let also $\boldsymbol{V}^{\mathrm{st}} = H^1_0(\Omega)^d$, $Q^{\mathrm{st}} = L^2(\Omega)/\mathbb{R}$, $\Psi^{\mathrm{st}} = H^1_0(\Omega)$ and define $\boldsymbol{V} = \boldsymbol{L}^2(0,T;\boldsymbol{V}^{\mathrm{st}})$, $Q = L^1(0,T;Q^{\mathrm{st}})$ (for example) and $\Psi = L^2(0,T;\Psi^{\mathrm{st}})$, where $L^p(0,T;X)$ stands of the space of functions such that their X norm in the spatial argument is an $L^p(0,T)$ function in time, that is, its p-th power is integrable if $1\le p<\infty$ or bounded if $p=\infty$.

The weak form of the problem consists in finding $(\boldsymbol{u},p,\vartheta)\in\boldsymbol{V}\times Q\times\Psi$ such that

$$(\partial_t\boldsymbol{u},\boldsymbol{v}) + (\boldsymbol{u}\cdot\nabla\boldsymbol{u},\boldsymbol{v}) + \nu(\nabla\boldsymbol{u},\nabla\boldsymbol{v}) - (p,\nabla\cdot\boldsymbol{v}) + \beta(\boldsymbol{g}\vartheta,\boldsymbol{v}) = \langle\boldsymbol{f},\boldsymbol{v}\rangle + \beta(\boldsymbol{g}\vartheta_0,\boldsymbol{v}) \quad (6.1)$$

$$(q,\nabla\cdot\boldsymbol{u}) = 0 \quad (6.2)$$

$$(\partial_t\vartheta,\psi) + (\boldsymbol{u}\cdot\nabla\vartheta,\psi) + \alpha(\nabla\vartheta,\nabla\psi) = (Q,\psi) \quad (6.3)$$

for all $(\boldsymbol{v},q,\psi)\in\boldsymbol{V}^{\mathrm{st}}\times Q^{\mathrm{st}}\times\Psi^{\mathrm{st}}$, where $(\cdot,\cdot)$ denotes the $L^2(\Omega)$ inner product and $(f,g) := \int_\Omega fg\,\mathrm{d}\Omega$ whenever functions f and g are such that the integral is well defined. The dimensionless numbers relevant in this problem are those already defined in chapter 2 and the Grashof number given by

$$\mathrm{Gr} = \frac{\beta|\boldsymbol{g}|l_0^3\Delta\vartheta}{\nu^2}$$

where l_0 is a characteristic length of the problem and $\Delta\vartheta$ a characteristic temperature difference. Note the relation $\mathrm{Ra} = \mathrm{Gr}\,\mathrm{Pr}$.

6.3 Multiscale approximation

Let us consider a finite element partition $\mathcal{P}_h = \{K\}$ of the computational domain Ω of n_{el} elements, from which we can construct finite element spaces for the velocity, pressure

and temperature in the usual manner. We will denote them by $\boldsymbol{V}_h \subset \boldsymbol{V}^{\mathrm{st}}$, $Q_h \subset Q^{\mathrm{st}}$ and $\Psi_h \subset \Psi^{\mathrm{st}}$, respectively, and, to simplify the exposition, we will assume that they are all built from continuous piecewise polynomials of the same degree k. The basic idea of the multiscale approach we will follow [78] is to split the continuous unknowns as

$$\boldsymbol{u} = \boldsymbol{u}_h + \widetilde{\boldsymbol{u}} \tag{6.4}$$

$$p = p_h + \widetilde{p} \tag{6.5}$$

$$\vartheta = \vartheta_h + \widetilde{\vartheta} \tag{6.6}$$

where the components with subscript h belong to the corresponding finite element spaces. The components with a tilde belong to any space such that its direct sum with the finite element space yields the functional space where the unknown is sought. For the moment, we leave it undefined. These additional components are what we will call *subscales*. Each particular variational multiscale method will depend on the way the subscales are approximated. However, our main focus in this work is *not* how to choose the space of subscales (in our case for velocity, pressure and temperature), but to explain the consequences *of considering these subscales time dependent*, and therefore requiring to be integrated in time. Likewise, we will keep the previous decomposition 6.4-6.6 in *all the terms of 6.1-6.3*. The only approximation we will make for the moment is to assume that the subscales vanish on the interelement boundaries, ∂K. This happens for example if they are approximated using bubble functions [4], or if one assumes that their Fourier modes correspond to high wave numbers, as it is explained in [29].

Substituting 6.4-6.6 into 6.1-6.3, taking the test functions in the corresponding finite element spaces and integrating some terms by parts, it is found that

$$\begin{aligned}(\partial_t \boldsymbol{u}_h, \boldsymbol{v}_h) + (\boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h) + \nu(\nabla \boldsymbol{u}_h, \nabla \boldsymbol{v}_h) - (p_h, \nabla \cdot \boldsymbol{v}_h) + \beta(\boldsymbol{g}\vartheta_h, \boldsymbol{v}_h) \\ - (\widetilde{\boldsymbol{u}}, \nu \Delta_h \boldsymbol{v}_h + \boldsymbol{u}_h \nabla \cdot \boldsymbol{v}_h) + (\partial_t \widetilde{\boldsymbol{u}}, \boldsymbol{v}_h) + (\widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h) - (\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{v}_h) \\ - (\widetilde{p}, \nabla \cdot \boldsymbol{v}_h) + \beta(\boldsymbol{g}\widetilde{\vartheta}, \boldsymbol{v}_h) = \langle \boldsymbol{f}, \boldsymbol{v}_h \rangle + \beta(\boldsymbol{g}\vartheta_0, \boldsymbol{v}_h)\end{aligned} \tag{6.7}$$

$$(q_h, \nabla \cdot \boldsymbol{u}_h) - (\widetilde{\boldsymbol{u}}, \nabla q_h) = 0 \tag{6.8}$$

$$\begin{aligned}(\partial_t \vartheta_h, \psi_h) + (\boldsymbol{u}_h \cdot \nabla \vartheta_h, \psi_h) + \alpha(\nabla \vartheta_h, \nabla \psi_h) - \left(\widetilde{\vartheta}, \alpha \Delta_h \psi_h + \boldsymbol{u}_h \cdot \nabla \psi_h\right) \\ + (\partial_t \widetilde{\vartheta}, \psi_h) + (\widetilde{\boldsymbol{u}} \cdot \nabla \vartheta_h, \psi_h) - \left(\widetilde{\vartheta}, \widetilde{\boldsymbol{u}} \cdot \nabla \psi_h\right) = (Q, \psi_h)\end{aligned} \tag{6.9}$$

which must hold for all test functions $(\boldsymbol{v}_h, q_h, \psi_h) \in \boldsymbol{V}_h \times Q_h \times \Psi_h$. The subindex h in the Laplacian denotes that it is evaluated elementwise.

The first row in 6.7 corresponds to the terms arising from the classical Galerkin approximation of the momentum equation (except for the term due to external forces). Once the velocity subscale is approximated, the first term of the second row provides stability of convection as usual in classical methods (see, for example, [28]). The rest of the terms in the second row and those in the third row are non-standard terms, in the sense

that they are usually neglected. One of our purposes here is to discuss the implications of these terms. The last row in 6.7 comes from the contribution of the pressure and temperature subscale and the contribution from the external forces. It is rather standard to take the pressure subscale into account, but to study the effect of the temperature subscale is one of the objectives of one of our numerical experiments.

In the left-hand-side of 6.8 the first term is the classical Galerkin contribution, whereas the second provides (pressure) stability once the velocity subscale is approximated.

Similar comments to those made for 6.7 apply to 6.9. The first three terms of the first row correspond to the classical Galerkin approximation (except for the heat source), the last term of the first row provide stability in convection dominated flows when the temperature subscale is approximated and, finally, the three terms in the left-hand-side of the second row are non-standard, and come from the fact that subscales are never neglected in the previous equations (except for the fact that they are assumed to vanish on the interelement boundaries, as it has been already mentioned).

Equations 6.7-6.9 can be understood as the projection of the original equations onto the finite element spaces of velocity, pressure and temperature. The equations for the subscales are obtained by projecting onto their corresponding spaces. If $\widetilde{P}$ denotes this projection onto any of these spaces, these equations are

$$\widetilde{P}\left[\partial_t \tilde{\boldsymbol{u}} + (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \widetilde{\boldsymbol{u}} - \nu \Delta \widetilde{\boldsymbol{u}} + \nabla \widetilde{p} + \beta \boldsymbol{g} \widetilde{\vartheta}\right] = \widetilde{P} \boldsymbol{R}_m \tag{6.10}$$

$$\widetilde{P}\left[\nabla \cdot \widetilde{\boldsymbol{u}}\right] = \widetilde{P} R_c \tag{6.11}$$

$$\widetilde{P}\left[\partial_t \widetilde{\vartheta} + (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \widetilde{\vartheta} - \alpha \Delta \widetilde{\vartheta}\right] = \widetilde{P} R_e \tag{6.12}$$

where

$$\begin{aligned}
\boldsymbol{R}_m &= \boldsymbol{f} + \beta \boldsymbol{g} \vartheta_0 - [\partial_t \boldsymbol{u}_h + (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \boldsymbol{u}_h - \nu \Delta_h \boldsymbol{u}_h + \nabla p_h + \beta \boldsymbol{g} \vartheta_h] \\
R_c &= -\nabla \cdot \boldsymbol{u}_h \\
R_e &= Q - [\partial_t \vartheta_h + (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla \vartheta_h - \alpha \Delta_h \vartheta_h],
\end{aligned}$$

are the residuals of the finite element unknowns in the momentum, continuity and energy equation, respectively. Equations 6.10-6.12 need to be solved within each element and, as we have assumed, considering homogeneous velocity and temperature Dirichlet boundary conditions.

It is not our purpose here to discuss how to approximate 6.10-6.11 which, in fact, is the essence of the different stabilized finite element methods that can be found in the literature. We will adopt a simple approximation that can be found, for example, in [29] and references therein. Our main concern, as in the reference just mentioned, is *to keep the time dependence of the subscales, as well their nonlinear effects.* When their time derivative is neglected, we will call them *quasi-static*, whereas otherwise we will call them *dynamic*.

Following the line of chapters 3, 4 and 5 now extended to thermally coupled flows, we propose to compute the subscales *within each element* of the finite element partition as solution to

$$\partial_t \widetilde{\boldsymbol{u}} + \tau_m^{-1} \widetilde{\boldsymbol{u}} = \widetilde{P} \boldsymbol{R}_m \tag{6.13}$$

$$\tau_c^{-1} \widetilde{p} = \widetilde{P} \left(R_c + \tau_m \partial_t R_p \right) \tag{6.14}$$

$$\partial_t \widetilde{\vartheta} + \tau_e^{-1} \widetilde{\vartheta} = \widetilde{P} R_e \tag{6.15}$$

where the isotropic *stabilization* parameters τ_m, τ_c and τ_e are computed as

$$\tau_m = \left(c_1 \frac{\nu}{h^2} + c_2 \frac{|\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}|}{h} \right)^{-1} \tag{6.16}$$

$$\tau_c = \frac{h^2}{c_1 \tau_m} = \nu + \frac{c_2}{c_1} h |\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}| \tag{6.17}$$

$$\tau_e = \left(c_1 \frac{\alpha}{h^2} + c_2 \frac{|\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}|}{h} \right)^{-1} \tag{6.18}$$

where h is the element size and c_1 and c_2 are algorithmic constants (we have adopted $c_1 = 4$ and $c_2 = 2$ in the numerical experiments).

The approximation adopted for the subscales could certainly be improved, for example by trying to relax the assumption that they vanish on the interelement boundaries or by trying to model the coupling between the three equations in play (momentum, continuity and heat) as done in chapter 4 for the Oseen equations. However, our interest here is only to analyze the effect of considering the subscales time dependent and taking into account their contribution in the nonlinear terms. In particular, it is important to remark that 6.13 is nonlinear, both because the velocity subscale contributes to the advection velocity and because the stabilization parameter τ_m depends also on the velocity subscale, as equation 6.15 and the stabilization parameter τ_e. Likewise, 6.13 depends on the temperature subscale, and therefore the velocity-temperature coupling is naturally accounted for.

Even though it is not our purpose to use an "accurate" approximation to the subscales like the one introduced in chapter 4, in some case we have found convenient to include the time derivative of R_c in the approximation 6.14 of the pressure subscale. This term was neglected in the previous chapter and in [29, 36], but in some situations it is crucial to improve pressure stability. This time derivative arises naturally if the second approximation to the subscales of chapter 5 is considered.

It is observed that in 6.13-6.15 we have kept the projections $\widetilde{P}$ in the right-hand-side terms. Basically, two different options can be considered. Classical stabilized finite element methods are recovered by taking $\widetilde{P} = I$ (the identity), whereas if $\widetilde{P} = P_h^{\perp} = I - P_h$, P_h being the L^2-projection onto the appropriate finite element space, the subscales turn out to be orthogonal to this finite element space. The resulting formulation is termed as orthogonal subscales stabilization (OSS) in [29].

The space-discrete formulation is now complete. However, contrary to what happens with quasi-static subscales and neglecting their nonlinear effects, now it is not possible to obtain a closed-form expression for these subscales and insert them into 6.7-6.9 in order to obtain a problem for the finite element components of velocity, pressure and temperature. Prior to discretizing in time, we cannot go any further than saying that the problem consists in solving 6.7-6.9 together with 6.13-6.15.

6.4 Temporal discretization

Any finite difference scheme can now be applied to discretize in time both equations 6.7-6.9 *and* equations 6.13-6.15. Obviously, space-time finite element discretizations are also possible. In order to make the exposition concise, we will restrict our attention to the trapezoidal rule.

Let δt be the time step size of a uniform partition of the time interval $[0,T]$, $0=t^0<t^1<...<t^N=T$. Functions approximated at time t^n will be identified with the superscript n. For a generic function f, we will use the notation $\delta f^n := f^{n+1}-f^n$, $\delta_t f^n = \delta f^n/\delta t$, $f^{n+\theta}=\theta f^{n+1}+(1-\theta)f^n$, $0\le\theta\le 1$.

The time discretization of 6.7-6.9 is standard and does not need any further explanation. Given $\boldsymbol{u}_h^n$, ϑ_h^n, $\widetilde{\boldsymbol{u}}^n$ and $\widetilde{\vartheta}^n$, it consists of solving the problem

$$
\begin{aligned}
&(\delta_t \boldsymbol{u}_h^n,\boldsymbol{v}_h)+\left(\boldsymbol{u}_h^{n+\theta}\cdot\nabla\boldsymbol{u}_h^{n+\theta},\boldsymbol{v}_h\right)+\nu(\nabla\boldsymbol{u}_h^{n+\theta},\nabla\boldsymbol{v}_h)-(p_h^{n+1},\nabla\cdot\boldsymbol{v}_h)+\beta(\boldsymbol{g}\vartheta_h^{n+\theta},\boldsymbol{v}_h)\\
&-\left(\widetilde{\boldsymbol{u}}^{n+\theta},\nu\Delta_h\boldsymbol{v}_h+\boldsymbol{u}_h^{n+\theta}\nabla\cdot\boldsymbol{v}_h\right)+(\delta_t\widetilde{\boldsymbol{u}}^n,\boldsymbol{v}_h)+\left(\widetilde{\boldsymbol{u}}^{n+\theta}\cdot\nabla\boldsymbol{u}_h^{n+\theta},\boldsymbol{v}_h\right)-\left(\widetilde{\boldsymbol{u}}^{n+\theta},\widetilde{\boldsymbol{u}}^{n+\theta}\cdot\nabla\boldsymbol{v}_h\right)\\
&-(\widetilde{p}^{n+1},\nabla\cdot\boldsymbol{v}_h)+\beta(\boldsymbol{g}\widetilde{\vartheta}^{n+\theta},\boldsymbol{v}_h)=\langle\boldsymbol{f},\boldsymbol{v}_h\rangle+\beta(\boldsymbol{g}\vartheta_0,\boldsymbol{v}_h) \qquad (6.19)\\
&(q_h,\nabla\cdot\boldsymbol{u}_h^{n+\theta})-(\widetilde{\boldsymbol{u}}^{n+\theta},\nabla q_h)=0 \qquad (6.20)\\
&(\delta_t\vartheta_h^n,\psi_h)+\left(\boldsymbol{u}_h^{n+\theta}\cdot\nabla\vartheta_h^{n+\theta},\psi_h\right)+\alpha(\nabla\vartheta_h^{n+\theta},\nabla\psi_h)-\left(\widetilde{\vartheta}^{n+\theta},\alpha\Delta_h\psi_h+\boldsymbol{u}_h^{n+\theta}\cdot\nabla\psi_h\right)\\
&+(\delta_t\widetilde{\vartheta}^n,\psi_h)+\left(\widetilde{\boldsymbol{u}}^{n+\theta}\cdot\nabla\vartheta_h^{n+\theta},\psi_h\right)-\left(\widetilde{\vartheta}^{n+\theta},\widetilde{\boldsymbol{u}}^{n+\theta}\cdot\nabla\psi_h\right)=(Q,\psi_h)\,, \qquad (6.21)
\end{aligned}
$$

which must hold for all test functions $(\boldsymbol{v}_h,q_h,\psi_h)\in\boldsymbol{V}_h\times Q_h\times\Psi_h$. Note that the pressure is considered approximated at time $n+1$. This avoids the need to deal with the pressure at a previous time step and does not modify the velocity approximation. As it is well known, the scheme is expected to be of second order if $\theta=1/2$ and of first order otherwise.

Equations 6.13-6.15 need also to be integrated in time. The simplest option is to use the same time discretization as for the finite element equations, which yields

$$
\begin{aligned}
\delta_t\widetilde{\boldsymbol{u}}^n+\frac{1}{\tau_m^{n+\theta}}\widetilde{\boldsymbol{u}}^{n+\theta}&=\widetilde{P}\boldsymbol{R}_m^{n+\theta} \qquad (6.22)\\
\frac{1}{\tau_c^{n+1}}\widetilde{p}^{n+1}&=\widetilde{P}\left(R_c^{n+1}+\tau_m^{n+1}\delta_t R_c^n\right) \qquad (6.23)\\
\delta_t\widetilde{\vartheta}^n+\frac{1}{\tau_e^{n+\theta}}\widetilde{\vartheta}^{n+\theta}&=\widetilde{P}R_e^{n+\theta} \qquad (6.24)
\end{aligned}
$$

However, we will consider two additional options. The first is that the time integration for the subscales could be less accurate than for the finite element equations 6.7-6.9 and still keep the same order of accuracy in time of the finite element solution. The formal idea to justify this is the following. From the expression of the stabilization parameters τ_m and τ_e in 6.16 and 6.18, respectively, it follows that they behave as the critical time steps of an explicit integration in time of the momentum and the heat equation [37]. Therefore, we may assume that they are of order $\mathcal{O}(\delta t)$. From 6.22 it follows that $\mathcal{O}(1)\delta\widetilde{\boldsymbol{u}}^{n+1} + \widetilde{\boldsymbol{u}}^{n+1} = \mathcal{O}(\delta t)\widetilde{P}(\boldsymbol{R}_m^{n+\theta})$, and thus we may conclude that $\widetilde{\boldsymbol{u}}^{n+1} = \mathcal{O}(\delta t)\widetilde{P}(\boldsymbol{R}_m^{n+1})$. If the residual of the finite element component is bounded, $|\widetilde{\boldsymbol{u}}^{n+1} - \widetilde{\boldsymbol{u}}^{n}| = \mathcal{O}(\delta t^2)$, and therefore evaluating the subscale at $n+1$, for example, in 6.19 instead of at $n+\theta$ introduces an error of order $\mathcal{O}(\delta t^2)$, which is the optimal error that can be reached with the trapezoidal rule (for $\theta = 1/2$). The same comments apply to 6.24 for the temperature subscale.

Considering the subscale equations integrated to first order and the finite element equations to second (or higher) is not particularly relevant in the case of the trapezoidal rule. However, if, for example, the second order backward-differencing (BDF) scheme is used, a first order integration of the equation for the subscales avoids the need to store them in two previous time steps. This storage is the most important cost of integrating the subscales in time. Another aspect to take into account is that the subscale approximation is not smooth, since the residual of the finite element components will be discontinuous across interelement boundaries. Thus, it seems reasonable to use a scheme as dissipative as possible to integrate the subscales in time. Further comments about this point are made in Section 4.

A first order time integration for the subscales is straightforward. Equations 6.22 and 6.24 have to be replaced by their counterparts for $\theta = 1$.

A third and final possibility that can be considered to integrate 6.13-6.15 in time is a combination of exact integration and approximation of the stabilization parameters and residuals at $t^{n+\theta}$. If this approximation is done, the equations for the velocity and temperature subscales are

$$\partial_t \widetilde{\boldsymbol{u}} + \frac{1}{\tau_m^{n+\theta}}\widetilde{\boldsymbol{u}} = \widetilde{P}\boldsymbol{R}_m^{n+\theta}$$
$$\partial_t \widetilde{\vartheta} + \frac{1}{\tau_e^{n+\theta}}\widetilde{\vartheta} = \widetilde{P}R_e^{n+\theta}$$

which can be integrated exactly, yielding

$$\widetilde{\boldsymbol{u}}^{n+1} = \left(\widetilde{\boldsymbol{u}}^{n} - \tau_m^{n+\theta}\widetilde{P}\boldsymbol{R}_m^{n+\theta}\right)\exp\left(-\frac{\delta t}{\tau_m^{n+\theta}}\right) + \tau_1^{n+\theta}\widetilde{P}\boldsymbol{R}_m^{n+\theta} \tag{6.25}$$

$$\widetilde{\vartheta}^{n+1} = \left(\widetilde{\vartheta}^{n} - \tau_e^{n+\theta}\widetilde{P}R_e^{n+\theta}\right)\exp\left(-\frac{\delta t}{\tau_e^{n+\theta}}\right) + \tau_3^{n+\theta}\widetilde{P}R_e^{n+\theta} \tag{6.26}$$

6.5 Main features of the formulation

The method described so far is an extension of the formulation proposed in chapter 5 to the case of thermally coupled flows using the Boussinesq approximation and therefore it is not necessary to repeat the same arguments again. Therefore, referring the redear to chapter 5 for their justification, let us briefly recall the fundamental features of the formulation and remark the differences that appear in the thermal case.

- The first point is the effect of considering the subscales dynamic, and therefore to deal with their time variation. Doing that the effect of time integration is now clear. Suppose for example that we are using 6.22-6.24 to integrate the subscales. Certainly, the effective stabilization parameters have to be modified (as it is done for example in [132, 138]), but when the steady-state is reached the subscale $\widetilde{\boldsymbol{u}}$ that is obtained satisfies

$$\left(\frac{1}{\beta\delta t}+\frac{1}{\tau_m}\right)\widetilde{\boldsymbol{u}}=\frac{1}{\beta\delta t}\widetilde{\boldsymbol{u}}+\boldsymbol{R}_m,$$

  from where

$$\widetilde{\boldsymbol{u}}=\tau_m\boldsymbol{R}_m,$$

  so that the usual expression employed for stationary problems is recovered. Numerical experiments also show that the temporal time integration is significantly improved eliminating oscillations originated by initial transients and minimizing numerical dissipation. The use of dynamic subscales also leads to the commutation of space discretization (understood as scale splitting) and time discretization. That is *time discretization + stabilization (scale splitting) = stabilization (scale splitting) + time discretization*. In what respects the time integration properties, the situation is similar for the energy equation.

- The second point is the effect of tracking the subscales along the nonlinear process and in the case of thermal problems along the coupling. On the one hand the tracking results in conservation properties. In the case of the incompressible Navier Stokes equations considered in chapter 5, this leads to global momentum conservation thanks to the term $\langle\widetilde{\boldsymbol{u}}\cdot\nabla\boldsymbol{u}_h,\boldsymbol{v}_h\rangle$. As shown in chapter 5 global momentum conservation holds if [84]

$$-\int_\Omega u_{h,1}\nabla\cdot\boldsymbol{u}_h\mathrm{d}\Omega+\int_\Omega\widetilde{\boldsymbol{u}}\cdot\nabla u_{h,1}\mathrm{d}\Omega=0.$$

  what is implied by the continuity equation provided $V_{h,1}\subseteqq Q_h$. This holds, in particular, for the "natural" choice $V_{h,1}=Q_h$. *For the standard Galerkin method, this condition is impossible to be satisfied*, since equal interpolation does not satisfy the inf-sup condition. In the same way if we consider $\psi_h=1$ in 6.9 and q is the

normal heat flux on the boundary Γ that results from considering the augmented problem [84] we have

$$\int_\Omega \partial_t \vartheta_h - \int_\Omega \vartheta_h \nabla \cdot \boldsymbol{u}_h + \int_\Omega \partial_t \widetilde{\vartheta} + \int_\Omega \widetilde{\boldsymbol{u}} \cdot \nabla \vartheta_h + \int_\Gamma \vartheta_h \boldsymbol{u}_n \cdot \boldsymbol{n} \mathrm{d}\Gamma = \int_\Omega Q + \int_\Gamma q \mathrm{d}\Gamma,$$

As in the case of the momentum equation, global energy conservation is obtained from the continuity equation provided $\Psi_h \subseteqq Q_h$ what holds if the temperature is interpolated in the same way as the pressure. As a conclusion, *the term* $\langle \tilde{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}_h, \boldsymbol{v}_h \rangle$ *provides global momentum conservation and the term* $\langle \tilde{\boldsymbol{u}}^{n+\theta} \cdot \nabla \vartheta_h, \psi_h \rangle$ *provides global energy conservation.*

On the other hand the tracking of subscales along the nonlinear and coupling processes opens the possibility of modelling turbulence. Some comments about this possibility have been made in chapter 5 where the reader is referred. Let us only mention that the formulation we propose would account for thermal turbulence in a very natural way. The traditional approach is to relate thermal turbulence to the mechanical one through the introduction of a turbulent Prandtl number whose physical meaning and adequate value are not well understood. This would be unnecessary with the approach presented here.

6.6 Numerical examples

In this section we present the results of two numerical tests involving two-dimensional thermally coupled flows. In both cases we have used $\widetilde{P} = I$ in 6.10-6.12, which corresponds to the most classical stabilized finite element methods.

In both numerical examples, our purpose is to compare the numerical performance of quasi-static subscales (QSS) and dynamic subscales (DS). To this end, we will proceed as follows. Two meshes will be considered in both examples, one that we will call "coarse" and another finer one. On both meshes we will present the results obtained using QSS and, only in the coarse mesh, the results obtained considering dynamic subscales. The goal is to show that DS yield better results than QSS on the coarse meshes by comparing both to the QSS results on the fine mesh, that we will call *reference* results. We anticipate that the conclusions of the following numerical experiments are

- The accuracy is higher using DS. This is reflected in particular by less damping of frequencies and amplitudes in the oscillating response of the flows considered.
- Stability is improved by using DS, particularly when subscales are integrated in time with a first order scheme. Some oscillations encountered with QSS are removed.

A full description of the iterative scheme developed for solving thermally coupled flows is presented in chapter 7, including different possibilities for the treatment of the

nonlinearities. In the examples presented here, the velocity-temperature coupling has been achieved using a block-iterative strategy, using also a nested iterative loop to solve the nonlinear Navier-Stokes equations within each coupling iteration. After solving for the finite element component of the velocity (temperature) the subgrid scale velocity iterating 6.22 (6.24). When the flow is fully developed, it converges very well, yielding fully converged subscales (with relative residuals of the order of 10^{-6}) with four or five iterations. Concerning the time integration schemes, the equations for the finite element unknowns have been integrated using the second order Crank-Nicolson scheme ($\theta = 1/2$ in 6.19-6.21), whereas the equations for the subscales have been integrated either using this same scheme or the first order version described in Section 3.1.

6.6.1 Thermoconvective instability of plane Poiseuille flow

The problem consists of a two-dimensional laminar flow in a horizontal channel occupying the domain $[0, 10]\times[0, 1]$ and suddenly heated from below. A parabolic inlet velocity profile is prescribed at $x = 0$, whereas the outlet is left free, i.e., the associated natural boundary condition is zero traction. The temperature is prescribed to $\vartheta = 1$ at the bottom wall $y = 0$ and to $\vartheta = 0$ at the top wall $y = 1$. The inlet and outlet are considered adiabatic.

This problem was solved in [46] as a benchmark for open boundary flows using a finite difference method and a fine grid. It can be considered as a model for several relevant engineering problems, such as the fabrication of microelectronic circuits using the chemical vapor deposition process (cf. [46], see references therein).

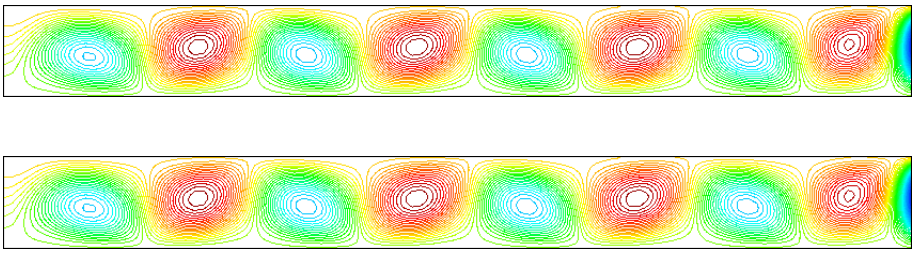

Figure 6.1: Streamlines at two different time steps for the plane Poiseuille flow example.

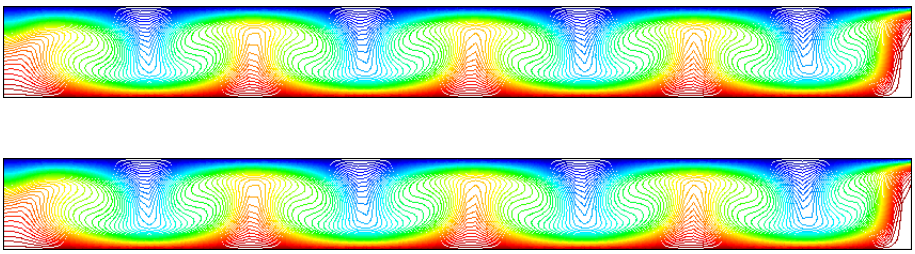

Figure 6.2: Temperature contours at two different time steps for the plane Poiseuille flow example.

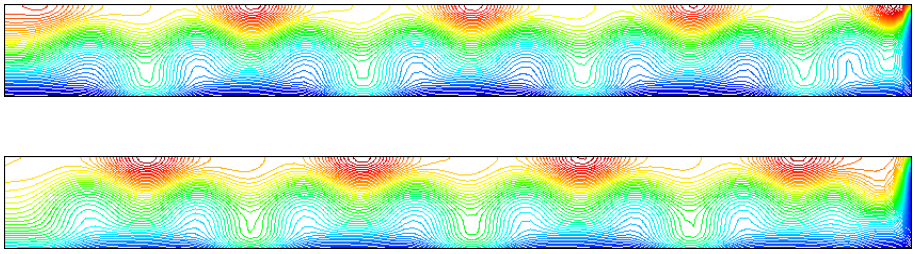

Figure 6.3: Pressure contours at two different time steps for the plane Poiseuille flow example.

The dimensionless parameters of the problem have been taken as $\mathrm{Re} = 10$, $\mathrm{Fr}^2 = 1/150$ and $\mathrm{Pe} = 40/9$ (the average inlet velocity, the height of the channel and the temperature difference between the top and bottom walls have been chosen as reference values for velocity, length and temperature, respectively). These parameters are the same as in [46] except for the Péclet number, which is slightly higher in that work ($\mathrm{Pe} = 20/3$). In both cases, these values result in a thermoconvective instability of the basic Poiseuille flow. The linear stability analysis of unstable stratified plane Poiseuille flow in a infinite horizontal channel can be found in [56]. It is shown there that the form of the instability could vary from traveling transverse waves to longitudinal rolls, with axes parallel to the main flow direction and thus leading to a three-dimensional flow pattern. Traveling transverse waves are found for small values of the Rayleigh number. This is the situation for the dimensionless parameters used here and therefore a two-dimensional calculation is possible. It should be remarked, however, that three-dimensional effects are in general very important for thermally coupled flows [94].

The domain $[0, 10] \times [0, 1]$ has been discretized using two uniform meshes of 16×40 and 50×100 bilinear elements, respectively. For the length of the channel considered, it is concluded in [46] that the numerical solution is not affected by the artificial boundary conditions for $2 \leq x \leq 8$.

Some results of the calculation on the fine mesh are shown in Figures 6.1, 6.2 and 6.3. They display the streamlines, temperature contours and pressure contours obtained at two time steps (roughly) half-a-period apart. The bad influence of the artificial boundary conditions can be observed, especially in what concerns the outlet wall. It is clear that the zero traction prescription does not reproduce the effect of an infinitely long channel. The proper evaluation of boundary conditions necessary for the numerical simulation of flows in infinite domains is an area that still deserves further of research.

The important point is the comparison of the results obtained using QSS and DS. To do this, we compare the evolution in time of velocity and pressure at the central point of the computational domain, $(x, y) = (5, 0.5)$. Results using $\delta t = 0.02$ on the fine mesh and $\delta t = 0.1$ on the coarse mesh are shown in Figure 6.4. For the DS case, two options have been considered, namely, a second order time integration of the subscales, labeled DS2 in

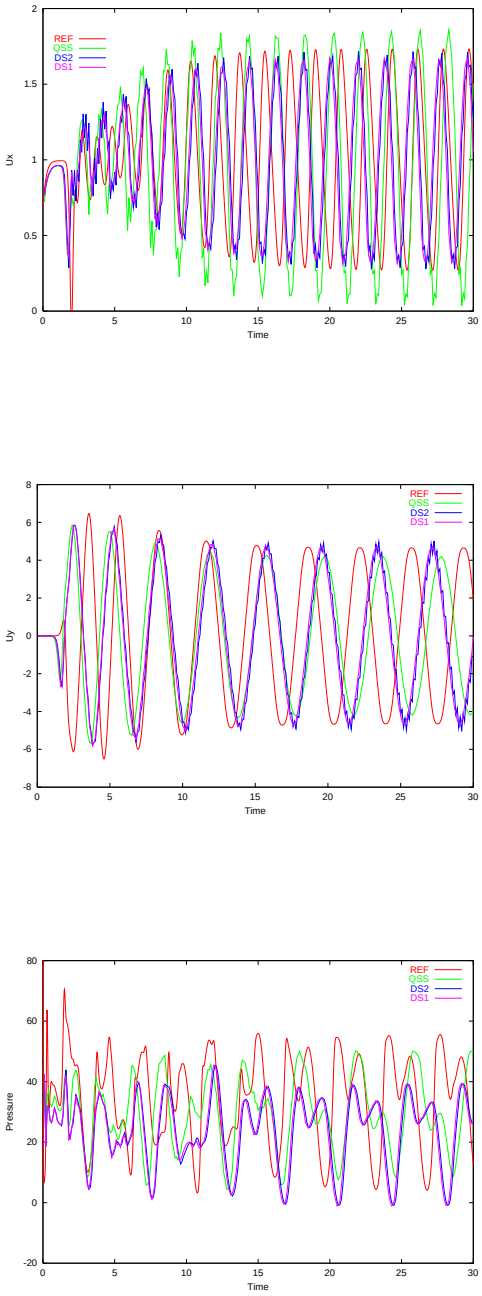

Figure 6.4: Time evolution at the central point for the plane Poiseuille flow example. Time step $\delta t = 0.1$. Top: horizontal velocity; Middle: Vertical velocity; Bottom: Pressure. REF: Reference solution; QSS: Solution with quasi-static subscales; DS2: Dynamic subscales with second order time integration; DS1: Dynamic subscales with first order time integration.

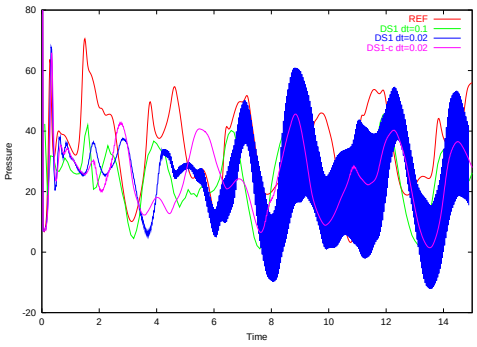

Figure 6.5: Pressure evolution in time at the central point for the plane Poiseuille flow. REF: Reference solution; DS1: Dynamic subscales with first order time integration without $\delta_t R_p^n$ in the pressure subscale. DS1-c: Same as DS1 but including $\delta_t R_p^n$ in the pressure subscale.

Figure 6.4, and a first order time integration, labeled DS1. From Figure 6.4 the following observations can be made:

- Results using DS1 and DS2 are very similar. This confirms the discussion of Section 2.3 about the feasibility of using DS1 and keeping the order of approximation.

- DS2 has spurious high frequency oscillations that are removed using DS1. This is to be expected, since it is known that the Crank-Nicolson scheme is unable to remove high frequencies. Our approximation to the subscales is non-smooth (they are discontinuous from element to element), and those high frequency components will be probably present.

- DS results are much more accurate than QSS, since they are closer to the reference results (obtained using QSS on the fine mesh).

- QSS results have some spurious oscillations in velocity that do not appear using DS. This is an important fact, since QSS are the results obtained with what can be considered a *standard* stabilized finite element method.

As a conclusion, results using DS1 seem to be excellent. Nevertheless, it is interesting to show in this example the effect of the term δR_p^n in 6.23. When $\delta t = 0.1$, this term is not important, but when $\delta t = 0.02$ its omission leads to a very important pressure oscillation from time step to time step using DS1. Figure 6.5 shows this oscillation, together with the results obtained including δR_p^n in 6.23, which are completely free of spurious oscillations.

6.6.2 Transient natural convection of low-Prandtl-number fluids

In this example, the transient convective motion of a fluid enclosed in a unit square cavity driven by a temperature gradient will be numerically analyzed. The left vertical wall is suddenly heated and maintained at a constant temperature, while the right vertical wall is maintained at the initial temperature. Horizontal walls are assumed to be adiabatic, i.e., the zero heat flux boundary condition is prescribed. Homogeneous Dirichlet boundary conditions are prescribed everywhere on the boundary for the velocity.

The only dimensionless parameters involved in the problem are the Prandtl number Pr and the Rayleigh number Ra or, equivalently, the Grashof number Gr. Numerical results will be presented for $\mathrm{Pr} = 0.005$ and the value $\mathrm{Gr} = 5 \times 10^6$.

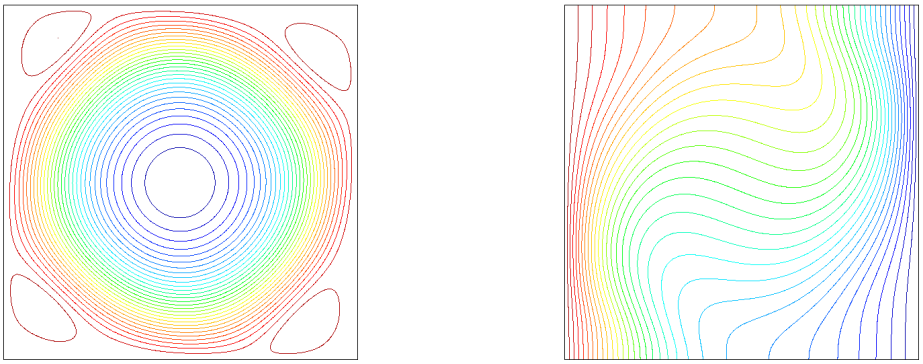

Figure 6.6: General streamline pattern (left) and temperature contours (right) for the flow in cavity at low Prandtl number.

The value $\mathrm{Pr} = 0.005$ is very small and not often encountered in common fluids. For example, the Prandtl number is 0.71 for air, 7.03 for water and 0.0249 for mercury (at 293 K). Small values of Pr are typical of liquid metals and semiconductors. The problem to be studied now is relevant to the solidification of ingots and casting, crystal growth from melts, material processing, nuclear reactor safety and other applications (cf. [112]).

Although the problem just described is a very popular test for thermally coupled flows when Pr is high, the interest for solving low-Prandtl-number flows is that this problem is not yet well understood. It is found that the flow exhibits a periodic oscillation when the Grashof number exceeds a critical value. In particular, for $\mathrm{Pr} = 0.005$ a steady-state solution is obtained for $\mathrm{Gr} = 3 \times 10^6$ but the solution bifurcates and for $\mathrm{Gr} = 5 \times 10^6$ an oscillatory flow field is found. For further information about this problem the reader is referred to [112], from where this problem has been taken. Our purpose here is to demonstrate the efficiency of the numerical method proposed in this work.

Two meshes of bilinear finite elements have been used in the calculations. The "coarse" one is made of 60×60 elements, refined near the walls of the cavity. The "fine" mesh is made of 180×180 elements, and it is also refined near the walls. The time step has been taken as $\delta t = 0.002$ in both cases. A remark is needed concerning the consequence of

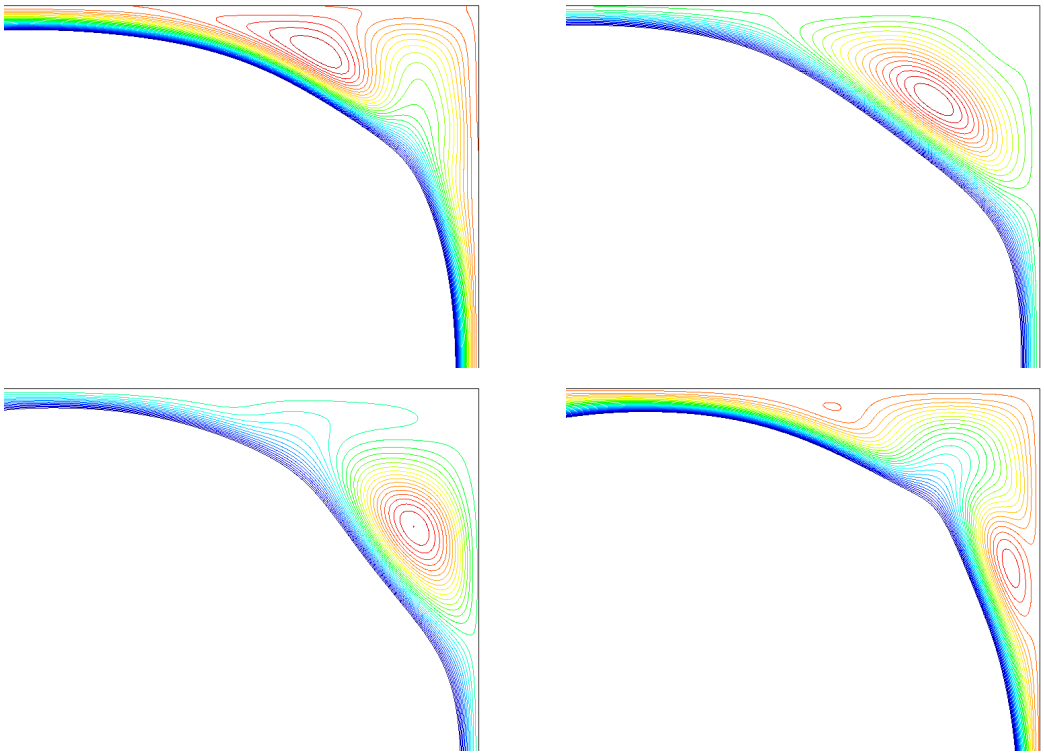

Figure 6.7: Evolution (from left to right and from top to bottom) of the streamlines at the rop right corner of the cavity for the flow in cavity at low Prandtl number.

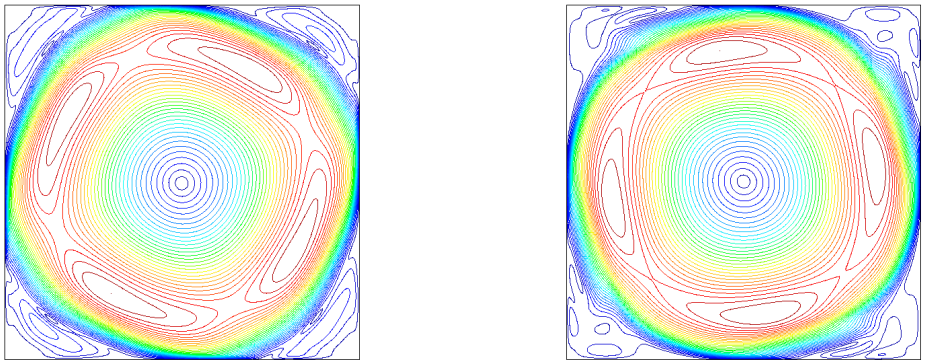

Figure 6.8: Velocity norm at two different time steps separated half a period for the flow in cavity at low Prandtl number.

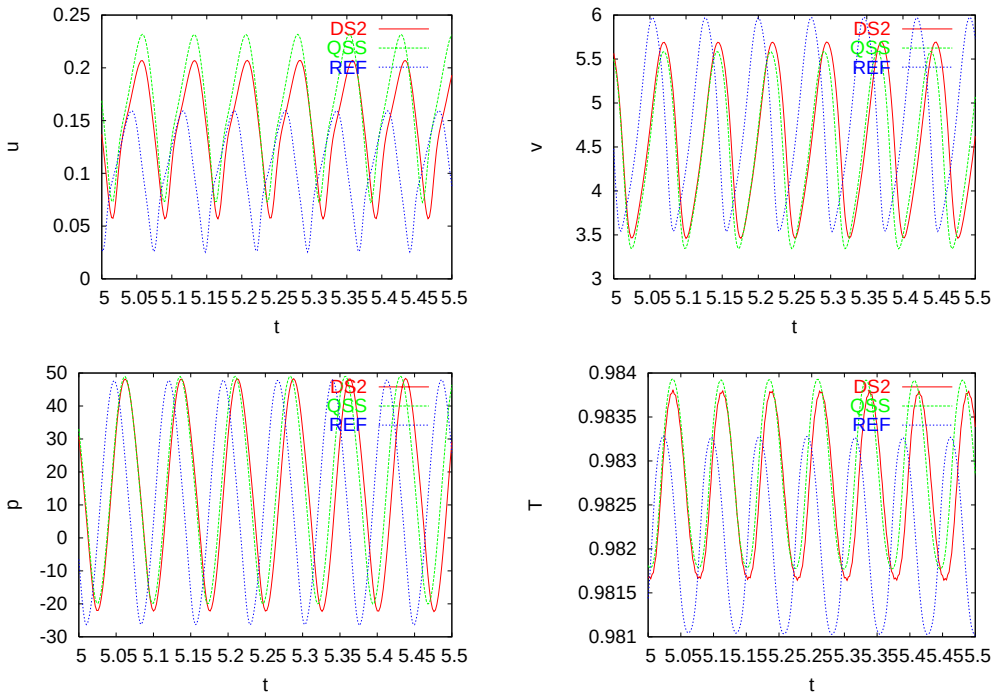

Figure 6.9: Comparison of results at point 1: $(x, y) = (0.006, 0.5)$ for the flow in cavity at low Prandtl number. u: Horizonal velocity; v: Vertical velocity; p: Pressure; T: Temperature.

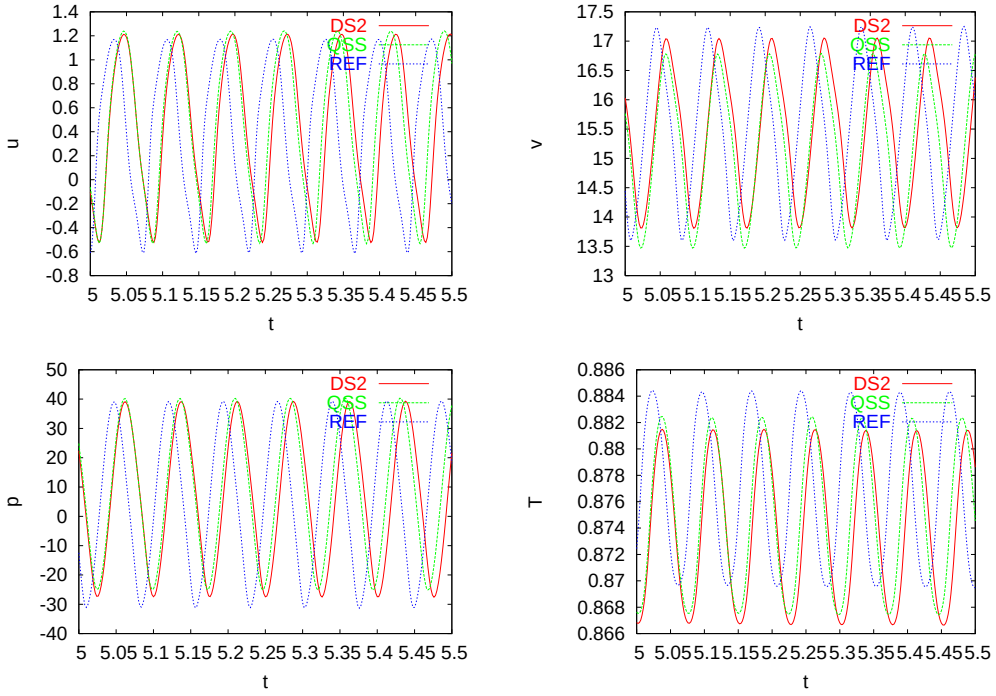

Figure 6.10: Comparison of results at point 2: $(x, y) = (0.0438, 0.5)$ for the flow in cavity at low Prandtl number. u: Horizonal velocity; v: Vertical velocity; p: Pressure; T: Temperature.

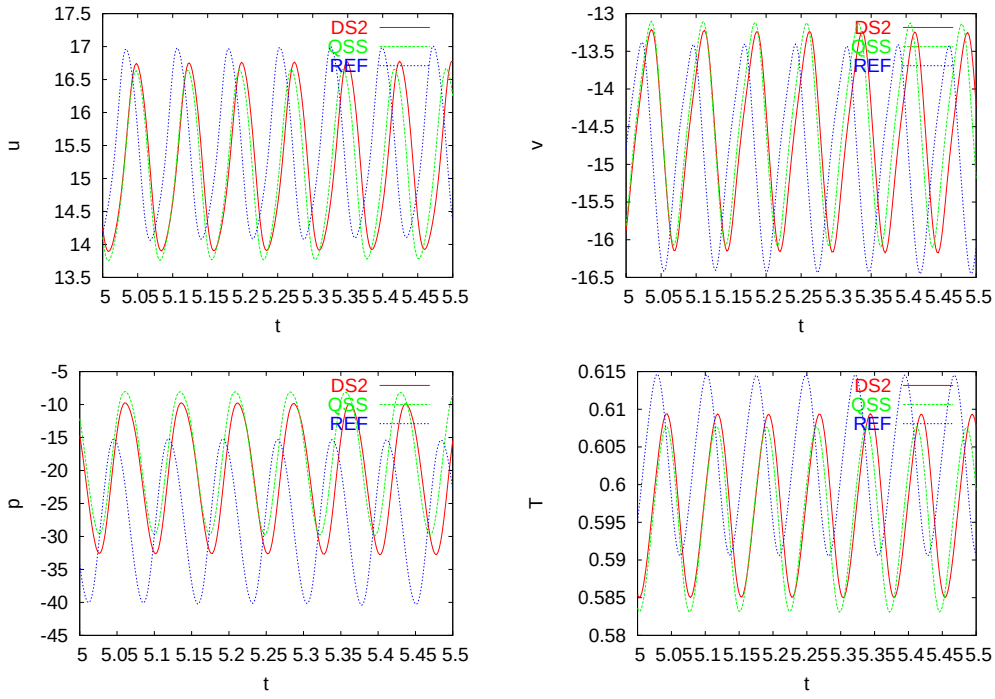

Figure 6.11: Comparison of results at point 6: $(x, y) = 0.773, 0.773)$ for the flow in cavity at low Prandtl number. u: Horizonal velocity; v: Vertical velocity; p: Pressure; T: Temperature.

this choice for the time integration of the Navier-Stokes and the temperature equations. The critical time step of the backward Euler scheme, obtained by taking $\theta = 0$ in 6.19-6.21, is approximately τ_1 for 6.19 and τ_3 for 6.21 [37]. Due to the low Prandtl number of the flow, the temperature equation is dominated by thermal diffusivity, whereas convective effects are important only in the Navier-Stokes equations. It turns out that the ratio $\delta t/\tau_1$ is 4.77 for the coarse mesh and 15.35 for the fine one, indicating that δt is comparable with τ_1. However, the ratio $\delta t/\tau_3$ is 222 for the coarse mesh and 2000 for the fine one. Therefore, the time step $\delta t = 0.002$ is very "large" for the time integration of the temperature equation and, as a consequence, not much influence is to be expected between quasi-static and dynamic subscales, particularly when diffusive effects dominate, as in boundary layers. Numerical results confirm this fact, as we shall show.

Let us discuss now the results of the numerical simulation. The general flow pattern is shown in Figure 6.6 at a time step when the flow is fully developed. It is observed that there is a main central vortex and also that vortices appear at each corner of the cavity. These small vortices move in clockwise sense, being created from flow detachment at the walls, growing and then collapsing against the walls. This evolution for the top right vortex can be observed in Figure 6.7. It is seen how the vortex is originated from the top wall, moves in the clockwise sense while grows, and then decreases until it reaches

the right wall. Before it completely disappears, a new vortex appears at the top wall. The contours of the velocity norm are plotted in Figure 6.8. These results correspond to time steps separated by half a period (approximately). They show that the main vortex pulsates, increasing and decreasing the flow magnitude periodically. All these results have been obtained with the fine mesh and QSS.

To compare the performance of QSS and DS we have considered three representative points. Point 1 is located at (0.006, 0.5), point 2 at (0.0438,0.5) and point 3 at (0.773,0.773). The first two points lie inside the boundary layer formed at the left wall, whereas the third one is placed at the top right position of the main vortex. Figures 6.9, 6.10 and 6.11 show the evolution in time of the flow variables (horizontal velocity, vertical velocity, pressure and temperature) at points 1, 2 and 6, respectively. From these pictures it is observed that all flow variables are more accurate using DS than QSS at points 1 and 6, whereas the results are inconclusive at point 2, where temperature seems to be slightly better using QSS (although the differences with DS are very small). The rest of flow variables are slightly better reproduced using DS. The explanation we give to this fact relies on the previous discussion about the size of the time step. As mentioned earlier, this time step is large for the heat equation, and thus QSS and DS should perform similarly, as it is observed in the numerical experiments. This is particularly so in boundary layers, since diffusive effects dominate there. At other sampling points of the computational domain, QSS performs consistently better than DS, in accordance with the results of the previous example. In this particular example, both the finite element equations and the equations for the subscales have been integrated in time with second order accuracy.

6.7 Conclusions

The aim of this chapter has been to explain how to deal with dynamic subscales in the finite element approximation of thermally coupled flows using the Boussinesq approximation. The space variation of the subscales is approximated in terms of the residual of the finite element unknowns in the classical way used in stabilized finite element methods, but now they are integrated in time.

From the conceptual point of view, the formulation presented has several benefits, inherited from the formulation applied to isothermal flows in chapter 5. In particular, global momentum conservation and global energy conservation is obtained. Additionally, in the case of thermally coupled flows the coupling of velocity and temperature subscales is dealt with in a natural way. The results of the numerical experiments conducted confirm the conclusions drawn for isothermal flows and that make the formulation particularly appealing:

- The formulation is more accurate than considering the subscales quasi-static.

- Some oscillations encountered using quasi-static subscales are removed.

The last item is especially significant when the subscales are integrated in time using a first order scheme, which avoids high frequency spurious oscillations in the tracking of the subscales in time.

Chapter 7

Numerical implementation aspects

In this chapter we present the strategy developed for the numerical solution of the stabilized finite element approximation of thermally coupled flows. The implementation algorithm is developed considering several possibilities for the solution of the discrete nonlinear problem. The full Newton linearization strategy gives rise to monolithic treatment of the coupling of variables whereas some fixed point schemes permit the segregated treatment of velocity-pressure and temperature. The first one turns out to be very efficient for steady-state problems and very robust when it is combined with a line search strategy that has been developed based on the Armijo rule. A segregated treatment of velocity-pressure and temperature happens to be more appropriate for transient problems.

7.1 Introduction

The approximated models considered in previous chapters can be written in a unified manner as a system of nonlinear convection-diffusion-reaction equations of the form

$$\boldsymbol{M}(\boldsymbol{U}_0)\frac{\partial \boldsymbol{U}}{\partial t} + \mathcal{L}(\boldsymbol{U};\boldsymbol{U}) = \boldsymbol{F} \quad \text{in } \Omega \tag{7.1}$$

where

$$\mathcal{L}(\boldsymbol{U}_0;\boldsymbol{U}) := \boldsymbol{A}_i(\boldsymbol{U}_0)\frac{\partial \boldsymbol{U}}{\partial x_i} - \frac{\partial}{\partial x_i}\left(\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}}{\partial x_j}\right) + \boldsymbol{S}(\boldsymbol{U}_0)\boldsymbol{U}$$

and $\boldsymbol{U} = (\boldsymbol{u}, p, \vartheta)$ is the vector of unknowns (velocity $\boldsymbol{u}$, pressure p and temperature ϑ), $\boldsymbol{F}$ is a known vector of $n_{\text{unk}} = \text{d} + 2$ components and $\boldsymbol{M}$, $\boldsymbol{A}_i$, $\boldsymbol{K}_{ij}$ and $\boldsymbol{S}$ are $n_{\text{unk}} \times n_{\text{unk}}$ matrices ($i, j = 1, ..., \text{d}$). The usual summation convention is implied in the last expression, with indices running from 1 to the number of space dimensions d and bold characters are used to denote vectors. When the arguments used to evaluate the matrices $\boldsymbol{M}$, $\boldsymbol{A}_i$ and $\boldsymbol{S}$ are clear we will omit them as well as the first argument in $\mathcal{L}$. We shall refer to the terms of the left-hand-side (LHS) of this equation as the temporal,

the convective, the diffusive and the reactive terms. The physical models presented in chapter 1 are written for the two-dimensional case (d = 2) as

- **Incompressible Navier Stokes equations**:

$$\boldsymbol{M} = \begin{bmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{A}_i(\boldsymbol{U}) = \begin{bmatrix} \rho u_i & 0 & \delta_{i1} \\ 0 & \rho u_i & \delta_{i2} \\ \delta_{i1} & \delta_{i2} & 0 \end{bmatrix}$$

$$\boldsymbol{K}_{ij} = \begin{bmatrix} \mu\delta_{ij} + \mu\delta_{i1}\delta_{j1} + \frac{2\mu}{3}\delta_{i1}\delta_{j1} & \mu\delta_{i2}\delta_{j1} + \frac{2\mu}{3}\delta_{i1}\delta_{j2} & 0 \\ \mu\delta_{i1}\delta_{j2} + \frac{2\mu}{3}\delta_{i2}\delta_{j1} & \mu\delta_{ij} + \mu\delta_{i2}\delta_{j2} + \frac{2\mu}{3}\delta_{i2}\delta_{j2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{S}(\boldsymbol{U}) = \boldsymbol{0}, \quad \boldsymbol{F} = \boldsymbol{0}$$

  where ρ is the density and μ the viscosity.

- **Boussinesq equations**:

$$\boldsymbol{M} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho c_p \end{bmatrix}, \quad \boldsymbol{A}_i(\boldsymbol{U}) = \begin{bmatrix} \rho u_i & 0 & \delta_{i1} & 0 \\ 0 & \rho u_i & \delta_{i2} & 0 \\ \delta_{i1} & \delta_{i2} & 0 & 0 \\ 0 & 0 & 0 & \rho c_p u_i \end{bmatrix}$$

$$\boldsymbol{K}_{ij} = \begin{bmatrix} \mu\delta_{ij} + \mu\delta_{i1}\delta_{j1} + \frac{2\mu}{3}\delta_{i1}\delta_{j1} & \mu\delta_{i2}\delta_{j1} + \frac{2\mu}{3}\delta_{i1}\delta_{j2} & 0 & 0 \\ \mu\delta_{i1}\delta_{j2} + \frac{2\mu}{3}\delta_{i2}\delta_{j1} & \mu\delta_{ij} + \mu\delta_{i2}\delta_{j2} + \frac{2\mu}{3}\delta_{i2}\delta_{j2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k \end{bmatrix}$$

$$\boldsymbol{S}(\boldsymbol{U}) = \begin{bmatrix} 0 & 0 & 0 & \rho\beta g_1 \\ 0 & 0 & 0 & \rho\beta g_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ Q \end{bmatrix}$$

  where β is the thermal expansion coefficient, c_p is the constant pressure specific heat and Q is a given external source of heat.

- **Low Mach number model**: we consider an ideal gas

$$\rho = \frac{p^{\text{th}}}{R\vartheta}$$

  what permits to write the continuity equation (see chapter 1) as

$$-\frac{\rho}{\vartheta}\frac{\partial\vartheta}{\partial t} + \frac{\rho}{p^{\text{th}}}\frac{dp^{\text{th}}}{dt} - \frac{\rho}{\vartheta}\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta + \rho\boldsymbol{\nabla}\cdot\boldsymbol{u} = 0$$

This is used to write the matrices that define the problem as

$$\boldsymbol{M}(\boldsymbol{U}) = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\vartheta} \\ 0 & 0 & 0 & \rho c_p \end{bmatrix}, \quad \boldsymbol{A}_i(\boldsymbol{U}) = \begin{bmatrix} \rho u_i & 0 & \delta_{i1} & 0 \\ 0 & \rho u_i & \delta_{i2} & 0 \\ \delta_{i1} & \delta_{i2} & 0 & -\frac{1}{\vartheta} u_i \\ 0 & 0 & 0 & \rho c_p u_i \end{bmatrix}$$

$$\boldsymbol{K}_{ij} = \begin{bmatrix} \mu\delta_{ij} + \mu\delta_{i1}\delta_{j1} + \frac{2\mu}{3}\delta_{i1}\delta_{j1} & \mu\delta_{i2}\delta_{j1} + \frac{2\mu}{3}\delta_{i1}\delta_{j2} & 0 & 0 \\ \mu\delta_{i1}\delta_{j2} + \frac{2\mu}{3}\delta_{i2}\delta_{j1} & \mu\delta_{ij} + \mu\delta_{i2}\delta_{j2} + \frac{2\mu}{3}\delta_{i2}\delta_{j2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k \end{bmatrix}$$

$$\boldsymbol{S}(\boldsymbol{U}) = \begin{bmatrix} 0 & 0 & 0 & \rho g_1 \\ 0 & 0 & 0 & \rho g_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{p^{\mathrm{th}}}\frac{dp^{\mathrm{th}}}{dt} \\ \frac{dp^{\mathrm{th}}}{dt} + Q \end{bmatrix}$$

Note that it is also possible to use the energy equation to write the continuity equation as

$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = -\frac{1}{\gamma p^{\mathrm{th}}}\frac{dp^{\mathrm{th}}}{dt} + \frac{\gamma - 1}{\gamma p^{\mathrm{th}}}\left[\boldsymbol{\nabla}\cdot\left(k\nabla\vartheta\right) + Q\right].$$

and that, as the density is temperature dependent, the temporal term is nonlinear.

The boundary conditions of these problems are

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{u}_d \quad \text{on } \Gamma_D^u \\ \vartheta &= \vartheta_d \quad \text{on } \Gamma_D^\vartheta \\ \boldsymbol{\sigma}\cdot\boldsymbol{n} &= \left(-p\boldsymbol{I} + 2\mu\boldsymbol{\varepsilon}\prime(\boldsymbol{u})\right)\cdot\boldsymbol{n} = \boldsymbol{t} \quad \text{on } \Gamma_N^u \\ \boldsymbol{q}\cdot\boldsymbol{n} &= -k\boldsymbol{n}\cdot\boldsymbol{\nabla}\theta = q_n \quad \text{on } \Gamma_N^\vartheta \end{aligned}$$

where Γ_D^α (Γ_N^α) is the part of the domain boundary where Dirichlet (Neumann) boundary conditions are given and $\Gamma = \partial\Omega = \overline{\Gamma_N^\alpha \cup \Gamma_D^\alpha}$, where α is either the velocity $\boldsymbol{u}$ or the temperature ϑ. In order to write boundary conditions in a unified manner we split matrices $\boldsymbol{A}_i$ as $\boldsymbol{A}_i = \boldsymbol{A}_i^{\mathrm{c}} + \boldsymbol{A}_i^{\mathrm{f}}$, where $\boldsymbol{A}_i^{\mathrm{c}}$ is the part of the convection matrices which *is not* integrated by parts and $\boldsymbol{A}_i^{\mathrm{f}}$ the part that *is* integrated by parts. This matrix contains pressure terms and in the case of the incompressible Navier Stokes equations is given by

$$\boldsymbol{A}_i^{\mathrm{f}} = \begin{bmatrix} 0 & 0 & \delta_{i1} \\ 0 & 0 & \delta_{i2} \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{A}_i^{\mathrm{c}} = \boldsymbol{A}_i - \boldsymbol{A}_i^{\mathrm{f}}$$

This permits to define the vector of fluxes in terms of the matrices $\boldsymbol{K}_{ij}$ and $\boldsymbol{A}_i^{\mathrm{f}}$ so in the simple case in which $\Gamma_N^u = \Gamma_N^\vartheta := \Gamma_N \subset \partial\Omega$. we can write Neumann conditions as

$$\boldsymbol{n}_i\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}}{\partial x_j} - \boldsymbol{n}_i\boldsymbol{A}_i^{\mathrm{f}}\boldsymbol{U} = \boldsymbol{T} \quad \text{in} \quad \Gamma_N$$

Initial conditions have to be appended to close the problem.

Let us denote by $\boldsymbol{\mathcal{W}}$ the functional space where the solution is to be sought, by $W^{m,p}(\omega)$ the usual Sobolev spaces and, in particular $H^m(\omega) := W^{m,2}(\omega)$ and by $L^2(\omega)$ the space of square integrable functions in a domain ω. In a steady state case, the velocity belongs to the space $\boldsymbol{V}^{\mathrm{st}} = \left\{\boldsymbol{u} \in [H^1(\Omega)]^{n_{\mathrm{sd}}} : \boldsymbol{u} = \boldsymbol{u}_d \,\mathrm{in}\, \Gamma_N^u\right\}$, the pressure to the space $Q^{\mathrm{st}} = L^2(\Omega)/\mathbb{R}$ and the temperature to the space $\Psi^{\mathrm{st}} = \left\{\vartheta \in H^1(\Omega) : \vartheta = \vartheta_d \,\mathrm{in}\, \Gamma_N^\vartheta\right\}$. When a transient problem defined in the interval $[0,T]$ is considered, the space of time dependent functions defined in a space X whose norm is $L^p(0,T)$ will be denoted by $L^p(0,T;X)$. Then the space $\boldsymbol{\mathcal{W}}$ is defined as $\boldsymbol{\mathcal{W}} = L^2\left(0,T,\boldsymbol{V}^{\mathrm{st}}\right) \times L^1\left(0,T,L^2(\Omega)\right) \times L^2\left(0,T,\Psi^{\mathrm{st}}\right)$ and $\boldsymbol{\mathcal{W}}_0$, the corresponding space of test functions which is given by $\boldsymbol{\mathcal{W}}_0 = \boldsymbol{V}_0 \times L^2(\Omega) \times \Psi_0$ where $\boldsymbol{V}_0 = \left\{\boldsymbol{u} \in [H^1(\Omega)]^{n_{\mathrm{sd}}} : \boldsymbol{u} = \boldsymbol{0} \,\mathrm{in}\, \Gamma_D^u\right\}$ and $\Psi_0 = \left\{\vartheta \in H^1(\Omega) : \vartheta = 0 \,\mathrm{in}\, \Gamma_D^\vartheta\right\}$.Then, the weak form of the problem consists in finding $\boldsymbol{U} \in \boldsymbol{\mathcal{W}}$ such that

$$B(\boldsymbol{U};\boldsymbol{U},\boldsymbol{V}) - L(\boldsymbol{V}) = 0 \qquad \forall \boldsymbol{V} \in \boldsymbol{\mathcal{W}}_0 \tag{7.2}$$

where the nonlinear form B and the linear form L are defined as

$$\begin{aligned} B(\boldsymbol{U}_0;\boldsymbol{U},\boldsymbol{V}) := & \int_\Omega \boldsymbol{V}^t \boldsymbol{M}(\boldsymbol{U}_0) \frac{\partial \boldsymbol{U}}{\partial t} + \int_\Omega \boldsymbol{V}^t \boldsymbol{A}_i^{\mathrm{c}}(\boldsymbol{U}_0) \frac{\partial \boldsymbol{U}}{\partial x_i} \\ & - \int_\Omega \frac{\partial}{\partial x_i}\left(\boldsymbol{V}^t \boldsymbol{A}_i^{\mathrm{f}}\right)\boldsymbol{U} + \int_\Omega \frac{\partial \boldsymbol{V}^t}{\partial x_i} \boldsymbol{K}_{ij}(\boldsymbol{U}_0) \frac{\partial \boldsymbol{U}}{\partial x_j} + \int_\Omega \boldsymbol{V}^t \boldsymbol{S}(\boldsymbol{U}_0)\boldsymbol{U} \end{aligned} \tag{7.3}$$

$$L(\boldsymbol{V}) := \int_\Omega \boldsymbol{V}^t \boldsymbol{F} + \int_\Gamma \boldsymbol{V}^t \boldsymbol{T}\, \mathrm{d}\Gamma \tag{7.4}$$

Note that the second term in 7.4 could be written as an integral over $\Gamma_N^u \cup \Gamma_N^\vartheta$ because $\boldsymbol{V}^t = \boldsymbol{0}$ in the rest of the domain boundary.

7.2 Discrete problem

We consider a finite element partition $\mathcal{P}_h = \{K\}$ of the computational domain Ω of n_{el} elements, from which we can construct finite element spaces for the velocity, pressure and temperature. We assume that they are all built from continuous piecewise polynomials of the same degree k. We denote by $\boldsymbol{\mathcal{W}}_h \subset \boldsymbol{\mathcal{W}}$ the approximating space, by $\boldsymbol{\mathcal{W}}_{0h} \subset \boldsymbol{\mathcal{W}}_0$ the space of test functions and by $\widetilde{\boldsymbol{\mathcal{W}}}$ the space of subscales. We also consider a uniform partition of the time interval $[0,T]$ of time step size δt. Functions approximated at time t^n will be identified with the superscript n. For a generic function f, we will use the notation $\delta f^n := f^{n+1} - f^n$, $\delta_t f^n = \delta f^n / \delta t$, $f^{n+\theta} = \beta f^{n+1} + (1-\theta) f^n$, $0 \le \theta \le 1$. Using this notation the fully discrete problem obtained using the variational multiscale formulation of [75] developed in the previous chapters can be written as follows. Given

order	1	2	Exact
τ_t	$\tau\left(1+\frac{1}{\frac{\Delta t}{\rho\tau}}\right)^{-1}$	$\tau\left(1+\frac{1}{\frac{\theta\Delta t}{\rho\tau}}\right)^{-1}$	$\tau\left(1-\exp\left[-\frac{\theta\Delta t}{\rho\tau}\right]\right)$
μ	$\frac{\rho}{\Delta t}$	$\frac{\rho}{\theta\Delta t}$	$\frac{1}{\tau_t}\left(1-\frac{\tau_t}{\tau}\right)$

Table 7.1: Integration parameters

$\boldsymbol{U}_h^n$ and $\widetilde{\boldsymbol{U}}^n$, find $\boldsymbol{U}_h^{n+\theta}$ and $\widetilde{\boldsymbol{U}}^{n+\theta}$ such that

$$\int_\Omega \boldsymbol{V}_h^t \boldsymbol{M}\delta_t \boldsymbol{U}_h^n + \int_\Omega \boldsymbol{V}_h^t \boldsymbol{A}_i^{\mathrm{c}} \frac{\partial \boldsymbol{U}_h^{n+\theta}}{\partial x_i} + \int_\Omega \frac{\partial \boldsymbol{V}_h^t}{\partial x_i}\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}_h^{n+\theta}}{\partial x_j} + \int_\Omega \boldsymbol{V}_h^t \boldsymbol{S}\boldsymbol{U}_h^{n+\theta}$$
$$-\int_\Omega \frac{\partial}{\partial x_i}\left(\boldsymbol{V}_h^t \boldsymbol{A}_i^{\mathrm{f}}\right)\boldsymbol{U}_h^{n+\theta} + \int_\Omega \boldsymbol{V}_h^t \boldsymbol{M}\delta_t\widetilde{\boldsymbol{U}}^n + \sum_K \int_K \left[\mathcal{L}^*\left(\boldsymbol{V}_h\right)\right]^t \widetilde{\boldsymbol{U}}^{n+\theta} = \int_\Omega \boldsymbol{V}^t \boldsymbol{F}^{n+\theta} \quad (7.5)$$

for any $\boldsymbol{V}_h \in \boldsymbol{\mathcal{W}}_{0h}$ Here and in what follows it is understood that matrices $\boldsymbol{M}$, $\boldsymbol{A}_i$, $\boldsymbol{S}$ are evaluated using $\boldsymbol{U}_h^{n+\theta} + \widetilde{\boldsymbol{U}}^{n+\theta}$. In 7.5, $\mathcal{L}^*$ is the adjoint of the differential operator $\mathcal{L}$ with homogeneous Dirichlet conditions given by

$$\mathcal{L}^*(\boldsymbol{U}_0;\boldsymbol{U}) := -\frac{\partial}{\partial x_i}\left[\boldsymbol{A}_i^t(\boldsymbol{U}_0)\boldsymbol{U}\right] - \frac{\partial}{\partial x_i}\left(\boldsymbol{K}_{ij}^t(\boldsymbol{U}_0)\frac{\partial \boldsymbol{U}}{\partial x_j}\right) + \boldsymbol{S}^t(\boldsymbol{U}_0)\boldsymbol{U}$$

(evaluated using $\boldsymbol{U}_0 = \boldsymbol{U}_h^{n+\theta} + \widetilde{\boldsymbol{U}}^{n+\theta}$ in 7.5). The subscale $\widetilde{\boldsymbol{U}}^{n+\theta}$ is found as the solution of nonlinear problem

$$\widetilde{\boldsymbol{U}}^{n+\theta} = \boldsymbol{\tau}_t \boldsymbol{R}^{n+\theta} + \boldsymbol{\mu}\boldsymbol{\tau}_t \widetilde{\boldsymbol{U}}^n \quad (7.6)$$

driven by the residual

$$\boldsymbol{R}^{n+\theta} := \boldsymbol{F}^{n+\theta} - \boldsymbol{M}\delta_t \boldsymbol{U}_h^n - \mathcal{L}(\boldsymbol{U}_h^{n+\theta})$$

The temporal derivative of the subscale is calculated as

$$\delta_t \widetilde{\boldsymbol{U}}^n = \boldsymbol{\mu}\boldsymbol{\tau}_t \boldsymbol{R}^{n+\theta} - \boldsymbol{\mu}\boldsymbol{\tau}^{-1}\boldsymbol{\tau}_t \widetilde{\boldsymbol{U}}^n$$

and the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\tau}_t$ depend on the time integration scheme used for the subscales evolution equations as explained in chapter 6 where three options are considered. They are defined as

$$\begin{aligned} \boldsymbol{\tau}_t &= \mathrm{diag}(\tau_{t1}, \tau_{t1}, \tau_2, \tau_{t3}) \\ \boldsymbol{\mu} &= \mathrm{diag}(\mu_1, \mu_1, 0, \mu_3) \end{aligned}$$

where τ_{t1} and τ_{t3} as well as μ_1 and μ_3 are defined in table 7.1

The parameters τ_1, τ_2 and τ_3 are computed as in previous chapters. In the case of an isotropic mesh they are given by

$$\tau_1 = \left[c_1\frac{\mu}{h^2} + c_2\frac{\rho\left|\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}\right|}{h}\right]^{-1}, \quad \tau_2 = \frac{h^2}{c_1\tau_1}, \quad \tau_3 = \left[c_1\frac{k}{h^2} + c_2\frac{\rho c_p\left|\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}\right|}{h}\right]^{-1}$$

where $c_1 = 4$ and $c_2 = 2$ for linear elements. Using these definitions the final problem to be solved can be written as follows. Given $\boldsymbol{U}_h^n$ and $\widetilde{\boldsymbol{U}}^n$, find $\boldsymbol{U}_h^{n+\theta}$ such that

$$\begin{aligned}
&\frac{1}{\theta\delta t}\int_\Omega \boldsymbol{V}_h^t \boldsymbol{M}\boldsymbol{U}_h^{n+\theta} + \int_\Omega \boldsymbol{V}_h^t \boldsymbol{A}_i^{\mathrm{c}} \frac{\partial \boldsymbol{U}_h^{n+\theta}}{\partial x_i} - \int_\Omega \frac{\partial}{\partial x_i}\left(\boldsymbol{V}_h^t \boldsymbol{A}_i^{\mathrm{f}}\right)\boldsymbol{U}_h^{n+\theta} \\
&+ \int_\Omega \frac{\partial \boldsymbol{V}_h^t}{\partial x_i}\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}_h^{n+\theta}}{\partial x_j} + \int_\Omega \boldsymbol{V}_h^t \boldsymbol{S}\boldsymbol{U}_h^{n+\theta} \\
&+ \sum_K \int_K \left[-\boldsymbol{\mu}\boldsymbol{M}\boldsymbol{V}_h - \mathcal{L}^*\left(\boldsymbol{V}_h\right)\right]^t \boldsymbol{\tau}_t \left[\frac{1}{\theta\delta t}\boldsymbol{M}\boldsymbol{U}_h^{n+\theta} + \mathcal{L}(\boldsymbol{U}_h^{n+\theta})\right] \\
&\qquad = \int_\Omega \boldsymbol{V}^t \boldsymbol{F}^{n+\theta} + \sum_K \int_K \left[-\boldsymbol{\mu}\boldsymbol{M}\boldsymbol{V}_h - \mathcal{L}^*\left(\boldsymbol{V}_h\right)\right]^t \boldsymbol{\tau}_t \boldsymbol{F}^{n+\theta} \\
&\qquad + \frac{1}{\theta\delta t}\int_\Omega \boldsymbol{V}_h^t \boldsymbol{M}\boldsymbol{U}_h^n + \frac{1}{\theta\delta t}\sum_K \int_K \left[-\boldsymbol{\mu}\boldsymbol{M}\boldsymbol{V}_h - \mathcal{L}^*\left(\boldsymbol{V}_h\right)\right]^t \boldsymbol{\tau}_t \boldsymbol{M}\boldsymbol{U}_h^n \\
&\qquad + \sum_K \int_K \left[\boldsymbol{\tau}^{-1}\boldsymbol{M}\boldsymbol{V}_h + \mathcal{L}^*\left(\boldsymbol{V}_h\right)\right]^t \boldsymbol{\mu}\boldsymbol{\tau}_t \widetilde{\boldsymbol{U}}^n
\end{aligned} \tag{7.7}$$

for any $\boldsymbol{V}_h \in \boldsymbol{\mathcal{W}}_{0h}$ and find $\widetilde{\boldsymbol{U}}^{n+\theta}$ such that

$$\widetilde{\boldsymbol{U}}^{n+\theta} = \boldsymbol{\tau}_t \boldsymbol{R}^{n+\theta} + \boldsymbol{\mu}\boldsymbol{\tau}_t \widetilde{\boldsymbol{U}}^n$$

Note that the explicit dependence of the subscales on the residual of the finite element component has been explicitly taken into account and has been assembled on the left hand side of the equation. However, the problem still depends on the subscales (and therefore on the residual of the finite element component) through the matrices $\boldsymbol{M}$, $\boldsymbol{A}_i$, $\boldsymbol{S}$ and the operators $\mathcal{L}$ and $\mathcal{L}^*$ and also through the stabilization parameters. The non linear treatment of this system is described in the following section.

7.2.1 Linearization and line search strategy

The discrete approximation described in the previous section leads to a highly nonlinear system of algebraic equations for the nodal values of $\boldsymbol{U}_h^{n+\alpha}$, which are denoted by the same character (but without the subscript h). This nonlinear problem can be written as

$$\left[\boldsymbol{L} + \boldsymbol{N}\left(\boldsymbol{U}\right)\right]\boldsymbol{U} = \boldsymbol{R}$$

where $\boldsymbol{L}$ is the linear part of the operator and $\boldsymbol{N}$ the nonlinear one and $\boldsymbol{R}$ the force vector. Therefore we look for the roots of the function

$$\boldsymbol{H}\left(\boldsymbol{U}\right) = \left[\boldsymbol{L} + \boldsymbol{N}\left(\boldsymbol{U}\right)\right]\boldsymbol{U} - \boldsymbol{R}$$

and we consider fixed points linearizations of the form

$$\boldsymbol{U}^k = \boldsymbol{G}\left(\boldsymbol{U}^{k-1}\right) \tag{7.8}$$

where

$$\boldsymbol{G}(\boldsymbol{U}) = \boldsymbol{D}^{-1}(\boldsymbol{D}\boldsymbol{U} - \boldsymbol{H}(\boldsymbol{U}))$$

for some matrix $\boldsymbol{D}$ to be defined in the following and which may depend on the iteration step. Then, using a superscript for the iteration counter the iterative scheme reads

$$\boldsymbol{D}\left(\boldsymbol{U}^{i+1} - \boldsymbol{U}^{i}\right) + \boldsymbol{H}\left(\boldsymbol{U}^{i}\right) = \boldsymbol{0}$$

or

$$\boldsymbol{D}\left(\boldsymbol{U}^{i+1} - \boldsymbol{U}^{i}\right) + \left[\boldsymbol{L} + \boldsymbol{N}(\boldsymbol{U})\right]\boldsymbol{U}^{i} - \boldsymbol{R} = \boldsymbol{0}$$

Different choices of $\boldsymbol{D}$ led to different schemes:

- The classical Picard scheme is obtained by taking

  $$\boldsymbol{D} = \left[\boldsymbol{L} + \boldsymbol{N}(\boldsymbol{U})\right]$$

  from where the problem to be solved is

  $$\left[\boldsymbol{L} + \boldsymbol{N}\left(\boldsymbol{U}^{i}\right)\right]\boldsymbol{U}^{i+1} = \boldsymbol{R}$$

- The Newton scheme is obtained taking

  $$\boldsymbol{D} = \boldsymbol{H}'(\boldsymbol{U})$$

  where $\boldsymbol{H}'$ is the Jacobian of $\boldsymbol{H}$, from where

  $$\boldsymbol{H}'\left(\boldsymbol{U}^{i}\right)\left(\boldsymbol{U}^{i+1} - \boldsymbol{U}^{i}\right) + \boldsymbol{H}\left(\boldsymbol{U}^{i}\right) = \boldsymbol{0}$$

  Sometimes a modified Newton scheme is obtained by taking

  $$\boldsymbol{D} = \boldsymbol{H}'\left(\boldsymbol{U}^{0}\right)$$

- In the case of a steady state problem (just to fix ideas) another option is to take

  $$\boldsymbol{D} = \frac{1}{\varepsilon}\boldsymbol{M}$$

  to obtain

  $$\frac{1}{\varepsilon}\boldsymbol{M}\left(\boldsymbol{U}^{i+1} - \boldsymbol{U}^{i}\right) + \left[\boldsymbol{L} + \boldsymbol{N}\left(\boldsymbol{U}^{i}\right)\right]\boldsymbol{U}^{i} = \boldsymbol{R}$$

  This scheme produces the same iterates that an explicit temporal integration of the equations, what shows how a temporal evolution can be considered as a fixed point scheme for the solution of a nonlinear problem.

- In a similar way, if we take

$$\boldsymbol{D} = \frac{1}{\varepsilon}\boldsymbol{M} + \left[\boldsymbol{L} + \boldsymbol{N}\left(\boldsymbol{U}^i\right)\right]$$

we obtain a semi-implicit temporal evolution

$$\frac{1}{\varepsilon}\boldsymbol{M}\left(\boldsymbol{U}^{i+1} - \boldsymbol{U}^i\right) + \left[\boldsymbol{L} + \boldsymbol{N}\left(\boldsymbol{U}^i\right)\right]\boldsymbol{U}^{i+1} = \boldsymbol{R}$$

The convergence rate of the method depends on how contractive the mapping $\boldsymbol{G}$ is. Precisely [98], if there exists $\alpha < 1$ such that

$$\|\boldsymbol{G}(\boldsymbol{U}) - \boldsymbol{G}(\boldsymbol{V})\| \leq \alpha \|\boldsymbol{U} - \boldsymbol{V}\|$$

the mapping $\boldsymbol{G}$ has only one fixed point $\boldsymbol{U}_*$ and the iterative scheme

$$\boldsymbol{U}^{i+1} = \boldsymbol{G}\left(\boldsymbol{U}^i\right)$$

converges at a rate given by the estimator

$$\left\|\boldsymbol{U}^i - \boldsymbol{U}_*\right\| \leq \frac{\alpha^i}{1-\alpha}\left\|\boldsymbol{U}^0 - \boldsymbol{U}^1\right\|$$

In particular, if the Jacobian of $\boldsymbol{G}$ is bounded we can take

$$\alpha = \|\boldsymbol{G}'(\boldsymbol{U})\|$$

and using 7.8 we have (for a fixed $\boldsymbol{D}$)

$$\boldsymbol{G}'(\boldsymbol{U}) = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{H}'(\boldsymbol{U})$$

The Newton method is based on a choice that makes $\boldsymbol{G}$ highly contractive but only in some neighborhood of the solution, which is the reason why it requires a good initial condition. If a temporal evolution is used to solve the problem, we have that $\boldsymbol{D}^{-1} = \varepsilon\boldsymbol{M}^{-1}$, which shows that, when $\varepsilon \to 0$, $\|\boldsymbol{G}'(\boldsymbol{U})\| \to 1$ making the iterative procedure very slow. Note that if

$$\boldsymbol{D} = \frac{1}{\varepsilon}\boldsymbol{M} + \boldsymbol{H}'(\boldsymbol{U})$$

when $\varepsilon \to 0$ we have that $\boldsymbol{D}^{-1} \to \varepsilon\boldsymbol{M}^{-1}$ and if $\varepsilon \to \infty$ we have $\boldsymbol{D}^{-1} \to [\boldsymbol{H}'(\boldsymbol{U})]^{-1}$ as in the Newton method.

The problem of the sensitivity of the Newton method with the initial condition can be partially solved using globally convergent methods (methods that converge for almost any initial guess) which can be developed by adding a line search strategy [123, 40, 92]. As a root of $\boldsymbol{H}$ is a minimum of the function

$$f(\boldsymbol{U}) = \frac{1}{2}\boldsymbol{H}(\boldsymbol{U}) \cdot \boldsymbol{H}(\boldsymbol{U}) \tag{7.9}$$

one may be tempted to apply a minimization algorithm to find the solution, but this is not a good idea because there could be a local minimum of f that is not a root of $\boldsymbol{H}$. However, this function is used to find the optimal parameter of advance. The direction of advance $\boldsymbol{P}$ is found solving the linear system

$$\boldsymbol{D}\boldsymbol{P} = -\boldsymbol{H}\left(\boldsymbol{U}^i\right)$$

and the next iterate is taken as

$$\boldsymbol{U}^{i+1} = \boldsymbol{U}^i + s\boldsymbol{P}$$

where s is the advancing parameter whose calculation is as follows. The step is accepted if the function f decreases at least a small fraction of the decrease given by a linear approximation at $s = 0$. This condition, known as Armijo rule, can be written as

$$f\left(\boldsymbol{U}^{k+1}\right) \leq f\left(\boldsymbol{U}^k\right) + \xi \boldsymbol{\nabla} f \cdot \left(\boldsymbol{U}^{k+1} - \boldsymbol{U}^k\right)$$

where ξ is a parameter of the method taken to be 10^{-4}, and prevents the algorithm to find a local minimum of f. This criterion is applied when a Newton type linearization of the problem is used because in this case $\boldsymbol{D} = \boldsymbol{H}'(\boldsymbol{U})$ and then

$$\boldsymbol{\nabla} f \cdot \boldsymbol{P} = \left[\boldsymbol{H}(\boldsymbol{U}) \cdot \boldsymbol{H}'(\boldsymbol{U})\right] \cdot \boldsymbol{P} = \left[\boldsymbol{H}(\boldsymbol{U}) \cdot \boldsymbol{H}'(\boldsymbol{U})\right] \cdot \left[-\boldsymbol{D}^{-1}\boldsymbol{H}(\boldsymbol{U})\right] = -\boldsymbol{H}(\boldsymbol{U}) \cdot \boldsymbol{H}(\boldsymbol{U})$$

In this case, one first tries $s = 1$ since if we are close to the solution using a Newton type linearization we will have a high rate of convergence (quadratic if the exact Jacobian is used). If the step is not accepted, a new value of s is tested. This value is found using a cubic model based on the values of $f\left(\boldsymbol{U}_h^k + s\boldsymbol{P}\right)$ previously computed [123, 40, 92] but it can be simply taken as a fraction of the previous one. This method can select a step that is too small (this happens when a local minimum of f has been found). In such a case, the method has to be restarted. We do so performing a Picard step, i.e. changing the searching direction.

When a Picard type scheme is used we accept the step when

$$f\left(\boldsymbol{U}^{k+1}\right) \leq f\left(\boldsymbol{U}^k\right)$$

Again we first try $s = 1$ and if the step is not accepted some smaller values of s are tested and the one that gives the minimum value of f is kept.

7.2.2 Linearized equations

In the previous development, matrix $\boldsymbol{D}$ is taken as an approximation to the exact derivative of the function $\boldsymbol{H}$. When we apply this to the flow equations we consider, we always evaluate the stabilization parameters as well as the adjoint operator using the previous iterate, but a full linearization of the operator $\mathcal{L}$ is considered. This linearization

can be written in terms of the linearized advection and reaction matrices, $\boldsymbol{A}_i^{\text{lin}}(\boldsymbol{U}_0)$ and $\boldsymbol{S}^{\text{lin}}(\boldsymbol{U}_0)$, as well as of the resulting forcing vector $\boldsymbol{F}^{\text{lin}}(\boldsymbol{U}_0)$. The expression of these matrices and vector is given below for different flow cases. Here we have explicitly displayed their dependency with respect to the known iterate $\boldsymbol{U}_h^{i-1}$ of $\boldsymbol{U}_h$. Having introduced these terms, the linearized differential operator applied to the finite element unknown is

$$\mathcal{L}^{\text{lin}}(\boldsymbol{U}_0;\boldsymbol{U}) := \boldsymbol{A}_i^{\text{lin}}(\boldsymbol{U}_0)\frac{\partial \boldsymbol{U}}{\partial x_i} - \frac{\partial}{\partial x_i}\left(\boldsymbol{K}_{ij}\frac{\partial \boldsymbol{U}}{\partial x_j}\right) + \boldsymbol{S}^{\text{lin}}(\boldsymbol{U}_0)\boldsymbol{U}$$

The fully discrete stabilized problem is given by 7.7 replacing $\mathcal{L}$ by $\mathcal{L}^{\text{lin}}$, $\boldsymbol{A}_i$ by $\boldsymbol{A}_i^{\text{lin}}$,$\boldsymbol{S}$ by $\boldsymbol{S}^{\text{lin}}$ and $\boldsymbol{F}$ by $\boldsymbol{F}^{\text{lin}}$. In this system it is understood that the stabilization parameters in matrix $\boldsymbol{\tau}$ and matrices $\boldsymbol{M}$, $\boldsymbol{A}_i^{\text{lin}}$, $\boldsymbol{S}^{\text{lin}}$ and $\boldsymbol{F}^{\text{lin}}$ are calculated using $\boldsymbol{U}_0 = \boldsymbol{U}_h^{n+\theta,i-1} + \widetilde{\boldsymbol{U}}^{n+\theta,i-1}$. After the discrete problem 7.7 is solved the subscale $\widetilde{\boldsymbol{U}}^{n+\theta,i}$ is computed and stored. Note that the subgrid problem 7.6 is also nonlinear and has to be iterated (at each point).

It remains to give the expression for $\boldsymbol{A}_i^{\text{lin}}$, $\boldsymbol{S}^{\text{lin}}$ and $\boldsymbol{F}^{\text{lin}}$. To this end, let us define a set of parameters λ_{ij} that can take the value 0 or 1. For $i = 1$ we will use them to write the linearized momentum equation, for $i = 2$ the continuity equation and for $i = 3$ the energy equation. The linearized matrices are given for each flow model as follows:

- Navier Stokes equations

$$\boldsymbol{A}_i^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \rho u_i & 0 & \delta_{i1} \\ 0 & \rho u_i & \delta_{i2} \\ \delta_{i1} & \delta_{i2} & 0 \end{bmatrix},$$

$$\boldsymbol{S}^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \lambda_{11}\rho\partial_1 u_1 & \lambda_{11}\rho\partial_2 u_1 & 0 \\ \lambda_{11}\rho\partial_1 u_2 & \lambda_{11}\rho\partial_2 u_2 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{F}^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \lambda_{11}\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}u_1 + g_1 \\ \lambda_{11}\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}u_2 + g_2 \\ 0 \end{bmatrix}.$$

- Boussinesq equations

$$\boldsymbol{A}_i^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \rho u_i & 0 & \delta_{i1} & 0 \\ 0 & \rho u_i & \delta_{i2} & 0 \\ \delta_{i1} & \delta_{i2} & 0 & 0 \\ 0 & 0 & 0 & \rho u_i \end{bmatrix},$$

$$\boldsymbol{S}^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \lambda_{11}\rho\partial_1 u_1 & \lambda_{11}\rho\partial_2 u_1 & 0 & \lambda_{13}\rho\beta g_1 \\ \lambda_{11}\rho\partial_1 u_2 & \lambda_{11}\rho\partial_2 u_2 & 0 & \lambda_{13}\rho\beta g_2 \\ 0 & 0 & 0 & 0 \\ \lambda_{31}\rho\partial_1\vartheta & \lambda_{31}\rho\partial_2\vartheta & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{F}^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \lambda_{11}\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}u_1 - (1-\lambda_{13})\rho\beta g_1\vartheta + \rho\beta\vartheta_0 g_1 \\ \lambda_{11}\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}u_2 - (1-\lambda_{13})\rho\beta g_1\vartheta + \rho\beta\vartheta_0 g_2 \\ 0 \\ \lambda_{31}\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta + Q \end{bmatrix}.$$

- Low Mach number equations

$$\boldsymbol{A}_i^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \rho u_i & 0 & \delta_{i1} & 0 \\ 0 & \rho u_i & \delta_{i2} & 0 \\ \delta_{i1} & \delta_{i2} & 0 & -\frac{\lambda_{21}}{\vartheta}u_i \\ 0 & 0 & 0 & \rho u_i \end{bmatrix},$$

$$\boldsymbol{S}^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} \lambda_{11}\rho\partial_1 u_1 & \lambda_{11}\rho\partial_2 u_1 & 0 & \frac{\rho}{\vartheta}\left(-\lambda_{12}\boldsymbol{u}\cdot\boldsymbol{\nabla}u_1 + \lambda_{13}g_1\right) \\ \lambda_{11}\rho\partial_1 u_2 & \lambda_{11}\rho\partial_2 u_2 & 0 & \frac{\rho}{\vartheta}\left(-\lambda_{12}\boldsymbol{u}\cdot\boldsymbol{\nabla}u_2 + \lambda_{13}g_2\right) \\ -\frac{\lambda_{22}}{\vartheta}\partial_1\vartheta & -\frac{\lambda_{22}}{\vartheta}\partial_2\vartheta & 0 & \frac{\lambda_{23}}{\vartheta^2}\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta \\ \lambda_{31}\rho\partial_1\vartheta & \lambda_{31}\rho\partial_2\vartheta & 0 & -\frac{\rho}{\theta}\lambda_{32}\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta \end{bmatrix},$$

$$\boldsymbol{F}^{\text{lin}}(\boldsymbol{U}) = \begin{bmatrix} (\lambda_{11}-\lambda_{12})\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}u_1 + (1+\lambda_{13})\rho g_1 \\ (\lambda_{11}-\lambda_{12})\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}u_2 + (1+\lambda_{13})\rho g_2 \\ (1-\lambda_{21}-\lambda_{22}+\lambda_{23})\frac{1}{\vartheta}\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta \\ (\lambda_{31}-\lambda_{32})\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta + Q \end{bmatrix}.$$

The parameters λ_{11} and λ_{12} correspond to the linearization of the convective term in the momentum equation ($\lambda_{11} = \lambda_{12} = 1$ would be Newton's method, whereas other options would be fixed point methods), whereas λ_{13} is used to decide whether the buoyancy term is treated in a coupled or in a block iterative way. Likewise, $\lambda_{2j}, j = 1, 2, 3$, determine both the linearization of the term $\frac{1}{\vartheta}\boldsymbol{u}\cdot\boldsymbol{\nabla}\vartheta$ ($\lambda_{2j} = 1$ would be full Newton's method) and the possibility to treat this term in a staggered way ($\lambda_{2j} = 0$). Finally, $\lambda_{3j}, j = 1, 2$, play the same role for the energy equation as $\lambda_{1j}, j = 1, 2$, for the momentum equation.

7.3 Numerical examples

In this section we present two examples. This first one is the natural convection in a two dimensional closed cavity and, as it is a well known benchmark for thermally coupled flows, we use it to test different numerical strategies proposed here. The second one is a two dimensional time dependent heated channel presented in [107] as a simplified version of what occurs in a chemical vapor deposition (CVD) reactor. CVD flow problems, reviewed in [90], present much of the physics of the Poiseuille-Rayleigh-Bénard (PRB) flow problem reviewed in [114], that consists of a channel with a prescribed Poiseuille velocity profile on the inlet and prescribed temperatures on the upper and lower walls. This example is included to illustrate the models considered as well as to point out the importance of outflow boundary conditions.

7.3.1 Natural convection in a cavity

The natural convection in a cavity is a standard benchmark for numerical methods on thermally coupled flows. It was initially devised for Boussinesq flows [38] and later for low Mach number flows [103]. The problem is sketched in Figure 7.1.

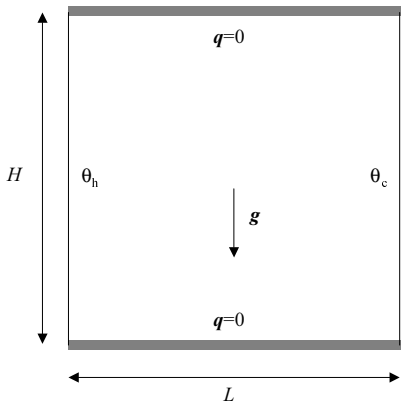

Figure 7.1: Geometry and boundary conditions of the natural convection in a cavity

First of all, let us mention the conditions for the validity of the approximations in this example. As this is a natural convection problem, a velocity scale must be chosen. Taking for example the viscous scale and using the benchmark specifications (see [103]) gives a Mach number of 2.2×10^{-5}, allowing the use of the zero Mach number equations. The conditions of applicability of the Boussinesq approximation need some care. In this case, the zero order temperature and density must be constants. In order to have this reference state, the (dimensionless) temperature difference between vertical walls must vanish. Finally, the Boussinesq number must tend to zero as fast as the Mach number (which is a restriction of the vertical scale of the problem). In the conditions of the benchmark, the Boussinesq number is 5.7×10^{-5} and is of the same order as the Mach number. Thus, the dimensionless parameters that define the problem are

$$\varepsilon = \frac{\vartheta_h - \vartheta_c}{\vartheta_h + \vartheta_c}, \quad \mathrm{A} = \frac{H}{L}$$
$$\mathrm{Pr} = \frac{c_p \mu}{k}, \quad \mathrm{Ra} = \mathrm{Pr}\frac{gL^3}{\nu^2}\varepsilon$$

where Pr is the Prandtl number and Ra is the Rayleigh number.

Let us first present results that show the physical behavior of the problem (see [23] for a full description of the physics of the problem). They have been obtained using a fine grid of 160×160 Q1 (bilinear) elements refined towards the walls. The steady state problem has been directly solved (without time advancing) with a convergence tolerance for the nonlinear process of 10^{-8}.

Figure 7.2 shows the streamline and temperature distribution obtained using the Boussinesq approximation for different Rayleigh numbers. For a Rayleigh number of 10^3 there is only one vortex that covers the whole domain. When the Rayleigh number is increased this vortex splits first in two and then the vortex distribution becomes more complex and the boundary layers on hot and cold walls become thinner.

When the low Mach number approximation is used similar results are obtained, but some differences are found. For a fixed Rayleigh number of 10^3, when the temperature increases, the central vortex moves to the right. This effect can be seen in table 7.2, where the position of the center of the vortex as a function of temperature difference is presented. For higher Rayleigh numbers the effect is similar: the flow is qualitatively similar although some differences appear when quantifying magnitudes.

ε	x coord.
0.0	0.50
0.2	0.54
0.4	0.58
0.6	0.63

Table 7.2: Evolution of the x-coordinate of the central vortex for $Ra = 10^3$ in terms of ϵ

An important difference between the Boussinesq and the low Mach number approximations is that the latter can describe phenomena related to the expansion of the flow. If a gas in a closed cavity is heated, basic thermodynamics implies that the pressure level should increase and this cannot be predicted using the Boussinesq approximation. In the case of the differentially heated cavity at $\varepsilon = 0.6$, the mean thermodynamic pressure normalized using the initial pressure is 0.856. This case was considered to test the mesh convergence of the proposed algorithm using graded meshes of 10×10, 20×20, 40×40 and 80×80 Q1 elements. Table 7.3 presents the thermodynamic pressure as a function of the mesh size. It is seen that the behavior is as expected and the results agree with those found in the literature. It is to be noted that the results presented in [72] correspond to a discretization with 855 556 degrees of freedom obtained by an adaptive procedure.

h	Ra $= 10^3$	Ra $= 10^4$	Ra $= 10^5$	Ra $= 10^6$
0.1000	0.8646791	0.8585644	0.8681438	-
0.0500	0.8603283	0.8497187	0.8567441	0.8670059
0.0250	0.8582884	0.8460002	0.8534048	0.8579742
0.0125	0.8574004	0.8445817	0.8524585	0.8567541
Reference[72]	-	-	-	0.856337

Table 7.3: h convergence of the thermodynamic pressure.

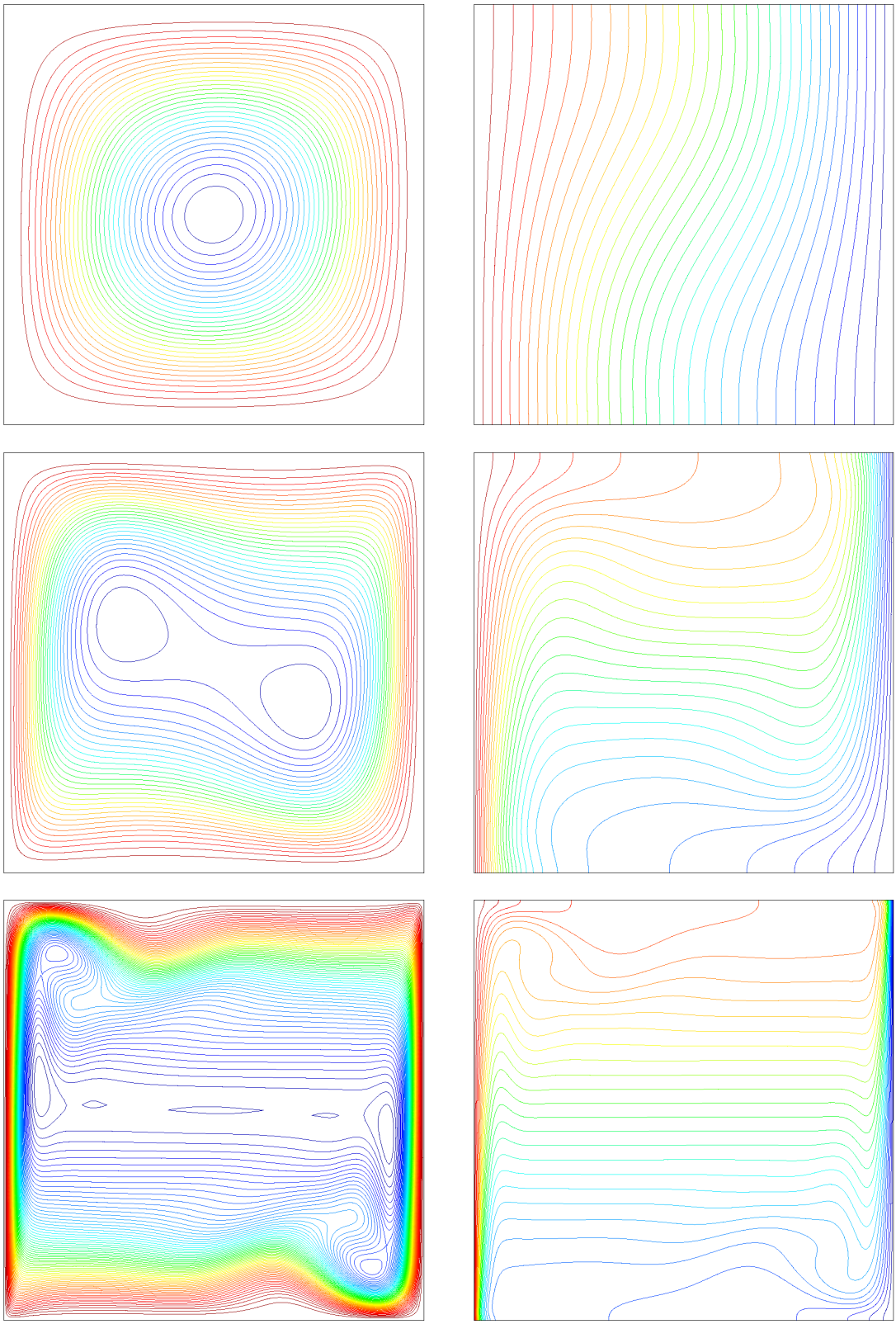

Figure 7.2: Streamline (left) and temperature (right) distribution obtained using the Boussinesq approximation at $\mathrm{Ra} = 10^3$ (top), $\mathrm{Ra} = 10^5$ (middle) and $\mathrm{Ra} = 10^7$ (bottom).

Let us now describe the non linear convergence of the iterative scheme when a direct steady state calculation is performed. A set of experiments were performed using the Boussinesq model for the different possible linearizations on a uniform mesh of 10×10 Q1 elements. Figure 7.3 shows the convergence of the algorithm for different Rayleigh numbers.

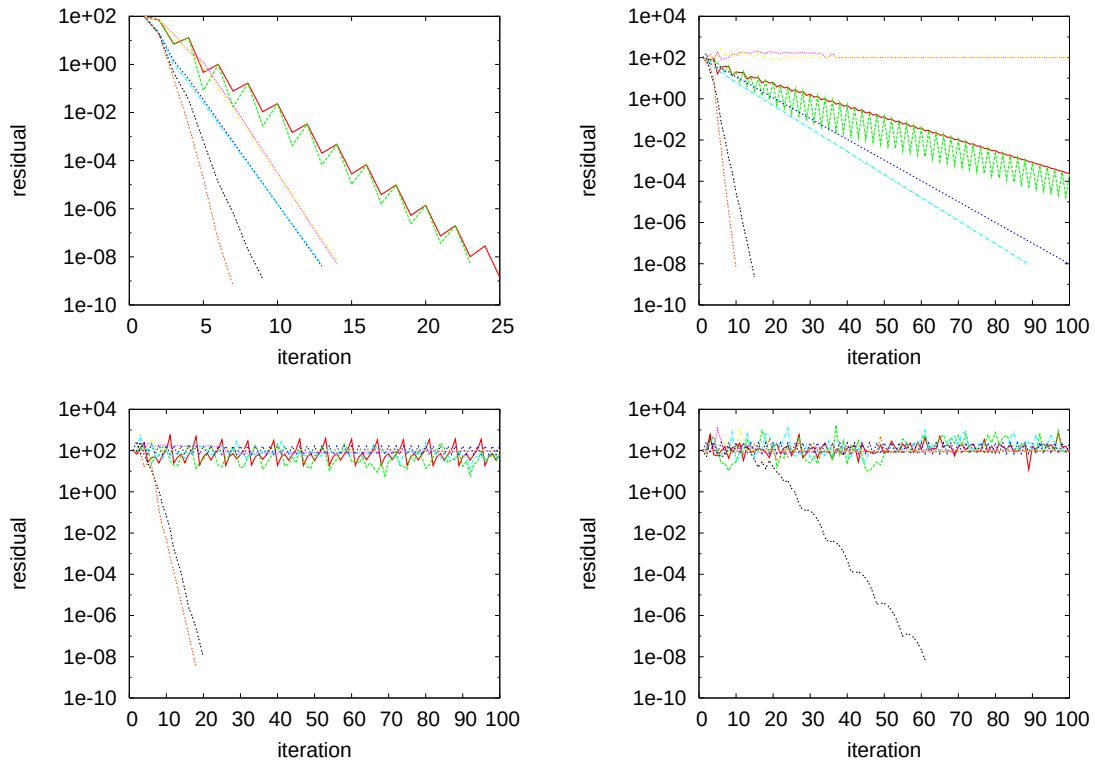

Figure 7.3: Non linear convergence for the 10×10 uniform mesh. From left to right and top to bottom $Ra = 10^3$, 10^4, 10^5 and 10^6. The curves corresponding to the linearization parameters λ_1, λ_2 and λ_3 are given as follows

From these experiments we conclude that the Newton scheme ($\lambda_{1,1} = 1$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 1$) is fastest, as expected, and the linearization $\lambda_{1,1} = 0$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 0$ is the most robust for this example (in the sense that it converges for higher values of the Rayleigh number). When the buoyancy term is treated in a coupled way (taking $\lambda_{1,3} = 1$) the convergence becomes monotone for $\mathrm{Ra} = 10^3$ and $\mathrm{Ra} = 10^4$ but it is also seen that what makes a big difference is to combine this treatment with a full Newton linearization of the convective term in the temperature equation (that is to take also $\lambda_{3,1} = 1$). Let us stress that although a Picard type linearization could be more robust when incompressible Navier Stokes are considered (which is also observed here when comparing $\lambda_{1,1} = 0$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 0$ and $\lambda_{1,1} = 1$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 1$), a Newton type linearization of the

velocity to the temperature coupling (in this case through the convective term in the temperature equation) is more robust.

Next we performed a set of experiments to test the behavior of the line search process described in section 7.2. For the full Newton linearization we performed computations for uniform meshes of 10×10, 20×20, 40×40 and 80×80 Q1 elements without a line search and using the Armijo rule described in section 7.2 (the experiments using the Armijo rule were also run on a uniform mesh of 160×160 Q1 elements). Tables 7.4 and 7.5 show the behavior of the iterative scheme by indicating the number of iterations needed when convergence is achieved.

elem.	$\mathrm{Ra} = 10^3$	$\mathrm{Ra} = 10^4$	$\mathrm{Ra} = 10^5$	$\mathrm{Ra} = 10^6$	$\mathrm{Ra} = 10^7$	$\mathrm{Ra} = 10^8$
10	7	10	18	-	-	-
20	6	9	14	20	-	-
40	6	9	14	-	-	-
80	6	8	13	-	-	-

Table 7.4: Number of iterations for the linearization $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 1$, without line search. The dash indicates divergence

elem.	$\mathrm{Ra} = 10^3$	$\mathrm{Ra} = 10^4$	$\mathrm{Ra} = 10^5$	$\mathrm{Ra} = 10^6$	$\mathrm{Ra} = 10^7$	$\mathrm{Ra} = 10^8$
10	8	13	17	22	96	64
20	8	12	13	17	34	162
40	8	12	13	14	17	*
80	9	12	13	14	16	24
160	7	13	14	16	17	26

Table 7.5: Number of iterations for the linearization $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 1$, using the Armijo rule. The star indicates lack of convergence

Very similar results are obtained for the linearization that corresponds to $\lambda_{1,1} = 0$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 1$. The conclusion to be drawn is that the use of the line search greatly improves the robustness of the iterative scheme. It is also to be mentioned that, although when a line search is not performed the linearization that corresponds to $\lambda_{1,1} = 0$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 1$ converges for some cases where the full Newton does not, when using the Armijo rule both linearizations behave identically. Moreover, the calculations where the scheme fails to converge are on coarse meshes and convergence is achieved for finer meshes (especially when refined meshes are used). The linearization scheme for the low Mach number model has also been tested in detail and the same behavior was observed: the full Newton scheme together with the Armijo rule converges for almost any case. In [107] difficulties to obtain convergence when performing calculations using the low Mach

number model have been reported even for a low Rayleigh number ($\mathrm{Ra} = 10^4$). In this reference, an ad hoc linearization of the system was performed to overcome this problem. We did not find this problems for low Rayleigh number. In [107] the problem is solved using a mixed finite element formulation, what could be the reason behind the difference in the behavior of the iterative algorithms.

We next consider the behavior of the line search process when a time dependent calculation is performed. A set of experiments were performed using the full Newton linearization and the Boussinesq model on a uniform mesh of 10×10 Q1 elements with and without line search. The number of iterations needed are shown in tables 7.6 and 7.7.

δt	$\mathrm{Ra} = 10^3$	$\mathrm{Ra} = 10^4$	$\mathrm{Ra} = 10^5$	$\mathrm{Ra} = 10^6$	$\mathrm{Ra} = 10^7$	$\mathrm{Ra} = 10^8$
10^0	7	10	18	35	-	-
10^{-1}	6	9	16	22	-	-
10^{-2}	5	8	12	20	-	-
10^{-3}	5	6	11	23	33	-
10^{-4}	4	5	7	13	53	204

Table 7.6: Number of iterations at the first time step for the linearization $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 1$, without line search. The dash indicates divergence

δt	$\mathrm{Ra} = 10^3$	$\mathrm{Ra} = 10^4$	$\mathrm{Ra} = 10^5$	$\mathrm{Ra} = 10^6$	$\mathrm{Ra} = 10^7$	$\mathrm{Ra} = 10^8$
10^0	7	13	17	22	118	*
10^{-1}	6	12	16	22	*	*
10^{-2}	5	8	13	22	45	*
10^{-3}	5	6	11	23	40	60
10^{-4}	4	5	7	13	53	52

Table 7.7: Number of iterations at the first time step for the linearization $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 1$, using the Armijo rule. The star indicates lack of convergence

As it could be expected, and these experiments confirm, less nonlinear iterations are required to achieve convergence when the time step is reduced. These experiments also show that still the line search algorithm is important when time steps are big and the non linearity is important. When the time step is reduced convergence is achieved without the need of relaxation.

We next consider the non linear convergence of the iterative scheme when a time dependent calculation is performed. A set of experiments were performed using the linearization that corresponds to ($\lambda_{1,1} = 0$, $\lambda_{1,3} = 0$, $\lambda_{3,1} = 0$) and the Boussinesq model on a uniform mesh of 10×10 Q1 elements.

δt	$\mathrm{Ra} = 10^3$	$\mathrm{Ra} = 10^4$	$\mathrm{Ra} = 10^5$	$\mathrm{Ra} = 10^6$	$\mathrm{Ra} = 10^7$	$\mathrm{Ra} = 10^8$
10^0	23	166	-	-	-	-
10^{-1}	17	85	-	-	-	-
10^{-2}	9	14	39	*	200	-
10^{-3}	6	8	13	32	*	*
10^{-4}	5	6	7	13	48	*

Table 7.8: Number of iterations at the first time step for the linearization $\lambda_1 = 0$, $\lambda_2 = 0$, $\lambda_3 = 0$, without line search. The dash indicates lack of convergence and the star lack of convergence on the first steps.

The main conclusion of these experiments is that when the time step is small, the number of iterations required by a Picard type linearization ($\lambda_{1,1} = 0$, $\lambda_{1,3} = 0$, $\lambda_{3,1} = 0$) or those required by a Newton type one ($\lambda_{1,1} = 1$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 1$) are similar. In this cases the Picard type linearization is preferred because it allows a splitting of the algebraic problem into a mechanical problem and a thermal problem. Note that this is interesting because it permits to modify an incompressible code to take thermal coupling into account and because it permits to reduce memory requirements storing smaller matrices (but note also that the time required to solve the linear system will be the same). This leads to three types of iterative coupling: the one that corresponds to $\lambda_{1,1} = 0$, $\lambda_{1,3} = 0$, $\lambda_{3,1} = 0$, that permits the parallel solution of the mechanical and the thermal problem, and those that correspond to $\lambda_{1,1} = 0$, $\lambda_{1,3} = 0$, $\lambda_{3,1} = 1$ and $\lambda_{1,1} = 0$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 0$ which result in a Gauss Seidel type scheme. The difference between the last two of them is which problem is solved first. It can be observed in figure 7.3 that the linearization that corresponds to $\lambda_{1,1} = 0$, $\lambda_{1,3} = 1$, $\lambda_{3,1} = 0$, that is the one in which the thermal problem is solved first, is more robust than the one that corresponds to $\lambda_{1,1} = 0$, $\lambda_{1,3} = 0$, $\lambda_{3,1} = 1$, that is the one in which the mechanical problem is solved first. Respect to this point let us also mention that in the case of the low Mach number system, it is quite important to solve for the temperature first because of the term $\partial_t \rho$ in the continuity equation. If the mechanical problem is solved first, in the first iteration of the time step, the approximation to $\partial_t \rho$ is zero as the initial guess is the temperature at the previous step and this needs to be corrected by the iterative coupling. This effect has been observed while solving the problem presented in the following section.

7.3.2 Time dependent heated channel

The problem studied here, sketched in Figure 7.4, is a channel whose length (L) is 5 times its height (H).

The inlet boundary conditions are given by a Poiseuille velocity profile and uniform

δt	$\mathrm{Ra} = 10^3$	$\mathrm{Ra} = 10^4$	$\mathrm{Ra} = 10^5$	$\mathrm{Ra} = 10^6$	$\mathrm{Ra} = 10^7$	$\mathrm{Ra} = 10^8$
10^0	14	132	-	-	-	-
10^{-1}	11	47	-	-	-	-
10^{-2}	8	11	25	-	-	-
10^{-3}	7	8	12	33	-	-
10^{-4}	5	5	7	13	79	*

Table 7.9: Number of iterations at the first time step for the linearization $\lambda_1 = 0$, $\lambda_2 = 0$, $\lambda_3 = 1$, without line search. The dash indicates lack of convergence and the star lack of convergence on the first steps.

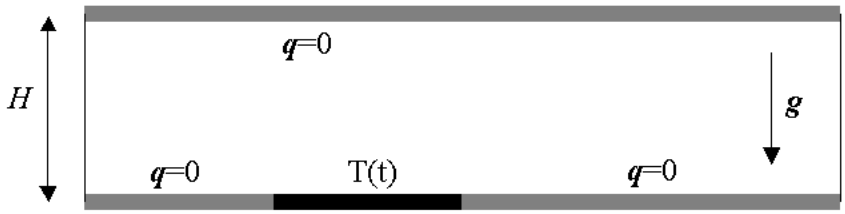

Figure 7.4: Geometry and boundary conditions of the time dependent heated channel (not to scale)

temperature. A non slip condition is prescribed on the upper and lower walls. Zero heat flux is prescribed on the upper wall and on the part of the lower wall where temperature is not prescribed, as indicated in Figure 7.4. A time dependent temperature is prescribed on a part of the lower wall of length $H/2$ located at a distance $H/2$ from the inlet. The prescribed (dimensionless) temperature rises from 1 at time $t = 0$ to 1.5 at time $t = 0.01$ and remains constant after that. The dimensionless parameters of the problem are

$$\mathrm{R} = \frac{\rho U L}{\mu} = 10, \quad \mathrm{Pr} = \frac{c_p \mu}{k} = 1$$
$$\varepsilon = \frac{\Delta T}{T_0} = 0.5, \quad \mathrm{Ra} = \mathrm{Pr}\frac{gL^3}{\nu^2}\varepsilon = 5 \times 10^4$$

The initial conditions are a Poiseuille velocity profile and a constant temperature on the whole domain. When the flow starts to heat (near the zone where the temperature is imposed) it goes up by buoyancy forces giving rise to two vortices, one before and the other after the heating zone. This is illustrated in Figure 7.5, where streamlines and temperature distribution obtained using the Boussinesq approximation are presented for different (early) times.

After this initial transient, the flow after the heating zone gradually rises its temperature and the second vortex (the one located after the heating zone) gradually

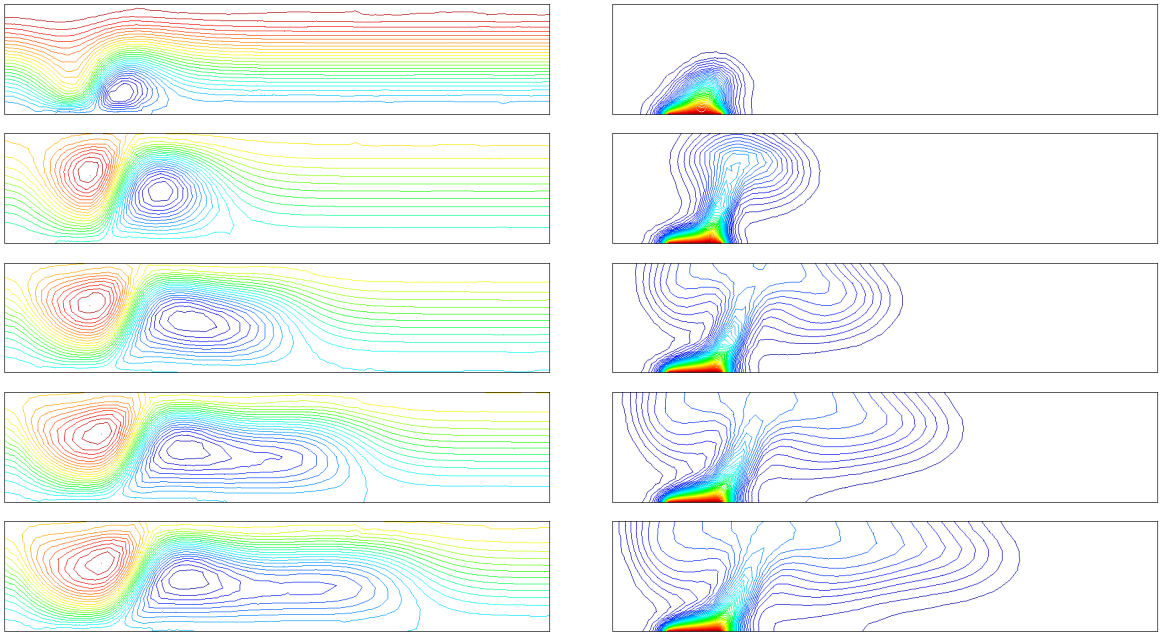

Figure 7.5: Streamlines and temperature distribution obtained using the Boussinesq approximation at times $t = 0.2$, $t = 0.4$, $t = 0.6$, $t = 0.8$ and $t = 1.0$

disappears. The final steady state is reached around $t = 12$. This behavior is also observed when the low Mach number approximation is used.

The first point we want to illustrate here is the influence of the output boundary condition. One possibility is to consider simply

$$\boldsymbol{t} \cdot \boldsymbol{n} = 0 \tag{7.10}$$

but it seems to be better to consider an "atmospheric stress condition" as the one suggested in [107], given by

$$\boldsymbol{t} \cdot \boldsymbol{n} = t_x = -\rho \, |\boldsymbol{g}| \, y \tag{7.11}$$

The final steady state obtained using these conditions is shown in figure 7.6.

7.4 Conclusions

We have implemented the discrete approximation to problems defined in chapter 1 as systems of second order equations. Different linearizations have been considered for the solution of the algebraic nonlinear problem. Using the well known example of the flow in a differentially heated cavity, we have shown that fully coupled schemes, in which a Newton type linearization is performed, are the best option for the solution of stationary problems or when the time step is large. In this cases, a line search strategy is very important to enhance the robustness of the scheme. Its cost is higher to that of forming the system of

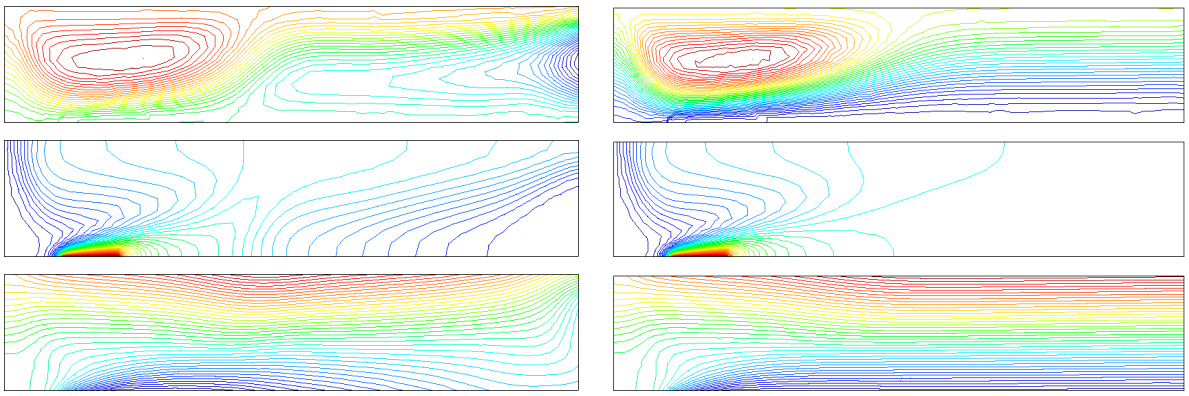

Figure 7.6: Steady state streamlines, temperature and pressure distributions obtained using boundary conditions 7.10 on the left and using boundary conditions 7.11 on the righ

equations and for this reason it is not convenient when small time steps are considered. In this case there is also a small difference between the full Newton linearization and the staggered approach in which the problem is split into a mechanical problem and a thermal one. It has also been shown that in this case it is better to solve the thermal problem first. A fractional step scheme, splitting also momentum and continuity equations, could be considered in this case, but this is a point that needs further research.

The zero Mach number model and the Boussinesq approximation have been considered to solve low speed thermally coupled flows. Both describe the basic mechanism of thermal coupling which is due to the dependence of the density on the temperature: when a fluid element is heated, it expands and moves up. However, they differ in the way they take into account the compressibility of the medium. While in the Boussinesq approximation the flow is incompressible, in the zero Mach number model the density distribution is predicted and the velocity field is affected by expansions or contractions due to heating. They also have different ranges of applicability: while the low Mach number approximation only requires a small Mach number, the Boussinesq model requires also a small Froude number and small heat sources.

Chapter 8

Thermal coupling of fluids and solids

In this chapter we analyze the problem of the thermal coupling of fluids and solids through a common interface. We state the global thermal problem in the whole domain, including the fluid part and the solid part. This global thermal problem presents discontinuous physical properties that depend on the solution of auxiliary problems on each part of the domain (a fluid flow problem and a solid state problem). We present a domain decomposition strategy to iteratively solve problems posed in both subdomains and discuss some implementation aspects of the algorithm. This domain decomposition framework is also used to revisit the use of wall function approaches used in this context.

8.1 Introduction

The problem we analyze in this chapter is that of the thermal coupling of fluids and solids. This problem is found in any engineering design in which a fluid is used to extract heat from a solid (refrigeration, ventilation, etc.). In fact many experimental correlations are available [86] in the form of convection coefficients. The objective of the present chapter is to present a domain decomposition approach that permits the separated treatment of a problem in the solid domain and of a problem in the fluid one. Let us emphasize that it is not our intention to use a domain decomposition strategy to perform parallel computations but to treat problems with different physics separately. Moreover, this domain decomposition approach will allow us to implement the thermal coupling problem in a master-slave algorithm (discussed below).

The model presented in section 8.2 is based on the solution of the thermal problem in the whole domain Ω that includes the solid subdomain Ω_{S} and the fluid subdomain Ω_{F}. The differential operators that describe the evolution of the temperature (ϑ) are different as a result of different physics and they depend on other variables that describe the state of each medium. On each subdomain the thermal problem could be coupled to other differential problems depending on the physical model used for the fluid and for

the solid. In the first case we may have a compressible flow or an incompressible one, a mix of species, chemical reactions, etc. In the second case we may have a purely thermal problem, a thermomechanical one or even a thermo-hygro-mechanical one as in [57]. Any model can be used on each subdomain but we will assume that the coupling between the fluid and solid is only due to heat exchange. This condition, in the case in which mechanical problems are solved on each subdomain, will be written explicitly indicating precisely the assumptions needed.

The numerical approximation of the fluid and solid problems is in general different. One important feature of our approach is that different numerical approximations could be used to solve each problem. In the case of the fluid we use a stabilized finite element formulation based on the subgrid scale concept. Each field is decomposed into a resolvable and a subgrid scale part according to the finite element partition, and the effect of the subgrid scale on the coarse scale is taken into account by an algebraic approximation. This approach allows us to deal with convection dominated problems and to use equal order interpolation of velocity and pressure which would lead to numerical instabilities when a standard Galerkin formulation is used. The Galerkin approximation is used to solve the solid problem. We describe the discrete formulation in section 8.3 but we emphasize once again that any other possibility could be used.

The possibility of using different models and different discrete approximations is not only theoretical but also practical. The coupling through the common interface between the solid and the fluid is accomplished by the transmission conditions, which we consider of Dirichlet/Neumann type. This leads to a non-overlapping domain decomposition problem that we implement in an iteration-by-subdomain strategy. The solution of each thermal problem, in the fluid and in the solid regions, and the transmission of boundary conditions from one domain to the other is done by a relatively small master code. This code, developed following the MPI 2 standard, is in charge of managing the subdomain iterative coupling and the time marching loops. In this way, each dedicated code acts as a slave and can be updated separately as only minor modifications are needed to change the information with the master code. These implementation aspects are described in section 8.4. The domain decomposition framework for the thermal coupling described, together with its implementation aspects and an interpretation of the use of wall function approaches, is the main contribution of this work.

Finally, the approach is illustrated in a simple one dimensional example and is applied to the simulation of a fire in a tunnel in section 8.5. Some conclusions are drawn in section 8.6.

8.2 Continuous problem

We consider a thermal problem in a domain Ω composed of two subdomains Ω_{S} and Ω_{F}, as illustrated in Figure 8.1 (left and center). We present first the problem in the whole domain considering discontinuous physical properties, which include the density (ρ), the specific heat (c_p) and the diffusion coefficient ($\kappa = \frac{k}{\rho c_p}$, where k is the thermal conductivity), as well as a velocity field ($\boldsymbol{v}$). This velocity field will be assumed to be solution of a mechanical problem defined also in the whole domain and having also discontinuous properties. The constitutive relations in the fluid and in the solid are different, the former relating the stress tensor ($\boldsymbol{\sigma}$) to the velocity gradients and the latter relating the stress tensor to the deformation gradient.

Once the problem in the whole domain has been written, we will present two different strategies for a domain decomposition approach to this problem. The first strategy presented consists of a standard non-overlapping domain decomposition of the problem into the fluid and solid subdomains. We will assume that the mechanical problem in the solid does not depend on that in the fluid, in a sense to be made precise later on. We will refer to this approach as *the full resolution strategy*.

The second strategy consists of a non-overlapping domain decomposition of the problem in three subdomains, one in the solid region and two in the fluid region, as illustrated in Figure 8.1 (right). One of the fluid subdomains will be a thin region of thickness δ near the solid surface and the other will be the rest of the fluid domain. The purpose of this second approach is to consider the problem of the strong boundary layers present in a turbulent flow using the wall function approach. An approximated solution of the problem in this thin region is written in terms of the wall function and an iteration strategy between the remaining subdomains is proposed. Therefore this second approach will also involve two subdomains. We again assume that the mechanical problems are uncoupled. We will refer to this approach as the *wall function strategy*.

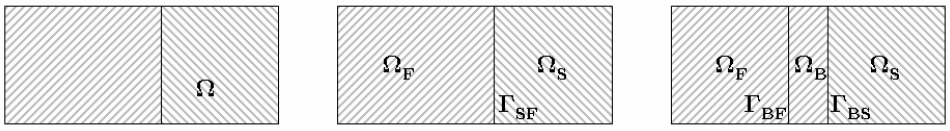

Figure 8.1: Domain of the problem

8.2.1 Problem definition in the whole domain

Strong form of the problem

The problem to be solved in Ω, an open domain in $\mathbb{R}^{\mathrm{d}}$ ($\mathrm{d} = 2, 3$ is the number of space dimensions) during the time interval $(0, t_{\mathrm{f}})$ is described by the equations of continuous

media. These equations could also be used as basis for more complex models [57]. These sets always contain an energy conservation statement that, under suitable assumptions, reads

$$\rho c_p \left(\partial_t \vartheta + \boldsymbol{v} \cdot \boldsymbol{\nabla} \vartheta\right) + \boldsymbol{\nabla} \cdot \boldsymbol{q} = Q \quad \text{in} \quad \Omega \times (0, t_\mathrm{f}) \tag{8.1}$$

Here Q is the external source of energy source per unit of mass and $\boldsymbol{q}$ is the internal heat flux vector. The velocity field $\boldsymbol{v}$ is the solution of a mechanical problem of the form

$$\rho \left(\partial_t \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{\nabla} \boldsymbol{v}\right) - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma} = \rho \boldsymbol{g} \quad \text{in} \quad \Omega \times (0, t_\mathrm{f})$$

Here $\boldsymbol{g}$ is the external source of momentum per unit of mass and σ the internal stress tensor. The parameters present in these equations (ρ and c_p) may be discontinuous across the surface $\Gamma_{\mathrm{SF}}(t)$, which is a moving surface separating the fluid and the solid. The constitutive equation for the internal heat flux is

$$\boldsymbol{q} = -k \boldsymbol{\nabla} \vartheta \tag{8.2}$$

where k may also be discontinuous across $\Gamma_{\mathrm{SF}}(t)$. The constitutive equation for the internal stress tensor will in general be different in both regions. In the case of the solid it will be related to the deformation tensor. In the case of an incompressible fluid it will be related to the velocity gradient and to another variable, the pressure (p), that will involve the solution of another equation, the conservation of mass.

The problem must be supplemented with appropriate boundary and initial conditions. Let us split the boundary $\partial\Omega$ of Ω into two parts as $\partial\Omega = \Gamma_{\mathrm{D}}^{\vartheta} \cup \Gamma_{\mathrm{N}}^{\vartheta}$, where $\Gamma_{\mathrm{D}}^{\vartheta}$ and $\Gamma_{\mathrm{N}}^{\vartheta}$ represent the part of the boundary where Dirichlet and Neumann boundary conditions for the temperature are prescribed, respectively (see Figure 8.2).

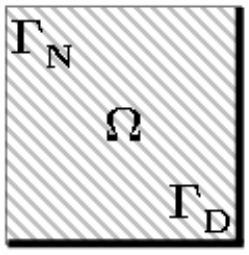

Figure 8.2: Boundary conditions of the problem

The initial and boundary conditions for the thermal problem can thus be written as

$$\begin{aligned} \vartheta &= \vartheta_0 \quad \text{in} \quad \Omega \times \{0\} \\ \vartheta &= \vartheta_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}^{\vartheta} \times (0, t_\mathrm{f}) \\ \boldsymbol{n} \cdot \boldsymbol{q} &= q_{\mathrm{N}} \quad \text{on} \quad \Gamma_{\mathrm{N}}^{\vartheta} \times (0, t_\mathrm{f}) \end{aligned}$$

where $\boldsymbol{n}$ is the normal exterior to the domain Ω. In turn, the initial and boundary conditions for the mechanical problem depend on the constitutive relation considered. In

the case of a solid they are usually written in terms of the displacement and in the case of the fluid in terms of the velocity. The purpose of writing the mechanical problem is to clearly specify the conditions under which it is uncoupled. Nevertheless, we will consider the weak form and the numerical approximation of the thermal problem only.

Weak form of the problem

In order to write the weak form of the problem we will not consider the mechanical problem and we will assume the velocity field to be given. Let us introduce some notation. We start introducing the functional spaces

$$H^1(\Omega) = \left\{ v \in L^2(\Omega) : \frac{\partial v}{\partial x_j} \in L^2(\Omega), j = 1, ..., n_{\text{sd}} \right\}$$

$$V^{\vartheta}(\Omega) = \left\{ v \in H^1(\Omega) : v = \vartheta_{\text{D}} \text{ in } \Gamma_{\text{D}}^{\vartheta} \right\}$$

$$V^0(\Omega) = \left\{ v \in H^1(\Omega) : v = 0 \text{ in } \Gamma_{\text{D}}^{\vartheta} \right\}$$

The scalar product in $L^2(\Omega)$ will be denoted by

$$(u, v)_{\Omega} := \int_{\Omega} uv \, d\Omega$$

and we will use the notation

$$\langle f, g \rangle_{\omega} := \int_{\omega} fg \, d\omega$$

when the functions f and g are not necessarily square integrable and ω is either a subdomain of Ω or part of the boundary $\partial\Omega$.

A weak formulation of the problem is obtained by integrating by parts the diffusive term in Equation 8.1 and using the constitutive Equation 8.2. Let us introduce the bilinear form

$$a(\vartheta, v) := (\rho c_p \partial_t \vartheta, v)_{\Omega} + (\rho c_p \boldsymbol{v} \cdot \boldsymbol{\nabla} \vartheta, v)_{\Omega} + (k \boldsymbol{\nabla} \vartheta, \boldsymbol{\nabla} v)_{\Omega}$$

(in fact, a is affine in the first argument, but we will omit this precision in the following) and the linear form

$$l(v) := \langle Q, v \rangle_{\Omega} + \langle q_{\text{N}}, v \rangle_{\Gamma_{\text{N}}^{\vartheta}}$$

The weak form of the problem consists in finding $\vartheta \in L^2(0, t_{\text{f}}; V^{\vartheta}) \cap L^{\infty}(0, t_{\text{f}}; L^2(\Omega))$ such that

$$a(\vartheta, v) = l(v) \quad \forall v \in V^0 \tag{8.3}$$

where $L^2(0, t_{\text{f}}; V^{\vartheta})$ is the set of functions whose norm in V^{ϑ} (which is the norm in $H^1(\Omega)$) is square integrable in time, and $L^{\infty}(0, t_{\text{f}}; L^2(\Omega))$ the set of functions whose norm in $L^2(\Omega)$ is bounded in time.

8.2.2 The full resolution strategy

Let us split the domain Ω into the solid and fluid subdomains, $\Omega_{\rm S}$ and $\Omega_{\rm F}$, as illustrated in Figure 8.1 (center) and let $\Gamma_{\rm SF}(t)$ be the interface between them. Let us also define

$$\begin{aligned}\Gamma^{\vartheta}_{{\rm D}i} &= \Gamma^{\vartheta}_{\rm D} \cap \partial\Omega_i \quad i = {\rm S}, {\rm F}\\ \Gamma^{\vartheta}_{{\rm N}i} &= \Gamma^{\vartheta}_{\rm N} \cap \partial\Omega_i \quad i = {\rm S}, {\rm F}\end{aligned}$$

Strong form of the problem

The strong form of the problem consists in finding temperatures $\vartheta_{\rm S}$ and $\vartheta_{\rm F}$, as well as velocities $\boldsymbol{v}_{\rm S}$ and $\boldsymbol{v}_{\rm F}$, such that

$$\begin{aligned}\rho_{\rm S} c_{p{\rm S}} \left(\partial_t \vartheta_{\rm S} + \boldsymbol{v}_{\rm S}\cdot\boldsymbol{\nabla}\vartheta_{\rm S}\right) + \boldsymbol{\nabla}\cdot\boldsymbol{q}_{\rm S} &= Q_{\rm S} \quad \text{in} \quad \Omega_{\rm S}\times(0,t_{\rm f})\\ \rho_{\rm S}\left(\partial_t \boldsymbol{v}_{\rm S} + \boldsymbol{v}_{\rm S}\cdot\boldsymbol{\nabla}\boldsymbol{v}_{\rm S}\right) - \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\rm S} &= \rho\boldsymbol{g} \quad \text{in} \quad \Omega_{\rm S}\times(0,t_{\rm f})\end{aligned}$$

and

$$\begin{aligned}\rho_{\rm F} c_{p{\rm F}} \left(\partial_t \vartheta_{\rm F} + \boldsymbol{v}_{\rm F}\cdot\boldsymbol{\nabla}\vartheta_{\rm F}\right) + \boldsymbol{\nabla}\cdot\boldsymbol{q}_{\rm F} &= Q_{\rm F} \quad \text{in} \quad \Omega_{\rm F}\times(0,t_{\rm f})\\ \rho_{\rm F}\left(\partial_t \boldsymbol{v}_{\rm F} + \boldsymbol{v}_{\rm F}\cdot\boldsymbol{\nabla}\boldsymbol{v}_{\rm F}\right) - \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\rm F} &= \rho\boldsymbol{g} \quad \text{in} \quad \Omega_{\rm F}\times(0,t_{\rm f})\end{aligned}$$

As it has been mentioned, the mechanical problem is written in abstract form, only for the purpose of presenting the assumptions required to have a coupled problem only for heat transfer.

The initial conditions of this problem are

$$\begin{aligned}\vartheta_{\rm S} &= \vartheta_0|_{\rm S} \quad \text{in} \quad \Omega_{\rm S}\times\{0\}\\ \vartheta_{\rm F} &= \vartheta_0|_{\rm F} \quad \text{in} \quad \Omega_{\rm F}\times\{0\}\end{aligned}$$

and its boundary conditions are

$$\begin{aligned}\vartheta_{\rm S} &= \vartheta_{\rm D}|_{\rm S} \quad \text{on} \quad \Gamma^{\vartheta}_{\rm DS}\times(0,t_{\rm f})\\ \boldsymbol{n}_{\rm S}\cdot\boldsymbol{q}_{\rm S} &= q_{\rm N}|_{\rm S} \quad \text{on} \quad \Gamma^{\vartheta}_{\rm NS}\times(0,t_{\rm f})\end{aligned}$$

$$\begin{aligned}\vartheta_{\rm F} &= \vartheta_{\rm D}|_{\rm F} \quad \text{on} \quad \Gamma^{\vartheta}_{\rm DF}\times(0,t_{\rm f})\\ \boldsymbol{n}_{\rm F}\cdot\boldsymbol{q}_{\rm F} &= q_{\rm N}|_{\rm F} \quad \text{on} \quad \Gamma^{\vartheta}_{\rm NF}\times(0,t_{\rm f})\end{aligned}$$

The conditions to be satisfied at the interface are the continuity of the temperatures and velocities as well as the normal components of heat fluxes and tractions, that read

$$\begin{aligned}\boldsymbol{v}_{\rm F}|_{\Gamma_{\rm SF}} = \boldsymbol{v}_{\rm S}|_{\Gamma_{\rm SF}} = \dot{\boldsymbol{u}}_{\rm S}|_{\Gamma_{\rm SF}}\\ \boldsymbol{n}_{\rm F}\cdot\boldsymbol{\sigma}_{\rm F}|_{\Gamma_{\rm SF}} = -\,\boldsymbol{n}_{\rm F}\cdot\boldsymbol{\sigma}_{\rm S}|_{\Gamma_{\rm SF}}\end{aligned}$$

and

$$\begin{aligned} \vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \\ \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{q}_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= -\boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{q}_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \end{aligned}$$

where now $\boldsymbol{n}_{\mathrm{S}}$ ($\boldsymbol{n}_{\mathrm{F}}$) is the normal exterior to the domain Ω_{S} (Ω_{F}). At this point we introduce the following assumptions:

1. The time derivative of the displacements of the solid is expected to be small compared to the dimensions of the solid itself and the fluid velocities, that is

$$\dot{\boldsymbol{u}}_{\mathrm{S}} \approx \mathbf{0}$$

2. The mechanical traction produced by the fluid on the solid is expected to be small, so that

$$\boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{\sigma}_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} \approx \mathbf{0}$$

With these approximations the interface conditions become

$$\begin{aligned} \boldsymbol{v}_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= \mathbf{0} \\ \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{\sigma}_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} &= \mathbf{0} \end{aligned}$$

and

$$\begin{aligned} \vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \\ \boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{q}_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= -\boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{q}_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \end{aligned}$$

and therefore the mechanical problems are uncoupled.

Weak form of the problem

Let us introduce the bilinear form a_{S} defined on the solid subdomain Ω_{S}

$$a_{\mathrm{S}}(\vartheta, v) := (\rho_{\mathrm{S}} c_{p\mathrm{S}} \partial_t \vartheta_{\mathrm{S}}, v)_{\Omega_{\mathrm{S}}} + (k_{\mathrm{S}} \boldsymbol{\nabla} \vartheta_{\mathrm{S}}, \boldsymbol{\nabla} v)_{\Omega_{\mathrm{S}}}$$

We have assumed that the advection term in the heat transport equation is negligible in the solid phase. Likewise, we define the bilinear form a_{F} defined on the fluid subdomain Ω_{F} as

$$a_{\mathrm{F}}(\vartheta, v) := (\rho_{\mathrm{F}} c_{p\mathrm{F}} \partial_t \vartheta_{\mathrm{F}}, v)_{\Omega_{\mathrm{F}}} + (\rho_{\mathrm{F}} c_{p\mathrm{F}} \boldsymbol{v}_{\mathrm{F}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{F}}, v)_{\Omega_{\mathrm{F}}} + (k_{\mathrm{F}} \boldsymbol{\nabla} \vartheta_{\mathrm{F}}, \boldsymbol{\nabla} v)_{\Omega_{\mathrm{F}}}$$

We also introduce the linear forms

$$l_{\mathrm{S}}(v) := \langle Q, v \rangle_{\Omega_{\mathrm{S}}} + \langle q_{\mathrm{N}}, v \rangle_{\Gamma^{\vartheta}_{\mathrm{NS}}}$$

and

$$l_{\mathrm{F}}(v) := \langle Q, v\rangle_{\Omega_{\mathrm{F}}} + \langle q_{\mathrm{N}}, v\rangle_{\Gamma^{\vartheta}_{\mathrm{NF}}}$$

The weak form of the problem consists in finding

$$\begin{aligned}\vartheta_{\mathrm{S}} &\in L^2(0,t_{\mathrm{f}};V^{\vartheta}_{\mathrm{S}}) \cap L^{\infty}(0,t_{\mathrm{f}};L^2(\Omega_{\mathrm{S}}))\\ \vartheta_{\mathrm{F}} &\in L^2(0,t_{\mathrm{f}};V^{\vartheta}_{\mathrm{F}}) \cap L^{\infty}(0,t_{\mathrm{f}};L^2(\Omega_{\mathrm{F}}))\end{aligned}$$

such that

$$\begin{aligned} a_{\mathrm{S}}(\vartheta,v) - \langle k_{\mathrm{S}}\boldsymbol{n}_{\mathrm{S}}\cdot\boldsymbol{\nabla}\vartheta_{\mathrm{S}}, v\rangle_{\Gamma_{\mathrm{SF}}} &= l(v) \quad \forall v \in V^0_{\mathrm{S}}\\ a_{\mathrm{F}}(\vartheta,v) - \langle k_{\mathrm{F}}\boldsymbol{n}_{\mathrm{F}}\cdot\boldsymbol{\nabla}\vartheta_{\mathrm{F}}, v\rangle_{\Gamma_{\mathrm{SF}}} &= l(v) \quad \forall v \in V^0_{\mathrm{F}}\\ \vartheta_{\mathrm{F}} &= \vartheta_{\mathrm{S}} \quad \text{on} \quad \Gamma_{\mathrm{SF}}\\ k_{\mathrm{F}}\boldsymbol{n}_{\mathrm{F}}\cdot\boldsymbol{\nabla}\vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= -\,k_{\mathrm{S}}\boldsymbol{n}_{\mathrm{S}}\cdot\boldsymbol{\nabla}\vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \quad \text{on} \quad \Gamma_{\mathrm{SF}}\end{aligned} \tag{8.4}$$

where the spaces $V^{\vartheta}_{\mathrm{S}}$ and V^0_{S} (resp. $V^{\vartheta}_{\mathrm{F}}$ and V^0_{F}) are defined in the same way as V^{ϑ} and V^0 but considering $\Gamma^{\vartheta}_{\mathrm{DS}}$ (resp. $\Gamma^{\vartheta}_{\mathrm{DF}}$) instead of $\Gamma^{\vartheta}_{\mathrm{D}}$.

8.2.3 The wall function strategy

Let us split the domain Ω into the solid subdomain Ω_{S}, a boundary layer in the fluid Ω_{B} and the rest of the fluid subdomain Ω_{F}, as illustrated in Figure 8.1 (right). Let also $\Gamma_{\mathrm{SB}}(t)$ be the interface between Ω_{S} and Ω_{B}, and $\Gamma_{\mathrm{BF}}(t)$ be the interface between Ω_{B} and Ω_{F}. Apart from the equations given in the previous subsection for the solid and the fluid subdomain, we now need to solve the problem on the boundary layer subdomain, which consists in finding a temperature ϑ_{B} and a velocity $\boldsymbol{v}_{\mathrm{B}}$ such that

$$\rho c_p\left(\partial_t\vartheta_{\mathrm{B}} + \boldsymbol{v}_{\mathrm{B}}\cdot\boldsymbol{\nabla}\vartheta_{\mathrm{B}}\right) + \boldsymbol{\nabla}\cdot\boldsymbol{q}_{\mathrm{B}} = q_{\mathrm{B}} \tag{8.5}$$

$$\rho\left(\partial_t\boldsymbol{v}_{\mathrm{B}} + \boldsymbol{v}_{\mathrm{B}}\cdot\boldsymbol{\nabla}\boldsymbol{v}_{\mathrm{B}}\right) - \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\mathrm{B}} = \rho\boldsymbol{g} \tag{8.6}$$

Under the same assumptions as in the previous subsection the interface conditions are

$$\begin{aligned}\boldsymbol{v}_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} &= \boldsymbol{0}\\ \boldsymbol{n}_{\mathrm{S}}\cdot\boldsymbol{\sigma}_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}} &= \boldsymbol{0}\end{aligned}$$

$$\begin{aligned}\vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} &= \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}}\\ \boldsymbol{n}_{\mathrm{B}}\cdot\boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} &= -\,\boldsymbol{n}_{\mathrm{S}}\cdot\boldsymbol{q}_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{v}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} &= \boldsymbol{v}_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}}\\ \boldsymbol{n}_{\mathrm{F}}\cdot\boldsymbol{\sigma}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} &= -\,\boldsymbol{n}_{\mathrm{B}}\cdot\boldsymbol{\sigma}_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}}\end{aligned}$$

$$\vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} = \vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}}$$
$$\boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{q}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} = -\,\boldsymbol{n}_{\mathrm{B}} \cdot \boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}}$$

If the boundary layer is of constant width, one may assume that its normal satisfies

$$\boldsymbol{n}_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} = -\,\boldsymbol{n}_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}}$$

The problem on the boundary layer subdomain is now approximately solved using the wall function approach, which is described next.

Wall function revisited

The so called wall function approach is a method for the approximate solution of the fluid mechanics problems with strong boundary layers. These boundary layers are removed from the computational domain and universal velocity profiles are used to define the boundary condition in terms of the boundary conditions on the solid surface, as shown in Figure 8.3.

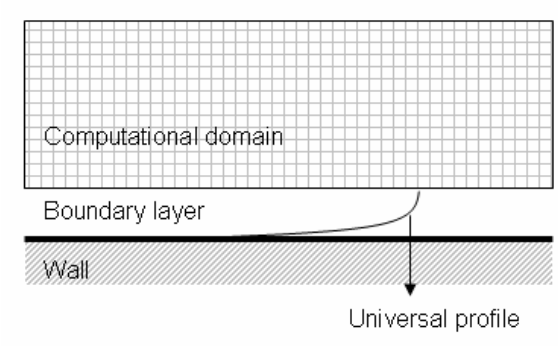

Figure 8.3: Wall function approach

This approximated solution is found assuming negligible inertial terms and external forces. From 8.5 and 8.6 it is seen that this yields

$$\begin{aligned} \boldsymbol{\nabla} \cdot \boldsymbol{q}_{\mathrm{B}} &= 0 \\ \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\mathrm{B}} &= 0 \end{aligned}$$

which imply constant stresses and heat fluxes across the boundary layer. We denote now by q the normal component of the heat flux (not to be confused with the heat sources introduced before)

$$q = \boldsymbol{n}_{\mathrm{B}} \cdot \boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}}$$

and by $\boldsymbol{t}$ the tangential stress

$$\boldsymbol{t} = \boldsymbol{n} \cdot \boldsymbol{\sigma}_{\mathrm{B}} - (\boldsymbol{n} \cdot \boldsymbol{\sigma}_{\mathrm{B}} \cdot \boldsymbol{n})\, \boldsymbol{n}|_{\Gamma_{\mathrm{SB}}}$$

evaluated at the wall. The mechanical problem is solved selecting a local coordinate system such that the first local direction is $\boldsymbol{t}$. If we denote by σ the norm of the tangential component of the stress (not to be confused with the full stress tensor), by u_{B} the component of the velocity in this system and by y the coordinate normal to the solid surface, the constitutive equations are

$$q = \left(k + k^t\right) \frac{d\vartheta_{\mathrm{B}}}{dy} \tag{8.7}$$

$$\sigma = \left(\mu + \mu^t\right) \frac{du_{\mathrm{B}}}{dy} \tag{8.8}$$

where μ^t and k^t are the turbulent viscosity and conductivity. Given the turbulent viscosity and conductivity we can integrate 8.7 and 8.8 to obtain the velocity and temperature profiles. The definition of the turbulent viscosity and conductivity is the definition of the model we are using to approximate the problem and is based on experimental correlations [89]. One of these models is the one that results in the logarithmic profiles for the velocity and temperature. This model is based on the existence of a zone near the wall, called laminar sublayer, in which the velocity is small and therefore also the local Reynolds number is. In the laminar sublayer the turbulent viscosity is neglected and we have

$$\frac{\sigma}{\rho} = \frac{\mu}{\rho} \frac{du_{\mathrm{B}}}{dy}$$

that is written in dimensionless form introducing the friction velocity $u_* = \sqrt{\sigma|_{y=0}/\rho}$. As the stress is constant across the boundary layer, we have

$$\frac{u_{\mathrm{B}}}{u_*} = \frac{\rho u_* y}{\mu}$$

On the other hand, in the turbulent region we may approximate $\mu^t = \rho\kappa y u_*$, κ being the Von Karman constant, and integrating 8.8 we have

$$\frac{u_{\mathrm{B}}}{u_*} = \frac{u_0}{u_*} + \frac{1}{\kappa} \ln\left(\frac{y}{y_0}\right)$$

where y_0 is the width of the laminar sublayer and u_0 the value of the velocity at this point. Defining the dimensionless velocity $u_{\mathrm{B}}^+ = u_{\mathrm{B}}/u_*$ and distance $y^+ = \rho y u^*/\mu$ and taking $y_0^+ = 11.6$ the final solution reads

$$u_{\mathrm{B}}^+ = \begin{cases} y^+ & \text{if} \quad y^+ < y_0^+ \\ \frac{1}{\kappa} \ln\left(y^+\right) + 5.5 & \text{if} \quad y^+ \geq y_0^+ \end{cases}$$

In the same way, integrating 8.7 we arrive to

$$\vartheta_{\mathrm{B}}^+ = \begin{cases} \mathrm{Pr}\, y^+ & \text{if} \quad y^+ < y_0^+ \\ \mathrm{Pr}^t \left[\frac{1}{\kappa} \ln\left(y^+\right) + \mathrm{P}_\theta\right] & \text{if} \quad y^+ \geq y_0^+ \end{cases}$$

where $\mathrm{Pr} := \kappa/\nu$ is the Prandtl number, Pr^t is the turbulent Prandtl number (which is part of the constitutive model), P_θ a function that gives the temperature jump across the laminar sublayer and the dimensionless temperature is defined as

$$\vartheta_{\mathrm{B}}^{+} = -\frac{\rho c_p u_*}{q}\left(\vartheta - \vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}}\right)$$

Strong form of the problem

Having an analytical solution to the problem in the boundary layer domain we can rewrite the complete problem in terms of two subdomains, the fluid (excluding the boundary layer) and the solid. To this end, let us remark that the solution of the thermal problem obtained using the wall function method is a constant heat flux and therefore

$$q = \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{q}_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}} = -\,\boldsymbol{n}_{\mathrm{B}} \cdot \boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} = \boldsymbol{n}_{\mathrm{B}} \cdot \boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}} = -\,\boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{q}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}}$$

This flux is proportional to the temperature jump across the layer

$$\begin{aligned} q &= \alpha\left(\vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}} - \vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}}\right) \\ &= \alpha\left(\vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} - \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}}\right) \end{aligned}$$

where

$$\alpha = \frac{\rho c_p u_*}{\vartheta^+}$$

and $\vartheta^+ = \vartheta_{\mathrm{B}}^{+}(\delta^+)$ is defined in terms of δ^+, the dimensionless boundary layer thickness. This parameter depends finally on the particular choice of the turbulent viscosity and conductivity of the wall function method. In the same way, as the tangential stress is constant across the boundary layer, we have

$$\boldsymbol{t}|_{\Gamma_{\mathrm{SB}}} = \boldsymbol{t}|_{\Gamma_{\mathrm{BF}}} = \beta\, \boldsymbol{v}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}}$$

for a certain parameter β, and the normal component of the velocity is set to zero

$$\boldsymbol{n} \cdot \boldsymbol{v}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} = 0$$

We can finally state the strong form of the problem as finding ϑ_{S} and ϑ_{F}, as well as $\boldsymbol{v}_{\mathrm{S}}$ and $\boldsymbol{v}_{\mathrm{F}}$, such that

$$\begin{aligned} \rho_{\mathrm{S}} c_{p\mathrm{S}} \partial_t \vartheta_{\mathrm{S}} + \boldsymbol{\nabla} \cdot \boldsymbol{q}_{\mathrm{S}} &= q_{\mathrm{S}} \quad \text{in} \quad \Omega_{\mathrm{S}} \times (0, t_{\mathrm{f}}) \\ \rho_{\mathrm{S}} \partial_t \boldsymbol{v}_{\mathrm{S}} - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\mathrm{S}} &= \rho \boldsymbol{g} \quad \text{in} \quad \Omega_{\mathrm{S}} \times (0, t_{\mathrm{f}}) \end{aligned}$$

and

$$\begin{aligned} \rho_{\mathrm{S}} c_{p\mathrm{S}} \left(\partial_t \vartheta_{\mathrm{F}} + \boldsymbol{v}_{\mathrm{F}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{F}}\right) + \boldsymbol{\nabla} \cdot \boldsymbol{q}_{\mathrm{F}} &= q_{\mathrm{F}} \quad \text{in} \quad \Omega_{\mathrm{F}} \times (0, t_{\mathrm{f}}) \\ \rho_{\mathrm{S}} \left(\partial_t \boldsymbol{v}_{\mathrm{F}} + \boldsymbol{v}_{\mathrm{F}} \cdot \boldsymbol{\nabla} \boldsymbol{v}_{\mathrm{F}}\right) - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\mathrm{F}} &= \rho \boldsymbol{g} \quad \text{in} \quad \Omega_{\mathrm{F}} \times (0, t_{\mathrm{f}}) \end{aligned}$$

and now the interface conditions become

$$\begin{aligned} \boldsymbol{t}|_{\Gamma_{\mathrm{BF}}} &= \beta\, \boldsymbol{v}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} \\ \boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{v}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} &= 0 \\ \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{\sigma}_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}} &= \mathbf{0} \end{aligned}$$

and

$$\begin{aligned} q &= \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{q}_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}} = -\,\boldsymbol{n}_{\mathrm{B}} \cdot \boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} = \boldsymbol{n}_{\mathrm{B}} \cdot \boldsymbol{q}_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}} = -\,\boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{q}_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} \\ &= \alpha \left(\vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{BF}}} - \vartheta_{\mathrm{B}}|_{\Gamma_{\mathrm{SB}}} \right) = \alpha \left(\vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{BF}}} - \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SB}}} \right) \end{aligned}$$

Finally, assuming the boundary layer thin we can write the final approximation as

$$\begin{aligned} q &= \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{q}_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} = -\,\boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{q}_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} \\ &= \alpha \left(\vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} - \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \right) \end{aligned}$$

which is a surface-convection-type boundary condition. As in [33], we have derived an expression for α based on the physical model being used (with a completely different meaning with respect to the mentioned reference).

Weak form of the problem

As in the previous subsection, let us introduce the bilinear form a_{S} defined on the solid subdomain Ω_{S}

$$a_{\mathrm{S}}(\vartheta, v) := (\rho_{\mathrm{S}} c_{p\mathrm{S}} \partial_t \vartheta_{\mathrm{S}}, v)_{\Omega_{\mathrm{S}}} + (k_{\mathrm{S}} \boldsymbol{\nabla} \vartheta_{\mathrm{S}}, \boldsymbol{\nabla} v)_{\Omega_{\mathrm{S}}}$$

and the bilinear form a_{F} defined on the fluid subdomain Ω_{F}

$$a_{\mathrm{F}}(\vartheta, v) := (\rho_{\mathrm{F}} c_{p\mathrm{F}} \partial_t \vartheta_{\mathrm{F}}, v)_{\Omega_{\mathrm{F}}} + (\rho_{\mathrm{F}} c_{p\mathrm{F}} \boldsymbol{v}_{\mathrm{F}} \boldsymbol{\cdot} \boldsymbol{\nabla} \vartheta_{\mathrm{F}}, v)_{\Omega_{\mathrm{F}}} + (k_{\mathrm{F}} \boldsymbol{\nabla} \vartheta_{\mathrm{F}}, \boldsymbol{\nabla} v)_{\Omega_{\mathrm{F}}}$$

as well as the linear forms

$$l_{\mathrm{S}}(v) := \langle q, v \rangle_{\Omega_{\mathrm{S}}} + \langle q_{\mathrm{N}}, v \rangle_{\Gamma^{\vartheta}_{\mathrm{NS}}}$$

$$l_{\mathrm{F}}(v) := \langle q, v \rangle_{\Omega_{\mathrm{F}}} + \langle q_{\mathrm{N}}, v \rangle_{\Gamma^{\vartheta}_{\mathrm{NF}}}$$

The weak form of the problem consists in finding

$$\begin{aligned} \vartheta_{\mathrm{S}} &\in L^2(0, t_{\mathrm{f}}; V^{\vartheta}_{\mathrm{S}}) \cap L^{\infty}(0, t_{\mathrm{f}}; L^2(\Omega_{\mathrm{S}})) \\ \vartheta_{\mathrm{F}} &\in L^2(0, t_{\mathrm{f}}; V^{\vartheta}_{\mathrm{F}}) \cap L^{\infty}(0, t_{\mathrm{f}}; L^2(\Omega_{\mathrm{F}})) \end{aligned}$$

such that

$$\begin{aligned} a_{\mathrm{S}}(\vartheta, v) - \langle k_{\mathrm{S}} \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{S}}, v \rangle_{\Gamma_{\mathrm{SF}}} &= l(v) \quad \forall v \in V^0_{\mathrm{S}} \\ a_{\mathrm{F}}(\vartheta, v) - \langle k_{\mathrm{F}} \boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{F}}, v \rangle_{\Gamma_{\mathrm{SF}}} &= l(v) \quad \forall v \in V^0_{\mathrm{F}} \\ k_{\mathrm{F}} \boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= \alpha \left(\vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} - \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \right) \quad \text{on} \quad \Gamma_{\mathrm{SF}} \\ k_{\mathrm{F}} \boldsymbol{n}_{\mathrm{F}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{F}}|_{\Gamma_{\mathrm{SF}}} &= -\,k_{\mathrm{S}} \boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{\nabla} \vartheta_{\mathrm{S}}|_{\Gamma_{\mathrm{SF}}} \quad \text{on} \quad \Gamma_{\mathrm{SF}} \end{aligned}$$

Comparing the weak form of this problem to 8.4 the only difference is a jump on the temperature proportional to the heat flux between domains. Our derivation allows us to give an interpretation to the surface convection coefficient α in terms of the wall function model used on the boundary layer subdomain.

8.3 Numerical approximation

Three different continuous problems have been described in section 8.2 but the first one, that consists of the solution of a global problem in the whole domain, was presented to define the problem we are facing and has not been actually implemented. The other two possibilities imply the solution of local thermal problems as well as local mechanical problems for the fluid and the solid. In this section we present the numerical approximation to the problem and we will concentrate on the thermal problem only. In the first subsection we will present the finite element discretization of the problem considering generically the domain Ω. This approximation could be applied on the whole domain but will be actually applied on each subdomain. A similar scheme is used to solve the mechanical problem on the fluid. Details on the finite element approximation to the Navier Stokes equation can be found in [29, 35, 36]. In the second subsection we describe an iterative strategy to solve the global thermal problem iteratively solving local problems on each subdomain.

8.3.1 Finite element approximation

The Galerkin finite element approximation of this problem is standard. Based on a partition of the domain $\mathcal{P}_h = \{K\}$ in n_{el} elements K, the space V^ϑ where the temperature is sought is approximated by a finite dimensional space V_h^ϑ (built using polynomials). If the space of test functions V^0 is approximated by V_h^0, defined in a similar way, the semi-discrete problem consists in finding $\vartheta_h \in L^2(0, t_{\text{f}}; V_h^\vartheta)$ such that

$$a\left(\vartheta_h, v_h\right) = l\left(v_h\right) \quad \forall v_h \in V_h^0$$

It is well known that this formulation is unstable when the convection dominates and therefore we employ a stabilized finite element formulation based on the subgrid scale method with an algebraic approximation to the subscales [75]. This method is based on a decomposition of the continuous space of the form (for simplicity consider homogeneous Dirichlet boundary conditions):

$$V^\vartheta = V_h^\vartheta \oplus \widetilde{V^\vartheta}$$

where $\widetilde{V^\vartheta}$ can be in principle any space to complete V_h^ϑ in V^ϑ. To fix ideas we may think $\widetilde{V^\vartheta}$ as the orthogonal complement of V_h^ϑ with respect to the L^2 inner product. Since $\widetilde{V^\vartheta}$ represents the component of V^ϑ which is not reproduced by the finite element space, we

call it the space of subscales or subgrid scales. The continuous Equation 8.3 can now be written as the system

$$a\left(\vartheta_h, v_h\right) + a(\widetilde{\vartheta}, v_h) = l\left(v_h\right) \quad \forall v_h \in V_h^0 \tag{8.9}$$

$$a\left(\vartheta_h, \widetilde{v}\right) + a(\widetilde{\vartheta}, \widetilde{v}) = l\left(\widetilde{v}\right) \quad \forall \widetilde{v} \in \widetilde{V^0} \tag{8.10}$$

After integration by parts within each element domain Equation 8.10 is equivalent to finding $\widetilde{\vartheta} \in \widetilde{V^\vartheta}$ such that

$$\rho c_p \left(\partial_t \widetilde{\vartheta} + \boldsymbol{v} \cdot \boldsymbol{\nabla} \widetilde{\vartheta}\right) - k\nabla^2 \widetilde{\vartheta} = R_h + \vartheta_{h,\text{ort}} \quad \text{in} \quad K, \tag{8.11}$$

$$\widetilde{\vartheta} = \widetilde{\vartheta}_{\text{ske}} \quad \text{on} \quad \partial K, \tag{8.12}$$

for any $K \in \mathcal{P}_h$, where

$$R_h := Q - \rho c_p \left(\partial_t \vartheta_h + \boldsymbol{v} \cdot \boldsymbol{\nabla} \vartheta_h\right) + k\nabla^2 \vartheta_h$$

is the residual of the finite element component. The function $\vartheta_{h,\text{ort}}$ is obtained from the condition that $\widetilde{\vartheta}$ must belong to $\widetilde{V^\vartheta}$ (and not to the whole space V^ϑ) and the function $\widetilde{\vartheta}_{\text{ske}}$, that we call the skeleton of $\widetilde{\vartheta}$, is defined on the element boundaries such that the normal component of the fluxes of ϑ is continuous across these boundaries [29]. Problem 8.9-8.10 is exactly equivalent to 8.9-8.11-8.12. The approximate problem is defined by the way in which Problem 8.11-8.12 is solved as well as by the way in which the functions $\vartheta_{h,\text{ort}}$ and $\widetilde{\vartheta}_{\text{ske}}$ are taken.

The simplest way to approximately solve Problem 8.11-8.12 is to neglect the time variation of $\widetilde{\vartheta}$ and to approximate the spatial differential operator $(\rho c_p \boldsymbol{v} \cdot \boldsymbol{\nabla} - k\nabla^2)$ by a parameter τ^{-1} that depends on the coefficients as [29]

$$\tau = \left[c_1 \frac{k}{h^2} + c_2 \frac{\rho c_p \left\|\boldsymbol{v}\right\|}{h}\right]^{-1},$$

where c_1 and c_2 are algorithmic constants that we take $c_1 = 4$ and $c_2 = 2$ for linear elements. These approximations give

$$\widetilde{\vartheta} \approx \tau R_h \tag{8.13}$$

Another possibility [35, 36] is to approximate the spatial differential operator but to allow the subscales to be time dependent. In this case, instead of 8.13 we have, at each point, an ordinary differential equation for the subscales evolution given by

$$\rho c_p \partial_t \widetilde{\vartheta} + \tau^{-1} \widetilde{\vartheta} = R_h \tag{8.14}$$

where the same expression for the stabilization parameter τ is used.

The approximation given by 8.13 and that given by 8.14 both have an implicit assumption on the function $\widetilde{\vartheta}_{\text{ske}}$ and the space $\widetilde{V^\vartheta}$, and therefore on the function $\vartheta_{h,\text{ort}}$.

This is not the only possibility as suggested in a previous work [29] where, in particular, the subscales are taken orthogonal to the finite element space. After integrating by parts within each element those terms that involve spatial derivatives of $\widetilde{\vartheta}$ in 8.9 and neglecting boundary terms, the final semi-discrete weak form of the problem reads: find $\vartheta_h \in C^0\left(0, t_\mathrm{f}; V_h^\vartheta\right)$ such that

$$a\left(\vartheta_h, v_h\right) + \left(\rho c_p \partial_t \widetilde{\vartheta}, v_h\right)_\Omega - \sum_K \left(\rho c_p \boldsymbol{v} \cdot \boldsymbol{\nabla} v_h + k \nabla^2 v_h, \widetilde{\vartheta}\right)_K = l\left(v_h\right) \quad \forall v_h \in V_h^0$$

where $\widetilde{\vartheta}$ is given either by 8.13 or by 8.14 (in the first case $\partial_t \widetilde{\vartheta} = 0$ is assumed).

The time discretization of the problem will be performed using the generalized trapezoidal rule, that is to say, a finite difference scheme. The fully discrete problem will be obtained by discretizing in time the semi-discrete problem. Obviously, it is also possible to start by discretizing first in time and then in space. We will use the first option but let us point out that if the subscales problem is solved using 8.14 the scheme is commutative [36]. Let us consider a uniform partition of the time interval $(0, t_\mathrm{f})$ of size δt and let us introduce the following notation

$$\begin{aligned} f^{n+\theta} &= \theta f^{n+1} + (1-\theta) f^n \\ \delta_t f^n &= \left(f^{n+1} - f^n\right)/\delta t = \left(f^{n+\theta} - f^n\right)/\left(\theta \delta t\right) \end{aligned}$$

where $0 < \theta \leq 1$. For $\theta = 1$ we obtain the backward Euler scheme, of first order, and for $\theta = 1/2$ the Crank-Nicolson scheme, of second order. Both are unconditionally stable. Let us define

$$\begin{aligned} a^h\left(\vartheta_h^{n+1}, v_h\right) = &\left(\rho^{n+\theta} c_p^{n+\theta} \delta_t \vartheta_h^n, v\right)_\Omega + \left(\rho^{n+\theta} c_p^{n+\theta} \boldsymbol{v}^{n+\theta} \cdot \boldsymbol{\nabla} \vartheta_h^{n+\theta}, v\right)_\Omega + \left(k^{n+\theta} \boldsymbol{\nabla} \vartheta_h^{n+\theta}, \boldsymbol{\nabla} v_h\right)_\Omega \\ &+ \left(\rho^{n+\theta} c_p^{n+\theta} \delta_t \widetilde{\vartheta}^n, v_h\right)_\Omega - \sum_K \left(\rho^{n+\theta} c_p^{n+\theta} \boldsymbol{v}^{n+\theta} \cdot \boldsymbol{\nabla} v_h + k^{n+\theta} \nabla^2 v_h, \widetilde{\vartheta}^{n+\theta}\right)_K \end{aligned}$$

where $\widetilde{\vartheta}^{n+1}$ is given by

- Quasi-static subscales:

$$\widetilde{\vartheta}^{n+\theta} \approx \tau^{n+\theta} R_h^{n+\theta}$$

- Dynamic subscales:

$$\rho^{n+\theta} c_p^{n+\theta} \delta_t \widetilde{\vartheta}^n + \tau^{-1} \widetilde{\vartheta}^{n+\theta} = R_h^{n+\theta}$$

The fully discrete problem consists in: for $n = 1, 2, ...$, find $\vartheta_h^{n+1} \in V_h^\vartheta$ such that

$$a^h\left(\vartheta_h^{n+1}, v_h\right) = \left\langle q^{n+\theta}, v_h \right\rangle_\Omega + \left\langle q_\mathrm{N}^{n+\theta}, v_h \right\rangle_{\Gamma_\mathrm{N}^\vartheta} \quad \forall v_h \in V_h^0$$

8.3.2 Coupling strategy

As mentioned before, we consider a geometric domain decomposition of the problem by means of a non-overlapping subdomain approach. Therefore, at each time step, we expect to construct the solution of the problem from the solution of local problems for the fluid and the structure using the interface conditions already described. This is carried out by iteratively solving local problems on each domain until convergence on the interface conditions is satisfied, that is to say, we use an iteration-by-subdomain strategy [74]. The choice of the boundary conditions of the local problems should be such that interface conditions presented in section 8.2 are satisfied when convergence is achieved. It is well known from the theory of domain decomposition methods that in the case of non-overlapping subdomains we can choose Dirichlet-Neumann(Robin), Neumann(Robin)-Dirichlet or Robin-Robin. Let us define a_{S}^h and a_{F}^h in the same way as a^h was defined in the previous subsection.

If we use the full resolution strategy and we apply Dirichlet boundary conditions to the solid and Neumann boundary conditions to the fluid, which according to [60, 127] is the most stable option, the coupling algorithm can be written as: for each time step n and each iteration i find $\vartheta_{\mathrm{S},h}^{n+1,i+1} \in V_{\mathrm{S},h}^{\vartheta}$ and $\vartheta_{\mathrm{F},h}^{n+1,i+1} \in V_{\mathrm{F},h}^{\vartheta}$ such that

$$a_{\mathrm{S}}^h\left(\vartheta_{\mathrm{S},h}^{n+1,i+1}, v_h\right) = \left\langle q^{n+\theta}, v_h\right\rangle_{\Omega} + \left\langle q_{\mathrm{N}}^{n+\theta}, v_h\right\rangle_{\Gamma_{\mathrm{N}}^{\vartheta}} \tag{8.15}$$

$$a_{\mathrm{F}}^h\left(\vartheta_{\mathrm{F},h}^{n+1,i+1}, v_h\right) = \left\langle q^{n+\theta}, v_h\right\rangle_{\Omega} + \left\langle q_{\mathrm{N}}^{n+\theta}, v_h\right\rangle_{\Gamma_{\mathrm{N}}^{\vartheta}} - \left\langle k_{\mathrm{S}}\boldsymbol{n}_{\mathrm{S}} \cdot \boldsymbol{\nabla}\vartheta_{\mathrm{S},h}^{n+1,i}, v_h\right\rangle_{\Gamma_{\mathrm{SF}}} \tag{8.16}$$

where $v_h \in V_{\mathrm{S},h}^0$ in 8.15 and in $v_h \in V_{\mathrm{F},h}^0$ in 8.16. It is understood that now these spaces $V_{\mathrm{S},h}^0$ and $V_{\mathrm{S},h}^{\vartheta}$ are constructed including Γ_{SF} in the Dirichlet part of the boundary in order to satisfy

$$\vartheta_{\mathrm{S},h}^{n+1,i+1} = \vartheta_{\mathrm{F},h}^{n+1,k} \quad \text{on} \quad \Gamma_{\mathrm{SF}}$$

We can take $k = i+1$ or $k = i$. In the first case the solution of this problems is sequential, that is, we solve first for the fluid and then for the solid, whereas in the second one it can be parallel.

If we use the wall function strategy, the coupling algorithm can be written as: for each time step and each iteration i find $\vartheta_{\mathrm{S},h}^{n+1,i+1} \in V_{\mathrm{S},h}^{\vartheta}$ and $\vartheta_{\mathrm{F},h}^{n+1,i+1} \in V_{\mathrm{F},h}^{\vartheta}$ such that

$$a_{\mathrm{S}}^h\left(\vartheta_{\mathrm{S},h}^{n+1,i+1}, v_h\right) = \left\langle q^{n+\theta}, v_h\right\rangle_{\Omega} + \left\langle q_{\mathrm{N}}^{n+\theta}, v_h\right\rangle_{\Gamma_{\mathrm{N}}^{\vartheta}} + \left\langle \alpha\left(\vartheta_{\mathrm{S},h}^{n+1,i+1} - \vartheta_{\mathrm{F},h}^{n+1,i}\right), v_h\right\rangle_{\Gamma_{\mathrm{SF}}} \tag{8.17}$$

$$a_{\mathrm{F}}^h\left(\vartheta_{\mathrm{F},h}^{n+1,i+1}, v_h\right) = \left\langle q^{n+\theta}, v_h\right\rangle_{\Omega} + \left\langle q_{\mathrm{N}}^{n+\theta}, v_h\right\rangle_{\Gamma_{\mathrm{N}}^{\vartheta}} + \left\langle \alpha\left(\vartheta_{\mathrm{F},h}^{n+1,i+1} - \vartheta_{\mathrm{S},h}^{n+1,k}\right), v_h\right\rangle_{\Gamma_{\mathrm{SF}}} \tag{8.18}$$

where $v_h \in V_{\mathrm{S},h}^0$ in 8.17 and in $v_h \in V_{\mathrm{F},h}^0$ in 8.18. Again we can take $k = i+1$ or $k = i$.

Apart from the fact that the physical models represented by System 8.15-8.16 and by System 8.17-8.18 are different, some conceptual differences have to be remarked. Firstly, it is observed that the imposition of the transmission conditions is "symmetric" for the fluid and the solid, contrary to the Dirichlet-Neumann conditions in 8.15-8.16. Secondly, 8.17-8.18 does not require the calculation of the normal heat fluxes from the solid to the fluid,

as needed in 8.16. This calculation is always involved in a finite element code, particularly for non-matching meshes between the fluid and the solid (see next section). Finally, in the limit $\alpha \to \infty$ it can be shown that the solution of System 8.17-8.18 converges to the solution of system 8.15-8.16, the convergence rate being α^{-1}. This can be proved using the analysis developed in [22]. Nevertheless, in our approach α has a physical meaning and, moreover, taking α large leads to ill-conditioning problems.

8.4 Implementation aspects

8.4.1 A master slave algorithm

One important point of the iteration-by-subdomain strategy proposed is that we already had programs that solve the fluid dynamics problem and the structural problem. Then a master/slave algorithm was implemented by developing a third code (the master code) in order to control the iterative process. The MPICH2 library, an implementation of the MPI-2 standard, provides functions for process communications that are used to interchange the data needed to apply boundary conditions on each dedicated (slave) code. Some minor modifications on these codes are needed in order to exchange data with the master. In order to perform a calculation, input data for each subproblem needs to be generated and the master code starts the calculation by starting the slave process (this is only possible under MPI-2 standard). During the calculation the master code needs to define the boundary conditions to be applied on each subproblem. The situation is illustrated in Figure 8.4.

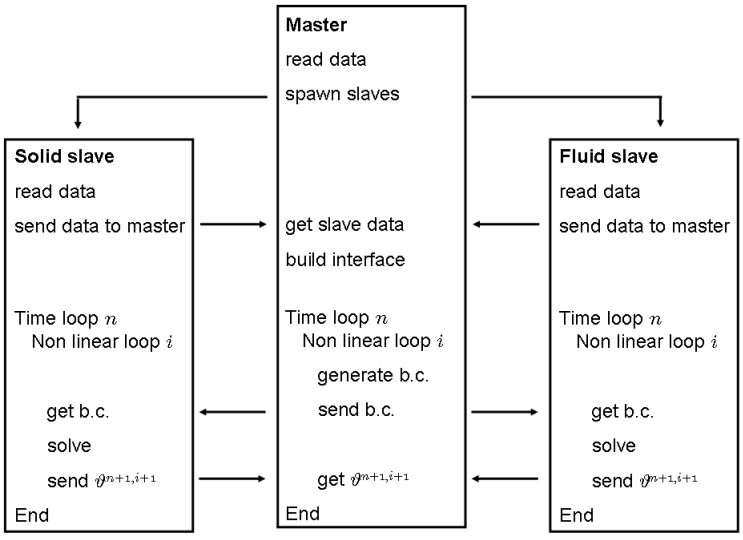

Figure 8.4: Master slave implementation

8.4.2 Boundary data interpolation

Another aspect of the implementation that deserves a comment is the interpolation of the boundary conditions to be applied in one subdomain from the results obtained in the other subdomain. For each interface node, this interpolation is performed finding, in the mesh of the other subdomain, the element in which it is located, the so called host element. The process is illustrated in Figure 8.5

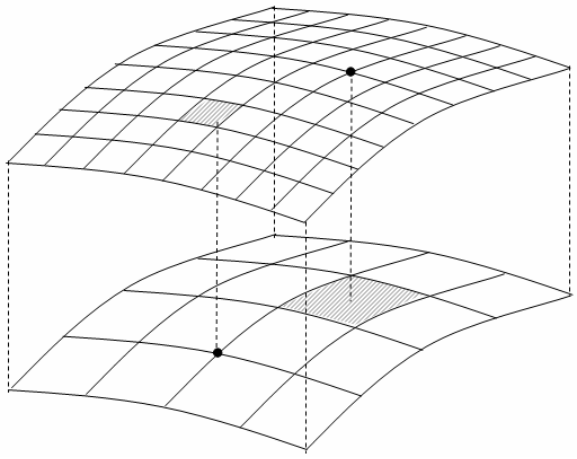

Figure 8.5: Boundary data interpolation

The element search strategy used in this work [74] is based on a quad-tree (oct-tree in 3D) algorithm. It consists of two steps: the preprocess in which a tree-like structure is built and a process in which the search is performed. In the preprocess the host computational domain is embedded in a box taking the maximum and minimum nodes coordinates to define its coordinates. This box is then subdivided recursively into 4 boxes (eight boxes in 3D) until each box contains a prescribed (small) number of elements. Once this preprocess has been performed, the process to search the host element of a given point is faster. Given the test point coordinates $\boldsymbol{x}$ we recursively locate the boxes it belongs to and we find a small number of elements in which the point must be. Then on each element we perform a local coordinates test. If the coordinates on the parent domain of the standard isoparametric mapping are denoted by $\boldsymbol{\xi}$, we have

$$\boldsymbol{x} = \sum_a N^a \left(\boldsymbol{\xi}\right) \boldsymbol{x}^a$$

and starting with $\boldsymbol{x}$ we solve this equation for $\boldsymbol{\xi}$ using a Newton-Raphson procedure. The solution permits us to determine if the point belongs to the element and if it is the case we already have the shape functions on the host mesh evaluated at that point. They are then used to interpolate the needed boundary data.

8.5 Numerical examples

In this section we present two numerical examples. The first one is a very simple one-dimensional example intended to show the role played by the wall function approach when very thin boundary layers are created. The second example is a practical application of the thermal coupling described in this chapter.

8.5.1 A one dimensional example

Assume we have two different materials F and S on domains $\Omega_{\mathrm{F}} = [-1, 0]$ and $\Omega_{\mathrm{S}} = [0, 1]$, with conductivities k_{F} and k_{S}, respectively defined by

$$\begin{aligned} k_{\mathrm{F}} &= C\frac{1 - e^{-\gamma}}{1 - e^{-\gamma} + \gamma e^{\gamma x}} \\ k_{\mathrm{S}} &= C\frac{1 - e^{-\gamma}}{1 - e^{-\gamma} + \gamma e^{-\gamma x}} \end{aligned}$$

where C and γ are constants. Both coefficients have a boundary layer near $x = 0$ and the constant γ is a measure of the boundary layer width. The coefficients are shown in Figure 8.6 for $\gamma = 10$ and $\gamma = 100$ and $C = 1$.

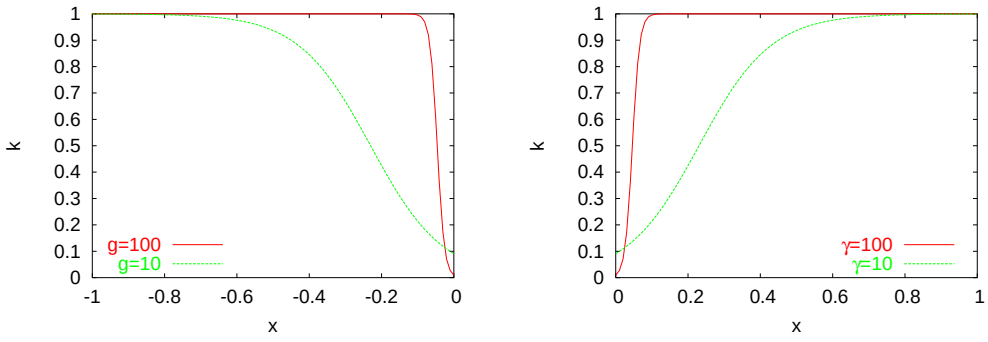

Figure 8.6: Thermal conductivity

The problem can be written as

$$\begin{aligned} -\frac{d}{dx}\left(k_{\mathrm{F}}\frac{d\vartheta}{dx}\right) &= Q_{\mathrm{F}} \quad \text{in} \quad \Omega_{\mathrm{F}} \\ -\frac{d}{dx}\left(k_{\mathrm{S}}\frac{d\vartheta}{dx}\right) &= Q_{\mathrm{S}} \quad \text{in} \quad \Omega_{\mathrm{S}} \end{aligned}$$

with the transmission conditions

$$-k_{\mathrm{F}}\frac{d\vartheta}{dx} = -k_{\mathrm{S}}\frac{d\vartheta}{dx} \quad \text{at} \quad x = 0$$

$$\vartheta_{\mathrm{F}}(0) = \vartheta_{\mathrm{S}}(0)$$

and the boundary conditions

$$\begin{aligned} \vartheta_{\mathrm{F}}(-1) &= 1 \\ \vartheta_{\mathrm{S}}(1) &= 0 \end{aligned}$$

The exact solution to this problem is

$$\begin{aligned} \vartheta_{\mathrm{F}}(x) &= \frac{1}{2}\left(-x + \frac{1-e^{\gamma x}}{1-e^{-\gamma}}\right) \\ \vartheta_{\mathrm{S}}(x) &= \frac{1}{2}\left(-x - \frac{1-e^{-\gamma x}}{1-e^{-\gamma}}\right) \end{aligned}$$

We have solved this problem in the case of $\gamma = 100$ using the first domain decomposition strategy using three different meshes of 10, 20 and 40 elements. The solution is compared to the analytical one in Figure 8.7. We have also solved this problem using the second approach using a mesh of 10 elements and the result is compared to the one obtained by the previous method and to the analytical solution in Figure 8.8.

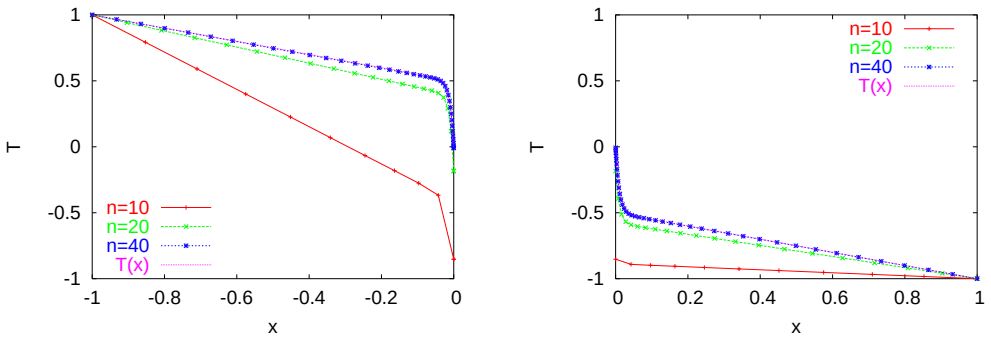

Figure 8.7: Finite element solution obtained using domain decomposition with two subdomains compared to the analytic one

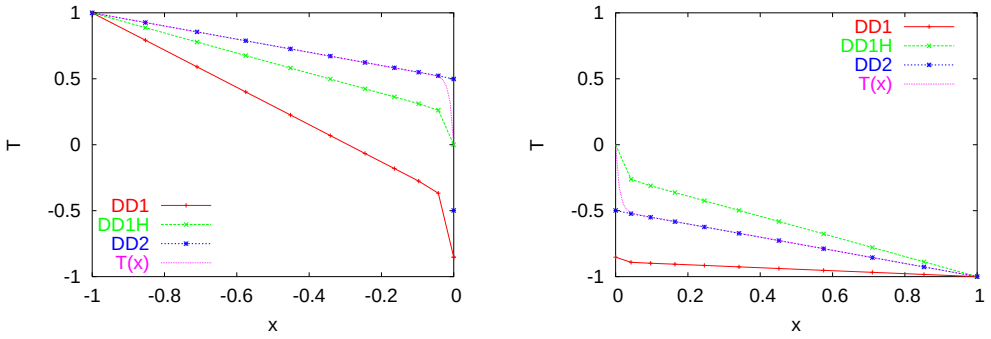

Figure 8.8: Finite element solution obtained using domain decomposition with three subdomains compared to the analytical one

It is clearly seen how the second method gives much better results in the case of a coarse discretization. The accuracy of this approach depends on the choice of the

coefficient α. The optimal value used here is found noting that, when $\gamma \to \infty$, the exact solution tends to

$$\begin{aligned} \vartheta_{\mathrm{F}}(x) &= -\frac{1}{2}x + \frac{1}{2} \\ \vartheta_{\mathrm{S}}(x) &= -\frac{1}{2}x - \frac{1}{2} \end{aligned}$$

and the conductions coefficients tend to 1 (except at $x = 0$ where both are 0) from where we obtain $\alpha = 1/2$.

8.5.2 A fire in a tunnel

A fire is a complex phenomenon whose detailed simulation involves many different aspects that we are not considering in this work. Here we have used a simple model that considers the fire as a source of heat, without taking into account the exact reactive mechanism, as this would imply a precise knowledge of the chemical components of the fuel. The heat released during a fire, which is between 1 MW and 100 MW, is partially dissipated by the flow and partially transported towards the concrete structure where it is finally dissipated. Thus, the heat transfer involves both the behavior of the fluid inside the tunnel and the structural behavior of the concrete and it is therefore necessary to solve a coupled problem.

We solve the problem using the low Mach number approximation to the compressible flow equations. This model takes into account the compressibility of the fluid but removes the acoustic modes [124]. Unlike the Boussinesq approximation, strong temperature and density gradients are allowed. The numerical treatment of the low Mach number equations is described in previous chapters.

The high Reynolds number of the problem implies the need of taking turbulence into account. We do this introducing a Smagorinsky eddy [130], which is defined by

$$\mu^t = \rho c_s \Delta^2 \left[\boldsymbol{\varepsilon}'(\boldsymbol{u}) : \boldsymbol{\varepsilon}'(\boldsymbol{u})\right]^{1/2},$$

where c_s is an empirical constant, Δ a characteristic length usually taken as the mesh size and $\boldsymbol{\varepsilon}'(\boldsymbol{u})$ is the deviatoric part of the rate of deformation tensor. A subgrid thermal conductivity is also added. It is defined in terms of the subgrid viscosity as

$$k^t = \frac{\mu^t c_p}{\mathrm{Pr}^t},$$

where Pr^t is the turbulent Prandtl number, which is assumed to be constant (and taken to be 0.5).

Two simulations were carried out considering heat sources of 10 MW and 30 MW, which correspond to a small size fire (a car for example) distributed in a volume of 8 m^3. Based on experimental results, a typical wind in a tunnel in absence of fire has a velocity

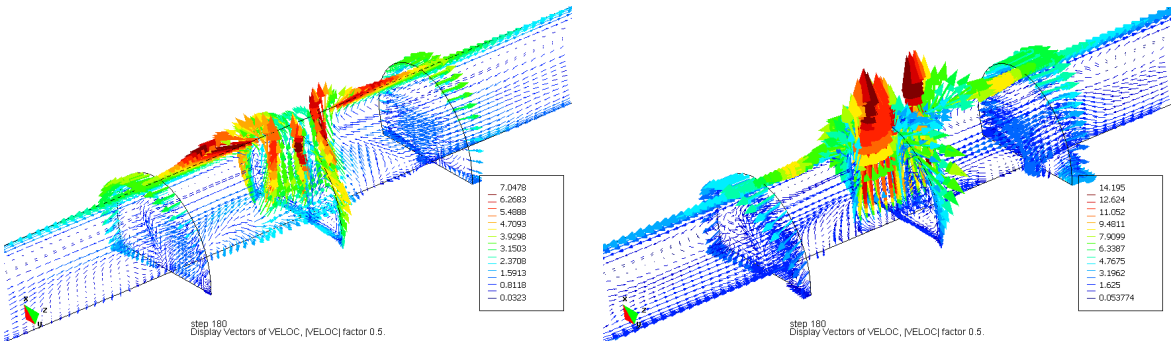

Figure 8.9: Velocity field at t = 180 s for Q = 1.25 MW/m^3 (top) and Q = 4.0 MW/m^3 (bottom)

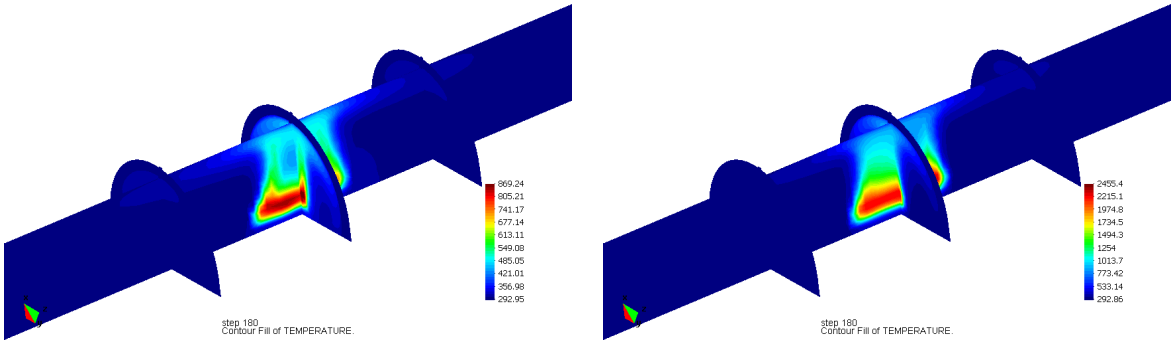

Figure 8.10: Temperature field at t = 180 s for Q = 1.25 MW/m^3 (top) and Q = 4.0 MW/m^3 (bottom)

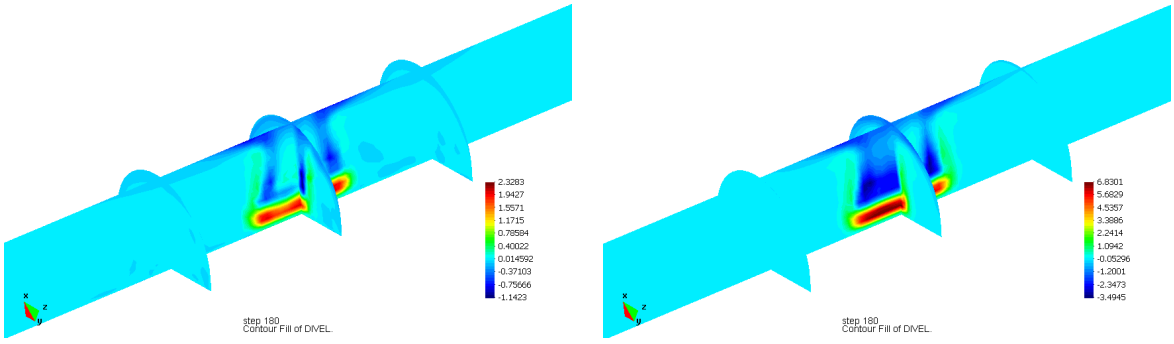

Figure 8.11: Divergence of the velocity field at t = 180 s for Q = 1.25 MW/m^3 (top) and Q = 4.0 MW/m^3(bottom)

of about 0.5 m/s. A preliminary calculation was performed to reproduce the initial state of a wind flowing through the tunnel which was obtained applying a pressure difference between the tunnel inlet and outlet. On the tunnel walls Neumann boundary conditions based on universal profiles were applied (wall laws). Boundary conditions for temperature were defined to reproduce the real situation as close as possible. On the tunnel walls a Robin type condition as in 8.17-8.18 was applied using a convection coefficient suggested by laboratory experiments and the temperature on the concrete walls was fixed. On the entrance and exit of the tunnel Neumann boundary conditions were considered.

The physics of the flow is quite complex and the temporal evolution is chaotic. When the heating starts, strong buoyancy forces determine the formation of a plume and recirculation zones that now, in contrast to the previous example, are fully tridimensional and of complex structure. In Figure 8.9 the velocity field at 3 minutes after the starting of the heating is shown and in Figure 8.10 the corresponding temperature field is shown. Both figures show a detail of the fire zone introducing cutting planes that intersect the fire zone. The heat source generates the plume that can be clearly observed in Figure 8.9, where an expansion of the flow is also apparent. This expansion is better shown in Figure 8.11, where contour lines of divergence of the velocity are shown. They have been obtained by projecting velocity gradients on the finite element space.

In both calculations we used a time step $\delta t = 1$ s. The nonlinear equations describing the flow are solved using two nested loops, an external global loop and internal loops for the momentum equations and for the temperature equation (which is non linear in the low Mach number case because of the dependence of the density on the temperature). The external loop is also used to account for the domain decomposition coupling. A maximum number of 5 iterations in the external loop were performed with a convergence tolerance of 10^{-3} for the velocity and of 10^{-4} for the temperature. In most steps 3 iterations were enough to achieve convergence and only in few steps the temperature residual after 5 iterations was around 0.2×10^{-3} (the velocity residual was always under the prescribed tolerance). The linear system has been solved using a GMRES method [129] preconditioned using an ILUT(nfill,thres) strategy described in [128], where nfill denotes the level of filling and acts as a memory limiter and thres is a threshold for the choice of filling elements and acts as a cpu time limiter. Several combinations of nfill and thres have been tested for the first time step of the problem and it was observed that, as expected, increasing the filling and reducing the threshold reduces the number of GMRES iterations needed to achieve convergence. The optimal compromise depends on the particular problem considered (including mesh size, initial and boundary conditions, etc.). Let us only point out that this method is more efficient for higher time steps. This is shown in figure 8.12, where the residuals after 100 iterations of GMRES as a function of the nfill parameter are shown for a threshold of 10^{-2}.

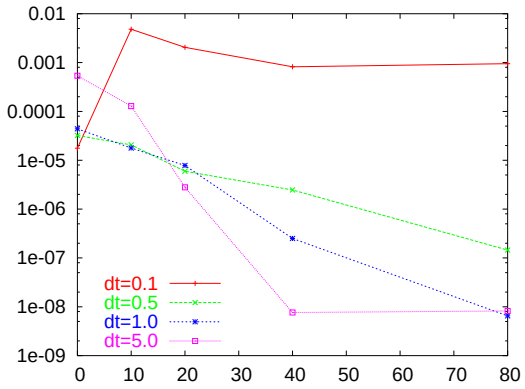

Figure 8.12: GMRES residuals for different time step sizes as a function of the ILUT filling

8.6 Conclusions

In this chapter we have described different aspects related to the numerical approximation of the thermal coupling between a fluid and a solid. Our basic strategy has been to pose the problem in a domain decomposition framework. This has allowed us to propose two alternatives to treat the interface coupling, namely, a classical one considering a perfect thermal contact (continuity of temperatures and heat flux) and another one based on the use of wall functions, which leads to a heat flux proportional to the temperature jump between the fluid and the solid. This surface-convection like transmission condition depends on a coefficient to which we have given an expression in terms of the parameters of the wall function approach. When this coefficient increases the perfect thermal contact condition is recovered.

We have discussed also the iteration-by-subdomain strategy we have implemented using a master-slave strategy. Again, the domain decomposition framework turns out to be crucial to formulate this (otherwise standard) iterative strategy.

From the practical point of view, we have found the algorithmic frame presented here very handful, easy to implement once the basic dedicated codes are available and, what is more important, robust (in accordance with results known from the literature). An application example of the overall formulation has been presented.

Chapter 9

Conclusions

In this chapter we present the conclusions and the possible research lines that could be followed starting from this thesis.

9.1 Achievements

The first objective of this work, to understand the derivation of the simplified models that describe low speed flows as well as the relation between them, has been achieved. The unified asymptotic approach presented in chapter 2 permits us to derive simplified models whose justification was separately known. Using these results we are in position of determining, a priori and in terms of dimensionless parameters, the range of applicability of each model. An important conclusion is that for the kind of problems we have in mind the Boussinesq approximation cannot be used and the low Mach number model is necessary.

The second and most important objective of this work was to develop a subgrid scale stabilized finite element formulation for the kind of problems we are considering. We started from the scalar convection diffusion reaction equation and we arrived to thermally coupled flows. The main conclusions are the following

- In chapter 3 the main contribution is the definition of a new stabilization parameter that improves the robustness of the numerical scheme when the mesh is anisotropic. It was also clearly demonstrated that the numerical scheme obtained using the standard definition with the minimum element length is unstable.

- In chapter 4 the contribution is the extension of results of chapter 3 to systems of equations. Two possible approximations of the solution of the fine scale problem have been proposed. The first one results in the usual diagonal approximation to the stabilization matrix whereas the second one results in a stabilization operator that consists in the usual term plus extra differential terms. The stability analysis

of the resulting scheme obtained has been performed. Numerical experiments for the diagonal approximation show that the OSS method performs much better than the ASGS method as it is much less sensitive to the choice of the element length. They also indicate that defining the stabilization paramters with the minimum element length results in an unstable scheme even for the Stokes problem. Using the maximum element length for this definition numerical oscilation are not found but the result deteriorates when the mesh is anisotropically refined.

- In chapter 5 the main contribution is to consider the subscale time dependent tracking it along the temporal evolution. This apparently natural approach leads to important improvements in the numerical scheme (stability for any time step, commutation of space and time discretizations). Tracking of subscales along nonlinear process permits global conservation of momentum. It also opens the door to the possibility of modelling turbulence.

- In chapter 6 the main contribution is the extension of the ideas of chapter 5 to thermally coupled flow tracking the subscales along the iterative coupling between equations leading to global energy conservation. The possibility of modelling turbulence is particularly appealing in this case due to the performance of turbulence models for thermal problems.

The third objective is to develop a finite element code to solve this problems. Apart from the discrete formulation of the problems, the final ingredient that we need is an algorithm for the solution of the discrete problem. In chapter 7 different linearization strategies are compared and the final algorithm is presented.

The fourth and last objective of this work is to apply the developed code to the problem of thermal coupling of fluids and solids. To achieve this goal, a coupling strategy based on a domain decomposition approach has been developed. This strategy implies the development of a small code to manage the coupling between the solid and the fluid. This development was applied to the problem described of a fire in a tunnel described above. Both the strategy and the application were described in chapter 8. As mentioned in the introduction, the work developed during this theses is the base for the following publications:

- Chapter2: "On the low Mach number and the Boussinesq approximations for low speed flows", J. Principe and R. Codina, Submitted.

- Chapter3: "The modelling of subgrid scales in the finite element approximation of convection diffusion reaction problems on anisotropic meshes", J. Principe and R. Codina, In preparation.

- Chapter 4:"The modelling of subgrid scales in the finite element approximation of incompressible flows", J. Principe and R. Codina, In preparation.
- Chapter 5: "Time dependent subscales in the stabilized finite element approximation of incompressible flow problems", R. Codina, J. Principe, O. Guasch and S. Badia, *Computer Methods in Applied Mechanics and Engineering*, 196 (2007), 2413-2430.
- Chapter 6: "Dynamic subscales in the finite element approximation of thermally coupled incompressible flows", R. Codina and J. Principe, *International Journal for Numerical Methods in Fluids*, 54 (2007), 707-730.
- Chapter 7: "A stabilized finite element approximation of low speed thermally coupled flows", J. Principe and R. Codina, *International Journal of Numerical Methods for Heat & Fluid Flow*, Accepted.
- Chapter 8: "A numerical approximation of the thermal coupling of fluids and solids", J. Principe and R. Codina, Submitted.

9.2 Future work

Several research lines emerge from this thesis. In the first important line is the modelling of the subgrid scales, what has been the main subject of this thesis but that is still not mature. As mentioned in chapters 3 and 4, three steps can be followed to build such a model

1. uncoupling of the fine scale problem into local (element) problems
2. approximation of the differential operator
3. choice of the space of subscales

The contribution of chapters 3 and 4 focus on step 2 and further research is needed to numerically evaluate the second approximation proposed for the Oseen problem. Stability of the resulting scheme has been shown but numerical results were only presented for the diagonal approximation. The better performance of the OSS method over the ASGS one has been shown but still several points need further examination such as, for example, the definition of the instability direction. Another point to investigate is the comparison of the projections $\widetilde{P}_\tau$ (the τ weighted projection) and $P_h^\perp$. Actually, the choice of the space of subscales could be performed a priori, that is, before step 1 and in this case the choice would influence the locality of the fine scale problem. In [83] it is shown that the $H_0^1(\Omega)$ projection leads to a local problem without any further approximation. A particular choice of the space of subscales is the space of bubble functions that also introduce a particular

projection. The equivalence of the variational multiscale method and the bubble function stabilization has been shown in [16] for the steady state problem. The concept of transient subgrid scale gives rise to the concept of transient bubble, something that, surprisingly, has not been considered up to date. When bubble functions are used, several terms defined as integrals over the element boundaries vanish. This is not the case if a general subgrid function is used although these terms are usually neglected. Considering these terms implies a definition of the subscale on the element boundaries something that we are currently investigating.

The second important line is the evaluation of the subgrid models for the solution of turbulent flow problems using the incompressible Navier Stokes equations as a continuous model without the use of a physical model for the turbulence effects. This line is intimately related to the previous one because the requirements the model should satisfy, detailed in chapter 5, can depend on the modelling of the fine scale problem. We already started research in this direction analyzing the dissipative structure of the orthogonal subscale model and showing it is capable of predicting backscatter [125]. Certainly much more numerical evidence is needed to analyze the performance of numerical methods in the solution of turbulent flow problem, specially regarding the prediction of mean flow features such as approximation of boundary layers, separation, etc. However we call the attention of the reader to other (more qualitative) features as well, such as the dimension of the global attractor or the prediction of transition and relaminarization. Finally let us also mention the potential of this approach for the analysis of thermal turbulence.

The third important line is the application of the techniques developed in this thesis to the class of problems that motivated it, the problem of simulation of fires. Although the physical mechanisms involved in a fire a really complex, a simulation with some degree of accuracy needs at least three ingredients. The first one is the simulation thermal turbulence that must be based on the solution of the low Mach number equations and not on the solution of the Boussinesq model. Needless to say that this point is related to the second research line mentioned above. The second ingredient is the simulation of the combustion process and third one the simulation of the radiative heat flux . The last two problems have not been even considered in this work and certainly need research.

Bibliography

[1] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics, A Wiley-Interscience Series of Texts, Monographs and Tracts. John Wiley & Sons, 2000. 45

[2] T. Apel and G. Lube. Anisotropic mesh refinement for a singularly perturbed reaction diffusion model problem. *Applied Numerical Mathematics*, 26:415–433, 1998. Available from: `http://dx.doi.org/10.1016/S0168-9274(97)00106-2`. 57, 87

[3] S. Badia and R. Codina. On a multiscale approach to the transient Stokes problem. Transient subscales and anisotropic space-time discretization. Submitted. Available from: `http://www.rmee.upc.es/homes/codina/`. 113

[4] C. Baiocchi, F. Brezzi, and L. Franca. Virtual bubbles and Galerkin/least-squares type methods (Ga.L.S). *Computer Methods in Applied Mechanics and Engineering*, 105:125–141, 1993. Available from: `http://dx.doi.org/10.1016/0045-7825(93)90119-I`. 40, 106, 130

[5] G. K. Batchelor. The conditions for dynamical similarity of motions of a frictionless perfect gas atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 79:224–235, 1953. 21, 29

[6] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, Cambridge, UK, 1967. 21

[7] S. E. Bechtel, M. G. Forest, F. J. Rooney, and Q. Wang. Thermal expansion models of viscous fluids based on limits of free energy. *Physics of Fluids*, 15(9):2681–2693, 2003. Available from: `http://link.aip.org/link/?PHF/15/2681/1`. 20

[8] R. Becker. *An Adaptive Finite Element Method for the Incompressible Navier-Stokes Equations on Time-dependent Domains*. PhD thesis, University of Heidelberg, 1995. 74, 75, 101

[9] R. Becker and R. Rannacher. Finite element solution of incompresible flow Navier-Stokes equations on anisotropically refined meshes. *Notes on Numerical Fluid Mechanics*, 49, 1995. 74, 75, 101

[10] J. Blasco. An anisotropic pressure-stabilized finite element method for incompressible flow problems. *Computers and Mathematics with Applications*, 53(6):895–909, Mar 2007. Available from: `http://dx.doi.org/10.1016/j.camwa.2006.12.022`. 74, 75, 101

[11] P. Bochev, M. Gunzburger, and R. Lehoucq. On stabilized finite element methods for the Stokes problem in the small time-step limit. *International Journal for Numerical Methods in Fluids*, 53:573–597, 2007. Available from: `http://dx.doi.org/10.1002/fld.1295`. 112, 116, 118

[12] P. A. Bois. Propagation linéare et non linéaire d'ondes atmosphériques. *Journal de Mécanique*, 15(5):781–811, 1976. 20, 21, 29, 31, 32, 33, 34, 35, 38

[13] P. A. Bois. Asymptotic aspects of the Boussinesq approximation for gases and liquids. *Geophysical and Astrophysical Fluid Dynamics*, 58:45–55, 1991. 21, 29, 31, 32, 33, 34, 35, 38

[14] J. Boussinesq. *Théorie analytique de la chaleur*, volume 2. Gauthier-Villars, Paris (France), 1903. Reproduction Bibliothèque Nationale de France, 1995. Available from: `http://gallica.bnf.fr`. 20

[15] F. Brezzi, M. Bristeau, L. Franca, M. Mallet, and G.Rogé. A relationship between stabilized finite element methods and the Galerkin method with bubble functions. *Computer Methods in Applied Mechanics and Engineering*, 96:117–129, 1992. Available from: `http://dx.doi.org/10.1016/0045-7825(92)90102-P`. 40

[16] F. Brezzi, L. P. Franca, T. J. R. Hughes, and A. Russo. $b = \int g$. *Computer Methods in Applied Mechanics and Engineering*, 145(3-4):329–339, Jun 1997. Available from: `http://dx.doi.org/10.1016/S0045-7825(96)01221-2`. 40, 106, 196

[17] F. Brezzi and A. Russo. Choosing bubbles for advection diffusion problems. *Mathematical Models and Methods in Applied Sciences*, 4(4):571–587, 1994. 40

[18] A. N. Brooks and T. J. R. Hughes. Streamline upwind/petrov-galerkin formulations for convection dominated flows with particular emphasis on the incompresible Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 32:199–259, 1982. Available from: `http://dx.doi.org/10.1016/0045-7825(82)90071-8`. 14, 39, 107

[19] E. Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Physical Review Series II*, 4(4):345–376, 1914. Available from: http://dx.doi.org/10.1103/PhysRev.4.345. 22

[20] E. Burman and M. Fernández. A finite element method with edge oriented stabilization for the time-dependent Navier-Stokes equations: space discretization and convergence. Submitted. Available from: http://www.inria.fr/rrrt/rr-5630.html. 110

[21] V. Calo. *Residual based mulstiscale turbulence modeling: finite volume simulations of bypass transition*. PhD thesis, Department of Civil and Environmental Engineering, Stanford University, 2004. Available from: http://users.ices.utexas.edu/~victor/. 51, 108, 111, 115

[22] T. Chacón Rebollo and E. Chacón Vera. Study of a non-overlapping domain decomposition method: Poisson and Stokes problems. *Applied Numerical Mathematics*, 48:169–194, 2004. Available from: http://dx.doi.org/10.1016/j.apnum.2003.06.003. 185

[23] D. R. Chenoweth and S. Paolucci. Natural convection in an enclosed vertical air layer with large horizontal temperature differences. *Journal of fluid mechanics*, 169:173–210, 1986. Available from: http://dx.doi.org/10.1017/S0022112086000587. 24, 36, 158

[24] R. Codina. Comparison of some finite element methods for solving the diffusion-convection-reaction equation. *Computer Methods in Applied Mechanics and Engineering*, 156:185–210, 1998. Available from: http://dx.doi.org/10.1016/S0045-7825(97)00206-5. 14, 40, 48, 49, 80

[25] R. Codina. On stabilized finite element methods for linear systems of convection-diffusion-reaction equations. *Computer Methods in Applied Mechanics and Engineering*, 188(1-3):61–82, Jul 2000. Available from: http://dx.doi.org/10.1016/S0045-7825(00)00177-8. 14

[26] R. Codina. Stabilization of incompressibility and convection through orthogonal sub-scales in finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 190(13-14):1579–1599, Dec 2000. Available from: http://dx.doi.org/10.1016/S0045-7825(00)00254-1. 75

[27] R. Codina. Pressure stability in fractional step finite element methods for incompressible flows. *Journal of Computational Physics*, 170:112–140, 2001. Available from: http://dx.doi.org/10.1006/jcph.2001.6725. 116

[28] R. Codina. A stabilized finite element method for generalized stationary incompressible flows. *Computer Methods in Applied Mechanics and Engineering*, 190(20-21):2681–2706, Feb 2001. Available from: `http://dx.doi.org/10.1016/S0045-7825(00)00254-1`. 14, 61, 75, 87, 105, 108, 130

[29] R. Codina. Stabilized finite element approximation of transient incompressible flows using orthogonal subscales. *Computer Methods in Applied Mechanics and Engineering*, 191(39-40):4295–4321, Aug 2002. Available from: `http://dx.doi.org/10.1016/S0045-7825(02)00337-7`. 45, 46, 48, 69, 75, 79, 80, 86, 105, 108, 110, 115, 122, 130, 131, 132, 181, 182, 183

[30] R. Codina and J. Blasco. A finite element formulation for the Stokes problem allowing equal velocity-pressure interpolation. *Computer Methods in Applied Mechanics and Engineering*, 143:373–391, 1997. Available from: `http://dx.doi.org/10.1016/S0045-7825(96)01154-1`. 75, 87, 105

[31] R. Codina and J. Blasco. Analysis of a pressure-stabilized finite element approximation of the stationary Navier-Stokes equations. *Numerische Mathematik*, 87:59–81, 2000. Available from: `http://dx.doi.org/10.1007/s002110000174`. 105

[32] R. Codina and J. Blasco. Analysis of a stabilized finite element approximation of the transient convection-diffusion-reaction equation using orthogonal subscales. *Computing and Visualization in Science*, 4:167–174, 2002. Available from: `http://dx.doi.org/10.1007/s007910100068`. 112

[33] R. Codina and G. Houzeaux. Numerical approximation of the heat transfer between domains separated by thin walls. *International Journal of Numerical Methods in Fluids*, 52:963–986, 2006. Available from: `http://dx.doi.org/10.1002/fld.1215`. 180

[34] R. Codina, E. Oñate, and M. Cervera. The intrinsic time for the streamline upwind / Petrov-Galerkin formulation using quadratic elements. *Computer Methods in Applied Mechanics and Engineering*, 94:239–262, 1992. Available from: `http://dx.doi.org/10.1016/0045-7825(92)90149-E`. 53

[35] R. Codina and J. Principe. Dynamic subscales in the finite element approximation of thermally coupled incompressible flows. *International journal for numerical methods in fluids*, 54(6-8):707–730, 2007. Available from: `http://dx.doi.org/10.1002/fld.1481`. 181, 182

[36] R. Codina, J. Principe, O. Guasch, and S. Badia. Time dependent subscales in the stabilized finite element approximation of incompressible flow problems.

Computer Methods in Applied Mechanics and Engineering, 196(21-24):2413–2430, 2007. Available from: `http://dx.doi.org/10.1016/j.cma.2007.01.002`. 113, 127, 132, 181, 182, 183

[37] R. Codina and O. Zienkiewicz. CBS versus GLS stabilization of the incompressible Navier-Stokes equations and the role of the time step as stabilization parameter. *Communications in Numerical Methods in Engineering*, 18:99–112, 2002. 134, 144

[38] G. D. V. Davis and I. Jones. Natural convection in a square cavity: a comparison exercise. *International Journal for numerical methods in fluids*, 3:227–248, 1983. 158

[39] P. de Sampaio, P. Hallak, A. Coutinho, and M. Pfeil. A stabilized finite element procedure for turbulent fluid-structure interaction using adaptive time-space refinement. *International Journal for Numerical Methods in Fluids*, 44:673–693, 2004. 115

[40] J. Dennis and R. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall series in computational mathematics. Prentice-Hall, New Jersey, first edition, 1983. 154, 155

[41] J. Donea. A Taylor-Galerkin method for convection transport problems. *International Journal for Numerical Methods in Engineering*, 20:101–119, 1984. 40

[42] J. Douglas and T. Russel. Numerical methods for convection dominated problems based on combining the method of characteristics with finite elements or finite difference procedures. *SIAM journal of numerical analysis*, 19:871–885, 1982. 40

[43] J. Douglas and J. Wang. An absolutely stabilized finite element method for the Stokes problem. *Mathematics of computation*, 52:495–508, 1989. 40

[44] J. A. Dutton and G. H. Fichtl. Approximate equations of motion for gases and liquids. *Journal of the atmospheric sciences*, 26:241–253, 1969. 21, 29, 35

[45] A. ElSheikh, S. Chidiac, and W. Smith. A posteriori error estimation based on numerical realization of the variational multiscale method. *Computer Methods in Applied Mechanics and Engineering*, In Press, 2008. 45

[46] G. Evans and S. Paolucci. The thermoconvective instability of plane Poiseuille flow heated from below: A proposed benchmark solution for open boundary flows. *International Journal for Numerical Methods in Fluids*, 11:1001–1013, 1990. 137, 138

[47] A. T. Fedorchenko. A model of unsteady subsonic flow with acoustics excluded. *Journal of Fluid Mechanics*, 334:135–155, 1997. 20

[48] L. Formaggia, S. Micheletti, and S. Perotto. Anisotropic mesh adaptation in computational fluid dynamics: Application to the advection-diffusion-reaction and the stokes problems. *Applied Numerical Mathematics*, 51(4):511–533, Dec 2004. Available from: http://dx.doi.org/10.1016/j.apnum.2004.06.007. 41

[49] L. Franca and C. Farhat. Bubble functions prompt unusual stabilized finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 123:299–308, 1995. Available from: http://dx.doi.org/10.1016/0045-7825(94)00721-X. 40

[50] L. Franca, S. Frey, and T. J. R. Hughes. Stabilized finite element methods: I. Application to the advective-diffusive model. *Computer Methods in Applied Mechanics and Engineering*, 95:253–276, 1992. Available from: http://dx.doi.org/10.1016/0045-7825(92)90143-8. 40

[51] L. Franca, G. Hauke, and A. Masud. Revisiting stabilized finite element methods for the advective-diffusive equation. *Computer Methods in Applied Mechanics and Engineering*, 195(13-16):1560–1572, Feb 2006. Available from: http://dx.doi.org/10.1016/j.cma.2005.05.028. 40

[52] L. Franca and A. Macedo. A two-level finite element method and its application to the Helmholtz equation. *International Journal for Numerical Methods in Engineering*, 43:23–32, 1998. 47

[53] L. Franca, A. Nesliturk, and M. Stynes. On the stability of residual-free bubbles for convection-diffusion problems and their approximation by a two-level finite element method. *Computer Methods in Applied Mechanics and Engineering*, 166(1-2):35–49, 1998. Available from: http://dx.doi.org/10.1016/S0045-7825(98)00081-4. 47

[54] L. Franca and A. Russo. Deriving upwinding, mass lumping and selective reduced integration by residual-free bubbles. *Applied Mathematics Letters*, 9(5):253–276, 1996. 40

[55] L. Franca and F. Valentin. On an improved unusual stabilized finite element method for the advective-reactive-diffusive equation. *Computer Methods in Applied Mechanics and Engineering*, 190(13-14):1785–1800, Dec 2000. Available from: http://dx.doi.org/10.1016/S0045-7825(00)00190-0. 40

[56] K. Gage and W. Reid. The stability of thermally stratified plane Poiseuille flow. *Journal of Fluid Mechanics*, 33:21–32, 1968. 138

[57] D. Gawin, F. Pesavento, and B. A. Schrefler. Simulation of damage-permeability coupling in hygro-thermo-mechanical analysis of concrete at high temperature. *Communications in numerical methods in engineering*, 18:113–119, 2002. 170, 172

[58] T. Gelhard, G. Lube, M. Olshanskii, and J. Starcke. Stabilized finite element schemes with LBB-stable elements for incompressible flows. *Journal of Computational and Applied Mathematics*, 177:243–267, 2005. 111, 112

[59] J. Gibbon and E. Titi. Attractor dimension and small length scale estimates for the three dimensional Navier-Stokes equations. *Nonlinearity*, 10:109–119, 1997. 115

[60] M. Giles. Stability analysis of numerical interface conditions in fluid-structure thermal analysis. *International Journal for Numerical Methods in Fluids*, 25:421–436, 1997. 184

[61] D. O. Gough. The anelastic approximation for thermal convection. *Journal of the atmospheric sciences*, 26:448–456, May 1969. 21, 29, 35

[62] V. Gravemeier. The variational mulstiscale method for laminar and turbulent flow. *Archives of Computational Mechanics–State of the Art Reviews*, 13:249–324, 2006. 115

[63] D. D. Gray and A. Giorgini. The validity of the Boussinesq approximation for liquids and gases. *International Journal of Heat and Mass Transfer*, 19(5):545–551, 1976. 20, 28, 37

[64] J. Guermond. Finite-element-based Faedo-Galerkin weak solutions to the Navier-Stokes equations in the three-dimensional torus are suitable. *Journal de Mathématiques Pures et Appliquées*, To appear. 115

[65] J. Guermond, J. Oden, and S. Prudhomme. Mathematical perspectives on large eddy simulation models for turbulent flows. *Journal of Mathematical Fluid Mechanics*, 6:194–248, 2004. 115

[66] J. Guermond and S. Prudhomme. On the construction of suitable solutions to the Navier-Stokes equations and questions regarding the definition of large eddy simulation. *Physica D*, 207:64–78, 2005. 115

[67] I. Harari and T. J. R. Hughes. What are c and h?: Inequalities for the analysis and design of finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 97(2):157–192, Jun 1992. Available from: `http://dx.doi.org/10.1016/0045-7825(92)90162-D`. 40, 49, 58, 67, 88

[68] G. Hauke. A simple subgrid scale stabilized method for the advection-diffusion-reaction equation. *Computer Methods in Applied Mechanics and Engineering*, 191(27-28):2925–2947, 2002. Available from: `http://dx.doi.org/10.1016/S0045-7825(02)00217-7`. 42

[69] G. Hauke, M. H. Doweidar, and M. Miana. The multiscale approach to error estimation and adaptivity. *Computer Methods in Applied Mechanics and Engineering*, 195(13-16):1573–1593, 2006. Available from: `http://dx.doi.org/10.1016/j.cma.2005.05.029`. 45

[70] G. Hauke, D. Fuster, and M. Doweidar. The multiscale approach to error estimation and adaptivity. *Computer Methods in Applied Mechanics and Engineering*, 2008. Available from: `http://dx.doi.org/10.1016/j.cma.2007.12.022`. 45

[71] G. Hauke and A. Garcia-Olivares. Variational subgrid scale formulations for the advection-diffusion-reaction equation. *Computer Methods in Applied Mechanics and Engineering*, 190(51-52):6847–6865, 2001. Available from: `http://dx.doi.org/10.1016/S0045-7825(01)00262-6`. 42

[72] V. Heuveline. On higher-order mixed fem for low Mach number flows: application to a natural convection benchmark problem. *International Journal for Numerical Methods in Fluids*, 41:1339–1356, 2003. 159

[73] J. Hoffman and C. Johnson. A new approach to computational turbulence modeling. *Computer Methods in Applied Mechanics and Engineering*, 195:2865–2880, 2006. 115

[74] G. Houzeaux. *A geometrical domain decomposition method in computational fluid dynamics*. PhD thesis, Escola Tècnica Superior d'Enginyers de Camins, Canals i Ports, Universitat Politècnica de Catalunya, Barcelona, 2002. 184, 186

[75] T. J. R. Hughes. Multiscale phenomena, Green's functions, the Dirichlet-to-Neuman formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Computer Methods in Applied Mechanics and Engineering*, 127:387–401, 1995. Available from: `http://dx.doi.org/10.1016/0045-7825(95)00844-9`. 14, 40, 41, 47, 48, 74, 75, 80, 105, 150, 181

[76] T. J. R. Hughes and A. N. Brooks. A multidimensional upwind scheme with no crosswind diffusion. In T. J. R. Hughes, editor, *FEM for convection dominated flows*, New York, 1979. ASME. 14, 39, 51

[77] T. J. R. Hughes, V. M. Calo, and G. Scovazzi. Variational and multiscale methods in turbulence. In W. Gutkowski and T. Kowalewski, editors, *Proceedings of the XXI international Congress of Theoretical and Applied Mechanics (IUTAM)*. Kluwer, 2004. 115

[78] T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J. Quincy. The variational multiscale method–a paradigm for computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 166:3–24, 1998. Available from: `http:`

//dx.doi.org/10.1016/S0045-7825(98)00079-6. 14, 40, 41, 44, 45, 46, 74, 75, 105, 127, 130

[79] T. J. R. Hughes, L. P. Franca, and M. Balestra. A new finite element formulation for computational fluid dynamics: V. Circumventing the Babuska-Brezzi condition: a stable Petrov-Galerkin formulation of the stokes problem accommodating equal-order interpolations. *Computer Methods in Applied Mechanics and Engineering*, 59(1):85–99, Nov 1986. Available from: http://dx.doi.org/10.1016/0045-7825(86)90025-3. 14

[80] T. J. R. Hughes, L. P. Franca, and G. M. Hulbert. A new finite element formulation for computational fluid dynamics: VIII. The galerkin/least-squares method for advective-diffusive equations. *Computer Methods in Applied Mechanics and Engineering*, 73(2):173–189, May 1989. Available from: http://dx.doi.org/10.1016/0045-7825(89)90111-4. 40

[81] T. J. R. Hughes and M. Mallet. A new finite element formulation for computational fluid dynamics: III. the generalized streamline operator for multidimensional advective-diffusive systems. *Computer Methods in Applied Mechanics and Engineering*, 58(3):305–328, Nov 1986. Available from: http://dx.doi.org/10.1016/0045-7825(86)90152-0. 41, 46

[82] T. J. R. Hughes, L. Mazzei, and K. E. Jansen. Large eddy simulation and the variational multiscale method. *Computing and Visualization in Science*, 3:47–59, 2000. 114

[83] T. J. R. Hughes and G. Sangalli. Variational multiscale analysis: the fine-scale Green's function, projection, optimization, localization, and stabilized methods. *SIAM Journal on Numerical Analysis*, To appear. 107, 195

[84] T. J. R. Hughes and G. N. Wells. Conservation properties for the galerkin and stabilised forms of the advection-diffusion and incompressible navier-stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 194(9-11):1141–1159, Mar 2005. Available from: http://dx.doi.org/10.1016/j.cma.2004.06.034. 113, 135, 136

[85] S. Idelsohn, N. Nigro, M. Storti, and G. Buscaglia. A Petrov-Galerkin formulation for advection-reaction-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 136:27–46, 1996. Available from: http://dx.doi.org/10.1016/0045-7825(96)01008-0. 111

[86] F. Incropera and D. Dewitt. *Fundamentals of Heat and Mass Transfer, Fourth edition.* John Wiley & Sons, 1996. 169

[87] K. Jansen, S. Collis, C. Whiting, and F. Shakib. A better consistency for low-order stabilized finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 174:153–170, 1999. 110

[88] K. Jansen, C. Whiting, and G. Hulbert. A generalized-α method for integrating the filtered Navier-Stokes equations with a stabilized finite element method. *Computer Methods in Applied Mechanics and Engineering*, 190:305–319, 2000. 110

[89] C. L. V. Jayatilleke. *The Influence of Prandtl Number and Surface Roughness on the Resistance of Laminar Sub-layer to Momentum and Heat Transfer*, volume 1 of *Progress in heat transfer*. Pergamon Press, Oxford, 1969. 178

[90] K. Jensen, E. Einset, and D. Fotiadis. Flow phenomena in chemical vapor deposition of thin films. *Annual review of fluid mechanics*, 23:197–233, 1991. 157

[91] C. Johnson, U. Nävert, and J. Pitkäranta. Finite element methods for linear hyperbolic equations. *Computer Methods in Applied Mechanics and Engineering*, 45:285–312, 1984. 39, 109

[92] C. T. Kelley. *Iterative methods for optimization*. SIAM, Philadelphia, 1999. 154, 155

[93] D. Kelly, S. Nakazawa, O. Zienkiewicz, and J. Heinrich. A note on upwinding and anisotropic balancing dissipation in finite element approximations to convective diffusion problems. *International journal for numerical methods in engineering*, 15:1705–1711, 1980. 14, 39, 51

[94] R. Kessler. Nonlinear transition in three-dimensional convection. *Journal of Fluid Mechanics*, 174:359–379, 1987. 138

[95] J. Kevorkian and J. D. Cole. *Perturbation Methods in Applied Mathematics*, volume 34 of *Applied mathematical sciences*. Springer-Verlag, New York, 1981. 25

[96] S. Klainerman and A. Majda. Compressible and incompressible fluids. *Communications on Pure and Applied Mathematics*, XXXV:629–651, 1982. 19

[97] R. Klein. Semi-implicit extension of a Godunov-type scheme based on low mach number asymptotics I: One-dimensional flow. *Journal of Computational Physics*, 121(2):213–237, Oct 1995. Available from: `http://dx.doi.org/10.1016/S0021-9991(95)90034-9`. 20, 25

[98] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, 1999. 154

[99] E. L. Koschmieder. *Bénard cells and Taylor vortices*. Cambridge monographs on mechanics and applied mathematics. Cambridge University Press, Cambridge, UK, 1993. 24

[100] V. I. Kovshov. Three approximations for thermal convection. *Soviet Astronomy*, 22:288–296, June 1978. 35

[101] L. Landau and E. Lifshitz. *Fluid Mechanics*, volume 6 of *Course of Theoretical Physics*. Pergamon Press, first edition, 1975. 36

[102] P. Le Quéré and M. Behnia. From onset to unsteadiness to chaos in a differentially heated cavity. *Journal of fluid mechanics*, 359:81–107, 1998. 24

[103] P. Le Quéré and P. Paillère. Modelling and simulation of natural flows with large temperature differences: a benchmark problem for low Mach number solvers. In *Workshop/12th CFD seminar*, pages 14–22, France, 2000. CEA Saclay. 158

[104] P. L. Lions. *Mathematical Topics in Fluid Dynamics, Volume 1. Incompressible Models*. Oxford University Press, 1996. 13, 21, 27

[105] P. L. Lions. *Mathematical Topics in Fluid Dynamics, Volume 2. Compressible Models*. Oxford University Press, 1998. 19

[106] A. Majda and J. Sethian. The derivation and numerical solution of the equations for zero Mach number combustion. *Combustion science and technology*, 42:185–205, 1985. 20

[107] M. Martinez and D. Gartling. A finite element method for low-speed compressible flows. *Computer Methods in Applied Mechanics and Engineering*, 193(21-22):1959–1979, May 2004. Available from: `http://dx.doi.org/10.1016/j.cma.2003.12.049`. 157, 162, 163, 166

[108] K. B. McGrattan, H. R. Baum, R. G. Rehm, A. Hamins, G. P. Forney, J. E. Floyd, and S. Hostikka. Fire dynamic simulator (version 2) - technical reference guide. Internal report, National Institute of Standards and Technology, November 2001. 16

[109] A. Meister. Asymptotic single and multiple scale expansions in the low Mach number limit. *SIAM Journal on Applied Mathematics*, 60(1):256–271, 1999. 25

[110] J. M. Mihaljan. A rigorous exposition of the Boussinesq approximations applicable to a thin layer of fluid. *Astrophysical Journal*, 136(3):1126–1133, 1962. Available from: `http://adsabs.harvard.edu/`. 20, 35, 36

[111] S. Mittal. On the performance of high aspect ratio elements for incompressible flows. *Computer Methods in Applied Mechanics and Engineering*, 188:269–287, 2000. 41, 53

[112] A. Mohamad and R. Viskanta. Transient natural convection of low-Prandtl-number fluids in a differentially heated cavity. *International Journal for Numerical Methods in Fluids*, 13:61–81, 1991. 141

[113] B. Müller. Low Mach number asymptotics of the Navier-Stokes equations. *Journal of Engineering Mathematics*, 34:97–109, 1998. 20, 25, 28

[114] X. Nicolas. Bibliographical review on the Poiseuille-Rayleigh-Bénard flows: the mixed convection flows in horizontal rectangular ducts heated from below. *International Journal of Thermal Sciences*, 41(10):961–1016, Oct 2002. Available from: `http://dx.doi.org/10.1016/S1290-0729(02)01374-1`. 24, 157

[115] Y. Ogura and N. A. Phillips. Scale analysis of deep and shallow convection in atmosphere. *Journal of the atmospheric sciences*, 19:173–179, 1962. 21, 29, 36

[116] E. Oñate. Derivation of stabilized equations for numerical solution of advective-diffusive transport and fluid flow problems. *Computer Methods in Applied Mechanics and Engineering*, 151(1-2):233–265, Jan 1998. Available from: `http://dx.doi.org/10.1016/S0045-7825(97)00119-9`. 41

[117] E. Oñate, J. García, and S. Idelsohn. Computation of the stabilization parameter for the finite element solution of advective-diffusive problems. *International Journal for Numerical Methods in Fluids*, 25(12):1385–1407, 1997. 41

[118] E. Oñate, F. Zarate, and S. R. Idelsohn. Finite element formulation for convective-diffusive problems with sharp gradients using finite calculus. *Computer Methods in Applied Mechanics and Engineering*, 195(13-16):1793–1825, Feb 2006. Available from: `http://dx.doi.org/10.1016/j.cma.2005.05.036`. 41

[119] S. Paolucci. On the filtering of sound from the Navier-Stokes equations. Technical Report 82-8257, Sandia National Laboratories, 1982. 20, 21, 28, 35

[120] R. Perez Cordon and M. G. Velarde. On the (non-linear) foundations of Boussinesq approximation applicable to a thin-layer of fluid. *Journal de physique*, 36(7):591–601, 1975. 20, 28, 35, 37

[121] M. Polner, J. van der Vegt, and R. van Damme. Analysis of stabilization operators for galerkin least-squares discretizations of the incompressible navier-stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 195(9-

12):982–1006, Feb 2006. Available from: http://dx.doi.org/10.1016/j.cma.2005.02.020. 74

[122] S. Pope. *Turbulent Flows*. Cambridge University Press, 2000. 114

[123] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, 2nd edition, 1992. 154, 155

[124] J. Principe and R. Codina. On the the low Mach number and the Boussinesq approximations for low speed flows. Submitted. 189

[125] J. Principe, R. Codina, and F. Henke. The dissipative structure of variational multiscale methods for incompressible flows. *Submitted to Computer Methods in Applied Mechanics and Engineering*, 2008. 196

[126] R. G. Rehm and H. R. Baum. The equations of motion for thermally driven bouyant flows. *Journal of research of the National Bureau of Standards*, 83(3):297–308, 1978. 20, 21, 26, 28

[127] B. Roe, A. Haselbacher, and P. Geubelle. Stability of fluid-structure thermal simulations on moving grids. *International Journal for Numerical Methods in Fluids*, 54:1097–1117, 2007. 184

[128] Y. Saad. Ilut: a dual threshold incomplete ilu factorization. *Numerical linear algebra with applications*, 1:387–402, 1994. 191

[129] Y. Saad. *Iterative methods for saprse linear systems*. PWS Publishing Company, Boston, 1st edition, 1996. 191

[130] P. Sagaut. *Large eddy simulation for incompressible flows*. Scientific Computing, Springer, 2001. 114, 189

[131] B. A. Schrefler, P. Brunello, D. Gawin, C. E. Majorana, and F. Pesavento. Concrete at high temperature with application to tunnel fire. *Computational Mechanics*, 29:43–51, 2002. 15, 16

[132] F. Shakib and T. J. R. Hughes. A new finite element formulation for computational fluid dynamics: IX. Fourier analysis of space-time galerkin/least-squares algorithms. *Computer Methods in Applied Mechanics and Engineering*, 87(1):35–58, May 1991. Available from: http://dx.doi.org/10.1016/0045-7825(91)90145-V. 109, 110, 111, 112, 135

[133] F. Shakib, T. J. R. Hughes, and Z. Johan. A new finite element formulation for computational fluid dynamics: X. the compressible euler and navier-stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 89(1-3):141–219, Aug 1991. Available from: `http://dx.doi.org/10.1016/0045-7825(91)90041-4`. 41, 46, 51

[134] S. A. Slimon, M. C. Soteriou, and D. W. Davis. Development of computational aeroacoustics equations for subsonic flows using a mach number expansion approach. *Journal of Computational Physics*, 159(2):377–406, Apr 2000. Available from: `http://dx.doi.org/10.1006/jcph.2000.6449`. 37

[135] E. A. Spiegel and G. Veronis. On the Boussinesq approximation for a compressible fluid. *Astrophysical Journal*, 131:442–447, 1960. Available from: `http://adsabs.harvard.edu/`. 20

[136] S. A. Suslov and S. Paolucci. Nonlinear stability of mixed convection flow under non-Boussinesq conditions. Part 1. Analysis and bifurcations. *Journal of Fluid Mechanics*, 398:61–86, Nov 1999. 36

[137] C. A. Taylor, T. J. R. Hughes, and C. K. Zarins. Finite element modeling of blood flow in arteries. *Computer Methods in Applied Mechanics and Engineering*, 158(1-2):155–196, May 1998. Available from: `http://dx.doi.org/10.1016/S0045-7825(98)80008-X`. 51

[138] T. Tezduyar and S. Sathe. Stabilization parameters in SUPG and PSPG formulations. *Journal of Computational and Applied Mechanics*, 4:71–88, 2003. 111, 112, 135

[139] T. E. Tezduyar. Computation of moving boundaries and interfaces and stabilization parameters. *International Journal for Numerical Methods in Fluids*, 43:555–575, 2003. 41

[140] T. E. Tezduyar, S. Mittal, S. E. Ray, and R. Shih. Incompressible flow computations with stabilized bilinear and linear equal-order-interpolation velocity-pressure elements. *Computer Methods in Applied Mechanics and Engineering*, 95(2):221–242, 1992. 41

[141] M. Van Dyke. *Perturbation methods in fluid mechanics*, volume 8 of *Applied mathematics and mechanics*. Academic Press, New York, 1964. 37

[142] T. Warburton and J. S. Hesthaven. On the constants in hp-finite element trace inverse inequalities. *Computer Methods in Applied Mechanics and Engineering*, 192(25):2765–2773, Jun 2003. Available from: `http://dx.doi.org/10.1016/S0045-7825(03)00294-9`. 88

[143] C. H. Whiting and K. E. Jansen. A stabilized finite element method for the incompressible navier-stokes equations using a hierarchical basis. *International Journal for Numerical Methods in Fluids*, 35(1):93–116, 2001. 51

[144] H. Xue, J. C. Ho, and Y. M. Cheng. Comparison of different combustion models in enclosure fire simulation. *Fire Safety Journal*, 36:37–54, 2001. 16

[145] G. P. Zank and W. H. Matthaeus. The Equations of Nearly Incompressible Fluids. I. Hydrodynamics, Turbulence, and Waves. *Physics of Fluids A*, 3:69–82, January 1991. 20

[146] R. K. Zeytounian. A rigoous derivation of the equations of compresible viscous fluid motion with gravity at low Mach number. *Archiwum Mechaniki Stosowanej*, 26:499–509, 1974. 21, 29, 31, 32, 33, 34, 35, 38

[147] R. K. Zeytounian. Sur une formulation asymptotique du problème de Rayleigh-Bénard via l'aproximation de Boussinesq pour les liquides dilatables. *Comptes Rendus de l'Académie des Sciences Paris*, 297(11):271–274, 1983. 21

[148] R. K. Zeytounian. *Asymptotic modelling of atmospheric flows*. Springer Verlag, Heidelberg, 1990. 13, 21, 25, 29, 31, 32, 33, 34, 35, 38

[149] R. K. Zeytounian. Joseph Boussinesq and his approximation: a contemporary view. *Comptes Rendus Mecanique*, 331(8):575–586, Aug 2003. Available from: `http://dx.doi.org/10.1016/S1631-0721(03)00120-7`. 20, 21, 29, 31, 32, 33, 34, 35, 38