Analysis of cell centred finite volume methods for incompressible fluid flows

R. Herbin [1]

Ecole de Printemps de Mécanique des Fluides Numérique,
Roscoff, Juin 2005

[1] Université de Provence, France (herbin@cmi.univ-mrs.fr)

Abstract

The aim of this review paper is to give some insight on some recent developments on the analysis of cell centred finite volume schemes for incompressible fluid flows. We first recall the cell centred finite volume scheme for convection diffusion equations, and give the mathematical tools which are needed for the convergence study. We then turn to the study of a recent scheme which was developed for anisotropic diffusion operators, and which is based on the construction of a discrete gradient. The third chapter is devoted to a colocated cell cantered finite volume method for the approximation of the incompressible Na-vier-Stokes equations posed on a 2D or 3D finite domain. We conclude with the discretization of Navier-Stokes equations with the full viscous stress tensor thanks to the previously mentioned discrete gradient.

Contents

1 Introduction **2**

1.1 Finite volume schemes for conservation laws . . . 2

1.2 Examples . . . 2

2 Cell centered finite volume schemes for convection–diffusion problems **5**

2.1 Meshes and discrete functional spaces . . . 5

2.2 The convection–diffusion problem and its finite volume discretization . . . 9

2.3 Existence and estimates . . . 12

2.4 Convergence . . . 13

2.5 Error estimate . . . 18

3 Discretization of the gradient, anisotropic diffusion problems **22**

3.1 Introduction . . . 22

3.2 A discrete gradient . . . 24

3.2.1 Discrete functional setting . . . 24

3.2.2 Definition of a discrete gradient . . . 25

3.3 Discretization of anisotropic problems . . . 27

3.3.1 The finite volume scheme . . . 27

3.3.2 Discrete $H^1(\Omega)$ estimate . . . 28

3.3.3 Convergence . . . 28

3.3.4 Error estimate . . . 30

3.4 Numerical results . . . 33

3.5 Conclusion . . . 34

4 A colocated scheme for the Navier-Stokes equations **36**

4.1 Spatial discretization and discrete functional analysis . . . 38

4.2 Approximation of the linear steady problem . . . 40

4.2.1 The Stokes problem . . . 40

4.2.2 The finite volume scheme . . . 41

4.2.3 Study of the scheme in the linear case . . . 41

4.2.4 An error estimate . . . 45

4.3 The finite volume scheme for the Navier-Stokes equations . . . 48

4.3.1 The steady-state case . . . 48

4.3.2 The transient case . . . 52

4.4 Stabilization by clustering . . . 55

4.5 Numerical results . . . 56

4.6 Discretization of the full viscous tensor . . . 57

4.7 Conclusions . . . 58

Chapter 1

Introduction

1.1 Finite volume schemes for conservation laws

The finite volume method is a discretization method which is well suited for the numerical simulation of various types (elliptic, parabolic or hyperbolic, for instance) of conservation laws; it has been extensively used in several engineering fields, such as fluid mechanics, heat and mass transfer, semi-conductor technologies or petroleum engineering. Some of the important features of the finite volume method are similar to those of the finite element method: it may be used on arbitrary geometries, using structured or unstructured meshes, and it leads to robust schemes. This feature is classical for inviscid flows leading to first order hyperbolic systems. The use of unstructured grids with finite volumes for viscous flows is more recent. An additional feature is the local conservativity of the numerical fluxes, that is the numerical flux is conserved from one discretization cell to its neighbour. This last feature makes the finite volume method quite attractive when modelling problems for which the flux is of importance, such as in the fields previously mentioned. The finite volume method is locally conservative because it is based on a "balance" approach: a local balance is written on each discretization cell which is often called "control volume"; by the divergence formula, an integral formulation of the fluxes over the boundary of the control volume is then obtained. The fluxes on the boundary are discretized with respect to the discrete unknowns.

Let us introduce the method more precisely on simple examples, and then give a description of the discretization of general conservation laws.

1.2 Examples

Two basic examples can be used to introduce the finite volume method. The analysis of the method for the second example, which is an elliptic equation, will be developed in details in Chapter 2, since it is one of the keys to the the analysis of the cell centred FV method for viscous flows.

Example 1.1 (Transport equation) Consider first the linear transport equation

$$\left\{ \begin{array}{lcl} u_t(x,t) + \mathrm{div}(\mathbf{v}u)(x,t) & = & 0, x \in \mathbb{R}^2, t \in \mathbb{R}_+, \\ u(x,0) = u_0(x), x \in \mathbb{R}^2 & & \end{array} \right. \tag{1.1}$$

where u_t denotes the time derivative of u, $\mathbf{v} \in C^1(\mathbb{R}^2, \mathbb{R}^2)$, and $u_0 \in L^\infty(\mathbb{R}^2)$. Let $\mathcal{M}$ be a mesh of $\mathbb{R}^2$ consisting of polygonal bounded convex subsets of $\mathbb{R}^2$ and let $K \in \mathcal{M}$ be a "control volume", that is an element of the mesh $\mathcal{M}$. Integrating the first equation of (1.1) over K yields the following "balance equation" over K:

$$\int_K u_t(x,t)dx + \int_{\partial K} \mathbf{v}(x,t) \cdot \mathbf{n}_K(x) u(x,t) \mathrm{d}\gamma(x) = 0, \forall t \in \mathbb{R}_+, \tag{1.2}$$

where $\mathbf{n}_K$ denotes the normal vector to ∂K, outward to K. Let $k \in \mathbb{R}_+^*$ be a constant time discretization step and let $t_n = nk$, for $n \in \mathbb{N}$. Writing equation (1.2) at time t_n, $n \in \mathbb{N}$ and discretizing the time partial derivative by the Euler explicit scheme suggests to find an approximation $u^{(n)}(x)$ of the solution of (1.1) at time t_n which satisfies the following semi-discretized equation:

$$\frac{1}{k}\int_K (u^{(n+1)}(x) - u^{(n)}(x))dx + \int_{\partial K} \mathbf{v}(x,t_n)\cdot\mathbf{n}_K(x)u^{(n)}(x)\mathrm{d}\gamma(x) = 0, \forall \mathbf{n} \in \mathbb{N}, \forall K \in \mathcal{M}, \tag{1.3}$$

where $\mathrm{d}\gamma$ denotes the one-dimensional Lebesgue measure on ∂K and $u^{(0)}(x) = u(x,0) = u_0(x)$. We need to define the discrete unknowns for the (finite volume) space discretization. We shall be concerned here principally with the so-called "cell-centred" finite volume method in which each discrete unknown is associated with a control volume, and which is widely used in industrial codes. Let $(u_K^{(n)})_{K\in\mathcal{M},n\in\mathbb{N}}$ denote the discrete unknowns. For $K \in \mathcal{M}$, let $\mathcal{E}_K$ be the set of edges which are included in ∂K, and for $\sigma \subset \partial K$, let $\mathbf{n}_{K,\sigma}$ denote the unit normal to σ outward to K. The second integral in (1.3) may then be split as:

$$\int_{\partial K} \mathbf{v}(x,t_n)\cdot\mathbf{n}_K(x)u^{(n)}(x)\mathrm{d}\gamma(x) = \sum_{\sigma\in\mathcal{E}_K}\int_\sigma \mathbf{v}(x,t_n)\cdot\mathbf{n}_{K,\sigma}u^{(n)}(x)\mathrm{d}\gamma(x); \tag{1.4}$$

for $\sigma \subset \partial K$, let

$$v_{K,\sigma}^{(n)} = \int_\sigma \mathbf{v}(x,t_n)\mathbf{n}_{K,\sigma}(x)\mathrm{d}\gamma(x).$$

Each term of the sum in the right-hand-side of (1.4) is then discretized as

$$F_{K,\sigma}^{(n)} = \begin{cases} v_{K,\sigma}^{(n)}u_K^{(n)} & \text{if } v_{K,\sigma}^{(n)} \geq 0,\\ v_{K,\sigma}^{(n)}u_L^{(n)} & \text{if } v_{K,\sigma}^{(n)} < 0,\end{cases} \tag{1.5}$$

where L denotes the neighbouring control volume to K with common edge σ. This "upstream" or "upwind" choice is classical for transport equations; it may be seen, from the mechanical point of view, as the choice of the "upstream information" with respect to the location of σ. This choice is crucial in the mathematical analysis; it ensures the stability properties of the finite volume scheme (see e.g. [24]). We have therefore derived the following finite volume scheme for the discretization of (1.1):

$$\begin{cases} \dfrac{|K|}{k}(u_K^{(n+1)} - u_K^{(n)}) + \displaystyle\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{(n)} = 0, \forall K \in \mathcal{M}, \forall n \in \mathbb{N},\\ u_K^{(0)} = \displaystyle\int_K u_0(x)dx,\end{cases} \tag{1.6}$$

where $|K|$ denotes the measure of the control volume K and $F_{K,\sigma}^{(n)}$ is defined in (1.5). This scheme is locally conservative in the sense that if σ is a common edge to the control volumes K and L, then $F_{K,\sigma} = -F_{L,\sigma}$. This property is important in several application fields; it will later be shown to be a key ingredient in the mathematical proof of convergence. Similar schemes for the discretization of linear or nonlinear hyperbolic equations are studied for instance in [24].

Example 1.2 (Stationary diffusion equation) Consider the basic diffusion equation

$$\begin{cases} -\Delta u = f \text{ on } \Omega =]0,1[\times]0,1[,\\ u = 0 \text{ on } \partial\Omega.\end{cases} \tag{1.7}$$

Let $\mathcal{M}$ be a rectangular mesh. Let us integrate the first equation of (1.7) over a control volume K of the mesh; with the same notations as in the previous example, this yields:

$$\sum_{\sigma\in\mathcal{E}_K}\int_\sigma -\nabla u(x)\cdot\mathbf{n}_{K,\sigma}\mathrm{d}\gamma(x) = \int_K f(x)dx. \tag{1.8}$$

For each control volume $K \in \mathcal{M}$, let x_K be the center of K. Let σ be the common edge between the control volumes K and L. The flux $-\int_\sigma \nabla u(x) \cdot \mathbf{n}_{K,\sigma} \mathrm{d}\gamma(x)$, may be approximated by the following finite difference approximation:

$$F_{K,\sigma} = -\frac{|\sigma|}{d_\sigma}(u_L - u_K), \tag{1.9}$$

where $(u_K)_{K\in\mathcal{M}}$ are the discrete unknowns and d_σ is the distance between x_K and x_L. This finite difference approximation of the first order derivative $\nabla u \cdot \mathbf{n}$ on the edges of the mesh (where $\mathbf{n}$ denotes the unit normal vector) is consistent: the truncation error on the flux is of order h, where h is the maximum length of the edges of the mesh. We may note that the consistency of the flux holds because for any $\sigma = \sigma_{KL}$ common to the control volumes K and L, the line segment $[x_K x_L]$ is perpendicular to $\sigma = \sigma_{KL}$. Indeed, this is the case here since the control volumes are rectangular. This property is satisfied by other meshes which will be studied hereafter. It is crucial for the discretization of diffusion operators.
In the case where the edge σ is part of the boundary, then d_σ denotes the distance between the centre x_K of the control volume K to which σ belongs and the boundary. The flux $-\int_\sigma \nabla u(x) \cdot \mathbf{n}_{K,\sigma} \mathrm{d}\gamma(x)$, is then approximated by

$$F_{K,\sigma} = \frac{|\sigma|}{d_\sigma} u_K, \tag{1.10}$$

Hence the finite volume scheme for the discretization of (1.7) is:

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = |K| f_K, \forall K \in \mathcal{M}, \tag{1.11}$$

where $F_{K,\sigma}$ is defined by (1.9) and (1.10), and f_K denotes (an approximation of) the mean value of f on K. We shall see later that the finite volume scheme is easy to generalise to a triangular mesh, whereas the finite difference method is not. As in the previous example, the finite volume scheme is locally conservative, since for any edge σ separating K from L, one has $F_{K,\sigma} = -F_{L,\sigma}$.

Chapter 2 below is concerned with the discretization of linear convection-diffusion operators by cell–centred finite volume schemes. For the sake of simplicity, we shall only deal here with Dirichlet boundary conditions, but indeed the analysis maybe extended to other types of boundary conditions, see [24], [41], [40].
In Chapter 3, we propose a discrete gradient operator which allows to treat, in particular, the case of anisotropic flows; it will be used in the next chapter to introduce the discretization of the viscous stress tensor for the Navier-Stokes equations.
Chapter 4 is devoted to a co–located finite volume scheme for the Navier-Stokes equations. We first study the convergence properties of the scheme for the linear Stokes equations, and then turn to the unsteady incompressible Navier–Stokes equations. We present a new stabilization technique by "clusters" which seems to be really efficient in applications. We end by the discretization of the viscous stress tensor in the compressible case.

Chapter 2

Cell centered finite volume schemes for convection–diffusion problems

2.1 Meshes and discrete functional spaces

Let us then turn to the discretization of convection-diffusion problems on general structured or non structured grids, consisting of any polygonal (recall that we shall call "polygonal" any polygonal domain of $\mathbb{R}^2$ or polyhedral domain or $\mathbb{R}^3$) control volumes (satisfying adequate geometrical conditions which are stated in the sequel) and not necessarily ordered in a Cartesian grid. The advantage of finite volume schemes using non structured meshes is clear for convection-diffusion equations. On one hand, the stability and convergence properties of the finite volume scheme (with an upstream choice for the convective flux) ensure a robust scheme for any admissible mesh as defined in Definition 2.1 page 5 below, without any need for refinement in the areas of a large convection flux. On the other hand, the use of a non structured mesh allows the computation of a solution for any shape of the physical domain.

We saw in Example 1.2 page 3 of Chapter 1 that a consistent discretization of the normal flux $-\nabla u \cdot \mathbf{n}$ over the interface of two control volumes K and L may be performed with a differential quotient involving values of the unknown located on the orthogonal line to the interface between K and L, on either side of this interface. This remark suggests the following definition of admissible finite volume meshes for the discretization of diffusion problems. We shall only consider here, for the sake of simplicity, the case of polygonal domains. The case of domains with a regular boundary does not introduce any supplementary difficulty other than complex notations. The definition of admissible meshes and notations introduced in this definition are illustrated in Figure 2.1

Definition 2.1 (Admissible meshes) Let Ω be an open bounded polygonal (polyhedral if $d = 3$) subset of $\mathbb{R}^d$, and $\partial\Omega = \overline{\Omega} \setminus \Omega$ its boundary. An admissible finite volume discretization of Ω, denoted by $\mathcal{D}$, is given by $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, where:

- $\mathcal{M}$ is a finite family of non empty open polygonal convex disjoint subsets of Ω (the "control volumes") such that $\overline{\Omega} = \cup_{K\in\mathcal{M}} \overline{K}$. For any $K \in \mathcal{M}$, let $\partial K = \overline{K} \setminus K$ be the boundary of K and $|K| > 0$ denote the area of K.

- $\mathcal{E}$ is a finite family of disjoint subsets of $\overline{\Omega}$ (the "edges" of the mesh), such that, for all $\sigma \in \mathcal{E}$, there exists a hyper-plane E of $\mathbb{R}^d$ and $K \in \mathcal{M}$ with $\overline{\sigma} = \partial K \cap E$ and σ is a non empty open subset of E. We then denote by $|\sigma| > 0$ the (d-1)-dimensional measure of σ. We assume that, for all $K \in \mathcal{M}$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \cup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$. It then results from the previous hypotheses that, for all $\sigma \in \mathcal{E}$, either $\sigma \subset \partial\Omega$ or there exists $(K, L) \in \mathcal{M}^2$ with $K \neq L$ such that $\overline{K} \cap \overline{L} = \overline{\sigma}$ in which case we shall write $\sigma = \sigma_{KL}$.

- $\mathcal{P}$ is a family of points of Ω indexed by $\mathcal{M}$, denoted by $\mathcal{P} = (x_K)_{K\in\mathcal{M}}$. The coordinates of x_K are denoted by $x_K^{(i)}$, $i = 1, \ldots, d$. The family $\mathcal{P}$ is such that, for all $K \in \mathcal{M}$, $x_K \in K$. Furthermore, for all $\sigma \in \mathcal{E}$ such that there exists $(K, L) \in \mathcal{M}^2$ with $\sigma = \sigma_{KL}$, it is assumed that the straight line (x_K, x_L) going through x_K and x_L is orthogonal to σ_{KL}. For all $K \in \mathcal{M}$ and all $\sigma \in \mathcal{E}_K$, let y_σ be the orthogonal projection of x_K on σ. We suppose that, for any $\sigma \subset \partial\Omega$, one has $y_\sigma \in \sigma$.

  Furthermore, we define the following notations:

  The mesh size is defined by: $h_{\mathcal{D}} = \sup\{\mathrm{diam}(K),\ K \in \mathcal{M}\}$.

  For any $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, $|K|$ is the d-dimensional Lebesgue measure of K (it is the area of K in the two-dimensional case and the volume in the three-dimensional case) and $|\sigma|$ the $(d-1)$-dimensional measure of σ.

  The set of interior (resp. boundary) edges is denoted by $\mathcal{E}_{\mathrm{int}}$ (resp. $\mathcal{E}_{\mathrm{ext}}$), that is $\mathcal{E}_{\mathrm{int}} = \{\sigma \in \mathcal{E};\ \sigma \not\subset \partial\Omega\}$ (resp. $\mathcal{E}_{\mathrm{ext}} = \{\sigma \in \mathcal{E};\ \sigma \subset \partial\Omega\}$). For a given control volume K, let $\mathcal{E}_{K,\mathrm{ext}} = \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}$.

  The set of neighbours of K is denoted by $\mathcal{N}_K$, that is $\mathcal{N}_K = \{L \in \mathcal{M};\ \exists \sigma \in \mathcal{E}_K,\ \overline{\sigma} = \overline{K} \cap \overline{L}\}$. For all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, we denote by $\mathbf{n}_{K,\sigma}$ the unit vector normal to σ outward to K.

  If $\sigma = \sigma_{KL}$, we set $\mathbf{n}_{KL} = \mathbf{n}_{K,\sigma_{KL}}$, and we denote by d_σ or $d_{\sigma_{KL}}$ the Euclidean distance between x_K and x_L (which is positive) and by $d_{K,\sigma}$ the distance from x_K to σ.

  If $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}$, let d_σ denote the Euclidean distance between x_K and y_σ (then, $d_\sigma = d_{K,\sigma}$).

  For any $\sigma \in \mathcal{E}$; the "transmissibility" through σ is defined by $\tau_\sigma = |\sigma|/d_\sigma$ if $d_\sigma \neq 0$.

  In some results and proofs given below, there are summations over $\sigma \in \mathcal{E}_0$, with $\mathcal{E}_0 = \{\sigma \in \mathcal{E};\ d_\sigma \neq 0\}$.

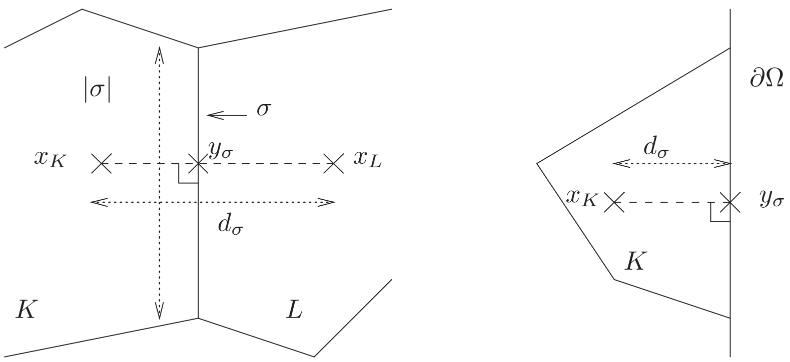

Figure 2.1: Admissible meshes

Remark 2.2 (i) The definition of y_σ for $\sigma \in \mathcal{E}_{\mathrm{ext}}$ requires that $y_\sigma \in \sigma$. However, in many cases, this condition may be relaxed. The condition $x_K \in \overline{K}$ may also be relaxed as described, for instance, in Example 2.3 below.
(ii) The condition $x_K \neq x_L$ if $\sigma = \sigma_{KL}$, is in fact quite easy to satisfy: two neighbouring control volumes K, L which do not satisfy it just have to be collapsed into a new control volume M with $x_M = x_K = x_L$, and the edge σ_{KL} removed from the set of edges. The new mesh thus obtained is admissible.

Example 2.3 (Triangular meshes) Let Ω be an open bounded polygonal subset of $\mathbb{R}^2$. Let $\mathcal{M}$ be a family of open triangular disjoint subsets of Ω such that two triangles having a common edge have also two common vertices. Assume that all angles of the triangles are less than $\pi/2$. This last condition is sufficient for the orthogonal bisectors to intersect inside each triangle, thus naturally defining the points

$x_K \in K$. One obtains an admissible mesh. In the case of an elliptic operator, the finite volume scheme defined on such a grid using differential quotients for the approximation of the normal flux yields a 4-point scheme [47]. This scheme does not lead to a finite difference scheme consistent with the continuous diffusion operator (using a Taylor expansion). The consistency is only verified for the approximation of the fluxes, but this, together with the conservativity of the scheme yields the convergence of the scheme, as it is proved below.
Note that the condition that all angles of the triangles are less than $\pi/2$ (which yields $x_K \in K$) may be relaxed (at least for the triangles the closure of which are in Ω) to the so called "strict Delaunay condition" which is that the closure of the circumscribed circle to each triangle of the mesh does not contain any other triangle of the mesh. For such a mesh, the point x_K (which is the intersection of the orthogonal bisectors of the edges of K) is not always in K, but the scheme (2.13)-(2.15) is convenient since (2.14) yields a consistent approximation of the diffusion fluxes and since the transmissibilities (denoted by $\tau_{\sigma_{KL}}$) are positive.

Example 2.4 (Voronoï meshes) Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$. An admissible finite volume mesh can be built by using the so called "Voronoï" technique. Let $\mathcal{P}$ be a family of points of $\overline{\Omega}$. For example, this family may be chosen as $\mathcal{P} = \{(k_1 h, \ldots, k_d h),\ k_1, \ldots k_d \in \mathbb{Z}\} \cap \Omega$, for a given $h > 0$. The control volumes of the Voronoï mesh are defined with respect to each point x of $\mathcal{P}$ by

$$K_x = \{y \in \Omega, |x-y| < |z-y|,\ \forall z \in \mathcal{P},\ z \neq x\},$$

where $|x-y|$ denotes the Euclidean distance between x and y. Voronoï meshes are admissible in the sense of Definition 2.1 if the assumption "on the boundary", namely part (v) of Definition 2.1, is satisfied. Indeed, this is true, in particular, if the number of points $x \in \mathcal{P}$ which are located on $\partial\Omega$ is "large enough". Otherwise, the assumption (v) of Definition 2.1 may be replaced by the weaker assumption "$d(y_\sigma, \sigma) \leq h_{\mathcal{D}}$ for any $\sigma \in \mathcal{E}_{\text{ext}}$" which is much easier to satisfy. Note also that a slight modification of the treatment of the boundary conditions in the finite volume scheme (2.16)-(2.19) page 11 allows us to obtain convergence and error estimates results (as in theorems 2.18 page 15 and 2.22 page 18) for all Voronoï meshes. This modification consists in replacing, for $K \in \mathcal{M}$ such that $\mathcal{E}_K \cap \mathcal{E}_{\text{ext}} \neq \emptyset$, the equation (2.16), associated to this control volume, by the equation $u_K = g(z_K)$, where z_K is some point on $\partial\Omega \cap \partial K$. In fact, Voronoï meshes often satisfy the following property:

$$\mathcal{E}_K \cap \mathcal{E}_{\text{ext}} \neq \emptyset \Rightarrow x_K \in \partial\Omega$$

and the mesh is therefore admissible in the sense of Definition 2.1 (then, the scheme (2.16)-(2.19) page 11 yields $u_K = g(x_K)$ if $K \in \mathcal{M}$ is such that $\mathcal{E}_K \cap \mathcal{E}_{\text{ext}} \neq \emptyset$).
An advantage of the Voronoï method is that it easily leads to meshes on non polygonal domains Ω.

Let us now introduce the space of piecewise constant functions associated to an admissible mesh and some "discrete H_0^1" norm for this space. This discrete norm will be used to obtain stability properties which are given by some estimates on the approximate solution of a finite volume scheme.

Definition 2.5 *Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, with $d \in \mathbb{N}_*$. Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be an admissible finite volume discretization of Ω in the sense of definition 2.1. We denote by $H_{\mathcal{D}}(\Omega) \subset L^2(\Omega)$ the space of functions which are piecewise constant on each control volume $K \in \mathcal{M}$. For all $w \in H_{\mathcal{D}}(\Omega)$ and for all $K \in \mathcal{M}$, we denote by w_K the constant value of w in K. The space $H_{\mathcal{D}}(\Omega)$ is embedded with the following Euclidean structure: For $(v,w) \in (H_{\mathcal{D}}(\Omega))^2$, we first define the following inner product (corresponding to Neumann boundary conditions)*

$$\langle v, w\rangle_{\mathcal{D}} = \frac{1}{2} \sum_{K \in \mathcal{M}} \sum_{L \in \mathcal{N}_K} \tau_{\sigma_{KL}} (v_L - v_K)(w_L - w_K), \tag{2.1}$$

and then another inner product (corresponding to Dirichlet boundary conditions)

$$[v, w]_{\mathcal{D}} = \langle v, w\rangle_{\mathcal{D}} + \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}} \frac{|\sigma|}{d_{K,\sigma}} v_K w_K. \tag{2.2}$$

Next, we define a semi-norm and a norm in $H_{\mathcal{D}}(\Omega)$ (thanks to the discrete Poincaré inequality (2.7) given below) by

$$|w|_{\mathcal{D}} = \left(\langle w, w\rangle_{\mathcal{D}}\right)^{1/2}, \qquad \|w\|_{\mathcal{D}} = \left([w, w]_{\mathcal{D}}\right)^{1/2}.$$

We define the interpolation operator $P_{\mathcal{D}} : C(\Omega) \to H_{\mathcal{D}}(\Omega)$ by

$$(P_{\mathcal{D}}\varphi)_K = \varphi(x_K), \text{ for all } K \in \mathcal{M}, \text{ for all } \varphi \in C(\Omega). \tag{2.3}$$

Similarly, for $u = (u^{(i)})_{i=1,\ldots,d} \in (H_{\mathcal{D}}(\Omega))^d$, $v = (v^{(i)})_{i=1,\ldots,d} \in (H_{\mathcal{D}}(\Omega))^d$ and $w = (w^{(i)})_{i=1,\ldots,d} \in (H_{\mathcal{D}}(\Omega))^d$, we define:

$$\|u\|_{\mathcal{D}} = \left(\sum_{i=1}^{d}[u^{(i)}, u^{(i)}]_{\mathcal{D}}\right)^{1/2}, \qquad [v, w]_{\mathcal{D}} = \sum_{i=1}^{d}[v^{(i)}, w^{(i)}]_{\mathcal{D}},$$

and $P_{\mathcal{D}} : C(\Omega)^d \to H_{\mathcal{D}}(\Omega)^d$ by $(P_{\mathcal{D}}\varphi)_K = \varphi(x_K)$, for all $K \in \mathcal{M}$, for all $\varphi \in C(\Omega)^d$.

Remark 2.6 *Note that for $u \in H_{\mathcal{D}}(\Omega)$, one has:*

$$\|u\|_{\mathcal{D}} = \Big(\sum_{\sigma\in\mathcal{E}}\tau_\sigma (D_\sigma u)^2\Big)^{\frac{1}{2}}, \tag{2.4}$$

where $\tau_\sigma = |\sigma|/d_\sigma$ and
$D_\sigma u = |u_K - u_L|$ if $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = \sigma_{KL}$,
$D_\sigma u = |u_K|$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$,
where u_K denotes the value taken by u on the control volume K and the sets $\mathcal{E}$, $\mathcal{E}_{\rm int}$, $\mathcal{E}_{\rm ext}$ and $\mathcal{E}_K$ are defined in Definition 2.1 page 5.

The discrete H_0^1 norm is used in the following sections to prove the convergence of finite volume schemes and, under some regularity conditions, to give error estimates. It is related to the H_0^1 norm, see the convergence of the norms in Theorem 2.18. One of the tools used below is the following "discrete Poincaré inequality" which may also be found in [66]:

Lemma 2.7 (Discrete Poincaré inequality) *Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, $d = 2$ or 3, $\mathcal{M}$ an admissible finite volume mesh in the sense of Definition 2.1 and $u \in H_{\mathcal{D}}(\Omega)$ (see Definition 2.5), then*

$$\|u\|_{L^2(\Omega)} \le \mathrm{diam}(\Omega)\|u\|_{\mathcal{D}}, \tag{2.5}$$

where $\|\cdot\|_{\mathcal{D}}$ is the discrete H_0^1 norm defined in Definition 2.5 page 7.

Remark 2.8 (Dirichlet condition on part of the boundary) *This lemma gives a discrete Poincaré inequality for Dirichlet boundary conditions on the boundary $\partial\Omega$. In the case of a Dirichlet condition on part of the boundary only, it is still possible to prove a Discrete boundary condition provided that the polygonal bounded open set Ω is also connex, thanks to Lemma 2.7 page 8 proven in the sequel.*

Proof of Lemma 2.7
For $\sigma \in \mathcal{E}$, define χ_σ from $\mathbb{R}^d \times \mathbb{R}^d$ to $\{0, 1\}$ by $\chi_\sigma(x, y) = 1$ if $\sigma \cap [x, y] \neq \emptyset$ and $\chi_\sigma(x, y) = 0$ otherwise.

Let $u \in H_{\mathcal{D}}(\Omega)$. Let $\mathbf{d}$ be a given unit vector. For all $x \in \Omega$, let $\mathcal{D}_x$ be the semi-line defined by its origin, x, and the vector $\mathbf{d}$. Let $y(x)$ such that $y(x) \in \mathcal{D}_x \cap \partial\Omega$ and $[x, y(x)] \subset \overline{\Omega}$, where $[x, y(x)] = \{tx + (1-t)y(x), t \in [0, 1]\}$ (i.e. $y(x)$ is the first point where $\mathcal{D}_x$ meets $\partial\Omega$).

Let $K \in \mathcal{M}$. For a.e. $x \in K$, one has

$$|u_K| \leq \sum_{\sigma\in\mathcal{E}} D_\sigma u \, \chi_\sigma(x, y(x)),$$

where the notations $D_\sigma u$ and u_K are defined in Definition 2.5 page 7. We write the above inequality for a.e $x \in \Omega$ and not for all $x \in \Omega$ in order to account for the cases where an edge or a vertex of the mesh is included in the semi-line $[x, y(x)]$; in both cases one may not write the above inequality, but there are only a finite number of edges and vertices, and since $\mathbf{d}$ is fixed, the above inequality may be written almost everywhere.
Let $c_\sigma = |\mathbf{d} \cdot \mathbf{n}_\sigma|$ (recall that $\xi \cdot \eta$ denotes the usual scalar product of ξ and η in $\mathbb{R}^d$). By the Cauchy-Schwarz inequality, the above inequality yields:

$$|u_K|^2 \leq \sum_{\sigma\in\mathcal{E}} \frac{(D_\sigma u)^2}{d_\sigma c_\sigma} \chi_\sigma(x, y(x)) \sum_{\sigma\in\mathcal{E}} d_\sigma c_\sigma \chi_\sigma(x, y(x)), \text{ for a.e. } x \in K. \tag{2.6}$$

Let us show that, for a.e. $x \in \Omega$,

$$\sum_{\sigma\in\mathcal{E}} d_\sigma c_\sigma \chi_\sigma(x, y(x)) \leq \mathrm{diam}(\Omega). \tag{2.7}$$

Let $x \in K$, $K \in \mathcal{M}$, such that $\sigma \cap [x, y(x)]$ contains at most one point, for all $\sigma \in \mathcal{E}$, and $[x, y(x)]$ does not contain any vertex of $\mathcal{M}$ (proving (2.7) for such points x leads to (2.7) a.e. on Ω, since $\mathbf{d}$ is fixed). There exists $\sigma_x \in \mathcal{E}_{\mathrm{ext}}$ such that $y(x) \in \sigma_x$. Then, using the fact that the control volumes are convex, one has:

$$\sum_{\sigma\in\mathcal{E}} \chi_\sigma(x, y(x)) d_\sigma c_\sigma = |(x_K - x_{\sigma_x}) \cdot \mathbf{d}|.$$

Since x_K and $x_{\sigma_x} \in \overline{\Omega}$ (see Definition 2.1), this gives (2.7).
Let us integrate (2.6) over Ω; (2.7) gives

$$\sum_{K\in\mathcal{M}} \int_K |u_K|^2 dx \leq \mathrm{diam}(\Omega) \sum_{\sigma\in\mathcal{E}} \frac{(D_\sigma u)^2}{d_\sigma c_\sigma} \int_\Omega \chi_\sigma(x, y(x)) dx.$$

Since $\int_\Omega \chi_\sigma(x, y(x)) dx \leq \mathrm{diam}(\Omega)|\sigma| c_\sigma$, this last inequality yields

$$\sum_{K\in\mathcal{M}} \int_K |u_K|^2 dx \leq (\mathrm{diam}(\Omega))^2 \sum_{\sigma\in\mathcal{E}} |D_\sigma u|^2 \frac{|\sigma|}{d_\sigma} dx.$$

Hence the result. ■

For the sake of completeness, we also recall the discrete Poincaré–Wirtinger inequality, and refer to [24] or [41] for its proof:

Lemma 2.9 (Poincaré–Wirtinger discrete inequality) *There exists $C_\Omega > 0$, only depending on Ω, such that*

$$\|w\|^2_{L^2(\Omega)} \leq C_\Omega |w|^2_{\mathcal{D}}, \ \forall w \in H_{\mathcal{D}}(\Omega) \text{ with } \int_\Omega w(x)\delta x = 0. \tag{2.8}$$

2.2 The convection–diffusion problem and its finite volume discretization

Let us consider here the following elliptic equation

$$-\Delta u(x) + \mathrm{div}(\mathbf{v}u)(x) + bu(x) = f(x), \quad x \in \Omega, \tag{2.9}$$

with Dirichlet boundary condition:

$$u(x) = g(x), \quad x \in \partial\Omega, \tag{2.10}$$

where

Assumption 2.10

1. *Ω is an open bounded polygonal subset of $\mathbb{R}^d$, $d = 2$ or 3,*
2. *$b \geq 0$,*
3. *$f \in L^2(\Omega)$,*
4. *$\mathbf{v} \in C^1(\overline{\Omega}, \mathbb{R}^d)$; $\mathrm{div}\mathbf{v} \geq 0$,*
5. *$g \in C(\partial\Omega, \mathbb{R})$ is such that there exists $\tilde{g} \in H^1(\Omega)$ such that $\overline{\gamma}(\tilde{g}) = g$ a.e. on $\partial\Omega$.*

Here, and in the sequel, "polygonal" is used for both $d = 2$ and $d = 3$ (meaning polyhedral in the latter case) and $\overline{\gamma}$ denotes the trace operator from $H^1(\Omega)$ into $L^2(\partial\Omega)$. Note also that "a.e. on $\partial\Omega$" is a.e. for the $d-1$-dimensional Lebesgue measure on $\partial\Omega$.
Under Assumption 2.10, by the Lax-Milgram theorem, there exists a unique variational solution $u \in H^1(\Omega)$ of Problem (2.9)-(2.10). (For the study of elliptic problems and their discretization by finite element methods, see e.g. [17] and references therein). This solution satisfies $u = w + \tilde{g}$, where $\tilde{g} \in H^1(\Omega)$ is such that $\overline{\gamma}(\tilde{g}) = g$, a.e. on $\partial\Omega$, and w is the unique function of $H^1_0(\Omega)$ satisfying

$$\begin{gathered}\int_\Omega \Big(\nabla w(x) \cdot \nabla\psi(x) + \mathrm{div}(\mathbf{v}w)(x)\psi(x) + bw(x)\psi(x)\Big)dx = \\ \int_\Omega \Big(-\nabla\tilde{g}(x) \cdot \nabla\psi(x) - \mathrm{div}(\mathbf{v}\tilde{g})(x)\psi(x) - b\tilde{g}(x)\psi(x) + f(x)\psi(x)\Big)dx, \ \forall\psi \in H^1_0(\Omega).\end{gathered} \tag{2.11}$$

Let $\mathcal{M}$ be an admissible mesh. Let us now define a finite volume scheme to discretize (2.9), (2.10) page 10.
Let

$$f_K = \frac{1}{|K|}\int_K f(x)dx, \forall K \in \mathcal{M}. \tag{2.12}$$

Let $(u_K)_{K\in\mathcal{M}}$ denote the discrete unknowns. In order to describe the scheme in the most general way, one introduces some auxiliary unknowns, namely the fluxes $F_{K,\sigma}$, for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, and some (expected) approximation of u in σ, denoted by u_σ, for all $\sigma \in \mathcal{E}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, let $\mathbf{n}_{K,\sigma}$ denote the normal unit vector to σ outward to K and $v_{K,\sigma} = \int_\sigma \mathbf{v}(x) \cdot \mathbf{n}_{K,\sigma}\mathrm{d}\gamma(x)$. Note that $\mathrm{d}\gamma$ is the integration symbol for the $(d-1)$-dimensional Lebesgue measure on the considered hyper-plane. In order to discretize the convection term $\mathrm{div}(\mathbf{v}(x)u(x))$ in a stable way, let us define the upstream choice $u_{\sigma,+}$ of u on an edge σ with respect to $\mathbf{v}$ in the following way. If $\sigma = \sigma_{KL}$, then $u_{\sigma,+} = u_K$ if $v_{K,\sigma} \geq 0$, and $u_{\sigma,+} = u_L$ otherwise; if $\sigma \subset K \cap \partial\Omega$, then $u_{\sigma,+} = u_K$ if $v_{K,\sigma} \geq 0$ and $u_{\sigma,+} = g(y_\sigma)$ otherwise.

Let us first assume that the points x_K are located in the interior of each control volume, and are therefore not located on the edges, hence $d_{K,\sigma} > 0$ for any $\sigma \in \mathcal{E}_K$, where $d_{K,\sigma}$ is the distance from x_K to σ. A finite volume scheme can be defined by the following set of equations:

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} + \sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma}u_{\sigma,+} + b|K|u_K = |K|f_K, \ \forall K \in \mathcal{M}, \tag{2.13}$$

$$F_{K,\sigma} = -\tau_{\sigma_{KL}}(u_L - u_K), \forall\sigma \in \mathcal{E}_{\mathrm{int}}, \text{ if } \sigma = \sigma_{KL}, \tag{2.14}$$

$$F_{K,\sigma} = -\tau_\sigma(g(y_\sigma) - u_K), \forall\sigma \in \mathcal{E}_{\mathrm{ext}} \text{ such that } \sigma \in \mathcal{E}_K. \tag{2.15}$$

In the general case, the centre of the cell may be located on an edge. This is the case for instance when constructing Voronoï meshes with some of the original points located on the boundary $\partial\Omega$. In this case, the following formulation of the finite volume scheme is valid, and is equivalent to the above scheme if no cell centre is located on an edge:

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} + \sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma}u_{\sigma,+} + b|K|u_K = |K|f_K,\ \forall K\in\mathcal{M}, \tag{2.16}$$

$$F_{K,\sigma} = -F_{L,\sigma},\ \forall\sigma\in\mathcal{E}_{\rm int},\ \text{if } \sigma=\sigma_{KL}, \tag{2.17}$$

$$F_{K,\sigma}d_{K,\sigma} = -|\sigma|(u_\sigma - u_K),\ \forall\sigma\in\mathcal{E}_K,\ \forall K\in\mathcal{M}, \tag{2.18}$$

$$u_\sigma = g(y_\sigma),\ \ \forall\sigma\in\mathcal{E}_{\rm ext}. \tag{2.19}$$

Note that (2.16)-(2.19) always lead, after an easy elimination of the auxiliary unknowns, to a linear system of N equations with N unknowns, namely the $(u_K)_{K\in\mathcal{M}}$, with $N = \text{card}(\mathcal{M})$.

Remark 2.11

1. Note that one may have, for some $\sigma\in\mathcal{E}_K$, $x_K\in\sigma$, and therefore, thanks to (2.18), $u_\sigma = u_K$.
2. The choice $u_\sigma = g(y_\sigma)$ in (2.19) needs some discussion. Indeed, this choice is possible since g is assumed to belong to $C(\partial\Omega, \mathbb{R})$ and then is everywhere defined on $\partial\Omega$. In the case where the solution to (2.9), (2.10) page 10 belongs to $H^2(\Omega)$ (which yields $g\in C(\partial\Omega,\mathbb{R})$), it is clearly a good choice since it yields the consistency of fluxes (even though an error estimate also holds with other choices for u_σ, the choice given below is, for instance, possible). If $g\in H^{1/2}$ (and not continuous), the value $g(y_\sigma)$ is not necessarily defined. Then, another choice for u_σ is possible, for instance,

$$u_\sigma = \frac{1}{|\sigma|}\int_\sigma g(x)\mathrm{d}\gamma(x).$$

With this latter choice for u_σ, a convergence result also holds, see Theorem [24] or [41].

For the sake of simplicity, it is assumed in Definition 2.1 that $x_K \neq x_L$, for all K, $L\in\mathcal{M}$. This condition may be relaxed; it simply allows an easy expression of the numerical flux $F_{K,\sigma} = -\tau_{\sigma_{KL}}(u_L - u_K)$ if $\sigma = \sigma_{KL}$.

Remark 2.12 (Weak formulation of the FV scheme) *Assume here (for the sake of simplicity) that $g = 0$. Then the scheme (2.16)-(2.19) (or equivalently (2.13)-(2.15) is equivalent to the following weak form, closer to that used in the finite element setting:*

$$\left\{\begin{array}{l} u\in H_{\mathcal{D}}(\Omega),\\ [u,\varphi]_{\mathcal{D}} + c_{\mathcal{D}}(u,\varphi) + \displaystyle\int_\Omega bu(x)\varphi(x)\,dx = \int_\Omega f(x)\varphi(x)dx,\ \forall\varphi\in H_{\mathcal{D}}(\Omega),\end{array}\right. \tag{2.20}$$

where $[\cdot,\cdot]_{\mathcal{D}}$ is an inner product on $H_{\mathcal{D}}$ defined in Definition 2.5, $c_{\mathcal{D}}$ is a (non–symmetric) bilinear form defined by:

$$c_{\mathcal{D}}(u,\varphi) = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}(v^+_{K,\sigma}u_K - v^-_{K,\sigma}u_L)\varphi_K, \tag{2.21}$$

with $v^+_{K,\sigma} = \max(v_{K,\sigma},0)$ and $v^-_{K,\sigma} = \min(v_{K,\sigma},0)$.
Indeed, let $\varphi\in H_{\mathcal{D}}(\Omega)$, mutiplying (2.13) by φ_K, summing over $K\in\mathcal{M}$, and remarking that $v^+_{K,\sigma}u_K - v^-_{K,\sigma}u_L = v_{K,\sigma}u_{\sigma,+}$ yields (2.20). Conversely, taking $\varphi = 1_K$ the characteristic function of K (that is $1_K(x) = 1$ if $x\in K$, and $1_K(x) = 0$ otherwise) in (2.20), one immediately obtains (2.13). Note that a similar weak formulation may also be obtained for non homogeneous Dirichlet boundary conditions.

2.3 Existence and estimates

Let us first prove the existence of the approximate solution and an estimate on this solution. This estimate ensures the stability of the scheme and will be obtained by using the discrete Poincaré inequality (2.5) and will yield convergence thanks to a compactness result (see Theorem 2.16 page 13).

Lemma 2.13 (Existence and estimate) *Under Assumptions 2.10, let $\mathcal{M}$ be an admissible mesh in the sense of Definition 2.1 page 5; there exists a unique solution $(u_K)_{K\in\mathcal{M}}$ to equations (2.16)-(2.19). Furthermore, assuming $g = 0$ and defining $u_{\mathcal{M}} \in H_{\mathcal{D}}(\Omega)$ (see Definition 2.5) by $u_{\mathcal{D}}(x) = u_K$ for a.e. $x \in K$, and for any $K \in \mathcal{M}$, the following estimate holds:*

$$\|u_{\mathcal{D}}\|_{\mathcal{D}} \leq \operatorname{diam}(\Omega)\|f\|_{L^2(\Omega)}, \tag{2.22}$$

where $\|\cdot\|_{\mathcal{D}}$ is the discrete H_0^1 norm defined in Definition 2.5.

Proof of Lemma 2.13

Equations (2.16)-(2.19) lead, after an easy elimination of the auxiliary unknowns, to a linear system of N equations with N unknowns, namely the $(u_K)_{K\in\mathcal{M}}$, with $N = \operatorname{card}(\mathcal{M})$.

Step 1 (existence and uniqueness)

Assume that $(u_K)_{K\in\mathcal{M}}$ satisfies this linear system with $g(y_\sigma) = 0$ for any $\sigma \in \mathcal{E}_{\text{ext}}$, and $f_K = 0$ for all $K \in \mathcal{M}$. Let us multiply (2.16) by u_K and sum over K; from (2.17) and (2.18) one deduces

$$b\sum_{K\in\mathcal{M}}|K|u_K^2 + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}F_{K,\sigma}u_K + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}v_{K,\sigma}u_{\sigma,+}u_K = 0, \tag{2.23}$$

which gives, reordering the summation over the set of edges

$$b\sum_{K\in\mathcal{M}}|K|u_K^2 + \sum_{\sigma\in\mathcal{E}}\tau_\sigma(D_\sigma u)^2 + \sum_{\sigma\in\mathcal{E}}v_\sigma\Big(u_{\sigma,+} - u_{\sigma,-}\Big)u_{\sigma,+} = 0, \tag{2.24}$$

where

$|D_\sigma u| = |u_K - u_L|$, if $\sigma = \sigma_{KL}$ and $|D_\sigma u| = |u_K|$, if $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$;

$v_\sigma = |\int_\sigma \mathbf{v}(x)\cdot\mathbf{n}\mathrm{d}\gamma(x)|$, $\mathbf{n}$ being a unit normal vector to σ;

$u_{\sigma,-}$ is the downstream value to σ with respect to $\mathbf{v}$, i.e. if $\sigma = \sigma_{KL}$, then $u_{\sigma,-} = u_K$ if $v_{K,\sigma} \leq 0$, and $u_{\sigma,-} = u_L$ otherwise; if $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$, then $u_{\sigma,-} = u_K$ if $v_{K,\sigma} \leq 0$ and $u_{\sigma,-} = u_\sigma$ if $v_{K,\sigma} > 0$.

Note that $u_\sigma = 0$ if $\sigma \in \mathcal{E}_{\text{ext}}$.

Now, remark that

$$\sum_{\sigma\in\mathcal{E}}v_\sigma u_{\sigma,+}(u_{\sigma,+} - u_{\sigma,-}) = \frac{1}{2}\sum_{\sigma\in\mathcal{E}}v_\sigma\Big((u_{\sigma,+} - u_{\sigma,-})^2 + (u_{\sigma,+}^2 - u_{\sigma,-}^2)\Big) \tag{2.25}$$

and, thanks to the assumption $\operatorname{div}\mathbf{v} \geq 0$,

$$\sum_{\sigma\in\mathcal{E}}v_\sigma(u_{\sigma,+}^2 - u_{\sigma,-}^2) = \sum_{K\in\mathcal{M}}\Big(\int_{\partial K}\mathbf{v}(x)\cdot\mathbf{n}_K\mathrm{d}\gamma(x)\Big)u_K^2 = \int_\Omega(\operatorname{div}\mathbf{v}(x))u_{\mathcal{D}}^2(x)dx \geq 0. \tag{2.26}$$

Hence,

$$b\|u_{\mathcal{D}}\|_{L^2(\Omega)}^2 + \|u_{\mathcal{D}}\|_{1,\mathcal{M}}^2 = b\sum_{K\in\mathcal{M}}|K|u_K^2 + \sum_{\sigma\in\mathcal{E}}\tau_\sigma(D_\sigma u)^2 \leq 0, \tag{2.27}$$

One deduces, from (2.27), that $u_K = 0$ for all $K \in \mathcal{M}$.

This proves the existence and the uniqueness of the solution $(u_K)_{K\in\mathcal{M}}$, of the linear system given by (2.16)-(2.19), for any $\{g(y_\sigma),\, \sigma \in \mathcal{E}_{\text{ext}}\}$ and $\{f_K,\, K \in \mathcal{M}\}$.

Step 2 (estimate)
Assume $g = 0$. Multiply (2.16) by u_K, sum over K; then, thanks to (2.17), (2.18), (2.25) and (2.26) one has

$$b\|u_{\mathcal{D}}\|^2_{L^2(\Omega)} + \|u_{\mathcal{D}}\|^2_{\mathcal{D}} \leq \sum_{K\in\mathcal{M}} |K| f_K u_K.$$

By the Cauchy-Schwarz inequality, this inequality yields

$$\|u_{\mathcal{D}}\|^2_{\mathcal{D}} \leq (\sum_{K\in\mathcal{M}} |K|u_K^2)^{\frac{1}{2}} (\sum_{K\in\mathcal{M}} |K|f_K^2)^{\frac{1}{2}} \leq \|f\|_{L^2(\Omega)}\|u_{\mathcal{D}}\|_{L^2(\Omega)}.$$

Thanks to the discrete Poincaré inequality (2.5), this yields $\|u_{\mathcal{D}}\|_{\mathcal{D}} \leq \|f\|_{L^2(\Omega)}\mathrm{diam}(\Omega)$, which concludes the proof of the lemma. ■

Let us now state a discrete positivity property which is satisfied by the scheme (2.16)-(2.19); this is an interesting stability property, even though it will not be used in the proofs of the convergence and error estimate.

Proposition 2.14 *Under Assumption 2.10 page 10, let $\mathcal{M}$ be an admissible mesh in the sense of Definition 2.1 page 5, let $(f_K)_{K\in\mathcal{M}}$ be defined by (2.12). If $f_K \geq 0$ for all $K \in \mathcal{M}$, and $g(y_\sigma) \geq 0$, for all $\sigma \in \mathcal{E}_{\mathrm{ext}}$, then the solution $(u_K)_{K\in\mathcal{M}}$ of (2.16)-(2.19) satisfies $u_K \geq 0$ for all $K \in \mathcal{M}$.*

Proof of Proposition 2.14
Assume that $f_K \geq 0$ for all $K \in \mathcal{M}$ and $g(y_\sigma) \geq 0$ for all $\sigma \in \mathcal{E}_{\mathrm{ext}}$. Let $a = \min\{u_K, K \in \mathcal{M}\}$. Let K_0 be a control volume such that $u_{K_0} = a$. Assume first that K_0 is an "interior" control volume, in the sense that $\mathcal{E}_K \subset \mathcal{E}_{\mathrm{int}}$, and that $u_{K_0} \leq 0$. Then, from (2.16),

$$\sum_{\sigma\in\mathcal{E}_{K_0}} F_{K_0,\sigma} + \sum_{\sigma\in\mathcal{E}_{K_0}} v_{K_0,\sigma} u_{\sigma,+} \geq 0; \tag{2.28}$$

since for any neighbour L of K_0 one has $u_L \geq u_{K_0}$, then, noting that $\mathrm{div}\mathbf{v} \geq 0$, one must have $u_L = u_{K_0}$ for any neighbour L of K. Hence, setting $B = \{K \in \mathcal{M},\ u_K = a\}$, there exists $K \in B$ such that $\mathcal{E}_K \not\subset \mathcal{E}_{\mathrm{int}}$, that is K is a control volume "neighbouring the boundary".
Assume then that K_0 is a control volume neighbouring the boundary and that $u_{K_0} = a < 0$. Then, for an edge $\sigma \in \mathcal{E}_{\mathrm{ext}} \cap \mathcal{E}_K$, relations (2.18) and (2.19) yield $g(y_\sigma) < 0$, which is in contradiction with the assumption. Hence Proposition 2.14 is proved. ■

Remark 2.15 This positivity property immediately yields the existence and uniqueness of the solution of the numerical scheme (2.16)-(2.19), which was proved directly in Lemma 2.13.

2.4 Convergence

Let us now show the convergence of approximate solutions obtained by the above finite volume scheme when the size of the mesh tends to 0. One uses Lemma 2.13 together with the following compactness theorem, which is a consequence of the Kolmogorov theorem (see [24] for its proof).

Theorem 2.16 *Let Ω be an open bounded set of $\mathbb{R}^d$ with a Lipschitz continuous boundary, $d \geq 1$, and $\{u_n,\ n \in \mathbb{N}\}$ a bounded sequence of $L^2(\Omega)$. For $n \in \mathbb{N}$, one defines $\tilde{u}_n$ by $\tilde{u}_n = u_n$ a.e. on Ω and $\tilde{u}_n = 0$ a.e. on $\mathbb{R}^d \setminus \Omega$. Assume that there exist $C \in \mathbb{R}$ and $\{h_n,\ n \in \mathbb{N}\} \subset \mathbb{R}_+$ such that $h_n \to 0$ as $n \to \infty$ and*

$$\|\tilde{u}_n(\cdot + \eta) - \tilde{u}_n\|^2_{L^2(\mathbb{R}^d)} \leq C|\eta|(|\eta| + h_n), \forall n \in \mathbb{N},\ \forall \eta \in \mathbb{R}^d. \tag{2.29}$$

Then, $\{u_n,\ n \in \mathbb{N}\}$ is relatively compact in $L^2(\Omega)$. Furthermore, if $u_n \to u$ in $L^2(\Omega)$ as $n \to \infty$, then $u \in H^1_0(\Omega)$.

In order to use Theorem 2.16, one needs the following lemma, which gives an estimate on the translations of the approximate solutions:

Lemma 2.17 (Estimate on the translations) *Let Ω be an open bounded set of $\mathbb{R}^d$, $d = 2$ or 3. Let $\mathcal{M}$ be an admissible mesh in the sense of Definition 2.1 page 5 and $u \in H_{\mathcal{D}}(\Omega)$ (see Definition 2.5). One defines $\tilde{u}$ by $\tilde{u} = u$ a.e. on Ω, and $\tilde{u} = 0$ a.e. on $\mathbb{R}^d \setminus \Omega$. Then there exists $C > 0$, only depending on Ω, such that*

$$\|\tilde{u}(\cdot + \eta) - \tilde{u}\|^2_{L^2(\mathbb{R}^d)} \le \|u\|^2_{1,\mathcal{M}} |\eta| (|\eta| + C\, h_{\mathcal{D}}), \forall \eta \in \mathbb{R}^d. \tag{2.30}$$

Proof of Lemma 2.17

For $\sigma \in \mathcal{E}$, define χ_σ from $\mathbb{R}^d \times \mathbb{R}^d$ to $\{0, 1\}$ by $\chi_\sigma(x, y) = 1$ if $[x, y] \cap \sigma \neq \emptyset$ and $\chi_\sigma(x, y) = 0$ if $[x, y] \cap \sigma = \emptyset$.

Let $\eta \in \mathbb{R}^d$, $\eta \neq 0$. One has

$$|\tilde{u}(x + \eta) - \tilde{u}(x)| \le \sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) |D_\sigma u|, \quad \text{for a.e. } x \in \Omega$$

(see Definition 2.5 page 7 for the definition of $D_\sigma u$).

This gives, using the Cauchy-Schwarz inequality,

$$|\tilde{u}(x + \eta) - \tilde{u}(x)|^2 \le \sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) \frac{|D_\sigma u|^2}{d_\sigma c_\sigma} \sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) d_\sigma c_\sigma, \quad \text{for a.e. } x \in \mathbb{R}^d, \tag{2.31}$$

where $c_\sigma = |\mathbf{n}_\sigma \cdot \frac{\eta}{|\eta|}|$, and $\mathbf{n}_\sigma$ denotes a unit normal vector to σ.

Let us now prove that there exists $C > 0$, only depending on Ω, such that

$$\sum_{\sigma \in \mathcal{E}} \chi_\sigma(x, x + \eta) d_\sigma c_\sigma \le |\eta| + C\, h_{\mathcal{D}}, \tag{2.32}$$

for a.e. $x \in \mathbb{R}^d$.

Let $x \in \mathbb{R}^d$ such that $\sigma \cap [x, x + \eta]$ contains at most one point, for all $\sigma \in \mathcal{E}$, and $[x, x + \eta]$ does not contain any vertex of $\mathcal{M}$ (proving (2.32) for such points x gives (2.32) for a.e. $x \in \mathbb{R}^d$, since η is fixed). Since Ω is not assumed to be convex, it may happen that the line segment $[x, x + \eta]$ is not included in $\overline{\Omega}$. In order to deal with this, let $y, z \in [x, x + \eta]$ such that $y \neq z$ and $[y, z] \subset \overline{\Omega}$; there exist $K, L \in \mathcal{M}$ such that $y \in \overline{K}$ and $z \in \overline{L}$. Hence,

$$\sum_{\sigma \in \mathcal{E}} \chi_\sigma(y, z) d_\sigma c_\sigma = |(y_1 - z_1) \cdot \frac{\eta}{|\eta|}|,$$

where $y_1 = x_K$ or y_σ with $\sigma \in \mathcal{E}_{\text{ext}} \cap \mathcal{E}_K$ and $z_1 = x_L$ or $y_{\tilde{\sigma}}$ with $\tilde{\sigma} \in \mathcal{E}_{\text{ext}} \cap \mathcal{E}_L$, depending on the position of y and z in $\overline{K}$ or $\overline{L}$ respectively.

Since $y_1 = y + y_2$, with $|y_2| \le h_{\mathcal{D}}$, and $z_1 = z + z_2$, with $|z_2| \le h_{\mathcal{D}}$, one has

$$|(y_1 - z_1) \cdot \frac{\eta}{|\eta|}| \le |y - z| + |y_2| + |z_2| \le |y - z| + 2\, h_{\mathcal{D}}$$

and

$$\sum_{\sigma \in \mathcal{E}} \chi_\sigma(y, z) d_\sigma c_\sigma \le |y - z| + 2\, h_{\mathcal{D}}. \tag{2.33}$$

Note that this yields (2.32) with $C = 2$ if $[x, x + \eta] \subset \overline{\Omega}$.

Since Ω has a finite number of sides, the line segment $[x, x+\eta]$ intersects $\partial\Omega$ a finite number of times; hence there exist $t_1, \ldots, t_n$ such that $0 \le t_1 < t_2 < \ldots < t_n \le 1$, $n \le N$, where N only depends on Ω (indeed, it is possible to take $N = 2$ if Ω is convex and N equal to the number of sides of Ω for a general Ω) and such that

$$\sum_{\sigma\in\mathcal{E}} \chi_\sigma(x, x+\eta) d_\sigma c_\sigma = \sum_{\substack{i=1,n-1\\ \text{odd} i}} \sum_{\sigma\in\mathcal{E}} \chi_\sigma(x_i, x_{i+1}) d_\sigma c_\sigma,$$

with $x_i = x + t_i\eta$, for $i = 1, \ldots, n$, $x_i \in \partial\Omega$ if $t_i \notin \{0, 1\}$ and $[x_i, x_{i+1}] \subset \overline{\Omega}$ if i is odd.
Then, thanks to (2.33) with $y = x_i$ and $z = x_{i+1}$, for $i = 1, \ldots, n-1$, one has (2.32) with $C = 2(N-1)$ (in particular, if Ω is convex, $C = 2$ is convenient for (2.32) and therefore for (2.30) as we shall see below).

In order to conclude the proof of Lemma 2.17, remark that, for all $\sigma \in \mathcal{E}$,

$$\int_{\mathbb{R}^d} \chi_\sigma(x, x+\eta) dx \le |\sigma| c_\sigma |\eta|.$$

Therefore, integrating (2.31) over $\mathbb{R}^d$ yields, with (2.32),

$$\|\tilde{u}(\cdot+\eta) - \tilde{u}\|^2_{L^2(\mathbb{R}^d)} \le \Big(\sum_{\sigma\in\mathcal{E}} \frac{|\sigma|}{d_\sigma} |D_\sigma u|^2\Big) |\eta| (|\eta| + C\, h_\mathcal{D}).$$

■

We are now able to state the convergence theorem.

Theorem 2.18 (Convergence, Dirichlet boundary conditions) *Assume items 1, 2, 3 and 4 of Assumption 2.10 page 10 and $g \in H^{1/2}(\partial\Omega)$. Let $\zeta \in \mathbb{R}_+$ and $M \in \mathbb{N}$ be given values. Let $\mathcal{M}$ be an admissible mesh (in the sense of Definition 2.1 page 5) such that $d_{K,\sigma} \ge \zeta \mathrm{diam}(K)$ for all control volumes $K \in \mathcal{M}$ and for all $\sigma \in \mathcal{E}_K$, and $\mathrm{card}(\mathcal{E}_K) \le M$ for all $K \in \mathcal{M}$. Let $(u_K)_{K\in\mathcal{M}}$ be the solution of the system given by equations (2.16)-(2.18) and*

$$u_\sigma = \frac{1}{|\sigma|} \int_\sigma g(x) \mathrm{d}\gamma(x),\ \ \forall \sigma \in \mathcal{E}_{\mathrm{ext}}. \tag{2.34}$$

(note that the proofs of existence and uniqueness of $(u_K)_{K\in\mathcal{M}}$ which were given in Lemma 2.13 page 12 remain valid). Define $u_\mathcal{M} \in H_\mathcal{D}(\Omega)$ by $u_\mathcal{D}(x) = u_K$ for a.e. $x \in K$ and for any $K \in \mathcal{M}$. Then, $u_\mathcal{M}$ converges, in $L^2(\Omega)$, to the unique variational solution $u \in H^1(\Omega)$ of Problem (2.9), (2.10) as $h_\mathcal{D} \to 0$. Furthermore, if $g = 0$, $\|u_\mathcal{M}\|_{1,\mathcal{M}}$ converges to $\|u\|_{H^1_0(\Omega)}$ as $h_\mathcal{D} \to 0$.

Remark 2.19
In Theorem 2.18, it is possible to prove convergence results when $f(x)$ (resp. $\mathbf{v}(x)$) is replaced by some nonlinear function $f(x, u(x))$, (resp. $\mathbf{v}(x, u(x))$ under adequate assumptions, see [25].

Proof of Theorem 2.18
For the sake of simplicity, we only give here the proof of convergence in the case of homogeneous Dirichlet boundary conditions, and refer to [24] and [25] for the general case.

Let Y be the set of approximate solutions, that is the set of $u_\mathcal{D}$ where $\mathcal{M}$ is an admissible mesh in the sense of Definition 2.1 page 5. First, we want to prove that $u_\mathcal{D}$ tends to the unique solution (in $H^1_0(\Omega)$) to (2.11) as $h_\mathcal{D} \to 0$.

Thanks to Lemma 2.13 and to the discrete Poincaré inequality (2.5), there exists $C_1 \in \mathbb{R}$, only depending on Ω and f, such that $\|u_\mathcal{M}\|_{1,\mathcal{M}} \le C_1$ and $\|u_\mathcal{M}\|_{L^2(\Omega)} \le C_1$ for all $u_\mathcal{D} \in Y$. Then, thanks to Lemma 2.17 and to the compactness result given in Theorem 2.16 page 13, the set Y is relatively compact in $L^2(\Omega)$ and any possible limit (in $L^2(\Omega)$) of a sequence $(u_{\mathcal{M}_n})_{n\in\mathbb{N}} \subset Y$ (such that $h_{\mathcal{D}^n)}) \to 0$) belongs to $H^1_0(\Omega)$. Therefore, thanks to the uniqueness of the solution (in $H^1_0(\Omega)$) of (2.11), it is sufficient to prove that if $(u_{\mathcal{M}_n})_{n\in\mathbb{N}} \subset Y$ converges towards some $u \in H_0^1(\Omega)$, in $L^2(\Omega)$, and $h_{\mathcal{D}^n)} \to 0$ (as $n \to \infty$), then u is the solution to (2.11). We prove this result below, omitting the index n, that is assuming $u_{\mathcal{D}} \to u$ in $L^2(\Omega)$ as $h_{\mathcal{D}} \to 0$.

Let $\psi \in C_c^\infty(\Omega)$ and let $h_{\mathcal{D}}$ be small enough so that $\psi(x) = 0$ if $x \in K$ and $K \in \mathcal{M}$ is such that $\partial K \cap \partial\Omega \neq \emptyset$. Taking $\varphi = P_{\mathcal{D}} psi$ in (2.20) (or equivalently multiplying (2.16) by $\psi(x_K)$ and summing the result over $K \in \mathcal{M}$) yields

$$T_1 + T_2 + T_3 = T_4, \tag{2.35}$$

with

$$\begin{aligned} T_1 &= b \sum_{K\in\mathcal{M}} |K| u_K \psi(x_K), \\ T_2 &= -\sum_{K\in\mathcal{M}} \sum_{L\in\mathcal{N}_K} \tau_{\sigma_{KL}} (u_L - u_K)\psi(x_K), \\ T_3 &= \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma} u_{\sigma,+} \psi(x_K), \\ T_4 &= \sum_{K\in\mathcal{M}} |K| \psi(x_K) f_K. \end{aligned}$$

First remark that, since $u_{\mathcal{M}}$ tends to u in $L^2(\Omega)$,

$$T_1 \to b \int_\Omega u(x)\psi(x)dx \text{ as } h_{\mathcal{D}} \to 0.$$

Similarly,

$$T_4 \to \int_\Omega f(x)\psi(x)dx \text{ as } h_{\mathcal{D}} \to 0.$$

Let us now turn to the study of T_2;

$$T_2 = -\sum_{\sigma_{KL}\in\mathcal{E}_{\text{int}}} \tau_{\sigma_{KL}} (u_L - u_K)(\psi(x_K) - \psi(x_L)).$$

Consider the following auxiliary expression:

$$\begin{aligned} T_2' &= \int_\Omega u_{\mathcal{M}}(x)\Delta\psi(x)dx \\ &= \sum_{K\in\mathcal{M}} u_K \int_K \Delta\psi(x)dx \\ &= \sum_{\sigma_{KL}\in\mathcal{E}_{\text{int}}} (u_K - u_L) \int_{\sigma_{KL}} \nabla\psi(x)\cdot\mathbf{n}_{K,L} \mathrm{d}\gamma(x). \end{aligned}$$

Since $u_{\mathcal{M}}$ converges to u in $L^2(\Omega)$, it is clear that T_2' tends to $\displaystyle\int_\Omega u(x)\Delta\psi(x)\ dx$ as $h_{\mathcal{D}}$ tends to 0. Define

$$R_{K,L} = \frac{1}{|\sigma_{KL}|} \int_{\sigma_{KL}} \nabla\psi(x)\cdot\mathbf{n}_{K,L}\mathrm{d}\gamma(x) - \frac{\psi(x_L) - \psi(x_K)}{d_{\sigma_{KL}}},$$

where $\mathbf{n}_{K,L}$ denotes the unit normal vector to σ_{KL}, outward to K, then

$$
\begin{aligned}
|T_2 + T_2'| &= |\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |\sigma_{KL}|(u_K - u_L)R_{K,L}| \\
&\le \Big[\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |\sigma_{KL}|\frac{(u_K-u_L)^2}{d_{\sigma_{KL}}} \sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |\sigma_{KL}| d_{\sigma_{KL}}(R_{K,L})^2\Big]^{1/2},
\end{aligned}
$$

Regularity properties of the function ψ give the existence of $C_2 \in \mathbb{R}$, only depending on ψ, such that $|R_{K,L}| \le C_2 h_{\mathcal{D}}$. Therefore, since

$$
\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |\sigma_{KL}| d_{\sigma_{KL}} \le d|\Omega|,
$$

from Estimate (2.22), we conclude that $T_2 + T_2' \to 0$ as $h_{\mathcal{D}} \to 0$.
Let us now show that T_3 tends to $-\int_\Omega \mathbf{v}(x)u(x)\nabla\psi(x)dx$ as $h_{\mathcal{D}} \to 0$. Let us decompose $T_3 = T_3' + T_3''$ where

$$
T_3' = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma}(u_{\sigma,+} - u_K)\psi(x_K)
$$

and

$$
T_3'' = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma} u_K \psi(x_K) = \int_\Omega \mathrm{div}\mathbf{v}(x) u_{\mathcal{D}}(x)\psi_{\mathcal{D}}(x)dx,
$$

where $\psi_{\mathcal{D}}$ is defined by $\psi_{\mathcal{D}}(x) = \psi(x_K)$ if $x \in K$, $K \in \mathcal{M}$. Since $u_{\mathcal{D}} \to u$ and $\psi_{\mathcal{D}} \to \psi$ in $L^2(\Omega)$ as $h_{\mathcal{D}} \to 0$ (indeed, $\psi_{\mathcal{D}} \to \psi$ uniformly on Ω as $h_{\mathcal{D}} \to 0$) and since $\mathrm{div}\mathbf{v} \in L^\infty(\Omega)$, one has

$$
T_3'' \to \int_\Omega \mathrm{div}\mathbf{v}(x)u(x)\psi(x)dx \text{ as } h_{\mathcal{D}} \to 0.
$$

Let us now rewrite T_3' as $T_3' = T_3''' + r_3$ with

$$
T_3''' = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} (u_{\sigma,+} - u_K)\int_\sigma \mathbf{v}(x)\cdot\mathbf{n}_{K,\sigma}\psi(x)\mathrm{d}\gamma(x)
$$

and

$$
r_3 = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} (u_{\sigma,+} - u_K)\int_\sigma \mathbf{v}(x)\cdot\mathbf{n}_{K,\sigma}(\psi(x_K) - \psi(x))\mathrm{d}\gamma(x).
$$

Thanks to the regularity of $\mathbf{v}$ and ψ, there exists C_3 only depending on $\mathbf{v}$ and ψ such that

$$
|r_3| \le C_3 h_{\mathcal{D}} \sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |u_K - u_L||\sigma_{KL}|,
$$

which yields, with the Cauchy-Schwarz inequality,

$$
|r_3| \le C_3 h_{\mathcal{D}} (\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} \tau_{\sigma_{KL}}|u_K - u_L|^2)^{\frac12} (\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |\sigma_{KL}| d_{\sigma_{KL}})^{\frac12},
$$

from which one deduces, with Estimate (2.22), that $r_3 \to 0$ as $h_{\mathcal{D}} \to 0$.
Next, remark that

$$
T_3''' = -\sum_{K\in\mathcal{M}} u_K \sum_{\sigma\in\mathcal{E}_K}\int_\sigma \mathbf{v}(x)\cdot\mathbf{n}_{K,\sigma}\psi(x)\mathrm{d}\gamma(x) = -\sum_{K\in\mathcal{M}} u_K \int_K \mathrm{div}(\mathbf{v}(x)\psi(x))dx.
$$

This implies (since $u_{\mathcal{D}} \to u$ in $L^2(\Omega)$) that $T_3''' \to -\int_\Omega \mathrm{div}(\mathbf{v}(x)\psi(x))u(x)dx$, so that T_3' has the same limit and $T_3 \to -\int_\Omega \mathbf{v}(x)\cdot\nabla\psi(x)u(x)dx$.

Hence, letting $h_{\mathcal{D}} \to 0$ in (2.35) yields that the function $u \in H_0^1(\Omega)$ satisfies

$$\int_\Omega \Big(bu(x)\psi(x) - u(x)\Delta\psi(x) - \mathbf{v}(x)u(x)\nabla\psi(x) - f(x)\psi(x)\Big) dx = 0,\ \forall \psi \in C_c^\infty(\Omega),$$

which, in turn, yields (2.11) thanks to the fact that $u \in H_0^1(\Omega)$, and to the density of $C_c^\infty(\Omega)$ in $H_0^1(\Omega)$. This concludes the proof of $u_\mathcal{D} \to u$ in $L^2(\Omega)$ as $h_\mathcal{D} \to 0$, where u is the unique solution (in $H_0^1(\Omega)$) to (2.11).

S Let us now prove that $\|u_\mathcal{D}\|_{1,\mathcal{M}}$ tends to $\|u\|_{H_0^1(\Omega)}$ in the pure diffusion case, i.e. assuming $b = 0$ and $\mathbf{v} = 0$. Since

$$\|u_\mathcal{D}\|_{1,\mathcal{M}}^2 = \int_\Omega f_\mathcal{D}(x)u_\mathcal{D}(x)dx \to \int_\Omega f(x)u(x)dx \text{ as } h_\mathcal{D} \to 0,$$

where $f_\mathcal{D}$ is defined from Ω to $\mathbb{R}$ by $f_\mathcal{D}(x) = f_K$ a.e. on K for all $K \in \mathcal{M}$, it is easily seen that

$$\|u_\mathcal{D}\|_{1,\mathcal{M}}^2 \to \int_\Omega f(x)u(x)dx = \|u\|_{H_0^1(\Omega)}^2 \text{ as } h_\mathcal{D} \to 0.$$

This concludes the proof of Theorem 2.18. ■

Note that above convergence result (and more particularly the convergence proof of T_2) also yield the following result, which will be used for the Stokes problem in Chapter 4.

Lemma 2.20 *Let $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ be a sequence of admissible discretizations of Ω in the sense of definition 2.1, such that $\lim\limits_{m\to\infty} h_{\mathcal{D}^{m})} = 0$. Let us assume that there exists $C > 0$ and a sequence $(u^{(m)})_{m\in\mathbb{N}}$ such that $u^{(m)} \in H_{\mathcal{D}^{(m)}}(\Omega)$ and $\|u^{(m)}\|_{\mathcal{D}_m} \le C$ for all $m \in \mathbb{N}$.*
Then, there exists $u \in H_0^1(\Omega)$ and a subsequence of $(u^{(m)})_{m\in\mathbb{N}}$, again denoted $(u^{(m)})_{m\in\mathbb{N}}$, such that:

1. *the sequence $(u^{(m)})_{m\in\mathbb{N}}$ converges in $L^2(\Omega)$ to u as $m \to +\infty$,*
2. *for all $\varphi \in C_c^\infty(\Omega)$, we have*

$$\lim_{m\to+\infty} [u^{(m)}, P_{\mathcal{D}_m}\varphi]_{\mathcal{D}_m} = \int_\Omega \nabla u(x)\cdot\nabla\varphi(x)\delta x, \tag{2.36}$$

Remark 2.21 (Consistency for the adjoint operator) The proof of Theorem 2.18 uses the property of consistency of the (diffusion) fluxes on the test functions. This property consists in writing the consistency of the fluxes for the adjoint operator to the discretized Dirichlet operator. This consistency is achieved thanks to that of fluxes for the discretized Dirichlet operator and to the fact that this operator is self adjoint. In fact, any discretization of the Dirichlet operator giving "L^2-stability" and consistency of fluxes on its adjoint, yields a convergence result. On the contrary, the error estimates proved in Section 2.5 directly use the consistency for the discretized Dirichlet operator itself.

2.5 Error estimate

Theorem 2.22 *Under Assumption 2.10 page 10, let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be an admissible discretization as defined in Definition 2.1 page 5 and $u_\mathcal{D} \in H_\mathcal{D}$ (see Definition 2.5 page 7) be defined a.e.in Ω by $u_\mathcal{D}(x) = u_K$ for a.e. $x \in K$, for all $K \in \mathcal{M}$, where $(u_K)_{K\in\mathcal{M}}$ is the solution to (2.16)-(2.19). Assume that the unique variational solution u of Problem (2.9)-(2.10) satisfies $u \in C^2(\overline{\Omega})$. Let, for each $K \in \mathcal{M}$, $e_K = u(x_K) - u_K$, and $e_\mathcal{D} \in H_\mathcal{D}$ defined by $e_\mathcal{D}(x) = e_K$ for a.e. $x \in K$, for all $K \in \mathcal{M}$.*
Then, there exists $C > 0$ only depending on u, $\mathbf{v}$ and Ω such that

$$\|e_\mathcal{D}\|_\mathcal{D} \le C h_\mathcal{D}, \tag{2.37}$$

where $\|\cdot\|_\mathcal{M}$ is the discrete H_0^1 norm defined in Definition 2.5,

$$\|e_{\mathcal{D}}\|_{L^2(\Omega)} \le Ch_{\mathcal{D}} \tag{2.38}$$

and

$$\begin{aligned}
&\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=\sigma_{KL}}} |\sigma| d_\sigma \Big(\frac{u_L-u_K}{d_\sigma} - \frac{1}{|\sigma|}\int_\sigma \nabla u(x)\cdot \mathbf{n}_{K,\sigma} \mathrm{d}\gamma(x)\Big)^2 + \\
&\sum_{\substack{\sigma\in\mathcal{E}_{\rm ext}\\ \sigma\in\overline{K}\cap\partial\Omega}} |\sigma| d_\sigma \Big(\frac{g(y_\sigma)-u_K}{d_\sigma} - \frac{1}{|\sigma|}\int_\sigma \nabla u(x)\cdot \mathbf{n}_{K,\sigma} \mathrm{d}\gamma(x)\Big)^2 \le Ch_{\mathcal{D}}^2.
\end{aligned} \tag{2.39}$$

Remark 2.23

1. Inequality (2.37) (resp. (2.38)) yields an estimate of order 1 for the discrete H_0^1 norm (resp. L^2 norm) of the error on the solution. Note also that, since $u \in C^1(\overline{\Omega})$, one deduces, from (2.38), the existence of C only depending on u and Ω such that $\|u - u_{\mathcal{D}}\|_{L^2(\Omega)} \le Ch_{\mathcal{D}}$. Inequality (2.39) may be seen as an estimate of order 1 for the L^2 norm of the flux.

2. The proof of Theorem 2.22 given below is close to that of error estimates for finite element schemes in the sense that it uses the coerciveness of the operator (the discrete Poincaré inequality) instead of the discrete maximum principle of Proposition 2.14 page 13 (which is used for error estimates with finite difference schemes).

3. A similar result holds if u is assumed to be in $H^2(\Omega)$ instead of $C^2(\Omega)$, with some minor additional assumption on the mesh. The proof is similar to the one given below, except for the consistency part, which is a bit more technical. We refer to [24] of [41], where other boundary conditions are also considered.

Proof of Theorem 2.22

Let $u_{\mathcal{D}} \in H_{\mathcal{D}}$ be defined a.e. in Ω by $u_{\mathcal{D}}(x) = u_K$ for a.e. $x \in K$, for all $K \in \mathcal{M}$, where $(u_K)_{K\in\mathcal{M}}$ is the solution to (2.16)-(2.19). Let us write the flux balance for any $K \in \mathcal{M}$;

$$\sum_{\sigma\in\mathcal{E}_K} \Big(\overline{F}_{K,\sigma} + \overline{V}_{K,\sigma}\Big) + b\int_K u(x)dx = \int_K f(x)dx, \tag{2.40}$$

where $\overline{F}_{K,\sigma} = -\int_\sigma \nabla u(x)\cdot\mathbf{n}_{K,\sigma}\mathrm{d}\gamma(x)$, and $\overline{V}_{K,\sigma} = \int_\sigma u(x)\mathbf{v}(x)\cdot\mathbf{n}_{K,\sigma}\mathrm{d}\gamma(x)$ are respectively the diffusion and convection fluxes through σ outward to K.
Let $F^\star_{K,\sigma}$ and $V^\star_{K,\sigma}$ be defined by

$$F^\star_{K,\sigma} = -\tau_{\sigma_{KL}}(u(x_L) - u(x_K)),\ \forall\sigma = \sigma_{KL} \in \mathcal{E}_K \cap \mathcal{E}_{\rm int},\ \forall K \in \mathcal{M},$$

$$F^\star_{K,\sigma} d(x_K, \sigma) = -|\sigma|(u(y_\sigma) - u(x_K)),\ \forall\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm ext},\ \forall K \in \mathcal{M},$$

$$V^\star_{K,\sigma} = v_{K,\sigma} u(x_{\sigma,+}), \forall\sigma \in \mathcal{E}_K,\ \forall K \in \mathcal{M},$$

where $x_{\sigma,+} = x_K$ (resp. x_L) if $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = \sigma_{KL}$ and $v_{K,\sigma} \ge 0$ (resp. $v_{K,\sigma} \le 0$) and $x_{\sigma,+} = x_K$ (resp. y_σ) if $\sigma = \mathcal{E}_K \cap \mathcal{E}_{\rm ext}$ and $v_{K,\sigma} \ge 0$ (resp. $v_{K,\sigma} \le 0$). Then, the consistency error on the diffusion and convection fluxes may be defined as

$$R_{K,\sigma} = \frac{1}{|\sigma|}(\overline{F}_{K,\sigma} - F^\star_{K,\sigma}), \tag{2.41}$$

$$r_{K,\sigma} = \frac{1}{|\sigma|}(\overline{V}_{K,\sigma} - V^\star_{K,\sigma}), \tag{2.42}$$

Thanks to the regularity of u and $\mathbf{v}$, there exists $C_1 \in \mathbb{R}$, only depending on u and $\mathbf{v}$, such that $|R_{K,\sigma}| + |r_{K,\sigma}| \leq C_1 h_{\mathcal{D}}$ for any $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$. For $K \in \mathcal{M}$, let

$$\rho_K = u(x_K) - (1/|K|) \int_K u(x) dx,$$

so that $|\rho_K| \leq C_2 h_{\mathcal{D}}$ with some $C_2 \in \mathbb{R}_+$ only depending on u.

Subs tract (2.16) to (2.40); thanks to (2.41) and (2.42), one has

$$\sum_{\sigma \in \mathcal{E}_K} \Big(G_{K,\sigma} + W_{K,\sigma} \Big) + b|K| e_K = b|K| \rho_K - \sum_{\sigma \in \mathcal{E}_K} |\sigma| (R_{K,\sigma} + r_{K,\sigma}), \tag{2.43}$$

where
$G_{K,\sigma} = F^{\star}_{K,\sigma} - F_{K,\sigma}$ is such that

$$G_{K,\sigma} = -\tau_{\sigma_{KL}}(e_L - e_K), \, \forall K \in \mathcal{M}, \, \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{int}}, \sigma = \sigma_{KL},$$

$$G_{K,\sigma} d(x_K, \sigma) = |\sigma| e_K, \, \forall K \in \mathcal{M}, \, \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}},$$

with $e_K = u(x_K) - u_K$, and $W_{K,\sigma} = V^{\star}_{K,\sigma} - V_{K,\sigma} = v_{K,\sigma}(u(x_{\sigma,+}) - u_{\sigma,+})$
Multiply (2.43) by e_K, sum for $K \in \mathcal{M}$, and note that

$$\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} G_{K,\sigma} e_K = \sum_{\sigma \in \mathcal{E}} |D_\sigma e|^2 \frac{|\sigma|}{d_\sigma} = \|e\|^2_{1,\mathcal{M}}.$$

Hence

$$\|e_{\mathcal{D}}\|^2_{1,\mathcal{M}} + \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} v_{K,\sigma} e_{\sigma,+} e_K + b \|e_{\mathcal{D}}\|^2_{L^2(\Omega)} \leq b \sum_{K \in \mathcal{M}} |K| \rho_K e_K - \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} |\sigma| (R_{K,\sigma} + r_{K,\sigma}) e_K, \tag{2.44}$$

where
$e_{\mathcal{D}} \in H_{\mathcal{D}}$, $e_{\mathcal{D}}(x) = e_K$ for a.e. $x \in K$ and for all $K \in \mathcal{M}$,
$|D_\sigma e| = |e_K - e_L|$, if $\sigma \in \mathcal{E}_{\text{int}}$, $\sigma = \sigma_{KL}$, $|D_\sigma e| = |e_K|$, if $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$,
$e_{\sigma,+} = u(x_{\sigma,+}) - u_{\sigma,+}$.
By Young's inequality, the first term of the left hand side satisfies:

$$\Big| \sum_{K \in \mathcal{M}} |K| \rho_K e_K \Big| \leq \frac{1}{2} \|e_{\mathcal{D}}\|^2_{L^2(\Omega)} + \frac{1}{2} C_2^2 (h_{\mathcal{D}})^2 \; |\Omega|. \tag{2.45}$$

Thanks to the assumption div$\mathbf{v} \geq 0$, one obtains, through a computation similar to (2.25)-(2.26) page 12 that

$$\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} v_{K,\sigma} e_{\sigma,+} e_K \geq 0.$$

Hence, (2.44) and (2.45) yield that there exists C_3 only depending on u, b and Ω such that

$$\|e_{\mathcal{D}}\|^2_{1,\mathcal{M}} + \frac{1}{2} b \|e_{\mathcal{D}}\|^2_{L^2(\Omega)} \leq C_3 (h_{\mathcal{D}})^2 - \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} |\sigma| (R_{K,\sigma} + r_{K,\sigma}) e_K, \tag{2.46}$$

Thanks to the property of conservativity, one has $R_{K,\sigma} = -R_{L,\sigma}$ and $r_{K,\sigma} = -r_{L,\sigma}$ for $\sigma \in \mathcal{E}_{\text{int}}$ such that $\sigma = \sigma_{KL}$. Let $R_\sigma = |R_{K,\sigma}|$ and $r_\sigma = |r_{K,\sigma}|$ if $\sigma \in \mathcal{E}_K$. Reordering the summation over the edges and from the Cauchy-Schwarz inequality, one then obtains

$$
\begin{gathered}
|\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}|\sigma|(R_{K,\sigma}+r_{K,\sigma})e_K| \le \sum_{\sigma\in\mathcal{E}}|\sigma|(D_\sigma e)(R_\sigma+r_\sigma) \le \\
\Big(\sum_{\sigma\in\mathcal{E}}\frac{|\sigma|}{d_\sigma}(D_\sigma e)^2\Big)^{\frac{1}{2}}\Big(\sum_{\sigma\in\mathcal{E}}|\sigma|d_\sigma(R_\sigma+r_\sigma)^2\Big)^{\frac{1}{2}}.
\end{gathered}
\tag{2.47}
$$

Now, since $|R_\sigma + r_\sigma| \le C_1 h_{\mathcal{D}}$ and since $\sum_{\sigma\in\mathcal{E}}|\sigma|d_\sigma = d$
$vert\Omega|$, (2.46) and (2.47) yield the existence of $C_4 \in \mathbb{R}_+$ only depending on $u, \mathbf{v}$ and Ω such that

$$
\|e_{\mathcal{D}}\|_{1,\mathcal{M}}^2 + \frac{1}{2}b\|e_{\mathcal{D}}\|_{L^2(\Omega)}^2 \le C_3(h_{\mathcal{D}})^2 + C_4 h_{\mathcal{D}}\|e\|_{1,\mathcal{M}}.
$$

Using again Young's inequality, there exists C_5 only depending on u, $\mathbf{v}$, b and Ω such that

$$
\|e_{\mathcal{D}}\|_{1,\mathcal{M}}^2 + b\|e_{\mathcal{D}}\|_{L^2(\Omega)}^2 \le C_5(h_{\mathcal{D}})^2. \tag{2.48}
$$

This inequality yields Estimate (2.37) and, in the case $b > 0$, Estimate (2.38). In the case where $b = 0$, one uses the discrete Poincaré inequality (2.5) and the inequality (2.48) to obtain

$$
\|e_{\mathcal{D}}\|_{L^2(\Omega)}^2 \le \mathrm{diam}(\Omega)^2 C_5(h_{\mathcal{D}})^2,
$$

which yields (2.38).

Remark now that (2.37) can be written

$$
\begin{gathered}
\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=\sigma_{KL}}} |\sigma|d_\sigma\Big(\frac{u_L-u_K}{d_\sigma} - \frac{u(x_L)-u(x_K)}{d_\sigma}\Big)^2 + \\
\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{ext}}\\ \sigma\in\overline{K}\cap\partial\Omega}} |\sigma|d_\sigma\Big(\frac{g(y_\sigma)-u_K}{d_\sigma} - \frac{u(y_\sigma)-u(x_K)}{d_\sigma}\Big)^2 \le (Ch_{\mathcal{D}})^2.
\end{gathered}
\tag{2.49}
$$

From Definition (2.41) and the consistency of the fluxes, one has

$$
\begin{gathered}
\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\ \sigma=\sigma_{KL}}} |\sigma|d_\sigma\Big(\frac{u(x_L)-u(x_K)}{d_\sigma} - \frac{1}{|\sigma|}\int_\sigma \nabla u(x)\cdot\mathbf{n}_{K,\sigma}\mathrm{d}\gamma(x)\Big)^2 + \\
\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{ext}}\\ \sigma\in\overline{K}\cap\partial\Omega}} |\sigma|d_\sigma\Big(\frac{u(y_\sigma)-u(x_K)}{d_\sigma} - \frac{1}{|\sigma|}\int_\sigma \nabla u(x)\cdot\mathbf{n}_{K,\sigma}\mathrm{d}\gamma(x)\Big)^2 = \\
\sum_{\sigma\in\mathcal{E}}|\sigma|d_\sigma R_\sigma^2 \le d \\
vert\Omega|C_1^2(h_{\mathcal{D}})^2.
\end{gathered}
\tag{2.50}
$$

Then (2.49) and (2.50) give (2.39). ∎

Chapter 3

Discretization of the gradient, anisotropic diffusion problems

Most of the notions presented in this chapter are studied in a joint work with Robert Eymard and Thierry Gallouët, which is submitted for publication [28].

3.1 Introduction

The approximation of convection diffusion problems in anisotropic media is an important issue in several engineering fields. Let us briefly review four particular situations where the discretization of a non diagonal second order operator is required:

1. In the case of a contaminant transported by a one-phase flow, one must account for the diffusion-dispersion operator $\mathrm{div}(\Lambda\nabla u)$, where the matrix $\Lambda(x) = \lambda(x)\mathrm{I}_d + \mu(x)\mathbf{q}(x)\cdot\mathbf{q}(x)^t$ depends on the space variable x and $\mathbf{q}(x)$ is the velocity of the fluid flow in the porous medium. The real parameter $\lambda(x)$ corresponds to a resulting isotropic diffusion term, including dispersion in the directions orthogonal to the flow, and the real parameter $\mu(x)$ to an additional diffusion in the direction of the flow [14]. The term $\mathbf{q}(x)$ is then given by $\mathbf{q}(x) = K(x)\nabla p(x)$, where $p(x)$ is a pressure and $K(x)$ another non diagonal matrix (the absolute permeability matrix, depending on the geological layers), and satisfies the incompressibility equation $\mathrm{div}\mathbf{q}(x) = 0$. In this coupled problem, one must simultaneously compute this pressure and the contaminant concentration $u(x)$.

2. In the study of under saturated flows in porous media (for example, air-water flows), two equations of conservation have to be solved, associated with two unknowns, pressure and saturation. These equations include nonlinear hyperbolic and degenerate parabolic terms with respect to the saturation unknown. As in the preceding case, one must discretize such terms as $\mathrm{div}\mathbf{q}(x) = \mathrm{div}(K(x)\nabla p(x))$, where again $K(x)$ is a non diagonal matrix depending on the geological layers.

3. In the case of the compressible Navier-Stokes equations, one has to discretize the viscous forces operator, which can be written under the form $a\Delta\mathbf{u} + b\nabla\mathrm{div}\mathbf{u}$ (a and b are deduced from the dynamic viscosity coefficients and $\mathbf{u}$ is the fluid velocity). In this problem, the term $\nabla\mathrm{div}\mathbf{u}$ involves all the cross derivatives $\partial^2_{ij}\mathbf{u}$.

4. Some problems arising in financial mathematics lead to anisotropic diffusion equations in high-dimensional domains (dimension equal to 5 or more for example). Under some assumptions on financial markets [54], the price of a European or an American option is obtained by solving a linear or nonlinear partial differential equation, involving the second order anisotropic diffusion matrix $\Lambda = \Sigma\Sigma^t$, where Σ is a real matrix.

All these cases involve a term under the form $\mathrm{div}(\Lambda\nabla u)$, where Λ is a (generally) non diagonal matrix depending on the space variable and u is a function of the space variable in steady problems, and of the space and time variables in transient problems. Finite element schemes are known to allow for an easy discretization of such a term on triangular or tetrahedral meshes [65]. However, in engineering situations such as the ones described above, one also has to discretize convection and reaction terms, and avoid numerical instabilities. Unfortunately, finite element methods (and more generally centered schemes) are known to generate instabilities on coarse grids, although some cures may be proposed, see [38; 3]; therefore a great many numerical codes [1; 2; 38; 52; 53] use finite volume or finite volume - finite element type schemes, which allow the implementation of discretization techniques (such as the classical upwind schemes) which prevent the apparition of instabilities. Let us also note that finite volume schemes are known for their simplicity of implementation, particularly so when discretizing coupled systems of equations of various nature.

Besides, we saw in the previous chapter that finite volume methods are well suited and convergent for a convection diffusion equation in the case where $\Lambda(x) = \lambda(x)\ \mathrm{I}_d$. But the situation is quite different in the case where the condition $\Lambda(x) = \lambda(x)\ \mathrm{I}_d$ no longer holds: only few of the actual discretization methods used for handling non diagonal second order terms on finite volume grids meet a full mathematical analysis of stability or convergence. Let us briefly review some of them. A first one consists in adapting the above orthogonality condition by stating that the line joining two cell centres is orthogonal to the interface between the two cells with respect to the dot product induced by the matrix Λ^{-1} (see [48] and [24] section 11 page 815). In the case of triangular grids, this yields a well defined scheme under some restriction on the allowed anisotropy for a given geometry, since the cell centre may be chosen as the intersection of the orthogonal bisectors of the triangle for the metric defined by Λ^{-1}. However, on meshes other than triangular, the scheme is not easily defined anymore, since there is no straightforward way to find the cell points satisfying the Λ^{-1} orthogonality property. Another method consists in defining the finite volume method as a dual method to a finite element one (for example, a P1 finite element [14] or a Crouzeix-Raviart one, see e.g. [35]).

Another possibility to derive a finite volume scheme on problems including anisotropic diffusion is to construct a local discrete gradient, allowing to get, at each edge σ of the mesh, a consistent approximate value for the flux $\int_\sigma (\Lambda(x)\nabla u(x)) \cdot \mathbf{n}_\sigma \mathrm{d}\gamma(x)$ involved in the finite volume scheme ($\mathbf{n}_\sigma$ is a unit vector normal to the edge σ, and $\mathrm{d}\gamma(x)$ is the $d-1$ Lebesgue measure on the edge σ). In two space dimensions, such a scheme was introduced in [18] on arbitrary meshes, but the proof of convergence was only possible on meshes close to parallelograms. Still in 2D, a technique using dual meshes is introduced in [50; 21], which generalises the idea of [59; 51] for div-curl problems to meshes with no orthogonality conditions; however the use of a dual mesh renders the scheme computationally expensive; moreover it does not seem to be easily extended to 3D. In [26], we used Raviart-Thomas shape functions, generalised to the case of any admissible mesh (again in the sense precised of [24], see also Definition 2.1 below), in order to define a discrete gradient for piecewise constant functions. The strong convergence of this discrete gradient was then shown in the case of the elliptic equation $-\Delta u = f$. A drawback of this definition was the difficulty to find an approximation of these generalised shape functions in other cases than triangles or rectangles.

We therefore propose in this paper a new cheap and simple method of constructing a discrete gradient for a piecewise constant function, on arbitrary admissible meshes in any space dimension (this method has been first introduced in [27]). We prove that the discrete gradients of any sequence of piecewise constant functions converging to some $u \in H_0^1(\Omega)$ weakly converges to ∇u in $L^2(\Omega)$. Moreover, the discrete gradient is shown to be consistent, in the sense that it satisfies a strong convergence property on the interpolation of regular function. In order to show the efficiency of this approximation method, we use this discrete gradient to design a scheme for the approximation of the weak solution $\bar{u}$ of the following diffusion problem with full anisotropic tensor:

$$\left\{\begin{array}{l} -\mathrm{div}(\Lambda\nabla\bar{u}) = f \text{ in } \Omega, \\ \bar{u} = 0 \text{ on } \partial\Omega, \end{array}\right. \tag{3.1}$$

under the following assumptions:

$$\Omega \text{ is an open bounded connected polygonal subset of } \mathbb{R}^d,\ d \in \mathbb{N}^\star, \tag{3.2}$$

$$\left\{\begin{array}{l} \Lambda \text{ is a measurable function from } \Omega \text{ to } \mathcal{M}_d(\mathbb{R}), \\ \text{ where } \mathcal{M}_d(\mathbb{R}) \text{ denotes the set of } d \times d \text{ matrices,} \\ \text{ such that for a.e. } x \in \Omega, \Lambda(x) \text{ is symmetric,} \\ \text{ and the set of its eigenvalues is included in } [\alpha(x), \beta(x)] \\ \text{ where } \alpha, \beta \in L^\infty(\Omega) \text{ are such that} \\ 0 < \alpha_0 \le \alpha(x) \le \beta(x) \text{ for a.e. } x \in \Omega, \end{array}\right. \tag{3.3}$$

and

$$f \in L^2(\Omega). \tag{3.4}$$

We give the classical weak formulation in the following definition.

Definition 3.1 (Weak solution) *Under hypotheses (3.2)-(3.4), we say that $\bar{u}$ is a weak solution of (3.1) if*

$$\left\{\begin{array}{l} \bar{u} \in H_0^1(\Omega), \\ \displaystyle\int_\Omega \Lambda(x)\nabla\bar{u}(x)\cdot\nabla v(x)\mathrm{d}x = \int_\Omega f(x)v(x)\mathrm{d}x, \quad \forall v \in H_0^1(\Omega). \end{array}\right. \tag{3.5}$$

Remark 3.2 *For the sake of clarity, we restrict ourselves here to the numerical analysis of Problem (3.1), however, the present analysis readily extends to convection-diffusion-reaction problems and coupled problems. Indeed, we emphasise that proofs of convergence or error estimate can easily be adapted to such situations, since the discretization methods of all these terms are independent of one another, and the treatment of convection and reaction term is well-known (see [41] or [24]).*

3.2 A discrete gradient

We present in this section a method for the approximation of the gradient of piecewise constant functions, in the case of grids satisfying some orthogonality condition as defined below.

3.2.1 Discrete functional setting

We shall work here with admissible discretizations in the sense of Definition 2.1 page 5 (see Figure 2.1 page 6), for which we need some additional notions: first present the following notion of admissible discretization, which is taken in [24]. The notations are summarised in Figure 2.1 for the particular case $d = 2$ (we recall that the case $d \ge 3$ is considered as well).

Definition 3.3 (Regularity of the mesh) *Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$, and let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be an admissible discretization of Ω in the sense of Definition 2.1 page 5. Then, for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, let:*

$$D_{K,\sigma} = \{tx_K + (1-t)y, t \in (0,1),\ y \in \sigma\},$$

For all $\sigma \in \mathcal{E}_{\rm int}$, let $K, L \in \mathcal{M}$ be such that $\sigma = \sigma_{KL}$, let $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$.
For all $\sigma \in \mathcal{E}_{\rm ext}$, let $K \in \mathcal{M}$ be such that $\sigma \in \mathcal{E}_K$; let $D_\sigma = D_{K,\sigma}$.
For all $\sigma \in \mathcal{E}$, let

$$x_\sigma = \frac{1}{|\sigma|}\int_\sigma x \ \mathrm{d}\gamma(x). \tag{3.6}$$

Finally let us define the regularity of the mesh $\mathcal{D}$ by the function $\theta_\mathcal{D}$ defined by

$$\theta_\mathcal{D} = \inf\left\{\frac{d_{K,\sigma}}{\mathrm{diam}(K)}, K \in \mathcal{M},\ \sigma \in \mathcal{E}_K\right\}. \tag{3.7}$$

Using the same notations as in Chapter 2, we define, for $(u, v) \in (H_{\mathcal{D}})^2$ and for any function $\alpha \in L^\infty(\Omega)$, the following symmetric bilinear form:

$$[u, v]_{\mathcal{D},\alpha} = \sum_{\sigma_{KL} \in \mathcal{E}_{\rm int}} \tau_{\sigma_{KL}} \alpha_{\sigma_{KL}} (u_L - u_K)(v_L - v_K) + \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_{K,\rm ext}} \tau_{K,\sigma} \alpha_\sigma u_K v_K, \tag{3.8}$$

where we set

$$\alpha_\sigma = \frac{1}{|D_\sigma|} \int_{D_\sigma} \alpha(x) \mathrm{d}x, \ \forall \sigma \in \mathcal{E}. \tag{3.9}$$

Remark 3.4 *One could also take, for α_σ, the harmonic averaging of the values in K and L when $\sigma = \sigma_{KL}$.*

Note that,

$$\left([u, u]_{\mathcal{D},1}\right)^{1/2} = \|u\|_{\mathcal{D}}$$

(where 1 denotes the constant function equal to 1), thanks to the discrete Poincaré inequality (2.7) page 8. Let us now give a relative compactness result, which is also partly stated in some other papers concerning finite volume methods [24; 49; 29], and which is useful in the proof of convergence.

Lemma 3.5 (Relative compactness in $L^2(\Omega)$) *Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$, $d \in \mathbb{N}^\star$ and let $(\mathcal{D}_n, u_n)_{n \in \mathbb{N}}$ be a sequence such that, for all $n \in \mathbb{N}$, $\mathcal{D}_n$ is an admissible finite volume discretization of Ω in the sense of Definition 3.3 and $u_n \in H_{\mathcal{D}_n}(\Omega)$ (cf Definition 3.6). Let us assume that $\lim_{n\to\infty} h_{\mathcal{D}_n} = 0$, and that there exists $C_1 > 0$ such that $\|u_n\|_{\mathcal{D}_n} \le C_1$, for all $n \in \mathbb{N}$.*
Then there exists a subsequence of $(\mathcal{D}_n, u_n)_{n \in \mathbb{N}}$, again denoted $(\mathcal{D}_n, u_n)_{n \in \mathbb{N}}$, and $\bar{u} \in H^1_0(\Omega)$ such that u_n tends to $\bar{u}$ in $L^2(\Omega)$ as $n \to +\infty$, and the inequality

$$\int_\Omega |\nabla \bar{u}(x)|^2 \mathrm{d}x \le \liminf_{n\to\infty} \|u_n\|^2_{\mathcal{D}_n} \tag{3.10}$$

holds. Moreover, for all function $\alpha \in L^\infty(\Omega)$, we have

$$\lim_{n\to\infty} [u_n, P_{\mathcal{D}_n} \varphi]_{\mathcal{D}_n,\alpha} = \int_\Omega \alpha(x) \nabla \bar{u}(x) \cdot \nabla \varphi(x) \mathrm{d}x, \ \forall \varphi \in \mathrm{C}^\infty_c(\Omega). \tag{3.11}$$

The proof of (3.11) is similar to the that of Lemma 2.20, that is, the term T_2 in the proof of Theorem 2.18 page 15. The proof or (3.10) can be seen in [49; 29].

3.2.2 Definition of a discrete gradient

We now define a discrete gradient for piecewise constant functions on an admissible discretization.

Definition 3.6 (Discrete gradient) *Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$, $d \in \mathbb{N}^\star$. Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be an admissible finite volume discretization of Ω in the sense of Definition 3.3. The main idea of the definition of our discrete gradient is to remark that for any vector $v \in \mathbb{R}^d$ and any point $x_0 \in \mathbb{R}^d$, one has:*

$$v = \frac{1}{|K|} \sum_{\sigma \in \mathcal{E}_K} |\sigma|(x_\sigma - x_0)\ (\mathbf{n}_{K,\sigma} \cdot v), \ \forall K \in \mathcal{M}, \ \forall x_0 \in \mathbb{R}^d, \ \forall v \in \mathbb{R}^d. \tag{3.12}$$

Hence, if we know the outward normal gradient $\varphi_{K,\sigma} = \nabla w \cdot \mathbf{n}$ of a given function along the edges of a control volume K, we may then reconstruct the whole gradient as:

$$\nabla w = \frac{1}{|K|} \sum_{\sigma \in \mathcal{E}_K} |\sigma|(x_\sigma - x_0)\ \varphi_{K,\sigma}.$$

Note that for this reconstruction, one may use any point x_0 of the cell K. However, in order to obtain a property of weak convergence of the discrete gradient, which will be crucial when using the discrete gradient for anisotropic problems, we need some consistency of the discrete gradient applied to test functions; this consistency holds if one chooses for x_0 a point x_K satisfying the usual orthogonality conditions, as we shall see later. Hence we define, for all $K \in \mathcal{M}$, for all $L \in \mathcal{N}_K$, the following (vector) coefficients:

$$A_{K,L} = \tau_{\sigma_{KL}}(x_{\sigma_{KL}} - x_K), \tag{3.13}$$

and for all $\sigma \in \mathcal{E}_{K,\text{ext}}$, we define

$$A_{K,\sigma} = \tau_\sigma(x_\sigma - x_K). \tag{3.14}$$

and the discrete gradient $\nabla_\mathcal{D} \ : H_\mathcal{D} \to H_\mathcal{D}^d$, for any $u \in H_\mathcal{D}$, by:

$$\left\{ \begin{array}{ll} \nabla_\mathcal{D} u(x) & = (\nabla_\mathcal{D} u)_K \\ & = \dfrac{1}{|K|} \left(\displaystyle\sum_{L\in\mathcal{N}_K} A_{K,L}\ (u_L - u_K) - \sum_{\sigma\in\mathcal{E}_{K,\text{ext}}} A_{K,\sigma}\ u_K \right), \\ & \textit{for a.e. } x \in K,\ \forall K \in \mathcal{M}. \end{array} \right. \tag{3.15}$$

Let us first state a bound for the $L^2(\Omega)^d$ norm of the discrete gradient of any element of $H_\mathcal{D}$, with respect to the $H_\mathcal{D}$ norm of this element. We refer to [28] for the proof of this result.

Lemma 3.7 (Bound for $\nabla_\mathcal{D} u$) *Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$, $d \in \mathbb{N}^\star$, let $\mathcal{D}$ be an admissible finite volume discretization of Ω in the sense of Definition 3.3 and let $\theta \in (0, \theta_\mathcal{D}]$. Then, there exists C_2 , only depending on d and θ, such that, for all $u \in H_\mathcal{D}$:*

$$\|\nabla_\mathcal{D} u\|_{L^2(\Omega)^d} \le C_2\, \|u\|_\mathcal{D}. \tag{3.16}$$

We now state a weak convergence property for the discrete gradient. We again refer to [28] for the proof.

Lemma 3.8 (Weak convergence of the discrete gradient)
Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$, $d \in \mathbb{N}^\star$, let $\mathcal{D}$ be an admissible finite volume discretization of Ω in the sense of Definition 3.3. We assume that there exist $u_\mathcal{D} \in H_\mathcal{D}$ and a function $\bar{u} \in H_0^1(\Omega)$ such that $u_\mathcal{D}$ tends to $\bar{u}$ in $L^2(\Omega)$ as $h_\mathcal{D}$ tends to 0 while $\|u_\mathcal{D}\|_\mathcal{D}$ remains bounded. Then $\nabla_\mathcal{D} u_\mathcal{D}$ weakly tends to $\nabla\bar{u}$ in $L^2(\Omega)^d$ as $h_\mathcal{D} \to 0$.

We now study, for a regular function φ, the strong convergence of the discrete gradient $\nabla_\mathcal{D} P_\mathcal{D}\varphi$ to $\nabla\varphi$. This is in fact a consistency property, which uses the identity (3.12).

Lemma 3.9 (Consistency property of the discrete gradient) *Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$, $d \in \mathbb{N}^\star$, let $\mathcal{D}$ be an admissible finite volume discretization in the sense of Definition 3.3 and let $\theta \in (0, \theta_\mathcal{D}]$. Let $\bar{u} \in \mathrm{C}^2(\overline{\Omega})$ be such that $\bar{u} = 0$ on the boundary of Ω. Then, there exists C_3 , only depending on Ω, θ and $\bar{u}$, such that:*

$$\|\nabla_\mathcal{D} P_\mathcal{D}\bar{u} - \nabla\bar{u}\|_{L^2(\Omega)^d} \le C_3\ \ h_\mathcal{D}. \tag{3.17}$$

(Recall that $P_\mathcal{D}$ is defined by (2.3) and $\nabla_\mathcal{D}$ in Definition 3.6.)

Remark 3.10 (Choice of the points x_K and x_σ) *Note that in the proof of Lemma 3.8, one is free to choose any point lying on σ_{KL} instead of $x_{\sigma_{KL}}$ in the definition of the coefficients $A_{K,L}$. However, we need this choice in the proof of the strong consistency of the discrete gradient (Lemma 3.9). Conversely, in the proof of Lemma 3.9, we could take any point of K instead of x_K in the definition of $A_{K,L}$. However, the choice of x_K is crucial in the proof of Lemma 3.8: indeed, one needs the property of consistency of the normal flux, which follows from the fact that $\mathbf{n}_{K,L} = \frac{x_L - x_K}{d_{\sigma_{KL}}}$.*

Let us end this section on the discrete gradient by stating a convergence property which allows the proof of convergence of the discrete gradient of the solution to the "classical" finite volume scheme presented in Chapter 2, in the case of the Laplace equation.

Lemma 3.11 (A sufficient condition for the strong convergence of the discrete gradient) *Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$, $d \in \mathbb{N}^\star$, let $\theta > 0$ and let $\mathcal{D}$ be an admissible finite volume discretizations in the sense of Definition 3.3, such that $\theta_{\mathcal{D}} \geq \theta$. Assume that there exists a function $u_{\mathcal{D}} \in H_{\mathcal{D}}$ and a function $\bar{u} \in H_0^1(\Omega)$ such that $u_{\mathcal{D}}$ tends to $\bar{u}$ in $L^2(\Omega)$ as $h_{\mathcal{D}}$ tend to 0. Assume also that there exists a function $\alpha \in L^\infty(\Omega)$ and $\alpha_0 > 0$ such that $\alpha(x) \geq \alpha_0$ for a.e. $x \in \Omega$ and $[u_{\mathcal{D}}, u_{\mathcal{D}}]_{\mathcal{D},\alpha}$ tends to $\int_\Omega \alpha(x)\nabla\bar{u}(x)^2 \mathrm{d}x$ as $h_{\mathcal{D}}$ tends to 0. Then $\nabla_{\mathcal{D}} u_{\mathcal{D}}$ tends to $\nabla\bar{u}$ in $L^2(\Omega)^d$ as $h_{\mathcal{D}}$ tends to 0.*

Again, we refer to [28] for the details of the proof of this result. Let us mention that it is a consequence of the bound on the gradient obtained in Lemma 3.7, and of the relative compactness in L^2 (Lemma (3.5)).

Remark 3.12 *Note that the proof of Lemma 3.11 does not use the weak convergence of the discrete gradient, and therefore any point of K can be taken instead of x_K in the definition of the coefficients $A_{K,L}$. We thus find that the average value in K of the gradient defined in [26] is also strongly convergent (the average of this gradient, defined by the generalized Raviart-Thomas basis functions, is obtained by replacing x_K by the barycenter of K in the definition of $A_{K,L}$). Note that the drawback of the generalisation of the Raviart-Thomas basis was the difficulty for computing approximate values of the gradients. This drawback no longer exists for an averaged gradient. Nevertheless, the properties of convergence of the finite volume method shown here for non isotropic problems are only proven for the choice (3.13) in the definition of $A_{K,L}$, and not for the Raviart-Thomas basis.*

3.3 Discretization of anisotropic problems

3.3.1 The finite volume scheme

Under hypotheses (3.2)-(3.4), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 3.3. The finite volume approximation to Problem (3.1) is given as the solution of the following equation:

$$\left\{ \begin{array}{l} u_{\mathcal{D}} \in H_{\mathcal{D}}, \\ \displaystyle\int_\Omega (\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}} u_{\mathcal{D}}(x) \cdot \nabla_{\mathcal{D}} v(x)\mathrm{d}x + [u_{\mathcal{D}}, v]_{\mathcal{D},\alpha} = \int_\Omega f(x)v(x)\mathrm{d}x, \ \forall v \in H_{\mathcal{D}}, \end{array} \right. \tag{3.18}$$

denoting by I_d the identity application of $\mathbb{R}^d$. The existence and the uniqueness of the solution $u_{\mathcal{D}}$ to (3.18) will be stated in Lemma 3.14. Note that in this formulation, we use the discrete gradient on part of the the operator only, while on a homogeneous part, we write the usual cell centred scheme. This needs to be done in order to obtain the stability of the scheme, that is some *a priori* estimate on the discrete solution. If we take $\alpha = 0$ in (3.18), we are no longer able to prove the discrete H^1 estimate (3.23) below. Taking for v the characteristic function of a control volume K in (3.18), we may note that Equation (3.18) is equivalent to finding the values $(u_K)_{K\in\mathcal{M}}$ (we again denote u_K instead of $(u_{\mathcal{D}})_K$), solution of the following system of equations:

$$\sum_{L\in\mathcal{N}_K} F_{KL} + \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}} F_{K\sigma} \ = \int_K f(x)\mathrm{d}x, \ \ \forall K \in \mathcal{M}, \tag{3.19}$$

where

$$F_{KL} = \tau_{\sigma_{KL}}\alpha_{\sigma_{KL}}(u_K - u_L) + \big(\{ \ \Lambda_L A_{LK} \cdot \nabla_{\mathcal{D}} u_L - \Lambda_K A_{KL} \cdot \nabla_{\mathcal{D}} u_K \ \big) \ \forall \sigma_{KL} \in \mathcal{E}_{\mathrm{int}}, \tag{3.20}$$

and

$$F_{K\sigma} = \tau_{K\sigma}\alpha_\sigma u_K + \Lambda_K A_{K\sigma} \cdot \nabla_{\mathcal{D}} u_K \ \forall \sigma \in \mathcal{E}_{K,\mathrm{ext}}. \tag{3.21}$$

In (3.20) and (3.21), the matrices $(\Lambda_K)_{K\in\mathcal{M}}$ are defined by:

$$\Lambda_K^{(ij)} = \frac{1}{|K|}\int_K (\Lambda(x) - \alpha(x)\mathrm{I}_d)\mathbf{e}^{(i)} \cdot \mathbf{e}^{(j)}\mathrm{d}x,\ i = 1,\ldots,d,\ j = 1,\ldots,d. \tag{3.22}$$

On can then complete the discrete expressions of F_{KL} and $F_{K\sigma}$ using Definition 3.6 for A_{KL} $A_{K\sigma}$, and $\nabla_{\mathcal{D}}u_K$ for all $K \in \mathcal{M}$, $L \in \mathcal{N}_K$ and $\sigma \in \mathcal{E}_K$.
This is indeed a finite volume scheme, since

$$F_{KL} = -F_{LK},\ \forall\sigma_{KL} \in \mathcal{E}_{\mathrm{int}}.$$

The existence of a solution to (3.18) will be proven below.

3.3.2 Discrete $H^1(\Omega)$ estimate

We now prove the following estimate:

Lemma 3.13 [Discrete H^1 estimate] *Under hypotheses (3.2)-(3.4), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 3.3. Let $u \in H_{\mathcal{D}}$ be a solution to (3.18). Then the following inequalities hold:*

$$\alpha_0\|u\|_{\mathcal{D}} \le \mathrm{diam}(\Omega)\|f\|_{(L^2(\Omega))^2}, \tag{3.23}$$

Proof We apply (3.18) setting $v = u$. We get

$$\int_\Omega (\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}u(x)\cdot\nabla_{\mathcal{D}}u(x)\mathrm{d}x + [u,u]_{\mathcal{D},\alpha} = \int_\Omega f(x)u(x)\mathrm{d}x,$$

which implies

$$\alpha_0[u,u]_{\mathcal{D}} \le \int_\Omega f(x)u(x)\mathrm{d}x.$$

Then the conclusion follows from the discrete Poincaré inequality (2.7).
We can now state the existence and the uniqueness of a discrete solution to (3.18).

Corollary 3.14 [Existence and uniqueness of a solution to the finite volume scheme] *Under hypotheses (3.2)-(3.4), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 3.3. Then there exists a unique $u_{\mathcal{D}}$ solution to (3.18).*

Proof System (3.18) is a linear system. Assume that $f = 0$. From the discrete Poincaré inequality (2.7), we get that $u = 0$. This proves that the linear system (3.18) is invertible.

3.3.3 Convergence

We have the following result, which states the convergence of the scheme (3.18).

Theorem 3.15 [Convergence of the finite volume scheme] *Under hypotheses (3.2)-(3.4), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 3.3. Let $u_{\mathcal{D}} \in H_{\mathcal{D}}(\Omega)$ be the solution to (3.18). Then*

- *$u_{\mathcal{D}}$ converges in $L^2(\Omega)$ to $\bar{u}$, weak solution of Problem (3.1) in the sense of Definition 4.6,*
- *the discrete gradient $\nabla_{\mathcal{D}}u_{\mathcal{D}}$ converges in $L^2(\Omega)^d$ to $\nabla\bar{u}$,*

as $h_{\mathcal{D}}$ tends to 0, under the condition that there exists $\theta > 0$ with $\theta_{\mathcal{D}} \ge \theta$.

PROOF We consider a sequence of admissible discretizations $(\mathcal{D}_n)_{n\in\mathbb{N}}$ such that $h_{\mathcal{D}_n}$ tend to 0 as $n\to\infty$ under the condition that there exists $\theta>0$ with $\theta_{\mathcal{D}_n}\ge\theta$ for all $n\in\mathbb{N}$. Thanks to Lemma 3.13, we can apply the compactness result (3.5), which gives the existence of a subsequence (again denoted $(\mathcal{D}_n)_{n\in\mathbb{N}}$), and of $\bar u\in H_0^1(\Omega)$ such that $u_{\mathcal{D}_n}$ (given by (3.18) with $\mathcal{D}=\mathcal{D}_n$) tends to $\bar u$ in $L^2(\Omega)$ as $n\to\infty$. Let $\varphi\in C_c^\infty(\Omega)$ be given, we choose $v=P_{\mathcal{D}_n}\varphi$ as test function in (3.18). We obtain

$$\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}_n}u_{\mathcal{D}_n}(x)\cdot\nabla_{\mathcal{D}_n}P_{\mathcal{D}_n}\varphi(x)\mathrm{d}x+[u_{\mathcal{D}_n},P_{\mathcal{D}_n}\varphi]_{\mathcal{D}_n,\alpha}=\int_\Omega f(x)P_{\mathcal{D}_n}\varphi(x)\mathrm{d}x. \tag{3.24}$$

We let $n\to\infty$ in (3.24). Thanks to Lemma 3.8 and Lemma 3.9 (which provide a weak/strong convergence result), we get that

$$\lim_{n\to\infty}\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}_n}u_{\mathcal{D}_n}(x)\cdot\nabla_{\mathcal{D}_n}P_{\mathcal{D}_n}\varphi(x)\mathrm{d}x=\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla\bar u(x)\cdot\nabla\varphi(x)\mathrm{d}x.$$

Using Lemma 3.5, we get that

$$\lim_{n\to\infty}[u_{\mathcal{D}_n},P_{\mathcal{D}_n}\varphi]_{\mathcal{D}_n,\alpha}=\int_\Omega\alpha(x)\nabla\bar u(x)\cdot\nabla\varphi(x)\mathrm{d}x.$$

Since it is easy to see that

$$\lim_{n\to\infty}\int_\Omega f(x)P_{\mathcal{D}_n}\varphi(x)\mathrm{d}x=\int_\Omega f(x)\varphi(x)\mathrm{d}x,$$

we thus get that any limit $\bar u$ of a subsequence of solutions satisfies (3.5) with $v=\varphi$. A classical density argument and the uniqueness of the solution to (3.5) permit to conclude to the convergence in $L^2(\Omega)$ of $u_{\mathcal{D}}$ to $\bar u$, weak solution of the problem in the sense of Definition 4.6, as $h_{\mathcal{D}}$ tends to 0, under the condition that there exists $\theta>0$ with $\theta_{\mathcal{D}}\ge\theta$. Let us now prove the strong convergence of $\nabla_{\mathcal{D}}u_{\mathcal{D}}$ to $\nabla\bar u$. We have, using (3.18) with $v=u_{\mathcal{D}}$,

$$\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\cdot\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\mathrm{d}x=\int_\Omega f(x)u_{\mathcal{D}}(x)\mathrm{d}x-[u_{\mathcal{D}},u_{\mathcal{D}}]_{\mathcal{D},\alpha}. \tag{3.25}$$

Thanks to Lemma 3.5, we have

$$\int_\Omega\alpha(x)\nabla\bar u(x)^2\mathrm{d}x\le\liminf_{h_{\mathcal{D}}\to0}[u_{\mathcal{D}},u_{\mathcal{D}}]_{\mathcal{D},\alpha},$$

and therefore, passing to the limit in (3.25), we get that

$$\limsup_{h_{\mathcal{D}}\to0}\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\cdot\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\mathrm{d}x\le\int_\Omega f(x)u_{\mathcal{D}}(x)\mathrm{d}x-\int_\Omega\alpha(x)\nabla\bar u(x)^2\mathrm{d}x.$$

We then have, letting $v=\bar u$ in (3.5),

$$\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla\bar u(x)\cdot\nabla\bar u(x)\mathrm{d}x=\int_\Omega f(x)\bar u(x)\mathrm{d}x-\int_\Omega\alpha(x)\nabla\bar u(x)^2\mathrm{d}x. \tag{3.26}$$

This leads to

$$\limsup_{h_{\mathcal{D}}\to0}\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\cdot\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\mathrm{d}x\le\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla\bar u(x)\cdot\nabla\bar u(x)\mathrm{d}x.$$

Using Lemma 3.8, which states the weak convergence of the gradient $\nabla_{\mathcal{D}}u_{\mathcal{D}}$ to $\nabla\bar u$, we get that

$$\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla\bar u(x)\cdot\nabla\bar u(x)\mathrm{d}x\le\liminf_{h_{\mathcal{D}}\to0}\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\cdot\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\mathrm{d}x.$$

The above inequalities yield

$$\lim_{h_{\mathcal{D}}\to 0}\int_\Omega (\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\cdot\nabla_{\mathcal{D}}u_{\mathcal{D}}(x)\mathrm{d}x = \int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla\bar u(x)\cdot\nabla\bar u(x)\mathrm{d}x. \tag{3.27}$$

From (3.25), (3.26) and (3.27), we thus obtain that

$$\lim_{h_{\mathcal{D}}\to 0}[u_{\mathcal{D}},u_{\mathcal{D}}]_{\mathcal{D},\alpha} = \int_\Omega \alpha(x)\nabla\bar u(x)^2\mathrm{d}x,$$

Therefore we can apply Lemma 3.11. This completes the proof of the strong convergence of the discrete gradient.

3.3.4 Error estimate

We now give an error estimate, assuming that the solution of (3.5) is in $\mathrm{C}^2(\overline{\Omega})$. Note that the same error estimate also holds under the weaker hypothesis that the solution of (3.5) is only in $H^2(\Omega)$, if $d\le 3$.

Theorem 3.16 (**C^2 error estimate**) *Assume hypotheses (3.2)-(3.4) and that Λ and α are of class C^1 on $\overline{\Omega}$. Let $\mathcal{D}$ be an admissible finite volume discretization (in the sense of Definition 3.3). Let $\theta\in(0,\theta_{\mathcal{D}}]$, where $\theta_{\mathcal{D}}$ is defined by (3.7). Let $u_{\mathcal{D}}\in H_{\mathcal{D}}$ be the solution of (3.18) and $\bar u\in H^1_0(\Omega)$ be the solution of (3.5). We assume that $\bar u\in\mathrm{C}^2(\overline{\Omega})$.*
Let us first assume that

$$\forall\sigma\in\mathcal{E}_{\mathrm{ext}},\int_\sigma \Lambda(x)\mathbf{n}_{\partial\Omega}(x)\cdot(x_\sigma-y_\sigma)\mathrm{d}\gamma(x)=0, \tag{3.28}$$

where $\mathbf{n}_{\partial\Omega}(x)$ is the unit normal vector to $\partial\Omega$ at point x, outward to Ω.
Then, there exists C_4 only depending on Ω, θ, α_0, α, β, Λ and $\|\bar u\|_{\mathrm{C}^2(\Omega)}$, such that:

$$\|u_{\mathcal{D}}-P_{\mathcal{D}}\bar u\|_{\mathcal{D}}\le C_4\,h_{\mathcal{D}}, \tag{3.29}$$

$$\|u_{\mathcal{D}}-\bar u\|_{L^2(\Omega)}\le C_4\,h_{\mathcal{D}}, \tag{3.30}$$

and

$$\|\nabla_{\mathcal{D}}u_{\mathcal{D}}-\nabla\bar u\|_{L^2(\Omega)^d}\le C_4\,h_{\mathcal{D}}. \tag{3.31}$$

Let us then assume that (3.28) no longer holds, then there exists C_5, only depending on Ω, θ, α, β, Λ and $\|\bar u\|_{H^2(\Omega)}$, such that (3.29), (3.30), (3.31) hold with $C_5\sqrt{h_{\mathcal{D}}}$ instead of $C_4\,h_{\mathcal{D}}$.

Remark 3.17 *Let us give some sufficient (and practical) conditions for (3.28) to hold :*

- *If the normal vector to $\partial\Omega$ is an eigenvector of $\Lambda(x)$ for a.e. $x\in\partial\Omega$, then (3.28) holds. Since this property is always satisfied in the isotropic case, the error estimate on the gradient (3.31) holds for the classical cell centred scheme, for any admissible mesh.*
- *If for all $\sigma\in\mathcal{E}_{\mathrm{ext}}$ with $\sigma\in\mathcal{E}_K$, the barycenter x_σ of σ is equal to the orthogonal projection z_σ of x_K on σ, then (3.28) holds. This hypothesis is naturally satisfied on admissible rectangular and triangular meshes.*

Note also that one could replace (3.28) by $|z_\sigma-x_\sigma|\le\frac{1}{\theta}\mathrm{diam}(K)(h_{\mathcal{D}})^{\frac12}$ for all $\sigma\in\mathcal{E}_{\mathrm{ext}}$.

Proof In the proof, we denote by C_i ($i\in\mathbb{N}$), various quantities only depending on Ω, θ, α_0, α, β, Λ and $\|\bar u\|_{\mathrm{C}^2(\Omega)}$.

Step 1. Let $v\in H_{\mathcal{D}}$. We first perform a computation of a consistency error, namely a bound for $|T_1(v)|$ where $T_1(v)$ is defined by:

$$\int_\Omega(\Lambda(x)-\alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}P_{\mathcal{D}}\bar u(x)\cdot\nabla_{\mathcal{D}}v(x)\mathrm{d}x+[P_{\mathcal{D}}\bar u,v]_{\mathcal{D},\alpha}=\int_\Omega f(x)v(x)\mathrm{d}x+T_1(v). \tag{3.32}$$

We first consider the second term of the left hand side of (3.32). Using classical consistency error (also used in the proof of Lemma 3.5), one has:

$$[P_{\mathcal{D}}\bar{u}, v]_{\mathcal{D},\alpha} = -\int_\Omega \mathrm{div}(\alpha\nabla\bar{u})(x)v(x)\mathrm{d}x + T_2(v), \tag{3.33}$$

with

$$|T_2(v)| \le \sum_{\sigma\in\mathcal{E}} |\sigma||R_\sigma|\delta_\sigma v,$$

where $\delta_\sigma v = |v_K - v_L|$ if $\sigma = \sigma_{KL}$ is an interior edge, $\delta_\sigma v = |v_K|$ is $\sigma \in \mathcal{E}_{\mathrm{ext}}$ and $|R_\sigma| \le C_6\, h_{\mathcal{D}}$. Using the Cauchy-Schwarz inequality, this leads to:

$$|T_2(v)| \le C_7\, h_{\mathcal{D}}\|v\|_{\mathcal{D}}. \tag{3.34}$$

We now consider the first term of the left hand side of (3.32). We have

$$\int_\Omega (\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}}P_{\mathcal{D}}\bar{u}(x)\cdot\nabla_{\mathcal{D}}v(x)\mathrm{d}x = T_3(v) + T_4(v), \tag{3.35}$$

with

$$T_3(v) = \int_\Omega (\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla\bar{u}(x)\cdot\nabla_{\mathcal{D}}v(x)\mathrm{d}x$$

and

$$|T_4(v)| \le C_8\, \|\nabla_{\mathcal{D}}P_{\mathcal{D}}\bar{u} - \nabla\bar{u}\|_{L^2(\Omega)^d}\|\nabla_{\mathcal{D}}v\|_{L^2(\Omega)^d}.$$

Using Lemma 3.9 and Lemma 3.7, we obtain

$$|T_4(v)| \le C_9\, h_{\mathcal{D}}\|v\|_{\mathcal{D}}. \tag{3.36}$$

We now compute $T_3(v)$. For $K\in\mathcal{M}$ and $\sigma\in\mathcal{E}$, let μ_K and μ_σ respectively be the mean values of $(\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla\bar{u}$ on K and σ:

$$\mu_K = \frac{1}{|K|}\int_K (\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla\bar{u}(x)\mathrm{d}x,\ \ \mu_\sigma = \frac{1}{|\sigma|}\int_\sigma (\Lambda(x) - \alpha(x)\mathrm{I}_d\nabla\bar{u}(x)\mathrm{d}\gamma(x).$$

The regularity of $\bar{u}$, Λ and α gives, for all $K\in\mathcal{M}$ and all $\sigma\in\mathcal{E}_K$ (recall that $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$):

$$|\mu_K - \mu_\sigma| \le C_{10}\, h_{\mathcal{D}}. \tag{3.37}$$

Indeed, C_{10} only depends on the L^∞-norms of Λ, α and $\nabla\bar{u}$ and on the L^∞-norms of the derivatives of Λ, α and $\nabla\bar{u}$.

We now use (3.37) in order to give a bound of $T_3(v)$ as a function of $h_{\mathcal{D}}$. Indeed, the definition of $\nabla_{\mathcal{D}}v$ leads to:

$$\left\{\begin{array}{l}
T_3(v) = \displaystyle\sum_{K\in\mathcal{M}} \mu_K\cdot|K|(\nabla_{\mathcal{D}}v)_K = \\
\displaystyle\sum_{K\in\mathcal{M}}\left(\sum_{L\in\mathcal{N}_K}\mu_K\cdot A_{K,L}\ (v_L - v_K) - \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}}\mu_K\cdot A_{K,\sigma}v_K\right) = \\
\displaystyle\sum_{K\in\mathcal{M}}\left(\sum_{L\in\mathcal{N}_K}\mu_{\sigma_{KL}}\cdot A_{K,L}\ (v_L - v_K) - \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}}\mu_\sigma\cdot A_{K,\sigma}v_K\right) + T_5(v),
\end{array}\right.$$

with

$$\left\{\begin{array}{l}
|T_5(v)| \le C_{10}\, h_{\mathcal{D}}\displaystyle\sum_{K\in\mathcal{M}}\left(\sum_{L\in\mathcal{N}_K}|A_{K,L}||v_L - v_K| + \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}}|A_{K,\sigma}||v_K|\right) \le \\
C_{10}\, h_{\mathcal{D}}\left(\displaystyle\sum_{\sigma=\sigma_{KL}\in\mathcal{E}_{\mathrm{int}}}(|A_{K,L}| + |A_{L,K}|)|v_L - v_K| + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}}|A_{K,\sigma}||v_K|\right).
\end{array}\right.$$

Since $A_{K,L} = \tau_{\sigma_{KL}}(x_{\sigma_{KL}} - x_K)$ and $A_{K,\sigma} = \tau_\sigma(x_\sigma - x_K)$, one deduces from the preceding inequality, thanks to the definition of $\theta_{\mathcal{D}}$ (which gives $\mathrm{d}(x_\sigma, x_K) \leq (d_{K,\sigma}/\theta)$ if $\sigma \in \mathcal{E}_K$) and using Cauchy-Schwarz Inequality:

$$|T_5(v)| \leq C_{11}\, h_{\mathcal{D}} \|v\|_{\mathcal{D}}. \tag{3.38}$$

We now remark that:

$$\left\{\begin{aligned} T_3(v) - T_5(v) = \sum_{K\in\mathcal{M}} \left(\sum_{L\in\mathcal{N}_K} \mu_{\sigma_{KL}} \cdot A_{K,L}\ (v_L - v_K) - \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}} \mu_\sigma \cdot A_{K,\sigma} v_K \right) = \\ \sum_{\sigma=\sigma_{KL}\in\mathcal{E}_{\mathrm{int}}} \mu_\sigma \cdot (x_L - x_K)\tau_\sigma(v_L - v_K) - \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}} \mu_\sigma \cdot (x_\sigma - x_K)\tau_\sigma v_K. \end{aligned}\right. \tag{3.39}$$

For $\sigma \in \mathcal{E}_{\mathrm{int}}$, one has $\sigma = \sigma_{KL}$ and $(x_L - x_K) = d_\sigma \mathbf{n}_{K,\sigma}$ where $\mathbf{n}_{K,\sigma}$ is the normal vector to σ exterior to K.

For $\sigma \in \mathcal{E}_{\mathrm{ext}}$, one has $\sigma \in \mathcal{E}_K$. Thanks to the fact that under homogeneous Dirichlet boundary conditions, the gradient of $\bar{u}$ is normal to the boundary, using Assumption (3.28), we get that

$$\mu_\sigma \cdot (x_\sigma - x_K)\tau_\sigma = \int_\sigma (\Lambda(x) - \alpha(x)\mathrm{I}_d \nabla\bar{u}(x) \cdot \mathbf{n}_{\partial\Omega}(x)\mathrm{d}\gamma(x).$$

Then, one deduces from (3.39):

$$T_3(v) - T_5(v) = -\int_\Omega \mathrm{div}((\Lambda - \alpha\mathrm{I}_d)\nabla\bar{u})(x)v(x)\mathrm{d}x. \tag{3.40}$$

Therefore, since $-\mathrm{div}(\Lambda\nabla\bar{u}) = f$, one has (3.32) with $T_1(v) = T_2(v) + T_4(v) + T_5(v)$. This gives, with (3.34), (3.36), (3.38):

$$|T_1(v)| \leq C_{12}\, h_{\mathcal{D}} \|v\|_{\mathcal{D}}. \tag{3.41}$$

This concludes Step 1.

Step 2.

Let $e_{\mathcal{D}} = P_{\mathcal{D}}\bar{u} - u_{\mathcal{D}}$ be the discrete discretization error. Using (3.32) and (3.18) give, for all $v \in H_{\mathcal{D}}$:

$$\int_\Omega (\Lambda(x) - \alpha(x)\mathrm{I}_d)\nabla_{\mathcal{D}} e_{\mathcal{D}}(x) \cdot \nabla_{\mathcal{D}} v(x)\mathrm{d}x + [e_{\mathcal{D}}, v]_{\mathcal{D},\alpha} = T_1(v).$$

Taking $v = e_{\mathcal{D}}$ in this formula gives, with (3.41), $[e_{\mathcal{D}}, e_{\mathcal{D}}]_{\mathcal{D},\alpha} \leq C_{12}\, h_{\mathcal{D}} \|e_{\mathcal{D}}\|_{\mathcal{D}}$ and then, with $C_{13} = C_{12}\,/\alpha_0$ (since $\alpha_0\|e_{\mathcal{D}}\|^2_{\mathcal{D}} \leq [e_{\mathcal{D}}, e_{\mathcal{D}}]_{\mathcal{D},\alpha}$):

$$\|e_{\mathcal{D}}\|_{\mathcal{D}} \leq C_{13}\, h_{\mathcal{D}}, \tag{3.42}$$

which is exactly (3.29).

Using the Discrete Poincaré Estimate (2.7) and the fact that $\bar{u} \in C(\overline{\Omega})$, one deduces (3.30) from (3.29).

The last estimate, Estimate (3.31), is a direct consequence of (3.42), (3.17) and (3.16). This concludes the first part of the theorem, *i.e.* assuming (3.28).

If $\mathcal{D}$ no longer satisfies the hypothesis (3.28), one has to replace (3.40) by:

$$T_3(v) - T_5(v) = -\int_\Omega \mathrm{div}((\Lambda - \alpha\mathrm{I}_d)\nabla\bar{u})(x)v(x)\mathrm{d}x + T_6(v),$$

where, recalling that by z_σ the orthogonal projection of x_K on σ (see Definition 3.3):

$$T_6(v) = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}} \mu_\sigma \cdot (z_\sigma - x_\sigma)\tau_\sigma v_K.$$

Thanks to the Cauchy-Schwarz inequality, we get

$$T_6(v)^2 \le \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}} \tau_\sigma \mu_\sigma^2 (\mathrm{diam}(K))^2 \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_{K,\mathrm{ext}}} \tau_\sigma v_K^2,$$

which leads to

$$T_6(v)^2 \le \frac{h_\mathcal{D}}{\theta} |\partial\Omega| \|\nabla \bar{u}\|_\infty^2 \|v\|_\mathcal{D}^2,$$

where $|\partial\Omega|$ is the $d-1$-dimensional Lebesgue measure of $\partial\Omega$. This gives (3.41) with $h_\mathcal{D}^{\frac{1}{2}}$ instead of $h_\mathcal{D}$. Following Step 2, this allows to conclude the proof.

An error estimate also holds when the solution of (3.5) is in $H^2(\Omega)$ instead of $\mathrm{C}^2(\overline{\Omega})$, in the case where the space dimension is lower or equal to 3, with an additional assumption on the mesh (the number of edges of a given cell must stay bounded as the mesh is refined). The proof of this estimate is similar to that of Theorem 3.16, once the following consistency lemma has been proven (see [28]):

Lemma 3.18 (Consistency of the gradient, $\bar{u} \in H^2(\Omega)$) *Under hypothesis (3.2), with $d \le 3$, let $\mathcal{D}$ be an admissible finite volume discretization in the sense of Definition 3.3, and let $\theta \in (0, \theta_\mathcal{D}]$. Let $\bar{u} \in H^2(\Omega) \cap H_0^1(\Omega)$. Then, there exists C_{14} , only depending on Ω, θ and $\bar{u}$, such that:*

$$\|\nabla_\mathcal{D}(P_\mathcal{D}\bar{u}) - \nabla\bar{u}\|_{L^2(\Omega)^d} \le C_{14}\, h_\mathcal{D} \|\bar{u}\|_{H^2(\Omega)}. \tag{3.43}$$

(Recall that $P_\mathcal{D}$ is defined in (2.3) and $\nabla_\mathcal{D}$ in Definition 3.6.)

3.4 Numerical results

The scheme was tried for various academic problems, for which the analytical solution is known. For the Laplace equation, we compared the classical cell centred scheme to the new scheme, which we shall call the gradient scheme in the sequel. First note that in the classical cell centred scheme, the equation relative to a given cell involves the neighbours of this cell, while in the gradient scheme, it involves the neighbours of this cell and the neighbours of the neighbours. Hence in the case of a rectangular (resp. parallelipedic) mesh, the classical cell centred scheme is a 5 points (resp. 7 points) scheme, while the gradient scheme is a 13 points (resp. 24 points) scheme, Similarly, if one uses a triangular (resp. tetrahedral) mesh the classical scheme is a 4 points (resp. 7 points) scheme, while the gradient scheme is a 10 points (resp. at most 17 points) scheme. Hence the gradient scheme is more expensive in terms of time and memory, although this is not so much, for example compared to the use of a Q^1 finite element in the case of a parallelipedic mesh, which leads to a 27 points scheme.
We tested the gradient scheme for some real anisotropic problems, the number of cells varying from 100 to 6400 in the rectangular meshes case (in fact, rectangles are squares), and from 700 to 17500 in the triangular meshes case. The convergence rates have been computed by fitting a less-square regression on the logarithmic values of the errors and of the characteristic size of the mesh.

The first case is an anisotropic homogeneous problem with diffusion matrix

$$\Lambda = \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}.$$

The second case is a rotating permeability field, that is, the diffusion matrix is constant in the (r, θ) coordinates and equal to $\Lambda_{r,\theta} = \begin{pmatrix} 10 & .2 \\ .2 & 10 \end{pmatrix}$. The exact solution is taken to be $u(x_1, x_2) = \frac{1}{2}\ln((x_1 - .5)^2 + (x_2 - 1.1)^2)$, on the domain $\Omega =]0,1[\times]0,1[$. The orders of convergence which were found are given Table 3.1.
Next, we tested different values of α to see how it affected the discretization error, on the first anisotropic case. Although the value of α does influence the resulting discretization error, the optimal value seems

h

	Case 1 homogeneous anisotropic		Case 2 heterogeneous anisotropic	
	Rectangles FV 13	Triangles VF10	Rectangles FV 13	Triangles VF10
u	2.00	2.0	2.2	2.0
∇u	1.00	1.0	1.4	1.3

Table 3.1: Rates of convergence of FV13 and FV10 in a homogeneous anisotropic case and in a heterogeneous anisotropic case

to be independent on the mesh, in both the triangular and rectangular cases, see Figure 3.1. Note that in the case of the error on the solution itself, the numerical optimal values for α are beyond the interval of convergence assumed in the theoretical analysis $(0, 1)$.
These numerical tests therefore indicate that this use of an discrete gradient in finite volume schemes leads to a correct numerical behaviour, indeed comparable with low degree finite element schemes on similar problems.
Finally, we replaced the point x_K by the center of gravity of cell K in the definition (3.13),(3.14) of the coefficients $A_{K,L}$. In this case, we recall (see Remark 3.12) that we obtain the discrete gradient based on the generalized Raviart-Thomas basis functions of [26]. Indeed, the tests performed with this scheme for Case 1 or Case 2 did not yield correct approximations of the solution nor of its gradient.

3.5 Conclusion

In this paper, we constructed a discrete gradient for piecewise constant functions. This discrete gradient revealed several advantages: it is easy and cheap to compute, and it provides simple schemes for the approximation of anisotropic diffusion convection problems. We showed a weak property convergence of this discrete gradient to the gradient of the limit of the considered functions, together with a consistency property, both leading to the strong convergence of the discrete solution and of its discrete gradient in the case of a Dirichlet problem with full matrix diffusion.
Since this notion of admissible mesh includes Voronoï meshes, which are more and more used in practice, and which seem to remain tractable even in high space dimension, applications to financial mathematics problems are being studied [8]. Applications to finite volume schemes for compressible Navier-Stokes equations are also expected to be successful, see chapter 4 and [67]. Further work includes a parametric study, and the generalisation to meshes without the orthogonality condition.

htb

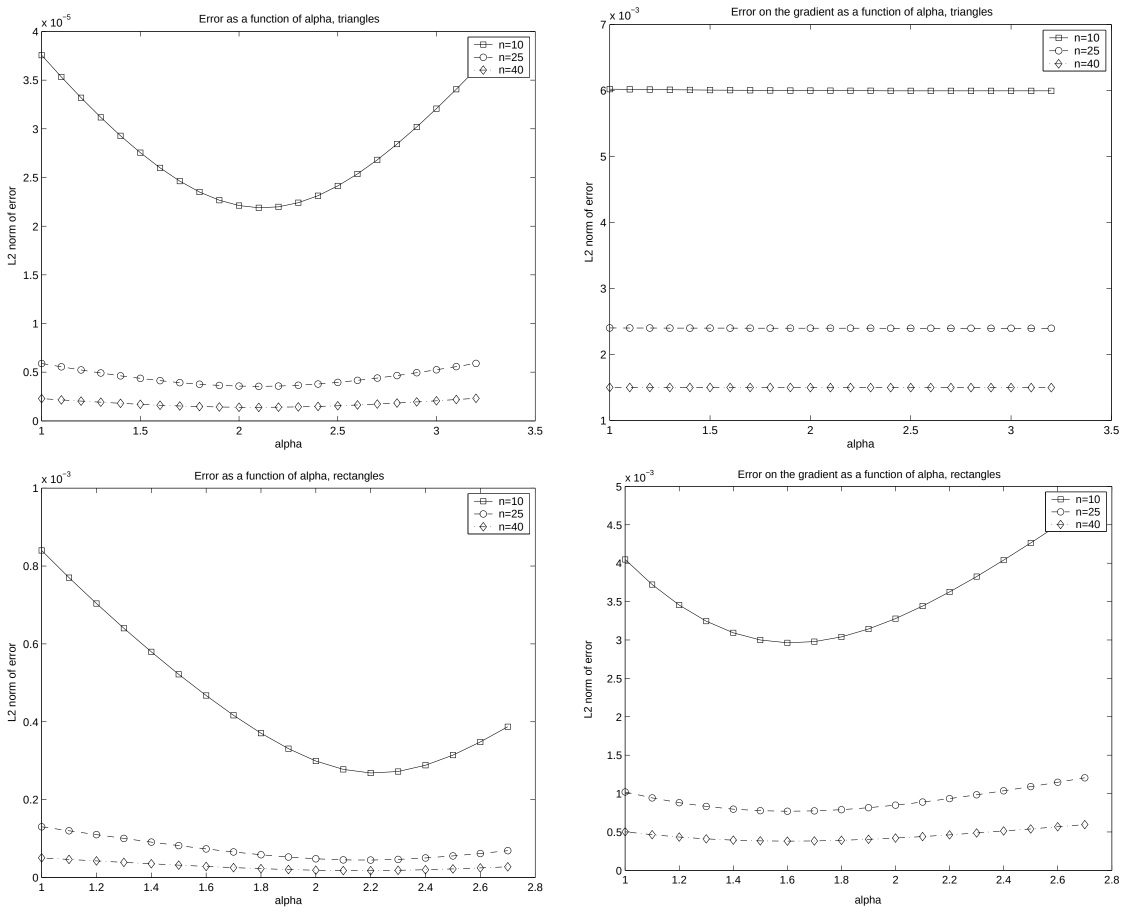

Figure 3.1: Diagrams of the errors on the solution (left) and its gradient (right) for various sizes of triangular (up) and rectangular (bottom) meshes, with respect to the value of the parameter α

Chapter 4

A colocated scheme for the Navier-Stokes equations

This chapter presents recent results obtained with Robert Eymard and Jean-Claude Latché [33]. We are interested here in finding an approximation of the fields $\bar{u} = (\bar{u}^{(i)})_{i=1,\ldots,d} : \Omega \times [0,T] \to \mathbb{R}^d$, and $\bar{p} : \Omega \times [0,T] \to \mathbb{R}$, weak solution to the incompressible Navier-Stokes equations which write:

$$\begin{cases} \partial_t \bar{u}^{(i)} - \nu \Delta \bar{u}^{(i)} + \partial_i \bar{p} + \displaystyle\sum_{j=1}^{d} \bar{u}^{(j)} \partial_j \bar{u}^{(i)} = f^{(i)} \text{ in } \Omega \times (0,T), \quad \text{for } i = 1,\ldots,d, \\ \mathrm{div}\bar{u} = \displaystyle\sum_{i=1}^{d} \partial_i \bar{u}^{(i)} = 0 \text{ in } \Omega \times (0,T). \end{cases} \tag{4.1}$$

with a homogeneous Dirichlet boundary condition for $\bar{u}$ and the initial condition

$$\left\{ \bar{u}^{(i)}(\cdot,0) = \bar{u}^{(i)}_{\mathrm{ini}} \text{ in } \Omega \text{ for } i = 1,\ldots,d. \right. \tag{4.2}$$

In the above equations, $\bar{u}^{(i)}$, $i = 1,\ldots,d$ denote the components of the velocity of a fluid which flows in a domain Ω during the time $(0,T)$, $\bar{p}$ denotes the pressure, $\nu > 0$ stands for the viscosity of the fluid. We make the following assumptions:

$$\Omega \text{ is a polygonal open bounded connected subset of } \mathbb{R}^d,\ d = 2 \text{ or } 3, \tag{4.3}$$

$$T > 0 \text{ is the finite duration of the flow}, \tag{4.4}$$

$$\nu \in (0,+\infty), \tag{4.5}$$

$$\bar{u}_{\mathrm{ini}} \in L^2(\Omega)^d, \tag{4.6}$$

$$f^{(i)} \in L^2(\Omega \times (0,T)), \quad \text{for } i = 1,\ldots,d. \tag{4.7}$$

We denote by $x = (x^{(i)})_{i=1,\ldots,d}$ any point of Ω, by $|.|$ the Euclidean norm in $\mathbb{R}^d$, *i.e.*: $|x|^2 = \displaystyle\sum_{i=1}^{d} (x^{(i)})^2$ and by δx the d-dimensional Lebesgue measure $\delta x = \delta x^{(1)} \ldots \delta x^{(d)}$.

The weak sense that we consider for the Navier-Stokes equations is the following.

Definition 4.1 (Weak solution for the transient Navier-Stokes equations)
Under hypotheses (4.3)-(4.7), let the function space $E(\Omega)$ be defined by:

$$E(\Omega) := \{\bar{v} = (\bar{v}^{(i)})_{i=1,\ldots,d} \in H^1_0(\Omega)^d, \mathrm{div}\bar{v} = 0 \ a.e. \ in \ \Omega\}. \tag{4.8}$$

Then $\bar{u}$ is called a weak solution of (4.1)-(4.2) if $\bar{u} \in L^2(0,T;E(\Omega)) \cap L^\infty(0,T;L^2(\Omega)^d)$ and:

$$\left\{\begin{array}{l} \forall \varphi \in L^2(0,T;E(\Omega)) \cap C_c^\infty(\Omega \times (-\infty,T))^d, \\ \\ \quad -\displaystyle\int_0^T \int_\Omega \bar{u}(x,t) \cdot \partial_t \varphi(x,t)\, \delta x\, \delta t - \int_\Omega \bar{u}_{\rm ini}(x) \cdot \varphi(x,0)\, \delta x \\ \quad +\nu \displaystyle\int_0^T \int_\Omega \nabla \bar{u}(x,t) : \nabla \varphi(x,t)\, \delta x\, \delta t \ + \int_0^T b(\bar{u}(\cdot,t), \bar{u}(\cdot,t), \varphi(\cdot,t))\, \delta t \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad = \displaystyle\int_0^T \int_\Omega f(x) \cdot \varphi(x,t)\, \delta x\, \delta t \end{array}\right. \tag{4.9}$$

where, for all $\bar{u}, \bar{v} \in H_0^1(\Omega)^d$ and for a.e. $x \in \Omega$, we use the following notation:

$$\nabla \bar{u}(x) : \nabla \bar{v}(x) = \sum_{i=1}^d \nabla \bar{u}^{(i)}(x) \cdot \nabla \bar{v}^{(i)}(x)$$

and where the trilinear form $b(.,.,.)$ is defined, for all $\bar{u}, \bar{v}, \bar{w} \in (H_0^1(\Omega))^d$, by

$$b(\bar{u}, \bar{v}, \bar{w}) = \sum_{k=1}^d \sum_{i=1}^d \int_\Omega \bar{u}^{(i)}(x) \partial_i \bar{v}^{(k)}(x) \bar{w}^{(k)}(x)\, \delta x. \tag{4.10}$$

Remark 4.2 *From (4.9), we get that a weak solution u of (4.1)-(4.2) in the sense of Definition 4.1 satisfies $\partial_t \bar{u} \in L^{4/d}(0,T;E(\Omega)')$, and is therefore a weak solution in the classical sense, such that $\bar{u}(\cdot,0)$ is the orthogonal L^2-projection of $\bar{u}_{\rm ini}$ on $\{\bar{v} \in L^2(\Omega)^d, {\rm div}\bar{v} = 0, {\rm trace}(\bar{v} \cdot n_{\partial\Omega}, \partial\Omega) = 0\}$ (see for example [66] or [12]).*

Numerical schemes for the Stokes equations and the Navier-Stokes equations have been extensively studied: see [42; 62; 63; 64; 44; 43] and references therein. Among different schemes, finite element schemes and finite volume schemes are frequently used for mathematical or engineering studies. An advantage of finite volume schemes is that the unknowns are approximated by piecewise constant functions: this makes it easy to take into account additional nonlinear phenomena or the coupling with algebraic or differential equations, for instance in the case of reactive flows; in particular, one can find in [62] the presentation of the classical finite volume scheme on rectangular meshes, which has been the basis of many industrial applications. However, the use of rectangular grids makes an important limitation to the type of domain which can be gridded and more recently, finite volume schemes for the Navier-Stokes equations on triangular grids have been presented: see for example [45] where the vorticity formulation is used and [11] where primal variables are used with a Chorin type projection method to ensure the divergence condition. Proofs of convergence for finite volume type schemes for the Stokes and steady-state Navier-Stokes equations are have recently been given for staggered grids [16], [45], [30], [31], [7], following the pioneering work of Nicolaides *et al.* [60], [61].

In this paper, we propose the mathematical and numerical analysis of a discretization method which uses the primitive variables, that is the velocity and the pressure, both approximated by piecewise constant functions on the cells of a 2D or 3D mesh. We emphasise that the approximate velocities and pressures are colocated, and therefore, no dual grid is needed. The only requirement on the mesh is a geometrical assumption needed for the consistency of the approximate diffusion flux (see [24] and section (4.1) for a precise definition of the admissible discretizations).

As far as we know, this work is a first proof of the convergence, of a finite volume scheme which is of large interest in industry. Indeed, industrial CFD codes (see e.g. [55], [4]) use colocated cell centred finite volume schemes; leaving aside implementation considerations, the principle of these schemes seems to differ from the present scheme only by the stabilization choice. The main reasons why this scheme is so popular in industry are:

- a colocated arrangement of the unknowns,
- a very cheap assembling step, (no numerical integration to perform)
- an easy coupling with other systems of equations.

The finite volume scheme studied here is based on three basic ingredients. First, a stabilization technique *à la* Brezzi-Pikäranta [13] is used to cope with the instability of colocated velocity/pressure approximation spaces. Second, the discretization of the pressure gradient in the momentum balance equation is performed to ensure, by construction, that it is the transpose of the divergence term of the continuity constraint. Finally, the contribution of the discrete nonlinear advection term to the kinetic energy balance vanishes for discrete divergence free velocity fields, as in the continuous case. These features appear to be essential in the proof of convergence.

We are then able to prove the stability of the scheme and the convergence of discrete solutions towards a solution of the continuous problem when the size of the mesh tends to zero, for the steady linear case (generalized Stokes problem), the stationary and the transient Navier-Stokes equations, in 2D and 3D. Our results are valid for general meshes, do not require any assumption on the regularity of the continuous solution nor, in the nonlinear case, any small data condition. We emphasise that the convergence of the fully discrete (time and space) approximation is proven here, using an original estimate on the time translates, which yields, combined with a classical estimate on the space translates, a sufficient relative compactness property.

An error analysis is performed in the steady linear case, under regularity assumptions on the solution. An error bound of order 0.5 with respect to the step size is obtained in the discrete H^1 norm and the L^2 norm for respectively the velocity and the pressure. Of course, this is probably not a sharp estimate, as can be seen from the numerical results shown in Section 4.5. Indeed, a better rate of convergence can be proved under additional assumptions on the mesh [34].

This paper is organised as follows. In section 4.1, we introduce the discretization tools together with some discrete functional analysis tools. Section 4.2 is devoted to the linear steady problem (Stokes problem), for which the finite volume scheme is given and convergence analysis and error estimates are detailed. The complete finite volume scheme for the nonlinear case is presented in section 4.3, in both the steady and transient cases. We then develop the analysis of its convergence to a weak solution of the continuous problem. We give some numerical results in section 4.5, and finally conclude with some remarks on open problems (section 4.7).

4.1 Spatial discretization and discrete functional analysis

Let $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ be an admissible discretization as defined in Definition 2.1 page 5. We shall measure the regularity of the mesh through the function $\mathrm{regul}(\mathcal{D})$ defined by

$$\begin{aligned} \mathrm{regul}(\mathcal{D}) = \inf \quad & \left\{ \tfrac{d_{K,\sigma}}{\mathrm{diam}(K)},\ K \in \mathcal{M},\ \sigma \in \mathcal{E}_K \right\} \\ & \cup \left\{ \tfrac{d_{K,\sigma_{KL}}}{d_{\sigma_{KL}}},\ K \in \mathcal{M},\ L \in \mathcal{N}_K \right\} \cup \left\{ \tfrac{1}{\mathrm{card}(\mathcal{E}_K)},\ K \in \mathcal{M} \right\}. \end{aligned} \tag{4.11}$$

We define a discrete divergence operator $\widetilde{\mathrm{div}}_{\mathcal{D}} : (H_{\mathcal{D}}(\Omega))^d \to H_{\mathcal{D}}(\Omega)$, by:

$$\widetilde{\mathrm{div}}_{\mathcal{D}}(u)(x) = \frac{1}{|K|} \sum_{L \in \mathcal{N}_K} \frac{1}{2}\, |\sigma_{KL}|\, \mathbf{n}_{KL} \cdot (u_K + u_L), \quad \text{for a.e. } x \in K, \forall K \in \mathcal{M}, \tag{4.12}$$

We then set $E_{\mathcal{D}}(\Omega) = \{u \in (H_{\mathcal{D}}(\Omega))^d, \mathrm{div}_{\mathcal{D}}(u) = 0\}$.

Remark 4.3 *In (4.12), we could replace the factor 1/2 by some coefficient $a_{KL} \geq 0$ such that $a_{KL} + a_{LK} = 1$. Such a choice, combined with the definition*

$$\widetilde{\mathrm{div}}_{\mathcal{D}}(u)(x) = \frac{1}{|K|} \sum_{L \in \mathcal{N}_K} (a_{KL}\ |\sigma_{KL}| \mathbf{n}_{KL} \cdot u_K - a_{LK}\ |\sigma_{KL}|\ \mathbf{n}_{LK} \cdot u_L),$$

produces the same results of convergence as those which are proven in this paper. On particular meshes, one can prove a better error estimate, choosing $a_{KL} = d(x_L, \sigma_{KL})/d_{KL}$ (see [34]). Nevertheless, in the general framework of this paper, other choices do not improve the convergence result and the error estimate. Therefore, we set in this paper $a_{KL} = 1/2$, which leads to the formula (4.12).
The advantage of this choice is that it leads to simpler notations and shorter equations.

The adjoint of this discrete divergence defines a discrete gradient $\widetilde{\nabla}\ : H_{\mathcal{D}}(\Omega) \to (H_{\mathcal{D}}(\Omega))^d$:

$$(\widetilde{\nabla} u)_K = \frac{1}{|K|} \sum_{L \in \mathcal{N}_K} \frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}(u_L - u_K),\ \forall K \in \mathcal{M},\ \forall u \in H_{\mathcal{D}}(\Omega). \tag{4.13}$$

Note that this gradient operator is quite different from the one defined in Chapter 3 (except in the particular cases of equilateral triangles and rectangles with edges bisecting the line segments $x_K x_K$). Indeed, the gradient defined by (3.13)–(3.14) is not well adapted to the Stokes problem because its adjoint operator does not present enough consistency with the continuous divergence operator. Conversely, the gradient defined by (4.13), which is used for the discretization of the pressure in the Navier–Stokes equations, is not well adapted to the discretization of anisotropic problems such as the one presented in Chapter 3, because it lacks convergence properties. Indeed, we shall see that all the convergence properties will be proven on the discrete divergence, using the fact that the operators are adjoint one to the other.
The operator $\widetilde{\nabla}$ then satisfies the following property.

Proposition 4.4 *Let $(\mathcal{D}^{(m)})_{m \in \mathbb{N}}$ be a sequence of admissible discretizations of Ω in the sense of Definition 2.1, such that $\lim_{m\to\infty} h_{\mathcal{D}^{m)}} = 0$. Let us assume that there exists $C > 0$ and $\alpha \in [0, 2)$ and a sequence $(u^{(m)})_{m \in \mathbb{N}}$ such that $u^{(m)} \in H_{\mathcal{D}^{(m)}}(\Omega)$ and $|u^{(m)}|^2_{\mathcal{D}_m} \leq C\ h^{-\alpha}_{\mathcal{D}^{m)}}$, for all $m \in \mathbb{N}$.*
Then the following property holds:

$$\lim_{m\to+\infty} \int_\Omega \Big(P_{\mathcal{D}_m}\varphi(x)\widetilde{\nabla}_{\mathcal{D}_m} u^{(m)}(x) + u^{(m)}(x)\nabla\varphi(x)\Big)\, dx = 0,\ \forall \varphi \in C_c^\infty(\Omega), \tag{4.14}$$

and therefore:

$$\lim_{m\to+\infty} \int_\Omega \widetilde{\nabla}_{\mathcal{D}_m} u^{(m)}(x) \cdot P_{\mathcal{D}_m}\psi(x)\delta x = 0,\ \forall \psi \in C_c^\infty(\Omega)^d \cap E(\Omega), \tag{4.15}$$

where $E(\Omega)$ is defined by (4.8).

Proof Let us assume the hypotheses of the above lemma, and let $i = 1, \ldots, d$ and $\varphi \in C_c^\infty(\Omega)$ be given. Let us study, for $m \in \mathbb{N}$, the term

$$T_7^{(m)} = \int_\Omega \Big(P_{\mathcal{D}_m}\varphi(x)\widetilde{\nabla}_{\mathcal{D}_m} u^{(m)}(x) + u^{(m)}(x)\nabla\varphi(x)\Big)\, dx.$$

From (4.13), we get that

$$T_7^{(m)} = \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}, \sigma = \sigma_{KL}} (u_L^{(m)} - u_K^{(m)})|\sigma_{KL}|\ R_{KL}^{(m)},$$

where

$$R_{KL}^{(m)} = \left(\frac{1}{2}(\varphi(x_K) + \varphi(x_L)) - \frac{1}{|\sigma_{KL}|}\int_{\sigma_{KL}} \varphi(x)\delta\gamma(x)\right)\mathbf{n}_{KL}.$$

Thanks to the Cauchy-Schwarz inequality,

$$|T_7^{(m)}|^2 \leq |u^{(m)}|^2_{\mathcal{D}_m} \sum_{\sigma \in \mathcal{E}_{\rm int}, \sigma = \sigma_{KL}} \left| R_{KL}^{(m)} \right|^2 |\sigma_{KL}| d_{KL}.$$

One has $\sum_{\sigma \in \mathcal{E}_{\rm int}, \sigma=\sigma_{KL}} |\sigma_{KL}| d_{KL} \leq d|\Omega|$. Thanks to the existence of $C_\varphi > 0$ which only depends on φ such that $|R_{KL}^{(m)}| \leq C_\varphi h_{\mathcal{D}^m)}$ and since $\alpha < 2$, we then get that

$$\lim_{m\to\infty} T_7^{(m)} = 0,$$

which yields (4.14).

Proposition 4.5 (Discrete Rellich theorem) *Let $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ be a sequence of admissible discretizations of Ω in the sense of definition 2.1, such that $\lim_{m\to\infty} h_{\mathcal{D}^m)} = 0$. Let us assume that there exists $C > 0$ and a sequence $(u^{(m)})_{m\in\mathbb{N}}$ such that $u^{(m)} \in H_{\mathcal{D}^{(m)}}(\Omega)$ and $\|u^{(m)}\|_{\mathcal{D}_m} \leq C$ for all $m \in \mathbb{N}$.*
Then, by Lemma 2.20, there exists $u \in H_0^1(\Omega)$ and a subsequence of $(u^{(m)})_{m\in\mathbb{N}}$, again denoted $(u^{(m)})_{m\in\mathbb{N}}$, such that the sequence $(u^{(m)})_{m\in\mathbb{N}}$ converges in $L^2(\Omega)$ to u as $m \to +\infty$; the sequence also satisfies: $\widetilde{\nabla}_{\mathcal{D}_m} u^{(m)}$ weakly converges to $\nabla\bar{u}$ in $L^2(\Omega)^d$ as $m \to +\infty$ and (4.14) holds.

Proof Since we have $|u^{(m)}|_{\mathcal{D}_m} \leq \|u^{(m)}\|_{\mathcal{D}_m}$, we can apply proposition 4.4, which gives the result.

4.2 Approximation of the linear steady problem

4.2.1 The Stokes problem

We first study the following linear steady problem: find an approximation of $\bar{u}$ and $\bar{p}$, weak solution to the generalized Stokes equations, which write:

$$\left\{ \begin{array}{l} \eta\bar{u} - \nu\Delta\bar{u} + \nabla\bar{p} = f \text{ in } \Omega \\ \mathrm{div}\bar{u} = 0 \text{ in } \Omega, \end{array} \right. \tag{4.16}$$

For this problem, the following assumptions are made:

$$\Omega \text{ is a polygonal open bounded connected subset of } \mathbb{R}^d,\ d = 2 \text{ or } 3 \tag{4.17}$$

$$\nu \in (0, +\infty),\ \eta \in [0, +\infty), \tag{4.18}$$

$$f \in L^2(\Omega)^d. \tag{4.19}$$

We then consider the following weak sense for problem (4.16).

Definition 4.6 (Weak solution for the steady Stokes equations) *Under hypotheses (4.17)-(4.19), let $E(\Omega)$ be defined by (4.8). Then $(\bar{u}, \bar{p})$ is called a weak solution of (4.16) (see e.g. [66] or [12]) if*

$$\left\{ \begin{array}{l} \bar{u} \in E(\Omega),\ \bar{p} \in L^2(\Omega) \text{ with } \int_\Omega \bar{p}(x)\delta x = 0, \\ \eta \displaystyle\int_\Omega \bar{u}(x) \cdot \bar{v}(x)\delta x + \nu \int_\Omega \nabla\bar{u}(x) : \nabla\bar{v}(x)\delta x - \\ \displaystyle\int_\Omega \bar{p}(x)\mathrm{div}\bar{v}(x)\delta x = \int_\Omega f(x)\cdot\bar{v}(x)\delta x,\ \forall \bar{v} \in H_0^1(\Omega)^d. \end{array} \right. \tag{4.20}$$

The existence and uniqueness of the weak solution of (4.16) in the sense of the above definition is a classical result (again, see e.g. [66] or [12]).

4.2.2 The finite volume scheme

Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1. It is then natural to write an approximate problem to the Stokes problem (4.20) in the following way.

$$\left\{\begin{aligned} & u \in E_{\mathcal{D}}(\Omega),\ p \in H_{\mathcal{D}}(\Omega) \text{ with } \int_\Omega p(x)\delta x = 0 \\ & \eta \int_\Omega u(x)\cdot v(x)\delta x + \nu[u,v]_{\mathcal{D}} \\ & \qquad - \int_\Omega p(x)\widetilde{\operatorname{div}}_{\mathcal{D}}(v)(x)\delta x = \int_\Omega f(x)\cdot v(x)\delta x \qquad \forall v \in H_{\mathcal{D}}(\Omega)^d \end{aligned}\right. \tag{4.21}$$

As we use a colocated approximation for the velocity and the pressure fields, the scheme must be stabilised. Using a non-consistent stabilization *à la* Brezzi-Pitkäranta [13], we then look for (u,p) such that

$$\left\{\begin{aligned} & (u,p) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega) \text{ with } \int_\Omega p(x)\delta x = 0 \\ & \eta \int_\Omega u(x)\cdot v(x)\delta x + \nu[u,v]_{\mathcal{D}} \\ & \qquad - \int_\Omega p(x)\widetilde{\operatorname{div}}_{\mathcal{D}}(v)(x)\delta x = \int_\Omega f(x)\cdot v(x)\delta x \qquad \forall v \in H_{\mathcal{D}}(\Omega)^d \\ & \int_\Omega \widetilde{\operatorname{div}}_{\mathcal{D}}(u)(x)q(x)\delta x = -\lambda\ h_{\mathcal{D}}^\alpha\ \langle p,q\rangle_{\mathcal{D}} \qquad \forall q \in H_{\mathcal{D}}(\Omega) \end{aligned}\right. \tag{4.22}$$

where $\lambda > 0$ and $\alpha \in (0,2)$ are adjustable parameters of the scheme which will have to be tuned in order to make a balance between accuracy and stability.

System (4.22) is equivalent to finding the family of vectors $(u_K)_{K\in\mathcal{M}} \subset \mathbb{R}^d$, and scalars $(p_K)_{K\in\mathcal{M}} \subset \mathbb{R}$ solution of the system of equations obtained by writing for each control volume K of $\mathcal{M}$:

$$\left\{\begin{aligned} & \eta\ |K|\ u_K - \nu \sum_{L\in\mathcal{N}_K} \tau_{\sigma_{KL}}(u_L - u_K) - \nu \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\text{ext}}} \frac{|\sigma|}{d_{K,\sigma}}(0-u_K) \\ & \qquad\qquad + \sum_{L\in\mathcal{N}_K} \frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\ (p_L - p_K) = \int_K f(x)\delta x \\ & \sum_{L\in\mathcal{N}_K} \frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(u_K + u_L) - \lambda\ h_{\mathcal{D}}^\alpha \sum_{L\in\mathcal{N}_K} \tau_{\sigma_{KL}}(p_L - p_K) = 0 \end{aligned}\right. \tag{4.23}$$

supplemented by the relation

$$\sum_{K\in\mathcal{M}} |K|\ p_K = 0 \tag{4.24}$$

Defining $p_\sigma = (p_K + p_L)/2$ if $\sigma = \sigma_{KL}$, and $p_\sigma = p_K$ if $\sigma \in \mathcal{E}_{\text{ext}} \cap \mathcal{E}_K$, and using the fact that $\sum_{\sigma\in\mathcal{E}_K} |\sigma|\mathbf{n}_{K,\sigma} = 0$, one notices that: $\sum_{L\in\mathcal{N}_K} \frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\ (p_L - p_K)$ is in fact equal to $\sum_{\sigma\in\mathcal{E}_K} |\sigma| p_\sigma \mathbf{n}_{K,\sigma}$, thus yielding a conservative form, which shows that (4.23) is indeed a finite volume scheme.

The existence of a solution to (4.22) will be proven below.

4.2.3 Study of the scheme in the linear case

We first prove a stability estimate for the velocity.

Proposition 4.7 (Discrete H^1 estimate on velocities) *Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of definition 2.1. Let $\lambda \in (0,+\infty)$ and $\alpha \in (0,2)$ be given. Let $(u,p) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega)$ be a solution to (4.22). Then the following inequalities hold:*

$$\nu\|u\|_{\mathcal{D}} \le \operatorname{diam}(\Omega)\|f\|_{(L^2(\Omega))^d}, \tag{4.25}$$

and

$$\nu\ \lambda\ h_{\mathcal{D}}^{\alpha}\ |p|_{\mathcal{D}}^2 \leq \mathrm{diam}(\Omega)^2 \|f\|_{(L^2(\Omega))^d}^2. \tag{4.26}$$

Proof We apply (4.22) setting $v = u$. We get

$$\eta \int_\Omega u(x)^2 \delta x + \nu \|u\|_{\mathcal{D}}^2 - \int_\Omega p(x) \widetilde{\mathrm{div}}_{\mathcal{D}}(u)(x)\delta x = \int_\Omega f(x)\cdot v(x)\delta x.$$

Since $\eta \geq 0$, the second equation of (4.22) with $q = p$ and Young's inequality yield that:

$$\left\{ \begin{array}{l} \eta \displaystyle\int_\Omega u(x)^2\delta x + \nu\|u\|_{\mathcal{D}}^2 + \lambda\ h_{\mathcal{D}}^{\alpha}\ |p|_{\mathcal{D}}^2 \leq \\ \dfrac{\mathrm{diam}(\Omega)^2}{2\nu}\|f\|_{(L^2(\Omega))^d}^2 + \dfrac{\nu}{2\mathrm{diam}(\Omega)^2}\|u\|_{(L^2(\Omega))^d}^2. \end{array} \right.$$

Using the Poincaré inequality (2.7) gives

$$\nu\|u\|_{\mathcal{D}}^2 + \lambda\ h_{\mathcal{D}}^{\alpha}\ |p|_{\mathcal{D}}^2 \leq \frac{\mathrm{diam}(\Omega)^2}{2\nu}\|f\|_{(L^2(\Omega))^d}^2 + \frac{\nu}{2}\|u\|_{\mathcal{D}}^2,$$

which leads to (4.25) and (4.26).

We can now state the existence and the uniqueness of a discrete solution to (4.22).

Corollary 4.8 [Existence and uniqueness of a solution to the finite volume scheme] *Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1. Let $\lambda \in (0, +\infty)$ and $\alpha \in (0, 2)$ be given. Then there exists a unique solution to (4.22).*

Proof System (4.22) is a linear system. Assume that $f = 0$. From propositions 4.7 and using (2.8), we get that $u = 0$ and $p = 0$. This proves that the linear system (4.22) is invertible.
We then prove the following strong estimate on the pressures.

Proposition 4.9 (L^2 estimate on pressures) *Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of definition 2.1 and let $\theta > 0$ be such that* $\mathrm{regul}(\mathcal{D}) > \theta$. *Let $\lambda \in (0, +\infty)$ and $\alpha \in (0, 2)$ be given. Let $(u, p) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega)$ be a solution to (4.22). Then there exists C_{15}, only depending on d, Ω, η, ν, λ, α and θ, and not on $h_{\mathcal{D}}$, such that the following inequality holds:*

$$\|p\|_{L^2(\Omega)} \leq C_{15}\, \|f\|_{(L^2(\Omega))^d}. \tag{4.27}$$

Proof We first apply a result by Nečas [57]: thanks to $\int_\Omega p(x)\delta x = 0$, there exists $C_{16} > 0$, which only depends on d and Ω, and $\bar{v} \in H_0^1(\Omega)^d$ such that $\mathrm{div}\bar{v}(x) = p(x)$ for a.e. $x \in \Omega$ and

$$\|\bar{v}\|_{H_0^1(\Omega)^d} \leq C_{16}\, \|p\|_{L^2(\Omega)}. \tag{4.28}$$

We then set

$$v_\sigma^{(i)} = \frac{1}{|\sigma|}\int_\sigma \bar{v}^{(i)}(x)\delta\gamma(x),\ \forall \sigma \in \mathcal{E},\ \forall i = 1, \ldots, d.$$

(note that $v_\sigma^{(i)} = 0$ for all $\sigma \in \mathcal{E}_{\mathrm{ext}}$ and $i = 1, \ldots, d$) and we define $v \in H_{\mathcal{D}}(\Omega)^d$ by

$$v_K^{(i)} = \frac{1}{|K|}\int_K \bar{v}^{(i)}(x)\delta x,\ \forall K \in \mathcal{M},\ \forall i = 1, \ldots, d.$$

Applying the results given p 777 in [24], we get that there exists $C_{17} > 0$, only depending on d and θ, such that

$$(v_K^{(i)} - v_\sigma^{(i)})^2 \leq C_{17}\, \frac{\mathrm{diam}(K)}{|\sigma|}\int_K (\nabla v^{(i)}(x))^2 \delta x, \tag{4.29}$$

and

$$\|v\|_{\mathcal{D}} \le C_{17}\,\|\bar{v}\|_{H^1_0(\Omega)^d}. \tag{4.30}$$

We then have

$$\int_\Omega p(x)\widetilde{\mathrm{div}}_{\mathcal{D}} v(x)\delta x = \sum_{K\in\mathcal{M}} p_K \sum_{L\in\mathcal{N}_K} \frac{1}{2}\,|\sigma_{KL}|\;\mathbf{n}_{KL}\cdot(v_K+v_L) = T_8 + T_9,$$

where

$$\left\{\begin{array}{rcl} T_8 &=& \displaystyle\sum_{K\in\mathcal{M}} p_K \sum_{L\in\mathcal{N}_K} 2A_{KL}\cdot v_{\sigma_{KL}} \\ &=& \displaystyle\sum_{K\in\mathcal{M}} p_K \sum_{L\in\mathcal{N}_K} \int_{\sigma_{KL}} \bar{v}(x)\cdot\mathbf{n}_{KL}\delta\gamma(x) \\ &=& \displaystyle\int_\Omega p(x)\mathrm{div}\bar{v}(x)\delta x = \|p\|^2_{L^2(\Omega)}, \end{array}\right.$$

and

$$\left\{\begin{array}{rcl} T_9 &=& \displaystyle\sum_{K\in\mathcal{M}} p_K \sum_{L\in\mathcal{N}_K} |\sigma_{KL}|\left(\frac{1}{2}(v_K+v_L)-v_{\sigma_{KL}}\right)\cdot\mathbf{n}_{KL} \\ &=& \displaystyle\sum_{\sigma=\sigma_{KL}\in\mathcal{E}_{\mathrm{int}}} |\sigma_{KL}|(p_K-p_L)\left(\frac{1}{2}(v_K+v_L)-v_{\sigma_{KL}}\right)\cdot\mathbf{n}_{KL}. \end{array}\right.$$

We then have, thanks to the Cauchy-Schwarz inequality

$$T_9^2 \le |p|^2_{\mathcal{D}} \sum_{\sigma=\sigma_{KL}\in\mathcal{E}_{\mathrm{int}}} |\sigma_{KL}| d_{KL}\left(\frac{1}{2}(v_K+v_L)-v_{\sigma_{KL}}\right)^2.$$

Applying Inequality (4.29) and thanks to $(\frac{1}{2}(v_K+v_L)-v_{\sigma_{KL}})^2 \le \frac{1}{2}((v_K-v_{\sigma_{KL}})^2+(v_L-v_{\sigma_{KL}})^2)$, we get that

$$T_9^2 \le |p|^2_{\mathcal{D}} \sum_{\sigma=\sigma_{KL}\in\mathcal{E}_{\mathrm{int}}} d_{KL} C_{17}\,h_{\mathcal{D}} \int_{K\cup L}\sum_{i=1}^d (\nabla v^{(i)}(x))^2\delta x.$$

This in turn implies the existence of $C_{18} > 0$, only depending on d and θ, such that

$$T_9^2 \le C_{18}\,h^2_{\mathcal{D}}|p|^2_{\mathcal{D}}\|\bar{v}\|^2_{H^1_0(\Omega)^d}.$$

Thanks to (4.28), we then get, gathering the previous results

$$\int_\Omega p(x)\widetilde{\mathrm{div}}_{\mathcal{D}} v(x)\delta x \ge \|p\|^2_{L^2(\Omega)} - C_{18}\,h_{\mathcal{D}}|p|_{\mathcal{D}} C_{16}\,\|p\|_{L^2(\Omega)}. \tag{4.31}$$

We then introduce v as a test function in (4.22). We get

$$\int_\Omega p(x)\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x = \eta\int_\Omega u(x)\cdot v(x)\delta x + \nu[u,v]_{\mathcal{D}} - \int_\Omega f(x)\cdot v(x)\delta x. \tag{4.32}$$

Applying the discrete Poincaré inequality, (4.30) and (4.31), we get the existence of C_{19}, only depending on d, Ω, f, η, ν, λ and θ, such that

$$\|p\|^2_{L^2(\Omega)} - C_{18}\,h_{\mathcal{D}}|p|_{\mathcal{D}} C_{16}\,\|p\|_{L^2(\Omega)} \le C_{19}\,\left(\|u\|_{\mathcal{D}} + \|f\|_{L^2(\Omega)^d}\right)\|p\|_{L^2(\Omega)}.$$

We now apply (4.25) and (4.26). Since $h^2_{\mathcal{D}} \le h^\alpha_{\mathcal{D}}\mathrm{diam}(\Omega)^{2-\alpha}$, the condition $\alpha \le 2$ suffices to produce (4.27) from the above inequality, a factor $1/\lambda$ being introduced in the expression of C_{15} (it is therefore not possible to let λ tend to 0 in (4.27)).
We then have the following result, which states the convergence of the scheme (4.22).

Proposition 4.10 (Convergence in the linear case) *Under hypotheses (4.17)-(4.19), let $(\bar{u},\bar{p})$ be the unique weak solution of the Stokes problem (4.16) in the sense of definition 4.6. Let $\lambda \in (0,+\infty)$, $\alpha \in (0,2)$ and $\theta > 0$ be given and let $\mathcal{D}$ be an admissible discretization of Ω in the sense of definition 2.1 such that* $\mathrm{regul}(\mathcal{D}) \geq \theta$. *Let $(u,p) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega)$ be the unique solution to (4.22).*
Then u converges to $\bar{u}$ in $(L^2(\Omega))^d$ and p weakly converges to $\bar{p}$ in $L^2(\Omega)$ as $h_{\mathcal{D}}$ tends to 0.

Proof Under the hypotheses of the above proposition, let $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ be a sequence of admissible discretizations of Ω in the sense of definition 2.1, such that $\lim_{m\to\infty} h_{\mathcal{D}^{m})} = 0$ and such that $\mathrm{regul}(\mathcal{D}^{(m)}) \geq \theta$, for all $m \in \mathbb{N}$.
Let $(u^{(m)},p^{(m)}) \in H_{\mathcal{D}^{(m)}}(\Omega)^d \times H_{\mathcal{D}^{(m)}}(\Omega)$ be given by (4.22) for all $m \in \mathbb{N}$. Let us prove the existence of a subsequence of $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ such that the corresponding sequence $(u^{(m)})_{m\in\mathbb{N}}$ converges in $(L^2(\Omega))^2$ to $\bar{u}$ and the sequence $(p^{(m)})_{m\in\mathbb{N}}$ weakly converges in $(L^2(\Omega))^2$ to $\bar{p}$, as $m\to\infty$. Then the proof is complete thanks to the uniqueness of $(\bar{u},\bar{p})$.
Using (4.25), we obtain (see [25], [24]) an estimate on the translates of $u^{(m)}$: for all $m \in \mathbb{N}$, there exists $C_{20} > 0$, only depending on Ω, ν, f and g such that

$$\left\{ \begin{array}{l} \int_\Omega (u^{(m,k)}(x+\xi) - u^{(m,k)}(x))^2 \delta x \leq C_{20}\, |\xi|(|\xi| + 4h_{\mathcal{D}^{m})}), \\ \text{for } k = 1,\ldots,d,\ \forall \xi \in \mathbb{R}^d, \end{array} \right. \tag{4.33}$$

where $u^{(m,k)}$ denotes the k-th component of $u^{(m)}$. We may then apply Kolmogorov's theorem, and obtain the existence of a subsequence of $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ and of $\bar{u} \in H_0^1(\Omega)^2$ such that $(u^{(m)})_{m\in\mathbb{N}}$ converges to $\bar{u}$ in $L^2(\Omega)^2$. Thanks to proposition 4.9, we extract from this subsequence another one (still denoted $u^{(m)}$) such that $(p^{(m)})_{m\in\mathbb{N}}$ weakly converges to some function $\bar{p}$ in $L^2(\Omega)$. In order to conclude the proof of the convergence of the scheme, there only remains to prove that $(\bar{u},\bar{p})$ is the solution of (4.20), thanks to the uniqueness of this solution.
Let $\varphi \in (C_c^\infty(\Omega))^d$. Let $m \in \mathbb{N}$ such that $\mathcal{D}^{(m)}$ belongs to the above extracted subsequence and let $(u^{(m)},p^{(m)})$ be the solution to (4.22) with $\mathcal{D} = \mathcal{D}^{(m)}$. We suppose that m is large enough and thus $h_{\mathcal{D}^{m})}$ is small enough to ensure for all $K \in \mathcal{M}$ such that $K\cap$ support(φ) $\neq \emptyset$, then $\partial K \cap \partial\Omega = \emptyset$ holds. Let us take $v = P_{\mathcal{D}^{(m)}}\varphi$ in (4.22). Applying proposition 4.5, we get

$$\lim_{n\to\infty} [u^{(m)}, P_{\mathcal{D}^{(m)}}\varphi]_{\mathcal{D}^{(m)}} = \int_\Omega \nabla\bar{u}(x) : \nabla\varphi(x)\delta x.$$

Moreover, it is clear that

$$\lim_{n\to\infty} \int_\Omega f(x)\cdot P_{\mathcal{D}^{(m)}}\varphi(x)\delta x = \int_\Omega f(x)\cdot\varphi(x)\delta x,$$

and

$$\lim_{n\to\infty} \eta \int_\Omega u^{(m)}(x)\cdot P_{\mathcal{D}^{(m)}}\varphi(x)\delta x = \eta\int_\Omega \bar{u}(x)\cdot\varphi(x)\delta x.$$

Thanks to the weak convergence of the sequence of approximate pressures, to (4.26) and to the hypothesis $\alpha < 2$, we now apply proposition 4.4, which gives

$$\lim_{n\to\infty} \int_\Omega p^{(m)}(x)\mathrm{div}_{\mathcal{D}^{(m)}}(P_{\mathcal{D}^{(m)}}\varphi)(x)\delta x = \int_\Omega \bar{p}(x)\mathrm{div}\varphi(x)\delta x. \tag{4.34}$$

The last step is to prove that $\mathrm{div}(\bar{u}) = 0$ a.e. in Ω. Let $\varphi \in C_c^\infty(\Omega)$ and let $m \in \mathbb{N}$ be given. Let us take $q = P_{\mathcal{D}^{(m)}}\varphi$ in (4.22). We get $T_{10}^{(m)} = -T_{11}^{(m)}$, where

$$T_{10}^{(m)} = \int_\Omega \mathrm{div}_{\mathcal{D}^{(m)}}(u^{(m)})(x) P_{\mathcal{D}^{(m)}}\varphi(x)\delta x.$$

and

$$T_{11}^{(m)} = \lambda\ h_{\mathcal{D}^{m})}^\alpha \langle p^{(m)}, P_{\mathcal{D}^{(m)}}\varphi\rangle_{\mathcal{D}}.$$

On the one hand, the third item of proposition 4.5 produces

$$\lim_{n\to\infty} T_{10}^{(m)} = \sum_{i=1}^{d} \int_{\Omega} \varphi(x)\partial_i \bar{u}^{(i)} \delta x.$$

On the other hand, using the Cauchy-Schwarz inequality, we get:

$$T_{11}^{(m)} \le \lambda h^{\alpha}_{\mathcal{D}^{m})} |p^{(m)}|_{\mathcal{D}} |P_{\mathcal{D}^{(m)}}\varphi|_{\mathcal{D}}$$

Therefore, thanks to (4.26) and to the regularity of φ (that implies that $|P_{\mathcal{D}^{(m)}}\varphi|_{\mathcal{D}}$ remains bounded independently on $h_{\mathcal{D}^{m})}$) we obtain $\lim_{n\to\infty} T_{11}^{(m)} = 0$. This in turn implies that:

$$\sum_{i=1}^{d} \int_{\Omega} \varphi(x)\partial_i \bar{u}^{(i)}(x) \delta x = 0, \text{ for all } \varphi \in C_c^{\infty}(\Omega), \tag{4.35}$$

which proves that $\bar{u} \in E(\Omega)$.

Remark 4.11 (Strong convergence of the pressure) *Note that the proof of the strong convergence of p to $\bar{p}$ is a straightforward consequence of the error estimate stated in Proposition 4.12 below, which holds under additional regularity hypotheses.*

4.2.4 An error estimate

We then have the following result, which states an error estimate for the scheme (4.22).

Proposition 4.12 (Error estimate in the linear case) *Under hypotheses (4.17)-(4.19), we assume that the weak solution $(\bar{u}, \bar{p})$ of the Stokes problem (4.16) in the sense of definition (4.6) is such that $(\bar{u}, \bar{p}) \in H^2(\Omega)^d \times H^1(\Omega)$. Let $\lambda \in (0, +\infty)$ and $\alpha \in (0, 2)$ be given, let $\mathcal{D}$ be an admissible discretization of Ω in the sense of definition 2.1 and let $\theta > 0$ such that* $\text{regul}(\mathcal{D}^{(m)}) \ge \theta$*. Let $(u, p) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega)$ be the solution to (4.22). Then there exists C_{21} , which only depends on d, Ω, ν, η and θ such that*

$$\|u - \bar{u}\|^2_{L^2(\Omega)} \le C_{21}\, \varepsilon(\lambda, h_{\mathcal{D}}, \bar{p}, \bar{u}), \tag{4.36}$$

$$\lambda\ h^{\alpha}_{\mathcal{D}}\ |p|^2_{\mathcal{D}} \le C_{21}\, \varepsilon(\lambda, h_{\mathcal{D}}, \bar{p}, \bar{u}) \tag{4.37}$$

$$\|p - \bar{p}\|^2_{L^2(\Omega)} \le C_{21}\, \varepsilon(\lambda, h_{\mathcal{D}}, \bar{p}, \bar{u}). \tag{4.38}$$

where

$$\left\{ \begin{array}{rl} \varepsilon(\lambda, h_{\mathcal{D}}, \bar{p}, \bar{u}) = & \max\left(\lambda h^{\alpha}_{\mathcal{D}}, \frac{1}{\lambda} h^{2-\alpha}_{\mathcal{D}}\right) \\ & \times \left(\|\bar{p}\|^2_{H^1(\Omega)} + \|\bar{u}\|^2_{H^2(\Omega)}\right). \end{array} \right. \tag{4.39}$$

Proof We define $(\hat{u}, \hat{p}) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega)$ by $\hat{u} = P_{\mathcal{D}}\bar{u}$, which means $\hat{u}_K = \bar{u}(x_k)$ for all $K \in \mathcal{M}$, and $\hat{p}_K = \frac{1}{|K|} \int_K \bar{p}(x)\delta x$ for all $K \in \mathcal{M}$. Integrating the first equation of (4.16) on $K \in \mathcal{M}$ gives

$$\eta \int_K \bar{u}(x)\delta x + \sum_{\sigma\in\mathcal{E}_K} \left(\left\{ \begin{array}{l} -\nu \int_\sigma \nabla\bar{u}(x) : \mathbf{n}_{K,\sigma}\delta\gamma(x) + \\ \int_\sigma \bar{p}(x)\mathbf{n}_{K,\sigma}\delta\gamma(x) \end{array} \right. \right) = \int_K f(x)\delta x. \tag{4.40}$$

We introduce, for $K \in \mathcal{M}$, $\varepsilon^u_K = \hat{u}_K - \frac{1}{|K|}\int_K \bar{u}(x)\delta x$, and, for $L \in \mathcal{N}_K$:
$R_{K,L} = \frac{1}{d_{\sigma_{KL}}}(\hat{u}_L - \hat{u}_K) - \frac{1}{|\sigma_{KL}|}\int_\sigma \nabla\bar{u}(x) : \mathbf{n}_{K,\sigma}\delta\gamma(x)$,
and for $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$, $R_{K,\sigma} = \frac{1}{d_{K,\sigma}}(0 - \hat{u}_K) - \frac{1}{|\sigma|}\int_\sigma \nabla\bar{u}(x) : \mathbf{n}_{K,\sigma}\delta\gamma(x)$;
moreover, we define for $L \in \mathcal{N}_K$: $\varepsilon^p_{\sigma_{KL}} = \frac{1}{2}(\hat{p}_K + \hat{p}_L) - \frac{1}{|\sigma_{KL}|}\int_{\sigma_{KL}} \bar{p}(x)\delta\gamma(x)$, and for $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$,

$\varepsilon^p_\sigma = \hat{p}_K - \frac{1}{|\sigma|}\int_\sigma \bar{p}(x)\delta\gamma(x)$. Using these notations and the relation $\sum_{\sigma\in\mathcal{E}_K}|\sigma|\mathbf{n}_{K,\sigma} = 0$, we get from (4.40)

$$\left\{\begin{array}{l} \eta\ |K|\hat{u}_K - \nu\left(\displaystyle\sum_{L\in\mathcal{N}_K}\tau_{\sigma_{KL}}(\hat{u}_L - \hat{u}_K) + \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}\frac{|\sigma|}{d_{K,\sigma}}(0-\hat{u}_K)\right) + \\ \displaystyle\sum_{L\in\mathcal{N}_K}\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\ (\hat{p}_L - \hat{p}_K) = \int_K f(x)\delta x + R_K, \end{array}\right.$$

with

$$\left\{\ R_K = \eta\ |K|\varepsilon^u_K - \nu\left(\sum_{L\in\mathcal{N}_K}|\sigma_{KL}|R_{K,L} + \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}|\sigma|R_{K,\sigma}\right) + \sum_{\sigma\in\mathcal{E}_K}|\sigma|\ \varepsilon^p_\sigma\mathbf{n}_{K,\sigma}.\right.$$

We then set $\delta u = \hat{u} - u$ and $\delta p = \hat{p} - p$. We then get, subtracting the first relation of the scheme (4.23) to the above equation,

$$\left\{\begin{array}{l} \eta\int_\Omega \delta u(x)v(x)\delta x + \nu[\delta u, v]_{\mathcal{D}} - \displaystyle\int_\Omega \delta p(x)\widetilde{\rm div}_{\mathcal{D}}(v)(x)\delta x = \\ \int_\Omega R(x)v\delta x,\ \forall v\in H_{\mathcal{D}}(\Omega)^d, \end{array}\right. \tag{4.41}$$

and, setting $v = \delta u$ in (4.41),

$$\eta\int_\Omega \delta u(x)^2\delta x + \nu\|\delta u\|^2_{\mathcal{D}} - \int_\Omega \delta p(x)\widetilde{\rm div}_{\mathcal{D}}(\delta u)(x)\delta x = \int_\Omega R(x)\delta u(x)\delta x.$$

We now integrate the second equation of (4.16) on $K\in\mathcal{M}$. This gives

$$\sum_{\sigma\in\mathcal{E}_K}\int_\sigma \bar{u}(x)\cdot\mathbf{n}_{K,\sigma}\delta\gamma(x) = 0,\ \forall K\in\mathcal{M}.$$

Using $\bar{u}\in H^1_0(\Omega)$, we then obtain

$$\sum_{L\in\mathcal{N}_K}\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(\hat{u}_K + \hat{u}_L) = \sum_{L\in\mathcal{N}_K}|\sigma_{KL}|\varepsilon^u_{\sigma_{KL}},\ \forall K\in\mathcal{M}$$

with

$$\varepsilon^u_{\sigma_{KL}} = \left(\frac{1}{2}(\hat{u}_K + \hat{u}_L) - \frac{1}{|\sigma_{KL}|}\int_{\sigma_{KL}}\bar{u}(x)\delta\gamma(x)\right)\cdot\mathbf{n}_{KL},\ \forall K\in\mathcal{M},\ \forall L\in\mathcal{N}_K.$$

We then give, subtracting the second relation of the scheme (4.23) to the above equation,

$$\int_\Omega \widetilde{\rm div}_{\mathcal{D}}(\delta u)(x)\delta p(x)\delta x = \lambda\ h^\alpha_{\mathcal{D}}\ \langle p, \hat{p} - p\rangle_{\mathcal{D}} + T_{12},$$

with

$$T_{12} = \sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}}|\sigma_{KL}|\varepsilon^u_{\sigma_{KL}}(\delta p_K - \delta p_L),$$

Gathering the above results, we get

$$\left\{\begin{array}{l} \eta\int_\Omega \delta u(x)^2\delta x + \nu\|\delta u\|^2_{\mathcal{D}} + \lambda\ h^\alpha_{\mathcal{D}}\ |p|^2_{\mathcal{D}} = \\ \lambda\ h^\alpha_{\mathcal{D}}\ \langle p, \hat{p}\rangle_{\mathcal{D}} + \int_\Omega R(x)\cdot\delta u(x)\delta x + T_{12}. \end{array}\right. \tag{4.42}$$

Let us study the terms at the right hand side of the above equation. We have, using the Young inequality,

$$\langle p, \hat{p}\rangle_{\mathcal{D}} \le \frac{1}{4}|p|^2_{\mathcal{D}} + |\hat{p}|^2_{\mathcal{D}} \le \frac{1}{4}|p|^2_{\mathcal{D}} + C_{22}\ \|\bar{p}\|^2_{H^1(\Omega)}. \tag{4.43}$$

We then study $\int_\Omega R(x)\cdot \delta u(x)\delta x = T_{13}+T_{14}+T_{15}$, with

$$T_{13}=\eta \int_\Omega \varepsilon^u(x)\cdot \delta u(x)\delta x,$$

$$T_{14}=\nu \sum_{K\in\mathcal{M}}\left(\sum_{L\in\mathcal{N}_K}|\sigma_{KL}|R_{K,L}+\sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}|\sigma|R_{K,\sigma}\right)\cdot \delta u_K,$$

and

$$T_{15}=\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}|\sigma|\ \varepsilon^p_\sigma \mathbf{n}_{K,\sigma}\cdot \delta u_K.$$

Thanks to interpolation results proven in [24] and to (2.7), we obtain

$$T_{13}\le C_{23}\, h_\mathcal{D}^2\|\bar u\|^2_{H^2(\Omega)}+\frac{\nu}{4}\|\delta u\|^2_\mathcal{D}, \tag{4.44}$$

$$T_{14}\le C_{24}\, h_\mathcal{D}^2\|\bar u\|^2_{H^2(\Omega)}+\frac{\nu}{4}\|\delta u\|^2_\mathcal{D}, \tag{4.45}$$

and

$$T_{15}\le C_{25}\, h_\mathcal{D}^2\|\bar p\|^2_{H^1(\Omega)}+\frac{\nu}{4}\|\delta u\|^2_\mathcal{D}. \tag{4.46}$$

We then study T_{12}. We have $T_{12}=T_{16}-T_{17}$ with

$$T_{16}=\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}}|\sigma_{KL}|\varepsilon^u_{\sigma_{KL}}(\hat p_K-\hat p_L),$$

which verifies

$$T_{16}\le C_{26}\, h_\mathcal{D}\left(\|\bar p\|^2_{H^1(\Omega)}+\|\bar u\|^2_{H^2(\Omega)}\right), \tag{4.47}$$

and

$$T_{17}=\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}}|\sigma_{KL}|\varepsilon^u_{\sigma_{KL}}(p_K-p_L),$$

which verifies

$$T_{17}\le \frac{1}{4}\lambda\ h_\mathcal{D}^\alpha\ |p|^2_\mathcal{D}+C_{27}\,\frac{1}{\lambda}h_\mathcal{D}^{2-\alpha}\|\bar u\|^2_{H^2(\Omega)}. \tag{4.48}$$

Gathering equations (4.42)-(4.48) gives

$$\|\delta u\|^2_\mathcal{D}+\lambda\ h_\mathcal{D}^\alpha\ |p|^2_\mathcal{D}\le C_{28}\,\varepsilon(\lambda,h_\mathcal{D},\bar p,\bar u),$$

where $\varepsilon(\lambda,h_\mathcal{D},\bar p,\bar u)$ is defined by (4.39). This in turn yields (4.36) and (4.37). We then again follow the method used in the proof of Proposition 4.9. Using $\int_\Omega \hat p(x)\delta x=0$ and therefore $\int_\Omega \delta p(x)\delta x=0$, let $\bar v\in H^1_0(\Omega)^d$ be given such that $\mathrm{div}\bar v(x)=\delta p(x)$ for a.e. $x\in\Omega$ and

$$\|\bar v\|_{H^1_0(\Omega)^d}\le C_{16}\,\|\delta p\|_{L^2(\Omega)}. \tag{4.49}$$

We again set

$$v_\sigma^{(i)}=\frac{1}{|\sigma|}\int_\sigma \bar v^{(i)}(x)\delta\gamma(x),\ \forall \sigma\in\mathcal{E},\ \forall i=1,\ldots,d.$$

and we define $v\in H_\mathcal{D}(\Omega)^d$ by

$$v_K^{(i)}=\frac{1}{|K|}\int_K \bar v^{(i)}(x)\delta x,\ \forall K\in\mathcal{M},\ \forall i=1,\ldots,d.$$

The same method gives

$$\left\{\begin{array}{ll} \|\delta p\|^2_{L^2(\Omega)} & \leq \displaystyle\int_\Omega \delta p(x)\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x + C_{18}\, h_{\mathcal{D}}|p|_{\mathcal{D}}\|\bar{v}\|_{H^1_0(\Omega)^d} \\ & \leq \displaystyle\int_\Omega \delta p(x)\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x + C_{29}\, h^2_{\mathcal{D}}|p|^2_{\mathcal{D}} + \frac{1}{4}\|\delta p\|^2_{L^2(\Omega)}. \end{array}\right.$$

We now use v as test function in (4.41). We get

$$\int_\Omega \delta p(x)\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x = \eta \int_\Omega \delta u(x)v(x)\delta x + \nu[\delta u, v]_{\mathcal{D}} + \int_\Omega R(x)v\delta x.$$

Gathering the two above inequalities, (4.44), (4.45), (4.46) and (4.49) produces

$$\left\{\begin{array}{ll} \|\delta p\|^2_{L^2(\Omega)} \leq & \dfrac{1}{2}\|\delta p\|^2_{L^2(\Omega)} + C_{30}\, h^2_{\mathcal{D}}\left(\|\bar{p}\|^2_{H^1(\Omega)} + \|\bar{u}\|^2_{H^2(\Omega)}\right) \\ & +C_{31}\, \|\delta u\|^2_{\mathcal{D}} + C_{29}\, h^2_{\mathcal{D}}|p|^2_{\mathcal{D}}. \end{array}\right.$$

Applying (4.36) and (4.37) gives (4.38).

Remark 4.13 *In the above result, it suffices to let $\alpha = 1$ to obtain the proof of an order $1/2$ for the convergence of the scheme. We recall that this result is not sharp, and that the numerical results show a much better order of convergence.*

4.3 The finite volume scheme for the Navier-Stokes equations

Before handling the transient nonlinear case, we first address in the following section the steady-state case.

4.3.1 The steady-state case

For the following continuous equations,

$$\left\{\begin{array}{l} \eta\bar{u}^{(i)} - \nu\Delta\bar{u}^{(i)} + \partial_i\bar{p} + \displaystyle\sum_{j=1}^d \bar{u}^{(j)}\partial_j\bar{u}^{(i)} = f^{(i)} \text{ in } \Omega, \quad \text{for } i = 1, \ldots, d, \\ \mathrm{div}\bar{u} = \displaystyle\sum_{i=1}^d \partial_i\bar{u}^{(i)} = 0 \text{ in } \Omega. \end{array}\right. \tag{4.50}$$

with a homogeneous Dirichlet boundary condition, we define the following weak sense.

Definition 4.14 (Weak solution for the steady Navier-Stokes equations) *Under hypotheses (4.17)-(4.19), let $E(\Omega)$ be defined by (4.8). Then $(\bar{u}, \bar{p})$ is called a weak solution of (4.50) if*

$$\left\{\begin{array}{l} \bar{u} \in E(\Omega),\ \bar{p} \in L^2(\Omega) \ \textit{with} \displaystyle\int_\Omega \bar{p}(x)\delta x = 0, \\ \eta \displaystyle\int_\Omega \bar{u}(x)\cdot\bar{v}(x)\delta x + \nu\int_\Omega \nabla\bar{u}(x) : \nabla\bar{v}(x)\delta x \\ \quad - \displaystyle\int_\Omega \bar{p}(x)\mathrm{div}\bar{v}(x)\delta x + b(\bar{u},\bar{u},\bar{v}) = \int_\Omega f(x)\cdot\bar{v}(x)\delta x \qquad \forall\bar{v} \in H^1_0(\Omega)^d, \end{array}\right. \tag{4.51}$$

where the trilinear form $b(.,.,.)$ is defined by (4.10).

We now give the finite volume scheme for this problem. Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1. We introduce Bernoulli's pressure $p+\frac{1}{2}u^2$ instead of p, again denoted by p, and for any real value $\lambda>0$ and $\alpha\in(0,2)$, we look for (u,p) such that

$$\left\{\begin{array}{ll}(u,p)\in H_{\mathcal{D}}(\Omega)^d\times H_{\mathcal{D}}(\Omega) \text{ with } \displaystyle\int_\Omega p(x)\delta x=0, & \\[2ex] \eta\displaystyle\int_\Omega u(x)\cdot v(x)\delta x+\nu[u,v]_{\mathcal{D}}+\frac{1}{2}\int_\Omega u(x)^2\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x & \\ \quad -\displaystyle\int_\Omega p(x)\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x+b_{\mathcal{D}}(u,u,v)=\int_\Omega f(x)\cdot v(x)\delta x & \forall v\in H_{\mathcal{D}}(\Omega)^d\\[2ex] \displaystyle\int_\Omega \widetilde{\mathrm{div}}_{\mathcal{D}}(u)(x)q(x)\delta x=-\lambda\ h_{\mathcal{D}}^\alpha\ \langle p,q\rangle_{\mathcal{D}} & \forall q\in H_{\mathcal{D}}(\Omega)\end{array}\right. \tag{4.52}$$

where, for $u,v,w\in H_{\mathcal{D}}(\Omega)$, we define the following approximation for $b(u,v,w)$

$$b_{\mathcal{D}}(u,v,w)=\frac{1}{2}\sum_{K\in\mathcal{M}}\sum_{L\in\mathcal{N}_K}(\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(u_K+u_L))\ ((v_L-v_K)\cdot w_K) \tag{4.53}$$

System (4.52) is equivalent to finding the family of vectors $(u_K)_{K\in\mathcal{M}}\subset\mathbb{R}^d$, and scalars $(p_K)_{K\in\mathcal{M}}\subset\mathbb{R}$ solution of the system of equations obtained by writing for each control volume K of $\mathcal{M}$:

$$\left\{\begin{array}{l}\eta\ |K|\ u_K-\nu\displaystyle\sum_{L\in\mathcal{N}_K}\tau_{\sigma_{KL}}(u_L-u_K)-\nu\sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}\frac{|\sigma|}{d_{K,\sigma}}(0-u_K)\\ +\displaystyle\sum_{L\in\mathcal{N}_K}(\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(\frac{1}{2}(u_K+u_L)))\ (u_L-u_K)\\ +\displaystyle\sum_{L\in\mathcal{N}_K}\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\ (p_L-p_K)-\frac{1}{2}\sum_{L\in\mathcal{N}_K}\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\ (u_L^2-u_K^2)=\int_K f(x)\delta x\\ \displaystyle\sum_{L\in\mathcal{N}_K}\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(u_K+u_L)-\lambda\ h_{\mathcal{D}}^\alpha\sum_{L\in\mathcal{N}_K}\tau_{\sigma_{KL}}(p_L-p_K)=0\end{array}\right. \tag{4.54}$$

supplemented by the relation:

$$\sum_{K\in\mathcal{M}}|K|\ p_K=0$$

Defining $\tilde p_K=p_K-u_K^2/2$ and $\tilde p_\sigma=(\tilde p_K+\tilde p_L)/2$ if $\sigma=\sigma_{KL}$, $\tilde p_\sigma=\tilde p_K$ if $\sigma\in\mathcal{E}_{\rm ext}\cap\mathcal{E}_K$, and using the fact that $\sum_{\sigma\in\mathcal{E}_K}|\sigma|\mathbf{n}_{K,\sigma}=0$, one again notices that: $\sum_{L\in\mathcal{N}_K}\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\ (\tilde p_L-\tilde p_K)$ is in fact equal to $\sum_{\sigma\in\mathcal{E}_K}|\sigma|\tilde p_\sigma\mathbf{n}_{K,\sigma}$, thus yielding a conservative form for the fifth and sixth terms of the left hand side of the discrete momentum equation in (4.54). Defining $u_\sigma=(u_K+u_L)/2$ if $\sigma=\sigma_{KL}$, $u_\sigma=0$ if $\sigma\in\mathcal{E}_{\rm ext}\cap\mathcal{E}_K$, one obtains that the nonlinear convective term $\sum_{L\in\mathcal{N}_K}(\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(\frac{1}{2}(u_K+u_L)))\ (u_L-u_K)$ is equal to $\sum_{\sigma\in\mathcal{E}_K}|\sigma|(\mathbf{n}_{K,\sigma}\cdot u_\sigma)u_\sigma-|K|u_K(\widetilde{\mathrm{div}}_{\mathcal{D}}u)_K$; one may note that $(\widetilde{\mathrm{div}}_{\mathcal{D}}u)_K=\sum_{\sigma\in\mathcal{E}_K}|\sigma|\mathbf{n}_{K,\sigma}\cdot u_\sigma$. Hence the nonlinear convective term is the sum of a conservative form and a source term due to the stabilization (this source term vanishes for a discrete divergence free function u).

Let us then study some properties of the trilinear form $b_{\mathcal{D}}$. First note that the quantity $b_{\mathcal{D}}(u,v,w)$ also writes

$$b_{\mathcal{D}}(u,v,w)=\frac{1}{2}\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}}(\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL}\cdot(u_K+u_L))\ ((v_L-v_K)\cdot(w_L+w_K)) \tag{4.55}$$

We thus get that, for all $u, v \in H_{\mathcal{D}}(\Omega)^d$,

$$\begin{cases} b_{\mathcal{D}}(u,v,v) & = \dfrac{1}{2} \displaystyle\sum_{\sigma_{KL} \in \mathcal{E}_{\text{int}}} (\frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL} \cdot (u_K + u_L))((v_L)^2 - (v_K)^2) \\ & = -\dfrac{1}{2} \displaystyle\int_\Omega v(x)^2\ \widetilde{\operatorname{div}}_{\mathcal{D}}(u)(x)\ \delta x \end{cases} \tag{4.56}$$

We get in particular, that, for all $u \in E_{\mathcal{D}}(\Omega)$, $b_{\mathcal{D}}(u,u,u) = 0$, which is the discrete equivalent of the continuous property.

Remark 4.15 [Upstream weighting versions of the scheme] *All the results of this paper are available, setting $F_{KL}(u) = \frac{1}{2}\ |\sigma_{KL}|\ \mathbf{n}_{KL} \cdot (u_K + u_L)$ and considering, for $u, v, w \in H_{\mathcal{D}}(\Omega)$,*

$$b_{\mathcal{D}}^{\text{ups}}(u,v,w) = b_{\mathcal{D}}(u,v,w) + \frac{1}{2} \sum_{\sigma_{KL} \in \mathcal{E}_{\text{int}}} \Theta_{KL} |F_{KL}(u)|\ (v_L - v_K) \cdot (w_L - w_K),$$

with, for example, $\Theta_{KL} = \max(1 - 2\nu\tau_{\sigma_{KL}}/|F_{KL}(u)|, 0)$. We then get, for all $u, v \in H_{\mathcal{D}}(\Omega)$, the inequality

$$b_{\mathcal{D}}^{\text{ups}}(u,v,v) \geq -\frac{1}{2} \int_\Omega v(x)^2 \widetilde{\operatorname{div}}_{\mathcal{D}}(u)(x) \delta x,$$

which is sufficient to get all the estimates of this paper, together with the convergence properties of the scheme. The use of such a local upwinding technique may be useful to avoid the development of nonphysical oscillations only where meshes are too coarse.

The following technical estimates are crucial to prove the convergence properties of the scheme.

Lemma 4.16 (Estimates on $b_{\mathcal{D}}(.,.,.)$) *Under hypotheses (4.3)-(4.7), let $\mathcal{D}$ be an admissible discretization of $\Omega \times (0,T)$ in the sense of definition 4.22, and $\theta > 0$ such that $\operatorname{regul}(\mathcal{D}) \geq \theta$. Then there exists $C_{32} > 0$ and $C_{33} > 0$, only depending on d, θ and Ω, such that*

$$b_{\mathcal{D}}(u,v,w) \leq C_{32}\ \|u\|_{L^4(\Omega)^d}\ \|v\|_{\mathcal{D}}\ \|w\|_{L^4(\Omega)^d} \leq C_{33}\ \|u\|_{\mathcal{D}}\ \|v\|_{\mathcal{D}}\ \|w\|_{\mathcal{D}}. \tag{4.57}$$

The proof of this estimate may be found in [33]; it uses the following discrete Sobolev inequality, which holds under the same regularity assumptions on the mesh (see proof in [19] or [24, pp. 790-791]):

$$\|u\|_{L^4(\Omega)} \leq C_{34}\ \|u\|_{\mathcal{D}}. \tag{4.58}$$

Remark 4.17 (Two dimensional case) *In the case $d = 2$, it may be proven setting $\alpha = 2, p = p' = 2$ in the proof p 791 of [24], that*

$$\|u\|_{L^4(\Omega)} \leq C_{35}\, \|u\|_{L^2(\Omega)}^{1/2} \|u\|_{\mathcal{D}}^{1/2}$$

and therefore, that there exists $C_{36} > 0$, only depending on d and Ω, such that

$$b_{\mathcal{D}}(u,v,w) \leq C_{36}\, \|v\|_{\mathcal{D}} \left(\|u\|_{\mathcal{D}}\ \|u\|_{L^2(\Omega)}\ \|w\|_{\mathcal{D}}\ \|w\|_{L^2(\Omega)}\right)^{1/2}.$$

This is a discrete analogue to the classical continuous estimate on the trilinear form.

The existence of a solution to the scheme (4.52) is obtained through a so-called "topological degree" argument. For the sake of completeness, we recall this argument (which was first used for numerical schemes in [23]) in the finite dimensional case in the following theorem and refer to [20] for the general case.

Theorem 4.18 (Application of the topological degree, finite dimensional case) *Let V be a finite dimensional vector space on $\mathbb{R}$ and g be a continuous function from V to V. Let us assume that there exists a continuous function F from $V \times [0,1]$ to V satisfying:*

1. *$F(\cdot, 1) = g$, $F(\cdot, 0)$ is an affine function.*
2. *There exists $R > 0$, such that for any $(v, \rho) \in V \times [0,1]$, if $F(v, \rho) = 0$, then $\|v\|_V \neq R$.*
3. *The equation $F(v, 0) = 0$ has a solution $v \in V$ such that $\|v\|_V < R$.*

Then there exists at least a solution $v \in V$ such that $g(v) = 0$ and $\|v\|_V < R$.

Here $g(v) = 0$ represents the nonlinear system (4.52), and we are now going to construct the function F and show the required estimates. Note that here, the use of Bernoulli's pressure leads to simpler calculations.

Proposition 4.19 (Discrete $H_0^1(\Omega)$ estimate on the velocities) *Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of $\Omega \times (0,T)$ in the sense of definition 4.22. Let $\lambda \in (0, +\infty)$ and $\alpha \in (0,2)$ be given. Let $\rho \in [0,1]$ be given and let $(u,p) \in (H_{\mathcal{D}}(\Omega))^d \times H_{\mathcal{D}}(\Omega)$, be a solution to the following system of equations (which reduces to (4.52) as $\rho = 1$ and to (4.22) as $\rho = 0$)*

$$\left\{\begin{array}{ll} (u,p) \in H_{\mathcal{D}}(\Omega)^d \times H_{\mathcal{D}}(\Omega) \text{ with } \displaystyle\int_\Omega p(x)\delta x = 0, & \\ \eta \displaystyle\int_\Omega u(x)\cdot v(x)\delta x + \nu[u,v]_{\mathcal{D}} + \frac{\rho}{2}\int_\Omega u(x)^2 \widetilde{\operatorname{div}}_{\mathcal{D}}(v)(x)\delta x & \\ +\rho\, b_{\mathcal{D}}(u,u,v) - \displaystyle\int_\Omega p(x)\widetilde{\operatorname{div}}_{\mathcal{D}}(v)(x)\delta x = \int_\Omega f(x)\cdot v(x)\delta x & \forall v \in H_{\mathcal{D}}(\Omega)^d \\ \displaystyle\int_\Omega \widetilde{\operatorname{div}}_{\mathcal{D}}(u)(x) q(x)\delta x = -\lambda\, h_{\mathcal{D}}^\alpha\, \langle p, q\rangle_{\mathcal{D}} & \forall q \in H_{\mathcal{D}}(\Omega) \end{array}\right. \tag{4.59}$$

Then u and p satisfy the following estimates, which are the same inequalities as obtained in the linear case (inequalities (4.25) and (4.26)):

$$\left\{\begin{array}{l} \nu\|u\|_{\mathcal{D}} \leq \operatorname{diam}(\Omega)\|f\|_{(L^2(\Omega))^d} \\ \nu\, \lambda\, h_{\mathcal{D}}^\alpha\, |p|_{\mathcal{D}}^2 \leq \operatorname{diam}(\Omega)^2 \|f\|_{(L^2(\Omega))^d}^2 \end{array}\right.$$

Proof The proof is similar to that of Proposition 4.7, using the property (4.56) on the discrete trilinear form.

We are now in position to prove the existence of at least one solution to scheme (4.52).

Proposition 4.20 (Existence of a discrete solution) *Under hypotheses (4.17)-(4.19), let $\mathcal{D}$ be an admissible discretization of $\Omega \times (0,T)$ in the sense of definition 4.22. Let $\lambda \in (0,+\infty)$ and $\alpha \in (0,2)$ be given. Then there exists at least one $(u,p) \in (H_{\mathcal{D}}(\Omega))^d \times H_{\mathcal{D}}(\Omega)$, solution to (4.52).*

Proof Let us define $V = \{(u,p) \in (H_{\mathcal{D}}(\Omega))^d \times H_{\mathcal{D}}(\Omega)$ s.t. $\int_\Omega p(x)\delta x = 0\}$. Consider the continuous application $F : V \times [0,1] \to V$ such that, for a given $(u,p) \in V$ and $\rho \in [0,1]$, $(\hat{u}, \hat{p}) = F(u,p,\rho)$ is defined by

$$\left\{\begin{array}{ll}\displaystyle\int_\Omega \hat{u}(x)\cdot v(x)\delta x & \displaystyle= \eta\int_\Omega u(x)\cdot v(x)\delta x + \nu[u,v]_{\mathcal{D}} - \int_\Omega p(x)\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x \\ & \displaystyle+\rho\left(\frac{1}{2}\int_\Omega u(x)^2\widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x)\delta x + b_{\mathcal{D}}(u,u,v)\right) \\ & \displaystyle-\int_\Omega f(x)\cdot v(x)\delta x \qquad\qquad\qquad\qquad \forall v\in H_{\mathcal{D}}(\Omega)^d \\ \displaystyle\int_\Omega \hat{p}(x)\cdot q(x)\delta x & \displaystyle= \int_\Omega \widetilde{\mathrm{div}}_{\mathcal{D}}(u)(x)q(x)\delta x + \lambda\ h_{\mathcal{D}}^\alpha\ \langle p,q\rangle_{\mathcal{D}} \qquad \forall q\in H_{\mathcal{D}}(\Omega).\end{array}\right.$$

It is easily checked that the two above relations define a one to one function $F(.,.,.)$. Indeed, the value of $\hat{u}_K^{(i)}$ and $\hat{p}_K$ for a given $K\in\mathcal{M}$ and $i=1,\ldots,d$ are readily obtained by setting $v^{(i)}=1_K$, $v^{(j)}=0$ for $j\neq i$, and $q=1_K$.
The application $F(.,.,.)$ is continuous, and, for a given (u,p) such that $F(u,p,\rho)=(0,0)$, we can apply Proposition 4.19 and (2.8), which prove that (u,p) is bounded independently on ρ. Since $F(u,p,0)$ is an affine function of (u,p) (indeed invertible, see corollary 4.8), we may apply Theorem 4.18 and conclude to the existence of at least one solution (u,p) to (4.52).

Theorem 4.21 (Convergence of the scheme) *Under hypotheses (4.17)-(4.19), let $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ be a sequence of admissible discretizations of Ω in the sense of definition 2.1, such that $h_{\mathcal{D}^{m)}}$ tends to 0 as $m\to\infty$ and such that there exists $\theta>0$ with $\mathrm{regul}(\mathcal{D}^{(m)})\geq\theta$, for all $m\in\mathbb{N}$. Let $\lambda\in(0,+\infty)$ and $\alpha\in(0,2)$ be given. Let, for all $m\in\mathbb{N}$, $(u^{(m)},p^{(m)})\in(H_{\mathcal{D}^{(m)}}(\Omega))^d\times H_{\mathcal{D}^{(m)}}(\Omega)$, be a solution to (4.52) with $\mathcal{D}=\mathcal{D}^{(m)}$. Then there exists a weak solution $(\bar{u},\bar{p})$ of (4.50) in the sense of definition 4.14 and a subsequence of $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$, again denoted $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$, such that the corresponding subsequence of solutions $(u^{(m)})_{m\in\mathbb{N}}$ converges to $\bar{u}$ in $L^2(\Omega)$ and $(p^{(m)}-\frac{1}{2}(u^{(m)})^2)_{m\in\mathbb{N}}$ weakly converges to $\bar{p}$ in $L^2(\Omega)$.*

The proof of this proposition follows that of Proposition 4.10, once proven (see [33]) the following estimate on the approximate pressure: if $(u,p)\in H_{\mathcal{D}}(\Omega)^d\times H_{\mathcal{D}}(\Omega)$ be a solution to (4.52). then there exists C_{37}, only depending on d, Ω, η, ν, λ, α and θ, and not on $h_{\mathcal{D}}$, such that the following inequality holds:

$$\|p\|_{L^2(\Omega)}\leq C_{37}\left(\|f\|_{(L^2(\Omega))^d}+\left(\|f\|_{(L^2(\Omega))^d}\right)^2\right) \tag{4.60}$$

Hence we have the same estimates as in the linear case are available in the steady nonlinear case, the proof of proposition 4.10 holds for all the terms of (4.51) which are present in (4.20). Thus there only remains to prove that for a given $\varphi\in(C_c^\infty(\Omega))^d$, as $m\to+\infty$:

$$T_{18}^{(m)}=\int_\Omega u^{(m)}(x)^2\mathrm{div}_{\mathcal{D}^{(m)}}(P_{\mathcal{D}^{(m)}}\varphi)(x)\delta x \quad\text{tends to}\quad \int_\Omega \bar{u}(x)^2\mathrm{div}\varphi(x)\delta x$$

and

$$T_{19}^{(m)}=b_{\mathcal{D}}(u^{(m)},u^{(m)},P_{\mathcal{D}^{(m)}}\varphi)\quad\text{tends to}\quad b(\bar{u},\bar{u},\varphi),$$

which are consequences of (4.5). We refer to [33] for the details.

4.3.2 The transient case

We now turn to the study of the finite volume scheme for the transient Navier-Stokes equations, the weak formulation of which is given in (4.1).
We first give the definition of an admissible discretization for a space-time domain.

Definition 4.22 (Admissible discretization, transient case) *Let Ω be an open bounded polygonal (polyhedral if $d = 3$) subset of $\mathbb{R}^d$, and $\partial\Omega = \overline{\Omega} \setminus \Omega$ its boundary, and let $T > 0$. An admissible finite volume discretization of $\Omega \times (0,T)$, denoted by $\mathcal{D}$, is given by $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P}, N)$, where $(\mathcal{M}, \mathcal{E}, \mathcal{P})$ is an admissible discretization of Ω in the sense of definition 2.1 and $N \in \mathbb{N}_\star$ is given. We then define $\delta t = T/N$, and we denote by $h_{\mathcal{D}} = \max(h_{(\mathcal{M},\mathcal{E},\mathcal{P})}, \delta t)$ and* $\mathrm{regul}(\mathcal{D}) = \mathrm{regul}(\mathcal{M}, \mathcal{E}, \mathcal{P})$.

Under hypotheses (4.3)-(4.7), let $\mathcal{D}$ be an admissible discretization of $\Omega \times (0,T)$ in the sense of definition 4.22 and let $\lambda \in (0, +\infty)$ and $\alpha \in (0, 2)$ be given. We write a Crank–Nicholson scheme for the time discretization, and follow the nonlinear steady–state case for the space discretization; the finite volume scheme for the approximation of the solution (4.1)–(4.2) is then:

$$\left\{ \begin{array}{l} u_0 \in H_{\mathcal{D}}(\Omega)^d, \\ u_{0,K} = \dfrac{1}{|K|} \displaystyle\int_K u_{\mathrm{ini}}(x) \delta x, \ \forall K \in \mathcal{M}, \end{array} \right. \tag{4.61}$$

and, again using Bernoulli's pressure $p + \frac{1}{2}u^2$ instead of p, again denoted by p,

$$\left\{ \begin{array}{l} (u_{n+1}, p_{n+\frac{1}{2}}) \in (H_{\mathcal{D}}(\Omega))^d \times H_{\mathcal{D}}(\Omega), \\ \int_\Omega p_{n+\frac{1}{2}}(x) \delta x = 0, \ u_{n+\frac{1}{2}} = \frac{1}{2}(u_{n+1} + u_n), \\ \displaystyle\int_\Omega (u_{n+1}(x) - u_n(x)) \cdot v(x) \delta x + \nu \delta t [u_{n+\frac{1}{2}}, v]_{\mathcal{D}} \\ -\delta t \displaystyle\int_\Omega p_{n+\frac{1}{2}}(x) \widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x) \delta x + \frac{\delta t}{2} \int_\Omega u_{n+\frac{1}{2}}(x)^2 \widetilde{\mathrm{div}}_{\mathcal{D}}(v)(x) \delta x \\ +\delta t b_{\mathcal{D}}(u_{n+\frac{1}{2}}, u_{n+\frac{1}{2}}, v) = \displaystyle\int_{n\delta t}^{(n+1)\delta t} \int_\Omega f(x,t) \cdot v(x) \delta x \delta t, \\ \displaystyle\int_\Omega \widetilde{\mathrm{div}}_{\mathcal{D}}(u_{n+\frac{1}{2}})(x) q(x) \delta x = -\lambda \ h_{\mathcal{D}}^\alpha \ \langle p_{n+\frac{1}{2}}, q \rangle_{\mathcal{D}}, \\ \forall v \in H_{\mathcal{D}}(\Omega)^d, \forall q \in H_{\mathcal{D}}(\Omega), \ \forall n \in \mathbb{N}. \end{array} \right. \tag{4.62}$$

In (4.62), we consider the approximation of $b_{\mathcal{D}}$ given by (4.53). We then define the set $H_{\mathcal{D}}(\Omega \times (0,T))$ of piecewise constant functions in each $K \times (n\delta t, (n+1)\delta t)$, $K \in \mathcal{M}$, $n \in \mathbb{N}$, and we define $(u, p) \in H_{\mathcal{D}}(\Omega \times (0,T))$ by

$$u(x,t) = u_{n+\frac{1}{2}}(x), \ \ \text{and } p(x,t) = p_{n+\frac{1}{2}}(x), \ \ \text{for a.e. } (x,t) \in \Omega \times (n\delta t, (n+1)\delta t), \ \forall n \in \mathbb{N}. \tag{4.63}$$

Remark 4.23 (Time discretization) *It is well known that the Crank–Nicholson discretization is implicit. If we use the θ scheme: $u_{n+\frac{1}{2}} = \theta u_{n+1} + (1-\theta) u_n$, with $\theta \in [\frac{1}{2}, 1]$, the convergence proof which follows applies with a few minor changes. Variable time steps may also be considered.*

Let us now prove the existence of at least one solution to scheme (4.61)-(4.63).

Proposition 4.24 (Existence of a discrete solution) *Under hypotheses (4.3)-(4.7), let $\mathcal{D}$ be an admissible discretization of $\Omega \times (0,T)$ in the sense of Definition 4.22. Let $\lambda \in (0,+\infty)$ and $\alpha \in (0,2)$ be given. Then there exists at least one $(u,p) \in (H_{\mathcal{D}}(\Omega \times (0,T)))^d \times H_{\mathcal{D}}(\Omega \times (0,T))$, solution to (4.61)-(4.63).*

PROOF We remark that, for a given $n = 0, \ldots, N-1$, taking as unknown $u_{n+\frac{1}{2}}$, and noting that $u_{n+1} = 2u_{n+\frac{1}{2}} - u_n$, Scheme (4.62) is under the same form as scheme (4.52), with $\eta = \frac{2}{\delta t}$ and with a term in u_n included in the right hand side. Therefore the existence of at least one solution follows from proposition 4.20.

We then have the following estimate.

Proposition 4.25 (Discrete $L^2(0,T; H_0^1(\Omega))$ estimate on velocities) *Under hypotheses (4.3)-(4.7), let $\mathcal{D}$ be an admissible discretization of $\Omega \times (0,T)$ in the sense of definition 4.22. Let $\lambda \in (0,+\infty)$ and*

$\alpha \in (0,2)$. *Let* $(u,p) \in (H_{\mathcal{D}}(\Omega \times (0,T)))^d \times H_{\mathcal{D}}(\Omega \times (0,T))$, *be a solution to (4.61)-(4.63). Then there exists* $C_{38} > 0$, *only depending on* d, Ω, ν, u_0, f, T *such that the following inequalities hold*

$$\|u\|_{L^\infty(0,T;L^2(\Omega)^d)} \le C_{38}\,, \tag{4.64}$$

$$\|u\|_{L^2(0,T;H_{\mathcal{D}}(\Omega)^d)} \le C_{38}\,, \tag{4.65}$$

and

$$\lambda\ h_{\mathcal{D}}^{\alpha} \sum_{n=0}^{N-1} \delta t |p_{n+\frac{1}{2}}|_{\mathcal{D}}^2 = \lambda\ h_{\mathcal{D}}^{\alpha} \int_0^T |p(\cdot,t)|_{\mathcal{D}}^2 \delta t \le C_{38}\,. \tag{4.66}$$

Proof Let $p = 1,\ldots,N$. We get, setting $v = u_{n+\frac{1}{2}}$ in the first equation of (4.62), summing on $K \in \mathcal{M}$ and $n = 0,\ldots,p-1$ in the first equation of (4.62) and using property (4.56),

$$\left\{ \begin{array}{l} \displaystyle \frac{1}{2}\sum_{n=0}^{p-1}\int_\Omega (u_{n+1}(x)^2 - u_n(x)^2)\delta x + \nu \sum_{n=0}^{p-1} \delta t [u_{n+\frac{1}{2}}, u_{n+\frac{1}{2}}]_{\mathcal{D}} - \\ \displaystyle \sum_{n=0}^{p-1} \delta t \int_\Omega p_{n+\frac{1}{2}}(x) \widetilde{\mathrm{div}}_{\mathcal{D}}(u_{n+\frac{1}{2}})(x)dx = \sum_{n=0}^{p-1} \int_{n\delta t}^{(n+1)\delta t} \int_\Omega f(x,t)\cdot u_{n+\frac{1}{2}}(x)\delta x \delta t, \end{array} \right.$$

This leads, setting $q = p_{n+\frac{1}{2}}$ in the second equation of (4.62), to

$$\left\{ \begin{array}{l} \displaystyle \frac{1}{2}\int_\Omega (u_p(x)^2 - u_0(x)^2)\delta x + \nu \sum_{n=0}^{p-1} \delta t [u_{n+\frac{1}{2}}, u_{n+\frac{1}{2}}]_{\mathcal{D}} + \\ \displaystyle \lambda\ h_{\mathcal{D}}^{\alpha} \sum_{n=0}^{p-1} \delta t |p_{n+\frac{1}{2}}|_{\mathcal{D}}^2 = \int_0^{p\delta t} \int_\Omega f(x,t)\cdot u(x,t) \delta x \delta t. \end{array} \right. \tag{4.67}$$

Setting $p = N$ in (4.67) gives (4.65) and (4.66). The discrete Poincaré inequality (2.7) and the inequality $\|u_0\|_{L^2(\Omega)^d} \le \|u_{\rm ini}\|_{L^2(\Omega)^d}$ give

$$\left\{ \ \|u_p\|_{L^2(\Omega)^d}^2 \le \frac{\mathrm{diam}(\Omega)^2}{2\nu} \|f\|_{L^2(\Omega\times(0,T))^d}^2 + \|u_{\rm ini}\|_{L^2(\Omega)^d}^2,\ \forall p = 1,\ldots,N, \right.$$

which proves (4.64), since $\|u_{n+\frac{1}{2}}\|_{L^2(\Omega)^d} \le \frac{1}{2}(\|u_n\|_{L^2(\Omega)^d} + \|u_{n+1}\|_{L^2(\Omega)^d})$ for all $n = 0,\ldots,N-1$.
In order to obtain some compactness, we use, as in the steady–state case, some estimates on translations. But now, we also need estimates on the time translations. We refer to [24] for an introduction to the estimate on the time translates for parabolic problems, and to [33] for the proof of the following proposition:

Proposition 4.26 (Space and time translate estimates) *Under hypotheses (4.3)-(4.7), let* $\mathcal{D}$ *be an admissible discretization of* $\Omega \times (0,T)$ *in the sense of definition 4.22. Let* $\lambda \in (0,+\infty)$, $\alpha \in (0,2)$ *and* $\theta > 0$, *such that* $\mathrm{regul}(\mathcal{D}) \ge \theta$. *Let* $(u,p) \in (H_{\mathcal{D}}(\Omega \times (0,T)))^d \times H_{\mathcal{D}}(\Omega \times (0,T))$, *be a solution to (4.61)-(4.63). We denote by* u *the extension in* $\mathbb{R}^d \times \mathbb{R}$ *of* u *by* 0 *outside of* $\Omega \times (0,T)$. *Then there exists* $C_{39} > 0$ *and* $C_{40} > 0$, *only depending on* d, Ω, ν, λ, α, u_0, f, θ *and* T *such that the following inequalities hold:*

$$\|u(\cdot+\xi,\cdot) - u\|_{L^2(\mathbb{R}^d\times\mathbb{R})}^2 \le C_{39}\,|\xi|(|\xi| + 4h_{\mathcal{D}}),\ \forall \xi \in \mathbb{R}^d, \tag{4.68}$$

and

$$\|u(\cdot,\cdot+\tau) - u\|_{L^1(\mathbb{R};L^2(\mathbb{R}^d))} \le C_{40}\,|\tau|^{1/2},\ \forall \tau \in \mathbb{R}. \tag{4.69}$$

We now have all the tools for the convergence of the approximate solutions:

Theorem 4.27 (Convergence of the scheme) *Under hypotheses (4.3)-(4.7), let $\theta > 0$ be given and let $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$ be a sequence of admissible discretizations of $\Omega \times (0,T)$ in the sense of definition 4.22, such that* $\mathrm{regul}(\mathcal{D}^{(m)}) \geq \theta$ *and $h_{\mathcal{D}^{m)}}$ tends to 0 as $m \to \infty$. Let $\lambda \in (0,+\infty)$ and $\alpha \in (0,2)$ be given. Let, for all $m \in \mathbb{N}$, $(u^{(m)}, p^{(m)}) \in (H_{\mathcal{D}^{(m)}}(\Omega \times (0,T)))^d \times H_{\mathcal{D}^{(m)}}(\Omega \times (0,T))$, be a solution to (4.61)-(4.63) with $\mathcal{D} = \mathcal{D}^{(m)}$. Then there exists a subsequence of $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$, again denoted $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$, such that the corresponding subsequence of solutions $(u^{(m)})_{m\in\mathbb{N}}$ converges in $L^2(\Omega \times (0,T))$ to a weak solution $\bar{u}$ of (4.1)-(4.2) in the sense of definition 4.1.*

Proof Let us assume the hypotheses of the theorem. Using translate estimates (4.68) and (4.69) in the space $L^1(\mathbb{R}^d \times \mathbb{R})$, we can apply Kolmogorov's theorem (see [24] for parabolic problems). We get that there exists $\bar{u} \in L^1(\Omega \times (0,T))$ and a subsequence of $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$, again denoted $(\mathcal{D}^{(m)})_{m\in\mathbb{N}}$, such that the corresponding subsequence of solutions $(u^{(m)})_{m\in\mathbb{N}}$ converges in $L^1(\Omega \times (0,T))$ to $\bar{u}$ as $m \to \infty$. Using (4.65), we get $\|u^{(m)}\|_{L^2(0,T;H_{\mathcal{D}_m}(\Omega))} \leq C_{38}$, for all $m \in \mathbb{N}$, which gives, using the discrete Sobolev inequalities, $\|u^{(m)}\|_{L^1(0,T;L^4(\Omega))} \leq C_{41}$, for all $m \in \mathbb{N}$. Using a classical result on spaces $L^p(0,T;L^q(\Omega))$, we get that $(u^{(m)})_{m\in\mathbb{N}}$ converges in $L^1(0,T;L^2(\Omega))$ to $\bar{u}$ as $m \to \infty$. Thanks to (4.64), we have $\|u^{(m)}\|_{L^\infty(0,T;L^2(\Omega)^d)} \leq C_{38}$, for all $m \in \mathbb{N}$. The same result on spaces $L^p(0,T;L^q(\Omega))$ implies that $(u^{(m)})_{m\in\mathbb{N}}$ converges in $L^2(0,T;L^2(\Omega))$ to $\bar{u}$ as $m \to \infty$. We can therefore pass to the limit in (4.68). The resulting inequality implies $\bar{u} \in L^2(0,T;H_0^1(\Omega)^d)$ (see [24]). Passing to the limit in (4.64) leads to $\bar{u} \in L^\infty(0,T;L^2(\Omega)^d)$.
The proof $\bar{u}$ is a weak solution of (4.1)-(4.2) in the sense of definition 4.1 is then obtained by passing to the limit in the scheme, we refer to [33] for the details.

Remark 4.28 *Using the above proof of convergence, we get the energy inequality for $d = 2$ or 3 from inequality (4.67), since we have the property*

$$\int_0^T \int_\Omega (\nabla u^{(i)}(x,t))^2 \delta x \delta t \leq \liminf_{m\to\infty} \sum_{n=0}^{N^{(m)}-1} \delta t [u^{(m,i)}_{n+\frac{1}{2}}, u^{(m,i)}_{n+\frac{1}{2}}]_{\mathcal{D}^{(m)}}$$

4.4 Stabilization by clustering

The drawback of the stabilization used in system (4.23) is that it yields some redistribution of the fluid mass over the whole domain. Moreover, in order to obtain convergence, one needs to let the stabilization parameter tend to 0. For both reasons, we replace here the system (4.23) by:

$$\left\{\begin{array}{l} \eta\,|K|\,u_K - \nu \displaystyle\sum_{L\in\mathcal{N}_K} \tau_{\sigma_{KL}}(u_L - u_K) - \nu \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\mathrm{ext}}} \frac{|\sigma|}{d_{K,\sigma}}(0 - u_K) \\ \qquad\qquad + \displaystyle\sum_{L\in\mathcal{N}_K} \frac{1}{2}\,|\sigma_{KL}|\,\mathbf{n}_{KL}\,(p_L - p_K) = \int_K f(x)\delta x \\ \displaystyle\sum_{L\in\mathcal{N}_K} \frac{1}{2}\,|\sigma_{KL}|\,\mathbf{n}_{KL}\cdot(u_K + u_L) - \sum_{L\in\mathcal{N}_K} \lambda_{KL}\tau_{\sigma_{KL}}(p_L - p_K) = 0, \end{array}\right. \tag{4.70}$$

where the parameters λ_{KL} are chosen according to the following method. The family of control volumes $\mathcal{M}$ is partitioned into disjoint clusters of neighbouring control volumes. These clusters are chosen such that the distance between two control volumes belonging to the same cluster is bounded by $Ch_{\mathcal{D}}$, where C is a given constant. Then the stabilising parameter λ_{KL} is chosen equal to some $\lambda > 0$ for any pair of neighbouring control volumes K and L belonging to the same cluster, 0 otherwise. The value λ is chosen large enough to prevent instabilities.
An example of such an algorithm for partitioning $\mathcal{M}$ is the following:

- select an order for the control volumes K_i, $i = 1, M$;

- for $i = 1$ to M, if K_i and all its neighbours do not yet belong to a cluster, initialise a new one by K_i and all its neighbours;
- for $i = 1$ to M, if K_i does not yet belong to one of the clusters, one of its neighbour does: include K_i in this cluster.

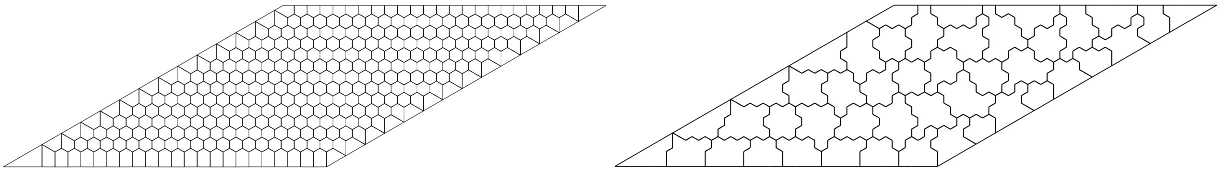

Figure 4.1: Voronoï mesh and clusters of controls volumes.

This stabilisation technique has been implemented with success for the transient Navier Stokes equations with or without the energy equation [15]. We shall present in a further paper the mathematical study of convergence of this stabilised scheme. Note that a crucial difference with the stabilisation of [13] is that there is no need to let λ tend to 0 with the size of the mesh, which means that the presence of a finite stabilization does not decrease the quality of the approximation. The numerical examples given below illustrate this kind of property.

4.5 Numerical results

An industrial implementation of a colocated finite volume scheme may be found in [4] for instance, where complex applications are considered. Focusing in this paper on properties of convergence and error estimates, some simple numerical experiments are described here to observe the convergence rate of Schemes (4.22) and (4.61)-(4.62) with respect to the space and time discretizations. To that purpose, we use a prototype code where the nonlinear equations are solved by an under relaxed Newton method, and the linear systems by a direct band Gaussian elimination solver. This code handles Stokes or Navier-Stokes problems with various boundary conditions, using non uniform rectangular or triangular meshes on general 2D polygonal domains.

The linear Stokes equations are first considered in the case $d = 2$, $\Omega = (0, 1) \times (0, 1)$, $\nu = 1$, and f is taken to satisfy (4.16) with a solution equal to

$$\left\{ \begin{array}{ll} \bar{u}^{(1)}(x^{(1)}, x^{(2)}) & = -\partial^{(2)}\Psi(x^{(1)}, x^{(2)}) \\ \bar{u}^{(2)}(x^{(1)}, x^{(2)}) & = \partial^{(1)}\Psi(x^{(1)}, x^{(2)}) \\ \bar{p}(x^{(1)}, x^{(2)}) & = 100\ \left((x^{(1)})^2 + (x^{(2)})^2\right), \end{array} \right.$$

denoting by $\Psi(x^{(1)}, x^{(2)}) = 1000\,[x^{(1)}(1-x^{(1)})x^{(2)}(1-x^{(2)})]^2$. The approximate solution (u, p) is computed with the scheme (4.22). The observed numerical order of convergence, considering the norms $\|u - P_{\mathcal{D}}\bar{u}\|_{L^2(\Omega)}$ and $\|p - P_{\mathcal{D}}\bar{p}\|_{L^2(\Omega)}$, is equal to 2 for the velocity components, and to 1 for the pressure in the cases of non uniform rectangular and square meshes (from 400 to 6400 grid blocks). Note that in these cases, there is apparently no need for a significant positive value of the stabilization coefficient λ. The observed numerical order of convergence is similar in the case of triangular meshes (from 1400 to 5600 grid blocks), but values such as $\lambda = 10^{-4}$, $\alpha = 1$ have to be used in order to avoid oscillations in the pressure field. This confirms that in the case of triangles, the approximate pressure space is too large to avoid stabilization. In fact, other tests were performed (e.g. the classical backward step) which show that stabilization is also needed in the case of rectangles when more severe problems are considered. Note that in industrial implementations, stabilization may be performed with other means, see [55], [4], (see also [11] in the triangular case).

We then proceed to a similar comparison in the case of transient nonlinear problems. Considering a transient adaptation of the above steady-state analytical solution, the continuous problem is then defined by zero initial and boundary conditions, $T = 0.1$, and the function f is taken to satisfy (4.1) with a solution equal to

$$\left\{\begin{array}{ll} \bar{u}^{(1)}(x^{(1)},x^{(2)},t) & = -t\ \partial^{(2)}\Psi(x^{(1)},x^{(2)}) \\ \bar{u}^{(2)}(x^{(1)},x^{(2)},t) & = t\ \partial^{(1)}\Psi(x^{(1)},x^{(2)}) \\ \bar{p}(x^{(1)},x^{(2)},t) & = 100\ t\ \left((x^{(1)})^2+(x^{(2)})^2\right), \end{array}\right.$$

with the same function Ψ as above. We again observe an order 2 of convergence of the approximate solution at times $t = .05$ and $t = .1$, when the space and the time discretizations are simultaneously modified with the same ratio (from $\delta t = 0.01$ to $\delta t = 0.0025$ as the size of the mesh is divided by 4). Similar observations are still valid for the classical Green-Taylor example.

4.6 Discretization of the full viscous tensor

Let us write the conservation of mass and momentum for a general compressible viscous flow:

$$\left\{\begin{array}{l} \partial_t\rho + \mathrm{div}(\rho u) = 0, \text{ in } \Omega\times(0,T), \\ \partial_t(\rho u) - \mathrm{div}\tau + \nabla p + \mathrm{div}(\rho u\otimes u) = \rho g \text{ in } \Omega\times(0,T), \end{array}\right. \tag{4.71}$$

with an energy equation, which we do not state here, and where the viscous stress tensor τ is defined by:

$$\tau = \mu(\nabla u + (\nabla u)^t - \frac{2}{3}\mathrm{div}u\, I_d) + \xi \mathrm{div}u I_d,$$

where μ and ξ are respectively the shear (resp. bulk) viscosities.
If the coefficients μ and ξ are assumed constant, then the divergence of the viscous stress tensor maybe written as:

$$\mathrm{div}\tau = \mu\Delta u + (\xi + \frac{\mu}{3})\nabla(\mathrm{div}u). \tag{4.72}$$

Let us then give a finite volume discretization of τ with the discrete derivatives introduced in Chapter 3 (see definition of the discrete gradient (3.13)–(3.15)).
For a given function $v \in (H_{\mathcal{D}})^2$, let $\widehat{\mathrm{div}}_{\mathcal{D}} \in H_{\mathcal{D}}$ be the function defined by:

$$(\widehat{\mathrm{div}}_{\mathcal{D}}v)|_K = \frac{1}{|K|}\sum_{L\in\mathcal{N}_K}\tau_{\sigma_{KL}}(v_L - v_K)\cdot(x_\sigma - x_K), \text{ for any } K\in\mathcal{M}. \tag{4.73}$$

Note that this discrete divergence operator is defined with the same discretization of derivatives as in the discrete gradient (3.13)–(3.15). However, it is not the discrete dual operator to this latter discrete gradient.
We may then define the discrete operator $\nabla_{\mathcal{D}}(\widehat{\mathrm{div}}_{\mathcal{D}})$ by:

$$(\nabla_{\mathcal{D}}(\widehat{\mathrm{div}}_{\mathcal{D}}v))|_K = \frac{1}{|K|}\sum_{L\in\mathcal{N}_K} G_{KL}, \tag{4.74}$$

where

$$G_{KL} = \tau_{\sigma_{KL}}\big((\widehat{\mathrm{div}}_{\mathcal{D}})v)|_K(x_{\sigma_{KL}} - x_K) + (\widehat{\mathrm{div}}_{\mathcal{D}})v)|_K(x_{\sigma_{KL}} - x_K)\big). \tag{4.75}$$

The complete discretization of (4.72) is then obtained by using the classical cell centred finite volume scheme described in Chapter 2 for the Laplace operator, and the formulae (4.73)–(4.75) for the discretization of the term $\nabla(\mathrm{div}u)$. We thus have an easy way of discretizing the full viscous tensor, using the discrete gradient introduced in Chapter 3. This scheme has been implemented successfully on test cases, and its convergence analysis is the object of ongoing work.

4.7 Conclusions

The above numerical results show that the theoretical error estimate which is proved in Section 4.2 for the linear Stokes equations is non optimal; a sharper estimate is currently being written [34] under more regularity assumptions on the mesh.

The proof of convergence of the full space-time discrete approximation of (4.1) given by (4.62) uses estimates on the time translates, which were introduced in the $L^2(\Omega \times (0,T))$ framework for the proof of convergence of the finite volume method for degenerate parabolic equations [36; 24] and used for several other cases, see e.g.[29]. A major difficulty which arises here is the handling on the nonlinear convective term, as in the continuous case, which leads us to establish an estimate on the time translates in $L^1(0,T;L^2(\Omega))$. This new technique may be used for parabolic problems with other type of nonlinearities.

We remarked that industrial codes use other types of stabilisations than the one used here. Further works will be devoted to the mathematical study of such stabilisations, for which, to our knowledge, no proof of convergence is known up to now. The cluster stabilization which has been introduced in our numerical code is also being studied.

Finally, let us mention ongoing work on the generalisation of the scheme studied here to the full transient Navier-Stokes equations including the energy balance, under the Boussinesq approximation, where the viscous stress tensor is discretized as shown in paragraph 4.6.

Bibliography

[1] I. Aavatsmark, T. Barkve, O. Boe and T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: Derivation of the methods. *SIAM Journal on Sc. Comp.*, 19 (1998), 1700–1716.

[2] I. Aavatsmark, T. Barkve, O. Boe and T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part II: Discussion and numerical results. *SIAM Journal on Sc. Comp.*, 19 (1998), 1717–1736.

[3] L. Angermann, A finite element method for the numerical solution of convection-dominated anisotropic diffusion equations. *Numer. Math.* 85 (2000), 175–195.

[4] F. Archambeau, N. Mehitoua and M. Sakiz, Code Saturne: A Finite Volume Code for Turbulent flows *Int. J. Finite Volumes, 23-Mar-2004, http://averoes.math.univ-paris13.fr/IJFV/*

[5] B.F. Armaly, F. Durst, J.C.F. Pereira and B. Schönung, "Experimental and Theoretical investigation of backward-facing step flow" *J. Fluid Mech.* **127**, 473–496, 1983.

[6] I.E. Barton: "A numerical study of flow over a confined backward-facing step", *Int. J. for Num. Meth. in Fluids*, **21**, 653–665, 1995.

[7] M. Benartzi, J.P. Croisille, M. Ben-Artzi, JP. Croisille, D. Fishelov, S. Trachtenberg, A pure-compact scheme for the streamfunction formulation of Navier-Stokes equations, *submitted*.

[8] J. Berton Comparaison de différentes méthodes pour apprécier les options américaines. *Thesis of Marne-la-Vallée university (France)*, in preparation (2005).

[9] Ph. Blanc, R. Eymard and R. Herbin, An error estimate for finite volume methods for the Stokes equations on equilateral triangular meshes, *Num. Meth. PDE*, Published Online: 8 Apr 2004

[10] Ph. Blanc, R. Eymard, & R. Herbin, A staggered finite volume scheme on general meshes for the generalized Stokes problem in two space dimensions, Int. J. Finite Volumes, 2005.

[11] S. Boivin, F. Cayré, J.M. Hérard, A finite volume method to solve the Navier-Stokes equations for incompressible flows on unstructured meshes, *Int. J. Therm. Sci.*, **38**, 806-825, 2000.

[12] F. Boyer, P. Fabrie, *Eléments d'analyse pour l'étude de quelques modèles d'écoulements de fluides visqueux incompressibles*, submitted.

[13] F. Brezzi, J. Pitkäranta, On the stabilization of finite element approximations of the Stokes equations. *Efficient solutions of elliptic systems*, 11–19, Kiel, 1984, *Notes Numer. Fluid Mech., Vieweg, Braunschweig,*, **10**, 1984.

[14] E. Chénier, R. Eymard and X. Nicolas, A Finite Volume Scheme for the Transport of Radionucleides *Porous Media: Simulation of Transport Around a Nuclear Waste Disposal Site: The COUPLEX Test Cases, Alain Bourgeat and Michel Kern eds, Computational Geosciences*, 8 (2004), 163–172.

[15] E. Chénier, R. Eymard and O. Touazi, Numerical results using a colocated finite volume scheme on unstructured grids for incompressible fluid flows, submitted.

[16] S.C. Chou, Analysis and convergence of a covolume method for the generalized Stokes problem. *Math.comp.*, **66**, 85–104, 1997.

[17] Ciarlet, P.G. (1991), *Basic error estimates for elliptic problems* in: *Handbook of Numerical Analysis* **II** (North-Holland,Amsterdam) 17-352.

[18] Y. Coudière, J.P. Vila and P. Villedieu, Convergence rate of a finite volume scheme for a two-dimensional convection-diffusion problem. *M2AN Math. Model. Numer. Anal.*, 33 (1999), 493–516.

[19] Y. Coudière, T. Gallouët and R. Herbin, Discrete Sobolev Inequalities and L^p Error Estimates for Finite Volume Solutions of Convection Diffusion Equations, *M2AN*, **35**,4 , 767–778, 2001.

[20] K. Deimling, *Nonlinear functional analysis*, Springer-Verlag, Berlin, 1985.

[21] K. Domelevo, P. Omnes, A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *submitted* (2005).

[22] R. Eymard and T. Gallouët, H-convergence and numerical schemes for elliptic equations *SIAM Journal on Numerical Analysis*, **41**, 2, 539–562, 2000.

[23] R. Eymard, T. Gallouët, M. Ghilani and R. Herbin, An error estimate for a finite volume scheme for a nonlinear hyperbolic equation on a triangular mesh, *IMA Journal of Numerical Analysis*, **18**, 563–594, 1998.

[24] R. Eymard, T. Gallouët and R. Herbin, Finite Volume Methods, *Handbook of Numerical Analysis*, Vol. VII, pp. 713–1020. Edited by P.G. Ciarlet and J.L. Lions (North Holland).

[25] R. Eymard, T. Gallouët and R. Herbin, Convergence of finite volume approximations to the solutions of semilinear convection diffusion reaction equations. *Numer. Math.*, 82 (1999), 91–116.

[26] R. Eymard, T. Gallouët and R. Herbin, Finite volume approximation of elliptic problems and convergence of an approximate gradient. *Appl. Num. Math.*, 37 (2001),31–53.

[27] R. Eymard, T. Gallouët and R. Herbin, A finite volume for anisotropic diffusion problems. *Comptes rendus à l'Académie des Sciences* , 339 (2004), 299–302.

[28] R. Eymard, T. Gallouët and R. Herbin, A cell-centered finite volume approximation for second order partial derivative operators with full matrix on unstructured meshes in any space dimension, submitted.

[29] R. Eymard, T. Gallouët, R. Herbin, A. Michel, Convergence of a finite volume scheme for nonlinear degenerate parabolic equations. *Num. Math*, 92 (2002), 41–82.

[30] R. Eymard and R. Herbin, A cell-centered finite volume scheme on general meshes for the Stokes equations in two dimensions, *C. R. Math. Acad. Sci., Mathématiques*, **337**, 2, 125–128, 2003.

[31] R. Eymard and R. Herbin, A staggered finite volume scheme on general meshes for the steady Navier-Stokes problem in two space dimensions, *preprint LATP http://www.cmi.univ-mrs.fr/LATP, see also Proceedings of the ENUMATH conference, Prague 2003.*

[32] R. Eymard, & R. Herbin, A staggered finite volume scheme on general meshes for the Navier-Stokes equations in two space dimensions, Int. J. Finite Volumes, 2005.

[33] R. Eymard, R. Herbin and J-C.Latché, Convergence analysis of a colocated finite volume scheme for the incompressible Navier-Stokes equations on general 2 or 3D meshes, submitted.

[34] R. Eymard, R. Herbin and J-C.Latché, An error estimate for a colocated scheme for the Stokes equations, in preparation.

[35] R. Eymard, D. Hilhorst, M. Vohralík, Combined finite volume-nonconforming/mixed-hybrid finite element scheme for degenerate parabolic problems, *submitted* (2004).

[36] R. Eymard, T. Gallouët, D. Hilhorst and Y. Naït Slimane, Finite volumes and nonlinear diffusion equations, RAIRO Modél. Math. Anal. Numér., 32, 1998, 6, 747–761.

[37] J.H. Ferziger, M. Peric, Computational methods for fluid dynamics, Springer, 2002.

[38] P.A. Forsyth, Control volume finite elements, A control volume finite element approach to NAPL groundwater contamination. *SIAM J. Sci. Stat. Comput.*, 12 (1991), 1029–1057.

[39] L.S.-K. Fung, L. Buchanan, and R. Sharma, Hybrid-CVFE Method for Flexible- Grid Reservoir Simulation. *Soc. Pet. Eng. J.*, 19 (1994), 188-199.

[40] E. Gallouët, R. Herbin, Axisymmetric finite volumes for the numerical simulation of CO_2 transport and assimilation in a leaf, submitted.

[41] T. Gallouët, R. Herbin and M.H. Vignal, Error estimates for the approximate finite volume solution of convection diffusion equations with general boundary conditions, *SIAM J. Numer. Anal.*, **37**, 6, 1935 - 1972, 2000.

[42] V. Girault, P.-A. Raviart, *Finite element methods for the Navier-Stokes equations: Theory and algorithms*, Springer, Berlin, 1986.

[43] R. Glowinski, Numerical Methods for fluids (Part3), *Handbook of Numerical Analysis*, Vol. IX. Edited by P.G. Ciarlet and J.L. Lions (North Holland), 2003.

[44] M.D. Gunzburger, *Finite element methods for viscous incompressible flows, A guide to theory, practice, and algorithms*, Computer Science qnd Scientific Computing, Academic Press 1989.

[45] M.D. Gunzburger and R.A Nicolaïdes *Incompressible computational fluid dynamics*, Cambridge University Press, 1993.

[46] F.H. Harlow, J.E. Welch, Numerical calculation of time dependent viscous incompressible flow of fluids with free surface, *Phys. Fluids*, **8**, 2182-2189, 1965.

[47] R. Herbin, An error estimate for a finite volume scheme for a diffusion-convection problem on a triangular mesh. *Num. Meth. P.D.E.* 11 (1995), 165-173.

[48] R. Herbin, Finite volume methods for diffusion convection equations on general meshes. *in Finite volumes for complex applications, Problems and Perspectives, F. Benkhaldoun and R. Vilsmeier eds, Hermes*, (1996) 153–160.

[49] R. Herbin, E. Marchand Finite volume approximation of a class of variational inequalities, IMA J. Numer. Anal. 21 (2001), no. 2, 553–585.

[50] F. Hermeline, A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys.* 160 (2000), 481–499

[51] X.H. Hu and R.A. Nicolaides, Covolume techniques for anisotropic media. *Numer. Math.* 61 (1992), 215–234.

[52] P. A. Jayantha and Ian W. Turner, A Second Order Finite Volume Technique for Simulating Transport in Anisotropic Media", *The Int. J. of Num. Met. for Heat and Fluid Flow*, 13 (2003), 31–56.

[53] P. A. Jayantha and I. W. Turner, A Second Order Control-Volume Finite-Element Least-Squares Strategy for Simulating Diffusion in Strongly Anisotropic Media. *J.Comp. Math.*, 23 (2005), 1–16.

[54] D. Lamberton and B. Lapeyre, An Introduction to Stochastic Calculus Applied to Finance. *Chapman and Hall*, (1995).

[55] S.R. Mathur and J.Y. Murthy, Pressure boundary conditions for incompressible flow using unstructured meshes, *Numer. Heat Transfer, Part B*, **32**, 283–298, 1997.

[56] I.D. Mishev, Finite volume methods on Voronoï meshes. *Num. Meth. P.D.E.*, 14 (1998), 193–212.

[57] J. Nečas, *Equations aux dérivées partielles*, Presses de l'Université de Montréal, 1965.

[58] J.C. Nedelec, *Notions sur les techniques d'éléments finis*, Mathématiques et applications Ellipses, 1980.

[59] R.A. Nicolaides, Direct discretization of planar div-curl problems. *SIAM J. Numer. Anal.* , 29 (1992), 32–56.

[60] R.A Nicolaïdes, Analysis and convergence of the MAC scheme I : The linear problem, *SIAM J. Numer. Anal.*, **29**, 1579–1591, 1992.

[61] R.A Nicolaïdes and X. Wu, Analysis and convergence of the MAC scheme II, Navier-Stokes equations, Math. Comp., **65**, 29–44, 1996.

[62] S.V. Patankar, *Numerical Heat Transfer and Fluid Flow*, Series in Computational Methods in Mechanics and Thermal Sciences, Minkowycz and Sparrow Eds. (Mc Graw Hill), 1980.

[63] R. Peyret and T. Taylor, *Computational methods for for fluid flow*, Springer, New-York, 1893.

[64] O. Pironneau, *Finite element methods for fluids*, John Wiley and sons, 1989.

[65] M. Putti and C. Cordes Finite Element Approximation of the Diffusion Operator on Tetrahedra. *SIAM Journal on Scientific Computing* 19 (1998), 1154–1168.

[66] R. Temam, *Navier-Stokes Equations, Studies in mathematics and its applications*, J.L. Lions, G. Papanicolaou, R.T. Rockafellar Editors, North-Holland, 1977.

[67] O. Touazi, Mise en oeuvre d'un schéma de volumes finis pour les équations de Navier-Stokes compressibles. *Thesis of Marne-la-Vallée university (France)*, in preparation (2007).

[68] S. Wang, Solving convection-dominated anisotropic diffusion equations by an exponentially fitted finite volume method. *Comput. Math. Appl.* 44 (2002), 1249–1265.