A discrete kinetic energy preserving convection operator for variable density flows on locally refined staggered meshes

J.-C. Latché[⋆], B. Piar[⋆], K. Saleh[♯]
[⋆] Institut de Radioprotection et de Sureté Nucléaire (IRSN),
13115 St-Paul-Lez-Durance, France,
[♯] Université de Lyon, CNRS UMR 5208, Université Lyon 1, Institut Camille Jordan,
43 bd 11 novembre 1918; F-69622 Villeurbanne cedex, France.
[jean-claude.latche, bruno.piar]@irsn.fr, saleh@math.univ-lyon1.fr

December 3, 2019

Abstract

In this paper, we build and analyze a scheme for the time-dependent variable density Navier-Stokes equations, able to cope with unstructured non-conforming meshes, with hanging nodes. The time advancement relies on a pressure correction algorithm and the space approximation is based on a low-order staggered non-conforming finite element, the so-called Rannacher-Turek element. The convection term in the momentum balance equation is discretized by a finite volume technique, and a careful construction of the fluxes, especially through non-conforming faces, ensures that the solution obeys a discrete kinetic energy balance. Its consistency is addressed by analyzing a model problem, namely the convection-diffusion equation, for which we theoretically establish a first order convergence in space for energy norms. This convergence order is also observed in the numerical experiments for the Navier-Stokes equations. Navier-Stokes equations, pressure correction scheme, finite volumes, finite elements, stability, kinetic energy, non-conforming local refinement.

1 Introduction

Let Ω be an open bounded connected subset of $\mathbb{R}^d$, with $d \in \{2,3\}$, which is supposed to be polygonal if $d=2$ and polyhedral if $d=3$. Let $T \in \mathbb{R}^+$. We address in this paper the following system of variable density unstationary Navier-Stokes equations:

$$\partial_t(\rho \boldsymbol{u}) + \mathbf{div}(\rho \boldsymbol{u} \otimes \boldsymbol{u}) - \mathrm{div}\big(\boldsymbol{\tau}(\boldsymbol{u})\big) + \boldsymbol{\nabla} p = 0, \qquad \text{on } \Omega \times (0,T), \tag{1.1a}$$

$$\partial_t \rho + \mathrm{div}(\rho \boldsymbol{u}) = 0, \qquad \text{on } \Omega \times (0,T). \tag{1.1b}$$

The unknowns $\boldsymbol{u} \in \mathbb{R}^d$ and $p \in \mathbb{R}$ are the velocity and the pressure in the flow. The density ρ is assumed to be a *known* positive function of $\Omega \times (0,T)$ (so Equation (1.1b) must be seen as a constraint on $\boldsymbol{u}$). The shear stress tensor $\boldsymbol{\tau}$ is given by:

$$\boldsymbol{\tau}(\boldsymbol{u}) = \mu(\boldsymbol{\nabla} \boldsymbol{u} + \boldsymbol{\nabla}^t \boldsymbol{u}) - \frac{2\mu}{3}\, \mathrm{div}\boldsymbol{u}\, I, \tag{1.2}$$

where μ is a positive parameter, possibly depending on $\boldsymbol{x}$. Consequently, we have:

$$\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla} \boldsymbol{u} \geq 0, \ \forall \boldsymbol{u} \in \mathbb{R}^d. \tag{1.3}$$

The two equations of system (1.1) respectively express the momentum balance and the mass conservation of the fluid. System (1.1) is supplemented with initial and boundary conditions:

$$\boldsymbol{u}|_{\partial\Omega} = \boldsymbol{u}_{\partial\Omega}, \qquad \boldsymbol{u}|_{t=0} = \boldsymbol{u}_0. \tag{1.4}$$

The present work is a continuation of a research program undertaken to develop staggered schemes satisfying a discrete kinetic energy balance [1, 10]. This point is crucial with respect to many issues: such a relation readily provides stability estimates, may be seen as a prerequisite for LES applications [4], and, last but not least, is a starting point for the extension of the schemes to compressible flows (compressible Navier-Stokes and Euler equations [13]). The difficulty lies in the definition of the velocity convection operator, which must be in some sense consistent with the discrete mass balance. More specifically, let us define the convection operator $\mathcal{C}$ as $\mathcal{C}: \ v \mapsto \partial_t(\rho v) + \mathbf{div}(\rho v \boldsymbol{u})$, in such a way that the i^{th} component of the convection term in the momentum balance equation reads:

$$\partial_t(\rho u_i) + \mathbf{div}(\rho u_i \boldsymbol{u}) = \mathcal{C}(u_i).$$

Then we observe that, thanks to the mass balance equation, $\mathcal{C}$ may be recast as a transport operator:

$$\mathcal{C}(v) = \rho\big(\partial_t v + \boldsymbol{u} \cdot \boldsymbol{\nabla} v\big),$$

and this equivalence between a divergence and transport form easily allows to show that (formally):

$$\mathcal{C}(v) \ \varphi'(v) = \mathcal{C}(\varphi(v)) = \partial_t(\rho\varphi(v)) + \mathbf{div}(\rho\varphi(v)\boldsymbol{u}),$$

for any regular real function φ. Taking $\varphi(v) = v^2/2$ yields the kinetic energy identity. We show in [15] that a discrete analogue of this computation holds for finite volume operators, with a similar interplay between the momentum and mass balances, although with some numerical dissipation and only for convex functions φ. Hence, even if, for unstructured quadrangular ($d = 2$) or hexahedric ($d = 3$) meshes, we base our space approximation on the Rannacher-Turek element, we are led to implement a finite volume discretization of the convection term in order to obtain a scheme satisfying a discrete kinetic energy balance. The resulting scheme thus mixes finite element and finite volume techniques, in the spirit of the algorithm developed and analyzed for incompressible flows in [24, 25]. It is based on a pressure-correction method (see *e.g.* [14]); its description, for variable density flows and conforming meshes, may be found in [1].

The objective of the present paper is to extend this scheme to non-conforming meshes. To this purpose, we have to define a discretization for the diffusion and for the convection terms, able to deal with faces with hanging nodes. For the diffusion term, we adopt the same finite element approach as for a regular mesh, and the extension to non-conforming meshes is easily obtained by slightly relaxing the weak continuity constraint across a face [3] (mean continuity is required only across the whole face containing the hanging node and not across each sub-face having this node as vertex). The construction of the convection operator is more intricate, and is the essential difficulty tackled in this work. Basically, the required consistency of the discrete mass and momentum balance equations is sufficient by itself to suggest a definition; this latter is, in some way, abstract, in the sense that it does not rely on a standard finite volume technique based on control volumes of known shape, but on a set of necessary algebraic equation, in a spirit reminiscent of mimetic schemes. The consistency of this operator then remains to be proved; to this purpose, we perform the error analysis of the scheme on a model problem, namely the convection-diffusion equation.

As mentioned above, in their version working on regular meshes (*i.e.* without hanging nodes), these discrete operators have already been used as building bricks in algorithms tackling various problems:

variable density low Mach number flows [1] (as here), barotropic and non-barotropic compressible flows [9, 15], drift-flux two-phase flow model [11, 12, 16]. The present extension thus opens the road to the implementation of local refinement in these schemes (and, as a matter of fact, this potentiality is already offered by our development platform, the open-source software library for fluid applications CALIF^{3}S [5], as a striking outcome of the use of Object Oriented Programming).

This paper is organized as follows. We first describe the space discretization (Section 2) then the proposed scheme (Section 3). Section 4 is devoted to the error analysis of the scheme on a model problem, and Section 5 gathers some numerical experiments which exemplifies this theoretical study. In Sections 2 and 3, we present the scheme as it is utilized in industrial applications at IRSN, through CALIF^{3}S. In particular, we get rid of restrictions on the mesh which are necessary for the theoretical analysis, but often overcome in practice, and the introduction of which is consequently postponed to Section 4.

2 Meshes and discretization spaces

Definition 2.1 (Unrefined mesh) A mesh $\mathcal{M}_0$ is said an unrefined mesh if it is a regular decomposition (in the usual sense of the finite element literature, see *e.g.* [7])) of the domain Ω either in quadrilaterals ($d = 2$) or hexahedra ($d = 3$). Each cell K of $\mathcal{M}_0$ is defined by the image by the standard Q_1 mapping associated with its vertices, denoted by $\mathcal{Q}_K$, of the unit square or cube $(0,1)^d$.

Let us now define a refinement process. In two dimensions, it consists in cutting a cell K in four sub-cells, which are defined as the image by $\mathcal{Q}_K$ of the four sub-squares of the unit square $(\alpha_1/2,\ (\alpha_1+1)/2)\times(\alpha_2/2,\ (\alpha_2+1)/2)$, for $(\alpha_1,\ \alpha_2)\in\{0,1\}^2$. In three dimensions, sub-cells are obtained by applying $\mathcal{Q}_K$ to the eight subset of the unit cube $(\alpha_1/2,\ (\alpha_1+1)/2)\times(\alpha_2/2,\ (\alpha_2+1)/2)\times(\alpha_3/2,\ (\alpha_3+1)/2)$, for $(\alpha_1,\ \alpha_2,\ \alpha_3)\in\{0,1\}^3$. The additional vertices produced by this process lie in the mid-point of a coarse edge in 2D, and at the center of a coarse face in 3D (precisely speaking, at the image by $\mathcal{Q}_K$ of the mass center of the associated face of the reference unit cube). The following lemma is essential for the well-posedness of the refinement process, and is a (not so easy) consequence of the properties of the Q_1 mapping.

Lemma 2.1 The sub-cells produced by the refinement process are themselves the image of the unit square or cube by the Q_1 mapping associated with their vertices.

Proof. We give the proof in two dimensions, the extension to $d = 3$ being easy although cumbersome. Let K be a quadrangle of vertices $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ and $\boldsymbol{a}_4$. The cell K is thus defined by:

$$K = \big\{\sum_{i=1}^{4} \varphi_i(\hat{\boldsymbol{x}})\ \boldsymbol{a}_i,\ \hat{\boldsymbol{x}}\in(0,1)^2\big\},$$

with $\varphi_1(\hat{\boldsymbol{x}}) = (1-\hat{x}_1)\,(1-\hat{x}_2)$, $\varphi_2(\hat{\boldsymbol{x}}) = \hat{x}_1\,(1-\hat{x}_2)$, $\varphi_3(\hat{\boldsymbol{x}}) = \hat{x}_1\,\hat{x}_2$ and $\varphi_4(\hat{\boldsymbol{x}}) = (1-\hat{x}_1)\,\hat{x}_2$. A fundamental property of the Q_1 mapping is that it may be obtained by 2 (in fact d) successive interpolations, performed in any order. Indeed, let us denote by $\mathcal{I}(\boldsymbol{a},\boldsymbol{b},s)$ the following interpolated point between $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\mathcal{I}(\boldsymbol{a},\boldsymbol{b},s) = (1-s)\,\boldsymbol{a} + s\boldsymbol{b}.$$

Then, by the definition of φ, we get:

$$K = \big\{\mathcal{I}\big(\mathcal{I}(\boldsymbol{a}_1,\boldsymbol{a}_4,\hat{x}_2),\ \mathcal{I}(\boldsymbol{a}_2,\boldsymbol{a}_3,\hat{x}_2),\ \hat{x}_1\big),\ \hat{\boldsymbol{x}}\in(0,1)^2\big\}.$$

In addition, we may check that a first property of the interpolation operator $\mathcal{I}$ is that, for any vector $\hat{\boldsymbol{x}}$ of $\mathbb{R}^2$:

$$\mathcal{I}\big(\mathcal{I}(\boldsymbol{a}_1,\boldsymbol{a}_4,\hat{x}_2),\ \mathcal{I}(\boldsymbol{a}_2,\boldsymbol{a}_3,\hat{x}_2),\ \hat{x}_1\big) = \mathcal{I}\big(\mathcal{I}(\boldsymbol{a}_1,\boldsymbol{a}_2,\hat{x}_1),\ \mathcal{I}(\boldsymbol{a}_4,\boldsymbol{a}_3,\hat{x}_1),\ \hat{x}_2\big).$$

Let $\boldsymbol{a}_{1,4}$ be defined by $\boldsymbol{a}_{1,4} = (\boldsymbol{a}_1 + \boldsymbol{a}_4)/2$. Then, it is easy to see that $\mathcal{I}$ also satisfies:

$$\big\{\mathcal{I}(\boldsymbol{a}_1, \boldsymbol{a}_4, s),\ s \in (0, 1/2)\big\} = \big\{\mathcal{I}(\boldsymbol{a}_1, \boldsymbol{a}_{1,4}, s),\ s \in (0, 1)\big\}.$$

We a similar definition for $\boldsymbol{a}_{1,3}$, we get for the first sub-cell of K, let us say K_1:

$$\begin{aligned} K_1 \quad &= \Big\{\mathcal{I}\Big(\mathcal{I}(\boldsymbol{a}_1, \boldsymbol{a}_4, \hat{x}_2),\ \mathcal{I}(\boldsymbol{a}_2, \boldsymbol{a}_3, \hat{x}_2),\ \hat{x}_1\Big),\ \hat{\boldsymbol{x}} \in (0, \frac{1}{2})^2\Big\} \\ &= \Big\{\mathcal{I}\Big(\mathcal{I}(\boldsymbol{a}_1, \boldsymbol{a}_{1,4}, \xi),\ \mathcal{I}(\boldsymbol{a}_2, \boldsymbol{a}_{2,3}, \xi),\ \hat{x}_1\Big),\ \hat{x}_1 \in (0, \frac{1}{2}),\ \xi \in (0, 1)\Big\}. \end{aligned}$$

We now permute the interpolations to obtain:

$$K_1 = \Big\{\mathcal{I}\Big(\mathcal{I}(\boldsymbol{a}_1, \boldsymbol{a}_2, \hat{x}_1),\ \mathcal{I}(\boldsymbol{a}_{1,4}, \boldsymbol{a}_{2,3}, \hat{x}_1),\ \xi\Big),\ \hat{x}_1 \in (0, \frac{1}{2}),\ \xi \in (0, 1)\Big\}.$$

Finally, with $\boldsymbol{a}_{1,2} = (\boldsymbol{a}_1 + \boldsymbol{a}_2)/2$ and $\boldsymbol{a}_{1,2,3,4} = (\boldsymbol{a}_1 + \boldsymbol{a}_2 + \boldsymbol{a}_3 + \boldsymbol{a}_4)/4$, we get:

$$K_1 = \Big\{\mathcal{I}\Big(\mathcal{I}(\boldsymbol{a}_1, \boldsymbol{a}_{1,2}, \eta),\ \mathcal{I}(\boldsymbol{a}_{1,4}, \boldsymbol{a}_{1,2,3,4}, \eta),\ \xi\Big),\ (\eta, \xi) \in (0, 1)^2\Big\}.$$

This means that K_1 may be obtained from the unit square by the Q_1 mapping associated with the vertices $\boldsymbol{a}_1$, $\boldsymbol{a}_{1,2}$, $\boldsymbol{a}_{1,2,3,4}$ and $\boldsymbol{a}_{1,4}$, which is the result we are searching for. The same conclusion may be obtained for the other sub-cells of K, by simple change of axes for the unit square. In three dimensions, the proof follows similar lines, but the interpolation operator $\mathcal{I}$ must be applied three times (along each of the coordinates) instead of twice. $\square$

We are now in position to define a locally refined mesh.

Definition 2.2 (Refined mesh) A locally refined mesh is obtained from an unrefined one by recursively splitting some cells by the above defined refinement process, in such a way that the number of hanging nodes per face is at most one (which means that the difference of level of refinement between two adjacent cells is at most one).

Examples of locally refined meshes are given on Figure 1

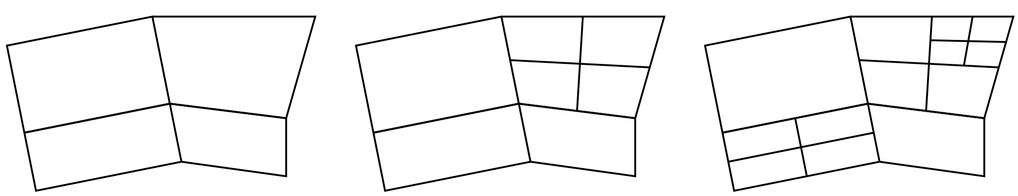


Figure 1: An exemple of admissible mesh refinement.

We denote by $\mathcal{E}(K)$ the set of the faces of an element $K \in \mathcal{M}$. We exclude the presence of a node in the interior of a face, *i.e.* we split an initial face in 2^{d-1} faces if one of the cells adjacent to the face is split. The number of faces, $N_K^{\mathcal{E}}$, of a cell K thus ranges between $2d$ and $2^d d$. Let $\mathcal{E} = \cup_{K \in \mathcal{M}} \mathcal{E}(K)$, $\mathcal{E}_{\rm ext} = \{\sigma \in \mathcal{E}, \sigma \subset \partial\Omega\}$ and $\mathcal{E}_{\rm int} = \mathcal{E} \setminus \mathcal{E}_{\rm ext}$. A face $\sigma \in \mathcal{E}_{\rm int}$ separating the cells K and L is denoted by $K|L$. Hereafter, $|\cdot|$ stands for the d- or $(d-1)$-dimensional measure of a subset of $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$ respectively.

For $\sigma \in \mathcal{E}(K)$, $\boldsymbol{n}_{K,\sigma}$ stands for a unit normal vector to σ outward K. In two dimensions and for plane faces in three dimensions, its definition is clear. For non-plane faces, many definitions are possible; here, we split the face in four 2-dimensional simplices the boundary of which joins the center of the face (with

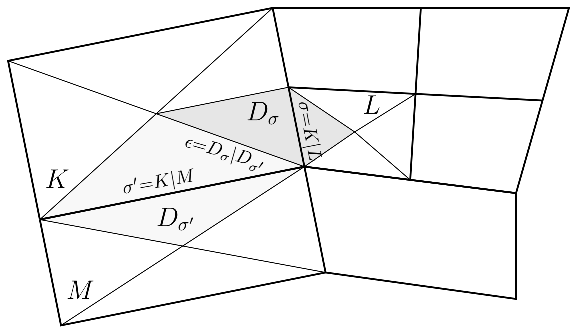


Figure 2: Notations for control volumes and diamond cells.

the same definition as before, namely the image by $\mathcal{Q}_K$ of the mass center of the associated face of the unit cube) and two vertices of the face, and use for $\boldsymbol{n}_{K,\sigma}$ the average of the unit normal vectors of these four simplices, weighted by their area.

We define a dual mesh associated with the faces $\mathcal{E}$ as follows. When $K \in \mathcal{M}$ is a rectangle or a cuboid, for $\sigma \in \mathcal{E}(K)$, we define the half-diamond cell $D_{K,\sigma}$ as the cone with basis σ and with vertex the mass center of K (see Fig. 2). We thus obtain a partition of K in $N_K^{\mathcal{E}}$ sub-volumes, each sub-volume having a measure $|D_{K,\sigma}|$ equal to $|K|/(2d)$, when σ has not been split, or $|K|/(2^d d)$ otherwise. We extend this definition to general quadrangles and hexahedra, by supposing that we have built a partition with the same connectivities and the same ratio between the volumes of the half-diamonds and of the cell. For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the dual (or diamond) cell D_σ associated with σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$. For $\sigma \in \mathcal{E}(K) \cap \mathcal{E}_{\rm ext}$, we define $D_\sigma = D_{K,\sigma}$. We denote by $\tilde{\mathcal{E}}(D_\sigma)$ the set of faces of D_σ, and by $\epsilon = D_\sigma | D_{\sigma'}$ the face separating two dual cells D_σ and $D_{\sigma'}$ (see Fig. 2).

The space discretization is staggered in the sense that the pressure and the velocity unknowns are discretized as piecewise constant functions respectively on the primal and dual mesh. In addition, in order for the algorithm to be further suitable for more general models including the density as unknown (*e.g.* the asymptotic model for anisothermal low Mach number flows, see Section 5.2), we suppose that ρ is approximated by a discrete function which is also piecewise constant on the primal mesh. Hence, the degrees of freedom for the pressure and the density are associated with the cells of the primal mesh: $\{p_K,\ K \in \mathcal{M}\}$ and $\{\rho_K,\ K \in \mathcal{M}\}$ while a discrete velocity field is associated with degrees of freedom localized at the cells of the dual mesh: $\{\boldsymbol{u}_\sigma,\ \sigma \in \mathcal{E}\}$.

3 The pressure correction scheme

3.1 General form of the scheme

Let us consider a uniform partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the time interval $(0,T)$, and let $\delta t = t_{n+1} - t_n$ for $n = 0, \ldots, N-1$ be the constant time step. The initial discrete pressure and velocity are defined as follows:

$$\begin{aligned} & p_K^0 = 0, && K \in \mathcal{M}, \\ & \boldsymbol{u}_\sigma^0 = \frac{1}{|\sigma|} \int_\sigma \boldsymbol{u}_0(\boldsymbol{x})\, {\rm d}\gamma(\boldsymbol{x}), && \sigma \in \mathcal{E}_{\rm int}. \end{aligned} \tag{3.1}$$

The Dirichlet boundary condition is taken into account by setting $\boldsymbol{u}_\sigma^n$ to the mean value of $\boldsymbol{u}_{\partial\Omega}$ over σ, for all $\sigma \in \mathcal{E}_{\rm ext}$ and all n in $\{0, 1, .., N\}$.

We also define discrete values of the density at the discrete time $n=-1$, which are needed by the algorithm, in order for this latter to satisfy a discrete kinetic energy balance for the scheme (see Section 3.4 below) :

$$\rho_K^{-1} = \frac{1}{|K|} \int_K \rho(\boldsymbol{x},0)\,\mathrm{d}\boldsymbol{x} \qquad K \in \mathcal{M}. \tag{3.2}$$

As usual [26, 6, 14], the pressure correction scheme is a two-step algorithm. Let us assume that $(\rho_K^{n-1})_{K\in\mathcal{M}} \subset \mathbb{R}$, $(\rho_K^n)_{K\in\mathcal{M}} \subset \mathbb{R}$, $(p_K^n)_{K\in\mathcal{M}} \subset \mathbb{R}$ and $(\boldsymbol{u}_\sigma^n)_{\sigma\in\mathcal{E}_{\rm int}} \subset \mathbb{R}^d$ are known families of real numbers. Then a time step computation consists in finding $(p_K^{n+1})_{K\in\mathcal{M}} \subset \mathbb{R}$ and $(\boldsymbol{u}_\sigma^{n+1})_{\sigma\in\mathcal{E}_{\rm int}} \subset \mathbb{R}^d$ through the following procedure:

Prediction step – Find $(\boldsymbol{u}_\sigma^{n+\frac12})_{\sigma\in\mathcal{E}_{\rm int}}$ such that:

$$\frac{1}{\delta t}(\rho_\sigma^n \boldsymbol{u}_\sigma^{n+\frac12} - \rho_\sigma^{n-1}\boldsymbol{u}_\sigma^n) + \frac{1}{|D_\sigma|} \sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}^n \boldsymbol{u}_\epsilon^{n+\frac12}$$
$$- (\mathrm{div}\boldsymbol{\tau}(\boldsymbol{u}))_\sigma^{n+\frac12} + \left(\frac{\rho_\sigma^n}{\rho_\sigma^{n-1}}\right)^{\frac12} (\boldsymbol{\nabla} p)_\sigma^n = 0, \quad \sigma \in \mathcal{E}_{\rm int}. \tag{3.3a}$$

Correction step – Find $(\boldsymbol{u}_\sigma^{n+1})_{\sigma\in\mathcal{E}_{\rm int}}$ and $(p_K^{n+1})_{K\in\mathcal{M}}$ such that:

$$\frac{1}{\delta t}\rho_\sigma^n\,(\boldsymbol{u}_\sigma^{n+1} - \boldsymbol{u}_\sigma^{n+\frac12}) + (\boldsymbol{\nabla} p)_\sigma^{n+1} - \left(\frac{\rho_\sigma^n}{\rho_\sigma^{n-1}}\right)^{\frac12} (\boldsymbol{\nabla} p)_\sigma^n = 0, \qquad \sigma \in \mathcal{E}_{\rm int}, \tag{3.3b}$$
$$\frac{1}{\delta t}(\rho_K^{n+1} - \rho_K^n) + \frac{1}{|K|}\sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}^{n+1} = 0, \qquad K \in \mathcal{M}. \tag{3.3c}$$

In the following sections, we give the definitions of the notations used in (3.3). In particular, Section 3.4 is devoted to the definition of the quantities ρ_σ and $F_{\sigma,\epsilon}$ which are built so as to ensure a discrete kinetic energy balance for the scheme.

3.2 Mass balance equation and pressure gradient

Equation (3.3c) is a discretization of the mass balance over the primal mesh, and $F_{K,\sigma}^{n+1}$ stands for the mass flux across σ outward K. On the primal mesh faces, $F_{K,\sigma}^{n+1}$ is given by:

$$F_{K,\sigma}^{n+1} = |\sigma|\ \widehat{\rho}_\sigma^{n+1}\ \boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}, \qquad \sigma \in \mathcal{E},$$

where $\widehat{\rho}_\sigma^{n+1}$ stands for an approximation of the density at the face σ. Note that, since the density is not an unknown and we do not need that the discrete mass balance equation ensures its positivity, any reasonable approximation may be used. Here, we choose a centered approximation at the internal faces of the domain (more precisely speaking, we compute $\widehat{\rho}_\sigma^{n+1}$ as the mean value of its approximation at the two neighbour cells of σ). For inflow external faces (*i.e.* a face where the flow is entering the domain), the quantity $\widehat{\rho}_\sigma^{n+1}$ is directly expressed from the data on ρ; for an outflow one, we take $\widehat{\rho}_\sigma^{n+1} = \rho_K^{n+1}$, where K is the cell adjacent to σ.

The pressure gradient is built as the dual operator of the discrete divergence, itself expressed from the already defined approximation of $\mathrm{div}(\rho\boldsymbol{u})$ by considering that the density is everywhere equal to 1. For $k=n,\, n+1$,

$$(\boldsymbol{\nabla} p)_\sigma^k = \frac{|\sigma|}{|D_\sigma|}(p_L^k - p_K^k)\,\boldsymbol{n}_{K,\sigma}, \qquad \sigma = K|L.$$

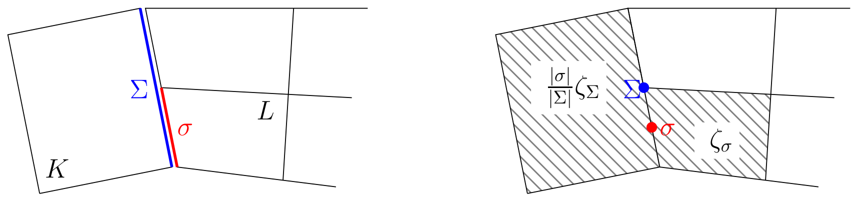


Figure 3: Piecewise definition of ζ_σ.

3.3 Discretization of the diffusion term

The discretization of the diffusion term relies on the so-called "rotated bi-linear element" introduced by Rannacher and Turek [23]. The reference element $\widehat{K}$ is the unit d-cube $(0,1)^d$, and the discrete functional space is:

$$\widetilde{Q}_1(\widehat{K}) = \operatorname{span}\left\{1,\ (\boldsymbol{x}_i)_{i=1,\ldots,d},\ (\boldsymbol{x}_i^2 - \boldsymbol{x}_{i+1}^2)_{i=1,\ldots,d-1}\right\}. \tag{3.4}$$

When no vertex of the face σ is a hanging node, we impose the jump of a discrete function through the face to have a zero mean value. When one vertex is a hanging node, we only impose to zero the integral of the jump through the initial coarse face containing σ. Hence, the set $\{\zeta_\sigma,\ \sigma \in \mathcal{E}\}$ of nodal functions associated with the Rannacher-Turek element is defined as follows.

Let $K \in \mathcal{M}$ and σ be a face of K. If σ is a whole side of K, the standard definition applies:

$$\begin{aligned}
&(i) && \zeta_\sigma|_K = \zeta_{\widehat{\sigma}} \circ \mathcal{Q}_K^{-1} \text{ where } \zeta_{\widehat{\sigma}} \text{ is some function of } \widetilde{Q}_1(\widehat{K}),\\
&(ii) && \frac{1}{|\sigma|}\int_\sigma \zeta_\sigma = 1 \text{ and, for all other sides } \Sigma' \text{ of } K,\ \int_{\Sigma'} \zeta_\sigma = 0.
\end{aligned} \tag{3.5}$$

Let us now suppose that σ is only a subset of a side of K, which we denote by Σ (which occurs when σ separates K form a cell which refinement level is equal to the refinement level of K plus 1, see Figure 3). Let ζ_Σ be the Rannacher-Turek usual shape function associated with Σ (*i.e.* the function satisfying the above definition (3.5), replacing σ by Σ). Then, we define ζ_σ on K by:

$$\zeta_\sigma(\boldsymbol{x}) = \frac{|\sigma|}{|\Sigma|}\,\zeta_\Sigma(\boldsymbol{x}).$$

Finally, of course, the nodal functions are local, in the sense that, for $\sigma \in \mathcal{E}$, the support of ζ_σ is reduced to the (one or two) cells adjacent to σ.

Finally, the discretization of the i^{th} component of the diffusion term reads:

$$-(\operatorname{div}\boldsymbol{\tau}(\boldsymbol{u}))_{\sigma,i}^{n+\frac{1}{2}} = \frac{1}{|D_\sigma|}\sum_{K\in\mathcal{M}}\int_K \sum_{\sigma'\in\mathcal{E}(K)}\ \sum_{j=1}^{d} \boldsymbol{u}_{\sigma',j}^{n+\frac{1}{2}}\ \boldsymbol{\tau}(\zeta_{\sigma'}\boldsymbol{e}^{(j)}) : \boldsymbol{\nabla}(\zeta_\sigma \boldsymbol{e}^{(i)})\,\mathrm{d}\boldsymbol{x},$$

where, for $1 \le j \le d$, $\boldsymbol{e}^{(j)}$ stands for the j^{th} vector of the canonical basis of $\mathbb{R}^d$.

3.4 The discrete momentum convection operator

In this section, we describe the approximation of the convection operator $\partial_t(\rho\boldsymbol{u}) + \operatorname{div}(\rho\boldsymbol{u}\otimes\boldsymbol{u})$ which appears in the momentum balance equation. It is of finite volume type, and takes the general form given by the first two terms of (3.3a).

The quantities ρ_σ^n and ρ_σ^{n-1} are approximations of the density on the dual cell D_σ at time t_n and t_{n-1} respectively, while the quantities $F_{\sigma,\epsilon}^n$ are mass fluxes across the edges of the dual cells. These quantities are built so that a finite volume discretization of the mass balance (1.1b) holds over the internal dual cells:

$$\frac{1}{\delta t}\ (\rho_\sigma^n - \rho_\sigma^{n-1}) + \frac{1}{|D_\sigma|} \sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}^n = 0, \qquad \sigma \in \mathcal{E}_{\rm int}. \tag{3.6}$$

This is crucial in order to reproduce, at the discrete level, the derivation of a kinetic energy balance equation, a consequence of which are discrete analogues of the usual $\mathrm{L}^\infty(\mathrm{L}^2)$- and $\mathrm{L}^2(\mathrm{H}^1)$- stability estimates for the velocity, if the density is assumed to be bounded by below by a positive constant. Indeed, the following result holds:

Proposition 3.1 (Discrete kinetic energy balance) Assume that the discrete densities ρ_σ^n, ρ_σ^{n-1} and the dual fluxes $F_{\sigma,\epsilon}^n$ are built so as to satisfy equation (3.6) for all $\sigma \in \mathcal{E}_{\rm int}$ and all n in $\{0, 1, .., N-1\}$. Then, any solution to the scheme (3.3) satisfies the following equality, for all $\sigma \in \mathcal{E}_{\rm int}$ and n in $\{0, 1, .., N\}$:

$$\begin{aligned}\frac{1}{2\delta t}\Big(\rho_\sigma^n |\boldsymbol{u}_\sigma^{n+1}|^2 - \rho_\sigma^{n-1}|\boldsymbol{u}_\sigma^n|^2\Big) + \frac{1}{2|D_\sigma|} \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)\\ \epsilon=D_\sigma|D_{\sigma'}}} F_{\sigma,\epsilon}^n\ \boldsymbol{u}_\sigma^{n+\frac12}\cdot \boldsymbol{u}_{\sigma'}^{n+\frac12}\\ -(\mathrm{div}\boldsymbol{\tau}(\boldsymbol{u}))_\sigma^{n+\frac12}\cdot\boldsymbol{u}_\sigma^{n+\frac12} + (\boldsymbol{\nabla} p)_\sigma^{n+1}\cdot \boldsymbol{u}_\sigma^{n+1} = -R_\sigma^{n+1} - P_\sigma^{n+1},\end{aligned} \tag{3.7}$$

where

$$R_\sigma^{n+1} = \frac{1}{2\delta t}\ \rho_\sigma^{n-1} |\boldsymbol{u}_\sigma^{n+\frac12} - \boldsymbol{u}_\sigma^n|^2, \qquad P_\sigma^{n+1} = \delta t\ \Big(\frac{1}{\rho_\sigma^n}|(\boldsymbol{\nabla} p)_\sigma^{n+1}|^2 - \frac{1}{\rho_\sigma^{n-1}}|(\boldsymbol{\nabla} p)^n|^2\Big).$$

Multiplying equation (3.7) by δt and summing over $\sigma \in \mathcal{E}_{\rm int}$ and $n = 0, .., M-1$, yields, for all M in $\{0, 1, .., N\}$:

$$\begin{aligned}\frac12 \sum_{\sigma\in\mathcal{E}_{\rm int}} |D_\sigma| \rho_\sigma^{M-1} |\boldsymbol{u}_\sigma^M|^2 + \sum_{n=0}^{M-1} \delta t \sum_{K\in\mathcal{M}} \int_K \boldsymbol{\tau}(\tilde{\boldsymbol{u}}^{n+\frac12}) : \boldsymbol{\nabla}\tilde{\boldsymbol{u}}^{n+\frac12} + \mathcal{R}^M + \mathcal{P}^M\\ = \frac12 \sum_{\sigma\in\mathcal{E}_{\rm int}} |D_\sigma| \rho_\sigma^{-1} |\boldsymbol{u}_\sigma^0|^2,\end{aligned} \tag{3.8}$$

where for all $n \in \mathbb{N}$, $\tilde{\boldsymbol{u}}^{n+\frac12} = \sum_{\sigma\in\mathcal{E}_{\rm int}} \boldsymbol{u}_\sigma^{n+\frac12}\zeta_\sigma$, and

$$\mathcal{R}^M = \sum_{n=0}^{M-1}\sum_{\sigma\in\mathcal{E}_{\rm int}} \rho_\sigma^{n-1}|\boldsymbol{u}_\sigma^{n+\frac12} - \boldsymbol{u}_\sigma^n|^2 \geq 0, \qquad \mathcal{P}^M = \delta t^2 \sum_{\sigma\in\mathcal{E}_{\rm int}} \frac{1}{\rho_\sigma^M}|(\boldsymbol{\nabla} p)_\sigma^M|^2 \geq 0.$$

Proof. We refer to [16, 13] where the proof of this result can be found in the context of a similar pressure correction scheme for Euler and Navier-Stokes equations. □

Let us now give the detailed construction of the dual densities and mass fluxes that ensures (3.6). For $\sigma \in \mathcal{E}_{\rm int}$ such that $\sigma = K|L$, the approximate densities on the dual cell D_σ are given by the following weighted average:

$$|D_\sigma|\ \rho_\sigma^k = \xi_K^\sigma |K|\ \rho_K^k + \xi_L^\sigma |L|\ \rho_L^k, \qquad \text{for } k = n-1 \text{ and } k = n, \tag{3.9}$$

where

$$\xi_K^\sigma = \frac{|D_{K,\sigma}|}{|K|}, \qquad K \in \mathcal{M},\ \sigma \in \mathcal{E}(K). \tag{3.10}$$

The set of dual fluxes $F^n_{\sigma,\epsilon}$ with ϵ included in the primal cell K, is computed by solving a linear system which right-hand side is a linear combination of the primal fluxes $(F^n_{K,\sigma})_{\sigma\in\mathcal{E}(K)}$, appearing in the discrete mass balance (1.1b) at the previous time step. More precisely, we have the following definition for the dual fluxes, in which we omit for short the time dependence (*i.e.* the superscript n).

Definition 1 (Definition of the dual fluxes from the primal ones) The fluxes through the faces of the dual mesh are defined so as to satisfy the following three constraints:

- *(H1)* – For all primal cell K in $\mathcal{M}$, the set $(F_{\sigma,\epsilon})_{\epsilon\subset K}$ of dual fluxes through faces included in K satisfies the following linear system:

$$F_{K,\sigma} + \sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\ \epsilon\subset K} F_{\sigma,\epsilon} = \xi_K^\sigma \sum_{\sigma'\in\mathcal{E}(K)} F_{K,\sigma'}, \qquad \forall\sigma\in\mathcal{E}(K). \tag{3.11}$$

- *(H2)* – The dual fluxes are conservative: $F_{\sigma,\epsilon} = -F_{\sigma',\epsilon}$ for all $\epsilon = D_\sigma|D'_\sigma$.
- *(H3)* – The dual fluxes are bounded with respect to the primal ones $(F_{K,\sigma})_{\sigma\in\mathcal{E}(K)}$:

$$|F_{\sigma,\epsilon}| \leq\ C \max\left\{|F_{K,\sigma}|,\ \sigma\in\mathcal{E}(K)\right\}, \qquad K\in\mathcal{M},\ \sigma\in\mathcal{E}(K),\ \epsilon\in\tilde{\mathcal{E}}(D_\sigma),\ \epsilon\subset K.$$

Appendix A provides the detailed construction of the dual fluxes (*i.e.* the way to find a solution to (3.11) satisfying the other two constraints). Owing to these definitions of the dual densities and mass fluxes, the mass balance on the dual cells (3.6) is an easy consequence of the mass balance on the cells of the primal mesh [1]:

$$\frac{1}{\delta t}\ (\rho_K^n - \rho_K^{n-1}) + \frac{1}{|K|} \sum_{\sigma\in\mathcal{E}(K)} F^n_{K,\sigma} = 0, \qquad K\in\mathcal{M}.$$

Since the mass balance is not yet solved when performing the prediction step, the mass fluxes $F^{n+1}_{K,\sigma}$ are not yet known. This is what leads us to perform a backward shift of the density in the prediction step, thus using ρ^n and ρ^{n-1} rather than the apparently more natural choice ρ^{n+1} and ρ^n.

To make the description of the scheme complete, it remains to define the velocity interpolates $\boldsymbol{u}_\epsilon^{n+\frac{1}{2}}$ at the internal dual faces. We chose a centered approximation:

$$\boldsymbol{u}_\epsilon^{n+\frac{1}{2}} = \frac{\boldsymbol{u}_\sigma^{n+\frac{1}{2}} + \boldsymbol{u}_{\sigma'}^{n+\frac{1}{2}}}{2}, \qquad \text{for } \epsilon = D_\sigma|D'_\sigma.$$

4 Study of the prediction step: Error analysis for the advection-diffusion equation

For the sake of simplicity, we propose to perform an error analysis of the scheme on the simplified model of the advection-diffusion equation. This scalar equation can be seen as a model problem for the prediction step of the method (for one component of the velocity), where the known density is assumed to be constant and the pressure gradient and the advective velocity are known given functions from the previous time step. Since the density is considered constant, for the kinetic energy identity to hold, we need to suppose that the given advection field $\boldsymbol{w}$ is divergence free; we assume in addition that $\boldsymbol{w}$ is a regular function, namely $\boldsymbol{w}\in \mathrm{C}^1(\bar{\Omega})^d$, and vanishes at the boundary $\partial\Omega$ of the computational domain. The problem we consider thus consists in finding a function u such that:

$$u + \mathrm{div}(u\boldsymbol{w}) - \boldsymbol{\Delta} u = f, \qquad \text{on } \Omega, \tag{4.1a}$$

$$u = 0, \qquad \text{on } \partial\Omega, \tag{4.1b}$$

where $f \in \mathrm{L}^2(\Omega)$. In order to write a weak formulation of (4.1), let us define the following forms corresponding respectively to the diffusion and convection terms:

$$a(u,v) = \int_\Omega \boldsymbol{\nabla} u \cdot \boldsymbol{\nabla} v, \qquad u,v \in \mathrm{H}^1_0(\Omega),$$
$$b(\boldsymbol{w},u,v) = \int_\Omega \mathrm{div}(u\boldsymbol{w})v, \quad u,v \in \mathrm{H}^1_0(\Omega).$$

We also introduce the inner product of $\mathrm{L}^2(\Omega)$ denoted $(.,.)$, and we are now in position to state the weak formulation of problem (4.1).

Definition 2 A weak solution of the advection-diffusion problem (4.1) is a function $u \in \mathrm{H}^1_0(\Omega)$ such that:

$$(u,v) + b(\boldsymbol{w},u,v) + a(u,v) = (f,v), \qquad \forall v \in \mathrm{H}^1_0(\Omega). \tag{4.2}$$

It is well known that, by the Lax-Milgram theorem, there exists a unique solution to (4.2).

4.1 Regularity of the mesh and approximation space

In addition to the definition of the mesh given in Section 2, we suppose that any quadrilaterals ($d = 2$) or hexahedra ($d = 3$) of the mesh are convex, which implies that their faces are hyperplanes of $\mathbb{R}^d$. For $K \in \mathcal{M}$, we denote by h_K the diameter of K. Similarly, we denote by h_σ the diameter of an edge $\sigma \in \mathcal{E}$. The size of the discretization is defined as usual by:

$$h = \max\{h_K, K \in \mathcal{M}\}.$$

For the consistency of the Rannacher-Turek finite element approximation of the diffusion term, we need a measure of the difference between the cells of $\mathcal{M}$ and parallelograms ($d = 2$) or parallelotopes ($d = 3$), as defined in [23]. For $K \in \mathcal{M}$, we denote by $\bar{\alpha}_K$ the maximum of the angles between the normal vectors of opposite faces, choosing for the latter the orientation which maximizes the angle, and set $\alpha_K = \pi - \bar{\alpha}_K$ (so $\alpha_K = 0$ if K is a parallelogram or a parallelotope, and $\alpha_K > 0$ otherwise). Then we define α_h as:

$$\alpha_h = \max\{\alpha_K,\ K \in \mathcal{M}\}. \tag{4.3}$$

For $K \in \mathcal{M}$, let $\{\boldsymbol{a}_{i,K},\ i = 1,..,2^d\}$ denote the vertices of K. Let $\boldsymbol{a}_{i,K}$ be one of these vertices and let $S_{i,K}$ be the simplex whose vertices are $\boldsymbol{a}_{i,K}$ and the d adjacent vertices to $\boldsymbol{a}_{i,K}$. We denote by $r_{i,K}$ the diameter of the largest ball included in $S_{i,K}$ and by r_K the real number given by $r_K = \min\{r_{i,K},\ i = 1,..,2^d\}$. We define the real number θ_h by:

$$\theta_h = \max\left\{\frac{h_K}{r_K},\ K \in \mathcal{M}\right\}. \tag{4.4}$$

In accordance with the velocity discretization described in Section 3, a discrete function v is associated with degrees of freedom $\{v_\sigma,\ \sigma \in \mathcal{E}_{\mathrm{int}}\}$ located at the internal faces, the values associated with the boundary faces $\mathcal{E} \in \mathcal{E}_{\mathrm{ext}}$ being set to zero, consistently with the boundary conditions (4.1b). For $K \in \mathcal{M}$, the restriction of a discrete function v to K belongs to the local Rannacher-Turek space: $v|_K = \widehat{v} \circ \mathcal{Q}_K^{-1}$ where $\widehat{v} \in \widetilde{Q}_1(\widehat{K})$ (see (3.4)). The last step to obtain a definition of the discrete approximation space is to state the continuity constraints satisfied by the discrete functions at the faces. Once again, when no vertex of the face is a hanging node, this issue is clear: as usual for the Rannacher-Turek element, we impose the jump through the face to have a zero mean value. Otherwise, we only impose to zero the integral of the jump through the initial coarse face. Hence the approximation space is given by

$$V_h = \big\{v(\boldsymbol{x}) = \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} v_\sigma\ \zeta_\sigma(\boldsymbol{x}), \quad (v_\sigma)_{\sigma \in \mathcal{E}_{\mathrm{int}}} \subset \mathbb{R}\big\}, \tag{4.5}$$

where the shape functions ζ_σ are defined in Section 3.1.

With the continuity requirements described above, the functions of V_h are discontinuous across each edge; the discretization is thus non-conforming in the sense that $V_h \not\subset \mathrm{H}^1_0(\Omega)$. Therefore we define the "broken" gradient $\boldsymbol{\nabla}_h v$ on V_h as the function of $\mathrm{L}^2(\Omega)^d$ which is equal to $\boldsymbol{\nabla} v|_K$ for all $K \in \mathcal{M}$. The corresponding broken Sobolev H^1 semi-norm is defined for $v \in V_h$ by:

$$\|v\|^2_{h,\mathrm{b}} = \int_\Omega |\boldsymbol{\nabla}_h v|^2 = \sum_{K\in\mathcal{M}} \int_K |\boldsymbol{\nabla} v|^2.$$

Thanks to the homogeneous Dirichlet boundary conditions, this defines a norm on V_h which is known to control the L^2-norm by a Poincaré inequality:

$$\|v\|_{\mathrm{L}^2(\Omega)} \leq \mathrm{diam}(\Omega)\ \|v\|_{h,\mathrm{b}}, \qquad \forall v \in V_h. \tag{4.6}$$

4.2 The staggered scheme for the advection-diffusion equation

Let us now adapt to the advection-diffusion problem the first step of the staggered scheme proposed in Section 3. For this purpose, we assume that the divergence of the velocity field $\boldsymbol{w}$ is discretized on the cells of the primal mesh by

$$\mathrm{div}(\boldsymbol{w})|_K \approx \frac{1}{|K|} \sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}(\boldsymbol{w}), \quad \text{with } \ F_{K,\sigma}(\boldsymbol{w}) = \int_\sigma \boldsymbol{w}\cdot n_{K,\sigma}\,\mathrm{d}\gamma, \tag{4.7}$$

and we define the dual fluxes $F_{\sigma,\epsilon}(\boldsymbol{w})$ from the primal ones following exactly the procedure described in Section 3.4. Adapting the staggered scheme to the advection-diffusion problem thus consists in finding a discrete function $u_h \in V_h$ such that:

$$(u_h, v)_h + b_h(\boldsymbol{w}, u_h, v) + a_h(u_h, v) = (f, v), \qquad \text{for all } v \in V_h, \tag{4.8}$$

where, for any $u,\ v \in V_h$,

$$(u, v)_h = \sum_{\sigma\in\mathcal{E}} |D_\sigma| u_\sigma v_\sigma, \tag{4.9}$$

$$b_h(\boldsymbol{w}, u, v) = \sum_{\sigma\in\mathcal{E}} v_\sigma \sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}(\boldsymbol{w}) u_\epsilon, \tag{4.10}$$

$$a_h(u, v) = \int_\Omega \boldsymbol{\nabla}_h u \cdot \boldsymbol{\nabla}_h v. \tag{4.11}$$

4.3 Stability properties of the scheme

One first important property of the scheme follows from the particular construction of the discrete convection term $b_h(\boldsymbol{w}, u, v)$. Indeed, since $\boldsymbol{w}$ is divergence-free, one has

$$\sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}(\boldsymbol{w}) = 0,$$

which, by construction, implies a similar discrete divergence-free property on the cells of the dual mesh:

$$\sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}(\boldsymbol{w}) = 0, \qquad \forall \sigma \in \mathcal{E}_{\mathrm{int}}. \tag{4.12}$$

This is crucial in order to reproduce, at the discrete level, the derivation of a stability estimate for the scheme, analogous to the kinetic energy balance equation in the context of the full Navier-Stokes equations. Indeed, by a computation which may be found in [1], Equation (4.12) yields:

$$b_h(\boldsymbol{w}, v, v) = 0, \qquad \forall v \in V_h. \tag{4.13}$$

Remark 4.1 At first glance, this relation is the usual well known antisymmetry of the centered convection operator with a divergence free advection field; note however that the scheme is posed on a dual mesh, which, if we restrict the set of dual cells to those associated with the internal faces, does not cover Ω, and that no balance equation is written on the remainder of the domain (*i.e.* the half-diamond-shaped volumes associated with the external faces). This necessitates a slight adaptation of the usual proof, which is performed in [1].

A first consequence of this antisymmetry property is the existence of a unique solution to the scheme (4.8). Let $\mathcal{A}_h$ be the following form defined on the finite dimensional space $V_h \times V_h$ by

$$\mathcal{A}_h(u, v) = (u, v)_h + b_h(\boldsymbol{w}, u, v) + a_h(u, v), \qquad u, v \in V_h.$$

The mapping Φ from V_h to V_h' which maps an element u to the form $\Phi(u) := \mathcal{A}_h(u, .)$ is clearly linear and satisfies ($\Phi(u) = 0 \Rightarrow u = 0$) since $\mathcal{A}_h(u, u) = (u, u)_h + \|u\|^2_{h,\mathrm{b}}$ for all $u \in V_h$ by (4.13). Hence, Φ is a one-to-one linear mapping which, in a finite dimensional context, is equivalent to the existence of a unique solution $u_h \in V_h$ which satisfies $\Phi(u_h) = f$ *i.e.* $\mathcal{A}_h(u, v) = (f, v)$ for all $v \in V_h$.

A second straightforward consequence is a stability estimate on the solution u_h. Taking u_h as a test function in (4.8) and using again (4.13), one gets $(u_h, u_h)_h + \|u_h\|^2_{h,\mathrm{b}} = (f, u_h)$. Applying Young's inequality in the right hand side term and recalling the Poincaré inequality (4.6), we obtain the stability estimate

$$2\,(u_h, u_h)_h + \|u_h\|^2_{h,\mathrm{b}} \leq \mathrm{diam}(\Omega)^2\ \|f\|^2_{\mathrm{L}^2(\Omega)}. \tag{4.14}$$

4.4 Error estimate

We may now state the main result of this section, which is an error estimate for the scheme (4.8).

Theorem 4.1 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω (in the sense of Definition 2.2) such that $\theta_h \leq \theta_0$, with θ_h defined by (4.4). Let $u_h \in V_h$ be the solution to the scheme (4.8). We assume that the solution u of the continuous problem (4.2) belongs to $\mathrm{H}^1_0(\Omega) \cap \mathrm{H}^2(\Omega)$. Then u_h satisfies:

$$\|u_h - u\|_{h,\mathrm{b}} \leq C\,(h + \alpha_h)\ \|u\|_{\mathrm{H}^2(\Omega)},$$

where C only depends on $\boldsymbol{w}$, Ω and θ_0.

Remark 4.2 (A particular construction of a regular sequence of discretizations) For $d = 2$, a sequence of discretizations satisfying $h \to 0$ and $\alpha_h \to 0$ is obtained by successively dividing each quadrangle in four sub-quadrangles, splitting it along the lines joining the mid-points of opposite faces. Unfortunately, the extension of this construction to the three-dimensional case is not straightforward, since this subdivision process may generate non-plane faces.

Remark 4.3 Theorem 4.1 shows that the accuracy of the scheme for the energy norm $\|.\|_{h,\mathrm{b}}$ is the same as that of the usual Rannacher-Turek approximation of the Stokes problem on non-refined meshes [23]. In particular, there is no loss in the convergence rate due to the non-conforming local refinement or the particular discretization of the convection term.

4.4.1 Preliminary lemmas

We begin with stating some technical lemmas, which will be useful in the proof of Theorem 4.1. We first introduce the following discrete H^1-norm on the space V_h:

$$\|v\|_{h,\mathrm{fv}}^2 = \sum_{K\in\mathcal{M}} h_K^{d-2} \sum_{\sigma,\sigma'\in\mathcal{E}(K)} |v_\sigma - v_{\sigma'}|^2, \tag{4.15}$$

which, by an easy computation, may be shown to be equivalent, over a regular sequence of discretizations such that $\max_h \theta_h \leq \theta_0$ for some $\theta_0 > 0$, to the usual finite volume H^1-norm. Lemma 4.1 shows that this H^1-norm is controlled by the broken-Sobolev H^1-norm.

LEMMA 4.1 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \leq \theta_0$, with θ_h defined by (4.4). Then, there exists C only depending on θ_0 such that:

$$\|v\|_{h,\mathrm{fv}} \leq C\ \|v\|_{h,\mathrm{b}}, \quad \forall v \in V_h.$$

Proof. For $K \in \mathcal{M}$, let $\widehat{v}$ be the function defined over $\widehat{K}$ by $\widehat{v}(\widehat{\boldsymbol{x}}) = v(\boldsymbol{x})$, where $\boldsymbol{x} \in K$ stands for the image of $\widehat{\boldsymbol{x}}$ by the Q_1 mapping $\mathcal{Q}_K$. By definition of the discretization space, we have $\widehat{v} \in \widetilde{Q}_1(\widehat{K})$. Now, since there exists only a bounded number of possible configurations for K (depending on the fact that its sides are split in 2^{d-1} faces or not), a finite dimensional argument for norms acting on $\widetilde{Q}_1(\widehat{K})$ shows that there exists a constant C such that:

$$\sum_{\sigma,\sigma'\in\mathcal{E}(K)} |v_\sigma - v_{\sigma'}|^2 \leq C \int_{\widehat{K}} |\boldsymbol{\nabla}\widehat{v}|^2.$$

We conclude the proof by invoking standard properties of the Q_1 mapping which enable to write:

$$\int_{\widehat{K}} |\boldsymbol{\nabla}\widehat{v}|^2 \leq C(\theta_0)\frac{h_K^2}{|K|}\int_K |\boldsymbol{\nabla} v|^2 \leq C'(\theta_0) h_K^{2-d} \int_K |\boldsymbol{\nabla} v|^2.$$

□

The following lemma compares piecewise constant approximates of a discrete function with the function itself. It is an easy consequence of the previous one.

LEMMA 4.2 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \leq \theta_0$, with θ_h defined by (4.4). For $v \in V_h$, let v_c stand for the piecewise constant function over each diamond cell D_σ and equal to v_σ. Then, there exists C, only depending on θ_0, such that:

$$\|v - v_\mathrm{c}\|_{\mathrm{L}^2(\Omega)} \leq C\, h\ \|v\|_{h,\mathrm{b}}.$$

Let now v_m be defined as a piecewise constant function over the primal mesh, the value over a cell K being a convex combination of $(v_\sigma)_{\sigma\in\mathcal{E}(K)}$. Then, once again, there exists C, only depending on θ_0, such that:

$$\|v - v_\mathrm{m}\|_{\mathrm{L}^2(\Omega)} \leq C\, h\ \|v\|_{h,\mathrm{b}}.$$

The following lemma is the analogue, for the Rannacher-Turek element and a locally refined mesh, of a well known result for the Crouzeix-Raviart element and regular meshes.

Lemma 4.3 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \leq \theta_0$, with θ_h defined by (4.4). Let $v \in V_h$. For $\sigma \in \mathcal{E}_{\rm int}$, let $[v]_\sigma$ denote the jump of v across σ (the orientation of which we do not need to precise), and, for $\sigma \in \mathcal{E}_{\rm ext}$, let $[v]_\sigma = v$. Then the following inequality holds:

$$\Big(\sum_{\sigma\in\mathcal{E}} \frac{1}{h_\sigma} \int_\sigma [v]_\sigma^2\Big)^{1/2} \leq C\ \|v\|_{h,\rm b}.$$

where C only depends on θ_0.

Proof. Let σ be a face of the mesh. By assumption, the integral of the jump of a discrete function v vanishes either through σ or through a coarse face Σ including σ. In any case, there exists $\boldsymbol{x}_\sigma \in \Sigma$ such that $[v]_\sigma(\boldsymbol{x}_\sigma) = 0$ and the distance between $\boldsymbol{x}_\sigma$ and any point of σ is lower than $C(\theta_0)\, h_\sigma$. We thus get:

$$\sum_{\sigma\in\mathcal{E}} \frac{1}{h_\sigma} \int_\sigma [v]_\sigma^2 = \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \frac{1}{h_\sigma} \int_\sigma \Big(\int_{\boldsymbol{x}_\sigma}^{\boldsymbol{x}} (\boldsymbol{\nabla} v|_K - \boldsymbol{\nabla} v|_L) \cdot \frac{\boldsymbol{x}-\boldsymbol{x}_\sigma}{|\boldsymbol{x}-\boldsymbol{x}_\sigma|}\Big)^2$$
$$+ \sum_{\sigma\in\mathcal{E}_{\rm ext}\cap\mathcal{E}(K)} \frac{1}{h_\sigma} \int_\sigma \Big(\int_{\boldsymbol{x}_\sigma}^{\boldsymbol{x}} \boldsymbol{\nabla} v|_K \cdot \frac{\boldsymbol{x}-\boldsymbol{x}_\sigma}{|\boldsymbol{x}-\boldsymbol{x}_\sigma|}\Big)^2.$$

The Cauchy-Schwarz inequality yields:

$$\sum_{\sigma\in\mathcal{E}} \frac{1}{h_\sigma} \int_\sigma [v]_\sigma^2 \leq \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \frac{1}{h_\sigma} \int_\sigma |\boldsymbol{x}-\boldsymbol{x}_\sigma| \int_{\boldsymbol{x}_\sigma}^{\boldsymbol{x}} |\boldsymbol{\nabla} v|_K|^2 + |\boldsymbol{\nabla} v|_L|$$
$$+ \sum_{\sigma\in\mathcal{E}_{\rm ext}\cap\mathcal{E}(K)} \frac{1}{h_\sigma} \int_\sigma |\boldsymbol{x}-\boldsymbol{x}_\sigma| \int_{\boldsymbol{x}_\sigma}^{\boldsymbol{x}} |\boldsymbol{\nabla} v|_K|^2.$$

Bounding $|\boldsymbol{x}-\boldsymbol{x}_\sigma|$ by $C(\theta_0)\, h_\sigma$ and switching the order of integration, we get:

$$\sum_{\sigma\in\mathcal{E}} \frac{1}{h_\sigma} \int_\sigma [v]_\sigma^2 \leq C(\theta_0) \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} \int_{\boldsymbol{x}_\sigma}^{\boldsymbol{x}} \int_\sigma |\boldsymbol{\nabla} v|_K|^2 + |\boldsymbol{\nabla} v|_L| + C(\theta_0) \sum_{\sigma\in\mathcal{E}_{\rm ext}\cap\mathcal{E}(K)} \int_{\boldsymbol{x}_\sigma}^{\boldsymbol{x}} \int_\sigma |\boldsymbol{\nabla} v|_K|^2,$$

and so, finally:

$$\sum_{\sigma\in\mathcal{E}} \frac{1}{h_\sigma} \int_\sigma [v]_\sigma^2 \leq C(\theta_0)^2 \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} h_\sigma \int_\sigma |\boldsymbol{\nabla} v|_K|^2.$$

We conclude the proof as for Lemma 4.1 by transporting the quantity $\int_\sigma |\boldsymbol{\nabla} v|_K|^2$ to the reference element $\widehat{K}$, invoking finite dimensional arguments to relate the integral of the squared gradient norm over the faces and over the element and, finally, standard properties of the Q_1 mapping. $\square$

The proof of the following trace lemma is an easy adaptation of a result which can be found in [8, appendix A].

Lemma 4.4 Let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω and K be a control volume of $\mathcal{M}$, and let σ be one of its faces. Then there exists C, only depending on d, such that the following inequality holds:

$$\|v\|_{\mathrm{L}^2(\sigma)} \leq C \frac{1}{r_K^{1/2}}\ \Big(\|v\|_{\mathrm{L}^2(K)} + h_K \|\nabla v\|_{\mathrm{L}^2(K)^d}\Big), \qquad \forall v \in \mathrm{H}^1(K).$$

We will also need the following Poincaré-Wirtinger inequality, which is proven for any convex domain K in [21, 2].

LEMMA 4.5 For all convex domain K of $\mathbb{R}^d$, $1 \le d \le 3$:

$$\|v - m_K(v)\|_{\mathrm{L}^2(K)} \le \frac{1}{\pi}\ h_K \|\nabla v\|_{\mathrm{L}^2(K)^d}, \qquad \forall v \in \mathrm{H}^1(K), \tag{4.16}$$

where $m_K(v)$ stands for the mean value of v over K.

We now give two corollaries of Lemmas 4.4 and 4.5; they both compare different mean values of functions of $\mathrm{H}^1(\Omega)$.

LEMMA 4.6 Let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω; let $K \in \mathcal{M}$ and let D be a subset of K. For all $v \in \mathrm{H}^1(K)$, we denote $m_K(v)$ and $m_D(v)$ the mean values of v on K and D respectively. Then:

$$|m_K(v) - m_D(v)| \le \frac{1}{\pi\,|D|^{\frac{1}{2}}}\ h_K\ \|\boldsymbol{\nabla} v\|_{\mathrm{L}^2(K)^d}, \qquad \forall v \in \mathrm{H}^1(K).$$

Proof. By the Cauchy-Schwarz inequality, we have:

$$|m_K(v) - m_D(v)| = \frac{1}{|D|}\ \Big|\int_D \big(v - m_K(v)\big)\Big| \le \frac{1}{|D|^{\frac{1}{2}}}\ \|v - m_K(v)\|_{\mathrm{L}^2(D)}.$$

Thus, since D is a subset of K, we have:

$$|m_K(v) - m_D(v)| \le \frac{1}{|D|^{\frac{1}{2}}}\ \|v - m_K(v)\|_{\mathrm{L}^2(K)},$$

and we conclude by (4.16), since K is convex. □

LEMMA 4.7 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \le \theta_0$, with θ_h defined by (4.4). Let $K \in \mathcal{M}$ be a control volume and let σ be a face of K. For all $v \in \mathrm{H}^1(K)$, we denote $m_\sigma(v)$ and $m_K(v)$ the mean values of v on σ and K respectively. Then, there exists C, only depending on d and θ_0 such that:

$$|m_\sigma(v) - m_K(v)| \le \frac{C}{r_K^{d/2}}\ h_K \|\nabla v\|_{\mathrm{L}^2(K)^d}, \qquad \forall v \in \mathrm{H}^1(K). \tag{4.17}$$

Proof. We have $|m_\sigma(v) - m_K(v)| \le |\sigma|^{-1} \int_\sigma |v - m_K(v)| \le |\sigma|^{-1/2} \|v - m_K(v)\|_{\mathrm{L}^2(\sigma)}$ by the Cauchy-Schwarz inequality. Invoking successively Lemma 4.4 and Lemma 4.5 and using the regularity of the mesh yields the result. □

We are now in position to prove the following technical lemma.

LEMMA 4.8 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \ge \theta_0$, where θ_h is defined by (4.4). We define $\mathcal{E}_\mathrm{c}$ as the set containing:

- the internal faces of the mesh which do not have as vertex a hanging node,
- the coarse faces, *i.e.* for any hanging node $\boldsymbol{b}$, the subset of a hyperplane made of the union of the faces having $\boldsymbol{b}$ as vertex.

Let $(a_\Sigma)_{\Sigma \in \mathcal{E}_\mathrm{c}}$ be a family of real numbers such that for all $\Sigma \in \mathcal{E}_\mathrm{c}$, $|a_\Sigma| \le 1$. Let v be a function of the Rannacher-Turek space V_h associated with $\mathcal{M}$, and, for $\Sigma \in \mathcal{E}_\mathrm{c}$ let $[v]_\Sigma$ be the jump of v through Σ. Then the following bound holds:

$$\sum_{\Sigma \in \mathcal{E}_\mathrm{c}} \Big|\int_\sigma a_\Sigma\ [v]_\Sigma\ g\Big| \le C\ h \|v\|_{h,\mathrm{b}}\ \|g\|_{\mathrm{H}^1_0(\Omega)}, \qquad \forall g \in \mathrm{H}^1_0(\Omega).$$

where the real number C only depends on θ_0 and d.

Proof. Let $\Sigma \in \mathcal{E}_{\rm c}$ and $v \in V_h$. Since the integral of the jump of v through Σ is zero, we have:

$$\int_\Sigma a_\Sigma \; [v]_\Sigma \, g = \int_\Sigma a_\Sigma \; [v]_\Sigma \, (g - g_\Sigma),$$

where g_Σ is any real number. By the Cauchy-Schwarz inequality in $\mathrm{L}^2(\Sigma)$, we thus get:

$$\sum_{\Sigma \in \mathcal{E}_{\rm c}} \Big| \int_\Sigma a_\Sigma \; [v]_\Sigma \, g \Big| \leq \sum_{\Sigma \in \mathcal{E}_{\rm c}} \big(\int_\Sigma [v]_\Sigma^2 \big)^{\frac{1}{2}} \; \big(\int_\Sigma (g - g_\Sigma)^2 \big)^{\frac{1}{2}}.$$

Let us now decompose the integral over Σ in integrals over the faces (in fact, for the coarse faces only), then use the concavity of the square root function, to obtain:

$$\sum_{\Sigma \in \mathcal{E}_{\rm c}} \Big| \int_\Sigma a_\Sigma \; [v]_\Sigma \, g \Big| \leq \sum_{\Sigma \in \mathcal{E}_{\rm c}} \Big(\sum_{\sigma \subset \Sigma} \big(\int_\sigma [v]_\sigma^2 \big)^{\frac{1}{2}} \Big) \; \big(\int_\Sigma (g - g_\Sigma)^2 \big)^{\frac{1}{2}}.$$

The discrete Cauchy-Schwarz inequality now yields:

$$\sum_{\Sigma \in \mathcal{E}_{\rm c}} \Big| \int_\Sigma a_\Sigma \; [v]_\Sigma \, g \Big| \leq \Big(\sum_{\Sigma \in \mathcal{E}_{\rm c}} \frac{1}{h_\sigma} \int_\sigma [v]_\sigma^2 \, \mathrm{d}\gamma \Big)^{\frac{1}{2}} \underbrace{\left(\sum_{\Sigma \in \mathcal{E}_{\rm c}} h_\Sigma \int_\Sigma (g - g_\Sigma)^2 \right)^{\frac{1}{2}}}_{T_1},$$

where h_Σ stands for the sum of the diameters of the faces included in Σ (which is equal to the diameter of Σ in two dimensions, and lower than four times this diameter in three dimensions). By Lemma 4.3, the first term of the latter product is bounded by $C(\theta_0) \, \|v\|_{h,\mathrm{b}}$. For the second one, for $\Sigma \in \mathcal{E}_{\rm c}$, let K_Σ be a cell of the mesh having Σ as a whole side (two choices are possible for a standard face, and only one for a coarse one). Applying the trace lemma 4.4, we get:

$$T_1^2 \leq C(d) \sum_{\Sigma \in \mathcal{E}_{\rm c}} \frac{h_\Sigma}{r_{K_\Sigma}} \; \Big(\|g - g_\Sigma\|^2_{\mathrm{L}^2(K_\Sigma)} + h^2_{K_\Sigma} \|\nabla g\|^2_{\mathrm{L}^2(K_\Sigma)^d} \Big).$$

Choosing for g_Σ the mean value of g on K_Σ and using (4.16), we thus get:

$$T_1^2 \leq C(d) \sum_{\Sigma \in \mathcal{E}_{\rm c}} (1 + \frac{1}{\pi^2}) \, \frac{h_\Sigma}{r_{K_\Sigma}} h^2_{K_\Sigma} \|\nabla g\|^2_{\mathrm{L}^2(K_\Sigma)^d} \leq C(d, \theta_0) \; h^2 \sum_{\Sigma \in \mathcal{E}_{\rm c}} \|\nabla g\|^2_{\mathrm{L}^2(K_\Sigma)^d}.$$

The result follows by observing that the H^1 semi-norm of g on a given cell K of the mesh appears at most $2d$ (the maximum number of sides of a cell K) times in the summation. $\square$

We define by r_h the following interpolation operator:

$$r_h : \quad \left| \begin{array}{lcl} \mathrm{H}^1_0(\Omega) & \longrightarrow & V_h \\ v & \mapsto & r_h v(\boldsymbol{x}) = \displaystyle\sum_{\sigma \in \mathcal{E}} |\sigma|^{-1} \Big(\int_\sigma v \, \mathrm{d}\gamma \Big) \; \zeta_\sigma(\boldsymbol{x}). \end{array} \right. \tag{4.18}$$

The stability and approximation properties of r_h are given in the following lemma.

Lemma 4.9 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \leq \theta_0$, with θ_h defined by (4.4). There exists C_1 and C_2, only depending on θ_0 such that

1. Stability:
$$\forall v \in \mathrm{H}^1_0(\Omega), \quad \|r_h v\|_{h,\mathrm{b}} \le C_1 \; \|\boldsymbol{\nabla} v\|_{\mathrm{L}^2(\Omega)^d}.$$

2. Approximation properties:
$$\begin{aligned}&\forall v \in \mathrm{H}^1_0(\Omega) \cap \mathrm{H}^2(\Omega),\ \forall K \in \mathcal{M},\\ &\qquad \|v - r_h v\|_{\mathrm{L}^2(K)} + h_K \, \|\boldsymbol{\nabla}(v - r_h v)\|_{\mathrm{L}^2(K)^d} \le C_2 \, h_K (h_K + \alpha_K) \, |v|_{\mathrm{H}^2(K)}.\end{aligned}$$

Proof. The stability property follows from usual estimates on the shape functions and from the trace lemma 4.4. We prove the approximation property. If there is no hanging node on the faces of K, r_h is the usual Rannacher-Turek interpolation operator, and the result is known. In the other case, for $d = 2$, let us suppose that an initial face Σ of K has been split: $\Sigma = \sigma_1 \cup \sigma_2$. Let $v \in \mathrm{H}^1_0(\Omega)$. Then we get for the part of the expansion of $r_h v$ associated with σ_1 and σ_2, let us say $r_h v^{\sigma_1 \cup \sigma_2}$:

$$\begin{aligned} r_h v^{\sigma_1 \cup \sigma_2}|_K &= \frac{1}{|\sigma_1|} \Big(\int_{\sigma_1} v \,\mathrm{d}\gamma \Big) \; \frac{|\sigma_1|}{|\Sigma|} \; \zeta_\Sigma(\boldsymbol{x}) + \frac{1}{|\sigma_2|} \Big(\int_{\sigma_2} v \,\mathrm{d}\gamma \Big) \; \frac{|\sigma_2|}{|\Sigma|} \; \zeta_\Sigma(\boldsymbol{x}) \\ &= \frac{1}{|\Sigma|} \; \Big(\int_{\Sigma} v \,\mathrm{d}\gamma \Big) \; \zeta_\Sigma(\boldsymbol{x}). \end{aligned}$$

Once again, we recognize the usual Rannacher-Turek interpolation operator. The same arguments readily extend to the 3D case. □

Remark 4.4 (*inf-sup* condition on locally refined meshes) The same computation shows that, as for the usual Rannacher-Turek approximation on regular meshes, the operator r_h, or, more precisely, its natural extension to vector-valued functions, is a Fortin operator (*i.e.* continuous from $\mathrm{H}^1_0(\Omega)^d$ to V_h^d endowed with the H^1-broken norm and such that $\int_\Omega q \,\mathrm{div}(\boldsymbol{u} - r_h \boldsymbol{u}) = 0$ for any discrete pressure function q); this implies that the *inf-sup* condition is satisfied by the pair of velocity and pressure approximation spaces also on locally refined meshes.

4.4.2 Estimates on the discrete convective term

The form b_h is a discretization of the convection term on the cells of the dual mesh. The analysis of the scheme actually requires an equivalent (or nearly equivalent) re-formulation of the form b_h on the cells of the primal mesh $\mathcal{M}$, that makes use of the primal fluxes $F_{K,\sigma}(\boldsymbol{w})$. Indeed, contrary to the dual fluxes $F_{\sigma,\epsilon}(\boldsymbol{w})$, the expression of $F_{K,\sigma}(\boldsymbol{w})$ with respect to the convection field $\boldsymbol{w}$ is quite simple (see (4.7)). This motivates the introduction of the following auxiliary form:

$$\widetilde{b}_h(\boldsymbol{w}, u, v) = \sum_{K \in \mathcal{M}} v_K \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}(\boldsymbol{w}) \; u_\sigma, \qquad u, v \in V_h, \tag{4.19}$$

where $v_K = \sum_{\sigma \in \mathcal{E}(K)} \xi_K^\sigma \, v_\sigma$ is a convex combination of $(v_\sigma)_{\sigma \in \mathcal{E}(K)}$, where the coefficient ξ_K^σ is equal to $1/(2d)$, when σ has not been split, or $1/(2^d d)$ otherwise. The following lemma provides a bound of the error made when replacing b_h by $\widetilde{b}_h$. It may be seen as a simplified version of a slightly more general result, dealing with density-dependent fluxes $F_{K,\sigma}$, which may be found in [18]; we however give its proof, for the sake of completeness.

Lemma 4.10 Let $\theta_0 > 0$ and let $\mathcal{M}$ be a locally refined mesh of the computational domain Ω such that $\theta_h \le \theta_0$, with θ_h defined by (4.4). There exists C, only depending on θ_0 such that:

$$|\widetilde{b}_h(\boldsymbol{w}, u, v) - b_h(\boldsymbol{w}, u, v)| \le C \, h \, \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d} \|u\|_{h,\mathrm{b}} \|v\|_{h,\mathrm{b}}, \qquad \forall u, v \in V_h. \tag{4.20}$$

Proof. Denote $R = b_h(\boldsymbol{w}, u, v) - \widetilde{b}_h(\boldsymbol{w}, u, v)$. In the expression (4.10) of $b_h(\boldsymbol{w}, u, v)$, for $\sigma = K|L$, let us split the sum over the fluxes through the faces of D_σ in the sum over the dual faces, on one side, included in K and, on the other side, included in L. We get by conservativity (*i.e.* using $F_{K,\sigma}(\boldsymbol{w}) = -F_{L,\sigma}(\boldsymbol{w})$):

$$b_h(\boldsymbol{w}, u, v) = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} v_\sigma \Big(F_{K,\sigma}(\boldsymbol{w})\ u_\sigma + \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\\ \epsilon\subset K,\ \epsilon=D_\sigma|D'_\sigma}} F_{\sigma,\epsilon}(\boldsymbol{w})\ \frac{u_\sigma + u_{\sigma'}}{2}\Big).$$

Let us write $b_h(\boldsymbol{w}, u, v) = T_1 + T_2$ with:

$$T_1 = \sum_{K\in\mathcal{M}} v_K \sum_{\sigma\in\mathcal{E}(K)} \Big(F_{K,\sigma}(\boldsymbol{w})\ u_\sigma + \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\\ \epsilon\subset K,\ \epsilon=D_\sigma|D'_\sigma}} F_{\sigma,\epsilon}(\boldsymbol{w})\ \frac{u_\sigma + u_{\sigma'}}{2}\Big),$$

$$T_2 = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} (v_\sigma - v_K) \Big(F_{K,\sigma}(\boldsymbol{w})\ u_\sigma + \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\\ \epsilon\subset K,\ \epsilon=D_\sigma|D'_\sigma}} F_{\sigma,\epsilon}(\boldsymbol{w})\ \frac{u_\sigma + u_{\sigma'}}{2}\Big).$$

By the conservativity assumption of the dual fluxes $(H2)$ (see Definition 1), we remark that $T_1 = \widetilde{b}_h(\boldsymbol{w}, u, v)$ so that $R = T_2$. Using now $(H1)$, we write $R = R_1 + R_2$ with:

$$R_1 = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} (v_\sigma - v_K) \Big(\sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\\ \epsilon\subset K,\ \epsilon=D_\sigma|D'_\sigma}} F_{\sigma,\epsilon}(\boldsymbol{w})\ \frac{u_{\sigma'} - u_\sigma}{2}\Big),$$

$$R_2 = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} (v_\sigma - v_K)\ u_\sigma\ \xi_K^\sigma \Big(\sum_{\sigma'\in\mathcal{E}(K)} F_{K,\sigma'}(\boldsymbol{w})\Big).$$

The assumption $(H3)$ yields $|F_{\sigma,\epsilon}(\boldsymbol{w})| \leq C\|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ h_K^{d-1}$. As a consequence, since v_K is a convex combination of the $(v_\sigma)_{\sigma\in\mathcal{E}(K)}$, we have for any $K \in \mathcal{M}$:

$$\Big| \sum_{\sigma\in\mathcal{E}(K)} (v_\sigma - v_K) \Big(\sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\\ \epsilon\subset K,\ \epsilon=D_\sigma|D'_\sigma}} F_{\sigma,\epsilon}(\boldsymbol{w})\ \frac{u_{\sigma'} - u_\sigma}{2}\Big)\Big| \leq$$
$$C\|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d} h \sum_{\sigma,\sigma',\sigma'',\sigma'''\in\mathcal{E}(K)} h_K^{d-2}\ |v_\sigma - v_{\sigma'}|\ |u_{\sigma''} - u_{\sigma'''}|,$$

and, for σ, $\sigma' \in \mathcal{E}(K)$, the quantity $|u_\sigma - u_{\sigma'}|$ (or $|v_\sigma - v_{\sigma'}|$) appears in the sum a finite number of times which depends of the dimension d. Hence, by the Cauchy-Schwarz inequality:

$$|R_1| \leq C\|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\|u\|_{h,\mathrm{fv}}\ \|v\|_{h,\mathrm{fv}} h \leq C'\|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\|u\|_{h,\mathrm{b}}\ \|v\|_{h,\mathrm{b}} h,$$

by Lemma 4.1. Let us now turn to R_2. By definition of v_K, $\sum_{\sigma\in\mathcal{E}(K)} \xi_K^\sigma\ (v_\sigma - v_K) = 0$, and we obtain that:

$$R_2 = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} (v_\sigma - v_K)\xi_K^\sigma\ (u_\sigma - u_K)\ \Big(\sum_{\sigma'\in\mathcal{E}(K)} F_{K,\sigma'}(\boldsymbol{w})\Big),$$

so, once again:

$$|R_2| \leq C\|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ h \sum_{K\in\mathcal{M}} h_K^{d-2} \sum_{\sigma\in\mathcal{E}(K)} |v_\sigma - v_K|\ |u_\sigma - u_K|,$$

and, by the same arguments as for the term R_1, we conclude the proof invoking Lemma 4.1. □

4.4.3 Proof of Theorem 4.1

Let u be the solution of the continuous problem (4.2) which we assume to belong to $\mathrm{H}^1_0(\Omega)\cap\mathrm{H}^2(\Omega)$ and let u_h be the solution of the scheme (4.8). By the triangle inequality, we have $\|u_h-u\|_{h,\mathrm{b}}\leq\|u_h-r_hu\|_{h,\mathrm{b}}+\|r_hu-u\|_{h,\mathrm{b}}$ where r_h is the interpolation operator defined in (4.18). The approximation property stated in Lemma 4.9 yields:

$$\|r_hu-u\|_{h,\mathrm{b}}=\Big(\sum_{K\in\mathcal{M}}\|\boldsymbol{\nabla}(u-r_hu)\|^2_{\mathrm{L}^2(K)^d}\Big)^{\frac12}\leq C_2(h+\alpha_h)|u|_{\mathrm{H}^2(\Omega)}. \tag{4.21}$$

To complete the proof, we now have to estimate the quantity $\|u_h-r_hu\|_{h,\mathrm{b}}$. To this purpose, we introduce $\|v\|_h^2=\mathcal{A}_h(v,v)=(v,v)_h+\|v\|^2_{h,\mathrm{b}}$, which defines a norm on V_h that controls the broken Sobolev H^1-norm $\|v\|_{h,\mathrm{b}}$. Hence, we have:

$$\|u_h-r_hu\|_{h,\mathrm{b}}\leq\|u_h-r_hu\|_h\leq\sup_{v\in V_h}\frac{\mathcal{A}_h(u_h-r_hu,v)}{\|v\|_h}.$$

Since u_h is the solution of the scheme (4.8), we have $\mathcal{A}_h(u_h,v)=(f,v)$ for all $v\in V_h$. In addition, u belongs to $\mathrm{H}^1_0(\Omega)\cap\mathrm{H}^2(\Omega)$, which implies that the strong form of the continuous problem, *i.e.* Equation (4.1a), holds in $\mathrm{L}^2(\Omega)$ and thus, since $V_h\subset\mathrm{L}^2(\Omega)$:

$$(u,v)+(\mathrm{div}(u\boldsymbol{w}),v)-(\boldsymbol{\Delta}u,v)=(f,v),\qquad\forall v\in V_h.$$

As a consequence, we have $\mathcal{A}_h(u_h-r_hu,v)=T_1+T_2+T_3$ where:

$$T_1=(u,v)-(r_hu,v)_h,\quad T_2=(\mathrm{div}(u\boldsymbol{w}),v)-b_h(\boldsymbol{w},r_hu,v),\quad T_3=(-\boldsymbol{\Delta}u,v)-a_h(r_hu,v).$$

Reaction term - The term T_1 may be split as follows:

$$T_1=\underbrace{(u,v-v_\mathrm{c})}_{T_{1,1}}\underbrace{-\sum_{\sigma\in\mathcal{E}}|D_\sigma|\;v_\sigma\Big((r_hu)_\sigma-\frac{1}{|D_\sigma|}\int_{D_\sigma}u\Big)}_{T_{1,2}},$$

where v_c stands for the piecewise constant function over each diamond cell and which takes the value v_σ over D_σ, for $\sigma\in\mathcal{E}$. By Lemma 4.2, we get for the first term at the right-hand side of the previous relation:

$$|T_{1,1}|=|(u,v-v_\mathrm{c})|\leq C\,h\;\|u\|_{\mathrm{L}^2(\Omega)}\;\|v\|_{h,\mathrm{b}}\leq C\,h\;\|u\|_{\mathrm{L}^2(\Omega)}\;\|v\|_h.$$

Let us now turn to the second term. We adopt the notation $m_\omega(u)=|\omega|^{-1}\int_\omega u\mathrm{d}\omega$ where ω is any measurable subset of $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$. By definition, we have $(r_hu)_\sigma=m_\sigma(u)$ and the second term may be decomposed as follows (invoking the boundary conditions to exclude the external faces from the summation):

$$\begin{aligned}|T_{1,2}|&=\Big|\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}}|D_\sigma|\;v_\sigma\big(m_\sigma(u)-m_{D_\sigma}(u)\big)\Big|\\&=\frac12\Big|\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\\sigma=K|L}}|D_\sigma|\;v_\sigma\big(m_\sigma(u)-m_{D_{K,\sigma}}(u)+m_\sigma(u)-m_{D_{L,\sigma}}(u)\big)\Big|\\&\leq\frac12\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\\sigma=K|L}}|D_\sigma|\,|v_\sigma|\,|m_\sigma(u)-m_{D_{K,\sigma}}(u)|+\frac12\sum_{\substack{\sigma\in\mathcal{E}_{\mathrm{int}}\\\sigma=K|L}}|D_\sigma|\,|v_\sigma|\,|m_\sigma(u)-m_{D_{L,\sigma}}(u)|.\end{aligned}$$

We only treat the first term since the following calculations are similar for the second one. We have $|m_\sigma(u)-m_{D_{K,\sigma}}(u)| \leq |m_\sigma(u)-m_K(u)|+|m_K(u)-m_{D_{K,\sigma}}(u)|$. Applying Lemma 4.6, we obtain $|m_K(u)-m_{D_{K,\sigma}}(u)| \leq 2\, r_K^{-d/2}\, \pi^{-1} h_K \|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(K)^d}$. Lemma 4.7 provides a similar control on $|m_\sigma(u)-m_K(u)|$ and we get, by the regularity of the mesh:

$$\begin{aligned}
|T_{1,2}| &\leq C\, h \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} |D_\sigma|^{1/2}\, |v_\sigma|\, \|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(K)^d} \\
&\leq C\, h \Big(\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} |D_\sigma|\, |v_\sigma|^2\Big)^{\frac{1}{2}} \Big(\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \|\boldsymbol{\nabla} u\|^2_{\mathrm{L}^2(K)^d}\Big)^{\frac{1}{2}} \\
&= C\, h\ (v,v)_h^{\frac{1}{2}} \Big(\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} \|\boldsymbol{\nabla} u\|^2_{\mathrm{L}^2(K)^d}\Big)^{\frac{1}{2}}.
\end{aligned}$$

Observing that the H^1 semi-norm of u on K appears at most $2^d d$ times in the summation, we obtain $|T_{1,2}| \leq Ch\|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(\Omega)^d}\|v\|_h$.

Convection term - Let us now turn to T_2. We may write $T_2 = T_{2,1} + R$ where:

$$T_{2,1} = (\mathrm{div}(u\boldsymbol{w}), v) - \widetilde{b}_h(\boldsymbol{w}, r_h u, v), \qquad R = \widetilde{b}_h(\boldsymbol{w}, r_h u, v) - b_h(\boldsymbol{w}, r_h u, v).$$

Applying Lemma 4.10 and invoking the stability property of Lemma 4.9, we get

$$|R| \leq C\, h\, \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ \|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(\Omega)^d}\|v\|_h.$$

For the term $T_{2,1}$, we have:

$$T_{2,1} = (\mathrm{div}(u\boldsymbol{w}), v - v_{\mathrm{m}}) - \sum_{K\in\mathcal{M}} v_K \sum_{\sigma\in\mathcal{E}(K)} \int_\sigma (u - (r_h u)_\sigma)\boldsymbol{w}\cdot\boldsymbol{n}_{K,\sigma},$$

where v_{m} stands for the piecewise constant function over each cell K and equal to v_K. By the Cauchy-Schwarz inequality and Lemma 4.2, the first term at the right-hand side of this relation may be bounded as follow:

$$|(\mathrm{div}(u\boldsymbol{w}), v - v_{\mathrm{m}})| \leq C\, h\, \|\boldsymbol{w}\|_{\mathrm{H}^1(\Omega)^d}\, \|u\|_{\mathrm{H}^1(\Omega)}\|v\|_h.$$

For the second term, we first remark that:

$$\begin{aligned}
\Big|\int_\sigma \big(u - (r_h u)_\sigma\big)\ \boldsymbol{w}\cdot\boldsymbol{n}_{K,\sigma}\Big| &\leq \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ |\sigma|^{1/2}\ \|u - m_\sigma(u)\|_{\mathrm{L}^2(\sigma)} \\
&\leq \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ \big(|\sigma|^{1/2}\ \|u - m_K(u)\|_{\mathrm{L}^2(\sigma)} + |\sigma|\ |m_K(u) - m_\sigma(u)|\big) \\
&\leq C\ \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ h_K^{d/2}\ \|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(K)^d},
\end{aligned}$$

by the regularity of the mesh and Lemmas 4.4, 4.5 and 4.7. Hence, reordering the summations, we thus get (invoking the boundary conditions to exclude the external faces from the summation):

$$\begin{aligned}
\Big| \sum_{K\in\mathcal{M}} v_K \sum_{\sigma\in\mathcal{E}(K)} \int_\sigma (u-(r_h u)_\sigma)\ \boldsymbol{w}\cdot\boldsymbol{n}_{K,\sigma}\Big| \\
&\leq C\ \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ h \sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} h_K^{d/2-1}\ \|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(K)^d}\ |v_K-v_L| \\
&\leq C\ \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ h\ \Big(\sum_{\sigma\in\mathcal{E}_{\rm int}} \|\boldsymbol{\nabla} u\|^2_{\mathrm{L}^2(K)^d}\Big)^{1/2} \Big(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} h_K^{d-2}\,|v_K-v_L|^2\Big)^{\frac12} \\
&\leq C\,2^d\,d\ \|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d}\ \|\boldsymbol{\nabla} u\|_{\mathrm{L}^2(\Omega)^d}\ h\ \Big(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} h_K^{d-2}|v_K-v_L|^2\Big)^{\frac12}.
\end{aligned}$$

Since v_K is a convex combination of $(v_\sigma)_{\sigma\in\mathcal{E}(K)}$, we have, possibly splitting a difference $v_\sigma - v_{\sigma'}$ as $v_\sigma - v_{\sigma'} = (v_\sigma - v_{K|L}) - (v_{\sigma'} - v_{K|L})$ if σ and σ' are not faces of the same element:

$$|v_K - v_L| \leq 2 \sum_{\sigma,\sigma'\in\mathcal{E}(K)} |v_\sigma - v_{\sigma'}| + 2 \sum_{\sigma,\sigma'\in\mathcal{E}(L)} |v_\sigma - v_{\sigma'}|.$$

Therefore, since the number of edges of an element is bounded, there exists a fixed real number C such that:

$$\Big(\sum_{\substack{\sigma\in\mathcal{E}_{\rm int}\\ \sigma=K|L}} h_K^{d-2}|v_K-v_L|^2\Big)^{1/2} \leq C\ \Big(\sum_{K\in\mathcal{M}} h_K^{d-2} \sum_{\sigma,\sigma'\in\mathcal{E}(K)} |v_\sigma - v_{\sigma'}|^2\Big)^{1/2},$$

and, invoking Lemma 4.1, this sum is bounded by $\|v\|_{h,\rm b}$, and thus by $\|v\|_h$. We finally get:

$$|T_{2,1}| \leq C\ \big(\|\boldsymbol{w}\|_{\mathrm{L}^\infty(\Omega)^d} + \|\boldsymbol{w}\|_{\mathrm{H}^1(\Omega)^d}\big)\ \|u\|_{\mathrm{H}^1(\Omega)}\ \|v\|_h,$$

where C only depends on θ_0. Combining this relation with the bound obtained for R yields that T_2 satisfies the same inequality.

Diffusion term - We finally turn to T_3. Integrating by parts, we get:

$$T_3 = \sum_{K\in\mathcal{M}} \int_K (\boldsymbol{\nabla} u - \boldsymbol{\nabla} r_h u)\cdot\boldsymbol{\nabla} v - \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} \int_\sigma v\ \boldsymbol{\nabla} u\cdot\boldsymbol{n}_{K,\sigma}.$$

Thanks to Lemma 4.9, we obtain for the first term, using first the Cauchy-Schwarz inequality in $\mathrm{L}^2(K)$ and then the discrete Cauchy-Schwarz inequality:

$$\Big|\sum_{K\in\mathcal{M}} \int_K (\boldsymbol{\nabla} u - \boldsymbol{\nabla} r_h u)\cdot\boldsymbol{\nabla} v\Big| \leq C\,(h+\alpha_h)\ \|u\|_{\mathrm{H}^2(\Omega)}\ \|v\|_{h,\rm b} \leq C\,(h+\alpha_h)\ \|u\|_{\mathrm{H}^2(\Omega)}\ \|v\|_h.$$

Reordering the sums in the second term yields:

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} \int_\sigma v\ \boldsymbol{\nabla} u\cdot\boldsymbol{n}_{K,\sigma} = \sum_{\sigma\in\mathcal{E}} \int_\sigma [v]_\sigma\ \boldsymbol{\nabla} u\cdot\boldsymbol{n}_\sigma,$$

where $[v]_\sigma$ and $\boldsymbol{n}_\sigma$ stand for the jump of v through σ and a normal vector to σ, with the same orientation. Since, on a coarse face (*i.e.* the subset of a hyperplane which consists in the union of the faces sharing the same hanging node), the normal is the same, we can group the terms in the above sum to obtain, with the notations of Lemma 4.8:

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} \int_\sigma v\, \boldsymbol{\nabla} u\cdot \boldsymbol{n}_{K,\sigma} = \sum_{\Sigma\in\mathcal{E}_{\rm c}} \int_\Sigma [v]_\Sigma\, \boldsymbol{\nabla} u\cdot \boldsymbol{n}_\Sigma.$$

We thus may apply Lemma 4.8 for $i=1,..,d$ with $a_\Sigma=\boldsymbol{n}^i_\Sigma$ and $g=\partial_i u$, and we get:

$$\Big|\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} \int_\sigma v\, \boldsymbol{\nabla} u\cdot \boldsymbol{n}_{K,\sigma}\Big| \leq C\, h\, \|u\|_{\mathrm{H}^1(\Omega)}\ \|v\|_h,$$

which provides the bound for T_3 which we are seeking.

Conclusion - Collecting the bounds for T_1, T_2 and T_3, we obtain that $\|u_h - r_h u\|_{h,\mathrm{b}} \leq C\,(h+\alpha_h)\ \|u\|_{\mathrm{H}^2(\Omega)}$ where C only depends on $\boldsymbol{w}$, Ω and θ_0. Combining this with equation (4.21) concludes the proof of Theorem 4.1.

5 Numerical tests

All computations presented in this section are performed with the CALIF^{3}S free component library for fluid flows computation developed at IRSN [5].

5.1 A stationary incompressible flow

We first assess the behavior of the proposed numerical scheme on an exact analytical solution to the stationary incompressible Navier-Stokes equations known as the Kovasznay flow [17]. The velocity and pressure fields are given by:

$$\boldsymbol{u} = \begin{bmatrix} 1-e^{\lambda x}\ \cos(2\pi y) \\ \dfrac{\lambda}{2\pi}\, e^{\lambda x}\ \sin(2\pi y)\end{bmatrix},\quad p=\frac{1}{2}\,(1-e^{2\lambda x}),\quad \lambda=\frac{1}{2\mu}-\big(\frac{1}{4\mu^2}+4\pi^2\big)^{1/2},$$

where μ stands for the viscosity of the flow, taken here as $\mu=1/40$. The computational domain is $\Omega=(-0.5,1)\times(-0.5,1.5)$. The mesh is built from a regular $n\times n$ grid, where we refine the sub-domain $\Omega_f=(-0.5,0.5)\times(-0.5,0.5)\cup(0.5,1)\times(0.5,1.5)$ by splitting each (square) cell included in Ω_f in four sub-squares. The solution is computed by the projection scheme, by letting a computed fictitious transient tend to the desired steady state. Boundary conditions are given by the analytical solution. The obtained numerical errors for various values of n are gathered in the following table, where $\boldsymbol{u}_{\rm exact}$ and $p_{\rm exact}$ stand for the exact velocity and pressure, respectively.

n	$\|\boldsymbol{u}-\boldsymbol{u}_{\rm exact}\|_{\mathrm{L}^2(\Omega)^d}$	$\|p-p_{\rm exact}\|_{\mathrm{L}^2(\Omega)}$
20	0.0384	0.0334
40	0.00825	0.0158
80	0.00211	0.00782
160	0.000544	0.00390

The observed order of convergence (in L^2-norm) is approximately 2 for the velocity and 1 for the pressure.

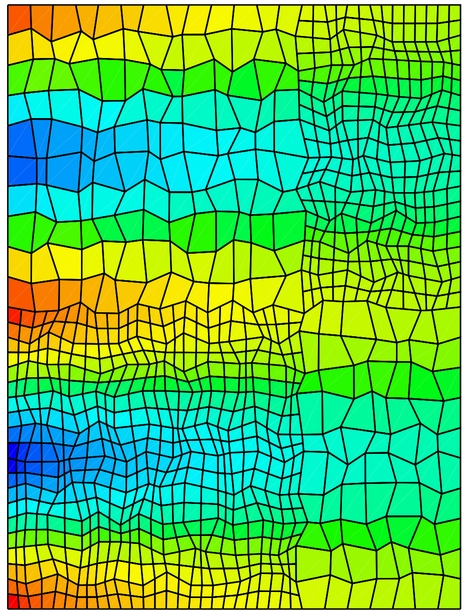


Figure 4: The unstructured mesh built by perturbation of the 20×20 uniform grid. The background is coloured as a function of the first component of the velocity.

We now confirm this behaviour on unstructured grids, which are obtained as follows. First, we perturb the 20×20 uniform grid, by moving each internal vertex of the mesh to a random position on a circle centered on its initial location and the radius of which is equal to 0.3 times the smallest distance between the considered vertex and its neighbours (so, for the specific unperturbed mesh used here, 1.5/20). Secondly, we build 3 refined meshes, by splitting each cell in 2, 4 and 8 respectively; by this process, the deviation from a parallelogram of each cell is divided by the same ratio, which ensures optimal convergence properties for the parametric version of the Rannacher-Turek element (see Remark 4.2). Finally, we apply local refinement to the same zones as previously. The coarsest of the obtained four meshes is plotted on Figure 4.

The obtained numerical errors are given in the following table. The orders of convergence are the same as in the uniform case.

n	$\Vert\boldsymbol{u} - \boldsymbol{u}_{\text{exact}}\Vert_{\mathrm{L}^2(\Omega)^d}$	$\Vert p - p_{\text{exact}}\Vert_{\mathrm{L}^2(\Omega)}$
20	0.0617	0.0406
40	0.0119	0.0179
80	0.00281	0.0087
160	0.000718	0.0043

The contour lines of the first component of the velocity are drawn on Fig. 5, for both meshes obtained from the 80×80 uniform grid. We may check that no spurious perturbation appears along the lines separating the refined and non-refined parts of the computational domain (in other words, the lines composed by the union of the faces including a hanging node). A careful examination is needed to observe that, as expected, contour lines are slightly more irregular in the unstructured case.

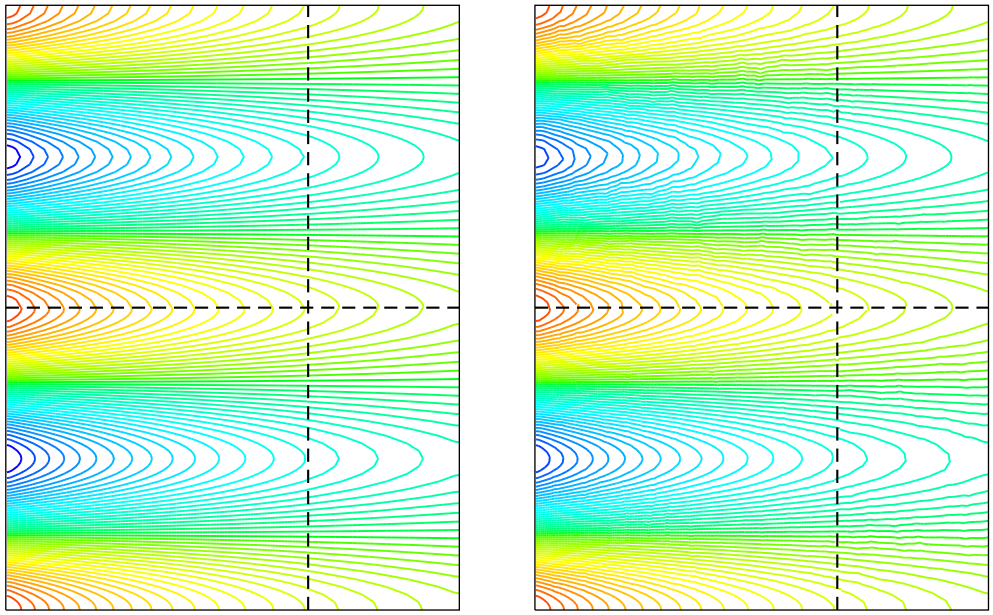


Figure 5: Contour lines of the field $\boldsymbol{u}_1$: 80×80 uniform grid (left) and perturbed mesh (right) built by perturbation of the 80×80 grid. The dashed lines materialize the boundary of the refined area (bottom-left and top-right sub-domains).

5.2 Low-Mach buoyant flows

We now turn to a natural convection flow, supposed to be represented by the asymptotic model for low Mach number flows [20], which combines System (1.1) with a balance equation for the energy:

$$\partial_t(\rho c_p \vartheta) + \operatorname{div}(\rho c_p \vartheta \boldsymbol{u}) - \operatorname{div}(\lambda \boldsymbol{\nabla} \vartheta) = 0, \tag{5.1}$$

where ϑ stands for the temperature, and c_p and λ are known positive real numbers (the heat capacity and the temperature diffusion coefficient, respectively). This system is complemented by an equation of state:

$$P_{th} = \rho R \vartheta, \tag{5.2}$$

where R stands for a constant specific to the gas under consideration (see its value below) and P_{th}, the so-called thermodynamical pressure, is supposed to depend only on time, or, in other words, to be constant in space, which has for effect to filter out the acoustic contributions from the flow field.

The computational domain is $\Omega = (0, L)^2$, the velocity and temperature are prescribed on the whole boundary; the velocity is set to zero and the temperature is given by:

$$\vartheta(\boldsymbol{x}) = \frac{L - x_1}{L}\, \vartheta_h + \frac{x_1}{L}\, \vartheta_c, \quad \vartheta_h = (1 + \varepsilon)\, \vartheta_0,\ \vartheta_c = (1 - \varepsilon) \vartheta_0,$$

with $\vartheta_0 = 600$ and $\epsilon = 0.6$. The left and right vertical boundaries are thus set at a constant (hot and cold, *i.e.* ϑ_h and ϑ_c, respectively) temperature, while this latter varies linearly along the horizontal boundaries, from ϑ_h at $x_1 = 0$ to ϑ_c at $x_1 = L$. Variations of temperature are too large for the Boussinesq approximation of (1.1)-(5.1) to be valid [19].

The initial thermodynamical pressure and temperature are set to $P_{th}(0) = 101325$ and $\vartheta(\boldsymbol{x}, 0) = \vartheta_0$, $\forall \boldsymbol{x} \in \Omega$, and the fluid is supposed to be initially at rest. The evolution of P_{th} with time must be given by and additional relation, which, in the case of impermeability boundaries as here, may just be the conservation of the total mass in the domain:

$$\int_\Omega \rho(\boldsymbol{x}, t)\,\mathrm{d}\boldsymbol{x} = \frac{P_{th}(t)}{R} \int_\Omega \frac{1}{\vartheta(\boldsymbol{x}, t)}\,\mathrm{d}\boldsymbol{x} = |\Omega|\ \rho_0,\ \text{with } \rho_0 = \frac{P_{th}(0)}{R\,\vartheta_0}. \tag{5.3}$$

A dimensional analysis shows that the flow is governed by two non-dimensional numbers, namely the Prandtl and the Rayleigh numbers, defined respectively by:

$$\mathrm{Pr} = \frac{\mu\, c_p}{\lambda}, \qquad \mathrm{Ra} = \frac{\rho_0^2\, c_p\, g\, (\vartheta_h - \vartheta_c)\, L^3}{\mu\, \lambda\, \vartheta_0}.$$

For the practical application performed here, we choose $g = 9.81$ and physical properties close to those of the air: $\mu = 1.68\,10^{-5}$, $c_p = \gamma R/(\gamma - 1)$, with $R = 287$. The Prandtl number is $\mathrm{Pr} = 0.71$, which allows to compute the thermal diffusion λ, and the size of the domain L will be determined as a function of the chosen Rayleigh number.

A stability analysis of this natural convection flow is given in [27]. It shows that the flow reaches a steady state up to a critical value of the Rayleigh number approximately equal to $\mathrm{Ra} = 2.1\,10^6$. Beyond this value, the flow remains time-dependent, with traveling waves along the boundaries, issued from exiting corners of the vertical boundary layers. Our aim here is to assess the capability of the proposed scheme to confirm this behaviour in the non-Boussinesq approach; an accurate determination of the critical Rayleigh number is however beyond the scope of this section.

For the solution (1.1)-(5.1), we implement a four step algorithm [4, Section 4]. We first solve (5.1) by a finite volume method, with an explicit MUSCL discretization of the convection term and an implicit approximation of the diffusion [22]. Then the thermodynamical pressure and the density are updated by (5.3) and (5.2). Finally, the velocity and the pressure are computed by the two-step pressure correction scheme (3.3).

The mesh used for this study is built as follows: we start from a 80×80 uniform grid, perform a first refinement step splitting each cell located in $\Omega \setminus (0.15\,L,\ 0.85\,L)^2$, a second one by subdividing once-again the cells in $\Omega \setminus (0.1\,L,\ 0.9\,L)^2$, and, finally, a third one by splitting the resulting cells in $\Omega \setminus (0.05\,L,\ 0.95\,L)^2$; at each refinement step, the cells are cut in 4, so *in fine* the characteristic sizes of the cells near the boundary are $\delta x_1 = \delta x_2 = L/640$. A mesh obtained by the same process from the 10×10 uniform grid is shown on Figure 6. The background color in this figure is the temperature; one can see that steep variations of the solution are concentrated in the boundary layers, which justifies the chosen refinement.

At $Ra = 2\,10^6$, the flow after 20 time units is stationary (at least approximately, see Figure 8 below). The obtained temperature and density fields are plotted on Figure 7. On can observe that the non-linearity of the equation of state makes that the flow looses its symmetry, and generates steep density gradients at the right boundary.

We then plot the evolution with time of the temperature at the location $\boldsymbol{x} = (0.1\,L,\ 0.92\,L)^t$, for $\mathrm{Ra} = 2\,10^6$ and $\mathrm{Ra} = 2.2\,10^6$. We observe that, for $\mathrm{Ra} = 2\,10^6$, the flow tends to a steady state, while an oscillatory (quasi-periodic) behaviour subsists at $\mathrm{Ra} = 2.2\,10^6$, which is consistent with the value for the critical Rayleigh known under the Boussinesq approximation [27].

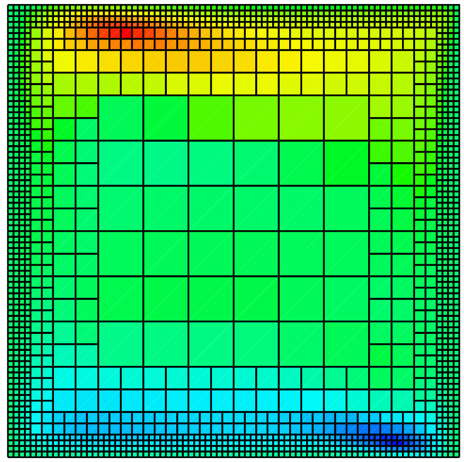


Figure 6: Typical (coarse) mesh and temperature.

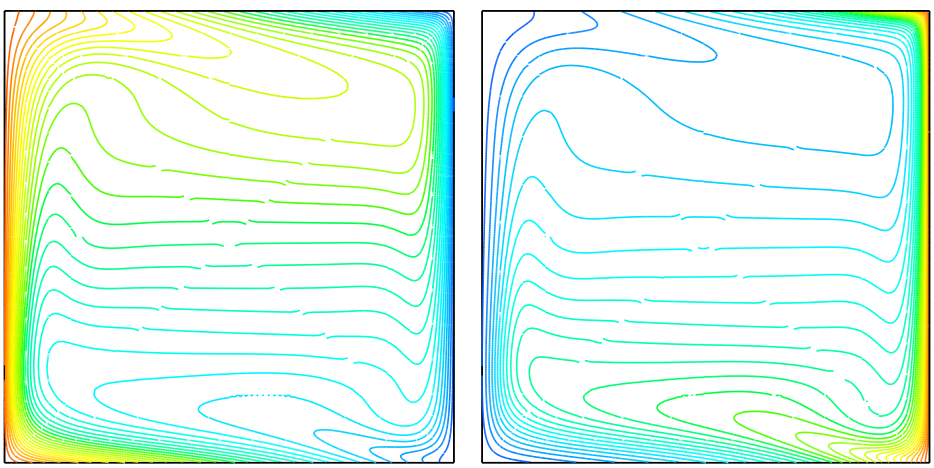
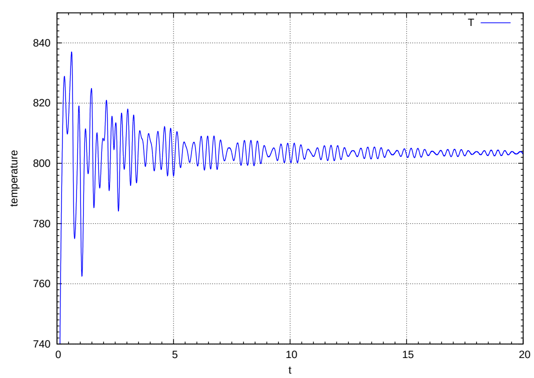
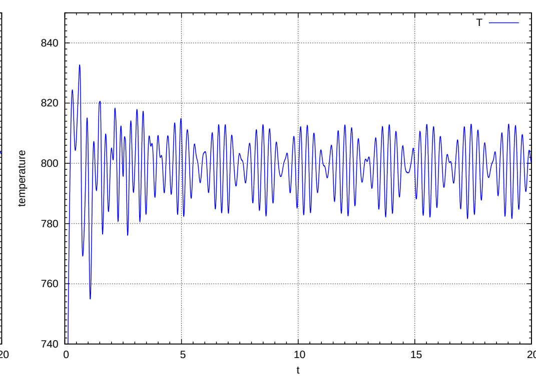


Figure 7: Temperature ($\vartheta \in (240,\ 960)$) and density ($\rho \in (0.33,\ 1.32)$) at the steady state, for $Ra = 2\,10^6$.



$\mathrm{Ra} = 2\,10^6$



$\mathrm{Ra} = 2.2\,10^6$

Figure 8: Temperature as a function of time at the location $\boldsymbol{x} = (0.1\,L,\ 0.92\,L)^t$, for different Rayleigh values.

A Construction of the dual fluxes

A.1 Dual fluxes for non-refined meshes

The system of equations (3.11) has an infinity of solutions, which makes necessary to impose in addition the constraint (H3). Since (3.11) is linear with respect to the $F_{\sigma,\epsilon}$, $\sigma \in \mathcal{E}(K)$, $\epsilon \in \tilde{\mathcal{E}}(D_\sigma)$, $\epsilon \subset K$, a solution of (3.11) may thus be expressed as:

$$F_{\sigma,\epsilon} = \sum_{\sigma' \in \mathcal{E}(K)} (\alpha_K)_\sigma^{\sigma'} F_{K,\sigma'}, \qquad \sigma \in \mathcal{E}(K), \epsilon \in \tilde{\mathcal{E}}(D_\sigma) \text{ and } \epsilon \subset K,$$

and (H3) is equivalent to requiring bounded coefficients $((\alpha_K)_\sigma^{\sigma'})_{\sigma,\sigma' \in \mathcal{E}(K)}$. In addition, since $\xi_K^\sigma = 1/(2d)$ for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$ (we recall that the mesh is not refined here), system (3.11) is completely independent from the cell K under consideration. We may thus consider a particular geometry for K, let us say $K = (0,1)^d$, and find an expression for the coefficients $((\alpha_K)_\sigma^{\sigma'})_{\sigma,\sigma' \in \mathcal{E}(K)}$ which we will apply to all the cells, thus automatically satisfying the constraint (H3). A technique for this computation is described in [1, Section 3.2]. The idea is to build a momentum field $\boldsymbol{w}$ with a constant divergence and such that

$$\int_\sigma \boldsymbol{w} \cdot \boldsymbol{n}_{K,\sigma} = F_{K,\sigma}, \qquad \forall \sigma \in \mathcal{E}(K).$$

Then, an easy computation shows that the following fluxes satisfy (3.11):

$$F_{\sigma,\epsilon} = \int_\epsilon \boldsymbol{w} \cdot \boldsymbol{n}_{\sigma,\epsilon}. \tag{A.1}$$

For $d = 2$, using the notations introduced in Figure 9, such a momentum field $\boldsymbol{w}$ is given by:

$$\boldsymbol{w}(x,y) = \begin{bmatrix} (1-x)\,(-F_{\mathrm{W}}) + x\,F_{\mathrm{E}} \\ (1-y)\,(-F_{\mathrm{S}}) + y\,F_{\mathrm{N}} \end{bmatrix}.$$

Using (A.1), we obtain:

$$F_{\sigma,\epsilon} = \alpha_{\mathrm{W}} F_{\mathrm{W}} + \alpha_{\mathrm{E}} F_{\mathrm{E}} + \alpha_{\mathrm{S}} F_{\mathrm{S}} + \alpha_{\mathrm{N}} F_{\mathrm{N}},$$

with the coefficients α_{W}, α_{E}, α_{S} and α_{N} given in table 1. The notation $F_{\mathrm{W|S}}$ for the dual flux means that one calculates the flux from the western (W) to the southern (S) region with this orientation.

$F_{\sigma,\epsilon}$	α_{W}	α_{E}	α_{S}	α_{N}
$F_{\mathrm{W\|S}}$	$-3/8$	$1/8$	$3/8$	$-1/8$
$F_{\mathrm{S\|E}}$	$-1/8$	$3/8$	$-3/8$	$1/8$
$F_{\mathrm{E\|N}}$	$1/8$	$-3/8$	$-1/8$	$3/8$
$F_{\mathrm{N\|W}}$	$3/8$	$-1/8$	$1/8$	$-3/8$

Table 1: Expression of the dual fluxes in 2D.

For $d = 3$, using the notations introduced in Figure 10, we may choose, for the constant divergence momentum field $\boldsymbol{w}$, the following expression:

$$\boldsymbol{w}(x,y,z) = \begin{bmatrix} (1-x)\,(-F_{\mathrm{W}}) + x\,F_{\mathrm{E}} \\ (1-y)\,(-F_{\mathrm{S}}) + y\,F_{\mathrm{N}} \\ (1-z)\,(-F_{\mathrm{B}}) + z\,F_{\mathrm{F}} \end{bmatrix}.$$

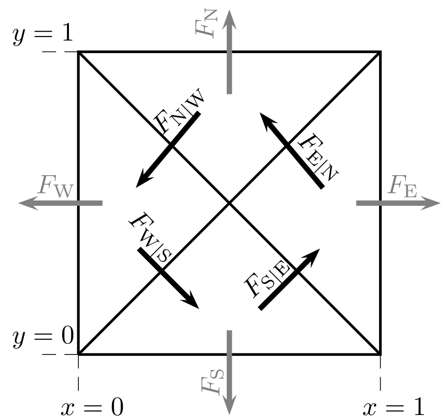


Figure 9: Notations for the primal and dual fluxes in 2D.

The dual fluxes may be expressed as linear combinations of the primal ones:

$$F_{\sigma,\epsilon} = \alpha_{\mathrm{W}} F_{\mathrm{W}} + \alpha_{\mathrm{E}} F_{\mathrm{E}} + \alpha_{\mathrm{S}} F_{\mathrm{S}} + \alpha_{\mathrm{N}} F_{\mathrm{N}} + \alpha_{\mathrm{B}} F_{\mathrm{B}} + \alpha_{\mathrm{F}} F_{\mathrm{F}}$$

where the coefficients α_{W}, α_{E}, α_{S}, α_{N}, α_{B}, α_{F} are given in table 2.

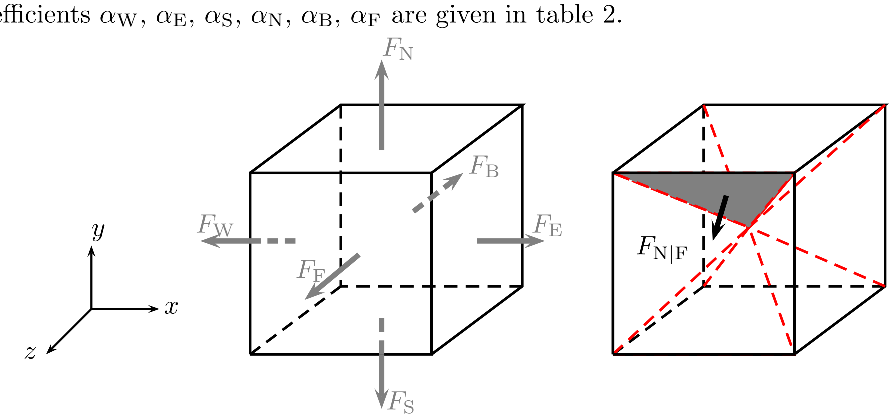


Figure 10: Notations for the primal and dual fluxes in 3D.

A.1.1 Dual fluxes for 2D-refined meshes

Here again, we may restrict the computation to square cells. In 2D, if a primal cell is surrounded with four refined cells, the half-diamond cells are obtained by splitting the cell in four sub-squares, each one being split in two triangles. Hence, eight dual fluxes must be computed; if some of the neighboring cells are not refined, one uses a *coarsening* procedure. We begin with computing the dual fluxes across the four sub-squares faces (solid gray color in Fig. 11) so that (3.11) holds, with $(F_{K,\sigma})_{\sigma\in\mathcal{E}(K)}$ denoted here by F_i $(4 \le i \le 11)$ and $F_{\sigma,\epsilon},\ \sigma \in \mathcal{E}(K), \epsilon \subset K$ denoted here by $\widetilde{F}_i$ $(4 \le i \le 7)$. The linear system to solve has a one dimensional kernel and a particular solution satisfying $(H3)$ is given by:

$F_{\sigma,\epsilon}$	α_{W}	α_{E}	α_{S}	α_{N}	α_{B}	α_{F}
$F_{\mathrm{F}\vert\mathrm{S}}$	0	0	$5/24$	$-1/24$	$1/24$	$-5/24$
$F_{\mathrm{S}\vert\mathrm{B}}$	0	0	$-5/24$	$1/24$	$5/24$	$-1/24$
$F_{\mathrm{B}\vert\mathrm{N}}$	0	0	$-1/24$	$5/24$	$-5/24$	$1/24$
$F_{\mathrm{N}\vert\mathrm{F}}$	0	0	$1/24$	$-5/24$	$-1/24$	$5/24$
$F_{\mathrm{W}\vert\mathrm{S}}$	$-5/24$	$1/24$	$5/24$	$-1/24$	0	0
$F_{\mathrm{S}\vert\mathrm{E}}$	$-1/24$	$5/24$	$-5/24$	$1/24$	0	0
$F_{\mathrm{E}\vert\mathrm{N}}$	$1/24$	$-5/24$	$-1/24$	$5/24$	0	0
$F_{\mathrm{N}\vert\mathrm{W}}$	$5/24$	$-1/24$	$1/24$	$-5/24$	0	0
$F_{\mathrm{F}\vert\mathrm{E}}$	$-1/24$	$5/24$	0	0	$1/24$	$-5/24$
$F_{\mathrm{E}\vert\mathrm{B}}$	$1/24$	$-5/24$	0	0	$5/24$	$-1/24$
$F_{\mathrm{B}\vert\mathrm{W}}$	$5/24$	$-1/24$	0	0	$-5/24$	$1/24$
$F_{\mathrm{W}\vert\mathrm{F}}$	$-5/24$	$1/24$	0	0	$-1/24$	$5/24$

Table 2: Expression of the dual fluxes in 3D.

$$
\begin{aligned}
\widetilde{F}_4 &= \frac{3}{8}\,(F_5 + F_6 - F_{11} - F_4) + \frac{1}{8}\,(F_7 + F_8 - F_9 - F_{10},)\\
\widetilde{F}_5 &= \frac{3}{8}\,(F_7 + F_8 - F_5 - F_6) + \frac{1}{8}\,(F_9 + F_{10} - F_{11} - F_4)\\
\widetilde{F}_6 &= \frac{3}{8}\,(F_9 + F_{10} - F_7 - F_8) + \frac{1}{8}\,(F_4 + F_{11} - F_5 - F_6)\\
\widetilde{F}_7 &= \frac{3}{8}\,(F_4 + F_{11} - F_9 - F_{10}) + \frac{1}{8}\,(F_5 + F_6 - F_7 - F_8)
\end{aligned}
$$

Then, the dual fluxes across the diagonal faces $\bar{F}_i$ $(0 \leq i \leq 3)$ (dashed gray color in Fig. 11) are computed by isolating the sub-squares and applying the procedure described above for the non-refined case. For instance, $\bar{F}_1 = F_{E|N} - F_{W|S}$, where $F_E := F_7$, $F_N := F_8$, $F_W := \widetilde{F}_6$, and $F_S = -\widetilde{F}_5$.

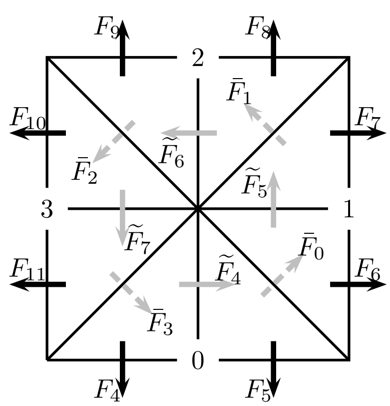


Figure 11: Dual fluxes for the neighboring cell of refined cells (2D case).

A.1.2 Dual fluxes for 3D-refined meshes

The procedure is the same as in the 2D-case. The first step consists in splitting the cube in eight sub-cubes and computing the dual fluxes across the faces of these sub-cubes. The formula of one of these intermediate fluxes $\widetilde{F}$ (see Fig. 12) is given by:

$$\begin{aligned} 24\,\widetilde{F} &= 7\,(\,F_{14} + F_{21} + F_{27}\,) - 7\,(\,F_{15} + F_{19} + F_{25}\,) \\ &+ 2\,(\,F_{7} + F_{20} + F_{26}\,) - 2\,(\,F_{6} + F_{18} + F_{24}\,) \\ &+ 2\,(\,F_{11} + F_{16} + F_{29}\,) - 2\,(\,F_{13} + F_{17} + F_{23}\,) \\ &+ (\,F_{9} + F_{10} + F_{28}\,) - (\,F_{8} + F_{12} + F_{22}\,) \end{aligned}$$

The computation of the other fluxes across the faces separating two sub-cubes is deduced by permutations of the indices.

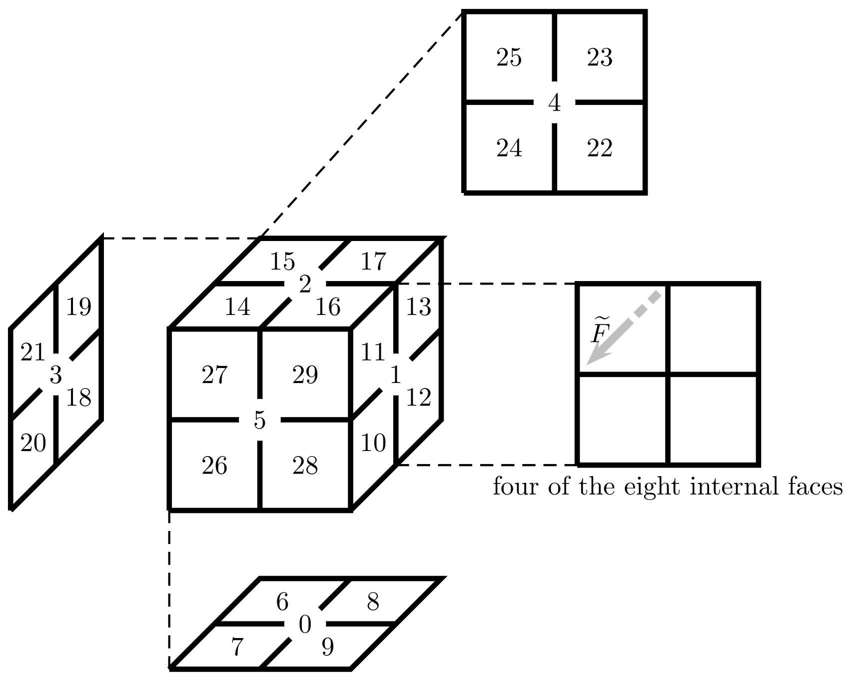


Figure 12: Intermediate dual fluxes for the neighboring cell of refined cells (3D case).

In the second step, each sub-cube is split in 3 half-diamonds of equal volumes. One obtains 24 half-diamonds and 48 internal half-diamond faces of two possible types (see Fig.13). The dual fluxes across these faces are obtained by isolating the sub-cubes and applying the procedure described above for the non-refined case, consisting in integrating the momentum field $\boldsymbol{w}$ over the faces of interest.

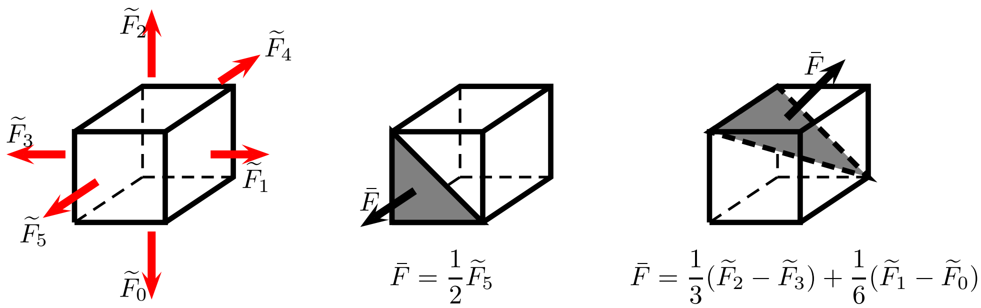


Figure 13: Two possible types of internal half-diamond faces (3D case).

References

[1] G. Ansanay-Alex, F. Babik, J.-C. Latché, and D. Vola. An L^2-stable approximation of the Navier-Stokes convection operator for low-order non-conforming finite elements. *International Journal for Numerical Methods in Fluids*, 66:555–580, 2011.

[2] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Journal for Analysis and its Applications*, 22:751–756, 2003.

[3] R. Becker and S. Mao. Quasi-optimality of adaptative nonconforming finite element methods for the Stokes equations. *SIAM Journal on Numerical Analysis*, 49:970–991, 2011.

[4] F. Boyer, F. Dardalhon, C. Lapuerta, and J.-C. Latché. Stability of a Crank-Nicolson pressure correction scheme based on staggered discretizations. *International Journal for Numerical Methods in Fluids*, 74:34–58, 2014.

[5] CALIF^{3}S. A software components library for the computation of reactive turbulent flows. `https://gforge.irsn.fr/gf/project/isis`.

[6] A.J. Chorin. Numerical solution of the Navier-Stokes equations. *Mathematics of Computation*, 22:745–762, 1968.

[7] P. G. Ciarlet. Basic error estimates for elliptic problems. In P. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis, Volume II*, pages 17–351. North Holland, 1991.

[8] R. Eymard, R. Herbin, J.-C. Latché, and B. Piar. Convergence analysis of a locally stabilized collocated finite volume scheme for incompressible flows. *Mathematical Modelling and Numerical Analysis*, 43:889–927, 2009.

[9] T. Gallouët, L. Gastaldo, R. Herbin, and J.-C. Latché. An unconditionally stable pressure correction scheme for compressible barotropic Navier-Stokes equations. *Mathematical Modelling and Numerical Analysis*, 42:303–331, 2008.

[10] L. Gastaldo, R. Herbin, W. Kheriji, C. Lapuerta, and J.-C. Latché. Staggered discretizations, pressure correction schemes and all speed barotropic flows. In *Finite Volumes for Complex Applications VI - Problems and Perspectives - Prague, Czech Republic*, volume 2, pages 39–56, 2011.

[11] L. Gastaldo, R. Herbin, and J.-C. Latché. An unconditionally stable finite element-finite volume pressure correction scheme for the drift-flux model. *Mathematical Modelling and Numerical Analysis*, 44:251–287, 2010.

[12] L. Gastaldo, R. Herbin, and J.-C. Latché. A discretization of phase mass balance in fractional step algorithms for the drift-flux model. *IMA Journal of Numerical Analysis*, 31:116–146, 2011.

[13] D. Grapsas, R. Herbin, W. Kheriji, and J.-C. Latché. An unconditionally stable finite element-finite volume pressure correction scheme for the compressible navier-stokes equations. *submitted*, 2014.

[14] J.L. Guermond, P. Minev, and J. Shen. An overview of projection methods for incompressible flows. *Computer Methods in Applied Mechanics and Engineering*, 195:6011–6045, 2006.

[15] R. Herbin, W. Kheriji, and J.-C. Latché. Consistent pressure correction staggered schemes for the shallow water and euler equations. *Under revision*, 2013.

[16] R. Herbin, W. Kheriji, and J.-C. Latché. Pressure correction staggered schemes for barotropic one-phase and two-phase flows. *Computers and Fluids*, 88:524–542, 2013.

[17] L.-I.-G. Kovasznay. Laminar flow behind a two-dimensional grid. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44:[58], 1948.

[18] J.-C. Latché and K. Saleh. A convergent staggered scheme for variable density incompressible Navier-Stokes equations. *submitted*, 2014.

[19] P. Le Quéré, C. Weisman, H. Paillère, J. Vierendeels, E. Dick, R. Becker, M. Braack, and J. Locke. Modelling of natural convection flows with large temperature differences: a benchmark problem for low Mach number solvers. Part 1. Reference solutions. *Mathematical Modelling and Numerical Analysis*, 2005.

[20] A. Majda and J. Sethian. The derivation and numerical solution of the equations for zero Mach number solution. *Combustion Science and Techniques*, 42:185–205, 1985.

[21] L.E. Payne and H.F. Weinberger. An optimal Poincaré-inequality for convex domains. *Archive for Rational Mechanics and Analysis*, 5:286–292, 1960.

[22] L. Piar, F. Babik, R. Herbin, and J.-C. Latché. A formally second order cell centered scheme for convection-diffusion equations on general grids. *International Journal for Numerical Methods in Fluids*, 71:873–890, 2013.

[23] R. Rannacher and S. Turek. Simple nonconforming quadrilateral Stokes element. *Numerical Methods for Partial Differential Equations*, 8:97–111, 1992.

[24] F. Schieweck and L. Tobiska. A nonconforming finite element method of upstream type applied to the stationary Navier-Stokes equation. *Mathematical Modelling and Numerical Analysis*, 23:627–647, 1989.

[25] F. Schieweck and L. Tobiska. An optimal order error estimate for an upwind discretization of the Navier-Stokes equations. *Numerical Methods for Partial Differential Equations*, 12:407–421, 1996.

[26] R. Temam. Sur l'approximation de la solution des équations de Navier-Stokes par la méthode des pas fractionnaires II. *Arch. Rat. Mech. Anal.*, 33:377–385, 1969.

[27] S. Xin and P. Le Quéré. Linear stability analyses of natural convection flows in a differentially heated square cavity with conducting horizontal walls. *Physics of Fluids*, 13:2529–2542, 2001.