A combined finite volume - finite element scheme for a low-Mach system involving a Joule term

Caterina Calgaro, Claire Colin, Emmanuel Creusé

A combined finite volume - finite element scheme for a low-Mach system involving a Joule term

Caterina Calgaro*, Claire Colin
Univ. Lille, CNRS, UMR 8524, Inria - Laboratoire Paul Painlevé, F-59000 Lille, France
and
Emmanuel Creusé
Univ. Polytechnique Hauts-de-France, EA 4015, LAMAV-FR CNRS 2956, F-59313 Valenciennes, France

Abstract

In this paper, we propose a combined finite volume - finite element scheme, for the resolution of a specific low-Mach model expressed in the velocity, pressure and temperature variables. The dynamic viscosity of the fluid is given by an explicit function of the temperature, leading to the presence of a so-called Joule term in the mass conservation equation. First, we prove a discrete maximum principle for the temperature. Second, the numerical fluxes defined for the finite volume computation of the temperature are efficiently derived from the discrete finite element velocity field obtained by the resolution of the momentum equation. Several numerical tests are presented to illustrate our theoretical results and to underline the efficiency of the scheme in term of convergence rates.

Keywords: Low Mach Model - Finite volume scheme - Finite element scheme - Joule term - Maximum principle
Mathematics Subject Classification: 35Q30 - 35Q35 - 65M08 - 65M60

1 Introduction

Variable density and low Mach numbers flows have been widely investigated for the last decades. Indeed, they arise in plenty of physical phenomena in which the sound wave speed is much faster than the convective characteristics of the fluid : flows in high-temperature gas reactors, meteorological flows, combustion processes and many others. In this work, we are interested in a specific model derived from the usual low-Mach one for a calorically perfect gas, coming from an asymptotic expansion of the variables with respect to the Mach number $\mathcal{M}$ in the Navier-Stokes compressible equations (see [34]). For the usual low-Mach model, the local-in-time existence of classical solutions in Sobolev spaces is established in [23]. As observed in [33], a small perturbation of a constant initial density provides a global existence of weak solutions in the two-dimensional case. The originality of the model considered here relies on the dynamic viscosity of the fluid being explicitly given as a function of the temperature, as introduced in [4] and generalized in [5]. In this recent work, the authors establish the global existence of weak solutions in the three-dimensional case with no smallness assumption on the initial velocity.

Thanks to a change of variables, we obtain a system in which the velocity field is divergence-free, leading in return to the presence of a non linear and so-called "Joule term" in the mass conservation equation expressed in term of the temperature. In [8], some theoretical results are obtained on the local-in-time existence of strong solutions in the three-dimensional case. We mention also [19] where the authors study the local and global existence in critical Besov spaces, assuming that the initial density is close to a constant and that the initial velocity is small enough. The formulation of this model is close to the so-called ghost effect system, considered in [30, 32], where

*Corresponding author : caterina.calgaro@univ-lille.fr

a thermal stress term is added to the right-hand-side of the momentum equation. The local well-posedness of classical solutions is established in [32] for 2D or 3D unbounded domains. A local well-posedness result for strong solutions is proved in [30], where the authors give also the existence and uniqueness of a global strong solution for the two-dimensional case.

From a numerical point of view, many authors compute flows at low-Mach number regime. We refer only to the so-called pressure-based methods, widely used to compute incompressible flows (see e.g. [1, 2, 20, 29, 31]), but there exists also the so-called density-based methods widely used to compute supersonic or transonic flows, and recently adapted in the case of low-Mach regime (see [27,28]). In previous contributions on incompressible variable density flows [6, 9, 10] or on low-Mach flows with large variations of temperature [7], a combined finite volume - finite element scheme was proposed. Based on a time splitting, this combined method allows to solve the mass conservation equation by a finite volume method, whereas the momentum equation associated with the divergence free constraint and the temperature one are solved by a finite element method. It allows, in particular, to preserve the constant density states and to ensure the discrete maximum principle. In the present work, following the same idea, the nonlinear temperature equation is solved by a finite volume method, whereas the velocity equation associated with the divergence free constraint is solved by a finite element one.

The main contribution of this paper is twofold. First, we prove a discrete maximum principle on the temperature (see Theorem 3.1), similarly to the solution behavior at the continuous level. Second, we establish a footbridge between the finite volume fluxes and the finite element velocity field (see subsection 3.4), to ensure the good consistency of the method.

The outline of the paper is the following. Section 2 briefly introduces the model derivation. Section 3 details the proposed combined finite volume - finite element scheme. The maximum principle is established (Theorem 3.1), some variants of the original scheme are proposed (subsection 3.3), and the link between the finite volume fluxes and the finite element velocity field is carefully explained (subsection 3.4). Finally, section 4 proposes several numerical tests to illustrate the obtained theoretical results, and to investigate the behavior of the scheme on a physical benchmark corresponding to the convection of the temperature in a cavity.

2 Model derivation

As already mentioned in the introduction, a low-Mach model is obtained by inserting the asymptotic expansions of the variables with respect to the Mach number $\mathcal{M}$ in the Navier-Stokes compressible equations (see for example [2, 20, 34]). One of the characteristics of the process is to consider the asymptotic expansion of the pressure π with respect to $\mathcal{M}$. Denoting $\mathbf{x} \in \mathbb{R}^d$ as the space variable and $t \in \mathbb{R}_*^+$ as the time one, we write

$$\pi(\mathbf{x}, t) = P(t) + \mathcal{M}^2 q(\mathbf{x}, t) + o(\mathcal{M}^2),$$

where P is called the thermodynamic pressure and q the dynamic pressure. Here, we assume that $P(t) = P_0 > 0$ is constant for all $t \geq 0$. The other variables considered here are the velocity $\mathbf{u}(\mathbf{x}, t)$, the density $\rho(\mathbf{x}, t)$ and the temperature $\theta(\mathbf{x}, t)$.

Let $\Omega \subset \mathbb{R}^d$ be an open polygonal domain with a boundary Γ and T a positive real. The continuity, momentum,

temperature and state equations in $Q_T = \Omega \times [0; T]$ for a calorically perfect gas are given by:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{2.1a}$$

$$\rho \partial_t \mathbf{u} + \rho \mathbf{u} \cdot \nabla \mathbf{u} + \nabla q - \nabla \cdot \left(\tilde{\mu} \left(2\, \mathbb{D}\mathbf{u} - \frac{2}{3} \nabla \cdot \mathbf{u} \mathrm{I} \right) \right) = \rho\, \mathbf{g}, \tag{2.1b}$$

$$\nabla \cdot \mathbf{u} = \frac{\gamma - 1}{\gamma\, P_0} \nabla \cdot (\kappa\, \nabla \theta), \tag{2.1c}$$

$$P_0 = R\, \rho\, \theta, \tag{2.1d}$$

where $\tilde{\mu}$ is the viscosity of the flow, κ is the heat conductivity which is assumed constant, R is the gas law constant and γ is the gas specific heat ratio. Here, $\mathbb{D}\mathbf{u} = (\nabla \mathbf{u} + \nabla^T \mathbf{u})/2$ denotes the deformation tensor, I the identity matrix and $\mathbf{g}(\mathbf{x}, t)$ the gravity field. In order to reduce the study to a system whose velocity is solenoidal, we define:

$$\mathbf{v} = \mathbf{u} - \lambda \nabla \theta,$$

where $\lambda = \dfrac{(\gamma - 1)\kappa}{\gamma\, P_0} > 0$ is a fixed constant. In addition, the density is eliminated from the equations thanks to the state equation (2.1d). Following the idea introduced in [5] where a particular relation between the density and the viscosity in the combustion model was introduced, we assume moreover that

$$\tilde{\mu}(\theta) = \frac{P_0}{R} \mu(\theta),$$

with

$$\mu(\theta) = -\lambda\, \ln \theta,$$

so that $\mu(\theta)$ is strictly positive if and only if $\theta \in (0; 1)$. If we denote by p the modified pressure, defined by:

$$p = \frac{R}{P_0} q + \lambda^2 \Delta \theta - \frac{2\, \lambda}{3} \mu(\theta)\, \Delta \theta,$$

we consequently obtain (see [8] for all details):

$$\partial_t \theta + \nabla \cdot (\theta\, \mathbf{v}) + 2\lambda |\nabla \theta|^2 - \lambda \nabla \cdot (\theta\, \nabla \theta) = 0, \tag{2.2a}$$

$$\frac{1}{\theta} \left(\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right) - \nabla \cdot (\mu(\theta) \mathbb{D}\mathbf{v}) + \frac{\lambda}{\theta} (\nabla \mathbf{v} - \nabla^T \mathbf{v}) \nabla \theta + \nabla p = \frac{1}{\theta} \mathbf{g}, \tag{2.2b}$$

$$\nabla \cdot \mathbf{v} = 0. \tag{2.2c}$$

The system (2.2) needs to be completed with suitable initial and boundary conditions. We set $\overline{\Gamma} = \overline{\Gamma^D} \cup \overline{\Gamma^N}$, with $\Gamma^D \cap \Gamma^N = \emptyset$. The initial conditions for the system (2.2) are given by:

$$\begin{aligned} \theta(\mathbf{x}, 0) &= \theta_0(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega, \\ \mathbf{v}(\mathbf{x}, 0) &= \mathbf{v}_0(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega, \end{aligned}$$

with:

$$0 < m \leq \min_{\mathbf{x} \in \Omega} \theta_0(\mathbf{x}) \leq \max_{\mathbf{x} \in \Omega} \theta_0(\mathbf{x}) \leq M < 1.$$

The boundary conditions on the temperature and the velocity are given by:

$$\begin{aligned} \nabla \theta(\mathbf{x}, t) \cdot \mathbf{n} &= 0, && \forall \mathbf{x} \in \Gamma^N, && \forall t \in [0; T], \\ \theta(\mathbf{x}, t) &= \theta_D(\mathbf{x}, t), && \forall \mathbf{x} \in \Gamma^D, && \forall t \in [0; T], \\ \mathbf{v}(\mathbf{x}, t) &= \mathbf{v}_D(\mathbf{x}, t), && \forall \mathbf{x} \in \Gamma, && \forall t \in [0; T]. \end{aligned}$$

The local existence of a regular solution to the problem (2.2) has been shown in [8] in the case of dimension $d = 3$, as well as the maximum principle for the temperature. Furthermore, the unique global strong solution can be proved

in the case of dimension $d = 2$ following the proof of Theorem 1.5 in [30] and assuming that for any fixed $\bar{\theta} > 0$, there exists a small constant $\delta(\bar{\theta}) > 0$ such that

$$\|\theta_0 - \bar{\theta}\|_{L^\infty(\Omega)} \leq \delta(\bar{\theta}). \tag{2.3}$$

In this paper, we will prove that there exists at least one solution to the discretized temperature equation, and that this solution verifies the maximum principle. The questions of unicity and convergence of the numerical scheme will be addressed in a further work.

3 A combined finite volume - finite element scheme

3.1 Spatial and temporal discretizations

3.1.1 The time splitting

The numerical scheme is based on a time splitting, allowing to solve the temperature equation with finite volumes, whereas the velocity equation is solved with finite elements, using the same strategy as in previous works [7, 9].

Let Δt be the time step and $t^n = n\,\Delta t$. Functions approximated at time t^n will be identified with superscript n. Assuming that θ^n and $\mathbf{v}^n$ are known:

1. We first compute the new temperature θ^{n+1} by solving the temperature equation (2.2a) using an Euler implicit scheme:

$$\begin{cases} \dfrac{\theta^{n+1} - \theta^n}{\Delta t} + \nabla \cdot (\theta^{n+1}\,\mathbf{v}^n) + 2\lambda |\nabla \theta^{n+1}|^2 - \lambda \nabla \cdot (\theta^{n+1}\,\nabla \theta^{n+1}) = 0, \\ \nabla \theta^{n+1}(\mathbf{x}) \cdot \mathbf{n} = 0, \quad \forall \mathbf{x} \in \Gamma^N, \\ \theta^{n+1}(\mathbf{x}) = \theta_D^{n+1}(\mathbf{x}), \quad \forall \mathbf{x} \in \Gamma^D, \end{cases} \tag{3.1}$$

   with $\theta_D^{n+1}(\mathbf{x}) = \theta_D(\mathbf{x}, t^{n+1})$. The non-linearities of Equation (3.1) are treated thanks to Newton's method.

2. Then we compute the new velocity $\mathbf{v}^{n+1}$ and the new pressure p^{n+1} by solving the momentum equation (2.2b)-(2.2c) by:

$$\begin{cases} \dfrac{1}{\theta^{n+1}} \left(\dfrac{\mathbf{v}^{n+1} - \mathbf{v}^n}{\Delta t} + (\mathbf{v}^n \cdot \nabla)\mathbf{v}^{n+1} \right) - \nabla \cdot (\mu(\theta^{n+1}) \mathbb{D}\mathbf{v}^{n+1}) + \dfrac{\lambda}{\theta^{n+1}} (\nabla \mathbf{v}^{n+1} - \nabla^T \mathbf{v}^{n+1})\, \nabla \theta^{n+1} \\ \qquad\qquad + \nabla p^{n+1} = \dfrac{1}{\theta^{n+1}} \mathbf{g}, \\ \nabla \cdot \mathbf{v}^{n+1} = 0, \\ \mathbf{v}^{n+1}(\mathbf{x}) = \mathbf{v}_D^{n+1}(\mathbf{x}), \quad \forall \mathbf{x} \in \Gamma. \end{cases} \tag{3.2}$$

   Here and as usual, the nonlinear term in the momentum equation (3.2) is linearized.

3.1.2 Description of the mesh and notations

From now, we fix $d = 2$, but our study can be easily extended to the three-dimensional case. The discretization in space is based on a triangulation $\mathcal{T}$ of the domain $\Omega \subset \mathbb{R}^2$ (the elements of $\mathcal{T}$ are called control volumes), a family $\mathcal{E}$ of edges and a set $\mathcal{P} = (\mathbf{x}_K)_{K \in \mathcal{T}}$ of points of Ω defining an admissible mesh in the sense of Definition 3.1 in [25]. We recall that the admissibility of $\mathcal{T}$ implies that the straight line between two neighboring centers of cells $\mathbf{x}_K$ and $\mathbf{x}_L$ is orthogonal to the edge $\sigma \in \mathcal{E}$ such that $\overline{\sigma} = \overline{K} \cap \overline{L}$ (and which is noted $\sigma = K|L$) in a point $\mathbf{x}_\sigma$ (see Figure 1). This condition is satisfied if all the angles of the triangles are less than $\frac{\pi}{2}$.

The set of interior (resp. boundary) edges is denoted by $\mathcal{E}^{\text{int}} = \{\sigma \in \mathcal{E} \ ; \ \sigma \notin \Gamma\}$ (resp. $\mathcal{E}^{\text{ext}} = \{\sigma \in \mathcal{E} \ ; \ \sigma \subset \Gamma\}$). Among the outer edges, there are $\mathcal{E}^N = \left\{\sigma \in \mathcal{E} \ ; \ \sigma \subset \Gamma^N\right\}$ and $\mathcal{E}^D = \left\{\sigma \in \mathcal{E} \ ; \ \sigma \subset \Gamma^D\right\}$. For all $K \in \mathcal{T}$, we denote by $\mathcal{E}_K = \left\{\sigma \in \mathcal{E} \ ; \ \sigma \subset \overline{K}\right\}$ the edges of K, $\mathcal{E}_K^{\text{int}} = \mathcal{E}^{\text{int}} \cap \mathcal{E}_K$, $\mathcal{E}_K^{\text{ext}} = \mathcal{E}^{\text{ext}} \cap \mathcal{E}_K$, $\mathcal{E}_K^N = \mathcal{E}_K^N \cap \mathcal{E}_K$ and $\mathcal{E}_K^D = \mathcal{E}_K^D \cap \mathcal{E}_K$.

The measure of $K \in \mathcal{T}$ is denoted by m_K and the length of σ by m_σ. For $\sigma \in \mathcal{E}^{\text{int}}$ such that $\sigma = K|L$, d_σ denotes the distance between $\mathbf{x}_K$ and $\mathbf{x}_L$ and $d_{K,\sigma}$ the distance between $\mathbf{x}_K$ and σ. For $\sigma \in \mathcal{E}_K^{\text{ext}}$, we note d_σ the distance between $\mathbf{x}_K$ and σ. For $\sigma \in \mathcal{E}$, The transmissibility coefficient is given by $\tau_\sigma = \dfrac{m_\sigma}{d_\sigma}$. Finally, for $\sigma \in \mathcal{E}_K$, we denote by $\mathbf{n}_{K,\sigma}$ the exterior unit normal vector to σ. The size of the mesh is given by:

$$h = \max_{K \in \mathcal{T}} \operatorname{diam}(K).$$

3.1.3 Spatial discretization

The piecewise constant temperature θ_h is computed with a cell-centered finite volume method described in section 3.2, so that $\theta_h \in \mathcal{X}(\mathcal{T})$ with :

$$\mathcal{X}(\mathcal{T}) = \left\{\theta \in L^2(\Omega) \ ; \ \forall K \in \mathcal{T} \ ; \ \theta_K = \theta|_K \in \mathbb{P}_0\right\}.$$

The velocity $\mathbf{v}_h$ is discretized with $\mathbb{P}_2$-Lagrange finite elements, and the pressure p_h with $\mathbb{P}_1$-Lagrange finite elements, so that they satisfy the usual LBB stability condition. The Degrees of Freedom of each variable are shown in the Figure 1.

The computation of the velocity and pressure by finite elements is usual, and we refer to our previous work for details [7,9]. One of the original points of the present work is the computation of the temperature by finite volumes. Indeed, we aim to develop a numerical scheme that ensures the discrete maximum principle property, see section 3.2. Another point of interest is the link between the two numerical methods, see section 3.4. Indeed, from the velocity field that has been computed by finite elements, we have to determine fluxes through the interfaces of the control volumes, which will be used for the computation of the temperature.

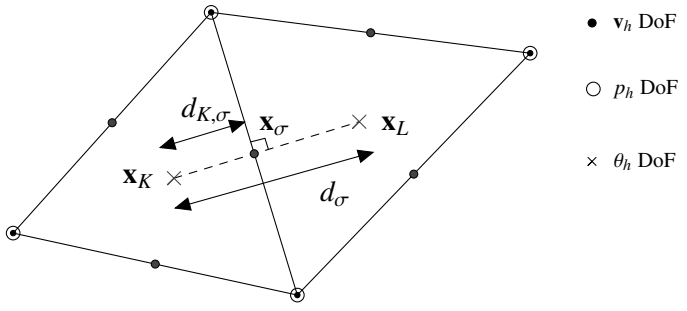


Figure 1: Degrees of Freedom (DoF) for each variable.

3.2 The finite volume scheme

Let $\mathbf{v}$ be a given velocity field and θ^n the previous approximate temperature, we aim at solving the temperature equation:

$$\begin{cases} \dfrac{\theta^{n+1}-\theta^n}{\Delta t} + \nabla\cdot(\theta^{n+1}\,\mathbf{v}) + 2\lambda|\nabla\theta^{n+1}|^2 - \lambda\nabla\cdot(\theta^{n+1}\nabla\theta^{n+1}) = 0, \quad \forall \mathbf{x}\in\Omega, \\ \nabla\theta^{n+1}(\mathbf{x})\cdot\mathbf{n} = 0, \quad \forall \mathbf{x}\in\Gamma^N, \\ \theta^{n+1}(\mathbf{x}) = \theta_D^{n+1}(\mathbf{x}), \quad \forall \mathbf{x}\in\Gamma^D. \end{cases} \tag{3.3}$$

Since we want to ensure maximum principle properties, we favor a finite volume method. The main difficulty comes from the term $|\nabla\theta^{n+1}|^2$, called the Joule term in the context of the electrical conductivity, see for example A. Bradji and R. Herbin [3] or the works from C. Chainais and her collaborators [13, 14]. Indeed, the definition of a discrete gradient is not straightforward, since the finite volume solution is piecewise constant. We can thus refer to the work of R. Eymard and his collaborators [24, 26] for some definitions of discrete gradients on an admissible mesh. Alternatively, K. Domelevo and P. Omnes [21], and C. Chainais [13], used a discrete gradient reconstruction following the idea from the paper of Y. Coudière, J.-P. Vila and P. Villedieu [18]. The principle of these schemes, valid on very general meshes, consists in the double resolution of the equations, on primal and dual meshes. Moreover in [22], J. Droniou and R. Eymard propose a scheme whose unknowns are the function, its gradient and flows. Therefore, the definition of a discrete gradient by mesh is intrinsic to the scheme.

The respect of the bounds, and in particular of the lower bound, is another difficulty related to the Joule term. Indeed, if we consider "close" models, like the equation of porous media, see for example the work of C. Cancès and C. Guichard [12], or the convection-diffusion equation involved in Khazhikhov-Smagulov models-type, see for example C. Calgaro, M. Ezzoug and E. Zahrouni [11], the maximum principle is obtained quite easily for one order schemes. Nevertheless, adding the positive term $|\nabla\theta^{n+1}|^2$ in the temperature equation prevents the scheme from directly obtaining the lower bound. Consequently, we will have to particularly pay attention to the discretization of this term.

We are looking for $\theta_h^{n+1} = (\theta_K^{n+1})_{K\in\mathcal{T}} \in \mathcal{X}(\mathcal{T})$ an approximated solution of $\theta(t^{n+1},.)$. The finite volume scheme is classically obtained by integrating equation (3.3) on a control volume K, that is:

$$m_K\frac{\theta_K^{n+1}-\theta_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma}^n\,\theta_{\sigma,+}^{n+1} + 2\lambda\, m_K\,\mathcal{J}_K(\theta_h^{n+1}) + \lambda\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}^{n+1} = 0 \quad \text{for all } K\in\mathcal{T}, \tag{3.4}$$

with

$$v_{K,\sigma}^n = \int_\sigma \mathbf{v}(\mathbf{x},t^n)\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma(\mathbf{x}). \tag{3.5}$$

Here, $\theta_{\sigma,+}^{n+1}$ is defined for $\sigma\in\mathcal{E}_K$ by:

$$\theta_{\sigma,+}^{n+1} = \begin{cases} \theta_K^{n+1} & \text{if } v_{K,\sigma}^n \geq 0, \\ \theta_{K,\sigma}^{n+1} & \text{otherwise,} \end{cases} \tag{3.6}$$

with

$$\theta_{K,\sigma}^{n+1} = \begin{cases} \theta_L^{n+1} & \text{for } \sigma\in\mathcal{E}_K^{\text{int}} \text{ such that } \sigma = K|L, \\ \theta_D^{n+1}(\mathbf{x}_\sigma) & \text{for } \sigma\in\mathcal{E}_K^D, \\ \theta_K^{n+1} & \text{for } \sigma\in\mathcal{E}_K^N. \end{cases}$$

The numerical flux $F_{K,\sigma}^{n+1}$ is an approximation of the exact flux of the diffusive term through the edge σ and is given by:

$$F_{K,\sigma}^{n+1} = \tau_\sigma\,\theta_\sigma^{n+1}(\theta_K^{n+1}-\theta_{K,\sigma}^{n+1}), \tag{3.7}$$

where we define θ_σ^{n+1}, an approximation of θ^{n+1} at $\mathbf{x}_\sigma$, by:

$$\theta_\sigma^{n+1} = \max(\theta_K^{n+1},\theta_{K,\sigma}^{n+1}). \tag{3.8}$$

Concerning the Joule term, $\mathcal{J}_K(\theta_h^{n+1})$ is an approximation of $\frac{1}{m_K}\int_K |\nabla\theta(t^{n+1},\mathbf{x})|^2 d\mathbf{x}$. We notice that $|\nabla\theta|^2 = \nabla\cdot(\theta\nabla\theta) - \theta\Delta\theta$, and we propose the following definition:

$$\mathcal{J}_K(\theta_h^{n+1}) = \frac{1}{m_K}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\,\overline{\theta}_\sigma^{n+1}\left(\theta_{K,\sigma}^{n+1}-\theta_K^{n+1}\right) - \frac{1}{m_K}\theta_K^{n+1}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma\left(\theta_{K,\sigma}^{n+1}-\theta_K^{n+1}\right),$$

where $\overline{\theta}_\sigma^{n+1}$ is another approximation of θ^{n+1} at $\mathbf{x}_\sigma$ (this choice will be justified later) defined this time by:

$$\overline{\theta}_\sigma^{n+1} = \frac{\theta_K^{n+1}+\theta_{K,\sigma}^{n+1}}{2}. \tag{3.9}$$

Finally we obtain:

$$\mathcal{J}_K(\theta_h^{n+1}) = \frac{1}{2\,m_K}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma(\theta_{K,\sigma}^{n+1}-\theta_K^{n+1})^2. \tag{3.10}$$

The rest of the section is devoted to the proof of the following result:

Theorem 3.1. *We assume that*

$$0 < m \le \theta_K^0 \le M, \quad \forall K\in\mathcal{T}, \tag{3.11}$$

and:

$$m \le \theta_D^n(\boldsymbol{x}_\sigma) \le M, \quad \forall\sigma\in\mathcal{E}^D,\ \forall n = 1,\cdots,N. \tag{3.12}$$

Then the scheme (3.4) admits at least one solution that satisfies:

$$m \le \theta_K^n \le M, \quad \forall K\in\mathcal{T},\ \forall n = 1,\cdots,N. \tag{3.13}$$

Proof. We start by studying the following intermediate scheme:

$$m_K\frac{\theta_K^{n+1}-\theta_K^n}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} v_{K,\sigma}^n\,\theta_{\sigma,+}^{n+1} + 2\lambda\,m_K\,\mathcal{J}_K(\theta_h^{n+1}) + \lambda\sum_{\sigma\in\mathcal{E}_K}\tilde{F}_{K,\sigma}^{n+1} = 0 \quad \text{for all } K\in\mathcal{T}, \tag{3.14}$$

with the modified numerical flux

$$\tilde{F}_{K,\sigma}^{n+1} = \tau_\sigma\,\tilde{\theta}_\sigma^{n+1}(\theta_K^{n+1}-\theta_{K,\sigma}^{n+1}), \tag{3.15}$$

and

$$\tilde{\theta}_\sigma^{n+1} = \max\left(0,\theta_\sigma^{n+1},\theta_\sigma^{n+1} - \left(\min(\theta_K^{n+1},\theta_{K,\sigma}^{n+1}) - m\right)\right), \quad \text{for } \sigma\in\mathcal{E}_K. \tag{3.16}$$

We underline that the definition (3.16) ensures that $\tilde{\theta}_\sigma^{n+1}\ge 0$. With this choice in the scheme (3.14), the diffusion term has been a little "inflated" in order to compensate the Joule term and to increase the stability, as it will be seen later. Note moreover that if θ_h^{n+1} is solution of (3.14) and satisfies (3.13), it is also solution of (3.4).

The schedule of proof is as follows. We first show in Lemma 3.2 that a solution of (3.14) satisfies the discrete maximum principle. Then, by an argument of topological degree, we prove the existence of a solution to (3.14) (see Lemma 3.3). As this solution satisfies the maximum principle, it is also a solution of (3.4) and this concludes the proof of Theorem 3.1.

Lemma 3.2. *We assume that* (3.11) *and* (3.12) *are satisfied. Let θ_h^{n+1} be a solution of* (3.14)*. Then θ_h^{n+1} satisfies* (3.13)*.*

Proof. For $n = 0$, the property (3.13) clearly holds because of (3.11). Suppose now that (3.13) holds at step n and assume that there exists $K_M\in\mathcal{T}$ such that:

$$\theta_{K_M}^{n+1} = \max_{K\in\mathcal{T}}\theta_K^{n+1} > M.$$

We write (3.14) for the control volume K_M and obtain:

$$\sum_{\sigma\in\mathcal{E}_{K_M}} v^n_{K_M,\sigma}\,\theta^{n+1}_{\sigma,+} + \lambda \sum_{\sigma\in\mathcal{E}_{K_M}} \tau_\sigma(\theta^{n+1}_{K_M,\sigma} - \theta^{n+1}_{K_M})^2 + \lambda \sum_{\sigma\in\mathcal{E}_{K_M}} \tau_\sigma\,\tilde{\theta}^{n+1}_\sigma(\theta^{n+1}_{K_M} - \theta^{n+1}_{K_M,\sigma}) = m_{K_M}\frac{\theta^n_{K_M} - \theta^{n+1}_{K_M}}{\Delta t}. \tag{3.17}$$

The right hand side of (3.17) is negative, whereas the left hand side is non-negative (indeed, the treatment of the first term is classical, see for instance [25], and the sign of other terms is obvious). We thus obtain a contradiction.

Assume now that there exists $K_m \in \mathcal{T}$ such that:

$$\theta^{n+1}_{K_m} = \min_{K\in\mathcal{T}} \theta^{n+1}_K < m. \tag{3.18}$$

We write (3.14) on K_m:

$$\sum_{\sigma\in\mathcal{E}_{K_m}} v^n_{K_m,\sigma}\,\theta^{n+1}_{\sigma,+} + \lambda \sum_{\sigma\in\mathcal{E}_{K_m}} \tau_\sigma(\theta^{n+1}_{K_m,\sigma} - \theta^{n+1}_{K_m})^2 + \lambda \sum_{\sigma\in\mathcal{E}_{K_m}} \tau_\sigma\,\tilde{\theta}^{n+1}_\sigma(\theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma}) = m_{K_m}\frac{\theta^n_{K_m} - \theta^{n+1}_{K_m}}{\Delta t}. \tag{3.19}$$

The right hand side of (3.19) is positive. Looking now at the left hand side, we notice that the first (see [25]) and third terms are non positive, whereas the second is non negative. Thus, to obtain a contradiction, we must reach a balance between the Joule term and the diffusive one. By noticing that:

$$\begin{aligned}
&\sum_{\sigma\in\mathcal{E}_{K_m}} \tau_\sigma(\theta^{n+1}_{K_m,\sigma} - \theta^{n+1}_{K_m})^2 + \sum_{\sigma\in\mathcal{E}_{K_m}} \tau_\sigma\,\tilde{\theta}^{n+1}_\sigma(\theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma}) \\
&= \sum_{\sigma\in\mathcal{E}_{K_m}} \tau_\sigma(\theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma} + \tilde{\theta}^{n+1}_\sigma)(\theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma}),
\end{aligned}$$

a sufficient condition to obtain the contradiction is to show that for all $\sigma \in \mathcal{E}_K$, $\theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma} + \tilde{\theta}^{n+1}_\sigma \geq 0$. Indeed it holds, because of (3.18) and Definition (3.16) of $\tilde{\theta}^{n+1}_\sigma$, we have:

$$\theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma} + \tilde{\theta}^{n+1}_\sigma = \theta^{n+1}_{K_m} - \theta^{n+1}_{K_m,\sigma} + \theta^{n+1}_{K_m,\sigma} - (\theta^{n+1}_{K_m} - m) = m \geq 0. \tag{3.20}$$

□

Lemma 3.3. *We assume that* (3.11) *and* (3.12) *are satisfied. Then the scheme* (3.14) *admits at least one solution.*

Proof. Lemma 3.3 is proved thanks to a topological degree argument (see for instance [25]). Let $\mu \in [0,1]$, we define $\theta^{n+1}_{h,\mu} = (\theta^{n+1}_{K,\mu})_{K\in\mathcal{T}}$ as the solution of the scheme: $\forall K \in \mathcal{T}$,

$$\begin{aligned}
&m_K\frac{\theta^{n+1}_{K,\mu} - \theta^n_K}{\Delta t} + \sum_{\sigma\in\mathcal{E}_K} v^n_{K,\sigma}\,\theta^{n+1}_{\sigma,+,\mu} + 2\mu\,\lambda\,m_K\,\mathcal{J}_K(\theta^{n+1}_h) + \mu\,\lambda \sum_{\sigma\in\mathcal{E}_K} \tilde{F}^{n+1}_{K,\sigma,\mu} \\
&+ 2(1-\mu)\lambda\,m_K\,\overline{\mathcal{J}}_K(\theta^{n+1}_{h,\mu}) + (1-\mu)\,\lambda \sum_{\sigma\in\mathcal{E}_K} \overline{F}^{n+1}_{K,\sigma,\mu} = 0,
\end{aligned} \tag{3.21}$$

where $\theta^{n+1}_{\sigma,+,\mu}$ and $\tilde{F}^{n+1}_{K,\sigma,\mu}$ are respectively defined by (3.6) and (3.15) by replacing θ^{n+1}_K and θ^{n+1}_L by $\theta^{n+1}_{K,\mu}$ and $\theta^{n+1}_{L,\mu}$; also $\overline{F}^{n+1}_{K,\sigma,\mu}$ is defined by (3.22) and $\overline{\mathcal{J}}_K(\theta^{n+1}_{h,\mu})$ by (3.23):

$$\overline{F}^{n+1}_{K,\sigma,\mu} = \tau_\sigma\,M\,(\theta^{n+1}_{K,\mu} - \theta^{n+1}_{K,\sigma,\mu}). \tag{3.22}$$

$$\overline{\mathcal{J}}_K(\theta^{n+1}_{h,\mu}) = \frac{1}{2\,m_K}\,(M-m) \sum_{\sigma\in\mathcal{E}_K} \tau_\sigma\,(\theta^{n+1}_{K,\sigma,\mu} - \theta^{n+1}_{K,\mu}). \tag{3.23}$$

Then (3.21) is equivalent to:

$$m_K \frac{\theta_{K,\mu}^{n+1} - \theta_K^n}{\Delta t} + \sum_{\sigma \in \mathcal{E}_K} v_{K,\sigma}^n \, \theta_{\sigma,+,\mu}^{n+1} + 2\mu \, \lambda \, m_K \, \mathcal{J}_K(\theta_{h,\mu}^{n+1}) + \mu \, \lambda \sum_{\sigma \in \mathcal{E}_K} \tilde{F}_{K,\sigma,\mu}^{n+1}$$
$$+ (1-\mu) \, \lambda \, m \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \left(\theta_{K,\mu}^{n+1} - \theta_{K,\sigma,\mu}^{n+1} \right) = 0,$$

and consequently, we can show as in the proof of Lemma 3.2 that for all $K \in \mathcal{T}$,

$$m \leq \theta_{K,\mu}^{n+1} \leq M. \tag{3.24}$$

We now define $\mathcal{K} = [m-1; M+1]^{\#\mathcal{T}}$ a compact of $\mathbb{R}^{\#\mathcal{T}}$, and $\mathcal{H} : [0,1] \times \mathbb{R}^{\#\mathcal{T}} \longrightarrow \mathbb{R}^{\#\mathcal{T}}$ by: $\forall K \in \mathcal{T}$,

$$\mathcal{H}_K(\mu, (\theta_K^{n+1})_{K \in \mathcal{T}}) = m_K \frac{\theta_{K,\mu}^{n+1} - \theta_K^n}{\Delta t} + \sum_{\sigma \in \mathcal{E}_K} v_{K,\sigma}^n \, \theta_{\sigma,+,\mu}^{n+1} + 2\mu \, \lambda \, m_K \, \mathcal{J}_K(\theta_h^{n+1}) + \mu \, \lambda \sum_{\sigma \in \mathcal{E}_K} \tilde{F}_{K,\sigma,\mu}^{n+1}$$
$$+ 2(1-\mu) \lambda \, m_K \, \overline{\mathcal{J}}_K(\theta_{h,\mu}^{n+1}) + (1-\mu) \, \lambda \sum_{\sigma \in \mathcal{E}_K} \overline{F}_{K,\sigma,\mu}^{n+1}.$$

Then, $\mathcal{H} \in C([0,1] \times \mathcal{K}, \mathbb{R}^{\#\mathcal{T}})$ and thanks to (3.24),

$$\mathcal{H}_K(\mu, (\theta_K^{n+1})_{K \in \mathcal{T}}) = 0$$

admits no solution on $\partial\mathcal{K}$. Consequently, its topological degree is independent of μ. As (3.21) admits a solution for $\mu = 0$ its topological degree is different from zero. We can now conclude that (3.21) has a solution for $\mu = 1$, and therefore (3.14) has at least one solution.

□

Lemma 3.2 and Lemma 3.3 conclude the proof of Theorem 3.1 since as already mentioned, if θ_h^{n+1} is solution of (3.14) and satisfies (3.13), it is also solution of (3.4). □

3.3 Variants of the scheme

In this subsection, several variants of the scheme (3.4), denoted in the following ($\mathcal{SD}_{\max}\mathcal{J}_{\text{cen}}$), are presented. The scheme ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}}$) (subsection 3.3.1) will also fulfill the discrete maximum principle without any condition, whereas the scheme ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}}$) (subsection 3.3.2) as well as ($\mathcal{SD}_{\max}\mathcal{J}_{\text{EGH}}$) and ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}}$) (subsection 3.3.3) will fulfill the discrete maximum principle under some restrictions on the temperature bounds m and M.

3.3.1 ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}}$)

We propose here a first variant of ($\mathcal{SD}_{\max}\mathcal{J}_{\text{cen}}$), which also leads to an unconditionally maximum principle. The idea is to consider the scheme (3.4), but with two differences compared to ($\mathcal{SD}_{\max}\mathcal{J}_{\text{cen}}$). On the one hand, the diffusion term is centered, by defining now θ_σ^{n+1} used in the definition of the diffusive flux $F_{K,\sigma}^{n+1}$ in (3.7) by:

$$\theta_\sigma^{n+1} = \frac{\theta_K^{n+1} + \theta_{K,\sigma}^{n+1}}{2} \tag{3.25}$$

instead of (3.8). On the other hand, an upwind in the Joule term gives us

$$\mathcal{J}_K(\theta_h^{n+1}) = \frac{1}{m_K} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \left((\theta_K^{n+1} - \theta_{K,\sigma}^{n+1})^+ \right)^2 \tag{3.26}$$

instead of (3.10), where we use the notation $a^+ = \max(0, a)$. In that case, the maximum principle occurs. Indeed, the proof of Theorem 3.1 can easily be adapted by noticing that $\left((\theta_K^{n+1} - \theta_{K,\sigma}^{n+1})^+ \right)^2 = 0$ if $\theta_K^{n+1} \leq \theta_{K,\sigma}^{n+1}$. Let us note

that the combination of (3.8) for the diffusion flux definition associated to (3.26) for the Joule term flux definition would also lead to a scheme with an unconditional maximum principle. Nevertheless, this choice would generate a loss of accuracy compared with ($\mathcal{SD}_{\max}\mathcal{J}_{\text{cen}}$) and ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}}$) which are both already unconditionally maximum-principle preserving. Since it is useless to upwind both the diffusion term and the Joule one, we did not considered it.

Remark 3.4. *Note that we could also have considered the following definition in the diffusion term:*

$$\theta_\sigma^{n+1} = \begin{cases} \dfrac{d_{L,\sigma}\,\theta_K^{n+1} + d_{K,\sigma}\,\theta_L^{n+1}}{d_\sigma}, & \text{for } \sigma = K|L, \\ \dfrac{\theta_K^{n+1} + \theta_{D,\sigma}^{n+1}}{2}, & \text{for } \sigma \in \mathcal{E}_K^D. \end{cases} \tag{3.27}$$

From the theoretical and numerical points of view, the cases (3.25) *and* (3.27) *give similar results.*

3.3.2 ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}}$)

A quite natural question, in order to increase the accuracy of the approximation, is to investigate the behaviour of the scheme when both flux are centered. Namely, it would consist in considering (3.25) for the definition of the diffusive flux $F_{K,\sigma}^{n+1}$ in (3.7), while using (3.10) for the Joule term flux definition. In that case, the upper bound can be obtained in the same way as in Lemma 3.2. Nevertheless, the obtention of the lower bound needs an additional assumption, so that the maximum principle can be ensured by a specific balance between the Joule term and the diffusion one. The definition of $\tilde{\theta}_\sigma^{n+1}$ in (3.16) has to be a little modified and given by :

$$\tilde{\theta}_\sigma^{n+1} = \max\left(0, \theta_\sigma^{n+1}, \theta_\sigma^{n+1} - \frac{3}{2}\left(\min(\theta_K^{n+1}, \theta_{K,\sigma}^{n+1}) - m\right)\right) \quad \text{for } \sigma \in \mathcal{E}_K,$$

so that the positivity of $\theta_{K_m}^{n+1} - \theta_{K_m,\sigma}^{n+1} + \tilde{\theta}_\sigma^{n+1}$ (see (3.20)) is ensured provided that $M \leq 3m$. As it will be illustrated in the numerical test 4.1 below, such a condition is necessary. This can seem strange from the physical point of view. As we see in the proof, it is necessary for technical reasons, but could also be linked to the fact that the global solution in time of the continuous model is ensured if the initial datum is not too far from a constant state (see (2.3)). Anyway, even with this restriction in mind, ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}}$) could be used for cases where temperature variations are low, to expect a better accuracy of the solution compared to the one obtained by ($\mathcal{SD}_{\max}\mathcal{J}_{\text{cen}}$) or ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}}$).

3.3.3 ($\mathcal{SD}_{\max}\mathcal{J}_{\text{EGH}}$) and ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}}$)

Finally, two other schemes are investigated, considering another way to define the piecewise discrete gradient in each control volume given by (see [26]) :

$$\mathcal{J}_K(\theta_h^{n+1}) = \left(\frac{1}{m_K}\sum_{\sigma\in\mathcal{E}_K}\tau_\sigma(\theta_K^{n+1} - \theta_{K,\sigma}^{n+1})(\mathbf{x}_\sigma - \mathbf{x}_K)\right)^2, \tag{3.28}$$

while considering either (3.8) or (3.25) for the computation of θ_σ^{n+1} arising in the definition of $F_{K,\sigma}^{n+1}$ given by (3.7), and leading respectively to the schemes denoted ($\mathcal{SD}_{\max}\mathcal{J}_{\text{EGH}}$) and ($\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}}$). Once again and in both cases, the maximum principle is ensured under a condition $M \leq C(\mathcal{T})\,m$ with $C(\mathcal{T}) \in (1,2)$, depending on geometrical characteristics of the mesh (see [17]).

In Section 4, we implement these different schemes and compare them on two main criteria: verification of the maximum principle and convergence rates. Table 1 summarizes the five considered schemes.

		Diffusion term $\theta_\sigma^{n+1} =$	
		$\frac{\theta_K^{n+1} + \theta_{K,\sigma}^{n+1}}{2}$	$\max(\theta_K^{n+1}, \theta_{K,\sigma}^{n+1})$
Joule Term	$\frac{1}{2\,m_K} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma (\theta_K^{n+1} - \theta_{K,\sigma}^{n+1})^2$	$(\mathcal{SD}_{\text{moy}} \mathcal{J}_{\text{cen}})$	$(\mathcal{SD}_{\max} \mathcal{J}_{\text{cen}})$
$\mathcal{J}_K(\theta_\mathcal{T}^{n+1}) =$	$\frac{1}{m_K} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma \left((\theta_K^{n+1} - \theta_{K,\sigma}^{n+1})^+ \right)^2$	$(\mathcal{SD}_{\text{moy}} \mathcal{J}_{\text{up}})$	$\times$
	$\left(\frac{1}{m_K} \sum_{\sigma \in \mathcal{E}_K} \tau_\sigma (\theta_K^{n+1} - \theta_{K,\sigma}^{n+1})(\mathbf{x}_\sigma - \mathbf{x}_K) \right)^2$	$(\mathcal{SD}_{\text{moy}} \mathcal{J}_{\text{EGH}})$	$(\mathcal{SD}_{\max} \mathcal{J}_{\text{EGH}})$

Table 1: Summary of considered schemes.

3.4 Coupling between finite elements and finite volumes

The resolution by a finite element method of (3.2) gives us the velocity field $\mathbf{v}_h^{n+1}$ which is $\mathbb{P}_2$ on each triangle $K \in \mathcal{T}$. Let us denote by $(\mathbf{v}_\sigma^{n+1})_{\sigma \in \mathcal{E}}$ the value of this velocity field at the center of edges. Thus, the local divergence constraint reads:

$$\sum_{\sigma \in \mathcal{E}_K} m_\sigma \, \mathbf{v}_\sigma^{n+1} \cdot \mathbf{n}_{K,\sigma} = \alpha_K, \quad \forall K \in \mathcal{T}. \tag{3.29}$$

The sequence of reals $(\alpha_K)_{K \in \mathcal{T}}$ is different from zero in general. Consequently, the velocity field $(\mathbf{v}_\sigma^{n+1})_{\sigma \in \mathcal{E}}$ is not divergence-free in the Finite Volume sense, and can not be used for the resolution of the temperature equation. Here we can not follow the idea of [9], adapted in [7] for the projection method, which consists in defining a constant velocity per triangle. Indeed, the unknowns for the temperature are located at the center of the cells, and no more at the vertices of the mesh.

For $\sigma \in \mathcal{E}^{\text{int}}$, we define arbitrarly K_σ^+ and K_σ^- the two triangles such that $\sigma = K_\sigma^+ | K_\sigma^-$. For $\sigma \in \mathcal{E}^{\text{ext}}$, we denote by K_σ^+ the triangle such that $\sigma \subset \partial K_\sigma^+$. Let $\mathbf{n}_\sigma$ the unit normal vector to σ getting out of K_σ^+, see Figure 2.

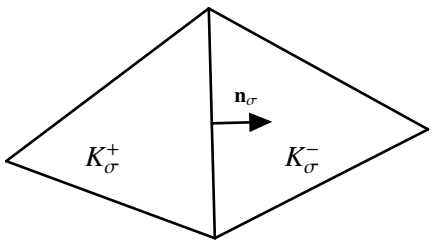


Figure 2: Neighboring triangles of σ and unit normal vector.

We also define $\forall K \in \mathcal{T}$ and $\forall \sigma \in \mathcal{E}_K$:

$$\varepsilon_{K,\sigma} = \begin{cases} 1 & \text{if } K = K_\sigma^+, \\ -1 & \text{if } K = K_\sigma^-. \end{cases}$$

By denoting $(f_\sigma) \in \mathbb{R}^{\#\mathcal{E}}$ the global numerical flux defined by:

$$f_\sigma = m_\sigma \, \mathbf{v}_\sigma^{n+1} \cdot \mathbf{n}_\sigma, \quad \forall \sigma \in \mathcal{E},$$

equation (3.29) can be written as:

$$\sum_{\sigma \in \mathcal{E}_K} \varepsilon_{K,\sigma} \, f_\sigma = \alpha_K, \quad \forall K \in \mathcal{T}. \tag{3.30}$$

We now approximate $(f_\sigma)_{\sigma\in\mathcal{E}}$ by $(\tilde{f}_\sigma)_{\sigma\in\mathcal{E}}$ in order to obtain the local divergence-free constraint

$$\sum_{\sigma \in \mathcal{E}_K} \varepsilon_{K,\sigma} \, \tilde{f}_\sigma = 0, \quad \forall K \in \mathcal{T}. \tag{3.31}$$

Then the fluxes $(\tilde{f}_\sigma)_{\sigma\in\mathcal{E}}$ are used in the cell-centered Finite Volume scheme (3.4) for the computation of the temperature field, by setting:

$$v_{K,\sigma}^{n+1} = \varepsilon_{K,\sigma} \, \tilde{f}_\sigma.$$

Practically, $(\tilde{f}_\sigma)_{\sigma\in\mathcal{E}}$ is computed as an approximation of $(f_\sigma)_{\sigma\in\mathcal{E}}$ in the discrete least-mean-squares sense, which fulfills the divergence-free constraint on each finite volume control. It consists in solving the global optimization problem given by:

$$(\tilde{f}_\sigma) = \underset{(\tilde{h}_\sigma)\in\mathbb{R}^{\#\mathcal{E}}}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{\sigma \in \mathcal{E}^{\text{int}}} w_\sigma \left(f_\sigma - \tilde{h}_\sigma \right)^2 \quad ; \quad \forall K \in \mathcal{T}, \quad \sum_{\sigma \in \mathcal{E}_K} \varepsilon_{K,\sigma} \, \tilde{h}_\sigma = 0 \right\}, \tag{3.32}$$

where $(w_\sigma)_{\sigma\in\mathcal{E}}$ is a sequence of strictly positive weights, which can be defined for example by $w_\sigma = 1$ or $w_\sigma = \frac{m_\sigma}{d_\sigma}$, $\forall \sigma \in \mathcal{E}$. Numerically, we observed similar behaviors for both possibilities. The solution of the problem (3.32) is given by the following theorem:

Theorem 3.5. *Let $\mathbf{M} \in \mathbb{R}^{\#\mathcal{T}\times\#\mathcal{T}}$ the M-matrix defined $\forall K, L \in \mathcal{T}$ by:*

$$\mathbf{M}_{KL} = \begin{cases} \displaystyle\sum_{\sigma \in \mathcal{E}_K} \frac{1}{w_\sigma} & \text{if } K = L, \\ \displaystyle -\frac{1}{w_\sigma} & \text{si } \overline{K} \cap \overline{L} = \sigma, \\ 0 & \text{otherwise,} \end{cases}$$

and $A = (\alpha_K)_{K\in\mathcal{T}} \in \mathbb{R}^{\#\mathcal{T}}$. Let $\Lambda = (\lambda_K)_{K\in\mathcal{T}} \in \mathbb{R}^{\#\mathcal{T}}$ be the solution of:

$$\mathbf{M}\,\Lambda = A. \tag{3.33}$$

Thus the solution of (3.32) *can be defined $\forall \sigma \in \mathcal{E}$ by:*

$$\tilde{f}_\sigma = f_\sigma - \frac{1}{w_\sigma} \left(\lambda_{K_\sigma^+} - \lambda_{K_\sigma^-} \right), \tag{3.34}$$

where for all $\sigma \in \mathcal{E}^{\text{ext}}$ such that $\sigma \subset \partial K$, we set $\lambda_{K_\sigma^-} = 0$.

Proof. With the following change of variables:

$$g_\sigma = \tilde{f}_\sigma - f_\sigma, \tag{3.35}$$

problem (3.32) writes:

$$(g_\sigma) = \underset{(h_\sigma)\in\mathbb{R}^{\#\mathcal{E}}}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{\sigma \in \mathcal{E}} w_\sigma h_\sigma^2 \quad ; \quad \forall K \in \mathcal{T}, \quad \sum_{\sigma \in \mathcal{E}_K} \varepsilon_{K,\sigma} \, h_\sigma + \alpha_K = 0 \right\}. \tag{3.36}$$

We define the lagrangian associated with (3.36) for $(h_\sigma) \in \mathbb{R}^{\#\mathcal{E}}$ and $(\mu_K) \in \mathbb{R}^{\#\mathcal{T}}$ by:

$$\mathcal{L}((h_\sigma),(\mu_K)) = \frac{1}{2}\sum_{\sigma\in\mathcal{E}} w_\sigma h_\sigma^2 + \sum_{K\in\mathcal{T}} \mu_K \left(\sum_{\sigma\in\mathcal{E}_K} \varepsilon_{K,\sigma}\, h_\sigma + \alpha_K \right).$$

We will show the existence of a saddle point of $\mathcal{L}$. Its first argument will therefore be the solution of the problem (3.36), while the second one corresponds to the Lagrange multiplier associated with the constraints, see for example [16]. We recall that if $((g_\sigma),(\lambda_K))$ is a saddle point of $\mathcal{L}$, then:

$$\sup_{(\mu_K)} \inf_{(h_\sigma)} \mathcal{L}((h_\sigma),(\mu_K)) = \mathcal{L}((g_\sigma),(\lambda_K)) = \inf_{(h_\sigma)} \sup_{(\mu_K)} \mathcal{L}((h_\sigma),(\mu_K)).$$

We start with computing $\mathcal{H}((\mu_K)) = \inf_{(h_\sigma)} \mathcal{L}((h_\sigma),(\mu_K))$. We easily verify that (g_σ) (which depends on (μ_K)) defined by:

$$g_\sigma = -\frac{1}{w_\sigma}\left(\mu_{K_\sigma^+} - \mu_{K_\sigma^-}\right), \quad \forall \sigma \in \mathcal{E} \tag{3.37}$$

is solution of

$$\frac{\partial \mathcal{L}}{\partial h_\sigma}((g_\sigma),(\mu_K)) = 0.$$

We thus obtain:

$$\begin{aligned}
\mathcal{H}((\mu_K)) &= \frac{1}{2}\sum_{\sigma\in\mathcal{E}} \frac{1}{w_\sigma}\left(\mu_{K_\sigma^+} - \mu_{K_\sigma^-}\right)^2 + \sum_{K\in\mathcal{T}} \mu_K \left(\sum_{\sigma\in\mathcal{E}_K} -\frac{\varepsilon_{K,\sigma}}{w_\sigma}\left(\mu_{K_\sigma^+} - \mu_{K_\sigma^-}\right) + \alpha_K \right) \\
&= -\frac{1}{2}\sum_{\sigma\in\mathcal{E}} \frac{1}{w_\sigma}\left(\mu_{K_\sigma^+} - \mu_{K_\sigma^-}\right)^2 + \sum_{K\in\mathcal{T}} \mu_K\, \alpha_K \\
&= -\frac{1}{2}\sum_{\sigma\in\mathcal{E}} \frac{1}{w_\sigma}\left(\mu_K - \mu_{K,\sigma}\right)^2 + \sum_{K\in\mathcal{T}} \mu_K\, \alpha_K,
\end{aligned}$$

with the notation

$$\mu_{K,\sigma} = \begin{cases} \mu_L & \text{if } \sigma \in \mathcal{E}_K^{\text{int}}, \quad \sigma = K|L, \\ 0 & \text{if } \sigma \in \mathcal{E}_K^{\text{ext}}. \end{cases} \tag{3.38}$$

We now compute $(\lambda_K) = \underset{(\mu_K)}{\text{argmax}}\, \mathcal{H}((\mu_K))$ by solving:

$$\frac{\partial(-\mathcal{H})}{\partial \mu_K}((\lambda_K)) = 0, \quad \forall K \in \mathcal{T},$$

which is equivalent, applying the notation (3.38) to $\lambda_{K,\sigma}$, to

$$\sum_{\sigma\in\mathcal{E}_K} \frac{1}{w_\sigma}\left(\lambda_K - \lambda_{K,\sigma}\right) - \alpha_K = 0.$$

Thus, (λ_K) is obtained by solving the linear system (3.33). We therefore deduce the expression of (g_σ) thanks to (3.37) and then $(\tilde{f}_\sigma)$ with (3.35).

□

4 Numerical simulations

4.1 Verification of the maximum principle

We previously proved that the schemes $(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{cen}})$ and $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}})$ satisfy the maximum principle, without any restriction on m and M. This first experiment illustrates that if M is too large compared to m, the other schemes do not respect the maximum principle. In this perspective, we consider only the temperature equation (2.2a) and set:

$$\mathbf{v}(\mathbf{x},t) = \mathbf{0}, \quad \forall \mathbf{x} \in \Omega, \quad \forall t \in [0,T].$$

The initial temperature is defined on $\Omega = [0;1]^2$ by:

$$\theta_0(\mathbf{x}) = 1, \quad \forall \mathbf{x} \in \Omega.$$

On all the boundaries, we impose Dirichlet conditions, see Figure 3. We denote by $\Gamma_H = \{1\} \times (0,3;0,7)$ and set $\forall t \in [0;T]$:

$$\theta_D(\mathbf{x},t) = \begin{cases} 1 & \text{if } \mathbf{x} \in \Gamma \backslash \Gamma_H \\ M & \text{if } \mathbf{x} \in \Gamma_H. \end{cases}$$

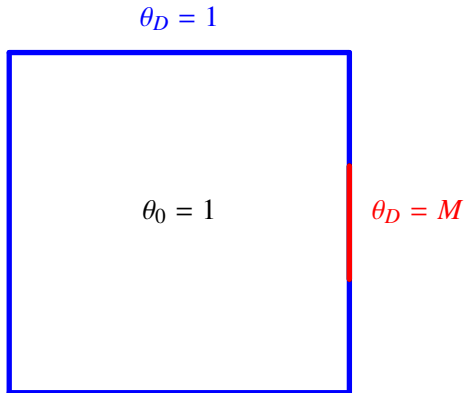


Figure 3: Initial and boundary conditions.

We run the simulations for different values of M on a triangulation $\mathcal{T}$ with mesh size $h = 7.25 \cdot 10^{-2}$ and report the results in the Table 2. We find numerically the results shown previously. On the one hand, the schemes

$M =$ \ Scheme	$(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}})$	$(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{EGH}})$	$(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}})$	$(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{cen}})$	$(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}})$
2	✓	✓	✓	✓	✓
3	×	✓	✓	✓	✓
4	×	✓	×	✓	✓
10	×	×	×	✓	✓
100	×	×	×	✓	✓

Table 2: Verification of the maximum principle according to M. ✓ : the maximum principle is satisfied, × :it is not.

$(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{cen}})$ and $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}})$ allow us to obtain the maximum principle whatever the value of M. On the other hand, the schemes $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}})$, $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}})$ and $(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{EGH}})$ do not give a solution that satisfies the maximum principle if M is too large.

4.2 Analytical benchmark

In this benchmark, we want to investigate the accuracy of the scheme described in section 3.2, depending on the discretization of the diffusive and Joule terms. We consider the domain $\Omega = [0;1]^2$. The simulations will be performed until the time $T = 0.1$. The exact solution is defined for $(x,y,t) \in \Omega \times [0;T]$ by:

$$\theta_{\text{ex}}(x,y,t) = \frac{1+t}{10\,\lambda\,\pi^2}\,(1,5+\cos\pi x)\,(1,5+\cos\pi y)\,,$$

with $\lambda = 2$. In (3.3), a source term is added accordingly. For the velocity, two cases are considered:

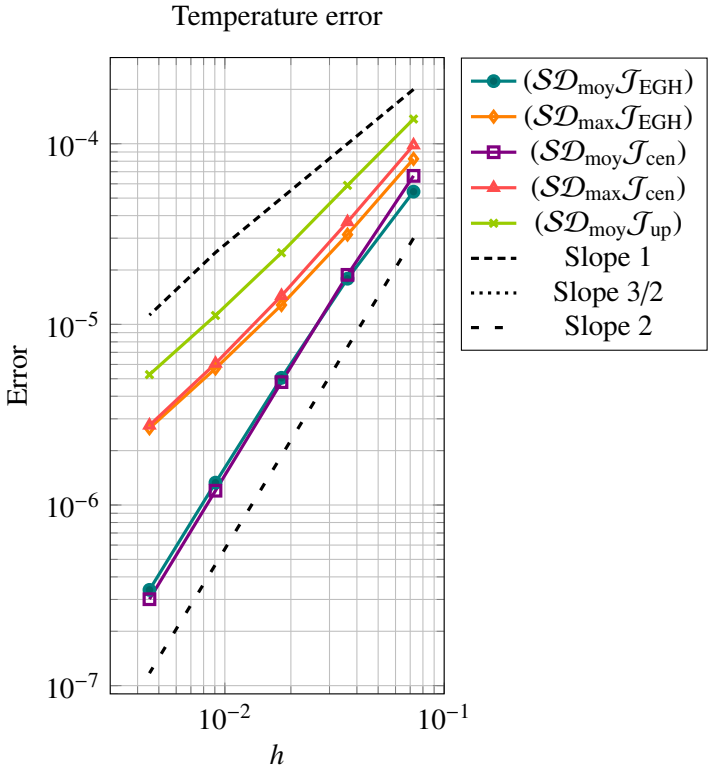
a) $\mathbf{v} = \mathbf{0}$,

b) $\mathbf{v}$ defined by

$$\mathbf{v}(x, y) = \frac{1}{\lambda \pi^2} \begin{pmatrix} -\dfrac{\sin(\pi y)}{1,5 + \cos(\pi y)} \\ \dfrac{\sin(\pi x)}{1,5 + \cos(\pi x)} \end{pmatrix}. \tag{4.1}$$

Here, Dirichlet conditions are imposed on all the boundaries, so that $\Gamma_D = \Gamma$. The temperature error in $L^\infty(0, T; L^2(\Omega))$ is plotted in Figure 4 as a function of the mesh size h, in log-log scale, for each scheme. We notice that without the convective term (case a)), the centered schemes $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}})$ and $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}})$ converge at order two, whereas the others are only at order one. Thus, the upwind choice in the diffusion term (schemes $(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{cen}})$ and $(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{EGH}})$) or in the Joule term (scheme $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{up}})$ degrades the rate of convergence to order one, but is necessary to obtain the maximum principle. Adding the convective term (case b)), the centered schemes $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{cen}})$ and $(\mathcal{SD}_{\text{moy}}\mathcal{J}_{\text{EGH}})$ converge at order 3/2. This decrease of convergence rate can be explained by the upwind treatment of the convective term in order to ensure the stability.

a) $\mathbf{v} = \mathbf{0}$



b) $\mathbf{v}$ given by (4.1)

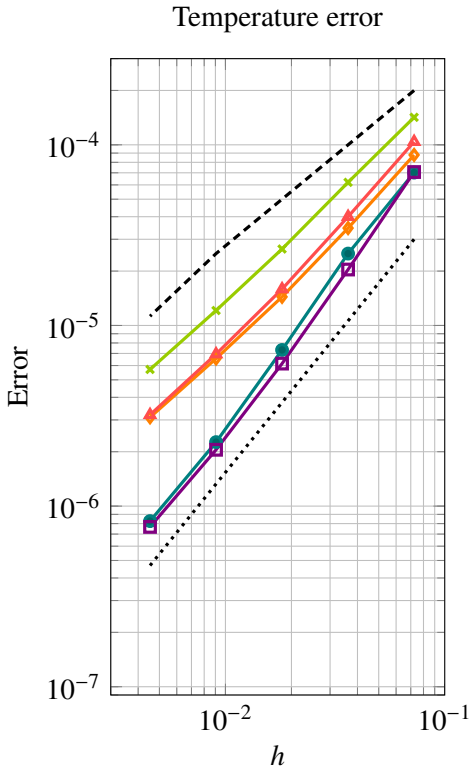


Figure 4: Errors in $L^\infty(0, T; L^2(\Omega))$ norm.

4.3 The natural convection in a cavity

We will now validate the scheme $(\mathcal{SD}_{\text{max}}\mathcal{J}_{\text{cen}})$ by coupling the temperature equation with the velocity one on the benchmark introduced in [1, 15, 29, 31] and used in [7] (slightly adapted because of the choice of the model (2.2)). We consider a square cavity $\Omega = [0, 1]^2$ containing a calorifically perfect gas, see Figure 5. The gas is initially at rest with uniform temperature:

$$\mathbf{u}_0 = \mathbf{0} \text{ and } \theta_0 = 0.5.$$

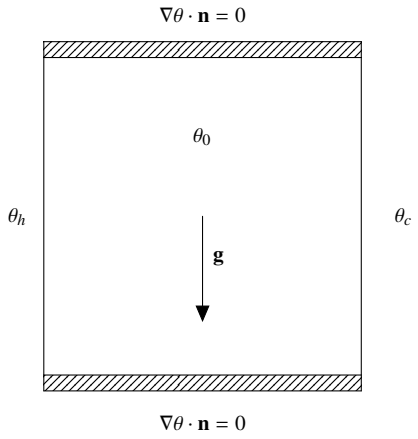


Figure 5: The differentially heated cavity.

Note that the temperature has been scaled to be between 0 and 1. A temperature of $\theta_h = \theta_0(1 + \varepsilon)$ (respectively $\theta_c = \theta_0(1 - \varepsilon)$) is imposed on the left (respectively right) wall with $\varepsilon = 0.01$ as in [29]. For this small temperature amplitude, the thermodynamic pressure can be approximated by a constant, where the author verifies numerically that $P(T) = P_0$ with an accuracy of 10^{-5}). The horizontal walls are insulated. Denoting by $\Gamma_N = [0, 1] \times \{0, 1\}$, we thus have:

$$\nabla\theta(\cdot, t) \cdot \mathbf{n}|_{\Gamma_N} \quad = \quad 0, \quad \forall t \in [0; T].$$

On all walls, the no-slip condition is imposed for the physical velocity $\mathbf{u}$, which gives for the solenoidal one:

$$\begin{aligned} &\mathbf{v}(\mathbf{x}, t) = \mathbf{0}, \quad \forall \mathbf{x} \in \Gamma_N, \quad \forall t \in [0; T], \\ &\mathbf{v}(\mathbf{x}, t) = \lambda \nabla\theta(\mathbf{x}, t), \quad \forall \mathbf{x} \in \Gamma_D, \quad \forall t \in [0; T]. \end{aligned}$$

We set $\lambda = 4,76 \cdot 10^{-2}$, which corresponds to a Rayleigh number of 10^{-5}, as it is the case in [31].

The time iterations are performed until the numerical steady state is reached, i.e. the relative errors for the solenoidal velocity and the temperature are less than 10^{-10}. The steady state is obtained approximatively at $T = 100$, using for instance a mesh $\mathcal{T}$ composed by 3584 triangles (corresponding to $h = 1.8 \cdot 10^{-2}$) and a time step Δt of the order of 10^{-2}. We represent in Figure 6 the streamlines of the velocity field and the contour lines of the temperature at steady state. The solutions obtained are close to the ones given in Figures 4 and 5 of [29]. We also verify that the maximum principle for the temperature is always respected.

Note that even though we have to solve a global optimization problem in order to obtain the fluxes used in the finite volume scheme, its cost is negligible. Indeed, the matrix of the linear system (3.33) is assembled a single time before the time loop. On the contrary, some matrices in the finite element step must be assembled at each time step. Note moreover that the cost of Newton's iterations for the temperature equation resolution is also negligible. Indeed, only two or three iterations are necessary in order to obtain the solution with an accuracy of 10^{-5}. With this choice, approximatively only 10% of the overall computation time is devoted to the resolution of the temperature equation. This ratio in the computation cost is quite the same when we consider more coarse or more fine meshes.

5 Conclusion

In this work we propose some finite volume schemes for the resolution of an unsteady convection-diffusion equation involving a Joule term. Several variants are investigated, depending on the way to discretize the diffusion term as

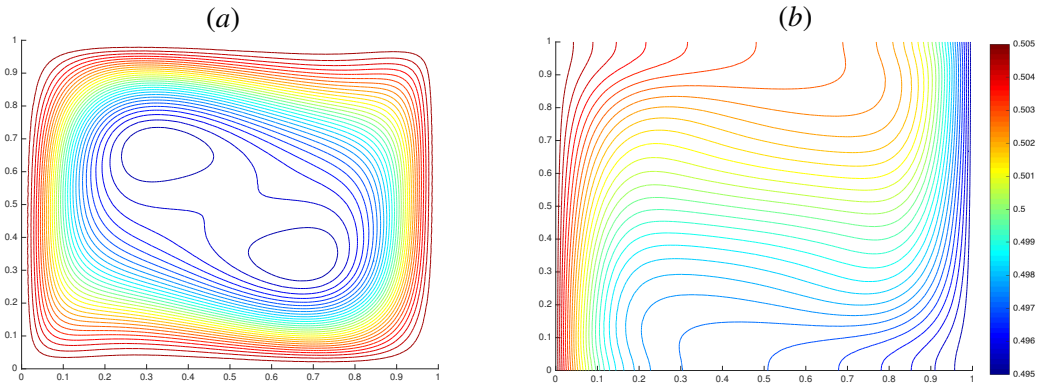


Figure 6: Stationnary solution. (*a*) Velocity, (*b*) Temperature.

well as the Joule one. A first main point is that two of these schemes verify the discrete maximum principle without any restriction on the data. Such schemes allow us to define a combined finite volume - finite element scheme for a low-Mach system, where the temperature is computed by the finite volume scheme whereas the velocity and pressure are approximated by a finite element one. The second point is that the finite volume velocity field, used for the convective term in the temperature equation, is computed from the finite element one by the resolution of a discrete least-mean-squares problem to ensure the local free divergence constraint. The numerical results illustrate the properties of the different schemes and confirm the relevance of the approach. The question of the convergence remains to be investigated, and will be addressed in a future work.

Acknowledgments

This work was supported in part by the Labex CEMPI (ANR-11-LABX-0007-01). The authors gratefully acknowledge Clément Cancès, Claire Chainais-Hillairet and Benoit Merlet for fruitful discussions about it.

References

[1] M. Avila, J. Principe, and R. Codina. A finite element dynamical nonlinear subscale approximation for the low Mach number flow equations. *J. Comput. Phys.*, 230(22):7988–8009, 2011.

[2] A. Beccantini, E. Studer, S. Gounand, J.-P. Magnaud, T. Kloczko, C. Corre, and S. Kudriakov. Numerical simulations of a transient injection flow at low Mach number regime. *Internat. J. Numer. Methods Engrg.*, 76(5):662–696, 2008.

[3] A. Bradji and R. Herbin. Discretization of coupled heat and electrical diffusion problems by finite-element and finite-volume methods. *IMA J. Numer. Anal.*, 28(3):469–495, 2008.

[4] D. Bresch, E. H. Essoufi, and M. Sy. Effect of density dependent viscosities on multiphasic incompressible fluid models. *J. Math. Fluid Mech.*, 9(3):377–397, 2007.

[5] D. Bresch, V. Giovangigli, and E. Zatorska. Two-velocity hydrodynamics in fluid mechanics: Part I. Well posedness for zero Mach number systems. *J. Math. Pures Appl. (9)*, 104(4):762–800, 2015.

[6] C. Calgaro, E. Chane-Kane, E. Creusé, and T. Goudon. L^∞-stability of vertex-based MUSCL finite volume schemes on unstructured grids: simulation of incompressible flows with high density ratios. *J. Comput. Phys.*, 229(17):6027–6046, 2010.

[7] C. Calgaro, C. Colin, and E. Creusé. A combined finite volumes - finite elements method for a low-Mach model. *Int. J. Numer. Methods Fluids*, 90(1):1–21, 2019.

[8] C. Calgaro, C. Colin, E. Creusé, and E. Zahrouni. Approximation by an iterative method of a low-Mach model with temperature dependant viscosity. *Math. Methods Appl. Sci.*, 42:250–271, 2019.

[9] C. Calgaro, E. Creusé, and T. Goudon. An hybrid finite volume-finite element method for variable density incompressible flows. *J. Comput. Phys.*, 227(9):4671–4696, 2008.

[10] C. Calgaro, E. Creusé, and T. Goudon. Modeling and simulation of mixture flows: application to powder-snow avalanches. *Comput. & Fluids*, 107:100–122, 2015.

[11] C. Calgaro, M. Ezzoug, and E. Zahrouni. Stability and convergence of an hybrid finite volume-finite element method for a multiphasic incompressible fluid model. *Commun. Pure Appl. Anal.*, 17(2):429–448, 2018.

[12] C. Cancès and C. Guichard. Convergence of a nonlinear entropy diminishing control volume finite element scheme for solving anisotropic degenerate parabolic equations. *Math. Comp.*, 85(298):549–580, 2016.

[13] C. Chainais-Hillairet. Discrete duality finite volume schemes for two-dimensional drift-diffusion and energy-transport models. *Internat. J. Numer. Methods Fluids*, 59(3):239–257, 2009.

[14] C. Chainais-Hillairet, Y.-J. Peng, and I. Violet. Numerical solutions of Euler-Poisson systems for potential flows. *Appl. Numer. Math.*, 59(2):301–315, 2009.

[15] D. R. Chenoweth and S. Paolucci. Natural convection in an enclosed vertical air layer with large horizontal temperature differences. *J. Fluid Mech*, 169:173–210, 1986.

[16] P. G. Ciarlet. *Introduction to numerical linear algebra and optimisation*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 1989. With the assistance of Bernadette Miara and Jean-Marie Thomas, Translated from the French by A. Buttigieg.

[17] C. Colin. *Analyse et simulation numérique par méthode combinée Volumes Finis - Eléments Finis de modèles de type Faible Mach*. PhD thesis, Université de Lille, 2019.

[18] Y. Coudière, J.-P. Vila, and P. Villedieu. Convergence rate of a finite volume scheme for a two-dimensional convection-diffusion problem. *M2AN Math. Model. Numer. Anal.*, 33(3):493–516, 1999.

[19] R. Danchin and X. Liao. On the well-posedness of the full low Mach number limit system in general critical Besov spaces. *Commun. Contemp. Math.*, 14(3):1250022, 47, 2012.

[20] S. Dellacherie. On a diphasic low Mach number system. *M2AN Math. Model. Numer. Anal.*, 39(3):487–514, 2005.

[21] K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN Math. Model. Numer. Anal.*, 39(6):1203–1249, 2005.

[22] J. Droniou and R. Eymard. A mixed finite volume scheme for anisotropic diffusion problems on any grid. *Numer. Math.*, 105(1):35–71, 2006.

[23] P. Embid. Well-posedness of the nonlinear equations for zero Mach number combustion. *Comm. Partial Differential Equations*, 12(11):1227–1283, 1987.

[24] R. Eymard and T. Gallouët. H-convergence and numerical schemes for elliptic equations. *SIAM J. Numer. Anal.*, (3):539–562, 2000.

[25] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.

[26] R. Eymard, T. Gallouët, and R. Herbin. A cell-centered finite-volume approximation for anisotropic diffusion operators on unstructured meshes in any space dimension. *IMA J. Numer. Anal.*, 26(2):326–353, 2006.

[27] R. Herbin, J.-C. Latché, and K. Saleh. Low Mach number limit of a pressure correction MAC scheme for compressible barotropic flows. In *Finite volumes for complex applications VIII—methods and theoretical aspects*, volume 199 of *Springer Proc. Math. Stat.*, pages 255–263. Springer, Cham, 2017.

[28] R. Herbin, J.-C. Latché, and K. Saleh. Low Mach number limit of some staggered schemes for compressible barotropic flows. *submitted*, 2019.

[29] V. Heuveline. On higher-order mixed FEM for low Mach number flows: application to a natural convection benchmark problem. *Internat. J. Numer. Methods Fluids*, 41(12):1339–1356, 2003.

[30] F. Huang and W. Tan. On the strong solution of the ghost effect system. *SIAM J. Math. Anal.*, 49(5):3496–3526, 2017.

[31] 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. I. Reference solutions. *M2AN Math. Model. Numer. Anal.*, 39(3):609–616, 2005.

[32] C.D. Levermore, W. Sun, and K. Trivisa. Local well-posedness of a ghost system effect. *Indiana Univ. Math. J.*, 60:517–576, 2011.

[33] P.-L Lions. *Mathematical topics in fluid mechanics. Vol. 2*, volume 10 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1998. Compressible models, Oxford Science Publications.

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