UNIVERSITÉ DE PROVENCE
U.F.R. de Mathématiques, Informatique et Mécanique
ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE
E.D numéro 184

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE PROVENCE

Discipline : MATHÉMATIQUES
Option : ANALYSE

présentée et soutenue publiquement

par

Walid KHERIJI

le 03 Novembre 2011

Titre :

MÉTHODES DE CORRECTION DE PRESSION POUR LES ÉQUATIONS DE NAVIER-STOKES COMPRESSIBLES

Directeur de thèse :
Pr. Raphaèle HERBIN

JURY

M. Frédéric COQUEL		Rapporteur
M. Hervé GUILLARD	INRIA, Sophia Antipolis	Rapporteur
M. Jean-Marc HÈRARD	EDF, Chatou	Examinateur
Mme Florence HUBERT	Maître de Conférences, Université de Provence	Examinateur
Mme Raphaèle HERBIN	Professeur, Université de Provence	Directeur de thèse
M. Jean-Claude LATCHÈ	Ingénieur de Recherche, IRSN, Cadarache	Encadrant

UNIVERSITÉ DE PROVENCE
U.F.R. de Mathématiques, Informatique et Mécanique
ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE
E.D numéro 184

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE PROVENCE

Discipline : MATHÉMATIQUES
Option : ANALYSE

présentée et soutenue publiquement

par

Walid KHERIJI

le 03 Novembre 2011

Titre :

MÉTHODES DE CORRECTION DE PRESSION POUR LES ÉQUATIONS DE NAVIER-STOKES COMPRESSIBLES

Directeur de thèse :
Pr. Raphaèle HERBIN

JURY

M. Frédéric COQUEL		Rapporteur
M. Hervé GUILLARD	INRIA, Sophia Antipolis	Rapporteur
M. Jean-Marc HÈRARD	EDF, Chatou	Examinateur
Mme Florence HUBERT	Maître de Conférences, Université de Provence	Examinateur
Mme Raphaèle HERBIN	Professeur, Université de Provence	Directeur de thèse
M. Jean-Claude LATCHÈ	Ingénieur de Recherche, IRSN, Cadarache	Encadrant

Table des matières

General synthesis

1 Introduction 1

2 Meshes and unknowns 3

3 Compressible barotropic Navier-Stokes equations 5

3.1 An implicit scheme scheme 7

3.2 A pressure correction scheme 13

4 Compressible Navier-Stokes equations 16

5 Euler equations 19

5.1 The discrete kinetic energy balance equation and the corrective source terms . . . 20

5.2 Passing to the limit in the scheme 22

5.3 Numerical tests 23

6 Conclusion and perspectives 24

7 APPENDIX – Some results associated to finite volume convection operators 26

Chapitre I
Pressure correction staggered schemes for barotropic monophasic and two-phase flows

I.1 Introduction 33

I.2 The pressure correction scheme 35

I.2.1 Time semi-discretization 35

I.2.2 Discrete spaces and unknowns 36

I.2.3 Discrete equations 38

I.3 Numerical experiments : Riemann problems 41

I.3.1 Sod shock tube problem 42
I.3.2 Two-fluid shock tube 48
I.4 A two-dimensional test case 57
I.5 Conclusion 61

Chapitre II
Consistent staggered schemes for compressible flows – Part I : barotopic Navier-Stokes equations.

II.1 Introduction 65
II.2 Meshes and unknowns 65
II.3 Some results associated to finite volume convection operators 66
II.4 An implicit scheme 69
II.4.1 The scheme 69
II.4.2 Estimates 71
II.4.3 Passing to the limit in the scheme 74
II.5 Pressure correction scheme 80
II.5.1 The scheme 80
II.5.2 Estimates 81
II.5.3 Passing to the limit in the scheme 83

Chapitre III
An unconditionally stable pressure correction scheme for compressible Navier-Stokes equations

III.1 Introduction 89
III.2 Meshes and unknowns 90
III.3 The time-implicit numerical scheme 92
III.3.1 Semi-discrete algorithm 92
III.3.2 The fully discrete algorithm and its first properties 93
III.4 Pressure correction scheme 99
III.4.1 Semi-discrete algorithm 99
III.4.2 Discrete algorithm 100
III.4.3 Properties of the scheme 101
III.5 Appendix : the MAC discretization of the dissipation term 103
III.5.1 The two-dimensional case 103
III.5.2 Extension to the three-dimensional case 106

Chapitre IV
Consistent staggered schemes for compressible flows – Part II : Euler equations.

IV.1 Introduction 111
IV.2 Meshes and unknowns 112
IV.3 An implicit scheme 113
IV.3.1 The scheme 113

IV.3.2 The discrete kinetic energy balance equation and the corrective source terms . . . 116
IV.3.3 Passing to the limit in the scheme 119
IV.4 A pressure correction scheme 125
IV.4.1 The scheme 125
IV.4.2 The discrete kinetic energy balance equation and the corrective source terms . . . 126
IV.4.3 Passing to the limit in the scheme 128
IV.5 Numerical tests 130

Annexe A The Riemann problem for the homegeneous model
A.1 The system of conservation laws and its mathematical properties 137
A.2 Solution of the Riemann problem 141
A.2.1 Sod shock tube 143
A.2.2 Two-fluid shock tube 144

Annexes

Annexe B Staggered discretizations, pressure corrections schemes and all speed barotropic flows 147
B.1 Introduction 149
B.2 The schemes : general form 149
B.2.1 Meshes and unknowns 149
B.2.2 The schemes 151
B.3 The stability issue and consequences 153
B.3.1 A stability result for the convection 153
B.3.2 Discretization of the convection term 154
B.4 Schemes and stability estimates 157
B.5 Euler equations and solutions with shocks 160
B.6 Discussion and perspectives 160

Annexe C Discretization of the viscous dissipation term with the MAC scheme 163
C.1 Introduction 165
C.2 Discretization of the dissipation term 166
C.2.1 The two-dimensional case 166
C.2.2 Extension to the three-dimensional case 169
C.3 A strong convergence result 170

Bibliographie

General synthesis

1 Introduction

This work was performed at the *Institut de Radioprotection et de Sûreté Nucléaire* (IRSN). The skill fields of IRSN cover all risks related to ionizing radiation, used in industry or medicine, or natural radiation. Specifically, IRSN carries expertises and conducts research in the domain of nuclear safety, protection against ionizing radiation, control and protection of nuclear materials and protection against malicious acts. An essential part of the safety analysis consists in studying the different situations that a nuclear reactor can face, from normal operation conditions to severe accidents. Many flows of interest in this context are compressible, either monophasic (hydrogen combustion, deflagration or detonation in the reactor containment in the late phases of severe accident scenarii, explosion of gaseous mixtures in industrial environment,. . .) or multi-phasic (primary accident depressurization, bubbling pools generated by the interaction between the molten structures of the vessel and the core and the concrete floor of the containment, once again in severe accidents late phases, . . .).

Our aim here is to contribute to the development of a class of schemes for the computation of compressible flows. The considered systems of governing equations are coupled and strongly nonlinear, and the (industrial) applications in view involve complex geometry and flows, possibly combining quasi-steady states with quick transient phases, with strong physical properties (in particular, density or compressibility) contrasts. Accordingly, the algorithms are developed so as to realize a compromise between two main requirements : preserve the stability in a wide range of Mach numbers and introduce sufficient decoupling to facilitate the resolution of discrete algebraic systems. Pressure correction methods seem to be a good choice to address these requirements. This class of schemes was first introduced in the framework of incompressible flows a long time ago [8, 67], and such algorithms are now quite widespread and well understood in this context (see, for example, [55] for an introduction and [29] for a review of most of the variants). Pressure correction schemes are less popular in the context of compressible flows, even though their application to compressible Navier-Stokes equations may also be traced back to the late sixties, with the seminal work of Harlow and Amsden [35, 36], who developped an iterative algorithm (the so-called ICE method) including an elliptic corrector step for the pressure. Later on, pressure correction equations appeared in numerical schemes proposed by several researchers, essentially in the finite-volume framework, using either a collocated [62, 15, 46, 64, 43, 56] or a staggered arrangement [7, 41, 42, 44, 3, 10, 69, 73, 74, 70, 72] of unknowns ; in the first case, some corrective actions are to be foreseen to avoid the usual odd-even decoupling of the pressure in the low Mach number regime. Some of these algorithms are essentially implicit, since the final stage of a time step involves the unknown at the end-of-step time level ; the end-of-step solution is then obtained by SIMPLE-like iterative processes [71, 44, 15, 46, 64, 43, 56]. The other schemes [41, 42, 62, 3, 10, 75, 69, 74, 70, 72] are predictor-corrector methods, where basically two steps are performed sequentially : first a semi-explicit decoupled prediction of the momentum or velocity (and possibly energy, for non-barotropic flows) and, second, a correction step where the end-of step pressure is evaluated and the momentum and velocity are corrected, as in projection methods for incompressible flows (see [8, 67] for the original papers, [55] for a comprehensive introduction and [29] for a review of most variants). The Characteristic-Based Split (CBS) scheme (see [60] for a recent review or [77] for the seminal paper) was developed in the finite-element context and belongs to this latter class of methods.

In this work, implicit-in-time discretizations are addressed for their (relative) simplicity in view of the theoretical studies ; however, non-iterative pressure correction schemes are our main concern for practical computations. We consider here staggered–in–space discretizations, with the aim to build schemes which are stable and accurate at all Mach numbers and, in particular, boil down to a usual algorithm for incompressible flows (or, more generally, for the asymptotic model of vanishing Mach number flows [54]) when the Mach number tends to zero. This last requirement also implies that, if we implement upwinding techniques (and we will have to for stability reasons), upwinding may have to be performed for each equation separately and with respect to the material velocity only. This is in contradiction with the most common strategy adopted for hyperbolic systems, where upwinding is built from the wave structure of the system (see *eg.* [68, 4] for surveys and [34, 33, 14] for analysis of these schemes at low Mach number), and yields algorithms which are used in practice (see, *eg.*, the so-called AUSM family of schemes [53, 52]), but sarcely studied from a theoretical point of view. One of our main concerns here will thus be to bring, as far as possible, theoretical arguments supporting our numerical developments. Let us first recall a (possible) common skeleton of convergence studies in the finite volume context [16]. The proof may usually be decomposed into three steps :

(i) The first step is to get the existence and some *a priori* estimates on the approximate solution, or, in other words, to obtain stability results for the scheme.

(ii) Next, up to the extraction of a subsequence, compactness arguments yield the existence of a (possibly weak) limit to a sequence of discrete solutions obtained with a sequence of discretizations the space step and, for unsteady problems, the time step of which tend to zero. At this point, *a priori* estimates may imply some regularity of the limit.

(iii) Finally, the fact that the limit is a solution to (a weak form) of the continuous problem is proven by passing to the limit in (a weak formulation of) the scheme.

For the problems studied here, namely the compressible Navier-Stokes or Euler equations, the realization of the complete program seems out of reach, due to the lack of control (Step (*i*)) of space translates of the unknown ; hence we obtain a convergence of sequence of discrete solutions (Step (*i*)) in a sense too weak to allow the passage to the limit in the scheme (Step (*iii*)). There is thus no hope at the present stage to prove the convergence of the schemes in the general cases (*i.e.* except for the barotropic viscous Navier-Stokes equations, see [51, 19, 61] for theoretical analysis of the continuous prolem and [21, 18, 17] for scheme convergence analysis in the simplified case of the steady Stokes problem), and our theoretical analyses are then necessarily somewhat incomplete. However, in both the barotropic and the non-barotropic cases, and at least for most variants of the schemes, we do get the following results :

(*i*) We show that the discrete solution satisfies discrete analogues of the estimates known in the continuous case : positivity of the density and, in the non-barotropic case, of the internal energy, decrease of the total energy, and, for the viscous barotropic flows, control of the velocity in the $\mathrm{L}^2(\mathrm{H}^1)$ norm. These estimates allow to prove the existence of at least one solution to the scheme, by topological degree arguments.

(*ii*) *Supposing the convergence of the scheme in strong enough norms*, we then show that the limits of sequences of solutions are weak solutions to the continuous problem, which may be seen (and is refered to hereafter) as a consistency property of the schemes.

Finally, we confort these theoretical experiments by numerical tests, performed with the open-source software ISIS [40], developed at IRSN on the basis of the software component library and programming environment PELICANS [63].

This paper is organized as follows. We first introduce the considered space discretizations (Section 2). Then we turn to the barotropic Navier-Stokes equations (Section 3), to the "complete" Navier-Stokes equations (Section 4), and, finally, to the Euler equations (Section 5) ; for each case, we present the schemes, summarize the theoretical results and the numerical tests.
In several theoretical developments, we are lead to use a derived form of a discrete finite volume convection operator (for instance, typically, a convection operator for the kinetic energy, possibly with residual terms, obtained from the finite volume discretization of the convection of the velocity components) ; an abstract presentation of such computations is given in the Appendix.

This organization closely follows the thesis one : barotropic flows are adressed in the first two chapters (numerical tests, focussed on the inviscid case, then theory), then a scheme for Navier-Stokes equations is presented and its stability is proven ; finally, we show how to adapt it to compute discontinuous solution of Euler equations.

2 Meshes and unknowns

Let the computational domain Ω be an open polygonal subset of $\mathbb{R}^d$, $d \leq 3$, and $\mathcal{M}$ be a partition of Ω, supposed to be regular in the usual sense of the finite element literature (*eg.* [9]). The cells may be :

- for a general domain Ω, either convex quadrilaterals ($d = 2$) or hexahedra ($d = 3$) or simplices, both types of meshes being possibly combined in a same mesh,

- for a domain the boundaries of which are hyperplanes normal to a coordinate axis, rectangles ($d = 2$) or rectangular parallelepipeds ($d = 3$) (the faces of which, of course, are then also necessarily normal to a coordinate axis).

By $\mathcal{E}$ and $\mathcal{E}(K)$ we denote the set of all $(d-1)$-faces σ of the mesh and of the element $K \in \mathcal{M}$ respectively. The set of edges included in the boundary of Ω is denoted by $\mathcal{E}_{\rm ext}$ and the set of internal ones (*i.e.* $\mathcal{E} \setminus \mathcal{E}_{\rm ext}$) is denoted by $\mathcal{E}_{\rm int}$; a face $\sigma \in \mathcal{E}_{\rm int}$ separating the cells K and L is denoted by $\sigma = K|L$. The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-measure of the face σ. For $1 \leq i \leq d$, we denote by $\mathcal{E}^{(i)} \subset \mathcal{E}$ the subset of the faces of $\mathcal{E}$ which are perpendicular to the i^{th} unit vector of the canonical basis of $\mathbb{R}^d$.

The space discretization is staggered, using either the Marker-And Cell (MAC) scheme [37, 36], or non-conforming low-order finite element approximations, namely the Rannacher and Turek element (RT) [65] for quadrilateral or hexahedric meshes, or the Crouzeix-Raviart (CR) element [11] for simplicial meshes.

For all these space discretizations, the degrees of freedom for the pressure, the density and the internal energy are associated to the cells of the mesh $\mathcal{M}$, and are denoted by :

$$\big\{p_K,\ \rho_K,\ e_K,\ K \in \mathcal{M}\big\}.$$

Let us then turn to the degrees of freedom for the velocity.

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – The degrees of freedom for the velocities are located at the center of the faces of the mesh, and we choose the version of the element where they represent the average of the velocity through a face. The set of degrees of freedom reads :
$$\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E},\ 1 \leq i \leq d\}.$$
- **MAC** discretization – The degrees of freedom for the i^{th} component of the velocity, defined at the centres of the face $\sigma \in \mathcal{E}^{(i)}$, are denoted by :
$$\big\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)},\ 1 \leq i \leq d\big\}.$$

For the definition of the schemes, we need a dual mesh which is defined as follows.

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – For the RT or CR discretization, the dual mesh is the same for all the velocity components. When $K \in \mathcal{M}$ is a simplex, a rectangles or a cuboid, for $\sigma \in \mathcal{E}(K)$, we define $D_{K,\sigma}$ as the cone with basis σ and with vertex the mass center of K. We thus obtain a partition of K in m sub-volumes, where m is the numbers of faces of the mesh, each sub-volume having the same measure $|D_{K,\sigma}| = |K|/m$. We extend this definition to general quadrangles and hexahedra, by supposing that we have built a partition still of equal-volume sub-cells, and with the same connectivities ; note that this is of course always possible, but that such a volume $D_{K,\sigma}$ may be no longer a cone, since, if K is far from a pallelogram, it may not be possible to built a cone having σ as basis, the opposite vertex lying in K and a volume equal to $|K|/m$. The volume $D_{K,\sigma}$ is referred to as the half-diamond cell associated to K and σ.
For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the diamond cell D_σ associated to σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external face $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}(K)$, D_σ is just the same volume as $D_{K,\sigma}$.

- **MAC** discretization – For the MAC scheme, the dual mesh depends on the component of the velocity. For each of them, its definition differs from the RT or CR one only by the choice of the half-diamond cell, which, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, is now the rectangle of basis σ and of measure $|D_{K,\sigma}|$ equal to half the measure of K.

We denote by $|D_\sigma|$ the measure of the dual cell $|D_\sigma|$, and by $\varepsilon = D_\sigma|D_{\sigma'}$ the face separating two diamond cells D_σ and $D_{\sigma'}$ (see Figure 1).

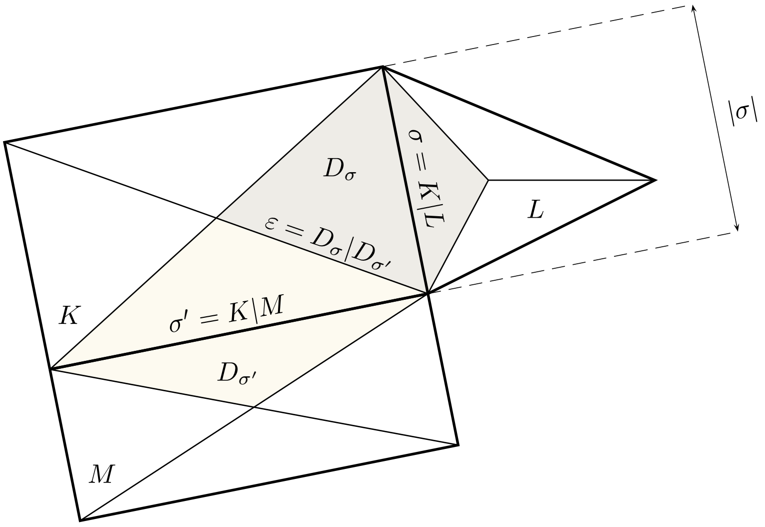

FIG. 1 – Primal and dual meshes for the Rannacher-Turek and Crouzeix-Raviart elements.

3 Compressible barotropic Navier-Stokes equations

The addressed problem in this section reads :

$$\partial_t \rho + \operatorname{div}(\rho\, \boldsymbol{u}) = 0, \tag{1a}$$

$$\partial_t(\rho\, \boldsymbol{u}) + \operatorname{div}(\rho\, \boldsymbol{u} \otimes \boldsymbol{u}) + \boldsymbol{\nabla} p - \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})) = 0, \tag{1b}$$

$$\rho = \wp(p), \tag{1c}$$

where t stands for the time, ρ, $\boldsymbol{u}$ and p are the density, velocity, and pressure in the flow, $\boldsymbol{\tau}(\boldsymbol{u})$ stands for the shear stress tensor, and the function $\wp$ is the equation of state. The problem is supposed to be posed over $\Omega \times (0,T)$, where $(0,T)$ is a finite time interval. This system must be supplemented by suitable boundary conditions, and initial conditions for ρ and $\boldsymbol{u}$, the initial condition for ρ being supposed positive. The closure relation for $\boldsymbol{\tau}(\boldsymbol{u})$ is assumed to be :

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

where μ stands for a non-negative parameter, possibly depending on $\boldsymbol{x}$. When the viscous term $\boldsymbol{\tau}(\boldsymbol{u})$ vanishes, the system (1) becomes hyperbolic.

Let us denote by E_c the kinetic energy $E_c = \frac{1}{2}\,\rho\,|\boldsymbol{u}|^2$. Taking the inner product of (1b) by $\boldsymbol{u}$ yields, after formal compositions of partial derivatives and using (1a) :

$$\partial_t E_c + \operatorname{div}\big(E_c\,\boldsymbol{u}\big) + \boldsymbol{\nabla} p \cdot \boldsymbol{u} = \operatorname{div}\big(\boldsymbol{\tau}(\boldsymbol{u})\big) \cdot \boldsymbol{u}. \tag{2}$$

This relation is refered to as the kinetic energy balance.

Let us now define the function $\mathcal{P}$, from $(0,+\infty)$ to $\mathbb{R}$, as a primitive of $s \mapsto \wp(s)/s^2$, where $\wp = \wp^{-1}$; this quantity is often called the elastic potential. Let $\mathcal{H}$ be the function defined by $\mathcal{H}(s) = s\mathcal{P}(s)$, $\forall s \in (0,+\infty)$; it may easily be checked that $\rho\mathcal{H}'(\rho) - \mathcal{H}(\rho) = \wp(\rho)$; therefore, by a formal computation detailed in the appendix (see Equation (34)), multiplying (1a) by $\mathcal{H}'(\rho)$ yields :

$$\partial_t\big(\mathcal{H}(\rho)\big) + \operatorname{div}\big(\mathcal{H}(\rho)\,\boldsymbol{u}\big) + p\,\operatorname{div}(\boldsymbol{u}) = 0. \tag{3}$$

Let us denote by $\mathcal{S}$ the quantity $\mathcal{S} = E_c + \mathcal{H}(\rho)$. Summing (2) and (3), we get :

$$\partial_t \mathcal{S} + \operatorname{div}\big((\mathcal{S}+p)\,\boldsymbol{u}\big) - \operatorname{div}\big(\boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{u}\big) = -\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u}. \tag{4}$$

This shows that, in the hyperbolic case, $\mathcal{S}$ is an entropy of the system, and an entropy solution to (1) is thus required to satisfy :

$$\int_0^T \int_\Omega \big[-\mathcal{S}\partial_t\varphi - (\mathcal{S}+p)\,\boldsymbol{u}\cdot\boldsymbol{\nabla}\varphi\big]\,\mathrm{d}\boldsymbol{x}\delta t \\ - \int_\Omega \mathcal{S}(\boldsymbol{x},0)\ \varphi(\boldsymbol{x},0)\,\mathrm{d}\boldsymbol{x} \leq 0, \quad \forall \varphi \in \mathrm{C}_c^\infty\big(\Omega\times[0,T)\big),\ \varphi \geq 0. \tag{5}$$

Then, formally, if we suppose that the velocity is prescribed to zero at the boundary (the normal velocity, in the hyperbolic case), integrating (4) yields, since the viscous dissipation term $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u}$ is non-negative :

$$\frac{d}{dt}\int_\Omega \big[\frac{1}{2}\,\rho\,|\boldsymbol{u}|^2 + \mathcal{H}(\rho)\big]\,\mathrm{d}\boldsymbol{x} \leq 0. \tag{6}$$

Since the function $\mathcal{P}$ is increasing, Inequality (6) provides an estimate of the solution.

We study two schemes for the numerical solution of System (1) which differs by the time discretization : the first one is implicit, and the second one is a non-iterative pressure-correction scheme introduced in [20]. This latter algorithm (and, by an easy extension, also the first one) was shown in [20] to have at least one solution, to provide solutions satisfying $\rho > 0$ (and so $p > 0$) and to be unconditionally stable, in the sense that its (their) solution(s) satisfies a discrete analogue of Inequality (6). The results presented in this section complement this work in several directions. For the implicit scheme :

- We first derive discrete analogues of (2) and (3), the first (local) balance equation, *i.e.* the discrete kinetic energy balance, being obtained on dual cells, and the second one, *i.e.* the elastic potential balance, on primal cells.

  These equations are used a first time to obtain the stability of the scheme by a simple integration in space (*i.e.* summation over the primal and dual control volumes).
- Second, in one space dimension and for the hyperbolic case, we prove that the limit of any convergent sequence of solutions to the scheme is a weak solution to the problem (in fact, satisfies the Rankine-Hugoniot conditions, and thus exhibits "correct" shocks).

- Finally, passing to the limit on the discrete kinetic energy and elastic potential balances, we show that such a limit also satisfies the entropy inequality (5).

For the pressure correction scheme, the results are essentially the same : the scheme is unconditionally stable, and the passage to the limit in the scheme shows that, in case of convergence, the predicted and end-of-step velocities necessarily tend to the same function, and that the limit is still a weak solution to the problem, satisfying the entropy inequality.

Numerical tests, performed with the pressure correction scheme, confort these theoretical results.

We first summarize in this section the obtained theoretical results (Sections 3.1 and 3.2.c). which are detailed in Chapter 2 of this document. Then we show results of a numerical test (Section 3.2.d), extracted from a more comprehensive study also addressing an extension of the scheme to two-phase flows, presented in Chapter 1 of this document.

3.1 An implicit scheme scheme

3.1.a 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 $n = 0, 1, \ldots, N-1$ be the constant time step.

We begin with the discretization of the mass balance equation (1a). For both the MAC and RT or CR discretizations, let us denote by $\boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}$ the outward normal velocity to the face σ of K, which is computed, for the RT and CR elements, by taking the inner product of the velocity at the face with the outward normal vector (as implied by the notation) and which is given, for the MAC scheme, by the value of the component of the velocity at the center of the face (up to a change of sign). The discrete equations are obtained by an upwind finite volume discretization and read :

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}\,(\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma} = 0, \qquad \text{with } F_{K,\sigma} = |\sigma|\ \boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}\ \widetilde{\rho}_\sigma^{n+1}, \tag{7}$$

and where $\widetilde{\rho}_\sigma^{n+1}$ is the upwind approximation of ρ^{n+1} at the face σ with respect to $\boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}$. This approximation ensures that $\rho^{n+1} > 0$ if $\rho^n > 0$ and if the density is prescribed to a positive value at inflow boundaries.

For both MAC and RT or CR discretizations, for $1 \le i \le d$ and $\sigma \in \mathcal{E}^{(i)}$, we denote by $(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}$ an approximation of the i-th component of the viscous term associated to σ, and we denote by $(\boldsymbol{\nabla} p^n)_{\sigma,i}$ the i-th component of the discrete pressure gradient at the face σ. With these notations, we are able to write the following general form of the approximation of the momentum balance equation :

$$\begin{aligned} \frac{|D_\sigma|}{\delta t}\,(\rho_\sigma^{n+1}\ \boldsymbol{u}_{\sigma,i}^{n+1} - \rho_\sigma^n\ \boldsymbol{u}_{\sigma,i}^n) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1}\ \boldsymbol{u}_{\varepsilon,i}^{n+1} \\ + |D_\sigma| (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} - |D_\sigma| (\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i} = 0, \end{aligned} \tag{8}$$

for $1 \le i \le d$, and for $\sigma \in \mathcal{E} \setminus \mathcal{E}_D$ in the case of the RT or CR discretizations, and $\sigma \in \mathcal{E}^{(i)} \setminus \mathcal{E}_D$ in the case of the MAC scheme. In this relation, ρ_σ^{n+1} and ρ_σ^n stand for an approximation of the density on the face

σ at time t^{n+1} and t^n respectively (which must not be confused with the upstream density $\widetilde{\rho}_\sigma$ used in the mass balance), $F^{n+1}_{\sigma,\varepsilon}$ is the discrete mass flux through the dual face ε outward D_σ, and $\boldsymbol{u}^{n+1}_{\varepsilon,i}$ stands for an approximation of $\boldsymbol{u}^{n+1}_i$ on ε which may be chosen either centred or upwind.

The finite element discretization of the i-th component of the pressure gradient term reads :

$$|D_\sigma|(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = -\sum_{M\in\mathcal{M}} \int_M p^{n+1}\ \mathrm{div}\boldsymbol{\varphi}^{(i)}_\sigma\ \mathrm{d}\boldsymbol{x},$$

with $\boldsymbol{\varphi}^{(i)}_\sigma$ reads $\boldsymbol{\varphi}^{(i)}_\sigma = \varphi_\sigma \boldsymbol{e}^{(i)}$, where φ_σ is the finite element shape function associated to σ and $\boldsymbol{e}^{(i)}$ stands for the i^{th} vector of the canonical basis of $\mathbb{R}^d$. Since the pressure is piecewise constant, using the definition of the RT or CR shape functions, an easy computation yields for an internal face $\sigma = K|L$:

$$|D_\sigma|(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = |\sigma|\ (p^{n+1}_L - p^{n+1}_K)\ \boldsymbol{n}_{K,\sigma}\cdot\boldsymbol{e}^{(i)},$$

and, for an external face $\sigma \in \mathcal{E}(K)\cap\mathcal{E}_{\rm ext}\setminus\mathcal{E}_D$:

$$|D_\sigma|(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = -|\sigma|\ p^n_K\ \boldsymbol{n}_{K,\sigma}\cdot\boldsymbol{e}^{(i)}.$$

These expressions coincide which the discrete gradient in the MAC discretization.

The finite element discretization of the viscous term $(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}$, associated to σ and to the component i, reads :

$$|D_\sigma|(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i} = -\mu\sum_{K\in\mathcal{M}}\int_K \nabla\boldsymbol{u}^{n+1}\cdot\nabla\boldsymbol{\varphi}^{(i)}_\sigma - \frac{\mu}{3}\sum_{K\in\mathcal{M}}\int_K \mathrm{div}\,\boldsymbol{u}^{n+1}\ \mathrm{div}\,\boldsymbol{\varphi}^{(i)}_\sigma.$$

The MAC discretization of this same viscous term is detailed in [2].

The main motivation to implement a finite volume approximation for the first two terms in (8) is to obtain a discrete equivalent of the kinetic energy balance (see next section). For this result to be valid, the necessary condition is that the convection operator vanishes for a constant velocity, *i.e.* that the following discrete mass balance over the diamond cells is satisfied [1, 20] :

$$\forall\sigma\in\mathcal{E}_{\rm int},\qquad \frac{|D_\sigma|}{\delta t}\ (\rho^{n+1}_\sigma - \rho^n_\sigma) + \sum_{\varepsilon\in\mathcal{E}(D_\sigma)} F^{n+1}_{\sigma,\varepsilon} = 0. \tag{9}$$

This governs the choice for the definition of the density approximation ρ_σ and the mass fluxes $F_{\sigma,\varepsilon}$. The density ρ_σ is defined by a weighted average : $\forall\sigma\in\mathcal{E}_{\rm int}$, $\sigma = K|L$, $|D_\sigma|\ \rho_\sigma = |D_{K,\sigma}|\ \rho_K + |D_{L,\sigma}|\ \rho_L$ and $\forall\sigma\in\mathcal{E}_{\rm ext}\setminus\mathcal{E}_D$, $\sigma\in\mathcal{E}(K)$, $\rho_\sigma = \rho_K$. For a dual edge ε included in the primal cell K, the flux $F_{\sigma,\varepsilon}$ is computed as a linear combination (with constant coefficients, *i.e.* independent of the edge and the cell) of the mass fluxes through the faces of K, *i.e.* the quantities $(F^{n+1}_{K,\sigma})_{\sigma\in\mathcal{E}(K)}$ appearing in the discrete mass balance (7). We do not give here this set of coefficients, and refer to [1, 38, 25] for a detailed construction of this approximation.

3.1.b Kinetic energy balance, elastic potential identity and stability

We begin by deriving a discrete kinetic energy balance equation. Let $\delta^{\rm up}$ be a coefficient defined by $\delta^{\rm up} = 1$ if an upwind discretization is used for the convection term in the momentum balance equation

(8) and $\delta^{\rm up}=0$ in the centered case. With this notation, the momentum balance equation reads :

$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1}\boldsymbol{u}_{\sigma,i}^{n+1}-\rho_\sigma^n\boldsymbol{u}_{\sigma,i}^n)+\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,F_{\sigma,\varepsilon}^{n+1}\,(\boldsymbol{u}_{\sigma,i}^{n+1}+\boldsymbol{u}_{\sigma',i}^{n+1})$$
$$+\,\delta^{\rm up}\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(\boldsymbol{u}_{\sigma,i}^{n+1}-\boldsymbol{u}_{\sigma',i}^{n+1})+|D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}-|D_\sigma|({\rm div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}=0.$$

Taking the inner product of this equation with the corresponding velocity unknown, *i.e.* $\boldsymbol{u}_{\sigma,i}^{n+1}$, yields $T_{\sigma,i}^{\rm conv}+T_{\sigma,i}^{\rm up}+T_{\sigma,i}^{p,\boldsymbol{\tau}}=0$, with :

$$\begin{aligned}
T_{\sigma,i}^{\rm conv} &= \Big[\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1}\boldsymbol{u}_{\sigma,i}^{n+1}-\rho_\sigma^n\boldsymbol{u}_{\sigma,i}^n)+\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,F_{\sigma,\varepsilon}^{n+1}\,(\boldsymbol{u}_{\sigma,i}^{n+1}+\boldsymbol{u}_{\sigma',i}^{n+1})\Big]\,\boldsymbol{u}_{\sigma,i}^{n+1},\\
T_{\sigma,i}^{\rm up} &= \delta^{\rm up}\Big[\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(\boldsymbol{u}_{\sigma,i}^{n+1}-\boldsymbol{u}_{\sigma',i}^{n+1})\Big]\,\boldsymbol{u}_{\sigma,i}^{n+1},\\
T_{\sigma,i}^{p,\boldsymbol{\tau}} &= |D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}\,\boldsymbol{u}_{\sigma,i}^{n+1}-|D_\sigma|({\rm div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}\,\boldsymbol{u}_{\sigma,i}^{n+1}.
\end{aligned}$$

Lemma .7.2, applied on the dual mesh, yields :

$$T_{\sigma,i}^{\rm conv}=\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho_\sigma^{n+1}(\boldsymbol{u}_{\sigma,i}^{n+1})^2-\rho_\sigma^n(\boldsymbol{u}_{\sigma,i}^n)^2\Big]+\frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}}F_{\sigma,\varepsilon}^{n+1}\,\boldsymbol{u}_{\sigma,i}^{n+1}\,\boldsymbol{u}_{\sigma',i}^{n+1}$$
$$+\,\frac{|D_\sigma|}{2\,\delta t}\,\rho_\sigma^n\,\big(\boldsymbol{u}_{\sigma,i}^{n+1}-\boldsymbol{u}_{\sigma,i}^n\big)^2.$$

Let us define $R_{\sigma,i}^{n+1}$ by the sum of $T_{\sigma,i}^{\rm up}$ and the last term of $T_{\sigma,i}^{\rm conv}$:

$$R_{\sigma,i}^{n+1}=\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\,\rho_\sigma^n\,\big(\boldsymbol{u}_{\sigma,i}^{n+1}-\boldsymbol{u}_{\sigma,i}^n\big)^2+\delta^{\rm up}\Big[\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(\boldsymbol{u}_{\sigma,i}^{n+1}-\boldsymbol{u}_{\sigma',i}^{n+1})\Big]\,\boldsymbol{u}_{\sigma,i}^{n+1}. \tag{10}$$

With this notation, we thus obtain the following relation :

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho_\sigma^{n+1}(\boldsymbol{u}_{\sigma,i}^{n+1})^2-\rho_\sigma^n(\boldsymbol{u}_{\sigma,i}^n)^2\Big]+\frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}}F_{\sigma,\varepsilon}^{n+1}\,\boldsymbol{u}_{\sigma,i}^{n+1}\,\boldsymbol{u}_{\sigma',i}^{n+1}$$
$$+\,|D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}\,\boldsymbol{u}_{\sigma,i}^{n+1}-|D_\sigma|({\rm div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}\,\boldsymbol{u}_{\sigma,i}^{n+1}=-R_{\sigma,i}^{n+1}. \tag{11}$$

We recognize at the left-hand side a discrete kinetic energy balance, *i.e.* a reasonable discretization of Equation (2), with a conservative finite volume discretization of the kinetic energy convection terms. The right-hand side consists in a numerical residual, the sign of which will be studied later.

We now turn to the elastic potential balance. Multiplying the discrete mass balance equation (7) by $\mathcal{H}'(\rho_K)$ and invoking Lemma .7.1 yields, $\forall K\in\mathcal{M}$:

$$\frac{|K|}{\delta t}\,(\mathcal{H}(\rho_K^{n+1})-\mathcal{H}(\rho_K^n))+\sum_{\sigma\in\mathcal{E}(K)}|\sigma|\,\big[\mathcal{H}(\rho_\sigma^{n+1})+p_K\big]\,\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}=-R_K^{n+1}, \tag{12}$$

with :

$$R_K^{n+1}=\frac{1}{2}\frac{|K|}{\delta t}\mathcal{H}''(\overline{\rho}_K^{n,n+1})(\rho_K^{n+1}-\rho_K^n)^2-\frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)}|\sigma|\,\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}\,\mathcal{H}''(\overline{\rho}_\sigma^{n+1})(\rho_\sigma^{n+1}-\rho_K^{n+1})^2,$$

where the quantity $\overline{\rho}_K^{n,n+1} \in [\min(\rho_K^{n+1},\rho_K^n),\max(\rho_K^{n+1},\rho_K^n)]$ and, for any face $\sigma \in \mathcal{E}(K)$, $\overline{\rho}_\sigma^{n+1} \in [\min(\rho_\sigma^{n+1},\rho_K^{n+1}),\max(\rho_\sigma^{n+1},\rho_K^{n+1})]$.

Equation (12) is a finite volume discretization of the (non conservative) elastic potential balance (3), with a non positive residual term, thanks to the fact that the function $\mathcal{H}$ is convex and that an upwind approximation of the density is used in the mass balance.

The stability of the scheme is then obtained by summing :

(i) Equation (11) over the components i and the faces $\sigma \in \mathcal{E}$ for the RT or CR discretizations, and over i and $\sigma \in \mathcal{E}^{(i)}$ for the MAC scheme,

(ii) Equation (12) over $K \in \mathcal{M}$,

(iii) and, finally, the two obtained relations.

Let us suppose that the velocity vanishes at the boundary, and let us then invoke three arguments. First, the discrete gradient and divergence operators are dual with respect to the L^2 inner product, in the sense that :

$$\sum_{i,\mathcal{E}} |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}\ \boldsymbol{u}_{\sigma,i}^{n+1} + \sum_{K\in\mathcal{M}} p_K \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\ \boldsymbol{u}_\sigma^{n+1}\cdot \boldsymbol{n}_{K,\sigma} = 0,$$

where the notation $\sum_{i,\mathcal{E}}$ means that we sum over the component index i and on $\sigma \in \mathcal{E}$ for the RT and CR discretizations, and on i and $\sigma \in \mathcal{E}^{(i)}$ for the MAC scheme. Second, we suppose that (see Section 4) :

$$\sum_{i,\mathcal{E}} |D_\sigma|\,(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}\ \boldsymbol{u}_{\sigma,i}^{n+1} \geq 0.$$

Third, reordering the summations yields, for the part of the remainder of the momentum balance equation associated to the upwinding :

$$\sum_{i,\mathcal{E}} T_{\sigma,i}^{\mathrm{up}} = \delta^{\mathrm{up}} \sum_{i,\bar{\mathcal{E}}\ (\varepsilon=D_\sigma|D_{\sigma'})} \frac{1}{2}\ |F_{\sigma,\varepsilon}^{n+1}|\ (\boldsymbol{u}_{\sigma,i}^{n+1}-\boldsymbol{u}_{\sigma',i}^{n+1})^2 \geq 0,$$

where the notation $\sum_{i,\bar{\mathcal{E}}\ (\varepsilon=D_\sigma|D_{\sigma'})}$ means that we perform the sum over i and the faces of the dual mesh associated to the component i of the velocity, and that, for a face ε in the sum, the two adjacent dual cells are denoted by D_σ and D'_σ. Finally, since the conservative fluxes vanish in the summation, we thus get :

$$\frac{1}{2}\sum_{i,\mathcal{E}} \frac{|D_\sigma|}{\delta t}\Big[\rho_\sigma^{n+1}(\boldsymbol{u}_{\sigma,i}^{n+1})^2 - \rho_\sigma^n(\boldsymbol{u}_{\sigma,i}^n)^2\Big] + \sum_{K\in\mathcal{M}} \frac{|K|}{\delta t}\ (\mathcal{H}(\rho_K^{n+1})-\mathcal{H}(\rho_K^n)) \leq 0, \qquad (13)$$

which is a discrete analogue to (6).

3.1.c Passing to the limit in the scheme (1D case)

We focus in this section on the inviscid 1D form of Problem (1), and show that, if a sequence of solutions is controlled in suitable norms and converges to a limit, this latter necessarily satisfies a (part of the) weak formulation of the continuous problem.

Let $(\mathcal{M}^{(m)},\delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that the time step $\delta t^{(m)}$ and the size $h^{(m)}$ of the mesh $\mathcal{M}^{(m)}$, defined by :

$$h^{(m)} = \sup_{K\in\mathcal{M}^{(m)}} \mathrm{diam}(K),$$

tend to zero as $m \to \infty$.

Let $\rho^{(m)}$, $p^{(m)}$ and $u^{(m)}$ be the solution given by the scheme with the mesh $\mathcal{M}^{(m)}$ and the time step $\delta t^{(m)}$, or, more precisely speaking, a 1D version of the scheme which may be obtained by taking the MAC variant, only one horizontal stripe of meshes, supposing that the vertical component of the velocity (the degree of freedom of which are located on the top and bottom boundaries) vanishes, and that the measure of the faces is equal to 1. To the discrete unknowns, we associate piecewise constant functions on time intervals and on primal or dual meshes, so the density $\rho^{(m)}$, the pressure $p^{(m)}$ and the velocity $u^{(m)}$ are defined almost everywhere on $\Omega \times (0,T)$ by :

$$\rho^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K \in \mathcal{M}} (\rho^{(m)})_K^n \; \mathcal{X}_K \; \mathcal{X}_{(n,n+1)}, \qquad p^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K \in \mathcal{M}} (p^{(m)})_K^n \; \mathcal{X}_K \; \mathcal{X}_{(n,n+1)},$$

$$u^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{\sigma \in \mathcal{E}} (u^{(m)})_\sigma^n \; \mathcal{X}_{D_\sigma} \; \mathcal{X}_{(n,n+1)},$$

where $\mathcal{X}_K$, $\mathcal{X}_{D_\sigma}$ and $\mathcal{X}_{(n,n+1)}$ stand for the characteristic function of K, D_σ and the interval (t^n, t^{n+1}) respectively.

We suppose a uniform control on the translates in space and time of the sequence of solutions, which we now state. For discrete function q and v defined on the primal and dual mesh, respectively, we define a discrete $\mathrm{L}^1\big((0,T); \mathrm{BV}(\Omega)\big)$ norm by :

$$\|q\|_{\mathcal{T},x,BV} = \sum_{n=0}^{N} \delta t \sum_{\sigma \in \mathcal{E},\ \sigma = K|L} |q_L^n - q_K^n|, \qquad \|v\|_{\mathcal{T},x,BV} = \sum_{n=0}^{N} \delta t \sum_{\varepsilon \in \bar{\mathcal{E}},\ \sigma = D_\sigma | D'_\sigma} |v_{\sigma'}^n - v_\sigma^n|,$$

and a discrete $\mathrm{L}^1\big(\Omega; \mathrm{BV}((0,T))\big)$ norm by :

$$\|q\|_{\mathcal{T},t,BV} = \sum_{K \in \mathcal{M}} h_K \sum_{n=0}^{N-1} |q_K^{n+1} - q_K^n|, \qquad \|v\|_{\mathcal{T},t,BV} = \sum_{\sigma \in \mathcal{E}} h_\sigma \sum_{n=0}^{N-1} |v_\sigma^{n+1} - v_\sigma^n|,$$

where, for $\sigma = K|L$, $h_\sigma = (h_K + h_L)/2$. We suppose the following uniform bounds of the sequence of solutions with respect to these two norms :

$$\|\rho^{(m)}\|_{\mathcal{T},x,BV} + \|p^{(m)}\|_{\mathcal{T},x,BV} + \|u^{(m)}\|_{\mathcal{T},x,BV} \leq C, \quad \forall m \in \mathbb{N}, \tag{14}$$

and :

$$\|u^{(m)}\|_{\mathcal{T},t,BV} \leq C, \quad \forall m \in \mathbb{N}. \tag{15}$$

A weak solution to the continuous problem satisfies, for any $\varphi \in \mathrm{C}_c^\infty\big([0,T) \times \Omega\big)$:

$$- \int_{\Omega \times (0,T)} \Big[\rho \, \partial_t \varphi + \rho \, u \, \partial_x \varphi \Big] \, \mathrm{d}x \delta t - \int_\Omega \rho(x,0) \, \varphi(x,0) \, \mathrm{d}x = 0, \tag{16a}$$

$$- \int_{\Omega \times (0,T)} \Big[\rho \, u \, \partial_t \varphi + (\rho \, u^2 + p) \, \partial_x \varphi \Big] \, \mathrm{d}\boldsymbol{x} \delta t - \int_\Omega \rho(x,0) \, u(x,0) \, \varphi(x,0) \, \mathrm{d}x = 0, \tag{16b}$$

$$\rho = \wp(p). \tag{16c}$$

Note that these relations are not sufficient to define a weak solution to the problem, since they do not imply anything about the boundary conditions. However, they allow to derive the Rankine-Hugoniot conditions ; so, if we show that they are satisfied by the limit of a sequence of solutions to the discrete problem, this implies, loosely speaking, that *the scheme computes the right shocks*, which is the result we are seeking. It is stated in the following theorem.

Theorem .3.1

Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence satisfies (15) and (14) and converges in $\mathrm{L}^r\big((0,T)\times\Omega\big)^3$, for $1 \leq r < \infty$, to $(\bar\rho,\, \bar p,\, \bar u) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^3$.

Then the limit $(\bar\rho,\, \bar p,\, \bar u)$ satisfies the system (16) and the entropy condition (5).

Proof The passage to the limit in the equations of the scheme is technical, but invokes rather standard arguments.

Obtaining the entropy condition is more intricate. We need to pass to the limit in the kinetic energy balance (11) and in the elastic potential balance (12) simultaneously. To this purpose, for $\varphi \in \mathrm{C}_c^\infty\big([0,T)\times\Omega\big)$, we define two interpolates : one is defined over the dual cells and is used as a test function for (11) and the second one is defined over the primal cells, and is used as a test function for (12). We then pass to the limit in the "differential terms" of these discrete equations, and disregard the non-negative residuals (at the left-hand side). A problem is posed by the residual associated to the upwinding, which reads :

$$R_\sigma = \Big[\sum_{\varepsilon=D_\sigma|D_{\sigma'}} \frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\;(\boldsymbol{u}_{\sigma,i}^{n+1} - \boldsymbol{u}_{\sigma',i}^{n+1})\Big]\;\boldsymbol{u}_{\sigma,i}^{n+1},$$

and the sign of which is unknown. To get an intuition of how to deal with this term, let us remark that it may be seen as a discrete analogue to a diffusion term $-\mu\,\Delta u\; u$ with a numerical viscosity μ tending to zero as the space step. Let us now compare this term to $\mu|\boldsymbol{\nabla} u|^2$, in the sense of distributions. For ψ a regular function with a compact support, remarking that $-\mu\, u\,\Delta u - \mu|\boldsymbol{\nabla} u|^2 = -\mathrm{div}(\mu u\,\boldsymbol{\nabla} u)$, we get :

$$\int_0^T\int_\Omega \big[-\mu\, u\,\Delta u - \mu|\boldsymbol{\nabla} u|^2\big]\psi\,\mathrm{d}\boldsymbol{x}\delta t = \int_0^T\int_\Omega \mu u\,\boldsymbol{\nabla} u\cdot\boldsymbol{\nabla}\psi\,\mathrm{d}\boldsymbol{x}\delta t \leq C_\psi\;\|u\|_{\mathrm{L}^\infty}\;\|u\|_{W^{1,1}}\mu,$$

and therefore, if $\|u\|_{\mathrm{L}^\infty}$ and $\|u\|_{W^{1,1}}$ are bounded, the difference between $-\mu\, u\,\Delta u$ and $\mu|\boldsymbol{\nabla} u|^2$ behaves like μ. Returning at the discrete level, this computation suggests that R_σ behaves at the limit as a dissipation term (*i.e.* a discrete equivalent of $\mu|\boldsymbol{\nabla} u|^2$), the sign of which is guaranteed. The same argument is used in a different way in the non-barotropic case : the "viscous term" R_σ is compensated in the internal energy balance by a "dissipation term" (see Section 5.1). □

Remark 1 (Control of the translates)

In the assumptions of the theorem .3.1, we can sharpen (14) and (15). Indeed, to prove that the limit is a weak solution, it is sufficient to have :

$$\lim_{m\to+\infty} h^{(m)}\Big[\|\rho^{(m)}\|_{\mathcal{T},x,BV} + \|p^{(m)}\|_{\mathcal{T},x,BV} + \|u^{(m)}\|_{\mathcal{T},x,BV}\Big] = 0.$$

In addition, this estimate may be proven (and not supposed) by adding to the scheme a numerical diffusion scaled by $(h^{(m)})^\beta$, with $0 < \beta < 2$. To obtain that the limit is the entropy weak solution, the following assumption is sufficient :

$$\lim_{m\to+\infty} \delta t\Big[\|u^{(m)}\|_{\mathcal{T},t,BV}\Big] = 0.$$

3.2 A pressure correction scheme

3.2.a The scheme

In this section, we derive the pressure correction scheme from the implicit scheme. The first step, as usual, is to compute a tentative velocity by solving the momentum balance equation with the begining-of-step pressure. Then, the velocity is corrected and the other variables are advanced in time, in the so-called correction step. For stability reasons, or, in other words, to be able to derive a kinetic energy balance, we need that the mass balance over the dual cells (9) holds ; since the mass balance is not yet solved when performing the prediction step, this leads us to do a time shift of the density at this step.

In the time semi-discrete setting, the proposed algorithm reads :
1 - **Pressure renormalization step** – Solve the following elliptic problem for $\tilde{p}^{n+1}$:

$$\mathrm{div}\Big[\frac{1}{\rho^n}\boldsymbol{\nabla}\tilde{p}^{n+1}\Big] = \mathrm{div}\Big[\frac{1}{(\rho^n\,\rho^n)^{1/2}}\boldsymbol{\nabla}p^n\Big] \tag{17}$$

2 - **Prediction step** – Solve the following semi-discrete linearized momentum balance equation for $\tilde{\boldsymbol{u}}^{n+1}$:

$$\frac{\rho^n\;\tilde{\boldsymbol{u}}^{n+1} - \rho^{n-1}\;\boldsymbol{u}^n}{\delta t} + \mathrm{div}(\rho^n\;\tilde{\boldsymbol{u}}^{n+1}\otimes\boldsymbol{u}^n) + \boldsymbol{\nabla}\tilde{p}^{n+1} - \mathrm{div}(\boldsymbol{\tau}(\tilde{\boldsymbol{u}}^{n+1})) = 0. \tag{18}$$

3 - **Correction step** – Solve (simultanuously) the following non linear equations for p^{n+1}, $\boldsymbol{u}^{n+1}$ and ρ^{n+1} :

$$\rho^n\;\frac{\boldsymbol{u}^{n+1} - \tilde{\boldsymbol{u}}^{n+1}}{\delta t} + \nabla(p^{n+1} - \tilde{p}^{n+1}) = 0, \tag{19a}$$

$$\frac{\rho^{n+1} - \rho^n}{\delta t} + \mathrm{div}(\rho^{n+1}\;\boldsymbol{u}^{n+1}) = 0, \tag{19b}$$

$$\rho^{n+1} = \wp(p^{n+1}). \tag{19c}$$

The solution of Step 3 is performed by combining equations (19a) and (19b), therefore obtaining a non-linear elliptic problem for the pressure, which reads in the time semi-discrete setting :

$$\frac{\wp(p^{n+1}) - \rho^n}{\delta t^2} - \mathrm{div}\big[\,\frac{\rho^{n+1}}{\rho^n}\,\nabla(p^{n+1} - \tilde{p}^{n+1})\,\big] = -\frac{1}{\delta t}\;\mathrm{div}(\rho^{n+1}\tilde{u}^{n+1}).$$

The fully discrete equations are obtained from the implicit scheme by a mere change in time levels, except for Equations (17) and (19a), which are new. The first one is obtained by using the discrete gradient and divergence operators already introduced, and reads :

$$\forall K \in \mathcal{M}, \qquad \sum_{\sigma=K|L} \frac{1}{\rho_\sigma^n}\frac{|\sigma|^2}{|D_\sigma|}\;(\tilde{p}_K^{n+1} - \tilde{p}_L^{n+1}) = \sum_{\sigma=K|L} \frac{1}{\sqrt{\rho_\sigma^n\,\rho_\sigma^{n-1}}}\frac{|\sigma|^2}{|D_\sigma|}\;(p_K^n - p_L^n)\,.$$

Relation (19a) is discretized similarly to the momentum balance (8), *i.e.* a finite volume technique is used for the unsteady term in both the MAC, RT and CR discretizations :

$$\frac{|D_\sigma|}{\delta t}\,\rho_\sigma^n\,(\boldsymbol{u}_{\sigma,i}^{n+1} - \tilde{\boldsymbol{u}}_{\sigma,i}^{n+1}) + |D_\sigma|\;\Big[(\boldsymbol{\nabla}p^{n+1})_{\sigma,i} - (\boldsymbol{\nabla}\tilde{p}^{n+1})_{\sigma,i}\Big] = 0,$$

for $1 \leq i \leq d$, and for $\sigma \in \mathcal{E} \setminus \mathcal{E}_D$ in the case of the RT or CR discretizations, and $\sigma \in \mathcal{E}^{(i)} \setminus \mathcal{E}_D$ in the case of the MAC scheme.

3.2.b Stability and kinetic energy balance equation

We repeat the process that we followed for the implicit scheme, to prove the stability of the scheme and derive a discrete kinetic energy balance equation. To this purpose, we multiply the velocity prediction equation by the corresponding degree of freedom of the predicted velocity $\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}$, to obtain :

$$\begin{aligned}\frac{|D_\sigma|}{\delta t}\,\big(\rho^n_\sigma \tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} - \rho^{n-1}_\sigma \boldsymbol{u}^n_{\sigma,i}\big)\,\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} + \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)} F^n_{\sigma,\varepsilon}\tilde{\boldsymbol{u}}^{n+1}_{\varepsilon,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}&\\ + |D_\sigma|\,(\boldsymbol{\nabla}\tilde{p}^{n+1})_{\sigma,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} - |D_\sigma|(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} = 0.& \quad (20)\end{aligned}$$

We then write the velocity correction equation as :

$$\Big[\frac{|D_\sigma|}{\delta t}\,\rho^n_\sigma\Big]^{1/2}\boldsymbol{u}^{n+1}_{\sigma,i} + \frac{\big[|D_\sigma|\;\delta t\big]^{1/2}}{(\rho^n_\sigma)^{1/2}}\,(\boldsymbol{\nabla}p^{n+1})_{\sigma,i} = \Big[\frac{|D_\sigma|}{\delta t}\,(\rho^n_\sigma\Big]^{1/2}\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} + \frac{\big[|D_\sigma|\;\delta t\big]^{1/2}}{(\rho^n_\sigma)^{1/2}}\,(\boldsymbol{\nabla}\tilde{p}^{n+1})_{\sigma,i},$$

and square this relation, sum with (20) and get, applying Lemma .7.2 (again on the dual mesh) to the first two terms of (20) :

$$\begin{aligned}\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho^n_\sigma(\boldsymbol{u}^{n+1}_{\sigma,i})^2 - \rho^{n-1}_\sigma(\boldsymbol{u}^n_{\sigma,i})^2\Big] + \frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}} F^n_{\sigma,\varepsilon}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma',i} + |D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}\;\boldsymbol{u}^{n+1}_{\sigma,i}&\\ - |D_\sigma|(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_{\sigma,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} + \frac{|D_\sigma|\;\delta t}{\rho^n_\sigma}\;\Big[|(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}|^2 - |(\boldsymbol{\nabla}\tilde{p}^{n+1})_{\sigma,i}|^2\Big] = R^{n+1}_{\sigma,i},& \quad (21)\end{aligned}$$

where $R^{n+1}_{\sigma,i}$ takes the same expression as in the implicit case (*i.e.* is given by Equation (10)), replacing $\boldsymbol{u}^{n+1}$ by $\tilde{\boldsymbol{u}}^{n+1}$. Summing Relation (21) over the components and edges, Relation (12) over the cells and finally the two resulting equations together yields :

$$\begin{aligned}\frac{1}{2}\sum_{i,\mathcal{E}}\frac{|D_\sigma|}{\delta t}\Big[\rho^{n+1}_\sigma(\boldsymbol{u}^{n+1}_{\sigma,i})^2 - \rho^n_\sigma(\boldsymbol{u}^n_{\sigma,i})^2\Big] + \sum_{K\in\mathcal{M}}\frac{|K|}{\delta t}\;(\mathcal{H}(\rho^{n+1}_K) - \mathcal{H}(\rho^n_K))&\\ + \sum_{i,\mathcal{E}}\frac{|D_\sigma|\;\delta t}{\rho^n_\sigma}\;\Big[|(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}|^2 - |(\boldsymbol{\nabla}\tilde{p}^{n+1})_{\sigma,i}|^2\Big] \le 0,&\end{aligned}$$

which would be a discrete analogue to (6), up to a detail : to obtain a difference of the same quantity taken at two consecutive time steps, we need to change $\rho^n_\sigma\;|(\boldsymbol{\nabla}\tilde{p}^{n+1})_{\sigma,i}|^2$ to $\rho^{n-1}_\sigma\;|(\boldsymbol{\nabla}p^n)_{\sigma,i}|^2$. This is the purpose of the pressure renormalization step, which was already introduced in [28] ; we finally get :

$$\begin{aligned}\frac{1}{2}\sum_{i,\mathcal{E}}\frac{|D_\sigma|}{\delta t}\Big[\rho^{n+1}_\sigma(\boldsymbol{u}^{n+1}_{\sigma,i})^2 - \rho^n_\sigma(\boldsymbol{u}^n_{\sigma,i})^2\Big] + \sum_{K\in\mathcal{M}}\frac{|K|}{\delta t}\;(\mathcal{H}(\rho^{n+1}_K) - \mathcal{H}(\rho^n_K))&\\ + \sum_{i,\mathcal{E}}|D_\sigma|\;\delta t\;\Big[\frac{1}{\rho^n_\sigma}|(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}|^2 - \frac{1}{\rho^{n-1}_\sigma}|(\boldsymbol{\nabla}p^n)_{\sigma,i}|^2\Big] \le 0.&\end{aligned}$$

Note that this inequality yields a control on (δt times) a H^1 discrete semi-norm of the pressure, conforting the robustness of the scheme, but also increasing its dissipation. In our numerical experiments, the pressure renormalization step did not appear to have a significant influence on the results, and was then systematically omitted.

3.2.c Passing to the limit in the scheme (1D case)

We obtain for the pressure correction scheme results which are similar to the implicit scheme ones. They are stated in the following theorem.

Theorem .3.2
Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, u^{(m)}, \tilde{u}^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence satisfies the control over the time and space estimates given by (14), (15) and :

$$\|\tilde{u}^{(m)}\|_{\mathcal{T},x,BV} \leq C, \quad \forall m \in \mathbb{N}.$$

We assume in addition that it converges in $\mathrm{L}^r\big((0,T)\times\Omega\big)^4$, for $1 \leq r < \infty$, to $(\bar{\rho}, \bar{p}, \bar{u}, \bar{\tilde{u}}) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^4$.

Then we have $\bar{\tilde{u}} = \bar{u}$, and the triplet $(\bar{\rho}, \bar{p}, \bar{u})$ satisfies the system (16) and the entropy condition (5).

3.2.d Numerical experiments

We now describe the behaviour of the pressure correction scheme for a Riemann problem, *i.e.* an inviscid monodimensional problem, the initial condition of which consists in two uniform left (L) and right (R) states, separated by a discontinuity, located by convention at the origin $\boldsymbol{x} = 0$. The two initial constant states are given by :

$$\begin{pmatrix} \rho \\ \boldsymbol{u} \end{pmatrix}_L = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} \rho \\ u \end{pmatrix}_R = \begin{pmatrix} 0.125 \\ 0 \end{pmatrix},$$

and the equation of state is given by $p = \rho$. The problem is posed over the interval $(-2, 3)$. The solution of this problem consists in a rarefaction wave travelling to the left and a shock travelling to the right.

The problem is solved with a one dimensional scheme, which may be obtained from the previous exposition by taking one horizontal stripe of meshes (of constant size) with the MAC discretization, and applying perfect slip boundary conditions at the top and bottom boundary.

On Figure 2, we show the solution at $t = 1$ obtained with various meshes and time steps. These latter parameters are adjusted to have $CFL = 1$, taking as reference velocity the sum of the maximum velocity $v = 1$ and the speed of sound $a = 1$. In these computations, we use a centred discretization of the convection term in the momentum balance equation, surprinsingly without observing any spurious oscillations. However, note that results obtained with the CR and RT discretizations (not shown here) differ in this respect : the introduction of a residual viscosity (either physical or by upwinding) is necessary to avoid the odd-even decoupling phenomenon, as usually observed with centred approximations of the convection operator.

We then report, on Figure 3, the obtained numerical error as a function of the time and space step. The observed order of convergence is close to 0.9, for both the velocity and the pressure.

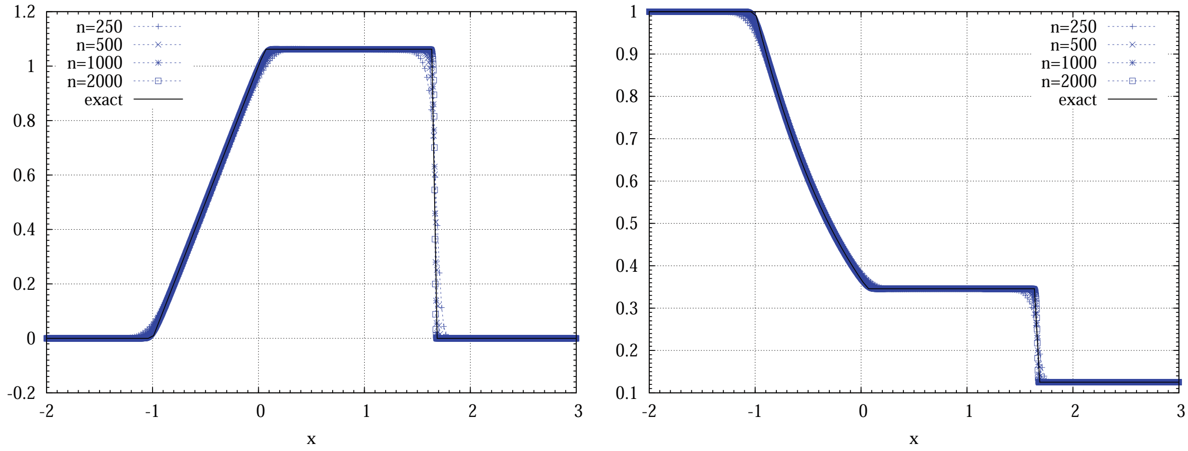

FIG. 2 – Sod shock tube problem – Centred scheme – Exact solution and numerical solution of the problem at $t = 1$ with CFL=1. Velocity (left) and pressure (right).

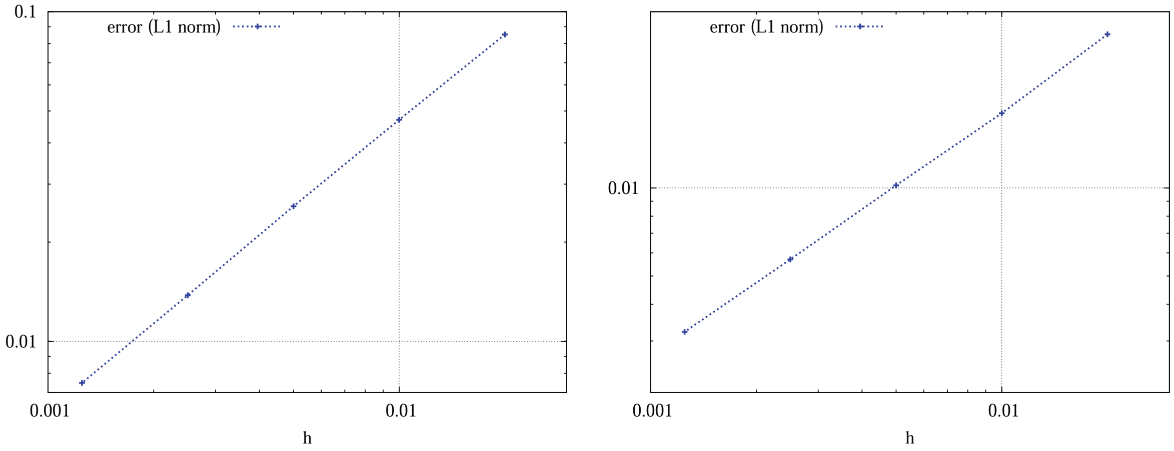

FIG. 3 – Sod shock tube problem – Centred scheme – L^1 norm of the error between the numerical solution and the exact solution at $t = 1$, as a function of the mesh (or time) step, for CFL=1. Velocity (left) and pressure (right).

4 Compressible Navier-Stokes equations

We now address the compressible Navier-Stokes equations (22).

$$\partial_t \rho + \operatorname{div}(\rho\, \boldsymbol{u}) = 0, \tag{22a}$$

$$\partial_t(\rho\, \boldsymbol{u}) + \operatorname{div}(\rho\, \boldsymbol{u} \otimes \boldsymbol{u}) + \boldsymbol{\nabla} p - \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})) = 0, \tag{22b}$$

$$\partial_t(\rho\, E) + \operatorname{div}(\rho\, E\, \boldsymbol{u}) + \operatorname{div}(p\, \boldsymbol{u}) + \operatorname{div}(\boldsymbol{q}) = \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u}) \cdot \boldsymbol{u}), \tag{22c}$$

$$\rho = \wp(p, e), \qquad E = \frac{1}{2}|\boldsymbol{u}|^2 + e, \tag{22d}$$

where E and e are the total energy and internal energy in the flow, and $\boldsymbol{q}$ stands for the heat conduction flux, assumed to be given by :

$$\boldsymbol{q} = -\lambda \boldsymbol{\nabla} e,$$

with $\lambda \geq 0$. We suppose that the equation of state may be set under the form $p = \wp(\rho, e)$ with $\wp(\cdot, 0) = 0$ and $\wp(0, \cdot) = 0$. This system must be complemented by suitable boundary conditions and initial conditions for u, ρ and e, which we suppose positive for the two latter unknowns.

Let us suppose that the solution is regular. Subtracting the kinetic energy balance equation from the total energy balance, we obtain the internal energy balance equation :

$$\partial_t(\rho e) + \operatorname{div}(\rho e \boldsymbol{u}) + p \operatorname{div}(\boldsymbol{u}) = \boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla} \boldsymbol{u}. \tag{23}$$

Since,

- (i) the viscous dissipation term $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla} \boldsymbol{u}$ is non-negative,
- (ii) thanks to the mass balance equation, the first two terms may be recast as a transport operator : $\partial_t(\rho e) + \operatorname{div}(\rho e \boldsymbol{u}) = \rho\,[\partial_t e + \boldsymbol{u} \cdot \boldsymbol{\nabla} e]$,
- (iii) and, finally, because, from the assumption on the equation of state, the pressure vanishes when $e = 0$,

this equation implies that e remains non-negative at all times.

In the framework of incompressible or low Mach number flows, the natural energy balance equation is the internal energy one (23), so discretizing (23) instead of the total energy balance (22c) is a reasonable choice in view to get an algorithm valid for all the flow regimes. In addition, it presents two advantages :

- first, its allow to avoid the space discretization of the total energy, which is rather unatural for staggered schemes since the velocity and the scalar variables are not colocated,
- second, a suitable discretization of (23) may yield, "by construction" of the scheme, the positivity of the internal energy.

However, integrating (22c) over Ω yields a stability estimate for the solution, which reads, if we suppose for short that $\boldsymbol{u}$ is prescribed to zero on the whole boundary $\partial\Omega$, and that the system is adiabatic, *i.e.* $\boldsymbol{\nabla q} \cdot \boldsymbol{n} = 0$ on $\partial\Omega$:

$$\frac{d}{dt} \int_\Omega \big[\frac{1}{2} \rho\, |\boldsymbol{u}|^2 + \rho e\big]\, \mathrm{d}\boldsymbol{x} \leq 0, \tag{24}$$

and we would like (an analogue of) this stability estimate to hold at the discrete level.

In fact, the bridge between the discretization of (23) and this latter inequality is once again the kinetic energy balance equation, and the tools developped in the previous sections will readily yield the desired stability result, if, at the discrete level,we are able :

- (i) to identify the integral of the dissipation term at the right-hand side of the discrete counterpart of (23) with what is obtained from the (discrete) L^2 inner product between the velocity and the diffusion term in the discrete momentum balance equation (22b).
- (ii) to prove that the right-hand side of (23) is non-negative in order to preserve the positivity of the internal energy.

Both properties are quite natural for finite element discretizations, but may be not so easy to obtain for the MAC scheme; for this latter case, a way to build an approximation of the viscous and dissipation terms to get this property is proposed in Chapter 3 (see also [2]).
Two unconditionally stable schemes for the compressible Navier-Stokes equations are built, on the basis of these arguments (Chapter 3) : the first one is implicit, and the second one, used in practice, is a pressure correction scheme. We only describe here this latter, which reads :

Pressure renormalization step – Solve for $\tilde p^{n+1}$:

$$\forall K \in \mathcal{M}, \qquad \sum_{\sigma=K|L} \frac{1}{\rho^n_\sigma} \frac{|\sigma|^2}{|D_\sigma|}\ (\tilde p^{n+1}_K - \tilde p^{n+1}_L) = \sum_{\sigma=K|L} \frac{1}{\sqrt{\rho^n_\sigma\, \rho^{n-1}_\sigma}} \frac{|\sigma|^2}{|D_\sigma|}\ (p^n_K - p^n_L)\,, \tag{25a}$$

Prediction step – Solve for $\tilde{\boldsymbol u}^{n+1}$:

$$\text{For } 1 \le i \le d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\begin{aligned} \frac{|D_\sigma|}{\delta t}(\rho^n_\sigma \tilde{\boldsymbol u}^{n+1}_{\sigma,i} - \rho^{n-1}_\sigma \boldsymbol u^n_{\sigma,i}) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F^n_{\sigma,\varepsilon} \tilde{\boldsymbol u}^{n+1}_{\varepsilon,i} + |D_\sigma|\, (\boldsymbol\nabla \tilde p^{n+1})_{\sigma,i} \\ -|D_\sigma|\, (\mathrm{div}\tau(\tilde{\boldsymbol u}^{n+1}))_{\sigma,i} = 0, \end{aligned} \tag{25b}$$

Correction step – Solve for ρ^{n+1}, p^{n+1}, e^{n+1} and $\boldsymbol u^{n+1}$:

$$\text{For } 1 \le i \le d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\frac{|D_\sigma|}{\delta t}\ \rho^n_\sigma\ (\boldsymbol u^{n+1}_{\sigma,i} - \tilde{\boldsymbol u}^{n+1}_{\sigma,i}) + |D_\sigma|\, \big[(\boldsymbol\nabla p^{n+1})_{\sigma,i} - (\boldsymbol\nabla \tilde p^{n+1})_{\sigma,i} \big] = 0, \tag{25c}$$

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho^{n+1}_K - \rho^n_K) + \sum_{\sigma \in \mathcal{E}(K)} F^{n+1}_{K,\sigma} = 0, \tag{25d}$$

$$\forall K \in \mathcal{M},$$

$$\begin{aligned} \frac{|K|}{\delta t}(\rho^{n+1}_K e^{n+1}_K - \rho^n_K e^n_K) + \sum_{\sigma \in \mathcal{E}(K)} F^{n+1}_{K,\sigma} e^{n+1}_\sigma + |K|\, \big(p^{n+1}\ \mathrm{div}\tilde{\boldsymbol u}^{n+1}\big)_K \\ -\mathrm{div}(\lambda \boldsymbol\nabla e)_K = |K|\, \big(\boldsymbol\tau(\tilde{\boldsymbol u}^{n+1}) : \boldsymbol\nabla \tilde{\boldsymbol u}^{n+1}\big)_K\,, \end{aligned} \tag{25e}$$

$$\forall K \in \mathcal{M}, \qquad p^{n+1}_K = (\gamma - 1)\ \rho^{n+1}_K\ e^{n+1}_K. \tag{25f}$$

The construction of this scheme relies on the same ingredients as in the barotropic case, in particular the time shift of the densities.

The equation (25e) is a approximation of the internal balance over the primal mesh K, which ensures the positivity of the internal energy, thanks to two essential arguments :

- first, the approximation of the convection operator $e \mapsto \partial_t(\rho e) + \mathrm{div}(\rho e \boldsymbol u)$ is upwind (*i.e.* $e^{n+1}_\sigma =$

e_K^{n+1} if $F_{K,\sigma}^{n+1} \geq 0$ and e_L^{n+1} otherwise) and this operator satisfies a consistency property with the mass balance $\partial_t \rho + \mathrm{div}(\rho \boldsymbol{u}) = 0$ which may be stated as the fact that it vanishes if e is constant.

This property is, of course, necessary for an operator to satisfy a discrete maximum principle (constants are necessarily solutions to an equation obeying a maximum principle...) ; it is also classically shown [50] to be sufficient.

- second, the internal energy balance is coupled to the algorithm in such a way that the pressure in the discretization of the term $p\,\mathrm{div}\boldsymbol{u}$ obeys the equation of state, and thus, in particular vanishes when $e < 0$ (see [58] for another pressure-correction algorithm using the same coupling).

The technique used to obtain this result is to define :

$$|K| \, \big(p^{n+1} \, \mathrm{div} \tilde{\boldsymbol{u}}^{n+1}\big)_K = \wp(\rho_K^{n+1}, (e_K^{n+1})^+) \sum_{\sigma \in \mathcal{E}(K)} |\sigma| \; \tilde{\boldsymbol{u}}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}, \tag{26}$$

where $(e_K^{n+1})^+$ stands for the positive part of e_K^{n+1}, *i.e.* $(e_K^{n+1})^+ = \max(e_K^{n+1}, 0)$. Testing then the internal energy balance by the negative part of e_K^{n+1}, designed by $(e_K^{n+1})^- = -\min(e_K^{n+1}, 0)$, and summing over $K \in \mathcal{M}$. Supposing, for short, that the normal velocity vanishes on the boundaries, Lemma .7.2 yields :

$$\sum_{K \in \mathcal{M}} \Big[\frac{|K|}{\delta t} \, \big(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n\big) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} e_\sigma^{n+1} \Big] \; (e_K^{n+1})^- \geq \\ \frac{-1}{\delta t} \sum_{K \in \mathcal{M}} \varrho_K^{n+1} \left[(e_K^{n+1})^-\right]^2 - \varrho_K^n \left[(e_K^n)^-\right]^2,$$

while, $\forall K \in \mathcal{M}$, $\big(p^{n+1} \, \mathrm{div}\tilde{\boldsymbol{u}}^{n+1}\big)_K \; (e_K^{n+1})^- = 0$ and the right-hand side is non-negative, which yields the result. A topological degree argument, applied to the algebraic system corresponding to the whole correction step, yields the existence of at least one solution and, since, for this solution, $e \geq 0$, $(e_K^{n+1})^+ = e_K^{n+1}$ and the discretization (26) is consistent.

The obtained stability result is stated in the following theorem.

Theorem .4.1

There exists a solution to the scheme which satisfies $\rho > 0$, $e > 0$ and for all $n \leq N$, the following inequality holds :

$$\sum_{K \in \mathcal{M}} |K| \; \rho_K^n e_K^n + \frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} |D_\sigma| \; \rho_\sigma^{n-1} |u_\sigma^n|^2 + \frac{\delta t^2}{2} |p^n|^2_{\rho^{n-1}, \, \mathcal{M}} \\ \leq \sum_{K \in \mathcal{M}} |K| \; \rho_K^0 e_K^0 + \frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} |D_\sigma| \; \rho_\sigma^{-1} |u_\sigma^0|^2 + \frac{\delta t^2}{2} |p^0|^2_{\rho^{-1}, \, \mathcal{M}},$$

where, for any discrete pressure q :

$$|q|^2_{\rho, \, \mathcal{M}} = \sum_{\sigma = K|L} \frac{1}{\rho_\sigma} \, \frac{|\sigma|^2}{|D_\sigma|} (p_L - p_K)^2.$$

5 Euler equations

For solutions with shocks, Equation (23) is not equivalent to (22c) ; more precisely speaking, one can show that, at a shock location, a positive measure should replace $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla} \boldsymbol{u}$ (which formally vanishes since $\mu = 0$) at the right-hand side of Equation (23). Discretizing (23) instead of (22c) may thus yield a scheme which does not compute the correct weak discontinuous solutions, the manifestation of this non-consistency being that the numerical solutions present shocks which do not satisfy the Rankine-Hugoniot conditions associated to (22c). The essential result of this section is to provide solutions to circumvent this problem.

This study is closely related to the analysis performed in the barotropic case. Indeed, it may be checked that the entropy of the barotropic problem takes an expression similar to the total energy E (in fact, if the equation of state in the barotropic case is derived by supposing that the flow is isentropic, we have the exact equality $\mathcal{H} = \rho e$) ; the elastic potential balance (in the barotropic case) plays the same role as the internal energy balance (in the non-barotropic case). The only difference is that the entropy condition is an inequality while the total energy is an equality : in other words, while, for the barotropic case, we just checked that residual terms were non-positive, we now have to ensure that they vanish with the discretization steps. To this purpose, we thus follow a strategy quite similar to Section 3 :

- Starting from the discrete momentum balance equation, with an *ad hoc* discretization of the convection operator, we derive a discrete kinetic energy balance ; residual terms are present in this relation, which do no tend to zero with space and time steps (they are the discrete manifestations of the the above mentioned measures).
- These residual terms are then compensated by source terms added to the internal energy balance.

We provide a theoretical justification of this process by showing that, in the 1D case, if the scheme is stable enough and converges to a limit (in a sense to be defined), this limit satisfies a weak form of (22c) which implies the correct Rankine-Hugoniot conditions. Then, we perform numerical tests which substantiate this analysis. Two different time discretizations are proposed in Chapter 4 : first, a fully implicit scheme (a solution to which may be rather difficult to obtain in practice) and, second, a pressure correction scheme (the algorithm indeed used in the tests presented here) ; we only present here the latter algorithm.

5.1 The discrete kinetic energy balance equation and the corrective source terms

We derive here a slightly different discrete kinetic energy balance than in Section 3.2.b. Our starting point, however, is still the velocity prediction step which we multiply by the corresponding unknown, *i.e.* Equation (20), which now reads, since, in the present algorithm, we omit the pressure renormalization step :

$$\frac{|D_\sigma|}{\delta t}\ \left(\rho^n_\sigma \tilde{\boldsymbol u}^{n+1}_{\sigma,i} - \rho^{n-1}_\sigma \boldsymbol u^n_{\sigma,i}\right)\ \tilde{\boldsymbol u}^{n+1}_{\sigma,i} + \sum_{\varepsilon\in\bar{\mathcal E}(D_\sigma)} F^n_{\sigma,\varepsilon}\tilde{\boldsymbol u}^{n+1}_{\varepsilon,i}\ \tilde{\boldsymbol u}^{n+1}_{\sigma,i} + |D_\sigma|\,(\boldsymbol\nabla p^n)_{\sigma,i}\ \tilde{\boldsymbol u}^{n+1}_{\sigma,i} = 0.$$

The next step is to multiply the velocity correction equation by $\tilde{\boldsymbol u}^{n+1}_{\sigma,i}$ and use the identity $2a(a-b) = a^2 + (a-b)^2 - b^2$ to get :

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho^n_\sigma(\boldsymbol u^{n+1}_{\sigma,i})^2 - \rho^n_\sigma(\tilde{\boldsymbol u}^{n}_{\sigma,i})^2\Big] + |D_\sigma|\,\Big[(\boldsymbol\nabla p^{n+1})_{\sigma,i} - (\boldsymbol\nabla p^n)_{\sigma,i}\Big]\ \tilde{\boldsymbol u}^{n+1}_{\sigma,i} \\ + \frac{|D_\sigma|}{2\,\delta t}\ \rho^n_\sigma\ \left(\boldsymbol u^{n+1}_{\sigma,i} - \tilde{\boldsymbol u}^{n+1}_{\sigma,i}\right)^2 = 0.$$

Invoking Lemma .7.2 for the first two terms of the first of these relations and summing with the second one yields :

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho^n_\sigma(\boldsymbol{u}^{n+1}_{\sigma,i})^2-\rho^{n-1}_\sigma(\boldsymbol{u}^n_{\sigma,i})^2\Big]+\frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}}F^n_{\sigma,\varepsilon}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma',i}$$
$$+\,|D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_{\sigma,i}\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}=R^{n+1}_{\sigma,i},\quad (27)$$

with :

$$R^{n+1}_{\sigma,i}=\frac{|D_\sigma|}{2\,\delta t}\;\rho^n_\sigma\;\big(\boldsymbol{u}^{n+1}_{\sigma,i}-\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}\big)^2-\frac{|D_\sigma|}{2\,\delta t}\;\rho^{n-1}_\sigma\;\big(\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}-\boldsymbol{u}^n_{\sigma,i}\big)^2$$
$$-\,\delta^{\rm up}\Big[\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\;|F^n_{\sigma,\varepsilon}|\;\big(\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}-\tilde{\boldsymbol{u}}^{n+1}_{\sigma',i}\big)\Big]\;\tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}.$$

We recognize at the left-hand side a conservative discrete kinetic energy balance. The next step is now to deal with the residual terms at the right-hand side, or, more precisely speaking, to somewhat compensate them by some source term which we introduced in the internal energy balance. Let us denote by S^{n+1}_K the source term in the balance over the cell K. We choose :

$$\forall K\in\mathcal{M},$$
$$S^{n+1}_K=\frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)}\frac{|D_{K,\sigma}|}{\delta t}\rho^{n-1}_K\;\big(\tilde{\boldsymbol{u}}^{n+1}_\sigma-\boldsymbol{u}^n_\sigma\big)^2-\frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)}\frac{|D_{K,\sigma}|}{\delta t}\rho^n_K\;\big(\boldsymbol{u}^{n+1}_\sigma-\tilde{\boldsymbol{u}}^{n+1}_\sigma\big)^2$$
$$+\,\delta^{\rm up}\sum_{\substack{\varepsilon\cap\bar{K}\neq\emptyset,\\ \varepsilon=D_\sigma|D_{\sigma'}}}\alpha_{K,\varepsilon}\;\frac{|F^{n+1}_{\sigma,\varepsilon}|}{2}(\boldsymbol{u}^{n+1}_\sigma-\boldsymbol{u}^{n+1}_{\sigma'})^2.\quad (28)$$

The coefficient $\alpha_{K,\varepsilon}$ is fixed to 1 if the face ε is included in K, and this is the only situation to consider for the RT and CR discretization. For the MAC scheme, some dual edges are included in the primal cells, whereas some lie on their boundary ; for ε on a cell boundary, we denote by $\mathcal{N}_\varepsilon$ the set of cells M such that $\bar{M}\cap\varepsilon\neq\emptyset$ (the cardinal of this set is always 4), and compute $\alpha_{K,\varepsilon}$ by :

$$\alpha_{K,\varepsilon}=\frac{|K|}{\sum_{M\in\mathcal{N}_\varepsilon}|M|}.$$

For a uniform grid, this formula yields $\alpha_{K,\varepsilon}=1/4$.

The expression of the terms $(S_K)_{K\in\mathcal{M}}$ is justified by the passage to the limit in the scheme (for a one-dimensional problem) performed in Section 5.2. Let us just here remark that :

$$\sum_{K\in\mathcal{M}}S^{n+1}_K+\sum_{\mathcal{E},i}R^{n+1}_{\sigma,i}=0,$$

which shows that the introduction of this term allows to recover the total energy balance over the whole computational domain Ω. Note however that, the term S^{n+1}_K may be negative, which we have indeed observed in computations, and so the above proof of the positivity of the internal energy is not valid here ; however, even in very severe cases (as, for instance, Test 3 of [68, chapter 4]), at least with a reasonable time step, we still obtained $e>0$.

Remark 2 (Form of the corrective source terms)
Comparing with the source term of the continuous internal energy balance (23), it is easy to identify in the last part of S_K the viscous dissipation associated to the numerical diffusion introduced by the upwinding. In fact, this analogy also holds for the first two terms : they are dissipation and antidissipation terms associated to the diffusion and antidiffusion introduced by the semi-implicit time discretization.
Note by the way that only a dissipation term is obtained for the implicit case (*i.e.* the corrective terms S_K^{n+1} are non-negative, see Chapter 4), and thus, for this time discretization, the positivity of the internal energy is ensured.

Remark 3 (On the necessity of the corrective source terms)
Let us consider a sequence of discretizations $(\mathcal{M}^{(m)},\ \delta t^{(m)})_{m\in\mathbb{N}}$, the space and time steps of which tend to zero, an associated sequence of discrete velocities $(\boldsymbol{u}^{(m)})_{m\in\mathbb{N}}$, and the corresponding sequence of (piecewise constant functions associated to the) corrective term $(S^{(m)})_{m\in\mathbb{N}}$. It may be checked that $S^{(m)}$ tends to zero in $\mathrm{L}^1(\Omega\times(0,T))$ as soon as the time and space derivatives of the functions $(\boldsymbol{u}^{(m)})_{m\in\mathbb{N}}$ are bounded in a strong enough norm, and in particular stronger than the BV norm (for instance, suppose that the jumps between two consecutive time steps and adjacent cells are bounded by δt and h respectively), *i.e.* everywhere the solution is regular. On the opposite, for a sequence $(\boldsymbol{u}^{(m)})_{m\in\mathbb{N}}$ obtained by projecting a discontinuous function $\boldsymbol{u}$, $S^{(m)}$ does not tend to zero.

5.2 Passing to the limit in the scheme

As for the barotropic equations, we now pass to the limit in the scheme.

We suppose given a sequence of meshes and time steps $(\mathcal{M}^{(m)},\delta t^{(m)})_{m\in\mathbb{N}}$, such that the time step and the size $h^{(m)}$ of the mesh $\mathcal{M}^{(m)}$, defined by :

$$h^{(m)} =\ \sup_{K\in\mathcal{M}^{(m)}} \mathrm{diam}(K),$$

tend to zero as $m\to\infty$.

Let $\rho^{(m)}$, $p^{(m)}$, $e^{(m)}$, $\tilde{u}^{(m)}$ and $u^{(m)}$ be the associated solution of the pressure correction scheme (25) with the mesh $\mathcal{M}^{(m)}$ and the time step $\delta t^{(m)}$ (or, more precisely speaking, as in the barotropic case, a 1D version of the scheme). To the discrete unknowns, we associate piecewise constant functions on time intervals and on primal or dual meshes :

$$\rho^{(m)}(x,t)=\sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}}(p^{(m)})_K^n\ \mathcal{X}_K\ \mathcal{X}_{(n,n+1)},\qquad p^{(m)}(x,t)=\sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}}(\rho^{(m)})_K^n\ \mathcal{X}_K\ \mathcal{X}_{(n,n+1)},$$
$$e^{(m)}(x,t)=\sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}}(e^{(m)})_K^n\ \mathcal{X}_K\ \mathcal{X}_{(n,n+1)},\qquad u^{(m)}(x,t)=\sum_{n=0}^{N-1}\sum_{\sigma\in\mathcal{E}}(u^{(m)})_\sigma^n\ \mathcal{X}_{D_\sigma}\ \mathcal{X}_{(n,n+1)},$$
$$\tilde{u}^{(m)}(x,t)=\sum_{n=0}^{N-1}\sum_{\sigma\in\mathcal{E}}(\tilde{u}^{(m)})_\sigma^n\ \mathcal{X}_{D_\sigma}\ \mathcal{X}_{(n,n+1)}.$$

We suppose that the sequence of discrete solutions $\big(\rho^{(m)},p^{(m)},e^{(m)},u^{(m)},\tilde{u}^{(m)}\big)_{m\in\mathbb{N}}$ is uniformly bounded in $\mathrm{L}^\infty\big((0,T)\times\Omega\big)$, *i.e.* :

$$|(\rho^{(m)})_K^n|+|(p^{(m)})_K^n|+|(e^{(m)})_K^n|\le C,\quad \forall K\in\mathcal{M}^{(m)},\ \text{for } 0\le n\le N^{(m)},\ \forall m\in\mathbb{N}, \tag{29}$$

and :

$$|(u^{(m)})^n_\sigma| + |(\tilde{u}^{(m)})^n_\sigma| \leq C, \quad \forall \sigma \in \mathcal{E}^{(m)},\ \text{for } 0 \leq n \leq N^{(m)},\ \forall m \in \mathbb{N}. \tag{30}$$

In addition, we also suppose the following uniform control on the translates in space and time :

$$\|\rho^{(m)}\|_{\mathcal{T},x,BV} + \|e^{(m)}\|_{\mathcal{T},x,BV} + \|u^{(m)}\|_{\mathcal{T},x,BV} + \|\tilde{u}^{(m)}\|_{\mathcal{T},x,BV} \leq C, \quad \forall m \in \mathbb{N}, \tag{31}$$

and :

$$\|\rho^{(m)}\|_{\mathcal{T},t,BV} + \|u^{(m)}\|_{\mathcal{T},t,BV} \leq C, \quad \forall m \in \mathbb{N}. \tag{32}$$

As in the barotropic case, we are not able to prove the estimates (29)–(32) for the solutions of the scheme, but such inequalities are satisfied by the "interpolation" of the solution to a Riemann problem, and are observed in computations (of course, as far as possible, *i.e.* with a limited sequence of meshes and time steps).

A weak solution to the continuous problem satisfies, for any $\varphi \in \mathrm{C}^\infty_c\big([0,T) \times \Omega\big)$:

$$-\int_{\Omega\times(0,T)} \Big[\rho\,\partial_t\varphi + \rho\,u\,\partial_x\varphi\Big]\,\mathrm{d}x\delta t - \int_\Omega \rho(x,0)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{33a}$$

$$-\int_{\Omega\times(0,T)} \Big[\rho\,u\,\partial_t\varphi + (\rho\,u^2 + p)\,\partial_x\varphi\Big]\,\mathrm{d}x\delta t - \int_\Omega \rho(x,0)\,u(x,0)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{33b}$$

$$-\int_{\Omega\times(0,T)} \Big[\rho\,E\,\partial_t\varphi + (\rho\,E + p)\,u\,\partial_x\varphi\Big]\,\mathrm{d}x\delta t - \int_\Omega \rho(x,0)\,E(x,0)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{33c}$$

$$\rho = \wp(p,e), \qquad E = \frac{1}{2}u^2 + e. \tag{33d}$$

Once again, since the test function φ vanishes at the boundary, these relations do not imply anything about the boundary conditions, but imply the Rankine-Hugoniot conditions. The scheme consistency result that we are seeking for is stated in the following theorem.

Theorem .5.1
Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, e^{(m)}, u^{(m)}, \tilde{u}^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence satisfies (29)–(32) and converges in $\mathrm{L}^r\big((0,T)\times\Omega\big)^5$, for $1 \leq r < \infty$, to a limit $(\bar{\rho}, \bar{p}, \bar{e}, \bar{u}, \bar{\tilde{u}}) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^5$.

Then $\bar{\tilde{u}} = \bar{u}$ and the limit $(\bar{\rho}, \bar{p}, \bar{e}, \bar{u})$ satisfies the system (33).

5.3 Numerical tests

We now assess the behaviour of the scheme on a one dimensional Riemann problem. We choose initial conditions such that the structure of the solution consists in two shock waves, separated by the contact discontinuity, with sufficiently strong shocks to allow to easily discrimate between convergence to the correct weak solution or not. These initial conditions are those proposed in [68, chapter 4], for the test referred to as Test 5 :

$$\text{left state : } \begin{bmatrix}\rho_L\\ u_L\\ p_L\end{bmatrix} = \begin{bmatrix}5.99924\\ 19.5975\\ 460.894\end{bmatrix} \qquad \text{right state : } \begin{bmatrix}\rho_R\\ u_R\\ p_R\end{bmatrix} = \begin{bmatrix}5.99242\\ -6.19633\\ 46.0950\end{bmatrix}$$

The problem is posed over $\Omega = (-0.5, 0.5)$, and the discontinuity is initially located at $x = 0$.

Since numerical experiments addressing barotropic flows (see Section 3.2.d) show that, at least for one dimensional computations, it is not necessary to use upwinding in the momentum balance equation, we only employ a centered approximation of the velocity at the dual faces.

The density fields obtained with $h = 1/2000$ (or a number of cells $n = 2000$) at $t = 0.035$, with and without assembling the corrective source term in the internal energy balance, are shown together with the analytical solution on Figure 4. The density and the pressure obtained, still with and without corrective terms, for various meshes, are plotted on Figure 5 and 6 respectively. For these computations, we take $\delta t = h/20$, which yields a CFL number, with respect to the material velocity only, close to one. The first conclusion is that both schemes seem to converge, but the corrective term is necessary to obtain the correct solution. In this case, for instance, we obtain the correct intermediate state for the pressure and velocity up to four digits in the essential part of the corresponding zone :

$$\text{(analytical) intermediate state :} \quad \begin{bmatrix} p^* \\ u^* \end{bmatrix} = \begin{bmatrix} 1691.65 \\ 8.68977 \end{bmatrix} \quad \text{for } x \in (0.028, 0.428)$$

$$\text{numerical results :} \quad \left| \begin{array}{l} p \in (1691.6, 1691.8) \\ u \in (8.689, 8.690) \end{array} \right. \quad \text{for } x \in (0.032, 0.417)$$

One can check that the solution obtained without the corrective term is not a weak solution to the Euler system.

We also observe that the scheme is rather diffusive, specially at the contact discontinuity, where the beneficial compressive effect of the shocks does not apply.

6 Conclusion and perspectives

We developed a class of schemes for barotropic and non-barotropic flows, based on staggered space discretizations and on a fractional time-stepping technique falling in the class of pressure correction methods. Upwinding is performed in an equation-by-equation way, and only with respect to the material velocity ; for non-barotropic equations, the energy equation is the internal energy balance. All of these characteristics ensure that the schemes boil down to usual incompressible flow solvers for a vanishing Mach number ; therefore they are hoped to be stable and accurate in the whole incompressible to compressible range. Numerical tests performed here focus on compressible flows, and assess the fact that weak solution to inviscid problems are correctly computed ; they are supported by theoretical arguments. These tests will be continued, adressing complex multi-dimensional geometries.

From an algorithmic point of view, let us first mention that, for high Mach number flows, explicit versions of these schemes are now under development [59] ; this would provide efficient algorithms (in particular, with an immediate construction of the fluxes at the cell faces), well suited to fast transient regimes, and offering, if necessary, the possibility of a partial implicitation without loosing any stability features (by the schemes studied in this work). In explicit schemes, less diffusive space discretizations, such as MUSCL-like or adaptative numerical viscosity [30, 31] techniques, are easy to implement ; this will be done in a near future.

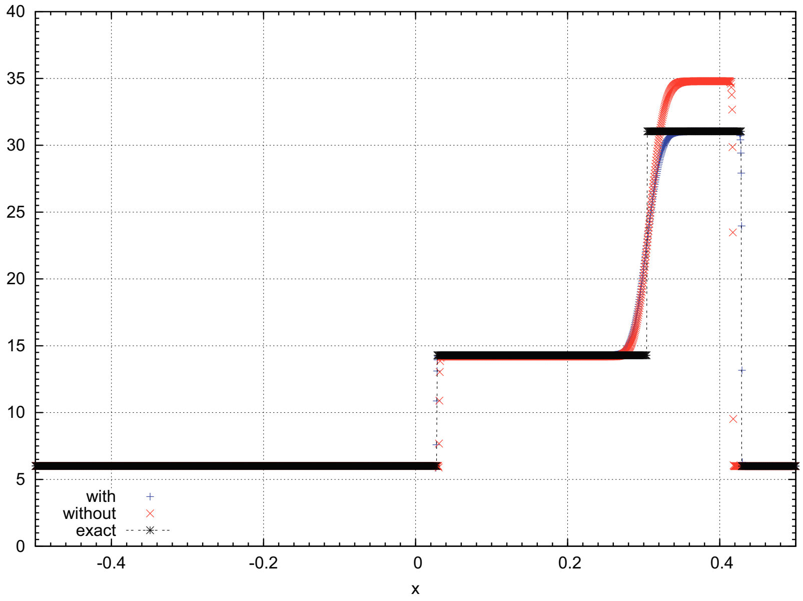

FIG. 4 – Test 5 of [68, chapter 4] - Density obtained with $n = 2000$ cells, with and without corrective source terms, and analytical solution.

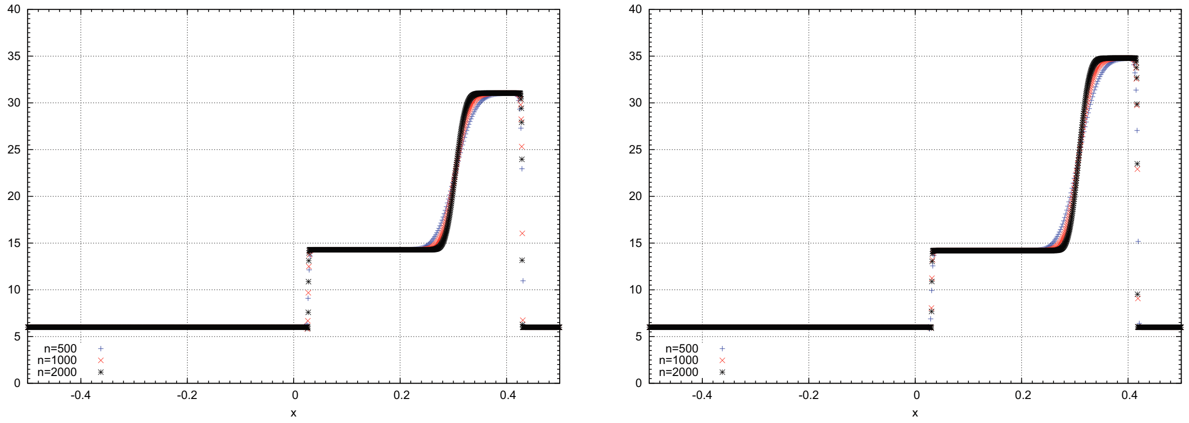

FIG. 5 – Test 5 of [68, chapter 4] - Density obtained with various meshes, with (left) and without (right) corrective source terms.

A lot of theoretical questions are suggested by the present study. First, the passage to the limit in the schemes in the multi-dimensional case raises only technical problems, which should not be so difficult to fix. A more intricate question is that of the boundary conditions, which was not addressed here (except through some numerical experiments described in Chapter 1) : the decoupling of pressure correction schemes is known to produce inherent spurious boundary conditions, the effect of which is extensively discussed for incompressible flows ; for compressible problems, this question seems to remain largely open, and should deserve to be studied in the future. We did not prove in this work that the solutions obtained

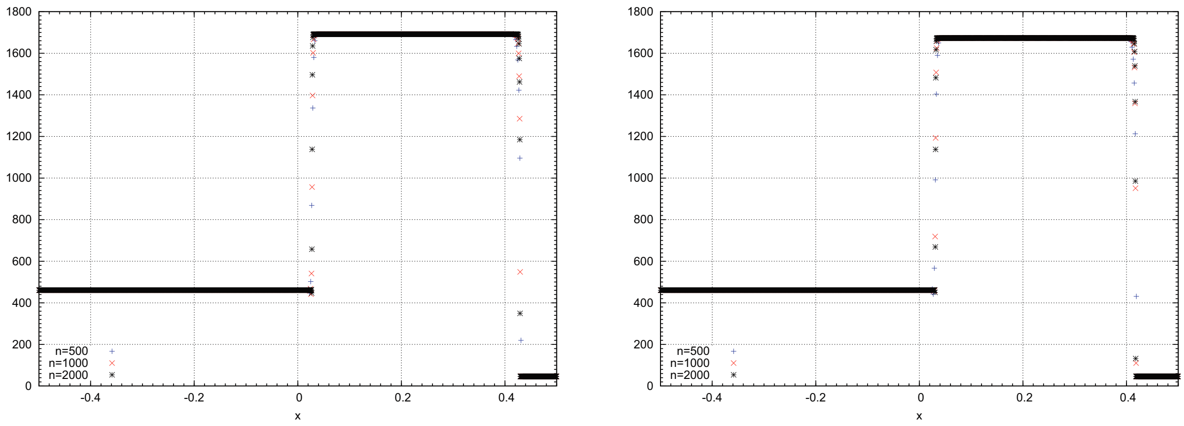

Fig. 6 – Test 5 of [68, chapter 4] - pressure obtained with various meshes, with (left) and without (right) corrective source terms.

for non-barotropic Euler equations, if they converge, tend to the entropy weak solution ; this is another issue to be addressed in the near future. Yet another topic is to analyse more indepth the behaviour of the proposed schemes in the low Mach number regime. In particular, since these algorithms satisfy stability estimates, it seems possible, at least with a fixed mesh and using the norm equivalence property of finite dimensional problems, to pass to the limit for a vanishing Mach number. This is interesting both theoretically and from an engineering point of view, to get some insight in what physical model is indeed solved by the code in such situations. Last but not least, we performed here some parts of the convergence analysis ; this should be continued as far as possible, in particular for barotropic viscous flows.

7 APPENDIX – Some results associated to finite volume convection operators

We gather in this section some results concerning the discretization by the finite volume method of two convection operators :

- the first one reads, at the continuous level, $\rho \rightarrow \mathcal{C}(\rho) = \partial_t \rho + \mathrm{div}(\rho \boldsymbol{u})$, where $\boldsymbol{u}$ stands for a given velocity field, which is not assumed to satisfy any divergence constraint,
- the second one is $z \rightarrow \mathcal{C}_\rho(z) = \partial_t(\rho z) + \mathrm{div}(\rho z \boldsymbol{u})$, where ρ and $\boldsymbol{u}$ stands for two given scalar and vector fields, which are supposed to satisfy $\partial_t \rho + \mathrm{div}(\rho \boldsymbol{u}) = 0$.

Multiplying these operators by functions depending on the unknown is currently used to obtain convection operators acting over different variables, possibly with residual terms : one may think, for instance, to the theory of renormalized solutions (for the first one), or, in mechanics, to the derivation of the so-called kinetic energy transport identity (for the second one). The results provided in this section are discrete variants of such relations.

We begin with a property of $\mathcal{C}$, which, at the continuous level, may be formally obtained as follows. Let ψ be a regular function from $(0,+\infty)$ to $\mathbb{R}$; then :

$$\begin{aligned}\psi'(\rho)\ \mathcal{C}(\rho) = \psi'(\rho)\,\partial_t(\rho) + \psi'(\rho)\,\boldsymbol{u}\cdot\boldsymbol{\nabla}\rho + \psi'(\rho)\,\rho\ \mathrm{div}\boldsymbol{u}\\ = \partial_t(\psi(\rho)) + \boldsymbol{u}\cdot\boldsymbol{\nabla}\psi(\rho) + \rho\,\psi'(\rho)\,\mathrm{div}\boldsymbol{u},\end{aligned}$$

so adding and subtracting $\psi(\rho)\,\mathrm{div}\boldsymbol{u}$ yields :

$$\psi'(\rho)\ \mathcal{C}(\rho) = \partial_t\big(\psi(\rho)\big) + \mathrm{div}\big(\psi(\rho)\boldsymbol{u}\big) + \big(\rho\psi'(\rho) - \psi(\rho)\big)\ \mathrm{div}\boldsymbol{u}. \tag{34}$$

Obtaining a proof of this last identity, in a weak sense and with minimal regularity assumptions for ρ and $\boldsymbol{u}$ and increasing properties of ψ is the object of the theory of renormalized solutions. The following lemma states a discrete analogue to (34).

Lemma .7.1
Let $K \in \mathcal{M}$. Let ρ_K^* and ρ_K be two positive real numbers. For $\sigma \in \mathcal{E}(K)$, let F_σ be a quantity associated to the face σ and the control volume K, defined by

$$\forall\sigma \in \mathcal{E}(K), \qquad F_\sigma = \rho_\sigma\ V_\sigma.$$

where ρ_σ and V_σ are a positive real number and a real number respectively, both associated to the edge σ. Let ψ be a twice continuously differentiable function, defined over $(0,+\infty)$.

Then the following identity holds :

$$\begin{aligned}\Big[\frac{|K|}{\delta t}\ (\rho_K - \rho_K^*) + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\Big]\psi^{'}(\rho_K) = \frac{|K|}{\delta t}\ \left[\psi(\rho_K) - \psi(\rho_K^*)\right] + \sum_{\sigma\in\mathcal{E}(K)} \psi(\rho_\sigma)\ V_\sigma\\ + \Big[\rho_K\psi^{'}(\rho_K) - \psi(\rho_K)\Big] \sum_{\sigma\in\mathcal{E}(K)} V_\sigma + R_{\sigma,\delta t}\end{aligned} \tag{35}$$

where

$$R_{\sigma,\delta t} = \frac{1}{2}\frac{|K|}{\delta t}\psi^{''}(\overline{\rho}_K)(\rho_K - \rho_K^*)^2 - \frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)} V_\sigma\ \psi''(\overline{\rho}_\sigma)(\rho_\sigma - \rho_K)^2, \tag{36}$$

and, $\forall\sigma \in \mathcal{E}(K)$, $\overline{\rho}_K \in [\min(\rho_K,\rho_K^*),\max(\rho_K,\rho_K^*)]$ and $\overline{\rho}_\sigma \in [\min(\rho_\sigma,\rho_K),\max(\rho_\sigma,\rho_K)]$.
If we now suppose that the function ψ is once differentiable and convex, and that $\rho_\sigma = \rho_K$ as soon as $V_\sigma \geq 0$, then expression (36) in no-more valid but the residual $R_{\sigma,\delta t}$ is non-negative.

We now turn to the second operator, for which we have, at the continuous level and formally, using twice the assumption $\partial_t\rho + \mathrm{div}(\rho\boldsymbol{u}) = 0$:

$$\begin{aligned}\psi'(z)\ \mathcal{C}_\rho(z) = \psi'(z)\big[\partial_t(\rho\, z) + \mathrm{div}(\rho\, z\,\boldsymbol{u})\big] = \psi'(z)\rho\big[\partial_t z + \boldsymbol{u}\cdot\boldsymbol{\nabla}z\big]\\ = \rho\big[\partial_t\psi(z) + \boldsymbol{u}\cdot\boldsymbol{\nabla}\psi(z)\big] = \partial_t\big(\rho\,\psi(z)\big) + \mathrm{div}\big(\rho\,\psi(z)\,\boldsymbol{u}\big).\end{aligned}$$

Taking for z a component of the velocity field and $\psi(s) = s^2/2$, this relation is the central argument used to derive the kinetic energy balance. The following lemma states a discrete counterpart of this identity.

LEMMA .7.2
Let $K \in \mathcal{M}$. Let ρ_K^* and ρ_K be two positive real numbers. For $\sigma \in \mathcal{E}(K)$, let F_σ be a quantity associated to the face σ, such that the following identity holds :

$$\frac{|K|}{\delta t}\,(\rho_K - \rho_K^*) + \sum_{\sigma \in \mathcal{E}(K)} F_\sigma = 0. \tag{37}$$

Let u_K^* and u_K be two real numbers, and, to each $\sigma \in \mathcal{E}(K)$, we associate a rela number u_σ. Let ψ be a twice continuously differentiable function, defined over $(0, +\infty)$. Then the following relation holds :

$$\begin{aligned} \Big[\frac{|K|}{\delta t}\,\big(\rho_K\, u_K - \rho_K^*\, u_K^*\big) + \sum_{\sigma \in \mathcal{E}(K)} F_\sigma\ u_\varepsilon\Big]\ \psi'(u_K) \\ = \frac{|K|}{\delta t}\,\big[\rho_K\, \psi(u_K) - \rho_K^*\, \psi(u_K^*)\big] + \sum_{\sigma \in \mathcal{E}(K)} F_\sigma\ \psi(u_\sigma) + R_{K,\delta t} \end{aligned} \tag{38}$$

where :

$$R_{K,\delta t} = \frac{1}{2}\frac{|K|}{\delta t}\rho_K^*\ \psi''(\overline{u}_K)(u_K - u_K^*)^2 - \frac{1}{2}\sum_{\sigma \in \mathcal{E}(K)} F_\sigma\ \psi''(\overline{u}_\sigma)\ (u_\sigma - u_K)^2, \tag{39}$$

with, $\overline{u}_K \in [\min(u_K, u_K^*), \max(u_K, u_K^*)]$ and, $\forall \sigma \in \mathcal{E}(K)$, $\overline{u}_\sigma \in [\min(u_\sigma, u_K), \max(u_\sigma, u_K)]$.
If we now suppose that the function ψ is once continuously differentiable and convex, and that $u_\sigma = u_K$ as soon as $F_\sigma \geq 0$, then expression (39) is no more valid but the residual $R_{\sigma,\delta t}$ is non-negative.

If we now take for ψ the function $\psi(s) = s^2/2$ and write, $\forall \sigma \in \mathcal{E}(K)$, $u_\sigma = (u_K + u_{\substack{K|\bullet \\ \sigma}})/2$ (or, in other words, define $u_{\substack{K|\bullet \\ \sigma}}$ as $u_{\substack{K|\bullet \\ \sigma}} = 2\, u_\sigma - u_K$), we get the following identity :

$$\begin{aligned} \Big[\frac{|K|}{\delta t}\,\big(\rho_K\, u_K - \rho_K^*\, u_K^*\big) + \sum_{\sigma \in \mathcal{E}(K)} F_\sigma\ u_\varepsilon\Big]\ u_K \\ = \frac{1}{2}\ \frac{|K|}{\delta t}\,\big[\rho_K\, u_K^2 - \rho_K^*\, (u_K^*)^2\big] + \sum_{\sigma \in \mathcal{E}(K)} F_\sigma\ u_K\ u_{\substack{K|\bullet \\ \sigma}} + R_{K,\delta t}, \end{aligned} \tag{40}$$

with $\quad R_{K,\delta t} = \dfrac{1}{2}\dfrac{|K|}{\delta t}\rho_K^*\ (u_K - u_K^*)^2.$

Chapitre

I Pressure correction staggered schemes for barotropic monophasic and two-phase flows

We assess in this paper the capability of a pressure correction scheme to compute irregular solutions (i.e. solutions with shocks) of the homogeneous model for barotropic two-phase flows. This scheme is designed to inherit the stability properties of the continuous problem : the unknowns (in particular the density and the dispersed phase mass fraction y) are kept within their physical bounds, and the entropy of the system is conserved, thus providing an unconditional stability property. In addition, the scheme keeps the velocity and pressure constant through contact discontinuities. These properties are obtained by coupling the mass balance and the transport equation for y in an original pressure correction step. The space discretization is staggered ; the numerical schemes which are considered are the Marker-And Cell (MAC) finite volume scheme and the nonconforming low-order Rannacher-Turek (RT) finite element approximation ; in either case, a finite volume technique is used for all convection terms. Numerical experiments performed here address the solution of various Riemann problems, often called in this context "shock tube problems". They show that, provided that a sufficient dissipation is introduced in the scheme, it converges to the (weak) solution of the continuous hyperbolic system. Observed orders of convergence as a function of the mesh and time step at constant CFL number vary with the studied case and the CFL number, and range from 0.5 *to* 1.5 *for the velocity and the pressure ; in most cases, the density and mass fraction converge with a* 0.5 *order. Finally, the scheme shows a satisfactory behaviour up to large CFL numbers.*

Plan du Chapitre I

I.1 Introduction **33**

I.2 The pressure correction scheme **35**

I.2.1 Time semi-discretization 35

I.2.2 Discrete spaces and unknowns 36

I.2.3 Discrete equations 38

I.3 Numerical experiments : Riemann problems **41**

I.3.1 Sod shock tube problem 42

I.3.2 Two-fluid shock tube 48

I.3.2.a First case : shock – contact discontinuity – shock 48

I.3.2.b Second case : rarefaction-contact discontinuity-rarefaction 55

I.4 A two-dimensional test case **57**

I.5 Conclusion **61**

I.1 Introduction

We address in this paper the following homogeneous two-phase flow model, describing for instance the flow of a mixture between a liquid phase and a gas phase :

$$\partial_t \rho + \operatorname{div}(\rho\, \boldsymbol{u}) = 0, \tag{I.1}$$

$$\partial_t z + \operatorname{div}(z\, \boldsymbol{u}) = 0, \tag{I.2}$$

$$\partial_t(\rho\, \boldsymbol{u}) + \operatorname{div}(\rho\, \boldsymbol{u} \otimes \boldsymbol{u}) + \nabla p - \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})) = 0, \tag{I.3}$$

where ∂_t is the time derivative, ρ, $\boldsymbol{u}$ and p are the (average) density, velocity and pressure in the flow, z stands for the partial density of the gas phase. The first two equations, (I.1) and (I.2), are the mixture and the gas mass balance equations respectively, and the third equation (I.3) is the mixture momentum balance equation. The tensor $\boldsymbol{\tau}$ is the viscous part of the stress tensor, given by the following expression :

$$\boldsymbol{\tau}(\boldsymbol{u}) = \mu\,(\boldsymbol{\nabla} \boldsymbol{u} + \boldsymbol{\nabla}^t \boldsymbol{u}) - \frac{2}{3}\,\mu\,(\operatorname{div}\boldsymbol{u})\; I. \tag{I.4}$$

The viscosity μ is supposed to be constant (in space), so this relation yields :

$$\operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})) = \mu \left[\Delta \boldsymbol{u} + \frac{1}{3}\, \boldsymbol{\nabla} \operatorname{div}\boldsymbol{u}\right]. \tag{I.5}$$

The problem is defined over an open bounded connected subset Ω of $\mathbb{R}^d$, $d \leq 3$, and over a finite time interval $(0,T)$. We suppose that suitable initial and boundary conditions are provided for ρ, $\boldsymbol{u}$ and z ; in particular, the prescribed values for ρ and z are supposed to be positive, and ρ, $\boldsymbol{u}$ and z are supposed to be prescribed at the inflow boundaries.

To close the problem, we need an additional relation, which results from the combination of the mixture equation of state and of phasic equations of state, *i.e.* relations satisfied by the density of each phase. Let us begin with the latter ones. The liquid density ρ_ℓ is supposed to be constant, and the gas density ϱ_g is assumed to depend on the pressure only :

$$\rho_g = \varrho_g(p), \tag{I.6}$$

where ϱ_g is defined and increasing over $[0, +\infty)$, $\varrho_g(0) = 0$ and $\lim_{s \to +\infty} \varrho_g(s) = +\infty$. Such a flow (*i.e.* a flow where phasic densities are functions of the pressure only) is usually referred to as a barotropic flow. Finally, the mixture equation of state is usually written :

$$\rho = (1-\alpha)\,\rho_\ell + \alpha\,\rho_g, \quad z = \alpha\rho_g = \rho y, \qquad \text{or} \quad \rho = \frac{1}{\dfrac{y}{\rho_g} + \dfrac{1-y}{\rho_\ell}}, \tag{I.7}$$

where α is called the void fraction (the volume of gas per specific volume), and $y = z/\rho$ is the gas mass fraction (the gas mass per specific mass). Note that Relation (I.7) may be recast as :

$$\rho = \left[1 - \frac{z}{\varrho_g(p)}\right] \rho_\ell + z,$$

which shows that it indeed provides a closure relation to the system (I.1)-(I.3), *i.e.* an additional relation involving only variables initially present in (I.1)-(I.3).

We now recall some estimates which are satisfied, at least formally, by the solution of System (I.1)-(I.3). Equation (I.1) shows that ρ remains non-negative at all time. Replacing z by ρy in the gas mass balance equation (I.2) and using the mass balance equation (I.1), we get :

$$\partial_t(\rho y) + \nabla \cdot (\rho y\, \boldsymbol{u}) = \rho\, \big(\partial_t y + \boldsymbol{u} \cdot \nabla y\big) = 0.$$

Let us suppose that ρ does not vanish (which is not necessarily true at the continuous level, since $\mathrm{div}\boldsymbol{u}$ is not bounded in $\mathrm{L}^\infty(\Omega)$, but will be true at the discrete level). Then this relation implies that y satisfies the transport equation $\partial_t y + \boldsymbol{u} \cdot \nabla y = 0$, and thereby it follows a maximum principle. Specifically, if the initial and boundary conditions for ρ and z are such that $y \in [\varepsilon, 1]$ at $t = 0$, where $0 < \varepsilon \leq 1$ (which excludes purely liquid zones at the initial time), we obtain that y remains in the interval $[\varepsilon, 1]$ at all times. From the second form of (I.7) and the fact that $\rho > 0$, we can deduce that $\rho \in [\min(\rho_\ell, \rho_g), \max(\rho_\ell, \rho_g)]$ and, now from the first form of (I.7), $\alpha \in (0, 1]$, so $\rho_g > 0$ and, since ϱ_g is one-to-one from $(0, +\infty)$ to itself, there exists a positive pressure p such that $\rho_g = \varrho_g(p)$.

Let us now define the function $\mathcal{P}$, from $(0, +\infty)$ to $\mathbb{R}$, as a primitive of $s \mapsto \wp_g(s)/s^2$, where $\wp_g = \varrho_g^{-1}$. Then, if we suppose that the velocity is prescribed to zero at the boundary, the solution to System (I.1)-(I.3) satisfies :

$$\frac{d}{dt} \int_\Omega \big[\frac{1}{2}\, \rho\, |\boldsymbol{u}|^2 + z\mathcal{P}(\varrho_g(p))\big]\, \mathrm{d}\boldsymbol{x} \leq 0. \tag{I.8}$$

The quantity $z\mathcal{P}(\varrho_g(p))$ is often called the Helmholtz energy, $\frac{1}{2}\, \rho\, |\boldsymbol{u}|^2$ the kinetic energy and their sum is the total energy of the system. Since the function $\mathcal{P}$ is increasing, Inequality (I.8) provides an estimate on the solution.

When $\mu = 0$, the viscous term $\tau(u)$ vanishes and the system (I.1)-(I.3) becomes hyperbolic, with a simple wave structure (see [32] for a comprehensive presentation). The solution to the Riemann problems always involves a contact discontinuity and two additional waves, which are either shock or rarefaction waves. Through the contact discontinuity, the pressure and velocity are kept constant, and z, ρ or y are discontinuous. The existence of this wave may be inferred by just checking that, provided this is consistent with initial and boundary conditions, a solution to the system with constant velocity and pressure exists : indeed, from the first form of (I.7), it may be seen that ρ and z are linked by an affine relation with constant (for a constant pressure) coefficients ; (I.1) and (I.2) then boil down to the same transport equation (with a constant velocity) and (I.3) is trivially satisfied.

Finally, note that, since y satisfies a transport equation, if $y = 1$ at the initial time (everywhere in domain) and at inflow boundaries (at all time), the solution satisfies $y = 1$ for all $\boldsymbol{x} \in \Omega$ and $t \in (0, T)$. In such a case, System (I.1)-(I.3) boils down to the governing equations of barotropic monophasic flows ; for the particuliar equation of state $\varrho_g(s) = s^{1/2}$, we recover in one or two dimensions the usual shallow-water equations.

The use of pressure correction schemes for compressible single phase flow seems to be widespread, see *eg.* [36] for the seminal work and [75] for a comprehensive introduction. Indeed, pressure correction schemes are often partly implicit, thereby ensuring some stability with respect to the time step together with introducing a decoupling of the equations which helps the numerical solution of the nonlinear sytems. Extensions to multi-phase flows are scarcer and seem to be restricted to iterative algorithms, often similar in spirit to the usual SIMPLE algorithm for incompressible flows [66, 57, 47]. In this paper, we perform a numerical study of a non-iterative pressure-correction scheme introduced in [26], based on a low order finite element and a finite volume discretization, which enjoys the following properties :

(i) the scheme has at least one solution, and any solution satisfies the above listed "discrete-maximum-based" estimates : $\rho > 0$, the gas mass fraction y satisfies a discrete maximum principle (so $0 < y \leq 1$), and $p > 0$.

(ii) the scheme is unconditionally stable, in the sense that its solution(s) satisfies a discrete analogue of Inequality (I.8),

(iii) the pressure and velocity are kept constant through contact discontinuities.

In addition, the scheme is conservative for ρ and z. As in the continuous case, thanks to the fact that y is kept constant if it is consistent with the initial and boundary conditions, it also cope with monophasic barotropic flows, a particuliar case of which may be formally identified to the shallow water equations. Finally, the scheme boils down to the usual projection scheme for incompressible flows (obtained in the present framework when $y = 0$ or, asymptotically, when the function ϱ_g becomes constant), and is indeed routinely used for the computation of low Mach number flows, as, for instance, classical bubble columns of chemical engineering processes or diphasic flows encountered in nuclear safety studies. Its accuracy was assessed for smooth solutions in [26], and the aim of the present paper is to check its convergence and accuracy in non-diffusive cases, for weak solutions with discontinuities.

The paper is organized as follows. We first present the scheme (Section I.2). Then we study several Riemann problems, first monophasic ($y = 1$) (Section I.3.1) then biphasic : in this latter case, we first address a flow which involves only a contact discontinuity and shocks (Section I.3.2.a), and finally a flow with rarefaction waves (Section I.3.2.b). Finally, we assess in Section I.4 the behaviour of the scheme on two-dimensional test cases.

I.2 The pressure correction scheme

I.2.1 Time semi-discretization

Let us consider a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the time interval $(0, T)$, which is supposed uniform. Let $\delta t = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N-1$ be the constant time step. In a time semi-discrete setting, denoting by ρ^{-1} and u^0 initial guesses for the density and velocity, the algorithm proposed in this paper is the following.

0 - Initialization – Compute ρ^0 by solving the following semi-discrete mass balance equation :

$$\frac{\rho^0 - \rho^{-1}}{\delta t} + \operatorname{div}(\rho^0 \boldsymbol{u}^0) = 0. \tag{I.9}$$

Then, for $n \geq 0$:

1 - Prediction step – Solve the following semi-discrete linearized momentum balance equation for $\tilde{\boldsymbol{u}}^{n+1}$:

$$\frac{\rho^n \ \tilde{\boldsymbol{u}}^{n+1} - \rho^{n-1} \ \boldsymbol{u}^n}{\delta t} + \operatorname{div}(\rho^n \ \tilde{\boldsymbol{u}}^{n+1} \otimes \boldsymbol{u}^n) + \nabla p^n - \operatorname{div}(\boldsymbol{\tau}(\tilde{\boldsymbol{u}}^{n+1})) = 0. \tag{I.10}$$

2 - Pressure correction step – Solve (simultanuously) the following non linear equations for p^{n+1}, $\boldsymbol{u}^{n+1}$,

ρ^{n+1} and z^{n+1} :

$$\rho^n \, \frac{\boldsymbol{u}^{n+1} - \tilde{\boldsymbol{u}}^{n+1}}{\delta t} + \nabla(p^{n+1} - p^n) = 0, \tag{I.11a}$$

$$\frac{\rho^{n+1} - \rho^n}{\delta t} + \operatorname{div}(\rho^{n+1} \, \boldsymbol{u}^{n+1}) = 0, \tag{I.11b}$$

$$\frac{z^{n+1} - z^n}{\delta t} + \operatorname{div}(z^{n+1} \, \boldsymbol{u}^{n+1}) = 0, \tag{I.11c}$$

$$\rho^{n+1} = \varrho(p^{n+1}, z^{n+1}) = z^{n+1}(1 - \frac{\varrho_g(p^{n+1})}{\rho_\ell}) + \rho_\ell. \tag{I.11d}$$

Step 1 is the usual prediction step for the velocity, which consists in solving the momentum balance equation (I.3) with the beginning-of-step pressure. Step 2 is the pressure correction step. Its resolution is performed by combining equations (I.11a) and (I.11b), therefore obtaining a non-linear elliptic problem for the pressure, which reads in the time semi-discrete setting :

$$\begin{aligned} &\frac{\rho^{n+1} - \rho^n}{\delta t^2} - \operatorname{div}\big[\, \frac{\rho^{n+1}}{\rho^n} \, \nabla(p^{n+1} - p^n) \,\big] = -\frac{1}{\delta t} \, \operatorname{div}(\rho^{n+1} \tilde{u}^{n+1}), \\ &\qquad \text{with } \rho^{n+1} = \varrho(p^{n+1}, z^{n+1}) = z^{n+1}(1 - \frac{\varrho_g(p^{n+1})}{\rho_\ell}) + \rho_\ell. \end{aligned} \tag{I.12}$$

Note that, for a given space discretization, this equation must be established at the algebraic level, with the discrete equivalent manipulations which were necessary to derive it at the continuous level (*i.e.* multiplying the first equation by ρ^{n+1}/ρ^n, taking its divergence and substracting to the second relation) [26].

Two features are unusual in this algorithm. The first one is the time-shift of the densities in the prediction step : thanks to this time-shift, the densities satisfy (I.11b) of the preceding correction step and therefore the convection operator vanishes for constant velocities (*i.e.* $\tilde{u}^{n+1} = 1$), which ensures the conservation of the kinetic energy [20, 1]. Second, the pressure correction step couples the mixture and dispersed phase mass balance equations (I.11b) and (I.11c) ; this coupling preserves the affine relation between ρ^{n+1} and z^{n+1} through the equation of state (I.11d), with coefficients only depending on the pressure (taken at the same time level). Thus, as in the continuous case, both equations boil down to only one relation when the pressure is constant ; consequently, the arguments necessary to obtain solutions with constant velocity and pressure (*i.e.* contact discontinuity waves) still hold at the discrete level.

I.2.2 Discrete spaces and unknowns

The scheme has been developed (and actually works) with unstructured (in particular simplicial) discretizations, and for 2D and 3D cases. We shall restrict ourselves here to 1D Riemann problems, and to L-shaped two-dimensional domains. Hence, for the sake of conciseness, we only describe here the case of structured meshes, using either a finite volume MAC or a Rannacher-Turek (RT) finite element discretization which we now present.

Let Ω be a rectangular domain of $\mathbb{R}^d$, $d = 2$ or 3, and let $\mathcal{M}$ be a decomposition of the domain Ω into rectangles or rectangular parallelepipeds, supposed to be regular in the usual sense of the finite element literature (e.g. [9]). By $\mathcal{E}$ and $\mathcal{E}(K)$ we denote the set of all faces σ of the mesh and of the element $K \in \mathcal{M}$ respectively. The set of faces included in the boundary of Ω is denoted by $\mathcal{E}_{\rm ext}$ and the set of internal faces (*i.e.* $\mathcal{E} \setminus \mathcal{E}_{\rm ext}$) is denoted by $\mathcal{E}_{\rm int}$. For each internal face of the mesh $\sigma = K|L$, $\boldsymbol{n}_{KL}$ stands for the normal vector to σ, oriented from K to L. The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-dimensional measure of the face σ. For $1 \leq i \leq d$, we denote by $\mathcal{E}^{(i)} \subset \mathcal{E}$ the subset of the faces of $\mathcal{E}$ which are perpendicular to the i^{th} unit vector of the canonical basis of $\mathbb{R}^d$.

For both MAC and RT discretizations, the degrees of freedom for the pressure, the density and the variables y and z are associated to the cells of the mesh $\mathcal{M}$: the degrees of freedom are therefore

$$\big\{p_K,\ \rho_K,\ y_K,\ z_K,\ K \in \mathcal{M}\big\}.$$

Let us then turn to the degrees of freedom for the velocity.

- **RT discretization.** The velocity is discretized using the so-called Rannacher–Turek (RT) element [65]. The approximation for the velocity is thus non-conforming (the discrete functions are discontinuous through a face, but the jump of their integral is imposed to be zero) ; the degrees of freedom for the velocities are located at the center of the faces of the mesh, and we choose the version of the element where they represent the average of the velocity through a face. The set of degrees of freedom reads :

  $$\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E},\ 1 \leq i \leq d\}.$$

  We denote by $\boldsymbol{\varphi}_\sigma^{(i)}$ the vector shape function associated to $\boldsymbol{u}_{\sigma,i}$, which, by definition, reads $\boldsymbol{\varphi}_\sigma^{(i)} = \varphi_\sigma\, \boldsymbol{e}^{(i)}$, where φ_σ is the RT scalar shape function and $\boldsymbol{e}^{(i)}$ is the $i^{\rm th}$ vector of the canonical basis of $\mathbb{R}^d$, and we define $\boldsymbol{u}_\sigma$ by $\boldsymbol{u}_\sigma = \sum_{i=1}^d \boldsymbol{u}_{\sigma,i}\, \boldsymbol{e}^{(i)}$. With these definitions, we have the identity :

  $$\boldsymbol{u}(\boldsymbol{x}) = \sum_{\sigma \in \mathcal{E}} \sum_{i=1}^{d} \boldsymbol{u}_{\sigma,i}\ \boldsymbol{\varphi}_\sigma^{(i)}(\boldsymbol{x}) = \sum_{\sigma \in \mathcal{E}} \boldsymbol{u}_\sigma\ \varphi_\sigma(\boldsymbol{x}), \qquad \text{for a.e. } \boldsymbol{x} \in \Omega.$$

- **MAC discretization.** The degrees of freedom for the i^{th} component of the velocity, defined at the centres of the face $\sigma \in \mathcal{E}^{(i)}$, are denoted by :

  $$\big\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)},\ 1 \leq i \leq d\big\}.$$

Let us now turn to the treatment of Dirichlet boundary conditions. Let $\mathcal{E}_D \subset \mathcal{E}_{\rm ext}$ be the set of faces located on the Dirichlet boundary, and let $\boldsymbol{u}_D$ be the prescribed value of the velocity on these faces. For the RT discretization, as usual in the finite element framework, these Dirichlet boundary conditions are built-in in the definition of the discrete space :

$$\forall \sigma \in \mathcal{E}_D, \text{ for } 1 \leq i \leq d, \qquad \boldsymbol{u}_{\sigma,i} = \frac{1}{|\sigma|} \int_\sigma \boldsymbol{u}_{D,i},$$

where $\boldsymbol{u}_{D,i}$ stands for the $i^{\rm th}$ component of $\boldsymbol{u}_D$. For the MAC scheme, the normal components of the velocity at the Dirichlet boundary are also prescribed :

$$\text{for } 1 \leq i \leq d,\ \forall \sigma \in \mathcal{E}_D \cap \mathcal{E}^{(i)}, \qquad \boldsymbol{u}_{\sigma,i} = \frac{1}{|\sigma|} \int_\sigma \boldsymbol{u}_{D,i},$$

while Dirichlet conditions for tangential components will be used, as usual for finite volumes, in the definition of the diffusion term.

I.2.3 Discrete equations

We now describe the space discretization of each equation of the time semi-discrete algorithm. We choose to present the equations of the projection step in their original form, *i.e.* before the derivation of the elliptic problem for the pressure, which is thoroughly described in [26]. Indeed, this latter step is purely algebraic, in the sense that it transforms a nonlinear algebraic system into another nonlinear algebraic system which is strictly equivalent, and thus has no impact on the properties of the scheme (besides, of course, the efficiency issue).

We begin with the discretization of the mass balance equations (I.11b) and (I.11c) of the projection step. For both the MAC and RT discretizations, let us denote by $\boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}$ the outward normal velocity to the face σ of K, which is computed, for the RT element, by taking the inner product of the velocity at the face with the outward normal vector (so exactly as said by the notation) and which is given, for the MAC scheme, by the value of the component of the velocity at the center of the face (up to a change of sign). Discrete equations are obtained by an upwind finite volume discretization and read :

$$\forall K \in \mathcal{M}, \qquad \begin{aligned} &\frac{|K|}{\delta t}\,(\rho_K^{n+1}-\rho_K^n) + \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\;\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}\;\rho_\sigma^{n+1} = 0,\\ &\frac{|K|}{\delta t}\,(z_K^{n+1}-z_K^n) + \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\;\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}\;z_\sigma^{n+1} = 0, \end{aligned} \tag{I.13}$$

where ρ_σ^{n+1} (resp. z_σ^{n+1}) is the upwind approximation of ρ^{n+1} (resp. z^{n+1}) at the face σ, the definition of which we now recall for the sake of completeness. For an internal face $\sigma = K|L$, ρ_σ^{n+1} (resp. z_σ^{n+1}) stands for ρ_K^{n+1} (resp. z_K^{n+1}) if $\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma} \geq 0$ and for ρ_L^{n+1} (resp. z_L^{n+1}) otherwise ; for an external face $\sigma \in \mathcal{E}(K)$, ρ_σ^{n+1} (resp. z_σ^{n+1}) is equal to ρ_K^{n+1} (resp. z_K^{n+1}) if the flow is directed outward Ω (*i.e.* $\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma} \geq 0$) or given by the boundary conditions otherwise. This approximation ensures that $\rho^{n+1} > 0$ if $\rho^n > 0$ and if the density is prescribed to a positive value at inflow boundaries. In addition, if we set $y_K^{n+1} = z_K^{n+1}/\rho_K^{n+1}$ and $y_K^n = z_K^n/\rho_K^n$, we may recast the second equation of (I.13) as :

$$\frac{|K|}{\delta t}\,(\rho_K^{n+1}y_K^{n+1}-\rho_K^n y_K^n) + \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\;\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}\;\rho_\sigma^{n+1}y_\sigma^{n+1} = 0, \tag{I.14}$$

where we recognize in y_σ^{n+1} the upwind approximation of y^{n+1} at the face σ. This relation thus yields that y^{n+1} satisfies a discrete maximum principle by standard arguments [50].

In the case of the MAC discretization, the velocity prediction equation is approximated by a finite volume technique over a dual mesh. For the RT discretization, the time derivative and convection terms are approximated by a similar finite volume technique, while the finite element formulation is used for the other terms. For each component of the velocity, a dual mesh of the computational domain Ω thus has to be defined :

- **RT discretization.** For the RT discretization, the dual mesh is the same for all the velocity components, and dual cells are chosen as follows. For any $K \in \mathcal{M}$ and any face $\sigma \in \mathcal{E}(K)$, let $D_{K,\sigma}$ be the cone with basis σ and with vertex the mass center of K. The volume $D_{K,\sigma}$ is referred to as the half-diamond cell associated to K and σ. For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the diamond cell D_σ associated to σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external face $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}(K)$, D_σ is just the same volume as $D_{K,\sigma}$.

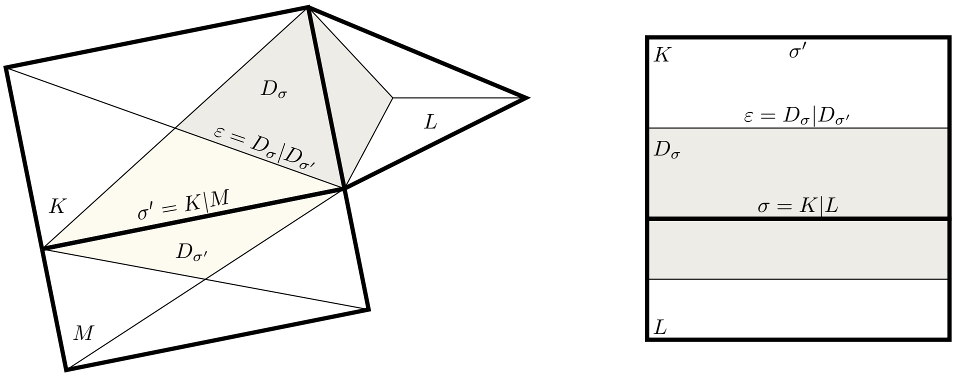

FIG. I.1 – Notations for control volumes and dual cells – Left : Finite Elements (the present sketch illustrates the possibility, implemented in the ISIS software, of mixing simplicial (Crouzeix-Raviart) and quadrangular cells, even if only rectangular cells are used in this paper) – Right : MAC discretization, dual cell for the y-compnenent of the velocity.

- **MAC discretization** For the MAC scheme, the dual mesh depends on the component of the velocity. For each of them, its definition differs from the RT one only by the choice of the half-diamond cell, which, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, is now the rectangle of basis σ and of measure $|D_{K,\sigma}|$ equal to half the measure of K.

We denote by $\varepsilon = D_\sigma|D_{\sigma'}$ the face separating two diamond cells D_σ and $D_{\sigma'}$ (see Figure I.1).

In both cases, for $1 \leq i \leq d$ and $\sigma \in \mathcal{E}^{(i)}$, we denote by $(\mathrm{div}\tau(\tilde{\boldsymbol{u}}^{n+1}))_{\sigma,i}$ an approximation of the i-th component of the viscous term associated to σ, and we denote by $(\boldsymbol{\nabla} p^n)_{\sigma,i}$ the i-th component of the discrete pressure gradient at the face σ. With these notations, we are able to write the following general form of the approximation to the momentum balance equation :

$$\begin{aligned}\frac{|D_\sigma|}{\delta t}\,(\bar\rho_\sigma^n\,\tilde{\boldsymbol{u}}_{\sigma,i}^{n+1} - \bar\rho_\sigma^{n-1}\,\boldsymbol{u}_{\sigma,i}^n) + \sum_{\varepsilon\in\mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^n\,\tilde{\boldsymbol{u}}_{\varepsilon,i}^{n+1}& \\ +|D_\sigma|(\boldsymbol{\nabla} p^n)_{\sigma,i} - |D_\sigma|(\mathrm{div}\tau(\tilde{\boldsymbol{u}}^{n+1}))_{\sigma,i} &= (\boldsymbol{f}^{n+1})_{\sigma,i},\end{aligned} \tag{I.15}$$

this equation being written for $1 \leq i \leq d$, $\sigma \in \mathcal{E} \setminus \mathcal{E}_D$ in the case of the RT discretization, and for $1 \leq i \leq d$, $\sigma \in \mathcal{E}^{(i)} \setminus \mathcal{E}_D$ for the MAC scheme. In this relation, $\bar\rho_\sigma^n$ and $\bar\rho_\sigma^{n-1}$ stand for an approximation of the density on the face σ at time t^n and t^{n-1} respectively (which must not be confused with the upstream density ρ_σ^n used in the mass balance), $F_{\sigma,\varepsilon}^n$ is the discrete mass flux through the dual face ε outward D_σ, and $\tilde{\boldsymbol{u}}_{\varepsilon,i}^{n+1}$ stands for an approximation of $\tilde{\boldsymbol{u}}_i^{n+1}$ on ε which may be chosen either centred or upwind. In the centered case, for an interior face $\varepsilon = D_\sigma|D_{\sigma'}$, we thus get $\tilde{\boldsymbol{u}}_{\varepsilon,i}^{n+1} = (\tilde{\boldsymbol{u}}_{\sigma,i}^{n+1} + \tilde{\boldsymbol{u}}_{\sigma',i}^{n+1})/2$ while, in the upwind case, we have $\tilde{\boldsymbol{u}}_{\varepsilon,i}^{n+1} = \tilde{\boldsymbol{u}}_{\sigma,i}^{n+1}$ if $F_{\sigma,\varepsilon}^n \geq 0$ and $\tilde{\boldsymbol{u}}_{\varepsilon,i}^{n+1} = \tilde{\boldsymbol{u}}_{\sigma',i}^{n+1}$ otherwise.

The quantity $(\boldsymbol{f}^{n+1})_{\sigma,i}$ is a forcing term, which, for our purpose here, does not vanish only on external faces where Neumann conditions are prescribed ; if this latter read $\boldsymbol{\tau} \cdot \boldsymbol{n} - p\boldsymbol{n} = \boldsymbol{f}$, we get :

$$(\boldsymbol{f}^{n+1})_{\sigma,i} = \frac{1}{\delta t} \int_{n\,\delta t}^{(n+1)\,\delta t} \int_\sigma \boldsymbol{f} \cdot \boldsymbol{e}^{(i)}\,\mathrm{d}\gamma.$$

The finite element discretization of the i-th component of the pressure gradient term reads :

$$|D_\sigma| (\boldsymbol{\nabla} p^n)_{\sigma,i} = - \sum_{M \in \mathcal{M}} \int_M p^n \ \mathrm{div} \boldsymbol{\varphi}_\sigma^{(i)} \ \mathrm{d}\boldsymbol{x}.$$

Since the pressure is piecewise constant, using the definition of the RT shape functions, an easy computation yields for an internal face $\sigma = K|L$:

$$|D_\sigma| (\boldsymbol{\nabla} p^n)_{\sigma,i} = |\sigma| \ (p_L^n - p_K^n) \ \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}^{(i)},$$

and, for an extermal face $\sigma \in \mathcal{E}(K) \cap \mathcal{E}_{\mathrm{ext}} \setminus \mathcal{E}_D$:

$$|D_\sigma| (\boldsymbol{\nabla} p^n)_{\sigma,i} = -|\sigma| \ p_K^n \ \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}^{(i)}.$$

These expressions coincide which the discrete gradient in the MAC discretization.

The finite element discretization of the viscous term $(\mathrm{div} \tau(\tilde{\boldsymbol{u}}^{n+1}))_{\sigma,i}$, associated to σ and to the component i, reads :

$$|D_\sigma| (\mathrm{div} \tau(\tilde{\boldsymbol{u}}^{n+1}))_{\sigma,i} = -\mu \sum_{K \in \mathcal{M}} \int_K \nabla \tilde{\boldsymbol{u}}^{n+1} \cdot \nabla \boldsymbol{\varphi}_\sigma^{(i)} - \frac{\mu}{3} \sum_{K \in \mathcal{M}} \int_K \mathrm{div} \, \tilde{\boldsymbol{u}}^{n+1} \ \mathrm{div} \, \boldsymbol{\varphi}_\sigma^{(i)}.$$

The MAC discretization of this same viscous term is detailed in [2].

The main motivation to implement a finite volume approximation for the first two terms is to obtain a discrete equivalent of the kinetic energy theorem, which holds in the case of homogeneous Dirichlet boundary conditions and reads :

$$\begin{aligned} \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \Big[\frac{|D_\sigma|}{\delta t} \ (\bar\rho_\sigma^n \ \tilde{\boldsymbol{u}}_\sigma^{n+1} - \bar\rho_\sigma^{n-1} \ \boldsymbol{u}_\sigma^n) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^n \ \tilde{\boldsymbol{u}}_\varepsilon^{n+1} \Big] \cdot \ \boldsymbol{u}_\sigma \geq \\ \frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\mathrm{int}}} \frac{|D_\sigma|}{\delta t} \ \big[\bar\varrho_\sigma^n \ |\tilde{\boldsymbol{u}}_\sigma^{n+1}|^2 - \bar\varrho_\sigma^{n-1} \ |\boldsymbol{u}_\sigma^n|^2 \big]. \end{aligned} \tag{I.16}$$

For this result to be valid, the necessary condition is that the convection operator vanishes for a constant velocity, *i.e.* that the following discrete mass balance over the diamond cells is satisfied [1, 20] :

$$\forall \sigma \in \mathcal{E}_{\mathrm{int}}, \qquad \frac{|D_\sigma|}{\delta t} \ (\bar\rho_\sigma^n - \bar\rho_\sigma^{n-1}) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^n = 0.$$

This governs the choice for the definition of the density approximation $\bar\rho_\sigma$ and the mass fluxes $F_{\sigma,\varepsilon}$. The density $\bar\rho_\sigma$ is defined by a weighted average : $\forall \sigma \in \mathcal{E}_{\mathrm{int}}$, $\sigma = K|L$, $|D_\sigma| \ \bar\rho_\sigma = |D_{K,\sigma}| \ \rho_K + |D_{L,\sigma}| \ \rho_L$ and $\forall \sigma \in \mathcal{E}_{\mathrm{ext}} \setminus \mathcal{E}_D$, $\sigma \in \mathcal{E}(K)$, $\bar\rho_\sigma = \rho_K$. The flux $F_{\sigma,\varepsilon}$ through the dual face ε of the half diamond cell $D_{K,\sigma}$ is computed as the flux through ε of a constant divergence lifting of the mass fluxes through the faces of the primal cell K, *i.e.* the quantities $(|\sigma| \boldsymbol{u}_\sigma \cdot \boldsymbol{n}_\sigma \, \rho_\sigma)_{\sigma \in \mathcal{E}(K)}$ appearing in the discrete mass balance (I.13). For a detailed construction of this approximation, we refer to [1, 38].

Equation (I.11a) is discretized similarly to the momentum balance (I.15), *i.e.* a finite volume technique is used for the unsteady term in the RT discretization. Hence, for both schemes, the discretization of (I.11a) reads :

$$\frac{|D_\sigma|}{\delta t} \, \rho_\sigma^n \, (\boldsymbol{u}_{\sigma,i}^{n+1} - \tilde{\boldsymbol{u}}_{\sigma,i}^{n+1}) + |D_\sigma| \, \Big[(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} - (\boldsymbol{\nabla} p^n)_{\sigma,i} \Big] = 0,$$

this equation being written for $1 \leq i \leq d$, $\sigma \in \mathcal{E} \setminus \mathcal{E}_D$ in the case of the RT discretization, and for $1 \leq i \leq d$, $\sigma \in \mathcal{E}^{(i)} \setminus \mathcal{E}_D$ for the MAC scheme.

I.3 Numerical experiments : Riemann problems

In this section, we assess the behaviour of the scheme for several 1D Riemann problems (often called also "shock tube problems") for the hyperbolic system (I.1)-(I.3) with $\mu = 0$ in (I.4), for which an analytical expression of the solution is known. We take benefit of the fact that the pressure correction scheme is able to keep $y = 1$ at any time, if the initial and boundary conditions allow, to first begin with a single phase flow, namely the solution of the so-called "Sod shock tube" problem. We then turn to two-phase flows, namely "two-fluid shock tube" model problems.

The computations presented here are performed with the ISIS code [40], built from the software component library PELICANS [63], both under development at IRSN and available as open-source softwares. The ISIS computer code is devoted to the solution of 2D or 3D problems (as the scheme presented in previous sections), so we are lead to define a fake 2D problem, designed to boil down to the addressed 1D Riemann problem. The domain Ω is rectangular, and the mesh is composed of only one horizontal stripe of cells (see Figures I.2– I.4).

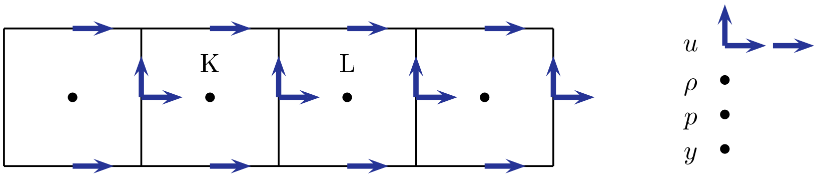

Fig. I.2 – Primal mesh and location of the unknowns for the Rannacher-Turek element, for a mesh consisting of only one stripe of cells, with homogeneous Neumann conditions at the bottom and left boundary, and a Dirichlet boundary condition at the left side of the computational domain.

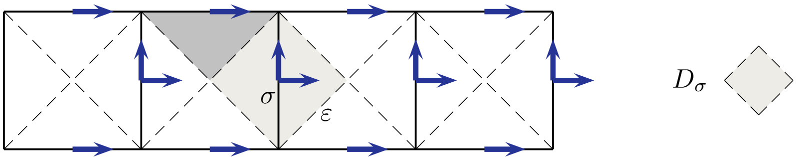

Fig. I.3 – Dual finite volume mesh for the Rannacher-Turek element, for a mesh consisting of only one stripe of cells.

In order to define a one-dimensional problem on this two dimensional domain, we impose a symmetry condition to the velocity at the top and bottom of the domain Ω (*i.e.* , with $\boldsymbol{u} = (u_1, u_2)$, we set $u_2 = 0$ and $\partial_{x_2} u_1 = 0$), which is satisfied by an horizontal flow invariant with respect to the second coordinate. An easy computation shows that, with the chosen mesh and boundary conditions, we obtain a discrete problem which exactly coincides with a 1D discretization for the MAC scheme ; this is clearly not the case for the RT element, since degrees of freedom for the horizontal velocity subsist at three different vertical locations.

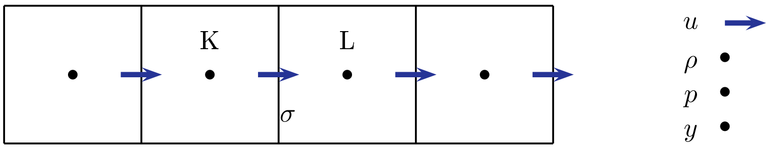

FIG. I.4 – Primal mesh and location of the unknowns for the MAC discretization, for a mesh consisting of only one stripe of cells, with homogeneous Neumann conditions at the bottom and left boundary, and a Dirichlet boundary condition at the left side of the computational domain.

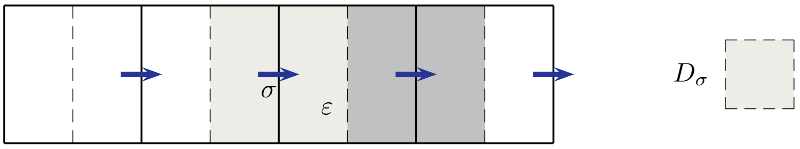

FIG. I.5 – Dual finite volume mesh for the MAC discretization, for a mesh consisting in only one stripe of meshes.

All the solutions computed in the following are such that the flow enters the domain on the left and leaves it on the right. So, at the left side of the domain, we impose $\boldsymbol{u} = (u_L, 0)$, $\rho = \rho_L$ and $z = z_L$; at the right hand side of the domain, we prescribe a Neumann boundary condition, with a surface forcing term equal to $-p_R\, \boldsymbol{n}$, where $\boldsymbol{n}$ is the unit outward normal vector to the boundary $\partial\Omega$.

As described above, for the velocity convection term in the momentum balance equation, the approximation of the velocity at the faces of the dual mesh (see Figures I.3 and I.5) may be chosen centred or upwind ; we will refer to the first option in the following as the centred variant (although upwinding is always used in the discrete mass balance equations), and to the second one as the upwind variant.

I.3.1 Sod shock tube problem

In order to simulate the Sod shock tube test, the gas mass fraction is set to $y \equiv 1$ (one-phase problem) ; we consider here the non-viscous homogeneous model resulting from Equations (I.1)–(I.3), with $\mu = 0$, with an equation of state where p is proportional to ρ (in fact, we take $p = \rho$), which corresponds to the isothermal Euler equations. The (1D) continuous problem is posed over the interval $(-2, 3)$ and, for the computation, we take $\Omega = (-2, 3) \times (0, 0.01)$. The two initial constant states are given by :

$$\begin{pmatrix} \rho \\ u \end{pmatrix}_L = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} \rho \\ u \end{pmatrix}_R = \begin{pmatrix} 0.125 \\ 0 \end{pmatrix}.$$

With this initial condition, the solution consists in a rarefaction wave travelling to the left and a shock travelling to the right.

We first address the results obtained with the RT discretization. The first outcome is that the scheme converges to the exact solution as soon as some diffusion is introduced in the momentum balance equation, either by adding a small artificial viscosity term to the centred approximation or by using the upwind scheme ; otherwise, *i.e.* with $\mu = 0$ and the centred variant, the usual (for a centred discretization of the advection operator) odd-even oscillations affect the computed velocity, and convergence is lost. More precisely, due to the particular structure of the mesh (see Figures I.2 and I.3), we observe in this latter case that the solution seems to result from the superposition of two different regular functions, one being associated to the degrees of freedom located at the intermediate elevation and the second one being associated to degrees of freedom located on the top and bottom boundaries ; surprinsingly, these two functions do not change when refining the mesh and time step with a constant CFL number. As an exemple of the numerical results obtained in convergent cases, the solution at $t = 1$ obtained with a mesh consisting of 2000 cells, $\delta t = 0.00125$ and a residual viscosity of $\mu = 0.001$ is presented in Figure I.6, together with the exact solution. Using $v = 1.6$ (which corresponds approximately to the velocity of the faster wave, namely the shock) as velocity range, these numerical parameters correspond to a CFL $= v\ \delta t/h = 0.8$.

Since combining a centred discretization of the momentum balance equation with the addition of an artificial viscosity may seem to be an appealing technique to avoid an excessive numerical dissipation (for instance, associated to an adjustment of μ as a function of the regularity of the solution, in the spirit of [30, 31]), we now investigate the influence of the value of μ on the accuracy of the centred scheme. We observe in Figure I.7 and Figure I.8 that taking a large viscosity yields inaccurate results, which is easily explained by the fact that the solved problem is too far from the original one. On the other hand, for too low values of the viscosity, oscillations appear, and the numerical error increases. In between, the error remains small, and one can remark that the optimal value for μ with respect to the L^1 norm of the error decreases with the time and space steps, as would be the numerical dissipation introduced by the upwinding technique. Comparing Figures I.7 and I.8, we note that the plateau is wider for CFL=9.6 than for CFL=0.8, but the overall shape of the curves remains essentially similar for both CFL numbers.

We end this study of the RT discretization by reporting the accuracy of the schemes as a function of the time and space step, with two constant CFL numbers. We study the centred scheme with $\mu = 0.001$ and the upwind scheme with $\mu = 0$ (we shall always set $\mu = 0$ for the upwind scheme hereafter). For the centred scheme, the observed orders of convergence (Figure I.9) are about 0.5 at CFL=0.8 and 1 at CFL=9.6 respectively, for both the velocity and the pressure. For the upwind variant, the order of convergence is 0.75 for any CFL.

With the MAC discretization, the behaviour is quite different, since the scheme seems to be convergent in its centred as well as in its upwind version, without needing the addition of any artificial viscosity. The solution at $t = 1$ obtained with the same parameters as for the RT discretization (*i.e.* 2000 cells, $\delta t = 0.00125$, so CFL $= v\ \delta t/h = 0.8$, and a residual viscosity of $\mu = 0.001$) is presented in Figure I.10. The influence of the addition of an artificial viscosity to the centred variant is shown in Figures I.11 and I.12. Finally, we once again assess the accuracy of the schemes as a function of the time and space step, with two constant CFL numbers (Figure I.13), all the computations being now performed with $\mu = 0$. Results seem to indicate that the order of convergence does not depend on the CFL number, neither on the upwind or centred choice : in all cases, the order of the convergence is close to 0.8, for the velocity and the pressure.

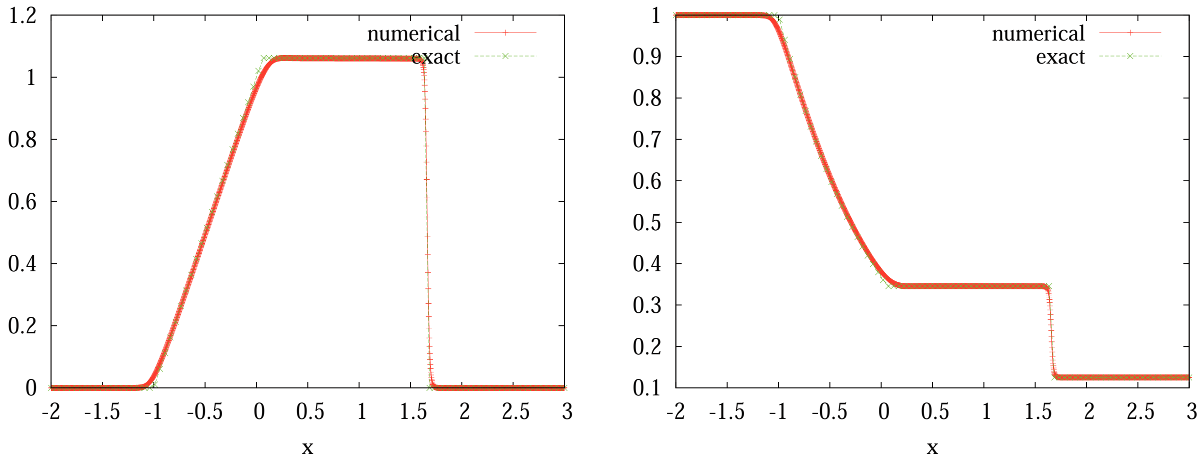

FIG. I.6 – Sod shock tube problem – Centred RT scheme – Numerical solution of the perturbed viscous problem at $t = 1$ with $\mu = 0.001$, 2000 cells, $\delta t = 0.00125$ (*i.e.* CFL=0.8). Velocity (left) and pressure (right).

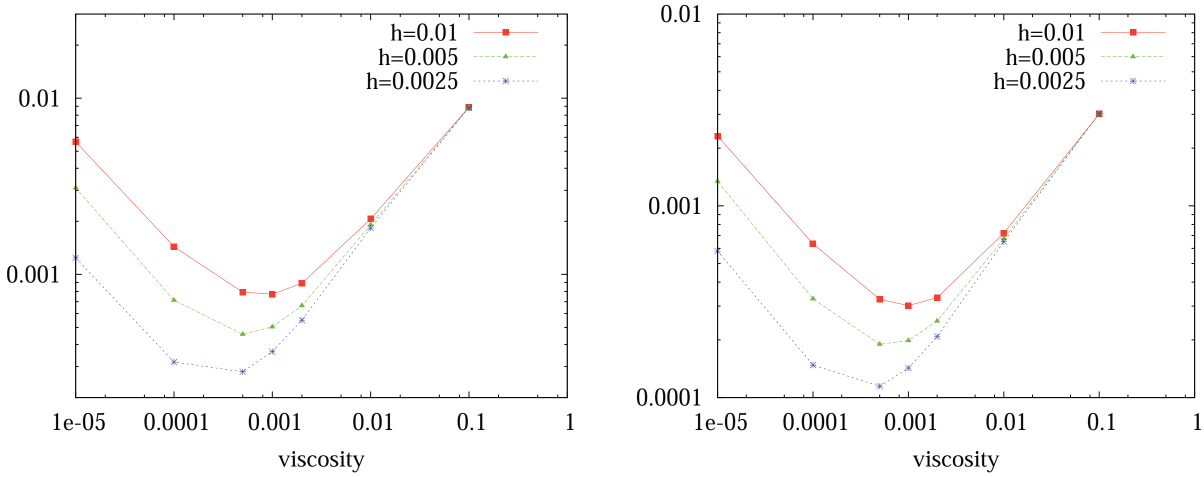

FIG. I.7 – Sod shock tube problem – Centred RT scheme – L^1 norm of the error between numerical solution of the perturbed viscous problem and exact solution of the inviscid problem at $t = 1$, for three meshes, as a function of the viscosity μ, with CFL $= 0.8$. Velocity (left) and pressure (right).

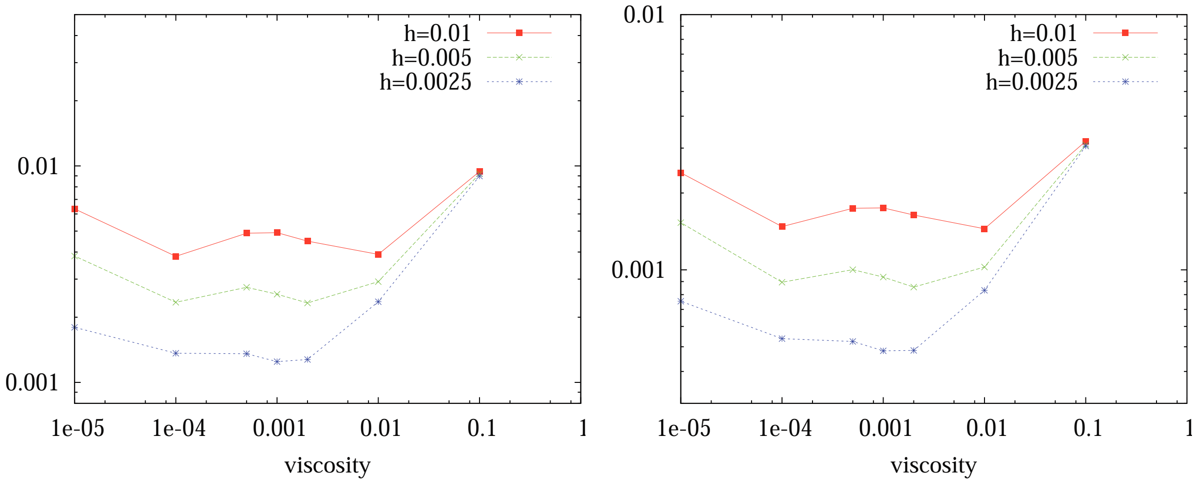

FIG. I.8 – Sod shock tube problem – Centred RT scheme – L^1 norm of the error between numerical solution of the perturbed viscous problem and exact solution of the inviscid problem at $t = 1$, for three meshes, as a function of the viscosity μ, with CFL$= 9.6$. Velocity (left) and pressure (right).

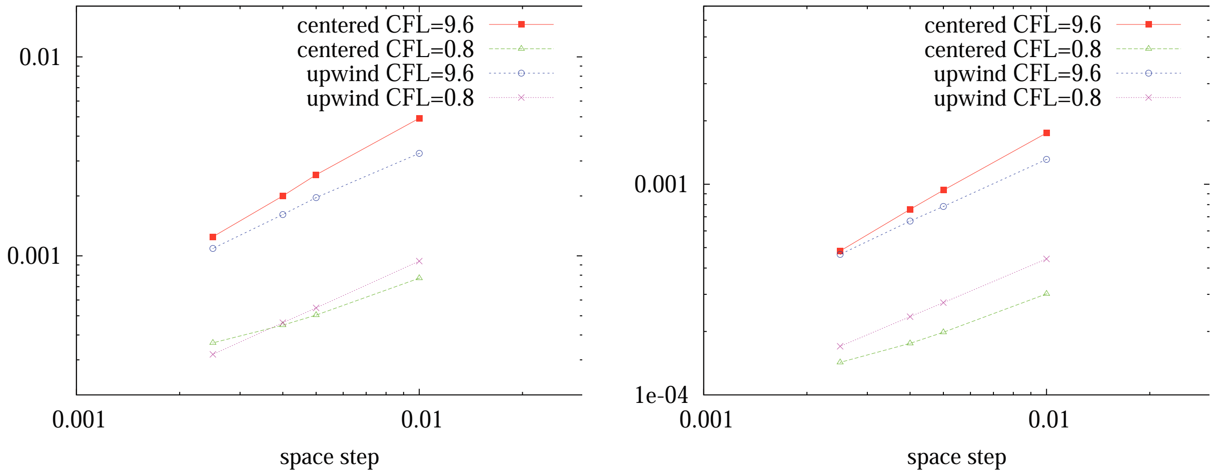

FIG. I.9 – Sod shock tube problem – Centred and upwind RT schemes – L^1 norm of the error between numerical solution of the perturbed viscous problem and exact solution of the inviscid problem at $t = 1$, as a function of the mesh (or time) step, for two fixed CFL numbers. In the centred case, the used artificial viscosity is $\mu = 0.001$, *i.e.* a value close to the one which yields the more accurate results. Velocity (left) and pressure (right).

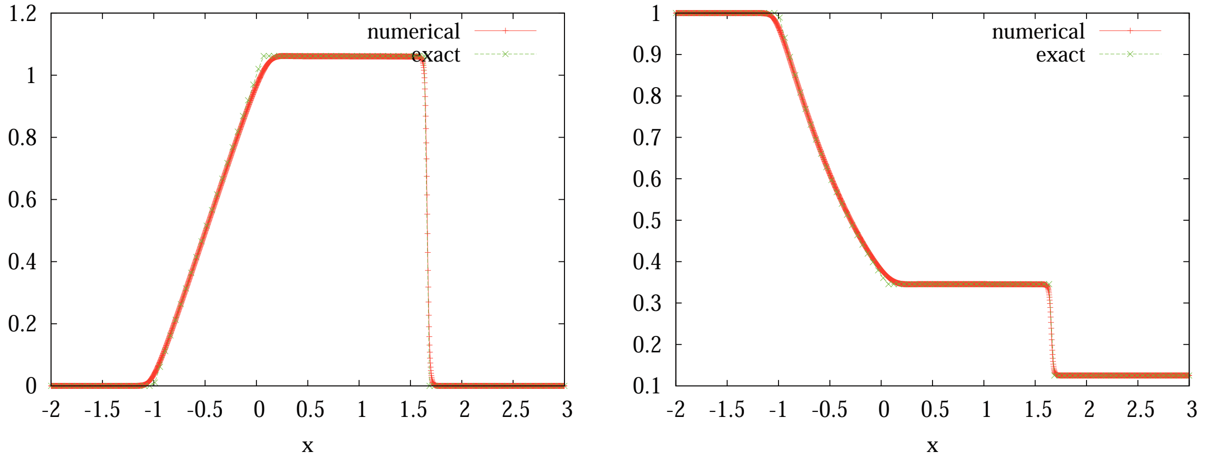

Fig. I.10 – Sod shock tube problem – Centred MAC scheme – Numerical solution of the perturbed viscous problem at $t = 1$ with $\mu = 0.001$, 2000 cells, $\delta t = 0.00125$ (*i.e.* CFL=0.8). Velocity (left) and pressure (right).

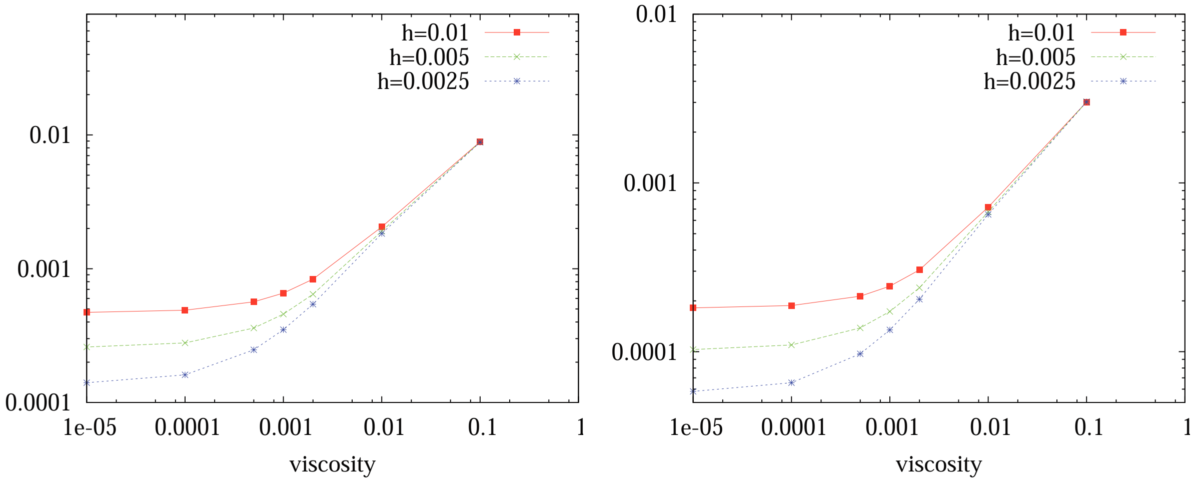

Fig. I.11 – Sod shock tube problem – Centred MAC scheme – L^1 norm of the error between numerical solution of the perturbed viscous problem and exact solution of the inviscid problem at $t = 1$, for three meshes, as a function of the viscosity μ, with CFL $= 0.8$. Velocity (left) and pressure (right).

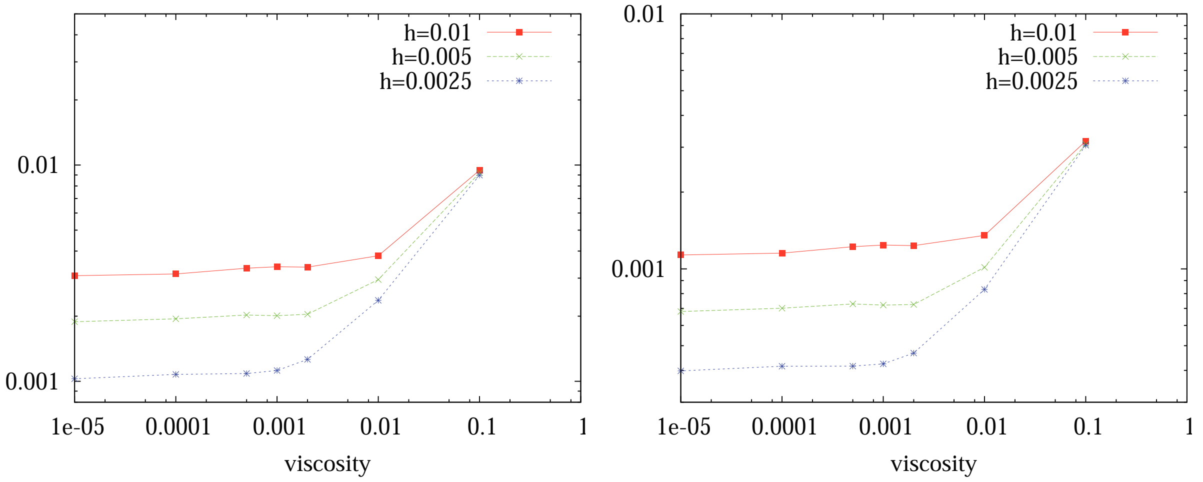

Fig. I.12 – Sod shock tube problem – Centred MAC scheme – L^1 norm of the error between numerical solution of the perturbed viscous problem and exact solution of the inviscid problem at $t = 1$, for three meshes, as a function of the viscosity μ, with CFL $= 9.6$. Velocity (left) and pressure (right).

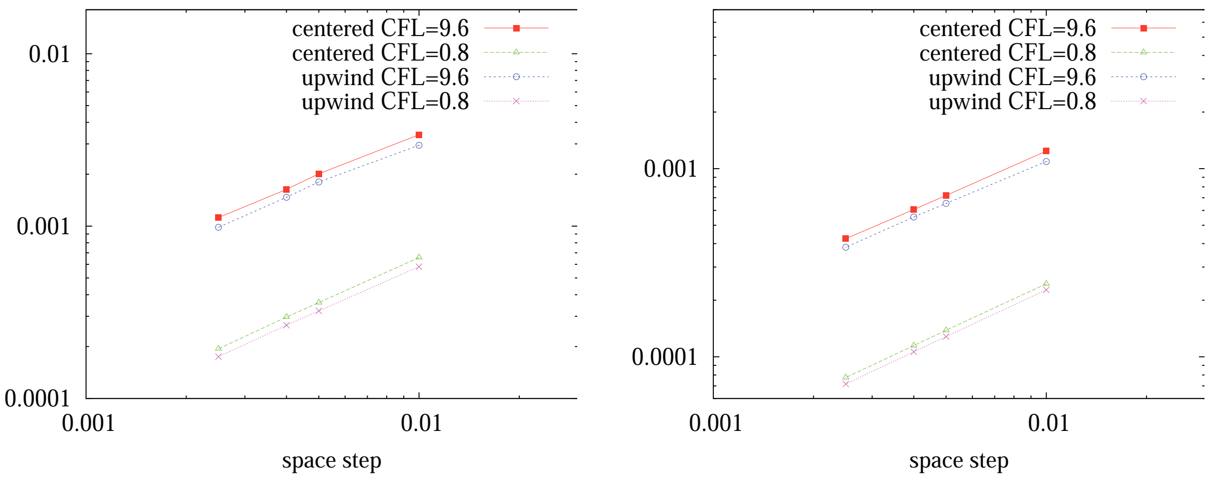

Fig. I.13 – Sod shock tube problem – Centred and upwind MAC scheme – L^1 norm of the error between the numerical solution and the exact solution at $t = 1$, as a function of the mesh (or time) step, for two fixed CFL numbers. Velocity (left) and pressure (right).

I.3.2 Two-fluid shock tube

We present here the numerical results for the two-fluid shock tube. The continuous problem is posed over $(-3,2)$ and we use a computational rectangular domain $\Omega = (-3,2) \times (0, 0.01)$. The equation of state is given by (I.7), with the following phasic equation of state for the gas phase :

$$p = 10\, \rho_g.$$

The constant liquid density is set to $\rho_\ell = 0.8$. We perform two tests, where the initial left and right constant states are chosen in order to yield two different flow structures : a contact discontinuity (in both cases), propagating between two shock waves in the first case, and two rarefaction waves in the second one.

I.3.2.a First case : shock – contact discontinuity – shock

The two initial constant states are given by :

$$\begin{pmatrix} \rho \\ \boldsymbol{u} \\ y \end{pmatrix}_L = \begin{pmatrix} 1 \\ 5 \\ 0.3 \end{pmatrix}, \qquad \begin{pmatrix} \rho \\ \boldsymbol{u} \\ y \end{pmatrix}_R = \begin{pmatrix} 2 \\ 1 \\ 0.8 \end{pmatrix}.$$

With this initial data, the exact solution consists in two shocks, the first one travelling to the left and the second one to the right, separated by a contact discontinuity slowly moving to the right.

The convergence behaviour of the schemes is quite similar to that of the one-phase case, namely convergence of the upwind scheme or of the centered scheme with a residual viscosity in both cases and non-convergence of the centered scheme with $\mu = 0$ and the RT element. A numerical solution given by the centred scheme at $t = 0.1$ with 5000 cells, $\delta t = 4.\, 10^{-5}$ and $\mu = 0.002$ is plotted in Figure I.14 and Figure I.15 for the RT and MAC discretization respectively, together with the exact solution. Taking $v = 18.16$ (the velocity of the fastest wave, namely the right shock), the CFL number for these numerical parameters is CFL$=v\, \delta t/h = 0.75$.

We then plot the solution obtained at $t = 0.1$ for various CFL numbers, with the centred schemes, 2500 cells and $\mu = 0.002$. We observe in Figure I.16, that with the RT discretization, the solution is qualitatively correct up to a CFL of the order of 20, and then strongly deteriorates, showing in particular wild velocity and pressure oscillations at the contact discontinuity. On the contrary, we observe more reasonable profiles for the MAC discretization in Figure I.17 : the oscillations only affect the pressure in the vicinity of the contact discontinuity for large CFL numbers (≥ 80). Note that, in any case, the structure of the solution seems to remain correct, *i.e.* we do not observe the apparition of spurious waves (for instance, non-entropic shocks), as often happens with a scheme without any numerical diffusion. .

We then assess the accuracy of the scheme as a function of the time and space step, with two constant CFL numbers, for the centred variants with $\mu = 0.002$ and for the upwind variant. The observed orders of convergence for the centred RT scheme (Figure I.18) are about 1.5 and 1. at CFL=0.75 and 9 respectively, for both velocity and pressure ; for ρ and y, the order of convergence is 0.7 and 0.5 respectively, for both CFL numbers. For the upwind RT scheme, the order of convergence is 1 for both the velocity and the pressure and 0.5 for ρ and y, at any CFL number. Again, for the MAC scheme (Figure I.19), the order of convergence does not seem to depend on the CFL number nor on the upwind or centred choice : in all cases, the order of convergence is 1 for both the velocity and the pressure and 0.5 for the density and the gas mass fraction. These behaviours are consistent with what is usually observed : the convergence order is about 1/2 for the variables which jump at the contact discontinuity, while, for the other variables which vary only at shocks, the compressive effect of the shocks counterbalances the diffusion of the scheme and yields a convergence order close to 1.

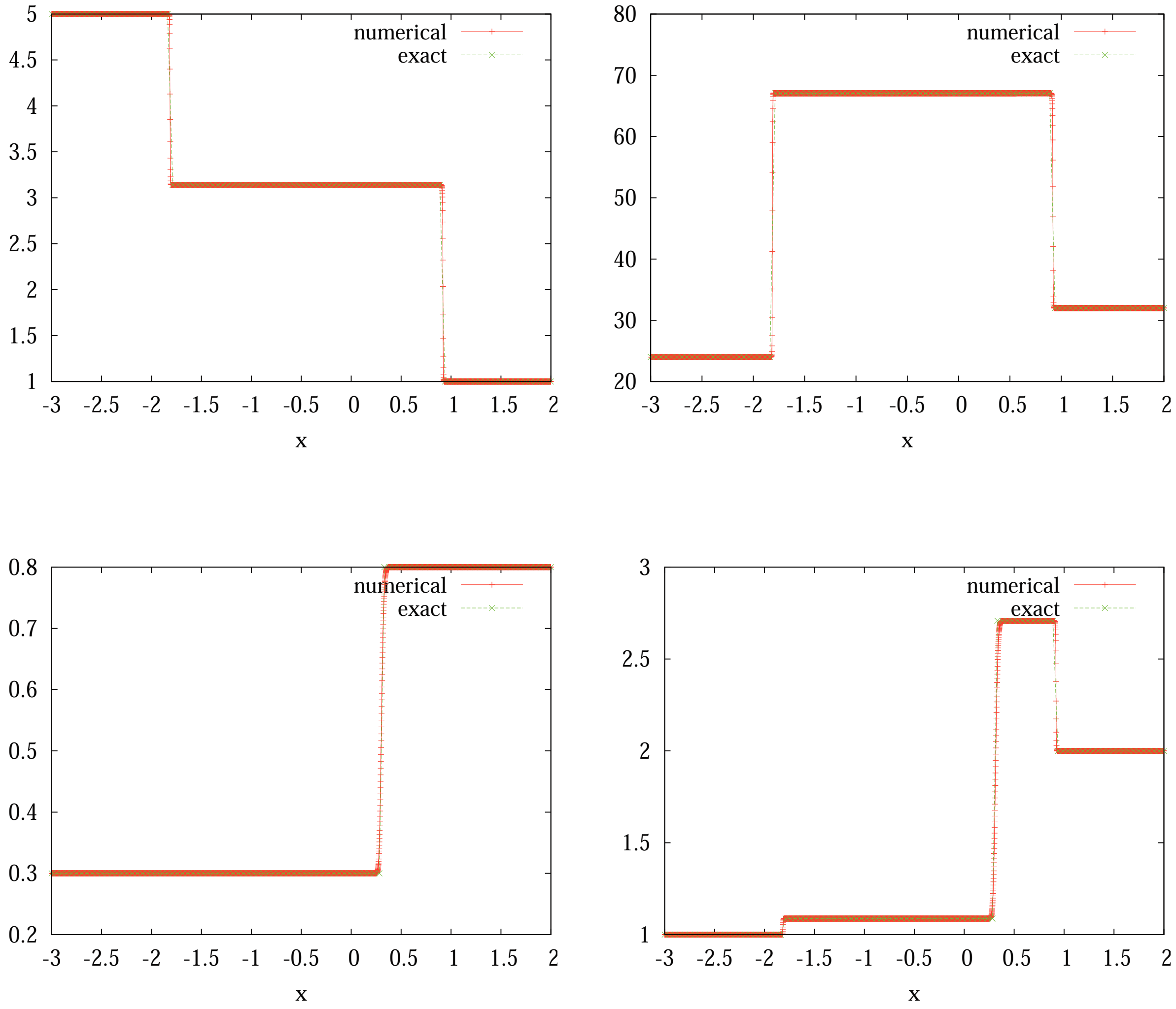

FIG. I.14 – Two-phase test : shock / contact discontinuity / shock – Centred RT scheme – Numerical solution at $t = 0.1$, with 5000 cells, $\delta t = 4.10^{-5}$ (so CFL= 0.75) and $\mu = 0.002$. Velocity (top left), pressure (top right), gas mass fraction (bottom left), density (bottom right).

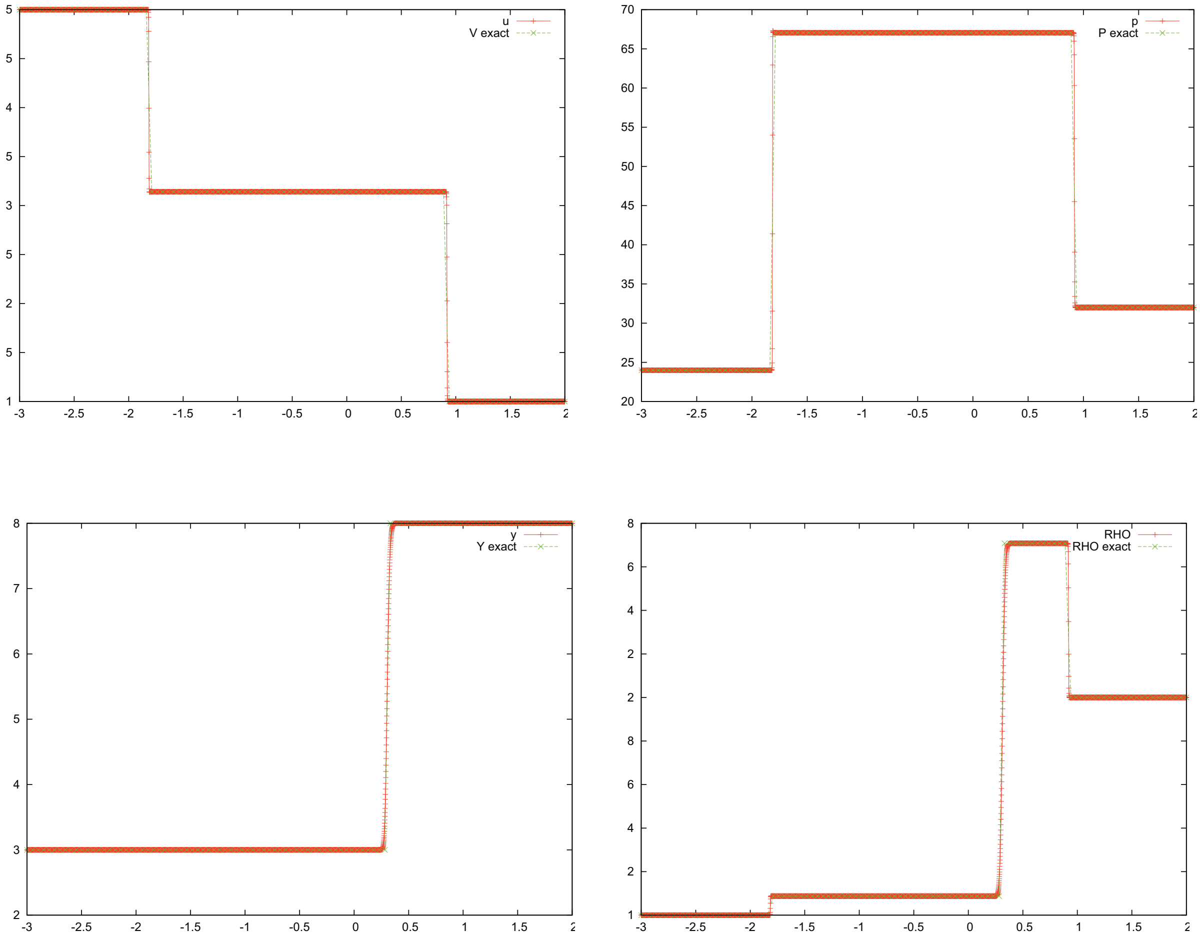

FIG. I.15 – Two-phase test : shock / contact discontinuity / shock – Centred MAC scheme – Numerical solution at $t = 0.1$, with 5000 cells, $\delta t = 4.\,10^{-5}$ (so CFL$=$ 0.75) and $\mu = 0.002$. Velocity (top left), pressure (top right), gas mass fraction (bottom left), density (bottom right).

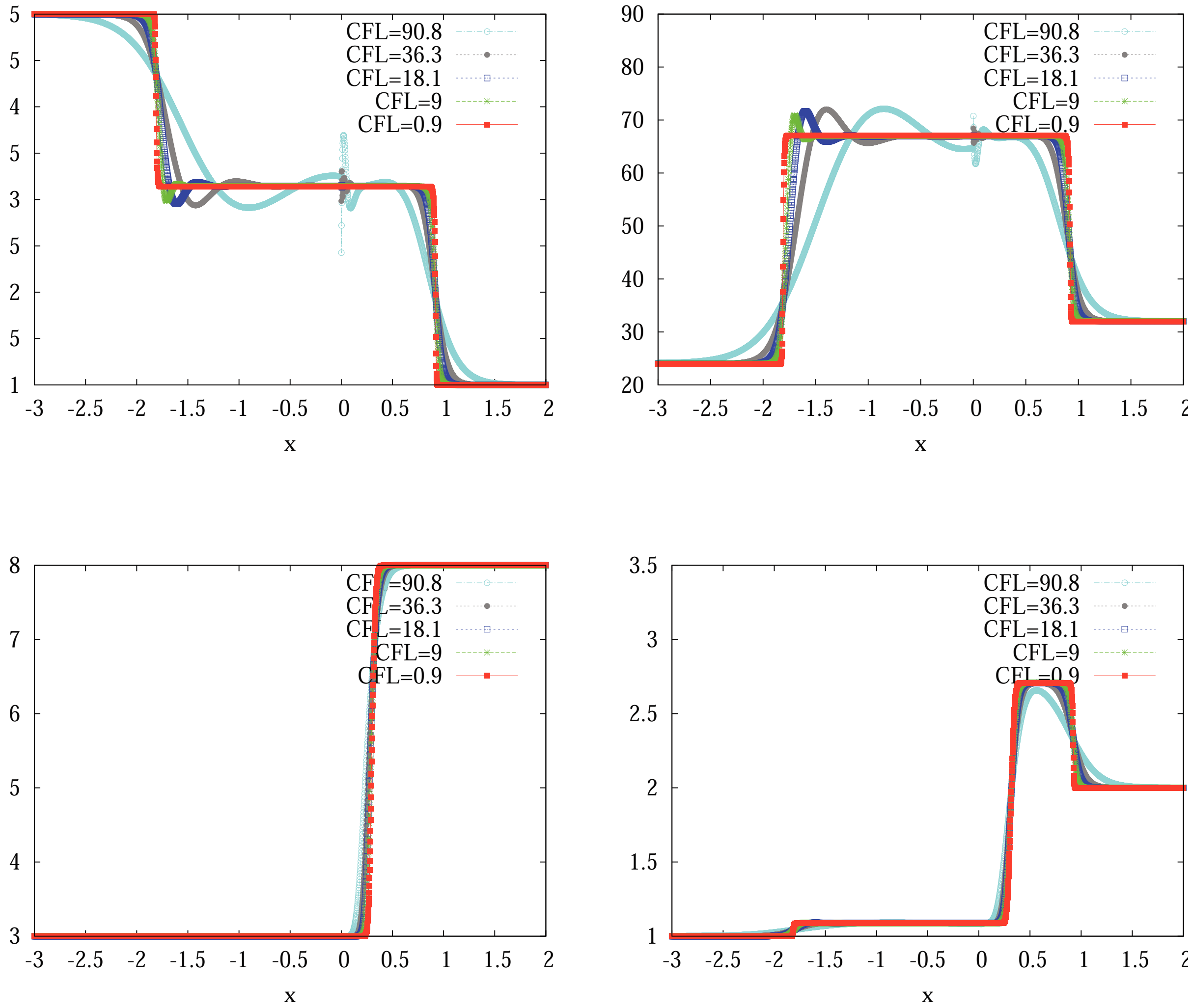

FIG. I.16 – Two-phase test : shock / contact discontinuity / shock – Centred RT scheme – Numerical solutions at $t = 0.1$ with 2500 cells, $\mu = 0.002$, for several CFL numbers. Velocity (top left), pressure (top right), gas mass fraction (bottom left), density (bottom right).

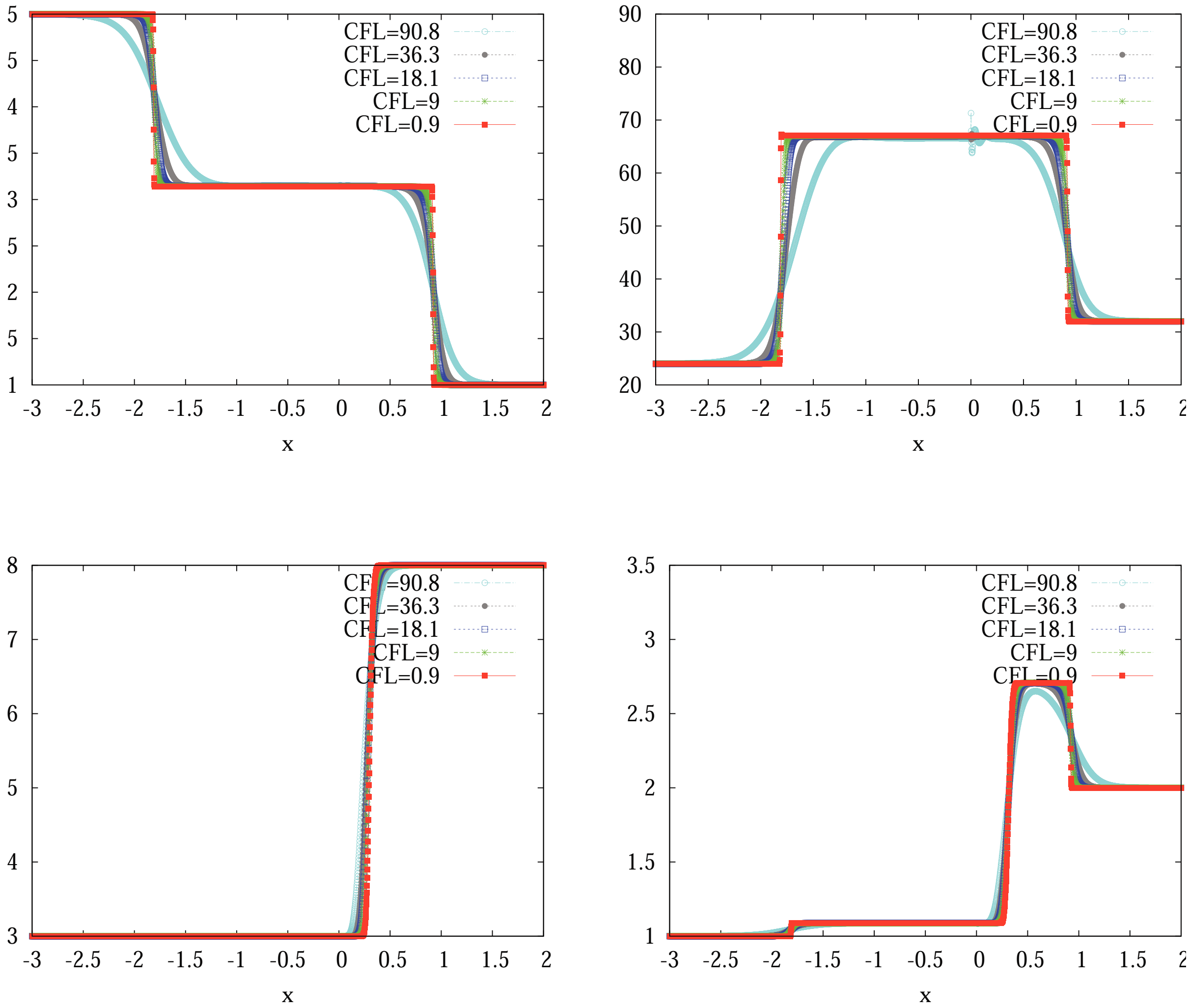

FIG. I.17 – Two-phase test : shock / contact discontinuity / shock – Centred MAC scheme – Numerical solutions at $t = 0.1$ with 2500 cells, $\mu = 0.002$, for several CFL numbers. Velocity (top left), pressure (top right), gas mass fraction (bottom left), density (bottom right).

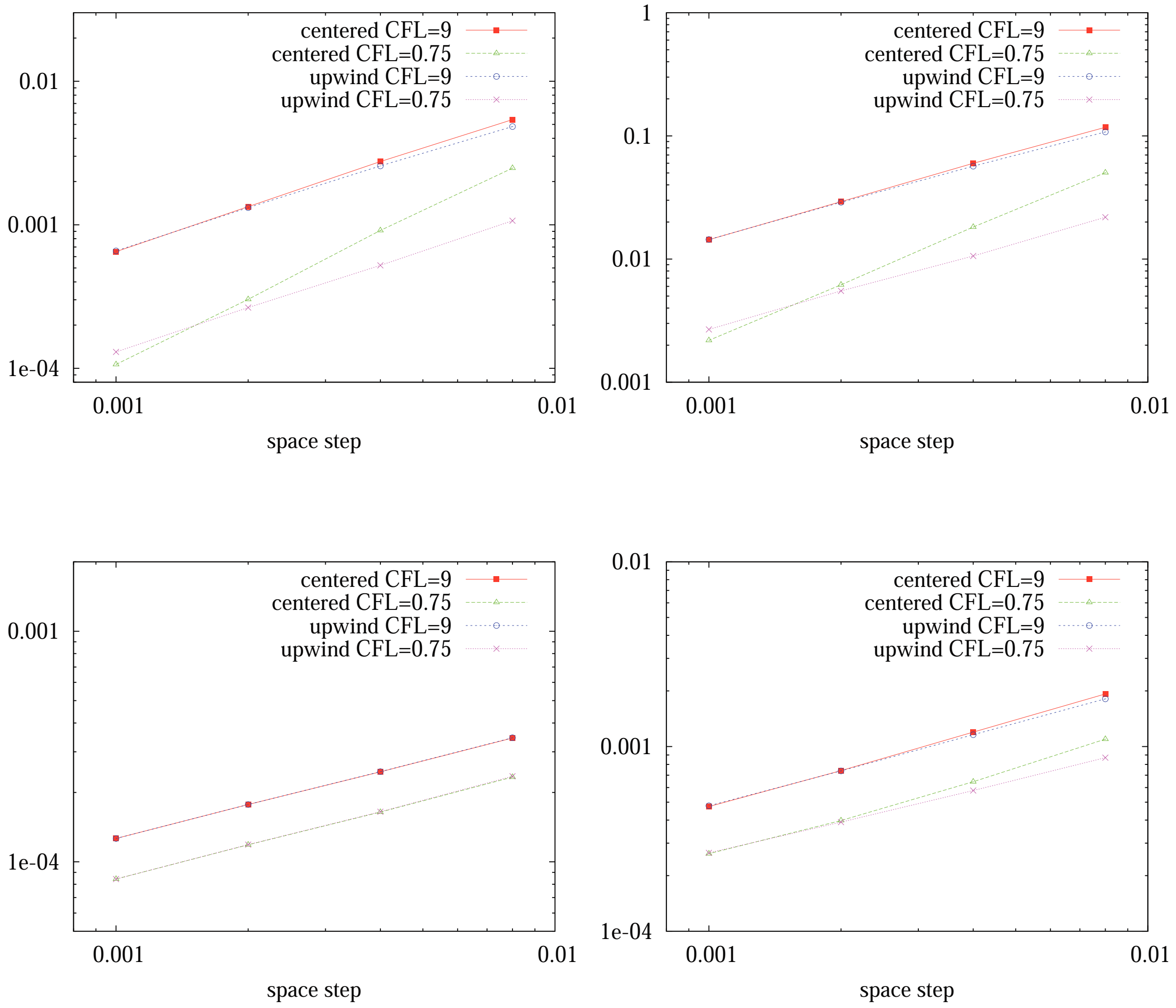

FIG. I.18 – Two-phase test : shock / contact discontinuity / shock – Centred and upwind RT schemes – L^1 norm of the error at $t = 0.1$ between the computed solution and the exact one, as a function of the mesh (or time) step, for two fixed CFL numbers. In the centred case, the used artificial viscosity is $\mu = 0.002$, *i.e.* a value close to the one which yields the more accurate results. Velocity (top left), pressure (top right), gas mass fraction (bottom left) and density (bottom right).

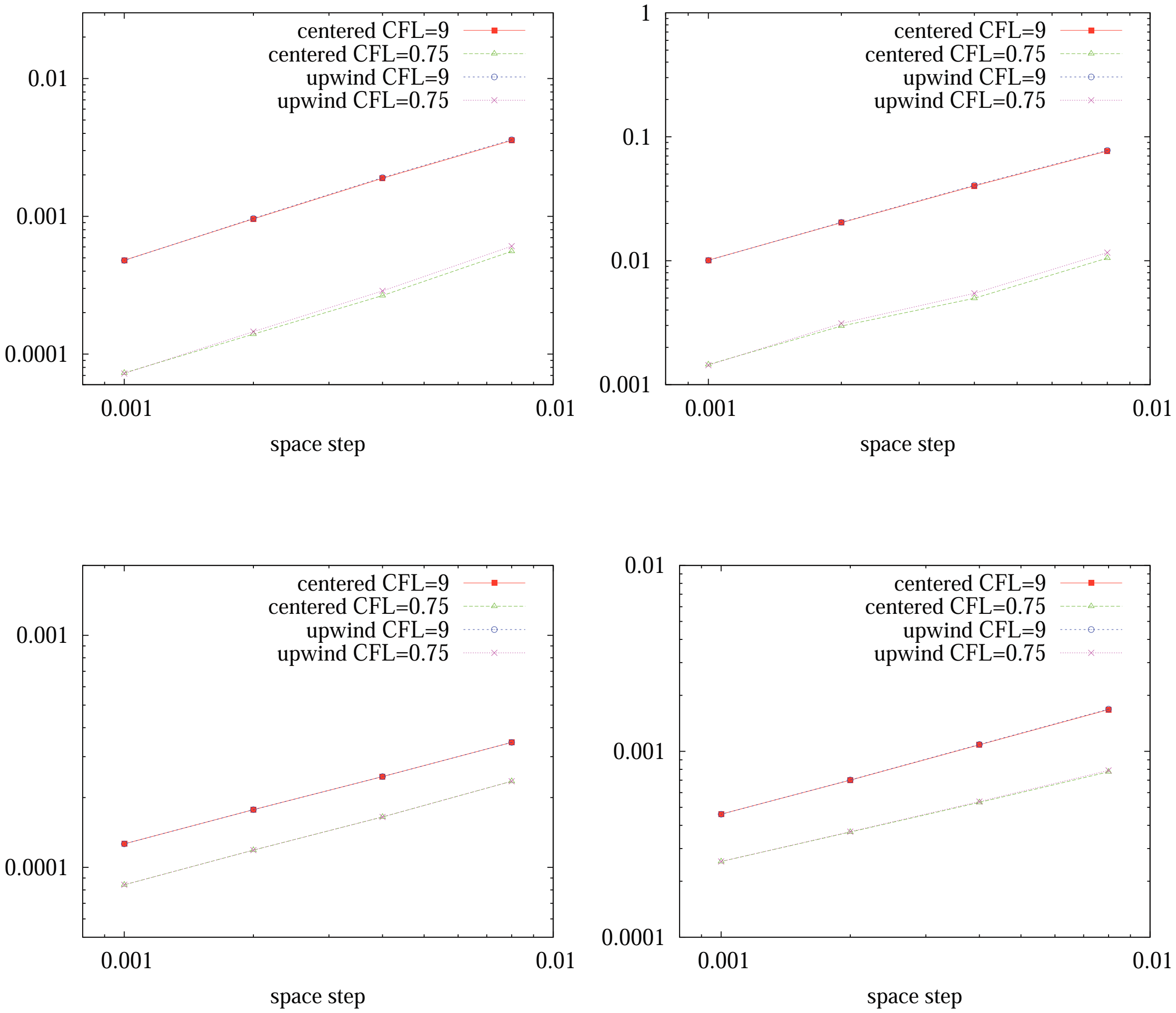

Fig. I.19 – Two-phase test : shock / contact discontinuity / shock – Centred and upwind MAC schemes – L^1 norm of the error at $t = 0.1$ between the computed solution and the exact one, as a function of the mesh (or time) step, for two fixed CFL numbers. In the centred case, the used artificial viscosity is $\mu = 0.002$, *i.e.* the same value as for the RT discretization. Velocity (top left), pressure (top right), gas mass fraction (bottom left) and density (bottom right).

I.3.2.b Second case : rarefaction-contact discontinuity-rarefaction

We conclude this study by the numerical simulation of a two-phase flow with rarefaction waves. The two initial constant states are given by :

$$\begin{pmatrix} \rho \\ \boldsymbol{u} \\ y \end{pmatrix}_L = \begin{pmatrix} 1 \\ 0 \\ 0.3 \end{pmatrix}, \qquad \begin{pmatrix} \rho \\ \boldsymbol{u} \\ y \end{pmatrix}_R = \begin{pmatrix} 2 \\ 2 \\ 0.8 \end{pmatrix}.$$

The numerical solutions at $t = 0.1$, obtained for 5000 cells, $\delta t = 0.0001$ and $\mu = 0.002$ with the RT and MAC centred schemes, presented in Figure I.20 (RT) and Figure I.21 respectively, are in close agreement with the exact solution.

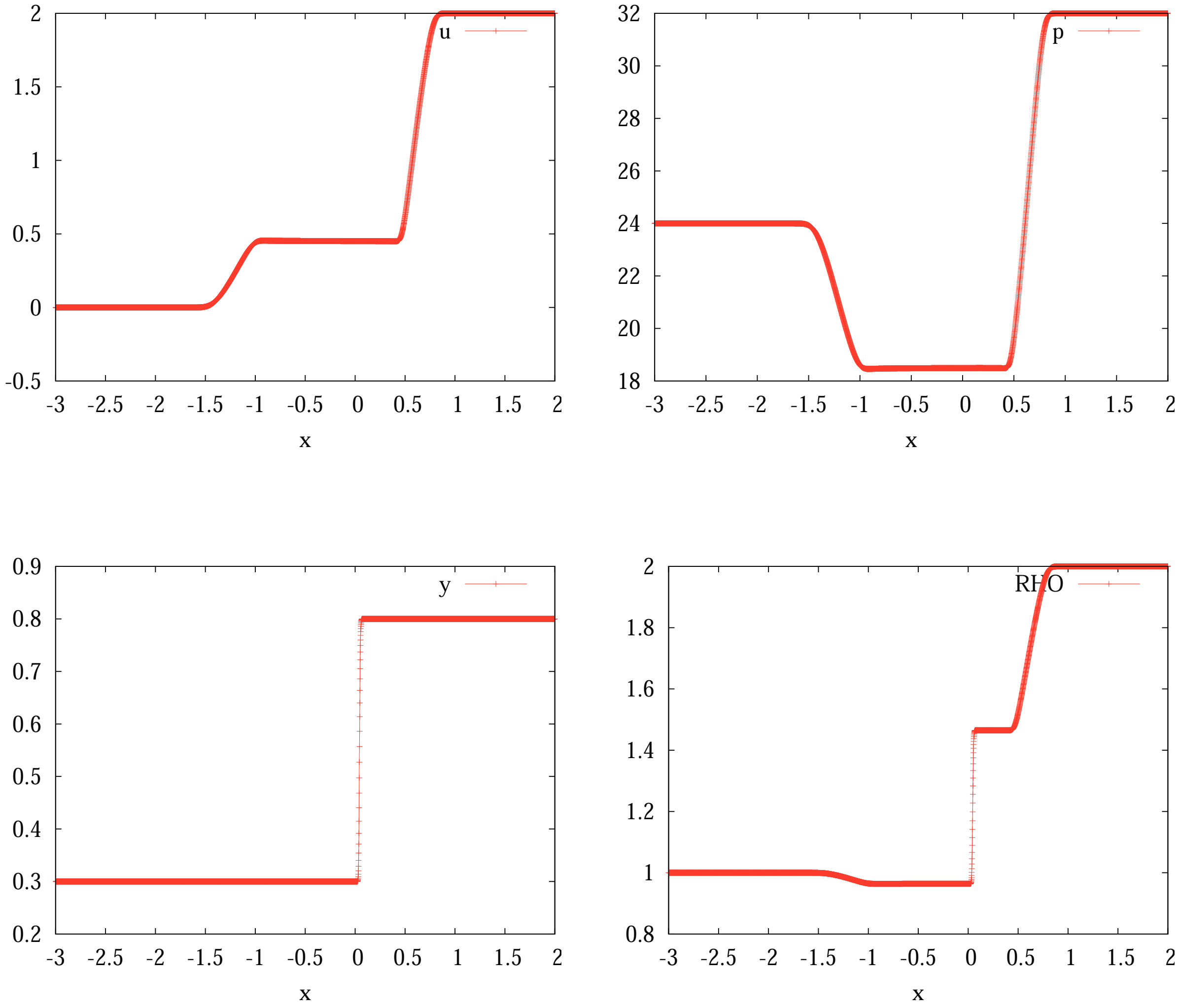

FIG. I.20 – Two-phase test : rarefaction wave / contact discontinuity / rarefaction wave – Centred RT scheme – Numerical solution at $t = 0.1$ with 5000 cells, $\delta t = 0.0001$ and $\mu = 0.002$. Velocity (top left), pressure (top right), gas mass fraction (bottom left), density (bottom right).

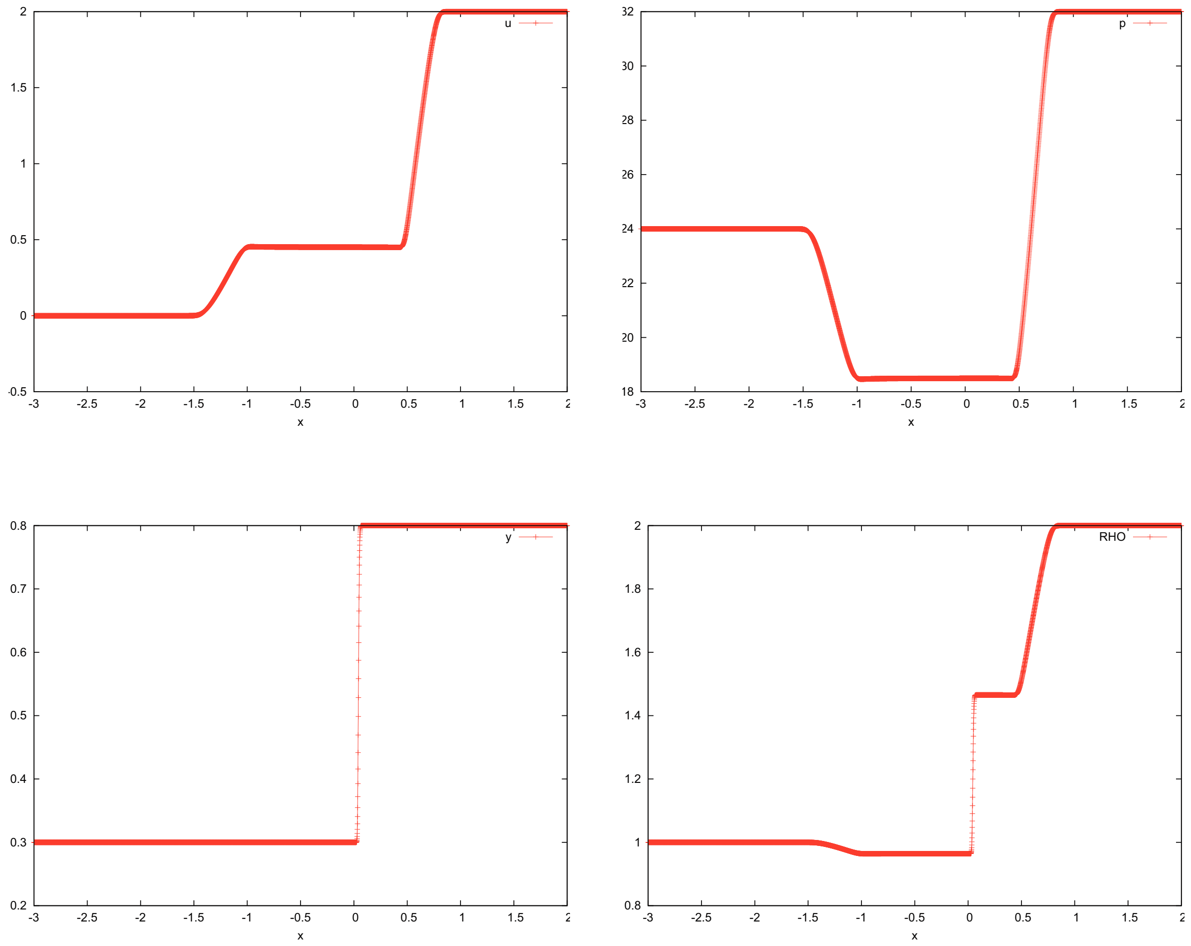

FIG. I.21 – Two-phase test : rarefaction wave / contact discontinuity / rarefaction wave – Centred MAC scheme – Numerical solution at $t = 0.1$ with 5000 cells, $\delta t = 0.0001$ and $\mu = 0.002$. Velocity (top left), pressure (top right), gas mass fraction (bottom left), density (bottom right).

I.4 A two-dimensional test case

We now turn to two-dimensional test cases. To this purpose, we use a "barotropic monophasic version", which we then extend to obtain a two-phase flow problem, of a test which is classical for (non-barotropic and monophasic) Euler equations [76, 31], and is often referred to as "the Mach 3 wind tunnel with step".

The flow enters through the left boundary a L-shaped domain, with a forward facing step, with the following geometry :

$$\Omega = (0,3) \times (0,1) \setminus (0.6,3) \times (0,0.2).$$

All the conservative variables, *i.e.* ρ and $\rho \boldsymbol{u}$ for the monophasic flow and ρ, ρy and $\rho \boldsymbol{u}$ in the two-phase case, are prescribed at the inlet (*i.e.* left) section. The equation of state of the fluid is $p = \rho$ and the inlet values are such that the Mach number is 3, which is obtained here by taking $\boldsymbol{u} = (3,0)^t$ and $\rho = 1$. On the top and bottom wall, we use homogeneous Neumann boundary conditions. The flow is free at the outflow (right) section, which means, since the resulting Mach number at this boundary is greater than one, that the three eigenvalues of the Jacobian matrix of the system are positive and therefore no boundary condition should be prescribed here ; however, since our discretization of the pressure gradient is centered and, less importantly, because we use a physical-like diffusion term in the momentum balance equation, we need an expression for the force which exerts at this surface, which we suppose given by :

$$\boldsymbol{\tau} \cdot \boldsymbol{n} - p\, \boldsymbol{n} = p_0\, \boldsymbol{n},$$

where p_0 is given the same value than the inlet pressure, *i.e.* $p_0 = 1$. We discuss later the effects of this boundary condition. The initial condition is $\boldsymbol{u} = (3,0)^t$ and $\rho = p = 1$.

The mesh is built from a regular $3n \times n$ Cartesian grid of the rectangle $(0,3) \times (0,1)$, suppressing meshes at the right bottom part of the domain (*i.e.* $(0.6,3) \times (0.,0.2)$) to take the step into account. The computations presented in this section are performed with the centered MAC scheme, and we use an artificial viscosity fixed at $\mu = 0.01$, which is in the range of what would be the numerical viscosity introduced by an upwinding technique, for the meshes used in this study.

The pressure field obtained with $n = 500$ (*i.e.* from a 1500×500 grid) and $\delta t = 2.\,10^{-3}$ is shown in Figure I.22. As in the non-barotropic case, we obtain a shock upflow the step, which propagates and reflects on the boundaries. Here, however, the shock moves slowly upward, while it is stationary in the non-barotropic case (so, contrary to this latter case, the flow is not steady at $t = 4$).

Besides the fact that we use, for numerical reasons, a non-physical outflow boundary condition, the time-splitting of pressure correction methods is also known to introduce spurious pressure boundary conditions. It is indeed the case for the present scheme, even if it is derived by an algebraic splitting (*i.e.* by discretizing first the equations up to obtain an implicit fully discrete scheme and then splitting in time, instead of first writting a split time semi-discrete algorithm with (artificial) boundary conditions explicitely stated at each step) : we show in [12] that the elliptic problem solved at the correction step for the pressure increment takes the form of a finite volume diffusion problem, with homogeneous Neumann conditions at the boundary where the velocity is prescribed and homogeneous Dirichlet conditions when the velocity is free. In the present case, it means in particular that the pressure suffers from a numerical boundary condition at the outlet section which tends to fix it at the initial value. Note, however, that

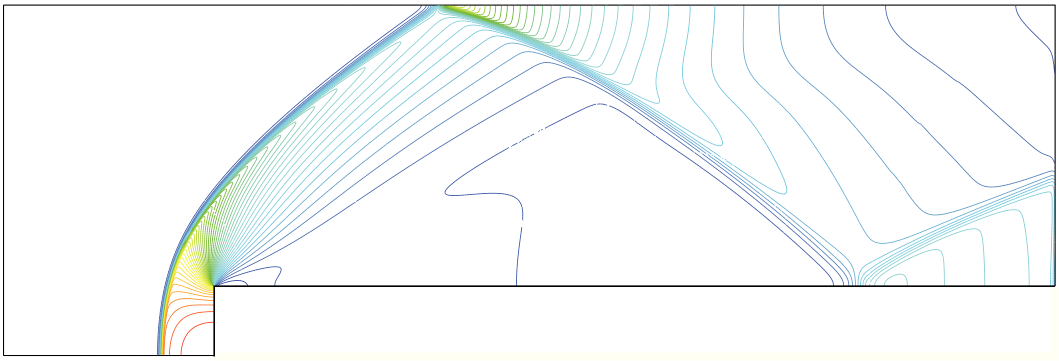

Fig. I.22 – Wind tunnel with step, monophasic case – Centred MAC scheme – Isolines of the pressure field obtained at t=4 with a 1500 × 500 mesh and $\delta t = 5.\,10^{-4}$. Minimal value (blue) : p= 0.8 – Maximal value (red) : p= 9.48

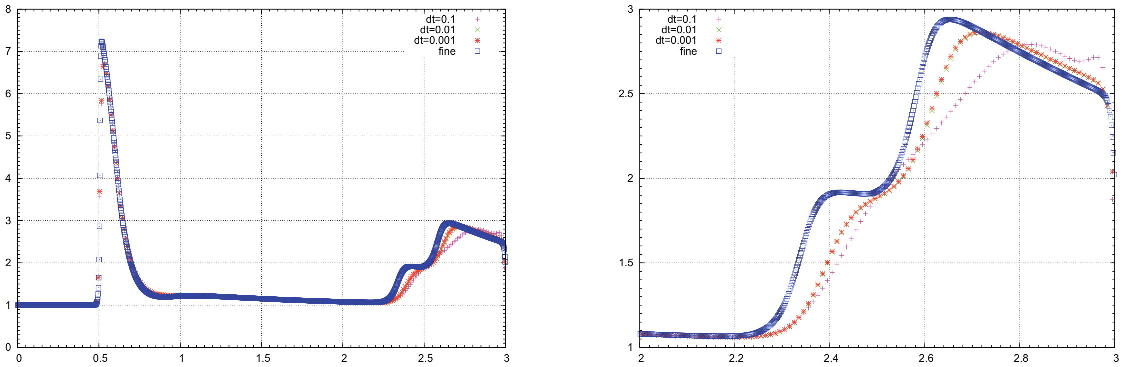

Fig. I.23 – Wind tunnel with step, monophasic case – Centred MAC scheme – Pressure obtained at t=4 along the line $x_2 = 0.3$, with an 300 × 100 mesh and various time steps, and with a 1500 × 500 mesh and $\delta t = 5.\,10^{-4}$.

this boundary condition is only prescribed in the "finite volume way" (*i.e.* through the expression of the flux), which may be seen as a penalization process with a $\delta t/h$ coefficient, so this condition is relaxed when this latter ratio is small [12]. We observe in Figure I.22 that this outlet condition indeed generates a pressure boundary layer. To investigate this phenomenon in more detail, we plot in Figure I.23 the value of the pressure along the $x_2 = 3$ line, for various meshes and time steps. We observe that the perturbation of the solution remains localized, and that, as wellknown for incompressible flow problems, the extension of the affected zone decreases with the space step. Besides, we also see that the computation is at least qualitatively correct for rather coarse meshes and time steps (using $v = |\boldsymbol{u}| + c = 4$ with c the speed of sound at the inlet section, $\delta t = 0.1$ (resp. $\delta t = 0.01$) corresponds with $n = 100$ to CFL=40 (resp. CFL=4)) ; in particular, convergence with respect to the time step seems to be reached, for this particuliar flow, for CFL numbers far greater than 1.

We now turn to a two-phase case, which is obtained by initializing the gas mass fraction by $y = 0.1$ for

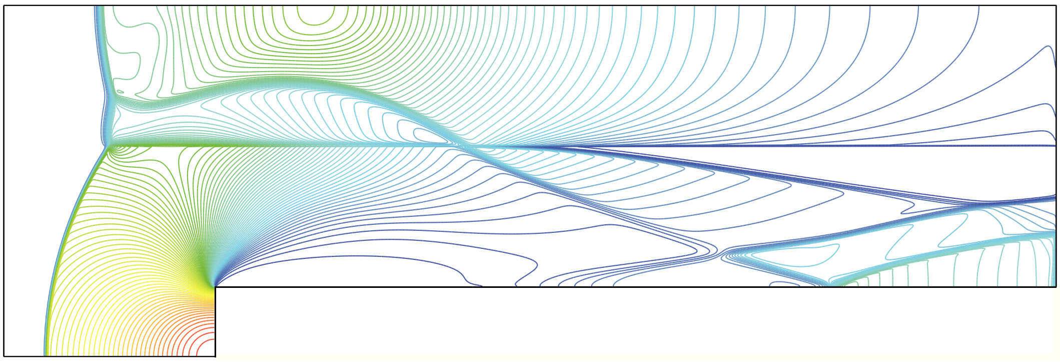

FIG. I.24 – Wind tunnel with step, non-uniform y case – Centred MAC scheme – Isolines of the pressure field obtained at t=1.6 with a 1500x500 mesh and $\delta t = 5.\,10^{-4}$. Minimal value (blue) : p= 0.095 – Maximal value (red) : p= 14.

$x_2 \leq 0.6$ and $y = 1$ in the rest of the domain. We choose $\rho_g = p$ and $\rho_\ell = 10$. We recall [32] that the speed of sound is given, in the two-phase case, by :

$$c^2 = \frac{\partial_p(\varrho_g)\ \rho_\ell\ z}{(\rho_\ell + z - \rho)^2},$$

with $z = \rho y$ and, here, $\partial_p(\varrho_g) = 1$. This relation shows that the speed of sound is lower for $y = 0.1$ than for $y = 1$, and we adjust the inlet velocity to keep the value of the Mach number at 3 in the two-phase zone. Inlet conditions are then given by $\boldsymbol{u} = (3c(y = 0.1), 0)^t)$, $y = 0.1$ for $x_2 \in (0, 0.6)$ and $y = 1$ for $x_2 \in (0.6, 1)$, and ρ given by the equation of state of the mixture with the local value of y and $p = 1$.

As a first step, we only solve the equations with $y(\boldsymbol{x})$ fixed at its initial value and independent of time (doing so,, we compute in fact a barotropic flow in a medium with a space-dependent equation of state). As a consequence of this change of equation of state, we observe that the shock moves upward more rapidly, and interaction with the inlet boudary conditions occurs as soon as $t \approx 2$; consequently, we restrict the time interval of computation, and stop at $t = 1.6$. The final time pressure, with a mesh built from the 1200×400 grid and $\delta t = 1.25\,10^{-3}$, is shown in Figure I.24. We observe that part of the pressure waves reflects at the y transition, the reflected wave propagating in a direction almost parallel to the transition (*i.e.* horizontal), thus giving a quite complicated structure.

Finally, we perform the same computation with the whole set of equations governing the two-phase flow. The obtained pressure, still at $t = 1.6$ and with the same grid, is shown in Figure I.25. The fist part of the flow shows some similarities with the previous computation, but the pressure evolution is quite different downstream, due to the fact that the liquid phase is now transported by the flow.

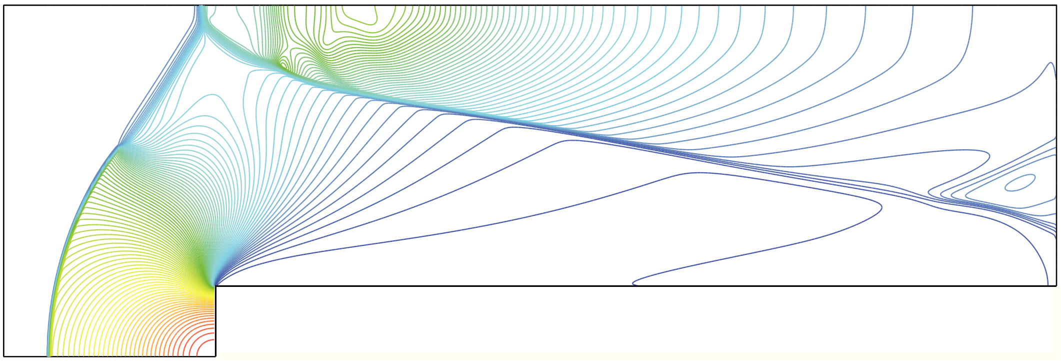

FIG. I.25 – Wind tunnel with step, two-phase case – Centred MAC scheme – Isolines of the pressure field obtained at t=1.6 with a 1500x500 mesh and $\delta t = 5.\,10^{-4}$. Minimal value (blue) : p= 0.086 – Maximal value (red) : p= 13.8.

I.5 Conclusion

In this paper, we have assessed the capability of a scheme issued from the incompressible flow context, namely a pressure correction scheme, to compute discontinuous solutions of hyperbolic systems. Numerical tests show that, provided that a sufficient numerical dissipation is introduced in the scheme, it converges to the (weak and entropic) solution to the continuous problem ; in addition, it shows a satisfactory behaviour up to large CFL numbers. Since the scheme boils down to a usual projection scheme when the density is constant, this approach yields an algorithm which is robust with respect to the flow Mach number, and the present solver is indeed now routinely used to compute viscous two-phase low Mach number flows, as bubble columns for instance.

The present work may be extended in various ways. First, the observed convergence can be conforted by theoretical arguments ; even if a complete convergence proof seems difficult at this time, because of the lack of compactness of sequences of discrete solutions due, in particular, to the absence of diffusion terms, it is possible to show, for monophasic flows, by passing to the limit in the scheme, that any limit of a convergent sequence is an entropy weak solution to the continuous problem (see next Chapter of this document).
Second, several variants of the scheme may be envisaged. The time discretization may be changed to an explicit one, to compute highly transient flows where a time-step limitation is not too stringent in practice ; such a scheme has been implemented, and first numerical result are promising. The extension of the above mentioned theoretical results to such a time discretization, under stability restrictions, seems possible. Third, in its present state, the scheme appears to be rather diffusive. Several directions exist to cure this problem. For instance, the artificial viscosity necessary for the scheme to converge could be monitored by *a posteriori* indicators, following the ideas developed in [30, 31]. Another route, especially for the explicit variant of the scheme, is to implement MUSCL techniques ; this work is underway.

Finally, let us mention that the present scheme has been extented to usual (*i.e.* non-batropic) Euler and Navier-Stokes equations, with a quite similar numerical behaviour and theoretical basis (see Chapters 3 and 4 of this document).

CHAPITRE

II Consistent staggered schemes for compressible flows – Part I : barotopic Navier-Stokes equations.

In this paper, we analyse the convergence of the pressure correction scheme introduced in [20] for the compressible barotropic Navier-Stokes equations in one space dimension. We use the staggered Marker And Cell (MAC) grid for the spatial discretization. We prove that if the solutions given by the version of the time-implicit and the pressure correction schemes converge to some limit, then this limit is an entropy weak solution of the continous problem.

Plan du Chapitre II

II.1 Introduction 65

II.2 Meshes and unknowns 65

II.3 Some results associated to finite volume convection operators 66

II.4 An implicit scheme 69

II.4.1 The scheme 69

II.4.2 Estimates 71

II.4.2.a The discrete kinetic energy balance equation 71

II.4.2.b The discrete elastic potential balance equation 72

II.4.2.c Stability estimates 72

II.4.3 Passing to the limit in the scheme 74

II.5 Pressure correction scheme 80

II.5.1 The scheme 80

II.5.2 Estimates 81

II.5.3 Passing to the limit in the scheme 83

II.1 Introduction

The problem addresed in this paper is the system of the so-called barotropic compressible Navier-Stokes equations, which reads :

$$\left|\begin{aligned} & \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\, u) = 0 \\ & \frac{\partial}{\partial t}(\rho\, u) + \operatorname{div}(\rho\, u \otimes u) + \nabla p = 0 \\ & p = \rho^{\gamma} \end{aligned}\right. \tag{II.1}$$

where t stands for the time, ρ, u and p are the (average) density, velocity and pressure in the flow. The three above equations are respectively the mass balance, the momentum balance and the equation of state of the fluid ; $\gamma \geqslant 1$ is a coefficient specific to the fluid considered. The problem is defined over an open bounded connected subset Ω of $\mathbb{R}$, and a finite time interval $[0, T)$.

II.2 Meshes and unknowns

Let $\mathcal{M}$ be a decomposition of the domain Ω, supposed to be regular in the usual sense of the finite element literature (*eg.* [9]). The cells may be :

- for a general domain Ω, either convex quadrilaterals ($d = 2$) or hexahedra ($d = 3$) or simplices, both type of cellsr being possibly combined in a same mesh,
- for a domain the boundaries of which are hyperplanes normal to a coordinate axis, rectangles ($d = 2$) or rectangular parallelepipeds ($d = 3$) (the faces of which, of course, are then also necessarily normal to a coordinate axis).

By $\mathcal{E}$ and $\mathcal{E}(K)$ we denote the set of all $(d-1)$-faces σ of the mesh and of the element $K \in \mathcal{M}$ respectively. The set of edges included in the boundary of Ω is denoted by $\mathcal{E}_{\rm ext}$ and the set of internal ones (*i.e.* $\mathcal{E} \backslash \mathcal{E}_{\rm ext}$) is denoted by $\mathcal{E}_{\rm int}$; a face $\sigma \in \mathcal{E}_{\rm int}$ separating the cells K and L is denoted by $\sigma = K|L$. The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-measure of the face σ. For $1 \leq i \leq d$, we denote by $\mathcal{E}^{(i)} \subset \mathcal{E}$ the subset of the faces of $\mathcal{E}$ which are perpendicular to the i^{th} unit vector of the canonical basis of $\mathbb{R}^d$.

The space discretization is staggered, using either the Marker-And Cell (MAC) scheme [37, 36], or non-conforming low-order finite element approximations, namely the Rannacher and Turek element (RT) [65] for quadrilateral or hexahedric meshes, or the Crouzeix-Raviart (CR) element [11] for simplicial meshes.

For all these space discretizations, the degrees of freedom for the pressure, the density and the internal energy are associated to the cells of the mesh $\mathcal{M}$, and are denoted by :

$$\big\{p_K,\ \rho_K,\ e_K,\ K \in \mathcal{M}\big\}.$$

Let us then turn to the degrees of freedom for the velocity.

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – The degrees of freedom for the velocities are located at the center of the faces of the mesh, and we choose the version of the element where they represent the average of the velocity through a face. The set of degrees of freedom reads :

$$\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E},\ 1 \leq i \leq d\}.$$

- **MAC** discretization – The degrees of freedom for the i^{th} component of the velocity, defined at the centres of the face $\sigma \in \mathcal{E}^{(i)}$, are denoted by :

$$\big\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)},\ 1 \leq i \leq d\big\}.$$

II.3 Some results associated to finite volume convection operators

We gather in this section some results concerning the discretization by the finite volume method of two convection operators :

- the first one reads, at the continuous level, $\rho \rightarrow \mathcal{C}(\rho) = \partial_t \rho + \mathrm{div}(\rho \boldsymbol{u})$, where $\boldsymbol{u}$ stands for a given velocity field, which is not assumed to satisfy any divergence constraint,
- the second one is $z \rightarrow \mathcal{C}_\rho(z) = \partial_t(\rho z) + \mathrm{div}(\rho z \boldsymbol{u})$, where ρ and $\boldsymbol{u}$ stands for two given scalar and vector fields, which are supposed to satisfy $\partial_t \rho + \mathrm{div}(\rho \boldsymbol{u}) = 0$.

Multiplying these operators by functions depending on the unknown is currently used to obtain convection operators acting over different variables, possibly with residual terms : one may think, for instance, to the theory of renormalized solutions (for the first one), or, in mechanics, to the derivation of the so-called kinetic energy transport identity (for the second one). The results provided in this section are discrete variants of such relations.

We begin with a property of $\mathcal{C}$, which, at the continuous level, may be formally obtained as follows. Let ψ be a regular function from $(0, +\infty)$ to $\mathbb{R}$; then :

$$\begin{aligned}\psi'(\rho)\ \mathcal{C}(\rho) = \psi'(\rho)\, \partial_t(\rho) + \psi'(\rho)\, \boldsymbol{u} \cdot \boldsymbol{\nabla} \rho + \psi'(\rho)\, \rho\ \mathrm{div}\boldsymbol{u} \\ = \partial_t(\psi(\rho)) + \boldsymbol{u} \cdot \boldsymbol{\nabla} \psi(\rho) + \rho\, \psi'(\rho)\, \mathrm{div}\boldsymbol{u},\end{aligned}$$

so adding and subtracting $\psi(\rho)\,\mathrm{div}\boldsymbol{u}$ yields :

$$\psi'(\rho)\ \mathcal{C}(\rho) = \partial_t\big(\psi(\rho)\big) + \mathrm{div}\big(\psi(\rho)\boldsymbol{u}\big) + \big(\rho\psi'(\rho) - \psi(\rho)\big)\ \mathrm{div}\boldsymbol{u}. \tag{II.2}$$

Obtaining a proof of this last identity, in a weak sense and with minimal regularity assumptions for ρ and $\boldsymbol{u}$ and increasing properties of ψ is the object of the theory of renormalized solutions. The following lemma states a discrete analogue to (II.2).

Lemma II.3.1
Let $K \in \mathcal{M}$. Let ρ_K^* and ρ_K be two positive real numbers. For $\sigma \in \mathcal{E}(K)$, let F_σ be a quantity associated to the face σ and the control volume K, defined by

$$\forall \sigma \in \mathcal{E}(K), \qquad F_\sigma = \rho_\sigma\ V_\sigma.$$

where ρ_σ and V_σ are a positive real number and a real number respectively, both associated to the edge σ. Let ψ be a twice continuously differentiable function, defined over $(0, +\infty)$.

Then the following identity holds :

$$\Big[\frac{|K|}{\delta t}\ (\rho_K-\rho_K^*)+\sum_{\sigma\in\mathcal{E}(K)} F_\sigma\Big]\psi'(\rho_K)=\frac{|K|}{\delta t}\ \big[\psi(\rho_K)-\psi(\rho_K^*)\big]+\sum_{\sigma\in\mathcal{E}(K)}\psi(\rho_\sigma)\ V_\sigma \\ +\Big[\rho_K\psi'(\rho_K)-\psi(\rho_K)\Big]\sum_{\sigma\in\mathcal{E}(K)} V_\sigma+R_{\sigma,\delta t} \qquad \text{(II.3)}$$

where

$$R_{\sigma,\delta t}=\frac{1}{2}\frac{|K|}{\delta t}\psi''(\overline{\rho}_K)(\rho_K-\rho_K^*)^2-\frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)} V_\sigma\ \psi''(\overline{\rho}_\sigma)(\rho_\sigma-\rho_K)^2,$$

and, $\forall\sigma\in\mathcal{E}(K)$, $\overline{\rho}_K\in[\min(\rho_K,\rho_K^*),\max(\rho_K,\rho_K^*)]$ and $\overline{\rho}_\sigma\in[\min(\rho_\sigma,\rho_K),\max(\rho_\sigma,\rho_K)]$. If we suppose that the function ψ is convex and that $\rho_\sigma=\rho_K$ as soon as $V_\sigma\geq 0$, then the residual $R_{\sigma,\delta t}$ is non-negative.

Proof Let be a twice continuously differentiable function, defined over $(0,+\infty)$, and $K\in\mathcal{M}$. We have :

$$\Big[\frac{|K|}{\delta t}\ (\rho_K-\rho_K^*)+\sum_{\sigma\in\mathcal{E}(K)} F_\sigma\Big]\ \psi'(\rho_K)=\frac{|K|}{\delta t}\ (\rho_K-\rho_K^*)\ \psi'(\rho_K)+\sum_{\sigma\in\mathcal{E}(K)}\psi(\rho_\sigma)\ V_\sigma \\ +\sum_{\sigma\in\mathcal{E}(K)}\Big[\rho_\sigma\psi'(\rho_K)-\psi(\rho_\sigma)\Big]\ V_\sigma.$$

By the regularity assumption for ψ, we may write Taylor expansions of ψ to obtain that there exists reals numbers $\overline{\rho}_K\in[\min(\rho_K,\rho_K^*),\max(\rho_K^*,\rho_K^*)]$ and, for all the faces $\sigma\in\mathcal{E}(K)$, $\overline{\rho}_\sigma\in[\min(\rho_\sigma,\rho_K),\max(\rho_\sigma,\rho_K)]$ such that :

$$(\rho_K-\rho_K^*)\psi'(\rho_K)=\psi(\rho_K)-\psi(\rho_K^*)+\frac{1}{2}\psi''(\overline{\rho}_K)(\rho_K-\rho_K^*)^2,$$
$$\rho_\sigma\psi'(\rho_K)-\psi(\rho_\sigma)=\rho_K\psi'(\rho_K)-\psi(\rho_K)-\frac{1}{2}\psi''(\overline{\rho}_\sigma)(\rho_\sigma-\rho_K)^2,$$

which yields the result. □

We now turn to the second operator, for which we have, at the continuous level and formally, using twice the assumption $\partial_t\rho+\mathrm{div}(\rho\boldsymbol{u})=0$:

$$\psi'(z)\ \mathcal{C}_\rho(z)=\psi'(z)\big[\partial_t(\rho\, z)+\mathrm{div}(\rho\, z\,\boldsymbol{u})\big]=\psi'(z)\rho\big[\partial_t z+\boldsymbol{u}\cdot\boldsymbol{\nabla} z\big] \\ =\rho\big[\partial_t\psi(z)+\boldsymbol{u}\cdot\boldsymbol{\nabla}\psi(z)\big]=\partial_t\big(\rho\,\psi(z)\big)+\mathrm{div}\big(\rho\,\psi(z)\,\boldsymbol{u}\big).$$

Taking for z a component of the velocity field, this relation is the central argument used to derive the kinetic energy balance. The following lemma states a discrete counterpart of this identity.

Lemma II.3.2

Let $K\in\mathcal{M}$. Let ρ_K^* and ρ_K be two positive real numbers. For $\sigma\in\mathcal{E}(K)$, let F_σ be a quantity associated to the face σ, such that the following identity holds :

$$\frac{|K|}{\delta t}\ (\rho_K-\rho_K^*)+\sum_{\sigma\in\mathcal{E}(K)} F_\sigma=0. \qquad \text{(II.4)}$$

Let u_K^* and u_K be two real numbers, and, to each $\sigma \in \mathcal{E}(K)$, we associate a rela number u_σ. Let ψ be a twice continuously differentiable function, defined over $(0,+\infty)$. Then the following relation holds :

$$\Big[\frac{|K|}{\delta t}\ (\rho_K\, u_K - \rho_K^*\, u_K^*) + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ u_\varepsilon\Big]\ \psi'(u_K) \\ = \frac{|K|}{\delta t}\ \big[\rho_K\, \psi(u_K) - \rho_K^*\, \psi(u_K^*)\big] + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ \psi(u_\sigma) + R_{K,\delta t} \qquad \text{(II.5)}$$

where :

$$R_{K,\delta t} = \frac{1}{2}\frac{|K|}{\delta t}\rho_K^*\ \psi''(\overline{u}_K)(u_K - u_K^*)^2 - \frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ \psi''(\overline{u}_\sigma)\ (u_\sigma - u_K)^2,$$

with, $\overline{u}_K \in [\min(u_K, u_K^*), \max(u_K, u_K^*)]$ and, $\forall \sigma \in \mathcal{E}(K)$, $\overline{u}_\sigma \in [\min(u_\sigma, u_K), \max(u_\sigma, u_K)]$. If we suppose that the function ψ is convex and that $u_\sigma = u_K$ as soon as $F_\sigma \geq 0$, then the residual $R_{\sigma,\delta t}$ is non-negative.

If we now take for ψ the function $\psi(s) = s^2/2$ and write, $\forall \sigma \in \mathcal{E}(K)$, $u_\sigma = (u_K + u_{K|\overset{\bullet}{\sigma}})/2$ (or, in other words, define $u_{K|\overset{\bullet}{\sigma}}$ as $u_{K|\overset{\bullet}{\sigma}} = 2\,u_\sigma - u_K$), we get the following identity :

$$\Big[\frac{|K|}{\delta t}\ (\rho_K\, u_K - \rho_K^*\, u_K^*) + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ u_\varepsilon\Big]\ u_K \\ = \frac{1}{2}\ \frac{|K|}{\delta t}\ \big[\rho_K\, u_K^2 - \rho_K^*\, (u_K^*)^2\big] + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ u_K\ u_{K|\overset{\bullet}{\sigma}} + R_{K,\delta t}, \qquad \text{(II.6)}$$

with $\quad R_{K,\delta t} = \dfrac{1}{2}\dfrac{|K|}{\delta t}\rho_K^*\ (u_K - u_K^*)^2.$

Proof Let ψ be a twice continuously differentiable function, defined over $(0,+\infty)$. Using Equation (II.4), we obtain :

$$T_K = \Big[\frac{|K|}{\delta t}\ (\rho_K u_K - \rho_K^* u_K^*) + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ u_\sigma\Big]\psi'(u_K) = \\ \Big[\frac{|K|}{\delta t}\ \rho_K^*\ (u_K - u_K^*) + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma(u_\sigma - u_K)\Big]\psi'(u_K).$$

By a Taylor expansion of ψ, then there exists a real number $\overline{u}_K \in [\min(u_K^*, u_K), \max(u_K^*, u_K)]$ such that :

$$\psi'(u_K)\big(u_K - u_K^*\big) = \psi(u_K) - \psi(u_K^*) + \frac{1}{2}\psi''(\overline{u}_K)\ \big(u_K - u_K^*\big)^2$$

Then, using once again (II.4), we have :

$$\begin{aligned} T_K &= \frac{|K|}{\delta t}\ \rho_K^*\ \big(\psi(u_K) - \psi(u_K^*)\big) + \frac{1}{2}\frac{|K|}{\delta t}\rho_K^*\,\psi''(\overline{u}_K)\ (u_K - u_K^*)^2 + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\ (u_\sigma - u_K)\,\psi'(u_K) \\ &= \frac{|K|}{\delta t}\ \big(\rho_K\psi(u_K) - \rho_K^*\psi(u_K^*)\big) + \sum_{\sigma\in\mathcal{E}(K)} F_\sigma\big[\psi(u_K) + \psi'(u_K)(u_\sigma - u_K)\big] \\ &\qquad + \frac{1}{2}\frac{|K|}{\delta t}\rho_K^*\,\psi''(\overline{u}_K)\ (u_K - u_K^*)^2. \end{aligned}$$

Once again by a Taylor expansion of ψ, for any face $\sigma \in \mathcal{E}(K)$, there exists a real number $\overline{u}_\sigma \in [\min(u_\sigma, u_K), \max(u_\sigma, u_K)]$ such that :

$$\psi(u_K) + \psi'\,(u_K)(u_\sigma - u_K) = \psi(u_\sigma) - \frac{1}{2}\psi''(\overline{u}_\sigma)\,(u_\sigma - u_K)^2.$$

Hence :

$$T_K = \frac{|K|}{\delta t}\;\rho_K^*\;\big(\psi(u_K) - \psi(u_K^*)\big) + \sum_{\sigma\in\sigma(K))} F_\sigma\;\psi(u_\sigma) + R_{K,\delta t},$$

where $R_{K,\delta t}$ is given by the expression given in the statement of the lemma. This yields the first assertion of the lemma ; the last two ones are straightforward consequences of this equality. □

II.4 An implicit scheme

II.4.1 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 $n = 0, 1, \ldots, N-1$ be the constant time step. We consider an implicit-in-time numerical scheme, which reads in its fully discrete form :

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}^{n+1} = 0, \tag{II.7a}$$

$$\text{For } 1 \leq i \leq d, \left|\begin{array}{l} \forall\sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall\sigma \in \mathcal{E} \text{ otherwise,} \end{array}\right.$$

$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1}\boldsymbol{u}_{\sigma,i}^{n+1} - \rho_\sigma^n\boldsymbol{u}_{\sigma,i}^n) + \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1}\boldsymbol{u}_{\varepsilon,i}^{n+1} + |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = 0, \tag{II.7b}$$

$$\forall K \in \mathcal{M}, \qquad p_K^{n+1} = \wp(\rho_K^{n+1}) = (\rho_K^{n+1})^\gamma \tag{II.7c}$$

Equation (II.7a) is obtained by discretization of the mass balance over the primal mesh, and $F_{K,\sigma}^{n+1}$ stands for the mass flux across σ outward K, given by :

$$\forall\sigma \in \mathcal{E}_{\rm int},\ \sigma = K|L, \qquad F_{K,\sigma}^{n+1} = |\sigma|\;\widetilde{\rho}_\sigma^{n+1}\;\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}.$$

In this relation, the notation $\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}$ stands for the approximation of the normal velocity to the face σ outward K. For the MAC discretization, this quantity is given (up, possibly, to a change of sign) by the velocity degree of freedom located at the face ; for the RT and CR discretizations, it is computed by taking the inner product of the (vector valued) velocity on σ, $\boldsymbol{u}_\sigma^{n+1}$, and the outward normal vector $\boldsymbol{n}_{K,\sigma}$ (*i.e.* doing exactly what the notation says). The density at the face $\sigma = K|L$, $\widetilde{\rho}_\sigma^{n+1}$, is approximated by the upwind technique :

$$\widetilde{\rho}_\sigma^{n+1} = \left|\begin{array}{ll} \rho_K^{n+1} & \text{if } \boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma} \geq 0, \\ \rho_L^{n+1} & \text{otherwise.} \end{array}\right.$$

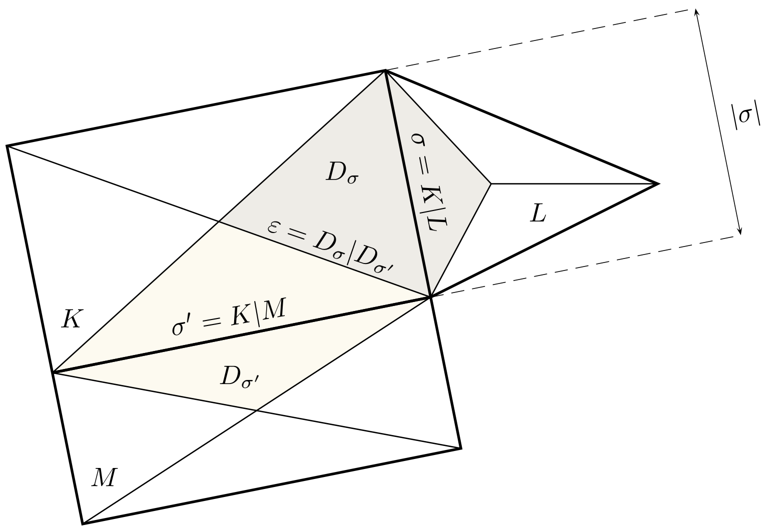

FIG. II.1 – Primal and dual meshes for the Rannacher-Turek and Crouzeix-Raviart elements.

We now turn to the discrete momentum balance (II.7b). For the MAC discretization, but also for the RT and CR discretization, the time derivative and convection terms are approximated in (II.7b) by a finite volume technique over a dual mesh, which we now define :

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – For the RT or CR discretization, the dual mesh is the same for all the velocity components. When $K \in \mathcal{M}$ is a simplex, a rectangles or a cuboid, for $\sigma \in \mathcal{E}(K)$, we define $D_{K,\sigma}$ as the cone with basis σ and with vertex the mass center of K. We thus obtain a partition of K in m sub-volumes, where m is the numbers of faces of the mesh, each sub-volume having the same measure $|D_{K,\sigma}| = |K|/m$. We extend this definition to general quadrangles and hexahedra, by supposing that we have built a partition still of equal-volume sub-cells, and with the same connectivities ; note that this is of course always possible, but that such a volume $D_{K,\sigma}$ may be no longer a cone, since, if K is far from a pallelogram, it may not be possible to built a cone having σ as basis, the opposite vertex lying in K and a volume equal to $|K|/m$. The volume $D_{K,\sigma}$ is referred to as the half-diamond cell associated to K and σ.
  For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the diamond cell D_σ associated to σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external face $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}(K)$, D_σ is just the same volume as $D_{K,\sigma}$.
- **MAC** discretization – For the MAC scheme, the dual mesh depends on the component of the velocity. For each of them, its definition differs from the RT or CR one only by the choice of the half-diamond cell, which, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, is now the rectangle of basis σ and of measure $|D_{K,\sigma}|$ equal to half the measure of K.

We denote by $|D_\sigma|$ the measure of the dual cell $|D_\sigma|$, and by $\varepsilon = D_\sigma|D_{\sigma'}$ the face separating two diamond cells D_σ and $D_{\sigma'}$ (see Figure II.1).

To make the discretization of the time derivative term complete, we must provide a definition for the ρ_σ^{n+1} and ρ_σ^n, which approximate the density on the edge σ at time t^{n+1} and t^n respectively. They are given by the following weighted average :

$$\forall \sigma \in \mathcal{E}_{\rm int},\ \sigma = K|L, \qquad |D_\sigma|\ \rho^n_\sigma = |D_{K,\sigma}|\ \rho^n_K + |D_{L,\sigma}|\ \rho^n_L. \tag{II.8}$$

We now turn to the convection term. The first task is to define the the discrete mass flux through the dual edge ε outward D_σ, denoted by $F^{n+1}_{\sigma,\varepsilon}$, the guideline for its construction being that we need a finite volume discretization of the mass balance equation over the diamond cells to hold :

$$\forall \sigma \in \mathcal{E}, \qquad |D_\sigma| \frac{\rho^{n+1}_\sigma - \rho^n_\sigma}{\delta t} + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F^{n+1}_{\sigma,\varepsilon} = 0, \tag{II.9}$$

in order to be be able to derive a discrete kinetic energy balance (see Section II.4.1 below). For a dual edge ε included in the primal cell K, this flux is computed as a linear combination (with constant coefficients, *i.e.* independent of the edge and the cell) of the mass fluxes through the faces of K, *i.e.* the quantities $(F^{n+1}_{K,\sigma})_{\sigma\in\mathcal{E}(K)}$ appearing in the discrete mass balance (II.7a). We do not give here this set of coefficients, and refer to [1, 38, 25] for a detailed construction of this approximation.

The quantity $\boldsymbol{u}^{n+1}_{\varepsilon,i}$ stands for an approximation of $\boldsymbol{u}^{n+1}_i$ on ε wich may be chosen centered or upwind, so, for $\varepsilon = D_\sigma|D_{\sigma'}$, reads :

$$\text{Centered case : } \boldsymbol{u}^{n+1}_{\varepsilon,i} = (\boldsymbol{u}^{n+1}_{\sigma,i} + \boldsymbol{u}^{n+1}_{\sigma',i})/2. \qquad \text{Upwind case : } \boldsymbol{u}^{n+1}_{\varepsilon,i} = \left| \begin{array}{ll} \boldsymbol{u}^{n+1}_{\sigma,i} & \text{if } F^{n+1}_{\sigma,\varepsilon} \geq 0, \\ \boldsymbol{u}^{n+1}_{\sigma',i} & \text{otherwise.} \end{array}\right.$$

The last term $(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}$ stands for the i-th component of the discrete pressure gradient at the face σ, which reads :

$$\text{for } \sigma \in \mathcal{E}_{\rm int},\ \sigma = K|L, \qquad (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = \frac{|\sigma|}{|D_\sigma|}\ (p^{n+1}_L - p^{n+1}_K)\ \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}^{(i)}.$$

II.4.2 Estimates

II.4.2.a The discrete kinetic energy balance equation

Let $\delta^{\rm up}$ be a coefficient defined by $\delta^{\rm up} = 1$ if an upwind discretization is used for the convection term in the momentum balance equation and $\delta^{\rm up} = 0$ in the centered case. With this notation, the momentum balance equation reads :

$$\begin{aligned} \frac{|D_\sigma|}{\delta t}(\rho^{n+1}_\sigma \boldsymbol{u}^{n+1}_{\sigma,i} - \rho^n_\sigma \boldsymbol{u}^n_{\sigma,i}) + \sum_{\varepsilon = D_\sigma|D_{\sigma'}} \frac{1}{2}\ F^{n+1}_{\sigma,\varepsilon}\ (\boldsymbol{u}^{n+1}_{\sigma,i} + \boldsymbol{u}^{n+1}_{\sigma',i}) \\ + \delta^{\rm up} \sum_{\varepsilon = D_\sigma|D_{\sigma'}} \frac{1}{2}\ |F^{n+1}_{\sigma,\varepsilon}|\ (\boldsymbol{u}^{n+1}_{\sigma,i} - \boldsymbol{u}^{n+1}_{\sigma',i}) + |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = 0. \end{aligned} \tag{II.10}$$

We multiply equation (II.10) by the corresponding velocity unknown $\boldsymbol{u}^{n+1}_{\sigma,i}$, which yields $T^{\rm conv}_{\sigma,i} + T^{\rm up}_{\sigma,i} + T^{\nabla}_{\sigma,i} = 0$, with :

$$\begin{aligned} T^{\rm conv}_{\sigma,i} &= \Big[\frac{|D_\sigma|}{\delta t}(\rho^{n+1}_\sigma \boldsymbol{u}^{n+1}_{\sigma,i} - \rho^n_\sigma \boldsymbol{u}^n_{\sigma,i}) + \sum_{\varepsilon = D_\sigma|D_{\sigma'}} \frac{1}{2}\ F^{n+1}_{\sigma,\varepsilon}\ (\boldsymbol{u}^{n+1}_{\sigma,i} + \boldsymbol{u}^{n+1}_{\sigma',i})\Big]\ \boldsymbol{u}^{n+1}_{\sigma,i}, \\ T^{\rm up}_{\sigma,i} &= \delta^{\rm up}\Big[\sum_{\varepsilon = D_\sigma|D_{\sigma'}} \frac{1}{2}\ |F^{n+1}_{\sigma,\varepsilon}|\ (\boldsymbol{u}^{n+1}_{\sigma,i} - \boldsymbol{u}^{n+1}_{\sigma',i})\Big]\ \boldsymbol{u}^{n+1}_{\sigma,i}, \\ T^{\nabla}_{\sigma,i} &= |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}\ \boldsymbol{u}^{n+1}_{\sigma,i}. \end{aligned}$$

From the identity (II.6), we get :

$$T^{\rm conv}_{\sigma,i} = \frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho_\sigma^{n+1}(\boldsymbol{u}_{\sigma,i}^{n+1})^2 - \rho_\sigma^n(\boldsymbol{u}_{\sigma,i}^n)^2\Big] + \frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}} F_{\sigma,\varepsilon}^{n+1}\ \boldsymbol{u}_{\sigma,i}^{n+1}\ \boldsymbol{u}_{\sigma',i}^{n+1} + \frac{|D_\sigma|}{2\,\delta t}\ \rho_\sigma^n\ \big(\boldsymbol{u}_{\sigma,i}^{n+1} - \boldsymbol{u}_{\sigma,i}^n\big)^2.$$

Let us define $R_{\sigma,i}$ by the sum of $-T^{\rm up}_{\sigma,i}$ and the opposite of the last term of $T^{\rm conv}_{\sigma,i}$:

$$R_{\sigma,i} = -\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\ \rho_\sigma^n\ \big(\boldsymbol{u}_{\sigma,i}^{n+1} - \boldsymbol{u}_{\sigma,i}^n\big)^2 - \delta^{\rm up}\Big[\sum_{\varepsilon=D_\sigma|D_{\sigma'}} \frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\ (\boldsymbol{u}_{\sigma,i}^{n+1} - \boldsymbol{u}_{\sigma',i}^{n+1})\Big]\ \boldsymbol{u}_{\sigma,i}^{n+1}.$$

With this notation, we thus obtain the following relation :

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho_\sigma^{n+1}(\boldsymbol{u}_{\sigma,i}^{n+1})^2 - \rho_\sigma^n(\boldsymbol{u}_{\sigma,i}^n)^2\Big] + \frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}} F_{\sigma,\varepsilon}^{n+1}\ \boldsymbol{u}_{\sigma,i}^{n+1}\ \boldsymbol{u}_{\sigma',i}^{n+1} + |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}\ \boldsymbol{u}_{\sigma,i}^{n+1} = R_{\sigma,i}. \tag{II.11}$$

II.4.2.b The discrete elastic potential balance equation

Let P the elastic potential which satisfies $P^{'}(z) = \frac{\wp(z)}{z^2}$. We multiply the discrete mass balance (II.7a) by $(\rho_K^{n+1}P(\rho_K^{n+1}))^{'}$; by Lemma II.3.1, we get :

$$|K|\frac{\rho_K^{n+1}P(\rho_K^{n+1}) - \rho_K^n P(\rho_K^n)}{\delta t} + \sum_{\sigma\in\mathcal{E}(K)} \rho_\sigma^{n+1}P(\rho_\sigma^{n+1})u_\sigma^{n+1}\cdot n_{K,\sigma} + p_K^{n+1}\sum_{\sigma\in\mathcal{E}(K)} u_\sigma^{n+1}\cdot n_{K,\sigma} = R_{\sigma,\delta t} \tag{II.12}$$

where

$$R_{\sigma,\delta t} = -\frac{1}{2}\frac{|K|}{\delta t}(\overline{\rho}_K^{n+1}P(\overline{\rho}_K^{n+1}))^{''}(\rho_K^{n+1} - \rho_K^n)^2 - \frac{1}{2}\sum_{\sigma=K|L} \delta_\sigma^{\rm up}|u_\sigma^{n+1}|(\overline{\rho}_\sigma^{n+1}P(\overline{\rho}_\sigma^{n+1}))^{''}(\rho_L^{n+1} - \rho_K^{n+1})^2 \leq 0,$$

$\overline{\rho}_K^{n+1} \in [\min(\rho_K^{n+1},\rho_K^n), \max(\rho_K^{n+1},\rho_K^n)]$ and $\overline{\rho}_\sigma^{n+1} \in [\min(\rho_\sigma^{n+1},\rho_K^{n+1}), \max(\rho_\sigma^{n+1},\rho_K^{n+1})]$ for all $\sigma \in \mathcal{E}(K)$, and $\delta_\sigma^{\rm up}$ is defined by $\delta_\sigma^{\rm up} = 1$ if $u_\sigma^{n+1}\cdot n_{K,\sigma} < 0$ and $\delta_\sigma^{\rm up} = 0$ otherwise.

II.4.2.c Stability estimates

Lemma II.4.1 (Stability of the advection operator)
Let $(\rho^*_\sigma)_{\sigma\in\mathcal{E}}$ and $(\rho_\sigma)_{\sigma\in\mathcal{E}}$ be two families of positive real number satisfying the following set of equations :

$$\forall\sigma\in\mathcal{E},\qquad \frac{|D_\sigma|}{\delta t}\,(\rho_\sigma - \rho^*_\sigma) + \sum_{\varepsilon\in\overline{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon} = 0 \tag{II.13}$$

where $F_{\sigma,\varepsilon}$ is a quantity associated to the edge ε and to the control volume D_σ ; we suppose that, for any internal edge $\varepsilon = D_{\sigma'}|D_\sigma$, $F_{\sigma,\varepsilon} = -F_{\sigma',\varepsilon}$. Let $(u^*_\sigma)_{\sigma\in\mathcal{E}}$ and $(u_\sigma)_{\sigma\in\mathcal{E}}$ be two families of real numbers. For any internal edge $\varepsilon = D_{\sigma'}|D_\sigma$, we define u_ε either by a centered approximation of u on ε : $u_\varepsilon = \frac{1}{2}(u_\sigma + u_{\sigma'})$, or by an upwind approximation of u on ε : $u_\varepsilon = u_\sigma$ if $F_{\sigma,\varepsilon} \geq 0$ and $u_\varepsilon = u_{\sigma'}$ otherwise. The following stability property holds :

$$\sum_{\sigma\in\mathcal{E}} u_\sigma \left[\frac{|D_\sigma|}{\delta t} \, (\rho_\sigma \, u_\sigma - \rho^*_\sigma \, u^*_\sigma) + \sum_{\varepsilon\in\overline{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon} \, u_\varepsilon \right] \geq \ \frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t} \, \left[\rho_\sigma \, u^2_\sigma - \rho^*_\sigma \, {u^*_\sigma}^2 \right] + \frac{1}{2} \sum_{\sigma\in\mathcal{E}} \tfrac{|D_\sigma|}{\delta t} \rho^*_\sigma \left[u_\sigma - u^*_\sigma \right]^2$$

Proposition II.4.2

Let $\gamma \geq 1$, $(u^n)_{0\leqslant n\leqslant N}$ and $(\rho^n)_{0\leqslant n\leqslant N}$ be a solution to the scheme. Let P be an elastic potential such that : $P^{'}(z) = z^{\gamma-2}$ ($i.e$ $P(z) = \frac{1}{\gamma-1} z^{\gamma-1}$ if $\gamma > 1$ and $P(z) = \log(z)$ if $\gamma = 1$. Then, for $\gamma > 1$, we have :

$$\frac{1}{\gamma-1} \sum_{K\in\mathcal{M}} |K| \, p^n_K + \frac{1}{2} \sum_{\sigma\in\mathcal{E}} |D_\sigma| \, \rho^n_\sigma \, |u^n_\sigma|^2 + \frac{1}{2} \sum_n \sum_{\sigma\in\mathcal{E}} |D_\sigma| \rho^n_\sigma \left[u^{n+1}_\sigma - u^n_\sigma \right]^2$$

$$+ \frac{\gamma}{2} \sum_n \delta t \left[\sum_{\sigma\in\mathcal{E}} |\sigma| \, \rho^{n+1}_{\sigma,\gamma} \, |u^{n+1}_\sigma| \, |\rho^{n+1}_K - \rho^{n+1}_L|^2 \right] \leq \frac{1}{2} \sum_{\sigma\in\mathcal{E}} |D_\sigma| \, \rho^0_\sigma \, |u^0_\sigma|^2 + \sum_{K\in\mathcal{M}} |K| \, \rho^0_K P(\rho^0_K),$$

$$\frac{1}{2} \sum_{\sigma\in\mathcal{E}} |D_\sigma| \, \rho^n_\sigma \, |u^n_\sigma|^2 + \frac{1}{2} \sum_n \sum_{\sigma\in\mathcal{E}} |D_\sigma| \rho^n_\sigma \left[u^{n+1}_\sigma - u^n_\sigma \right]^2$$

$$+ \frac{\gamma}{2} \sum_n \delta t \left[\sum_{\sigma\in\mathcal{E}} |\sigma| \, \rho^{n+1}_{\sigma,\gamma} \, |u^{n+1}_\sigma| \, |\rho^{n+1}_K - \rho^{n+1}_L|^2 \right] \leq \frac{1}{2} \sum_{\sigma\in\mathcal{E}} |D_\sigma| \, \rho^0_\sigma \, |u^0_\sigma|^2 + \sum_{K\in\mathcal{M}} |K| \, \rho^0_K P(\rho^0_K) + \frac{|\Omega|}{e},$$

where $\rho^{n+1}_{\sigma,\gamma} = \ \min \ \left((\rho^{n+1}_K)^{\gamma-2}, (\rho^{n+1}_L)^{\gamma-2} \right)$.

Proof For any $\gamma \geq 1$, let $\phi(z) = zP(z)$ be a continuously twice-differentiable function from $(0,\infty)$ to $\mathbb{R}$. there exists a real number $\overline{\rho}^{n+1}_{\sigma,\gamma} \in [\min(\rho^{n+1}_K, \rho^{n+1}_L), \max(\rho^{n+1}_K, \rho^{n+1}_L)]$ such that :

$$\sum_{K\in\mathcal{M}} \frac{|K|}{\delta t} \left[\rho^{n+1}_K P(\rho^{n+1}_K) - \rho^n_K P(\rho^n_K) \right] + \sum_{K\in\mathcal{M}} p^{n+1}_K \sum_{\sigma\in\mathcal{E}(K)} |\sigma| u^{n+1}_\sigma \cdot n_{K,\sigma}$$
$$+ \frac{1}{2} \sum_{\sigma=K|L} |\sigma| \phi^{''}(\overline{\rho}^{n+1}_{\sigma,\gamma}) |u^{n+1}_\sigma| |\rho^{n+1}_K - \rho^{n+1}|^2 \leq 0. \quad \text{(II.14)}$$

Multiplying the momentum balance equation of the scheme by the corresponding unknown of the the velocity on the corresponding edge and summing over the edges, we obtain :

$$\sum_{\sigma\in\mathcal{E}} u^{n+1}_\sigma \left[\frac{|D_\sigma|}{\delta t} \, (\rho^{n+1}_\sigma \, u^{n+1}_\sigma - \rho^n_\sigma \, u^n_\sigma) + \sum_{\varepsilon\in\overline{\mathcal{E}}(D_\sigma)} F^{n+1}_{\sigma,\varepsilon} \, u^{n+1}_\varepsilon \right] - \sum_{K\in\mathcal{M}} p^{n+1}_K \sum_{\sigma\in\mathcal{E}(K)} |\sigma| u^{n+1}_\sigma \cdot n_{K,\sigma} = 0.$$

By Lemma II.4.1 we get :

$$\sum_{K\in\mathcal{M}} p^{n+1}_K \sum_{\sigma\in\mathcal{E}(K)} |\sigma| u^{n+1}_\sigma \cdot n_{K,\sigma} \geq \frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t} \, \left[\rho^{n+1}_\sigma \, |u^{n+1}_\sigma|^2 - \rho^n_\sigma \, |u^n_\sigma|^2 \right]$$
$$+ \frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t} \rho^n_\sigma \left[u^{n+1}_\sigma - u^n_\sigma \right]^2 \quad \text{(II.15)}$$

Thanks to (II.14) and (II.15) we obtain :

$$\sum_{K\in\mathcal{M}} |K|\, \rho_K^n P(\rho_K^n) + \frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^n\, |u_\sigma^n|^2 + \frac{1}{2}\sum_n \sum_{\sigma\in\mathcal{E}} |D_\sigma| \rho_\sigma^n \left[u_\sigma^{n+1} - u_\sigma^n\right]^2$$

$$+\frac{\gamma}{2}\sum_n \delta t \left[\sum_{\sigma\in\mathcal{E}} |\sigma|\, (\overline{\rho}_{\sigma,\gamma}^{n+1})^{\gamma-2}\, |u_\sigma^{n+1}|\, |\rho_K^{n+1} - \rho_L^{n+1}|^2\right] \le \frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^0\, |u_\sigma^0|^2 + \sum_{K\in\mathcal{M}} |K|\, \rho_K^0 P(\rho_K^0)$$

Let $\rho_{\sigma,\gamma}^{n+1} = \ \min\ \big((\rho_K^{n+1})^{\gamma-2}, (\rho_L^{n+1})^{\gamma-2}\big)$, thus $\rho_{\sigma,\gamma}^{n+1} \le (\overline{\rho}_{\sigma,\gamma}^{n+1})^{\gamma-2}$, and for $\gamma > 1$ we obtain :

$$\frac{1}{\gamma-1}\sum_{K\in\mathcal{M}} |K|\, p_K^n + \frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^n\, |u_\sigma^n|^2 + \frac{1}{2}\sum_n \sum_{\sigma\in\mathcal{E}} |D_\sigma| \rho_\sigma^n \left[u_\sigma^{n+1} - u_\sigma^n\right]^2$$

$$+\frac{\gamma}{2}\sum_n \delta t \left[\sum_{\sigma\in\mathcal{E}} |\sigma|\, \rho_{\sigma,\gamma}^{n+1}\, |u_\sigma^{n+1}|\, |\rho_K^{n+1} - \rho_L^{n+1}|^2\right] \le \frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^0\, |u_\sigma^0|^2 + \sum_{K\in\mathcal{M}} |K|\, \rho_K^0 P(\rho_K^0).$$

For $\gamma = 1$, $zP(z) = z\log(z) \geqslant -\frac{1}{e}$ thus :

$$\frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^n\, |u_\sigma^n|^2 + \frac{1}{2}\sum_n \sum_{\sigma\in\mathcal{E}} |D_\sigma| \rho_\sigma^n \left[u_\sigma^{n+1} - u_\sigma^n\right]^2$$

$$+\frac{\gamma}{2}\sum_n \delta t \left[\sum_{\sigma\in\mathcal{E}} |\sigma|\, \rho_{\sigma,\gamma}^{n+1}\, |u_\sigma^{n+1}|\, |\rho_K^{n+1} - \rho_L^{n+1}|^2\right]$$

$$\le \frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^0\, |u_\sigma^0|^2 + \sum_{K\in\mathcal{M}} |K|\, \rho_K^0 P(\rho_K^0) + \frac{|\Omega|}{e}.$$

□

II.4.3 Passing to the limit in the scheme

The objective of this section is to show, in the one dimensional case, that, if a sequence of solutions is controlled in suitable norms and converges to a limit, this latter necessarily satisfies a (part of the) weak formulation of the continuous problem.

We suppose given a sequence of meshes and time steps $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$, such that the time step and the size $h^{(m)}$ of the mesh $\mathcal{M}^{(m)}$, defined by :

$$h^{(m)} = \ \sup_{K\in\mathcal{M}^{(m)}} \operatorname{diam}(K),$$

tend to zero as $m \to \infty$.

Let $\rho^{(m)}$, $p^{(m)}$ and $u^{(m)}$ be the solution given by the scheme (II.7) with the mesh $\mathcal{M}^{(m)}$ and the time step $\delta t^{(m)}$, or, more precisely speaking, a 1D version of the scheme which may be obtained by taking the MAC variant, only one horizontal stripe of meshes, supposing that the vertical component of the velocity (the degree of freedom of which are located on the top and bottom boundaries) vanishes, and that the measure of the faces is equal to 1. To the discrete unknowns, we associate piecewise constant functions on time intervals and on primal or dual meshes, so the density $\rho^{(m)}$, the pressure $p^{(m)}$ and the velocity $u^{(m)}$ are defined almost everywhere on $\Omega \times (0, T)$ by :

$$\rho^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{M}} (p^{(m)})_K^n\ \mathcal{X}_K\ \mathcal{X}_{(n,n+1)}, \quad lp^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{M}} (\rho^{(m)})_K^n\ \mathcal{X}_K\ \mathcal{X}_{(n,n+1)},$$

$$u^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{\sigma\in\mathcal{E}} (u^{(m)})_\sigma^n \ \mathcal{X}_{D_\sigma} \ \mathcal{X}_{(n,n+1)},$$

where $\mathcal{X}_K$, $\mathcal{X}_{D_\sigma}$ and $\mathcal{X}_{(n,n+1)}$ stand for the characteristic function of K, D_σ and the interval $(n, n+1)$ respectively.

We suppose that the sequence $\big(\rho^{(m)}, p^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ converges.

A weak solution to the continuous problem satisfies, for any $\varphi \in \mathrm{C}_c^\infty\big((0,T)\times\Omega\big)$:

$$-\int_{\Omega\times(0,T)} \Big[\rho\,\partial_t\varphi + \rho\,u\,\partial_x\varphi\Big]\,\mathrm{d}\boldsymbol{x} = 0, \tag{II.16a}$$

$$-\int_{\Omega\times(0,T)} \Big[\rho\,u\,\partial_t\varphi + (\rho\,u^2+p)\,\partial_x\varphi\Big]\,\mathrm{d}\boldsymbol{x} = 0, \tag{II.16b}$$

$$p = \rho^\gamma. \tag{II.16c}$$

Note that these relations are not sufficient to define a weak solution to the problem, since they do not imply anything about the initial and boundary conditions. However, they allow to derive the Rankine-Hugoniot conditions ; so, if we show that they are satisfied by the limit of a sequence of solutions to the discrete problem, this implies, loosely speaking, that *the scheme computes the right shock velocities*, which is the result we are searching for. It is stated in the following theorem.

Theorem II.4.3

Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence converges in $\mathrm{L}^p\big((0,T)\times\Omega\big)^4$, for $1 \le p < \infty$, to $(\bar\rho,\, \bar p,\, \bar u) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^4$.

Then the limit $(\bar\rho,\, \bar p,\, \bar u)$ satisfies the system (II.16).

Proof Let $(\mathcal{M}_m, \delta t_m)_{m\in\mathbb{N}}$ be a sequence of meshes and time step. Let $\varphi \in C_c^\infty(\Omega\times[0,T))$, we define φ_K^n and φ_σ^n by :

$$\varphi_K^n = \varphi(x_K, t^n) \qquad \text{and} \qquad \varphi_\sigma^n = \varphi(x_\sigma, t^n).$$

with those notations we define these discrete functions $\varphi_{\mathcal{M}}^m$, $\varphi_{\mathcal{E}}^m$, $\eth_t\varphi_{\mathcal{M}}^m$, $\eth_t\varphi_{\mathcal{E}}^m$, $\eth_x\varphi_{\mathcal{M}}^m$, $\eth_x\varphi_{\mathcal{E}}^m$ by :

$$\varphi_{\mathcal{M}}^m = \sum_{n=0}^{N-1}\left(\sum_{K\in\mathcal{M}} \varphi_K^n 1_K\right) 1_{[t^n,t^{n+1}]}, \qquad \varphi_{\mathcal{E}}^m = \sum_{n=0}^{N-1}\left(\sum_{\sigma\in\mathcal{E}} \varphi_\sigma^n 1_{D_\sigma}\right) 1_{[t^n,t^{n+1}]}.$$

$$\eth_t\varphi_{\mathcal{M}}^m = \frac{\varphi_K^{n+1}-\varphi_K^n}{\delta t} \ \text{ over each } \ K\times[t^n,t^{n+1}], \qquad \eth_t\varphi_{\mathcal{E}}^m = \frac{\varphi_\sigma^{n+1}-\varphi_\sigma^n}{\delta t} \ \text{ over each } \ D_\sigma\times[t^n,t^{n+1}]$$

$$\eth_x\varphi_{\mathcal{M}}^m = \frac{\varphi_L^{n+1}-\varphi_K^{n+1}}{|d_\sigma|} \ \text{ over each } \ D_\sigma\times[t^n,t^{n+1}], \qquad \eth_x\varphi_{\mathcal{E}}^m = \frac{\varphi_\sigma^{n+1}-\varphi_{\sigma'}^{n+1}}{|d_{\sigma,\sigma'}|} \ \text{ over each } \ K\times[t^n,t^{n+1}].$$

We multiply the first equation of the scheme by $\delta t\,\varphi_K^n$, and sum the result on $n \in \{0, ..., N-1\}$ and $K \in \mathcal{M}$, to obtain :

$$\underbrace{\sum_{n=0}^{N}\sum_{K\in\mathcal{M}} |K|(\rho_K^{n+1}-\rho_K^n)\varphi_K^n}_{T_1^m} + \underbrace{\sum_{n=0}^{N}\delta t\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}^{n+1}\varphi_K^n}_{T_2^m} = 0.$$

We begin by study T_1^m :

$$T_1^m = -\sum_{n=1}^{N}\delta t \sum_{K\in\mathcal{M}} |K|\,\rho_K^{n+1}\,\frac{\varphi_K^{n+1}-\varphi_K^n}{\delta t} - \sum_{K\in\mathcal{M}} |K|\,(\rho^0)_K\,(\varphi^0)_K$$

$$\begin{aligned} T_1^m &= -\sum_{n=1}^{N}\sum_{K\in\mathcal{M}}\int_{t^n}^{t^{n+1}}\int_K \rho_m \eth_t\varphi_{\mathcal{M}}^m - \sum_{K\in\mathcal{M}}\int_K (\rho^0)_m\,(\varphi^0)_{\mathcal{M}}^m = -\int_0^T\int_\Omega \rho_m\eth_t\varphi_{\mathcal{M}}^m - \int_\Omega (\rho^0)_m\,(\varphi^0)_{\mathcal{M}}^m \\ &= -\int_0^T\int_\Omega \rho_m\partial_t\varphi - \int_\Omega (\rho^0)_m\varphi^0 + \underbrace{\int_0^T\int_\Omega \rho_m(\partial_t\varphi - \eth_t\varphi_{\mathcal{M}}^m) + \int_\Omega (\rho^0)_m(\varphi_0 - (\varphi^0)_{\mathcal{M}}^m)}_{R_m} \end{aligned}$$

thanks to the fact that $\rho_m \in L^1(\Omega\times[0,T])$ and $\varphi \in C_c^\infty(Q)$, we obtain :

$$|R_m| \leq \parallel \rho_m \parallel_{L^1(\Omega\times[0,T])}\parallel \partial_t\varphi - \eth_t\varphi_{\mathcal{M}}^m \parallel_{L^\infty(\Omega\times[0,T])} + \parallel (\rho^0)_m \parallel_{L^1(\Omega\times[0,T])}\parallel \varphi_0 - (\varphi^0)_{\mathcal{M}}^m \parallel_{L^\infty(\Omega\times[0,T])},$$

passing to the limit in T_1^m we obtain

$$T_1^m \longrightarrow -\int_0^T\int_\Omega \rho\partial_t\varphi - \int_\Omega \rho_0\,\varphi_0 \text{ as } m\to\infty.$$

We now study T_2^m

$$T_2^m = \sum_{n=0}^{N}\delta t\sum_{\sigma=K|L} |D_\sigma|\widetilde{\rho}_\sigma^{n+1}\; u_\sigma^{n+1}\cdot n_{K,\sigma}\frac{(\varphi_K^n-\varphi_L^n)}{|d_\sigma|}$$

$\widetilde{\rho}_\sigma^{n+1}$ calculated by the upwind thechnique, we can to choose the orientation of $n_{K,\sigma}$ such that $\widetilde{\rho}_\sigma^{n+1} = \rho_K$ on the other hand we have $|D_\sigma| = \frac{|K|}{2} + \frac{|L|}{2}$, so that :

$$\begin{aligned} T_2^m \quad &= \sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}\left(\frac{|K|}{2}\rho_K^{n+1} + \frac{|L|}{2}\rho_L^{n+1}\right)\; u_\sigma^{n+1}\cdot n_{K,\sigma}\frac{(\varphi_K^n-\varphi_L^n)}{|d_\sigma|} \\ &+\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}\frac{|L|}{2}\left(\rho_K^{n+1}-\rho_L^{n+1}\right)\; u_\sigma^{n+1}\cdot n_{K,\sigma}\frac{(\varphi_K^n-\varphi_L^n)}{|d_\sigma|} \end{aligned}$$

Therefore

$$\begin{aligned} T_2^m = \sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}|D_\sigma|\rho_\sigma^{n+1}\; u_\sigma^{n+1}\cdot n_{K,\sigma}\frac{(\varphi_K^n-\varphi_L^n)}{|d_\sigma|} \\ +\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}\frac{|L|}{2}\left(\rho_K^{n+1}-\rho_L^{n+1}\right)\; u_\sigma^{n+1}\cdot n_{K,\sigma}\frac{(\varphi_K^n-\varphi_L^n)}{|d_\sigma|}, \end{aligned}$$

which may be written :

$$T_2^m = \underbrace{-\sum_{n=0}^{N}\sum_{\sigma}\int_{t^n}^{t^{n+1}}\int_{D_\sigma}\rho_m\,u_m\eth_x\varphi_{\mathcal{E}}^m}_{T_{2,1}^m} + \underbrace{\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}\frac{|L|}{2}(\rho_K^{n+1}-\rho_L^{n+1})\; u_\sigma^{n+1}\cdot n_{K,\sigma}\frac{(\varphi_K^n-\varphi_L^n)}{|d_\sigma|}}_{R_m}.$$

Let us first study $T_{2,1}^m$:

$$T_{2,1}^m = -\int_0^T\int_\Omega \rho_m u_m\partial_x\varphi + \int_0^T\int_\Omega \rho_m u_m(\partial_x\varphi - \eth_x\varphi_{\mathcal{E}}^m).$$

Using the fact that $\rho_m u_m \in L^1(\Omega \times [0,T])$, and passing to the limit in $T^m_{2,1}$, we obtain :

$$T^m_{2,1} \longrightarrow -\int_0^T \int_\Omega \rho u \partial_x \varphi \text{ as } m \to \infty.$$

Let us now study R_m.

$$\begin{aligned}
\mid R_m \mid \quad & \leq C_\varphi \sum_{n=0}^{N} \delta t \sum_{\sigma = K|L} |D_\sigma| \mid \rho_K^{n+1} - \rho_L^{n+1} \mid \mid u_\sigma^{n+1} \mid \\
& \leq C_\varphi \sqrt{h} \sum_{n=0}^{N} \delta t \left[\left(\sum_{\sigma=K|L} |\sigma| \left|u_\sigma^{n+1}\right| \left|\rho_K^{n+1} - \rho_L^{n+1}\right|^2 \right)^{1/2} \left(\sum_{\sigma=K|L} |D_\sigma| \left|u_\sigma^{n+1}\right| \right)^{1/2} \right] \\
& \leq C_\varphi \sqrt{h} \parallel u_m \parallel^{\frac{1}{2}}_{L^1([0,T),L^1(\Omega))} \left[\sum_{n=0}^{N} \delta t \sum_{\sigma=K|L} |\sigma| \left|u_\sigma^{n+1}\right| \left|\rho_K^{n+1} - \rho_L^{n+1}\right|^2 \right]^{\frac{1}{2}} \longrightarrow 0 \text{ as } m \to \infty.
\end{aligned}$$

Therefore, we obtain :

$$T_2^m \longrightarrow -\int_0^T \int_\Omega \rho u \partial_x \varphi \text{ as } m \to \infty.$$

Now multiply Equation (II.7b) of the scheme by $\delta t\, \varphi^n_\sigma$, and sum the result on $n \in \{0, ..., N-1\}$ and $\sigma = K|L$; we obtain :

$$\underbrace{\sum_{n=0}^{N-1} \delta t \sum_\sigma |D_\sigma| (\rho_\sigma^{n+1} u_\sigma^{n+1} - \rho_\sigma^n u_\sigma^n) \varphi_\sigma^n}_{T_1^m} + \underbrace{\sum_{n=0}^{N-1} \delta t \sum_\sigma \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma\varepsilon}^{n+1} u_\varepsilon^{n+1} \varphi_\sigma^n}_{T_2^m}$$
$$+ \underbrace{\sum_{n=0}^{N-1} \delta t \sum_\sigma |D_\sigma| \frac{p_L^{n+1} - p_K^{n+1}}{d_{\sigma\sigma'}} \varphi_\sigma^n}_{T_3^m} = 0.$$

Let us first study T_1^m.

$$\begin{aligned}
T_1^m \quad & = -\sum_{n=0}^{N} \delta t \sum_\sigma |D_\sigma| \rho_\sigma^{n+1} u_\sigma^{n+1} \frac{\varphi_\sigma^{n+1} - \varphi_\sigma^n}{\delta t} - \sum_\sigma |D_\sigma| (\rho^0)_\sigma (u^0)_\sigma (\varphi^0)_\sigma \\
& = -\sum_{n=1}^{N} \sum_\sigma \int_{t^n}^{t^{n+1}} \int_{D_\sigma} \rho_m u_m \eth_t \varphi^m_{\mathcal{E}} - \sum_\sigma \int_{D_\sigma} (\rho^0)_m (u^0)_m \, (\varphi^0)^m_{\mathcal{E}} \\
& = -\textstyle\int_0^T \int_\Omega \rho_m u_m \eth_t \varphi^m_{\mathcal{E}} - \int_\Omega (\rho^0)_m (u^0)_m \, (\varphi^0)^m_{\mathcal{E}}.
\end{aligned}$$

Threfore

$$T_1^m = -\int_0^T \int_\Omega \rho_m u_m \partial_t \varphi - \int_\Omega (\rho^0)_m (u^0)_m \varphi^0$$
$$+ \underbrace{\int_0^T \int_\Omega \rho_m u_m (\partial_t \varphi - \eth_t \varphi^m_{\mathcal{E}}) + \int_\Omega (\rho^0)_m (u^0)_m (\varphi_0 - (\varphi^0)^m_{\mathcal{E}})}_{R_m}.$$

We have $\rho_m u_m \in L^1(I \times [0,T])$, and therefore, $R_m \longrightarrow 0$ as $m \to \infty$ and :

$$T_1^m \longrightarrow -\int_0^T \int_\Omega \rho u \partial_t \varphi - \int_\Omega \rho_0 u_0 \varphi_0 \text{ as } m \to \infty$$

Let us now study T_3^m.

$$\begin{aligned} T_3^m = -\sum_{n=0}^{N} \delta t \sum_K |K| p_K^{n+1} \frac{\varphi_{\sigma'}^n - \varphi_\sigma^n}{|K|} \quad &= -\sum_{n=1}^{N} \sum_K \int_{t^n}^{t^{n+1}} \int_K p_m \eth_x \varphi_{\mathcal{M}}^m = -\int_0^T \int_\Omega p_m \eth_x \varphi_{\mathcal{M}}^m \\ &= -\int_0^T \int_\Omega p_m \partial_t \varphi + \underbrace{\int_0^T \int_\Omega p_m (\partial_x \varphi - \eth_x \varphi_{\mathcal{M}}^m)}_{R_m}. \end{aligned}$$

Thanks to the fact that $p_m \in L^1(\Omega \times [0,T])$, $R_m \longrightarrow 0$ as $m \to \infty$, and thus :

$$T_3^m \longrightarrow -\int_0^T \int_\Omega p \partial_x \varphi \text{ as } m \to \infty.$$

Finally, we study we reorder the sum in T_2^m. We have

$$F_{\sigma,\varepsilon}^{n+1} = \frac{F_{\sigma',K}^{n+1} - F_{K,\sigma}^{n+1}}{2} \text{ and } u_\varepsilon^{n+1} = \frac{u_\sigma^{n+1} + u_{\sigma'}^{n+1}}{2}.$$

Therefore,

$$\begin{aligned} T_2^m \quad &= -\frac{1}{4} \sum_{n=0}^{N} \delta t \sum_K (F_{K,\sigma}^{n+1} - F_{\sigma',K}^{n+1})(u_\sigma^{n+1} + u_{\sigma'}^{n+1})(\varphi_\sigma^n - \varphi_{\sigma'}^n) \\ &= -\frac{1}{4} \sum_{n=0}^{N} \delta t \sum_K (|\sigma|\, \widetilde{\rho}_\sigma^{n+1} u_\sigma^{n+1} + |\sigma^{'}|\, \widetilde{\rho}_{\sigma'}^{n+1} u_{\sigma'}^{n+1})(u_\sigma^{n+1} + u_{\sigma'}^{n+1})(\varphi_\sigma^n - \varphi_{\sigma'}^n) \\ &= \underbrace{-\sum_{n=0}^{N} \delta t \sum_K |K| \rho_K^{n+1} \frac{(u_\sigma^{n+1})^2 + (u_{\sigma'}^{n+1})^2}{2} \frac{(\varphi_\sigma^n - \varphi_{\sigma'}^n)}{d_{\sigma\sigma'}}}_{T_{2,1}^m} + R_m, \end{aligned}$$

where $\sigma = K|L$, $\sigma^{'} = M|K$ and R_m reads :

$$\begin{aligned} R_m = -\frac{1}{4} \sum_{n=0}^{N} \delta t \sum_K |\sigma| \Big[&(\widetilde{\rho}_\sigma^{n+1} u_\sigma^{n+1} + \widetilde{\rho}_{\sigma'}^{n+1} u_{\sigma'}^{n+1})(u_\sigma^{n+1} + u_{\sigma'}^{n+1}) \\ &-2\rho_K^{n+1} \Big((u_\sigma^{n+1})^2 + (u_{\sigma'}^{n+1})^2 \Big) \Big] (\varphi_\sigma^n - \varphi_{\sigma'}^n). \end{aligned}$$

Let us study $T_{2,1}^m$,

$$\begin{aligned} T_{2,1}^m = -\sum_{n=1}^{N} \sum_K \int_{t^n}^{t^{n+1}} \int_K \rho_m u_m^2 \eth_x \varphi_{\mathcal{M}}^m &= -\int_0^T \int_\Omega \rho_m u_m^2 \eth_x \varphi_{\mathcal{M}}^m \\ &= -\int_0^T \int_\Omega \rho_m u_m^2 \partial_x \varphi + \underbrace{\int_0^T \int_\Omega \rho_m u_m^2 (\partial_x \varphi - \eth_x \varphi_{\mathcal{M}}^m)}_{R_m^{'}}. \end{aligned}$$

We have $\rho_m u_m^2 \in L^1(\Omega \times [0,T])$, so $R_m^{'} \longrightarrow 0$ as $m \to \infty$ thus :

$$T_{2,1}^m \longrightarrow -\int_0^T \int_\Omega \rho u^2 \partial_x \varphi \text{ as } m \to \infty$$

We now study R_m. Expanding the quantity $2\rho_K^{n+1}\Big((u_\sigma^{n+1})^2 + (u_{\sigma'}^{n+1})^2\Big)$ by the fact that $2(a^2+b^2) = (a+b)^2 + (a-b)^2$, thus the term R_m reads :

$$\begin{aligned} R_m = \quad & -\frac{1}{4}\underbrace{\sum_{n=0}^{N} \delta t \sum_K |\sigma| \Big[\Big((\widetilde{\rho}_\sigma^{n+1} - \rho_K^{n+1}) u_\sigma^{n+1} + (\widetilde{\rho}_{\sigma'}^{n+1} - \rho_K^{n+1}) u_{\sigma'}^{n+1}\Big)(u_\sigma^{n+1} + u_{\sigma'}^{n+1})\Big] (\varphi_\sigma^n - \varphi_{\sigma'}^n)}_{R_{m_1}} \\ & + \frac{1}{4}\underbrace{\sum_{n=0}^{N} \delta t \sum_K |\sigma|\, \rho_K^{n+1} \,(u_\sigma^{n+1} - u_{\sigma'}^{n+1})^2\,(\varphi_\sigma^n - \varphi_{\sigma'}^n)}_{R_{m_2}} \end{aligned}$$

First we study R_{m_1} :

$$\begin{aligned} R_{m_1} = \sum_{n=0}^{N} \delta t \sum_K |\sigma| \,\big[\big((\rho_L^{n+1} - \rho_K^{n+1}) \min(u_\sigma^{n+1}, 0) \\ + (\rho_M^{n+1} - \rho_K^{n+1}) \max(u_{\sigma'}^{n+1}, 0)\Big)\,(u_\sigma^{n+1} + u_{\sigma'}^{n+1})\Big]\,(\varphi_\sigma^n - \varphi_{\sigma'}^n) \end{aligned}$$

Therefore :

$$\begin{aligned} |R_{m_1}| \quad & \leqslant C_\varphi \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L} |\rho_K^{n+1} - \rho_L^{n+1}||u_\sigma^{n+1}| \Big(|K||u_\sigma^{n+1}| + |K||u_{\sigma'}^{n+1}| + |L||u_\sigma^{n+1}| + |L||u_{\sigma''}^{n+1}|\Big) \\ & \leqslant C_\varphi \sum_{n=0}^{N} \delta t \sum_{\sigma=K|L} |D_\sigma||\rho_K^{n+1} - \rho_L^{n+1}||u_\sigma^{n+1}| \Big(|u_\sigma^{n+1}| + |u_{\sigma'}^{n+1}| + |u_{\sigma''}^{n+1}|\Big). \end{aligned}$$

Therefore, by the Cauchy-Scharz inequality, we get that :

$$|R_{m_1}| \leqslant \sqrt{h}\, C_\varphi \;\|\; u_m \;\|_{L^3([0,T],L^3(\Omega))}^{\frac{3}{2}} \left[\sum_{n=0}^{N} \delta t \sum_{\sigma=K|L} |\sigma||u_\sigma^{n+1}|\,\big|\rho_K^{n+1} - \rho_L^{n+1}\big|^2\right]^{\frac{1}{2}}.$$

Now it is easy to see that

$$\begin{aligned} |R_{m_2}| \quad & \leqslant h^2\, C_\varphi \left[\sum_{n=0}^{N} \delta t \sum_K |K| \rho_K^{n+1} \left(\frac{u_\sigma^{n+1} - u_{\sigma'}^{n+1}}{d_{\sigma,\sigma'}}\right)^2\right] \\ & \leqslant h^2\, C_\varphi \;\|\; \rho_m \;\|_{L^\infty([0,T]\times\Omega))} \|\; u_m \;\|_{L^2([0,T],H^1(\Omega))}^2 \\ & \leqslant h^{2-\frac{d}{p}}\, C_\varphi \;\|\; \rho_m \;\|_{L^p([0,T]\times\Omega))} \|\; u_m \;\|_{L^2([0,T],H^1(\Omega))}^2 . \end{aligned}$$

Finally, we conclude that :

$$R_m \longrightarrow 0 \text{ as } m \to \infty.$$

□

Theorem II.4.4
Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence converges in $\mathrm{L}^p\big((0,T)\times\Omega\big)^4$, for $1 \le p < \infty$, to $(\bar{\rho},\, \bar{p},\, \bar{u}) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^4$.

Then the limit $(\bar{\rho},\, \bar{p},\, \bar{u})$ is an entropy solution of the continous problem (II.1) in the following weak sense :

$$-\int_0^T \int_\Omega S\, \partial_t\varphi - \int_0^T \int_\Omega (S+u)\, p\, \partial_x\varphi - \int_\Omega S_0\, \varphi_0 \le 0$$

where S is the entropy of the system (II.1), is given by :

$$S = \frac{1}{2}\rho u^2 + \rho P(\rho),\ P(.) \text{ is the "elastic potentiel" wich satisfies } P^{'}(z) = \frac{\rho^\gamma}{\rho^2}.$$

Proof we multiply the discret equation (II.11) by $\delta t \varphi_\sigma^n$ and summing over the edges σ and n, in other hand we multiply the discret equation (II.12) by $\delta t \varphi_K^n$, and summing over the primal mesh K and n. Summing both results and we tend m to infinity we obtain :

$$-\int_0^T \int_\Omega S\, \partial_t\varphi - \int_0^T \int_\Omega (S+u)\, p\, \partial_x\varphi - \int_\Omega S_0\, \varphi_0 \le 0$$

where S is the entropy of the system (II.1), is given by :

$$S = \frac{1}{2}\rho u^2 + \rho P(\rho).$$

□

II.5 Pressure correction scheme

II.5.1 The scheme

We derive in this section a pressure correction numerical scheme from the implicit scheme (II.7). The first step, is a renormalizationof the pressure the interset of which is clarified only by the stability analysis. The next step, as usual, is to compute a tentative velocity by solving the momentum balance equation with the begining-of-step pressure. Then, the velocity is corrected and the other variables are advanced in time, here, which is less standard, by a single coupled step ; this is motivated by stability reasons detailed in [20]. Still for stability reasons, or, in other words, to be able to derive a kinetic energy balance, we need that the mass balance over the dual cells (II.9) to hold ; since the mass balance is not yet solved when performing the prediction step, this leads us to perform a time shift of the density at this step.

The algorithm reads :

Renormalization step – Solve for $\tilde{\boldsymbol{u}}^{n+1}$:

$$\text{For } 1 \le i \le d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\sum_{\sigma=K|L} \frac{1}{\rho_\sigma^n} \frac{|\sigma|^2}{|D_\sigma|} \left(\tilde{p}_K^{n+1} - \tilde{p}_L^{n+1} \right) = \sum_{\sigma=K|L} \frac{1}{\sqrt{\rho_\sigma^n\, \rho_\sigma^{n-1}}} \frac{|\sigma|^2}{|D_\sigma|} \left(p_K^n - p_L^n \right), \tag{II.17a}$$

Prediction step – Solve for $\tilde{p}^{n+1}$:

$$\text{For } 1 \le i \le d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\frac{|D_\sigma|}{\delta t} (\rho_\sigma^n \tilde{\boldsymbol{u}}_{\sigma,i}^{n+1} - \rho_\sigma^{n-1} \boldsymbol{u}_{\sigma,i}^n) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^n \tilde{\boldsymbol{u}}_{\varepsilon,i}^{n+1} + |D_\sigma| \, (\boldsymbol{\nabla} p^n)_{\sigma,i} = 0, \tag{II.17b}$$

Correction step – Solve for ρ^{n+1}, p^{n+1}, e^{n+1} and $\boldsymbol{u}^{n+1}$:

$$\text{For } 1 \le i \le d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\frac{|D_\sigma|}{\delta t} \; \rho_\sigma^n \; (\boldsymbol{u}_{\sigma,i}^{n+1} - \tilde{\boldsymbol{u}}_{\sigma,i}^{n+1}) + |D_\sigma| \left[(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} - (\boldsymbol{\nabla} p^n)_{\sigma,i} \right] = 0, \tag{II.17c}$$

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t} (\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} = 0, \tag{II.17d}$$

$$\forall K \in \mathcal{M}, \qquad p_K^{n+1} = (\rho_K^{n+1})^\gamma. \tag{II.17e}$$

II.5.2 Estimates

PROPOSITION II.5.1

Let $\gamma \geq 1$, $(u^n)_{0 \leqslant n \leqslant N}$ and $(\rho^n)_{0 \leqslant n \leqslant N}$ be a solution to the scheme. Let P be an elastic potential such that : $P^{'}(z) = z^{\gamma-2}(i.e.P(z) = \frac{1}{\gamma-1} z^{\gamma-1}$ if $\gamma > 1$ and $P(z) = \log(z)$ if $\gamma = 1)$. Then, for $\gamma > 1$:

$$\begin{aligned}
&\frac{1}{\gamma-1} \sum_{K \in \mathcal{M}} |K| \, p_K^n + \frac{1}{2} \sum_{\sigma \in \mathcal{E}} |D_\sigma| \, \rho_\sigma^n \, |u_\sigma^{n+1}|^2 + \frac{1}{2} \sum_n \sum_{\sigma \in \mathcal{E}} |D_\sigma| \rho_\sigma^n \left[\tilde{u}_\sigma^{n+1} - u_\sigma^n \right]^2 \\
&+ \frac{\gamma}{2} \sum_n \delta t \left[\sum_{\sigma \in \mathcal{E}} |\sigma| \, \rho_{\sigma,\gamma}^{n+1} \, |u_\sigma^{n+1}| \, |\rho_K^{n+1} - \rho_L^{n+1}|^2 \right] + \delta t^2 \sum_{\sigma \in \mathcal{E}} \frac{|D_\sigma|}{\rho_\sigma^n} \left(\frac{p_L^{n+1} - p_K^{n+1}}{d_\sigma} \right)^2 \\
&\leq \frac{1}{2} \sum_{\sigma \in \mathcal{E}} |D_\sigma| \, \rho_\sigma^0 \, |u_\sigma^0|^2 + \sum_{K \in \mathcal{M}} |K| \, \rho_K^0 P(\rho_K^0),
\end{aligned}$$

and for $\gamma = 1$:

$$\frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^n\, |u_\sigma^{n+1}|^2 + \frac{1}{2}\sum_{n} \sum_{\sigma\in\mathcal{E}} |D_\sigma| \rho_\sigma^n \left[\tilde{u}_\sigma^{n+1} - u_\sigma^n\right]^2$$
$$+\frac{\gamma}{2}\sum_{n} \delta t \left[\sum_{\sigma\in\mathcal{E}} |\sigma|\, \rho_{\sigma,\gamma}^{n+1}\, |u_\sigma^{n+1}|\, |\rho_K^{n+1} - \rho_L^{n+1}|^2\right]$$
$$+\delta t^2 \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\rho_\sigma^n} \left(\frac{p_L^{n+1} - p_K^{n+1}}{d_\sigma}\right)^2 \le \frac{1}{2}\sum_{\sigma\in\mathcal{E}} |D_\sigma|\, \rho_\sigma^0\, |u_\sigma^0|^2 + \sum_{K\in\mathcal{M}} |K|\, \rho_K^0 P(\rho_K^0) + \frac{|\Omega|}{e}.$$

where $\rho_{\sigma,\gamma}^{n+1} = \ \min\ \left((\rho_K^{n+1})^{\gamma-2}, (\rho_L^{n+1})^{\gamma-2}\right)$.

Proof For any $\gamma \geq 1$, let $\phi(z) = zP(z)$ be a continuously twice-differentiable function from $(0,\infty)$ to $\mathbb{R}$. there exists a real number $\overline{\rho}_{\sigma,\gamma}^{n+1} \in [\min(\rho_K^{n+1}, \rho_L^{n+1}), \max(\rho_K^{n+1}, \rho_L^{n+1})]$ such that :

$$\sum_{K\in\mathcal{M}} \frac{|K|}{\delta t}\Big[\rho_K^{n+1} P(\rho_K^{n+1}) - \rho_K^n P(\rho_K^n)\Big] + \sum_{K\in\mathcal{M}} p_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma| u_\sigma^{n+1} \cdot n_{K,\sigma}$$
$$+ \frac{1}{2} \sum_{\sigma=K|L} |\sigma| \phi^{''}(\overline{\rho}_{\sigma,\gamma}^{n+1}) |u_\sigma^{n+1}| |\rho_K^{n+1} - \rho^{n+1}|^2 \leq 0. \qquad \text{(II.18)}$$

Multiplying the momentum balance equation of the scheme by the corresponding unknown of the the velocity on the corresponding edge and summing over the edges, we obtain :

$$\sum_{\sigma\in\mathcal{E}} \tilde{u}_\sigma^{n+1} \left[\frac{|D_\sigma|}{\delta t}\,(\rho_\sigma^n\, \tilde{u}_\sigma^{n+1} - \rho_\sigma^{n-1}\, u_\sigma^n) + \sum_{\varepsilon\in\overline{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^n\, \tilde{u}_\varepsilon^{n+1}\right] - \sum_{K\in\mathcal{M}} \tilde{p}_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma| \tilde{u}_\sigma^{n+1} \cdot n_{K,\sigma} = 0$$

By Lemma II.4.1 we get :

$$\frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t}\ \left[\rho_\sigma^n\, |\tilde{u}_\sigma^{n+1}|^2 - \rho_\sigma^{n-1}\, |u_\sigma^n|^2\right] + \frac{1}{2}\sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t} \rho_\sigma^{n-1} \left[\tilde{u}_\sigma^{n+1} - u_\sigma^n\right]^2$$
$$- \sum_{K\in\mathcal{M}} \tilde{p}_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma| \tilde{u}_\sigma^{n+1} \cdot n_{K,\sigma} \leqslant 0 \qquad \text{(II.19)}$$

Squaring the pressure correction equation, multiplying by $\dfrac{\delta t}{2\,\rho_\sigma^n}$ and summing over the edges, we get :

$$\frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t}\ \rho_\sigma^n\, |u_\sigma^{n+1}|^2 - \sum_{K\in\mathcal{M}} p_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma| u_\sigma^{n+1} \cdot n_{K,\sigma} + \frac{1}{2}\ \sum_{\sigma\in\mathcal{E}} \delta t\, |D_\sigma| \frac{1}{\rho_\sigma^n} \left(\frac{p_L^{n+1} - p_K^{n+1}}{d_\sigma}\right)^2$$
$$= \frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t}\ \rho_\sigma^n\, |\tilde{u}_\sigma^{n+1}|^2 - \sum_{K\in\mathcal{M}} \tilde{p}_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma| \tilde{u}_\sigma^{n+1} \cdot n_{K,\sigma} + \frac{1}{2}\ \sum_{\sigma\in\mathcal{E}} \delta t\, |D_\sigma| \frac{1}{\rho_\sigma^n} \left(\frac{\tilde{p}_L^{n+1} - \tilde{p}_K^{n+1}}{d_\sigma}\right)^2 \qquad \text{(II.20)}$$

Using the equation (II.19) in (II.20), we get :

$$\frac{1}{2} \sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t}\ \left[\rho_\sigma^n\, |u_\sigma^{n+1}|^2 - \rho_\sigma^{n-1}\, |u_\sigma^n|^2\right] + \frac{1}{2}\sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t} \rho_\sigma^{n-1} \left[\tilde{u}_\sigma^{n+1} - u_\sigma^n\right]^2$$
$$+\frac{1}{2}\ \sum_{\sigma\in\mathcal{E}} \delta t\, |D_\sigma| \frac{1}{\rho_\sigma^n} \left[\left(\frac{p_L^{n+1} - p_K^{n+1}}{d_\sigma}\right)^2 - \left(\frac{\tilde{p}_L^{n+1} - \tilde{p}_K^{n+1}}{d_\sigma}\right)^2\right] - \sum_{K\in\mathcal{M}} p_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma| u_\sigma^{n+1} \cdot n_{K,\sigma} \leqslant 0$$

Using the next equation of the pressure correction scheme, we show that :

$$\sum_{\sigma\in\mathcal{E}} \delta t\,|D_\sigma| \frac{1}{\rho_\sigma^n}\left(\frac{\tilde{p}_L^{n+1}-\tilde{p}_K^{n+1}}{d_\sigma}\right)^2 \le \sum_{\sigma\in\mathcal{E}} \delta t\,|D_\sigma| \frac{1}{\rho_\sigma^{n-1}}\left(\frac{p_L^{n}-p_K^{n}}{d_\sigma}\right)^2$$

Then

$$\begin{aligned}
&\frac{1}{2}\sum_{\sigma\in\mathcal{E}} \frac{|D_\sigma|}{\delta t}\ \left[\rho_\sigma^n\,|u_\sigma^{n+1}|^2-\rho_\sigma^{n-1}\,|u_\sigma^n|^2\right]+\frac{1}{2}\sum_{\sigma\in\mathcal{E}}\frac{|D_\sigma|}{\delta t}\rho_\sigma^{n-1}\left[\tilde{u}_\sigma^{n+1}-u_\sigma^n\right]^2\\
&+\frac{1}{2}\sum_{\sigma\in\mathcal{E}}\delta t\,|D_\sigma|\left[\frac{1}{\rho_\sigma^n}\left(\frac{p_L^{n+1}-p_K^{n+1}}{d_\sigma}\right)^2-\frac{1}{\rho_\sigma^{n-1}}\left(\frac{p_L^{n}-p_K^{n}}{d_\sigma}\right)^2\right]\leqslant\sum_{K\in\mathcal{M}}p_K^{n+1}\sum_{\sigma\in\mathcal{E}(K)}|\sigma|u_\sigma^{n+1}\cdot n_{K,\sigma}
\end{aligned} \tag{II.21}$$

Thanks to (II.18) and (II.21) we obtain :

$$\begin{aligned}
&\sum_{K\in\mathcal{M}}|K|\,\rho_K^nP(\rho_K^n)+\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,\rho_\sigma^n\,|u_\sigma^{n+1}|^2+\frac{1}{2}\sum_n\sum_{\sigma\in\mathcal{E}}|D_\sigma|\rho_\sigma^{n-1}\left[\tilde{u}_\sigma^{n+1}-u_\sigma^n\right]^2\\
&+\frac{\gamma}{2}\sum_n\delta t\left[\sum_{\sigma\in\mathcal{E}}|\sigma|\,(\overline{\rho}_{\sigma,\gamma}^{n+1})^{\gamma-2}\,|u_\sigma^{n+1}|\,|\rho_K^{n+1}-\rho_L^{n+1}|^2\right]+\frac{1}{2}\,\delta t^2\sum_{\sigma\in\mathcal{E}}\frac{|D_\sigma|}{\rho_\sigma^n}\left(\frac{p_L^{n+1}-p_K^{n+1}}{d_\sigma}\right)^2\\
&\qquad\le\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,\rho_\sigma^0\,|u_\sigma^0|^2+\sum_{K\in\mathcal{M}}|K|\,\rho_K^0P(\rho_K^0)
\end{aligned}$$

Let $\rho_{\sigma,\gamma}^{n+1} =\ \min\ \big((\rho_K^{n+1})^{\gamma-2},(\rho_L^{n+1})^{\gamma-2}\big)$, thus $\rho_{\sigma,\gamma}^{n+1}\le(\overline{\rho}_{\sigma,\gamma}^{n+1})^{\gamma-2}$, for $\gamma>1$ we obtain :

$$\begin{aligned}
&\frac{1}{\gamma-1}\sum_{K\in\mathcal{M}}|K|\,p_K^n+\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,\rho_\sigma^n\,|u_\sigma^{n+1}|^2+\frac{1}{2}\sum_n\sum_{\sigma\in\mathcal{E}}|D_\sigma|\rho_\sigma^n\left[\tilde{u}_\sigma^{n+1}-u_\sigma^n\right]^2\\
&+\frac{\gamma}{2}\sum_n\delta t\left[\sum_{\sigma\in\mathcal{E}}|\sigma|\,\rho_{\sigma,\gamma}^{n+1}\,|u_\sigma^{n+1}|\,|\rho_K^{n+1}-\rho_L^{n+1}|^2\right]+\delta t^2\sum_{\sigma\in\mathcal{E}}\frac{|D_\sigma|}{\rho_\sigma^n}\left(\frac{p_L^{n+1}-p_K^{n+1}}{d_\sigma}\right)^2\\
&\le\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,\rho_\sigma^0\,|u_\sigma^0|^2+\sum_{K\in\mathcal{M}}|K|\,\rho_K^0P(\rho_K^0)
\end{aligned}$$

for $\gamma=1$, $zP(z)=z\log(z)\geqslant-\frac{1}{e}$ thus :

$$\begin{aligned}
&\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,\rho_\sigma^n\,|u_\sigma^{n+1}|^2+\frac{1}{2}\sum_n\sum_{\sigma\in\mathcal{E}}|D_\sigma|\rho_\sigma^n\left[\tilde{u}_\sigma^{n+1}-u_\sigma^n\right]^2+\frac{\gamma}{2}\sum_n\delta t\left[\sum_{\sigma\in\mathcal{E}}|\sigma|\,\rho_{\sigma,\gamma}^{n+1}\,|u_\sigma^{n+1}|\,|\rho_K^{n+1}-\rho_L^{n+1}|^2\right]\\
&+\delta t^2\sum_{\sigma\in\mathcal{E}}\frac{|D_\sigma|}{\rho_\sigma^n}\left(\frac{p_L^{n+1}-p_K^{n+1}}{d_\sigma}\right)^2\le\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,\rho_\sigma^0\,|u_\sigma^0|^2+\sum_{K\in\mathcal{M}}|K|\,\rho_K^0P(\rho_K^0)+\frac{|\Omega|}{e}.
\end{aligned}$$

□

II.5.3 Passing to the limit in the scheme

Theorem II.5.2

Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)},\delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)},p^{(m)},u^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence converges in $\mathrm{L}^p\big((0,T)\times\Omega\big)^4$, for $1\le p<\infty$, to $(\bar{\rho},\,\bar{p},\,\bar{u})\in\mathrm{L}^\infty\big((0,T)\times\Omega\big)^4$.

Then the limit $(\bar{\rho},\,\bar{p},\,\bar{u})$ satisfies the system (II.16).

Proof We sum the pressure correction equation and the momentum balance equation of scheme we obtain :

$$|D_\sigma|\frac{\rho_\sigma^n u_\sigma^{n+1}-\rho_\sigma^{n-1}u_\sigma^n}{\delta t}+\sum_{\varepsilon\in\mathcal{E}(D_\sigma)} F^n_{\sigma,\varepsilon}\tilde{u}_\varepsilon^{n+1}+|D_\sigma|\frac{p_L^{n+1}-p_K^{n+1}}{d_\sigma}-h^\alpha\sum_{\varepsilon=D_{\sigma'}|D_\sigma}|\varepsilon|\frac{\tilde{u}_\sigma^{n+1}-\tilde{u}_{\sigma'}^{n+1}}{d_{\sigma\sigma'}}=0. \quad \text{(II.22)}$$

Now we multiply the equation II.22 by $\delta t\,\varphi_\sigma^n$, and sum the result on $n\in\{0,...,N-1\}$ and $\sigma=K|L$, we obtain :

$$\underbrace{\sum_{n=0}^{N-1}\sum_{\sigma}|D_\sigma|(\rho_\sigma^n u_\sigma^{n+1}-\rho_\sigma^{n-1}u_\sigma^n)\varphi_\sigma^n}_{T_1^m}+\underbrace{\sum_{n=0}^{N-1}\delta t\sum_{\sigma}\sum_{\varepsilon\in\mathcal{E}(D_\sigma)}F^n_{\sigma\varepsilon}\tilde{u}_\varepsilon^{n+1}\varphi_\sigma^n}_{T_2^m}+\underbrace{\sum_{n=0}^{N-1}\delta t\sum_{\sigma}|D_\sigma|\frac{p_L^{n+1}-p_K^{n+1}}{d_{\sigma\sigma'}}\varphi_\sigma^n}_{T_3^m}=0$$

First we study T_1^m

$$T_1^m=\underbrace{\sum_{n=0}^{N-1}\sum_{\sigma}|D_\sigma|(\rho_\sigma^{n+1}u_\sigma^{n+1}-\rho_\sigma^n u_\sigma^n)\varphi_\sigma^n}_{T_{1,1}^m}+\underbrace{\sum_{n=0}^{N-1}\sum_{\sigma}|D_\sigma|\left[u_\sigma^n(\rho_\sigma^n-\rho_\sigma^{n-1})-u_\sigma^{n+1}(\rho_\sigma^{n+1}-\rho_\sigma^n)\right]\varphi_\sigma^n}_{R_m}$$

as for the implicit scheme we show that :

$$T_{1,1}^m\longrightarrow-\int_0^T\int_\Omega\rho u\partial_t\varphi-\int_\Omega\rho_0u_0\varphi_0 \text{ as } m\to\infty,$$

$$R_m=\underbrace{\sum_{n=0}^{N-1}\sum_{\sigma}|D_\sigma|\,u_\sigma^{n+1}(\rho_\sigma^{n+1}-\rho_\sigma^n)(\varphi_\sigma^{n+1}-\varphi_\sigma^n)}_{R_2^m}+\underbrace{\sum_{\sigma}|D_\sigma|\,u_\sigma^0(\rho_\sigma^0-\rho_\sigma^{-1})\varphi_\sigma^0}_{R_1^m}$$

$$\begin{aligned}|R_1^m| &\leq C_\varphi\sum_{n=0}^{N-1}\delta t\sum_{\sigma}|D_\sigma|\,|u_\sigma^{n+1}|\,|\rho_\sigma^{n+1}-\rho_\sigma^n|\\ &\leqslant\sqrt{\delta t}\,C_\varphi\left[\sum_{n=0}^{N-1}\delta t\sum_{\sigma=K|L}|D_\sigma|\,|u_\sigma^{n+1}|\right]^{\frac{1}{2}}\left[\sum_{n=0}^{N-1}\sum_{\sigma=K|L}|D_\sigma|\,|u_\sigma^{n+1}|\,|\rho_\sigma^{n+1}-\rho_\sigma^n|^2\right]^{\frac{1}{2}}\end{aligned}$$

$$\text{So, } R_1^m\longrightarrow 0,\ \text{ as } m\to\infty$$

$$R_2^m=-\delta t\sum_{\sigma\in\mathcal{E}}\sum_{\varepsilon\in\mathcal{E}(D_\sigma)}F^0_{\sigma\varepsilon}\,u_\sigma^0\,\varphi_\sigma^0\longrightarrow 0,\ \text{ as } m\to\infty$$

as for the implicit scheme we show that :

$$T_3^m\longrightarrow-\int_0^T\int_\Omega p\partial_x\varphi \text{ as } m\to\infty.$$

Now we study T_2^m

$$\begin{aligned}T_2^m &=-\frac{1}{4}\sum_{n=0}^{N-1}\delta t\sum_{K}\left(F^n_{K,\sigma}-F^n_{K,\sigma'}\right)\left(\tilde{u}_\sigma^{n+1}+\tilde{u}_{\sigma'}^{n+1}\right)(\varphi_\sigma^n-\varphi_{\sigma'}^n)\\ &=-\frac{1}{4}\sum_{n=0}^{N-1}\delta t\sum_{K}\left(|\sigma|\widetilde{\rho}_\sigma^n u_\sigma^n+|\sigma^{'}|\widetilde{\rho}_{\sigma'}^n u_{\sigma'}^n\right)\left(\tilde{u}_\sigma^{n+1}+\tilde{u}_{\sigma'}^{n+1}\right)(\varphi_\sigma^n-\varphi_{\sigma'}^n)\\ &=\underbrace{-\sum_{n=0}^{N}\delta t\sum_{K}|K|\,\rho_K^n\,\frac{u_\sigma^n\,\tilde{u}_\sigma^{n+1}+u_{\sigma'}^n\,\tilde{u}_{\sigma'}^{n+1}}{2}\,\frac{(\varphi_\sigma^n-\varphi_{\sigma'}^n)}{d_{\sigma\sigma'}}}_{T_{2,1}^m}+R_m\end{aligned}$$

where $\sigma = K|L$, $\sigma^{'} = M|K$ and R_m reads :

$$R_m = -\frac{1}{4}\sum_{n=0}^{N}\delta t\sum_{K}|\sigma|\left[(\widetilde{\rho}^n_\sigma u^n_\sigma + \widetilde{\rho}^n_{\sigma'} u^n_{\sigma'})(\tilde{u}^{n+1}_\sigma + \tilde{u}^{n+1}_{\sigma'}) - 2\rho^n_K\left(u^n_\sigma\,\tilde{u}^{n+1}_\sigma + u^n_{\sigma'}\,\tilde{u}^{n+1}_{\sigma'}\right)\right](\varphi^n_\sigma - \varphi^n_{\sigma'}).$$

First we study $T^m_{2,1}$,

$$\begin{aligned} T^m_{2,1} &= -\sum_{n=1}^{N}\sum_{K}\int_{t^n}^{t^{n+1}}\int_K \rho_m u_m^2 \eth_x \varphi^m_{\mathcal{M}} \\ &= -\int_0^T\int_\Omega \rho_m u_m^2 \eth_x\varphi^m_{\mathcal{M}} = -\int_0^T\int_\Omega \rho_m u_m^2\partial_x\varphi + \underbrace{\int_0^T\int_\Omega \rho_m u_m^2(\partial_x\varphi - \eth_x\varphi^m_{\mathcal{M}})}_{R'_m}. \end{aligned}$$

We have $\rho_m u_m^2 \in L^1(\Omega\times[0,T])$, so $R^{'}_m \longrightarrow 0$ as $m \to \infty$ thus :

$$T^m_{2,1} \longrightarrow -\int_0^T\int_\Omega \rho u^2\partial_x\varphi \text{ as } m\to\infty$$

We now study R_m. Expanding the quantity $2\,\rho^n_K\left(u^n_\sigma\,\tilde{u}^{n+1}_\sigma + u^n_{\sigma'}\,\tilde{u}^{n+1}_{\sigma'}\right)$ by the fact that $2\,(a\,b\,+\,c\,d) = (a+c)(b+d) + (a-c)(b-d)$, thus the term R_m reads :

$$\begin{aligned} R_m = \quad & \underbrace{-\frac{1}{4}\sum_{n=0}^{N}\delta t\sum_{K}|\sigma|\left[\left((\widetilde{\rho}^n_\sigma - \rho^n_K)u^n_\sigma + (\widetilde{\rho}^n_{\sigma'} - \rho^n_K)u^n_{\sigma'}\right)(\tilde{u}^{n+1}_\sigma + \tilde{u}^{n+1}_{\sigma'})\right](\varphi^n_\sigma - \varphi^n_{\sigma'})}_{R^m_1} \\ & \underbrace{+\frac{1}{4}\sum_{n=0}^{N}\delta t\sum_{K}|\sigma|\,\rho^n_K\,(u^n_\sigma - u^n_{\sigma'})(\tilde{u}^{n+1}_\sigma - \tilde{u}^{n+1}_{\sigma'})\,(\varphi^n_\sigma - \varphi^n_{\sigma'})}_{R^m_2}. \end{aligned}$$

First we study R^m_1 :

$$R^m_1 = \sum_{n=0}^{N}\delta t\sum_{K}|\sigma|\left[\left((\rho^n_L - \rho^n_K)\min(u^n_\sigma,0) + (\rho^n_M - \rho^n_K)\max(u^n_{\sigma'},0)\right)(\tilde{u}^{n+1}_\sigma + \tilde{u}^{n+1}_{\sigma'})\right](\varphi^n_\sigma - \varphi^n_{\sigma'})$$

$$\begin{aligned} |R^m_1| \quad & \leqslant C_\varphi\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}|\rho^n_K - \rho^n_L||u^n_\sigma|\left(|K||\tilde{u}^{n+1}_\sigma| + |K||\tilde{u}^{n+1}_{\sigma'}| + |L||\tilde{u}^{n+1}_\sigma| + |L||\tilde{u}^{n+1}_{\sigma''}|\right) \\ & \leqslant C_\varphi\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}|D_\sigma||\rho^n_K - \rho^n_L||u^n_\sigma|\left(|\tilde{u}^{n+1}_\sigma| + |\tilde{u}^{n+1}_{\sigma'}| + |\tilde{u}^{n+1}_{\sigma''}|\right) \\ & \leqslant \sqrt{h}\,C_\varphi\left[\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}|\sigma||u^n_\sigma|\,|\rho^n_K - \rho^n_L|^2\right]^{\frac{1}{2}}\left[\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}|D_\sigma||u^n_\sigma|\left(|\tilde{u}^{n+1}_\sigma| + |\tilde{u}^{n+1}_{\sigma'}| + |\tilde{u}^{n+1}_{\sigma''}|\right)^2\right]^{\frac{1}{2}} \\ & \leqslant \sqrt{h}\,C_\varphi\ \parallel u_m \parallel^{\frac{3}{2}}_{L^3([0,T],L^3(\Omega))}\left[\sum_{n=0}^{N}\delta t\sum_{\sigma=K|L}|\sigma||u^n_\sigma|\,|\rho^n_K - \rho^n_L|^2\right]^{\frac{1}{2}}. \end{aligned}$$

Now it is easy to see that

$$\begin{aligned}
|R_2^m| &\leqslant h^2\, C_\varphi \left[\sum_{n=0}^{N} \delta t \sum_K |K| \rho_K^n \left(\frac{u_\sigma^n - u_{\sigma'}^n}{d_{\sigma,\sigma'}}\right)^2\right]^{\frac{1}{2}} \left[\sum_{n=0}^{N} \delta t \sum_K |K| \rho_K^n \left(\frac{\tilde{u}_\sigma^{n+1} - \tilde{u}_{\sigma'}^{n+1}}{d_{\sigma,\sigma'}}\right)^2\right]^{\frac{1}{2}} \\
&\leqslant h^2\, C_\varphi \;\|\, \rho_m \,\|_{L^\infty([0,T]\times\Omega)} \|\, u_m \,\|^2_{L^2([0,T],H^1(\Omega))} \\
&\leqslant h^{2-\frac{d}{p}}\, C_\varphi \;\|\, \rho_m \,\|_{L^p([0,T]\times\Omega)} \|\, u_m \,\|^2_{L^2([0,T],H^1(\Omega))} \;.
\end{aligned}$$

Finally, we conclude that :

$$R_m \longrightarrow 0 \text{ as } m \rightarrow \infty$$

□

Chapitre

III

An unconditionally stable pressure correction scheme for compressible Navier-Stokes equations

In this paper we present a pressure correction scheme which is an extension to the full compressible Navier-Stokes equations of a scheme which was recently introduced for the compressible barotropic Navier-Stokes equations [20] and for the drift-flux model [26] . The space discretization is staggered, using either the Marker-And Cell (MAC) sheme or a nonconforming low-order finite element approximation ; general quandrangular or triangular meshes may thus be considered. The pressure correction scheme is shown to preserve the stability properties of the continuous problem, irrespectively of the space and time steps. To ensure the positivity of the energy, a key ingredient is to couple the mass and energy balance in the projection step. The existence of a solution to each step of the scheme is proven.

Plan du Chapitre III

III.1 Introduction **89**

III.2 Meshes and unknowns **90**

III.3 The time-implicit numerical scheme **92**

III.3.1 Semi-discrete algorithm 92

III.3.2 The fully discrete algorithm and its first properties 93

III.3.2.a Mass balance 93

III.3.2.b Momentum balance 94

III.3.2.c Energy balance 96

III.3.2.d Discrete EOS 97

III.3.2.e Existence of a solution to the fully discrete scheme 97

III.4 Pressure correction scheme **99**

III.4.1 Semi-discrete algorithm 99

III.4.2 Discrete algorithm 100

III.4.3 Properties of the scheme 101

III.5 Appendix : the MAC discretization of the dissipation term **103**

III.5.1 The two-dimensional case 103

III.5.2 Extension to the three-dimensional case 106

III.1 Introduction

The main object of this paper is to study the behaviour of a pressure correction scheme for the full compressible Navier-Stokes equations, with a low order finite element- finite volume discretization or with the MAC scheme. In particular, we wish to design a scheme for which we are able to prove the existence of a solution at each step of the scheme, and such that the approximate density and internal energy thus obtained are non-negative and the approximate total energy is controled. Let us consider the compressible Navier-Stokes equations, which may be written as :

$$\partial_t\rho + \operatorname{div}(\rho\,\boldsymbol{u}) = 0, \tag{III.1a}$$

$$\partial_t(\rho\,\boldsymbol{u}) + \operatorname{div}(\rho\,\boldsymbol{u}\otimes\boldsymbol{u}) + \boldsymbol{\nabla} p - \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})) = 0, \tag{III.1b}$$

$$\partial_t(\rho\,E) + \operatorname{div}(\rho\,E\,\boldsymbol{u}) + \operatorname{div}(p\,\boldsymbol{u}) + \operatorname{div}(\boldsymbol{q}) = \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})\cdot\boldsymbol{u}), \tag{III.1c}$$

$$E = \frac{1}{2}|\boldsymbol{u}|^2 + e, \tag{III.1d}$$

$$\rho = \wp(p,e). \tag{III.1e}$$

where t stands for the time, ρ, $\boldsymbol{u}$, p, E and e are the density, velocity, pressure, total energy and internal energy of the flow, $\boldsymbol{\tau}(\boldsymbol{u})$ stands for the shear stress tensor, which satisfies

$$\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{u} \geq 0, \forall \boldsymbol{u} \in \mathbb{R}^d, \tag{III.2}$$

$\boldsymbol{q}$ stands for the heat diffusion flux, and the function $\wp$ is the equation of state (EOS). The problem is supposed to be posed over $\Omega \times (0,T)$, where Ω is an open bounded connected subset of $\mathbb{R}^d$, $d \leq 3$ and $(0,T)$ is a finite time interval. This system must be supplemented by suitable boundary conditions, initial conditions and closure relations. For the sake of simplicity, we shall assume that $\boldsymbol{u}$ is prescribed to zero on the whole boundary $\partial\Omega$, and that the system is adiabatic, *i.e.* $\boldsymbol{\nabla q}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. The initial conditions for ρ, e are assumed to be positive ; finally, the closure relations for $\boldsymbol{\tau}(\boldsymbol{u})$ and for q, are given by :

$$\boldsymbol{\tau}(\boldsymbol{u}) = \mu(\boldsymbol{\nabla u} + \boldsymbol{\nabla}^t\boldsymbol{u}) - \frac{2\mu}{3}\operatorname{div}\boldsymbol{u}\,I, \quad \boldsymbol{q} = -\lambda\boldsymbol{\nabla} e, \tag{III.3}$$

where λ and μ are two positive parameters, possibly depending on x. In the sequel, we shall assume $\lambda = 1$ for the sake of simplicity. Let us suppose, again for the sake of simplicity, that $\boldsymbol{u}$ is prescribed to zero on the whole boundary $\partial\Omega$, and that the system is adiabatic, *i.e.* $\boldsymbol{\nabla q}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$.

Replacing the total energy E by its expression (III.1d) in the total energy equation (III.1c), we obtain :

$$\partial_t(\rho e) + \operatorname{div}(\rho e\boldsymbol{u}) + p\operatorname{div}\boldsymbol{u} + \operatorname{div}(\boldsymbol{q}) + \frac{1}{2}\partial_t(\rho\,|\boldsymbol{u}|^2) + \frac{1}{2}\operatorname{div}(\rho\,|\boldsymbol{u}|^2\,\boldsymbol{u}) + \boldsymbol{\nabla} p\cdot\boldsymbol{u} = \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u}).\boldsymbol{u}). \tag{III.4}$$

Noting that we have :

$$\begin{aligned}
\tfrac{1}{2}\partial_t(\rho\,|\boldsymbol{u}|^2) + \tfrac{1}{2}\operatorname{div}(\rho\,|\boldsymbol{u}|^2\,\boldsymbol{u}) &= \tfrac{|\boldsymbol{u}|^2}{2}\,[\partial_t(\rho) + \operatorname{div}(\rho\boldsymbol{u})] + \rho\boldsymbol{u}\cdot\partial_t(\boldsymbol{u}) + \rho\,|\boldsymbol{u}|^2\operatorname{div}(\boldsymbol{u}) \\
&= [\rho\partial_t(\boldsymbol{u}) + \partial_t(\rho)\boldsymbol{u} + \operatorname{div}(\rho\boldsymbol{u})\boldsymbol{u} + \rho\operatorname{div}(\boldsymbol{u})\boldsymbol{u}]\cdot\boldsymbol{u} \\
&= [\partial_t(\rho\,\boldsymbol{u}) + \operatorname{div}(\rho\,\boldsymbol{u}\otimes\boldsymbol{u})]\cdot\boldsymbol{u}
\end{aligned}$$

we get from the total energy equation (III.4) and from the momentum balance equation (III.1b) :

$$\partial_t(\rho e) + \operatorname{div}(\rho e \boldsymbol{u}) - \triangle e + p \operatorname{div}(\boldsymbol{u}) = \boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla} \boldsymbol{u}. \tag{III.5}$$

Formally, taking the inner product of (III.1b) with $\boldsymbol{u}$ and integrating over Ω, integrating (III.5) over Ω, and summing both relations yields the stability estimate :

$$\frac{d}{dt} \int_\Omega \big[\frac{1}{2} \rho \, |\boldsymbol{u}|^2 + \rho e\big] \, \mathrm{d}\boldsymbol{x} \leq 0. \tag{III.6}$$

Since we assume the initial condition for ρ to be positive, the mass balance (III.1a) formally implies that the density ρ remains positive.
We assume that the equation of state (III.1e) is such that there exists a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $p = f(\rho, e)$ with $f(\cdot, 0) = 0$ and $f(0, \cdot) = 0$, which we prolong by continuity to :

$$p = f(\rho, e) \text{ with } f(\rho, e) = 0 \, \forall \rho \leq 0 \text{ or } e \leq 0. \tag{III.7}$$

Equation (III.5) then implies (thanks to (III.2)) that the internal energy e remains positive (again at least formally), and so (III.6) yields a control on the unknown $\boldsymbol{u}$. Mimicking this computation at the discrete level necessitates to check some arguments, among them :

(i) a discrete counterpart to the relation :

$$\int_\Omega \big[\partial_t(\rho \boldsymbol{u}) + \operatorname{div}(\rho \boldsymbol{u} \otimes \boldsymbol{u})\big] \cdot \boldsymbol{u} \, \mathrm{d}\boldsymbol{x} = \frac{d}{dt} \int_\Omega \frac{1}{2} \rho \, |\boldsymbol{u}|^2 \, \mathrm{d}\boldsymbol{x}.$$

(ii) the equality of the integral of the dissipation term at the right-hand side of the discrete counterpart of (III.5) and the (discrete) L^2 inner product between the velocity and the diffusion term in the discrete momentum balance equation (III.13).

(iii) the non-negativity of the right-hand side of (III.5) in order, to preserve the positivity of the internal energy.

The point (i) is extensively discussed in [25] (see also [38]), and will not be treated here.

III.2 Meshes and unknowns

Let $\mathcal{M}$ be a discretization mesh of the domain Ω consisting of discretization cells which are either convex quadrilaterals ($d = 2$) or hexahedra ($d = 3$), or simplices. If the shape of Ω allows, we whall consider rectangular cells ($d = 2$) or rectangular parallelepipedic cells ($d = 3$). By $\mathcal{E}$ and $\mathcal{E}(K)$ we denote the set of all edges σ of the mesh and of the element $K \in \mathcal{M}$ respectively. The set of edges included in the boundary of Ω is denoted by $\mathcal{E}_{\text{ext}}$ and the set of internal edges ($i.e.$ $\mathcal{E} \setminus \mathcal{E}_{\text{ext}}$) is denoted by $\mathcal{E}_{\text{int}}$. The mesh $\mathcal{M}$ is supposed to be regular in the usual sense of the finite selement literature (e.g. [9]), and, in particular, it satisfies the following properties : .2

$\bar{\Omega} = \bigcup_{K \in \mathcal{M}} \bar{K}$

for K, $L \in \mathcal{M}$, the intersection $\bar{K} \cap \bar{L}$ is either reduced to the empty set, or to a vertex if $d = 2$ and a segment if $d = 3$, or else it is (the closure of) a common $(d-1)$-edge of K and L, denoted by $K|L$.

For each internal edge of the mesh $\sigma = K|L$, $\boldsymbol{n}_{KL}$ stands for the normal vector to σ, oriented from K to L. The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-dimensional measure of the face σ. For any $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, we denote by $d_{K,\sigma}$ the Euclidean distance between the center x_K of the mesh and the edge σ. For any $\sigma \in \mathcal{E}$, we define $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$, if $\sigma \in \mathcal{E}_{\text{int}}$ and $d_\sigma = d_{K,\sigma}$ if $\sigma \in \mathcal{E}_{\text{ext}}$. For any σ and $\sigma^{'}$ elements of $\mathcal{E}$, we denote by $d_{\sigma\sigma'}$ the Euclidean distance between σ and $\sigma^{'}$.

The space discretization is staggered, using either the Marker-And Cell (MAC) scheme [37, 36], or non-conforming low-order finite element approximations, namely the Rannacher and Turek (RT) element [65] for quadrilateral or hexahedric meshes, or the Crouzeix-Raviart (CR) element [11] for simplicial meshes.

For all discretizations (MAC, RT and CR), the degrees of freedom for the pressure, the density and the internal energy are associated to the cells of the mesh $\mathcal{M}$. The degrees of freedom are therefore :

$$\big\{p_K,\ \rho_K,\ e_K,\ K \in \mathcal{M}\big\}.$$

The approximate density, pressure and internal energy therefore belong to the space L_h of piecewise constant functions :

$$L_h = \big\{q_h \in L^2(\Omega)\ :\ q_h|_K = \text{ constant}, \forall K \in \mathcal{M}\big\}.$$

For $1 \le i \le d$, the degrees of freedom for the i^{th} component of the velocity are associated to a subset of $\mathcal{E}$, denoted by $\mathcal{E}^{(i)} \subset \mathcal{E}$, and are denoted by

$$\big\{u_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)}\big\}.$$

The definition of the sets $\mathcal{E}^{(i)}$ depends on the choice of the discretization :

- **MAC discretization.** In this case the set $\mathcal{E}^{(i)}$ is the set of edges that are orthogonal to the i-th basis vector $\boldsymbol{e}^{(i)}$.
- **RT and CR discretization.** In this case the set $\mathcal{E}^{(i)}$ is the whole set $\mathcal{E}_{\text{int}}$, and the degrees of freedom $u_{\sigma,i}$ are the components of the velocities with respect to the finite element shape functions. More precisely :
  - The reference element $\widehat{K}$ for the Rannacher-Turek rotated bilinear element is the unit d-cube (with edges parallel to the coordinate axes). The discrete functional space on $\widehat{K}$ is $\tilde{Q}_1(\widehat{K})^d$, where $\tilde{Q}_1(\widehat{K})$ is defined as follows :

    $$\tilde{Q}_1(\widehat{K}) = \text{span}\big\{1,\ (x_i)_{i=1,\dots,d},\ (x_i^2 - x_{i+1}^2)_{i=1,\dots,d-1}\big\}.$$

  - The reference element for the Crouzeix-Raviart is the unit d-simplex and the discrete functional space is the space P_1 of affine polynomials.

  The mapping from the reference element to the actual one is, for the Rannacher-Turek element, the standard Q_1 mapping and, for the Crouzeix-Raviart element, the standard affine mapping. The discrete space W_h is then defined as follows :

$$\begin{aligned} W_h = \{\boldsymbol{u} \in (L^2(\Omega))^d\ :\ & \boldsymbol{u}|_K \in W(K)^d, \forall K \in \mathcal{M}, \\ & \int_\sigma \boldsymbol{u}|_K \,\mathrm{d}\gamma = \int_\sigma \boldsymbol{u}|_L \,\mathrm{d}\gamma\, \forall \sigma = K|L \in \mathcal{E}_{\text{int}} \text{ and } \int_\sigma \boldsymbol{u} \,\mathrm{d}\gamma = 0,\ \forall \sigma \in \mathcal{E}_{\text{ext}}\ \}. \end{aligned}$$

where $W(K)$ is the space of functions on K generated by the reference element and the above described mapping. We define $\boldsymbol{u}_\sigma = \sum_{i=1}^d u_{\sigma,i}\, \boldsymbol{e}^{(i)}$ where $\boldsymbol{e}^{(i)}$ is the i^{th} vector of the canonical basis of $\mathbb{R}^d$.

In order to write a discrete momentum conservation, we need to introduce a dual mesh. For any $K \in \mathcal{M}$ and any face $\sigma \in \mathcal{E}(K)$, let $D_{K,\sigma}$ be the cone with basis σ and with vertex the mass center of K in both the RT and CR cases and let $D_{K,\sigma}$ be the rectangle of basis σ and of measure $|D_{K,\sigma}|$ equal to half the measure of K in the MAC case. The volume $D_{K,\sigma}$ is referred to as the half-diamond cell associated to K and σ. For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the diamond cell D_σ associated to σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external edge $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}(K)$, D_σ is set identical to $D_{K,\sigma}$. We denote by $\varepsilon = D_\sigma | D_{\sigma'}$ the face separating two diamond cells D_σ and $D_{\sigma'}$ (see Figure III.1).

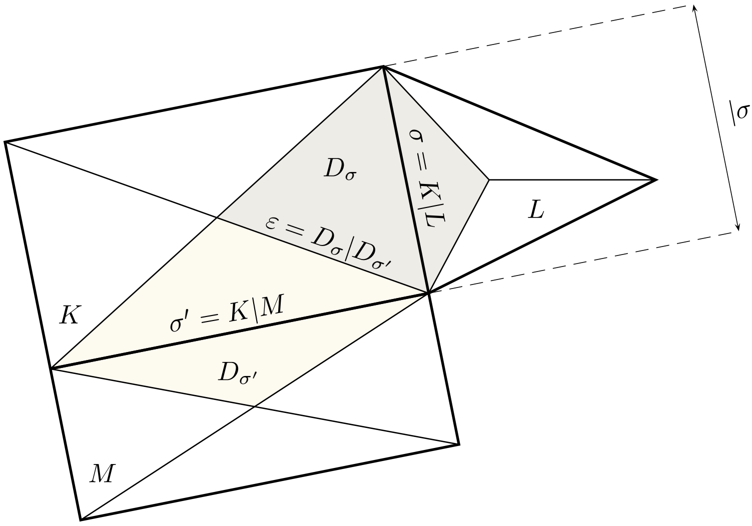
$|\sigma|$
D_σ
$\sigma = K|L$
L
$\varepsilon = D_\sigma|D_{\sigma'}$
K
$\sigma' = K\backslash M$
$D_{\sigma'}$
M

Fig. III.1 – Rannacher-Turek and Crouzeix-Raviart elements.

III.3 The time-implicit numerical scheme

III.3.1 Semi-discrete algorithm

Let us consider a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the time interval $(0, T)$, which, for the sake of simplicity, we suppose uniform. Let $\delta t = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N-1$ be the constant time step. In a time semi-discrete setting, the implicit-in-time numerical scheme reads.

$$\frac{\rho^{n+1}-\rho^n}{\delta t}+\mathrm{div}(\rho^{n+1}\,\boldsymbol{u}^{n+1})=0 \tag{III.8a}$$

$$\frac{\rho^{n+1}\,\boldsymbol{u}^{n+1}-\rho^n\,\boldsymbol{u}^n}{\delta t}+\mathrm{div}(\rho^{n+1}\,\boldsymbol{u}^{n+1}\otimes\boldsymbol{u}^{n+1})+\boldsymbol{\nabla}p^{n+1}-\mathrm{div}\tau(\boldsymbol{u}^{n+1})=0 \tag{III.8b}$$

$$\frac{\rho^{n+1}\,e^{n+1}-\rho^n\,e^n}{\delta t}+\mathrm{div}(\rho^{n+1}\,e^{n+1}\,\boldsymbol{u}^{n+1})-\triangle e^{n+1}+p^{n+1}\,\mathrm{div}(\boldsymbol{u}^{n+1})=\boldsymbol{\tau}(\boldsymbol{u}^{n+1}):\boldsymbol{\nabla}\boldsymbol{u}^{n+1} \tag{III.8c}$$

$$p^{n+1}=\wp(e^{n+1},\rho^{n+1}). \tag{III.8d}$$

III.3.2 The fully discrete algorithm and its first properties

Let us now give the space discretization of the various steps of the algorithm (III.8).

III.3.2.a Mass balance

The mass balance equation (III.8a) is always discretized by an upwind finite-volume technique in order to ensure the positivity of the density ; more precisely, the discretized mass balance reads :

$$\forall K\in\mathcal{M},\qquad \frac{|K|}{\delta t}(\rho_K^{n+1}-\rho_K^n)+\sum_{\sigma\in\mathcal{E}(K)}F_{K,\sigma}^{n+1}=0, \tag{III.9}$$

where $F_{K,\sigma}^{n+1}$ stands for the numerical mass flux across σ outward K. On the internal edges, the numerical flux is defined by :

$$\forall\sigma\in\mathcal{E}_{\rm int},\ \sigma=K|L,\ F_{K,\sigma}^{n+1}=|\sigma|\,\widetilde{\rho}_\sigma^{n+1}\ \boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}, \tag{III.10}$$

where $\boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}$ is the approximation of the normal velocity to the face σ outward K, and $\widetilde{\rho}_\sigma^{n+1}$ is the upwind density at the edge $\sigma=K|L$, that is :

$$\widetilde{\rho}_\sigma^{n+1}=\left|\begin{array}{l}\rho_K^{n+1}\ \text{ if } \boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}\geq 0,\\ \rho_L^{n+1}\ \text{ otherwise .}\end{array}\right. \tag{III.11}$$

Since $\boldsymbol{u}$ is assumed to be equal to 0 on the boundary, we impose :

$$\forall\sigma\in\mathcal{E}_{\rm ext},\ \sigma=K|L,\ F_{K,\sigma}^{n+1}=0, \tag{III.12}$$

As mentioned previously, with such an upwind discretization, we get the positivity of the density :

Lemma III.3.1 (Positivity of the density)
(see e.g. [27, Lemma 2.1]) Let $(\boldsymbol{u}_\sigma^{n+1})_{\sigma in\mathcal{E}_{\rm int}}$ be a given discrete velocity field, let $(\rho_K^n)_{K\in\mathcal{M}}$ be a discrete density field for a given $n\in\mathbb{N}$. Assume that $\rho_K^n\geq 0\ \forall K\in\mathcal{M}$. If a family $(\rho_K^{n+1})_{K\in\mathcal{M}}$ satisfies (III.9)–(III.11), then $\rho_K^{n+1}\geq 0$, $\forall K\in\mathcal{M}$.

III.3.2.b Momentum balance

Because of the choice of a staggered discretization, the momentum equation is discretized on a dual mesh, the dual cells of which are related to the faces where the velocity unknowns are located. On rectangular grids, it is approximated by the MAC scheme. Otherwise we use a combined finite volume – finite element method with low-degree finite elements for the diffusive terms, Crouzeix-Raviart element for simplicial meshes, Rannacher-Turek element [65] for quadrangles and hexahedra, and with a finite volume technique on the dual mesh for the time derivative term and convection term. The fully discretized momentum balance equations read, for $1 \leq i \leq d$, $\forall \sigma \in \mathcal{E}^{(i)}$:

$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1} u_{\sigma,i}^{n+1} - \rho_\sigma^n u_{\sigma,i}^n) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1} u_{\varepsilon,i}^{n+1} + |D_\sigma| \, (\boldsymbol{\nabla} p^{n+1})_\sigma^{(i)} - |D_\sigma| \, (\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_\sigma^{(i)} = 0, \quad \text{(III.13)}$$

where ρ_σ^{n+1} (resp. ρ_σ^n) stands for an approximation of the density on the edge σ at time t^{n+1} (resp. t^n), $F_{\sigma,\varepsilon}^{n+1}$ is the discrete mass flux through the dual edge ε outward D_σ, $u_{\varepsilon,i}^{n+1}$ stands for an approximation of u_i^{n+1} on ε, $(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_\sigma^{(i)}$ is an approximation of the i-th component of the viscous term associated to σ, and $(\boldsymbol{\nabla} p^{n+1})_\sigma^{(i)}$ is the i-th component of the discrete gradient of the pressure p at the face σ. Let us give some details on these approximations.

Discrete dual densities and mass fluxes The approximate densities ρ_σ^{n+1} and discrete mass fluxes on the dual edges are chosen such the following discrete mass balance over the dual cells is satisfied :

$$\forall \sigma \in \mathcal{E}_{\text{int}}, \qquad \frac{|D_\sigma|}{\delta t} \, (\rho_\sigma^{n+1} - \rho_\sigma^n) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1} = 0, \quad \text{(III.14)}$$

This relationship may be obtained from the primal mass balance (III.9) by defining ρ_σ^n as a weighted average with respect to the primal unknowns :

$$\forall \sigma \in \mathcal{E}_{\text{int}}, \, \sigma = K|L, \qquad |D_\sigma| \; \rho_\sigma^n = |D_{K,\sigma}| \; \rho_K^n + |D_{L,\sigma}| \; \rho_L^n, \quad \text{(III.15)}$$

choosing for the discrete flux $F_{\sigma,\varepsilon}^{n+1}$ through the dual face ε of the half dual cell D_σ the value of the flux through ε of a constant divergence lifting of the mass fluxes $(|\sigma| \boldsymbol{u}_\sigma \cdot \boldsymbol{n}_\sigma \rho_\sigma)_{\sigma \in \mathcal{E}}$ through the faces of the primal cell K ; for a detailed construction of this approximation, we refer to [25, 1] in the finite element case in $2D$, and to [38] in the MAC case. The additional unknowns $u_{\varepsilon,i}^{n+1}$ may be chosen centered or upwind. In the centered case, for an internal side $\varepsilon = D_\sigma | D_{\sigma'}$, we thus get $u_{\varepsilon,i}^{n+1} = (u_{\sigma,i}^{n+1} + u_{\sigma',i}^{n+1})/2$ while, in the upwind case, we have $u_{\varepsilon,i}^{n+1} = u_{\sigma,i}^{n+1}$ if $F_{\sigma,\varepsilon}^{n+1} \geq 0$ and $u_{\varepsilon,i}^{n+1} = u_{\sigma',i}^{n+1}$ otherwise.

Because the velocity unknowns are located on the edges, the dual discrete balance equation (III.14) is crucial in order to obtain the following stability result, which is a discrete equivalent of the kinetic energy theorem :

$$\begin{aligned} \sum_{i=1}^{d} \sum_{\sigma \in \mathcal{E}_{\text{int}}^{(i)}} \Big[\frac{|D_\sigma|}{\delta t} \; (\rho_\sigma^n \; u_{\sigma,i}^{n+1} - \rho_\sigma^n \; u_{\sigma,i}^n) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1} \; u_{\varepsilon,i}^{n+1} \Big] u_{\sigma,i}^{n+1} \geq \\ \frac{1}{2} \sum_{i=1}^{d} \sum_{\sigma \in \mathcal{E}_{\text{int}}^{(i)}} \frac{|D_\sigma|}{\delta t} \; \big[\rho_\sigma^{n+1} \; |u_{\sigma,i}^{n+1}|^2 - \rho_\sigma^n \; |u_{\sigma,i}^n|^2 \big]. \end{aligned} \quad \text{(III.16)}$$

We refer to [27] and [38] for the proof of this result in the finite element case and MAC case respectively.

Viscous term The MAC discretization of the dissipation term $\big(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla u}^{n+1}\big)_K$ associated to K is detailed in the appendix (see formula (III.44), see also [2]), and the following property is satisfied :

$$\big(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla u}^{n+1}\big)_K \geq 0. \tag{III.17}$$

It is clear that (III.17) also holds with a low order finite element discretization. Multiplying the approximation of the viscous term by the corresponding unknown of the velocity $u_{\sigma,i}^{n+1}$ and summing over the edges and the components, we obtain :

$$\sum_{i=1}^{d} \sum_{\sigma\in\mathcal{E}^{(i)}} |D_\sigma|\,(\mathrm{div}\boldsymbol{\tau}(\boldsymbol{u}^{n+1}))_\sigma^{(i)} u_{\sigma,i}^{n+1} = -\sum_{K\in\mathcal{M}} |K|\ \big(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla u}^{n+1}\big)_K. \tag{III.18}$$

This equality is the analogue of $\displaystyle\int_\Omega \mathrm{div}\boldsymbol{\tau}(\boldsymbol{u})\cdot\boldsymbol{u} = -\int_\Omega \boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla u}$. For the proof of the property (III.17) and the equality (III.18), we refer to [2].

Pressure gradient term The finite element discretization for the pressure gradient term at the internal face $\sigma = K|L$ reads :

$$|D_\sigma|(\boldsymbol{\nabla} p^{n+1})_\sigma^{(i)} = -\sum_{K\in\mathcal{M}} \int_K p^{n+1}\ \mathrm{div}\boldsymbol{\varphi}_\sigma^{(i)}\,\mathrm{d}\boldsymbol{x},\ i = 1,\ldots,d.$$

where $\boldsymbol{\varphi}_\sigma^{(i)} = \varphi_\sigma\,\boldsymbol{e}^{(i)}$ and where φ_σ is the scalar finite element basis function (for the CR finite element, it is affine on each element, equal to 1 at the center of σ and equal to 0 at the center of all other edges). Since the pressure is piecewise constant, the transposed of the discrete gradient operator takes the form of the finite volume standard discretization of the divergence based on the finite element mesh, which coincides with the MAC discretization of the divergence ; indeed, the previous relation can be rewritten as follows :

$$|D_\sigma|(\boldsymbol{\nabla} p^{n+1})_\sigma^{(i)} = -\sum_{K\in\mathcal{M}} \int_K p^{n+1}\ \mathrm{div}\boldsymbol{\varphi}_\sigma^{(i)}\,\mathrm{d}\boldsymbol{x} = |\sigma|\ (p_L^{n+1} - p_K^{n+1})\ \boldsymbol{n}_{K,\sigma}\cdot\boldsymbol{e}^{(i)}. \tag{III.19}$$

Multiplying this equality by $u_{\sigma,i}^{n+1}$ and summing over the edges and the components, we obtain :

$$\begin{aligned}\sum_{i=1}^{d} \sum_{\sigma\in\mathcal{E}^{(i)}} &|D_\sigma|\ (\boldsymbol{\nabla} p^{n+1})_\sigma^{(i)}\ u_{\sigma,i}^{n+1} \\ &= -\sum_{K\in\mathcal{M}} p_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\ \boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma} = -\sum_{K\in\mathcal{M}} |K| p_K^{n+1}(\mathrm{div}\boldsymbol{u}^{n+1})_K,\end{aligned} \tag{III.20}$$

where we have introduced the discrete divergence

$$(\mathrm{div}\boldsymbol{u}^{n+1})_K = \frac{1}{|K|}\sum_{\sigma\in\mathcal{E}(K)} |\sigma|\ \boldsymbol{u}_\sigma^{n+1}\cdot\boldsymbol{n}_{K,\sigma}.$$

This equality is also valid in the case of the MAC discretization ; it is the discrete analogue to $\displaystyle\int_\Omega \boldsymbol{\nabla} p\cdot\boldsymbol{u} = -\int_\Omega p\,\mathrm{div}(\boldsymbol{u})$. The finite element discretization for this term reads :

$$|D_\sigma|(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_\sigma^{(i)} = -\sum_{K\in\mathcal{M}} \int_K \tau(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla}\boldsymbol{\varphi}_\sigma^{(i)}\,\mathrm{d}\boldsymbol{x}.$$

III.3.2.c Energy balance

The internal energy equation (III.5) is discretized in a similar way to the momentum equation. The resulting discrete internal energy equation reads :

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} e_\sigma^{n+1} \\ + \sum_{\substack{\sigma \in \mathcal{E}(K) \\ \sigma = K|L}} |\sigma| \frac{e_K^{n+1} - e_L^{n+1}}{d_\sigma} + |K| p_K^{n+1} (\mathrm{div}(\boldsymbol{u}^{n+1}))_K = |K| \left(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla} \boldsymbol{u}^{n+1}\right)_K, \qquad \text{(III.21)}$$

where e_σ^{n+1} is the internal energy at the edge $\sigma = K|L$, computed with the upwind technique :

$$e_\sigma^{n+1} = \left| \begin{array}{l} e_K^{n+1}, \ \text{ if } \boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma} \geq 0, \\ e_L^{n+1}, \ \text{ otherwise.} \end{array} \right. \qquad \text{(III.22)}$$

Again, this upwind choice allows to to ensure the positivity of the internal energy, as we shall prove thanks to the following lemma, which states the stability of an appropriate discretization of a convection operator ; it gives, in particular, the conservation of the kinetic energy (III.16) which we mentioned earlier (see also [20]).

Lemma III.3.2
[27, Lemma 2.2] Let $(\rho_K)_{K \in \mathcal{M}}$ and $(\rho_K^*)_{K \in \mathcal{M}}$ be two families of real numbers satisfying the following set of equations :

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K - \rho_K^*) + \sum_{\sigma = K|L} F_{K,\sigma} = 0$$

where $F_{K,\sigma}$ is a quantity associated to the edge σ and to the control volume K. We suppose that, for any internal $\sigma = K|L$, $F_{K,\sigma} = -F_{L,\sigma}$. Let $(z_K)_{K \in \mathcal{M}}$ and $(z_K^*)_{K \in \mathcal{M}}$ be two families of real numbers. Then the following stability property holds :

$$- \sum_{K \in \mathcal{M}} y_K^- \Big[\frac{|K|}{\delta t}(\rho_K z_K - \rho_K^* z_K^*) + \sum_{\sigma = K|L} (F_{K,\sigma}^+ z_K - F_{K,\sigma}^- z_L) \Big] \geq \\ \frac{1}{2} \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} \, \big[\rho_K (z_K^-)^2 - \rho_K^* ((z_K^*)^-)^2\big].$$

Let us now state the positivity result. Note that this result together with the positivity of the density (Lemma (III.3.1)) and the stability inequality (III.16) are *a priori* result since we have not yet shown the existence of a solution to the scheme (III.9)–(III.19), (III.21)–(III.22). In fact, these *a priori* estimates are used to prove the existence of a solution in Section III.3.2.e below.

Lemma III.3.3 (Positivity of the internal energy)
Under assumption (III.7) and (III.2), let $n \in \mathbb{N}$, let $(\rho_K^n, \boldsymbol{u}_K^n, e_K^n)_{K \in \mathcal{M}} \in \mathbb{R}^{\mathrm{card}\mathcal{M}} \times (\mathbb{R}^{\mathrm{card}\mathcal{E}})^d \times \mathbb{R}^{\mathrm{card}\mathcal{M}})$ and assume that $e_K^n \geq 0 \ \forall K \in \mathcal{M}$; let $(\rho_K^{n+1}, \boldsymbol{u}_K^{n+1}, e_K^{n+1})_{K \in \mathcal{M}}$ satisfy (III.9)–(III.19), (III.21)–(III.22), then $e_K^{n+1} \geq 0 \ \forall K \in \mathcal{M}$.

Proof For $n \in \mathbb{N}$ we assume that $e_K^n \geq 0$ for all $K \in \mathcal{M}$. Multiplying the discrete internal energy equation (III.21) by $(-(e_K^{n+1})^-)$, using the fact that e_σ^{n+1} is given by the upwind choice (III.22) and summing over the mesh, we obtain :

$$\underbrace{-\sum_{K\in\mathcal{M}} (e_K^{n+1})^- \left[\frac{|K|}{\delta t}\rho_K^{n+1}e_K^{n+1} - \rho_K^n e_K^n + \sum_{\sigma=K|L}((F_{K,\sigma}^{n+1})^+ e_K^{n+1} - (F_{K,\sigma}^{n+1})^- e_L^{n+1})\right]}_{E_1}$$

$$\underbrace{-\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}(K)} |\sigma| \frac{e_K^{n+1} - e_L^{n+1}}{d_\sigma}(e_K^{n+1})^-}_{E_2} \underbrace{- \sum_{K\in\mathcal{M}} |K| p_K^{n+1}(\mathrm{div} u^{n+1})_K (e_K^{n+1})^-}_{E_3}$$

$$= \underbrace{-\sum_{K\in\mathcal{M}} |K| \left(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla}\boldsymbol{u}^{n+1}\right)_K (e_K^{n+1})^-}_{E_4}.$$

By virtue of Lemma III.3.2, the first term E_1 can be estimated as follows :

$$E_1 \geq \frac{1}{2}\sum_{K\in\mathcal{M}} \frac{|K|}{\delta t}\left[\rho_K^{n+1}((e_K^{n+1})^-)^2 - \rho_K^n((e_K^n)^-)^2\right] = \frac{1}{2}\sum_{K\in\mathcal{M}}\frac{|K|}{\delta t}\rho_K^{n+1}((e_K^{n+1})^-)^2.$$

Thanks to (III.7), we have $E_3 = 0$ and thanks to (III.17), we have $E_4 \leq 0$. Reordering the sum in the term E_2, we obtain :

$$E_2 = -\sum_{\sigma=K|L}\frac{|\sigma|}{d_\sigma}[e_K^{n+1} - e_L^{n+1}][(e_K^{n+1})^- - (e_L^{n+1})^-].$$

Since the function $x \mapsto x^-$ is non-increasing, we obtain that $E_2 \geq 0$; Gathering all the terms, we obtain :

$$0 \leq \frac{1}{2}\sum_{K\in\mathcal{M}}\frac{|K|}{\delta t}\rho_K^{n+1}((e_K^{n+1})^-)^2 \leq 0$$

which shows that $(e_K^{n+1})^- = 0$, for all $K \in \mathcal{M}$; this concludes the proof. □

III.3.2.d Discrete EOS

Finally, the equation of state (III.1e) is easily discretized by

$$\rho_K^n = \wp(p_K^n, e_K^n),\ \forall K \in \mathcal{M},\ \forall n \in \mathbb{N}. \tag{III.23}$$

III.3.2.e Existence of a solution to the fully discrete scheme

We recall the following theorem, which is a consequence of the topological degree theory, see *e.g.* [13], and which is a very powerful tool for the proof of existence of non linear systems arising from the discretization of non linear partial differential equations (see [16] for other examples of its use).

Theorem III.3.4 (Application of the topological degree, finite dimensional case)

Let V be a finite dimensional vector space on $\mathbb{R}$, $\|.\|$ a norm on V, let f be a continuous function from V to V and let $R > 0$. Let us assume that there exists a continuous function $F : V \times [0,1] \to V$ satisfying :

(*i*) $F(.,1) = f$,

(ii) $\forall \alpha \in [0,1]$, if $v \in V$ is such that $F(v,\alpha)=0$ then $v \in B_R = \{v \in V \ ; \|v\| < R\}$,

(iii) the topological degree of $F(.,0)$ with respect to 0 and B_R is equal to $d_0 \neq 0$.

Then the topological degree of $F(.,1)$ with respect to 0 and to B_R is also equal to $d_0 \neq 0$; consequently, there exists at least a solution $v \in B_R$ such that $f(v)=0$.

Theorem III.3.5 (Existence of a solution to the implicit scheme)

Under assumption (III.7) and (III.2), assume that $\rho^0_K > 0$ and $e^0_K > 0$, for all $K \in \mathcal{M}$. There exists a solution $(\rho^n_K, \boldsymbol{u}^n_K, e^n_K)_{K\in\mathcal{M},\ n\le N}$ to the implicit scheme (III.9)–(III.19), (III.21)–(III.22) which satisfies $\rho^n_K > 0$, $e^n_K > 0$ for all $K \in \mathcal{M}$ and $n \le N$, and such that the following inequality holds for all $n \le N$:

$$\sum_{K\in\mathcal{M}} |K|\ \rho^n_K e^n_K + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} |D_\sigma|\ \rho^n_\sigma\ |u^n_\sigma|^2 \le \sum_{K\in\mathcal{M}} |K|\ \rho^0_K e^0_K + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} |D_\sigma|\ \rho^0_\sigma\ |u^0_\sigma|^2. \qquad \text{(III.24)}$$

Proof

Let us first show that under the assumptions of the theorem, if the family $(\rho^n_K, \boldsymbol{u}^n_K, e^n_K)_{K\in\mathcal{M},\ n\le N}$ satisfies (III.9)–(III.19), (III.21)–(III.22), then the inequality (III.24) holds. Multiplying the discrete momentum balance equation (III.13) by the corresponding unknown of the velocity $u^{n+1}_{\sigma,i}$ and summing over the edges and the components i, by virtue of the stability of the discrete advection operator (III.16) and the equality (III.20) and (III.18) we obtain :

$$\frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} \frac{|D_\sigma|}{\delta t}\ \left[\rho^{n+1}_\sigma\ |\boldsymbol{u}^{n+1}_\sigma|^2 - \rho^n_\sigma\ |\boldsymbol{u}^n_\sigma|^2\right] - \sum_{K\in\mathcal{M}} |K| p^{n+1}_K ({\rm div}(\boldsymbol{u}^{n+1}))_K \\ + \sum_{K\in\mathcal{M}} |K| \left(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}) : \boldsymbol{\nabla}\boldsymbol{u}^{n+1}\right)_K \le 0. \qquad \text{(III.25)}$$

Noting that, by conservativity, $F_{K,\sigma} = -F_{L,\sigma}$ for $\sigma = K|L$, and that

$$\sum_{K\in\mathcal{M}} \sum_{\substack{\sigma\in\mathcal{E}(K)\\ \sigma=K|L}} |\sigma| \frac{e^{n+1}_K - e^{n+1}_L}{d_\sigma} = 0,$$

and summing (III.25) with the sum of the discrete internal energy equation (III.21) over $K \in \mathcal{M}$, we get :

$$\sum_{K\in\mathcal{M}} |K|\ \left[\rho^{n+1}_K\ e^{n+1}_K - \rho^n_K\ e^n_K\right] + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} |D_\sigma|\ \left[\rho^{n+1}_\sigma\ |\boldsymbol{u}^{n+1}_\sigma|^2 - \varrho^n_\sigma\ |\boldsymbol{u}^n_\sigma|^2\right] \le 0,$$

which concludes the proof of (III.24).

Let us now show the existence of a solution to the sheme (III.9)–(III.19), (III.21)–(III.23). For $\alpha \in [0,1]$ and a fixed $n \in \mathbb{N}$, we consider the following discrete set of equations, for $K \in \mathcal{M}$ and $\mathcal{E} in \mathcal{E}_{\rm int}$. For $\alpha = 0$, it is an invertible linear system, and for $\alpha = 1$, it is the fully discrete scheme (III.9)–(III.19), (III.21)–(III.23).

$$
\begin{aligned}
&\frac{|K|}{\delta t}(\rho_K^{n+1}-\rho_K^n)+\alpha\sum_{\sigma\in\mathcal{E}(K)}F_{K,\sigma}^{n+1}=0,\\
&\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1}u_{\sigma,i}^{n+1}-\rho_\sigma^n u_{\sigma,i}^n)+\alpha\left(\sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)}F_{\sigma,\varepsilon}^{n+1}u_{\varepsilon,i}^{n+1}.+|D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_\sigma^{(i)}-|D_\sigma|\,(\mathrm{div}\tau(\boldsymbol{u}^{n+1}))_\sigma^{(i)}\right)=0,\\
&\frac{|K|}{\delta t}(\rho_K^{n+1}e_K^{n+1}-\rho_K^n e_K^n)+\alpha\left[\sum_{\sigma\in\mathcal{E}(K)}F_{K,\sigma}^{n+1}e_\sigma^{n+1}+\sum_{\substack{\sigma\in\mathcal{E}(K)\\ \sigma=K|L}}|\sigma|\frac{e_K^{n+1}-e_L^{n+1}}{d_\sigma}+|K|p_K^{n+1}(\mathrm{div}(\boldsymbol{u}^{n+1}))_K\right]\\
&\qquad\qquad\qquad\qquad=\alpha|K|\left(\boldsymbol{\tau}(\boldsymbol{u}^{n+1}):\boldsymbol{\nabla}\boldsymbol{u}^{n+1}\right)_K,\forall K\in\mathcal{M},\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\rho_K^{n+1}=\wp(p_K^{n+1},e_K^{n+1}).
\end{aligned}
$$

By the same analysis that we performed for the study of the scheme (III.9)–(III.19), (III.21)–(III.23), any family $(\rho_K^{n+1})_{K\in\mathcal{M}}$ and $(e_K^{n+1})_{K\in\mathcal{M}}$ satisfying the above scheme is such that ρ_K^{n+1} and e_K^{n+1} are positive for all K. Moreover, the conservativity of the mass balance discretization yields that

$$\sum_{K\in\mathcal{M}}|K|\rho_K^{n+1}=\sum_{K\in\mathcal{M}}|K|\rho_K^n$$

which yields an L^∞ bound on the family $(\rho_K^{n+1})_{K\in\mathcal{M}}$; finally, the above stability result (III.16) also holds ; we thus have an uniform control over the families of real numbers $(\rho_K)_{K\in\mathcal{M}}$, $(\rho_K e_K)_{K\in\mathcal{M}}$ and vectors $(\rho_\sigma\boldsymbol{u}_\sigma)_{\sigma\in\mathcal{E}_{int}}$. For $\alpha=0$, the system is linear and invertible with respect to these unknowns. We conclude thanks to a topological degree argument. □

III.4 Pressure correction scheme

III.4.1 Semi-discrete algorithm

A pressure correction numerical scheme is obtained by complementing the scheme presented in the preceding section by an incremental projection method. Writing this algorithm in a semi-discrete time setting, this yields the following three steps :

1 - solve for $\tilde{p}^{n+1}$

$$\mathrm{div}\left(\frac{1}{\rho^n}\boldsymbol{\nabla}\tilde{p}^{n+1}\right)=\mathrm{div}\left(\frac{1}{\sqrt{\rho^n}\sqrt{\rho^{n-1}}}\boldsymbol{\nabla}p^n\right)\tag{III.26}$$

2 - Solve for $\tilde{u}^{n+1}$:

$$\frac{\rho^n\,\tilde{\boldsymbol{u}}^{n+1}-\rho^{n-1}\,\boldsymbol{u}^n}{\delta t}+\mathrm{div}(\rho^n\,\boldsymbol{u}^n\otimes\tilde{\boldsymbol{u}}^{n+1})+\boldsymbol{\nabla}\tilde{p}^{n+1}-\mathrm{div}\tau(\tilde{\boldsymbol{u}}^{n+1})=0\tag{III.27}$$

3 - Solve for p^{n+1}, u^{n+1}, ρ^{n+1} and e^{n+1} :

$$\rho^n \, \frac{\boldsymbol{u}^{n+1} - \tilde{\boldsymbol{u}}^{n+1}}{\delta t} + \boldsymbol{\nabla}(p^{n+1} - \tilde{p}^{n+1}) = 0 \tag{III.28a}$$

$$\frac{\rho^{n+1} - \rho^n}{\delta t} + \operatorname{div}(\rho^{n+1} \, \boldsymbol{u}^{n+1}) = 0 \tag{III.28b}$$

$$\frac{\rho^{n+1} \, e^{n+1} - \rho^n \, e^n}{\delta t} + \operatorname{div}(\rho^{n+1} \, e^{n+1} \, \boldsymbol{u}^{n+1}) - \triangle e^{n+1} + p^{n+1} \operatorname{div}(\boldsymbol{u}^{n+1}) = \boldsymbol{\tau}(\tilde{\boldsymbol{u}}^{n+1}) : \boldsymbol{\nabla}\tilde{\boldsymbol{u}}^{n+1} \tag{III.28c}$$

$$p^{n+1} = \wp(e^{n+1}, \rho^{n+1}). \tag{III.28d}$$

The first step is a renormalization of the pressure which is used in the stability analysis.

The second step is a classical semi-implicit solution of the momentum blance equation to obtain a predicted velocity.

Finally the last step is an original nonlinear pressure correction step, which couples the mass balance equation (III.28b) with the internal energy balance equation (III.28c). This coupling is important to ensure the positivity of the energy : indeed, in the proof of Lemma aIII.3.3, we used the fact that the pressure vanishes in the term $p^{n+1} \operatorname{div}(\boldsymbol{u}^{n+1})$ when e^{n+1} is negative.

III.4.2 Discrete algorithm

The space discretization is again staggered, using either the Marker–And–Cell (MAC) scheme, or nonconforming low-order finite element approximations.

The finite element discretization for the pressure prediction step at the internal face $\sigma = K|L$ reads :

$$\sum_{K \in \mathcal{M}} \int_K \frac{1}{\rho^n} \boldsymbol{\nabla}\tilde{p}^{n+1} \cdot \; \boldsymbol{\nabla}\boldsymbol{\varphi}_\sigma^{(i)} \, \mathrm{d}\boldsymbol{x} = \sum_{K \in \mathcal{M}} \int_K \frac{1}{\sqrt{\rho^n}\sqrt{\rho^{n-1}}} \boldsymbol{\nabla} p^n \cdot \; \boldsymbol{\nabla}\boldsymbol{\varphi}_\sigma^{(i)} \, \mathrm{d}\boldsymbol{x}$$

which coincides with the MAC discretization and may be rewritten as follows :

$$\forall K \in \mathcal{M}, \qquad \sum_{\sigma = K|L} \frac{|\sigma|^2}{|D_\sigma|} \, \frac{1}{\rho_\sigma^n} \, (\tilde{p}_K^{n+1} - \tilde{p}_L^{n+1}) = \sum_{\sigma = K|L} \frac{|\sigma|^2}{|D_\sigma|} \, \frac{1}{\sqrt{\rho_\sigma^n \; \rho_\sigma^{n-1}}} \, (p_K^n - p_L^n) \tag{III.29}$$

The discretization of projection equation (III.28a) is consistent with that of the momentum balance (III.27) , *i.e.* we use a mass lumping technique for the unsteady term in both cases and a standard finite element formulation for the gradient of the pressure increment in the finite element case :

$$\forall \sigma \in \mathcal{E}_{\rm int}, \text{ for } 1 \leq i \leq d, \qquad \frac{|D_\sigma|}{\delta t} \, \rho_\sigma^n \, (u_{\sigma,i}^{n+1} - \tilde{u}_{\sigma,i}^{n+1}) - \sum_{K \in \mathcal{M}} \int_K (p^{n+1} - \tilde{p}^{n+1}) \, \nabla \cdot \boldsymbol{\varphi}_\sigma^{(i)} \; dx = 0,$$

which can be rewritten as follows :

$$\forall \sigma \in \mathcal{E}_{\rm int}, \; \sigma = K|L, \qquad \frac{|D_\sigma|}{\delta t} \, \rho_\sigma^n \, (u_\sigma^{n+1} - \tilde{u}_\sigma^{n+1}) + |\sigma| \left[(p_L^{n+1} - \tilde{p}_L^{n+1}) - (p_K^{n+1} - \tilde{p}_K^{n+1}) \right] \boldsymbol{n}_{KL} = 0 \tag{III.30}$$

We may then write the general form of the fully discrete to the pressure correction scheme :

1 - Renormalization step : $\forall K \in \mathcal{M}$, ,

$$\sum_{\sigma=K|L} \frac{|\sigma|^2}{|D_\sigma|}\ \frac{1}{\rho_\sigma^n}\ (\tilde{p}_K^{n+1} - \tilde{p}_L^{n+1}) = \sum_{\sigma=K|L} \frac{|\sigma|^2}{|D_\sigma|}\ \frac{1}{\sqrt{\rho_\sigma^n\ \rho_\sigma^{n-1}}}\ (p_K^n - p_L^n) \tag{III.31}$$

2 - Velocity prediction step : for $1 \leq i \leq d$, and for any $\sigma \in \mathcal{E}_{\text{int}}^{(i)}$,

$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma^n \tilde{u}_{\sigma,i}^{n+1} - \rho_\sigma^{n-1} u_{\sigma,i}^n) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^n \tilde{u}_{\varepsilon,i}^{n+1} + |D_\sigma|\,(\boldsymbol{\nabla} \tilde{p}^{n+1})_\sigma^{(i)} - |D_\sigma|\,(\mathrm{div}\tau(\tilde{\boldsymbol{u}}^{n+1}))_\sigma^{(i)} = 0, \tag{III.32}$$

where $F_{\sigma,\varepsilon}^n$ is the dual edge mass flux.

3 - Projection step :

$$\begin{aligned} &\frac{|D_\sigma|}{\delta t}\ \rho_\sigma^n\,(u_{\sigma,i}^{n+1} - \tilde{u}_{\sigma,i}^{n+1}) + |\sigma|\,\left[(p_L^{n+1} - \tilde{p}_L^{n+1}) - (p_K^{n+1} - \tilde{p}_K^{n+1})\right] \boldsymbol{n}_{KL} = 0, \\ &\qquad\qquad\qquad\qquad\qquad \text{for } 1 \leq i \leq d \text{ and } \sigma = K|L \in \mathcal{E}_{\text{int}}^{(i)}, \end{aligned} \tag{III.33a}$$

$$\frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} = 0, \quad \forall K \in \mathcal{M}, \tag{III.33b}$$

$$\begin{aligned} &\frac{|K|}{\delta t}(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} e_\sigma^{n+1} - \sum_{\substack{\sigma \in \mathcal{E}(K), \\ \sigma = K|L}} |\sigma| \frac{e_L^{n+1} - e_K^{n+1}}{d_\sigma} \\ &\qquad + |K| p_K^{n+1} (\mathrm{div}(\boldsymbol{u}^{n+1}))_K = |K| \left(\boldsymbol{\tau}(\tilde{\boldsymbol{u}}^{n+1}) : \boldsymbol{\nabla} \tilde{\boldsymbol{u}}^{n+1}\right)_K, \quad \forall K \in \mathcal{M}, \end{aligned} \tag{III.33c}$$

$$p_K^{n+1} = \wp(e_K^{n+1}, \rho_K^{n+1}), \forall K \in \mathcal{M}, \tag{III.33d}$$

where $F_{K,\sigma}^{n+1}$ is the primal edge mass flux.

III.4.3 Properties of the scheme

THEOREM III.4.1

There exists a solution to the scheme (III.31)–(III.33d), which satisfies $\rho_K^n > 0$, $e_K^n > 0$ for all $K \in \mathcal{M}$ and $n \in \mathbb{N}$, and such that the following inequality holds :

$$\begin{aligned} &\sum_{K \in \mathcal{M}} |K|\ \rho_K^n e_K^n + \frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\text{int}}} |D_\sigma|\ \rho_\sigma^{n-1} |u_\sigma^n|^2 + \frac{\delta t^2}{2} |p^n|_{\rho^{n-1},\ \mathcal{M}}^2 \leq \\ &\qquad \sum_{K \in \mathcal{M}} |K|\ \rho_K^0 e_K^0 + \frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\text{int}}} |D_\sigma|\ \rho_\sigma^{-1} |u_\sigma^0|^2 + \frac{\delta t^2}{2} |p^0|_{\rho^{-1},\ \mathcal{M}}^2, \end{aligned} \tag{III.34}$$

where

$$|q|_{\rho,\ \mathcal{M}}^2 = \sum_{\sigma=K|L} \frac{1}{\rho_\sigma}\ \frac{|\sigma|^2}{|D_\sigma|} (p_L - p_K)^2.$$

Proof

The positivity of the density ρ_K^{n+1} and the internal energy e_K^{n+1} and the existence of a solution are obtained by repeating arguments similar to those invoked in the implicit-in-time scheme. There remains to prove that (III.34) holds.

Multiplying the discrete momentum balance equation (III.32) by the corresponding unknown of the velocity $\tilde{u}_{\sigma,i}^{n+1}$ and summing and the components i and their associated edges, by virtue of the stability of the discrete advection operator (III.16) and the equalities (III.20) and (III.18) we obtain :

$$\frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} \frac{|D_\sigma|}{\delta t}\left[\rho_\sigma^n\ |\tilde{\boldsymbol{u}}_\sigma^{n+1}|^2 - \rho_\sigma^{n-1}\ |\boldsymbol{u}_\sigma^n|^2\right] - \sum_{K\in\mathcal{M}} |K|\tilde{p}_K^{n+1}(\mathrm{div}(\tilde{\boldsymbol{u}}^{n+1}))_K \\ + \sum_{K\in\mathcal{M}} |K|\left(\boldsymbol{\tau}(\tilde{\boldsymbol{u}}^{n+1}) : \boldsymbol{\nabla}\tilde{\boldsymbol{u}}^{n+1}\right)_K \leq 0.$$

Summing the discrete internal energy equation (III.33c) over the cells $K\in\mathcal{M}$ and with the previous relation, we obtain :

$$\sum_{K\in\mathcal{M}} \frac{|K|}{\delta t}\left[\rho_K^{n+1}\quad e_K^{n+1} - \rho_K^n\ e_K^n\right] + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} \frac{|D_\sigma|}{\delta t}\left[\rho_\sigma^n\ |\tilde{\boldsymbol{u}}_\sigma^{n+1}|^2 - \rho_\sigma^{n-1}\ |\boldsymbol{u}_\sigma^n|^2\right] \\ + \sum_{K\in\mathcal{M}} |K|p_K^{n+1}(\mathrm{div}(\boldsymbol{u}^{n+1}))_K - \sum_{K\in\mathcal{M}} |K|\tilde{p}_K^{n+1}(\mathrm{div}(\tilde{\boldsymbol{u}}^{n+1}))_K \leq 0 \tag{III.35}$$

Reordering the first relation of the projection step (III.33a) and multiplying by $(\rho_\sigma^n)^{-1/2}$ we obtain, for $1\leq i\leq d$ and $\sigma = K|L\in\mathcal{E}_{\rm int}^{(i)}$:

$$\frac{|D_\sigma|}{\delta t}\sqrt{\rho_\sigma^n}\, u_{\sigma,i}^{n+1} + |\sigma|\frac{1}{\sqrt{\rho_\sigma^n}}(p_L^{n+1} - p_K^{n+1})\boldsymbol{n}_{KL} = \frac{|D_\sigma|}{\delta t}\sqrt{\rho_\sigma^n}\, \tilde{u}_{\sigma,i}^{n+1} + |\sigma|\frac{1}{\sqrt{\rho_\sigma^n}}(\tilde{p}_L^{n+1} - \tilde{p}_K^{n+1})\boldsymbol{n}_{KL}$$

Squaring the previous relation and multiplying by $\delta t/2|D_\sigma|$ and summing over the edges and the component i, we obtain :

$$\frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} \frac{|D_\sigma|}{\delta t}\left[\rho_\sigma^n\ |\boldsymbol{u}_\sigma^{n+1}|^2 - \rho_\sigma^n\ |\tilde{\boldsymbol{u}}_\sigma^{n+1}|^2\right] - \sum_{K\in\mathcal{M}} |K|p_K^{n+1}(\mathrm{div}(\boldsymbol{u}^{n+1}))_K \\ + \sum_{K\in\mathcal{M}} |K|\tilde{p}_K^{n+1}(\mathrm{div}(\tilde{\boldsymbol{u}}^{n+1}))_K + \frac{\delta t}{2}\sum_{\sigma=K|L}\frac{1}{\rho_\sigma^n}\frac{|\sigma|^2}{|D_\sigma|}(p_L^{n+1} - p_K^{n+1})^2 \\ - \frac{\delta t}{2}\sum_{\sigma=K|L}\frac{1}{\rho_\sigma^n}\frac{|\sigma|^2}{|D_\sigma|}(\tilde{p}_L^{n+1} - \tilde{p}_K^{n+1})^2 = 0$$

Summing this last relation with (III.35) and multiplying by δt, we get :

$$\sum_{K\in\mathcal{M}} |K|\left[\rho_K^{n+1}\quad e_K^{n+1} - \rho_K^n\ e_K^n\right] + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\rm int}} |D_\sigma|\left[\rho_\sigma^n\ |\boldsymbol{u}_\sigma^{n+1}|^2 - \rho_\sigma^{n-1}\ |\boldsymbol{u}_\sigma^n|^2\right] \\ +\frac{\delta t^2}{2}\underbrace{\sum_{\sigma=K|L}\frac{1}{\rho_\sigma^n}\frac{|\sigma|^2}{|D_\sigma|}(p_L^{n+1} - p_K^{n+1})^2}_{|p^{n+1}|^2_{\rho^n,\ \mathcal{M}}} - \frac{\delta t^2}{2}\underbrace{\sum_{\sigma=K|L}\frac{1}{\rho_\sigma^n}\frac{|\sigma|^2}{|D_\sigma|}(\tilde{p}_L^{n+1} - \tilde{p}_K^{n+1})^2}_{|\tilde{p}^{n+1}|^2_{\rho^n,\ \mathcal{M}}} = 0$$

Thanks to the renormalization step (III.31), we have :

$$|\tilde{p}^{n+1}|^2_{\rho^n,\ \mathcal{M}} \leq |p^n|^2_{\rho^{n-1},\ \mathcal{M}}$$

which concludes the proof. □

III.5 Appendix : the MAC discretization of the dissipation term

III.5.1 The two-dimensional case

Let us first propose a discretization for the diffusion term $-\mathrm{div}(\tau(u))$ in the momentum equation (III.1b).. We begin with the x-component of the velocity, for which we write a balance equation on $K^x_{i-\frac{1}{2},j} = (x_{i-1},\, x_i) \times (y_{j-\frac{1}{2}},\, y_{j+\frac{1}{2}})$ (see Figures III.2 and III.3 for the notations).

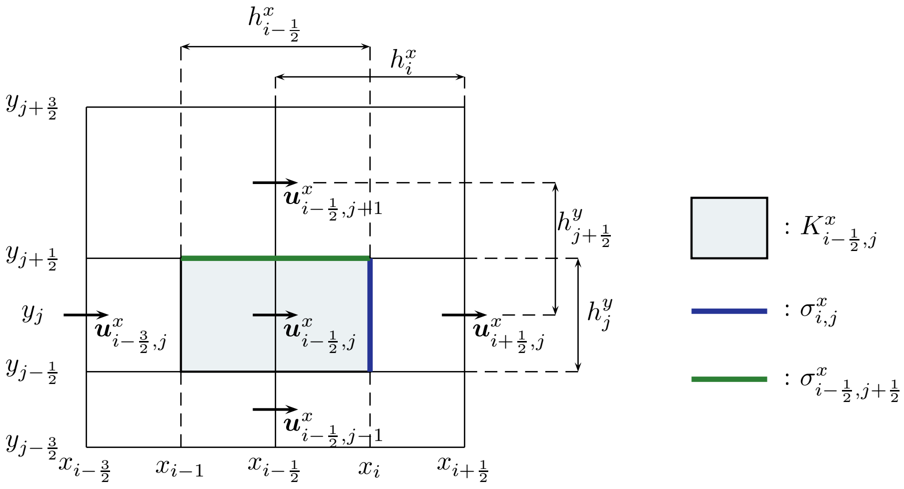
$h^x_{i-\frac{1}{2}}$
h^x_i
$y_{j+\frac{3}{2}}$
$\boldsymbol{u}^x_{i-\frac{1}{2},j+1}$
$h^y_{j+\frac{1}{2}}$
$y_{j+\frac{1}{2}}$
y_j
$\boldsymbol{u}^x_{i-\frac{3}{2},j}$
$\boldsymbol{u}^x_{i-\frac{1}{2},j}$
$\boldsymbol{u}^x_{i+\frac{1}{2},j}$
h^y_j
$y_{j-\frac{1}{2}}$
$\boldsymbol{u}^x_{i-\frac{1}{2},j-1}$
$y_{j-\frac{3}{2}}$
$x_{i-\frac{3}{2}}$
x_{i-1}
$x_{i-\frac{1}{2}}$
x_i
$x_{i+\frac{1}{2}}$
: $K^x_{i-\frac{1}{2},j}$
: $\sigma^x_{i,j}$
: $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$

FIG. III.2 – Dual cell for the x-component of the velocity

Integrating the x component of the momentum balance equation over $K^x_{i-\frac{1}{2},j}$, we get for the diffusion term :

$$\bar{T}^{\mathrm{dif}}_{i-\frac{1}{2},j} = -\Big[\int_{K^x_{i-\frac{1}{2},j}} \mathrm{div}\big[\boldsymbol{\tau}(\boldsymbol{u})\big]\,\mathrm{d}\boldsymbol{x}\Big] \cdot \boldsymbol{e}^{(x)} = -\Big[\int_{\partial K^x_{i-\frac{1}{2},j}} \boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{n}\,\mathrm{d}\gamma\Big] \cdot \boldsymbol{e}^{(x)}, \tag{III.36}$$

where $\boldsymbol{e}^{(x)}$ stands for the first vector of the canonical basis of $\mathbb{R}^2$. We denote by $\sigma^x_{i,j}$ the right face of $K^x_{i-\frac{1}{2},j}$, *i.e.* $\sigma^x_{i,j} = \{x_i\} \times (y_{j-\frac{1}{2}},\, y_{j+\frac{1}{2}})$. Splitting the boundary integral in (III.36), the part of $\bar{T}^{\mathrm{dif}}_{i-\frac{1}{2},j}$ associated to $\sigma^x_{i,j}$, also referred to as the viscous flux through $\sigma^x_{i,j}$, reads :

$$-\Big[\int_{\sigma^x_{i,j}} \boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{n}\,\mathrm{d}\gamma\Big] \cdot \boldsymbol{e}^{(x)} = -2\int_{\sigma^x_{i,j}} \mu\ \partial_x \boldsymbol{u}^x\,\mathrm{d}\gamma + \frac{2}{3}\int_{\sigma^x_{i,j}} \mu\ (\partial_x \boldsymbol{u}^x + \partial_y \boldsymbol{u}^y)\,\mathrm{d}\gamma,$$

and the usual finite difference technique yields the following approximation for this term :

$$-\frac{4}{3}\int_{\sigma^x_{i,j}} \mu\ \partial_x \boldsymbol{u}^x\,\mathrm{d}\gamma + \frac{2}{3}\int_{\sigma^x_{i,j}} \mu\ \partial_y \boldsymbol{u}^y\,\mathrm{d}\gamma$$
$$\approx -\frac{4}{3}\,\mu_{i,j}\ \frac{h^y_j}{h^x_i}\ (\boldsymbol{u}^x_{i+\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j}) + \frac{2}{3}\,\mu_{i,j}\ \frac{h^y_j}{h^y_j}\ (\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j-\frac{1}{2}}), \tag{III.37}$$

where $\mu_{i,j}$ is an approximation of the viscosity at the face $\sigma^x_{i,j}$. Similarly, let $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}} = (x_{i-1},\, x_i)\times\{y_{j+\frac{1}{2}}\}$

be the top edge of the cell. Then :

$$
\begin{aligned}
-\Big[\int_{\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}} \boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{n}\,\mathrm{d}\gamma\Big]\cdot \boldsymbol{e}^{(x)} &= -\int_{\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}} \mu\ (\partial_y \boldsymbol{u}^x + \partial_x \boldsymbol{u}^y)\,\mathrm{d}\gamma \\
&\approx -\mu_{i-\frac{1}{2},j+\frac{1}{2}}\ \Big[\frac{h^x_{i-\frac{1}{2}}}{h^y_{j+\frac{1}{2}}}\ (\boldsymbol{u}^x_{i-\frac{1}{2},j+1} - \boldsymbol{u}^x_{i-\frac{1}{2},j}) + \frac{h^x_{i-\frac{1}{2}}}{h^x_{i-\frac{1}{2}}}\ (\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i-1,j+\frac{1}{2}})\Big],
\end{aligned}
$$

where $\mu_{i-\frac{1}{2},j+\frac{1}{2}}$ stands for an approximation of the viscosity at the edge $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$.

Let us now multiply each discrete equation for $\boldsymbol{u}^x$ by the corresponding degree of freedom of a velocity field $\boldsymbol{v}$ (*i.e.* the balance over $K^x_{i-\frac{1}{2},j}$ by $\boldsymbol{v}^x_{i-\frac{1}{2},j}$) and sum over i and j. The viscous flux at the face $\sigma^x_{i,j}$ appears twice in the sum, once multiplied by $\boldsymbol{v}^x_{i-\frac{1}{2},j}$ and the second one by $-\boldsymbol{v}^x_{i+\frac{1}{2},j}$, and the corresponding term reads :

$$
\begin{aligned}
T^{\mathrm{dis}}_{i,j}(\boldsymbol{u},\boldsymbol{v}) &= \mu_{i,j}\ \Big[-\frac{4}{3}\ \frac{h^y_j}{h^x_i}\ (\boldsymbol{u}^x_{i+\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j}) + \frac{2}{3}\ \frac{h^y_j}{h^y_j}\ (\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j-\frac{1}{2}})\Big]\ (\boldsymbol{v}^x_{i-\frac{1}{2},j} - \boldsymbol{v}^x_{i+\frac{1}{2},j}) \\
&= \mu_{i,j}\ h^y_j h^x_i\ \Big[\frac{4}{3}\ \frac{\boldsymbol{u}^x_{i+\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j}}{h^x_i} - \frac{2}{3}\ \frac{\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j-\frac{1}{2}}}{h^y_j}\Big]\ \frac{\boldsymbol{v}^x_{i+\frac{1}{2},j} - \boldsymbol{v}^x_{i-\frac{1}{2},j}}{h^x_i}. \qquad \text{(III.38)}
\end{aligned}
$$

Similarly, the term associated to $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ appears multiplied by $\boldsymbol{v}^x_{i-\frac{1}{2},j}$ and by $-\boldsymbol{v}^x_{i-\frac{1}{2},j+1}$, and we get :

$$
\begin{aligned}
T^{\mathrm{dis}}_{i-\frac{1}{2},j+\frac{1}{2}}(\boldsymbol{u},\boldsymbol{v}) = \mu_{i-\frac{1}{2},j+\frac{1}{2}}\ & h^x_{i-\frac{1}{2}} h^y_{j+\frac{1}{2}} \\
& \Big[\frac{\boldsymbol{u}^x_{i-\frac{1}{2},j+1} - \boldsymbol{u}^x_{i-\frac{1}{2},j}}{h^y_{j+\frac{1}{2}}} + \frac{\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i-1,j+\frac{1}{2}}}{h^x_{i-\frac{1}{2}}}\Big]\ \frac{\boldsymbol{v}^x_{i-\frac{1}{2},j+1} - \boldsymbol{v}^x_{i-\frac{1}{2},j}}{h^y_{j+\frac{1}{2}}}. \qquad \text{(III.39)}
\end{aligned}
$$

Let us now define the discrete gradient of the velocity as follows :

- The derivatives involved in the divergence, $\partial^{\mathcal{M}}_x \boldsymbol{u}^x$ and $\partial^{\mathcal{M}}_y \boldsymbol{u}^y$, are defined over the primal cells by :

$$
\partial^{\mathcal{M}}_x \boldsymbol{u}^x(\boldsymbol{x}) = \frac{\boldsymbol{u}^x_{i+\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j}}{h^x_i},\quad \partial^{\mathcal{M}}_y \boldsymbol{u}^y(\boldsymbol{x}) = \frac{\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j-\frac{1}{2}}}{h^y_j},\quad \forall \boldsymbol{x} \in K_{i,j}. \qquad \text{(III.40)}
$$

- For the other derivatives, we introduce another mesh which is vertex-centred, and we denote by K^{xy} the generic cell of this new mesh, with $K^{xy}_{i+\frac{1}{2},j+\frac{1}{2}} = (x_i, x_{i+1}) \times (y_j, y_{j+1})$. Then, $\forall \boldsymbol{x} \in K^{xy}_{i+\frac{1}{2},j+\frac{1}{2}}$:

$$
\partial^{\mathcal{M}}_y \boldsymbol{u}^x(\boldsymbol{x}) = \frac{\boldsymbol{u}^x_{i+\frac{1}{2},j+1} - \boldsymbol{u}^x_{i+\frac{1}{2},j}}{h^y_{j+\frac{1}{2}}},\quad \partial^{\mathcal{M}}_x \boldsymbol{u}^y(\boldsymbol{x}) = \frac{\boldsymbol{u}^y_{i+1,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j+\frac{1}{2}}}{h^x_{i+\frac{1}{2}}}. \qquad \text{(III.41)}
$$

With this definition, we get :

$$
T^{\mathrm{dis}}_{i,j}(\boldsymbol{u},\boldsymbol{v}) = \mu_{i,j} \int_{K_{i,j}} \Big[\frac{4}{3}\partial^{\mathcal{M}}_x \boldsymbol{u}^x - \frac{2}{3}\partial^{\mathcal{M}}_y \boldsymbol{u}^y\Big]\partial^{\mathcal{M}}_x \boldsymbol{v}^x\,\mathrm{d}\boldsymbol{x},
$$

and :

$$
T^{\mathrm{dis}}_{i-\frac{1}{2},j+\frac{1}{2}}(\boldsymbol{u},\boldsymbol{v}) = \mu_{i-\frac{1}{2},j+\frac{1}{2}} \int_{K^{xy}_{i-\frac{1}{2},j+\frac{1}{2}}} (\partial^{\mathcal{M}}_y \boldsymbol{u}^x + \partial^{\mathcal{M}}_x \boldsymbol{u}^y)\ \partial^{\mathcal{M}}_y \boldsymbol{v}^x\,\mathrm{d}\boldsymbol{x}.
$$

Let us now perform the same operations for the y-component of the velocity. Doing so, we are lead to introduce an approximation of the viscosity at the edge $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}} = \{x_{i-\frac{1}{2}}\} \times (y_j,\, y_{j+1})$ (see Figure III.3).

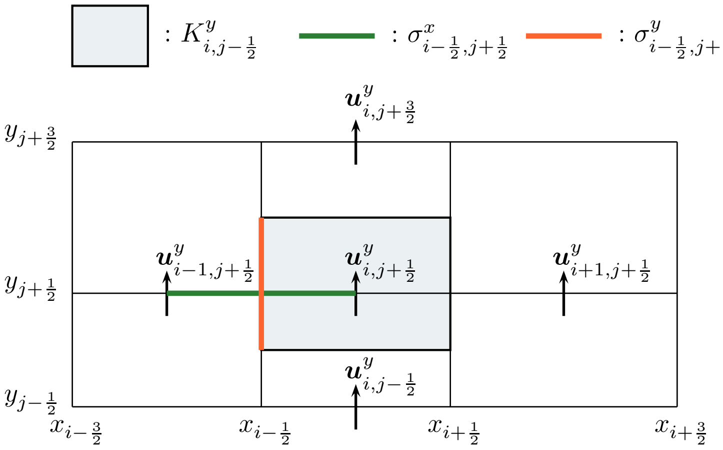
$: K^y_{i,j-\frac{1}{2}}$
$: \sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$
$: \sigma^y_{i-\frac{1}{2},j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i,j+\frac{3}{2}}$
$y_{j+\frac{3}{2}}$
$\boldsymbol{u}^y_{i-1,j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i,j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i+1,j+\frac{1}{2}}$
$y_{j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i,j-\frac{1}{2}}$
$y_{j-\frac{1}{2}}$
$x_{i-\frac{3}{2}}$
$x_{i-\frac{1}{2}}$
$x_{i+\frac{1}{2}}$
$x_{i+\frac{3}{2}}$

FIG. III.3 – Dual cell for the y-component of the velocity

Let us suppose that we take the same approximation as on $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$. Then, the same argument yields that multiplying each discrete equation for $\boldsymbol{u}^x$ and for $\boldsymbol{u}^y$ by the corresponding degree of freedom of a velocity field $\boldsymbol{v}$, we obtain a dissipation term which reads :

$$T^{\text{dis}}(\boldsymbol{u},\boldsymbol{v}) = \int_\Omega \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \boldsymbol{\nabla}^{\mathcal{M}} \boldsymbol{v} \,\mathrm{d}\boldsymbol{x}, \tag{III.42}$$

where $\boldsymbol{\nabla}^{\mathcal{M}}$ is the discrete gradient defined by (III.40)-(III.41) and $\boldsymbol{\tau}^{\mathcal{M}}$ the discrete tensor :

$$\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) = \begin{bmatrix} 2\mu\, \partial^{\mathcal{M}}_x \boldsymbol{u}_x & \mu^{xy}\,(\partial^{\mathcal{M}}_y \boldsymbol{u}_x + \partial^{\mathcal{M}}_x \boldsymbol{u}_y) \\ \mu^{xy}\,(\partial^{\mathcal{M}}_y \boldsymbol{u}_x + \partial^{\mathcal{M}}_x \boldsymbol{u}_y) & 2\mu\, \partial^{\mathcal{M}}_y \boldsymbol{u}_y \end{bmatrix} - \frac{2}{3}\,\mu\,(\partial^{\mathcal{M}}_x \boldsymbol{u}_x + \partial^{\mathcal{M}}_y \boldsymbol{u}_y)\, I, \tag{III.43}$$

where μ is the viscosity defined on the primal mesh by $\mu(\boldsymbol{x}) = \mu_{i,j},\ \forall \boldsymbol{x} \in K_{i,j}$ and μ^{xy} is the viscosity defined on the vertex-centred mesh, by $\mu(\boldsymbol{x}) = \mu_{i+\frac{1}{2},j+\frac{1}{2}},\ \forall \boldsymbol{x} \in K^{xy}_{i+\frac{1}{2},j+\frac{1}{2}}$.

Now the form (III.42) suggests a natural to discretize the viscous dissipation term in the internal energy balance in order for the consistency property (ii) to hold. Indeed, if we simply set on each primal cell $K_{i,j}$:

$$(\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u})_{i,j} = \frac{1}{|K_{i,j}|} \int_{K_{i,j}} \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \boldsymbol{\nabla}^{\mathcal{M}} \boldsymbol{u} \,\mathrm{d}\boldsymbol{x}, \tag{III.44}$$

then, thanks to (III.42), the property (ii) which reads :

$$T^{\text{dis}}(\boldsymbol{u},\boldsymbol{u}) = \sum_{i,j} |K_{i,j}|\ (\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u})_{i,j}.$$

holds. Furthermore, we get from Definition (III.43) that $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})(\boldsymbol{x})$ is a symmetrical tensor, for any i,j and $\boldsymbol{x} \in K_{i,j}$, and therefore an elementary algebraic argument yields :

$$\begin{aligned}(\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u})_{i,j} &= \frac{1}{|K_{i,j}|} \int_{K_{i,j}} \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \boldsymbol{\nabla}^{\mathcal{M}} \boldsymbol{u} \,\mathrm{d}\boldsymbol{x} \\ &= \frac{1}{2\,|K_{i,j}|} \int_{K_{i,j}} \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \left[\boldsymbol{\nabla}^{\mathcal{M}} \boldsymbol{u} + (\boldsymbol{\nabla}^{\mathcal{M}} \boldsymbol{u})^t\right] \mathrm{d}\boldsymbol{x} \geq 0.\end{aligned}$$

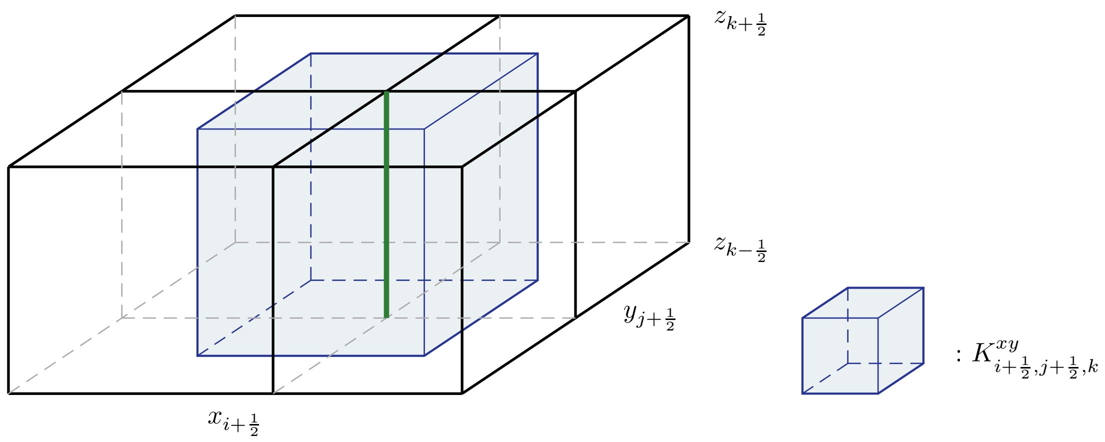

FIG. III.4 – The xy-staggered cell $K^{xy}_{i+\frac{1}{2},j+\frac{1}{2},k}$, used in the definition of $\partial_y^{\mathcal{M}}\boldsymbol{u}^x$, $\partial_x^{\mathcal{M}}\boldsymbol{u}^y$, and $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})_{x,y} = \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})_{y,x}$.

Remark 4 (Approximation of the viscosity)
Note that, for the symmetry of $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})$ to hold, the choice of the same viscosity at the edges $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ and $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}}$ is crucial even though other choices may appear natural. Assuming for instance the viscosity to be a function of an additional variable defined on the primal mesh, the following construction seems reasonable :

1. define a constant value for μ on each primal cell,
2. associate a value of μ to the primal edges, by taking the average between the value at the adjacent cells,
3. finally, split the integral of the shear stress over $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ in two parts, one for the part included in the (top) boundary of $K_{i-1,j}$ and the second one in the boundary of $K_{i,j}$.

Then the viscosities on $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ and $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}}$ coincide only for uniform meshes, and, in the general case, the symmetry of $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})$ is lost.

III.5.2 Extension to the three-dimensional case

Extending the computations of the preceding section to three space dimensions yields the following construction.

- First, define three new meshes, which are "edge-centred" : $K^{xy}_{i+\frac{1}{2},j+\frac{1}{2},k} = (x_i, x_{i+1}) \times (y_i, y_{j+1}) \times (z_{k-\frac{1}{2}}, z_{k+\frac{1}{2}})$ is staggered from the primal mesh $K_{i,j,k}$ in the x and y direction (see Figure III.4), $K^{xz}_{i+\frac{1}{2},j,k+\frac{1}{2}}$ in the x and z direction, and $K^{yz}_{i,j+\frac{1}{2},k+\frac{1}{2}}$ in the y and z direction.
- The partial derivatives of the velocity components are then defined as piecewise constant functions, the value of which is obtained by natural finite differences :
  - for $\partial_x^{\mathcal{M}}\boldsymbol{u}^x$, $\partial_y^{\mathcal{M}}\boldsymbol{u}^y$ and $\partial_z^{\mathcal{M}}\boldsymbol{u}^z$, on the primal mesh,
  - for $\partial_y^{\mathcal{M}}\boldsymbol{u}^x$ and $\partial_x^{\mathcal{M}}\boldsymbol{u}^y$ on the cells $(K^{xy}_{i+\frac{1}{2},j+\frac{1}{2},k})$,

  - for $\partial_z^{\mathcal{M}} \boldsymbol{u}^x$ and $\partial_x^{\mathcal{M}} \boldsymbol{u}^z$ on the cells $(K^{xz}_{i+\frac{1}{2},j,k+\frac{1}{2}})$,
  - for $\partial_y^{\mathcal{M}} \boldsymbol{u}^z$ and $\partial_z^{\mathcal{M}} \boldsymbol{u}^y$ on the cells $(K^{yz}_{i,j+\frac{1}{2},k+\frac{1}{2}})$.
- Then, define four families of values for the viscosity field, μ, μ^{xy}, μ^{xz} and μ^{yz}, associated to the primal and the three edge-centred meshes respectively.
- The shear stress tensor is obtained by the extension of (III.43) to $d = 3$.
- And, finally, the dissipation term is given by (III.44).

CHAPITRE

IV Consistent staggered schemes for compressible flows – Part II : Euler equations.

In this paper, we propose an implicit scheme and a pressure correction scheme for the Euler equations, based on space discretizations of staggered type : MAC scheme or low-order (Rannacher-Turek or Crouzeix-Raviart) finite elements. Both schemes rely on the discretization of the internal energy balance equation, which offers two main advantages : first, we avoid the space discretization of the total energy, the expression of which involves cell-centered and face-centered variables ; second, we obtain algorithms which boil down to usual schemes in the incompressible limit. To obtain correct weak solutions (in particular, with shocks satisfying the Rankine-Hugoniot conditions), we need to introduce a source term in the internal energy balance, which we build as follows. We first derive a discrete kinetic energy balance. This relation involves source terms, which are then, in some way, compensated in the internal energy balance. Since the kinetic and internal energy equation are associated to the primal and dual mesh respectively, they cannot be summed to obtain a total energy balance. However, we theoretically prove, in the 1D case, that, if the scheme converges, the limit indeed satisfies a weak form of this latter equation. Finally, we present numerical results which confort this theory.

Plan du Chapitre IV

IV.1 Introduction **111**

IV.2 Meshes and unknowns **112**

IV.3 An implicit scheme **113**

IV.3.1 The scheme 113

IV.3.2 The discrete kinetic energy balance equation and the corrective source terms . 116

IV.3.3 Passing to the limit in the scheme 119

IV.4 A pressure correction scheme **125**

IV.4.1 The scheme 125

IV.4.2 The discrete kinetic energy balance equation and the corrective source terms . 126

IV.4.3 Passing to the limit in the scheme 128

IV.5 Numerical tests **130**

IV.1 Introduction

Let us consider the compressible Navier-Stokes equations, which reads :

$$\partial_t\rho + \operatorname{div}(\rho\,\boldsymbol{u}) = 0, \tag{IV.1a}$$

$$\partial_t(\rho\,\boldsymbol{u}) + \operatorname{div}(\rho\,\boldsymbol{u}\otimes\boldsymbol{u}) + \boldsymbol{\nabla}p - \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})) = 0, \tag{IV.1b}$$

$$\partial_t(\rho\,E) + \operatorname{div}(\rho\,E\,\boldsymbol{u}) + \operatorname{div}(p\,\boldsymbol{u}) = \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u})\cdot\boldsymbol{u}), \tag{IV.1c}$$

$$p = (\gamma-1)\,\rho\,e, \qquad E = \frac{1}{2}|\boldsymbol{u}|^2 + e, \tag{IV.1d}$$

where t stands for the time, ρ, $\boldsymbol{u}$, p, E and e are the density, velocity, pressure, total energy and internal energy in the flow, $\boldsymbol{\tau}(\boldsymbol{u})$ stands for the shear stress tensor, and $\gamma > 1$ is a coefficient specific to the considered fluid. The problem is supposed to be posed over $\Omega\times(0,T)$, where Ω is a open bounded connected subset of $\mathbb{R}^d$, $d\leq 3$ and $(0,T)$ is a finite time interval. This system must be complemented by suitable boundary conditions, and initial conditions for ρ, e and $\boldsymbol{u}$, which are positive for ρ and e. The closure relation for $\boldsymbol{\tau}(\boldsymbol{u})$ is assumed to be :

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

where μ stand for a (possibly depending on $\boldsymbol{x}$) non-negative parameter.

We suppose, for the sake of simplicity, that $\boldsymbol{u}$ is prescribed to zero on the whole boundary $\partial\Omega$.

Let us suppose that the solution is regular. Taking the inner product of the momentum balance equation (IV.1b) by $\boldsymbol{u}$ and using the mass balance equation, we obtain the so-called the kinetic energy balance equation :

$$\frac{1}{2}\partial_t(\rho\,|\boldsymbol{u}|^2) + \frac{1}{2}\operatorname{div}(\rho\,|\boldsymbol{u}|^2\boldsymbol{u}) + \boldsymbol{\nabla}p\cdot\boldsymbol{u} = \operatorname{div}(\boldsymbol{\tau}(\boldsymbol{u}))\cdot\boldsymbol{u}. \tag{IV.3}$$

Subtracting this relation from the total energy balance, we obtain the internal energy balance equation :

$$\partial_t(\rho e) + \operatorname{div}(\rho e\boldsymbol{u}) + p\operatorname{div}(\boldsymbol{u}) = \boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u}. \tag{IV.4}$$

Since,

- (i) from Equation (IV.2) (and from thermodynamical arguments), $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u} \geq 0$,
- (ii) thanks to the mass balance equation, the first two terms in the left-hand side of (IV.4) may be recast as a transport operator : $\partial_t(\rho e) + \operatorname{div}(\rho e\boldsymbol{u}) = \rho\,[\partial_t e + \boldsymbol{u}\cdot\boldsymbol{\nabla}e]$,
- (iii) and, finally, because, from the equation of state, the pressure vanishes when $e = 0$,

this equation implies that e remains non-negative at all times.

The aim of this paper is to build a numerical scheme for the Euler equations (*i.e.* System (IV.1) with $\mu = 0$) based on staggered space discretizations, the motivation for this choice being that we would like to obtain a scheme taht is stable and accurate at all Mach numbers, and, in particular, boils down to a usual scheme for incompressible flows (or, more generally, for the asymptotic model of vanishing Mach number flows [54]) when the Mach number tends to zero. In incompressible models, the natural energy balance equation is the internal energy equation (IV.4). In addition, discretizing (IV.4) instead of the total energy balance (IV.1c) presents two advantages :

- first, it avoids the space discretization of the total energy, which is rather unatural for staggered schemes since the degrees of freedom for the velocity and the scalar variables are not colocated,
- second, a suitable discretization of (IV.4) may yield, "by construction" of the scheme, the positivity of the internal energy.

However, for solutions with shocks, Equation (IV.4) is not equivalent to (IV.1c) ; more precisely speaking, one can show that, at the locations of shocks, positive measures should replace $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla} \boldsymbol{u}$ (which formally vanishes since $\mu = 0$) at the right-hand side of Equation (IV.4). Discretizing (IV.4) instead of (IV.1c) may thus yield a scheme which does not compute the correct weak discontinuous solutions, the manifestation of this non-consistency being that the numerical solutions present shocks which are not consistent with the Rankine-Hugoniot conditions associated to (IV.1c). The essential result of this paper is to provide solutions to circumvent this problem. To this purpose, we closely mimick the formal computation performed above :

- Starting from the discrete momentum balance equation, with an *ad hoc* discretization of the convection operator, we derive a discrete kinetic energy balance ; residual terms are present in this relation, which do no tend to zero with space and time step (they are the discrete manifestations of the the above mentionned measures).
- These residual terms are then compensated by source terms in the internal energy balance.

We provide a theoretical justification of this process by showing that, in the 1D case, if the scheme is stable enough and converges to a limit (in a sense to be defined), this limit satisfies a weak form of (IV.1c) which implies the correct Rankine-Hugoniot conditions. Then, we perform numerical tests which substantiate this analysis. Two different time discretizations are proposed : first, a fully implicit scheme (a solution to which may be rather difficult to obtain in practice) and, second, a pressure correction scheme (the algorithm which is indeed used in the tests presented here, and in the industrial open-source code ISIS [40], developed at IRSN on the basis of the software components library PELICANS [63]). Let us mention also that fully explicit versions may be built, and are now under study [59].

This paper is organized as follows. We begin by describing the space discretizations (Section IV.2). We then study the implicit scheme (Section IV.3) : we first give the general form of the algorithm (Section IV.3.1), then derive the kinetic energy balance and deduce the source terms to be included in the internal energy balance (Section IV.3.2), and, finally, we pass to the limit in the scheme to prove (in 1D) the consistency of the scheme (Section IV.3.3). Section IV.4 follows the same lines for the pressure correction scheme. Finally, we present some numerical tests in Section IV.5.

IV.2 Meshes and unknowns

Let $\mathcal{M}$ be a decomposition of the domain Ω, supposed to be regular in the usual sense of the finite element literature (*eg.* [9]). The cells may be :

- for a general domain Ω, either convex quadrilaterals ($d = 2$) or hexahedra ($d = 3$) or simplices, both type of discretizations being possibly combined in a same mesh,
- for a domain the boundaries of which are hyperplanes normal to a coordinate axis, rectangles ($d = 2$) or rectangular parallelepipeds ($d = 3$) (the faces of which, of course, are then also necessarily

normal to a coordinate axis).

By $\mathcal{E}$ and $\mathcal{E}(K)$ we denote the set of all $(d-1)$-faces σ of the mesh and of the element $K \in \mathcal{M}$ respectively. The set of edges included in the boundary of Ω is denoted by $\mathcal{E}_{\text{ext}}$ and the set of internal ones (*i.e.* $\mathcal{E} \backslash \mathcal{E}_{\text{ext}}$) is denoted by $\mathcal{E}_{\text{int}}$; a face $\sigma \in \mathcal{E}_{\text{int}}$ separating the cells K and L is denoted by $\sigma = K|L$. The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-measure of the face σ. For $1 \leq i \leq d$, we denote by $\mathcal{E}^{(i)} \subset \mathcal{E}$ the subset of the faces of $\mathcal{E}$ which are perpendicular to the i^{th} unit vector of the canonical basis of $\mathbb{R}^d$.

The space discretization is staggered, using either the Marker-And Cell (MAC) scheme [37, 36], or non-conforming low-order finite element approximations, namely the Rannacher and Turek element (RT) [65] for quadrilateral or hexahedric meshes, or the Crouzeix-Raviart (CR) element [11] for simplicial meshes.

For all these space discretizations, the degrees of freedom for the pressure, the density and the internal energy are associated to the cells of the mesh $\mathcal{M}$, and are denoted by :

$$\big\{p_K,\ \rho_K,\ e_K,\ K \in \mathcal{M}\big\}.$$

Let us then turn to the degrees of freedom for the velocity.

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – The degrees of freedom for the velocities are located at the center of the faces of the mesh, and we choose the version of the element where they represent the average of the velocity through a face. The set of degrees of freedom reads :

$$\{u_{\sigma,i},\ \sigma \in \mathcal{E},\ 1 \leq i \leq d\}.$$

- **MAC** discretization – The degrees of freedom for the i^{th} component of the velocity, defined at the centres of the face $\sigma \in \mathcal{E}^{(i)}$, are denoted by :

$$\big\{u_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)},\ 1 \leq i \leq d\big\}.$$

IV.3 An implicit scheme

IV.3.1 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 $n = 0, 1, \ldots, N-1$ be the constant time step. We consider an implicit-in-time numerical scheme, which reads in its fully discrete form :

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} = 0, \tag{IV.5a}$$

$$\text{For } 1 \le i \le d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1} u_{\sigma,i}^{n+1} - \rho_\sigma^n u_{\sigma,i}^n) + \sum_{\varepsilon \in \tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1} u_{\varepsilon,i}^{n+1} + |D_\sigma| \, (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = 0, \tag{IV.5b}$$

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} e_\sigma^{n+1} + |K| \, p_K^{n+1} \, (\mathrm{div}(\boldsymbol{u}^{n+1}))_K = S_K^{n+1}, \tag{IV.5c}$$

$$\forall K \in \mathcal{M}, \qquad p_K^{n+1} = (\gamma - 1) \rho_K^{n+1} \, e_K^{n+1}. \tag{IV.5d}$$

Equation (IV.5a) is obtained by the discretization of the mass balance over the primal mesh, and $F_{K,\sigma}^{n+1}$ stands for the mass flux across σ outward K, given by :

$$\forall \sigma \in \mathcal{E}_{\mathrm{int}}, \, \sigma = K|L, \qquad F_{K,\sigma}^{n+1} = |\sigma| \; \tilde{\rho}_\sigma^{n+1} \; \boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}.$$

In this relation, the notation $\boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}$ stands the approximation of the normal velocity to the face σ outward K. For the MAC discretization, this quantity is given (up, possibly, to a change of sign) by the velocity degree of freedom located at the face ; for the RT and CR discretizations, it is computed by taking the inner product of the (vector valued) velocity on σ, $\boldsymbol{u}_\sigma^{n+1}$, and the outward normal vector $\boldsymbol{n}_{K,\sigma}$ (*i.e.* doing exactly what the notation says). The density at the face $\sigma = K|L$, $\tilde{\rho}_\sigma^{n+1}$, is approximated by the upwind technique :

$$\tilde{\rho}_\sigma^{n+1} = \left| \begin{array}{ll} \rho_K^{n+1} & \text{if } \boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma} \ge 0, \\ \rho_L^{n+1} & \text{otherwise.} \end{array} \right.$$

We now turn to the discrete momentum balance (IV.5b). For the MAC discretization, but also for the RT and CR discretization, the time derivative and convection terms are approximated in (IV.5b) by a finite volume technique over a dual mesh, which we now define :

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – For the RT or CR discretization, the dual mesh is the same for all the velocity components. When $K \in \mathcal{M}$ is a simplex, a rectangles or a cuboid, for $\sigma \in \mathcal{E}(K)$, we define $D_{K,\sigma}$ as the cone with basis σ and with vertex the mass center of K. We thus obtain a partition of K in m sub-volumes, where m is the number of faces of the mesh, each sub-volume having the same measure $|D_{K,\sigma}| = |K|/m$. We extend this definition to general quadrangles and hexahedra, with a partition still of equal-volume sub-cells, and with the same connectivities ; note that this is of course always possible, but that such a volume $D_{K,\sigma}$ may be no longer a cone : indeed, if K is far from a pallelogram, it may not be possible to built a cone having σ as basis, the opposite vertex lying in K and a volume equal to $|K|/m$. The volume $D_{K,\sigma}$ is refered to as the half-diamond cell associated to K and σ.

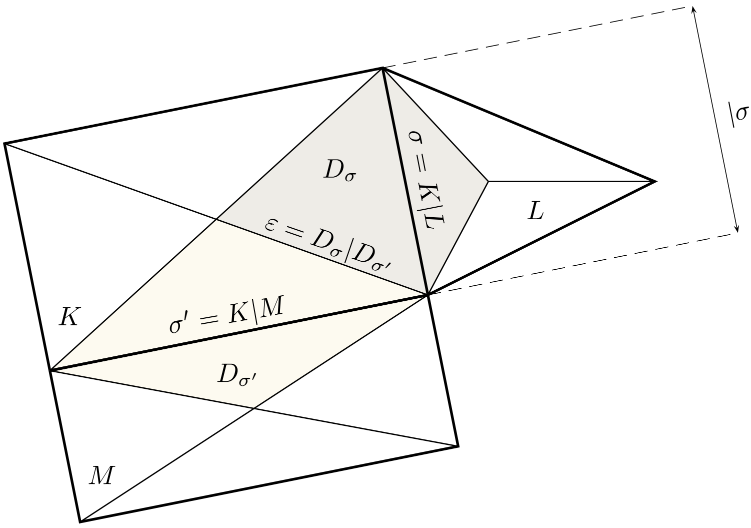

FIG. IV.1 – Primal and dual meshes for the Rannacher-Turek and Crouzeix-Raviart elements.

For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the diamond cell D_σ associated to σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external face $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}(K)$, D_σ is just the same volume as $D_{K,\sigma}$.

- **MAC** discretization – For the MAC scheme, the dual mesh depends on the component of the velocity. For each of them, its definition differs from the RT or CR dual mesh only by the choice of the half-diamond cell, which, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, is now the rectangle of basis σ and of measure $|D_{K,\sigma}|$ equal to half the measure of K.

We denote by $|D_\sigma|$ the measure of the dual cell $|D_\sigma|$, and by $\varepsilon = D_\sigma|D_{\sigma'}$ the face separating two diamond cells D_σ and $D_{\sigma'}$ (see Figure IV.1).

To make the discretization of the time derivative term complete, we must provide a definition for the ρ_σ^{n+1} and ρ_σ^n, which approximate the density on the edge σ at time t^{n+1} and t^n respectively. They are given by the following weighted average :

$$\forall \sigma \in \mathcal{E}_{\rm int},\ \sigma = K|L, \qquad |D_\sigma|\ \rho_\sigma^n = |D_{K,\sigma}|\ \rho_K^n + |D_{L,\sigma}|\ \rho_L^n. \tag{IV.6}$$

We now turn to the convection term. The first task is to define the discrete mass flux through the dual edge ε outward D_σ, denoted by $F_{\sigma,\varepsilon}^{n+1}$, the guideline for its construction being that we need a finite volume discretization of the mass balance equation over the diamond cells to hold :

$$\forall \sigma \in \mathcal{E}, \qquad |D_\sigma| \frac{\rho_\sigma^{n+1} - \rho_\sigma^n}{\delta t} + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon}^{n+1} = 0, \tag{IV.7}$$

in order to be be able to derive a discrete kinetic energy balance (see Section IV.3.2 below). For a dual edge ε included in the primal cell K, this flux is computed as a linear combination (with constant coefficients, *i.e.* independent of the edge and the cell) of the mass fluxes through the faces of K, *i.e.* the quantities $(F_{K,\sigma}^{n+1})_{\sigma \in \mathcal{E}(K)}$ appearing in the discrete mass balance (IV.5a). We do not give here this set of coefficients, and refer to [1, 38, 25] for a detailed construction of this approximation.

The quantity $u_{\varepsilon,i}^{n+1}$ stands for an approximation of u_i^{n+1} on ε wich may be chosen centered or upwind, so, for $\varepsilon = D_\sigma|D_{\sigma'}$, reads :

$$\text{Centered case : } u_{\varepsilon,i}^{n+1} = (u_{\sigma,i}^{n+1} + u_{\sigma',i}^{n+1})/2. \qquad \text{Upwind case : } u_{\varepsilon,i}^{n+1} = \left| \begin{array}{ll} u_{\sigma,i}^{n+1} & \text{if } F_{\sigma,\varepsilon}^{n+1} \geq 0, \\ u_{\sigma',i}^{n+1} & \text{otherwise.} \end{array} \right.$$

The last term $(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}$ stands for the i-th component of the discrete pressure gradient at the face σ, which reads :

$$\text{for } \sigma \in \mathcal{E}_{\rm int},\ \sigma = K|L, \qquad (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = \frac{|\sigma|}{|D_\sigma|}\ (p_L^{n+1} - p_K^{n+1})\ \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}^{(i)}.$$

Finally, Equation (IV.5c) is a approximation of the internal balance over the primal mesh K. To ensure the positivity of the convection operator [50], we use an upwinding technique for this term :

$$e_\sigma^{n+1} = \left| \begin{array}{ll} e_K^{n+1} & \text{if } F_{K,\sigma}^{n+1} \geq 0, \\ e_L^{n+1} & \text{otherwise.} \end{array} \right.$$

The divergence of the velocity, $\bigl(\mathrm{div}(\boldsymbol{u}^{n+1})\bigr)_K$, is discretized ar follows :

$$\bigl(\mathrm{div}(\boldsymbol{u}^{n+1})\bigr)_K = \frac{1}{|K|} \sum_{\sigma \in \mathcal{E}(K)} |\sigma|\ \boldsymbol{u}_\sigma^{n+1} \cdot \boldsymbol{n}_{K,\sigma}.$$

Note that this definition implies that the discrete gradient and divergence operators are dual with respect to the L^2 inner product :

$$\sum_{K \in \mathcal{M}} |K|\ p_K\ \bigl(\mathrm{div}(\boldsymbol{u})\bigr)_K + \sum_{\mathcal{E},i} |D_\sigma|\ u_{\sigma,i}\ (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = 0,$$

where the notation $\sum_{\mathcal{E},i}$ means that the summation is performed for $1 \leq i \leq d$ and, for a given index i, on $\sigma \in \mathcal{E}^{(i)}$ for the MAC scheme and on $\sigma \in \mathcal{E}$ for the RT or CR discretization. The right-hand side, S_K^{n+1}, is derived using consistency arguments in the next section.

IV.3.2 The discrete kinetic energy balance equation and the corrective source terms

Let $\delta^{\rm up}$ be a coefficient defined by $\delta^{\rm up} = 1$ if an upwind discretization is used for the convection term in the momentum balance equation and $\delta^{\rm up} = 0$ in the centered case. With this notation, the momentum balance equation reads :

$$\begin{aligned} \frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1} u_{\sigma,i}^{n+1} - \rho_\sigma^n u_{\sigma,i}^n) + \sum_{\varepsilon = D_\sigma|D_{\sigma'}} \frac{1}{2}\ F_{\sigma,\varepsilon}^{n+1}\ (u_{\sigma,i}^{n+1} + u_{\sigma',i}^{n+1}) \\ + \delta^{\rm up} \sum_{\varepsilon = D_\sigma|D_{\sigma'}} \frac{1}{2}\ |F_{\sigma,\varepsilon}^{n+1}|\ (u_{\sigma,i}^{n+1} - u_{\sigma',i}^{n+1}) + |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = 0. \end{aligned} \tag{IV.8}$$

We begin with deriving a discrete kinetic energy balance equation. To this purpose, we multiply equation (IV.8) by the corresponding velocity unknown $u_{\sigma,i}^{n+1}$, which yields $T_{\sigma,i}^{\rm conv}+T_{\sigma,i}^{\rm up}+T_{\sigma,i}^{\nabla}=0$, with :

$$\begin{aligned}
T_{\sigma,i}^{\rm conv} &= \Bigl[\frac{|D_\sigma|}{\delta t}(\rho_\sigma^{n+1}u_{\sigma,i}^{n+1}-\rho_\sigma^n u_{\sigma,i}^n)+\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,F_{\sigma,\varepsilon}^{n+1}\,(u_{\sigma,i}^{n+1}+u_{\sigma',i}^{n+1})\Bigr]\,u_{\sigma,i}^{n+1},\\
T_{\sigma,i}^{\rm up} &= \delta^{\rm up}\Bigl[\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(u_{\sigma,i}^{n+1}-u_{\sigma',i}^{n+1})\Bigr]\,u_{\sigma,i}^{n+1},\\
T_{\sigma,i}^{\nabla} &= |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}\,u_{\sigma,i}^{n+1}.
\end{aligned}$$

Lemma II.3.2 of Chapter 2 yields :

$$T_{\sigma,i}^{\rm conv}=\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Bigl[\rho_\sigma^{n+1}(u_{\sigma,i}^{n+1})^2-\rho_\sigma^n(u_{\sigma,i}^n)^2\Bigr]+\frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}}F_{\sigma,\varepsilon}^{n+1}\,u_{\sigma,i}^{n+1}\,u_{\sigma',i}^{n+1}
+\frac{|D_\sigma|}{2\,\delta t}\,\rho_\sigma^n\,\bigl(u_{\sigma,i}^{n+1}-u_{\sigma,i}^n\bigr)^2.$$

Let us define $R_{\sigma,i}^{n+1}$ by the sum of $-T_{\sigma,i}^{\rm up}$ and the opposite of the last term of this equation :

$$R_{\sigma,i}^{n+1}=-\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\,\rho_\sigma^n\,\bigl(u_{\sigma,i}^{n+1}-u_{\sigma,i}^n\bigr)^2-\delta^{\rm up}\Bigl[\sum_{\varepsilon=D_\sigma|D_{\sigma'}}\frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(u_{\sigma,i}^{n+1}-u_{\sigma',i}^{n+1})\Bigr]\,u_{\sigma,i}^{n+1}.$$

With this notation, we thus obtain the following relation :

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Bigl[\rho_\sigma^{n+1}(u_{\sigma,i}^{n+1})^2-\rho_\sigma^n(u_{\sigma,i}^n)^2\Bigr]+\frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}}F_{\sigma,\varepsilon}^{n+1}\,u_{\sigma,i}^{n+1}\,u_{\sigma',i}^{n+1}
+|D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}\,u_{\sigma,i}^{n+1}=R_{\sigma,i}^{n+1}. \qquad \text{(IV.9)}$$

We recognize at the left-hand side a conservative discrete kinetic energy balance. The next step is now to deal with the residual term at the right-hand side. To this purpose, our guideline is to recover a consistent discretization of the total energy balance. The first idea to do this could be just to sum the (discrete) kinetic energy balance with the internal energy balance : we show in [59] that it is indeed possible for a colocated discretization But here, we face the fact that the kinetic energy balance is associated to the dual mesh, while the internal energy balance is discretized on the primal one. The way to circumvent this difficulty is to remark that we do not really need a discrete total energy balance ; in fact, we only need to recover (a weak form of) this equation when the mesh and time steps tend to zero. To this purpose, we choose S_K^{n+1} in such a way to somewhat compensate the terms $(R_{\sigma,i}^{n+1})$:

$\forall K\in\mathcal{M}$,

$$S_K^{n+1}=\frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)}\frac{|D_{K,\sigma}|}{\delta t}\rho_K^n\,\bigl(\boldsymbol{u}_\sigma^{n+1}-\boldsymbol{u}_\sigma^n\bigr)^2+\delta^{\rm up}\sum_{\substack{\varepsilon\cap\bar K\neq\emptyset,\\ \varepsilon=D'_\sigma|D_{\sigma''}}}\alpha_{K,\varepsilon}\,\frac{|F_{\sigma,\varepsilon}^{n+1}|}{2}(\boldsymbol{u}_{\sigma'}^{n+1}-\boldsymbol{u}_{\sigma''}^{n+1})^2. \qquad \text{(IV.10)}$$

The coefficient $\alpha_{K,\varepsilon}$ is fixed to 1 if the face ε is included in K, and this is the only situation to consider for the RT and CR discretization. For the MAC scheme, some dual edges are included in the primal cells, but some lie on their boundary ; for ε being in the latter case, we denote by $\mathcal{N}_\varepsilon$ the set of cells M such that $\bar{M} \cap \varepsilon \neq \emptyset$ (the cardinal of this set being always 4), and compute $\alpha_{K,\varepsilon}$ by :

$$\alpha_{K,\varepsilon} = \frac{|K|}{\sum_{M \in \mathcal{N}_\varepsilon} |M|}.$$

For a uniform grid, this formula yields $\alpha_{K,\varepsilon} = 1/4$.

The expression of the $(S_K^{n+1})_{K\in\mathcal{M}}$ is justified by the passage to the limit in the scheme (for a one-dimensional problem) performed in Section IV.3.3. However, its expression may be anticipated, making the following remarks. First, we note that :

$$\sum_{K\in\mathcal{M}} S_K^{n+1} + \sum_{\mathcal{E},i} R_{\sigma,i}^{n+1} = 0. \tag{IV.11}$$

Indeed, the first part of S_K, thanks to the expression (IV.6) of the density at the face ρ_σ, results from a dispatching of the first part of the residual over the two adjacent cells :

$$-\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\,\rho_\sigma^n\,\bigl(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\bigr)^2 = \underbrace{-\frac{1}{2}\,\frac{|D_{K,\sigma}|}{\delta t}\,\rho_K^n\,\bigl(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\bigr)^2}_{\text{affected to K}} \underbrace{-\frac{1}{2}\,\frac{|D_{L,\sigma}|}{\delta t}\,\rho_L^n\,\bigl(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\bigr)^2}_{\text{affected to L}}.$$

For the second part of the remainder (or of S_K^{n+1}), a standard reordering of the sum yields :

$$\sum_{\mathcal{E},i}\ \sum_{\varepsilon=D_\sigma|D_{\sigma'}} \frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(u_{\sigma,i}^{n+1} - u_{\sigma',i}^{n+1})\Big]\,u_{\sigma,i}^{n+1} = \sum_{\bar{\mathcal{E}},i\ (\varepsilon=D_\sigma|D_{\sigma'})} \frac{1}{2}\,|F_{\sigma,\varepsilon}^{n+1}|\,(u_{\sigma,i}^{n+1} - u_{\sigma',i}^{n+1})^2,$$

where the notation $\sum_{\bar{\mathcal{E}},i\ (\varepsilon=\sigma|\sigma')}$ means that we perform the sum over the components $1 \le i \le d$ and the faces of the dual mesh associated to the component i, and that the dual cells separated by a a generic dual face ε in the summation are denoted by D_σ and $D_{\sigma'}$.

However, we may wonder why we do not use in S_K^{n+1} the expression of this term as it is written in the remainder,*i.e.* , in other words, use the numerical diffusion mutiplied by $\boldsymbol{u}$ instead of the dissipation. A first answer is that we mimick what happens at the continuous level : the term which appears in the kinetic energy balance is $\mathrm{div}\bigl(\boldsymbol{\tau}(\boldsymbol{u})\bigr) \cdot \boldsymbol{u}$ and the corresponding term in the internal energy balance is $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla u}$. A more involved argument is that the expression in S_K^{n+1} provides a positive source term to the internal energy balance, and we may hope that the difference between both expressions tends to zero (because the numerical diffusion tends to zero) in the sense of distributions. To have an intuition of this fact, let us consider the toy elliptic problem, posed over Ω :

$$v - \mu\Delta v = f,$$

where μ is a positive parameter and $f \in \mathrm{L}^2(\Omega)$. Assuming homogeneous Dirichlet boundary conditions, we obtain by standard variational arguments $\|v\|_{\mathrm{L}^2(\Omega)} + \mu^{1/2}\|\boldsymbol{\nabla} v\| \le C$, with C only depending on Ω and f. We thus get, with $\varphi \in \mathrm{C}_c^\infty(\Omega)$:

$$\int_\Omega \Bigl[\mu(\Delta v)v + \mu|\boldsymbol{\nabla} v|^2\Bigr]\varphi\,\mathrm{d}\boldsymbol{x} = \mu\int_\Omega \mathrm{div}(v\boldsymbol{\nabla} v)\varphi\,\mathrm{d}\boldsymbol{x} = -\mu\int_\Omega v\boldsymbol{\nabla} v\cdot\boldsymbol{\nabla}\varphi\,\mathrm{d}\boldsymbol{x},$$

and so, finally, by the Cauchy-Schwarz inequality :

$$\Bigl|\int_\Omega \Bigl[\mu(\Delta u)u + \mu|\boldsymbol{\nabla} u|^2\Bigr]\varphi\,\mathrm{d}\boldsymbol{x}\Bigr| \le C\|\boldsymbol{\nabla}\varphi\|_{\mathrm{L}^\infty(\Omega)}\ \mu^{1/2}.$$

A discrete analogue of this simple computation is used to pass to the limit in the scheme in the next section (with a control on the unknown assumed and not proven).

Since, in the equation of state, the pressure vanishes for $e = 0$, and that S_K^{n+1} is a non-negative continuous function of the unknowns ρ, $\boldsymbol{u}$ and p, adapting the proof of Chapter 3 to cope with this additional term, we obtain that the scheme admits at least one solution, which satisfies $p \geq 0$, $\rho \geq 0$ and $e \geq 0$. In addition, Equation (IV.11) shows that the scheme conserves the total energy.

IV.3.3 Passing to the limit in the scheme

The objective of this section is to show, in the one dimensional case, that, if a sequence of solutions is controlled in suitable norms and converges to a limit, this latter necessarily satisfies a (part of the) weak formulation of the continuous problem.

We suppose given a sequence of meshes and time steps $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$, such that the time step and the size $h^{(m)}$ of the mesh $\mathcal{M}^{(m)}$, defined by :

$$h^{(m)} = \ \sup_{K\in\mathcal{M}^{(m)}} h_K,$$

tend to zero as $m \to \infty$, where h_K stands for the diameter of K. Note that, since we are dealing with a 1D problem, $h_K = |K|$.

Let $\rho^{(m)}$, $p^{(m)}$, $e^{(m)}$ and $u^{(m)}$ be the solution given by the scheme (IV.5) with the mesh $\mathcal{M}^{(m)}$ and the time step $\delta t^{(m)}$, or, more precisely speaking, a 1D version of the scheme which may be obtained by taking the MAC variant, only one horizontal stripe of meshes, supposing that the vertical component of the velocity (the degree of freedom of which are located on the top and bottom boundaries) vanishes, and that the measure of the faces is equal to 1. To the discrete unknowns, we associate piecewise constant functions on time intervals and on primal or dual cells, so the density $\rho^{(m)}$, the pressure $p^{(m)}$, the internal energy $e^{(m)}$ and the velocity $u^{(m)}$ are defined almost everywhere on $\Omega \times (0,T)$ by :

$$\rho^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{M}} \rho_K^n \ \mathcal{X}_K \ \mathcal{X}_{(n,n+1)}, \qquad p^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{M}} p_K^n \ \mathcal{X}_K \ \mathcal{X}_{(n,n+1)},$$
$$e^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{K\in\mathcal{M}} e_K^n \ \mathcal{X}_K \ \mathcal{X}_{(n,n+1)}, \qquad u^{(m)}(x,t) = \sum_{n=0}^{N-1} \sum_{\sigma\in\mathcal{E}} u_\sigma^n \ \mathcal{X}_{D_\sigma} \ \mathcal{X}_{(n,n+1)},$$

where $\mathcal{X}_K$, $\mathcal{X}_{D_\sigma}$ and $\mathcal{X}_{(n,n+1)}$ stand for the characteristic function of K, D_σ and the interval $(n, n+1)$ respectively, and, for short, we have dropped the superscript $^{(m)}$ in $\mathcal{M}^{(m)}$, $\mathcal{E}^{(m)}$, $N^{(m)}$ and in the local values of the discrete functions.

We suppose that the sequence $\big(\rho^{(m)}, p^{(m)}, e^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ is uniformly bounded in $\mathrm{L}^\infty\big((0,T)\times\Omega\big)$, *i.e.* :

$$|(\rho^{(m)})_K^n| + |(p^{(m)})_K^n| + |(e^{(m)})_K^n| \leq C, \quad \forall K \in \mathcal{M}^{(m)}, \text{ for } 0 \leq n \leq N^{(m)}, \ \forall m \in \mathbb{N}, \tag{IV.12}$$

and :

$$|(u^{(m)})_\sigma^n| \leq C, \quad \forall \sigma \in \mathcal{E}^{(m)}, \text{ for } 0 \leq n \leq N^{(m)}, \ \forall m \in \mathbb{N}. \tag{IV.13}$$

We also suppose a uniform control on the translates in space and time, which we now state. For discrete function q and v defined on the primal and dual mesh, respectively, we define a discrete $\mathrm{L}^1\big((0,T);\mathrm{BV}(\Omega)\big)$ norm by :

$$\|q\|_{\mathcal{T},x,BV} = \sum_{n=0}^{N} \delta t \sum_{\sigma\in\mathcal{E},\ \sigma=K|L} |q_L^n - q_K^n|, \qquad \|v\|_{\mathcal{T},x,BV} = \sum_{n=0}^{N} \delta t \sum_{\varepsilon\in\bar{\mathcal{E}},\ \sigma=D_\sigma|D'_\sigma} |v_{\sigma'}^n - v_\sigma^n|,$$

and a discrete $\mathrm{L}^1\big(\Omega;\mathrm{BV}((0,T))\big)$ norm by :

$$\|q\|_{\mathcal{T},t,BV} = \sum_{K\in\mathcal{M}} h_K \sum_{n=0}^{N-1} |q_K^{n+1} - q_K^n|, \qquad \|v\|_{\mathcal{T},t,BV} = \sum_{\sigma\in\mathcal{E}} h_\sigma \sum_{n=0}^{N-1} |v_\sigma^{n+1} - v_\sigma^n|,$$

where, for $\sigma = K|L$, $h_\sigma = (h_K + h_L)/2$. We suppose that the sequence of solutions satisfies the following uniform bounds with respect to these two norms :

$$\|\rho^{(m)}\|_{\mathcal{T},x,BV} + \|e^{(m)}\|_{\mathcal{T},x,BV} + \|u^{(m)}\|_{\mathcal{T},x,BV} \leq C, \quad \forall m \in \mathbb{N}. \tag{IV.14}$$

and :

$$\|u^{(m)}\|_{\mathcal{T},t,BV} \leq C, \quad \forall m \in \mathbb{N}, \tag{IV.15}$$

Of course, we are not able to prove the estimates (IV.12)–(IV.14) for the solutions of the scheme ; however, such inequalities are satisfied by the "interpolation" (for instance, by taking the cell average) of the solution to a Riemann problem, and are observed in computations (of course, as far as possible, *i.e.* with a limited sequence of meshes and time steps).

A weak solution to the continuous problem satisfies, for any $\varphi \in \mathrm{C}_c^\infty\big([0,T) \times \Omega\big)$:

$$-\int_{\Omega\times(0,T)} \Big[\rho\,\partial_t\varphi + \rho\,u\,\partial_x\varphi\Big]\,\mathrm{d}x - \int_\Omega \rho(x,0)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{IV.16a}$$

$$-\int_{\Omega\times(0,T)} \Big[\rho\,u\,\partial_t\varphi + (\rho\,u^2 + p)\,\partial_x\varphi\Big]\,\mathrm{d}\boldsymbol{x} - \int_\Omega \rho(x,0)\,u(x,0)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{IV.16b}$$

$$-\int_{\Omega\times(0,T)} \Big[\rho\,E\,\partial_t\varphi + (\rho\,E + p)\,u\,\partial_x\varphi\Big]\,\mathrm{d}\boldsymbol{x} - \int_\Omega \rho(x,0)\,E(x,0)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{IV.16c}$$

$$p = (\gamma-1)\rho\,e, \qquad E = \frac{1}{2}u^2 + e. \tag{IV.16d}$$

Note that these relations are not sufficient to define a weak solution to the problem, since they do not imply anything about the boundary conditions. However, they allow to derive the Rankine-Hugoniot conditions ; so, if we show that they are satisfied by the limit of a sequence of solutions to the discrete problem, this implies, loosely speaking, that *the scheme computes the right shocks*, which is the result we seek. It is stated in the following theorem.

Theorem IV.3.1

Let Ω be an open bounded interval of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, e^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence satisfies (IV.12)–(IV.14) and converges in $\mathrm{L}^r\big((0,T)\times\Omega\big)^4$, for $1 \leq r < \infty$, to $(\bar\rho,\, \bar p,\, \bar e,\, \bar u) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^4$.

Then the limit $(\bar\rho,\, \bar p,\, \bar e,\, \bar u)$ satisfies the system (IV.16).

Proof The fact that the limit ($\bar{\rho}$, $\bar{p}$, $\bar{u}$) satisfies (IV.16a) and (IV.16b) is proven in Chapter 2, using milder estimates and convergence assumptions. On the other hand, the fact that ($\bar{\rho}$, $\bar{p}$, $\bar{e}$, $\bar{u}$) satisfies the equation of state is straightforward, in view of the supposed convergence. We thus only need to prove that ($\bar{\rho}$, $\bar{p}$, $\bar{e}$, $\bar{u}$) satisfies (IV.16c).

Let $\varphi \in C_c^\infty(\Omega \times [0,T))$. Let $m \in \mathbb{N}$, $\mathcal{M}^{(m)}$ and $\delta t^{(m)}$ be given. Dropping for short the superscript ${}^{(m)}$, we define $\varphi_{\mathcal{M}}$ and $\varphi_{\mathcal{E}}$, an interpolate of φ on the primal and dual mesh respectively, by :

$$\varphi_{\mathcal{M}} = \sum_{n=0}^{N-1} \sum_{K \in \mathcal{M}} \varphi_K^n \ \mathcal{X}_K \ \mathcal{X}_{(t^n,t^{n+1})}, \qquad \varphi_{\mathcal{E}} = \sum_{n=0}^{N-1} \sum_{\sigma \in \mathcal{E}} \varphi_\sigma^n \ \mathcal{X}_{D_\sigma} \ \mathcal{X}_{(t^n,t^{n+1})}, \tag{IV.17}$$

where, for $1 \leq n \leq N$, $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we set :

$$\varphi_K^n = \varphi(x_K, t^n) \qquad \text{and} \qquad \varphi_\sigma^n = \varphi(x_\sigma, t^n),$$

with x_K the mass center of K and x_σ the abscissa of the face σ. We also define the time discrete derivative of these discrete functions by :

$$\begin{aligned} \eth_t \varphi_{\mathcal{M}} &= \sum_{n=0}^{N-1} \sum_{K \in \mathcal{M}} \frac{\varphi_K^{n+1} - \varphi_K^n}{\delta t} \ \mathcal{X}_K \ \mathcal{X}_{(t^n,t^{n+1})}, \\ \eth_t \varphi_{\mathcal{E}} &= \sum_{n=0}^{N-1} \sum_{\sigma \in \mathcal{E}} \frac{\varphi_\sigma^{n+1} - \varphi_\sigma^n}{\delta t} \ \mathcal{X}_{D_\sigma} \ \mathcal{X}_{(t^n,t^{n+1})}, \end{aligned} \tag{IV.18}$$

and their space discrete derivatives :

$$\begin{aligned} \eth_x \varphi_{\mathcal{M}} &= \sum_{n=0}^{N-1} \sum_{\sigma \in \mathcal{E},\ \sigma = K < L} \frac{\varphi_L^n - \varphi_K^n}{d_\sigma} \ \mathcal{X}_{D_\sigma} \ \mathcal{X}_{(t^n,t^{n+1})}, \\ \eth_x \varphi_{\mathcal{E}} &= \sum_{n=0}^{N-1} \sum_{K \in \mathcal{M},\ K = <\sigma,\sigma'>} \frac{\varphi_{\sigma'}^n - \varphi_\sigma^n}{h_K} \ \mathcal{X}_K \ \mathcal{X}_{(t^n,t^{n+1})}, \end{aligned} \tag{IV.19}$$

where the notation $\sigma = K < L$ means that $\sigma = K|L$ with the orientation $x_K < x_L$, $K =< \sigma, \sigma' >$ means that $K = (x_\sigma, x_{\sigma'})$, with $x_\sigma < x_{\sigma'}$ and, for $\sigma = K|L$, $d_\sigma = (h_K + h_L)/2$. Finally, we define $\eth \varphi_{\mathcal{M},\mathcal{E}}$ by :

$$\eth_x \varphi_{\mathcal{M},\mathcal{E}} = \sum_{n=0}^{N-1} \sum_{K \in \mathcal{M},\ K = <\sigma,\sigma'>} \frac{\varphi_K^n - \varphi_\sigma^n}{h_K/2} \ \mathcal{X}_{D_{K,\sigma}} \ \mathcal{X}_{(t^n,t^{n+1})}, + \frac{\varphi_{\sigma'}^n - \varphi_K^n}{h_K/2} \ \mathcal{X}_{D_{K,\sigma'}} \ \mathcal{X}_{(t^n,t^{n+1})}. \tag{IV.20}$$

Thanks to the regularity of φ, the piecewise constant functions $\varphi_{\mathcal{M}}$, $\varphi_{\mathcal{E}}$, $\eth_t \varphi_{\mathcal{M}}$, $\eth_t \varphi_{\mathcal{E}}$, $\eth_x \varphi_{\mathcal{M}}$, $\eth_x \varphi_{\mathcal{E}}$ and $\eth_x \varphi_{\mathcal{M},\mathcal{E}}$ converge in $\mathrm{L}^r\big(\Omega \times (0,T)\big)$, for $r \geq 1$ (including $r = +\infty$), to φ, φ, $\partial_t \varphi$, $\partial_t \varphi$, $\partial_x \varphi$, $\partial_x \varphi$ and $\partial_x \varphi$ respectively.

On one hand, let us multiply the discrete kinetic energy equation (IV.9) by $\delta t \, \varphi_\sigma^n$ and sum over the edges and the time steps. On the other hand, let us multiply the discrete internal energy equation (IV.5c) by $\delta t \, \varphi_K^n$, and sum over the primal cells and the time steps. Finally, let us sum the two obtained relations.

We get :

$$T_1^{(m)}+T_2^{(m)}+T_3^{(m)}+T_4^{(m)}+T_5^{(m)}=R^{(m)}, \qquad \text{with :}$$

$$T_1^{(m)}=\frac{1}{2}\sum_{n=0}^{N-1}\delta t\sum_{\sigma\in\mathcal{E}}\frac{|D_\sigma|}{\delta t}\big[\rho_\sigma^{n+1}(u_\sigma^{n+1})^2-\rho_\sigma^n(u_\sigma^n)^2\big]\ \varphi_\sigma^n,$$

$$T_2^{(m)}=\sum_{n=0}^{N-1}\delta t\sum_{K\in\mathcal{M}}\frac{|K|}{\delta t}\big[\rho_K^{n+1}e_K^{n+1}-\rho_K^n e_K^n\big]\ \varphi_K^n,$$

$$T_3^{(m)}=\frac{1}{2}\sum_{n=0}^{N-1}\delta t\sum_{\sigma\in\mathcal{E}}\ \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma),\ \varepsilon=D_\sigma|D_{\sigma'}}F_{\sigma,\varepsilon}^{n+1}\ u_{\sigma'}^{n+1}\ u_\sigma^{n+1}\ \varphi_\sigma^n,$$

$$T_4^{(m)}=\sum_{n=0}^{N-1}\delta t\sum_{K}\sum_{\sigma\in\mathcal{E}(K)}F_{K,\sigma}^{n+1}e_\sigma^{n+1}\ \varphi_K^n,$$

$$T_5^{(m)}=\sum_{n=0}^{N-1}\delta t\sum_{\sigma\in\mathcal{E}}|D_\sigma|\,(\boldsymbol{\nabla}p^{n+1})_\sigma\ u_\sigma^{n+1}\ \varphi_\sigma^n+\sum_{n=0}^{N-1}\delta t\sum_{K\in\mathcal{M}}|K|\ p_K^{n+1}\,(\mathrm{div}(u^{n+1}))_K\ \varphi_K^n,$$

$$R^{(m)}=\sum_{n=0}^{N-1}\delta t\sum_{\sigma\in\mathcal{E}}R_\sigma^{n+1}\ \varphi_\sigma^n+\sum_{n=0}^{N-1}\delta t\sum_{K\in\mathcal{M}}S_K^{n+1}\ \varphi_K^n.$$

We first study $T_1^{(m)}$. Since the support of φ is compact in $\Omega\times(0,T)$, for space and time steps small enough (or, equivalently, m large enough), the interpolates of φ vanish for $n=N$, and, at any time, on the cells and faces located in a neighbourhood of the boundaries ; we suppose that it is the case for the element m of the sequence under consideration, for the term $T_1^{(m)}$ as well as for the remainder of the proof. Reordering of the sums and then using the definition (IV.6) of the density at the edges, we thus get :

$$\begin{aligned}T_1^{(m)}&=-\frac{1}{2}\sum_{n=0}^{N-1}\delta t\sum_{\sigma\in\mathcal{E}}|D_\sigma|\ \rho_\sigma^{n+1}\ (u_\sigma^{n+1})^2\ \frac{\varphi_\sigma^{n+1}-\varphi_\sigma^n}{\delta t}-\frac{1}{2}\sum_{\sigma\in\mathcal{E}}|D_\sigma|\ \rho_\sigma^0\ (u_\sigma^0)^2\ \varphi_\sigma^0\\&=-\frac{1}{2}\int_0^T\int_\Omega\rho^{(m)}\ (u^{(m)})^2\ \eth_t\varphi_{\mathcal{E}}\,\mathrm{d}x\delta t-\frac{1}{2}\int_\Omega\rho^{(m)}(x,0)\ (u^{(m)}(x,0))^2\ \varphi_{\mathcal{E}}(x,0)\,\mathrm{d}x.\end{aligned}$$

Since, by assumption, the sequence of discrete solutions and of interpolates converge in $\mathrm{L}^r\big(\Omega\times(0,T)\big)$ for $r\geq 1$, and by definition of the initial conditions, we get :

$$\lim_{m\longrightarrow+\infty}T_1^{(m)}=-\frac{1}{2}\int_0^T\int_\Omega\bar\rho\ (\bar u)^2\partial_t\varphi\,\mathrm{d}x\delta t-\frac{1}{2}\int_\Omega\bar\rho(x,0)\ (\bar u(x,0)^2\ \varphi(x,0)\,\mathrm{d}x.$$

By a similar computation, we get for $T_2^{(m)}$:

$$\begin{aligned}T_2^{(m)}&=-\sum_{n=0}^{N-1}\delta t\sum_{\sigma\in\mathcal{E}}|K|\ \rho_K^{n+1}\ e_K^{n+1}\ \frac{\varphi_K^{n+1}-\varphi_K^n}{\delta t}-\sum_{\sigma\in\mathcal{E}}|K|\ \rho_K^0\ e_K^0\ \varphi_K^0\\&=-\int_0^T\int_\Omega\rho^{(m)}\ e^{(m)}\ \eth_t\varphi_{\mathcal{M}}\,\mathrm{d}x\delta t-\int_\Omega\rho^{(m)}(x,0)\,e^{(m)}(x,0)\,\varphi_{\mathcal{M}}(x,0)\,\mathrm{d}x,\end{aligned}$$

and therefore :

$$\lim_{m\longrightarrow+\infty}T_2^{(m)}=-\int_0^T\int_\Omega\bar\rho\,\bar e\,\partial_t\varphi\,\mathrm{d}x\delta t-\int_\Omega\bar\rho(x,0)\,\bar e(x,0)\,\varphi(x,0)\,\mathrm{d}x.$$

Let us now turn to $T_3^{(m)}$. For $K =< \sigma, \sigma' >$ and ε the dual face included in K, the dual mass flux reads :

$$F_{\sigma,\varepsilon}^{n+1} = \frac{1}{2}\,(F_{K,\sigma'}^{n+1} - F_{K,\sigma}^{n+1}). \tag{IV.21}$$

We thus get, reordering the sums :

$$\begin{aligned}
T_3^{(m)} &= -\frac{1}{4}\sum_{n=0}^{N-1}\delta t \sum_{K\in\mathcal{M},\,K=<\sigma,\sigma'>} \big(F_{K,\sigma}^{n+1} - F_{K,\sigma'}^{n+1}\big)\; u_{\sigma'}^{n+1}\, u_{\sigma}^{n+1}\; \big(\varphi_\sigma^n - \varphi_{\sigma'}^n\big)\\
&= -\frac{1}{4}\sum_{n=0}^{N-1}\delta t \sum_{K\in\mathcal{M},\,K=<\sigma,\sigma'>} \big(\widetilde{\rho}_\sigma^{n+1} u_\sigma^{n+1} + \widetilde{\rho}_{\sigma'}^{n+1} u_{\sigma'}^{n+1}\big)\; u_{\sigma'}^{n+1}\, u_{\sigma}^{n+1}\; \big(\varphi_\sigma^n - \varphi_{\sigma'}^n\big)\\
&= -\frac{1}{2}\sum_{n=0}^{N-1}\delta t \sum_{K\in\mathcal{M},\,K=<\sigma,\sigma'>} |K|\;\rho_K^{n+1}\;\frac{(u_\sigma^{n+1})^3 + (u_{\sigma'}^{n+1})^3}{2}\;\frac{\varphi_\sigma^n - \varphi_{\sigma'}^n}{h_K} + \mathcal{R}_3^{(m)}.
\end{aligned}$$

Let us denote by $\mathcal{T}_3^{(m)}$ the first term. We have :

$$\mathcal{T}_3^{(m)} = -\frac{1}{2}\int_0^T\int_\Omega \rho^{(m)}\,(u^{(m)})^3\,\eth_x\varphi_{\mathcal{E}}\;\mathrm{d}x\delta t, \qquad \text{so} \qquad \lim_{m\longrightarrow+\infty}\mathcal{T}_3^{(m)} = -\frac{1}{2}\int_0^T\int_\Omega \bar\rho\,\bar u^3\,\partial_x\varphi\;\mathrm{d}x\delta t.$$

The residual term $\mathcal{R}_3^{(m)}$ reads :

$$\begin{aligned}
\mathcal{R}_3^{(m)} = -\frac{1}{4}\sum_{n=0}^{N-1}\delta t &\sum_{K\in\mathcal{M},\,K=<\sigma,\sigma'>}\\
&\Big[\big(\widetilde{\rho}_\sigma^{n+1} u_\sigma^{n+1} + \widetilde{\rho}_{\sigma'}^{n+1} u_{\sigma'}^{n+1}\big)\; u_{\sigma'}^{n+1}\, u_\sigma^{n+1} - \rho_K^{n+1}\Big((u_\sigma^{n+1})^3 + (u_{\sigma'}^{n+1})^3\Big)\Big]\;(\varphi_\sigma^n - \varphi_{\sigma'}^n)
\end{aligned} \tag{IV.22}$$

Expanding the quantity $(u_\sigma^{n+1})^3 + (u_{\sigma'}^{n+1})^3$ thanks to the identity $a^3 + b^3 = (a+b)(ab + (a-b)^2)$, and then reordering the sums, we obtain $\mathcal{R}_3^{(m)} = \mathcal{R}_{3,1}^{(m)} + \mathcal{R}_{3,2}^{(m)}$ with :

$$\begin{aligned}
\mathcal{R}_{3,1}^{(m)} &= -\frac{1}{4}\sum_{n=0}^{N-1}\delta t \sum_{K\in\mathcal{M},\,K=<\sigma,\sigma'>} \Big[(\widetilde{\rho}_\sigma^{n+1} - \rho_K^{n+1})\,u_\sigma^{n+1} + (\widetilde{\rho}_{\sigma'}^{n+1} - \rho_K^{n+1})\,u_{\sigma'}^{n+1}\Big]\\
&\qquad\qquad\qquad\qquad\qquad\qquad u_\sigma^{n+1} u_{\sigma'}^{n+1}\;(\varphi_\sigma^n - \varphi_{\sigma'}^n)\\
\mathcal{R}_{3,2}^{(m)} &= \frac{1}{4}\sum_{n=0}^{N}\delta t \sum_K |\sigma|\,\rho_K^{n+1}\,(u_\sigma^{n+1} + u_{\sigma'}^{n+1})\,(u_\sigma^{n+1} - u_{\sigma'}^{n+1})^2\,(\varphi_\sigma^n - \varphi_{\sigma'}^n)
\end{aligned}$$

In the term $\mathcal{R}_{3,1}^{(m)}$, the differences $\widetilde{\rho}_\sigma^{n+1} - \rho_K^{n+1}$ and $\widetilde{\rho}_{\sigma'}^{n+1} - \rho_K^{n+1}$ either vanish or compare the density in two adjacent cells. We thus get :

$$|\mathcal{R}_{3,1}^{(m)}| \leq h\,C_\varphi\;\|u^{(m)}\|_{\mathrm{L}^\infty}^3\;\|\rho^{(m)}\|_{\mathcal{T},x,BV},$$

and $\mathcal{R}_{3,1}^{(m)}$ tends to zero when m tends to $+\infty$. By similar arguments :

$$|\mathcal{R}_{3,2}^{(m)}| \leq h\,C_\varphi\;\|\rho^{(m)}\|_{\mathrm{L}^\infty}\;\|u^{(m)}\|_{\mathrm{L}^\infty}^2\;\|u^{(m)}\|_{\mathcal{T},x,BV},$$

and thus $\mathcal{R}_{3,2}^{(m)}$ also tends to zero when m tends to $+\infty$.

Expressing the mass fluxes as a function of the unknowns in $T_4^{(m)}$, we get, choosing for $\sigma = K|L$ the orientation such that $F_{K,\sigma}^{n+1} \geq 0$, so $\widetilde{\rho}_\sigma^{n+1} = \rho_K^{n+1}$ and $e_\sigma^{n+1} = e_K^{n+1}$:

$$T_4^{(m)} = \sum_{n=0}^{N-1}\delta t \sum_{\sigma\in\mathcal{E},\,\sigma=K\to L} |D_\sigma|\,\rho_K^{n+1}\,e_K^{n+1}\,u_\sigma^{n+1}\;n_\sigma\;\frac{(\varphi_K^n - \varphi_L^n)}{|d_\sigma|},$$

where the quantity n_σ is equal to $+1$ if $x_L \geq x_K$ and to -1 otherwise and the notation $\sigma = K \rightarrow L$ means that $\sigma = K|L$, with a flow leaving K and entering L. We decompose $T_4^{(m)} = \mathcal{T}_4^{(m)} + \mathcal{R}_4^{(m)}$, with :

$$\begin{aligned}
\mathcal{T}_4^{(m)} &= -\sum_{n=0}^{N-1} \delta t \sum_{\sigma\in\mathcal{E},\, \sigma=K\rightarrow L} \Big[|D_{K,\sigma}|\, \rho_K^{n+1}\, e_K^{n+1} + |D_{L,\sigma}|\, \rho_L^{n+1}\, e_L^{n+1}\Big]\, u_\sigma^{n+1}\, \frac{\varphi_L^n - \varphi_K^n}{d_\sigma}\, n_\sigma, \\
\mathcal{R}_4^{(m)} &= \sum_{n=0}^{N-1} \delta t \sum_{\sigma\in\mathcal{E},\, \sigma=K\rightarrow L} |D_{L,\sigma}|\, \Big[\rho_K^{n+1}\, e_K^{n+1} - \rho_L^{n+1}\, e_L^{n+1}\Big]\, u_\sigma^{n+1}\, \frac{\varphi_K^n - \varphi_L^n}{d_\sigma}\, n_\sigma.
\end{aligned}$$

We have :

$$\mathcal{T}_4^{(m)} = -\int_0^T \int_\Omega \rho^{(m)}\, e^{(m)}\, u^{(m)}\, \eth_x \varphi_{\mathcal{M}}\, \mathrm{d}x\delta t, \qquad \text{so} \quad \lim_{m\longrightarrow+\infty} \mathcal{T}_4^{(m)} = -\int_0^T \int_\Omega \bar\rho\, \bar e\, \bar u\, \partial_x \varphi\, \mathrm{d}x\delta t.$$

Expanding the quantity $(\rho_K^{n+1}\, e_K^{n+1} - \rho_L^{n+1}\, e_L^{n+1})$ in the residual term $\mathcal{R}_4^{(m)}$ thanks to the identity $2(ab - cd) = (a + c)(b - d) + (b + d)(a - c)$, we get :

$$|\mathcal{R}_4^{(m)}| \leq C_\varphi h\, \|u^{(m)}\|_{\mathrm{L}^\infty} \big[\|\rho^{(m)}\|_{\mathrm{L}^\infty}\, \|e^{(m)}\|_{\mathcal{T},x,BV} + \|e^{(m)}\|_{\mathrm{L}^\infty}\, \|\rho^{(m)}\|_{\mathcal{T},x,BV}\big],$$

so that $\mathcal{R}_4^{(m)}$ tends to zero when m tends to $+\infty$.

The term $T_5^{(m)}$ reads :

$$\begin{aligned}
T_5^{(m)} &= \sum_{n=0}^{N-1} \delta t \sum_{\sigma\in\mathcal{E}} |D_\sigma|\, (\boldsymbol{\nabla} p^{n+1})_\sigma\, u_\sigma^{n+1}\, \varphi_\sigma^n + \sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M}} |K|\, p_K^{n+1}\, (\mathrm{div}(u^{n+1}))_K\, \varphi_K^n \\
&= \sum_{n=0}^{N-1} \delta t\, \Big[\sum_{\sigma\in\mathcal{E},\, \sigma=K<L} (p_L^{n+1} - p_K^{n+1})\, u_\sigma^{n+1}\, \varphi_\sigma^n + \sum_{K\in\mathcal{M},\, K=<\sigma,\sigma'>} p_K^{n+1}\, (u_{\sigma'}^{n+1} - u_\sigma^{n+1})\, \varphi_K^n\Big] \\
&= \sum_{n=0}^{N-1} -\delta t \sum_{K\in\mathcal{M}\, K=<\sigma,\sigma'>} p_K^{n+1}\, (u_{\sigma'}^{n+1}\, \varphi_{\sigma'}^n - u_\sigma^{n+1}\, \varphi_\sigma^n) + p_K^{n+1}\, (u_\sigma^{n+1} - u_{\sigma'}^{n+1})\, \varphi_K^n
\end{aligned}$$

We thus obtain :

$$\begin{aligned}
T_5^{(m)} &= -\sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M},\, K=<\sigma,\sigma'>} \frac{h_K}{2}\, p_K^{n+1}\, u_\sigma^{n+1}\, \frac{\varphi_K^n - \varphi_\sigma^n}{h_K/2} + \frac{h_K}{2}\, p_K^{n+1}\, u_{\sigma'}^{n+1}\, \frac{\varphi_{\sigma'}^n - \varphi_K^n}{h_K/2} \\
&= -\int_0^T \int_\Omega p^{(m)}\, u^{(m)}\, \eth_x \varphi_{\mathcal{M},\mathcal{E}}\, \mathrm{d}x\delta t.
\end{aligned}$$

and so :

$$\lim_{m\longrightarrow+\infty} T_5^{(m)} = -\int_0^T \int_\Omega \bar p\, \bar u\, \partial_x \varphi\, \mathrm{d}x\delta t.$$

Finally we study $R^{(m)}$, which we decompose in $R^{(m)} = R_c^{(m)} + R_{up}^{(m)}$, the first part gathering the terms which are not linked to a possible upwinding. We have for this residual :

$$\begin{aligned}
R_c^{(m)} &= \frac{1}{2} \sum_{n=0}^{N-1} \Big[\sum_{\sigma\in\mathcal{E}} -|D_\sigma|\, \rho_\sigma^n\, (u_\sigma^{n+1} - u_\sigma^n)^2\, \varphi_\sigma^n + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} |D_{K,\sigma}|\, \rho_K^n\, (u_\sigma^{n+1} - u_\sigma^n)^2\, \varphi_K^n\Big] \\
&= \frac{1}{2} \sum_{n=0}^{N-1} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}(K)} |D_{K,\sigma}|\, \rho_K^n\, (\boldsymbol{u}_\sigma^{n+1} - \boldsymbol{u}_\sigma^n)^2\, (\varphi_K^n - \varphi_\sigma^n).
\end{aligned}$$

We thus obtain :

$$R_c^{(m)} \leq h\, C_\varphi\ \|\rho^{(m)}\|_{\mathrm{L}^\infty}\ \|u^{(m)}\|_{\mathrm{L}^\infty}\ \|u^{(m)}\|_{\mathcal{T},t,BV},$$

and $R^{(m)}$ tends to zero when $m \to \infty$. We now turn to the upwind case. The corresponding terms read :

$$R_{up}^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \Big[\sum_{\sigma\in\mathcal{E}} \ \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma),\, \varepsilon=D_\sigma|D_{\sigma'}} -|F_{\sigma,\varepsilon}^{n+1}|\ (u_\sigma^{n+1} - u_{\sigma'}^{n+1})\ u_\sigma^{n+1}\ \varphi_\sigma^n \\ + \sum_{K\in\mathcal{M}} \ \sum_{\varepsilon\subset K,\, \varepsilon=D_\sigma|D_{\sigma'}} |F_{\sigma,\varepsilon}^{n+1}|\ (u_\sigma^{n+1} - u_{\sigma'}^{n+1})^2\ \varphi_K^n \Big]$$

As explained at the end of Section IV.3.2, the general idea is now to recast this term as a discrete version of the integral over space and time of a quantity of the form $-u\, \partial_x u\, \partial_x \varphi$ scaled by a numerical viscosity vanishing with the space step ; then, the supposed controls on the solution imply that the term tends to zero. We thus reorder the sums in $R_{up}^{(m)}$, which yields :

$$R_{up}^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M},\ K=<\sigma,\sigma'>} \big|F_{\sigma,D_\sigma|D_{\sigma'}}^{n+1}\big|\ (u_\sigma^{n+1} - u_{\sigma'}^{n+1})\ (u_{\sigma'}^{n+1}\varphi_{\sigma'}^n - u_\sigma^{n+1}\varphi_\sigma^n) \\ + \big|F_{\sigma,D_\sigma|D_{\sigma'}}^{n+1}\big|\ (u_{\sigma'}^{n+1} - u_\sigma^{n+1})^2 \varphi_K^n,$$

and thus :

$$R_{up}^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M},\ K=<\sigma,\sigma'>} \big|F_{\sigma,D_\sigma|D_{\sigma'}}^{n+1}\big|\ (u_\sigma^{n+1} - u_{\sigma'}^{n+1}) \Big[u_\sigma^{n+1}(\varphi_K^n - \varphi_\sigma^n) + u_{\sigma'}^{n+1}(\varphi_{\sigma'}^n - \varphi_K^n) \Big]$$

We thus get, using the definition (IV.21) of the mass fluxes at the dual faces :

$$|R_{up}^{(m)}| \leq h\, C_\varphi\ \|\rho^{(m)}\|_{\mathrm{L}^\infty}\ \|u^{(m)}\|_{L^\infty}^2\ \|u^{(m)}\|_{\mathcal{T},x,BV},$$

which yields the desired control.

Gathering the expression of the limits of each of the terms $T_1^{(m)}$ to $T_5^{(m)}$ and $R^{(m)}$ concludes the proof.
□

IV.4 A pressure correction scheme

IV.4.1 The scheme

We derive in this section a pressure correction numerical scheme from the implicit scheme (IV.5). The first step, as usual, is to compute a tentative velocity by solving the momentum balance equation with the begining-of-step pressure. Then, the velocity is corrected and the other variables are advanced in time, here, which is less standard, by a single coupled step ; this is motivated by stability reasons detailed in Chapter 3. Still for stability reasons, or, in other words, to be able to derive a kinetic energy balance, we need that a mass balance over the dual cells (IV.7) holds ; since the mass balance is not yet solved when performing the prediction step, this leads us to perform a time shift of the density at this step.

The algorithm reads :

Prediction step – Solve for $\tilde{\boldsymbol{u}}^{n+1}$:

$$\text{For } 1 \leq i \leq d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma^n \tilde{u}_{\sigma,i}^{n+1} - \rho_\sigma^{n-1} u_{\sigma,i}^n) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^n \tilde{u}_{\varepsilon,i}^{n+1} + |D_\sigma| \, (\boldsymbol{\nabla} p^n)_{\sigma,i} = 0, \tag{IV.23a}$$

Correction step – Solve for ρ^{n+1}, p^{n+1}, e^{n+1} and $\boldsymbol{u}^{n+1}$:

$$\text{For } 1 \leq i \leq d, \quad \left| \begin{array}{l} \forall \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \forall \sigma \in \mathcal{E} \text{ otherwise,} \end{array} \right.$$

$$\frac{|D_\sigma|}{\delta t} \, \rho_\sigma^n \, (u_{\sigma,i}^{n+1} - \tilde{u}_{\sigma,i}^{n+1}) + |D_\sigma| \left[(\boldsymbol{\nabla} p^{n+1})_{\sigma,i} - (\boldsymbol{\nabla} p^n)_{\sigma,i} \right] = 0, \tag{IV.23b}$$

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} = 0, \tag{IV.23c}$$

$\forall K \in \mathcal{M}$,

$$\frac{|K|}{\delta t}(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^{n+1} e_\sigma^{n+1} + |K| \, p_K^{n+1} \, \big(\mathrm{div}(\tilde{\boldsymbol{u}}^{n+1})\big)_K = S_K^{n+1}, \tag{IV.23d}$$

$$\forall K \in \mathcal{M}, \qquad p_K^{n+1} = (\gamma - 1) \, \rho_K^{n+1} \, e_K^{n+1}. \tag{IV.23e}$$

IV.4.2 The discrete kinetic energy balance equation and the corrective source terms

We repeat the same process that we have followed for the implicit scheme, to determine the numerical term source S_K^n. We thus begin with deriving the discrete kinetic energy equation. To this purpose, we sum the momentum balance equation (IV.23a) with the velocity correction equation (IV.23b), which yields :

$$\frac{|D_\sigma|}{\delta t} \, \big(\rho_\sigma^n u_{\sigma,i}^{n+1} - \rho_\sigma^{n-1} u_{\sigma,i}^n\big) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^n \tilde{u}_{\varepsilon,i}^{n+1} + |D_\sigma| \, (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} = 0.$$

Multiplying this equation by the corresponding degree of freedom of the predicted velocity $\tilde{u}_{\sigma,i}^{n+1}$, we obtain :

$$\frac{|D_\sigma|}{\delta t} \, \big(\rho_\sigma^n u_{\sigma,i}^{n+1} - \rho_\sigma^{n-1} u_{\sigma,i}^n\big) \, \tilde{u}_{\sigma,i}^{n+1} + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}^n \tilde{u}_{\varepsilon,i}^{n+1} \, \tilde{u}_{\sigma,i}^{n+1} + |D_\sigma| \, (\boldsymbol{\nabla} p^{n+1})_{\sigma,i} \, \tilde{u}_{\sigma,i}^{n+1} = 0.$$

Let us recast the first two terms of this equation as $T^{(1)}_{\sigma,i}+T^{(2)}_{\sigma,i}$, with :

$$\begin{aligned} T^{(1)}_{\sigma,i} &= \frac{|D_\sigma|}{\delta t}\,\left(\rho^n_\sigma \tilde u^{n+1}_{\sigma,i} - \rho^{n-1}_\sigma u^n_{\sigma,i}\right)\,\tilde u^{n+1}_{\sigma,i} + \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)} F^n_{\sigma,\varepsilon}\tilde u^{n+1}_{\varepsilon,i}\ \tilde u^{n+1}_{\sigma,i},\\ T^{(2)}_{\sigma,i} &= \frac{|D_\sigma|}{\delta t}\ \rho^n_\sigma\ \left(u^{n+1}_{\sigma,i} - \tilde u^{n+1}_{\sigma,i}\right)\ \tilde u^{n+1}_{\sigma,i}. \end{aligned}$$

The term $T^{(1)}_{\sigma,i}$ has the structure which allows to apply Lemma II.3.2 of Chapter 2, and we get :

$$T^{(1)}_{\sigma,i} = \frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho^n_\sigma(\tilde u^{n+1}_{\sigma,i})^2 - \rho^{n-1}_\sigma(u^n_{\sigma,i})^2\Big] + \frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}} F^n_{\sigma,\varepsilon}\ \tilde u^{n+1}_{\sigma,i}\ \tilde u^{n+1}_{\sigma',i} + R^{(1)}_{\sigma,i},$$

with :

$$R^{(1)}_{\sigma,i} = \frac{|D_\sigma|}{2\,\delta t}\ \rho^{n-1}_\sigma\ \left(\tilde u^{n+1}_{\sigma,i} - u^n_{\sigma,i}\right)^2 + \delta^{\rm up}\Big[\sum_{\varepsilon=D_\sigma|D_{\sigma'}} \frac{1}{2}\,|F^n_{\sigma,\varepsilon}|\ \left(\tilde u^{n+1}_{\sigma,i} - \tilde u^{n+1}_{\sigma',i}\right)\Big]\ \tilde u^{n+1}_{\sigma,i}.$$

Using the identity $2\,(a-b)\,a = a^2 - b^2 + (a-b)^2$, valid for any real numbers a and b, we get for T_2 :

$$T^{(2)}_{\sigma,i} = \frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho^n_\sigma(u^{n+1}_{\sigma,i})^2 - \rho^n_\sigma(\tilde u^n_{\sigma,i})^2\Big] + R^{(2)}_{\sigma,i},$$

with :

$$R^{(2)}_{\sigma,i} = -\frac{|D_\sigma|}{2\,\delta t}\ \rho^n_\sigma\ \left(u^{n+1}_{\sigma,i} - \tilde u^{n+1}_{\sigma,i}\right)^2.$$

Summing, we get the discrete kinetic energy balance equation :

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\Big[\rho^n_\sigma(u^{n+1}_{\sigma,i})^2 - \rho^{n-1}_\sigma(u^n_{\sigma,i})^2\Big] + \frac{1}{2}\sum_{\varepsilon=D_\sigma|D_{\sigma'}} F^n_{\sigma,\varepsilon}\ \tilde u^{n+1}_{\sigma,i}\ \tilde u^{n+1}_{\sigma',i} \\ + |D_\sigma|\,(\boldsymbol{\nabla} p^{n+1})_{\sigma,i}\ \tilde u^{n+1}_{\sigma,i} = R^n_{\sigma,i}, \qquad \text{(IV.24)}$$

with :

$$R^{n+1}_{\sigma,i} = -R^{(1)}_{\sigma,i} - R^{(2)}_{\sigma,i}.$$

By the same arguments as in the implicit case, we get :

$$\forall K\in\mathcal{M},\\ S^{n+1}_K = \frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)} \frac{|D_{K,\sigma}|}{\delta t}\rho^{n-1}_K\ |\tilde{\boldsymbol u}^{n+1}_\sigma - \boldsymbol u^n_\sigma|^2 - \frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)} \frac{|D_{K,\sigma}|}{\delta t}\rho^n_K\ |\boldsymbol u^{n+1}_\sigma - \tilde{\boldsymbol u}^{n+1}_\sigma|^2 \\ + \delta^{\rm up} \sum_{\substack{\varepsilon\cap\bar K\neq\emptyset,\\ \varepsilon=D_\sigma|D_{\sigma'}}} \alpha_{K,\varepsilon}\ \frac{|F^{n+1}_{\sigma,\varepsilon}|}{2}|\boldsymbol u^{n+1}_\sigma - \boldsymbol u^{n+1}_{\sigma'}|^2. \qquad \text{(IV.25)}$$

Note that, now, the term S_K may be negative, which we have indeed observed in computations ; however, even in very severe cases (as, for instance, Test 3 of [68, chapter 4]), at least with a reasonable time step, we still obtained a positive internal energy.

IV.4.3 Passing to the limit in the scheme

As for the implicit scheme, we show in this section, in the one dimensional case, that, if a sequence of solutions is controlled in suitable norms and converges to a limit, this limit necessarily satisfies a (part of the) weak formulation of the continuous problem.

Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $\delta t^{(m)}$ and $h^{(m)}$ tend to zero as $m \to \infty$. Let $\rho^{(m)}$, $p^{(m)}$, $e^{(m)}$, $\tilde u^{(m)}$ and $u^{(m)}$ be the associated solution of the pressure correction scheme (IV.23), obtained, as in the implicit case, with the 1D version of the scheme. We suppose that this solution satisfies similar controls as in the case of the implicit scheme, so, in addition of the already written bounds for $\rho^{(m)}$, $p^{(m)}$, $e^{(m)}$ and $u^{(m)}$, we also assume :

$$|(\tilde u^{(m)})^n_\sigma| \leq C, \quad \forall \sigma \in \mathcal{E}^{(m)}, \text{ for } 0 \leq n \leq N^{(m)}, \ \forall m \in \mathbb{N},$$
$$\text{and} \quad \|\tilde u^{(m)}\|_{\mathcal{T},x,BV} \leq C, \quad \|\rho^{(m)}\|_{\mathcal{T},t,BV} \leq C, \quad \forall m \in \mathbb{N}. \qquad \text{(IV.26)}$$

Note that we do not need any control on $\|\tilde u^{(m)}\|_{\mathcal{T},t,BV}$. Then we get the following "passage to the limit" theorem.

Theorem IV.4.1
Let Ω be an open bounded interval of of $\mathbb{R}$. Let $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ be a sequence of meshes and time steps, such that $h^{(m)}$ and $\delta t^{(m)}$ tend to zero as m tends to infinity. Let $\big(\rho^{(m)}, p^{(m)}, e^{(m)}, \tilde u^{(m)}, u^{(m)}\big)_{m\in\mathbb{N}}$ be the corresponding sequence of solutions. We suppose that this sequence satisfies (IV.12)–(IV.14) and (IV.26) and converges in $\mathrm{L}^p\big((0,T)\times\Omega\big)^5$, for $1 \leq p < \infty$, to $(\bar\rho,\, \bar p,\, \bar e,\, \bar{\tilde u},\, \bar u) \in \mathrm{L}^\infty\big((0,T)\times\Omega\big)^5$.

Then $\bar{\tilde u} = \bar u$ and $(\bar\rho,\, \bar p,\, \bar e,\, \bar u)$ satisfies the system (IV.16).

Proof . Let $\varphi \in \mathrm{C}^\infty_c(\Omega \times (0,T))$. Let $m \in \mathbb{N}$, $\mathcal{M}^{(m)}$ and $\delta t^{(m)}$ be given, and let the interpolates, and time and space discrete derivatives of φ associated to this discretization be defined, as in the implicit scheme, by (IV.17), (IV.18), (IV.19) and (IV.20).

We begin with checking that $\bar{\tilde u} = \bar u$. To this purpose, it is sufficient to note that the correction step yields :

$$|D_\sigma|\ |u^{n+1}_\sigma - \tilde u^{n+1}_\sigma| \leq \delta t\ |p_L - p_K|, \qquad \forall \sigma = K|L \in \mathcal{E}, \text{ and for } 0 \leq n \leq N-1,$$

so :

$$\|u^{(m)} - \tilde u^{(m)}\|_{\mathrm{L}^1} \leq \delta t\ \|p^{(m)}\|_{\mathcal{T},x,BV}$$

which, passing to the limit when $m \to +\infty$, yields the result.

We now turn to the proof that the limit satisfies (IV.16). On one hand, lets us multiply the discrete kinetic energy equation (IV.24) by $\delta t\, \varphi^n_\sigma$ and sum over the edges and the time steps. On the other hand, let us multiply the discrete internal energy equation (IV.23d) by $\delta t\, \varphi^n_K$, and sum over the primal celles and the time steps. Finally, let us sum the two obtained relations. We get :

$$T_1^{(m)} + T_2^{(m)} + T_3^{(m)} + T_4^{(m)} + T_5^{(m)} = R^{(m)}, \qquad \text{with :}$$

$$T_1^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{\sigma \in \mathcal{E}} \frac{|D_\sigma|}{\delta t} \left[\rho_\sigma^n (u_\sigma^{n+1})^2 - \rho_\sigma^{n-1} (u_\sigma^n)^2\right] \varphi_\sigma^n,$$

$$T_3^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{\sigma \in \mathcal{E}} \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma),\ \varepsilon = D_\sigma | D_{\sigma'}} F_{\sigma,\varepsilon}^n \ \tilde{u}_{\sigma'}^{n+1} \ \tilde{u}_\sigma^{n+1} \ \varphi_\sigma^n,$$

$$R^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{\sigma \in \mathcal{E}} R_\sigma^{n+1} \ \varphi_\sigma^n + \sum_{n=0}^{N-1} \delta t \sum_{K \in \mathcal{M}} S_K^{n+1} \ \varphi_K^n,$$

the terms $T_2^{(m)}$, $T_4^{(m)}$ and $T_5^{(m)}$ being the same as in the implicit scheme.

The passage to the limit in the term $T_1^{(m)}$ is done as in the implicit case, just remarking that $\rho^{(m)}(\cdot, \cdot - \delta t)$ strongly converges to $\bar{\rho}$. For the term $T_3^{(m)}$, still by a computation similar to the implicit case, we get :

$$T_3^m = -\frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{K \in \mathcal{M},\ K = <\sigma,\sigma'>} |K| \ \rho_K^n \ \frac{u_\sigma^n (\tilde{u}_\sigma^{n+1})^2 + u_{\sigma'}^n (\tilde{u}_{\sigma'}^{n+1})^2}{2} \ \frac{\varphi_\sigma^n - \varphi_{\sigma'}^n}{h_K} + \mathcal{R}_3^{(m)}.$$

Let us denote by $\mathcal{T}_3^{(m)}$ the first term. We get :

$$\mathcal{T}_3^{(m)} = -\frac{1}{2} \int_0^T \int_\Omega \rho^{(m)}(x, t - \delta t) \ u^{(m)}(x, t - \delta t) \ \tilde{u}^{(m)}(x,t)^2 \, \mathrm{d}x \delta t,$$

so :

$$\lim_{m \longrightarrow +\infty} \mathcal{T}_3^{(m)} = -\frac{1}{2} \int_0^T \int_\Omega \bar{\rho} \, \bar{u}^3 \, \partial_x \varphi \, \mathrm{d}x \delta t.$$

The residual term $\mathcal{R}_3^{(m)}$ reads :

$$\mathcal{R}_3^{(m)} = -\frac{1}{4} \sum_{n=0}^{N-1} \delta t \sum_{K \in \mathcal{M},\ K = <\sigma,\sigma'>} \left[\left(\widetilde{\rho}_\sigma^n u_\sigma^n + \widetilde{\rho}_{\sigma'}^n u_{\sigma'}^n\right) \tilde{u}_{\sigma'}^{n+1} \, \tilde{u}_\sigma^{n+1} - \rho_K^n \left(u_\sigma^n \, (\tilde{u}_\sigma^{n+1})^2 + u_{\sigma'}^n \, (\tilde{u}_{\sigma'}^{n+1})^2\right)\right] (\varphi_\sigma^n - \varphi_{\sigma'}^n).$$

We thus get :

$$\mathcal{R}_3^{(m)} = -\frac{1}{4} \sum_{n=0}^{N-1} \delta t \sum_{K \in \mathcal{M},\ K = <\sigma,\sigma'>} \Big[\underbrace{\left(\widetilde{\rho}_\sigma^n \tilde{u}_{\sigma'}^{n+1} - \rho_K^n \tilde{u}_\sigma^{n+1}\right)}_{\mathcal{D}_1} u_\sigma^n \, \tilde{u}_\sigma^{n+1} + \underbrace{\left(\widetilde{\rho}_{\sigma'}^n \tilde{u}_\sigma^{n+1} - \rho_K^n \tilde{u}_{\sigma'}^{n+1}\right)}_{\mathcal{D}_2} u_{\sigma'}^n \, \tilde{u}_{\sigma'}^{n+1}\Big] (\varphi_\sigma^n - \varphi_{\sigma'}^n).$$

Using the identity $2(ab - cd) = (a - c)(b + d) + (a + c)(b - d)$ for $\mathcal{D}_1$ and $\mathcal{D}_1$, we conclude that :

$$|\mathcal{R}_3^{(m)}| \leq h \, C_\varphi \ \|u^{(m)}\|_{\mathrm{L}^\infty} \ \|\tilde{u}^{(m)}\|_{\mathrm{L}^\infty} \ \Big[\|\rho^{(m)}\|_{\mathcal{T},x,BV} \ \|\tilde{u}^{(m)}\|_{\mathrm{L}^\infty} + \|\rho^{(m)}\|_{\mathrm{L}^\infty} \ \|\tilde{u}^{(m)}\|_{\mathcal{T},x,BV}\Big],$$

and thus $\mathcal{R}_3^{(m)}$ tends to zero when m tends to $+\infty$.

Finally we study $R^{(m)}$, which we split in $R^{(m)} = R_c^{(m)} + R_{up}^{(m)}$, the first part, namely $R_c^{(m)}$, gathering the terms which are not associated to the upwinding :

$$R_c^{(m)} = -\frac{1}{2}\sum_{n=0}^{N-1}\sum_{\sigma\in\mathcal{E}} |D_\sigma| \Big[\rho_\sigma^{n-1}\,(\tilde u_\sigma^{n+1} - u_\sigma^n)^2 - \rho_\sigma^n\,(\tilde u_\sigma^{n+1} - u_\sigma^{n+1})^2\Big]\varphi_\sigma^n \\ + \frac{1}{2}\sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}(K)} |D_{K,\sigma}| \Big[\rho_K^{n-1}\,(\tilde u_\sigma^{n+1} - u_\sigma^n)^2 - \rho_K^n\,(\tilde u_\sigma^{n+1} - u_\sigma^{n+1})^2\Big]\,\varphi_K^n$$

Thanks to the definition of the density on the edges, we get :

$$R_c^{(m)} = \frac{1}{2}\sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}(K)} |D_{K,\sigma}| \Big[\rho_K^{n-1}\,(\tilde u_\sigma^{n+1} - u_\sigma^n)^2 - \rho_K^n\,(\tilde u_\sigma^{n+1} - u_\sigma^{n+1})^2\Big]\,(\varphi_K^n - \varphi_\sigma^n),$$

so :

$$|R_c^{(m)}| \le h\,C_\varphi \sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}(K)} |D_{K,\sigma}| \Big|\rho_K^{n-1}\,(\tilde u_\sigma^{n+1} - u_\sigma^n)^2 - \rho_K^n\,(\tilde u_\sigma^{n+1} - u_\sigma^{n+1})^2\Big|.$$

Developping :

$$\rho_K^{n-1}\,(\tilde u_\sigma^{n+1} - u_\sigma^n)^2 - \rho_K^n\,(\tilde u_\sigma^{n+1} - u_\sigma^{n+1})^2 = \\ (\rho_K^{n-1} - \rho_K^n)\,(\tilde u_\sigma^{n+1} - u_\sigma^n)^2 + \rho_K^n\,(u_\sigma^{n+1} - u_\sigma^n)(2\tilde u_\sigma^{n+1} - u_\sigma^n - u_\sigma^{n+1})$$

yields :

$$|R_c^{(m)}| \le h\,C_\varphi\ \|\rho^{(m)}\|_{\mathcal{T},t,BV}\ \Big(\|u^{(m)}\|_{\mathrm{L}^\infty}^2 + \|\tilde u^{(m)}\|_{L^\infty}^2\Big) \\ + \|\rho^{(m)}\|_{\mathrm{L}^\infty}\ \|u^{(m)}\|_{\mathcal{T},t,BV}\ \Big(\|u^{(m)}\|_{\mathrm{L}^\infty} + \|\tilde u^{(m)}\|_{L^\infty}\Big),$$

and so $R_c^{(m)}$ tends to zero as m tends to $+\infty$. Replacing $u^{(m)}$ by $\tilde u^{(m)}$, the term $R_{up}^{(m)}$ takes the same expression as in the implicit case, and so also tends to zero. Gathering all the limits yields the result we are seeking. □

IV.5 Numerical tests

In this section, we assess the behaviour of the scheme on a one dimensional Riemann problem. We choose initial conditions such that the structure of the solution consists in two shock waves, separated by the contact discontinuity, with sufficiently strong shocks to allow to easily discrimate between convergence to the correct weak solution or not. These initial conditions are those proposed in [68, chapter 4], for the test refered to as Test 5 :

$$\text{left state : } \begin{bmatrix}\rho_L\\ u_L\\ p_L\end{bmatrix} = \begin{bmatrix}5.99924\\ 19.5975\\ 460.894\end{bmatrix} \qquad \text{right state : } \begin{bmatrix}\rho_R\\ u_R\\ p_R\end{bmatrix} = \begin{bmatrix}5.99242\\ -6.19633\\ 46.0950\end{bmatrix}$$

The problem is posed over $\Omega = (-0.5, 0.5)$, and the discontinuity is initially located at $x = 0$.

We obtain a one dimensional scheme by simply taking one horizontal stripe of meshes (of constant size) with the MAC discretization, and applying perfect slip boundary conditions at the top and bottom boundary. At the other boundaries, since, in this test, the flow is entering the domain, the solution is prescribed (which, in fact, is unimportant, the solution being constant at any time in a sufficiently large neighbourhood of these boundaries). Passed numerical experiments addressing barotropic flows (see Chapter 1) showed that, at least for one dimensional computations with schemes similar to the one under study here, it was not necessary to use upwinding in the momentum balance equation ; consequently, we only employ a centered approximation of the velocity at the dual edges.

The computations are performed with the open-source software ISIS [40], developed at IRSN on the basis of the software component library and programming environment PELICANS [63].

The density fields obtained with $h = 1/2000$ (or a number of cells $n = 2000$) at $t = 0.035$, with and without assembling the corrective source term in the internal energy balance $(S_K)_{K\in\mathcal{M}}$, together with the analytical solution, are shown on Figure IV.2. The density and the pressure obtained, still with and without corrective terms, for various meshes, are plotted on Figure IV.3 and IV.3 respectively. For these computations, we take $\delta t = h/20$, which yields a cfl number, with respect to the material velocity only, close to one. The first conclusion is that both schemes seem to converge, but the corrective term is necessary to obtain the correct solution. In this case, for instance, we obtain the correct intermediate state for the pressure and velocity up to four digits in the essential part of the corresponding zone :

$$\text{(analytical) intermediate state : } \quad \begin{bmatrix} p^* \\ u^* \end{bmatrix} = \begin{bmatrix} 1691.65 \\ 8.68977 \end{bmatrix} \quad \text{for } x \in (0.028, 0.428)$$

$$\text{numerical results : } \quad \left| \begin{array}{l} p \in (1691.6, 1691.8) \\ u \in (8.689, 8.690) \end{array} \right. \quad \text{for } x \in (0.032, 0.417)$$

Without corrective term, one can check that the obtained solution is not a weak solution to the Euler system : indeed, the Rankine-Hugoniot condition applied to the total energy balance, with the states obtained numerically, yields a right shock velocity slightly greater than the analytical solution one, while the same shock velocity obtained numerically is clearly lower.

We also observe that the scheme is rather diffusive, specially for representing the contact discontinuity, where the beneficial compressive effect of the shocks does not apply. More accurate variants may certainly be derived, using for instance MUSCL-like techniques. Finally, let us also mention that a fully explicit version of the scheme is currently under testing.

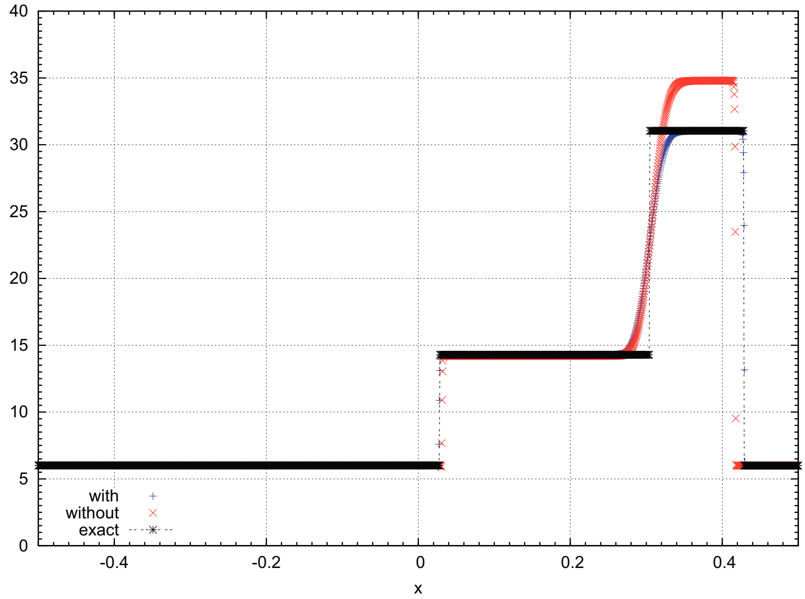

FIG. IV.2 – Test 5 of [68, chapter 4] - Density obtained with $n = 2000$ cells, with and without corrective source terms, and analytical solution.

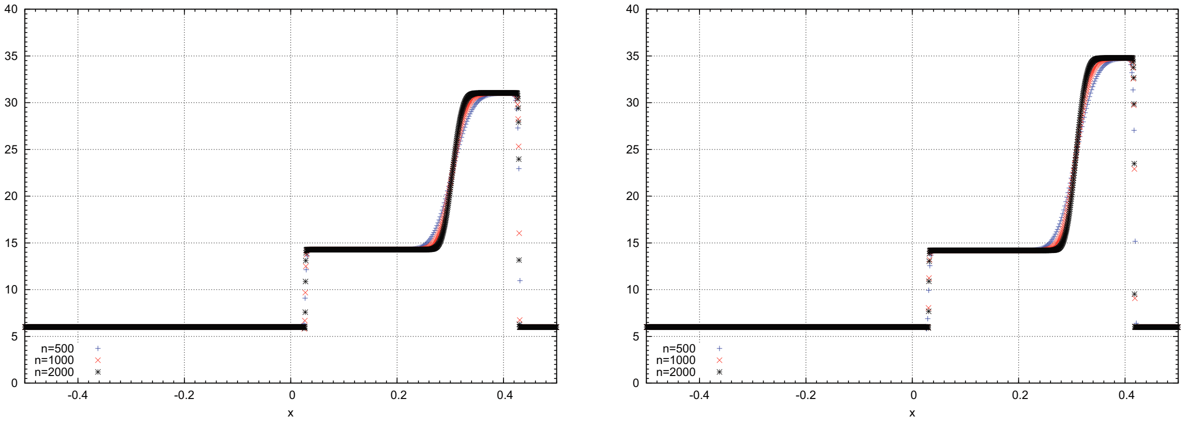

FIG. IV.3 – Test 5 of [68, chapter 4] - Density obtained with various meshes, with (left) and without (right) corrective source terms.

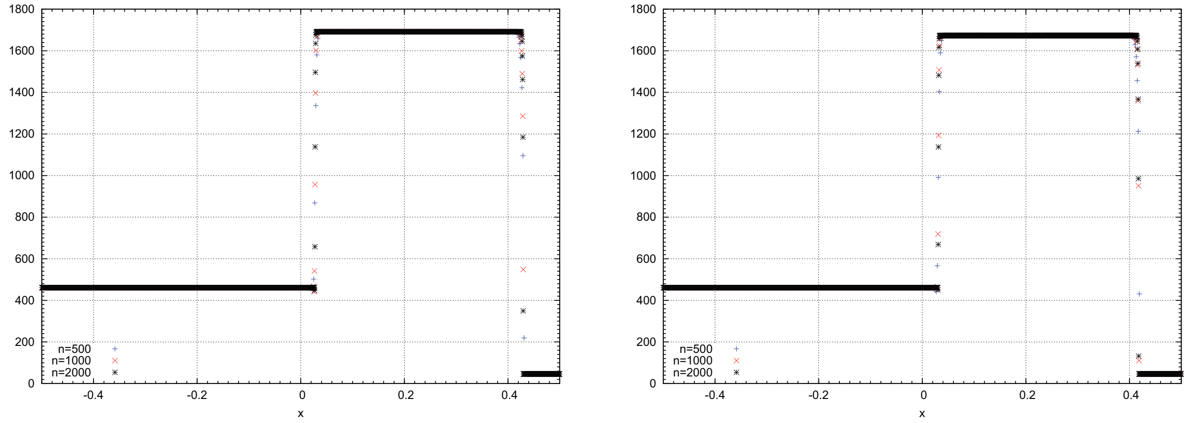

Fig. IV.4 – Test 5 of [68, chapter 4] - pressure obtained with various meshes, with (left) and without (right) corrective source terms.

Annexe

A The Riemann problem for the homegeneous model

A.1 The system of conservation laws and its mathematical properties

The model – We address in this section a model for two-phase flows (without phase change), which reads, in the one-dimensionnal case :

$$\left|\begin{array}{l} \partial_t \rho + \partial_x(\rho u) = 0, \\[2ex] \partial_t(\rho u) + \partial_x(\rho u^2 + p) = 0, \\[2ex] \partial_t(\rho y) + \partial_x(\rho y u) = 0, \end{array}\right. \tag{A.1}$$

where u stands for the fluid velocity, p for the pressure, ρ for the fluid density and y stands for the gas mass fraction. This system must be complemented by an equation of state, which takes the form :

$$\rho = \frac{1}{\dfrac{y}{\rho_g} + \dfrac{1-y}{\rho_l}}, \tag{A.2}$$

where ρ_g and ρ_l stand for the (phasic) gas and liquid density respectively. We assume that the liquid density ρ_l is constant, and that the gas phase obeys the perfect gas law $\rho_g = p/(RT)$, where where R is the gas constant and T is the absolute temperature. This allows to compute the pressure from relation (A.2), in order to obtain an expression of the quantity $\partial_x p$ as a function of the conservative variables :

$$p = \frac{RT\rho_l\ (\rho y)}{\rho_l + (\rho y) - \rho}. \tag{A.3}$$

We define $q = \rho u$, $z = \rho y$, $U = (\rho,\ q,\ z)^t$. With these definitions, the system (A.1) reads :

$$\partial_t U + \partial_x\big(F(U)\big) = 0, \quad \text{with } F(U) = (q,\ \frac{q^2}{\rho} + p(\rho, z),\ z)^t, \quad p(\rho, z) = \frac{RT\rho_l z}{\rho_l + z - \rho}. \tag{A.4}$$

We suppose that the variable U belongs to the convex subset of $\mathbb{R}^3$ (sometimes referred to as the set of states of the system) :

$$\mathcal{C} = \big\{(\rho, q, z) \in \mathbb{R}^3,\ \rho > 0,\ 0 < z \leq \rho,\ \rho_l + z - \rho > 0\big\},$$

which ensures in particular that the equation of state makes sense. For a regular solution, System (A.4) may be set in non-conservative form :

$$\partial_t U + \partial_x F(U) = \partial_t U + A(U) \cdot \partial_x U = 0,$$

with :

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -\dfrac{q^2}{\rho^2} + \partial_\rho p & \dfrac{2q}{\rho} & \dfrac{q^2}{\rho^2} + \partial_z p \\ -\dfrac{qz}{\rho^2} & \dfrac{z}{\rho} & \dfrac{q}{\rho} \end{pmatrix}, \tag{A.5}$$

and, by (A.3) :

$$\partial_\rho p = \frac{RT\rho_l z}{(\rho_l + z - \rho)^2}, \qquad \partial_z p = \frac{RT\rho_l\,(\rho_l - \rho)}{(\rho_l + z - \rho)^2}.$$

Hyperbolicity, eigenvalues and eigenvectors of the system

Definition A.1.1 (hyperbolic problems)
Let $\mathcal{C}$ be an open subset of $\mathbb{R}^n$. We consider the nonlinear system of conservation laws :

$$\partial_t U + B(U) \cdot \partial_x \big(F(U,x,t) \big) = 0, \qquad x \in \mathbb{R},\ t > 0, \tag{A.6}$$

where $U \in \mathcal{C}$ is a vector function of x an t, and $F \in \mathbb{R}^n$ stands for a regular vector function depending on U as well as, possibly, on x and t. We denote by A the $n \times n$-matrix associated to the differential of F with respect to U ; A depends on U as well as, possibly, on x and t. System (A.6) *is said to be hyperbolic if, for each x, t and U, all the eigenvalues of the matrix A belong to $\mathbb{R}$:*

$$\lambda_1(U) \leq \lambda_2(U) \leq \ldots \leq \lambda_n(U).$$

To each eigenvalue $\lambda_k(U)$, we associate an eigenvector $r_k(U)$:

$$A(U) \cdot r_k(U) = \lambda_k(U)\ r_k(U).$$

The k^{th} characteristic field is said to be genuinely nonlinear if :

$$D\lambda_k(U) \cdot r_k(U) \neq 0, \qquad \forall U \in \mathcal{C},$$

where D stands for the differential operator in $\mathbb{R}^n$ (i.e. $D\lambda_k$ stands for the derivative of λ_k with respect to U). The k^{th} characteristic field is said to be linearly degenerate if :

$$D\lambda_k(U) \cdot r_k(U) = 0, \qquad \forall U \in \mathcal{C}.$$

Returning, to alleviate notations, to non-conservative variables, the matrix A reads :

$$A = \begin{pmatrix} 0 & 1 & 0 \\ \partial_\rho p - u^2 & 2u & u^2 + \partial_z p \\ -uy & y & u \end{pmatrix}. \tag{A.7}$$

We denote by a the positive real number such that :

$$a^2 = \partial_\rho p + y\, \partial_z p = \frac{RT\rho_l\, z}{(\rho_l + z - \rho)^2}.$$

This quantity a is referred to as the sound velocity of the mixture. The matrix A has three eigenvalues, which, with this notation, read :

$$\lambda_1(U) = u - a, \qquad \lambda_2(U) = u, \qquad \lambda_3(U) = u + a,$$

and the system is thus hyperbolic. The corresponding eigenvectors are given by :

$$r_1 = \begin{pmatrix} 1 \\ u - a \\ y \end{pmatrix}, \qquad r_2 = \begin{pmatrix} 1 \\ u \\ z/(\rho - \rho_l) \end{pmatrix}, \qquad r_3 = \begin{pmatrix} 1 \\ u + a \\ y \end{pmatrix}.$$

A tedious but straightforward computation shows that the characteristic fields associated to the eigenvalues λ_1 and λ_3 are genuinely nonlinear, while the characteristic field associated to λ_2 is linearly degenerate.

Riemman invariants

Definition A.1.2 (Riemann invariants)
Fo $1 \leq k \leq n$, a smooth function $W : \mathcal{C} \to \mathbb{R}$ is called a k-Riemann invariant if it satisfies :

$$DW(U) \cdot r_k(U) = 0, \ \forall U \in \mathcal{C}.$$

A k-Rieman invariant W is constant on a curve $V : \xi \in \mathbb{R} \to V(\xi) \in \mathbb{R}^n$ if :

$$\frac{d}{d\xi} W(V(\xi)) = DW(V(\xi)).V'(\xi) = 0, \tag{A.8}$$

which holds if V is an integral curve of r_k, i.e. satisfies that $V'(\xi)$ is colinear to $r_k(V(\xi))$. There exist locally $(n-1)$ k-Rieman invariants whose gradients are linearly independent.

Let us now search for the Riemann invariants associated of the system equation (A.1). According to the definition A.1.2, we have two Riemann invariants for each eigenvalue of A.
- the 1-Riemann invariants are :

$$W_{1,1} = y, \qquad W_{1,2} = u + \sqrt{RTy} \log(\frac{\rho - z}{\rho_l + z - \rho}). \tag{A.9}$$

- the 2-Riemann invariants are given by :

$$W_{2,1} = u, \qquad W_{2,1} = p. \tag{A.10}$$

- the 3-Riemann invariants are :

$$W_{3,1} = y, \qquad W_{3,2} = u - \sqrt{RTy} \log(\frac{\rho - z}{\rho_l + z - \rho}). \tag{A.11}$$

Rarefaction waves – System (A.4) satisfies the property of self-similarity, *i.e.* is invariant under the transormation $t \mapsto \alpha t, \ x \mapsto \alpha x, \ \alpha > 0$. If the intial data of the problem is also invariant under the transformation $x \mapsto \alpha x$, we thus conclude that a regular solution to (A.4) must satisfy $U(x,t) = U(\alpha x, \alpha t)$ whatever $\alpha > 0$ may be, *i.e.* $U(x,t) = U(x/t)$, for $t > 0$. Such a (regular) solution is called a rarefaction wave.

Let $\xi = x/t$, and $V(\xi) = U(x/t)$, and substitute this expression for U in (A.4), to obtain :

$$A(V) \ V'(\xi) = \xi \ V'(\xi).$$

We deduce for this relation that either V' is zero, which corresponds to the trivial case of a constant state, or this vector is necessarily colinear to an eigenvector $r_k(V)$ of the matrix $A(V)$:

$$V'(\xi) = \beta \ r_k\big(V(\xi)\big), \qquad \lambda_k\big(V(\xi)\big) = \xi. \tag{A.12}$$

The rarefaction waves are always associated to genuinely nonlinear fields, so, for the problem at hand, there are two possible families of rarefaction waves, the first one associated to λ_1 and the second one to λ_3 ; a solution of the first class is called a 1-rarefaction wave, and a solution of the second class is called a 3-rarefraction wave. Thanks to (A.8), Riemann invariants are kept constant in rarefaction waves, so an 1-wave satisfies :

$$W_{1,1}(x,t) = y = cste, \qquad W_{1,2}(x,t) = u + \sqrt{RTy} \log(\frac{\rho - z}{\rho_l + z - \rho}) = cste, \tag{A.13}$$

and a 3-wave satisfies :

$$W_{3,1}(x,t) = y = cste, \qquad W_{3,2}(x,t) = u - \sqrt{RTy}\log(\frac{\rho - z}{\rho_l + z - \rho}) = cste. \tag{A.14}$$

Discontinuous solutions and entropy condition – It is wellknown that hyperbolic problems do not always have continuous solutions. This leads to introduce the notion of "weak solution", defined as a solution in the distribution sense of the problem in conservative form, here System (A.1). Let us suppose that such a solution is piecewise constant, consisting (locally) in two constant states separated by a discontinuity. Exploiting the definition of weak solutions yields algebraic relations (one per equation) which links the jump through the discontinuity of the solution, the associated fluxes and the velocity of the discontinuity, defined by $\sigma = d(x_s)/dt$, where x_s stands for the discontinuity location ; such a relation is called a Rankine-Hugoniot condition, and the system constituted by these relations reads :

$$\sigma[U] = [F(U)], \tag{A.15}$$

where $[U]$ (resp. $[F(U)]$) stands for the jump of U (resp. $F(U)$) through the discontinuity. Unfortunately, this algebraic system is not sufficient to ensure the uniqueness of the solution (of course, in the class of piecewise constant functions). Hence, we need to introduce some criterion that enables us to choose the "physically relevent" solution among all the weak solutions of the problem. This criterion is called the "Lax entropy conditions".

DEFINITION A.1.3 (Lax entropy conditions)
Let U be defined by $U = U^L$ if $x < \sigma t$, and $U = U^R$ if $x > \sigma t$, where U^L and U^R are two constant states (i.e. two constant vectors of $\mathbb{R}^n$). We say that the discontinuity satisfies the Lax entropy conditions if there exists an index $k \in \{1, 2 \ldots, n\}$ such that we have either :

$$\lambda_{k-1}(U^L) < \sigma < \lambda_k(U^L) \text{ and } \lambda_k(U^R) < \sigma < \lambda_{k+1}(U^R), \tag{A.16}$$

if the k^{th} characteristic field is genuinely nonlinear (setting, in this relation, $\lambda_0 = -\infty$ and $\lambda_{n+1} = +\infty$), or :

$$\lambda_k(U^L) = \sigma = \lambda_k(U^R) \tag{A.17}$$

if the k^{th} characteristic field is linearly degenerate.

Shocks – For System (A.1), we have two class of discontinuities associated to genuinely nonlinear fields, which we call shocks : a 1-shock is associated to λ_1 and a 3-shock is associated to λ_3. Exploiting the Rankine-Hugoniot and Lax entropy conditions, we find that the quantity y is left constant through the shocks, and that a state (u, p) may be connected to U^L and U^R respectively by a 1-shock wave and a 3-shock wave if :

– 1-shock wave :

$$u = u^L - \frac{p - p^L}{\sqrt{p}}\sqrt{\frac{\rho_l RT\, y^L}{\rho_l RT\ \rho^L\, y^L + p^L\, \rho^L\,(1 - y^L)}}, \quad p^L \le p. \tag{A.18}$$

– 3-shock wave :

$$u = u^R + \frac{p - p^R}{\sqrt{p}}\sqrt{\frac{\rho_l RT\, y^R}{\rho_l RT\ \rho^R\, y^R + p^R\, \rho^R\,(1 - y^R)}}, \quad p^R \le p. \tag{A.19}$$

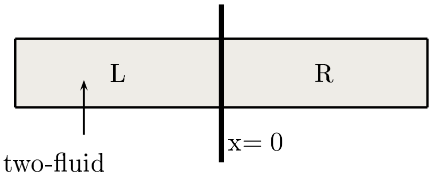

Fig. A.1 – Geometry of tube

Contact discontinuity – The possible discontinuity associated to the linearly degenerated field is called a contact discontinuity. The Riemann invariants are known to be kept constant through such a discontinuity, so such is the associated eingenvalue of the system. The Lax condition thus implies that the velocity of the discontinuity is necessarily equal to this constant value. Here, the contact discontinuity is associated to the second eigenvalue $\lambda_2 = u$, and thus, thanks to (A.10) :

$$\sigma = U^L = U^R, \quad p^L = p^R.$$

A.2 Solution of the Riemann problem

A Riemann problem consists in searching for the solution to an hyperbolic problem with a piecewise constant initial data, with a single discontinuity, usually located at the origin. For fluid mechanics problem, it is often called a "shock tube problem", since it can be thought of as an infinitely long (in order to avoid reflections) tube where the left and the right regions are separated by a diaphragm, and filled by the same fluid in two different physical states. At the bursting of the diaphragm, the discontinuity between the two initial states breaks into leftward and rightward moving waves, wich are separated by a contact surface.

For the system under consideration, according to the wave structure described in the previous section, each wave pattern is composed by a contact discontinuity (C) in the middle, and a shock (S) or a rarefaction wave (R) at the left and right hand sides separating uniform states (see Figure A.2). All the available combinations produce four wave patterns ; RCR, RCS, SCR,SCS, which are self-similar, that is only depend on x/t.

Let U_L be the left state and U_R be the right one. The unknown region between the left and right waves is divided by the middle wave (contact discontinuity) into two intermediate states U_1 and U_2 such that :

- U_2 is connected to the state U_L by a 1-wave,
- U_1 is connected to the state U_2 by a 2-wave,
- U_R is connected to the state U_1 by a 3-wave.

We use the fact that the pressure p and the velocity u are constant through the contact discontinuity, and the gas mass fraction y is a Riemann invariant for both the 1-wave and the 3-wave, to obtain :

$$u_1 = u_2 = u^*, \qquad p_1 = p_2 = p^*, \qquad y_1 = y_L, \qquad y_2 = y_R.$$

where (p^*, u^*) stands for the pressure and velocity in the two intermediate states. The problem thus boils down to determine this pair of values. To this purpose, we use the results obtained in the previous section :

- 1-wave – If $p^* > p_L$, the 1-wave is a shock, and the pair (u^*, p^*) satisfies (A.18) with $U^L = U_L$:

$$u^* = u_L - \frac{p^* - p_L}{\sqrt{p^*}} \sqrt{\frac{\rho_l RT\, y_L}{\rho_l RT\ \rho_L\, y_L + p_L\, \rho_L\, (1 - y_L)}}, \quad p^* \geq p_L.$$

  Otherwise, the 1-wave is a rarefaction wave, and, using the expression (A.9) of the second associated Riemann invariant, we obtain that (u^*, p^*) satisfies :

$$u^* = u_L + \sqrt{RTy_L}\ \log(\frac{p_L}{p^*}), \qquad \text{with } p^* \leq p_L.$$

  Equation (A.18) and this latter relation define a curve $\mathcal{C}_1$ in the plane (u, p), representative of a function of p^*, $p^* \in (0, +\infty)$; for $y_L > 0$, this function is strictly increasing, and one-to-one from $(0, +\infty)$ to $\mathbb{R}$ ($\lim_{p^* \to 0} u^* = +\infty$, and $\lim_{p^* \to +\infty} u^* = -\infty$).

- 3-wave – Similarly, If $p^* > p_L$, the 3-wave is a shock, and the pair (u^*, p^*) satisfies (A.19) with $U^R = U_R$:

$$u^* = u_R + \frac{p^* - p_R}{\sqrt{p^*}} \sqrt{\frac{\rho_l RT\, y_R}{\rho_l RT\ \rho_R\, y_R + p_R\, \rho_R\, (1 - y_R)}}, \quad p^* \geq p_R.$$

  Otherwise, the 3-wave is a rarefaction wave, and, using the expression (A.11) of the second associated Riemann invariant, we obtain that (u^*, p^*) satisfies :

$$u^* = u_R - \sqrt{RTy_R}\ \log(\frac{p_R}{p^*}), \qquad \text{with } p^* \leq p_R.$$

  Equation (A.19) and this latter relation also define a curve $\mathcal{C}_3$ in the plane (u, p), representative of a function of p^*, $p^* \in (0, +\infty)$; for $y_R > 0$, this function is strictly decreasing, and one-to-one from $(0, +\infty)$ to $\mathbb{R}$ ($\lim_{p^* \to 0} u^* = -\infty$, and $\lim_{p^* \to +\infty} u^* = +\infty$).

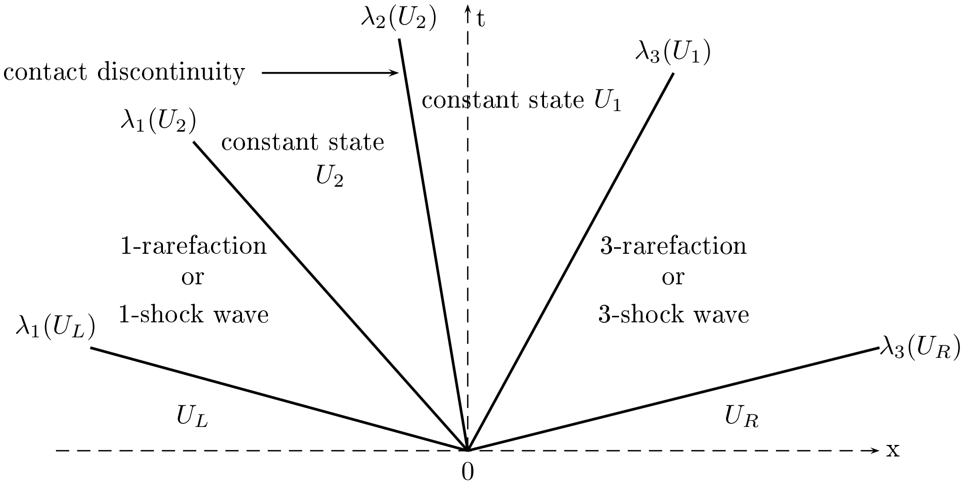

FIG. A.2 – Solution of the Riemann problem in (x, t) space.

The pair (u^*, p^*) is located at the (unique) intersection of the curves $\mathcal{C}_1$ and $\mathcal{C}_3$.

We give below two exemples of application of this strategy to find the solution of particular Riemann problems.

A.2.1 Sod shock tube

We assume here that the gas mass fraction is set to $y \equiv 1$ (one phase problem) ; the equation of state is given by $p = \rho RT$, and the two-phase problem just boils down to the isothermal Euler equations. The two initial constant states are given by :

$$\begin{pmatrix} \rho \\ \boldsymbol{u} \end{pmatrix}_L = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} \rho \\ u \end{pmatrix}_R = \begin{pmatrix} 0.125 \\ 0 \end{pmatrix}.$$

The parameters R and T are adjusted to produce $RT = 1$.

We start by determining the intermediate states by drawing the set of accessible states from the left and right in the space (p, u) (see Figure A.3), and we determine the intersection :

$$(p^*, u^*) = (0.34, 1.06).$$

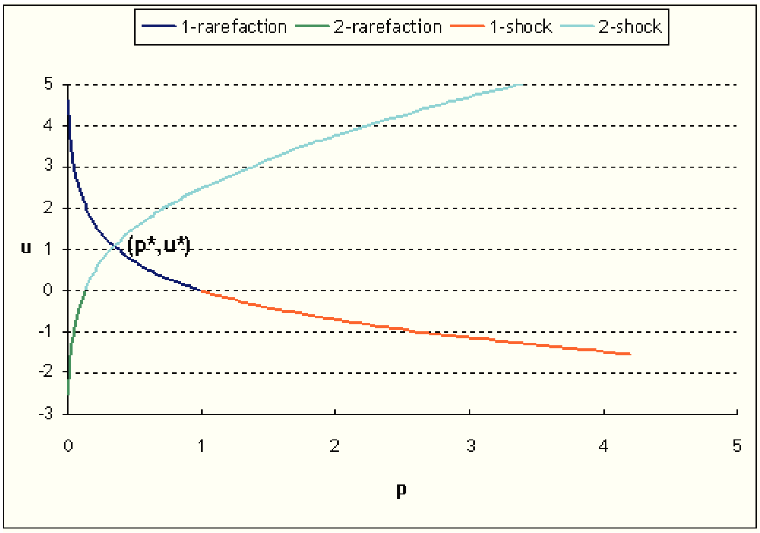

FIG. A.3 – curves of the shock-rarefaction in the space (p, u)

The wave structure of this system consists in a shock travelling to the right with a velocity equal to 1.66 and a rarefaction wave travelling to the left, which reads :

$$p(x) = \frac{1}{e^{x+1}} \qquad \text{and} \qquad u(x) = x + 1, \qquad \text{for } -t \leqslant x \leqslant 0.061\,t.$$

This solution is drawn in Fig A.4.

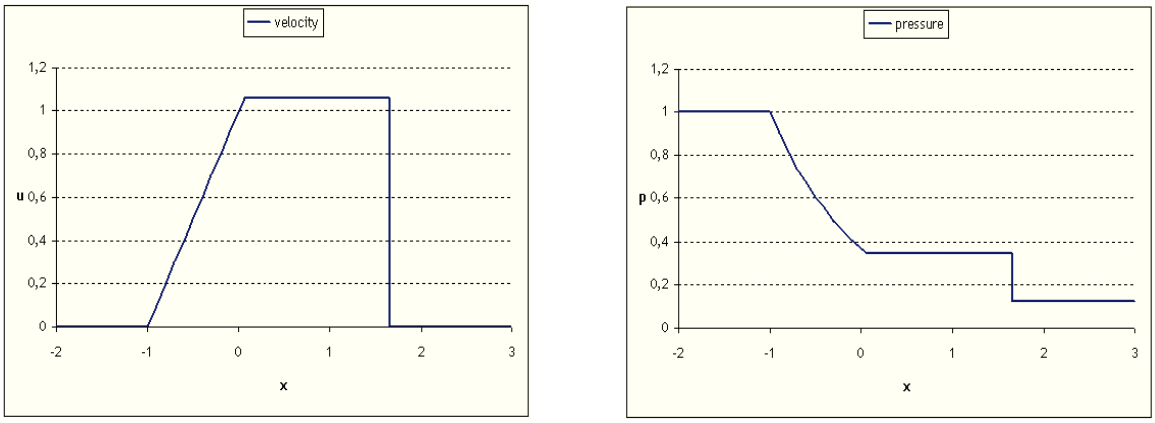

FIG. A.4 – Sod shock tube problem – Exact solution at $t = 1$

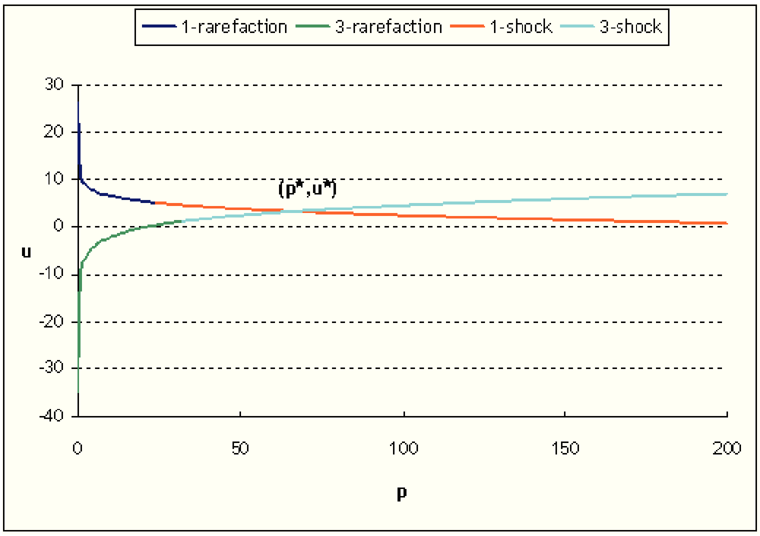

FIG. A.5 – Curves of the shock-rarefaction in the space (p, u)

A.2.2 Two-fluid shock tube

We now address a two-phase problem, the equation of state (A.3) of which we recall :

$$p = \frac{RT\rho_\ell \rho y}{\rho_\ell + \rho y - \rho}.$$

The parameters R and T are adjusted to produce $RT = 10$ and the liquid density is constant and set to $\rho_\ell = 0.8$. The two initial constant states are given by :

$$\begin{pmatrix} \rho \\ u \\ y \end{pmatrix}_L = \begin{pmatrix} 1. \\ 5. \\ 0.3 \end{pmatrix}, \qquad \begin{pmatrix} \rho \\ u \\ y \end{pmatrix}_R = \begin{pmatrix} 2. \\ 1. \\ 0.8 \end{pmatrix}.$$

We determine the intermediates states by drawing the set of accessible states from the left and right in the space (p, u) (see Figure A.5), and then compute the intersection :

$$(u^*, p^*) = (3.14\,,\,67.06).$$

The associated waves are a 1-shock and a 3-shock. The wave structure of this system thus consists in a shock wave travelling to the left and a shock wave travelling to the right, separated by a contact discontinuity in the middle, see Figure A.6. The solution reads :

- (u_L, p_L) is connected to (u^*, p^*) by 1-shock with a shock velocity equal to -18.16, and (u^*, p^*) is connected to u_R by a 3-shock with a shock velocity equal to 9.18,
- $y = y_L$ up to the contact discontinuity, and then equal to y_R ; the contact discontinuity velocity is equal to $u^* = 3.14$.
- ρ_L is connected to ρ_1 by the 1-shock (shock velocity equal to -18.16), then ρ_1 is connected to ρ_2 by the contact discontinuity (velocity equal to $u^* = 3.14$), then, finally, ρ_2 is connected to ρ_R by the 3-shock (shock velocity equal to 9.18).

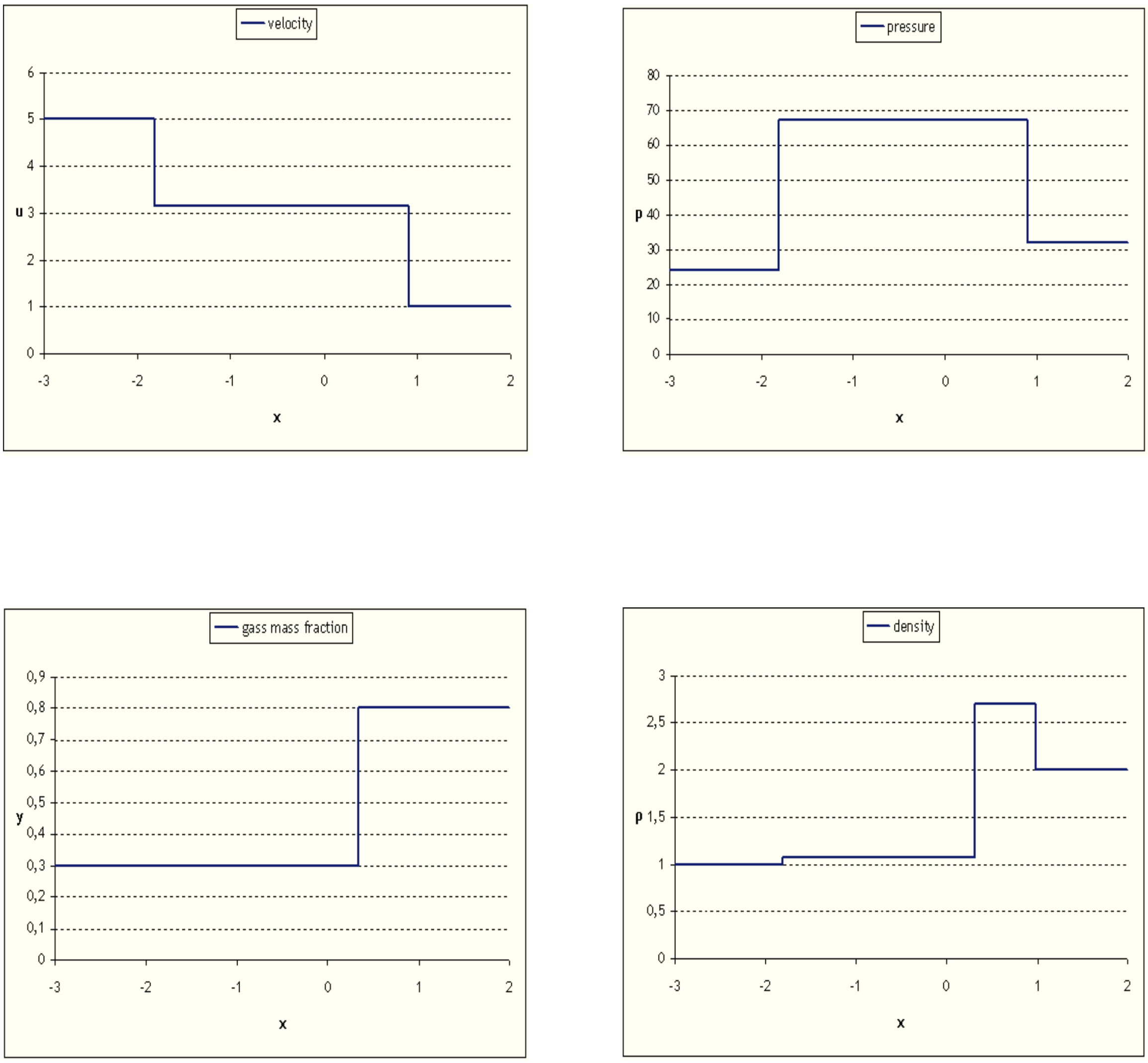

FIG. A.6 – Two-phase : shock – contact discontinuity – shock - Exact solution at $t = 0.1$.

Annexe

B Staggered discretizations, pressure corrections schemes and all speed barotropic flows

We present in this paper a class of schemes for the solution of the barotropic Navier-Stokes equations. These schemes work on general meshes, preserve the stability properties of the continuous problem, irrespectively of the space and time steps, and boil down, when the Mach number vanishes, to discretizations which are standard (and stable) in the incompressible framework. Finally, we show that they are able to capture solutions with shocks to the Euler equations.

Plan du Chapitre B

B.1 Introduction **149**

B.2 The schemes : general form **149**

B.2.1 Meshes and unknowns 149

B.2.2 The schemes 151

B.3 The stability issue and consequences **153**

B.3.1 A stability result for the convection 153

B.3.2 Discretization of the convection term 154

B.3.2.a MAC scheme 154

B.3.2.b Rannacher-Turek element 155

B.4 Schemes and stability estimates **157**

B.5 Euler equations and solutions with shocks **160**

B.6 Discussion and perspectives **160**

B.1 Introduction

The problem addressed in this paper is the system of the so-called barotropic compressible Navier-Stokes equations, which reads :

$$\partial_t \bar{\rho} + \operatorname{div}(\bar{\rho}\bar{\boldsymbol{u}}) = 0, \tag{B.1a}$$

$$\partial_t(\bar{\rho}\bar{\boldsymbol{u}}) + \operatorname{div}(\bar{\rho}\bar{\boldsymbol{u}} \otimes \bar{\boldsymbol{u}}) + \boldsymbol{\nabla}\bar{p} - \operatorname{div}(\boldsymbol{\tau}(\bar{\boldsymbol{u}})) = 0, \tag{B.1b}$$

$$\bar{\rho} = \wp(\bar{p}), \tag{B.1c}$$

where t stands for the time, $\bar{\rho}$, $\bar{\boldsymbol{u}}$ and $\bar{p}$ are the density, velocity and pressure in the flow, and $\boldsymbol{\tau}(\boldsymbol{u})$ stands for the shear stress tensor. The function $\wp(\cdot)$ is the equation of state used for the modelling of the particular flow at hand, which may be the actual equation of state of the fluid or may result from assumptions concerning the flow ; typically, laws as $\wp(\bar{p}) = \bar{p}^{1/\gamma}$, where γ is a coefficient which is specific to the considered fluid, are obtained by making the assumption that the flow is isentropic. This system of equations is posed over $\Omega \times (0,T)$, where Ω is a domain of $\mathbb{R}^d$, $d \leq 3$ supposed to be polygonal ($d = 2$) or polyhedral ($d = 3$), and the final time T is finite. We suppose that the boundary of Ω is split into $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$, and we suppose that the velocity and density are prescribed on $\partial\Omega_D$, while Neumann boundary conditions are prescribed on $\partial\Omega_D$. The flow is assumed to enter the domain on $\partial\Omega_D$ and to leave it on Ω_N. This system must be supplemented by initial conditions for $\bar{\rho}$ and $\bar{\boldsymbol{u}}$.

The objective of this paper is to present a class of schemes which enjoy three essential features. First, these schemes work on quite general two and three dimensional meshes, including locally refined non-conforming (*i.e.* with hanging nodes) discretizations. Second, they respect the (expected) stability properties of the continuous problem at hand, irrespectively of the space and time step : positivity of the density, conservation of mass, energy inequality. Third, they boil down, for vanishing Mach numbers, to usual stable coupled or pressure correction schemes, which means that the discretization enjoys a discrete *inf-sup* condition. Note, even if this aspect is left beyond the scope of this paper, that this implies that a control of the pressure will be obtained through a control of its gradient ; this property is used as a central argument to obtain convergence results on model problems [21, 18, 17].

This paper is organized as follows. First, we describe the general form of the schemes (Section B.2). Then we show how stability requirements are taken into account to design the discretization of the velocity convection term (Section B.3). The final expression for the schemes is given in Section B.4, and their stability properties are stated. Finally, we discuss their capability to capture solutions of the Euler equations with shocks (Section B.5).

B.2 The schemes : general form

B.2.1 Meshes and unknowns

A finite volume mesh of Ω is defined by a set $\mathcal{M}$ of non–empty convex open disjoint subsets K of Ω (the control volumes), such that $\bar{\Omega} = \bigcup_{K \in \mathcal{M}} \bar{K}$. We denote by $\mathcal{E}$ the set of edges (in 2D) or faces (in 3D), by $\mathcal{E}(K) \subset \mathcal{E}$ the set of faces of the cell $K \in \mathcal{M}$, by $\mathcal{E}_{\text{ext}}$ and $\mathcal{E}_{\text{int}}$ the set of boundary and interior faces, respectively. The set of external faces $\mathcal{E}_{\text{ext}}$ is split in $\mathcal{E}_N$ and $\mathcal{E}_D$, which stand for the set of the faces included in $\partial\Omega_N$ and $\partial\Omega_D$, respectively. Each internal face, denoted by $\sigma \in \mathcal{E}_{\text{int}}$, is supposed to have exactly two neighboring cells, say $K,\ L \in \mathcal{M}$, and $\bar{K} \cap \bar{L} = \bar{\sigma}$ which we denote by $\sigma = K|L$. By analogy, we write $\sigma = K|\text{ext}$ for an external face σ of K, even if this notation is somewhat incorrect, since K may have more than one external edge. The mesh $\mathcal{M}$ will be referred to hereafter as the "primal mesh".

The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-measure of the face σ.

Then, for $\sigma \in \mathcal{E}$ and $K \in \ mesh$ such that $\sigma \in \mathcal{E}(K)$ (in fact, the only cell if $\sigma \in \mathcal{E}_{\text{ext}}$ and one among the two possible ones if $\sigma \in \mathcal{E}_{\text{int}}$), we denote by $D_{K,\sigma}$ a subvolume of K having σ as a face (see Figure B.1), and by $|D_{K,\sigma}|$ the measure of $D_{K,\sigma}$. For $\sigma \in \mathcal{E}_{\text{int}}$, $\sigma = K|L$, we set $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$, so $|D_\sigma| = |D_{K,\sigma}| + |D_{L,\sigma}|$ (see Figure B.1), and for $\sigma \in \mathcal{E}_{\text{ext}}, \sigma = K|\text{ext}$, $D_\sigma = D_{K,\sigma}$, so $|D_\sigma| = |D_{K,\sigma}|$. The set of faces of the dual cell D_σ is denoted by $\bar{\mathcal{E}}(D_\sigma)$, and the face separating two adjacent dual cells D_σ and $D_{\sigma'}$ is denoted by $\varepsilon = \sigma|\sigma'$.

For $1 \le i \le d$, the degree of freedom for the i^{th} component of the velocity are assumed to be associated to a subset of $\mathcal{E}$, denoted by $\mathcal{E}^{(i)} \subset \mathcal{E}$, and are denoted by :

$$\big\{\boldsymbol{u}_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)}\big\}.$$

The sets of internal, external, Neumann and Dirichlet faces associated to the component i are denoted by $\mathcal{E}^{(i)}_{\text{int}}$, $\mathcal{E}^{(i)}_{\text{ext}}$, $\mathcal{E}^{(i)}_{N}$ and $\mathcal{E}^{(i)}_{D}$ (so, for instance, $\mathcal{E}^{(i)}_{\text{int}} = \mathcal{E}_{\text{int}} \cap \mathcal{E}^{(i)}$). We consider the following assumption :

$$\text{(H1)} \qquad \text{for } 1 \le i \le d,\ \forall K \in \mathcal{M}, \qquad \cup_{\sigma \in \mathcal{E}^{(i)} \cap \mathcal{E}(K)} \overline{D}_{K,\sigma} = \overline{K}$$

$$\text{and} \qquad \sum_{\sigma \in \mathcal{E}^{(i)} \cap \mathcal{E}(K)} |D_{K,\sigma}| = |K|,$$

which means that the volumes $D_{K,\sigma}$, $\sigma \in \mathcal{E}^{(i)}$, are disjoint, and that, for $1 \le i \le d$, $(D_\sigma)_{\sigma \in \mathcal{E}^{(i)}}$ is a partition of Ω. The sets of faces, internal faces and Neumann faces of this dual mesh are denoted by $\bar{\mathcal{E}}^{(i)}$, $\bar{\mathcal{E}}^{(i)}_{\text{int}}$ and $\bar{\mathcal{E}}^{(i)}_{N}$ respectively.

We suppose that the degrees of freedom for the pressure and the density are associated to primal cells, so they read

$$\big\{p_K,\ K \in \mathcal{M}\big\}, \quad \big\{\rho_K,\ K \in \mathcal{M}\big\}.$$

We denote by $\boldsymbol{V}$ the approximation space for the velocity, by $\boldsymbol{V}^{(i)}$, $1 \le i \le d$ the approximation spaces for the velocity components and by Q the approximation space for the pressure and the density, and we identify the discrete functions to their degrees of freedom :

$$\forall \boldsymbol{v} \in \boldsymbol{V},\ \boldsymbol{v}_i \in \boldsymbol{V}^{(i)},\ 1 \le i \le d \text{ and } \boldsymbol{v}_i = (\boldsymbol{v}_{\sigma,i})_{\sigma \in \mathcal{E}^{(i)}}; \quad \forall q \in Q,\ q = (q_K)_{K \in \mathcal{M}}.$$

For the velocity, since the concerned degrees of freedom at located on the boundary, the Dirichlet boundary conditions are enforced in the approximation space :

$$\text{For } 1 \le i \le d,\ \forall \sigma \in \mathcal{E}^{(i)}_{D}, \quad \boldsymbol{u}_{\sigma,i} = \frac{1}{|\sigma|} \int_\sigma \boldsymbol{u}_{D,i}\, \mathrm{d}\gamma,$$

where $\boldsymbol{u}_{D,i}$ stands for the i^{th} component of the prescribed velocity.

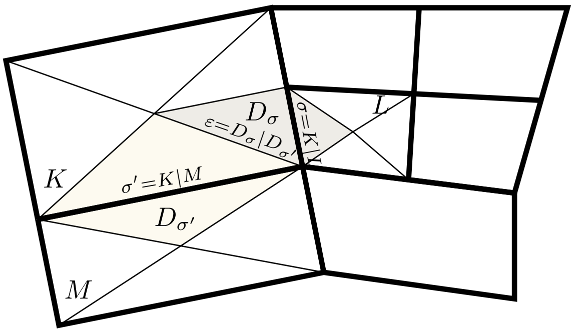

FIG. B.1 – Notations for control volumes and diamond cells.

B.2.2 The schemes

We now introduce the following notations and assumptions :

- for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, by $\boldsymbol{u}\cdot\boldsymbol{n}_{K,\sigma}$, we denote an approximation of the normal velocity to the face σ outward K,
- for $\boldsymbol{v} \in \boldsymbol{V}$, $1 \le i \le d$ and $\sigma \in \mathcal{E}^{(i)}$, we denote by $(\mathrm{div}\tau(\boldsymbol{v}))_\sigma^{(i)}$ an approximation of the viscous term associated to σ and to the component i, and we suppose that the following assumption is satisfied :

$$\text{(H2)} \qquad \sum_{i=1}^{d} \sum_{\sigma\in\mathcal{E}^{(i)}} |D_\sigma|\ (\mathrm{div}\tau(\boldsymbol{v}))_\sigma^{(i)}\ \boldsymbol{v}_{\sigma,i} \ge 0.$$

- for $q \in Q$, $1 \le i \le d$ and $\sigma \in \mathcal{E}^{(i)}$, we denote by $(\boldsymbol{\nabla} q)_\sigma^{(i)}$ the component i of the discrete gradient of q at the face σ, and we suppose that the following assumption is satisfied for any $q \in Q$ and $\boldsymbol{v} \in \boldsymbol{V}$:

$$\text{(H3)} \qquad \sum_{i=1}^{d} \sum_{\sigma\in\mathcal{E}^{(i)}} |D_\sigma|\ (\boldsymbol{\nabla} q)_\sigma^{(i)}\ \boldsymbol{v}_{\sigma,i} = \sum_{K\in\mathcal{M}} q_K \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\ \boldsymbol{v}\cdot\boldsymbol{n}_{K,\sigma}.$$

With these notations, we are able to write the general form of the implicit scheme :

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K - \rho_K^*) + \sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma} = 0. \tag{B.2a}$$

$$\text{For } 1 \le i \le d,\ \forall \sigma \in \mathcal{E}_{\mathrm{int}}^{(i)} \cup \mathcal{E}_N^{(i)},$$
$$\frac{|D_\sigma|}{\delta t}(\rho_\sigma \boldsymbol{u}_{\sigma,i} - \rho_\sigma^* \boldsymbol{u}_{\sigma,i}^*) + \sum_{\varepsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}\boldsymbol{u}_{\varepsilon,i} \tag{B.2b}$$
$$+|D_\sigma|\,(\boldsymbol{\nabla} p)_\sigma^{(i)} + |D_\sigma|\,(\mathrm{div}\tau(\boldsymbol{u}))_\sigma^{(i)} = 0,$$

$$\forall K \in \mathcal{M}, \qquad \rho_K = \wp(p_K), \tag{B.2c}$$

where $F_{K,\sigma}$ stands for the mass flux leaving K through σ, ρ_σ stands for an approximation of the density at the face, and $F_{\sigma,\varepsilon}$ is a mass flux leaving D_σ through ε. For the flux $F_{K,\sigma}$ at the internal edge $\sigma = K|L$, we choose an upwind approximation of the density :

$$F_{K,\sigma} = |\sigma|\ \boldsymbol{u}\cdot\boldsymbol{n}_{K,\sigma}\ \rho_\sigma^{\mathrm{up}}, \quad \text{with } \rho_\sigma^{\mathrm{up}} = \rho_K \text{ if } F_{K,\sigma} \ge 0,\ \rho_\sigma^{\mathrm{up}} = \rho_L \text{ otherwise.} \tag{B.3}$$

On $\sigma \in \mathcal{E}_D$, the density $\rho_\sigma^{\rm up}$ is given by the boundary condition, and, on $\sigma \in \mathcal{E}_N$, $\sigma = K|{\rm ext}$, $\rho_\sigma^{\rm up} = \rho_K$, which, since the flow is supposed to enter the domain on $\partial\Omega_D$ and to leave the domain on $\partial\Omega_N$, is consistent with the upwind choice. For the velocity components at the dual edges, $\boldsymbol{u}_{\varepsilon,i}$, we choose either the centered or upwind approximation on the internal faces, and the value at the face for the outflow ones.

A pressure correction scheme is obtained from (B.2) by splitting the resolution in two steps :

1- Velocity prediction step – Solve for $\tilde{\boldsymbol{u}} \in \boldsymbol{V}$ the momentum balance equation with the beginning-of-step pressure :

$$\begin{aligned}
&\text{For } 1 \leq i \leq d,\ \forall \sigma \in \mathcal{E}_{\rm int}^{(i)} \cup \mathcal{E}_N^{(i)}, \\
&\qquad \frac{|D_\sigma|}{\delta t}(\rho_\sigma \tilde{\boldsymbol{u}}_{\sigma,i} - \rho_\sigma^* \boldsymbol{u}_{\sigma,i}^*) + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon} \tilde{\boldsymbol{u}}_{\varepsilon,i} \\
&\qquad\qquad + |D_\sigma|\,(\boldsymbol{\nabla} p^*)_\sigma^{(i)} + |D_\sigma|\,({\rm div}\tau(\tilde{\boldsymbol{u}}))_\sigma^{(i)} = 0,
\end{aligned} \tag{B.4}$$

2 - Correction step – Solve for $\boldsymbol{u} \in \boldsymbol{V}$ and $p \in Q$:

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho_K - \rho_K^*) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma} = 0. \tag{B.5a}$$

$$\begin{aligned}
&\text{For } 1 \leq i \leq d,\ \forall \sigma \in \mathcal{E}_{\rm int}^{(i)} \cup \mathcal{E}_N^{(i)}, \\
&\qquad \frac{|D_\sigma|}{\delta t}\, \rho_\sigma\, (\boldsymbol{u}_{\sigma,i} - \tilde{\boldsymbol{u}}_{\sigma,i}) + |D_\sigma|\, \big(\boldsymbol{\nabla}(p - p^*)\big)_\sigma^{(i)} = 0,
\end{aligned} \tag{B.5b}$$

$$\forall K \in \mathcal{M}, \qquad \rho_K = \wp(p_K), \tag{B.5c}$$

The equations of the correction step are combined to produce a nonlinear parabolic problem for the pressure, which reads, $\forall K \in \mathcal{M}$:

$$\begin{aligned}
\frac{|K|}{\delta t}\, \big(\wp(p_K) - \rho_K^*\big) + \sum_{\sigma = K|L} \frac{\rho_\sigma^{\rm up}}{\rho_\sigma}\, \frac{|\sigma|^2}{|D_\sigma|}(\phi_K - \phi_L) + \sum_{\sigma \in \mathcal{E}(K) \cap \mathcal{E}_N} \frac{\rho_\sigma^{\rm up}}{\rho_\sigma}\, \frac{|\sigma|^2}{|D_\sigma|}\phi_K \\
= \frac{1}{\delta t} \sum_{\sigma \in \mathcal{E}(K)} |\sigma|\, \rho_\sigma^{\rm up} \tilde{\boldsymbol{u}} \cdot \boldsymbol{n}_{K,\sigma},
\end{aligned} \tag{B.6}$$

where $\phi \in Q$ is defined by $\phi = p - p^*$. Note that the second and third terms at the left-hand side look like a finite volume discretization of a diffusion operator, with homogeneous Neumann boundary conditions on $\mathcal{E}_D$ and Dirichlet boundary conditions on $\mathcal{E}_N$ for the pressure increment, as usual in pressure correction schemes (see [12] for a discussion on the effect on these spurious boundary conditions).

The standard discretizations entering the present framework are either low-degree non-conforming finite elements, namely the Crouzeix-Raviart element [11] for simplicial meshes or the Rannacher-Turek element [65] for quadrangles and hexahedra, or, for structured cartesian grids, the MAC scheme [37, 36]. We describe here the construction of the diffusion and pressure gradient term for the finite element schemes, supposing for the sake of simplicity that the velocity obeys homogeneous Dirichlet boundary conditions on $\partial\Omega$. Let $\sigma \in \mathcal{E}_{\rm int}$ and φ_σ be the finite element shape function associated to σ. In Rannacher-Turek or

Crouzeix-Raviart elements, a degree of freedom for each component of the velocity is associated to each edge, so $\mathcal{E}_{\rm int}^{(i)} = \mathcal{E}_{\rm int}$, for $1 \leq i \leq d$. Let $1 \leq i \leq d$ be given, let $\boldsymbol{e}^{(i)}$ be the i^{th} vector of the canonical basis of $\mathbb{R}^d$ and let us define $\boldsymbol{\varphi}_\sigma^{(i)}$ by :

$$\boldsymbol{\varphi}_\sigma^{(i)} = \varphi_\sigma \, \boldsymbol{e}^{(i)}.$$

Then the usual finite element discretization reads, for a constant viscosity Newtonian fluid (so supposing $\mathrm{div}\boldsymbol{\tau}(\boldsymbol{u}) = \mu\Delta\boldsymbol{u} + (\mu/3)\boldsymbol{\nabla}\mathrm{div}(\boldsymbol{u})$, with μ the viscosity) :

$$(\mathrm{div}\boldsymbol{\tau}(\boldsymbol{u}))_\sigma^{(i)} = \sum_{K\in\mathcal{M}} \mu \int_K \boldsymbol{\nabla}\boldsymbol{u} : \boldsymbol{\nabla}\boldsymbol{\varphi}_\sigma^{(i)} \,\mathrm{d}\boldsymbol{x} + \frac{u}{3}\int_K \mathrm{div}\boldsymbol{u}\ \mathrm{div}\boldsymbol{\varphi}_\sigma^{(i)} \,\mathrm{d}\boldsymbol{x}.$$

The pressure gradient term at the internal face $\sigma = K|L$ reads :

$$(\boldsymbol{\nabla}p)_\sigma^{(i)} = \sum_{K\in\mathcal{M}} \int_K p \ \mathrm{div}\boldsymbol{\varphi}_\sigma^{(i)} \,\mathrm{d}\boldsymbol{x} = |\sigma| \ (p_L - p_K) \ \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}^{(i)}.$$

B.3 The stability issue and consequences

B.3.1 A stability result for the convection

At the continuous level, let us assume that the mass balance $\partial_t\rho + \mathrm{div}(\boldsymbol{\beta}) = 0$ holds, with $\boldsymbol{\beta}$ a regular vector-valued function. Then, for all scalar regular functions u and v, we have :

$$\int_\Omega \big[\partial_t(\rho u) + \mathrm{div}(u\boldsymbol{\beta})\big] \ v \,\mathrm{d}\boldsymbol{x} = \\ \int_\Omega \big[\partial_t(\rho u) - \frac{1}{2}\,(\partial_t\rho)\,u\big] \ v \,\mathrm{d}\boldsymbol{x} + s(u,v) + \frac{1}{2}\int_{\partial\Omega_N} u\,v\boldsymbol{\beta}\cdot\boldsymbol{n} \,\mathrm{d}\gamma \quad \text{(B.7)}$$

where s is the following skew-symmetric bilinear form :

$$s(u,v) = \frac{1}{2}\int_\Omega v\boldsymbol{\beta}\cdot\boldsymbol{\nabla}u \,\mathrm{d}\boldsymbol{x} - \frac{1}{2}\int_\Omega u\boldsymbol{\beta}\cdot\boldsymbol{\nabla}v \,\mathrm{d}\boldsymbol{x}.$$

Taking $u = v = \boldsymbol{u}_i$ and summing over i, the first term gives the time derivative of the kinetic energy, the second one vanishes and the last one corresponds to the kinetic energy flux through the boundary of the domain. The following Lemma, proven in [48], states a discrete counterpart of this computation (see also [1] and [22] for a direct estimate of the kinetic energy, for an implicit and explicit scheme respectively).

Lemma B.3.1

Let us suppose that, for an index i, $1 \leq i \leq d$, the following discrete mass balance holds over the dual cells associated to the i^{th} component of the velocity :

$$\forall\sigma \in \mathcal{E}_{\rm int}^{(i)} \cup \mathcal{E}_N^{(i)}, \qquad \frac{|D_\sigma|}{\delta t}(\rho_\sigma - \rho_\sigma^*) + \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon} = 0. \quad \text{(B.8)}$$

Let $u, v \in V^{(i)}$, and let us suppose that these discrete functions obey homogeneous Dirichlet boundary. Then we have :

$$\sum_{\mathcal{E}\in\mathcal{E}_{\rm int}^{(i)}\cup\mathcal{E}_N^{(i)}} v_\sigma \ \Big[\frac{|D_\sigma|}{\delta t}(\rho_\sigma u_\sigma - \rho_\sigma^* u_\sigma^*) + \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)} F_{\sigma,\varepsilon}\boldsymbol{u}_\varepsilon\Big] \\ \geq T_{\Omega,\mathrm{k}}(u,v) + T_{\Omega,\mathrm{s}}(u,v) + T_{\partial\Omega}(u,v), \quad \text{(B.9)}$$

with :

$$\begin{aligned}
T_{\Omega,\mathrm{k}}(u,v) &= \sum_{\mathcal{E}\in\mathcal{E}^{(i)}_{\mathrm{int}}\cup\mathcal{E}^{(i)}_N} \frac{|D_\sigma|}{\delta t}\,(\rho_\sigma u_\sigma - \rho^*_\sigma u^*_\sigma)\,v_\sigma - \frac{1}{2}\,(\rho_\sigma - \rho^*_\sigma)\,u_\sigma\,v_\sigma,\\
T_{\Omega,\mathrm{s}}(u,v) &= S(u,v) - S(v,u), \qquad S(u,v) = \frac{1}{2} \sum_{\varepsilon\in\bar{\mathcal{E}}^{(i)}_{\mathrm{int}},\ \varepsilon=D_\sigma|D_{\sigma'}} F_{\sigma,\varepsilon}\,v_\varepsilon\,(u_{\sigma'} - u_\sigma),\\
T_{\partial\Omega}(u,v) &= \frac{1}{2} \sum_{\varepsilon\in\bar{\mathcal{E}}^{(i)}_N,\ \sigma=D_\sigma|\mathrm{ext}} F_{\sigma,\varepsilon}\,u_\varepsilon\,v_\varepsilon.
\end{aligned}$$

Of course, $T_{\Omega,\mathrm{s}}(u,u) = 0$, and an easy computation shows that :

$$T_{\Omega,\mathrm{k}}(u,v) = \frac{1}{2\delta t} \sum_{\mathcal{E}\in\mathcal{E}^{(i)}_{\mathrm{int}}\cup\mathcal{E}^{(i)}_N} |D_\sigma|\,\bigl[\rho_\sigma u^2_\sigma - \rho^*_\sigma (u^*_\sigma)^2\bigr].$$

Applying Lemma B.3.1 to each component of the velocity, the obtained term is thus the discrete time-derivative of the kinetic energy, and may be used to obtain stability estimates for the scheme (see Section B.4).

Remark 5 (Non-homogeneous Dirichlet boundary conditions)
The limitation to homogeneous Dirichlet boundary conditions may be seen, from the proof, to stem from the fact that no balance equation is written on the dual cells associated to edges lying on $\partial\Omega_D$. The problem thus may be fixed by keeping these degrees of freedom and using a penalization technique.

Remark 6 (Artificial boundary conditions)
Lemma B.3.1 may be used to derive artificial boundary conditions allowing the flow to enter the domain through $\partial\Omega_N$, by first collecting the boundary terms in the variational form of the momentum balance equation (*i.e.* adding to $T_{\partial\Omega}(u,v)$ the terms issued from the diffusion and the pressure gradient) and then imposing that the result may be written as a linear form acting on the test function (see [6] for a similar development in the incompressible case). The so-built boundary condition is observed in practice to give quite good results when modelling external flows [48].

B.3.2 Discretization of the convection term

The problem to tackle is now the following one : on one side, the discrete mass balance over the dual cells (B.8) is necessary for the stability of the scheme ; on the other side, the mass balance is only written by the scheme(s) for the primal cells (Equation (B.2a) or (B.5a)). We are thus lead to express the mass fluxes $(F_{\sigma,\varepsilon})$ through the dual edges as a function of the mass fluxes $(F_{K,\sigma})$ through the primal ones, in such a way that the discrete balance over the primal cells implies the same property over the dual ones. We describe in this section how this may be done, first for the MAC (structured) mesh and second for the Rannacher-Turek element on general quadrangles.

B.3.2.a MAC scheme

We describe a possible construction of the momentum convection operator for the MAC scheme [38]. In two space dimensions and with the local notations introduced on Figure B.2, the mass balance on the

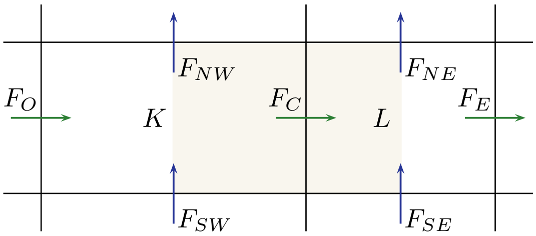

FIG. B.2 – Local notations for the definition of the mass fluxes at the dual edges with the MAC scheme.

primal cells reads :

$$\begin{aligned} K: \quad & \frac{|K|}{\delta t}\ (\varrho_K - \varrho_K^*) - F_O - F_{SW} + F_C + F_{NW} = 0, \\ L: \quad & \frac{|L|}{\delta t}\ (\varrho_L - \varrho_L^*) - F_C - F_{SE} + F_E + F_{NE} = 0. \end{aligned}$$

Multiplying both equations by $1/2$ and summing them yields, for $\sigma = K|L$:

$$\frac{|D_\sigma|}{\delta t}\ (\varrho_\sigma - \varrho_\sigma^*) \\ - \frac{1}{2}\,\big[F_W + F_C\big] - \frac{1}{2}\,\big[F_{SW} + F_{SE}\big] + \frac{1}{2}\,\big[F_C + F_E\big] + \frac{1}{2}\,\big[F_{NW} + F_{NE}\big] = 0, \quad \text{(B.10)}$$

with the usual definition of the dual cell D_σ, which implies that $|D_{K,\sigma}| = |K|/2$ and $|D_{L,\sigma}| = |L|/2$, and with the following definition of the density on the face :

$$|D_\sigma|\ \varrho_\sigma = |D_{K,\sigma}|\ \varrho_K + |D_{L,\sigma}|\ \varrho_L. \quad \text{(B.11)}$$

Equation (B.10) thus suggests the following definition for the mass fluxes at the dual faces :

$$\begin{aligned} &\text{left face :} && F_{\sigma,\varepsilon} = -\frac{1}{2}\,\big[F_W + F_C\big]; && \text{right face :} && F_{\sigma,\varepsilon} = \frac{1}{2}\,\big[F_C + F_E\big]; \\ &\text{bottom face :} && F_{\sigma,\varepsilon} = -\frac{1}{2}\,\big[F_{SW} + F_{SE}\big]; && \text{top face :} && F_{\sigma,\varepsilon} = \frac{1}{2}\,\big[F_{NW} + F_{NE}\big]. \end{aligned}$$

Note that this definition is rather non-standard : for instance, the flux at the left face of $D_{K|L}$, which is included in K, may involve densities of the neighbouring primal cells. The extension of the above construction to the three-dimensional case is straightforward.

B.3.2.b Rannacher-Turek element

A construction similar to the MAC scheme one may be performed for rectangular meshes. For K and L two neighboring cells of $\mathcal{M}$, the half-diamond cell $D_{K,\sigma}$ (resp. $D_{L,\sigma}$) associated to the common face $\sigma = K|L$ is defined as the cone having the mass center of K (resp. L) as a vertex and σ as basis, the density ρ_σ is defined by the weighted average (B.11), and the dual mass fluxes are obtained by multiplying the mass balances over K and L by $1/4$ and summing. With the local notations of Figure B.3, this yields, for the dual mass fluxes, expressions of the form :

$$F_{\sigma,\varepsilon} = -\frac{1}{8}F_W + \frac{3}{8}F_N - \frac{3}{8}F_E + \frac{1}{8}F_S. \quad \text{(B.12)}$$

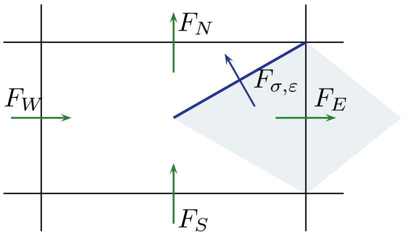

FIG. B.3 – Local notations for the definition of the mass fluxes at the dual edges with the Rannacher-Turek element

We now explain how to extend this formulation to general meshes.

Let us suppose that we are able to define the fluxes through the dual faces in such a way that :

(A1) The mass balance over the half-diamond cells is proportional to the mass balance over the primal cells, in the following sense :

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}(K), \qquad F_{K,\sigma} + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma),\ \varepsilon \subset K} F_{\sigma,\varepsilon} = \xi_K^\sigma \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma},$$

with, for any cell $K \in \mathcal{M}$, $\displaystyle\sum_{\sigma \in \mathcal{E}(K)} \xi_K^\sigma = 1$ and, for any $\sigma \in \mathcal{E}(K)$, $\xi_K^\sigma \geq 0$.

(A2) The dual fluxes are conservative, *i.e.* for any dual face $\varepsilon = D_\sigma | D'_\sigma$, we have $F_{\sigma,\varepsilon} = -F_{\sigma',\varepsilon}$.

(A3) The dual fluxes are bounded with respect to the $(F_{K,\sigma})_{\sigma \in \mathcal{E}(K)}$:

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}(K),\ \forall \varepsilon \in \bar{\mathcal{E}}(D_\sigma) \quad |F_{\sigma,\varepsilon}| \leq C\ \max\Big\{|F_{K,\sigma}|,\ \sigma \in \mathcal{E}(K)\Big\}.$$

In addition, let us define $|D_{K,\sigma}|$ as :

$$|D_{K,\sigma}| = \xi_K^\sigma\ |K|, \tag{B.13}$$

and ρ_σ, once again, by the weighted average (B.11). Then the dual fluxes satisfy the required mass balance. Indeed, for $\sigma \in \mathcal{E}_{\text{int}}$, $\sigma = K|L$, we have :

$$\begin{aligned}
&\frac{|D_\sigma|}{\delta t}\ (\rho_\sigma - \rho_\sigma^*) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\sigma,\varepsilon} \\
&= \frac{|D_{K,\sigma}|}{\delta t}\ (\rho_K - \rho_K^*) + F_{K,\sigma} + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma),\ \varepsilon \subset K} F_{\sigma,\varepsilon} \\
&\qquad + \frac{|D_{L,\sigma}|}{\delta t}\ (\rho_L - \rho_L^*) + F_{L,\sigma} + \sum_{\varepsilon \in \bar{\mathcal{E}}(D_\sigma),\ \varepsilon \subset L} F_{\sigma,\varepsilon} \\
&= \xi_K^\sigma\ \Big[\frac{|K|}{\delta t}\ (\rho_K - \rho_K^*) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}\Big] + \xi_L^\sigma\ \Big[\frac{|L|}{\delta t}\ (\rho_L - \rho_L^*) + \sum_{\sigma \in \mathcal{E}(L)} F_{L,\sigma}\Big] = 0.
\end{aligned}$$

A similar computation leads to the same conclusion for the (half-)dual cells associated to the Neumann boundary faces.

The next issue is to check whether Assumptions (A1)-(A3) are sufficient for the consistency of the scheme. In this respect, the following lemma brings a decisive argument.

Lemma B.3.2
Let Assumptions (A1)-(A3) hold. For $v \in V$ and $K \in \mathcal{M}$, let v_K be defined by $v_K = \sum_{\sigma\in\mathcal{E}(K)} \xi_K^\sigma\, v_\sigma$. Let $u \in V$, and $R(u,v)$ be the quantity defined by :

$$R(u,v) = \sum_{\sigma\in\mathcal{E}_{\rm int}} v_\sigma \sum_{\substack{\varepsilon\in\bar{\mathcal{E}}(D_\sigma),\\ \varepsilon=D_\sigma|D'_\sigma}} F_{\sigma,\varepsilon}\, \frac{u_\sigma + u_{\sigma'}}{2} - \sum_{K\in\mathcal{M}} v_K \sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}\; u_\sigma.$$

Let us suppose that the primal fluxes are associated to a convection momentum field $\boldsymbol{\beta}$, *i.e.* :

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}(K), \quad F_{K,\sigma} = |\sigma|\ \boldsymbol{\beta}_\sigma \cdot \boldsymbol{n}_{K,\sigma}.$$

For the schemes used here, of course, $\boldsymbol{\beta}$ is a combination of the density and the velocity, as introduced in Section B.2 and made precise in Section B.4. Then there exists C depending only on the regularity of the mesh such that :

$$|R(u,v)| \leq C\ h\ \|\boldsymbol{\beta}\|_{l^\infty}\ \|u\|_1\ \|v\|_1,$$

with $\|\boldsymbol{\beta}\|_{l^\infty} = \max_{\sigma\in\mathcal{E}} |\boldsymbol{\beta}_\sigma|$ and the discrete H^1-norm on the dual mesh is defined by :

$$\forall v \in V, \quad \|v\|_1 = \sum_{K\in\mathcal{M}} h_K^{d-2} \sum_{\sigma,\sigma'\in\mathcal{E}(K)} (u_\sigma - u_{\sigma'})^2.$$

The quantity $R(u,v)$ compares two discrete analogue to $\int_\Omega v\,{\rm div}(u\boldsymbol{\beta})\,{\rm d}\boldsymbol{x}$, the first one being defined with the divergence taken over the dual meshes, and the second one with the divergence over the primal cells. The discrete H^1-norm of the solution is controlled by the diffusion term. Thus, when making a convergence or error analysis study in the linear case (*i.e.* with a given regular convection field $\boldsymbol{\beta}$), Lemma B.3.2 allows to replace the first formulation by the second one, thus substituting well defined quantities to quantities only defined through (A1)-(A3). It is used in [39] to prove that the scheme is first-order for the stationary convection-diffusion equation. The convergence for the constant density Navier-Stokes equations (so with $\boldsymbol{\beta} = \boldsymbol{u}$ has also been proven, controlling now $\|\boldsymbol{u}\|_{l^\infty}$ by $\|\boldsymbol{u}\|_1$ thanks to an inverse inequality.

The last task is now to build fluxes satisfying (A1)-(A3), which is easily done by choosing $\xi_K^\sigma = 1/4$, and keeping for the expression of the dual fluxes as a function of the primal ones the same linear combination (B.12) as in the rectangular case. Note that this implicitly implies that the geometrical definition of the dual cells has been generalized, since it is not possible in general to split a (even convex) quadrangle in four simplices of same measure. Extension to three dimensions only needs to deal with the rectangular parallelepipedic case, which is quite simple [1]. Finding directly a solution to (A1)-(A3) may also be an alternative route, to deal with more complex cases, as done in [39] to extend the scheme to locally refined non-conforming grids.

B.4 Schemes and stability estimates

To obtain the complete formulation of the considered schemes, we now have to fix the time-marching procedure. This is straightforwart for the implicit scheme, and we concentrate here on the pressure correction scheme. The problem which we face in this case is that the mass balance is not yet solved when dealing with the prediction step. In our implementations [40], it is circumvented by just shifting in time the density ρ_σ, and the mass balance on the dual cells is recovered from the mass balance on the primal cells at the previous time step. This has essentially two drawbacks. First, the trick indeed works only if the time step is constant; when it changes, one has to choose between loosing stability or consistency (locally in time, so fortunately, without observed impact in practice). Second, the scheme is only first order in time.

In addition, stability seems to require an initial pressure renormalization step, which is an algebraic variant of the one introduced in [28]. It seems however that this step may be omitted in practice.

The algorithm (keeping in this presentation the pressure renormalization step) reads, assuming that $\boldsymbol{u}^n$, p^n, ρ^n and the family $(F^n_{K,\sigma})$ are known :

1- Pressure renormalization step – Let $(\lambda_\sigma)_{\sigma\in\mathcal{E}_{\rm int}}$ be a family of positive real numbers, and let $-{\rm div}(\lambda\boldsymbol{\nabla})_\mathcal{M}$ be the discrete elliptic operator from Q to Q defined by, $\forall K \in \mathcal{M}$ and $q \in Q$:

$$\big[-{\rm div}(\lambda\boldsymbol{\nabla})_\mathcal{M}(q)\big]_K = \sum_{\sigma=K|L} \lambda_\sigma\ \frac{|\sigma|^2}{|D_\sigma|}(q_K - q_L) + \sum_{\sigma\in\mathcal{E}_N,\sigma=K|{\rm ext}} \lambda_\sigma\ \frac{|\sigma|^2}{|D_\sigma|} q_K.$$

Then $\tilde p^{n+1} \in Q$ is given by :

$$-{\rm div}(\frac{1}{\rho^n}\boldsymbol{\nabla})_\mathcal{M}\ (\tilde p^{n+1}) = -{\rm div}(\frac{1}{[\rho^n\,\rho^{n-1}]^{1/2}}\boldsymbol{\nabla})_\mathcal{M}\ (p^n), \tag{B.14}$$

the weights $(\rho^n_\sigma)_{\sigma\in\mathcal{E}_{\rm int}\cup\mathcal{E}_N}$ and $(\rho^{n-1}_\sigma)_{\sigma\in\mathcal{E}_{\rm int}\cup\mathcal{E}_N}$ being the densities involved in the time-derivative term of the momentum balance equation (next step of the algorithm).

2- Velocity prediction step – Solve for $\tilde{\boldsymbol{u}}^{n+1} \in \boldsymbol{V}$, for $1 \le i \le d$ and $\forall\sigma \in \mathcal{E}^{(i)}_{\rm int} \cup \mathcal{E}^{(i)}_N$:

$$\frac{|D_\sigma|}{\delta t}(\rho^n_\sigma \tilde{\boldsymbol{u}}^{n+1}_{\sigma,i} - \rho^{n-1}_\sigma \boldsymbol{u}^n_{\sigma,i}) + \sum_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)} F^n_{\sigma,\varepsilon}\tilde{\boldsymbol{u}}^{n+1}_{\varepsilon,i} \\ + |D_\sigma|\,(\boldsymbol{\nabla}\tilde p^{n+1})^{(i)}_\sigma + |D_\sigma|\,({\rm div}\tau(\tilde{\boldsymbol{u}}^{n+1}))^{(i)}_\sigma = 0, \tag{B.15}$$

where the quantity $(F^n_{\sigma,\varepsilon})_{\varepsilon\in\bar{\mathcal{E}}(D_\sigma)}$ are built as explained in the previous section, from the primal fluxes at time t^n.

3 - Correction step – Solve for $\boldsymbol{u}^{n+1} \in \boldsymbol{V}$ and $p^{n+1} \in Q$:

$$\forall K \in \mathcal{M}, \qquad \frac{|K|}{\delta t}(\rho^{n+1}_K - \rho^n_K) + \sum_{\sigma\in\mathcal{E}(K)} F^{n+1}_{K,\sigma} = 0. \tag{B.16a}$$

$$\text{For } 1 \le i \le d,\ \forall\sigma \in \mathcal{E}^{(i)}_{\rm int} \cup \mathcal{E}^{(i)}_N, \\ \frac{|D_\sigma|}{\delta t}\,\rho^n_\sigma\,(\boldsymbol{u}^{n+1}_{\sigma,i} - \tilde{\boldsymbol{u}}^{n+1}_{\sigma,i}) + |D_\sigma|\,\big(\boldsymbol{\nabla}(p^{n+1} - \tilde p^{n+1})\big)^{(i)}_\sigma = 0, \tag{B.16b}$$

$$\forall K \in \mathcal{M}, \qquad \rho^{n+1}_K = \wp(p^{n+1}_K), \tag{B.16c}$$

The algorithm must be initialized by the data of $\boldsymbol{u}^0 \in \boldsymbol{V}$, $\rho^{-1} \in Q$, $\rho^0 \in Q$ satisfying the discrete mass balance equation, and with the corresponding mass fluxes $(F^0_{K,\sigma})$. A possible way to obtain these quantities is to evaluate $\boldsymbol{u}^0$ and ρ^{-1} from the initial conditions, and, as a preliminary step, to solve for ρ^0 the mass balance equation.

The upwinding in the discretization of the mass balance equation has for consequence that any density appearing in the algorithm is positive (provided that the initial density is positive). The existence and uniqueness of a solution to Steps 1 and 2 is then clear : these are linear problems with coercive operators (for Step 2, thanks to the stability of the convection term). The existence of a solution to Step 3 may be obtained by a Brouwer fixed point argument, using the fact that the conservativity of the mass balance yields an estimate for ρ, so for p, and finally for $\boldsymbol{u}$ (in any norm, since we work on finite dimensional spaces). The algorithm is thus well-posed.

Let us now turn to the energy estimate. At the continuous level, this relation is obtained for the barotropic Navier-Stokes equations by choosing the velocity $\boldsymbol{u}$ in the variational form of the momentum balance equation, writing the convection term as the time derivative of the kinetic energy, and setting the pressure work, namely $-\int_\Omega p\,\mathrm{div}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{x}$, under a convenient form. This is done by the following formal computation. Let $b(\cdot)$ be a regular function from $(0,+\infty)$ to $\mathbb{R}$, and let us multiply the mass balance by $b'(\rho)$:

$$b'(\rho)\ \Big[\partial_t \rho + \mathrm{div}(\rho\,\boldsymbol{u})\Big] = 0.$$

Using :

$$\begin{aligned} b'(\rho)\mathrm{div}(\rho\,\boldsymbol{u}) = b'(\rho)[\boldsymbol{u}\cdot\boldsymbol{\nabla}\rho + \rho\mathrm{div}(\boldsymbol{u})] = \boldsymbol{u}\cdot\boldsymbol{\nabla} b(\rho) + \rho b'(\rho)\mathrm{div}(\boldsymbol{u}) \\ = \mathrm{div}(b(\rho)\boldsymbol{u}) + \big[\rho b'(\rho) - b(\rho)\big]\,\mathrm{div}(\boldsymbol{u}), \end{aligned}$$

we get :

$$\partial_t\big[b(\rho)\big] + \mathrm{div}\big[b(\rho)\,\boldsymbol{u}\big] + \big[\rho b'(\rho) - b(\rho)\big]\,\mathrm{div}(\boldsymbol{u}) = 0.$$

Choosing now the function $b(\cdot)$ in such a way that $\rho b'(\rho) - b(\rho) = \wp^{-1}(p)$, integrating over Ω and using the boundary conditions yields :

$$-\int_\Omega p\,\mathrm{div}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{x} = \frac{d}{dt}\int_\Omega b(\rho)\,\mathrm{d}\boldsymbol{x}.$$

The following lemma [20] states a discrete counterpart of this computation.

Lemma B.4.1

Let $b(\cdot)$ be a regular convex function from $(0,+\infty)$ to $\mathbb{R}$, and $(\rho^\star_K)_{K\in\mathcal{M}}$ be a positive family of real numbers. Then, with the upwind discretization (B.3) of the mass balance equation, the family $(\rho_K)_{K\in\mathcal{M}}$ is also positive, and we get :

$$\sum_{K\in\mathcal{M}} b'(\rho_K)\Big[\frac{|K|}{\delta t}(\rho_K - \rho^*_K) + \sum_{\sigma\in\mathcal{E}(K)} F_{K,\sigma}\Big] \geq \frac{1}{\delta t}\sum_{K\in\mathcal{M}} |K|\ \big[b(\rho_K) - b(\rho^*_K)\big].$$

We are now in position to state the following stability result.

Theorem B.4.2

The scheme (B.14)-(B.16) satisfies the following energy identity, for $1 \leq n \leq N$:

$$\begin{aligned} \sum_{i=1}^{d}\sum_{\sigma\in\mathcal{E}^{(i)}_{\rm int}} |\sigma|\,\rho_\sigma^{n-1}\ (\boldsymbol{u}^n_{\sigma,i})^2 + \delta t\sum_{k=1}^{n}\sum_{\sigma\in\mathcal{E}^{(i)}} |D_\sigma|\ (\mathrm{div}\tau(\boldsymbol{u}^k))^{(i)}_\sigma\ \boldsymbol{u}^k_{\sigma,i} \\ + \sum_{K\in\mathcal{M}} |K|\ b(\rho^n_K) \leq \sum_{i=1}^{d}\sum_{\sigma\in\mathcal{E}^{(i)}_{\rm int}} |\sigma|\,\rho_\sigma^{(-1)}\ (\boldsymbol{u}^0_{\sigma,i})^2 + \sum_{K\in\mathcal{M}} |K|\ b(\rho^0_K) \end{aligned}$$

The proof of this theorem is based on Lemma B.3.1 and Lemma B.4.1, and may be found, for the essential arguments, in [20].

Remark 7
Let us suppose that the equation of state reads $p = \rho^\gamma$, with $\gamma \in (1, +\infty)$. Then an easy computation yields $b(\rho) = \rho^\gamma/(\gamma - 1) = p/(\gamma - 1)$. Theorem B.4.2 thus yields an estimate for the pressure in $\mathrm{L}^\infty(0, T; \mathrm{L}^1)$-norm. Note that this estimate is however not sufficient to ensure that a sequence of pressures obtained as discrete solutions converges to a function, which explains that the pressure has to be controlled from estimates of its gradient, in convergence studies of numerical schemes as well as in mathematical analysis of the continuous problem [51].

B.5 Euler equations and solutions with shocks

In this section we briefly discuss the capability of the considered numerical schemes to compute irregular (*i.e.* with discontinuities) solutions of inviscid flows.

The results obtained with the above described pressure correction scheme for the so-called one-dimensional Sod shock-tube problem are displayed on Figure B.4 (see [45] for a more detailed presentation). From numerical experiments, it seems that this scheme converges when the velocity space translates are controlled, either by upwinding the discretization of the velocity convection term, or by keeping a residual viscosity in the (discrete) momentum balance equation. Numerical experiments reported in [45] (addressing also an extension of this algorithm to the barotropic homogeneous two-phase flow model [26]) confirm the stability of the scheme, and show that the qualitative behaviour of the solution is captured up to very large values of the CFL number (typically, in the range of 50).

From the theoretical point of view, for Euler equations (*i.e.* , precisely speaking, with a diffusion vanishing with the space step), the control that we are able to prove on the solution of course does not yield (weak or strong) convergence in strong enough norms to pass to the limit in the scheme. We can however prove the following result : supposing convergence for the density in $\mathrm{L}^p(\Omega)$, $p \in [1, +\infty)$ and for the velocity in $\mathrm{L}^p(\Omega)$, $p \in [1, 3]$, it is possible to pass to the limit in the discrete equations, provided that the viscosity vanishes as h^α, $\alpha \in (0, 2)$ for both the implicit and the pressure correction scheme. In this case, the limit of a sequence of discrete solutions is proven to satisfy the weak form of the Euler equations, and so, in particular, the Rankine-Hugoniot conditions at the shocks.

B.6 Discussion and perspectives

The analysis of the schemes presented here has been undertaken, for the present time for model stationary problems : in [21, 18], we prove the convergence for the Crouzeix-Raviart discretization of the Stokes equations (however, with the addition, for technical reasons, of a stabilization term) ; in [17], we prove the same result for the (this time, standard) MAC scheme. Extension, still for the MAC discretization, to the stationary Navier-Stokes equations is underway.

From a practical point of view, a next step for the barotropic Navier-Stokes equations should be to derive an upwind explicit version of the scheme presented here ; in this direction, an extension of Lemma B.3.1 (stability of the velocity convection term) to the explicit case may be found in [22].

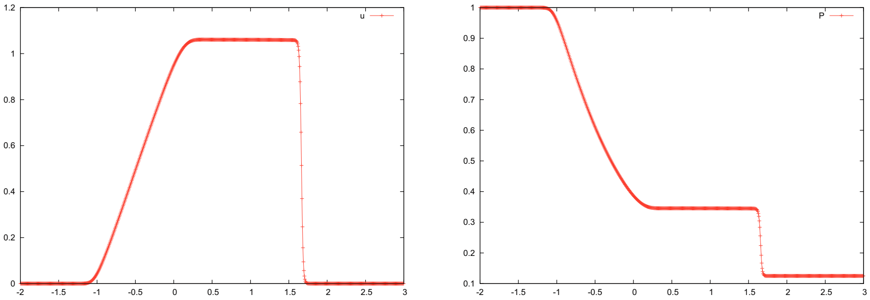

FIG. B.4 – Solution for the Sod shock-tube problem, obtained with a uniform mesh of 800 cells, with a residual viscosity – *left :* velocity, *right :* pressure.

The main objective is however to deal with the full (*i.e.* non barotropic, so including an energy balance) Navier-Stokes equations. An unconditionally stable pressure correction scheme has been derived for this problem (see Chapter 3 of this document), but extensive tests of this scheme remain to be done. In particular, stability requires that the internal energy remains non-negative (in practice, positive), and the way we obtained this property was to solve the internal energy balance, with a scheme able to preserve the sign of the unknown . . .but it is commonly agreed that, for the scheme to converge toward the correct weak solution, a conservative discretization of the total energy balance should be used. The actual occurrence of this problem, and the possibility to circumvent it, possibly by adding stabilizing viscous terms, will deserve investigations in the next future ; a preliminary step on this route may be found in [23].

Annexe

C Discretization of the viscous dissipation term with the MAC scheme

We propose a discretization for the MAC scheme of the viscous dissipation term $\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u}$ (where $\boldsymbol{\tau}(\boldsymbol{u})$ stands for the shear stress tensor associated to the velocity field $\boldsymbol{u}$), which is suitable for the approximation of this term in a conservation equation for a scalar variable. This discretization enjoys the property that the integral over the computational domain Ω of the (discrete) dissipation term is equal to what is obtained when taking the inner product of the (discrete) momentum balance equation by $\boldsymbol{u}$ and integrating over Ω. As a consequence, it may be used as an ingredient to obtain an unconditionally stable scheme for the compressible Navier-Stokes equations. It is also shown, in some model cases, to ensure the strong convergence in L^1 of the dissipation term.

Plan du Chapitre C

C.1 Introduction **165**

C.2 Discretization of the dissipation term **166**

C.2.1 The two-dimensional case 166

C.2.2 Extension to the three-dimensional case 169

C.3 A strong convergence result **170**

C.1 Introduction

Let us consider the compressible Navier-Stokes equations, which may be written as :

$$\partial_t\rho + \mathrm{div}(\rho\boldsymbol{u}) = 0, \tag{C.1a}$$

$$\partial_t(\rho\boldsymbol{u}) + \mathrm{div}(\rho\boldsymbol{u}\otimes\boldsymbol{u}) + \boldsymbol{\nabla}p - \mathrm{div}(\boldsymbol{\tau}(\boldsymbol{u})) = 0, \tag{C.1b}$$

$$\partial_t(\rho e) + \mathrm{div}(\rho e\boldsymbol{u}) + p\mathrm{div}\boldsymbol{u} + \mathrm{div}(\boldsymbol{q}) = \boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u}, \tag{C.1c}$$

$$\rho = \wp(p,e), \tag{C.1d}$$

where t stands for the time, ρ, $\boldsymbol{u}$, p and e are the density, velocity, pressure and internal energy in the flow, $\boldsymbol{\tau}(\boldsymbol{u})$ stands for the shear stress tensor, $\boldsymbol{q}$ for the temperature diffusion flux, and the function $\wp$ is the equation of state. This system of equations is posed over $\Omega\times(0,T)$, where Ω is a domain of $\mathbb{R}^d$, $d\leq 3$. This system must be supplemented by a closure relation for $\boldsymbol{\tau}(\boldsymbol{u})$ and for q, assumed to be :

$$\boldsymbol{\tau}(\boldsymbol{u}) = \mu(\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}^t\boldsymbol{u}) - \frac{2\mu}{3}\,\mathrm{div}\boldsymbol{u}\,I, \quad \boldsymbol{q} = -\lambda\boldsymbol{\nabla}e, \tag{C.2}$$

where μ and λ stand for two (possibly depending on $\boldsymbol{x}$) positive parameters.

Let us suppose, for the sake of simplicity, that $\boldsymbol{u}$ is prescribed to zero on the whole boundary, and that the system is adiabatic, *i.e.* $\boldsymbol{\nabla q}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. Then, formally, taking the inner product of (C.1b) with $\bar{\boldsymbol{u}}$ and integrating over Ω, integrating (C.1c) over Ω, and, finally, summing both relations yields the stability estimate :

$$\frac{d}{dt}\int_\Omega \big[\frac{1}{2}\rho\,|\boldsymbol{u}|^2 + \rho e\big]\,\mathrm{d}\boldsymbol{x} \leq 0. \tag{C.3}$$

If we suppose that the equation of state may be set under the form $p = f(\rho,e)$ with $f(\cdot,0) = 0$ and $f(0,\cdot) = 0$, Equation (C.1c) implies that e remains positive (still at least formally), and so (C.3) yields a control on the unknown. Mimicking this computation at the discrete level necessitates to check some arguments, among them :

(*i*) to have at disposal a discrete counterpart to the relation :

$$\int_\Omega \big[\partial_t(\rho\boldsymbol{u}) + \mathrm{div}(\rho\boldsymbol{u}\otimes\boldsymbol{u})\big]\cdot\boldsymbol{u}\,\mathrm{d}\boldsymbol{x} = \frac{d}{dt}\int_\Omega \frac{1}{2}\rho\,|\boldsymbol{u}|^2\,\mathrm{d}\boldsymbol{x}.$$

(*ii*) to identify the integral of the dissipation term at the right-hand side of the discrete counterpart of (C.1c) with what is obtained from the (discrete) L^2 inner product between the velocity and the diffusion term in the discrete momentum balance equation (C.1b).

(*iii*) to be able to prove that the right-hand side of (C.1c) is non-negative, to preserve the positivity of the internal energy.

The point (*i*) is extensively discussed in [25] (see also [38]), and will not be treated here. Describing a way, implemented in the ISIS free software developed at IRSN [40], to obtain the two other issues with the usual Marker and Cell (MAC) discretization [37, 36] is the objective of this paper. We complete the presentation by showing how (*ii*) may also be used, in some model problems, to prove the convergence in L^1 of the dissipation term.

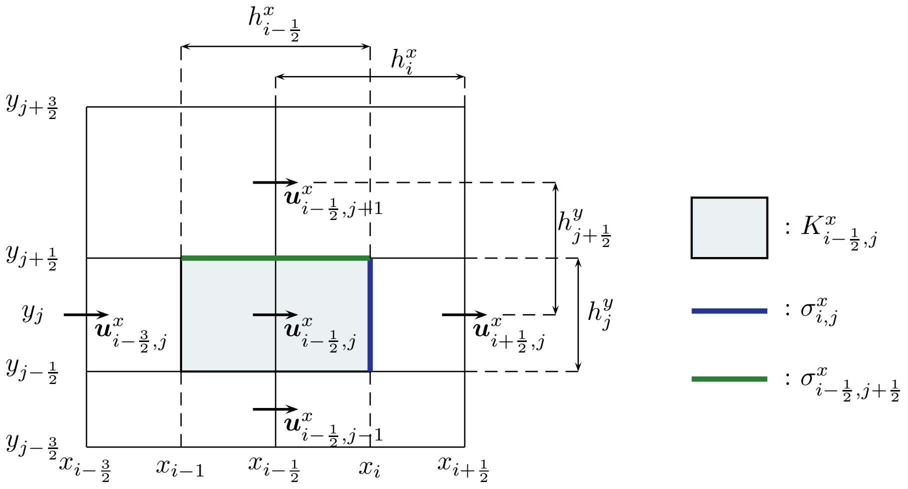
$h^x_{i-\frac{1}{2}}$
h^x_i
$y_{j+\frac{3}{2}}$
$\boldsymbol{u}^x_{i-\frac{1}{2},j+1}$
$h^y_{j+\frac{1}{2}}$
$y_{j+\frac{1}{2}}$
y_j
$\boldsymbol{u}^x_{i-\frac{3}{2},j}$
$\boldsymbol{u}^x_{i-\frac{1}{2},j}$
$\boldsymbol{u}^x_{i+\frac{1}{2},j}$
h^y_j
$y_{j-\frac{1}{2}}$
$\boldsymbol{u}^x_{i-\frac{1}{2},j-1}$
$y_{j-\frac{3}{2}}$
$x_{i-\frac{3}{2}}$
x_{i-1}
$x_{i-\frac{1}{2}}$
x_i
$x_{i+\frac{1}{2}}$
$: K^x_{i-\frac{1}{2},j}$
$: \sigma^x_{i,j}$
$: \sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$

FIG. C.1 – Dual cell for the x-component of the velocity

C.2 Discretization of the dissipation term

C.2.1 The two-dimensional case

Let us begin with a two-dimensional case, for the sake of simplicity, let us suppose that :

$$\boldsymbol{\tau}(\boldsymbol{u}) = \mu(\boldsymbol{x})(\boldsymbol{\nabla}\boldsymbol{u} + \boldsymbol{\nabla}^t\boldsymbol{u}),$$

the extension of the present material to the other term in (C.2) being straightforward.

The first step is to propose a discretization for the diffusion term in the momentum equation. We begin with the x-component of the velocity, for which we write a balance equation on $K^x_{i-\frac{1}{2},j} = (x_{i-1},\, x_i) \times (y_{j-\frac{1}{2}},\, y_{j+\frac{1}{2}})$ (see Figure C.1 for the notations). Integrating the projection of the momentum balance equation onto O_x over $K^x_{i-\frac{1}{2},j}$, we get for the diffusion term :

$$\bar{T}^{\mathrm{dif}}_{i-\frac{1}{2},j} = -\Big[\int_{K^x_{i-\frac{1}{2},j}} \operatorname{div}\big[\boldsymbol{\tau}(\boldsymbol{u})\big]\,\mathrm{d}\boldsymbol{x}\Big] \cdot \boldsymbol{e}^{(x)} = -\Big[\int_{\partial K^x_{i-\frac{1}{2},j}} \boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{n}\,\mathrm{d}\gamma\Big] \cdot \boldsymbol{e}^{(x)}, \tag{C.4}$$

where $\boldsymbol{e}^{(x)}$ stands for the first vector of the canonical basis of $\mathbb{R}^2$. We denote by $\sigma^x_{i,j}$ the left face of $K^x_{i-\frac{1}{2},j}$, *i.e.* $\sigma^x_{i,j} = \{x_i\} \times (y_{j-\frac{1}{2}},\, y_{j+\frac{1}{2}})$. Splitting the boundary integral in (C.4), the part of $\bar{T}^{\mathrm{dif}}_{i-\frac{1}{2},j}$ associated to $\sigma^x_{i,j}$, also referred to as the viscous flux through $\sigma^x_{i,j}$, reads :

$$-\Big[\int_{\sigma^x_{i,j}} \boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{n}\,\mathrm{d}\gamma\Big] \cdot \boldsymbol{e}^{(x)} = -2\int_{\sigma^x_{i,j}} \mu\ \partial_x \boldsymbol{u}^x\,\mathrm{d}\gamma,$$

and the usual finite difference technique yields the following approximation for this term :

$$-2\int_{\sigma^x_{i,j}} \mu\ \partial_x \boldsymbol{u}^x\,\mathrm{d}\gamma \approx 2\mu_{i,j}\ \frac{h^y_j}{h^x_i}\ (\boldsymbol{u}^x_{i-\frac{1}{2},j} - \boldsymbol{u}^x_{i+\frac{1}{2},j}),$$

where $\mu_{i,j}$ is an approximation of the viscosity at the face $\sigma^x_{i,j}$. Similarly, let $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}} = (x_i,\, x_{i+1})\times\{y_{j+\frac{1}{2}}\}$ be the top edge of the cell. Then :

$$
\begin{aligned}
-\Big[\int_{\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}} \boldsymbol{\tau}(\boldsymbol{u})\ \boldsymbol{n}\,\mathrm{d}\gamma\Big]\cdot \boldsymbol{e}^{(x)} &= -\int_{\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}} \mu\ (\partial_y \boldsymbol{u}^x + \partial_x \boldsymbol{u}^y)\,\mathrm{d}\gamma \\
&\approx \mu^{xy}_{i-\frac{1}{2},j+\frac{1}{2}}\ \Big[\frac{h^x_{i-\frac{1}{2}}}{h^y_{j+\frac{1}{2}}}\ (\boldsymbol{u}^x_{i-\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j+1}) + \frac{h^x_{i-\frac{1}{2}}}{h^x_{i-\frac{1}{2}}}\ (\boldsymbol{u}^y_{i-1,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j+\frac{1}{2}})\Big],
\end{aligned}
$$

where $\mu^{xy}_{i-\frac{1}{2},j+\frac{1}{2}}$ stands for an approximation of the viscosity at the edge $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$.

Let us now multiply each discrete equation for $\boldsymbol{u}^x$ by the corresponding degree of freedom of a velocity field $\boldsymbol{v}$ (*i.e.* the balance over $K^x_{i-\frac{1}{2},j}$ by $\boldsymbol{v}^x_{i-\frac{1}{2},j}$) and sum over i and j. The viscous flux at the face $\sigma^x_{i,j}$ appears twice in the sum, once multiplied by $\boldsymbol{u}^x_{i-\frac{1}{2},j}$ and the second one by $-\boldsymbol{u}^x_{i+\frac{1}{2},j}$, and the corresponding term reads :

$$
\begin{aligned}
T^{\mathrm{dis}}_{i,j}(\boldsymbol{u},\boldsymbol{v}) &= 2\mu_{i,j}\ \frac{h^y_j}{h^x_i}\ (\boldsymbol{u}^x_{i-\frac{1}{2},j} - \boldsymbol{u}^x_{i+\frac{1}{2},j})\ (\boldsymbol{v}^x_{i-\frac{1}{2},j} - \boldsymbol{v}^x_{i+\frac{1}{2},j}) \\
&= 2\mu_{i,j}\ h^y_j h^x_i\ \frac{\boldsymbol{u}^x_{i-\frac{1}{2},j} - \boldsymbol{u}^x_{i+\frac{1}{2},j}}{h^x_i}\ \frac{\boldsymbol{v}^x_{i-\frac{1}{2},j} - \boldsymbol{v}^x_{i+\frac{1}{2},j}}{h^x_i}. \qquad \text{(C.5)}
\end{aligned}
$$

Similarly, the term associated to $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ appears multiplied by $\boldsymbol{v}^x_{i-\frac{1}{2},j}$ and $-\boldsymbol{v}^x_{i-\frac{1}{2},j+1}$, and we get :

$$
\begin{aligned}
T^{\mathrm{dis}}_{i-\frac{1}{2},j+\frac{1}{2}}(\boldsymbol{u},\boldsymbol{v}) &= \mu^{xy}_{i-\frac{1}{2},j+\frac{1}{2}}\ h^x_{i-\frac{1}{2}} h^y_{j+\frac{1}{2}} \\
&\Big[\frac{\boldsymbol{u}^x_{i-\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j+1}}{h^y_{j+\frac{1}{2}}} + \frac{\boldsymbol{u}^y_{i-1,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j+\frac{1}{2}}}{h^x_{i-\frac{1}{2}}}\Big]\ \frac{\boldsymbol{v}^x_{i-\frac{1}{2},j} - \boldsymbol{v}^x_{i-\frac{1}{2},j+1}}{h^y_{j+\frac{1}{2}}}. \qquad \text{(C.6)}
\end{aligned}
$$

Let us now define the discrete gradient of the velocity as follows :

- The derivatives involved in the divergence, $\partial^{\mathcal{M}}_x \boldsymbol{u}^x$ and $\partial^{\mathcal{M}}_y \boldsymbol{u}^y$, are defined over the primal cells by :

$$
\partial^{\mathcal{M}}_x \boldsymbol{u}^x(\boldsymbol{x}) = \frac{\boldsymbol{u}^x_{i+\frac{1}{2},j} - \boldsymbol{u}^x_{i-\frac{1}{2},j}}{h^x_i},\quad \partial^{\mathcal{M}}_y \boldsymbol{u}^y(\boldsymbol{x}) = \frac{\boldsymbol{u}^y_{i,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j-\frac{1}{2}}}{h^y_j},\quad \forall \boldsymbol{x}\in K_{i,j}. \qquad \text{(C.7)}
$$

- For the other derivatives, we introduce another mesh which is vertex-centered, and we denote by K^{xy} the generic cell of this new mesh, with $K^{xy}_{i+\frac{1}{2},j+\frac{1}{2}} = (x_i, x_{i+1})\times(y_j, y_{j+1})$. Then :

$$
\partial^{\mathcal{M}}_y \boldsymbol{u}^x(\boldsymbol{x}) = \frac{\boldsymbol{u}^x_{i+\frac{1}{2},j+1} - \boldsymbol{u}^x_{i+\frac{1}{2},j}}{h^y_{j+\frac{1}{2}}},\quad \partial^{\mathcal{M}}_x \boldsymbol{u}^y(\boldsymbol{x}) = \frac{\boldsymbol{u}^y_{i+1,j+\frac{1}{2}} - \boldsymbol{u}^y_{i,j+\frac{1}{2}}}{h^x_{i+\frac{1}{2}}},\quad \forall \boldsymbol{x}\in K^{xy}_{i+\frac{1}{2},j+\frac{1}{2}}. \qquad \text{(C.8)}
$$

With this definition, we get :

$$
T^{\mathrm{dis}}_{i,j}(\boldsymbol{u},\boldsymbol{v}) = 2\mu_{i,j}\int_{K_{i,j}} \partial^{\mathcal{M}}_x \boldsymbol{u}^x\ \partial^{\mathcal{M}}_x \boldsymbol{v}^x\,\mathrm{d}\boldsymbol{x},
$$

and :

$$
T^{\mathrm{dis}}_{i-\frac{1}{2},j+\frac{1}{2}}(\boldsymbol{u},\boldsymbol{v}) = \mu^{xy}_{i-\frac{1}{2},j+\frac{1}{2}}\int_{K^{xy}_{i-\frac{1}{2},j+\frac{1}{2}}} (\partial^{\mathcal{M}}_y \boldsymbol{u}^x + \partial^{\mathcal{M}}_x \boldsymbol{u}^y)\ \partial^{\mathcal{M}}_y \boldsymbol{v}^x\,\mathrm{d}\boldsymbol{x}.
$$

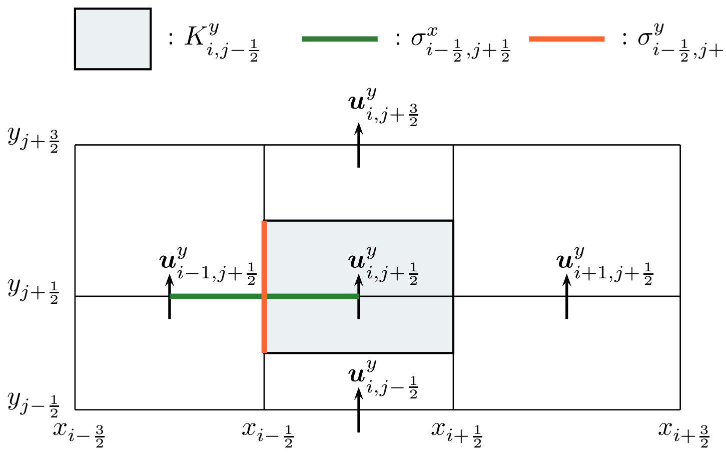
: $K^y_{i,j-\frac{1}{2}}$
: $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$
: $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i,j+\frac{3}{2}}$
$y_{j+\frac{3}{2}}$
$\boldsymbol{u}^y_{i-1,j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i,j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i+1,j+\frac{1}{2}}$
$y_{j+\frac{1}{2}}$
$\boldsymbol{u}^y_{i,j-\frac{1}{2}}$
$y_{j-\frac{1}{2}}$
$x_{i-\frac{3}{2}}$
$x_{i-\frac{1}{2}}$
$x_{i+\frac{1}{2}}$
$x_{i+\frac{3}{2}}$

FIG. C.2 – Dual cell for the y-component of the velocity

Let us now perform the same operations for the y-component of the velocity. Doing so, we are lead to introduce an approximation of the viscosity at the edge $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}} = \{x_{i+\frac{1}{2}}\} \times (y_j,\, y_{j+1})$ (see Figure C.2). Let us suppose that we take the same approximation as on $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$. Then, the same arguments yield that multiplying each discrete equation for $\boldsymbol{u}^x$ and for $\boldsymbol{u}^y$ by the corresponding degree of freedom of a velocity field $\boldsymbol{v}$, we obtain a dissipation term which reads :

$$T^{\mathrm{dis}}(\boldsymbol{u},\boldsymbol{v}) = \int_\Omega \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \boldsymbol{\nabla}^{\mathcal{M}}\boldsymbol{v}\,\mathrm{d}\boldsymbol{x}, \tag{C.9}$$

with the above defined gradient and :

$$\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) = \begin{bmatrix} 2\mu\;\partial^{\mathcal{M}}_x \boldsymbol{u}_x & \mu^{xy}\;(\partial^{\mathcal{M}}_y \boldsymbol{u}_x + \partial^{\mathcal{M}}_x \boldsymbol{u}_y) \\ \mu^{xy}\;(\partial^{\mathcal{M}}_y \boldsymbol{u}_x + \partial^{\mathcal{M}}_x \boldsymbol{u}_y) & 2\mu\;\partial^{\mathcal{M}}_y \boldsymbol{u}_y \end{bmatrix}, \tag{C.10}$$

where μ is the viscosity defined on the primal mesh by $\mu(\boldsymbol{x}) = \mu_{i,j},\ \forall \boldsymbol{x} \in K_{i,j}$ and μ^{xy} is the viscosity defined on the vertex-centered mesh, by $\mu(\boldsymbol{x}) = \mu_{i+\frac{1}{2},j+\frac{1}{2}},\ \forall \boldsymbol{x} \in K^{xy}_{i+\frac{1}{2},j+\frac{1}{2}}$.

Then, finally, to discretize the viscous dissipation term in the internal energy balance, we just set on each primal cell $K_{i,j}$:

$$(\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u})_{i,j} = \frac{1}{|K_{i,j}|} \int_{K_{i,j}} \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \boldsymbol{\nabla}^{\mathcal{M}}\boldsymbol{u}\,\mathrm{d}\boldsymbol{x}, \tag{C.11}$$

which, thanks to (C.9), yields the consistency property (ii) we are searching for, namely :

$$T^{\mathrm{dis}}(\boldsymbol{u},\boldsymbol{u}) = \sum_{i,j} |K_{i,j}|\;(\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u})_{i,j}.$$

In addition, we get from Definition (C.10) that $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})(\boldsymbol{x})$ is a symmetrical tensor, for any i,j and $\boldsymbol{x} \in K_{i,j}$, so an elementary algebraic argument yields :

$$\begin{aligned}(\boldsymbol{\tau}(\boldsymbol{u}) : \boldsymbol{\nabla}\boldsymbol{u})_{i,j} &= \frac{1}{|K_{i,j}|} \int_{K_{i,j}} \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \boldsymbol{\nabla}^{\mathcal{M}}\boldsymbol{u}\,\mathrm{d}\boldsymbol{x} \\ &= \frac{1}{2\,|K_{i,j}|} \int_{K_{i,j}} \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u}) : \left[\boldsymbol{\nabla}^{\mathcal{M}}\boldsymbol{u} + (\boldsymbol{\nabla}^{\mathcal{M}}\boldsymbol{u})^t\right] \mathrm{d}\boldsymbol{x} \geq 0.\end{aligned}$$

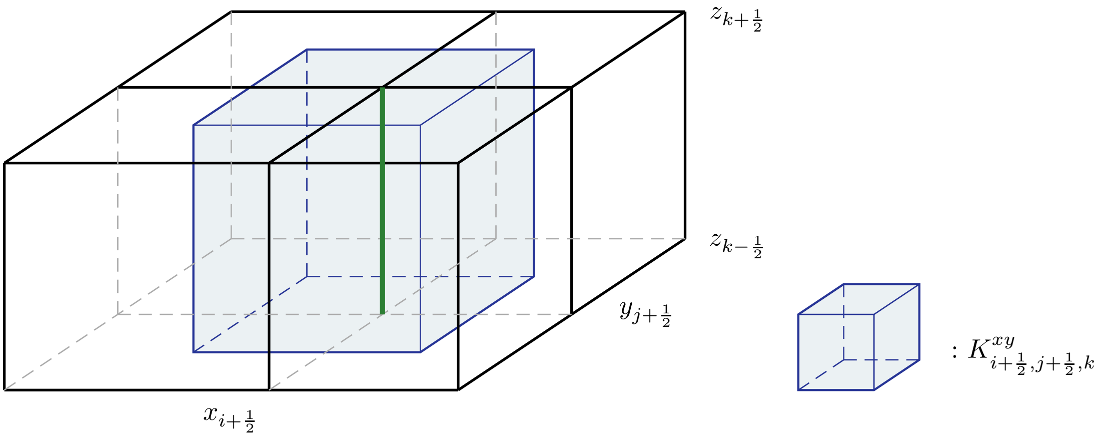

FIG. C.3 – The xy-staggered cell $K^{xy}_{i+\frac{1}{2},j+\frac{1}{2},k}$, used in the definition of $\partial^{\mathcal{M}}_y \boldsymbol{u}^x$, $\partial^{\mathcal{M}}_x \boldsymbol{u}^y$, and $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})_{x,y} = \boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})_{y,x}$.

Remark 8 (Approximation of the viscosity)
Note that, for the symetry of $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})$ to hold, the choice of the same viscosity at the edges $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ and $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}}$ is crucial. . .and that other choices may appear natural. For instance, suppose that the viscosity is a function of the temperature ; then the following construction is reasonable :

1. define, from the discrete temperature, a constant value for μ over the primal meshes,
2. associate a value of μ to the primal edges, by taking the average between the value at the adjacent cells,
3. finally, split the integral of the shear stress over $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ in two parts, one for the part included in the (top) boundary of $K_{i-1,j}$ and the second one in the boundary of $K_{i,j}$.

Then the viscosities on $\sigma^x_{i-\frac{1}{2},j+\frac{1}{2}}$ and $\sigma^y_{i-\frac{1}{2},j+\frac{1}{2}}$ coincide only for uniform meshes, and, in the general case, the symetry of $\boldsymbol{\tau}^{\mathcal{M}}(\boldsymbol{u})$ is lost.

C.2.2 Extension to the three-dimensional case

Extending the computations of the preceding section to dimension three yields the following construction.

- First, define three new meshes, which are "edge centered" : $K^{xy}_{i+\frac{1}{2},j+\frac{1}{2},k} = (x_i, x_{i+1} \times (y_i, y_j + 1) \times (z_{k-\frac{1}{2}}, z_{k+\frac{1}{2}})$ is staggered from the primal mesh $K_{i,j,k}$ in the x and y direction (see Figure C.3), $K^{xz}_{i+\frac{1}{2},j,k+\frac{1}{2}}$ in the x and z direction, and $K^{yz}_{i,j+\frac{1}{2},k+\frac{1}{2}}$ in the y and z direction.
- The partial derivatives of the velocity components are then defined as piecewise constant functions, the value of which is obtained by natural finite differences :
  - for $\partial^{\mathcal{M}}_x \boldsymbol{u}^x$, $\partial^{\mathcal{M}}_y \boldsymbol{u}^y$ and $\partial^{\mathcal{M}}_z \boldsymbol{u}^z$, on the primal mesh,
  - for $\partial^{\mathcal{M}}_y \boldsymbol{u}^x$ and $\partial^{\mathcal{M}}_x \boldsymbol{u}^y$ on the cells $(K^{xy}_{i+\frac{1}{2},j+\frac{1}{2},k})$,
  - for $\partial^{\mathcal{M}}_z \boldsymbol{u}^x$ and $\partial^{\mathcal{M}}_x \boldsymbol{u}^z$ on the cells $(K^{xz}_{i+\frac{1}{2},j,k+\frac{1}{2}})$,
  - for $\partial^{\mathcal{M}}_y \boldsymbol{u}^z$ and $\partial^{\mathcal{M}}_z \boldsymbol{u}^y$ on the cells $(K^{yz}_{i,j+\frac{1}{2},k+\frac{1}{2}})$.

- We then define four families of values for the viscosity field, μ, μ^{xy}, μ^{xz} and μ^{yz}, associated to the primal and the three edge centered meshes respectively.
- The shear stress tensor is obtained by the extension of (C.10) to $d = 3$.
- And, finally, the dissipation term is given by (C.11).

C.3 A strong convergence result

We finally conclude this paper by showing how the consistency property (*ii*) may be used, in some particular case, to obtain the strong convergence of the dissipation term. To this purpose, let us just address first the model problem :

$$-\Delta \underline{u} = \underline{f} \text{ in } \Omega = (0,1) \times (0,1), \qquad \underline{u} = 0 \text{ on } \partial\Omega, \tag{C.12}$$

with $\underline{u}$ and $\underline{f}$ two scalar functions, $\underline{f} \in \mathrm{L}^2(\Omega)$. Let us suppose that this problem is discretized by the usual finite volume technique, with the uniform MAC mesh associated to the x-component of the velocity. We define a discrete function as a piecewise constant function, vanishing on the left and right sides of the domain (so on the left and right stripes of (half-)staggered meshes adjacent to these boundaries), and we define the discrete H^1-norm of a discrete function v by :

$$\|v\|_1^2 = \int_\Omega (\partial_x^{\mathcal{M}} v)^2 + (\partial_y^{\mathcal{M}} v)^2 \,\mathrm{d}\boldsymbol{x}.$$

Let $(\mathcal{M}^{(n)})_{n\in\mathbb{N}}$ be a sequence of such meshes, with a step h^n tending to zero, and $(u^{(n)})_{n\in\mathbb{N}}$ the corresponding sequence of discrete solutions. Then, with the variational technique employed in the preceding section (*i.e.* multiplying each discrete equation by the corresponding equation and summing), we get, with the usual discretization of the right-hand side :

$$\|u^{(n)}\|_1^2 = \int_\Omega (\partial_x^{\mathcal{M}} u^{(n)})^2 + (\partial_y^{\mathcal{M}} u^{(n)})^2 \,\mathrm{d}\boldsymbol{x} = \int_\Omega \underline{f} u^{(n)} \,\mathrm{d}\boldsymbol{x}. \tag{C.13}$$

Since the discrete H^1-norm controls the L^2-norm (*i.e.* a discrete Poincaré inequality holds, [16]), this yields a uniform bound for the sequence $(u^{(n)})_{n\in\mathbb{N}}$ in discrete H^1-norm. We know [16] that this implies that the sequence $(u^{(n)})_{n\in\mathbb{N}}$ converges in $\mathrm{L}^2(\Omega)$ to a function $\bar{u} \in \mathrm{H}^1_0(\Omega)$, and that the discrete derivatives $(\partial_x^{\mathcal{M}} u^{(n)})_{n\in\mathbb{N}}$ and $(\partial_y^{\mathcal{M}} u^{(n)})_{n\in\mathbb{N}}$ weakly converge in $\mathrm{L}^2(\Omega)$ to $\partial_x \bar{u}$ and $\partial_y \bar{u}$ respectively. This allows to pass to the limit in the scheme, and we obtain that $\bar{u}$ satisfies the continuous equation (C.12), so, taking $\bar{u}$ as test function in the variational form of (C.12) :

$$\int_\Omega (\partial_x \bar{u})^2 + (\partial_y \bar{u})^2 \,\mathrm{d}\boldsymbol{x} = \int_\Omega \underline{f} \bar{u} \,\mathrm{d}\boldsymbol{x}.$$

But, passing to the limit in (C.13), we get :

$$\lim_{n\mapsto\infty} \int_\Omega (\partial_x^{\mathcal{M}} u^{(n)})^2 + (\partial_y^{\mathcal{M}} u^{(n)})^2 \,\mathrm{d}\boldsymbol{x} = \lim_{n\mapsto\infty} \int_\Omega \underline{f} u^{(n)} \,\mathrm{d}\boldsymbol{x} = \int_\Omega \underline{f} \bar{u} \,\mathrm{d}\boldsymbol{x},$$

which, comparing to the preceding relation, yields :

$$\lim_{n\to\infty} \int_\Omega (\partial_x^{\mathcal{M}} u^{(n)})^2 + (\partial_y^{\mathcal{M}} u^{(n)})^2 \,\mathrm{d}\boldsymbol{x} = \int_\Omega (\partial_x \bar{u})^2 + (\partial_y \bar{u})^2 \,\mathrm{d}\boldsymbol{x}.$$

So the discrete gradient weakly converges and its norm converges to the norm of the limit : the discrete gradient strongly converges in $\mathrm{L}^2(\Omega)^2$ to the gradient of the solution. Let us now imagine that Equation (C.12) is coupled to a balance equation for another variable, the right-hand side of which is $|\boldsymbol{\nabla}\underline{u}|^2$; this situation occurs in several physical situations, as the modelling of Joule effect [5], or RANS turbulence models [49, 24]. Then using the expression (C.11) for the discretization of the dissipation term in the cell K, which reads here :

$$\big(|\boldsymbol{\nabla} u^{(n)}|^2\big)_K = \frac{1}{|K|}\int_K (\partial_x^{\mathcal{M}} u^{(n)})^2 + (\partial_y^{\mathcal{M}} u^{(n)})^2 \,\mathrm{d}\boldsymbol{x},$$

yields a convergent right-hand side, in the sense that, for any regular function $\varphi \in \mathrm{C}_c^\infty(\Omega)$, we have :

$$\lim_{n\to\infty} \sum_K \int_K \big(|\boldsymbol{\nabla}\boldsymbol{u}^{(n)}|^2\big)_K \varphi \,\mathrm{d}\boldsymbol{x} = \int_\Omega |\boldsymbol{\nabla}\underline{u}|^2 \varphi \,\mathrm{d}\boldsymbol{x}.$$

(A declination of) this argument has been used to prove the convergence of numerical schemes in [5, 49, 24].

Bibliographie

[1] G. Ansanay-Alex, F. Babik, J.-C. Latché, and D. Vola. An L2-stable approximation of the Navier-Stokes convection operator for low-order non-conforming finite elements. *International Journal for Numerical Methods in Fluids*, online (2010).

[2] R. Babik, R. Herbin, W. Kheriji, and J.-C. Latché. Discretization of the viscous dissipation term with the MAC scheme. In *Finite Volumes for Complex Applications VI - Problems and Perspectives - Prague, Czech Republic*, 2011.

[3] H. Bijl and P. Wesseling. A unified method for computing incompressible and compressible flows in boundary-fitted coordinates. *Journal of Computational Physics*, 141 :153–173, 1998.

[4] F. Bouchut. *Nonlinear Stability of finite volume methods for hyperbolic conservation laws.* Birkhauser, 2004.

[5] A. Bradji and R. Herbin. Discretization of the coupled heat and electrical diffusion problems by the finite element and the finite volume methods. *IMA Journal of Numerical Analysis*, 28 :469–495, 2008.

[6] C.H. Bruneau and P. Fabrie. Effective downstream boundary conditions for incompressible Navier-Stokes equations. *International Journal for Numerical Methods in Fluids*, 99 :693–705, 1994.

[7] V. Casulli and D. Greenspan. Pressure method for the numerical solution of transient, compressible fluid flows. *International Journal for Numerical Methods in Fluids*, 4 :1001–1012, 1984.

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

[9] P. G. Ciarlet. Handbook of numerical analysis volume II : Finite elements methods – 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.

[10] P. Colella and K. Pao. A projection method for low speed flows. *Journal of Computational Physics*, 149 :245–269, 1999.

[11] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *Revue Française d'Automatique, Informatique et Recherche Opérationnelle (R.A.I.R.O.)*, R-3 :33–75, 1973.

[12] F. Dardalhon, J.-C. Latché, and S. Minjeaud. Analysis of a projection method for low-order nonconforming finite elements. *to appear in IMA Journal of Numerical Analysis*, 2011.

[13] K. Deimling. *Nonlinear Functional Analysis.* Springer, New-York, Berlin, 1985.

[14] S. Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *Journal of Computational Physics*, 229 :978–1016, 2010.

[15] I. Demirdžić, Ž. Lilek, and M. Perić. A collocated finite volume method for predicting flows at all speed. *International Journal for Numerical Methods in Fluids*, 16 :1029–1050, 1993.

[16] R. Eymard, T Gallouët, and R. Herbin. Finite volume methods. In P. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis, Volume VII*, pages 713–1020. North Holland, 2000.

[17] R. Eymard, T. Gallouët, R. Herbin, and J.-C. Latché. Convergence of the MAC scheme for the compressible Stokes equations. *SIAM Journal on Numerical Analysis*, 48 :2218–2246, 2010.

[18] R. Eymard, T. Gallouët, R. Herbin, and J.-C. Latché. A convergent finite element-finite volume scheme for the compressible Stokes problem. Part II : the isentropic case. *Mathematics of Computation*, 79 :649–675, 2010.

[19] E. Feireisl. Dynamics of viscous compressible flows. volume 26 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, 2004.

[20] 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.

[21] T. Gallouët, R. Herbin, and J.-C. Latché. A convergent finite element-finite volume scheme for the compressible Stokes problem. Part I : the isothermal case. *Mathematics of Computation*, 267 :1333–1352, 2009.

[22] T. Gallouët, R. Herbin, and J.-C. Latché. Kinetic energy control in explicit finite-volume discretizations of the incompressible and compressible Navier-Stokes equations. *International Journal of Finite Volumes*, 2, 2010.

[23] T. Gallouët, R. Herbin, J.-C. Latché, and T.T. Nguyen. Playing with burgers equation. In *Finite Volumes for Complex Applications VI - Problems and Perspectives - Prague, Czech Republic*, 2011.

[24] T. Gallouët and J.-C. Latché. Compactness of discrete approximate solutions to parabolic PDEs - Application to a turbulence model. *to appear in Communications on Pure and Applied Analysis*, 2011.

[25] 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.

[26] 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.

[27] L. Gastaldo, R. Herbin, and J.-C. Latché. A discretization of the phase mass balance in fractional step algorithms for the drift-flux model. *IMA J.Numer. Anal.*, 2011.

[28] J.-L. Guermond and L. Quartapelle. A projection FEM for variable density incompressible flows. *Journal of Computational Physics*, 165 :167–188, 2000.

[29] 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.

[30] J.L Guermond and R. Pasquetti. Entropy-based nonlinear viscosity for Fourier approximations of conservation laws. *Comptes Rendus de l'Académie des Sciences de Paris – Série I – Analyse Numérique*, 346 :801–806, 2008.

[31] J.L Guermond, R. Pasquetti, and B. Popov. Entropy viscosity method for nonlinear conservation laws. *Journal of Computational Physics*, 230 :4248–4267, 2011.

[32] H. Guillard and F. Duval. A Darcy law for the drift velocity in a two-phase flow model. *Journal of Computational Physics*, 224 :288–313, 2007.

[33] H. Guillard and A. Murrone. On the behavior of upwind schemes in the low Mach number limit : II. Godunov type schemes. *Computer & Fluids*, 33(4) :655–675, may 2004.

[34] H. Guillard and C. Viozat. On the behavior of upwind schemes in the low Mach number limit. *Computer & Fluids*, 28 :63–86, 1999.

[35] F.H. Harlow and A.A. Amsden. Numerical calculation of almost incompressible flow. *Journal of Computational Physics*, 3 :80–93, 1968.

[36] F.H. Harlow and A.A. Amsden. A numerical fluid dynamics calculation method for all flow speeds. *Journal of Computational Physics*, 8 :197–213, 1971.

[37] F.H. Harlow and J.E. Welsh. Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. *Physics of Fluids*, 8 :2182–2189, 1965.

[38] R. Herbin and J.-C. Latché. Kinetic energy control in the MAC discretization of the compressible Navier-Stokes equations. *International Journal of Finites Volumes*, 7, 2010.

[39] R. Herbin, J.-C. Latché, and B. Piar. A finite-element finite-volume face centred scheme with non-conforming local refinement. I – Convection-diffusion equation. *in preparation*, 2011.

[40] ISIS. A CFD computer code for the simulation of reactive turbulent flows.
`https ://gforge.irsn.fr/gf/project/isis`.

[41] R.I. Issa. Solution of the implicitly discretised fluid flow equations by operator splitting. *Journal of Computational Physics*, 62 :40–65, 1985.

[42] R.I. Issa, A.D. Gosman, and A.P. Watkins. The computation of compressible and incompressible recirculating flows by a non-iterative implicit scheme. *Journal of Computational Physics*, 62 :66–82, 1986.

[43] R.I. Issa and M.H. Javareshkian. Pressure-based compressible calculation method utilizing total variation diminishing schemes. *AIAA Journal*, 36 :1652–1657, 1998.

[44] K.C. Karki and S.V. Patankar. Pressure based calculation procedure for viscous flows at all speeds in arbitrary configurations. *AIAA Journal*, 27 :1167–1174, 1989.

[45] W. Kheriji, R. Herbin, and J.-C. Latché. Numerical tests of a new pressure correction scheme for the homogeneous model. In *ECCOMAS CFD 2010, Lisbon, Portugal*, 2010.

[46] M.H. Kobayashi and J.C.F. Pereira. Characteristic-based pressure correction at all speeds. *AIAA Journal*, 34 :272–280, 1996.

[47] D. Kuzmin and S. Turek. Numerical simulation of turbulent bubbly flows. In *3rd International Symposium on Two-Phase Flow Modelling and Experimentation, Pisa, 22-24 September*, 2004.

[48] C. Lapuerta and J.-C. Latché. Discrete artificial boundary conditions for compressible external flows. *in preparation*, 2011.

[49] A. Larcher and J.-C. Latché. Convergence analysis of a finite element-finite volume scheme for a RANS turbulence model. *submitted*, 2011.

[50] B. Larrouturou. How to preserve the mass fractions positivity when computing compressible multi-component flows. *Journal of Computational Physics*, 95 :59–84, 1991.

[51] P.-L. Lions. Mathematical topics in fluid mechanics – volume 2 – compressible models. volume 10 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, 1998.

[52] M.-S. Liou. A sequel to AUSM, part II :AUSM+-up. *Journal of Computational Physics*, 214 :137–170, 2006.

[53] M.-S. Liou and C.J. Steffen. A new flux splitting scheme. *Journal of Computational Physics*, 107 :23–39, 1993.

[54] 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.

[55] M. Marion and R. Temam. Navier-Stokes equations : Theory and approximation. In P. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis, Volume VI.* North Holland, 1998.

[56] F. Moukalled and M. Darwish. A high-resolution pressure-based algorithm for fluid flow at all speeds. *Journal of Computational Physics*, 168 :101–133, 2001.

[57] F. Moukalled, M. Darwish, and B. Sekar. A pressure-based algorithm for multi-phase flow at all speeds. *Journal of Computational Physics*, 190 :550–571, 2003.

[58] K. Nerinckx, J. Vierendeels, and E. Dick. A Mach-uniform algorithm : coupled versus segregated approach. *Journal of Computational Physics*, 224 :314–331, 2007.

[59] T.T. Nguyen, R. Herbin, and J.-C. Latché. An explicit staggered scheme for euler equations. *in preparation*, 2011.

[60] P. Nithiarasu, R. Codina, and O.C. Zienkiewicz. The Characteristic-Based Split (CBS) scheme – A unified approach to fluid dynamics. *International Journal for Numerical Methods in Engineering*, 66 :1514–1546, 2006.

[61] A. Novotný and I. Straškraba. Introduction to the mathematical theory of compressible flow. volume 27 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, 2004.

[62] G. Patnaik, R.H. Guirguis, J.P. Boris, and E.S. Oran. A barely implicit correction for flux-corrected transport. *Journal of Computational Physics*, 71 :1–20, 1987.

[63] PELICANS. Collaborative development environment.
`https ://gforge.irsn.fr/gf/project/pelicans`.

[64] E.S. Politis and K.C. Giannakoglou. A pressure-based algorithm for high-speed turbomachinery flows. *International Journal for Numerical Methods in Fluids*, 25 :63–80, 1997.

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

[66] B. Spalding. Numerical computation of multiphase flow and heat transfer. In *Recent Advances in Numerical Methods in Fluids – Volume 1*, pages 139–168, 1980.

[67] 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.

[68] E. Toro. *Riemann solvers and numerical methods for fluid dynamics – A practical introduction (third edition).* Springer, 2009.

[69] D.R. Van der Heul, C. Vuik, and P. Wesseling. Stability analysis of segregated solution methods for compressible flow. *Applied Numerical Mathematics*, 38 :257–274, 2001.

[70] D.R. Van der Heul, C. Vuik, and P. Wesseling. A conservative pressure-correction method for flow at all speeds. *Computer & Fluids*, 32 :1113–1132, 2003.

[71] J.P. Van Dormaal, G.D. Raithby, and B.H. McDonald. The segregated approach to predicting viscous compressible fluid flows. *Transactions of the ASME*, 109 :268–277, 1987.

[72] D. Vidović, A. Segal, and P. Wesseling. A superlinearly convergent Mach-uniform finite volume method for the Euler equations on staggered unstructured grids. *Journal of Computational Physics*, 217 :277–294, 2006.

[73] C. Wall, C.D. Pierce, and P. Moin. A semi-implicit method for resolution of acoustic waves in low Mach number flows. *Journal of Computational Physics*, 181 :545–563, 2002.

[74] I. Wenneker, A. Segal, and P. Wesseling. A Mach-uniform unstructured staggered grid method. *International Journal for Numerical Methods in Fluids*, 40 :1209–1235, 2002.

[75] P. Wesseling. Principles of computational fluid dynamics. volume 29 of *Springer Series in Computational Mathematics*. Springer, 2001.

[76] P. Woodward and P. Colella. The numerical simulation of two-dimensional fluid flow with strong shocks. *Journal of Computational Physics*, 54 :115–173, 1984.

[77] O.C. Zienkiewicz and R. Codina. A general algorithm for compressible and incompressible flow – Part I. The split characteristic-based scheme. *International Journal for Numerical Methods in Fluids*, 20 :869–885, 1995.

Méthodes de correction de pression pour les écoulements compressibles

Résumé : Cette thèse porte sur le développement de schémas semi-implicites à pas fractionnaires pour les équations de Navier-Stokes compressibles ; ces schémas entrent dans la classe des méthodes de correction de pression. La discrétisation spatiale choisie est de type "à mailles décalées : éléments finis mixtes non conformes (éléments finis de Crouzeix-Raviart ou Rannacher-Turek) ou schéma MAC classique. Une discrétisation en volumes finis décentrée amont du bilan de masse garantit la positivité de la masse volumique. La positivité de l'énergie interne est obtenue en discrétisant le bilan d'énergie interne continu , par une méthode de volumes finis décentrée amont, enfin, et en couplant ce bilan d'énergie interne discret à l'étape de correction de pression. On effectue une discrétisation particulière en volumes finis sur un maillage dual du terme de convection de vitesse dans le bilan de quantité de mouvement et une étape de renormalisation de la pression ; ceci permet de garantir le contrôle au cours du temps de l'intégrale de l'énergie totale sur le domaine. L'ensemble de ces estimations a priori implique en outre, par un argument de degré topologique, l'existence d'une solution discrète.

L'application de ce schéma aux équations d'Euler pose une difficulté supplémentaire. En effet, l'obtention de vitesses de choc correctes nécessite que le schéma soit consistant avec l'équation de bilan d'énergie totale, propriété que nous obtenons comme suit. Tout d'abord, nous établissons un bilan discret (local) d'énergie cinétique. Ce dernier comporte des termes sources, que nous compensons ensuite dans le bilan d'énergie interne. Les équations d'énergie cinétique et interne sont associées au maillage dual et primal respectivement, et ne peuvent donc être additionnées pour obtenir un bilan d'énergie totale ; cette dernière équation est toutefois retrouvée, sous sa forme continue, à convergence : si nous supposons qu'une suite de solutions discrètes converge lorsque le pas de temps et d'espace tendent vers 0,, nous montrons en effet, en 1D au moins, que la limite en satisfait une forme faible. Ces résultats théoriques sont confortés par des tests numériques.

Des résultats similaires sont obtenus pour les équations de Navier-Stokes barotropes.

Mots clefs : Méthodes de correction de pression, Navier-Stokes compressibles, schéma MAC, éléments finis mixtes non conformes, stabilité, convergence, tests numériques.

Pressure correction schemes for compressible flows

Abstract : This thesis is concerned with the development of semi-implicit fractional step schemes, for the compressible Navier-Stokes equations ; these schemes are part of the class of the pressure correction methods. The chosen spatial discretisation is staggered : non conforming mixed finite elements (Crouzeix-Raviart or Rannacher-Turek) or the classic MAC scheme. An upwind finite volume discretisation of the mass balanced guarantees the positivity of the density. The positivity of the internal energy is obtained by discretising the internal energy balance by an upwind finite volume scheme and by coupling the discrete internal energy balance with the pressure correction step. A special finite volume discretisation on dual cells is performed for the convection term in the momentum balance equation, along with a renormalisation of the pressure ; this allows to guarantee the control in time of integral of the total energy over the domain. All these a priori estimates implies lead to the existence of a discrete solution by by a topological degree argument.

The application of this scheme the equations of Euler yields an additional difficulty. Indeed, obtaining correct shock speeds requires that the scheme be consistent with the total energy balance,, property which we obtain as follows. First of all, a local discrete kinetic energy balance is established ; it contains source terms which are compensated by adding some source terms in the internal energy balance. . The kinetic and internal energy equations are associated with the dual and primal meshes respectively, and thus cannot be added to obtain a balance total energy balance ; its continuous counterpart is however recovered at the limit : if we suppose that a sequence of discrete solutions converges when the space and time steps tend to 0, we indeed show, in 1D at least, that the limit satisfies a weak form of the equation. These theoretical results are comforted by numerical tests.

Similar results are obtained for the barotropic Navier–Stokes equations.

Key words :Pressure correction scheme, compressible Navier-Stokes, MAC scheme, mixed non-conforming finite elements, stability, convergence, numerical tests.

Discipline - Spécialité doctoral : Mathématique.

Adresse de laboratoires : IRSN/DPAM/SEMIC/LIMSI/, BP 3, 13115 St-Paul-Lez-Durance,
LATP, 39 rue F. Joliot Curie, 13453 Marseille cedex 13.