

## AN UNCONDITIONALLY STABLE PRESSURE CORRECTION SCHEME FOR THE COMPRESSIBLE BAROTROPIC NAVIER-STOKES EQUATIONS

THIERRY GALLOUËT<sup>1</sup>, LAURA GASTALDO<sup>2</sup>, RAPHAËLE HERBIN<sup>3</sup>  
AND JEAN-CLAUDE LATCHÉ<sup>4</sup>

**Abstract.** We present in this paper a pressure correction scheme for the barotropic compressible Navier-Stokes equations, which enjoys an unconditional stability property, in the sense that the energy and maximum-principle-based *a priori* estimates of the continuous problem also hold for the discrete solution. The stability proof is based on two independent results for general finite volume discretizations, both interesting for their own sake: the  $L^2$ -stability of the discrete advection operator provided it is consistent, in some sense, with the mass balance and the estimate of the pressure work by means of the time derivative of the elastic potential. The proposed scheme is built in order to match these theoretical results, and combines a fractional-step time discretization of pressure-correction type with a space discretization associating low order non-conforming mixed finite elements and finite volumes. Numerical tests with an exact smooth solution show the convergence of the scheme.

**Mathematics Subject Classification.** 35Q30, 65N12, 65N30, 76M25.

Received February 23, 2007.

### 1. INTRODUCTION

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

$$\left\{ \begin{array}{l} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0 \\ \frac{\partial}{\partial t}(\rho u) + \nabla \cdot (\rho u \otimes u) + \nabla p - \nabla \cdot \tau(u) = f_v \\ \rho = \varrho(p) \end{array} \right. \quad (1.1)$$

where  $t$  stands for the time,  $\rho$ ,  $u$  and  $p$  are the density, velocity and pressure in the flow,  $f_v$  is a forcing term and  $\tau(u)$  stands for the shear stress tensor. The function  $\varrho(\cdot)$  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

---

*Keywords and phrases.* Compressible Navier-Stokes equations, pressure correction schemes.

<sup>1</sup> Université de Provence, France. [gallouet@cmi.univ-mrs.fr](mailto:gallouet@cmi.univ-mrs.fr)

<sup>2</sup> Institut de Radioprotection et de Sécurité Nucléaire (IRSN), France. [laura.gastaldo@irsn.fr](mailto:laura.gastaldo@irsn.fr)

<sup>3</sup> Université de Provence, France. [herbin@cmi.univ-mrs.fr](mailto:herbin@cmi.univ-mrs.fr)

<sup>4</sup> Institut de Radioprotection et de Sécurité Nucléaire (IRSN), France. [jean-claude.latche@irsn.fr](mailto:jean-claude.latche@irsn.fr)

concerning the flow; typically, laws as  $\varrho(p) = p^{1/\gamma}$ , where  $\gamma$  is a coefficient that 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  $\Omega$  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. It must be supplemented by boundary conditions and by an initial condition for  $\rho$  and  $u$ .

The development of pressure correction techniques for compressible Navier-Stokes equations may be traced back to the seminal work of Harlow and Amsden [20,21] in the late sixties, who developed 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 [10,23,26,30,33,34] or a staggered arrangement [2,4,7,22,24,25,37,38,40–42] 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 [10,23,25,26,30,34,39]. The other schemes [2,7,22,24,33,37,38,40,42,43] 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 [5,36] for the original papers, [29] for a comprehensive introduction and [19] for a review of most variants). The Characteristic-Based Split (CBS) scheme (see [31] for a recent review or [44] for the seminal paper), developed in the finite-element context, belongs to this latter class of methods.

Our aim in this paper is to propose and study a non-iterative pressure correction scheme for the solution of (1.1). In addition, this method is designed so as to be stable in the low Mach number limit, since our final goal is to apply it to simulate through a drift-flux approach a class of bubbly flows encountered in nuclear safety studies, where pure liquid (incompressible) and pure gaseous (compressible) zones may coexist. To this purpose, we use a low order mixed finite element approximation, which meets the two following requirements: to allow a natural discretization of the viscous terms and to provide a spatial discretization that is intrinsically stable (*i.e.* without the adjunction of stabilization terms to circumvent the so-called *inf-sup* or BB condition) in the incompressible limit.

In this work, a special attention is payed to stability issues. To be more specific, let us recall the *a priori* estimates associated to problem (1.1) with a zero forcing term, *i.e.* estimates which should be satisfied by any possible regular solution [15,28,32]:

$$\left\{ \begin{array}{ll}
 (i) & \rho(x, t) > 0, & \forall x \in \Omega, \forall t \in (0, T) \\
 (ii) & \int_{\Omega} \rho(x, t) \, dx = \int_{\Omega} \rho(x, 0) \, dx, & \forall t \in (0, T) \\
 (iii) & \frac{1}{2} \frac{d}{dt} \int_{\Omega} \rho(x, t) u(x, t)^2 \, dx + \frac{d}{dt} \int_{\Omega} \rho(x, t) P(\rho(x, t)) \, dx \\
 & \quad + \int_{\Omega} \tau(u(x, t)) : \nabla u(x, t) \, dx = 0, & \forall t \in (0, T).
 \end{array} \right. \tag{1.2}$$

In the latter relation,  $P(\cdot)$ , the “elastic potential”, is a function derived from the equation of state, which satisfies:

$$P'(z) = \frac{\varphi(z)}{z^2} \tag{1.3}$$

where  $\varphi(\cdot)$  is the inverse function of  $\varrho(\cdot)$ , *i.e.* the function giving the pressure as a function of the density. The usual choice for  $P(\cdot)$  is, provided that this expression makes sense:

$$P(z) = \int_0^z \frac{\varphi(s)}{s^2} \, ds. \tag{1.4}$$

For these estimates to hold, the condition (1.2)-(i) must be satisfied by the initial condition; note that a non-zero forcing term  $f_v$  in the momentum balance would make an additional term appear at the right hand side of relation (1.2)-(iii). This latter estimate is obtained from the second relation of (1.1) (i.e. the momentum balance) by taking the inner product by  $u$  and integrating over  $\Omega$ . This computation then involves two main arguments which read:

$$\begin{aligned}
 (i) \quad \text{Stability of the advection operator:} \quad & \int_{\Omega} \left[ \frac{\partial}{\partial t}(\rho u) + \nabla \cdot (\rho u \otimes u) \right] \cdot u \, dx = \frac{1}{2} \frac{d}{dt} \int_{\Omega} \rho |u|^2 \, dx \\
 (ii) \quad \text{Stability due to the pressure work:} \quad & \int_{\Omega} -p \nabla \cdot u \, dx = \frac{d}{dt} \int_{\Omega} \rho(x, t) P(\rho(x, t)) \, dx.
 \end{aligned}
 \tag{1.5}$$

Note that the derivation of both relations rely on the mass balance equation.

This paper is organized as follows.

We first derive a bound similar to (1.5)-(ii) for a given class of spatial discretizations; the latter are introduced in Section 2.1 and the desired stability estimate (Thm. 2.1) is stated and proven in Section 2.2. We then show that this result allows to prove the existence of a solution for a fairly general class of discrete compressible flow problems. Section 2 gathers this whole study, and constitutes the first part of this paper.

In a second part (Sect. 3), we turn to the derivation of a pressure correction scheme, the solution of which satisfies a discrete equivalent of the whole set of *a priori* estimates (1.2). To this purpose, besides Theorem 2.1, we need as a second key ingredient a discrete version of the bound (1.5)-(i) relative to the stability of the advection operator, which is stated and proven in Section 3.2 (Thm. 3.1). We then derive a fully discrete algorithm which is designed to meet the assumptions of these theoretical results, and establish its stability. Moreover, numerical experiments show that, for smooth solutions, this scheme converges as expected, namely with first order in time convergence for all the variables and first to second order in space in  $L^2$  and discrete  $L^2$  norm for the velocity and the pressure, respectively.

## 2. ANALYSIS OF A CLASS OF DISCRETE PROBLEMS

The class of problems addressed in this section can be seen as the class of discrete systems obtained by space discretization with low-order non-conforming finite elements of continuous problems of the following form:

$$\left\{ \begin{array}{ll} Au + \nabla p = f_v & \text{in } \Omega \\ \frac{\varrho(p) - \rho^*}{dt} + \nabla \cdot (\varrho(p) u) = 0 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{array} \right. \tag{2.1}$$

where  $A$  stands for an abstract elliptic operator and the forcing term  $f_v$  and the density field  $\rho^*$  are known quantities. The unknowns of the problem are the velocity  $u$  and the pressure  $p$ ; the function  $\varrho(\cdot)$  stands for the equation of state. The domain  $\Omega$  is a polygonal ( $d = 2$ ) or polyhedral ( $d = 3$ ) open, bounded and connected subset of  $\mathbb{R}^d$ . Of course, at the continuous level, this statement of the problem should be completed by a precise definition of the functional spaces in which the velocity and the pressure are sought, together with regularity assumptions on the data. This is out of the scope here, since system (2.1) is only given to fix ideas; indeed, the aim here is to prove some mathematical properties of the discrete problem, namely to establish some *a priori* estimates for its solution and to prove that this nonlinear problem admits solutions for fairly general equations of state.

This section is organized as follows. We begin by describing the considered discretization and precisely stating the discrete problem at hand. Then we prove, for the chosen particular discretization, a fundamental result which is a discrete analogue of the elastic potential identity (1.5)-(ii). The next section is devoted to the proof of the existence of a solution, and we finally conclude by giving some practical examples of application of the abstract theory developed here.

### 2.1. The discrete problem

Let  $\mathcal{M}$  be a decomposition of the domain  $\Omega$  either into convex quadrilaterals ( $d = 2$ ) or hexahedra ( $d = 3$ ) or in simplices. By  $\mathcal{E}$  and  $\mathcal{E}(K)$  we denote the set of all  $(d - 1)$ -edges  $\sigma$  of the mesh and of the element  $K \in \mathcal{M}$  respectively. The set of  $(d - 1)$ -edges included in the boundary of  $\Omega$  is denoted by  $\mathcal{E}_{\text{ext}}$  and the set of internal ones (*i.e.*  $\mathcal{E} \setminus \mathcal{E}_{\text{ext}}$ ) is denoted by  $\mathcal{E}_{\text{int}}$ . The decomposition  $\mathcal{M}$  is supposed to be regular in the usual sense of the finite element literature (*e.g.* [6]), and, in particular,  $\mathcal{M}$  satisfies the following properties:  $\bar{\Omega} = \bigcup_{K \in \mathcal{M}} \bar{K}$ ; if  $K, L \in \mathcal{M}$ , then  $\bar{K} \cap \bar{L}$  is reduced to the empty set, to a vertex or (if  $d = 3$ ) to a segment, or  $\bar{K} \cap \bar{L}$  is (the closure of) a common  $(d - 1)$ -edge of  $K$  and  $L$ , which is denoted by  $K|L$ . For each internal edge of the mesh  $\sigma = K|L$ ,  $n_{KL}$  stands for the normal vector of  $\sigma$ , oriented from  $K$  to  $L$ . By  $|K|$  and  $|\sigma|$  we denote the measure, respectively, of  $K$  and of the edge  $\sigma$ .

The space discretization relies either on the so-called “rotated bilinear element”/ $P_0$  introduced by Rannacher and Turek [35] for quadrilateral or hexahedric meshes, or on the Crouzeix-Raviart element (see [8] for the seminal paper and, for instance [12], pp. 199–201, for a synthetic presentation) for simplicial meshes. The reference element  $\hat{K}$  for the rotated bilinear element is the unit  $d$ -cube (with edges parallel to the coordinate axes); the discrete functional space on  $\hat{K}$  is  $\tilde{Q}_1(\hat{K})^d$ , where  $\tilde{Q}_1(\hat{K})$  is defined as follows:

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

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. For both velocity elements used here, the degrees of freedom are determined by the following set of nodal functionals:

$$\{F_{\sigma,i}, \sigma \in \mathcal{E}(K), i = 1, \dots, d\}, \quad F_{\sigma,i}(v) = |\sigma|^{-1} \int_{\sigma} v_i \, d\gamma. \tag{2.2}$$

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. Finally, in both cases, the continuity of the average value of discrete velocities (*i.e.*, for a discrete velocity field  $v$ ,  $F_{\sigma,i}(v)$ ,  $1 \leq i \leq d$ ) across each edge of the mesh is required, thus the discrete space  $W_h$  is defined as follows:

$$W_h = \{ v_h \in L^2(\Omega)^d : v_h|_K \in W(K)^d, \forall K \in \mathcal{M}; F_{\sigma,i}(v_h) \text{ continuous across each edge } \sigma \in \mathcal{E}_{\text{int}}, 1 \leq i \leq d; F_{\sigma,i}(v_h) = 0, \forall \sigma \in \mathcal{E}_{\text{ext}}, 1 \leq i \leq d \}$$

where  $W(K)$  is the space of functions on  $K$  generated by  $\tilde{Q}_1(\hat{K})$  through the  $Q_1$  mapping from  $\hat{K}$  to  $K$  for the Rannacher-Turek element and the space of affine functions on  $K$  for the Crouzeix-Raviart element. For both the Rannacher-Turek and Crouzeix-Raviart discretizations, the pressure is approximated by the space  $L_h$  of piecewise constant functions:

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

Since only the continuity of the integral over each edge of the mesh is imposed, the velocities are discontinuous through each edge; the discretization is thus nonconforming in  $H^1(\Omega)^d$ . These pairs of approximation spaces for the velocity and the pressure are *inf-sup* stable, in the usual sense for “piecewise  $H^1$ ” discrete velocities, *i.e.* there exists  $c_{\text{is}} > 0$  possibly depending on the regularity of the shape of the cells but not on their size such that:

$$\forall p \in L_h, \quad \sup_{v \in W_h} \frac{\int_{\Omega,h} p \nabla \cdot v \, dx}{\|v\|_{1,b}} \geq c_{\text{is}} \|p - \bar{p}\|_{L^2(\Omega)}$$

where  $\bar{p}$  is the mean value of  $p$  over  $\Omega$ , the symbol  $\int_{\Omega,h}$  stands for  $\sum_{K \in \mathcal{M}} \int_K$  and  $\|\cdot\|_{1,b}$  stands for the broken Sobolev  $H^1$  semi-norm:

$$\|v\|_{1,b}^2 = \sum_{K \in \mathcal{M}} \int_K |\nabla v|^2 \, dx = \int_{\Omega,h} |\nabla v|^2 \, dx.$$

From definition (2.2), each velocity degree of freedom can be univocally associated to an element edge. Therefore we shall use hereafter, somewhat improperly, the expression “velocity on the edge  $\sigma$ ” to name the velocity vector defined by the degrees of freedom of the velocity components associated to  $\sigma$ . In addition, the velocity degrees of freedom are indexed by the number of the component and the associated edge, thus the set of velocity degrees of freedom reads:

$$\{v_{\sigma,i}, \sigma \in \mathcal{E}_{\text{int}}, 1 \leq i \leq d\}.$$

We define  $v_\sigma = \sum_{i=1}^d v_{\sigma,i} e_i$  where  $e_i$  is the  $i^{\text{th}}$  vector of the canonical basis of  $\mathbb{R}^d$ . We denote by  $\varphi_\sigma^{(i)}$  the vector shape function associated to  $v_{\sigma,i}$ , which, by the definition of the considered finite elements, reads:

$$\varphi_\sigma^{(i)} = \varphi_\sigma e_i$$

where  $\varphi_\sigma$  is a scalar function. Similarly, each degree of freedom for the pressure is associated to a mesh  $K$ , and the set of pressure degrees of freedom is denoted by  $\{p_K, K \in \mathcal{M}\}$ .

For any  $K \in \mathcal{M}$ , let  $\rho_K^*$  be a quantity approximating a known density  $\rho^*$  on  $K$ . The family of real numbers  $(\rho_K^*)_{K \in \mathcal{M}}$  is supposed to be positive. The discrete problem considered in this section reads:

$$\left\{ \begin{array}{ll} a(u, \varphi_\sigma^{(i)}) - \int_{\Omega,h} p \nabla \cdot \varphi_\sigma^{(i)} = \int_{\Omega} f_v \cdot \varphi_\sigma^{(i)} \, dx, & \forall \sigma \in \mathcal{E}_{\text{int}}, \text{ for } 1 \leq i \leq d \\ \frac{|K|}{\delta t} (\varrho(p_K) - \rho_K^*) + \sum_{\sigma=K|L} v_{\sigma,K}^+ \varrho(p_K) - v_{\sigma,K}^- \varrho(p_L) = 0, & \forall K \in \mathcal{M} \end{array} \right. \tag{2.3}$$

where  $v_{\sigma,K}^+$  and  $v_{\sigma,K}^-$  stand respectively for  $v_{\sigma,K}^+ = \max(v_{\sigma,K}, 0)$  and  $v_{\sigma,K}^- = -\min(v_{\sigma,K}, 0)$  with  $v_{\sigma,K} = |\sigma| u_\sigma \cdot n_{KL} = v_{\sigma,K}^+ - v_{\sigma,K}^-$ . The first equation is the standard finite element discretization of the first equation of (2.1), provided that the bilinear form  $a(\cdot, \cdot)$  is related to the operator  $A$  by a relation of the form:

$$a(v, w) = \int_{\Omega} Av \cdot w \, dx$$

where  $v$  and  $w$  are regular functions vanishing on the boundary (while this identity generally does not hold for functions of  $W_h$ ). Since the pressure is piecewise constant, the finite element discretization of the second relation of (2.1), *i.e.* the mass balance, is similar to a finite volume formulation, in which we introduce the standard first-order upwinding. The bilinear form  $a(\cdot, \cdot)$  is supposed to be elliptic on  $W_h$ , *i.e.* to be such that the following property holds:

$$\exists c_a > 0 \text{ such that, } \forall v \in W_h, \quad a(v, v) \geq c_a \|v\|_*^2$$

where  $\|\cdot\|_*$  is a norm over  $W_h$ . We denote by  $\|\cdot\|^*$  its dual norm with respect to the  $L^2(\Omega)^d$  inner product, defined by:

$$\forall v \in W_h, \quad \|v\|^* = \sup_{w \in W_h} \frac{\int_{\Omega} v \cdot w \, dx}{\|w\|_*}.$$

### 2.2. On the pressure control induced by the pressure forces work

The aim of this subsection is to prove that the discretization at hand satisfies a stability bound which can be seen as the discrete analogue of equation (1.5)-(ii), which we recall here:

$$-\int_{\Omega} p \nabla \cdot u \, dx = \frac{d}{dt} \int_{\Omega} \rho P(\rho) \, dx, \quad \text{where } P(\cdot) \text{ is such that } P'(z) = \frac{\wp(z)}{z^2}.$$

The formal computation which allows to derive this estimate in the continuous setting is the following. The starting point is the mass balance, which is multiplied by the derivative of  $z \mapsto z P(z)$  taken at  $\rho$ , denoted by  $[\rho P(\rho)]'$ :

$$[\rho P(\rho)]' \left[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) \right] = 0.$$

This relation yields:

$$\frac{\partial [\rho P(\rho)]}{\partial t} + [\rho P(\rho)]' [u \cdot \nabla \rho + \rho \nabla \cdot u] = 0. \tag{2.4}$$

And thus:

$$\frac{\partial [\rho P(\rho)]}{\partial t} + u \cdot \nabla [\rho P(\rho)] + [\rho P(\rho)]' \rho \nabla \cdot u = 0.$$

Developing the derivative, we get:

$$\frac{\partial [\rho P(\rho)]}{\partial t} + \underbrace{u \cdot \nabla [\rho P(\rho)] + \rho P(\rho) \nabla \cdot u}_{\nabla \cdot (\rho P(\rho) u)} + \underbrace{\rho^2 P'(\rho) \nabla \cdot u}_{p \nabla \cdot u} = 0 \tag{2.5}$$

and the result follows by integration in space, thanks to the fact that the velocity vanishes at the boundary. We are going here to reproduce this computation at the discrete level.

**Theorem 2.1** (stability due to the pressure work). *Let us suppose that the equation of state  $\varrho(\cdot)$  is defined over  $[0, +\infty)$ . Let  $P(\cdot)$  be an elastic potential (i.e. a function satisfying (1.3)) such that the function  $f : (0, +\infty) \rightarrow \mathbb{R}$  defined by  $f(z) = z P(z)$  is once continuously differentiable and strictly convex. Let  $(p_K)_{K \in \mathcal{M}}$  satisfy the second relation of (2.3). For any  $K \in \mathcal{M}$ , we suppose that  $p_K > 0$  and we define  $\rho_K$  by  $\rho_K = \varrho(p_K)$ ; we also recall that, by assumption,  $\rho_K^* > 0$ . Then the following estimate holds:*

$$-\int_{\Omega, h} p \nabla \cdot u \, dx = \sum_{K \in \mathcal{M}} -p_K \sum_{\sigma=K|L} v_{\sigma, K} \geq \frac{1}{\delta t} \sum_{K \in \mathcal{M}} |K| [\rho_K P(\rho_K) - \rho_K^* P(\rho_K^*)]. \tag{2.6}$$

*Proof.* Let us write the divergence term in the discrete mass balance over  $K$  (i.e. the second relation of (2.3)) under the following form:

$$\sum_{\sigma=K|L} \rho_{\sigma} v_{\sigma, K}$$

where  $\rho_{\sigma}$  is either  $\rho_K$  if  $v_{\sigma, K} \geq 0$  or  $\rho_L$  if  $v_{\sigma, K} \leq 0$ . Multiplying this term by  $f'(\rho_K)$ , we obtain:

$$T_{\text{div}, K} = f'(\rho_K) \sum_{\sigma=K|L} \rho_{\sigma} v_{\sigma, K} = f'(\rho_K) \left[ \sum_{\sigma=K|L} (\rho_{\sigma} - \rho_K) v_{\sigma, K} + \rho_K \sum_{\sigma=K|L} v_{\sigma, K} \right].$$

This latter form of  $T_{\text{div}, K}$  may be compared to equation (2.4): up to the multiplication by  $1/|K|$ , the first summation in the right hand side is the analogue of  $u \cdot \nabla \rho$  and the second one to  $\rho \nabla \cdot u$ .

Developing the derivative of  $f(\cdot)$ , we then obtain a discrete analogue of the corresponding terms in relation (2.5):

$$T_{\text{div},K} = f'(\rho_K) \sum_{\sigma=K|L} (\rho_\sigma - \rho_K) v_{\sigma,K} + \rho_K P(\rho_K) \sum_{\sigma=K|L} v_{\sigma,K} + \rho_K^2 P'(\rho_K) \sum_{\sigma=K|L} v_{\sigma,K}. \tag{2.7}$$

By definition (1.3) of  $P(\cdot)$ , the last term is equal to  $p_K \sum_{\sigma=K|L} v_{\sigma,K}$ . The process will be completed if we put the first two terms in divergence form. To this end, let us sum up the quantities  $T_{\text{div},K}$  over  $K \in \mathcal{M}$  and reorder the summation:

$$\sum_{K \in \mathcal{M}} T_{\text{div},K} = \sum_{K \in \mathcal{M}} p_K \sum_{\sigma=K|L} v_{\sigma,K} + \sum_{\sigma \in \mathcal{E}_{\text{int}}} T_{\text{div},\sigma} \tag{2.8}$$

where, if  $\sigma = K|L$ :

$$T_{\text{div},\sigma} = v_{\sigma,K} [f(\rho_K) + f'(\rho_K)(\rho_\sigma - \rho_K) - f(\rho_L) - f'(\rho_L)(\rho_\sigma - \rho_L)].$$

In this relation, there are two possible choices for the orientation of  $\sigma$ , *i.e.*  $K|L$  or  $L|K$ ; we suppose that the chosen orientation is such that  $v_{\sigma,K} \geq 0$ . Let  $\bar{\rho}_\sigma$  be defined by:

$$\left| \begin{array}{ll} \text{if } \rho_K \neq \rho_L: & f(\rho_K) + f'(\rho_K)(\bar{\rho}_\sigma - \rho_K) = f(\rho_L) + f'(\rho_L)(\bar{\rho}_\sigma - \rho_L) \\ \text{otherwise:} & \bar{\rho}_\sigma = \rho_K = \rho_L. \end{array} \right. \tag{2.9}$$

As the function  $f(\cdot)$  is supposed to be once continuously differentiable and strictly convex, the technical Lemma 2.3 proven hereafter applies and  $\bar{\rho}_\sigma$  is uniquely defined and satisfies  $\bar{\rho}_\sigma \in [\min(\rho_K, \rho_L), \max(\rho_K, \rho_L)]$ . By definition, the choice  $\rho_\sigma = \bar{\rho}_\sigma$  is such that the term  $T_{\text{div},\sigma}$  vanishes, and thus, with this centred choice for  $\rho_\sigma$ , the first two terms of equation (2.7) are a conservative approximation of the quantity  $\nabla \cdot (\rho P(\rho))u$  appearing in equation (2.5), with the following expression for the flux:

$$F_{\sigma,K} = [f(\rho)]_\sigma v_{\sigma,K}, \quad \text{with:} \quad [f(\rho)]_\sigma = f(\rho_K) + f'(\rho_K)(\bar{\rho}_\sigma - \rho_K) = f(\rho_L) + f'(\rho_L)(\bar{\rho}_\sigma - \rho_L).$$

Now, for any choice of  $\rho_\sigma$ , we have:

$$T_{\text{div},\sigma} = v_{\sigma,K} (\rho_\sigma - \bar{\rho}_\sigma) (f'(\rho_K) - f'(\rho_L)).$$

With the orientation taken for  $\sigma$ , an upwind choice for  $\rho_\sigma$  yields:

$$T_{\text{div},\sigma} = v_{\sigma,K} (\rho_K - \bar{\rho}_\sigma) (f'(\rho_K) - f'(\rho_L))$$

and, using the fact that  $f'(\cdot)$  is an increasing function since  $f(\cdot)$  is convex and that  $\min(\rho_K, \rho_L) \leq \bar{\rho}_\sigma \leq \max(\rho_K, \rho_L)$ , it is easily seen that  $T_{\text{div},\sigma}$  is non-negative.

Multiplying by  $f'(\rho_K)$  the mass balance over each cell  $K$  and summing for  $K \in \mathcal{M}$  thus yields, invoking equation (2.8):

$$- \sum_{K \in \mathcal{M}} p_K \sum_{\sigma=K|L} v_{\sigma,K} = R + \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} f'(\rho_K) (\rho_K - \rho_K^*) \tag{2.10}$$

where  $R$  is non-negative, and the result follows invoking once again the convexity of  $f(\cdot)$ . □

**Remark 2.2** (on a non-dissipative scheme). The preceding proof shows that, for a scheme to conserve the energy (*i.e.* to obtain a discrete equivalent of (1.2)-(iii)), besides other arguments, the choice of  $\bar{\rho}_\sigma$  given by (2.9) for the density at the edge of the control volume in the discretization of the flux in the mass balance seems to be mandatory; any other choice leads to an artificial (*i.e.* due to the numerical scheme) dissipation or production in the work of the pressure forces. Note however that, this discretization being essentially of centered type, the positivity of the density is not warranted in this case.

In the course of the preceding proof, we used the following technical lemma.

**Lemma 2.3.** *Let  $g(\cdot)$  be a strictly convex and once continuously derivable real function over an open interval  $I \subset \mathbb{R}$ . Let  $\rho_1 \in I$  and  $\rho_2 \in I$  be two distinct real numbers. Then the following relation:*

$$g(\rho_1) + g'(\rho_1)(\bar{\rho} - \rho_1) = g(\rho_2) + g'(\rho_2)(\bar{\rho} - \rho_2) \quad (2.11)$$

uniquely defines the real number  $\bar{\rho}$ . In addition, we have  $\bar{\rho} \in [\min(\rho_1, \rho_2), \max(\rho_1, \rho_2)]$ .

*Proof.* Without loss of generality, let us suppose that  $\rho_1 < \rho_2$ . Reordering equation (2.11), we get:

$$g(\rho_1) + g'(\rho_1)(\rho_2 - \rho_1) - g(\rho_2) = (\bar{\rho} - \rho_2) [g'(\rho_2) - g'(\rho_1)].$$

Since  $g(\cdot)$  is strictly convex,  $g'(\rho_2) - g'(\rho_1)$  does not vanish, and therefore the latter equation proves that  $\bar{\rho}$  is uniquely defined. In addition, for the same reason, the left hand side of this relation is negative and  $g'(\rho_2) - g'(\rho_1)$  is positive, thus we have  $\bar{\rho} < \rho_2$ . Reordering equation (2.11) yields:

$$g(\rho_2) + g'(\rho_2)(\rho_1 - \rho_2) - g(\rho_1) = (\bar{\rho} - \rho_1) [g'(\rho_1) - g'(\rho_2)]$$

which, considering the signs of the left hand side and of  $g'(\rho_1) - g'(\rho_2)$ , implies  $\bar{\rho} > \rho_1$ .  $\square$

### 2.3. Existence of a solution

The aim of this section is to prove the existence of a solution to the discrete problem (2.3). It follows from a topological degree argument, linking by a homotopy the problem at hand to a linear system.

This section begins with a lemma which is used to obtain a positive lower bound for the pressure in the sequel.

**Lemma 2.4.** *Let  $(p_K^*)_{K \in \mathcal{M}}$  and  $(p_K)_{K \in \mathcal{M}}$  be two families of real numbers such that:*

$$\forall K \in \mathcal{M}, \quad |K| \frac{\varphi_1(p_K) - \varphi_1(p_K^*)}{\delta t} + \sum_{\sigma=K|L} v_{\sigma,K}^+ \varphi_2(p_K) - v_{\sigma,K}^- \varphi_2(p_L) = 0 \quad (2.12)$$

where  $\varphi_1(\cdot)$  is an increasing function and  $\varphi_2(\cdot)$  is a non-decreasing and non-negative function. Suppose that there exists  $\bar{p}$  such that:

$$\varphi_1(\bar{p}) + \delta t \varphi_2(\bar{p}) \max \left[ 0, \max_{K \in \mathcal{M}} \left( \frac{1}{|K|} \sum_{\sigma=K|L} v_{\sigma,K} \right) \right] = \min_{K \in \mathcal{M}} [\varphi_1(p_K^*)]. \quad (2.13)$$

Then,  $\forall K \in \mathcal{M}$ ,  $p_K$  satisfies  $p_K \geq \bar{p}$ .

*Proof.* Let us assume that there exists a cell  $\bar{K}$  such that  $p_{\bar{K}} = \min_{K \in \mathcal{M}} p_K < \bar{p}$ . Multiplying by  $\delta t/|\bar{K}|$  the relation (2.12) written for  $K = \bar{K}$ , we get:

$$\varphi_1(p_{\bar{K}}) + \frac{\delta t}{|\bar{K}|} \sum_{\sigma=\bar{K}|L} \left[ v_{\sigma,\bar{K}}^+ \varphi_2(p_{\bar{K}}) - v_{\sigma,\bar{K}}^- \varphi_2(p_L) \right] = \varphi_1(p_{\bar{K}}^*). \quad (2.14)$$

Then, subtracting equation (2.13), we have:

$$\begin{aligned} \varphi_1(p_{\bar{K}}) - \varphi_1(\bar{p}) + \frac{\delta t}{|\bar{K}|} \sum_{\sigma=\bar{K}|L} \left[ v_{\sigma,\bar{K}}^+ \varphi_2(p_{\bar{K}}) - v_{\sigma,\bar{K}}^- \varphi_2(p_L) \right] \\ - \delta t \varphi_2(\bar{p}) \max \left[ 0, \max_{K \in \mathcal{M}} \left( \frac{1}{|K|} \sum_{\sigma=K|L} v_{\sigma,K} \right) \right] = \varphi_1(p_{\bar{K}}^*) - \min_{K \in \mathcal{M}} [\varphi_1(p_K^*)] \geq 0. \end{aligned}$$

The previous relation can be written as  $T_1 + T_2 + T_3 \geq 0$  with:

$$\begin{aligned} T_1 &= \varphi_1(p_{\bar{K}}) - \varphi_1(\bar{p}) \\ T_2 &= \delta t \varphi_2(p_{\bar{K}}) \left[ \frac{1}{|\bar{K}|} \sum_{\sigma=\bar{K}|L} v_{\sigma,\bar{K}} \right] - \delta t \varphi_2(\bar{p}) \max \left[ 0, \max_{K \in \mathcal{M}} \left( \frac{1}{|K|} \sum_{\sigma=K|L} v_{\sigma,K} \right) \right] \\ T_3 &= \frac{\delta t}{|\bar{K}|} \sum_{\sigma=\bar{K}|L} v_{\sigma,K}^- (\varphi_2(p_{\bar{K}}) - \varphi_2(p_L)). \end{aligned}$$

Since  $\varphi_1(\cdot)$  is an increasing function and, by assumption,  $p_{\bar{K}} < \bar{p}$ , we have  $T_1 < 0$ . Similarly,  $0 \leq \varphi_2(p_{\bar{K}}) \leq \varphi_2(\bar{p})$  and the discrete divergence over  $\bar{K}$  (i.e.  $1/|\bar{K}| \sum_{\sigma=\bar{K}|L} v_{\sigma,\bar{K}}$ ) is necessarily smaller than the maximum of this quantity over the cells of the mesh, thus  $T_2 \leq 0$ . Finally, since, by assumption,  $p_{\bar{K}} \leq p_L$  for any neighbouring cell  $L$  of  $\bar{K}$ ,  $\varphi_2(\cdot)$  is a non-decreasing function and  $v_{\sigma,K}^- \geq 0$ ,  $T_3 \leq 0$ . We thus obtain a contradiction with the fact that  $T_1 + T_2 + T_3 \geq 0$ , which proves that  $p_K \geq \bar{p}$ ,  $\forall K \in \mathcal{M}$ .  $\square$

We now state the abstract theorem which will be used hereafter; this result follows from standard arguments of the topological degree theory (see [9] for an exposition of the theory and [13] for another utilisation for the same objective as here, namely the proof of existence of a solution to a numerical scheme).

**Theorem 2.5** (a result from the topological degree theory). *Let  $N$  and  $M$  be two positive integers and  $V$  be defined as follows:*

$$V = \{(x, y) \in \mathbb{R}^N \times \mathbb{R}^M \text{ such that } y > 0\}$$

where, for any real number  $c$ , the notation  $y > c$  means that each component of  $y$  is greater than  $c$ . Let  $b \in \mathbb{R}^N \times \mathbb{R}^M$  and  $f(\cdot)$  and  $F(\cdot, \cdot)$  be two continuous functions respectively from  $V$  and  $V \times [0, 1]$  to  $\mathbb{R}^N \times \mathbb{R}^M$  satisfying:

- (i)  $F(\cdot, 1) = f(\cdot)$ ;
- (ii)  $\forall \alpha \in [0, 1]$ , if  $v \in V$  is such that  $F(v, \alpha) = b$  then  $v \in W$ , where  $W$  is defined as follows:

$$W = \{(x, y) \in \mathbb{R}^N \times \mathbb{R}^M \text{ s.t. } \|x\| < C_1 \text{ and } \epsilon < y < C_2\}$$

with  $C_1, C_2$  and  $\epsilon$  three positive constants and  $\|\cdot\|$  a norm defined over  $\mathbb{R}^N$ ;

- (iii) the topological degree of  $F(\cdot, 0)$  with respect to  $b$  and  $W$  is equal to  $d_0 \neq 0$ .

Then the topological degree of  $F(\cdot, 1)$  with respect to  $b$  and  $W$  is also equal to  $d_0 \neq 0$ ; consequently, there exists at least a solution  $v \in W$  such that  $f(v) = b$ .

We are now in position to prove the existence of a solution to the discrete problem (2.3), for fairly general equations of states.

**Theorem 2.6** (existence of a solution). *Let us suppose that the equation of state  $\varrho(\cdot)$  is such that:*

- (1)  $\varrho(\cdot)$  is defined and increasing over  $[0, +\infty)$ ,  $\varrho(0) = 0$  and  $\lim_{z \rightarrow +\infty} \varrho(z) = +\infty$ ;
- (2) there exists an elastic potential  $P(\cdot)$  (i.e. a function satisfying (1.3)) such that the function  $f : (0, +\infty) \rightarrow \mathbb{R}$  defined by  $f(z) = z P(z)$  is once continuously differentiable, strictly convex and  $f(z) \geq -C_P$ ,  $\forall z \in (0, +\infty)$ , where  $C_P$  is a non-negative constant.

In addition, we recall that, in the discrete problem at hand (2.3),  $\rho_K^* > 0, \forall K \in \mathcal{M}$ .

Then problem (2.3) admits at least one solution  $(u_\sigma, p_K)_{\sigma \in \mathcal{E}_{\text{int}}, K \in \mathcal{M}}$ , and any solution is such that,  $\forall K \in \mathcal{M}$ ,  $p_K$  is positive.

*Proof.* Let  $N = d \text{ card}(\mathcal{E}_{\text{int}})$  and  $M = \text{card}(\mathcal{M})$ . We identify the space of discrete velocities  $W_h$  and pressures  $L_h$  to  $\mathbb{R}^N$  and  $\mathbb{R}^M$  respectively and, keeping the same notation for the discrete functions and the associated vectors of degree of freedom, we define  $V$  by:

$$V = \{(u, p) \in \mathbb{R}^N \times \mathbb{R}^M \text{ such that } p > 0\}.$$

Let the mapping  $F : V \times [0, 1] \rightarrow \mathbb{R}^N \times \mathbb{R}^M$  be given by:

$$F(u, p, \alpha) = \begin{cases} v_{\sigma,i} = a(u, \varphi_\sigma^{(i)}) - \alpha \int_{\Omega,h} p \nabla \cdot \varphi_\sigma^{(i)} dx - \int_\Omega f_v \cdot \varphi_\sigma^{(i)} dx, & \forall \sigma \in \mathcal{E}_{\text{int}}, \text{ for } 1 \leq i \leq d \\ q_K = \frac{|K|}{\delta t} [\varrho(p_K) - \varrho(p_K^*)] + \alpha \sum_{\sigma=K|L} v_{\sigma,K}^+ \varrho(p_K) - v_{\sigma,K}^- \varrho(p_L), & \forall K \in \mathcal{M} \end{cases} \quad (2.15)$$

where,  $\forall K \in \mathcal{M}$ ,  $p_K^*$  is chosen such that  $\rho_K^* = \varrho(p_K^*)$ ; note that, by assumption,  $\varrho(\cdot)$  is one to one from  $(0, +\infty)$  to  $(0, +\infty)$ , so the previous definition makes sense. The problem  $F(u, p, 1) = 0$  is exactly system (2.3).

The present proof is obtained by applying Theorem 2.5 with  $b = 0$ ; we are thus going to show that any solution of  $F(u, p, \alpha) = 0$  satisfies suitable *a priori* estimates. To this purpose, we progress as follows. First, Lemma 2.4 shows that the pressure is positive, thus Theorem 2.1 applies, and we obtain a control on  $u$  in the discrete norm associated to  $a(\cdot, \cdot)$ , uniform with respect to  $\alpha$ . Since we work in a finite dimensional space, we then obtain a control on  $p$  in  $L^\infty$  by using the conservativity of the system of equations. For the same reason, the control on  $u$  yields a bound in  $L^\infty$  of the value of the discrete divergence, which is shown to allow, by Lemma 2.4, to bound  $p$  away from zero independently of  $\alpha$ . The proof finally ends by examining the properties of the system  $F(u, p, 0) = 0$ .

**Step 1.**  $\alpha \in (0, 1]$ ,  $\|\cdot\|_*$  estimate for the velocity.

Applying Lemma 2.4 to the second equation  $F(u, p, \alpha) = 0$  (i.e. the relation obtained by setting  $q_K = 0$  in (2.15)) with  $\varphi_1(\cdot) = \varphi_2(\cdot) = \varrho(\cdot)$ , we get:

$$\forall K \in \mathcal{M}, \quad p_K \geq \bar{p}_\alpha \quad (2.16)$$

where  $\bar{p}_\alpha$  is given by:

$$\varrho(\bar{p}_\alpha) = \frac{\min_{K \in \mathcal{M}} \varrho(p_K^*)}{1 + \delta t \max_{K \in \mathcal{M}} \left[ 0, \frac{\alpha}{|K|} \sum_{\sigma=K|L} v_{\sigma,K} \right]}.$$

Note that  $\bar{p}_\alpha$  is well defined since, by assumption,  $\varrho(\cdot)$  is one to one from  $(0, +\infty)$  to  $(0, +\infty)$ , and  $\bar{p}_\alpha > 0$ , for any discrete velocity field  $u$ . The pressure is thus positive. Setting now  $v_\sigma = 0$  in (2.15), multiplying

the corresponding equation by  $u_{\sigma,i}$  and summing over  $\sigma \in \mathcal{E}_{\text{int}}$  and  $1 \leq i \leq d$  yields the following equation:

$$a(u, u) - \alpha \int_{\Omega, h} p \nabla \cdot u \, dx = \int_{\Omega} f_v \cdot u \, dx.$$

Since the pressure is positive, by a computation very similar to the proof of Theorem 2.1, we see that, from the second relation of (2.15) with  $q_K = 0$ :

$$-\alpha \int_{\Omega, h} p \nabla \cdot u \, dx \geq \frac{1}{\delta t} \sum_{K \in \mathcal{M}} |K| [\rho_K P(\rho_K) - \rho_K^* P(\rho_K^*)]$$

where  $\rho_K = \varrho(p_K)$ . By the stability of the bilinear form  $a(\cdot, \cdot)$  and Young's inequality, we thus get:

$$\underbrace{\frac{c_a}{2} \|u\|_*^2}_{T_1} + \underbrace{\frac{1}{\delta t} \sum_{K \in \mathcal{M}} |K| \rho_K P(\rho_K)}_{T_2} \leq \frac{1}{2c_a} \|f_v\|_*^2 + \frac{1}{\delta t} \sum_{K \in \mathcal{M}} |K| \rho_K^* P(\rho_K^*). \tag{2.17}$$

By assumption,  $T_2 \geq -C_p |\Omega|$  and we thus get the following estimate on the discrete norm of the velocity:

$$\|u\|_* \leq C_1 \tag{2.18}$$

where  $C_1$  only depends on the data of the problem, *i.e.* the bilinear form  $a(\cdot, \cdot)$ ,  $f_v$ ,  $\rho^*$ , the mesh and  $\delta t$  and not on  $\alpha$ .

**Step 2.**  $\alpha \in (0, 1]$ ,  $L^\infty$  estimate for the pressure.

Let us now turn to the estimate of the pressure. By conservativity of the discrete mass balance, it is easily seen that:

$$\sum_{K \in \mathcal{M}} |K| \varrho(p_K) = \sum_{K \in \mathcal{M}} |K| \rho_K^*.$$

Since each term in the sum on the left hand side is non-negative, we thus have:

$$\forall K \in \mathcal{M}, \quad \varrho(p_K) \leq \frac{1}{\min_{K \in \mathcal{M}} (|K|)} \sum_{K \in \mathcal{M}} |K| \rho_K^*$$

which, since by assumption  $\lim_{z \rightarrow +\infty} \varrho(z) = +\infty$ , yields:

$$\forall K \in \mathcal{M}, \quad p_K \leq C_2 \tag{2.19}$$

where  $C_2$  only depends on the data of the problem.

**Step 3.**  $\alpha \in (0, 1]$ ,  $p$  bounded away from zero.

We now exploit the estimate (2.16). As  $\alpha \leq 1$ , we get:

$$\varrho(\bar{p}_\alpha) \geq \frac{\min_{K \in \mathcal{M}} \varrho(p_K^*)}{1 + \delta t \max_{K \in \mathcal{M}} \left[ 0, \frac{1}{|K|} \sum_{\sigma=K|L} v_{\sigma, K} \right]}$$

and, by equivalence of norms in a finite dimensional space, the bound (2.18) also yields a bound in the  $L^\infty$  norm and, finally, an upper bound for the denominator of the fraction at the right hand side of this relation. We thus get, still since  $\varrho(\cdot)$  is increasing on  $(0, +\infty)$ , that,  $\forall \alpha \in (0, 1]$ ,  $\bar{p}_\alpha \geq \epsilon_1$ , and, finally:

$$\forall K \in \mathcal{M}, \quad p_K \geq \epsilon_1 \tag{2.20}$$

where  $\epsilon_1$  only depends on the data.

**Step 4.** Conclusion.

For  $\alpha = 0$ , the system  $F(u, p, 0) = 0$  reads:

$$\left\{ \begin{array}{ll} a(u, \varphi_\sigma^{(i)}) = \int_\Omega f_v \cdot \varphi_\sigma^{(i)} dx & \forall \sigma \in \mathcal{E}_{\text{int}}, 1 \leq i \leq d \\ \varrho(p_K) = \varrho(p_K^*) & \forall K \in \mathcal{M}. \end{array} \right.$$

Since  $\varrho(\cdot)$  is one to one from  $(0, +\infty)$  to  $(0, +\infty)$  and thanks to the stability of the bilinear form  $a(\cdot, \cdot)$ , this system has one and only one solution (which, for the pressure, reads of course  $p_K = p_K^*$ ,  $\forall K \in \mathcal{M}$ ), which satisfies:

$$\|u\|_* \leq C_3, \quad \epsilon_2 = \min_{K \in \mathcal{M}} p_K^* \leq p \leq \max_{K \in \mathcal{M}} p_K^* = C_4. \tag{2.21}$$

Let  $W$  be defined by:

$$W = \left\{ (u, p) \in \mathbb{R}^N \times \mathbb{R}^M \text{ such that } \|u\|_* < 2 \max(C_1, C_3) \text{ and } \frac{1}{2} \min(\epsilon_1, \epsilon_2) < p < 2 \max(C_2, C_4) \right\}.$$

We now need to prove that the topological degree  $d_0$  of  $F(\cdot, \cdot, 0)$  with respect to 0 and  $W$  is not zero. Let us first suppose that the function  $\varrho(\cdot)$  is continuously differentiable and that its derivative is positive over  $(0, +\infty)$ . The Jacobian matrix of the system  $F(u, p, 0) = 0$  is block diagonal: the first block, associated to the first relation, is constant (this part of the system is linear) and non-singular; the second one, associated to the second relation, is diagonal, and each diagonal entry is equal to the derivative of  $\varrho(\cdot)$ , taken at the considered point. The determinant of this Jacobian matrix thus does not vanish for the solution of the system, and  $d_0 \neq 0$ . This proof can then be extended to a continuous increasing function  $\varrho(\cdot)$  by a regularization technique. Hence, finally, by inequalities (2.18), (2.19), (2.20) and (2.21), Theorem 2.5 applies, which concludes the proof.  $\square$

**2.4. Some cases of application**

First of all, let us give some examples for the bilinear form  $a(\cdot, \cdot)$ , for which the theory developed in this work holds. The first of them is:

$$a(u, v) = \int_\Omega u \cdot v \, dx, \quad \|u\|_* = \|u\|_{L^2(\Omega)^d}, \quad \|f_v\|^* = \|f_v\|_{L^2(\Omega)^d}.$$

This choice for  $a(\cdot, \cdot)$  yields a discrete Darcy-like problem which is, up to numerical integration technicalities, the projection step arising in the pressure correction scheme which is considered in the present paper (see Sect. 3). Note that, in this case, the boundary condition  $u \in H_0^1(\Omega)^d$  does not make sense at the continuous level; in addition, the considered discretization is known to be not consistent enough to yield convergence (see Rem. 3.6 hereafter) for the Darcy problem.

The bilinear form associated to the Stokes problem provides another example of application. It may read in this case:

$$a(u, v) = \int_{\Omega, h} \nabla u \cdot \nabla v \, dx, \quad \|u\|_* = |u|_{H^1(\Omega)^d}$$

or, without additional theoretical difficulties:

$$a(u, v) = \mu \int_{\Omega, h} \nabla u \cdot \nabla v \, dx + \frac{\mu}{3} \int_{\Omega, h} (\nabla \cdot u) (\nabla \cdot v) \, dx$$

this latter form, where the real number  $\mu > 0$  is the viscosity, corresponding to the physical shear stress tensor expression for a compressible flow of a constant viscosity Newtonian fluid.

In addition, consider a time step of a (semi-)implicit time discretization of the unsteady Navier-Stokes equations, in which case  $a(\cdot, \cdot)$  and  $f_v$  read:

$$a(u, v) = \frac{1}{\delta t} \int_{\Omega} \rho u \cdot v \, dx + \int_{\Omega, h} \nabla \cdot (\rho w \otimes u) \cdot v \, dx + \mu \int_{\Omega, h} \nabla u \cdot \nabla v \, dx + \frac{\mu}{3} \int_{\Omega, h} (\nabla \cdot u) (\nabla \cdot v) \, dx$$

$$f_v = \frac{1}{\delta t} \rho^* u^* + f_{v,0}$$

where  $f_{v,0}$  is the physical forcing term,  $\rho^*$  and  $u^*$  stand for known density and velocity fields and  $w$  is an advection field, which may be  $u$  itself or be derived from the velocity obtained at the previous time steps. Let us suppose that the following identity holds:

$$\frac{1}{\delta t} \int_{\Omega} (\rho u - \rho^* u^*) \cdot u \, dx + \int_{\Omega, h} \nabla \cdot (\rho w \otimes u) \cdot u \, dx \geq \frac{1}{2\delta t} \left[ \int_{\Omega} \rho |u|^2 - \int_{\Omega} \rho^* |u^*|^2 \right]$$

which is the discrete counterpart of equation (1.5)-(i). The algorithm considered in this paper provides an example where this condition is verified (see Sect. 3). Then the present theory applies with few modifications: in the proof of existence of Theorem 2.6, the right hand side of the preceding equation must be multiplied by the homotopy parameter  $\alpha$  (and thus this term vanishes at  $\alpha = 0$ , which yields the problem considered in Step 4 above); the (uniform with respect to  $\alpha$ ) stability in Step 1 stems from the diffusion term, and Steps 2 and 3 remain unchanged.

Note finally that, in the steady state case, an additional constraint is needed for the problem to have a chance to be well posed, namely to impose the total mass  $M$  of fluid in the computational domain to a given value. This constraint can be simply enforced by solving an approximate mass balance which reads:

$$c(h) \left[ \rho - \frac{M}{|\Omega|} \right] + \nabla \cdot \rho u = 0$$

where  $|\Omega|$  stands for the measure of  $\Omega$ ,  $h$  is the spatial discretization step and  $c(h) > 0$  must tend to zero with  $h$ , fast enough to avoid any loss of consistency. With this form of the mass balance, the theory developed here directly applies to this case too, provided that the corresponding unsteady-like term is also introduced in the momentum balance equation.

Examining now the assumptions for the equation of state in Theorem 2.6, we see that our results hold with equations of state of the form:

$$\varrho(p) = p^{1/\gamma} \quad \text{or, equivalently} \quad \rho = p^\gamma, \quad \text{where } \gamma > 1.$$

In this case, the elastic potential is given by equation (1.4), which yields:

$$P(\rho) = \frac{1}{\gamma - 1} \rho^{\gamma-1}, \quad \rho P(\rho) = \frac{1}{\gamma - 1} \rho^\gamma \quad \left( = \frac{1}{\gamma - 1} p \right).$$

The same conclusion still holds with  $\gamma = 1$  (*i.e.*  $p = \rho$ ), with  $P(\rho) = \log(\rho)$  satisfying equation (1.3). The case  $\gamma > 1$  is for instance encountered for isentropic perfect gas flows, whereas  $\gamma = 1$  corresponds to the

isothermal case. It is worth noting that this range of application is larger than what is known for the continuous case, for which the existence of a solution is known only in the case  $\gamma > d/2$  [15,28,32].

### 3. A PRESSURE CORRECTION SCHEME

In this section, we build a pressure correction numerical scheme for the solution of the compressible barotropic Navier-Stokes equations (1.1), based on the low order non-conforming finite element spaces used in the previous section, namely the Crouzeix-Raviart or Rannacher-Turek elements.

The presentation is organized as follows. First, we write the scheme in the time semi-discrete setting (Sect. 3.1). Then we prove a general stability estimate which applies to the discretization by a finite volume technique of the convection operator (Sect. 3.2). The proposed scheme is built in such a way that the assumptions of this stability result hold (Sect. 3.3); this implies first a prediction of the density, as a non-standard first step of the algorithm and, second, a discretization of the convection terms in the momentum balance equation by a finite volume technique which is especially designed to this purpose. The discretization of the projection step (Sect. 3.4) also combines the finite element and finite volume methods, in such a way that the theory developed in Section 2 applies; in particular, the proposed discretization allows to take benefit of the pressure or density control induced by the pressure work, *i.e.* to apply Theorem 2.1. The remaining steps of the algorithm are described in Section 3.5 and an overview of the scheme is given in Section 3.6. The following section (Sect. 3.7) is devoted to the proof of the stability of the algorithm. Finally, we shortly address some implementation difficulties (Sect. 3.8), then we provide some numerical tests (Sect. 3.9) which are performed to assess the time and space convergence of the scheme.

#### 3.1. Time semi-discrete formulation

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

$$1 - \text{Solve for } \tilde{\rho}^{n+1}: \quad \frac{\tilde{\rho}^{n+1} - \rho^n}{\delta t} + \nabla \cdot (\tilde{\rho}^{n+1} u^n) = 0. \quad (3.1)$$

$$2 - \text{Solve for } \tilde{p}^{n+1}: \quad -\nabla \cdot \left( \frac{1}{\tilde{\rho}^{n+1}} \nabla \tilde{p}^{n+1} \right) = -\nabla \cdot \left( \frac{1}{\sqrt{\tilde{\rho}^{n+1} \tilde{\rho}^n}} \nabla p^n \right). \quad (3.2)$$

$$3 - \text{Solve for } \tilde{u}^{n+1}: \\ \frac{\tilde{\rho}^{n+1} \tilde{u}^{n+1} - \rho^n u^n}{\delta t} + \nabla \cdot (\tilde{\rho}^{n+1} u^n \otimes \tilde{u}^{n+1}) + \nabla \tilde{p}^{n+1} - \nabla \cdot \tau(\tilde{u}^{n+1}) = f_v^{n+1}. \quad (3.3)$$

$$4 - \text{Solve for } \bar{u}^{n+1}, p^{n+1}, \rho^{n+1}: \\ \left| \begin{array}{l} \tilde{\rho}^{n+1} \frac{\bar{u}^{n+1} - \tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1} - \tilde{p}^{n+1}) = 0 \\ \frac{\varrho(p^{n+1}) - \rho^n}{\delta t} + \nabla \cdot (\varrho(p^{n+1}) \bar{u}^{n+1}) = 0 \\ \rho^{n+1} = \varrho(p^{n+1}). \end{array} \right. \quad (3.4)$$

$$5 - \text{Compute } u^{n+1} \text{ given by: } \quad \sqrt{\rho^{n+1}} u^{n+1} = \sqrt{\tilde{\rho}^{n+1}} \bar{u}^{n+1}. \quad (3.5)$$

The first step is a prediction of the density, used for the discretization of the time derivative of the momentum. As remarked by Bijl and Wesseling [2] and Wesseling [43], this step can be avoided when solving the Euler equations: in this case, the mass flowrate may be chosen as an unknown, using the explicit velocity as an advective field in the discretization of the convection term in the momentum balance; the velocity is then updated by dividing by the density at the end of the time step. For viscous flows, if the discretization of the diffusion term is chosen to be implicit, both the mass flowrate and the velocity appear as unknowns

in the momentum balance; this seems to impede the use of this trick. Let us emphasize that the special way that step one is carried out (*i.e.* solving a discretization of the mass balance instead as, for instance, performing a Richardson’s extrapolation) is crucial for the stability.

Likewise, the second step is a renormalization of the pressure the interest of which is clarified only by the stability analysis. A similar technique has already been introduced by Guermond and Quartapelle for variable density incompressible flows [18].

Step 3 consists in a classical semi-implicit solution of the momentum balance equation to obtain a predicted velocity.

Step 4 is a nonlinear pressure correction step, which degenerates in the usual projection step as used in incompressible flow solvers when the density is constant (*e.g.* [29]). Taking the divergence of the first relation of (3.4) and using the second one to eliminate the unknown velocity  $\bar{u}^{n+1}$  yields a non-linear elliptic problem for the pressure. This computation is formal in the semi-discrete formulation, but, of course, is necessarily made clear at the algebraic level, as described in Section 3.8. Once the pressure is computed, the first relation yields the updated velocity and the third one gives the end-of-step density.

Finally, Step 5 is a renormalization of the velocity, once again useful for stability reasons.

### 3.2. Stability of the advection operator: a finite-volume result

The aim of this section is to state and prove a discrete analogue to the stability identity (1.5)-(i), which may be written for any sufficiently regular functions  $\rho$ ,  $z$  and  $u$  as follows:

$$\int_{\Omega} \left[ \frac{\partial \rho z}{\partial t} + \nabla \cdot (\rho z u) \right] z \, dx = \frac{1}{2} \frac{d}{dt} \int_{\Omega} \rho z^2 \, dx$$

and which holds if the velocity  $u$  vanishes at the boundary of the computational domain  $\Omega$  and provided that the following balance is satisfied by  $\rho$  and  $u$ :

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

As stated in the introduction, applying this identity to each component of the velocity yields the central argument of the proof of the kinetic energy theorem.

The discrete analogue to this identity follows. This result is presented in a general algebraic setting, with no reference to the underlying partial differential equation (see Rem. 3.2 hereafter for a clarification of this link); note however that, in the following relations, the sum of the fluxes is restricted to the internal edges of the mesh, which implicitly reflects the fact that the normal velocity is supposed to be zero at the boundary.

**Theorem 3.1** (stability of the advection operator). *Let  $(\rho_K^*)_{K \in \mathcal{M}}$  and  $(\rho_K)_{K \in \mathcal{M}}$  be two families of positive real numbers satisfying the following set of equations:*

$$\forall K \in \mathcal{M}, \quad \frac{|K|}{\delta t} (\rho_K - \rho_K^*) + \sum_{\sigma=K|L} F_{\sigma,K} = 0 \tag{3.6}$$

where  $F_{\sigma,K}$  is a quantity associated to the edge  $\sigma$  and to the control volume  $K$ ; we suppose that, for any internal edge  $\sigma = K|L$ ,  $F_{\sigma,K} = -F_{\sigma,L}$ . Let  $(z_K^*)_{K \in \mathcal{M}}$  and  $(z_K)_{K \in \mathcal{M}}$  be two families of real numbers. For any internal edge  $\sigma = K|L$ , we define  $z_{\sigma}$  either by  $z_{\sigma} = \frac{1}{2}(z_K + z_L)$ , or by  $z_{\sigma} = z_K$  if  $F_{\sigma,K} \geq 0$  and  $z_{\sigma} = z_L$  otherwise. The first choice is usually referred to as the “centered choice”, the second one as “the upwind choice”. In both cases, the following stability property holds:

$$\sum_{K \in \mathcal{M}} z_K \left[ \frac{|K|}{\delta t} (\rho_K z_K - \rho_K^* z_K^*) + \sum_{\sigma=K|L} F_{\sigma,K} z_{\sigma} \right] \geq \frac{1}{2} \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} [\rho_K z_K^2 - \rho_K^* z_K^{*2}]. \tag{3.7}$$

*Proof.* We write:

$$\sum_{K \in \mathcal{M}} z_K \left[ \frac{|K|}{\delta t} (\rho_K z_K - \rho_K^* z_K^*) + \sum_{\sigma=K|L} F_{\sigma,K} z_\sigma \right] = T_1 + T_2$$

where  $T_1$  and  $T_2$  read:

$$T_1 = \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} z_K (\rho_K z_K - \rho_K^* z_K^*), \quad T_2 = \sum_{K \in \mathcal{M}} z_K \left[ \sum_{\sigma=K|L} F_{\sigma,K} z_\sigma \right].$$

The first term reads:

$$T_1 = \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} [z_K^2 (\rho_K - \rho_K^*) + \rho_K^* z_K (z_K - z_K^*)].$$

Developing the last term by the identity  $a(a-b) = \frac{1}{2}(a^2 + (a-b)^2 - b^2)$ , we get:

$$T_1 = \underbrace{\sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} z_K^2 (\rho_K - \rho_K^*)}_{T_{1,1}} + \underbrace{\frac{1}{2} \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} \rho_K^* (z_K^2 - z_K^{*2})}_{T_{1,2}} + \underbrace{\frac{1}{2} \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} \rho_K^* (z_K - z_K^*)^2}_{T_{1,3}}.$$

The last term, namely  $T_{1,3}$ , is always non-negative and can be seen as a dissipation associated to the backward time discretization of equation (3.7). We now turn to  $T_2$ :

$$T_2 = \underbrace{\sum_{K \in \mathcal{M}} z_K^2 \left[ \sum_{\sigma=K|L} F_{\sigma,K} \right]}_{T_{2,1}} + \underbrace{\sum_{K \in \mathcal{M}} z_K \left[ \sum_{\sigma=K|L} F_{\sigma,K} (z_\sigma - z_K) \right]}_{T_{2,2}}.$$

The first term, namely  $T_{2,1}$ , will cancel with  $T_{1,1}$  by equation (3.6). The second term reads, developing as previously the quantity  $z_K (z_\sigma - z_K)$ :

$$T_{2,2} = -\frac{1}{2} \sum_{K \in \mathcal{M}} z_K^2 \left[ \sum_{\sigma=K|L} F_{\sigma,K} \right] - \underbrace{\frac{1}{2} \sum_{K \in \mathcal{M}} \left[ \sum_{\sigma=K|L} F_{\sigma,K} [(z_\sigma - z_K)^2 - z_\sigma^2] \right]}_{T_{2,3}}.$$

Reordering the sum in the last term, we have, as  $F_{\sigma,K} = -F_{\sigma,L}$ :

$$T_{2,3} = \frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\text{int}} (\sigma=K|L)} F_{\sigma,K} [(z_\sigma - z_K)^2 - (z_\sigma - z_L)^2].$$

This expression can easily be seen to vanish with the centered choice. With the upwind choice, supposing without loss of generality that we have chosen for the edge  $\sigma = K|L$  the orientation such that  $F_{\sigma,K} \geq 0$ , we get, as  $z_\sigma = z_K$ :

$$T_{2,3} = -\frac{1}{2} \sum_{\sigma \in \mathcal{E}_{\text{int}} (\sigma=K|L)} F_{\sigma,K} (z_K - z_L)^2 \leq 0.$$

We thus have, by equation (3.6):

$$T_{2,2} \geq -\frac{1}{2} \sum_{K \in \mathcal{M}} z_K^2 \left[ \sum_{\sigma=K|L} F_{\sigma,K} \right] = \frac{1}{2} \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} z_K^2 (\rho_K - \rho_K^*)$$

and thus:

$$T_1 + T_2 \geq \frac{1}{2} \sum_{K \in \mathcal{M}} \frac{|K|}{\delta t} \left[ z_K^2 (\rho_K - \rho_K^*) + \rho_K^* (z_K^2 - z_K^{*2}) \right]$$

which concludes the proof. □

**Remark 3.2.** Equation (3.6) can be seen as a discrete mass balance, with  $F_{\sigma,K}$  standing for the mass flux across the edge  $\sigma$ , and the right hand side of (3.7) may be derived by the multiplication by  $z_K$  and summation over the control volumes of the transport terms in a discrete balance equation for the quantity  $\rho z$ , reading:

$$\forall K \in \mathcal{M}, \quad \frac{|K|}{\delta t} (\rho_K z_K - \rho_K^* z_K^*) + \sum_{\sigma=K|L} F_{\sigma,K} z_\sigma + \dots [\text{possible diffusion terms}] \dots = 0.$$

In this context, the relation (3.6) is known to be exactly the compatibility condition which ensures a discrete maximum principle for the solution  $z$  of this transport equation, provided that the upwind choice (or any monotone choice) is made for the expression of  $z_\sigma$  [27]. We proved here that the same compatibility condition ensures a  $L^2$  stability for  $\rho^{1/2} z$ .

### 3.3. Space discretization of the density prediction and the momentum balance equation

The main difficulty in the discretization of the momentum balance equation is to build a discrete convection operator which enjoys the stability property (1.5)-(i). To this purpose, we derive for this term a finite volume discretization which satisfies the assumptions of Theorem 3.1.

The natural space discretization for the density is the same as for the pressure, *i.e.* piecewise constant functions over each element. This legitimates a standard mass lumping technique for the time derivative term, since no additional accuracy seems to be expected from a more complex numerical integration. Note that, for the Crouzeix-Raviart element in two dimensions, the mass matrix is genuinely diagonal. Let the quantity  $|D_\sigma|$  be defined as follows:

$$|D_\sigma| \stackrel{\text{def}}{=} \int_\Omega \varphi_\sigma \, dx > 0. \tag{3.8}$$

For any  $\sigma \in \mathcal{E}$  and any control volume  $K$  adjacent to  $\sigma$ , let  $D_{K,\sigma}$  be the cone of basis  $\sigma$  and having the mass center of  $K$  as opposite vertex. The volume  $D_{K,\sigma}$  is referred to as the half-diamond cell associated to  $\sigma$  and  $K$  (see Fig. 1) and the measure of  $D_{K,\sigma}$  is denoted by  $|D_{K,\sigma}|$ . For  $\sigma \in \mathcal{E}$ , we define the diamond cell  $D_\sigma$  associated to  $\sigma$  by  $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$  if  $\sigma \in \mathcal{E}_{\text{int}}$ ,  $\sigma = K|L$ , and, if  $\sigma \in \mathcal{E}_{\text{ext}}$ , by  $D_\sigma = D_{K,\sigma}$  where  $K$  is the only control volume adjacent to  $\sigma$ . For the Crouzeix-Raviart element,  $|D_\sigma|$  can be identified to the measure of the diamond cell  $D_\sigma$  associated to  $\sigma$ . The same property holds for the Rannacher-Turek element in the case of rectangles ( $d = 2$ ) or cuboids ( $d = 3$ ), which are the only cases considered here, even though extensions to non-perpendicular grids are probably possible.

The discretization of the term  $\rho^n u^n$  thus leads, in the equations associated to the velocity on an internal edge  $\sigma$ , to an expression of the form  $\rho_\sigma^n u_\sigma^n$ , where  $\rho_\sigma^n$  results from an average of the values taken by the density in the two elements adjacent to  $\sigma$ , weighted by the measure of the half-diamonds:

$$\forall \sigma \in \mathcal{E}_{\text{int}}, \quad |D_\sigma| \rho_\sigma^n = |D_{K,\sigma}| \rho_K^n + |D_{L,\sigma}| \rho_L^n. \tag{3.9}$$

This definition naturally extends to any external edge  $\sigma$  by  $\rho_\sigma^n = \rho_K^n$ , where  $K$  is the control volume which is adjacent to  $\sigma$ .

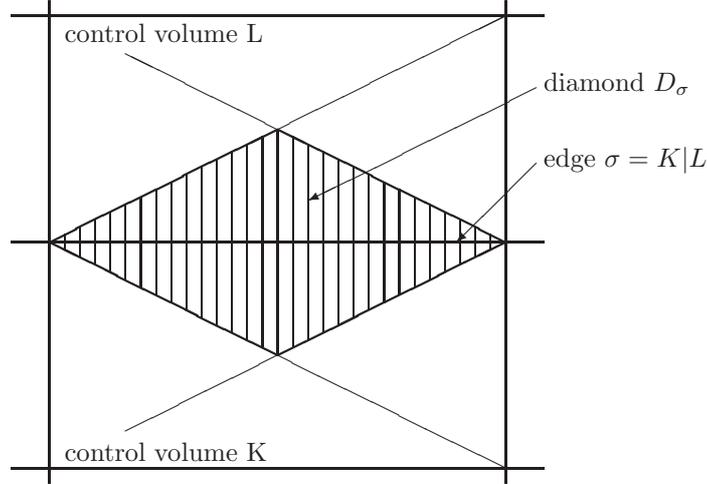


FIGURE 1. Dual finite volume mesh: the so-called “diamond cells”.

In order to satisfy the compatibility condition which was introduced in the previous section, a prediction of the density is first performed, by a finite volume discretization of the mass balance equation, taking the diamond cells as control volumes:

$$\frac{|D_\sigma|}{\delta t} (\tilde{\rho}_\sigma^{n+1} - \rho_\sigma^n) + \sum_{\varepsilon \in \mathcal{E}(D_\sigma)} F_{\varepsilon, \sigma}^{n+1} = 0, \quad \forall \sigma \in \mathcal{E} \quad (3.10)$$

where  $\mathcal{E}(D_\sigma)$  is the set of the edges of  $D_\sigma$  and  $F_{\varepsilon, \sigma}$  stands for the mass flux across  $\varepsilon$  outward  $D_\sigma$ . This latter quantity is expressed as follows:

$$F_\varepsilon^{n+1} = |\varepsilon| u_\varepsilon^n \cdot n_{\varepsilon, \sigma} \tilde{\rho}_\varepsilon^{n+1}$$

where  $|\varepsilon|$  is the measure of  $\varepsilon$ ,  $n_{\varepsilon, \sigma}$  is the normal to  $\varepsilon$  outward  $D_\sigma$ , the velocity  $u_\varepsilon^n$  is obtained by interpolation of  $u^n$  at the center of  $\varepsilon$  (using the standard finite element expansion) and  $\tilde{\rho}_\varepsilon^{n+1}$  is the density at the edge, calculated by the standard upwinding technique (*i.e.* either  $\tilde{\rho}_\sigma^{n+1}$  if  $u_\varepsilon^n \cdot n_{\varepsilon, \sigma} \geq 0$  or  $\tilde{\rho}_{\sigma'}^{n+1}$  otherwise, with  $\sigma'$  such that  $\varepsilon$  separates  $D_\sigma$  and  $D_{\sigma'}$ , which we denote by  $\varepsilon = D_\sigma | D_{\sigma'}$ ).

The discretization of the convection terms of the momentum balance equation is built from relation (3.10), according to the structure which is necessary to apply Theorem 3.1. This yields the following discrete momentum balance equation:

$$\begin{aligned} \frac{|D_\sigma|}{\delta t} (\tilde{\rho}_\sigma^{n+1} \tilde{u}_{\sigma, i}^{n+1} - \rho_\sigma^n u_{\sigma, i}^n) + \sum_{\substack{\varepsilon \in \mathcal{E}(D_\sigma) \\ (\varepsilon = D_\sigma | D_{\sigma'})}} \frac{1}{2} F_{\varepsilon, \sigma}^{n+1} (\tilde{u}_{\sigma, i}^{n+1} + \tilde{u}_{\sigma', i}^{n+1}) \\ + \int_{\Omega, h} \tau(\tilde{u}^{n+1}) : \nabla \varphi_\sigma^{(i)} dx - \int_{\Omega, h} \tilde{p}^{n+1} \nabla \cdot \varphi_\sigma^{(i)} dx = \int_{\Omega} f_v^{n+1} \cdot \varphi_\sigma^{(i)}, \quad \forall \sigma \in \mathcal{E}_{\text{int}}, \text{ for } 1 \leq i \leq d \end{aligned} \quad (3.11)$$

where  $\varphi_\sigma^{(i)}$  stands for the vector shape function associated to  $v_{\sigma, i}$ , which reads  $\varphi_\sigma^{(i)} = \varphi_\sigma e_i$  with  $e_i$  the  $i^{\text{th}}$  vector of the canonical basis of  $\mathbb{R}^d$  and  $\varphi_\sigma$  the scalar shape function, and where the notation  $\int_{\Omega, h}$  means  $\sum_{K \in \mathcal{M}} \int_K$ .

Note that, for Crouzeix-Raviart elements, a combined finite volume/finite element method, similar to the technique employed here for the discretization of the momentum balance, has already been analysed for a transient non-linear convection-diffusion equation by Feistauer and co-workers [1, 11, 16].

### 3.4. Space discretization of the projection step

The fully discrete projection step of the proposed algorithm reads:

$$\left\{ \begin{aligned} |D_\sigma| \frac{\tilde{\rho}_\sigma^{n+1}}{\delta t} (\bar{u}_{\sigma,i}^{n+1} - \tilde{u}_{\sigma,i}^{n+1}) - \int_{\Omega,h} (p^{n+1} - \tilde{p}^{n+1}) \nabla \cdot \varphi_\sigma^{(i)} dx &= 0, & \forall \sigma \in \mathcal{E}_{\text{int}}, \text{ for } 1 \leq i \leq d \\ \frac{|K|}{\delta t} (\varrho(p_K^{n+1}) - \rho_K^n) + \sum_{\sigma=K|L} (v_{\sigma,K}^+)^{n+1} \varrho(p_K^{n+1}) - (v_{\sigma,K}^-)^{n+1} \varrho(p_L^{n+1}) &= 0, & \forall K \in \mathcal{M} \end{aligned} \right. \quad (3.12)$$

where  $(v_{\sigma,K}^+)^{n+1}$  and  $(v_{\sigma,K}^-)^{n+1}$  stand respectively for  $\max(v_{\sigma,K}^{n+1}, 0)$  and  $-\min(v_{\sigma,K}^{n+1}, 0)$  with  $v_{\sigma,K}^{n+1} = |\sigma| \bar{u}_\sigma^{n+1} \cdot n_{KL}$ . The first (vector) equation may be seen as the finite element discretization of the first relation of the projection step (3.4), with the same lumping of the mass matrix for the Rannacher-Turek element as in the prediction step. As the pressure is piecewise constant, the finite element discretization of the second relation of (3.4), *i.e.* the mass balance, is equivalent to a finite volume formulation, in which we introduce the standard first-order upwinding. Exploiting the expression of the velocity and pressure shape functions, the first set of relations of this system can be alternatively written as follows:

$$|D_\sigma| \frac{\tilde{\rho}_\sigma^{n+1}}{\delta t} (\bar{u}_\sigma^{n+1} - \tilde{u}_\sigma^{n+1}) + |\sigma| [(p_L^{n+1} - \tilde{p}_L^{n+1}) - (p_K^{n+1} - \tilde{p}_K^{n+1})] n_{KL} = 0, \quad \forall \sigma \in \mathcal{E}_{\text{int}}, \sigma = K|L \quad (3.13)$$

or, in an algebraic setting:

$$\frac{1}{\delta t} M_{\tilde{\rho}^{n+1}} (\bar{u}^{n+1} - \tilde{u}^{n+1}) + B^t (p^{n+1} - \tilde{p}^{n+1}) = 0. \quad (3.14)$$

In this relation,  $M_w$  stands for the diagonal mass matrix weighted by  $(w_\sigma)_{\sigma \in \mathcal{E}_{\text{int}}}$  (so, for  $1 \leq i \leq d$  and  $\sigma \in \mathcal{E}_{\text{int}}$ , the corresponding entry on the diagonal of  $M_{\tilde{\rho}^{n+1}}$  reads  $(M_{\tilde{\rho}^{n+1}})_{\sigma,i} = |D_\sigma| \tilde{\rho}_\sigma^{n+1}$ ),  $B^t$  is the matrix of  $\mathbb{R}^{(dN) \times M}$ , where  $N$  is the number of internal edges (*i.e.*  $N = \text{card}(\mathcal{E}_{\text{int}})$ ) and  $M$  is the number of control volumes in the mesh (*i.e.*  $M = \text{card}(\mathcal{M})$ ), associated to the gradient operator; consequently, the matrix  $B$  is associated to the opposite of the discrete divergence operator. Throughout this section, we use the same notation for the discrete function (defined as usual in the finite element context by its expansion using the shape functions) and for the vector gathering the degrees of freedom; so, in relation (3.14),  $\bar{u}$  (respectively  $\tilde{u}$ ) stands for the vector of  $\mathbb{R}^{dN}$  components  $\bar{u}_{\sigma,i}$  (respectively  $\tilde{u}_{\sigma,i}$ ),  $\sigma \in \mathcal{E}_{\text{int}}$ ,  $1 \leq i \leq d$  and  $p$  (respectively  $\tilde{p}$ ) stands for the vector of  $\mathbb{R}^M$  of components  $p_K$  (respectively  $\tilde{p}_K$ ),  $K \in \mathcal{M}$ . Both forms (3.13) and (3.14) are used hereafter.

We have the following existence result.

**Proposition 3.3.** *Let the equation of state  $\varrho(\cdot)$  be defined and increasing over  $[0, +\infty)$ , and be such that  $\varrho(0) = 0$ ,  $\lim_{z \rightarrow +\infty} \varrho(z) = +\infty$  and such that there exists an elastic potential function  $P(\cdot)$  (*i.e.* a function satisfying (1.3)) such that the function  $z \mapsto z P(z)$  is bounded from below in  $(0, +\infty)$ , once continuously differentiable and strictly convex. Let us suppose that  $\tilde{\rho}_\sigma^{n+1} > 0, \forall \sigma \in \mathcal{E}_{\text{int}}$ . Then the nonlinear system (3.12) admits at least one solution and any possible solution is such that  $p_K > 0, \forall K \in \mathcal{M}$  (and thus  $\rho_K > 0, \forall K \in \mathcal{M}$ ).*

*Proof.* The theory of Section 2 applies, with:

$$\|\bar{u}^{n+1}\|_*^2 = \sum_{\sigma \in \mathcal{E}_{\text{int}}} \frac{|D_\sigma|}{\delta t} \tilde{\rho}_\sigma^{n+1} |\bar{u}_\sigma^{n+1}|^2.$$

This yields both the existence of a solution and the positivity of the pressure. □

In view of this result and of the form of the discrete density prediction (3.10), the property  $\tilde{\rho}^{n+1} > 0$  is satisfied by induction at any time step of the computation (provided, of course, that the initial density is positive everywhere).

We finish this section by some remarks concerning the projection step at hand.

**Lemma 3.4.** *The following identity holds for each discrete pressure  $q \in L_h$ :*

$$\forall K \in \mathcal{M}, \quad (\text{B M}_{\tilde{\rho}^{n+1}}^{-1} \text{B}^t q)_K = \sum_{\sigma=K|L} \frac{1}{\tilde{\rho}_\sigma^{n+1}} \frac{|\sigma|^2}{|D_\sigma|} (q_K - q_L).$$

*Proof.* Let  $q \in L_h$  be given. By relation (3.13), we have:

$$(\text{B}^t q)_{\sigma,i} = |\sigma| (q_L - q_K) n_{KL} \cdot e_i.$$

Let  $1_K \in L_h$  be the characteristic function of  $K$ . Denoting by  $(\cdot, \cdot)$  the standard Euclidean inner product, by the previous relation and the definition of the lumped velocity mass matrix, we obtain:

$$\begin{aligned} (\text{B M}_{\tilde{\rho}^{n+1}}^{-1} \text{B}^t q, 1_K) &= (\text{M}_{\tilde{\rho}^{n+1}}^{-1} \text{B}^t q, \text{B}^t 1_K) = \sum_{\sigma \in \mathcal{E}_{\text{int}}} \sum_{i=1}^d \frac{1}{\tilde{\rho}_\sigma^{n+1} |D_\sigma|} (\text{B}^t q)_{\sigma,i} (\text{B}^t 1_K)_{\sigma,i} \\ &= \sum_{\sigma=K|L} \sum_{i=1}^d \frac{1}{\tilde{\rho}_\sigma^{n+1} |D_\sigma|} [|\sigma| (q_L - q_K) n_{KL} \cdot e_i] [-|\sigma| n_{KL} \cdot e_i] \end{aligned}$$

which, remarking that  $\sum_{i=1}^d (n_{KL} \cdot e_i)^2 = 1$ , yields the result.  $\square$

**Remark 3.5** (on spurious pressure boundary conditions). In the context of projection methods for incompressible flow, it is known that spurious boundary conditions are to be imposed to the pressure in the projection step, in order to make the definition of this step of the algorithm complete. These boundary conditions are explicit when the process to derive the projection step is first to pose the elliptic problem for the pressure at the time semi-discrete level and then discretize it in space; for instance, with a constant density equal to one and prescribed velocity boundary conditions on  $\partial\Omega$ , the semi-discrete projection step would take the form:

$$\begin{cases} -\Delta (p^{n+1} - \tilde{p}^{n+1}) = -\frac{1}{\delta t} \nabla \cdot \tilde{u}^{n+1} & \text{in } \Omega \\ \nabla (p^{n+1} - \tilde{p}^{n+1}) \cdot n = 0 & \text{on } \partial\Omega. \end{cases}$$

When the elliptic problem for the pressure is built at the algebraic level, the boundary conditions for the pressure are somehow hidden in the discrete operator  $\text{B M}^{-1} \text{B}^t$ . Lemma 3.4 shows that this matrix takes the form of a finite-volume Laplace discrete operator, with homogeneous Neumann boundary conditions, *i.e.* the same boundary conditions as in the time semi-discrete problem above stated.

**Remark 3.6** (on the non-consistency of the discretization at hand for the Darcy problem). Considering the semi-discrete problem (3.4), in the case of a constant density equal to one, one may expect to recover a consistent discretization of a Poisson problem with homogeneous Neumann boundary conditions, as stated above. The following example shows that this route is misleading. Let us take for the mesh a uniform square grid of step  $h$ . The coefficient  $|\sigma|^2/|D_\sigma|$  can be easily evaluated, and we obtain:

$$(\text{B M}^{-1} \text{B}^t q)_K = d \sum_{\sigma=K|L} \frac{|\sigma|}{h} (q_K - q_L)$$

that is the usual finite volume Laplace operator, but multiplied by the space dimension  $d$ . This result is of course consistent with (and gives some insight in) the well-known non-consistency of the Rannacher-Turek element for the Darcy problem; similar examples could also be given for simplicial grids, with the Crouzeix-Raviart element.

### 3.5. Renormalization steps

The pressure renormalization (Step 2 of the algorithm) reads, in an algebraic setting:

$$\mathbf{B} \mathbf{M}_{\tilde{\rho}^{n+1}}^{-1} \mathbf{B}^t \tilde{p}^{n+1} = \mathbf{B} \mathbf{M}_{\sqrt{\tilde{\rho}^{n+1} \tilde{\rho}^n}}^{-1} \mathbf{B}^t p^n. \tag{3.15}$$

Note that, at the first time step, the quantity  $\tilde{\rho}^0$  must be defined; it can for instance be computed from the initial density (defined on the control volumes of  $\mathcal{M}$ ) by equation (3.9). In view of the expression of these operators provided by Lemma 3.4, this relation equivalently reads:

$$\sum_{\sigma=K|L} \frac{1}{\tilde{\rho}_\sigma^{n+1}} \frac{|\sigma|^2}{|D_\sigma|} (\tilde{p}_K^{n+1} - \tilde{p}_L^{n+1}) = \sum_{\sigma=K|L} \frac{1}{\sqrt{\tilde{\rho}_\sigma^{n+1} \tilde{\rho}_\sigma^n}} \frac{|\sigma|^2}{|D_\sigma|} (p_K^n - p_L^n), \quad \forall K \in \mathcal{M}. \tag{3.16}$$

As  $\mathbf{B}^t$  and  $\mathbf{B}$  stands for respectively the discrete gradient and (opposite of the) divergence operator, this system can be seen as a discretization of the semi-discrete expression of Step 2; note however, as shown in Remark 3.6, that this discretization is non-consistent.

The velocity renormalization (Step 5 of the algorithm) simply reads:

$$\forall \sigma \in \mathcal{E}_{\text{int}}, \quad \sqrt{\rho_\sigma^{n+1}} u_\sigma^{n+1} = \sqrt{\tilde{\rho}_\sigma^{n+1}} \bar{u}_\sigma^{n+1} \quad \text{or} \quad \mathbf{M}_{\sqrt{\rho^{n+1}}} u^{n+1} = \mathbf{M}_{\sqrt{\tilde{\rho}^{n+1}}} \bar{u}^{n+1}. \tag{3.17}$$

### 3.6. An overview of the algorithm

To sum up, the algorithm considered in this section is the following one:

- (1) *Prediction of the density.* The density on the edges at  $t^n$ ,  $(\rho_\sigma^n)_{\sigma \in \mathcal{E}}$ , being given by (3.9), compute  $(\tilde{\rho}^{n+1})_{\sigma \in \mathcal{E}}$  by the upwind finite volume discretization of the mass balance over the diamond cells (3.10).
- (2) *Renormalization of the pressure.* Compute a renormalized pressure  $(\tilde{p}_K^{n+1})_{K \in \mathcal{M}}$  by equation (3.16).
- (3) *Prediction of the velocity.* Compute  $(\tilde{u}_\sigma^{n+1})_{\sigma \in \mathcal{E}_{\text{int}}}$  by equation (3.11), obtained by a finite volume discretization of the transport terms over the diamond cells and a finite element discretization of the other terms.
- (4) *Projection step.* Compute  $(\bar{u}_\sigma^{n+1})_{\sigma \in \mathcal{E}_{\text{int}}}$  and  $(p_K^{n+1})_{K \in \mathcal{M}}$  from equation (3.12), obtained by a finite element discretization of the velocity correction equation and an upwind finite volume discretization of the mass balance (over the elements  $K \in \mathcal{M}$ ).
- (5) *Renormalization of the velocity.* Compute  $(u_\sigma^{n+1})_{\sigma \in \mathcal{E}_{\text{int}}}$  from equation (3.17).

The existence of a solution to Step 4 is proven above; the other problems are linear, and their well-posedness follows from standard coercivity arguments, using the fact that the discrete densities (*i.e.*  $\rho^n$  and  $\tilde{\rho}^{n+1}$ ) are positive, provided that this property is satisfied by the initial condition.

### 3.7. Stability analysis

In this section, we use the following discrete norm and semi-norm:

$$\begin{aligned} \forall v \in W_h, \quad & \|v\|_{h,\tilde{\rho}}^2 = \sum_{\sigma \in \mathcal{E}_{\text{int}}} |D_\sigma| \tilde{\rho}_\sigma |v_\sigma|^2 \\ \forall q \in L_h, \quad & |q|_{h,\tilde{\rho}}^2 = \sum_{\sigma \in \mathcal{E}_{\text{int}}, \sigma=K|L} \frac{1}{\tilde{\rho}_\sigma} \frac{|\sigma|^2}{|D_\sigma|} (q_K - q_L)^2 \end{aligned} \tag{3.18}$$

where  $\tilde{\rho} = (\tilde{\rho}_\sigma)_{\sigma \in \mathcal{E}_{\text{int}}}$  is a family of positive real numbers. The function  $\|\cdot\|_{h,\tilde{\rho}}^2$  defines a norm over  $W_h$ , and  $|\cdot|_{h,\tilde{\rho}}$  can be seen as a weighted version (with mesh dependent weights) of the  $H^1$  semi-norm classical

in the finite volume context [14]. The links between this latter semi-norm and the problem at hand are clarified in the following lemma, which is a straightforward consequence of Lemma 3.4.

**Lemma 3.7.** *The following identity holds for each discrete pressure  $q \in L_h$ :*

$$(\mathbf{B} \mathbf{M}_{\bar{\rho}}^{-1} \mathbf{B}^t q, q) = |q|_{h, \bar{\rho}}^2.$$

We are now in position to state the stability of the scheme under consideration.

**Theorem 3.8** (stability of the scheme). *Let the equation of state  $\varrho(\cdot)$  be defined and increasing over  $[0, +\infty)$ , and be such that  $\varrho(0) = 0$ ,  $\lim_{z \rightarrow +\infty} \varrho(z) = +\infty$  and that there exists an elastic potential function  $P(\cdot)$  (i.e. a function satisfying (1.3)) such that the function  $f: z \mapsto z P(z)$  is bounded from below in  $(0, +\infty)$ , once continuously differentiable and strictly convex. Let  $(\tilde{u}^n)_{0 \leq n \leq N}$ ,  $(u^n)_{0 \leq n \leq N}$ ,  $(p^n)_{0 \leq n \leq N}$  and  $(\rho^n)_{0 \leq n \leq N}$  be the solution to the considered scheme, with a zero forcing term. Then the following bound holds for  $0 \leq n < N$ :*

$$\begin{aligned} \frac{1}{2} \|u^{n+1}\|_{h, \rho^{n+1}}^2 + \int_{\Omega} \rho^{n+1} P(\rho^{n+1}) \, dx + \delta t \sum_{k=1}^{n+1} \int_{\Omega, h} \nabla \tilde{u}^k : \tau(\tilde{u}^k) \, dx + \frac{\delta t^2}{2} |p^{n+1}|_{h, \bar{\rho}^{n+1}}^2 \\ \leq \frac{1}{2} \|u^0\|_{h, \rho^0}^2 + \int_{\Omega} \rho^0 P(\rho^0) \, dx + \frac{\delta t^2}{2} |p^0|_{h, \bar{\rho}^0}^2. \end{aligned} \quad (3.19)$$

*Proof.* Multiplying each equation of the Step 3 of the scheme (3.11) by the corresponding unknown (i.e. the corresponding component of the velocity  $\tilde{u}^{n+1}$  on the corresponding edge  $\sigma$ ) and summing over the edges and the components yields, by virtue of the stability of the discrete advection operator (Thm. 3.1):

$$\frac{1}{2\delta t} \|\tilde{u}^{n+1}\|_{h, \bar{\rho}^{n+1}}^2 - \frac{1}{2\delta t} \|u^n\|_{h, \rho^n}^2 + \int_{\Omega, h} \tau(\tilde{u}^{n+1}) : \nabla \tilde{u}^{n+1} \, dx - \int_{\Omega, h} \tilde{p}^{n+1} \nabla \cdot \tilde{u}^{n+1} \, dx \leq 0. \quad (3.20)$$

On the other hand, reordering equation (3.14) and multiplying by  $\mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2}$  (recall that  $\mathbf{M}_{\bar{\rho}^{n+1}}$  is diagonal), we obtain:

$$\frac{1}{\delta t} \mathbf{M}_{\bar{\rho}^{n+1}}^{1/2} \bar{u}^{n+1} + \mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2} \mathbf{B}^t p^{n+1} = \frac{1}{\delta t} \mathbf{M}_{\bar{\rho}^{n+1}}^{1/2} \tilde{u}^{n+1} + \mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2} \mathbf{B}^t \tilde{p}^{n+1}.$$

Squaring this relation gives:

$$\begin{aligned} \left( \frac{1}{\delta t} \mathbf{M}_{\bar{\rho}^{n+1}}^{1/2} \bar{u}^{n+1} + \mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2} \mathbf{B}^t p^{n+1}, \frac{1}{\delta t} \mathbf{M}_{\bar{\rho}^{n+1}}^{1/2} \bar{u}^{n+1} + \mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2} \mathbf{B}^t p^{n+1} \right) = \\ \left( \frac{1}{\delta t} \mathbf{M}_{\bar{\rho}^{n+1}}^{1/2} \tilde{u}^{n+1} + \mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2} \mathbf{B}^t \tilde{p}^{n+1}, \frac{1}{\delta t} \mathbf{M}_{\bar{\rho}^{n+1}}^{1/2} \tilde{u}^{n+1} + \mathbf{M}_{\bar{\rho}^{n+1}}^{-1/2} \mathbf{B}^t \tilde{p}^{n+1} \right) \end{aligned}$$

which reads:

$$\begin{aligned} \frac{1}{\delta t^2} (\mathbf{M}_{\bar{\rho}^{n+1}} \bar{u}^{n+1}, \bar{u}^{n+1}) + (\mathbf{M}_{\bar{\rho}^{n+1}}^{-1} \mathbf{B}^t p^{n+1}, \mathbf{B}^t p^{n+1}) + \frac{2}{\delta t} (\bar{u}^{n+1}, \mathbf{B}^t p^{n+1}) = \\ \frac{1}{\delta t^2} (\mathbf{M}_{\bar{\rho}^{n+1}} \tilde{u}^{n+1}, \tilde{u}^{n+1}) + (\mathbf{M}_{\bar{\rho}^{n+1}}^{-1} \mathbf{B}^t \tilde{p}^{n+1}, \mathbf{B}^t \tilde{p}^{n+1}) + \frac{2}{\delta t} (\tilde{u}^{n+1}, \mathbf{B}^t \tilde{p}^{n+1}). \end{aligned}$$

Multiplying by  $\delta t/2$ , remarking that,  $\forall v \in W_h$ ,  $(\mathbf{M}_{\bar{\rho}^{n+1}} v, v) = \|v\|_{h, \bar{\rho}^{n+1}}^2$  and that, thanks to Lemma 3.7,  $\forall q \in L_h$ ,  $(\mathbf{M}_{\bar{\rho}^{n+1}}^{-1} \mathbf{B}^t q, \mathbf{B}^t q) = (\mathbf{B} \mathbf{M}_{\bar{\rho}^{n+1}}^{-1} \mathbf{B}^t q, q) = |q|_{h, \bar{\rho}^{n+1}}^2$ , we get:

$$\begin{aligned} \frac{1}{2\delta t} \|\bar{u}^{n+1}\|_{h, \bar{\rho}^{n+1}}^2 + \frac{\delta t}{2} |p^{n+1}|_{h, \bar{\rho}^{n+1}}^2 + (\bar{u}^{n+1}, \mathbf{B}^t p^{n+1}) \\ - \frac{1}{2\delta t} \|\tilde{u}^{n+1}\|_{h, \bar{\rho}^{n+1}}^2 - \frac{\delta t}{2} |\tilde{p}^{n+1}|_{h, \bar{\rho}^{n+1}}^2 - (\tilde{u}^{n+1}, \mathbf{B}^t \tilde{p}^{n+1}) = 0. \end{aligned} \quad (3.21)$$

The quantity  $-(\tilde{u}^{n+1}, B^t \tilde{p}^{n+1})$  is nothing more than the opposite of the term  $\int_{\Omega, h} \tilde{p}^{n+1} \nabla \cdot \tilde{u}^{n+1} dx$  appearing in (3.20), so, when summing (3.20) and (3.21), these terms cancel, leading to:

$$\begin{aligned} \frac{1}{2\delta t} \|\bar{u}^{n+1}\|_{h, \tilde{\rho}^{n+1}}^2 - \frac{1}{2\delta t} \|u^n\|_{h, \rho^n}^2 + \int_{\Omega, h} \tau(\tilde{u}^{n+1}) : \nabla \tilde{u}^{n+1} dx \\ + \frac{\delta t}{2} |p^{n+1}|_{h, \tilde{\rho}^{n+1}}^2 - \frac{\delta t}{2} |\tilde{p}^{n+1}|_{h, \tilde{\rho}^{n+1}}^2 + (\bar{u}^{n+1}, B^t p^{n+1}) \leq 0. \end{aligned}$$

Finally,  $(\bar{u}^{n+1}, B^t p^{n+1})$  is precisely the pressure work which can be bounded by the time derivative of the elastic potential, as stated in Theorem 2.1:

$$\begin{aligned} \frac{1}{2\delta t} \|\bar{u}^{n+1}\|_{h, \tilde{\rho}^{n+1}}^2 + \int_{\Omega, h} \tau(\tilde{u}^{n+1}) : \nabla \tilde{u}^{n+1} dx + \frac{\delta t}{2} |p^{n+1}|_{h, \tilde{\rho}^{n+1}}^2 - \frac{\delta t}{2} |\tilde{p}^{n+1}|_{h, \tilde{\rho}^{n+1}}^2 \\ + \frac{1}{\delta t} \int_{\Omega} \rho^{n+1} P(\rho^{n+1}) dx \leq \frac{1}{2\delta t} \|u^n\|_{h, \rho^n}^2 + \frac{1}{\delta t} \int_{\Omega} \rho^n P(\rho^n) dx. \end{aligned} \tag{3.22}$$

The proof then ends by using the renormalization steps (Steps 2 and 5 of the algorithm). Step 2 reads in an algebraic setting:

$$B M_{\tilde{\rho}^{n+1}}^{-1} B^t \tilde{p}^{n+1} = B M_{\tilde{\rho}^{n+1}}^{-1/2} M_{\tilde{\rho}^n}^{-1/2} B^t p^n.$$

Multiplying by  $\tilde{p}^{n+1}$ , we obtain:

$$\left( M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1}, M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1} \right) = \left( M_{\tilde{\rho}^n}^{-1/2} B^t p^n, M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1} \right)$$

and thus, by Cauchy-Schwarz inequality:

$$\left( M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1}, M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1} \right) \leq \left( M_{\tilde{\rho}^n}^{-1/2} B^t p^n, M_{\tilde{\rho}^n}^{-1/2} B^t p^n \right)^{1/2} \left( M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1}, M_{\tilde{\rho}^{n+1}}^{-1/2} B^t \tilde{p}^{n+1} \right)^{1/2}.$$

This relation yields  $|\tilde{p}^{n+1}|_{h, \tilde{\rho}^{n+1}}^2 \leq |p^n|_{h, \tilde{\rho}^n}^2$ . In addition, Step 5 of the algorithm gives  $\|\bar{u}^{n+1}\|_{h, \tilde{\rho}^{n+1}}^2 = \|\bar{u}^{n+1}\|_{h, \tilde{\rho}^{n+1}}^2$ . Using these two relations in (3.22), we get:

$$\begin{aligned} \frac{1}{2\delta t} \|u^{n+1}\|_{h, \rho^{n+1}}^2 + \int_{\Omega, h} \tau(\tilde{u}^{n+1}) : \nabla \tilde{u}^{n+1} dx + \frac{\delta t}{2} |p^{n+1}|_{h, \tilde{\rho}^{n+1}}^2 + \frac{1}{\delta t} \int_{\Omega} \rho^{n+1} P(\rho^{n+1}) dx \\ \leq \frac{1}{2\delta t} \|u^n\|_{h, \rho^n}^2 + \frac{1}{\delta t} \int_{\Omega} \rho^n P(\rho^n) dx + \frac{\delta t}{2} |p^n|_{h, \tilde{\rho}^n}^2 \end{aligned}$$

and the estimate of Theorem 3.8 follows by summing over the time steps. □

**Remark 3.9** (entropy conservation). In this paper, we employ the terminology of the mathematical analysis of the compressible Navier-Stokes equations, as can be found in [15,28,32]. In this context, the function  $P(\cdot)$  is referred to as the elastic potential, and the total energy of the system is the sum of the kinetic energy and of the integral of  $\rho P(\rho)$  over the fluid domain. In the literature concerned with hyperbolic problems, the same quantity is referred to as the entropy of the system, and Theorem 3.8 states that the considered pressure correction scheme preserves the entropy.

**Remark 3.10** (on the upwinding of the mass balance discretization, the *inf-sup* stability of the discretization and the appearance of spurious pressure wiggles.). In the scheme considered in this section, the upwinding in the discretization of mass balance controls the onset of density oscillations. As long as the pressure and the density are linked by an increasing function, that is as long as the flow remains compressible with a reasonable equation of state, it is probably sufficient to prevent pressure oscillations. Besides, the fourth term of the left hand side

of (3.19), *i.e.* the term involving  $|p^{n+1}|_{h,\tilde{\rho}^{n+1}}^2$ , provides a control on the discrete  $H^1$ -like semi-norm of the pressure, at least for large time steps, and therefore also produces an additional pressure smearing. However, it comes up in the analysis as the composition of the discrete divergence with the discrete gradient; consequently, one will obtain such a smoothing effect only for *inf-sup* stable discretizations. Note also that, even for steady state problems, some authors recommend the use of stable approximation space pairs to avoid pressure wiggles [3,17].

**Remark 3.11** (on a different projection step). Some authors propose a different projection step [2,43], which reads in the time semi-discrete setting:

$$\begin{cases} \frac{\varrho(p^{n+1}) u^{n+1} - \tilde{\rho}^{n+1} \tilde{u}^{n+1}}{\delta t} + \nabla(p^{n+1} - \tilde{p}^{n+1}) = 0 \\ \frac{\varrho(p^{n+1}) - \rho^n}{\delta t} + \nabla \cdot (\varrho(p^{n+1}) u^{n+1}) = 0. \end{cases}$$

Considering this system, one may be tempted by the following line of thought: choosing  $q^{n+1} = \varrho(p^{n+1}) u^{n+1}$  as variable, taking the discrete divergence of the first equation and using the second one will cause the convection term of the mass balance to disappear from the discrete elliptic problem for the pressure, whatever the discretization of this term (and, in particular, the choice of the density at the edges) may be. Consequently, the equation for the pressure will be free of the non-linearities induced by the upwinding and the dependency of the convected density on the pressure, while one still may hope to obtain a positive upwind (with respect to the density) scheme. In fact, this last point is incorrect. To be valid, it would necessitate that, from any solution  $(q^{n+1}, p^{n+1})$ , one could be able to compute a velocity field  $u^{n+1}$  by dividing  $q^{n+1}$  by the density of the control volume located upstream with respect to  $u^{n+1}$ . Unfortunately, it is not always possible to obtain this upstream value; for instance, if for two neighbouring control volumes  $K$  and  $L$ ,  $\rho_K < 0$ ,  $\rho_L > 0$  and  $q^{n+1} \cdot n_{K|L} > 0$ , neither the choice of  $K$  nor  $L$  for the upstream control volume is valid. Consequently, with this discretization, we are no longer able to guarantee the positivity of  $\rho$  or the absence of oscillations. However, as explained above, if the density remains positive, we will have a smearing of pressure or density wiggles due to the fact that the discretization is *inf-sup* stable.

### 3.8. Implementation

The implementation of the first three steps (3.1)–(3.3) of the numerical scheme is standard, and we therefore only describe here in details the fourth step, that is the projection step. The precise algebraic formulation of the system (3.4) reads:

$$\begin{cases} \frac{1}{\delta t} M_{\tilde{\rho}^{n+1}} (\bar{u}^{n+1} - \tilde{u}^{n+1}) + B^t (p^{n+1} - \tilde{p}^{n+1}) = 0 \\ \frac{1}{\delta t} R (\varrho(p^{n+1}) - \rho^n) - B Q_{\rho^{n+1}}^{\text{up}} \bar{u}^{n+1} = 0 \end{cases} \quad (3.23)$$

where  $M_{\tilde{\rho}^{n+1}}$  and  $Q_{\rho^{n+1}}^{\text{up}}$  are two diagonal matrices; for the first one, we recall that the entry corresponding to an edge  $\sigma \in \mathcal{E}_{\text{int}}$ ,  $\sigma = K|L$  is computed by multiplying the measure of the diamond associated to  $\sigma$  by the predicted density (at the edge center)  $\tilde{\rho}_\sigma^{n+1}$ ; in the second one, the same entry is obtained by just taking the density at  $t^{n+1}$  in the element located upstream of  $\sigma$  with respect to  $\bar{u}^{n+1}$ , *i.e.* either  $\varrho(p_K^{n+1})$  or  $\varrho(p_L^{n+1})$ . Note that these definitions can be extended in a straightforward way to the boundary edges, if the velocity is not prescribed to zero on the boundary of the computational domain. The matrix  $R$  is diagonal and, for any  $K \in \mathcal{M}$ , its entry  $R_K$  is the measure of the element  $K$ . For the sake of simplicity, we suppose for the moment that the equation of state is linear:

$$\varrho(p^{n+1}) = \frac{\partial \varrho}{\partial p} p^{n+1}.$$

The elliptic problem for the pressure is obtained by multiplying the first relation of (3.23) by  $\mathbf{B} \mathbf{Q}_{\rho^{n+1}}^{\text{up}} (\mathbf{M}_{\bar{\rho}^{n+1}})^{-1}$  and using the second one. This equation reads:

$$\left[ \mathbf{L} + \frac{\partial \varrho}{\partial p} \frac{1}{\delta t^2} \mathbf{R} \right] p^{n+1} = \mathbf{L} \tilde{p}^{n+1} + \frac{1}{\delta t^2} \mathbf{R} \rho^n + \frac{1}{\delta t} \mathbf{B} \mathbf{Q}_{\rho^{n+1}}^{\text{up}} \tilde{u}^{n+1} \quad (3.24)$$

where  $\mathbf{L} = \mathbf{B} \mathbf{Q}_{\rho^{n+1}}^{\text{up}} (\mathbf{M}_{\bar{\rho}^{n+1}})^{-1} \mathbf{B}^t$  can be viewed, for the discretization at hand, as a finite volume discrete approximation of the Laplace operator with Neumann boundary conditions (when the velocity is prescribed at the boundary), weighted by a mesh-dependent coefficient and the densities ratio (see Rems. 3.5 and 3.6). We recall that, by a calculation similar to the proof of Lemma 3.4, this matrix can be evaluated directly in the ‘‘finite volume way’’, by the following relation, valid for each element  $K$ :

$$(\mathbf{L} p^{n+1})_K = \sum_{\sigma=K|L} \frac{\rho_{\text{up},\sigma}}{\tilde{\rho}_\sigma^{n+1}} \frac{|\sigma|^2}{|D_\sigma|} (p_K^{n+1} - p_L^{n+1})$$

where  $\rho_{\text{up},\sigma}$  stands for the upwind density associated to the edge  $\sigma$ . Provided that  $p^{n+1}$  is known, the first relation of (3.23) gives us the updated value of the velocity:

$$\bar{u}^{n+1} = \tilde{u}^{n+1} - \delta t (\mathbf{M}_{\bar{\rho}^{n+1}})^{-1} \mathbf{B}^t (p^{n+1} - \tilde{p}^{n+1}). \quad (3.25)$$

In order to preserve the positivity of the density, it is necessary to use the value of the density upwinded with respect to  $\bar{u}^{n+1}$  in the mass balance; therefore, equations (3.24) and (3.25) are not decoupled, in contrast with what happens in usual projection methods. We thus implement the following iterative algorithm:

Initialization:  $p_0^{n+1} = \tilde{p}^{n+1}$  and  $\bar{u}_0^{n+1} = \tilde{u}^{n+1}$ .

Step 4.1. Solve for  $p_{k+1/2}^{n+1}$ :

$$\left[ \mathbf{L} + \frac{\partial \varrho}{\partial p} \frac{1}{\delta t^2} \mathbf{R} \right] p_{k+1/2}^{n+1} = \mathbf{L} \tilde{p}^{n+1} + \frac{1}{\delta t^2} \mathbf{R} \rho^n + \frac{1}{\delta t} \mathbf{B} \mathbf{Q}_{\rho^{n+1}}^{\text{up}} \tilde{u}^{n+1}$$

where the density in  $\mathbf{L}$  and  $\mathbf{Q}_{\rho^{n+1}}^{\text{up}}$  is evaluated at  $p_k^{n+1}$  and the upwinding in  $\mathbf{Q}_{\rho^{n+1}}^{\text{up}}$  is performed with respect to  $\bar{u}_k^{n+1}$ .

Step 4.2. Compute  $p_{k+1}^{n+1}$  as  $p_{k+1}^{n+1} = \alpha p_{k+1/2}^{n+1} + (1 - \alpha) p_k^{n+1}$ .

Step 4.3. Compute  $\bar{u}_{k+1}^{n+1}$  as:

$$\bar{u}_{k+1}^{n+1} = \tilde{u}^{n+1} - \delta t (\mathbf{M}_{\bar{\rho}^{n+1}})^{-1} \mathbf{B}^t (p_{k+1}^{n+1} - \tilde{p}^{n+1}).$$

Convergence criteria:  $\max [\|p_{k+1}^{n+1} - p_k^{n+1}\|, \|u_{k+1}^{n+1} - u_k^{n+1}\|] < \varepsilon$ .

The second step of the previous algorithm is a relaxation step which can be performed to ensure convergence; however, in the tests presented hereafter, we use  $\alpha = 1$  and obtain convergence in a few iterations (typically less than 5). When the equation of state is nonlinear, Step 4.1 is replaced by one iteration of Newton’s algorithm.

### 3.9. Numerical experiments

In this section, we describe numerical experiments which are performed to assess the behaviour of the pressure correction scheme presented in this paper, in particular the convergence rate with respect to the space and time discretizations.

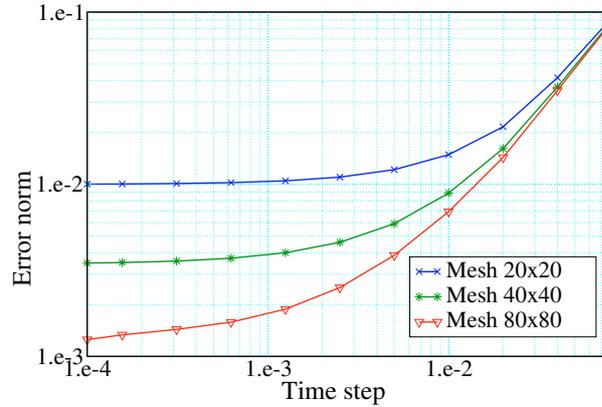


FIGURE 2. Velocity error as a function of the time step.

With  $\Omega = (0, 1) \times \left(-\frac{1}{2}, \frac{1}{2}\right)$ , we choose for the momentum and density the following expressions:

$$\rho u = -\frac{1}{4} \cos(\pi t) \begin{bmatrix} \sin(\pi x_1) \\ \cos(\pi x_2) \end{bmatrix}, \quad \rho = 1 + \frac{1}{4} \sin(\pi t) [\cos(\pi x_1) - \sin(\pi x_2)].$$

These functions satisfy the mass balance equation; for the momentum balance, we add the corresponding right hand side. In this latter equation, the divergence of the stress tensor is given by:

$$\nabla \cdot \tau(u) = \mu \Delta u + \frac{\mu}{3} \nabla \nabla \cdot u, \quad \mu = 10^{-2}$$

and, in the discrete momentum balance equation (3.11), we use instead of  $\int_{\Omega, h} \tau(\tilde{u}^{n+1}) : \nabla \varphi_\sigma^{(i)} dx$  the expression:

$$\mu \int_{\Omega, h} \left[ \nabla \tilde{u}^{n+1} : \nabla \varphi_\sigma^{(i)} + \frac{1}{3} (\nabla \cdot \tilde{u}^{n+1}) (\nabla \cdot \varphi_\sigma^{(i)}) \right] dx$$

which ensures that this term is coercive. The pressure is linked to the density by the following equation of state:

$$p = \wp(\rho) = \frac{\rho - 1}{\gamma \text{Ma}^2}, \quad \gamma = 1.4, \text{Ma} = 0.5$$

where the parameter Ma corresponds to the characteristic Mach number.

We use in these tests a special numerical integration of the forcing term of the momentum balance, which is designed to ensure that the discretization of a gradient is indeed a discrete gradient (*i.e.* if the forcing term  $f_v$  can be recast under the form  $f_v = \nabla g_v$ , the discrete right hand side of the momentum balance belongs to the range of  $B^t$ ).

Velocity and pressure errors obtained at  $t = 0.5$ , in respectively  $L^2$  and discrete  $L^2$  norms and as a function of the time step, are drawn on respectively Figures 2 and 3, for  $20 \times 20$ ,  $40 \times 40$  and  $80 \times 80$  uniform meshes (so using the Rannacher-Turek element). For large time steps, these curves show a decrease corresponding to approximately a first order convergence in time for the velocity and the pressure, until a plateau is reached, due to the fact that errors are bounded from below by the residual spatial discretization error. For both velocity and pressure, the value of the errors on this plateau show a space convergence order (in  $L^2$  norm) between 1 and 2.

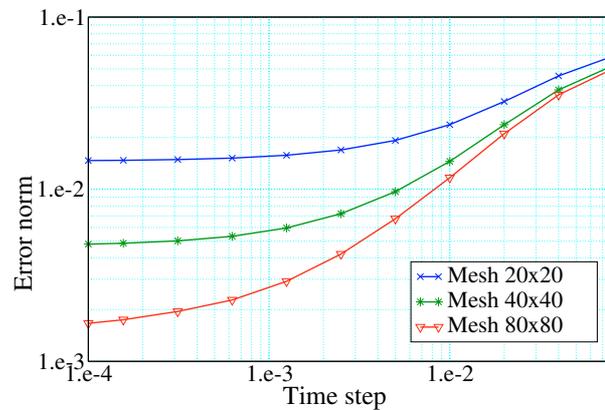


FIGURE 3. Pressure error as a function of the time step.

#### 4. CONCLUSION

We presented in this paper a numerical scheme for the barotropic Navier-Stokes compressible equations, based on a pressure-correction time stepping algorithm. For the space discretization, it combines low-order non-conforming mixed finite elements with finite volumes; in the incompressible limit, one recovers a classical projection scheme based on an *inf-sup* stable pair of approximation spaces for the velocity and the pressure. This scheme is proven to enjoy an unconditional stability property: irrespectively of the time step, the discrete solution obeys the *a priori* estimates associated to the continuous problem, *i.e.* strict positivity of the density, bounds in the  $L^\infty$ -in-time norm of the quantity  $\int_\Omega \rho u^2 dx$  and  $\int_\Omega \rho P(\rho) dx$  and in the  $L^2$ -in-time norm of the viscous dissipation  $\int_\Omega \tau(u) : \nabla u dx$ . To our knowledge, this result is the first one of this type for barotropic compressible flows.

However, the scheme presented here is by no means “the ultimate scheme” for the solution to the compressible Navier-Stokes equations. It should rather be seen as an example of application (and probably one of the less sophisticated ones) of the mathematical arguments developed to obtain stability, namely Theorems 2.1 (discrete elastic potential identity) and 3.1 (stability of the advection operator), and our hope is that these two ingredients could be used as such or adapted in the future to study other algorithms. For instance, a computation close to the proof of Theorem 3.8 (and even simpler) would yield the stability of the fully implicit scheme; adding to this latter algorithm a prediction step for the density (as performed here) would also allow to linearize (once again as performed here) the convection operator without loss of stability. A stable pressure-correction scheme avoiding this prediction step can also be obtained, and is currently under tests at IRSN for the computation of compressible bubbly flows. Besides these variants, less diffusive schemes should certainly be sought. Finally, the proposed scheme is currently the object of more in-depth numerical studies including, in particular, problems admitting less smooth solutions than the test presented here.

#### REFERENCES

- [1] P. Angot, V. Dolejší, M. Feistauer and J. Felcman, Analysis of a combined barycentric finite volume-nonconforming finite element method for nonlinear convection-diffusion problems. *Appl. Math.* **4** (1998) 263–310.
- [2] H. Bijl and P. Wesseling, A unified method for computing incompressible and compressible flows in boundary-fitted coordinates. *J. Comp. Phys.* **141** (1998) 153–173.
- [3] M.O. Bristeau, R. Glowinski, L. Dutto, J. Périaux and G. Rogé, Compressible viscous flow calculations using compatible finite element approximations. *Internat. J. Numer. Methods Fluids* **11** (1990) 719–749.
- [4] V. Casulli and D. Greenspan, Pressure method for the numerical solution of transient, compressible fluid flows. *Internat. J. Numer. Methods Fluids* **4** (1984) 1001–1012.
- [5] A.J. Chorin, Numerical solution of the Navier-Stokes equations. *Math. Comp.* **22** (1968) 745–762.

- [6] P.G. Ciarlet, Finite elements methods – Basic error estimates for elliptic problems, in *Handbook of Numerical Analysis II*, P. Ciarlet and J.-L. Lions Eds., North Holland (1991) 17–351.
- [7] P. Colella and K. Pao, A projection method for low speed flows. *J. Comp. Phys.* **149** (1999) 245–269.
- [8] M. Crouzeix and P.-A. Raviart, Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *RAIRO Anal. Numér.* **7** (1973) 33–75.
- [9] K. Deimling, *Nonlinear Functional Analysis*. Springer, New-York (1980).
- [10] I. Demirdžić, Ž. Lilek and M. Perić, A collocated finite volume method for predicting flows at all speeds. *Internat. J. Numer. Methods Fluids* **16** (1993) 1029–1050.
- [11] V. Dolejší, M. Feistauer, J. Felcman and A. Kliková, Error estimates for barycentric finite volumes combined with nonconforming finite elements applied to nonlinear convection-diffusion problems. *Appl. Math.* **47** (2002) 301–340.
- [12] A. Ern and J.-L. Guermond, *Theory and practice of finite elements, Applied Mathematical Sciences* **159**. Springer (2004).
- [13] R. Eymard, T. Gallouët, M. Ghilani and R. Herbin, Error estimates for the approximate solutions of a nonlinear hyperbolic equation given by finite volume schemes. *IMA J. Numer. Anal.* **18** (1998) 563–594.
- [14] R. Eymard, T. Gallouët and R. Herbin, Finite volume methods, in *Handbook of Numerical Analysis VII*, P. Ciarlet and J.-L. Lions Eds., North Holland (2000) 713–1020.
- [15] E. Feireisl, *Dynamics of viscous compressible flows, Oxford Lecture Series in Mathematics and its Applications* **6**. Oxford University Press (2004).
- [16] H. Feistauer, J. Felcman and I. Straškraba, *Mathematical and computational methods for compressible flows, Oxford Science Publications*. Clarendon Press (2003).
- [17] M. Fortin, H. Manouzi and A. Soulaimani, On finite element approximation and stabilization methods for compressible viscous flows. *Internat. J. Numer. Methods Fluids* **17** (1993) 477–499.
- [18] J.-L. Guermond and L. Quartapelle, A projection FEM for variable density incompressible flows. *J. Comp. Phys.* **165** (2000) 167–188.
- [19] J.-L. Guermond, P. Mineev and J. Shen, An overview of projection methods for incompressible flows. *Comput. Methods Appl. Mech. Engrg.* **195** (2006) 6011–6045.
- [20] F.H. Harlow and A.A. Amsden, Numerical calculation of almost incompressible flow. *J. Comp. Phys.* **3** (1968) 80–93.
- [21] F.H. Harlow and A.A. Amsden, A numerical fluid dynamics calculation method for all flow speeds. *J. Comp. Phys.* **8** (1971) 197–213.
- [22] R.I. Issa, Solution of the implicitly discretised fluid flow equations by operator splitting. *J. Comp. Phys.* **62** (1985) 40–65.
- [23] R.I. Issa and M.H. Javareshkian, Pressure-based compressible calculation method utilizing total variation diminishing schemes. *AIAA J.* **36** (1998) 1652–1657.
- [24] R.I. Issa, A.D. Gosman and A.P. Watkins, The computation of compressible and incompressible recirculating flows by a non-iterative implicit scheme. *J. Comp. Phys.* **62** (1986) 66–82.
- [25] K.C. Karki and S.V. Patankar, Pressure based calculation procedure for viscous flows at all speeds in arbitrary configurations. *AIAA J.* **27** (1989) 1167–1174.
- [26] M.H. Kobayashi and J.C.F. Pereira, Characteristic-based pressure correction at all speeds. *AIAA J.* **34** (1996) 272–280.
- [27] B. Larrouturou, How to preserve the mass fractions positivity when computing compressible multi-component flows. *J. Comp. Phys.* **95** (1991) 59–84.
- [28] P.-L. Lions, *Mathematical topics in fluid mechanics, Volume 2: Compressible models, Oxford Lecture Series in Mathematics and its Applications* **10**, Oxford University Press (1998).
- [29] M. Marion and R. Temam, Navier-Stokes equations: Theory and approximation, in *Handbook of Numerical Analysis VI*, P. Ciarlet and J.-L. Lions Eds., North Holland (1998).
- [30] F. Moukalled and M. Darwish, A high-resolution pressure-based algorithm for fluid flow at all speeds. *J. Comp. Phys.* **168** (2001) 101–133.
- [31] P. Nithiarasu, R. Codina and O.C. Zienkiewicz, The Characteristic-Based Split (CBS) scheme – a unified approach to fluid dynamics. *Internat. J. Numer. Methods Engrg.* **66** (2006) 1514–1546.
- [32] A. Novotný and I. Straškraba, *Introduction to the mathematical theory of compressible flow, Oxford Lecture Series in Mathematics and its Applications* **27**. Oxford University Press (2004).
- [33] G. Patnaik, R.H. Guirguis, J.P. Boris and E.S. Oran, A barely implicit correction for flux-corrected transport. *J. Comp. Phys.* **71** (1987) 1–20.
- [34] E.S. Politis and K.C. Giannakoglou, A pressure-based algorithm for high-speed turbomachinery flows. *Internat. J. Numer. Methods Fluids* **25** (1997) 63–80.
- [35] R. Rannacher and S. Turek, Simple nonconforming quadrilateral Stokes element. *Numer. Methods Partial Differential Equations* **8** (1992) 97–111.
- [36] 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** (1969) 377–385.
- [37] D.R. van der Heul, C. Vuik and P. Wesseling, Stability analysis of segregated solution methods for compressible flow. *Appl. Numer. Math.* **38** (2001) 257–274.

- [38] D.R. van der Heul, C. Vuik and P. Wesseling, A conservative pressure-correction method for flow at all speeds. *Comput. Fluids* **32** (2003) 1113–1132.
- [39] J.P. Van Dormaal, G.D. Raithby and B.H. McDonald, The segregated approach to predicting viscous compressible fluid flows. *Trans. ASME* **109** (1987) 268–277.
- [40] D. Vidović, A. Segal and P. Wesseling, A superlinearly convergent Mach-uniform finite volume method for the Euler equations on staggered unstructured grids. *J. Comput. Phys.* **217** (2006) 277–294.
- [41] C. Wall, C.D. Pierce and P. Moin, A semi-implicit method for resolution of acoustic waves in low Mach number flows. *J. Comp. Phys.* **181** (2002) 545–563.
- [42] I. Wenneker, A. Segal and P. Wesseling, A Mach-uniform unstructured staggered grid method. *Internat. J. Numer. Methods Fluids* **40** (2002) 1209–1235.
- [43] P. Wesseling, *Principles of computational fluid dynamics, Springer Series in Computational Mathematics* **29**. Springer (2001).
- [44] O.C. Zienkiewicz and R. Codina, A general algorithm for compressible and incompressible flow – Part I. The split characteristic-based scheme. *Internat. J. Numer. Methods Fluids* **20** (1995) 869–885.