Construction of a low Mach finite volume scheme for the isentropic Euler system with porosity

Stéphane Dellacherie*, Jonathan Jung† and Pascal Omnes‡

April 13, 2021

Abstract

Classical finite volume schemes for the Euler system are not accurate at low Mach number and some fixes have to be used and were developed in a vast literature over the last two decades. The question we are interested in in this article is: What about if the porosity is no longer uniform? We first show that this problem may be understood on the linear wave equation taking into account porosity. We explain the influence of the cell geometry on the accuracy property at low Mach number. In the triangular case, the stationary space of the Godunov scheme approaches well enough the continuous space of constant pressure and divergence-free velocity, while this is not the case in the Cartesian case. On Cartesian meshes, a fix is proposed and accuracy at low Mach number is proved to be recovered. Based on the linear study, a numerical scheme and a low Mach fix for the non-linear system, with a non-conservative source term due to the porosity variations, is proposed and tested.

1 Introduction

In this paper, we are interested in low Mach compressible fluid flows in porous media. In industrial processes, porous media are used to simulate the flow in a nuclear reactor core. The porosity appears because there are section reductions in a nuclear reactor core. Moreover, if we want to simulate an accidental scenario, we sometimes need to take into account the compressibility effects. Another class of problems motivated by industrial consideration is the simulation of a gas flow across a grid. Since the grid is in general too small to be meshed, a homogenization process is used to model the interactions betwe en the grid and the flow [41]. Then, we consider the barotropic Euler equation. Since the porosity is not constant, a non-conservative term appears in the equations during the homogenization process [4] and the equations write

$$\begin{cases} \partial_t(\alpha\rho) + \nabla_{\mathbf{x}} \cdot (\alpha\rho\mathbf{u}) = 0, \\ \partial_t(\alpha\rho\mathbf{u}) + \nabla_{\mathbf{x}} \cdot (\alpha\rho\mathbf{u} \otimes \mathbf{u}) + \nabla_{\mathbf{x}}(\alpha p) = p\nabla_{\mathbf{x}}\alpha. \end{cases} \tag{1}$$

In (1), $t \geq 0$ and $\mathbf{x} \in \Omega$ are respectively time and space variables and $\alpha(\mathbf{x})$ is the porosity. We suppose here that $\alpha(\mathbf{x})$ is known and does not depend on time. Unknowns ρ, $\mathbf{u}$ and $p(\rho)$ are respectively the density, the velocity and the pressure of the fluid. The pressure law satisfies $p'(\rho) > 0$. System (1) is a non conservative hyperbolic system [27] with eigenvalues in direction $\mathbf{n}$ given by $\mathbf{u}\cdot\mathbf{n} - c$, $\mathbf{u}\cdot\mathbf{n}$ and $\mathbf{u}\cdot\mathbf{n} + c$. Studies of flows in a variable cross section duct consider the same model and variations of the cross section are modeled through (possibly discontinuous) changes in porosity.

The non-conservative term in (1) introduces mathematical and numerical difficulties. In [11], the authors give a mathematical sense to the non-conservative product and introduce some schemes, named well-balanced schemes, that solve correctly the non-conservative term [20, 8]. The treatment of the non-conservative term is essential to preserve steady states solutions. In this paper, we propose a well-balanced scheme that exactly preserves steady solutions over time in one space dimension [21, 26]. The proposed scheme is based on a VFRoe scheme, established in [17] for

*Hydro-Québec - TransÉnergie et Équipement, DCMÉ, Prévisions de contrôle du réseau, Édifice Jean-Lesage, 75 boulevard René-Lévesque Ouest, Montréal (Qc), H2Z 1A4, Canada, **dellacherie.stephane@hydroquebec.com**.

†Université de Pau et des Pays de l'Adour, E2S UPPA, CNRS, INRIA, équipe CAGIRE, LMAP, Pau, France, **jonathan.jung@univ-pau.fr**.

‡Université Paris-Saclay, CEA, Service de Thermo-hydraulique et de Mécanique des Fluides, 91191, Gif-sur-Yvette, France and Université Sorbonne Paris Nord, LAGA, CNRS UMR 7539, Institut Galilée, 99 Av. J.-B. Clément, 93430 Villetaneuse, France, **pascal.omnes@cea.fr**.

the shallow water equation with topography and derived for the Euler system with porosity in [40, 39]. The VFRoe solver consists into a local linearization of a Riemann problem which is simpler to handle since it only deals with linear problems and avoids the complex exact resolution of the Riemann problem with porosity jump. This construction allows to easily build schemes that exactly preserve one dimensional steady states. To our knowledge, no generalization for purely multidimensional problems exists and the behavior of the numerical scheme with respect to multi-dimensional steady solutions must be studied on a case-by-case basis. In this paper, we focus on the behavior of the scheme in the low Mach limit and, as we will see, steady solutions of the numerical scheme will play a determining role.

Finite volume Godunov type schemes applied to the compressible Euler system with uniform porosity are known to be inaccurate at low Mach number [23, 13]. Indeed, they do not allow to recover the incompressible limit as the Mach number tends to zero. Over the two last decades, a large amount of work has been dedicated to deriving fixes for the uniform porosity case: [23, 28, 29, 13, 37, 12, 15, 33, 7, 24, 5]. Some recent works have been done on low Mach fix for non-conservative systems, we refer to [3, 2, 45] for the Euler equation with gravity or to [35, 34] for two-phases flows. In these last studies, flux preconditioning techniques, initially proposed by Turkel [46], are applied and quadrangular meshes are considered. Here, we propose to also study the behavior of the numerical scheme on triangular meshes. Indeed, it was shown that in the uniform porosity case, if the mesh is composed of triangles in 2D or tetrahedra in 3D, the accuracy at low Mach number with the Roe scheme is recovered [38, 16, 22]. To our knowledge, this is the first study on the behavior of classical schemes at low Mach number on triangular meshes for non-conservative systems.

In this article, we study the accuracy, at low Mach number, on triangular and Cartesian meshes, of a numerical scheme for the non conservative system (1). Since the accuracy problem appears also in the linear case, we base our study on the linear wave equation with porosity. The low Mach accuracy problem is then understood and fixed in the linear case for Cartesian meshes, and the reason for its correct behavior on triangular meshes is underlined. In particular, preliminary results obtained in [14] based on a modified equation approach are extended to the discrete Cartesian case. Based on the linear study, a well-balanced scheme accurate at low Mach number for the non-linear system (1) is proposed and numerical tests are performed. They confirm that both the non corrected and corrected schemes are able to recover the low Mach asymptotics on triangular meshes, while this is the case only for the corrected scheme on Cartesian meshes.

2 Low Mach limit and wave equation with porosity

2.1 Low Mach limit

To study the behavior of system (1) at low Mach number, four characteristic scales are supposed to be known : a time scale t_0, a density scale ρ_0, a velocity scale u_0 and a porosity scale α_0. Then, the following dimensionless variables are defined

$$\tilde{t} = \frac{t}{t_0}, \quad \tilde{\rho} = \frac{\rho}{\rho_0}, \quad \tilde{\mathbf{u}} = \frac{\mathbf{u}}{u_0}, \quad \tilde{\alpha} = \frac{\alpha}{\alpha_0} \tag{2}$$

It is natural to scale the length by $L_0 = u_0 \times t_0$, the sound speed by $c_0^2 = p'(\rho_0)$ and the pressure by $p_0 = \rho_0 c_0^2$. If the corresponding dimensionless variables are used, system (1) reads

$$\begin{cases} \partial_{\tilde{t}}(\tilde{\alpha}\tilde{\rho}) + \nabla_{\tilde{\mathbf{x}}} \cdot (\tilde{\alpha}\tilde{\rho}\tilde{\mathbf{u}}) = 0, \\ \partial_{\tilde{t}}(\tilde{\alpha}\tilde{\rho}\tilde{\mathbf{u}}) + \nabla_{\tilde{\mathbf{x}}} \cdot (\tilde{\alpha}\tilde{\rho}\tilde{\mathbf{u}} \otimes \tilde{\mathbf{u}}) + \dfrac{\tilde{\alpha}}{M^2}\nabla_{\tilde{\mathbf{x}}}\tilde{p} = 0 \end{cases} \tag{3}$$

with $\tilde{\mathbf{x}} = \mathbf{x}/L_0$, $\tilde{p} = p/p_0$, and where $M = u_0/c_0$ is the so-called Mach number.

2.2 Formal asymptotic expansion when the Mach number goes to 0

We are interested in the solutions of (3) when $M \to 0$. We recall formally the theoretical results of [25] in order to take the porosity into account. All the variables of the system, $\varphi \in \{\tilde{\alpha}, \tilde{\rho}, \tilde{\mathbf{u}}\}$, are developed as power series of the Mach number M:

$$\tilde{\varphi}(\tilde{\mathbf{x}}, \tilde{t}, M) = \sum_{n=0}^{N} M^n \tilde{\varphi}^{(n)}(\tilde{\mathbf{x}}, \tilde{t}) + \mathcal{O}\left(M^{N+1}\right). \tag{4}$$

Assumption 2.1. *Concerning α, we assume that $\alpha(\boldsymbol{x})$ is a function that takes its values in $[\alpha_{\min}, 1]$, where $\alpha_{\min} > 0$ is a constant independent of the Mach number M. This implies that*

$$\tilde{\alpha}^{(0)} \neq 0.$$

The case $\tilde{\alpha}^{(0)} = 0$ is out of the scope of this paper.

By injecting these quantities in (3), the momentum equation at order M^{-2} and M^{-1} gives

$$\nabla_{\tilde{\mathbf{x}}} \tilde{p}^{(0)} = \nabla_{\tilde{\mathbf{x}}} \tilde{p}^{(1)} = 0 \tag{5}$$

and then, since p is a regular function of ρ, this leads to

$$\tilde{\rho}^{(0)}(\tilde{\mathbf{x}}, \tilde{t}) = \tilde{\rho}^{(0)}(\tilde{t}) \quad \text{and} \quad \tilde{\rho}^{(1)}(\tilde{\mathbf{x}}, \tilde{t}) = \tilde{\rho}^{(1)}(\tilde{t}). \tag{6}$$

At order M^0, we get

$$\begin{cases} \partial_{\tilde{t}}(\tilde{\alpha}\tilde{\rho})^{(0)} + \nabla_{\tilde{\mathbf{x}}} \cdot (\tilde{\alpha}\tilde{\rho}\tilde{\mathbf{u}})^{(0)} = 0, \\ \partial_{\tilde{t}}(\tilde{\alpha}\tilde{\rho}\tilde{\mathbf{u}})^{(0)} + \nabla_{\tilde{\mathbf{x}}} \cdot \left((\tilde{\alpha}\tilde{\rho}\tilde{\mathbf{u}})^{(0)} \otimes \tilde{\mathbf{u}}^{(0)} \right) + \tilde{\alpha}^{(0)} \nabla_{\tilde{\mathbf{x}}} \tilde{p}^{(2)} = 0. \end{cases}$$

Then, if the initial and boundary conditions are *well prepared* in the sense that

$$\begin{cases} \tilde{\rho}(\tilde{t} = 0, \tilde{\mathbf{x}}, M) = \tilde{\rho}_0 + \mathcal{O}\left(M^2\right), & \text{where } \tilde{\rho}_0(\tilde{\mathbf{x}}) = \tilde{\rho}_0 \in \mathbb{R}^{+*} \\ (\tilde{\alpha}\tilde{\mathbf{u}})(\tilde{t} = 0, \tilde{\mathbf{x}}, M) = (\tilde{\alpha}\tilde{\mathbf{u}})_0^{(0)} + \mathcal{O}\left(M\right), & \text{where } \nabla \cdot (\tilde{\alpha}\tilde{\mathbf{u}})_0^{(0)} = 0 \end{cases}$$

and if on the domain boundary $\tilde{\rho}^{(0)}$ (resp. $\tilde{\rho}^{(1)}$) is uniformly and constantly equals to $\tilde{\rho}_0$ (resp. 0) and if $\int_{\partial\Omega} (\tilde{\alpha}\tilde{\mathbf{u}})^{(0)} \cdot \mathbf{n} = 0$, the solution of (3) satisfies

$$\begin{cases} \tilde{\rho}(\tilde{t}, \tilde{\mathbf{x}}, M) = \tilde{\rho}_0 + \mathcal{O}\left(M^2\right), \\ (\tilde{\alpha}\tilde{\mathbf{u}})(\tilde{t}, \tilde{\mathbf{x}}, M) = (\tilde{\alpha}\tilde{\mathbf{u}})^{(0)}(\tilde{t}, \tilde{\mathbf{x}}) + \mathcal{O}\left(M\right) \end{cases} \tag{7}$$

where $(\tilde{\rho}^{(2)}, \tilde{\mathbf{u}}^{(0)})$ satisfies

$$\begin{cases} \nabla \cdot (\tilde{\alpha}\tilde{\mathbf{u}})^{(0)} = 0, \\ \partial_{\tilde{t}} \tilde{\mathbf{u}}^{(0)} + \left(\tilde{\mathbf{u}}^{(0)} \cdot \nabla_{\tilde{\mathbf{x}}} \right) \tilde{\mathbf{u}}^{(0)} + \nabla_{\tilde{\mathbf{x}}} \tilde{\rho}^{(2)} = 0. \end{cases} \tag{8}$$

Note that in order to obtain the second equation in (8), we have chosen $\tilde{\rho}_0 = 1$, which is always possible up to a change of density scale from ρ_0 to $\rho_0\tilde{\rho}_0$. Equations (7) mean that at low Mach number, if the initial and boundary conditions are *well prepared*, the solution of the compressible Euler system with porosity (3) is close to the solution of the incompressible Euler equation with porosity (8). Results (7) are formally proven here. For a rigorous proof in the uniform porosity case, we refer to [42, 30].

For classical finite volume schemes, relations (7) are not always satisfied at the discrete level: this is the so-called accuracy problem at low Mach number, which expresses that a spurious component $\tilde{\rho}^{(1)} \neq 0$ could be introduced at the discrete level [23] due to numerical approximations. In the current contribution, we consider that a numerical scheme is *accurate at low Mach number* for system (1) if relations (7) are satisfied at the discrete level.

2.3 Wave equation with porosity

To study the low Mach behavior, we change the variables to symmetrize the problem.

2.3.1 Model

For this purpose, we set the reference sound speed to $1/M$ and we define $r(\tilde{t}, \tilde{\mathbf{x}})$ such that

$$\tilde{\rho}(\tilde{t}, \tilde{\mathbf{x}}) = \tilde{\rho}_0 \left(1 + Mr(\tilde{t}, \tilde{\mathbf{x}})\right) \tag{9}$$

where formally $Mr \ll 1$. By injecting (9) in (3), we obtain the system

$$\begin{cases} \partial_{\tilde{t}}(\tilde{\alpha}r) + \nabla_{\tilde{\mathbf{x}}} \cdot (\tilde{\alpha}r\tilde{\mathbf{u}}) + \dfrac{1}{M} \nabla_{\tilde{\mathbf{x}}} \cdot (\tilde{\alpha}\tilde{\mathbf{u}}) = 0, \\ \partial_{\tilde{t}}(\tilde{\alpha}\tilde{\mathbf{u}}) + (\tilde{\mathbf{u}} \cdot \nabla_{\tilde{\mathbf{x}}})(\tilde{\alpha}\tilde{\mathbf{u}}) + \dfrac{\tilde{\alpha}}{M} \dfrac{\tilde{p}'\left(\tilde{\rho}_0\left(1 + Mr\right)\right)}{1 + Mr} \nabla_{\tilde{\mathbf{x}}} r = 0. \end{cases}$$

Linearizing around $(r, \tilde{\mathbf{u}}) = (0, 0)$, taking into account that $\tilde{p}'(\tilde{\rho}_0) = 1$ when $\tilde{\rho}_0 = 1$ as explained above and simplifying the notation by removing all the $\tilde{\cdot}$, we obtain the linear wave equation with porosity

$$\partial_t(\alpha q) + \frac{L_\alpha}{M}(q) = 0 \tag{10}$$

where

$$q = \begin{pmatrix} r \\ \mathbf{u} \end{pmatrix} \quad \text{and} \quad L_\alpha(q) = a_\star \begin{pmatrix} \nabla \cdot (\alpha \mathbf{u}) \\ \alpha \nabla r \end{pmatrix}$$

and $a_\star = 1$.

2.3.2 Weighted incompressible space $\mathcal{E}_\alpha$ and acoustic space $\mathcal{E}_\alpha^\perp$

We are interested in the properties of System (10) solved on a torus $\mathbb{T} \subset \mathbb{R}^{d \in \{1,2,3\}}$ with periodic boundary conditions. For this, we assume that α is a periodic function on $\mathbb{T}$ and we define the weighted Hilbert space

$$L^2_\alpha(\mathbb{T})^{1+d} := \left\{ q := (r, \mathbf{u})^T \Big| \int_{\mathbb{T}} r^2 \alpha d\mathbf{x} + \int_{\mathbb{T}} |\ \mathbf{u}\ |^2\ \alpha d\mathbf{x} < +\infty \right\}$$

endowed with the scalar product

$$\langle q_1, q_2 \rangle_\alpha = \int_{\mathbb{T}} r_1 r_2 \alpha d\mathbf{x} + \int_{\mathbb{T}} \mathbf{u}_1 \cdot \mathbf{u}_2 \alpha d\mathbf{x}. \tag{11}$$

Of course, the space L^2_α should not be mistaken for the acoustic operator L_α. We also define the spaces $H^1_\alpha(\mathbb{T})$ and $H^2_\alpha(\mathbb{T})$ that are generalizations of $H^1(\mathbb{T})$ and $H^2(\mathbb{T})$ to weighted spaces. We note that since $\alpha(\mathbf{x}) \in [\alpha_{\min}, 1]$ with $\alpha_{\min} > 0$, the functions α and $\frac{1}{\alpha}$ are in $L^\infty(\mathbb{T})$, and we have $L^2_\alpha(\mathbb{T}) = L^2(\mathbb{T})$, $H^1_\alpha(\mathbb{T}) = H^1(\mathbb{T})$ and $H^2_\alpha(\mathbb{T}) = H^2(\mathbb{T})$. Nevertheless, we keep the index α to define these spaces to refer to the scalar product (11). At last, we define the space

$$\mathcal{E}_\alpha := \left\{ q = (r, \mathbf{u})^T \in L^2_\alpha(\mathbb{T})^{1+d} \quad \Big| \quad \nabla r = 0 \text{ and } \nabla \cdot (\alpha \mathbf{u}) = 0 \right\} = \text{Ker}\, L_\alpha. \tag{12}$$

When $\alpha = 1$, $\mathcal{E}_\alpha$ is named the incompressible space (see [13]). We have the following result:

Lemma 2.2. *We have*

$$\begin{aligned} \mathcal{E}_\alpha^\perp &= \left\{ q = (r, \boldsymbol{u})^T \in L^2_\alpha(\mathbb{T})^{1+d} \Big| \int_{\mathbb{T}} r \alpha d\boldsymbol{x} = 0 \text{ and } \exists \phi \in H^1_\alpha(\mathbb{T}),\ \boldsymbol{u} = \nabla \phi \right\}, \\ \mathcal{E}_\alpha &\oplus \mathcal{E}_\alpha^\perp = L^2_\alpha(\mathbb{T})^{1+d}. \end{aligned} \tag{13}$$

In other words, any $q = (r, \boldsymbol{u})^T \in L^2_\alpha(\mathbb{T})^{1+d}$ can be decomposed into

$$q = \hat{q} + q^\perp \tag{14}$$

where $\hat{q} = (\hat{r}, \hat{\boldsymbol{u}})^T \in \mathcal{E}_\alpha$ and $q^\perp = (r^\perp, \boldsymbol{u}^\perp)^T \in \mathcal{E}_\alpha^\perp$ and this decomposition is unique and orthogonal with respect to the scalar product defined by (11).

We call $\mathcal{E}_\alpha^\perp$ the acoustic space. This is a generalization of the Hodge decomposition. Decomposition (14) defines an orthogonal projection

$$\begin{aligned} \mathbb{P}_\alpha :\ L^2_\alpha(\mathbb{T})^{1+d} &\longrightarrow \mathcal{E}_\alpha \\ q &\longmapsto \mathbb{P}_\alpha q := \hat{q}. \end{aligned} \tag{15}$$

2.3.3 Properties of the linear wave equation with porosity

We now detail some properties of the linear wave equation with porosity. These properties will not be always satisfied in the discrete case.

Lemma 2.3. *Let $q(t, \boldsymbol{x})$ be the solution of* (10) *on $\mathbb{T} \subset \mathbb{R}^{d \in \{1,2,3\}}$ with initial condition q^0. Then:*

1. $q^0 \in \mathcal{E}_\alpha \Longrightarrow q(t \geq 0) = q^0 \in \mathcal{E}_\alpha$;
2. $q^0 \in \mathcal{E}_\alpha^\perp \Longrightarrow q(t \geq 0) \in \mathcal{E}_\alpha^\perp$.

For all $q \in L^2_\alpha(\mathbb{T})^{1+d}$, we now define the energy $E_\alpha := \langle q, q \rangle_\alpha$. The following lemma is an extension of the energy conservation property of the classical linear wave equation:

Lemma 2.4. *Let $q(t, \boldsymbol{x})$ be the solution of* (10) *on $\mathbb{T} \subset \mathbb{R}^{d \in \{1,2,3\}}$. Then, for all $t \geq 0$,*

$$E_\alpha(t \geq 0) = E_\alpha(t = 0).$$

2.3.4 The low Mach asymptotics

With Lemma 2.3 and by linearity, we get that if $q(t,\mathbf{x})$ is the solution of (10) on $\mathbb{T} \subset \mathbb{R}^{d\in\{1,2,3\}}$ with initial condition q^0, then

$$\|q^0 - \mathbb{P}_\alpha q^0\| = \mathcal{O}(M) \qquad \Longrightarrow \qquad \forall t \geq 0, \quad \|q - \mathbb{P}_\alpha q\|(t) = \mathcal{O}(M). \tag{16}$$

We note that since $\mathbb{P}_\alpha q^0 \in \mathcal{E}_\alpha$ is a stationary solution of (10), then $\mathbb{P}_\alpha q = \mathbb{P}_\alpha q^0$; hence (16) can be written as

$$\|q^0 - \mathbb{P}_\alpha q^0\| = \mathcal{O}(M) \qquad \Longrightarrow \qquad \forall t \geq 0, \quad \|q - \mathbb{P}_\alpha q^0\|(t) = \mathcal{O}(M). \tag{17}$$

In fact, (16) is a version of (7) for the linear case. Indeed, the left condition in (16) just means that the initial condition is *well-prepared*. In the non-linear case, the projection $\mathbb{P}_\alpha q$ in the incompressible space $\mathcal{E}_\alpha$ is replaced by the incompressible solution of (8).

In this article, we consider that a numerical scheme for the linear system (10) is *accurate at low Mach number* if (17) is satisfied at the discrete level. We will study this property on Cartesian and triangular meshes.

3 Godunov scheme for the linear wave equation with porosity and its kernels

In [16, 15], we explained the satisfactory behavior of the Godunov scheme at low Mach number on triangular meshes and its wrong behavior on Cartesian rectangular meshes on the Euler system without porosity (α uniformly equal to 1) by studying the kernel of the discrete spatial operator associated to the Godunov scheme. We also remarked that the accuracy of the Godunov scheme at low Mach number on Cartesian meshes can be recovered by deleting the diffusion term on the velocity field in the Godunov scheme. In [14] we discussed the case with porosity with the help of the modified equation approach; the limitations of this approach is that it only gives hints (but does not provide with a complete proof) on what happens on Cartesian meshes, and does not apply to triangular meshes. Our aim here is to analyse the behavior of the schemes on triangular and rectangular Cartesian meshes by directly studying them rather than their modified equations.

We now recall the Godunov scheme for the linear wave equation with porosity, recall why the study of its kernel is so important to study its low Mach accuracy and compute explicitly its kernels on triangular and Cartesian rectangular meshes. In particular, we underline that the kernel is strongly linked to the numerical dissipation of the Godunov scheme.

3.1 Godunov scheme

Let us suppose that the domain $\mathbb{T} \subset \mathbb{R}^2$ is discretized by N cells Ω_i. Let Γ_{ij} be the common edge of the two neighboring cells Ω_i and Ω_j and $\mathbf{n}_{ij}$ the unit vector normal to Γ_{ij} pointing from Ω_i to Ω_j. We assume that the data α, and the unknowns r and $\mathbf{u}$ are defined on the cells Ω_i in the following way

$$\alpha_i = \frac{1}{|\Omega_i|}\int_{\Omega_i} \alpha d\mathbf{x}, \quad r_i \approx \frac{1}{|\Omega_i|}\int_{\Omega_i} r d\mathbf{x}, \quad \mathbf{u}_i \approx \frac{1}{|\Omega_i|}\int_{\Omega_i} \mathbf{u} d\mathbf{x},$$

and then set $(\alpha r)_i = \alpha_i r_i$ and $(\alpha \mathbf{u})_i = \alpha_i \mathbf{u}_i$.

The semi-discrete Godunov scheme applied to the resolution of the linear wave equation is obtained by integrating (10) over each cell Ω_i and then solving a Riemann problem on each Γ_{ij} to express interface fluxes as functions of cell-centered values. Details are provided in [14]. This results in

$$\begin{cases} \dfrac{d}{dt}(\alpha r)_i + \dfrac{a_\star}{2M}\dfrac{1}{|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\Big[\big((\alpha\mathbf{u})_i + (\alpha\mathbf{u})_j\big)\cdot\mathbf{n}_{ij} + \alpha_{ij}(r_i - r_j)\Big] = 0, \\ \dfrac{d}{dt}(\alpha \mathbf{u})_i + \dfrac{a_\star}{2M}\dfrac{\alpha_i}{|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\left[r_i + r_j + \frac{\kappa}{\alpha_{ij}}\big((\alpha\mathbf{u})_i - (\alpha\mathbf{u})_j\big)\cdot\mathbf{n}_{ij}\right]\mathbf{n}_{ij} = 0 \end{cases} \tag{18}$$

with $\kappa = 1$ and where α_{ij} is a mean-value of α on Γ_{ij} which depends on (α_i, α_j) (*e.g.* arithmetic or harmonic mean). The numerical flux in (19) is non-conservative because of the term α_i that multiplies the flux on the momentum equation. Moreover, it is easy to prove the following properties:

Remark 3.1. *The numerical scheme* (18) *is well-balanced in the sense that it preserves exactly the one-dimensional steady states* ($r = \text{cte}, \alpha\boldsymbol{u} = \text{cte}$).

Remark 3.2. *The numerical scheme* (18) *can also be viewed as the VFRoe scheme [18, 6] obtained with the variables* $(\alpha, r, \alpha\boldsymbol{u})$ *for system* (10) *where the linearized Riemann problem is solved considering that* α *satisfies* $\partial_t \alpha = 0$.

Scheme (18) can be written in compact form

$$\begin{cases} \dfrac{d}{dt}(\alpha q_h) + \dfrac{\mathbb{L}^h_{\kappa,\alpha}}{M}(q_h) = 0, \\ q_h(t=0) = q_h^0, \end{cases} \qquad \text{with } q_h := \begin{pmatrix} r_i \\ \mathbf{u}_i \end{pmatrix}_{1 \le i \le N} \tag{19}$$

where the subscript $\cdot_h$ recalls that (19) comes from a spatial discretization of (10).

3.2 The low Mach problem

We want to study whether the Godunov scheme is *accurate at low Mach number* in the sense that it satisfies a version of (17) at the discrete level. Then, discrete incompressible spaces $\mathcal{E}^h_\alpha$ and $\left(\mathcal{E}^h_\alpha\right)^\perp$ and a discrete orthogonal projection $\mathbb{P}^h_\alpha$ have to be define on triangular or Cartesian meshes. Moreover, the key points to obtain (17) at the continuous level are that $\mathcal{E}_\alpha = \operatorname{Ker} L_\alpha$ and that (10) conserves energy (see Lemma 2.4). Then, the relationship between the discrete incompressible space $\mathcal{E}^h_\alpha$ and the kernel of the Godunov scheme $\operatorname{Ker} \mathbb{L}^h_\alpha$ have to be studied. The following theorem explains why this study is so important:

Theorem 3.3. *Suppose that system* (19) *is well-posed in such a way that* $\|q_h(t)\| \le C\|q_h^0\|$ *for any* $t \ge 0$, *where* C *is a positive constant independent of the Mach number* M *and suppose moreover that* $\mathcal{E}^h_\alpha \subseteq Ker\, \mathbb{L}^h_\alpha$. *Then, we have*

$$\|q_h^0 - \mathbb{P}^h_\alpha q_h^0\| = \mathcal{O}(M) \qquad \Longrightarrow \qquad \forall t \ge 0, \quad \|q_h - \mathbb{P}^h_\alpha q_h^0\| = \mathcal{O}(M)$$

For a proof, we refer to [13, 15]. In Theorem 3.3, system (19) is assumed to be well-posed. In particular, stability will be studied in more details in section 4. In the current section, we focus on the kernel of the Godunov scheme on Cartesian and triangular meshes.

3.3 Kernels of the Godunov scheme

We first study the discrete kernel of the Godunov scheme ($\kappa = 1$ in (18)) on different types of meshes and of its low-Mach modification ($\kappa = 0$) on Cartesian rectangular meshes. The kernel $\operatorname{Ker} \mathbb{L}^h_{\kappa,\alpha}$ of the discrete acoustic operator $\mathbb{L}^h_{\kappa,\alpha}$ is defined by

$$\operatorname{Ker} \mathbb{L}^h_{\kappa,\alpha} = \Bigg\{ q_h := \begin{pmatrix} r_i \\ \mathbf{u}_i \end{pmatrix}_i \in \mathbb{R}^{3N} \Bigg| \, \forall i, \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \Big[\big((\alpha\mathbf{u})_i + (\alpha\mathbf{u})_j \big) \cdot \mathbf{n}_{ij} + \alpha_{ij}(r_i - r_j) \Big] = 0$$
$$\text{and } \forall i, \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left[r_i + r_j + \frac{\kappa}{\alpha_{ij}} \Big((\alpha\mathbf{u})_i - (\alpha\mathbf{u})_j \Big) \cdot \mathbf{n}_{ij} \right] \mathbf{n}_{ij} = 0 \Bigg\}. \tag{20}$$

On any type of mesh we have the following result, whose proof is postponed to Appendix A:

Lemma 3.4.

$$Ker\ \mathbb{L}^h_{\kappa>0,\alpha} \ = \ \left\{ q_h := \begin{pmatrix} r_i \\ \boldsymbol{u}_i \end{pmatrix}_i \in \mathbb{R}^{3N} \Big| \exists c \in \mathbb{R}, \forall i, r_i = c \ and \ (\alpha\boldsymbol{u})_i \cdot \boldsymbol{n}_{ij} = (\alpha\boldsymbol{u})_j \cdot \boldsymbol{n}_{ij} \right\} \tag{21}$$

and

$$Ker \mathbb{L}^h_{\kappa=0,\alpha} = \Bigg\{ q_h := \begin{pmatrix} r_i \\ \boldsymbol{u}_i \end{pmatrix}_i \in \mathbb{R}^{3N} \Big| \exists c \in \mathbb{R}, \forall i, r_i = c$$
$$and \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \Big((\alpha\boldsymbol{u})_i + (\alpha\boldsymbol{u})_j \Big) \cdot \boldsymbol{n}_{ij} = 0 \Bigg\}. \tag{22}$$

Moreover, we have

$$Ker \mathbb{L}^h_{\kappa>0,\alpha} \subsetneq Ker \mathbb{L}^h_{\kappa=0,\alpha}.$$

3.3.1 Kernel on a triangular mesh

We now study some particular properties of the behavior of the Godunov scheme on a triangular mesh. Especially, we study the relation between the kernel of the Godunov scheme on a triangular mesh and a discrete version of the space $\mathcal{E}_\alpha$ defined by (12).

Construction of $\mathcal{E}_\alpha^{h,\triangle}$ et $\left(\mathcal{E}_\alpha^{h,\triangle}\right)^\perp$ We construct an accurate discrete version of the well-prepared subspace $\mathcal{E}_\alpha$ defined by (12). Let us suppose that all Ω_i are triangles arranged so that the computational domain is periodic. Moreover, let us denote by V_h the standard P^1 (first-order polynomial functions) Lagrange finite element space associated with this triangular mesh

$$V_h := \left\{\psi_h \in C^0(\mathbb{T}), \psi_h \text{ periodic on } \mathbb{T} \text{ such that } \forall \Omega_i :\ (\psi_h)_{|\Omega_i} \in P^1(\Omega_i)\right\}. \tag{23}$$

Let us also denote by W_h the nonconforming Crouzeix-Raviart P^1 finite element space associated with this triangular mesh

$$W_h := \{\phi_h \in L^2(\mathbb{T}), \phi_h \text{ periodic on } \mathbb{T} \text{ such that } \forall \Omega_i :\ (\phi_h)_{|\Omega_i} \in P^1(\Omega_i) \text{ and } \phi_h \text{ is continuous at the edge midpoints}\}.$$

Note that since the functions in V_h (resp. W_h) are P^1 on each cell, their curls (resp. their gradients) are constant vectors on each cell. Let us also define the discrete vector subspace

$$\mathcal{E}_\alpha^{h,\triangle} = \left\{ q_h := \begin{pmatrix} r_i \\ \mathbf{u}_i \end{pmatrix}_i \in \mathbb{R}^{3N} \Big| \exists (a,b,c,\psi_h) \in \mathbb{R}^3 \times V_h,\ \forall i \in [\![1,N]\!],\ r_i = c \text{ and } (\alpha\mathbf{u})_i = \begin{pmatrix} a \\ b \end{pmatrix} + (\nabla \times \psi_h)_{|\Omega_i} \right\}. \tag{24}$$

Then, we define the space of constant piecewise functions

$$l^2_\alpha(\mathbb{T})^3 := \left\{ q_h := \begin{pmatrix} r_i \\ \mathbf{u}_i \end{pmatrix}_i \in \mathbb{R}^{3N} \Big| \sum_i |\Omega_i|(r_i^2 + |\mathbf{u}_i|^2)\alpha_i < +\infty \right\}$$

endowed with the scalar product (11) which may be written for $(q_h)_1$ and $(q_h)_2$ in $l^2_\alpha(\mathbb{T})^3$ as

$$\langle (q_h)_1, (q_h)_2 \rangle_{\alpha,h} = \sum_i |\Omega_i| \left[(r_1)_i (r_2)_i + (\mathbf{u}_1)_i \cdot (\mathbf{u}_2)_i \right] \alpha_i. \tag{25}$$

Adapting the proof of Theorem 4.1 in [1] (see also [32]) to the case of periodic elements in V_h and W_h and weighted spaces, we may prove the following lemma:

Lemma 3.5. *Assume that $(\Omega_i)_{i=1\cdots N}$ is a triangular periodic mesh of a rectangular domain with no internal holes. For any $(r,\boldsymbol{u})^T \in \mathbb{R}^{3N}$, there exist unique $(a,b) \in \mathbb{R}^2$, a unique $\psi_h \in V_h$ and a unique $\phi_h \in W_h$ with $\int_{\mathbb{T}} \psi_h(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{T}} \phi_h(\boldsymbol{x})d\boldsymbol{x} = 0$, such that on any Ω_i, we have*

$$\begin{pmatrix} r_i \\ \boldsymbol{u}_i \end{pmatrix} = \begin{pmatrix} \bar{r} \\ \dfrac{1}{\alpha_i}\begin{pmatrix} a \\ b \end{pmatrix} + \dfrac{1}{\alpha_i}(\nabla \times \psi_h)_{|\Omega_i} \end{pmatrix} + \begin{pmatrix} r_i - \bar{r} \\ (\nabla \phi_h)_{|\Omega_i} \end{pmatrix} \tag{26}$$

with $\bar{r} = \dfrac{\sum_i |\Omega_i|\alpha_i r_i}{\sum_i |\Omega_i|\alpha_i}$. Moreover this decomposition is orthogonal for the scalar product (25).

Proof. We firstly prove the orthogonality of decomposition (26). The orthogonality between $\bar{r}$ and $r - \bar{r}$ is obvious because, by definition of $\bar{r}$ we have:

$$\langle \bar{r}, r - \bar{r} \rangle_{\alpha,h} = \sum_i |\Omega_i|\alpha_i \bar{r}(r-\bar{r})_i = \bar{r}\left(\sum_i |\Omega_i|\alpha_i r_i - \bar{r} \sum_i |\Omega_i|\alpha_i \right) = 0.$$

Now, we prove the orthogonality for the decomposition of $\mathbf{u}$. For any $(a,b)^T \in \mathbb{R}^2$ and $\phi_h \in W_h$ (then $\nabla\phi_h$ is a constant vector on each cell Ω_i), we have:

$$\begin{aligned} \left\langle \frac{1}{\alpha}\begin{pmatrix} a \\ b \end{pmatrix}, \nabla\phi_h \right\rangle_{\alpha,h} &= \begin{pmatrix} a \\ b \end{pmatrix} \cdot \sum_i |\Omega_i| (\nabla\phi_h)_{|\Omega_i} = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \sum_i \int_{\Omega_i} (\nabla\phi_h)_{|\Omega_i} d\mathbf{x} \\ &= \begin{pmatrix} a \\ b \end{pmatrix} \cdot \sum_i \int_{\partial\Omega_i} \phi_h \mathbf{n} d\sigma = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \sum_{\Gamma_{ij}} \int_{\Gamma_{ij}} [\phi_h]_{ij} \mathbf{n}_{ij} d\sigma \end{aligned}$$

where $[\phi_h]_{ij}$ denotes the jump of ϕ_h through the edge Γ_{ij}. To obtain the last equality, we used the fact that each interface Γ_{ij} contributes twice in the sum over the cell boundaries. Since ϕ_h is a P^1 function, its integral on the edge Γ_{ij} is equal to the length $|\Gamma_{ij}|$ multiplied by the value of ϕ_h at its midpoint. Thus, since ϕ_h is continuous at the edge midpoints, we have $\int_{\Gamma_{ij}}[\phi_h]_{ij}\mathbf{n}_{ij}d\sigma = 0$ on any edge, which proves the orthogonality between the field $\frac{1}{\alpha}(a,b)^T$ and the gradient of any element in W_h. Moreover, for any $\psi_h \in V_h$ and $\phi_h \in W_h$ (then $\nabla \times \psi_h$ and $\nabla\phi_h$ are constant vectors on each cell Ω_i), it holds that

$$\begin{aligned}\left\langle \frac{1}{\alpha}\nabla\times\psi_h, \nabla\phi_h \right\rangle_{\alpha,h} &= \sum_i |\Omega_i|(\nabla\times\psi_h)_{|\Omega_i}\cdot(\nabla\phi_h)_{|\Omega_i} = \sum_i \int_{\Omega_i}(\nabla\times\psi_h)_{|\Omega_i}\cdot(\nabla\phi_h)_{|\Omega_i}d\mathbf{x} \\ &= \sum_i \int_{\partial\Omega_i}(\phi_h)_{|\Omega_i}(\nabla\times\psi_h)_{|\Omega_i}\cdot\mathbf{n}d\sigma - \sum_i\int_{\Omega_i}(\phi_h)_{|\Omega_i}\nabla\cdot(\nabla\times\psi_h)_{|\Omega_i}d\mathbf{x}.\end{aligned}$$

Since $\nabla\cdot(\nabla\times) = 0$, the second sum vanishes. Moreover, denoting by $\mathbf{t}$ a unit vector such that $(\mathbf{n},\mathbf{t})$ is a direct orthonormal system, the equality $(\nabla\times\psi_h)\cdot\mathbf{n} = (\nabla\psi_h)\cdot\mathbf{t}$ and the fact that $\nabla\psi_h\cdot\mathbf{t}$ is continuous along any interface Γ_{ij} (since $\psi_h \in V_h$ is a P^1 nodal Lagrange function) imply that

$$\left\langle \frac{1}{\alpha}\nabla\times\psi_h, \nabla\phi_h \right\rangle_{\alpha,h} = \sum_{\Gamma_{ij}}\int_{\Gamma_{ij}}\nabla\psi_h\cdot\mathbf{t}_{ij}\,[\phi_h]_{ij}\,d\sigma.$$

But on Γ_{ij}, the product $\nabla\psi_h\cdot\mathbf{t}_{ij}\,[\phi_h]_{ij}$ is a P^1 function, and its integral over Γ_{ij} is equal to the length $|\Gamma_{ij}|$ multiplied by the value of this function at its midpoint. Thus, since ϕ_h is continuous at the midpoint, then $\int_{\Gamma_{ij}}\nabla\psi_h\cdot\mathbf{t}_{ij}\,[\phi_h]_{ij}\,d\sigma = 0$ on any edge, which proves orthogonality between $\frac{1}{\alpha}\nabla\times\psi_h$ and $\nabla\phi_h$. Then, the orthogonality of the decomposition is proved.

We secondly prove the existence and the uniqueness of decomposition (26). For r there is no difficulty. Thus, we only consider the decomposition for $\mathbf{u}$. We have to prove that the function L defined by

$$\begin{array}{rcl} L:\ \mathbb{R}^2\times V_h^0\times W_h^0 & \to & \mathbb{R}^{2N} \\ \left(\begin{pmatrix} a \\ b \end{pmatrix}, \psi_h, \phi_h\right) & \mapsto & \frac{1}{\alpha}\begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha}\nabla\times\psi_h + \nabla\phi_h \end{array} \tag{27}$$

is bijective, where $V_h^0 = \left\{\psi_h \in V_h \middle| \int_{\mathbb{T}}\psi_h d\mathbf{x} = 0\right\}$ and $W_h^0 = \left\{\phi_h \in W_h \middle| \int_{\mathbb{T}}\phi_h d\mathbf{x} = 0\right\}$. Firstly, we prove injectivity. As L is a linear function, we just have to prove that

$$L\left(\begin{pmatrix} a \\ b \end{pmatrix}, \psi_h, \phi_h\right) = 0 \Longrightarrow \left(\begin{pmatrix} a \\ b \end{pmatrix}, \psi_h, \phi_h\right) = 0.$$

Assume that for all $i \in [\![1,N]\!]$, $\frac{1}{\alpha_i}\begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha_i}(\nabla\times\psi_h)_{|\Omega_i} + (\nabla\phi_h)_{|\Omega_i} = 0$. By the orthogonality that we proved above, this implies

$$\forall i \in [\![1,N]\!]: \quad \begin{cases} \frac{1}{\alpha_i}\begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha_i}(\nabla\times\psi_h)_{|\Omega_i} = 0, \\ (\nabla\phi_h)_{|\Omega_i} = 0, \end{cases}$$

which implies that

$$\forall i \in [\![1,N]\!]:\ \exists\gamma_i\in\mathbb{R},\ \exists\beta_i\in\mathbb{R},\ \forall(x,y)\in\Omega_i, \quad \begin{cases} (\psi_h)_{|\Omega_i}(x,y) = bx - ay + \beta_i, \\ (\phi_h)_{|\Omega_i}(x,y) = \gamma_i. \end{cases}$$

Since ϕ_h is continuous at the edge midpoints and since ψ_h is continuous on $\mathbb{T}$, $(\gamma_i)_{i=1\cdots N}$ and $(\beta_i)_{i=1\cdots N}$ do not depend on i. Then, we have

$$\exists(\beta,\gamma)\in\mathbb{R}^2,\ \forall i\in[\![1,N]\!],\ \forall(x,y)\in\mathbb{T}: \quad \begin{cases} \psi_h(x,y) = bx - ay + \beta, \\ \phi_h(x,y) = \gamma. \end{cases}$$

Since $\int_{\mathbb{T}} \phi_h d\mathbf{x} = 0$, we obtain $\phi_h = 0$. Since $\psi_h(x,y) = bx - ay + \beta$ is periodic on $\mathbb{T}$, we have $a = b = 0$ which implies that $\psi_h = \beta$. And since $\int_{\mathbb{T}} \psi_h d\mathbf{x} = 0$, we obtain $\psi_h = 0$. The conclusion is that $\left(\begin{pmatrix} a \\ b \end{pmatrix}, \psi_h, \phi_h \right) = 0$ and the function L in injective.

To prove surjectivity, we prove that $\dim\left(\mathbb{R}^2 \times V_h^0 \times W_h^0\right) = \dim\left(\mathbb{R}^{2N}\right) = 2N$. Any function $\psi_h \in V_h$ is completely and uniquely determined by its values at the V independent nodes of the mesh, which implies that $\dim V_h = V$. Moreover the vanishing mean-value of ψ_h implies a constraint that links the values on the various nodes. Thus, we have $\dim V_h^0 = V - 1$. On the other hand, any $\phi_h \in W_h$ is completely and uniquely determined by its values at the E independent edge midpoints of the mesh, then $\dim W_h = E$. Moreover, the vanishing mean-value of ϕ_h implies a constraint that links the values on the various edges. Thus, we have $\dim W_h^0 = E - 1$. To summarize, we have

$$\dim\left(\mathbb{R}^2 \times V_h^0 \times W_h^0\right) = 2 + \dim(V_h^0) + \dim(W_h^0) = 2 + (V-1) + (E-1) = V + E.$$

Now, in a triangular periodic mesh of a rectangular domain with no internal holes, it is well known that $E + V = 2N$ (proof by recurrence on the number of cells using the Descartes-Euler formula for a periodic domain), which proves the bijectivity of the function L. □

Corollary 3.6. *We have*

$$\left(\mathcal{E}_\alpha^{h,\triangle}\right)^\perp = \left\{ q_h := \begin{pmatrix} r_i \\ \boldsymbol{u}_i \end{pmatrix}_i \in \mathbb{R}^{3N} \middle| \sum_i |\Omega_i| \alpha_i r_i = 0 \text{ and } \exists \phi_h \in W_h, \forall i \in [\![1,N]\!], \boldsymbol{u}_i = (\nabla \phi_h)_{|\Omega_i} \right\}.$$

Let us underline that Lemma 3.5 with Corollary 3.6 is the discrete version of Lemma 2.2 on a triangular mesh.

A first explanation of the satisfying behavior of the Godunov scheme on triangular meshes Here, we prove that, on triangular meshes, the kernel of the Godunov scheme corresponds exactly to the discretized space $\mathcal{E}_\alpha^{\triangle}$. This property shows that the discrete stationary space discretizes well the continuous one. This gives a (partial) explanation of the satisfactory behavior of the Godunov scheme on a triangular mesh.

Proposition 3.7. *Assume that $(\Omega_i)_{i=1\cdots N}$ is a triangular periodic mesh of a rectangular domain with no internal holes. We have*

$$Ker\, \mathbb{L}_{\kappa=1,\alpha}^{h,\triangle} = \mathcal{E}_\alpha^{h,\triangle}.$$

Proof. We firstly prove that $\mathcal{E}_\alpha^{h,\triangle} \subset \mathrm{Ker}\, \mathbb{L}_{\kappa=1,\alpha}^{h}$. Let $q_h \in \mathcal{E}_\alpha^{h,\triangle}$. There exists $(a,b,c) \in \mathbb{R}^3$ and $\psi_h \in V_h$ such that (see (24))

$$\forall i \in [\![1,N]\!]: \quad \begin{cases} r_i = c, \\ (\alpha \mathbf{u})_i = \begin{pmatrix} a \\ b \end{pmatrix} + (\nabla \times \psi_h)_{|\Omega_i} \end{cases}$$

that is to say

$$\forall i \in [\![1,N]\!]: \quad (\alpha \mathbf{u})_i \cdot \mathbf{n}_{ij} = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \mathbf{n}_{ij} + (\nabla \times \psi_h)_{|\Omega_i} \cdot \mathbf{n}_{ij} = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \mathbf{n}_{ij} + (\nabla \psi_h)_{|\Omega_i} \cdot \mathbf{t}_{ij}$$

where $\mathbf{t}_{ij}$ is a unit vector such that $(\mathbf{n}_{ij}, \mathbf{t}_{ij})$ is a direct orthonormal system. As already explained in the proof of Lemma 3.5, the fact that ψ_h is a P^1 Lagrange function implies that $(\nabla \psi_h)_{|\Omega_i} \cdot \mathbf{t}_{ij} = (\nabla \psi_h)_{|\Omega_j} \cdot \mathbf{t}_{ij}$ and as a consequence the continuity of $(\alpha \mathbf{u}) \cdot \mathbf{n}$ through the cell edges. This means that $q_h \in \mathrm{Ker}\, \mathbb{L}_{\kappa=1,\alpha}^{h,\triangle}$ by using (21). Now, we prove that $\mathrm{Ker}\, \mathbb{L}_{\kappa=1,\alpha}^{h,\triangle} \subset \mathcal{E}_\alpha^{h,\triangle}$. Let $q_h \in \mathrm{Ker}\, \mathbb{L}_{\kappa=1,\alpha}^{h,\triangle}$. Since $q_h \in \mathrm{Ker}\, \mathbb{L}_{\kappa=1,\alpha}^{h,\triangle}$, there exists $c \in \mathbb{R}$ such that for all i, $r_i = c$. With Lemma 3.5, we have

$$\mathbf{u}_i = \left(\frac{1}{\alpha_i} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha_i} (\nabla \times \psi_h)_{|\Omega_i} \right) + (\nabla \phi_h)_{|\Omega_i}$$

for some $(a, b, \psi_h, \phi_h) \in \mathbb{R}^2 \times V_h \times W_h$, this decomposition being orthogonal. Thus, we just have to prove that $\nabla \phi_h = 0$. By orthogonality, we have

$$\begin{aligned} \sum_i |\Omega_i| \alpha_i \left\| (\nabla \phi_h)_{|\Omega_i} \right\|^2 &= \langle \mathbf{u}, (\nabla \phi_h) \rangle_{\alpha,h} \\ &= \sum_i |\Omega_i| (\alpha \mathbf{u})_i \cdot (\nabla \phi_h)_{|\Omega_i} = \sum_i (\alpha \mathbf{u})_i \cdot \int_{\Omega_i} (\nabla \phi_h)_{|\Omega_i} d\mathbf{x} \end{aligned}$$

because $\alpha\mathbf{u}$ and $\nabla\phi_h$ are constant on each triangle Ω_i. Then, we can write

$$\begin{aligned}\sum_i |\Omega_i|\alpha_i \left\|(\nabla\phi_h)_{|\Omega_i}\right\|^2 &= \sum_i (\alpha\mathbf{u})_i \cdot \int_{\partial\Omega_i} (\phi_h)_{|\Omega_i}\mathbf{n}_{ij}d\sigma = \sum_i \int_{\partial\Omega_i} (\alpha\mathbf{u})_i \cdot (\phi_h)_{|\Omega_i}\mathbf{n}_{ij}d\sigma \\ &= \sum_{\Gamma_{ij}} \int_{\Gamma_{ij}} \left[(\phi_h)_{|\Omega_i}(\alpha\mathbf{u})_i \cdot \mathbf{n}_{ij} + (\phi_h)_{|\Omega_j}(\alpha\mathbf{u})_j \cdot \mathbf{n}_{ji}\right] d\sigma.\end{aligned}$$

Since $q_h \in \operatorname{Ker} \mathbb{L}^{h,\triangle}_{\kappa=1,\alpha}$, we have $(\alpha\mathbf{u})_i \cdot \mathbf{n}_{ij} = (\alpha\mathbf{u})_j \cdot \mathbf{n}_{ij}$ and we denote by $(\alpha u_n)_{ij}$ this common value. Thus

$$\sum_i |\Omega_i|\alpha_i \left\|(\nabla\phi_h)_{|\Omega_i}\right\|^2 = \sum_{\Gamma_{ij}} (\alpha u_n)_{ij} \cdot \int_{\Gamma_{ij}} [\phi_h]_{ij} d\sigma,$$

where $[\phi_h]_{ij}$ denotes the jump of ϕ_h through the edge Γ_{ij}. As already explained in the proof of Lemma 3.5, the fact that ϕ_h is a P^1 function, which is continuous at the edge midpoints implies that $\int_{\Gamma_{ij}} [\phi_h]_{ij} d\sigma = 0$ on any edge, which proves

$$\sum_i |\Omega_i|\alpha_i \left\|(\nabla\phi_h)_{|\Omega_i}\right\|^2 = 0$$

that is to say for all $i \in [\![1, N]\!]$, $(\nabla\phi_h)_{|\Omega_i} = 0$. This proves that $q_h \in \mathcal{E}^{h,\triangle}_\alpha$. □

3.3.2 Kernel on a Cartesian mesh

We now study some particular properties of the behavior of the Godunov scheme on a rectangular uniform Cartesian mesh. Especially, we study the relation between the kernels of the standard Godunov scheme ($\kappa = 1$) and of its modification ($\kappa = 0$) on a uniform Cartesian mesh and a discrete version of the space $\mathcal{E}_\alpha$ defined by (12).

Construction of $\mathcal{E}^{h,\square}_\alpha$ and $\left(\mathcal{E}^{h,\square}_\alpha\right)^\perp$ We construct an accurate discrete version of the well-prepared subspace $\mathcal{E}_\alpha$ defined by (12). Suppose that the computational domain is a rectangle and that the mesh is made up of $N_x \times N_y$ rectangles of constant size $\Delta x \times \Delta y$ where N_x and N_y are the numbers of cells in the x and y directions. In what follows, we shall suppose that both N_x and N_y are odd. Indeed, if this is not the case, the situation is a little more involved due to even/odd decoupling which may produce checkerboard modes. We introduce the following operators, which are accurate approximations of their continuous counterparts:

$$\mathbf{rot}_{2h} : \mathbb{R}^{N_xN_y} \mapsto \left(\mathbb{R}^{N_xN_y}\right)^2 \text{ with } (\mathbf{rot}_{2h}\psi)_{i,j} := \begin{pmatrix} \dfrac{\psi_{i,j+1} - \psi_{i,j-1}}{2\Delta y} \\ -\dfrac{\psi_{i+1,j} - \psi_{i-1,j}}{2\Delta x} \end{pmatrix},$$

$$\mathbf{grad}_{2h} : \mathbb{R}^{N_xN_y} \mapsto \left(\mathbb{R}^{N_xN_y}\right)^2 \text{ with } (\mathbf{grad}_{2h}\psi)_{i,j} := \begin{pmatrix} \dfrac{\psi_{i+1,j} - \psi_{i-1,j}}{2\Delta x} \\ \dfrac{\psi_{i,j+1} - \psi_{i,j-1}}{2\Delta y} \end{pmatrix}$$

In these definitions, it is implicitly meant that $(\psi_{i,j}) \in \mathbb{R}^{N_xN_y}$ is periodic, that is to say

$$\begin{cases} \forall i \in [\![1, N_x]\!], & \psi_{i,0} = \psi_{i,N_y} \quad \text{and} \quad \psi_{i,N_y+1} = \psi_{i,1}, \\ \forall j \in [\![1, N_y]\!], & \psi_{0,j} = \psi_{N_x,j} \quad \text{and} \quad \psi_{N_x+1,j} = \psi_{1,j}. \end{cases} \tag{28}$$

Let us now define the following subspace, which is an accurate discrete version of $\mathcal{E}_\alpha$ defined by (12).

$$\mathcal{E}^{h,\square}_\alpha = \left\{ q_h := \begin{pmatrix} r_{i,j} \\ \mathbf{u}_{i,j} \end{pmatrix} \in \mathbb{R}^{3N_xN_y} \middle| \exists\, (a, b, c, (\psi_{i,j})) \in \mathbb{R}^3 \times \mathbb{R}^{N_xN_y},\ \forall (i, j) \in [\![1, N_x]\!] \times [\![1, N_y]\!], \right.$$
$$\left. r_{i,j} = c \text{ and } (\alpha\mathbf{u})_{i,j} = \begin{pmatrix} a \\ b \end{pmatrix} + (\mathbf{rot}_{2h}\psi)_{i,j} \right\}. \tag{29}$$

We shall also need the following weighted discrete scalar product:

$$\langle (q_h)_1, (q_h)_2 \rangle_{\alpha,h} = \sum_{i,j} |\Omega_{i,j}|\alpha_{i,j} \left[(r_1)_{i,j}(r_2)_{i,j} + (\mathbf{u}_1)_{i,j} \cdot (\mathbf{u}_2)_{i,j}\right]. \tag{30}$$

We introduce in the following lemma a discrete Hodge decomposition for a collocated Cartesian mesh with periodic boundary conditions. The orthogonality is to be understood with respect to the discrete scalar product defined by (30). Nicolaides [31] also proved some kind of similar result but did not consider periodic boundary conditions, weighted spaces and collocated meshes (he did the proof for a staggered mesh). The proof presented here does not use the same techniques as Nicolaides'.

Lemma 3.8. *For any* $(r, \boldsymbol{u})^T \in \mathbb{R}^{3N_xN_y}$, *there exists unique* $(a, b, (\psi_{i,j}), (\phi_{i,j}))) \in \mathbb{R}^2 \times \mathbb{R}^{N_xN_y} \times \mathbb{R}^{N_xN_y}$, *with* $\sum_{i,j}|\Omega_{i,j}|\psi_{i,j} = \sum_{i,j}|\Omega_{i,j}|\phi_{i,j} = 0$, *such that for all* $(i,j) \in [\![1, N_x]\!] \times [\![1, N_y]\!]$,

$$\begin{pmatrix} r_{i,j} \\ \boldsymbol{u}_{i,j} \end{pmatrix} = \begin{pmatrix} \bar{r} \\ \frac{1}{\alpha_{i,j}} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha_{i,j}} (\boldsymbol{rot}_{2h}\psi)_{i,j} \end{pmatrix} + \begin{pmatrix} r_{i,j} - \bar{r} \\ (\boldsymbol{grad}_{2h}\phi)_{i,j} \end{pmatrix} \tag{31}$$

with $\bar{r} = \dfrac{\sum_{i,j}|\Omega_{i,j}|\alpha_{i,j}r_{i,j}}{\sum_{i,j}|\Omega_{i,j}|\alpha_{i,j}}$. *Moreover this decomposition is orthogonal for the scalar product* (30).

Proof. Let us first prove orthogonality. The orthogonality between $\bar{r}$ and $r - \bar{r}$ is obvious. Now we prove the orthogonality for the decomposition of $\mathbf{u}$. We have, for any $(a, b)^T \in \mathbb{R}^2$ and periodic sequence $(\phi_{i,j})_{i,j} \in \mathbb{R}^{N_xN_y}$ in the sense of (28)

$$\begin{aligned} \left\langle \frac{1}{\alpha} \begin{pmatrix} a \\ b \end{pmatrix}, (\mathbf{grad}_{2h}\phi) \right\rangle_{\alpha,h} &= \sum_{i,j} |\Omega_{i,j}|\alpha_{i,j} \frac{1}{\alpha_{i,j}} \begin{pmatrix} a \\ b \end{pmatrix} \cdot (\mathbf{grad}_{2h}\phi)_{i,j} \\ &= \frac{1}{2} \begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} \Delta y \sum_{i,j} (\phi_{i+1,j} - \phi_{i-1,j}) \\ \Delta x \sum_{i,j} (\phi_{i,j+1} - \phi_{i,j-1}) \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0 \end{aligned}$$

because of (28).
Moreover, for any $\psi_{i,j} \in \mathbb{R}^{N_xN_y}$ and $\phi_{i,j} \in \mathbb{R}^{N_xN_y}$ periodic in the sense of (28),

$$\begin{aligned} &\left\langle \frac{1}{\alpha}\mathbf{rot}_{2h}\psi, \mathbf{grad}_{2h}\phi \right\rangle_{\alpha,h} \\ &= \frac{1}{4}\sum_{i,j} (\psi_{i,j+1} - \psi_{i,j-1})(\phi_{i+1,j} - \phi_{i-1,j}) - (\psi_{i+1,j} - \psi_{i-1,j})(\phi_{i,j+1} - \phi_{i,j-1}) = 0 \end{aligned}$$

because of the periodicity of $(\phi_{i,j})$ and (ψ_{ij}). Then, orthogonality of the decomposition is proved.

We shall now prove existence and uniqueness of decomposition (31). For r there is no problem, so we only consider the equation in $\mathbf{u}$. We have to prove that the function L defined by

$$\begin{aligned} L :\ \mathbb{R}^2 \times \mathbb{R}_0^{N_xN_y} \times \mathbb{R}_0^{N_xN_y} &\to (\mathbb{R}^2)^{N_xN_y} \\ \left(\begin{pmatrix} a \\ b \end{pmatrix}, (\psi_{i,j}), (\phi_{i,j}) \right) &\mapsto \left(\frac{1}{\alpha} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha}(\mathbf{rot}_{2h}\psi) + (\mathbf{grad}_{2h}\phi) \right)_{i,j} \end{aligned} \tag{32}$$

is bijective, where $\mathbb{R}_0^{N_xN_y} = \left\{ (\psi_{ij}) \in \mathbb{R}^{N_xN_y} \;\middle|\; \sum_{i,j}|\Omega_{i,j}|\psi_{i,j} = 0 \right\}$. Firstly, we prove injectivity. As L is a linear function, we just have to prove that

$$L\left(\begin{pmatrix} a \\ b \end{pmatrix}, (\psi_{i,j}), (\phi_{i,j}) \right) = 0 \quad \Longrightarrow \quad \begin{cases} \begin{pmatrix} a \\ b \end{pmatrix} = 0, \\ \forall (i,j) \in [\![1, N_x]\!] \times [\![1, N_y]\!],\ \psi_{i,j} = \phi_{i,j} = 0. \end{cases}$$

Assume that for all $(i,j) \in [\![1, N_x]\!] \times [\![1, N_y]\!]$,

$$\left(\frac{1}{\alpha} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha}(\mathbf{rot}_{2h}\psi) + (\mathbf{grad}_{2h}\phi) \right)_{i,j} = 0.$$

By the orthogonality property proved above, this implies

$$\forall (i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!], \quad \begin{pmatrix} a \\ b \end{pmatrix} + (\mathbf{rot}_{2h}\psi)_{i,j} = 0 \quad \text{and} \quad (\mathbf{grad}_{2h}\phi)_{i,j} = 0$$

$$\implies \quad \forall (i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!], \quad \begin{cases} \psi_{i,j+1} - \psi_{i,j-1} = -2a\Delta y, \\ \psi_{i+1,j} - \psi_{i-1,j} = 2b\Delta x, \\ \phi_{i+1,j} - \phi_{i-1,j} = 0, \\ \phi_{i,j+1} - \phi_{i,j-1} = 0. \end{cases}$$

Then, for all $i \in [\![1,N_x]\!]$, $(\psi_{i,j})_{2j}$ is an arithmetic sequence of step $-2a\Delta y$. By periodicity, we deduce that $a = 0$. We obtain that $\psi_{i,j+1} = \psi_{i,j-1}$ for all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$. Then, because N_y is odd, this implies that

$$\forall i \in [\![1,N_x]\!],\ \exists \beta_i \in \mathbb{R},\ \forall j \in [\![1,N_y]\!],\ \psi_{i,j} = \beta_i.$$

Note that if N_y were not odd, there would be an even/odd decoupling here (there would exist constants β_i^{odd} and β_i^{even} such that $\psi_{i,2j} = \beta_i^{\text{even}}$ and $\psi_{i,2j+1} = \beta_i^{\text{odd}}$).

In the same way, it holds that $b = 0$ and that

$$\forall j \in [\![1,N_y]\!],\ \exists \gamma_j \in \mathbb{R},\ \forall i \in [\![1,N_x]\!],\ \psi_{i,j} = \gamma_j.$$

Both equalities on $\psi_{i,j}$ can happen simultaneously only if the values do not depend on i and j, and thus $\psi_{i,j}$ is constant. Since $\sum_{i,j} \Delta x \Delta y \psi_{i,j} = 0$ we obtain

$$\forall (i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!],\ \psi_{i,j} = 0.$$

Similarly, we obtain for all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$, $\phi_{i,j} = 0$ and the function L is injective. Moreover, injectivity and the following space dimension equality ensure bijectivity:

$$\dim\left(\mathbb{R}^2 \times \mathbb{R}_0^{N_x N_y} \times \mathbb{R}_0^{N_x N_y}\right) = 2 + 2\dim\left(\mathbb{R}_0^{N_x N_y}\right) = 2 + 2(N_x N_y - 1) = 2N_x N_y.$$

□

A first explanation of the wrong behavior of the Godunov scheme on a Cartesian mesh
In this section, we show that the kernel of the standard ($\kappa = 1$) Godunov scheme is not an accurate approximation of the kernel of the continuous wave equation. On the other hand, we show that the kernel of the modified ($\kappa = 0$) Godunov scheme does approximate correctly the kernel of the continuous wave equation.

Proposition 3.9. *Assume that $(\Omega_{i,j})_{i=1\cdots N_x,\, j=1\cdots N_y}$ is a Cartesian periodic mesh of a rectangular domain with no internal holes. We have*

$$Ker\,\mathbb{L}^{h,\square}_{\kappa>0,\alpha} \subsetneq \mathcal{E}^{h,\square}_{\alpha}$$

with

$$Ker\,\mathbb{L}^{h,\square}_{\kappa>0,\alpha} = \left\{ q_h := \begin{pmatrix} r_i \\ \boldsymbol{u_i} \end{pmatrix} \in \mathbb{R}^{3N_xN_y} \middle| \exists\, (c,(a_j),(b_i)) \in \mathbb{R} \times \mathbb{R}^{N_y} \times \mathbb{R}^{N_x}, \right.$$
$$\left. \forall (i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!], \quad r_{i,j} = c \text{ and } (\alpha \boldsymbol{u})_{i,j} = \begin{pmatrix} a_j \\ b_i \end{pmatrix} \right\}. \quad (33)$$

On the other hand, we have

$$Ker\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha} = \mathcal{E}^{h,\square}_{\alpha}.$$

Proof. Starting from (21), equality (33) is readily obtained. Indeed, considering vertical edges of the Cartesian rectangular mesh, (21) implies that $(\alpha u_x)_{i,j}$ is constant along the x-direction; then this quantity depends only on j. In the same way, considering horizontal edges, it follows that $(\alpha u_y)_{i,j}$ is constant along the y-direction, and thus depends only on i. We clearly see that $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa>0,\alpha}$ is a very poor approximation of its discrete counterpart, since it contains only velocity fields whose horizontal (resp. vertical) component depends only (up to the factor α) on the vertical (resp. horizontal) coordinate.

Concerning the inclusion $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa>0,\alpha} \subsetneq \mathcal{E}^{h,\square}_{\alpha}$, it will be a consequence of the proof that $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha} = \mathcal{E}^{h,\square}_{\alpha}$ since Lemma 3.4 implies that $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa>0,\alpha} \subsetneq \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha}$.

Let us now turn to the case $\kappa = 0$. We first prove that $\mathcal{E}^{h,\square}_{\alpha} \subset \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha}$. Let $q_h \in \mathcal{E}^{h,\square}_{\alpha}$. There exist $(a,b,c) \in \mathbb{R}^3$ and $(\psi_{i,j}) \in \mathbb{R}^{N_xN_y}$ such that

$$\forall (i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!], \quad r_{i,j} = c \quad \text{and} \quad (\alpha \mathbf{u})_{i,j} = \begin{pmatrix} a \\ b \end{pmatrix} + (\mathbf{rot}_{2h}\psi)_{i,j}.$$

Denoting by $\Omega_{(k,\ell)}$ one of the four neighbors of $\Omega_{i,j}$ and $\mathbf{n}_{(i,j)(k,\ell)}$ the corresponding unit exterior normal vector on their interface, we have for all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$,

$$\begin{aligned}
&\sum_{\Gamma_{(i,j)(k,\ell)} \subset \partial\Omega_{i,j}} |\Gamma_{(i,j)(k,\ell)}| \left((\alpha\mathbf{u})_{i,j} + (\alpha\mathbf{u})_{k,\ell} \right) \cdot \mathbf{n}_{(i,j)(k,\ell)} \\
= \;& \Delta y \left((\alpha\mathbf{u})_{i,j} + (\alpha\mathbf{u})_{i+1,j} \right) \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \Delta y \left((\alpha\mathbf{u})_{i,j} + (\alpha\mathbf{u})_{i-1,j} \right) \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\
& + \Delta x \left((\alpha\mathbf{u})_{i,j} + (\alpha\mathbf{u})_{i,j+1} \right) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Delta x \left((\alpha\mathbf{u})_{i,j} + (\alpha\mathbf{u})_{i,j-1} \right) \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} \\
= \;& \Delta y \left((\alpha\mathbf{u})_{i+1,j} - (\alpha\mathbf{u})_{i-1,j} \right) \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \Delta x \left((\alpha\mathbf{u})_{i,j+1} - (\alpha\mathbf{u})_{i,j-1} \right) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
= \;& a\Delta y + \frac{\psi_{i+1,j+1} - \psi_{i+1,j-1}}{2} - \left(a\Delta y + \frac{\psi_{i-1,j+1} - \psi_{i-1,j-1}}{2} \right) \\
& + \left(b\Delta x - \frac{\psi_{i+1,j+1} - \psi_{i-1,j+1}}{2} \right) - \left(b\Delta x - \frac{\psi_{i+1,j-1} - \psi_{i-1,j-1}}{2} \right) \\
= \;& 0
\end{aligned}$$

and then $q_h \in \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha}$.

Now, we prove that $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha} \subset \mathcal{E}^{h,\square}_{\alpha}$. Let $q_h \in \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha}$. This first implies that there exists $c \in \mathbb{R}$ such that for all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$, $r_{i,j} = c$. Next, using Lemma 3.8, we can write

$$\forall (i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!] \quad , \quad \mathbf{u}_{i,j} = \frac{1}{\alpha_{i,j}} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha_{i,j}} (\mathbf{rot}_{2h}\psi)_{i,j} + (\mathbf{grad}_{2h}\phi)_{i,j} \tag{34}$$

for some $(a,b,(\psi_{i,j}),(\phi_{i,j})) \in \mathbb{R}^2 \times \left(\mathbb{R}^{N_xN_y}\right)^2$ and this decomposition is orthogonal. We need to prove that $(\mathbf{grad}_{2h}\phi)_{i,j} = 0$ for all (i,j). By orthogonality in (34), we have

$$\sum_{i,j} \Delta x \Delta y \alpha_{i,j} \left| (\mathbf{grad}_{2h}\phi)_{i,j} \right|^2 = \langle \mathbf{u}, \mathbf{grad}_{2h}\phi \rangle_{\alpha,h} = \sum_{i,j} \Delta x \Delta y \alpha_{i,j} \mathbf{u}_{i,j} \cdot \begin{pmatrix} \dfrac{\phi_{i+1,j} - \phi_{i-1,j}}{2\Delta x} \\ \dfrac{\phi_{i,j+1} - \phi_{i,j-1}}{2\Delta y} \end{pmatrix}.$$

Expanding the dot product and rearranging the sum through changes of indexes in order to factorize by $\phi_{i,j}$, we obtain, using periodicity to handle the boundary terms and the fact that, for any given cell $\Omega_{i,j}$ we have $\sum_{\Gamma_{(i,j)(k,\ell)} \subset \partial\Omega_{i,j}} |\Gamma_{(i,j)(k,\ell)}| \, \mathbf{n}_{(i,j)(k,\ell)} = 0$:

$$\begin{aligned}
\|(\mathbf{grad}_{2h}\phi)\|^2_{\alpha,h} &= \frac{1}{2} \sum_{i,j} \big(\Delta y \left[(\alpha u_x)_{i-1,j} - (\alpha u_x)_{i+1,j} \right] + \Delta x \left[(\alpha u_y)_{i,j-1} - (\alpha u_y)_{i,j+1} \right] \big) \phi_{i,j} \\
&= \frac{1}{2} \sum_{i,j} \phi_{i,j} \sum_{\Gamma_{(i,j)(k,\ell)} \subset \partial\Omega_{i,j}} |\Gamma_{(i,j)(k,\ell)}| \, (\alpha\mathbf{u})_{k,\ell} \cdot \mathbf{n}_{(i,j)(k,\ell)} \\
&= \frac{1}{2} \sum_{i,j} \phi_{i,j} \sum_{\Gamma_{(i,j)(k,\ell)} \subset \partial\Omega_{i,j}} |\Gamma_{(i,j)(k,\ell)}| \left((\alpha\mathbf{u})_{i,j} + (\alpha\mathbf{u})_{k,\ell} \right) \cdot \mathbf{n}_{(i,j)(k,\ell)} = 0,
\end{aligned}$$

because $q_h \in \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=0,\alpha}$. This means that for all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$, $(\mathbf{grad}_{2h}\phi)_{i,j} = 0$ and then (34) gives us for all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$,

$$\mathbf{u}_{i,j} = \frac{1}{\alpha_{i,j}} \begin{pmatrix} a \\ b \end{pmatrix} + \frac{1}{\alpha_{i,j}} (\mathbf{rot}_{2h}\psi)_{i,j}.$$

Then, we have $q_h \in \mathcal{E}^{h,\square}_{\alpha}$. □

On Cartesian meshes, this proves that deleting the diffusion term on the velocity field ($\kappa = 0$) allows to recover a kernel that is an accurate approximation of its continuous counterpart.

4 Right or wrong behavior of the Godunov scheme in the linear discrete case

We now study the *low Mach accuracy* of the Godunov scheme in the sense that the numerical solution (19) satisfies a discrete version of (17). As explained in subsection 3.2, the two key points to prove this kind of property is that the kernel of the scheme satisfies $\mathcal{E}_\alpha^h \subset \operatorname{Ker} \mathbb{L}_\alpha^h$ and that system (19) is well-posed (see Theorem 3.3). The study of the kernel was performed in section 3, $\kappa > 0$ and $\kappa = 0$ define discrete operators $\mathbb{L}_{\kappa,\alpha}$ whose kernels are very different. We now prove the well-posed property (l^2_α-stability) and then study the *low Mach accuracy* of the Godunov scheme on Cartesian rectangular and on triangular meshes.

4.1 l^2_α-stability of the Godunov scheme

We now prove stability of the semi-discrete scheme both when $\kappa = 0$ and when $\kappa > 0$. This property is essential in the sequel. Let us define the energy

$$E_h := \|q_h\|^2_{l^2_\alpha} = \sum_i |\Omega_i| \alpha_i \left(r_i^2 + |\mathbf{u}_i|^2\right). \tag{35}$$

Theorem 4.1. *Let $(r_i, \boldsymbol{u}_i)$ be the solution of the semi-discrete scheme* (18)*. We have*

$$\frac{d}{dt} E_h = \frac{d}{dt} \|q_h\|^2_{l^2_\alpha} = -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left(\alpha_{ij} |r_i - r_j|^2 + \frac{\kappa}{\alpha_{ij}} \left| \left((\alpha \boldsymbol{u})_i - (\alpha \boldsymbol{u})_j \right) \cdot \boldsymbol{n}_{ij} \right|^2 \right). \tag{36}$$

Then, for $\kappa \geq 0$ the Godunov scheme is dissipative since

$$\frac{d}{dt} E_h = \frac{d}{dt} \|q_h\|^2_{l^2_\alpha} \leq 0.$$

Proof. We multiply the first equation of (18) with $2|\Omega_i| r_i$ and sum with respect to i. Since α_i does not depend on time, we obtain

$$\begin{aligned}
\frac{d}{dt} \sum_i |\Omega_i| \alpha_i r_i^2 &= -\frac{a_\star}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left(\left((\alpha \mathbf{u})_i + (\alpha \mathbf{u})_j \right) \cdot \mathbf{n}_{ij} r_i + \alpha_{ij} (r_i - r_j) r_i \right), \\
&= -\frac{a_\star}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left((\alpha \mathbf{u})_j \cdot \mathbf{n}_{ij} r_i + \alpha_{ij} (r_i - r_j) r_i \right), \\
&= -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left((\alpha \mathbf{u})_j \cdot \mathbf{n}_{ij} r_i + (\alpha \mathbf{u})_i \cdot \mathbf{n}_{ji} r_j + \alpha_{ij} (r_i - r_j) r_i + \alpha_{ji} (r_j - r_i) r_j \right), \\
&= -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left(\left(r_i (\alpha \mathbf{u})_j - r_j (\alpha \mathbf{u})_i \right) \cdot \mathbf{n}_{ij} + \alpha_{ij} |r_i - r_j|^2 \right).
\end{aligned}$$

Taking the scalar product of the second equation of (18) with $2|\Omega_i| \mathbf{u}_i$ and summing with respect to i, we obtain

$$\begin{aligned}
\frac{d}{dt} \sum_i |\Omega_i| \alpha_i |\mathbf{u}_i|^2 &= -\frac{a_\star}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \alpha_i \left(r_i + r_j + \frac{\kappa}{\alpha_{ij}} \left((\alpha \mathbf{u})_i - (\alpha \mathbf{u})_j \right) \cdot \mathbf{n}_{ij} \right) \mathbf{u}_i \cdot \mathbf{n}_{ij}, \\
&= -\frac{a_\star}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left(r_j + \frac{\kappa}{\alpha_{ij}} \left((\alpha \mathbf{u})_i - (\alpha \mathbf{u})_j \right) \cdot \mathbf{n}_{ij} \right) (\alpha \mathbf{u})_i \cdot \mathbf{n}_{ij}, \\
&= -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left(r_j (\alpha \mathbf{u})_i \cdot \mathbf{n}_{ij} + r_i (\alpha \mathbf{u})_j \cdot \mathbf{n}_{ji} \right) \\
&\quad -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \frac{\kappa}{\alpha_{ij}} \Big(\left((\alpha \mathbf{u})_i - (\alpha \mathbf{u})_j \right) \cdot \mathbf{n}_{ij} (\alpha \mathbf{u})_i \cdot \mathbf{n}_{ij} \\
&\qquad + \left((\alpha \mathbf{u})_j - (\alpha \mathbf{u})_i \right) \cdot \mathbf{n}_{ji} (\alpha \mathbf{u})_j \cdot \mathbf{n}_{ji} \Big), \\
&= -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left(r_j (\alpha \mathbf{u})_i - r_i (\alpha \mathbf{u})_j \right) \cdot \mathbf{n}_{ij} \\
&\quad -\frac{a_\star}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \frac{\kappa}{\alpha_{ij}} \left| \left((\alpha \mathbf{u})_i - (\alpha \mathbf{u})_j \right) \cdot \mathbf{n}_{ij} \right|^2.
\end{aligned}$$

By summing $\frac{d}{dt}\sum_i |\Omega_i|\alpha_i r_i^2$ and $\frac{d}{dt}\sum_i |\Omega_i|\alpha_i |\mathbf{u}_i|^2$, we obtain (36). □

4.2 The triangular mesh case

In Section 2.3.4, we saw that in the continuous setting and for low values of the Mach number, the solution of the continuous system remains close to an incompressible state for any $t > 0$, if this was the case at the initial time $t = 0$. This section will show that this is also the case at the discrete level for the numerical solution of the Godunov scheme applied to the linear wave equation with porosity on triangular meshes. This explains the satisfactory behavior of this scheme on this type of meshes.

We saw that the Godunov scheme ($\kappa = 1$) preserves an incompressible state $q_h^0 \in \mathcal{E}_\alpha^{h,\triangle}$. We want to study the impact of a perturbation of order M in $\left(\mathcal{E}_\alpha^{h,\triangle}\right)^\perp$ on the initial condition. With the orthogonal decomposition proved in Lemma 3.5 and with definition (24), we can define an orthogonal projection

$$\mathbb{P}_\alpha^{h,\triangle} : \quad l^2_\alpha(\mathbb{T})^3 \to \mathcal{E}_\alpha^{h,\triangle}.$$

The theorem that follows expresses the fact that any perturbation of order M in the orthogonal space at the initial time will not affect the solution over time more than its original size:

Theorem 4.2. *Let $q_h(t)$ be an approximate solution of (10) given by the Godunov scheme (19) with the initial condition q_h^0. On triangular meshes with $\kappa = 1$, for all $q_h^0 \in l^2_\alpha(\mathbb{T})^3$, we have*

$$\forall C_1 > 0, \ \left(\left\|q_h^0 - \mathbb{P}_\alpha^{h,\triangle} q_h^0\right\|_{l^2_\alpha} = C_1 M\right) \implies \left(\forall t \geq 0, \quad \left\|q_h - \mathbb{P}_\alpha^{h,\triangle} q_h^0\right\|_{l^2_\alpha}(t) \leq C_1 M\right). \quad (37)$$

Proof. By linearity of $\mathbb{L}^{h,\triangle}_{\kappa=1,\alpha}$, the solution q_h of scheme (19) with initial condition q_h^0 can be written as

$$q_h = q_{h,1} + q_{h,2}$$

where $q_{h,1}$ is the solution of (19) with initial condition $q_{h,1}(t = 0, \mathbf{x}) = (q_h^0 - \mathbb{P}_\alpha^{h,\triangle} q_h^0)(\mathbf{x})$ and $q_{h,2}$ is the solution of (19) with initial condition $q_{h,2}(t = 0, \mathbf{x}) = \mathbb{P}_\alpha^{h,\triangle} q_h^0(\mathbf{x})$. We have

$$\forall t \geq 0, \quad \left\|q_h - \mathbb{P}_\alpha^{h,\triangle} q_h^0\right\|_{l^2_\alpha}(t) = \left\|q_{h,1} + q_{h,2} - \mathbb{P}_\alpha^{h,\triangle} q_h^0\right\|_{l^2_\alpha}(t) \leq \|q_{h,1}\|_{l^2_\alpha}(t) + \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\triangle} q_h^0\right\|_{l^2_\alpha}(t).$$

Because the Godunov scheme (19) is dissipative when $\kappa \geq 0$ (see Theorem 4.1), we have $\|q_{h,1}\|_{l^2_\alpha}(t) \leq \|q_{h,1}\|_{l^2_\alpha}(0)$. Moreover, since $\mathbb{P}_\alpha^{h,\triangle} q_h^0 \in \mathcal{E}_\alpha^{h,\triangle} = \operatorname{Ker} \mathbb{L}^{h,\triangle}_{\kappa=1,\alpha}$, the initial condition for $q_{h,2}$ is in the kernel of $\mathbb{L}^{h,\triangle}_{\kappa=1,\alpha}$, and therefore $q_{h,2}$ is stationary: for all $t > 0$ we have $q_{h,2}(t) = \mathbb{P}_\alpha^{h,\triangle} q_h^0$. We obtain

$$\forall t \geq 0, \quad \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\triangle} q^0\right\|_{l^2_\alpha}(t) = 0$$

and (38) becomes

$$\forall t \geq 0, \quad \left\|q_h - \mathbb{P}_\alpha^{h,\triangle} q_h^0\right\|_{l^2_\alpha}(t) \leq \|q_{h,1}\|_{l^2_\alpha}(0) = C_1 M$$

if $\|q_h^0 - \mathbb{P}_\alpha^{h,\triangle} q_h^0\|_{l^2_\alpha} = C_1 M$. □

4.3 The Cartesian mesh case

We saw that the Godunov scheme ($\kappa = 1$) on Cartesian meshes does not preserve an incompressible state $q_h^0 \in \mathcal{E}_\alpha^{h,\square}$, but it preserves it if we delete the numerical diffusion on the velocity by setting $\kappa = 0$. From Lemma 3.8, we can define an orthogonal projection

$$\mathbb{P}_\alpha^{h,\square} : \quad l^2_\alpha(\mathbb{T})^3 \to \mathcal{E}_\alpha^{h,\square}.$$

We want to study the evolution over time of the initial condition when it consists in the sum of an element in the discrete incompressible space $\left(\mathcal{E}_\alpha^{h,\square}\right)$ and of a perturbation of order M in $\left(\mathcal{E}_\alpha^{h,\square}\right)^\perp$.

This will give an explanation of the wrong behavior of the standard ($\kappa = 1$) Godunov scheme on a Cartesian mesh and of the satisfactory behavior of the modified ($\kappa = 0$) scheme. Moreover, since completely deleting the numerical diffusion by setting $\kappa = 0$ was shown in [15] to present stability issues in the non-linear case, we shall also study the intermediate case $\kappa = M$.

4.3.1 Explanation of the wrong behavior of the Godunov scheme on a Cartesian mesh

The next theorem shows that for the standard Godunov scheme ($\kappa = 1$) on Cartesian meshes, starting from a perturbation of an incompressible field, the numerical solution will substantially deviate from the initial condition after a short time that scales like $\mathcal{O}(M)$, when the space discretization parameters $(\Delta x, \Delta y)$ are larger than the Mach number.

Theorem 4.3. *Let $q_h(t)$ be the solution of the Godunov scheme (19) with initial condition q_h^0 on a Cartesian mesh with discretization parameters $(\Delta x, \Delta y)$. Then, when $\kappa = 1$, there exists $C_2 > 0$, depending only on α, $a_\star$ and on $\mathbb{T}$ such that for almost all $q_h^0 \in l^2_\alpha(\mathbb{T})^3$ and for all $C_1 > 0$, there exists C_3 depending only on (C_1, q_h^0) such that for any $M \leq \frac{C_3}{C_1}\min(\Delta x, \Delta y)$ we have*

$$\left\|q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha} = C_1 M \implies \forall t \geq C_2 M, \ \left\|q_h - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha}(t) \geq C_3 \min(\Delta x, \Delta y). \quad (38)$$

Proof. By linearity of $\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$, the solution q_h of the Godunov scheme (19) with initial condition q_h^0 can be written as

$$q_h = q_{h,1} + q_{h,2}$$

where $q_{h,1}$ is the solution of

$$\begin{cases} \partial_t(\alpha q_{h,1}) + \dfrac{\mathbb{L}^{h,\square}_{\kappa=1,\alpha}}{M}(q_{h,1}) = 0, \\ q_{h,1}(t=0) = q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0 \end{cases} \quad (39)$$

and $q_{h,2}$ is the solution of

$$\begin{cases} \partial_t(\alpha q_{h,2}) + \dfrac{\mathbb{L}^{h,\square}_{\kappa=1,\alpha}}{M}(q_{h,2}) = 0, \\ q_{h,2}(t=0) = \mathbb{P}_\alpha^{h,\square} q_h^0. \end{cases} \quad (40)$$

We have

$$\begin{aligned} \forall t \geq 0, \quad \left\|q_h - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha}(t) &= \left\|q_{h,1} + q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha}(t) \\ &\geq \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha}(t) - \|q_{h,1}\|_{l^2_\alpha}(t) \\ &\geq \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha}(t) - \|q_{h,1}\|_{l^2_\alpha}(0) \\ &\geq \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha}(t) - \left\|q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha} \end{aligned} \quad (41)$$

because scheme (19) is dissipative when $\kappa \geq 0$ (see Theorem 4.1). If $\left\|q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l^2_\alpha} = C_1 M$, then (41) shows that we need to find a lower bound for the function

$$t \mapsto \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q^0\right\|_{l^2_\alpha}(t)$$

where $q_{h,2}$ is the solution of (40). Before proceeding to the detailed proof of this proposition, let us briefly mention the ideas behind it: the initial condition of (40) will be diffused by the operator $\dfrac{\mathbb{L}^{h,\square}_{\kappa=1,\alpha}}{M}$ onto its orthogonal projection in the kernel $\mathrm{Ker}(\mathbb{L}^{h,\square}_{\kappa=1,\alpha})$ (this orthogonal projection is denoted by $\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$ in the sequel) and we shall prove that the solution of (40) will tend to $\mathbb{P}^{h,\square}_{\kappa=1,\alpha}(\mathbb{P}^{h,\square}_\alpha q_h^0)$ exponentially fast with a convergence rate that depends on $\frac{\min(\Delta x, \Delta y)}{M}$. As a consequence, after a time that scales like $\mathcal{O}(M)$, the solution of (40) will be close enough to its projection, and thus far enough from the initial condition. To prove this in detail, we shall follow the lines below:

1. we write
$$q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0 = q_{h,2} - \mathbb{P}^{h,\square}_{\kappa=1,\alpha}\mathbb{P}_\alpha^{h,\square} q_h^0 + \mathbb{P}^{h,\square}_{\kappa=1,\alpha}\mathbb{P}_\alpha^{h,\square} q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0,$$

2. we verify that $\hat{q}_h := q_{h,2} - \mathbb{P}^{h,\square}_{\kappa=1,\alpha}\mathbb{P}_\alpha^{h,\square} q_h^0$ is solution of (19) and that $\hat{q}_h(t) \in \mathrm{Ker}\,\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$, for all $t \geq 0$,

3. we use an energy estimate for solutions of (19) and a discrete Poincaré-Wirtinger inequality for $\hat{q}_h$ that is satisfied on $\mathrm{Ker}\,\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$, to estimate how fast $\hat{q}_h$ tends to 0,
4. we obtain (38) by considering times of order M.

In order to obtain these results, we first prove a series of Lemmas. We start by some considerations on the orthogonal projection onto $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$:

Lemma 4.4. *The operator*

$$\mathbb{P} \,:\, l^2_\alpha(\mathbb{T})^{1+d} \;\to\; Ker\,\mathbb{L}^{h,\square}_{\kappa=1,\alpha} \tag{42}$$

$$q_h = \begin{pmatrix} r_{i,j} \\ \boldsymbol{u}_{i,j} \end{pmatrix}_{i,j} \;\mapsto\; \begin{pmatrix} \dfrac{1}{\sum\limits_{k,\ell}\alpha_{k,\ell}\Delta x\Delta y}\sum\limits_{k,\ell} r_{k,\ell}\alpha_{k,\ell}\Delta x\Delta y \\ \dfrac{1}{\alpha_{i,j}\sum\limits_{k=1}^{N_x}\frac{1}{\alpha_{k,j}}\Delta x}\sum\limits_{k=1}^{N_x}(u_x)_{k,j}\Delta x \\ \dfrac{1}{\alpha_{i,j}\sum\limits_{\ell=1}^{N_y}\frac{1}{\alpha_{i,\ell}}\Delta y}\sum\limits_{\ell=1}^{N_y}(u_y)_{i,\ell}\Delta y \end{pmatrix}_{i,j} \tag{43}$$

is the orthogonal projection $\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$ on $Ker\,\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$. Moreover, if q_h is a solution of (19) on $\mathbb{T}$ with initial condition q^0_h, then:

$$q^0_h \in Ker\,\mathbb{P}^{h,\square}_{\kappa=1,\alpha} \Longrightarrow q_h(t\geq 0) \in Ker\,\mathbb{P}^{h,\square}_{\kappa=1,\alpha}. \tag{44}$$

Proof. Recalling that $\mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$ is characterized by (33), it is first straightforward to see that $\mathbb{P}(q_h) \in \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$. Next, lengthy but easy algebra leads to $\langle (q_h - \mathbb{P}q_h), s_h\rangle_{\alpha,h} = 0$ for all $s_h \in \mathrm{Ker}\,\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$. These two properties exactly prove that $\mathbb{P} = \mathbb{P}^{h,\square}_{\kappa=1,\alpha}$.

Moreover, when $q^0_h \in \mathrm{Ker}\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$, then (43) implies that $s(0) = s_{x,j}(0) = s_{y,i}(0) = 0$ for all (i,j) where

$$s(t) := \sum_{k,\ell} r_{k,\ell}(t)\alpha_{k,\ell}\Delta x\Delta y \;\;,\;\; s_{x,j}(t) := \sum_{k=1}^{N_x}(u_x)_{k,j}(t)\Delta x \;\;,\;\; s_{y,i}(t) := \sum_{\ell=1}^{N_y}(u_y)_{i,\ell}(t)\Delta y.$$

Then it suffices to prove that $\frac{ds}{dt}(t) = \frac{ds_{x,j}}{dt}(t) = \frac{ds_{y,i}}{dt}(t) = 0$ for all time and all (i,j) to obtain (44). As far as s is concerned, this is a direct consequence of the conservativity of fluxes in the first equation of (18). As far as $s_{x,j}$ is concerned, extracting the x component of (18) and specializing to a Cartesian mesh, we get

$$\frac{d(u_x)_{i,j}}{dt} + \frac{a_\star}{2M\Delta x}(F_{i,i+1,j} - F_{i,i-1,j}) = 0$$

with $F_{i,i+1,j} = (r_{i,j} + r_{i+1,j}) + \frac{1}{\alpha_{i+\frac{1}{2},j}}(\alpha_{i,j}(u_x)_{i,j} - \alpha_{i+1,j}(u_x)_{i+1,j})$ and $F_{i,i-1,j} = (r_{i,j} + r_{i-1,j}) - \frac{1}{\alpha_{i-\frac{1}{2},j}}(\alpha_{i,j}(u_x)_{i,j} - \alpha_{i-1,j}(u_x)_{i-1,j})$, where we recall that $\alpha_{i-\frac{1}{2},j}$ is the value of the porosity at the interface between cells $\Omega_{i-1,j}$ and $\Omega_{i,j}$. Multiplying the equality above with (Δx), summing over i and noting that $F_{i-1,i,j} = F_{i,i-1,j}$, we obtain with periodic boundary conditions that $\frac{ds_{x,j}}{dt}(t) = 0$ for all $t \geq 0$ and all j. The same kind of proof applies for the vertical component $s_{y,i}$. □

We now write a discrete Poincaré-Wirtinger inequality for a function $q_h \in \mathrm{Ker}\,\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$.

Lemma 4.5. *There exists a constant $K_\alpha(\mathbb{T}) > 0$ depending on $\mathbb{T}$ and α such that for any $q_h := (r, u_x, u_y)^T_h \in Ker\,\mathbb{P}^{h,\square}_{\kappa=1,\alpha}$*

$$\|q_h\|^2_{l^2_\alpha} \leq \frac{K_\alpha(\mathbb{T})}{\min(\Delta x,\Delta y)}\left[\sum_{i,j}\Delta x\alpha_{i,j-\frac{1}{2}}\left|r_{i,j} - r_{i,j-1}\right|^2 + \sum_{i,j}\Delta y\alpha_{i-\frac{1}{2},j}\left|r_{i,j} - r_{i-1,j}\right|^2\right.$$
$$\left. + \sum_{i,j}\Delta y\left|(\alpha u_x)_{i,j} - (\alpha u_x)_{i-1,j}\right|^2\frac{1}{\alpha_{i-\frac{1}{2},j}} + \sum_{i,j}\Delta x\left|(\alpha u_y)_{i,j} - (\alpha u_y)_{i,j-1}\right|^2\frac{1}{\alpha_{i,j-\frac{1}{2}}}\right]. \tag{45}$$

Proof. Let $q_h = (r, u_x, u_y)_h^T \in \operatorname{Ker} \mathbb{P}^{h,\square}_{\kappa=1,\alpha}$. This implies that $\sum_{i,j} \Delta x \Delta y \alpha_{i,j} r_{i,j} = 0$. Using the discrete weighted Poincaré-Wirtinger inequality (see Proposition B.2 of Appendix B) on r with $(\mu_{i,j}, \mu_{i-\frac{1}{2},j}, \mu_{i,j-\frac{1}{2}}) = (\alpha_{i,j}, \alpha_{i-\frac{1}{2},j}, \alpha_{i,j-\frac{1}{2}})$, we obtain

$$\sum_{i,j} \Delta x \Delta y \alpha_{i,j} r_{i,j}^2 \leq 2\|\alpha\|_\infty^2 \left\|\frac{1}{\alpha}\right\|_\infty^2 \frac{L_x^2 + L_y^2}{\min(\Delta x, \Delta y)} \left[\sum_{i,j} \Delta x \alpha_{i,j-\frac{1}{2}} \left|r_{i,j} - r_{i,j-1}\right|^2 + \sum_{i,j} \Delta y \alpha_{i-\frac{1}{2},j} \left|r_{i,j} - r_{i-1,j}\right|^2\right]. \quad (46)$$

As far as u_x is concerned since for all $j \in [\![1, N_y]\!]$, $0 = \sum_{k=1}^{N_x} (u_x)_{k,j} \Delta x = \sum_{k=1}^{N_x} (\alpha u_x)_{k,j} \frac{1}{\alpha_{k,j}} \Delta x$, by applying the 1D discrete Poincaré inequality (see Proposition B.1 of Appendix B) to the sequence $k \mapsto (\alpha u_x)_{k,j}$ with the weights $(\mu_k, \mu_{k-\frac{1}{2}}) = (\frac{1}{\alpha_{k,j}}, \frac{1}{\alpha_{k-\frac{1}{2},j}})$, we obtain for all $j \in [\![1, N_y]\!]$

$$\sum_{k=1}^{N_x} \Delta x \frac{1}{\alpha_{k,j}} (\alpha u_x)_{k,j}^2 \leq \|\alpha\|_\infty \left\|\frac{1}{\alpha}\right\|_\infty \frac{L_x^2}{\Delta x} \sum_{k=1}^{N_x} \left|(\alpha u_x)_{k,j} - (\alpha u_x)_{k-1,j}\right|^2 \frac{1}{\alpha_{k-\frac{1}{2},j}}.$$

By multiplying by Δy and by summing over j, we have

$$\sum_{i,j} \Delta x \Delta y \alpha_{i,j} (u_x)_{i,j}^2 \leq \|\alpha\|_\infty \left\|\frac{1}{\alpha}\right\|_\infty \frac{L_x^2}{\Delta x} \sum_{i,j} \Delta y \left|(\alpha u_x)_{i,j} - (\alpha u_x)_{i-1,j}\right|^2 \frac{1}{\alpha_{i-\frac{1}{2},j}}. \quad (47)$$

The same analysis holds for u_y such that for all $i \in [\![1, N_x]\!]$, $0 = \sum_{\ell=1}^{N_y} (u_y)_{i,\ell} \Delta y = \sum_{\ell=1}^{N_y} (\alpha u_y)_{i,\ell} \frac{1}{\alpha_{i,\ell}} \Delta y$, and we finally obtain

$$\sum_{i,j} \Delta x \Delta y \alpha_{i,j} (u_y)_{i,j}^2 \leq \|\alpha\|_\infty \left\|\frac{1}{\alpha}\right\|_\infty \frac{L_y^2}{\Delta y} \sum_{i,j} \Delta x \left|(\alpha u_y)_{i,j} - (\alpha u_y)_{i,j-1}\right|^2 \frac{1}{\alpha_{i,j-\frac{1}{2}}}. \quad (48)$$

With (46), (47) and (48), the result follows. □

To prove inequality (38), we first prove the following Lemma which shows that $q_{h,2}$ tends exponentially fast to the projection of its initial condition on $\operatorname{Ker} \mathbb{L}^{h,\square}_{\kappa=1,\alpha}$ (Items 2. and 3. above):

Lemma 4.6. *There exists a constant $K_\alpha(\mathbb{T}) > 0$ depending on $\mathbb{T}$ and α such that*

$$\forall t \geq 0, \quad \left\|q_{h,2} - \mathbb{P}^{h,\square}_{\kappa=1,\alpha} \mathbb{P}^{h,\square}_{\alpha} q_h^0\right\|_{l^2_\alpha}(t) \leq \left\|(1 - \mathbb{P}^{h,\square}_{\kappa=1,\alpha}) \circ \mathbb{P}^{h,\square}_{\alpha} q_h^0\right\|_{l^2_\alpha} \exp\left(-\frac{a_\star \min(\Delta x, \Delta y)}{2M K_\alpha(\mathbb{T})} t\right). \quad (49)$$

Proof. Let us define $\hat{q}_h = q_{h,2} - \mathbb{P}^{h,\square}_{\kappa=1,\alpha} \mathbb{P}^{h,\square}_{\alpha} q_h^0 := (\hat{r}, \hat{\mathbf{u}})_h^T$. The idea is to apply the energy estimate of Theorem 4.1 to $\hat{q}_h$ and then the Poincaré inequality of Lemma 4.5. For this, we first remark that $\hat{q}_h$ satisfies (19). Indeed, $q_{h,2}$ satisfies (40), and $\mathbb{P}^{h,\square}_{\kappa=1,\alpha} \mathbb{P}^{h,\square}_{\alpha} q_h^0$ does not depend on time and is in the kernel of $\mathbb{L}^{h,\square}_{\kappa=1,\alpha}$. Then, $\hat{q}_h$ is solution of (19) and with Theorem 4.1, we have

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|\hat{q}_h\|^2_{l^2_\alpha}(t) &= -\frac{a_\star}{2M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left(\alpha_{ij} |\hat{r}_i - \hat{r}_j|^2 + \frac{1}{\alpha_{ij}} \left|\left((\alpha\hat{\mathbf{u}})_i - (\alpha\hat{\mathbf{u}})_j\right) \cdot \mathbf{n}_{ij}\right|^2\right) \\
&= -\frac{a_\star}{2M} \left[\sum_{i,j} \Delta x \alpha_{i,j-\frac{1}{2}} \left|\hat{r}_{i,j} - \hat{r}_{i,j-1}\right|^2 + \sum_{i,j} \Delta y \alpha_{i-\frac{1}{2},j} \left|\hat{r}_{i,j} - \hat{r}_{i-1,j}\right|^2 \right. \\
&\quad \left. + \sum_{i,j} \Delta y \left|(\alpha\hat{u}_x)_{i,j} - (\alpha\hat{u}_x)_{i-1,j}\right|^2 \frac{1}{\alpha_{i-\frac{1}{2},j}} + \sum_{i,j} \Delta x \left|(\alpha\hat{u}_y)_{i,j} - (\alpha\hat{u}_y)_{i,j-1}\right|^2 \frac{1}{\alpha_{i,j-\frac{1}{2}}}\right]. \quad (50)
\end{aligned}$$

Moreover, the initial condition of $\hat{q}_h$ is $\mathbb{P}_\alpha^{h,\square} q_h^0 - \mathbb{P}_{\kappa=1,\alpha}^{h,\square} \mathbb{P}_\alpha^{h,\square} q_h^0$, which belongs to $\operatorname{Ker} \mathbb{P}_{\kappa=1,\alpha}^{h,\square}$. Thus, applying (44) of Lemma 4.4, it follows that $\hat{q}_h(t)$ belongs to $\operatorname{Ker} \mathbb{P}_{\kappa=1,\alpha}^{h,\square}$ for all $t \geq 0$ and we can apply Lemma 4.5 to estimate the right-hand side of (50). This leads to

$$\frac{1}{2}\frac{d}{dt}\left\|\hat{q}_h\right\|_{l_\alpha^2}^2(t) \quad \leq \quad -\frac{a_\star \min(\Delta x, \Delta y)}{2MK_\alpha(\mathbb{T})}\left\|\hat{q}_h\right\|_{l_\alpha^2}^2(t).$$

Then Applying Grönwall's lemma, we obtain (49) because $\hat{q}_h^0 = (1 - \mathbb{P}_{\kappa=1,\alpha}^{h,\square}) \circ \mathbb{P}_\alpha^{h,\square} q_h^0$. □

Now, we are able to prove Theorem 4.3 (Item 4. above). By applying Lemma 4.6, we have for all $t \geq 0$

$$\begin{aligned}\left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) &\geq \left\|\mathbb{P}_\alpha^{h,\square} q_h^0 - \mathbb{P}_{\kappa=1,\alpha}^{h,\square}\mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) - \left\|q_{h,2} - \mathbb{P}_{\kappa=1,\alpha}^{h,\square}\mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t), \\ &\geq \left\|(\mathrm{Id} - \mathbb{P}_{\kappa=1,\alpha}^{\mathrm{h},\square}) \circ \mathbb{P}_\alpha^{\mathrm{h},\square} \mathrm{q}_\mathrm{h}^0\right\| \left(1 - \exp\left(-\frac{a_\star \min(\Delta x, \Delta y)}{2MK_\alpha(\mathbb{T})} t\right)\right). \quad (51)\end{aligned}$$

Since the right-hand side of (51) is a growing function of time, we can obtain a lower bound by evaluating it at any time; we set $C = \left\|(\mathrm{Id} - \mathbb{P}_{\kappa=1,\alpha}^{\mathrm{h},\square}) \circ \mathbb{P}_\alpha^{\mathrm{h},\square} \mathrm{q}_\mathrm{h}^0\right\|$ and choose $t_0 = C_2 M$ with $C_2 = \frac{K_\alpha(\mathbb{T})}{a_\star}$ and we obtain: $\forall t \geq C_2 M$ it holds that

$$\left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) \geq C\left(1 - \exp\left(-\frac{\min(\Delta x, \Delta y)}{2}\right)\right). \quad (52)$$

Using that $1 - \exp(-x/2) \geq x/3$ for $x \in [0;1]$, Eq. (52) implies that for $\min(\Delta x, \Delta y) \leq 1$

$$\forall t \geq C_2 M, \quad \left\|q_{h,2} - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) \geq \frac{C}{3}\min(\Delta x, \Delta y). \quad (53)$$

In the sequel, we assume that C is strictly positive, which is the case for almost all functions $q_h^0 \in l_\alpha^2(\mathbb{T})^3$. Let us now suppose that

$$C_1 M \leq C_3 \min(\Delta x, \Delta y) \quad \text{with} \quad C_3 = \frac{C}{6},$$

then we obtain from (41) and (53) that

$$\forall t \geq C_2 M, \quad \left\|q_h - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) \quad \geq \quad C_3 \min(\Delta x, \Delta y)$$

for any $M \leq \dfrac{C_3}{C_1}\min(\Delta x, \Delta y)$. □

Theorem 4.3 tells us that the wrong behavior of the standard Godunov scheme is due at the same time to a wrong kernel (the image of $(\mathrm{Id} - \mathbb{P}_{\kappa=1,\alpha}^{\mathrm{h},\square}) \circ \mathbb{P}_\alpha^{\mathrm{h},\square}$ is "too large") and to a fast diffusion rate, at least proportional to $\frac{\min(\Delta x, \Delta y)}{M}$. There are thus two options to propose a correction to this scheme, namely restoring a correct kernel by setting $\kappa = 0$ or drastically diminishing the diffusion rate by setting $\kappa = M$. If none of these solutions is used, then a possible (but expensive) solution is to choose $(\Delta x, \Delta y)$ of the size of M. These three possibilities are studied in the next subsections.

4.3.2 Correction of the Godunov scheme on a Cartesian mesh

Theorem 4.7. *Let $q_h(t)$ be a solution of scheme (19) with initial condition q_h^0.*

1. *With $\kappa = 0$, for all $q_h^0 \in l_\alpha^2(\mathbb{T})^3$, and all $C_1 > 0$ we have*

$$\left\|q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2} = C_1 M \Longrightarrow \forall t \geq 0, \left\|q_h - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) \leq C_1 M. \quad (54)$$

2. *When $\kappa = M$, for all $q_h^0 \in l_\alpha^2(\mathbb{T})^3$ and all $C_1, C_2 > 0$, there exists $C_3(C_1, C_2, q_h^0) > 0$ such that*

$$\left\|q_h^0 - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2} = C_1 M \Longrightarrow \forall t \in [0; C_2 M], \left\|q_h - \mathbb{P}_\alpha^{h,\square} q_h^0\right\|_{l_\alpha^2}(t) \leq C_3 M \quad (55)$$

where C_3 does not depend on M.

Proof. By linearity of $\mathbb{L}^{h,\Box}_{\kappa,\alpha}$, the solution q_h of (10) given by scheme (19) with initial condition q_h^0 can be written as

$$q_h = q_{h,1} + q_{h,2}$$

where $q_{h,1}$ is the solution of

$$\begin{cases} \partial_t(\alpha q_{h,1}) + \dfrac{\mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M}(q_{h,1}) = 0, \\ q_{h,1}(t=0) = q_h^0 - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \end{cases} \tag{56}$$

and $q_{h,2}$ is the solution of

$$\begin{cases} \partial_t(\alpha q_{h,2}) + \dfrac{\mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M}(q_{h,2}) = 0, \\ q_{h,2}(t=0) = \mathbb{P}^{h,\Box}_{\alpha} q_h^0. \end{cases} \tag{57}$$

We have

$$\begin{aligned} \forall t \geq 0, \quad \left\| q_h - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha}(t) &= \left\| q_{h,1} + q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha}(t) \\ &\leq \|q_{h,1}\|_{l^2_\alpha}(t) + \left\| q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha}(t) \\ &\leq \|q_{h,1}\|_{l^2_\alpha}(0) + \left\| q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha}(t) \\ &\leq \left\| q_h^0 - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha} + \left\| q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha}(t) \end{aligned} \tag{58}$$

because scheme (19) is dissipative when $\kappa \geq 0$ (see Theorem 4.1). If $\left\| q_h^0 - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha} = C_1 M$, then (58) shows that we need to find an upper bound for the function

$$t \mapsto \left\| q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q^0 \right\|_{l^2_\alpha}(t)$$

where $q_{h,2}$ is the solution of (57).
Assume that $\kappa = 0$. Since $\mathbb{P}^{h,\Box}_{\alpha} q_h^0 \in \mathcal{E}^{h,\Box}_{\alpha} = \operatorname{Ker} \mathbb{L}^{h,\Box}_{\kappa=0,\alpha}$ we have $\mathbb{L}^{h,\Box}_{\kappa=0,\alpha}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right) = 0$ and $q_{h,2}(t) = \mathbb{P}^{h,\Box}_{\alpha} q_h^0$ for all $t \geq 0$ is the solution of (57). Then of course $\left\| q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q^0 \right\|_{l^2_\alpha}(t) = 0$ and (58) reduces to

$$\forall t \geq 0, \quad \left\| q_h - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha}(t) \leq C_1 M$$

if $\left\| q_h^0 - \mathbb{P}^{h,\Box}_{\alpha} q_h^0 \right\|_{l^2_\alpha} = C_1 M$. The first point is proved.
Assume now that $\kappa > 0$. We have $\mathbb{P}^{h,\Box}_{\alpha} q_h^0 \in \mathcal{E}^{h,\Box}_{\alpha} \supsetneq \operatorname{Ker} \mathbb{L}^{h,\Box}_{\kappa,\alpha}$ and we can have $\mathbb{L}^{h,\Box}_{\kappa,\alpha}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right) \neq 0$. Since $\mathbb{P}^{h,\Box}_{\alpha} q_h^0 \in \mathcal{E}^{h,\Box}_{\alpha} = \operatorname{Ker} \mathbb{L}^{h,\Box}_{\kappa=0,\alpha}$, we have $\mathbb{L}^{h,\Box}_{\kappa=0,\alpha}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right) = 0$ and thus

$$\partial_t\left(\alpha \mathbb{P}^{h,\Box}_{\alpha} q_h^0\right) + \frac{\mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right) = \frac{\mathbb{L}^{h,\Box}_{\kappa,\alpha} - \mathbb{L}^{h,\Box}_{\kappa=0,\alpha}}{M}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right).$$

Combining this with (57), setting $q_h^\star := q_{h,2} - \mathbb{P}^{h,\Box}_{\alpha} q_h^0$ and using linearity, we deduce that

$$\partial_t\left(\alpha q_h^\star\right) + \frac{\mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M} q_h^\star = \frac{\mathbb{L}^{h,\Box}_{\kappa=0,\alpha} - \mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right). \tag{59}$$

Taking the weighted scalar product of (59) with $\frac{q_h^\star}{\alpha}$, we obtain

$$\left\langle \partial_t\left(\alpha q_h^\star\right), \frac{q_h^\star}{\alpha} \right\rangle_{\alpha,h} + \left\langle \frac{\mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M} q_h^\star, \frac{q_h^\star}{\alpha} \right\rangle_{\alpha,h} = \left\langle \frac{\mathbb{L}^{h,\Box}_{\kappa=0,\alpha} - \mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right), \frac{q_h^\star}{\alpha} \right\rangle_{\alpha,h}. \tag{60}$$

It follows from the proof of Theorem 4.1 that

$$\left\langle \frac{\mathbb{L}^{h}_{\kappa,\alpha}}{M} q_h^\star, \frac{q_h^\star}{\alpha} \right\rangle_{\alpha,h} = \frac{a_\star}{M} \sum_{\Gamma_{ij}} \left(\alpha_{ij} |r_i^\star - r_j^\star|^2 + \frac{\kappa}{\alpha_{ij}} \left| \left((\alpha \mathbf{u}^\star)_i - (\alpha \mathbf{u}^\star)_j \right) \cdot \mathbf{n}_{ij} \right|^2 \right) \geq 0.$$

Thus, from (60) and using the Cauchy-Schwarz inequality, we obtain

$$\frac{1}{2}\frac{d}{dt}\|q_h^\star\|^2_{l^2_\alpha}(t) \leq \left\| \frac{\mathbb{L}^{h,\Box}_{\kappa=0,\alpha} - \mathbb{L}^{h,\Box}_{\kappa,\alpha}}{M\alpha}\left(\mathbb{P}^{h,\Box}_{\alpha} q_h^0\right) \right\|_{l^2_\alpha} \|q_h^\star\|_{l^2_\alpha}(t)$$

which leads to

$$\frac{d}{dt}\left\|q_h^\star\right\|_{l^2_\alpha}(t) \leq \left\| \frac{\mathbb{L}^{h,\square}_{\kappa=0,\alpha} - \mathbb{L}^{h,\square}_{\kappa,\alpha}}{M\alpha}\left(\mathbb{P}^{h,\square}_\alpha q_h^0\right)\right\|_{l^2_\alpha}. \tag{61}$$

For any q, a direct calculation shows that $(\mathbb{L}^{h,\square}_{\kappa=0,\alpha} - \mathbb{L}^{h,\square}_{\kappa,\alpha})q$ is proportional to κ and does not depend on M. So (61) and the fact that $q_h^\star(t=0)=0$ show that $\exists \hat{C}_3(\alpha, q_h^0, \Delta x, \Delta y, a_\star)$ such that

$$\left\|q_h^\star\right\|_{l^2_\alpha}(t) \leq \hat{C}_3 \frac{\kappa}{M} t \ \ \forall t \geq 0. \tag{62}$$

Then from (58) and (62) when $\kappa = M$, we obtain (55) with $C_3 = C_1 + C_2\hat{C}_3$. □

Remark 4.8. *It is important to stress that the constant C_3 in item 2 of Theorem 4.7 depends on a concept of discrete smoothness for q_h^0 detailed in the next subsection and that, in the worst case, it may behave proportionally to $\frac{1}{\min(\Delta x, \Delta y)}$.*

4.3.3 The case of a very fine mesh

We observe that if the right-hand side of (61) can be bounded by $C\kappa\max(\Delta x, \Delta y)$ with C not depending on $(\kappa, M, \Delta x, \Delta y)$, then we shall also have a bound of the type (55) if $\kappa = 1$ (uncorrected Godunov scheme) and $\max(\Delta x, \Delta y) \leq C_0 M$. For this, we introduce the definition of discrete regularity:

Definition 4.9. *Let $q_h := (r_h, u_{h,x}, u_{h,y})$ be a family of discrete fields parameterized by $(\Delta x, \Delta y)$; then we define $H^2_\alpha(\mathbb{T})^3$ to be the set of families of discrete fields such that*

$$\|q_h\|_{H^2_\alpha(\mathbb{T})} := \sup_{\Delta x>0, \Delta y>0} \left\|\delta^{c,x}\left(\frac{1}{\alpha}\delta^{s,x}(\alpha u_{h,x})\right)\right\|_{l^2_\alpha(\mathbb{T})} + \left\|\delta^{c,y}\left(\frac{1}{\alpha}\delta^{s,y}(\alpha u_{h,y})\right)\right\|_{l^2_\alpha(\mathbb{T})} < +\infty$$

with the following definitions for the centered and staggered finite differences in the horizontal and vertical directions

$$(\delta^{c,x}v)_{i,j} := \frac{(v_{i+\frac{1}{2},j} - v_{i-\frac{1}{2},j})}{\Delta x} \ , \ (\delta^{s,x}z)_{i+\frac{1}{2},j} := \frac{(z_{i+1,j} - z_{i,j})}{\Delta x} \ ,$$

$$(\delta^{c,y}w)_{i,j} := \frac{(w_{i,j+\frac{1}{2}} - w_{i,j-\frac{1}{2}})}{\Delta y} \ , \ (\delta^{s,y}z)_{i,j+\frac{1}{2}} := \frac{(z_{i,j+1} - z_{i,j})}{\Delta y} \ .$$

This concept allows us to prove that with discrete regular initial conditions, refining the mesh is also a possibility to obtain acceptable results on an $\mathcal{O}(M)$ time scale when using the standard Godunov scheme. Indeed, the following theorem holds:

Theorem 4.10. *Let $q_h(t)$ be a solution of scheme (19) with initial condition q_h^0. When $\kappa = 1$, for all q_h^0 such that $\mathbb{P}^{h,\square}_\alpha q_h^0 \in H^2_\alpha(\mathbb{T})^3$, and all $C_0, C_1, C_2 > 0$, there exists $C_3(C_0, C_1, C_2, q_h^0) > 0$ that does not depend on M, Δx and Δy such that*

$$\begin{cases} \Delta x \leq C_0 M, \\ \Delta y \leq C_0 M, \\ \left\|q_h^0 - \mathbb{P}^{h,\square}_\alpha q_h^0\right\|_{l^2_\alpha} = C_1 M \end{cases} \Longrightarrow \forall t \in [0; C_2 M], \left\|q_h - \mathbb{P}^{h,\square}_\alpha q_h^0\right\|_{l^2_\alpha}(t) \leq C_3 M. \tag{63}$$

Proof. For any $q_h := (r_h, u_{h,x}, u_{h,y})_{i,j}$, a direct calculation shows that

$$\left(\frac{(\mathbb{L}^{h,\square}_{\kappa=0,\alpha} - \mathbb{L}^{h,\square}_{\kappa,\alpha})}{M\alpha} q\right)_{i,j} = \frac{a_\star \kappa}{2M}\begin{pmatrix} 0 \\ \Delta x\left(\delta^{c,x}\left[\frac{1}{\alpha}\delta^{s,x}(\alpha u_{h,x})\right]\right)_{i,j} \\ \Delta y\left(\delta^{c,y}\left[\frac{1}{\alpha}\delta^{s,y}(\alpha u_{h,y})\right]\right)_{i,j} \end{pmatrix}. \tag{64}$$

Therefore, if $\mathbb{P}^{h,\square}_\alpha q_h^0 \in H^2_\alpha(\mathbb{T})^3$, then (61) and the fact that $q_h^\star(t=0)=0$ show that

$$\left\|q_h^\star\right\|_{l^2_\alpha}(t) \leq \frac{a_\star \kappa}{2M}\|\mathbb{P}^{h,\square}_\alpha q_h^0\|_{H^2_\alpha(\mathbb{T})}\max(\Delta x, \Delta y) t \ \ \forall t \geq 0. \tag{65}$$

Then from (58) and (65) when $\kappa = 1$ and $\max(\Delta x, \Delta y) \leq C_0 M$, we obtain (63) with $C_3 = C_1 + \frac{a_\star}{2} C_0 C_2 \|\mathbb{P}^{h,\square}_\alpha q_h^0\|_{H^2_\alpha(\mathbb{T})}$. □

5 Numerical results on the wave equation

In this section, we perform some numerical simulations on the linear wave equation with porosity (10) using the Godunov scheme (18). The aim is to illustrate all the theoretical results of the article. A 2D periodic domain $\mathbb{T} = [0,1[\times[0,1[$ is considered. All simulations were run with an Euler explicit time stepping with a CFL number of 0.4. The parameters $a_\star$ and M are set to $a_\star = 1$ and $M = 10^{-4}$. We consider a regular Cartesian mesh containing 1 600 cells ($\Delta x = \Delta y = 0.025$) and an unstructured triangular mesh containing 2 326 cells generated by `GMSH` [19].

5.1 A stationary case

We firstly illustrate the influence of the mesh type (triangular or Cartesian) on the kernel of the Godunov scheme. The initial condition $q^0 = (r^0, \mathbf{u}^0)^T$ is chosen such that $q^0 \in \mathcal{E}_\alpha$. We take

$$\begin{cases} r^0(x,y) = 1, \\ \mathbf{u}^0(x,y) = \dfrac{\nabla \times \psi}{\alpha}(x,y) \end{cases} \tag{66}$$

where

$$\alpha(x,y) = \frac{1}{2} + \frac{1}{2}\exp\left(-\frac{\|(x,y)-(0.5,0.5)\|^2}{0.25^2}\right), \tag{67}$$

$$\psi(x,y) = \frac{1}{\pi}\sin^2(\pi x)\sin^2(\pi y). \tag{68}$$

This expression of ψ corresponds to the "vortex in a box" test case of [9]. We note that ψ is very important from a numerical point of view because it allows to define q_h^0 such that $q_h^0 \in \mathcal{E}_\alpha^{h,\square}$ on Cartesian meshes and such that $q_h^0 \in \mathcal{E}_\alpha^{h,\triangle}$ on triangular meshes. Since $q^0 \in \mathcal{E}_\alpha$, the field q defined by

$$\forall t \geq 0,\ \forall (x,y) \in \mathbb{T}, \quad q(t,x,y) = q^0(x,y) \tag{69}$$

is solution of the linear wave equation with porosity (10). We study if (69) is or is not satisfied at the discrete level when we solve system (10) with Godunov's scheme (18) on a Cartesian or a triangular mesh with $\kappa = 0$ or $\kappa = 1$.

In Figure 1, we plot the norm of $\alpha\mathbf{u}$ obtained after 1 000 iterations on Cartesian and triangular meshes with $\kappa = 1$ and $\kappa = 0$. The solution is preserved over time on triangular meshes with $\kappa = 1$ and $\kappa = 0$ but is also preserved over time on Cartesian meshes with $\kappa = 0$. This result illustrates Proposition 3.7 and Proposition 3.9.

5.2 A well-prepared initial condition

We now consider a well-prepared initial condition. It means that the initial condition can be split into two components, a component in the kernel $\mathcal{E}_\alpha$ plus a component of order M in the orthogonal set to the kernel, $\mathcal{E}_\alpha^\perp$. We illustrate the theoretical results Theorem 4.3 and Theorem 4.7 on the evolution with respect to time of the deviation $\left\|q_h - \mathbb{P}_\alpha^{h,\triangle \text{ or } \square} q_h^0\right\|_{l^2_\alpha}$ with the different schemes on triangular and Cartesian meshes. The initial condition q_h^0 is given by

$$q_h^0 = M q_{h,1}^0 + q_{h,2}^0$$

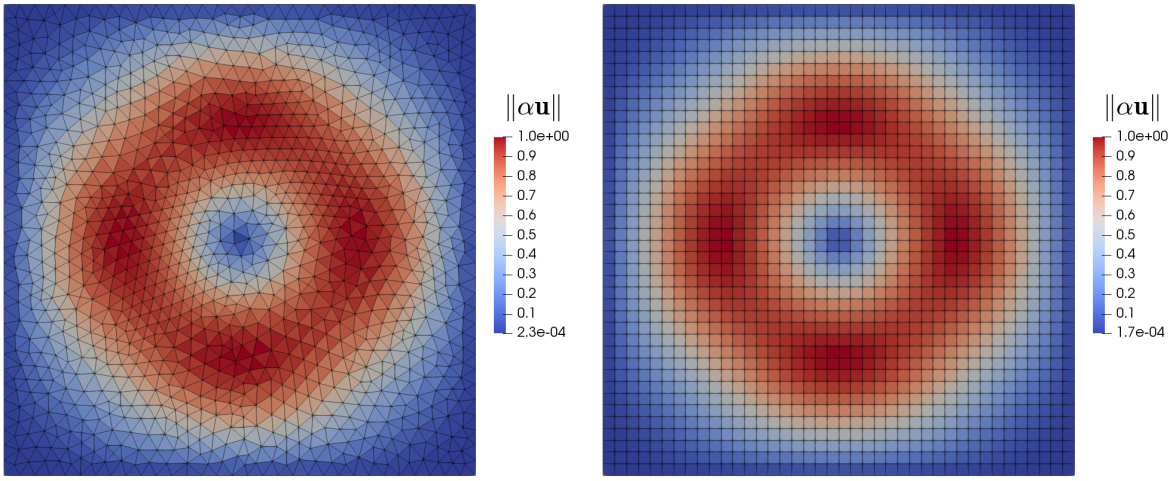
where $q_{h,2}^0 \in \mathcal{E}_\alpha^{h,\square \text{ or } \triangle}$ is given by (66) and $q_{h,1}^0 \in \left(\mathcal{E}_\alpha^{h,\square \text{ or } \triangle}\right)^\perp$ satisfying $\|q_{h,1}^0\|_{l^2_\alpha} = 1$. More precisely, we take $q_{h,1}^0 = \bar{q}_{h,1} / \|\bar{q}_{h,1}\|_{l^2_\alpha}$ with

$$\begin{cases} \bar{r}_{h,1}^0(x,y) = \left(\dfrac{\sin(2\pi x)\cos(2\pi y)}{\alpha(x,y)}\right)_h, \\ \bar{\mathbf{u}}_{h,1}^0 = \nabla_h \phi_h \end{cases}$$

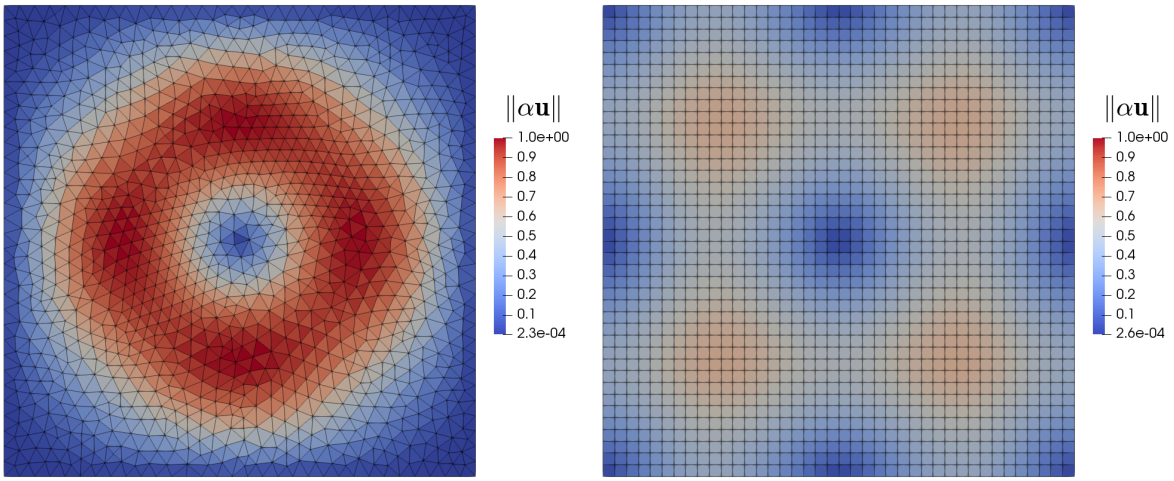
where

$$\phi_h(x,y) = \left(\sin(2\pi x)\cos(2\pi y)\right)_h.$$

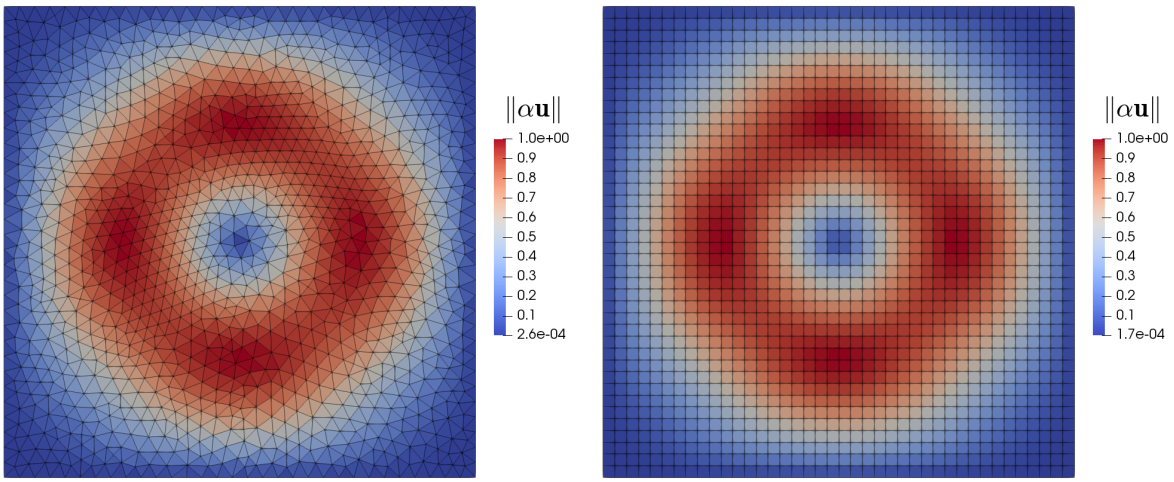
The discrete field $r_{h,1}$ is defined at the cell centers, and so is ϕ_h on Cartesian meshes; but on a triangular mesh $\phi_h \in W_h$, then ϕ_h is defined at the edge midpoints.



Reference solution (initial condition), triangles (left) and Cartesian (right)



Triangles, $\kappa = 1$ Cartesian, $\kappa = 1$



Triangles, $\kappa = 0$ Cartesian, $\kappa = 0$

Figure 1: Norm of $\alpha\mathbf{u}$ obtained after 1 000 iterations with the Godunov scheme (18) with $\kappa = 1$ and $\kappa = 0$ on triangular and Cartesian meshes.

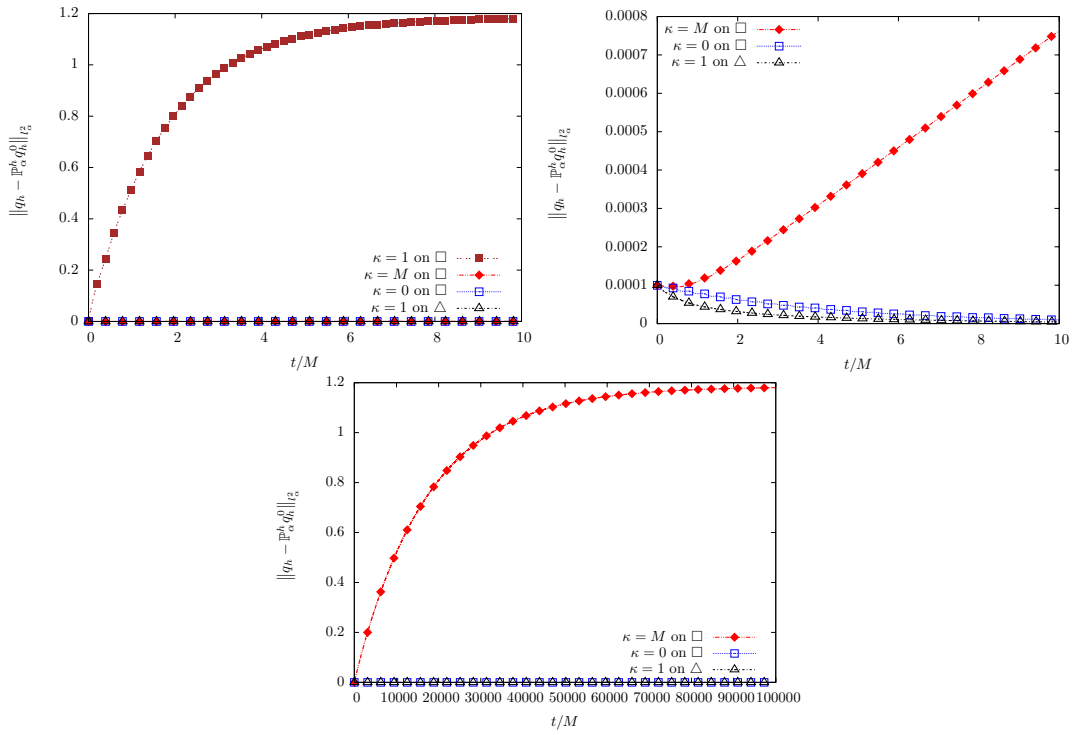


Figure 2: Norm of the deviation $\|q_h - \mathbb{P}_\alpha^{h,\triangle \text{ or } \square} q_h^0\|_{l^2_\alpha}(t)$ obtained with the Godunov scheme with $\kappa = 1$, $\kappa = M$ and $\kappa = 0$ on a Cartesian mesh and with $\kappa = 1$ on a triangular mesh for times of order M: $0 \leq t/M \leq 10$ (top) and for long times: $0 \leq t/M \leq 10/M$ (bottom).

In Figure 2, we plot the evolution with respect to time of the deviation $\left\|q_h^0 - \mathbb{P}_\alpha^{h,\triangle \text{ or } \square} q_h^0\right\|_{l^2_\alpha}$ with scheme (18) with $\kappa = 1$, $\kappa = M$ and $\kappa = 0$ on Cartesian meshes and for $\kappa = 1$ on triangular meshes. The Godunov scheme ($\kappa = 1$) on a triangular mesh is accurate at low Mach number. Indeed, the deviation remains of order M, even for long times (see Theorem 4.2). The Cartesian case is very different. The Godunov scheme ($\kappa = 1$) on a Cartesian mesh is not accurate at low Mach number, it introduces a deviation greater than $\Delta x = \Delta y = 0.025$, even for a time of order M (see the case $\kappa = 1$ on a Cartesian mesh in Theorem 4.3). The scheme has to be corrected on a Cartesian mesh at low Mach number. With $\kappa = 0$, the deviation remains of order M, even for long times (see the case $\kappa = 0$ in Theorem 4.7). With $\kappa = M$, the deviation remains of order M for times of order M (see the case $\kappa = M$ in Theorem 4.7), but this is not the case for long times (times of order one).

6 The non linear case

6.1 Numerical schemes

Since α is regular and does not depend on time, we can write system (1) as

$$\partial_t \mathbf{W} + \nabla \cdot \mathbf{f}(\mathbf{W}) = S(\mathbf{W}) \nabla \alpha \tag{70}$$

where $\mathbf{W} = (\alpha, \alpha\rho, \alpha\rho\mathbf{u})^T$ and the flux $\mathbf{f}$ and the source term $S(\mathbf{W})$ are given by

$$\mathbf{f}(\mathbf{W}) = \begin{pmatrix} 0 \\ \alpha\rho\mathbf{u} \\ \alpha\rho\mathbf{u} \otimes \mathbf{u} + \alpha p \mathrm{I} \end{pmatrix} \qquad S(\mathbf{W}) = \begin{pmatrix} 0 \\ 0 \\ p \end{pmatrix}.$$

The numerical scheme for system (70) is given by

$$\frac{\mathbf{W}_i^{n+1} - \mathbf{W}_i^n}{\Delta t} + \frac{1}{|\Omega_i|} \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| F_-(\mathbf{W}_i^n, \mathbf{W}_j^n, \mathbf{n}_{ij}) = 0 \tag{71}$$

where $\mathbf{W} = (\alpha, \alpha\rho, \alpha\rho\mathbf{u})^T$ and F_- is the non conservative numerical flux. In this paper, we use two different fluxes, a VFRoe flux and the well-balanced Lax-Friedrich scheme of [26].

6.2 Well-balanced Lax-Friedrich scheme of [26]

The well-balanced Lax-Friedrich scheme of [26] allows to maintain equilibrium states and is easy to implement. The non-conservative numerical flux is given by

$$F_-^{\text{LF-WB}}(\mathbf{W}_L, \mathbf{W}_R, \mathbf{n}) = F^{\text{LF}}(\mathbf{W}_L, \mathbf{W}_R^-, \mathbf{n}) \tag{72}$$

where F^{LF} corresponds to the standard Lax-Friedrich numerical flux

$$F^{\text{LF}}(\mathbf{W}_L, \mathbf{W}_R, \mathbf{n}) = \frac{\mathbf{f}(\mathbf{W}_L) + \mathbf{f}(\mathbf{W}_R)}{2} \cdot \mathbf{n} - \frac{\max\limits_{i\in\{L,R\}} (|\mathbf{u}_i \cdot \mathbf{n}| + c_i)}{2} (\mathbf{W}_R - \mathbf{W}_L)$$

and the state $\mathbf{W}_R^- = (\alpha, \alpha\rho, \alpha\rho\mathbf{u})_R^-$ is defined such that

$$\begin{cases} \alpha_R^- = \alpha_L, \\ (\alpha\rho\mathbf{u})_R^- = (\alpha\rho\mathbf{u})_R, \\ \dfrac{\|\mathbf{u}_R^-\|^2}{2} + h(\rho_R^-) = \dfrac{\|\mathbf{u}_R\|^2}{2} + h(\rho_R) \end{cases}$$

where $h(\rho) = \kappa\gamma\rho^{\gamma-1}/(\gamma-1)$. For the existence and uniqueness of W_R^-, we refer to [26].

6.3 VFRoe scheme

We want to write a non-linear scheme that is consistent with the study we did in the linear case (see section 3 and section 4). We recall that in the linear case, Godunov's scheme (18) can be interpreted as a VFRoe scheme [18, 6] in variables $(\alpha, r, \alpha\mathbf{u})$ (see Remark 3.2). The VFRoe solver consists in a local linearization of a Riemann problem which is simpler to handle since it only deals with linear problems and avoids the complex exact resolution of the Riemann problem with porosity jump. Then, in the non-linear case, we write a VFRoe scheme in variables $\mathbf{Y} = (\alpha, \rho, \alpha\rho\mathbf{u})^T$. Another advantage of this set of variables is that we get a scheme that is well-balanced in the sense that it exactly preserves the one-dimensional steady states. For the VFRoe scheme (with another set of variables) applied to system (70), we refer to [40, 39]. The VFRoe numerical flux is given by

$$F_-^{\text{VFRoe}}(\mathbf{W}_L, \mathbf{W}_R, \mathbf{n}) = \mathbf{f}(R(0^-, \mathbf{Y}_L, \mathbf{Y}_R, \mathbf{n})) \cdot \mathbf{n} \tag{73}$$

where $R(0^-, \mathbf{Y}_L, \mathbf{Y}_R, \mathbf{n})$ corresponds to the solution in $\xi/t = 0^-$ of the linearized Riemann problem that is detailed in Appendix C.

The VFRoe solver considered does not allow to treat the resonant cases when eigenvalues $\lambda_1 = \mathbf{u}\cdot\mathbf{n} - c$ or $\lambda_2 = \mathbf{u}\cdot\mathbf{n} + c$ vanish. The resonant cases are out the scope of this paper since we focus here on the accuracy at low Mach number. For the resonant cases, we refer to [10].

6.4 All-Mach VFRoe scheme

The flux in the all-Mach VFRoe scheme is given by

$$F_-^{\text{AM-VFRoe}}(\mathbf{W}_i, \mathbf{W}_j, \mathbf{n}) = F_-^{\text{VFRoe}}(\mathbf{W}_i, \mathbf{W}_j, \mathbf{n}) + (\theta_{ij} - 1)\frac{\alpha_i \hat{c}_{ij}}{\hat{\alpha}_{ij}} \begin{pmatrix} 0 \\ [((\alpha\rho\mathbf{u})_i - (\alpha\rho\mathbf{u})_j)\cdot\mathbf{n}]\,\mathbf{n} \end{pmatrix} \tag{74}$$

where $\hat{\alpha}_{ij}$ and $\hat{c}_{ij}$ correspond to VFRoe average states (see (80) in the Appendix) and $\theta_{ij} = \min(1, \max(M_i, M_j)) = \min(1, \max(\|\mathbf{u}_i\|/c_i, \|\mathbf{u}_j\|/c_j))$. We remark that in (74), we recover the classical VFRoe scheme if $\theta_{ij} = 1$. This means that we correct the numerical flux only if both states $\mathbf{W}_i$ and $\mathbf{W}_j$ are subsonic.

6.5 Numerical results

We perform a one dimensional test to check the robustness of the low Mach corrected scheme but also the capability of the scheme to maintain equilibrium states across a discontinuous cross-section. Indeed, since the low Mach correction reduces the numerical diffusion of the scheme, stability of this scheme for unsteady low Mach flow has to be tested. Moreover, it is well-known that schemes which do not maintain the equilibrium states may give unsatisfactory results when refining the mesh [26], so that the well-balanced property also has to be tested. Then, we perform a two dimensional test to check the low Mach accuracy of the different schemes on triangular and Cartesian meshes.

For all simulations, we use the following pressure law $p(\rho) = \kappa\rho^\gamma$ where $\kappa = 1$ and $\gamma = 1.5$ and $CFL = 0.4$.

6.5.1 A one dimensional unsteady subsonic flow

Let us denote $\mathbf{U} = (\alpha, \rho, u)$. The initial condition is a Riemann problem where the left state $\mathbf{U}_L$ and the right state $\mathbf{U}_R$ are given by

$$\mathbf{U}_L = (1, 1, 0.001), \qquad \mathbf{U}_R = (0.75, 0.3, 0.005).$$

The domain is $[0, 1]$ and the discontinuity in the initial condition is set to $x = 0.5$. The exact solution is 1-rarefaction followed by a stationary contact, then followed by a 3-shock. For an exact solution, we refer to [27]. The Mach number of the solution varies from 4×10^{-4} to 0.85 and then allows to test the robustness of the all-Mach VFRoe scheme. Moreover, since α is discontinuous between $\mathbf{U}_L$ and $\mathbf{U}_R$, we also test the capability of the scheme to preserve the two invariants of the stationary contact $\alpha\rho u$ and $u^2/2 + h(\rho)$ where $h(\rho) = \kappa\gamma\rho^{\gamma-1}/(\gamma - 1)$.

In Figure 3, we plot the porosity (or cross-section) α, the density ρ, the velocity u, the Mach number, $\alpha\rho u$ and $u^2/2 + h(\rho)$ at time $t = 0.25$ obtained with the well-balanced Lax-Friedrich scheme [26], the VFRoe scheme and the all-Mach VFRoe scheme. The all-Mach VFRoe scheme is stable. In fact, as for the constant porosity case, numerical tests seem to show that the all-Mach scheme is stable under a degenerated CFL condition which is exactly the half of the classical one (see [13, 5] for more details). This justifies why all numerical results are obtained with $CFL = 0.4$. As expected, the all-Mach VFRoe scheme is the least diffusive scheme and the well-balanced Lax-Friedrich scheme is the most diffusive one. Looking at the stationary contact in $x = 0.5$, we remark that the two invariants of the stationary contact $\alpha\rho u$ and $u^2/2 + h(\rho)$ are preserved across the discontinuity of α. Then, the VFRoe and all-Mach VFRoe schemes are also well-balanced, like the well-balanced Lax-Friedrich scheme.

6.5.2 Two-dimensional low Mach flow

We consider a two-dimensional low Mach vortex flow. Domain, meshes and boundary conditions are the same as for the wave equation (see section 5). The initial condition is an exact, steady and regular solution of the incompressible system (8). Then, (7) tells us that the solution of (3) will remain close to the initial condition since the latter solves (8) for all times. Note that in order to build an exact solution of the incompressible system (8), we adapted the isentropic vortex solution of [43, 44] to the case of variable porosity fields. The initial condition is given by

$$\begin{cases} \alpha = \alpha_0 \tilde{\alpha}_0^{(0)} = \alpha_0 \Omega^{c_1}, \\ \rho^0 = \rho_0 \left(\tilde{\rho}_0^{(0)} + M^2 \tilde{\rho}_0^{(2)} \right) = \rho_0 \left(1 + M^2 c_2 \Omega^{c_3} \right), \\ \mathbf{u}^0 = u_0 \tilde{\mathbf{u}}_0^{(0)} = u_0 \dfrac{\nabla \times \psi}{\alpha} \end{cases}$$

where $\Omega = \exp\left(-\dfrac{(x-0.5)^2 + (y-0.5)^2}{2 \times R^2} \right)$, $R = 0.15$, $\alpha_0 = 1$, $\rho_0 = 1$, $u_0 = M \times c(\rho_0)$, $c_1 = 0.25$, $c_3 = 2 \times (1 - c_1)$, $c_2 = -1/(\alpha_0^2 c_3)$ and $\psi = R\Omega$. We can easily check that $(\tilde{\rho}_0^{(2)}, \tilde{\mathbf{u}}_0^{(0)})$ is regular and satisfies

$$\begin{cases} \nabla \cdot (\tilde{\alpha}\tilde{\mathbf{u}})_0^{(0)} = 0, \\ \left(\tilde{\mathbf{u}}_0^{(0)} \cdot \nabla_{\tilde{\mathbf{x}}} \right) \tilde{\mathbf{u}}_0^{(0)} + \nabla_{\tilde{\mathbf{x}}} \tilde{\rho}_0^{(2)} = 0. \end{cases}$$

We firstly study from a numerical point of view if the background (order 0 in the asymptotic expansion) steady incompressible solution is preserved over time and secondly if the different schemes are accurate at low Mach number in the sense that the amplitude of the perturbation with respect to the background incompressible solution satisfies (7) at the discrete level.

In Figure 4 and Figure 5, we plot the norm of $\alpha\mathbf{u}$ obtained at time $t = 2$s with $M = 10^{-4}$ on Cartesian and triangular meshes with the well-balanced Lax-Friedrich, the VFRoe and the all-Mach VFRoe schemes. The incompressible steady velocity seems to be preserved over time with the all-Mach VFRoe scheme on triangular and Cartesian meshes and with the VFRoe scheme on triangular meshes. With the other schemes, the solution is extremely diffused. Note that the accuracy problem of the Lax-Friedrich scheme at low Mach number on triangular mesh was already illustrated in [36] for the uniform porosity case.

In Figure 6, we study the low Mach accuracy of the different numerical schemes in the sense that we check whether (7) is or is not satisfied at the discrete level. For that, we study the amplitude of the deviation of the numerical solution from the incompressible solution (which is the initial condition) with respect to the Mach number. We plot the norm of the deviation for the dimensionless density $\tilde{\rho}$ and the dimensionless field $\tilde{\alpha}\tilde{\mathbf{u}}$ for Mach numbers M ranging from 10^{-1}

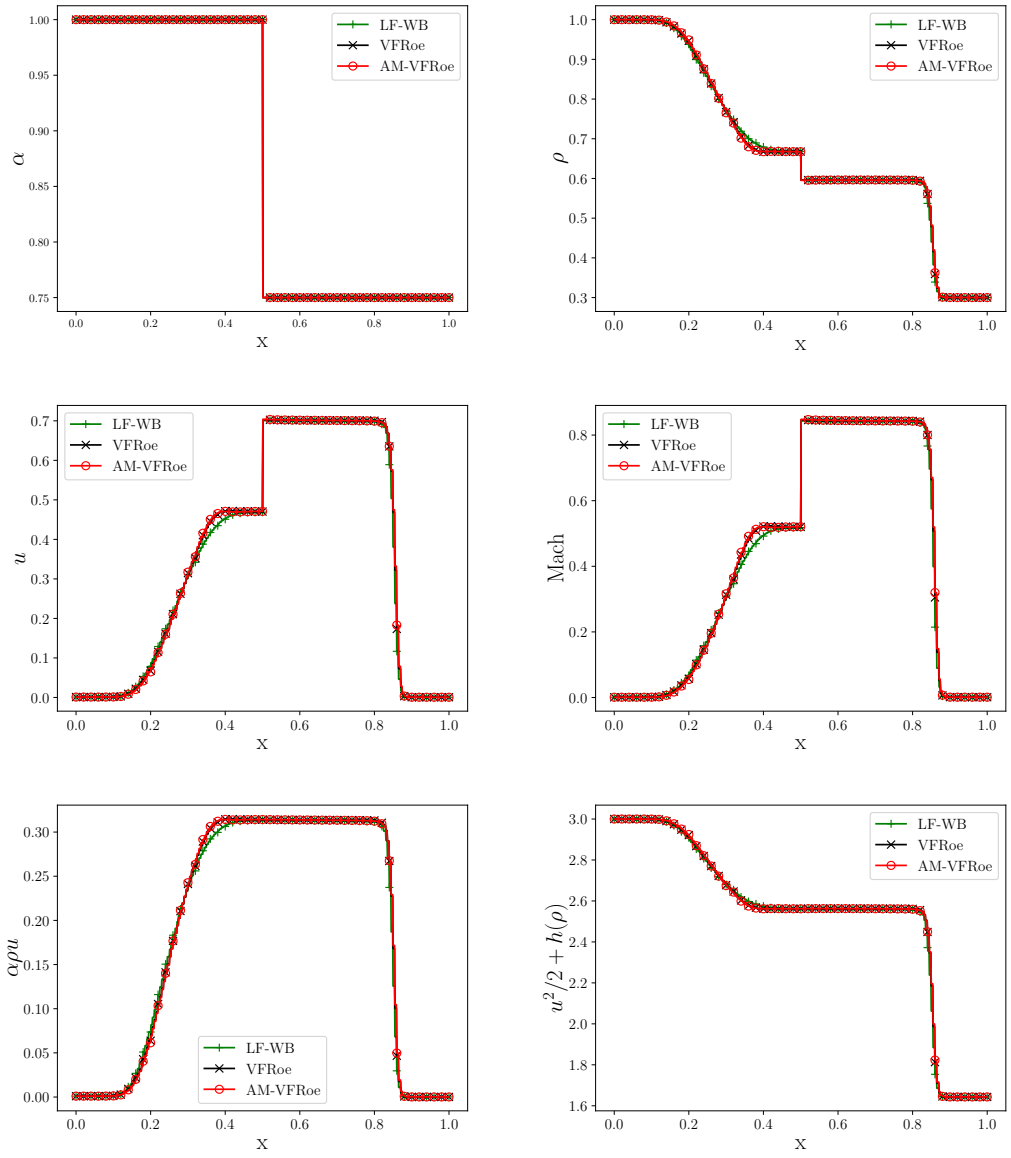
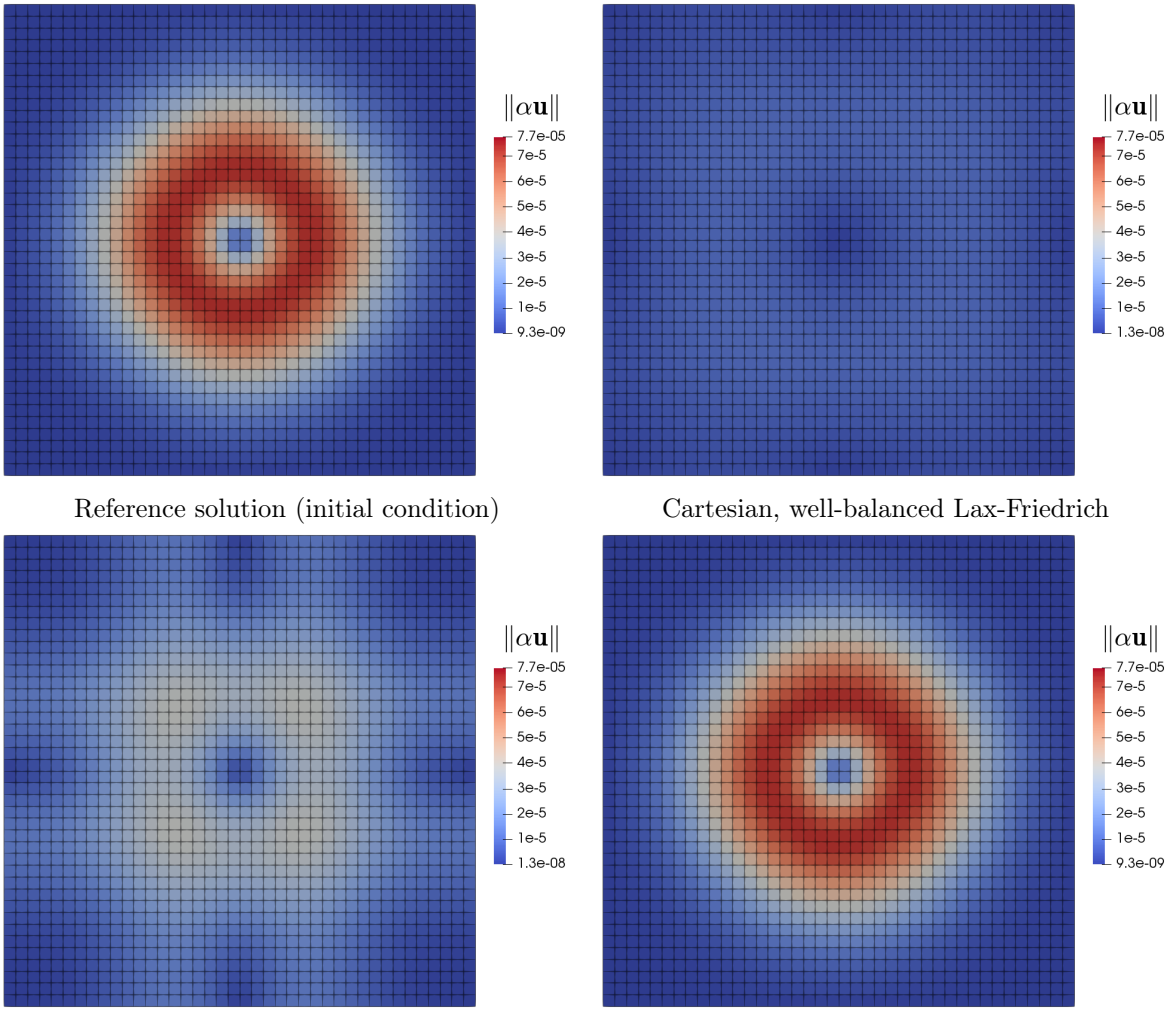


Figure 3: One dimensional unsteady subsonic flow : porosity (or cross section) α, density ρ, velocity u, Mach number, $\alpha\rho u$ and $u^2/2 + h(\rho)$ obtained at time $t = 0.25$ with the well-balanced Lax-Friedrich scheme (referenced by LF-WB), the VFRoe scheme and the all-Mach VFRoe scheme (referenced by AM-VFRoe). The mesh contains 200 cells.



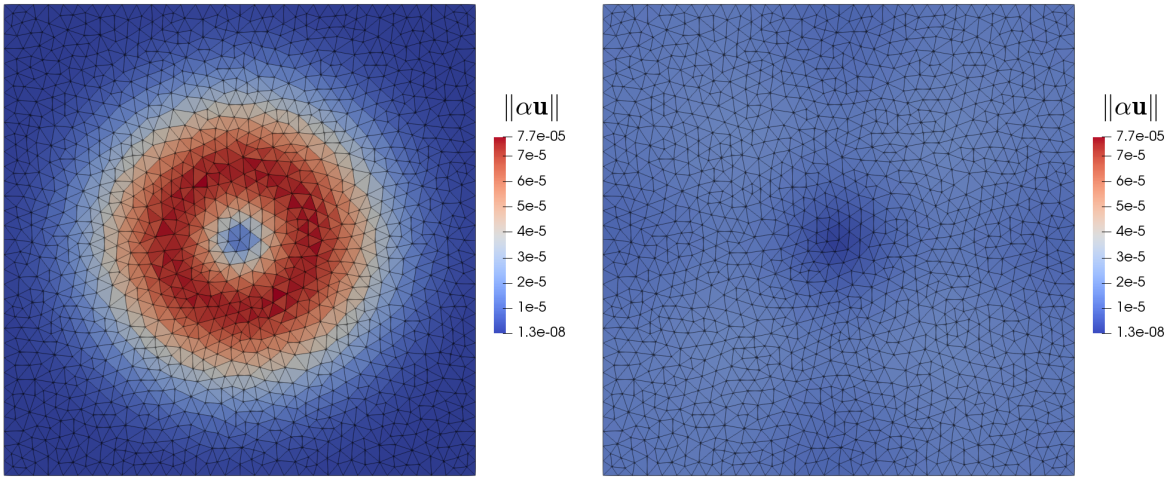
Reference solution (initial condition)

Cartesian, well-balanced Lax-Friedrich

Cartesian, VFRoe

Cartesian, all-Mach VFRoe

Figure 4: Norm of $\alpha\mathbf{u}$ obtained at final time $t = 2$s with the well-balanced Lax-Friedrich scheme (72), the VFRoe scheme (73) and the all-Mach VFRoe scheme (74) on Cartesian meshes with an initial Mach number $M = 10^{-4}$.



Reference solution (initial condition)

Triangles, well-balanced Lax-Friedrich

Triangles, VFRoe

Triangles, all-Mach VFRoe

Figure 5: Norm of $\alpha\mathbf{u}$ obtained at final time $t = 2$s with the well-balanced Lax-Friedrich scheme (72), the VFRoe scheme (73) and the all-Mach VFRoe scheme (74) on triangular meshes with an initial Mach number $M = 10^{-4}$.

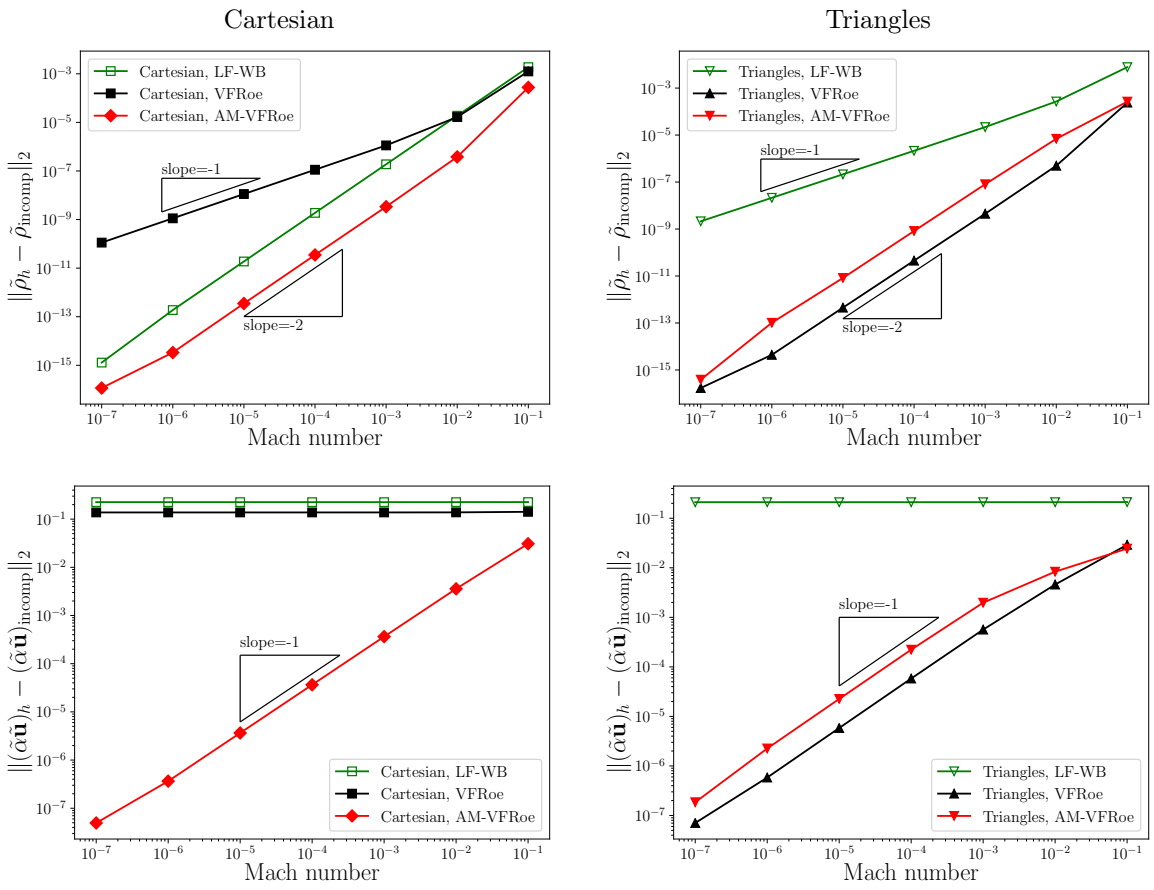


Figure 6: Norm of the deviation from the incompressible solution for the dimensionless density and dimensionless field $\tilde{\alpha}\tilde{\mathbf{u}}$ at final time $t = 2$s for Mach numbers M ranging from 10^{-1} to 10^{-7} with the well-balanced Lax-Friedrich scheme (72), the VFRoe scheme (73) and the all-Mach VFRoe scheme (74) on Cartesian and triangular meshes.

to 10^{-7}. Recall that $\tilde{\alpha}$, $\tilde{\rho}$ and $\tilde{\mathbf{u}}$ are defined by (2) where here $\alpha_0 = \rho_0 = 1$ and $u_0 = c(\rho_0) \times M$. Note that it is very important to initialize the field $\alpha\mathbf{u}$ and compute the incompressible solution by using (24) for triangular meshes and (29) for Cartesian meshes, in which the discrete values of ψ are interpolated from the analytical expression of ψ, because otherwise (i.e. if the discrete values of $\alpha\mathbf{u}$ are initialized directly from their analytical expression) an error of the order of the space step will be introduced and will hide the deviation that scales like the Mach number. We observe that the VFRoe scheme is accurate at low Mach number on triangles while the all-Mach VFRoe scheme is the only one which is accurate on Cartesian and triangular meshes. Indeed, for these schemes, the density deviation is of order M^2 and the velocity deviation is of order M, as expected. The well-balanced Lax-Friedrich scheme and the VFRoe scheme on Cartesian meshes are not accurate at low Mach number because their velocity deviation is of order M^0, and, moreover, the density deviation of the VFRoe scheme scales like M.

7 Conclusion

In this article, we proposed a well-balanced compressible scheme accurate at low Mach number for the Euler equations with porosity. The proposed scheme is based on the study that is performed on the linear wave equation with porosity. Indeed, the low Mach accuracy problem of the Godunov scheme can be understood and cured in the linear case. For this, we extended the discrete Hodge decomposition of [16] to a weighted L^2 space in order to take into account the porosity, and we extended to the discrete level the properties that were proven by studying the modified equation related to the Godunov scheme in [14]. We enlightened the influence of the cell geometry on the accuracy of this scheme. In the triangular case, the stationary space of the Godunov scheme approaches well enough the continuous space of constant pressures and divergence-free velocity fields (up to the porosity factor), while this is not the case in the Cartesian case. On Cartesian

meshes, we have to delete the usual numerical diffusion on the velocity field to preserve constant pressure fields and divergence-free velocity fields (up to the porosity factor). Moreover, as the aim was to design an all Mach regime scheme, the correction that is introduced varies continuously with respect to the Mach number. As a result, on Cartesian meshes, we propose to multiply the numerical diffusion on the velocity field by the Mach number M when M is smaller then 1. We check with numerical tests that this corrected scheme is accurate at low Mach number. Note that these conclusions are only valid when the boundary conditions are periodic: non-periodic boundary conditions may require additional analysis that was not performed in the present work.

The proposed non-linear scheme is based on a VFRoe solver and is a non linear extension of the Godunov scheme proposed for the linear case. The VFRoe solver avoids the exact resolution of a Rieamnn problem with variable porosity and is easy to implement. Like in the linear case, the VFRoe scheme for the non linear system is not accurate at low Mach number on Cartesian meshes but is accurate at low Mach number on triangular meshes. Based on the linear study, a fix is proposed for Cartesian meshes. This fix is easy to implement, requires only the modification of a few lines of code and allows to recover the accuracy at low Mach number on Cartesian meshes.

Further research could be driven by the following issues : First, if the porosity α is discontinuous, then care must be taken in the interpretation and the numerical treatment of System (1). Such questions are dealt with for example in [27] and we note that, in the particular case of Section 6.5.1, the scheme proposed in the present work computes a relevant numerical approximation. A second topic that needs to be studied is the extension of the approach presented here to the full Euler system with energy balance.

Appendices

A Kernel of the Godunov and modified Godunov schemes on general meshes

We prove Lemma 3.4:

Proof. The proof uses the fact that for any $q_h \in \operatorname{Ker} \mathbb{L}^h_{\kappa,\alpha}$ defined by (20), we have

$$\sum_{\Gamma_{ij}} |\Gamma_{ij}| \left[\alpha_{ij}(r_i - r_j)^2 + \frac{\kappa}{\alpha_{ij}} \Big(\big((\alpha \mathbf{u})_i - (\alpha \mathbf{u})_j \big) \cdot \mathbf{n}_{ij} \Big)^2 \right] = 0. \tag{75}$$

This equality is implied by the energy estimate (36) since any element in the kernel is stationary.

Since for all i, j we have $\alpha_{ij} > 0$, (75) leads to the fact $r_i = r_j$ for all neighboring cells (i, j) and thus

$$\exists c \in \mathbb{R},\ \forall i \in [\![1, N]\!], \quad r_i = c. \tag{76}$$

If $\kappa > 0$, we also deduce from (75) that

$$\forall i \in [\![1, N]\!],\ \forall j \in \{\text{neighboring cell of } i\}, \quad (\alpha \mathbf{u})_i \cdot \mathbf{n}_{ij} = (\alpha \mathbf{u})_j \cdot \mathbf{n}_{ij},$$

which allows to write (21). If $\kappa = 0$, we can only deduce (76) from (75). Nevertheless, by injecting $r_i = c$ in the first relation of (20), we obtain

$$\forall i \in [\![1, N]\!],\ \forall j \in \{\text{neighboring cell of } i\}, \quad \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \big((\alpha \mathbf{u})_i + (\alpha \mathbf{u})_j \big) \cdot \mathbf{n}_{ij} = 0,$$

which allows to write (22).

Let us prove that $\operatorname{Ker} \mathbb{L}^h_{\kappa>0,\alpha} \subsetneq \operatorname{Ker} \mathbb{L}^h_{\kappa=0,\alpha}$. Let $q_h \in \operatorname{Ker} \mathbb{L}^h_{\kappa>0,\alpha}$. We have for all $i \in [\![1, N]\!]$

$$\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \big((\alpha \mathbf{u})_i + (\alpha \mathbf{u})_j \big) \cdot \mathbf{n}_{ij} = \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \big((\alpha \mathbf{u})_i + (\alpha \mathbf{u})_i \big) \cdot \mathbf{n}_{ij} = 2(\alpha \mathbf{u})_i \cdot \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \mathbf{n}_{ij} = 0$$

and then $q_h \in \operatorname{Ker} \mathbb{L}^h_{\kappa=0,\alpha}$. □

B Discrete Poincaré-Wirtinger inequalities with weights

Proposition B.1. *(one-dimension) Let $\mathbb{T} =]a_1, b_1[$ be covered by a uniform rectangular mesh with N_x cells of size $\Delta x := \dfrac{b_1 - a_1}{N_x} = \dfrac{L_x}{N_x}$. Let $(\mu_i)_{1\leq i\leq N_x}$ be a positive sequence. Let $(\mu_{i-\frac{1}{2}})_{1\leq i\leq N_x}$ be a strictly positive sequence. Set $\|\mu\|_\infty = \sup_i(\mu_i)$ and $\left\|\dfrac{1}{\mu}\right\|_\infty = \sup_i\left(\mu_{i+\frac{1}{2}}^{-1}\right)$. Then, for any $(\phi_i)_{1\leq i\leq N_x}$ (for which we set for the sake of periodicity $\phi_0 = \phi_{N_x}$), there holds*

$$\sum_{i=1}^{N_x} \Delta x \mu_i \left(\phi_i - \bar{\phi}\right)^2 \leq \|\mu\|_\infty \left\|\frac{1}{\mu}\right\|_\infty \frac{L_x^2}{\Delta x} \sum_{k=1}^{N_x} |\phi_k - \phi_{k-1}|^2 \mu_{k-\frac{1}{2}}$$

where $\bar{\phi} = \dfrac{1}{L_\mu}\sum_{j=1}^{N_x} \Delta x \mu_j \phi_j$ with $L_\mu = \sum_{j=1}^{N_x} \Delta x \mu_j$.

Proof. For all $i \in [\![1, N_x]\!]$, we have

$$\begin{aligned}
\left|\phi_i - \bar{\phi}\right| &= \frac{1}{L_\mu}\left|\sum_{j=1}^{N_x} \Delta x \mu_j (\phi_i - \phi_j)\right| \\
&= \frac{1}{L_\mu}\left|\sum_{j=1}^{N_x} \Delta x \mu_j \sum_{k=j+1}^{i} (\phi_k - \phi_{k-1})\right| \leq \frac{1}{L_\mu}\sum_{j=1}^{N_x} \Delta x \mu_j \sum_{k=1}^{N_x} |\phi_k - \phi_{k-1}| = \sum_{k=1}^{N_x} |\phi_k - \phi_{k-1}|
\end{aligned}$$

and then, using the discrete Cauchy-Schwarz inequality, we get

$$\left(\phi_i - \bar{\phi}\right)^2 \leq N_x \sum_{k=1}^{N_x} |\phi_k - \phi_{k-1}|^2 \leq N_x \left\|\frac{1}{\mu}\right\|_\infty \sum_{k=1}^{N_x} \mu_{k-\frac{1}{2}} |\phi_k - \phi_{k-1}|^2 .$$

By multiplying by $\mu_i \Delta x$ and by summing over i, we have, since $L_x = N_x \Delta x$

$$\begin{aligned}
\sum_{i=1}^{N_x} \Delta x \mu_i \left(\phi_i - \bar{\phi}\right)^2 &\leq N_x \left\|\frac{1}{\mu}\right\|_\infty \sum_{i=1}^{N_x} \Delta x \mu_i \sum_{k=1}^{N_x} \mu_{k-\frac{1}{2}} |\phi_k - \phi_{k-1}|^2 \\
&\leq \|\mu\|_\infty \left\|\frac{1}{\mu}\right\|_\infty \frac{L_x^2}{\Delta x} \sum_{k=1}^{N_x} \mu_{k-\frac{1}{2}} |\phi_k - \phi_{k-1}|^2 .
\end{aligned}$$

□

Proposition B.2. *(two-dimensions) Let $\mathbb{T} =]a_1, b_1[\times]a_2, b_2[$ be covered by a uniform rectangular mesh with cell sizes $\Delta x \times \Delta y$ where $\Delta x = \dfrac{b_1 - a_1}{N_x} = \dfrac{L_x}{N_x}$ and $\Delta y = \dfrac{b_2 - a_2}{N_y} = \dfrac{L_y}{N_y}$. Let $(\mu_i)_{1\leq i\leq N_x, 1\leq j\leq N_y}$, $(\mu_{i-\frac{1}{2},j})_{1\leq i\leq N_x, 1\leq j\leq N_y}$ and $(\mu_{i,j-\frac{1}{2}})_{1\leq i\leq N_x, 1\leq j\leq N_y}$ be three strictly positive sequences. Set $\|\mu\|_\infty = \sup_{i,j}(\mu_{i,j})$ and $\left\|\dfrac{1}{\mu}\right\|_\infty = \sup_{i,j}\left(\mu_{i-\frac{1}{2},j}^{-1}, \mu_{i,j-\frac{1}{2}}^{-1}, \mu_{i,j}^{-1}\right)$. Then, for any $(\phi_{i,j})_{1\leq i\leq N_x, 1\leq j\leq N_y}$ (for which we set for the sake of periodicity $\phi_{0,j} = \phi_{N_x,j}$ and $\phi_{i,0} = \phi_{i,N_y}$), there holds*

$$\begin{aligned}
\sum_{i,j} \Delta x \Delta y \mu_{i,j} \left(\phi_{i,j} - \bar{\phi}\right)^2 \leq 2\|\mu\|_\infty^2 \left\|\frac{1}{\mu}\right\|_\infty^2 \left(\frac{L_x^2}{\Delta x} + \frac{L_y^2}{\Delta y}\right) \Bigg[&\sum_{i,j} \Delta x \mu_{i,j-\frac{1}{2}} |\phi_{i,j} - \phi_{i,j-1}|^2 \\
&+ \sum_{i,j} \Delta y \mu_{i-\frac{1}{2},j} |\phi_{i,j} - \phi_{i-1,j}|^2 \Bigg]
\end{aligned}$$

where $\bar{\phi} = \dfrac{1}{A_\mu}\sum_{k,\ell} \Delta x \Delta y \mu_{k,\ell} \phi_{k,\ell}$ with $A_\mu = \sum_{k,\ell} \Delta x \Delta y \mu_{k,\ell}$.

Proof. For all $(i,j) \in [\![1,N_x]\!] \times [\![1,N_y]\!]$, we have

$$\begin{aligned}
|\phi_{i,j} - \bar{\phi}| &= \left| \frac{1}{A_\mu} \sum_{k,\ell} \Delta x \Delta y \mu_{k,\ell} \left(\phi_{i,j} - \phi_{k,\ell}\right) \right| \\
&\leq \frac{1}{A_\mu} \sum_{k,\ell} \Delta x \Delta y \mu_{k,\ell} \left(|\phi_{i,j} - \phi_{i,\ell}| + |\phi_{i,\ell} - \phi_{k,\ell}| \right) \\
&\leq \frac{1}{A_\mu} \sum_{k,\ell} \Delta x \Delta y \mu_{k,\ell} \left(\sum_n |\phi_{i,n} - \phi_{i,n-1}| + \sum_m |\phi_{m,\ell} - \phi_{m-1,\ell}| \right).
\end{aligned}$$

We have to take into account that the sum over n does not depend on (k,ℓ), while the sum over m depends on ℓ but not on k. This provides:

$$\begin{aligned}
|\phi_{i,j} - \bar{\phi}| &\leq \sum_n |\phi_{i,n} - \phi_{i,n-1}| + \frac{1}{A_\mu} N_x \Delta x \Delta y \|\mu\|_\infty \sum_\ell \sum_m |\phi_{m,\ell} - \phi_{m-1,\ell}| \\
|\phi_{i,j} - \bar{\phi}|^2 &\leq 2 \left(\sum_n |\phi_{i,n} - \phi_{i,n-1}| \right)^2 + 2 \left(\frac{1}{A_\mu} N_x \Delta x \Delta y \|\mu\|_\infty \sum_{\ell,m} |\phi_{m,\ell} - \phi_{m-1,\ell}| \right)^2 . \quad (77)
\end{aligned}$$

The first term in the right-hand side of (77) depends on i but not on j, while the second does not depend on (i,j). This implies, on the one hand

$$\begin{aligned}
\sum_{i,j} \Delta x \Delta y \mu_{i,j} \left(\sum_n |\phi_{i,n} - \phi_{i,n-1}| \right)^2 &\leq \Delta x \Delta y \|\mu\|_\infty N_y \sum_i \left(\sum_n |\phi_{i,n} - \phi_{i,n-1}| \right)^2 \\
&\leq \Delta x \Delta y \|\mu\|_\infty N_y^2 \sum_{i,n} |\phi_{i,n} - \phi_{i,n-1}|^2 \\
&\leq \frac{L_y^2}{\Delta y} \|\mu\|_\infty \left\| \frac{1}{\mu} \right\|_\infty \sum_{i,n} \Delta x \mu_{i,n-\frac{1}{2}} |\phi_{i,n} - \phi_{i,n-1}|^2 \quad (78)
\end{aligned}$$

and, on the other hand

$$\begin{aligned}
&\sum_{i,j} \Delta x \Delta y \mu_{i,j} \left(\frac{1}{A_\mu} N_x \Delta x \Delta y \|\mu\|_\infty \sum_{\ell,m} |\phi_{m,\ell} - \phi_{m-1,\ell}| \right)^2 \\
&\leq \frac{1}{A_\mu} L_x^2 \Delta y^2 \|\mu\|_\infty^2 \left(\sum_{\ell,m} |\phi_{m,\ell} - \phi_{m-1,\ell}| \right)^2 \\
\leq\ & \frac{1}{L_x L_y} \left\| \frac{1}{\mu} \right\|_\infty L_x^2 \Delta y \, \|\mu\|_\infty^2 \, N_x N_y \left\| \frac{1}{\mu} \right\|_\infty \sum_{m,l} \Delta y \mu_{m-\frac{1}{2},\ell} \, |\phi_{m,\ell} - \phi_{m-1,\ell}|^2 \\
\leq\ & \frac{L_x^2}{\Delta x} \|\mu\|_\infty^2 \left\| \frac{1}{\mu} \right\|_\infty^2 \sum_{m,l} \Delta y \mu_{m-\frac{1}{2},\ell} \, |\phi_{m,\ell} - \phi_{m-1,\ell}|^2 . \quad (79)
\end{aligned}$$

The result follows from (77), (78), (79) and from the fact that $\|\mu\|_\infty \left\| \frac{1}{\mu} \right\|_\infty \geq 1$. □

C VFRoe scheme

We now detail how we obtain the solution $R(0^-, \mathbf{Y}_L, \mathbf{Y}_R, \mathbf{n})$ in $\xi/t = 0^-$ of the linearized Riemann problem that is used to compute the VFRoe flux (73). Since the variables used for the VFRoe scheme are $\mathbf{Y} = (\alpha, \rho, \alpha\rho\mathbf{u})^T$, we write system (70) as

$$\partial_t \mathbf{Y} + \sum_{i \in \{x,y,z\}} B_i(\mathbf{Y}) \partial_i \mathbf{Y} = 0$$

where

$$\sum_{i \in \{x,y,z\}} B_i(\mathbf{Y}) \mathbf{n}_i = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbf{n}^T/\alpha \\ -\rho \mathbf{u}\mathbf{u} \cdot \mathbf{n} & -\alpha \mathbf{u}\mathbf{u} \cdot \mathbf{n} + \alpha c^2 \mathbf{n} & \mathbf{u} \cdot \mathbf{n} I_d + \mathbf{u} \otimes \mathbf{n} \end{bmatrix}.$$

Then, $R(0^-, \mathbf{Y}_L, \mathbf{Y}_R, \mathbf{n})$ corresponds to the solution in $\xi/t = 0^-$ of the linearized Riemann problem

$$\partial_t \mathbf{Y} + (B(\hat{\mathbf{Y}}) \cdot \mathbf{n}) \partial_\xi \mathbf{Y} = 0 \quad \text{with} \quad \mathbf{Y}(\xi, t=0) = \left\{ \begin{array}{ll} \mathbf{Y}_L & \text{if } \xi < 0, \\ \mathbf{Y}_R & \text{elsewhere} \end{array} \right.$$

where $B(\mathbf{Y}) \cdot \mathbf{n} = \sum_{i \in \{x,y,z\}} B_i(\mathbf{Y}) \mathbf{n}_i$ and $\hat{\mathbf{Y}}$ is defined by

$$\hat{\alpha} = \frac{\alpha_L + \alpha_R}{2}, \quad \hat{\rho} = \frac{\rho_L + \rho_R}{2}, \quad \hat{\mathbf{u}} = \frac{\mathbf{u}_L + \mathbf{u}_R}{2} \quad \text{and} \quad \hat{c} = c(\hat{\rho}). \tag{80}$$

Eigenvalues of $B(\mathbf{Y}) \cdot \mathbf{n}$ are $\lambda_0 = 0$, $\lambda_1 = \mathbf{u} \cdot \mathbf{n} - c$, $\lambda_2 = \mathbf{u} \cdot \mathbf{n} + c$, $\lambda_3 = \lambda_4 = \mathbf{u} \cdot \mathbf{n}$ and the associated left $\mathbf{l}_i$ and right $\mathbf{r}_i$ eigenvectors are

$$\begin{array}{rcl} \mathbf{l}_0(\mathbf{Y}) & = & \dfrac{1}{\alpha\left(c^2 - (\mathbf{u}\cdot\mathbf{n})^2\right)} \left(1, 0, 0, 0\right), \\ \mathbf{l}_1(\mathbf{Y}) & = & \dfrac{1}{2\alpha c} \left(\dfrac{\rho(\mathbf{u}\cdot\mathbf{n})^2}{\mathbf{u}\cdot\mathbf{n} - c}, \alpha(\mathbf{u}\cdot\mathbf{n} + c), -\mathbf{n}^T \right), \\ \mathbf{l}_2(\mathbf{Y}) & = & -\dfrac{1}{2\alpha c} \left(\dfrac{\rho(\mathbf{u}\cdot\mathbf{n})^2}{\mathbf{u}\cdot\mathbf{n} + c}, \alpha(\mathbf{u}\cdot\mathbf{n} - c), -\mathbf{n}^T \right), \\ \mathbf{l}_3(\mathbf{Y}) & = & \left(-\rho\mathbf{u}\cdot\mathbf{t}^a, -\alpha\mathbf{u}\cdot\mathbf{t}^a, (\mathbf{t}^a)^T\right), \\ \mathbf{l}_4(\mathbf{Y}) & = & \left(-\rho\mathbf{u}\cdot\mathbf{t}^b, -\alpha\mathbf{u}\cdot\mathbf{t}^b, (\mathbf{t}^b)^T\right) \end{array}$$

and

$$[\mathbf{r}_0|\mathbf{r}_1|\mathbf{r}_2|\mathbf{r}_3|\mathbf{r}_4](\mathbf{Y}) = \left[\begin{array}{ccccc} \alpha\left(c^2 - (\mathbf{u}\cdot\mathbf{n})^2\right) & 0 & 0 & 0 & 0 \\ \rho(\mathbf{u}\cdot\mathbf{n})^2 & 1 & 1 & 0 & 0 \\ \alpha\rho c^2(\mathbf{u} - (\mathbf{u}\cdot\mathbf{n})\mathbf{n}) & \alpha(\mathbf{u} - c\mathbf{n}) & \alpha(\mathbf{u} + c\mathbf{n}) & \mathbf{t}^a & \mathbf{t}^b \end{array} \right].$$

The solution $R(0^-, \mathbf{Y}_L, \mathbf{Y}_R, \mathbf{n})$ is given by

$$R(0^-, \mathbf{Y}_L, \mathbf{Y}_R, \mathbf{n}) = \mathbf{Y}_L + \sum_{\lambda_i < 0} \mathbf{l}_i(\hat{\mathbf{Y}})(\mathbf{Y}_R - \mathbf{Y}_L)\mathbf{r}_i(\hat{\mathbf{Y}}).$$

References

[1] Douglas N. Arnold and Richard S. Falk. A uniformly accurate finite element method for the Reissner–Mindlin plate. *SIAM J. Numer. Anal.*, 26:1276–1290, 1989.

[2] Wasilij Barsukow, Philipp VF Edelmann, Christian Klingenberg, Fabian Miczek, and Friedrich K Röpke. A numerical scheme for the compressible low-Mach number regime of ideal fluid dynamics. *Journal of Scientific Computing*, 72(2):623–646, 2017.

[3] Wasilij Barsukow, Philipp VF Edelmann, Christian Klingenberg, and Friedrich K Röpke. A low Mach Roe-type solver for the Euler equations allowing for gravity source terms. *ESAIM: Proceedings and Surveys*, 58:27–39, 2017.

[4] Jacob Bear and Yehuda Bachmat. *Introduction to modeling of transport phenomena in porous media*, volume 4. Springer Science & Business Media, 2012.

[5] Pascal Bruel, Simon Delmas, Jonathan Jung, and Vincent Perrier. A low Mach correction able to deal with low Mach acoustics. *Journal of Computational Physics*, 378:723–759, 2019.

[6] Thierry Buffard, Thierry Gallouët, and Jean-Marc Hérard. A sequel to a rough Godunov scheme: application to real gases. *Computers & fluids*, 29(7):813–847, 2000.

[7] Christophe Chalons, Mathieu Girardin, and Samuel Kokh. An all-regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes. *Communications in Computational Physics*, 20(1):188–233, 2016.

[8] Stéphane Clain and David Rochette. First-and second-order finite volume methods for the one-dimensional nonconservative Euler system. *Journal of computational Physics*, 228(22):8214–8248, 2009.

[9] Phillip Colella and Karen Pao. A projection method for low speed flows. *Journal of Computational Physics*, 149(2):245–269, 1999.

[10] Dao Huy Cuong and Mai Duc Thanh. A Godunov-type scheme for the isentropic model of a fluid flow in a nozzle with variable cross-section. *Applied Mathematics and Computation*, 256:602–629, 2015.

[11] Gianni Dal Maso, Philippe G Lefloch, and François Murat. Definition and weak stability of nonconservative products. *Journal de mathématiques pures et appliquées*, 74(6):483–548, 1995.

[12] Pierre Degond and Min Tang. All speed scheme for the low Mach number limit of the isentropic Euler equations. *Communications in Computational Physics*, 10(1):1–31, 2011.

[13] Stéphane Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *Journal of Computational Physics*, 4(229):978–1016, 2010.

[14] Stéphane Dellacherie, Jonathan Jung, and Pascal Omnes. Preliminary results for the study of the Godunov scheme applied to the linear wave equation with porosity at low Mach number. *ESAIM: Proceedings and Surveys*, 52:105–126, 2015.

[15] Stéphane Dellacherie, Jonathan Jung, Pascal Omnes, and Pierre-Arnaud Raviart. Construction of modified Godunov type schemes accurate at any Mach number for the compressible Euler system. *Math. Models Methods Appl. Sci.*, 26:2525–2615, 2016.

[16] Stéphane Dellacherie, Pascal Omnes, and Felix Rieper. The influence of cell geometry on the Godunov scheme applied to the linear wave equation. *Journal of Computational Physics*, 229(14):5315–5338, 2010.

[17] Thierry Gallouët, Jean-Marc Hérard, and Nicolas Seguin. Some approximate Godunov schemes to compute shallow-water equations with topography. *Computers & Fluids*, 32(4):479–513, 2003.

[18] Thierry Gallouet and Jean Marie Masella. Un schéma de Godunov approché. *Comptes rendus de l'Académie des sciences Paris Série 1*, 323:77–84, 1996.

[19] Christophe Geuzaine and Jean-François Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre-and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, 2009.

[20] Laurent Gosse. A well-balanced scheme using non-conservative products designed for hyperbolic systems of conservation laws with source terms. *Mathematical Models and Methods in Applied Sciences*, 11(02):339–365, 2001.

[21] Joshua M Greenberg and Alain-Yves LeRoux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM Journal on Numerical Analysis*, 33(1):1–16, 1996.

[22] Hervé Guillard. On the behavior of upwind schemes in the low Mach number limit. IV: P0 approximation on triangular and tetrahedral cells. *Computers & fluids*, 38(10):1969–1972, 2009.

[23] Hervé Guillard and Cécile Viozat. On the behaviour of upwind schemes in the low Mach number limit. *Computers & fluids*, 28(1):63–86, 1999.

[24] David Iampietro, Frédéric Daude, Pascal Galon, and Jean-Marc Hérard. A Mach-sensitive splitting approach for Euler-like systems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 52:207–253, 2018.

[25] Sergiu Klainerman and Andrew Majda. Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. *Communications on pure and applied Mathematics*, 34(4):481–524, 1981.

[26] Dietmar Kröner and Mai Duc Thanh. Numerical solutions to compressible flows in a nozzle with variable cross-section. *SIAM Journal on Numerical Analysis*, 43(2):796–824, 2005.

[27] Philippe G Lefloch and Mai Duc Thanh. The Riemann problem for fluid flows in a nozzle with discontinuous cross-section. *Communications in Mathematical Sciences*, 1(4):763–797, 2003.

[28] Xue-song Li and Chun-wei Gu. An all-speed Roe-type scheme and its asymptotic analysis of low Mach number behaviour. *Journal of Computational Physics*, 227(10):5144–5159, 2008.

[29] Xue-song Li, Chun-wei Gu, and Jian-zhong Xu. Development of Roe-type scheme for all-speed flows based on preconditioning method. *Computers & Fluids*, 38(4):810–817, 2009.

[30] Andrew Majda. *Compressible fluid flow and systems of conservation laws in several space variables*, volume 53. Springer Science & Business Media, 2012.

[31] Roy A. Nicolaides. Analysis and convergence of the MAC scheme. I: The linear problem. *SIAM J. Numer. Anal.*, 29:1579–1591, 1992.

[32] Roy A. Nicolaides. Direct discretization of planar div-curl problems. *SIAM J. Numer. Anal.*, 29:32–56, 1992.

[33] K. Oß wald, A. Siegmund, P. Birken, V. Hannemann, and A. Meister. L^2Roe: a low dissipation version of Roe's approximate Riemann solver for low Mach numbers. *Internat. J. Numer. Methods Fluids*, 81(2):71–86, 2016.

[34] Marica Pelanti. Low Mach number preconditioning techniques for Roe-type and HLLC-type methods for a two-phase compressible flow model. *Applied Mathematics and Computation*, 310:112–133, 2017.

[35] Marica Pelanti and Keh-Ming Shyue. A Roe-type scheme with low Mach number preconditioning for a two-phase compressible flow model with pressure relaxation. *Bulletin of the Brazilian Mathematical Society, New Series*, 47(2):655–669, 2016.

[36] Felix Rieper. On the dissipation mechanism of upwind-schemes in the low Mach number regime: A comparison between Roe and HLL. *Journal of Computational Physics*, 229(2):221–232, 2010.

[37] Felix Rieper. A low-Mach number fix for Roe's approximate Riemann solver. *Journal of Computational Physics*, 230(13):5263–5287, 2011.

[38] Felix Rieper and Georg Bader. The influence of cell geometry on the accuracy of upwind schemes in the low Mach number regime. *Journal of Computational Physics*, 228(8):2918–2933, 2009.

[39] David Rochette and Stéphane Clain. Two-dimensional computation of gas flow in a porous bed characterized by a porosity jump. *Journal of Computational Physics*, 219(1):104–119, 2006.

[40] David Rochette, Stéphane Clain, and Thierry Buffard. Numerical scheme to complete a compressible gas flow in variable porosity media. *International Journal of Computational Fluid Dynamics*, 19(4):299–309, 2005.

[41] David Rochette, Stéphane Clain, and François Gentils. Numerical investigations on the pressure wave absorption and the gas cooling interacting in a porous filter, during an internal arc fault in a medium-voltage cell. *IEEE transactions on power delivery*, 23(1):203–212, 2007.

[42] Steven Schochet. Fast singular limits of hyperbolic PDEs. *Journal of differential equations*, 114(2):476–512, 1994.

[43] Chi-Wang Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In *Advanced numerical approximation of nonlinear hyperbolic equations*, pages 325–432. Springer, 1998.

[44] Seth C Spiegel, HT Huynh, and James R DeBonis. A survey of the isentropic Euler vortex problem using high-order methods. In *22nd AIAA Computational Fluid Dynamics Conference*, AIAA paper 2015-2444, 2015.

[45] Andrea Thomann, Gabriella Puppo, and Christian Klingenberg. An all speed second order well-balanced IMEX relaxation scheme for the Euler equations with gravity. *Journal of Computational Physics*, 420:109723, 2020.

[46] Eli Turkel. Preconditioning techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*, 31(1):385–416, 1999.