Preliminary results for the study of the Godunov Scheme Applied to the Linear Wave Equation with Porosity at Low Mach Number

Stéphane Dellacherie, Jonathan Jung, Pascal Omnes

Preliminary results for the study of the Godunov Scheme Applied to the Linear Wave Equation with Porosity at Low Mach Number

Stéphane Dellacherie[*], Jonathan Jung[†] and Pascal Omnes[‡]

December 3, 2014

Abstract

We introduce continuous tools to study the low Mach number behaviour of the Godunov scheme applied to the linear wave equation with porosity on cartesian meshes. More precisely, we extend the Hodge decomposition to a weighted L^2 space in the continuous case and we study the properties of the modified equation associated to this Godunov scheme. This allows to partly explain the inaccuracy of the Godunov scheme at low Mach number on cartesian meshes and to propose two corrections: a first one named *low Mach* and a second one named *all Mach*. These results are preliminary since it remains to prove them in the discrete case.

1 Linear wave equation with porosity

The dimensionless barotropic Euler system with porosity may be written as

$$\begin{cases} \partial_t(\alpha\rho) + \nabla\cdot(\alpha\rho\mathbf{u}) = 0, \\ \\ \partial_t(\alpha\rho\mathbf{u}) + \nabla\cdot(\alpha\rho\mathbf{u}\otimes\mathbf{u}) + \dfrac{\alpha}{M^2}\nabla p = 0. \end{cases} \tag{1}$$

In (1), M is the Mach number that is supposed to be small and $\alpha(\mathbf{x})$ is the porosity, $t \geq 0$ and $\mathbf{x} \in \Omega$ are respectively the time and space variables. The quantities ρ, $\mathbf{u}$ and $p(\rho)$ are respectively the density, the velocity field and the pressure law of the fluid. We assume that the pressure law satisfies $p'(\rho) > 0$ and that $\alpha(\mathbf{x})$ is a known function that takes its values in $[\alpha_{\min}, 1]$, where $\alpha_{\min} > 0$ is a constant which does not depend on M.

When the geometry is 2D or 3D and the cells of the mesh are not triangular (in 2D) nor tetrahedral (in 3D), finite volume Godunov type schemes applied to (1) with periodic boundary conditions are known to be inaccurate at low Mach number when $\nabla\alpha = 0$ [1, 5, 3], contrarily to staggered schemes on cartesian meshes [5, 6]. To better understand this behaviour when $\nabla\alpha \neq 0$ and to propose a *low Mach correction* (when it is necessary) and an *all Mach correction* (when we want to recover the Godunov scheme for Mach numbers of order one) in the spirit of what is done in [3], we introduce tools adapted to a linearization of (1) around $(\rho = \rho_\star, \mathbf{u} = 0)$ when Ω is periodic and we extend to the 2D/3D case some 1D results proposed in [4]. For this purpose, we set the reference sound speed to $\frac{a_\star}{M}$ $(a_\star^2 = p'(\rho_\star))$ and we define $r(t, \mathbf{x})$ such as

$$\rho(t,\mathbf{x}) := \rho_\star\left(1 + \frac{M}{a_\star} r(t,\mathbf{x})\right) \tag{2}$$

[*] CEA, DEN, DM2S, STMF F-91191, Gif-sur-Yvette, France and Université Pierre et Marie Curie, LRC Manon and LJLL, 4 place Jussieu, 75252 Paris, cedex 05, France, stephane.dellacherie@cea.fr

[†] EFREI, 30-32 avenue de la République, 94800 Villejuif, France and Université Pierre et Marie Curie, LRC Manon and LJLL, 4 place Jussieu, 75252 Paris, cedex 05, France, jonathan.jung@ljll.math.upmc.fr

[‡] CEA, DEN, DM2S, STMF F-91191, Gif-sur-Yvette, France and Université Paris 13, Sorbonne Paris Cité, LAGA, CNRS UMR 7539, 99, Avenue J.-B. Clément F-93430 Villetaneuse Cedex, France, pascal.omnes@cea.fr

where formally $\frac{M}{a_\star} r \ll 1$. By injecting (2) in (1), we obtain the system

$$\begin{cases} \partial_t(\alpha r) + \nabla\cdot(\alpha r\mathbf{u}) + \dfrac{a_\star}{M}\nabla\cdot(\alpha\mathbf{u}) = 0, \\ \partial_t(\alpha\mathbf{u}) + (\alpha\mathbf{u}\cdot\nabla)\mathbf{u} + \dfrac{\alpha}{M}\dfrac{p'\left(\rho_\star\left(1+\frac{M}{a_\star}r\right)\right)}{a_\star\left(1+\frac{M}{a_\star}r\right)}\nabla r = 0. \end{cases}$$

By linearizing around $(r,\mathbf{u}) = (0,0)$, we obtain the linear wave equation with porosity

$$\partial_t(\alpha q) + \frac{L_\alpha}{M}(q) = 0 \quad \text{where} \quad q = \begin{pmatrix} r \\ \mathbf{u} \end{pmatrix} \quad \text{and} \quad L_\alpha(q) = a_\star \begin{pmatrix} \nabla\cdot(\alpha\mathbf{u}) \\ \alpha\nabla r \end{pmatrix}. \tag{3}$$

2 Weighted spaces $\mathcal{E}_\alpha$ and $\mathcal{E}_\alpha^\perp$

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

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

endowed with the scalar product

$$\langle q_1, q_2\rangle_\alpha = \int_{\mathbb{T}} r_1 r_2 \alpha dx + \int_{\mathbb{T}} \mathbf{u}_1\cdot\mathbf{u}_2\alpha dx. \tag{4}$$

Of course, the space L^2_α must not be confused with the acoustic operator L_α. We use the same notation to 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 (4). At last, we define the space

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

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

Lemma 2.1. *We have*

$$\begin{aligned} \mathcal{E}_\alpha^\perp &= \left\{ q = (r,\boldsymbol{u})^T \in L^2_\alpha\left(\mathbb{T}\right)^{1+d} \Big| \int_{\mathbb{T}} r\alpha dx = 0 \text{ and } \exists\phi\in H^1_\alpha\left(\mathbb{T}\right),\ \boldsymbol{u} = \nabla\phi \right\}, & (5) \\ \mathcal{E}_\alpha &\oplus \mathcal{E}_\alpha^\perp = L^2_\alpha\left(\mathbb{T}\right)^{1+d}. & (6) \end{aligned}$$

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

$$q = \hat{q} + q^\perp \tag{7}$$

where $\hat{q} = (\hat{r},\hat{\boldsymbol{u}})^T \in \mathcal{E}_\alpha$ and $q^\perp = (r^\perp,\boldsymbol{u}^\perp)^T \in \mathcal{E}_\alpha^\perp$, this decomposition is unique and orthogonal with respect to the scalar product defined by (4).

We call $\mathcal{E}_\alpha^\perp$ the acoustic space. This is a generalization of the Hodge decomposition to the weighted space $L^2_\alpha\left(\mathbb{T}\right)^{1+d}$. The decomposition (7) defines the orthogonal projection

$$\begin{aligned} \mathbb{P}_\alpha :\ L^2_\alpha\left(\mathbb{T}\right)^{1+d} &\longrightarrow \mathcal{E}_\alpha \\ q &\longmapsto \mathbb{P}_\alpha q := \hat{q}. \end{aligned} \tag{8}$$

Proof. We firstly prove (5). We note A the space

$$A := \left\{ q = (r,\mathbf{u})^T \in L^2_\alpha\left(\mathbb{T}\right)^{1+d} \quad \Big| \quad \int_{\mathbb{T}} r\alpha dx = 0 \text{ and } \exists\phi\in H^1_\alpha\left(\mathbb{T}\right),\ \mathbf{u} = \nabla\phi \right\}.$$

Firstly, we prove that $A \subset \mathcal{E}_\alpha^\perp$ and, secondly, we prove that $\mathcal{E}_\alpha^\perp \subset A$. Let $q_1 = (r_1, \mathbf{u}_1)^T \in A$. For all $q_2 = (r_2, \mathbf{u}_2)^T \in \mathcal{E}_\alpha$, we have

$$\begin{aligned} \langle q_1, q_2 \rangle_\alpha &= \int_{\mathbb{T}} r_1 r_2 \alpha dx + \int_{\mathbb{T}} \mathbf{u}_1 \cdot \mathbf{u}_2 \alpha dx = r_2 \int_{\mathbb{T}} r_1 \alpha dx + \int_{\mathbb{T}} \nabla \phi_1 \cdot \mathbf{u}_2 \alpha dx \\ &= 0 + \int_{\partial \mathbb{T}} \phi_1 (\alpha \mathbf{u}_2) \cdot \mathbf{n} d\sigma - \int_{\mathbb{T}} \phi_1 \nabla \cdot (\alpha \mathbf{u}_2) dx = \int_{\partial \mathbb{T}} \phi_1 (\alpha \mathbf{u}_2) \cdot \mathbf{n} d\sigma = 0 \end{aligned}$$

because $q_2 \in \mathcal{E}_\alpha$ and $\phi_1(\alpha \mathbf{u}_2)$ is periodic. This proves that $A \subset \mathcal{E}_\alpha^\perp$. Let $q_1 = (r_1, \mathbf{u}_1)^T \in \mathcal{E}_\alpha^\perp$. For all $q_2 = (r_2, \mathbf{u}_2)^T \in \mathcal{E}_\alpha$, we have

$$\langle q_1, q_2 \rangle_\alpha = 0 \quad \Longrightarrow \quad \int_{\mathbb{T}} r_1 r_2 \alpha dx + \int_{\mathbb{T}} \mathbf{u}_1 \cdot \mathbf{u}_2 \alpha dx = 0 \quad \Longrightarrow \quad r_2 \int_{\mathbb{T}} r_1 \alpha dx + \int_{\mathbb{T}} \mathbf{u}_1 \cdot (\alpha \mathbf{u}_2) dx = 0.$$

Then $\int_{\mathbb{T}} r_1 \alpha dx = 0$ and $\int_{\mathbb{T}} \mathbf{u}_1 \cdot (\alpha \mathbf{u}_2) dx = 0$ for all $\mathbf{u}_2 \in L^2_\alpha(\mathbb{T})^d$ such that $\nabla \cdot (\alpha \mathbf{u}_2) = 0$. Moreover, since α and $\frac{1}{\alpha}$ are in $L^\infty(\mathbb{T})$, the last equality is equivalent to $\int_{\mathbb{T}} \mathbf{u}_1 \cdot \tilde{\mathbf{u}}_2 dx = 0$ for all $\tilde{\mathbf{u}}_2 \in \mathcal{E} := \left\{ \mathbf{u} \in L^2(\mathbb{T})^d | \nabla \cdot \mathbf{u} = 0 \right\}$. Therefore, $\mathbf{u}_1 \in \mathcal{E}^\perp$ for the classical $L^2(\mathbb{T})^d$ scalar product. It is a classical result that

$$\mathcal{E}^\perp = \left\{ \mathbf{u} \in L^2(\mathbb{T})^d | \exists \phi \in H^1(\mathbb{T}), \mathbf{u} = \nabla \phi \right\}.$$

This implies that $\exists \phi_1 \in H^1_\alpha(\mathbb{T})$ such that $\mathbf{u}_1 = \nabla \phi_1$, which allows to write that $\mathcal{E}_\alpha^\perp \subset A$. To conclude, we have $\mathcal{E}_\alpha^\perp = A$.
Now, we prove (6). Since the inclusion $\subset$ is trivial, we just have to prove that $\mathcal{E}_\alpha \oplus \mathcal{E}_\alpha^\perp \supset L^2_\alpha(\mathbb{T})^{1+d}$. Let $q = (r, \mathbf{u})^T \in L^2_\alpha(\mathbb{T})^{1+d}$. We dissociate the construction of $(\hat{r}, r^\perp)$ from that of $(\hat{\mathbf{u}}, \mathbf{u}^\perp)$. For r, we can define $\hat{r} = \frac{1}{\int_{\mathbb{T}} \alpha dx} \int_{\mathbb{T}} r \alpha dx$ since $\alpha \geq \alpha_{\min} > 0$ implies that $\int_{\mathbb{T}} \alpha d\mathbf{x} > 0$. Moreover, $\alpha \in]0, 1]$ implies that $\int_{\mathbb{T}} \alpha d\mathbf{x} \leq \|\alpha\|_\infty |\mathbb{T}| < +\infty$ and with the Cauchy-Schwarz inequality, we obtain

$$\left| \int_{\mathbb{T}} r \alpha d\mathbf{x} \right| \leq \left(\int_{\mathbb{T}} r^2 \alpha d\mathbf{x} \right)^{\frac{1}{2}} \left(\int_{\mathbb{T}} \alpha d\mathbf{x} \right)^{\frac{1}{2}} \leq \|r\|_{L^2_\alpha} \|\alpha\|_\infty^{\frac{1}{2}} |\mathbb{T}|^{\frac{1}{2}} < +\infty.$$

Then, since

$$\int_{\mathbb{T}} (r - \hat{r}) \alpha dx = \int_{\mathbb{T}} r \alpha dx - \hat{r} \int_{\mathbb{T}} \alpha dx = 0,$$

we can write $r = \hat{r} + (r - \hat{r})$ with $\nabla \hat{r} = 0$ and $\int_{\mathbb{T}} (r - \hat{r}) \alpha dx = 0$ which gives the decomposition for r. For $\mathbf{u}$, the construction is slightly more difficult. We want to construct $\hat{\mathbf{u}}$ and $\nabla \phi$ such that $\mathbf{u} = \hat{\mathbf{u}} + \nabla \phi$ with $\nabla \cdot (\alpha \hat{\mathbf{u}}) = 0$. It is then sufficient to prove that there exists $\phi \in H^1_\alpha(\mathbb{T})$ such that

$$\begin{cases} \nabla \cdot (\alpha \nabla \phi) = \nabla \cdot (\alpha \mathbf{u}), \\ \displaystyle\int_{\mathbb{T}} \phi \alpha dx = 0 \end{cases} \tag{9}$$

and to set $\hat{\mathbf{u}} = \mathbf{u} - \nabla \phi$. We note $H^1_{\alpha,0}(\mathbb{T}) \subset H^1_\alpha(\mathbb{T})$ the subset of functions ϕ such that $\int_{\mathbb{T}} \phi \alpha dx = 0$. We write (9) under variational form:

$$\textit{Find } \phi \in H^1_{\alpha,0}(\mathbb{T}) \quad \textit{such that} \quad \forall \psi \in H^1_{\alpha,0}(\mathbb{T}): \quad a(\phi, \psi) := \int_{\mathbb{T}} (\alpha \nabla \phi) \cdot \nabla \psi d\mathbf{x} = \int_{\mathbb{T}} (\alpha \mathbf{u}) \cdot \nabla \psi d\mathbf{x} =: L(\psi).$$

Moreover, L is a continuous linear functional on the Hilbert space $H^1_{\alpha,0}(\mathbb{T})$ because

$$|L(\psi)| = \left| \int_{\mathbb{T}} \sqrt{\alpha} \mathbf{u} \cdot \sqrt{\alpha} \nabla \psi d\mathbf{x} \right| \leq \|\mathbf{u}\|_{L^2_\alpha} \|\nabla \psi\|_{L^2_\alpha} \leq \|\mathbf{u}\|_{L^2_\alpha} \|\psi\|_{H^1_\alpha}.$$

By using a similar argument, we also prove that $a(\cdot, \cdot)$ is a symmetric bilinear form that is continuous on $H^1_{\alpha,0}(\mathbb{T})$. To prove the coercivity of a, we use a generalization of the Poincaré-Wirtinger inequality to the probabilistic measure $\mu := \frac{\alpha}{\int_{\mathbb{T}} \alpha d\mathbf{x}}$ on the convex space $\mathbb{T}$ (see Appendix). As μ and $\frac{1}{\mu}$ are in $L^\infty(\mathbb{T})$, for $\phi \in H^1_\alpha(\mathbb{T})$ and $\bar{\phi} := \frac{1}{\int_{\mathbb{T}} \alpha d\mathbf{x}} \int_{\mathbb{T}} \phi \alpha d\mathbf{x}$, we have

$$\int_{\mathbb{T}} |\phi - \bar{\phi}|^2 \mu d\mathbf{x} \leq 2\text{diam}(\mathbb{T})^2 \|\mu\|_\infty \left\| \frac{1}{\mu} \right\|_\infty \int_{\mathbb{T}} |\nabla \phi|^2 \mu d\mathbf{x} \tag{10}$$

where $\text{diam}(\mathbb{T}) := \sup\limits_{(x,y)\in\mathbb{T}^2} |x-y|$, which is equivalent to $\displaystyle\int_{\mathbb{T}} |\phi - \bar{\phi}|^2 \alpha d\mathbf{x} \leq 2\text{diam}(\mathbb{T})^2 \|\alpha\|_\infty \left\| \frac{1}{\alpha} \right\|_\infty \int_{\mathbb{T}} |\nabla \phi|^2 \alpha d\mathbf{x}$. Thus, we can write that

$$C_\alpha(\mathbb{T}) \int_{\mathbb{T}} |\phi - \bar{\phi}|^2 \alpha d\mathbf{x} \leq \int_{\mathbb{T}} |\nabla \phi|^2 \alpha d\mathbf{x} \tag{11}$$

with $C_\alpha(\mathbb{T}) := \dfrac{1}{2\text{diam}(\mathbb{T})^2 \|\alpha\|_\infty \left\| \frac{1}{\alpha} \right\|_\infty} > 0$. For $\phi \in H^1_{\alpha,0}(\mathbb{T})$, we have $\bar{\phi} = 0$ and we can write

$$\begin{aligned} a(\phi,\phi) &= \int_{\mathbb{T}} |\nabla \phi|^2 \alpha d\mathbf{x} = \frac{1}{2} \int_{\mathbb{T}} |\nabla \phi|^2 \alpha d\mathbf{x} + \frac{1}{2} \int_{\mathbb{T}} |\nabla \phi|^2 \alpha d\mathbf{x} \\ &\geq \frac{1}{2} \int_{\mathbb{T}} |\nabla \phi|^2 \alpha d\mathbf{x} + \frac{C_\alpha(\mathbb{T})}{2} \int_{\mathbb{T}} |\phi|^2 \alpha d\mathbf{x} \geq \frac{1}{2} \min\left(1, C_\alpha\left(\mathbb{T}\right)\right) \|\phi\|^2_{H^1_\alpha} \end{aligned}$$

which means that $a(\cdot,\cdot)$ is coercive. Then, by applying the Lax-Milgram theorem, we obtain the existence of a unique function ϕ in $H^1_{\alpha,0}(\mathbb{T})$ such that $\forall \psi \in H^1_{\alpha,0}(\mathbb{T})$, $a(\phi,\psi) = L(\psi)$ that is to say

$$\forall \psi \in H^1_{\alpha,0}(\mathbb{T}): \quad \int_{\mathbb{T}} (\alpha \nabla \phi) \cdot \nabla \psi d\mathbf{x} = \int_{\mathbb{T}} (\alpha \mathbf{u}) \cdot \nabla \psi d\mathbf{x}. \tag{12}$$

We note $D(\mathbb{T})$ the set of functions $C^\infty(\mathbb{T})$ with a compact support. For all ψ in $D(\mathbb{T})$, the function

$$\tilde{\psi} := \psi - \bar{\psi} = \psi - \frac{\displaystyle\int_{\mathbb{T}} \psi \alpha d\mathbf{x}}{\displaystyle\int_{\mathbb{T}} \alpha d\mathbf{x}}$$

is in $H^1_{\alpha,0}(\mathbb{T})$. Then, $\tilde{\psi}$ satisfies (12). And since $\nabla\tilde{\psi} = \nabla\psi$, we have

$$\forall \psi \in D(\mathbb{T}): \quad \int_{\mathbb{T}} (\alpha \nabla \phi) \cdot \nabla \psi d\mathbf{x} = \int_{\mathbb{T}} (\alpha \mathbf{u}) \cdot \nabla \psi d\mathbf{x}$$

that is to say

$$\forall \psi \in D(\mathbb{T}): \quad \langle -\nabla \cdot (\alpha \nabla \phi), \psi \rangle_{D,D'} = \langle -\nabla \cdot (\alpha \mathbf{u}), \psi \rangle_{D,D'}.$$

In other words, $\nabla \cdot (\alpha \nabla \phi) = \nabla \cdot (\alpha \mathbf{u})$ in the sense of distribution. By setting $\hat{\mathbf{u}} = \mathbf{u} - \nabla\phi$, we obtain $\mathbf{u} = \hat{\mathbf{u}} + \nabla\phi$ with $\nabla \cdot (\alpha \hat{\mathbf{u}}) = 0$. Thus $L^2_\alpha(\mathbb{T})^{1+d} \subset \mathcal{E}_\alpha \oplus \mathcal{E}_\alpha^\perp$. □

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

1) $\forall q^0 \in \mathcal{E}_\alpha: \ q(t \geq 0) \in \mathcal{E}_\alpha$.

2) $\forall q^0 \in \mathcal{E}_\alpha^\perp: \ q(t \geq 0) \in \mathcal{E}_\alpha^\perp$.

Proof. The first point is a direct consequence of the expression of System (3) because $\mathcal{E}_\alpha = Ker\, L_\alpha$ and then for all $t \geq 0$, $q(t) = q^0$. Let $q^0 = \left(r^0, \mathbf{u}^0\right)^T \in \mathcal{E}_\alpha^\perp$. We have $\alpha q(t) = \alpha q^0 - \dfrac{1}{M}\displaystyle\int_0^t L_\alpha(q) d\tau$. Then, for all $\tilde{q} = (\tilde{r}, \tilde{\mathbf{u}})^T \in \mathcal{E}_\alpha$, we have

$$\langle q, \tilde{q}\rangle_\alpha \quad = \quad \langle q^0, \tilde{q}\rangle_\alpha - \frac{1}{M}\int_0^t \int_{\mathbb{T}} L_\alpha(q)\cdot \tilde{q}\, d\mathbf{x} d\tau$$

with

$$\begin{aligned}\int_{\mathbb{T}} L_\alpha(q)\cdot\tilde{q}\, d\mathbf{x} &= a_\star \int_{\mathbb{T}} \big(\nabla\cdot(\alpha\mathbf{u})\tilde{r} + \alpha\nabla r\cdot\tilde{\mathbf{u}}\big) d\mathbf{x} \;=\; -a_\star\int_{\mathbb{T}}\big((\alpha\mathbf{u})\cdot\nabla\tilde{r} + r\nabla\cdot(\alpha\tilde{\mathbf{u}})\big)d\mathbf{x} \\ &= -\int_{\mathbb{T}} q\cdot L_\alpha(\tilde{q})\, d\mathbf{x} \;=\; 0\end{aligned}$$

because $\tilde{q} \in \mathcal{E}_\alpha = Ker\, L_\alpha$. Then, for all $\tilde{q} = (\tilde{r}, \tilde{\mathbf{u}})^T \in \mathcal{E}_\alpha$, $\langle q, \tilde{q}\rangle_\alpha = \langle q^0, \tilde{q}\rangle_\alpha = 0$ which means that $q \in \mathcal{E}_\alpha^\perp$. □

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 3.2. *Let $q(t, \boldsymbol{x})$ be the solution of* (3) *on $\mathbb{T} \subset \mathbb{R}^{d\in\{1,2,3\}}$. Then:*

$$\forall t \geq 0: \quad E_\alpha(t \geq 0) = E_\alpha(t = 0).$$

Proof. For a solution $q = (r, \mathbf{u})^T$ of System (3), we have

$$\partial_t(\alpha q) + \frac{L_\alpha}{M}(q) = 0 \;\Rightarrow\; \frac{1}{2}\frac{d}{dt}\langle q, q\rangle_\alpha + \left\langle \frac{q}{\alpha}, \frac{L_\alpha}{M}(q)\right\rangle_\alpha = 0 \;\Rightarrow\; \frac{d}{dt}E_\alpha(t) = 0$$

because

$$\left\langle \frac{q}{\alpha}, \frac{L_\alpha}{M}(q)\right\rangle_\alpha \;=\; \frac{a_\star}{M}\int_{\mathbb{T}}\Big(\frac{r}{\alpha}\nabla\cdot(\alpha\mathbf{u}) + \frac{\mathbf{u}}{\alpha}\cdot(\alpha\nabla r)\Big)\alpha d\mathbf{x} \;=\; \frac{a_\star}{M}\int_{\partial\mathbb{T}} r(\alpha\mathbf{u})\cdot\mathbf{n} d\sigma \;=\; 0 \tag{13}$$

by using the periodicity of $\mathbb{T}$. □

4 Godunov scheme with porosity

We construct the Godunov scheme with porosity.

4.1 Finite volume scheme and Riemann problem

Let us suppose that the domain $\mathbb{T} \subset \mathbb{R}^{d\in\{1,2,3\}}$ is discretized by N cells Ω_i. Let Γ_{ij} be the common edge (in 2D and common face in 3D) 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 quantities $(\alpha, \alpha r, \alpha\mathbf{u})$ are defined on the cells Ω_i by

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

The semi-discrete finite volume scheme applied to the resolution of the linear wave equation with porosity (3) is given by

$$\begin{cases} \dfrac{d}{dt}(\alpha r)_i + \dfrac{a_\star}{M}\dfrac{1}{|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\left(\alpha\mathbf{u}\cdot\mathbf{n}\right)_{ij} = 0, \\ \dfrac{d}{dt}(\alpha\mathbf{u})_i + \dfrac{a_\star}{M}\dfrac{\alpha_i}{|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}| r_{ij}\mathbf{n}_{ij} = 0. \end{cases} \tag{14}$$

The Godunov approach consists in defining $\Big(r_{ij}, (\alpha\mathbf{u}\cdot\mathbf{n})_{ij}\Big)$ as the solution of the $1D$ Riemann problem in the $\mathbf{n}_{ij}$ direction on $\xi/t = 0$

$$\begin{cases} \alpha_{ij}\partial_t r_\xi + \dfrac{a_\star}{M}\partial_\xi\big((\alpha u)_\xi\big) = 0, \\[2ex] \partial_t\big((\alpha u)_\xi\big) + \dfrac{a_\star}{M}\alpha_{ij}\partial_\xi r_\xi = 0, \\[2ex] \Big(r_\xi, (\alpha u)_\xi\Big)(t=0,\xi) = \begin{cases} \big(r_i, (\alpha\mathbf{u})_i\cdot\mathbf{n}_{ij}\big) \text{ if } \xi < 0, \\[2ex] \big(r_j, (\alpha\mathbf{u})_j\cdot\mathbf{n}_{ij}\big) \text{ otherwise} \end{cases} \end{cases} \tag{15}$$

where ξ is the coordinate in the $\mathbf{n}_{ij}$ direction and α_{ij} is a mean value of α on Γ_{ij} which depends on (α_i, α_j) (*e.g.* $\alpha_{ij} = \frac{\alpha_i+\alpha_j}{2}$).

4.2 Solution of the Riemann problem

We explicit the solution of the Riemann problem (15). By a simple scaling argument, the solution of (15) is a function only of ξ/t. We set $U = (r, J)^T$ where $J = \alpha u$ and we write (15) under the form

$$\partial_t U + A\partial_\xi U = 0, \quad \text{where} \quad A = \begin{bmatrix} 0 & \dfrac{a_\star}{M\alpha_{ij}} \\ \dfrac{a_\star\alpha_{ij}}{M} & 0 \end{bmatrix}.$$

System (15) is hyperbolic and the matrix A admits the two distinct eigenvalues $\lambda_1 = -\dfrac{a_\star}{M} < \lambda_2 = \dfrac{a_\star}{M}$. The solution $R(U_i, U_j, \xi/t)$ of (15) is under the form

$$R(U_i, U_j, \xi/t) = \begin{cases} U_i, \text{ if } \xi/t < \lambda_1, \\[2ex] U^\star, \text{ if } \lambda_1 < \xi/t < \lambda_2, \\[2ex] U_j, \text{ if } \xi/t > \lambda_2, \end{cases}$$

with

$$U_i := \big(r_i, (\alpha\mathbf{u})_i\cdot\mathbf{n}_{ij}\big) \quad \text{and} \quad U_j := \big(r_j, (\alpha\mathbf{u})_j\cdot\mathbf{n}_{ij}\big)$$

and where we have to find $U^\star$. We use the Riemann invariants to explicit $U^\star = (r^\star, J^\star)^T$. We can prove that $\mathbf{v}_1 = (1, -\alpha_{ij})^T$ (resp. $\mathbf{v}_2 = (1, \alpha_{ij})^T$) is an eigenvector of A associated to λ_1 (resp. λ_2) and that $R_1 = J + \alpha_{ij}r$ is a 1-Riemann invariant and $R_2 = J - \alpha_{ij}r$ is a 2-Riemann invariant. As a Riemann invariant is constant through a linearly degenerate wave, we obtain

$$\begin{cases} J^\star + \alpha_{ij}r^\star = J_i + \alpha_{ij}r_i, \\[2ex] J^\star - \alpha_{ij}r^\star = J_j - \alpha_{ij}r_j \end{cases} \implies \begin{cases} r^\star = \dfrac{r_i + r_j}{2} + \dfrac{1}{2\alpha_{ij}}(J_i - J_j), \\[2ex] J^\star = \dfrac{J_i + J_j}{2} + \dfrac{\alpha_{ij}}{2}(r_i - r_j). \end{cases} \tag{16}$$

4.3 The Godunov scheme

Finally, setting in (14) $r_{ij} = r^\star$ and $(\alpha\mathbf{u}\cdot\mathbf{n})_{ij} = J^\star$ given by (16), the Godunov scheme is given by

$$\begin{cases} \dfrac{d}{dt}(\alpha r)_i + \dfrac{a_\star}{2M}\dfrac{1}{|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\Big[\big((\alpha\mathbf{u})_i + (\alpha\mathbf{u})_j\big)\cdot\mathbf{n}_{ij} + \alpha_{ij}(r_i - r_j)\Big] = 0, \\ \dfrac{d}{dt}(\alpha\mathbf{u})_i + \dfrac{a_\star}{2M}\dfrac{\alpha_i}{|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left[r_i + r_j + \dfrac{\kappa}{\alpha_{ij}}\big((\alpha\mathbf{u})_i - (\alpha\mathbf{u})_j\big)\cdot\mathbf{n}_{ij}\right]\mathbf{n}_{ij} = 0 \end{cases} \tag{17}$$

where $\kappa = 1$. We introduce the parameter κ because this parameter will be important in the sequel.

5 Kernel of the first order modified equation on a cartesian mesh

To understand the behaviour of the Godunov scheme at low Mach number, a first step is to study the kernel of the spatial operator associated to the modified equation related to the Godunov scheme. Indeed, we will see that this kernel is strictly included in the kernel of the acoustic operator in (3) which is exactly equal to $\mathcal{E}_\alpha$. As a consequence, the Godunov scheme does not preserve some states in $\mathcal{E}_\alpha$.

5.1 First order modified equation on a cartesian mesh

We suppose for the sake of simplicity that the space dimension is 2. Assume that the mesh is cartesian with the space step Δx (*resp.* Δy) in the x (*resp.* y) direction. The subscript (i,j) defines the center of each cell of the cartesian mesh, $\left(i \pm \frac{1}{2}, j\right)$ and $\left(i, j \pm \frac{1}{2}\right)$ defining the interfaces of the cell (i,j). The Godunov scheme (17) can be written with

$$\begin{aligned}
\frac{d}{dt}(\alpha r)_{i,j} + \frac{a_\star}{M}\frac{(\alpha u_x)_{i+1,j} - (\alpha u_x)_{i-1,j}}{2\Delta x} + \frac{a_\star}{M}\frac{(\alpha u_y)_{i,j+1} - (\alpha u_y)_{i,j-1}}{2\Delta y}
&= \frac{a_\star}{2M\Delta x}\left(\alpha_{i+\frac{1}{2},j}\left(r_{i+1,j} - r_{i,j}\right) - \alpha_{i-\frac{1}{2},j}\left(r_{i,j} - r_{i-1,j}\right)\right) \\
&\quad + \frac{a_\star}{2M\Delta y}\left(\alpha_{i,j+\frac{1}{2}}\left(r_{i,j+1} - r_{i,j}\right) - \alpha_{i,j-\frac{1}{2}}\left(r_{i,j} - r_{i,j-1}\right)\right), \\
\frac{d}{dt}(\alpha u_x)_{i,j} + \frac{a_\star}{M}\alpha_{i,j}\frac{r_{i+1,j} - r_{i-1,j}}{2\Delta x} &= \kappa\frac{a_\star}{2M\Delta x}\alpha_{i,j}\left(\frac{1}{\alpha_{i+\frac{1}{2},j}}\left((\alpha u_x)_{i+1,j} - (\alpha u_x)_{i,j}\right)\right. \\
&\quad \left. - \frac{1}{\alpha_{i-\frac{1}{2},j}}\left((\alpha u_x)_{i,j} - (\alpha u_x)_{i-1,j}\right)\right), \\
\frac{d}{dt}(\alpha u_y)_{i,j} + \frac{a_\star}{M}\alpha_{i,j}\frac{r_{i,j+1} - r_{i,j-1}}{2\Delta y} &= \kappa\frac{a_\star}{2M\Delta y}\alpha_{i,j}\left(\frac{1}{\alpha_{i,j+\frac{1}{2}}}\left((\alpha u_y)_{i,j+1} - (\alpha u_y)_{i,j}\right)\right. \\
&\quad \left. - \frac{1}{\alpha_{i,j-\frac{1}{2}}}\left((\alpha u_y)_{i,j} - (\alpha u_y)_{i,j-1}\right)\right)
\end{aligned}$$

with $\kappa = 1$. The first order modified equation associated to this scheme is given by

$$\partial_t(\alpha q) + \frac{\mathcal{L}_{\kappa,\alpha}}{M}(q) = 0 \tag{18}$$

where $\mathcal{L}_{\kappa,\alpha} = L_\alpha - MB_{\kappa,\alpha}$ with

$$L_\alpha(q) = a_\star\begin{pmatrix} \nabla\cdot(\alpha\mathbf{u}) \\ \alpha\nabla r \end{pmatrix} \quad \text{and} \quad B_{\kappa,\alpha}(q) = \begin{pmatrix} \frac{a_\star\Delta x}{2M}\partial_x(\alpha\partial_x r) + \frac{a_\star\Delta y}{2M}\partial_y(\alpha\partial_y r) \\ \kappa\alpha\frac{a_\star\Delta x}{2M}\partial_x\left(\frac{1}{\alpha}\partial_x(\alpha u_x)\right) \\ \kappa\alpha\frac{a_\star\Delta y}{2M}\partial_y\left(\frac{1}{\alpha}\partial_y(\alpha u_y)\right) \end{pmatrix}.$$

5.2 Kernel of the modified equation and energy relation

We study the kernel of the spatial operator associated to the modified equation (18). The structure of the kernel depends on the value of κ. The kernel for the Godunov scheme ($\kappa = 1$) is different from the incompressible space $\mathcal{E}_\alpha$. Indeed:

Lemma 5.1. *1. If $\kappa > 0$, we have*

$$Ker\,\mathcal{L}_{\kappa>0,\alpha} = \left\{q := (r, \boldsymbol{u})^T | \nabla r = 0 \text{ and } \partial_x(\alpha u_x) = \partial_y(\alpha u_y) = 0\right\} \subsetneq \mathcal{E}_\alpha. \tag{19}$$

2. If $\kappa = 0$, we have

$$Ker\,\mathcal{L}_{\kappa=0,\alpha} = \mathcal{E}_\alpha.$$

Proof. If $\kappa = 0$, we easily obtain that Ker $\mathcal{L}_{\kappa=0,\alpha} = \mathcal{E}_\alpha$. Let us now suppose that $\kappa > 0$. By using (13), we can write that

$$\left\langle \frac{q}{\alpha}, \frac{\mathcal{L}_{\kappa,\alpha}}{M}(q) \right\rangle_\alpha = \left\langle \frac{q}{\alpha}, B_{\kappa,\alpha}(q) \right\rangle_\alpha .$$

Let us choose $q := (r, \mathbf{u})^T \in$ Ker $\mathcal{L}_{\kappa,\alpha}$. In that case, we deduce from the previous equality that

$$\left\langle \frac{q}{\alpha}, B_{\kappa,\alpha}(q) \right\rangle_\alpha = 0.$$

On the other hand, we have

$$-\left\langle \frac{q}{\alpha}, B_{\kappa,\alpha}(q) \right\rangle_\alpha = \frac{a_\star \Delta x}{2M} \|\partial_x r\|^2_{L^2_\alpha} + \frac{a_\star \Delta y}{2M} \|\partial_y r\|^2_{L^2_\alpha} + \kappa \frac{a_\star \Delta x}{2M} \left\| \frac{\partial_x(\alpha u_x)}{\alpha} \right\|^2_{L^2_\alpha} + \kappa \frac{a_\star \Delta y}{2M} \left\| \frac{\partial_y(\alpha u_y)}{\alpha} \right\|^2_{L^2_\alpha} . \quad (20)$$

This allows to write that

$$\|\partial_x r\|^2_{L^2_\alpha} = \|\partial_y r\|^2_{L^2_\alpha} = \left\| \frac{\partial_x(\alpha u_x)}{\alpha} \right\|^2_{L^2_\alpha} = \left\| \frac{\partial_y(\alpha u_y)}{\alpha} \right\|^2_{L^2_\alpha} = 0$$

that is to say $\nabla r = 0$ and $\partial_x(\alpha u_x) = \partial_y(\alpha u_y) = 0$. This proves that Ker $\mathcal{L}_{\kappa,\alpha} \subset \mathcal{A}$ with

$$\mathcal{A} := \left\{ q := (r, \mathbf{u})^T | \nabla r = 0 \text{ and } \partial_x(\alpha u_x) = \partial_y(\alpha u_y) = 0 \right\}.$$

Let us now suppose that $q \in \mathcal{A}$. In that case, we have $q \in \mathcal{E}_\alpha =$ Ker L_α and $q \in$ Ker $B_{\kappa,\alpha}$ that is to say $q \in$ Ker $(L_\alpha - M B_{\kappa,\alpha}) =$ Ker $\mathcal{L}_{\kappa,\alpha}$. Thus, we have also $\mathcal{A} \subset$ Ker $\mathcal{L}_{\kappa,\alpha}$. □

If the kernel depends on the value of κ, System (18) is dissipative for all $\kappa \geq 0$. Indeed:

Lemma 5.2. *Let $q(t, \boldsymbol{x})$ be the solution of* (18) *on $\mathbb{T} \subset \mathbb{R}^2$. If $\kappa \geq 0$, System* (18) *is dissipative. That is to say:*

$$\forall t \geq 0 : \quad \frac{d}{dt} E_\alpha(t) \leq 0 \quad \textit{where} \quad E_\alpha(t) := \|q\|^2_{L^2_\alpha} .$$

Proof. We have $\dfrac{1}{2}\dfrac{d}{dt} E_\alpha(t) = -\left\langle \dfrac{q}{\alpha}, \dfrac{\mathcal{L}_{\alpha,\kappa}}{M}(q) \right\rangle_\alpha$. And since $\left\langle \dfrac{q}{\alpha}, \dfrac{\mathcal{L}_{\alpha,\kappa}}{M}(q) \right\rangle_\alpha = -\left\langle \dfrac{q}{\alpha}, B_{\kappa,\alpha}(q) \right\rangle_\alpha$, we obtain by using (20)

$$\frac{1}{2}\frac{d}{dt} E_\alpha(t) = -\frac{a_\star}{2M} \left(\Delta x \|\partial_x r\|^2_{L^2_\alpha} + \Delta y \|\partial_y r\|^2_{L^2_\alpha} + \kappa \Delta x \left\| \frac{\partial_x(\alpha u_x)}{\alpha} \right\|^2_{L^2_\alpha} + \kappa \Delta y \left\| \frac{\partial_y(\alpha u_y)}{\alpha} \right\|^2_{L^2_\alpha} \right) . \quad (21)$$

This equality allows to write that $\dfrac{d}{dt} E_\alpha(t) \leq 0$ for any $\kappa \geq 0$. □

6 Explanation of the inaccuracy of the Godunov scheme on a cartesian mesh at low Mach number by using the modified equation

We studied the kernel Ker $\mathcal{L}_{\kappa,\alpha}$ of the spatial operator associated to the Godunov scheme ($\kappa = 1$). This kernel is a subset of the incompressible space $\mathcal{E}_\alpha$. As a consequence, the Godunov scheme does not preserve any incompressible state $q \in \mathcal{E}_\alpha$. However, if we delete the numerical diffusion of the Godunov scheme on the velocity field by setting $\kappa = 0$, the kernel of the modified equation is exactly the incompressible space $\mathcal{E}_\alpha$. Thus, all the incompressible states $q \in \mathcal{E}_\alpha$ will be preserved over time. Nevertheless, the knowledge of Ker $\mathcal{L}_{\kappa,\alpha}$ gives only partial informations on the time behaviour of the solution of (18). In the sequel, we give the definition of an accurate scheme at low Mach number when its first order modified equation is (18) and we prove that the Godunov scheme is not accurate at low Mach number when $M \ll \min(\Delta x, \Delta y)$.

6.1 Definition of an accurate scheme at low Mach number

We propose the following definition in order to clearly define an accurate scheme at low Mach number:

Definition 6.1. *Scheme* (17) *is accurate at low Mach number if the solution $q(t,\boldsymbol{x})$ of the modified equation* (18) *related to this scheme satisfies*

$$\forall (C_1,C_2)\in \left(\mathbb{R}_*^+\right)^2,\ \exists C_3>0 \text{ such that } \|q^0-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}=C_1M \\ \Longrightarrow\ \forall t\in[0,C_2M],\ \|q-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t)\le C_3M. \quad (22)$$

We underline that C_3 does not depend on M and we recall that $\mathbb{P}_\alpha$ is the orthogonal projection on $\mathcal{E}_\alpha$ defined by (8)*.*

This definition is justified by the fact that the solution of the linear wave equation (3) satisfies (22) (see [3] for an accurate justification of this definition).

6.2 Inaccuracy of the Godunov scheme at low Mach number

The following theorem – written in 2D for the sake of simplicity, the 3D case being similar – explains why the Godunov scheme applied to the linear wave equation with porosity on a cartesian mesh is not accurate at low Mach number:

Theorem 6.2. *When*

$$\min(\Delta x,\Delta y)\le \sqrt{2}\,diam(\mathbb{T})\sqrt{\|\alpha\|_\infty\cdot\left\|\frac{1}{\alpha}\right\|_\infty}, \quad (23)$$

for almost all initial conditions $q^0\in L^2_\alpha(\mathbb{T})^3$, the solution $q(t,\boldsymbol{x})$ of (18) *with $\kappa=1$ verifies:*

$$\exists (C_2,C_3)\in(\mathbb{R}_*^+)^2 \text{ such that } \forall C_1>0,\ \|q^0-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}=C_1M \\ \Longrightarrow\quad \forall t\ge C_2M,\ \|q-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t)\ge C_3\min(\Delta x,\Delta y) \quad (24)$$

for any $M\le \dfrac{C_3}{C_1}\min(\Delta x,\Delta y)$, C_2 and C_3 being positive parameters that do not depend on M, Δx and Δy.

This result – which is a generalization of Theorem 3.1 in [3] obtained with a constant porosity – shows that the Godunov scheme is not accurate at low Mach number when $M\ll\min(\Delta x,\Delta y)$ (for almost all initial condition q^0) since it does not verify (22). Let us note that (23) is verified when $\min(\Delta x,\Delta y)\le\sqrt{2\alpha_{\min}}\mathrm{diam}(\mathbb{T})$ because

$$0<\alpha_{\min}\le\alpha_{\max}\le 1\ \Longrightarrow\ 0<\alpha_{\min}\le\sqrt{\alpha_{\max}}\ \Longrightarrow\ \sqrt{\alpha_{\min}}\le\sqrt{\frac{\alpha_{\max}}{\alpha_{\min}}}=\sqrt{\|\alpha\|_\infty\cdot\left\|\frac{1}{\alpha}\right\|_\infty}$$

and that $\min(\Delta x,\Delta y)\le\sqrt{2\alpha_{\min}}\mathrm{diam}(\mathbb{T})$ is easily satisfied (we underline that $\alpha_{\min}$ is of order one in the sense that $M\ll\alpha_{\min}$).

6.3 Proof of Theorem 6.2

By linearity, the solution $q(t,\mathbf{x})$ of (18) with the initial condition q^0 can be written as $q(t,\mathbf{x})=q_1(t,\mathbf{x})+q_2(t,\mathbf{x})$, where q_1 is solution of

$$\begin{cases}\partial_t(\alpha q_1)+\dfrac{\mathcal{L}_{\kappa,\alpha}}{M}(q_1)=0,\\[2ex] q_1(t=0,\mathbf{x})=(q^0-\mathbb{P}_\alpha q^0)(\mathbf{x})\end{cases} \quad (25)$$

and q_2 is the solution of

$$\begin{cases}\partial_t(\alpha q_2)+\dfrac{\mathcal{L}_{\kappa,\alpha}}{M}(q_2)=0,\\[2ex] q_2(t=0,\mathbf{x})=\mathbb{P}_\alpha q^0(\mathbf{x}).\end{cases} \quad (26)$$

We have

$$\begin{aligned}\forall t \geq 0, \quad \|q-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t) &= \|q_1+q_2-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t) \;\geq\; \|q_2-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t) - \|q_1\|_{L^2_\alpha}(t) \\ &\geq \|q_2-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t) - \|q_1\|_{L^2_\alpha}(0) \;=\; \|q_2-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t) - \|q^0-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}\end{aligned} \tag{27}$$

because Equation (18) is dissipative when $\kappa \geq 0$ (see Lemma 5.2). Then, if $\|q^0-\mathbb{P}_\alpha q^0\|_{L^2_\alpha} = C_1 M$, we only have to study the function $t \mapsto \|q_2-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t)$, where q_2 is the solution of (26). The idea is to find a lower bound for the function $t \mapsto \|q_2-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}(t)$. To do this, we need some tools:

- a projection $\mathbb{P}_{\kappa=1,\alpha}$ on Ker $\mathcal{L}_{\kappa=1,\alpha}$ where Ker $\mathbb{P}_{\kappa=1,\alpha}$ is invariant for Equation (18) (in the sense of (28)),
- we write
$$q_2-\mathbb{P}_\alpha q^0 \;=\; q_2-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0 + \mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0 - \mathbb{P}_\alpha q^0,$$
- we verify that $\mathbb{P}_{\kappa=1,\alpha}\left(q_2-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\right)=0$,
- we use a Poincaré-Wirtinger inequality valid on Ker $\mathbb{P}_{\kappa=1,\alpha}$ for $q_2-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0$,
- we verify that $q_2-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0$ is solution of (18),
- we obtain the rate of dissipation of $\|q_2-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\|$ to zero by applying the Grönwall's lemma.

Lemma 6.3. *The function*

$$\begin{array}{ccc} L^2_\alpha(\mathbb{T})^3 & \to & Ker\,\mathcal{L}_{\kappa=1,\alpha} \\ q=\begin{pmatrix} r \\ \boldsymbol{u}\end{pmatrix} & \mapsto & \begin{pmatrix} \dfrac{1}{\int_\mathbb{T}\alpha dxdy}\displaystyle\int_\mathbb{T} r\alpha dxdy \\ \dfrac{1}{\alpha\int_{a_1}^{b_1}\frac{1}{\alpha(x,y)}dx}\displaystyle\int_{a_1}^{b_1} u_x(x,y)dx \\ \dfrac{1}{\alpha\int_{a_2}^{b_2}\frac{1}{\alpha(x,y)}dy}\displaystyle\int_{a_2}^{b_2} u_y(x,y)dy \end{pmatrix}\end{array}$$

defines a projection $\mathbb{P}_{\kappa=1,\alpha}$. Moreover, if $q(t,\boldsymbol{x})$ is the solution of (18) *on $\mathbb{T}$ with initial condition q^0:*

$$\forall q^0 \in Ker\,\mathbb{P}_{\kappa=1,\alpha}: \quad q(t\geq 0) \in Ker\,\mathbb{P}_{\kappa=1,\alpha}. \tag{28}$$

Proof. Recall that $\mathbb{T}=[a_1,b_1]\times[a_2,b_2]$. It is easy to prove that $\mathbb{P}_{\kappa=1,\alpha}\circ\mathbb{P}_{\kappa=1,\alpha}=\mathbb{P}_{\kappa=1,\alpha}$. This proves that $\mathbb{P}_{\kappa=1,\alpha}$ is a projector. Moreover, we have Im $\mathbb{P}_{\kappa=1,\alpha}\subset$ Ker $\mathcal{L}_{\kappa=1,\alpha}$ because for all $q=(r,\mathbf{u})^T\in L^2_\alpha(\mathbb{T})^{1+d}$,

$$\begin{aligned}\nabla\left(\frac{1}{\int_\mathbb{T}\alpha dxdy}\int_\mathbb{T} r\alpha dxdy\right) &= 0,\\ \partial_x\left(\alpha\frac{1}{\alpha\int_{a_1}^{b_1}\frac{1}{\alpha(x,y)}dx}\int_{a_1}^{b_1}u_x(x,y)dx\right) &= \partial_x\left(\frac{1}{\int_{a_1}^{b_1}\frac{1}{\alpha(x,y)}dx}\int_{a_1}^{b_1}u_x(x,y)dx\right) \;=\; 0,\\ \partial_y\left(\alpha\frac{1}{\alpha\int_{a_2}^{b_2}\frac{1}{\alpha(x,y)}dy}\int_{a_2}^{b_2}u_y(x,y)dy\right) &= \partial_y\left(\frac{1}{\int_{a_2}^{b_2}\frac{1}{\alpha(x,y)}dy}\int_{a_2}^{b_2}u_y(x,y)dy\right) \;=\; 0.\end{aligned}$$

Let $q(t,\mathbf{x})$ be a solution of (18) on $\mathbb{T}$ with initial condition $q^0\in$ Ker $\mathbb{P}_{\kappa=1,\alpha}$. By integrating the first equation of system (18) on $\mathbb{T}$, we obtain

$$\frac{d}{dt}\int_\mathbb{T} r\alpha dxdy + \frac{a_\star}{M}\int_\mathbb{T}\nabla\cdot(\alpha\mathbf{u})dxdy = \int_\mathbb{T}\Big(\frac{a_\star\Delta x}{2M}\partial_x(\alpha\partial_x r)+\frac{a_\star\Delta y}{2M}\partial_y(\alpha\partial_y r)\Big)dxdy \;\Rightarrow\; \frac{d}{dt}\int_\mathbb{T}\alpha r dxdy = 0$$

by periodicity. Then, if $\int_\mathbb{T}\alpha r^0 dxdy=0$, it is the case at any time. Moreover, since α does not depend on time, we can write the second equation of (18) under the form

$$\partial_t u_x + \frac{a_\star}{M}\partial_x r = \kappa\frac{a_\star\Delta x}{2M}\partial_x\left(\frac{1}{\alpha}\partial_x(\alpha u_x)\right).$$

By integrating on $[a_1, b_1]$, we obtain

$$\partial_t \int_{a_1}^{b_1} u_x(x,y)dx + \frac{a_\star}{M}\int_{a_1}^{b_1} \partial_x r dx = \kappa \frac{a_\star \Delta x}{2M}\int_{a_1}^{b_1} \partial_x \left(\frac{1}{\alpha}\partial_x(\alpha u_x)\right) dx \;\Rightarrow\; \partial_t \int_{a_1}^{b_1} u_x(x,y)dx = 0$$

by periodicity. Then, if $\int_{a_1}^{b_1} u_x^0(x,y)dx = 0$, it is the case at any time. We apply the same technique for u_y and we obtain that (28) is satisfied. □

We now write a Poincaré-Wirtinger inequality for a function $q \in \text{Ker}\,\mathbb{P}_{\kappa=1,\alpha}$:

Lemma 6.4. *For any $q := (r, u_x, u_y)^T \in Ker\,\mathbb{P}_{\kappa=1,\alpha}$ such that $(r, \alpha u_x, \alpha u_y)^T \in H_1(\mathbb{T})^3$, we have*

$$\|q\|^2_{L^2_\alpha} \le K_\alpha(\mathbb{T})^2 \left(\|\nabla r\|^2_{L^2_\alpha} + \left\|\frac{\partial_x(\alpha u_x)}{\alpha}\right\|^2_{L^2_\alpha} + \left\|\frac{\partial_y(\alpha u_y)}{\alpha}\right\|^2_{L^2_\alpha} \right) \tag{29}$$

with $K_\alpha(\mathbb{T}) = \sqrt{2}diam(\mathbb{T})\sqrt{||\alpha||_\infty \cdot ||\frac{1}{\alpha}||_\infty}$.

Proof. Let $q = (r, u_x, u_y)^T \in \text{Ker}\,\mathbb{P}_{\kappa=1,\alpha}$. Since $\int_\mathbb{T} r\alpha dxdy = 0$, by using the weighted Poincaré-Wirtinger inequality (11) on r, we obtain

$$\|r\|_{L^2_\alpha} \le \frac{1}{\sqrt{C_\alpha(\mathbb{T})}}\|\nabla r\|_{L^2_\alpha} \tag{30}$$

where $\sqrt{C_\alpha(\mathbb{T})} = 1/K_\alpha(\mathbb{T})$. Moreover, since for all $y \in [a_2, b_2]$, we have $0 = \int_{a_1}^{b_1} u_x(x,y)dx = \int_{a_1}^{b_1}(\alpha u_x)(x,y)\frac{1}{\alpha}dx$, by applying the weighted Poincaré-Wirtinger inequality (10) to the function $x \mapsto (\alpha u_x)(x,y)$ with the weight $\mu = \frac{1}{\alpha}$, we obtain

$$\int_{a_1}^{b_1} |(\alpha u_x)(x,y)|^2 \frac{1}{\alpha}dx \le \frac{1}{C_\alpha(\mathbb{T})}\int_{a_1}^{b_1} |\partial_x(\alpha u_x)(x,y)|^2 \frac{1}{\alpha}dx$$

that is to say

$$\int_{a_1}^{b_1} |u_x(x,y)|^2 \alpha dx \le \frac{1}{C_\alpha(\mathbb{T})}\int_{a_1}^{b_1} \left|\frac{\partial_x(\alpha u_x)(x,y)}{\alpha}\right|^2 \alpha dx.$$

Thus, by integrating over $[a_2, b_2]$, we find

$$\|u_x\|^2_{L^2_\alpha}(t) \le \frac{1}{C_\alpha(\mathbb{T})}\left\|\frac{\partial_x(\alpha u_x)}{\alpha}\right\|^2_{L^2_\alpha}(t). \tag{31}$$

We apply the same analysis for u_y such that $0 = \int_{a_2}^{b_2} u_y(x,y)dy = \int_{a_2}^{b_2}(\alpha u_y)(x,y)\frac{1}{\alpha}dy$, which gives

$$\|u_y\|^2_{L^2_\alpha}(t) \le \frac{1}{C_\alpha(\mathbb{T})}\left\|\frac{\partial_y(\alpha u_y)}{\alpha}\right\|^2_{L^2_\alpha}(t). \tag{32}$$

We obtain (29) with (30), (31) and (32). □

To prove inequality (24), we firstly have to prove the following lemma:

Lemma 6.5. *There exists a constant $K_\alpha(\mathbb{T}) > 0$ depending on $\mathbb{T}$ and α such that*

$$\forall t \ge 0, \quad \left\|q_2 - \mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t) \le \left\|(1 - \mathbb{P}_{\kappa=1,\alpha}) \circ \mathbb{P}_\alpha q^0\right\|_{L^2_\alpha} \exp\left(-\frac{a_\star \min(\Delta x, \Delta y)}{2MK_\alpha(\mathbb{T})^2}t\right). \tag{33}$$

Proof. Let us define $\hat{q} = q_2 - \mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0 =: (\hat{r}, \hat{\mathbf{u}})^T$. The idea is to apply the inequality of Lemma 6.4 combined with the equality (21) in the proof of Lemma 5.2. For this, we firstly prove that $\hat{q}$ satisfies (18). Since q_2 satisfies (18), $\hat{q}$ satisfies (18) if and only if $\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0$ satisfies (18). Since $\text{Im}\,\mathbb{P}_{\kappa=1,\alpha} \subset \text{Ker}\,\mathcal{L}_{\kappa=1,\alpha}$, we have

$\mathcal{L}_{\kappa=1,\alpha}\left(\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\right)=0$ and then $\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0$ satisfies (18). Then, $\hat{q}$ is solution of (18) and by using (21), we obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\hat{q}\|^2_{L^2_\alpha}(t) &= -\frac{a_\star}{2M}\left(\Delta x\,\|\partial_x\hat{r}\|^2_{L^2_\alpha}+\Delta y\,\|\partial_y\hat{r}\|^2_{L^2_\alpha}(t)+\kappa\Delta x\left\|\frac{\partial_x(\alpha\hat{u}_x)}{\alpha}\right\|^2_{L^2_\alpha}(t)+\kappa\Delta y\left\|\frac{\partial_y(\alpha\hat{u}_y)}{\alpha}\right\|^2_{L^2_\alpha}(t)\right)\\ &\le -\frac{a_\star}{2M}\min(\Delta x,\Delta y)\left(\|\nabla\hat{r}\|^2_{L^2_\alpha}(t)+\left\|\frac{\partial_x(\alpha\hat{u}_x)}{\alpha}\right\|^2_{L^2_\alpha}(t)+\left\|\frac{\partial_y(\alpha\hat{u}_y)}{\alpha}\right\|^2_{L^2_\alpha}(t)\right)\end{aligned}\tag{34}$$

since $\kappa=1$. Since $\mathbb{P}_{\kappa=1,\alpha}\hat{q}(t=0)=\mathbb{P}_{\kappa=1,\alpha}\left(q_2^0-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\right)=\mathbb{P}_{\kappa=1,\alpha}\circ\mathbb{P}_\alpha q^0-\mathbb{P}_{\kappa=1,\alpha}\circ\mathbb{P}_\alpha q^0=0$ and since $\mathrm{Ker}\,\mathbb{P}_{\kappa=1,\alpha}$ is invariant for Equation (18) (in the sense of (28)), we have $\hat{q}(t\ge 0)\in\mathrm{Ker}\,\mathbb{P}_{\kappa=1,\alpha}$. Thus, we can apply Lemma 6.4 to $\hat{q}$ and we obtain from (34)

$$\frac{1}{2}\frac{d}{dt}\|\hat{q}\|^2_{L^2_\alpha}(t)\;\le\;-\frac{a_\star\min(\Delta x,\Delta y)}{2MK_\alpha(\mathbb{T})^2}\|\hat{q}\|^2_{L^2_\alpha}(t).$$

By applying the Grönwall's lemma, we obtain (33) because $\hat{q}(t=0)=(1-\mathbb{P}_{\kappa=1,\alpha})\circ\mathbb{P}_\alpha q^0$. □

Now, we are able to prove Theorem 6.2. By applying Lemma 6.5, we have for all $t\ge 0$:

$$\begin{aligned}\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t) &\ge \left\|\mathbb{P}_\alpha q^0-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}-\left\|q_2-\mathbb{P}_{\kappa=1,\alpha}\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\\ &\ge \left\|(1-\mathbb{P}_{\kappa=1,\alpha})\circ\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}\left(1-\exp\left(-\frac{a_\star\min(\Delta x,\Delta y)}{2MK_\alpha(\mathbb{T})^2}t\right)\right).\end{aligned}$$

By noting that $1-\exp(-x/2)\ge x/3$ for $x\in[0,1]$, we have

$$\forall t\le\frac{MK_\alpha(\mathbb{T})^2}{a_\star\min(\Delta x,\Delta y)}:\quad 1-\exp\left(-\frac{a_\star\min(\Delta x,\Delta y)}{2MK_\alpha(\mathbb{T})^2}t\right)\ge\frac{a_\star\min(\Delta x,\Delta y)}{3MK_\alpha(\mathbb{T})^2}t.$$

Thus, we have also

$$\forall t\in\left[\frac{MK_\alpha(\mathbb{T})}{a_\star},\frac{MK_\alpha(\mathbb{T})^2}{a_\star\min(\Delta x,\Delta y)}\right]:\quad 1-\exp\left(-\frac{a_\star\min(\Delta x,\Delta y)}{2MK_\alpha(\mathbb{T})^2}t\right)\ge\frac{\min(\Delta x,\Delta y)}{3K_\alpha(\mathbb{T})}$$

when $\min(\Delta x,\Delta y)\le K_\alpha(\mathbb{T})$. Moreover, we have

$$\forall t\ge\frac{MK_\alpha(\mathbb{T})^2}{a_\star\min(\Delta x,\Delta y)}:\quad 1-\exp\left(-\frac{a_\star\min(\Delta x,\Delta y)}{2MK_\alpha(\mathbb{T})^2}t\right)\ge 1-\frac{1}{\sqrt{e}}$$

Then, if $\min(\Delta x,\Delta y)\le K_\alpha(\mathbb{T})$, we have

$$\forall t\ge\frac{MK_\alpha(\mathbb{T})}{a_\star}:\quad 1-\exp\left(-\frac{a_\star\min(\Delta x,\Delta y)}{2MK_\alpha(\mathbb{T})^2}t\right)\ge\min\left(\frac{\min(\Delta x,\Delta y)}{3K_\alpha(\mathbb{T})},1-\frac{1}{\sqrt{e}}\right)=\frac{\min(\Delta x,\Delta y)}{3K_\alpha(\mathbb{T})}$$

because $\frac{1}{3}\le 1-\frac{1}{\sqrt{e}}$. Then, if $\min(\Delta x,\Delta y)\le K_\alpha(\mathbb{T})$, we have

$$\begin{aligned}\forall t\ge C_2M:\quad\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t) &\ge \left\|(1-\mathbb{P}_{\kappa=1,\alpha})\circ\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}\frac{\min(\Delta x,\Delta y)}{3K_\alpha(\mathbb{T})}\\ &\ge C\min(\Delta x,\Delta y)\end{aligned}\tag{35}$$

with $C_2=\dfrac{K_\alpha(\mathbb{T})}{a_\star}$ and $C=\dfrac{\|(1-\mathbb{P}_{\kappa=1,\alpha})\circ\mathbb{P}_\alpha q^0\|_{L^2_\alpha}}{3K_\alpha(\mathbb{T})}$. In the sequel, we suppose that C is strictly positive, which is the case for all function $q^0\in L^2_\alpha(\mathbb{T})^3$ such that $\mathbb{P}_\alpha q^0\notin\mathrm{Ker}\,\mathcal{L}_{\kappa=1,\alpha}$. Moreover, since Equation (18) is dissipative when $\kappa\ge 0$ (see Lemma 5.2), we can write that

$$C_1M=\|q_1\|_{L^2_\alpha}(0)\ge\|q_1\|_{L^2_\alpha}(t).$$

Let us now suppose that $C_1M\le C\min(\Delta x,\Delta y)$. Then, by using (27) and (35), we find

$$\forall t\ge C_2M:\quad\left\|q-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\;\ge\;\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)-\|q_1\|_{L^2_\alpha}(t)\;\ge\;C\min(\Delta x,\Delta y)-C_1M\ge 0.$$

Let us now suppose that $C_1M\le C_3\min(\Delta x,\Delta y)$ with $C_3=\dfrac{C}{2}$. This gives

$$\forall t\ge C_2M:\quad\left\|q-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\quad\ge\quad C_3\min(\Delta x,\Delta y)$$

for any $M\le\dfrac{C_3}{C_1}\min(\Delta x,\Delta y)$. This concludes the proof of Theorem 6.2.

7 Low Mach and all Mach corrections for the Godunov scheme to be accurate at low Mach number on a cartesian mesh

Theorem 6.2 shows that the Godunov scheme is not accurate (in the sense of Definition 6.1) at low Mach number when $M \ll \min(\Delta x, \Delta y)$. We now prove that the Godunov scheme is accurate when $\max(\Delta x, \Delta y)$ is of the order of M. Of course, this condition on the mesh is too expensive for practical applications. To overcome this difficulty, we propose a *low Mach correction* which allows to recover the accuracy when $M \ll \min(\Delta x, \Delta y)$. At last, we propose an *all Mach correction* which allows to recover the accuracy at low Mach number when $M \ll \min(\Delta x, \Delta y)$ and the Godunov scheme when the Mach number is of order one (when the porosity is constant, we show in [3] that this *all Mach correction* may be more robust than the *low Mach correction* when it is applied in the linear case with linear convection and in the non-linear case (1)). These three points are detailed in the following theorem:

Theorem 7.1. *Let $q(t, \boldsymbol{x})$ be the solution of* (18) *with the initial condition q^0. We have:*

1) The Godunov scheme – obtained with $\kappa = 1$ in (17) *– is accurate at low Mach number when*

$$\max(\Delta x, \Delta y) = \mathcal{O}(M).$$

More precisely:

$$\forall (C_0, C_1, C_2) \in (\mathbb{R}_*^+)^3,\ \exists C_3 > 0 \textit{ such that } \begin{cases} \Delta x \leq C_0 M, \\ \Delta y \leq C_0 M, \\ \left\| q^0 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha} = C_1 M \end{cases}$$

$$\Longrightarrow \quad \forall t \in [0, C_2 M],\ \left\| q - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \leq C_3 M \tag{36}$$

where C_3 does not depend on M, Δx and Δy.

2) The low Mach Godunov scheme – obtained with $\kappa = 0$ in (17) *– is accurate at low Mach number. More precisely:*

$$\forall C_1 \in \mathbb{R}_*^+, \quad \left\| q^0 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha} = C_1 M \quad \Longrightarrow \quad \forall t \geq 0,\ \left\| q - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \leq C_1 M. \tag{37}$$

3) The all Mach Godunov scheme – obtained with $\kappa = \min(1, M)$ in (17) *– is accurate at low Mach number. More precisely:*

$$\forall (C_1, C_2) \in (\mathbb{R}_*^+)^2,\ \exists C_3 > 0 \textit{ such that } \left\| q^0 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha} = C_1 M$$
$$\Longrightarrow \forall t \in [0, C_2 M],\ \left\| q - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \leq C_3 M \tag{38}$$

where C_3 does not depend on M.

Proof. By linearity, the solution $q(t, \mathbf{x})$ of (18) with the initial condition q^0 can be written as

$$q(t, \mathbf{x}) = q_1(t, \mathbf{x}) + q_2(t, \mathbf{x})$$

where q_1 is the solution of (25) and q_2 is the solution of (26). We have

$$\begin{aligned} \forall t \geq 0, \quad \left\| q - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \quad &= \quad \left\| q_1 + q_2 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \ \leq \ \|q_1\|_{L^2_\alpha}(t) + \left\| q_2 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \\ &\leq \quad \left\| q^0 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha} + \left\| q_2 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t) \end{aligned} \tag{39}$$

because Equation (18) is dissipative when $\kappa \geq 0$ (see Lemma 5.2). Then, if $\left\| q^0 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha} = C_1 M$, we just have to study the function $t \mapsto \left\| q_2 - \mathbb{P}_\alpha q^0 \right\|_{L^2_\alpha}(t)$, where q_2 is the solution of (26). Since $\mathbb{P}_\alpha q^0 \in \mathcal{E}_\alpha = \text{Ker } L_\alpha$, we have $L_\alpha(\mathbb{P}_\alpha q^0) = 0$ and

$$\partial_t(\alpha \mathbb{P}_\alpha q^0) + \frac{L_\alpha}{M}(\mathbb{P}_\alpha q^0) = 0,$$

which implies, by using (26), that

$$\partial_t\left(\alpha(q_2-\mathbb{P}_\alpha q^0)\right)+\frac{L_\alpha}{M}(q_2-\mathbb{P}_\alpha q^0)=B_{\kappa,\alpha}(q_2-\mathbb{P}_\alpha q^0)+B_{\kappa,\alpha}(\mathbb{P}_\alpha q^0). \tag{40}$$

By multiplying (40) with $(q_2-\mathbb{P}_\alpha q^0)$, by integrating over $\mathbb{T}$ and by using (13), we obtain

$$\left\langle\frac{q_2-\mathbb{P}_\alpha q^0}{\alpha},\partial_t\left(\alpha(q_2-\mathbb{P}_\alpha q^0)\right)\right\rangle_\alpha+0\;=\;\left\langle\frac{q_2-\mathbb{P}_\alpha q^0}{\alpha},B_{\kappa,\alpha}(q_2-\mathbb{P}_\alpha q^0)\right\rangle_\alpha+\left\langle\frac{q_2-\mathbb{P}_\alpha q^0}{\alpha},B_{\kappa,\alpha}(\mathbb{P}_\alpha q^0)\right\rangle_\alpha.$$

Since (20) allows to write that $\left\langle\frac{q_2-\mathbb{P}_\alpha q^0}{\alpha},B_{\kappa,\alpha}(q_2-\mathbb{P}_\alpha q^0)\right\rangle_\alpha\leq 0$, we obtain

$$\frac{1}{2}\frac{d}{dt}\left\|q_2-\mathbb{P}_\alpha q^0\right\|^2_{L^2_\alpha}(t)\;\leq\;\left\langle\frac{q_2-\mathbb{P}_\alpha q^0}{\alpha},B_{\kappa,\alpha}(\mathbb{P}_\alpha q^0)\right\rangle_\alpha\;\leq\;\left\|\frac{B_{\kappa,\alpha}(\mathbb{P}_\alpha q^0)}{\alpha}\right\|_{L^2_\alpha}\cdot\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)$$

that is to say

$$\frac{d}{dt}\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\leq\left\|\frac{B_{\kappa,\alpha}(\mathbb{P}_\alpha q^0)}{\alpha}\right\|_{L^2_\alpha}. \tag{41}$$

Since $\nabla\left(\mathbb{P}_\alpha r^0\right)=0$, we deduce from (20) that

$$\left\|\frac{B_{\kappa,\alpha}(\mathbb{P}_\alpha q^0)}{\alpha}\right\|_{L^2_\alpha}\leq\max\left(\left|\kappa\frac{a_\star\Delta x}{2M}\right|,\left|\kappa\frac{a_\star\Delta y}{2M}\right|\right)\cdot\left(\left\|\partial_x\left(\frac{\partial_x(\alpha\mathbb{P}_\alpha u_x^0)}{\alpha}\right)\right\|_{L^2_\alpha}+\left\|\partial_y\left(\frac{\partial_y(\alpha\mathbb{P}_\alpha u_y^0)}{\alpha}\right)\right\|_{L^2_\alpha}\right).$$

Thus, by using the fact that $\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(0)=0$, we obtain by using (41)

$$\forall t\in[0,C_2M],\quad\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\leq C_2M\max\left(\left|\kappa\frac{a_\star\Delta x}{2M}\right|,\left|\kappa\frac{a_\star\Delta y}{2M}\right|\right)\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0)$$

where $\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0):=\left(\left\|\partial_x\left(\frac{\partial_x(\alpha\mathbb{P}_\alpha u_x^0)}{\alpha}\right)\right\|_{L^2_\alpha}+\left\|\partial_y\left(\frac{\partial_y(\alpha\mathbb{P}_\alpha u_y^0)}{\alpha}\right)\right\|_{L^2_\alpha}\right)$. Thus, when $\|q^0-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}=C_1M$, by using (39), we obtain

$$\forall t\in[0,C_2M],\quad\left\|q-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\leq M\left(C_1+C_2\kappa\frac{a_\star}{2M}\max\left(\Delta x,\Delta y\right)\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0)\right). \tag{42}$$

Let us now suppose that $\kappa=1$. In that case, Inequality (42) becomes

$$\forall t\in[0,C_2M],\quad\left\|q-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\leq M\left(C_1+C_2\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0)\frac{a_\star}{2M}\max(\Delta x,\Delta y)\right)$$

which allows to obtain (36) with $C_3=C_1+C_2\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0)\frac{a_\star}{2}C_0$ when $\Delta x\leq C_0M$ and $\Delta y\leq C_0M$. We now assume that $\kappa=M$. In this case, (42) can be written as

$$\forall t\in[0,C_2M],\quad\left\|q-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\leq M\left(C_1+C_2\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0)\frac{a_\star}{2}\max(\Delta x,\Delta y)\right)$$

which allows to obtain (38) with $C_3=C_1+C_2\mathcal{C}_\alpha(\mathbb{P}_\alpha q^0)\frac{a_\star}{2}\max(\Delta x,\Delta y)$. When $\kappa=0$, we have

$$\left\|\frac{B_{\kappa=0,\alpha}(\mathbb{P}_\alpha q^0)}{\alpha}\right\|_{L^2_\alpha}=0.$$

Then, we deduce from (41) that

$$\frac{d}{dt}\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)=0$$

which implies that $\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)=0$ for any non-negative time since $\left\|q_2-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(0)=0$. As a consequence, we deduce from (39) that

$$\forall t\geq 0,\quad\left\|q-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}(t)\leq\left\|q^0-\mathbb{P}_\alpha q^0\right\|_{L^2_\alpha}.$$

which gives (37) since $\|q^0-\mathbb{P}_\alpha q^0\|_{L^2_\alpha}=C_1M$. □

8 Numerical results

We illustrate Theorem 6.2 and Theorem 7.1 with an initial condition q^0. We choose an initial condition q^0 such that $q^0 = Mq_1^0 + q_2^0$ where $q_1^0 \in \mathcal{E}_\alpha^\perp$, $\|q_1^0\|_{L^2_\alpha} = 1$ and $q_2^0 \in \mathcal{E}_\alpha$. The function $q_1^0 := (r_1^0, \mathbf{u}_1^0)^T$ is given by $q_1^0 = \dfrac{\bar{q}_1}{\|\bar{q}_1\|_{L^2_\alpha}}$ with

$$\begin{cases} \bar{r}_1^0(x,y) = \dfrac{\sin(2\pi x)\cos(2\pi y)}{\alpha(x,y)}, \\ \\ \bar{\mathbf{u}}_1^0 = \nabla\phi \end{cases} \qquad \text{where} \qquad \begin{cases} \alpha(x,y) = \dfrac{1}{2} + \dfrac{1}{4}\sin(\pi x)\sin(2\pi y), \\ \\ \phi(x,y) = \sin(2\pi x)\cos(2\pi y). \end{cases}$$

The functions α, r_1 and ϕ are defined at the cell center. The function $q_2^0 := (r_2^0, \mathbf{u}_2^0)^T$ is given by

$$\begin{cases} r_2^0 = 1, \\ \\ \mathbf{u}_2^0 = \dfrac{\nabla \times \psi}{\alpha}, \end{cases} \qquad \text{where} \quad \psi(x,y) = \frac{1}{\pi}\sin^2(\pi x)\sin^2(\pi y).$$

By construction, we have $q_1^0 \in \mathcal{E}_\alpha^\perp$ with $\|q_1^0\|_{L^2_\alpha} = 1$ and $q_2 \in \mathcal{E}_\alpha$. Moreover, we choose the parameters $a_\star = 1$, $M = 10^{-2}$ and $CFL = 0.4$ where $\Delta t = CFL \times a_\star \frac{\min(\Delta x, \Delta y)}{M}$. We compare the results obtained with the Godunov scheme ($\kappa = 1$), the all Mach scheme ($\kappa = M$) and the low Mach scheme ($\kappa = 0$).

In Figure 1, Figure 2 and Figure 3, we plot the norm of $\alpha\mathbf{u}$ in each cell at the initial time and at the final time $t_{\text{final}} = M = 10^{-2}$. On a 30×30 cartesian mesh, the solution given by the Godunov scheme ($\kappa = 1$) is very diffused over time while the solution on a 300×300 cartesian mesh seems to be close to the initial condition (see Figure 1 and Figure 2). These numerical results illustrate the inaccuracy of the Godunov scheme ($\kappa = 1$) at low Mach number when $M \ll \min(\Delta x, \Delta y)$ (see Theorem 6.2) and its good behaviour if we use a very fine mesh (*i.e.* such that $\min(\Delta x, \Delta y) = \mathcal{O}(M)$: see Point 1 of Theorem 7.1). Moreover, the *low Mach Godunov scheme* and the *all Mach Godunov scheme* allow to keep the accuracy at low Mach number even when $M \ll \min(\Delta x, \Delta y)$ (see Points 2 and 3 of Theorem 7.1) since the numerical solutions given by these schemes are near the initial condition (see Figure 1 and Figure 3).

9 Conclusion

We proposed a *low Mach correction* and an *all Mach correction* for the Godunov scheme applied on a cartesian mesh to the linear wave equation with porosity. These corrections have been justified by studying the time behaviour of a solution of the first order modified equations associated to these schemes. It remains to justify these corrections in the discrete cartesian case. The triangular case is also important since we know that the Godunov scheme with a constant porosity is accurate at low Mach number on a triangular (or tetrahedral) mesh [7, 5]. These two points are studied in [2].

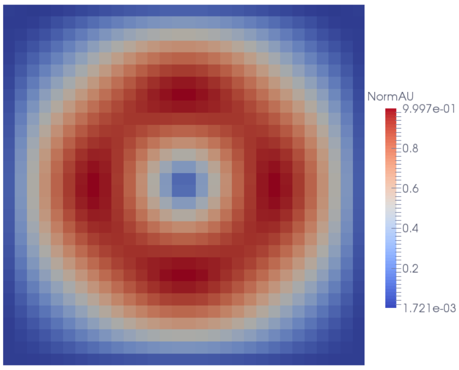

Figure 1: Norm of the velocity $\alpha\mathbf{u}$ at initial time.

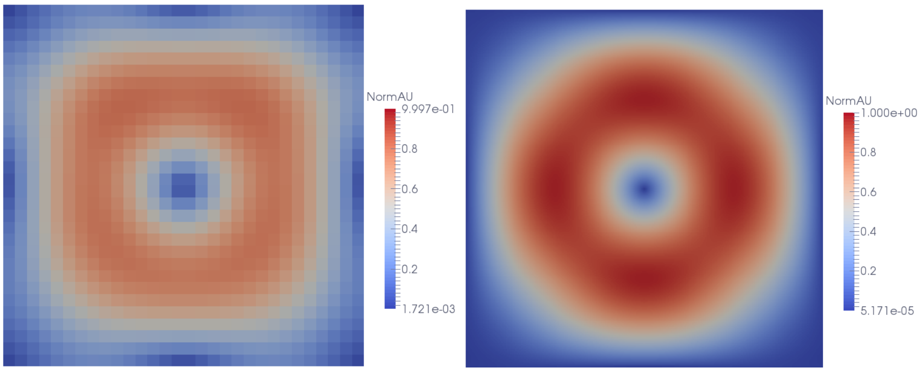

Figure 2: Norm of the velocity $\alpha\mathbf{u}$ at final time $t_{\text{final}} = M = 10^{-2}$ with the Godunov scheme ($\kappa = 1$) on a 30×30 cartesian mesh (left picture) and on a 300×300 cartesian mesh (right picture). On the coarse mesh, the velocity field is very diffused over time while it seems to be close to the initial condition (see Figure 1) for a fine mesh. These numerical results illustrate the inaccuracy of the Godunov scheme ($\kappa = 1$) at low Mach number when $M \ll \min(\Delta x, \Delta y)$ (see Theorem 6.2) and its good behaviour if we use a very fine mesh *i.e.* such that $\max(\Delta x, \Delta y) = \mathcal{O}(M)$ (see Point 1 of Theorem 7.1).

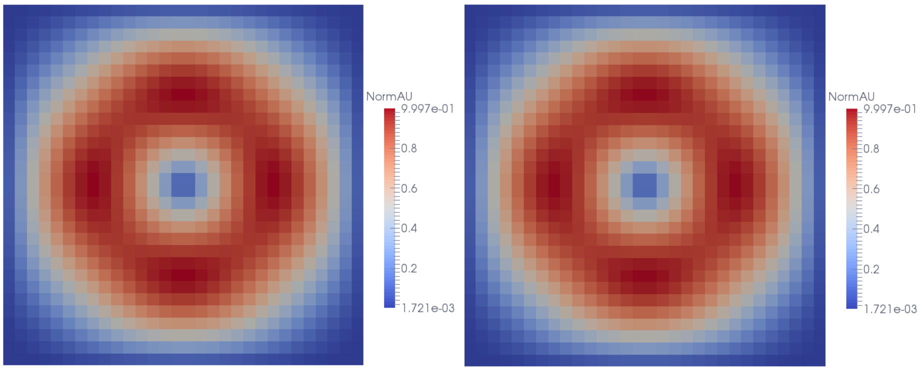

Figure 3: Norm of the velocity $\alpha\mathbf{u}$ at final time $t_{\text{final}} = M = 10^{-2}$ with the *low Mach Godunov scheme* (left picture, $\kappa = 0$) and the *all Mach Godunov scheme* (right picture, $\kappa = M$) on a 30×30 cartesian mesh. The *low Mach Godunov scheme* ($\kappa = 0$) and the *all Mach Godunov scheme* ($\kappa = M$) allow to keep the accuracy at low Mach number even when $M \ll \min(\Delta x, \Delta y)$ (see Points 2 and 3 of Theorem 7.1) since the numerical solutions given by these schemes are near the initial condition (see Figure 1).

Appendix

Poincaré-Wirtinger inequality for weighted space

Proposition. *Assume that Ω is an open convex bounded space in $\mathbb{R}^{d\in\{1,2,3\}}$ and that μ is a probabilistic measure on Ω such that μ and $\dfrac{1}{\mu}$ are in $L^\infty(\Omega)$. Then, we have:*

$$\forall \phi \in H^1(\Omega): \qquad \int_\Omega |\phi(\boldsymbol{x}) - \bar{\phi}|^2 \mu(\boldsymbol{x}) d\boldsymbol{x} \leq 2 diam(\Omega)^2 \|\mu\|_\infty \left\| \frac{1}{\mu} \right\|_\infty \int_\Omega |\nabla \phi|^2 \mu(\boldsymbol{x}) d\boldsymbol{x} \tag{43}$$

where $\bar{\phi} := \displaystyle\int_\Omega \phi(\boldsymbol{x})\mu(\boldsymbol{x})d\boldsymbol{x}$ and $diam(\Omega) := \displaystyle\sup_{(\boldsymbol{x},\boldsymbol{y})\in\Omega^2} |\boldsymbol{x} - \boldsymbol{y}|$.

Proof. The proof is done for $d \in \{1,2,3\}$. We have for all $(\mathbf{x},\mathbf{y}) \in \Omega^2$

$$\begin{aligned}
\phi(\mathbf{x}) - \phi(\mathbf{y}) = \int_0^1 \nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big) \cdot (\mathbf{x} - \mathbf{y}) dt \implies \phi(\mathbf{x}) - \bar{\phi} &= \int_\Omega \int_0^1 \nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big) \cdot (\mathbf{x} - \mathbf{y}) dt \mu(\mathbf{y}) d\mathbf{y} \\
\implies \big(\phi(\mathbf{x}) - \bar{\phi}\big)^2 &\leq \int_\Omega \int_0^1 \Big(\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big) \cdot (\mathbf{x} - \mathbf{y})\Big)^2 dt \mu(\mathbf{y}) d\mathbf{y} \\
\implies \big(\phi(\mathbf{x}) - \bar{\phi}\big)^2 &\leq \int_\Omega \int_0^1 \Big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\Big|^2 |\mathbf{x} - \mathbf{y}|^2 dt \mu(\mathbf{y}) d\mathbf{y}
\end{aligned}$$

with a Jensen inequality (with the function squared) and a Cauchy-Schwartz inequality. By multiplying by $\mu(\mathbf{x})$ and by integrating on Ω, we find

$$\int_\Omega \big(\phi(\mathbf{x}) - \bar{\phi}\big)^2 \mu(\mathbf{x}) d\mathbf{x} \leq \int_\Omega \int_\Omega \int_0^1 \big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\big|^2 |\mathbf{x} - \mathbf{y}|^2 dt \mu(\mathbf{y}) d\mathbf{y} \mu(\mathbf{x}) d\mathbf{x}$$

which implies that

$$\int_\Omega \big(\phi(\mathbf{x}) - \bar{\phi}\big)^2 \mu(\mathbf{x}) d\mathbf{x} \leq \mathrm{diam}(\Omega)^2 \int_\Omega \int_\Omega \int_0^1 \big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\big|^2 dt \mu(\mathbf{y}) d\mathbf{y} \mu(\mathbf{x}) d\mathbf{x}.$$

We split the integral by integrating on $\big[0, \frac{1}{2}\big]$ and on $\big[\frac{1}{2}, 1\big]$. This gives

$$\begin{aligned}
\int_\Omega \big(\phi(\mathbf{x}) - \bar{\phi}\big)^2 \mu(\mathbf{x}) d\mathbf{x} \leq \mathrm{diam}(\Omega)^2 \int_\Omega \int_\Omega \bigg(&\int_0^{\frac{1}{2}} \big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\big|^2 dt \\
&+ \int_{\frac{1}{2}}^1 \big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\big|^2 dt \bigg) \mu(\mathbf{y}) d\mathbf{y} \mu(\mathbf{x}) d\mathbf{x} \\
\leq \mathrm{diam}(\Omega)^2 \|\mu\|_\infty \bigg(&\int_\Omega \int_\Omega \int_0^{\frac{1}{2}} \big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\big|^2 dt d\mathbf{x} \mu(\mathbf{y}) d\mathbf{y} \\
&+ \int_\Omega \int_\Omega \int_{\frac{1}{2}}^1 \big|\nabla\phi\big((1-t)\mathbf{x} + t\mathbf{y}\big)\big|^2 dt d\mathbf{y} \mu(\mathbf{x}) d\mathbf{x} \bigg) \\
= \mathrm{diam}(\Omega)^2 \|\mu\|_\infty \bigg(&\int_\Omega \int_0^{\frac{1}{2}} \int_{\omega(t,\mathbf{y})} \big|\nabla\phi\left(\mathbf{z}\right)\big|^2 \frac{d\mathbf{z}}{1-t} dt \mu(\mathbf{y}) d\mathbf{y} + \int_\Omega \int_{\frac{1}{2}}^1 \int_{\omega(t,\mathbf{x})} \big|\nabla\phi\left(\mathbf{z}\right)\big|^2 \frac{d\mathbf{z}}{t} dt \mu(\mathbf{x}) d\mathbf{x} \bigg)
\end{aligned}$$

where $\omega(t,\mathbf{x})$ and $\omega(t,\mathbf{y})$ are included in Ω. Thus, by replacing $\omega(t,\mathbf{x})$ and $\omega(t,\mathbf{y})$ with Ω, and since $\frac{1}{1-t} \geq 1$ when $t \in \big[0, \frac{1}{2}\big]$ and $\frac{1}{t} \geq 1$ when $t \in \big[\frac{1}{2}, 1\big]$, we can write

$$\begin{aligned}
\int_\Omega \big(\phi(\mathbf{x}) - \bar{\phi}\big)^2 \mu(\mathbf{x}) d\mathbf{x} \quad &\leq \quad \mathrm{diam}(\Omega)^2 \|\mu\|_\infty \int_\Omega \big|\nabla\phi\left(\mathbf{z}\right)\big|^2 d\mathbf{z} \bigg(\int_\Omega \int_0^{\frac{1}{2}} \frac{dt}{1-t} \mu(\mathbf{y}) d\mathbf{y} + \int_\Omega \int_{\frac{1}{2}}^1 \frac{dt}{t} \mu(\mathbf{x}) d\mathbf{x} \bigg) \\
&\leq \quad \mathrm{diam}(\Omega)^2 \|\mu\|_\infty \int_\Omega \big|\nabla\phi\left(\mathbf{z}\right)\big|^2 d\mathbf{z} \bigg(\int_\Omega \mu(\mathbf{y}) d\mathbf{y} + \int_\Omega \mu(\mathbf{x}) d\mathbf{x} \bigg)
\end{aligned}$$

which finally gives

$$\int_\Omega \left(\phi(\mathbf{x}) - \bar{\phi}\right)^2 \mu(\mathbf{x}) d\mathbf{x} \leq 2\,\mathrm{diam}(\Omega)^2 \|\mu\|_\infty \int_\Omega \left|\nabla\phi\left(\mathbf{z}\right)\right|^2 d\mathbf{z}.$$

We obtain (43) by using the previous inequality and by noting that for all $\mathbf{z} \in \Omega$, we have $1 \leq \left\|\frac{1}{\mu}\right\|_\infty \mu(\mathbf{z})$. □

References

[1] Stéphane Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *J. Comp. Phy.*, 4(229):978–1016, 2010.

[2] Stéphane Dellacherie, Jonathan Jung, and Pascal Omnes. An all Mach correction for the Godunov scheme applied to the linear wave equation with porosity. *In preparation*, 2014.

[3] 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. *Preprint 2013, http://hal.archives-ouvertes.fr/hal-00776629. In preparation*, 2014.

[4] Stéphane Dellacherie and Pascal Omnes. On the Godunov scheme applied to the variable cross-section linear equation. *Finite Volumes for Complex Applications VI (FVCA6). Problems and Perspectives*, (4):313–321, 2011.

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

[6] Yann Moguen, Stéthane Dellacherie, Pascal Bruel, and Erick Dick. Momentum interpolation for quasi one-dimensional unsteady low Mach number flows with acoustics. *11th World Congress on Computational Mechanics (WCCM XI), ECCOMAS*, 2014.

[7] Felix Rieper and Bader Georg. The influence of cell geometry on the accuracy of upwind schemes in the low Mach number regime. *J. Comp. Phy.*, (228):2918–2934, 2009.