HAL open science

Construction of modified Godunov type schemes accurate at any Mach number for the compressible Euler system

Stéphane Dellacherie, Jonathan Jung, Pascal Omnes, Pierre-Arnaud Raviart

▶ **To cite this version:**

Stéphane Dellacherie, Jonathan Jung, Pascal Omnes, Pierre-Arnaud Raviart. Construction of modified Godunov type schemes accurate at any Mach number for the compressible Euler system. 2013. hal-00776629v2

HAL Id: hal-00776629

https://hal.science/hal-00776629v2

Preprint submitted on 22 Mar 2015 (v2), last revised 15 Aug 2016 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Construction of modified Godunov type schemes accurate at any Mach number for the compressible Euler system

Stéphane Dellacherie[1,2], Jonathan Jung[2,3], Pascal Omnes[1,4] and Pierre-Arnaud Raviart[2]

[1] *Commissariat à l'Énergie Atomique et aux Énergies Alternatives, CEA,DEN,DM2S,STMF, F-91191 Gif-sur-Yvette, France.*

[2] *Université Paris 6, LRC Manon, Laboratoire J.L. Lions, 4 place Jussieu, 75005 Paris, France.*

[3] *EFREI, 30-32 avenue de la République, 94800 Villejuif, France*

[4] *Université Paris 13, LAGA, CNRS UMR 7539, Institut Galilée, 99, Avenue J.-B. Clément F-93430 Villetaneuse Cedex, France.*

Abstract

Through a linear analysis, we show how to modify Godunov type schemes applied to the compressible Euler system to make them accurate at any Mach number. This allows to propose *all Mach Godunov type schemes*. A linear stability result is proposed and a formal asymptotic analysis justifies the construction in the barotropic case when the Godunov type scheme is a Roe scheme. We also underline that we may have to introduce a cut-off in the *all Mach correction* to avoid the creation of non-entropic shock waves in the case of the Roe scheme.

Key words:
Compressible Euler system, linear wave equation, low Mach number flow, Godunov scheme, Roe scheme.

1. Introduction

In many situations, the Mach number in the nuclear core of a pressurized water reactor is close to zero. This implies that the acoustic waves are often not crucial in the mass, momentum and energy balances to model the thermalhydraulics in the nuclear core. As a consequence, a low Mach number model as the one proposed in [1, 2] can be a correct approach, such a model being free of any acoustic waves. Nevertheless, in some accidental situations, the Mach number is not always and/or not everywhere close to zero, which implies that acoustic waves (which can be rarefaction and/or shock waves) cannot be neglected. The simplest model which can model low Mach flows as well as rarefaction and/or shock waves is the

Email address: `stephane.dellacherie@cea.fr` (Stéphane Dellacherie[1,2], Jonathan Jung[2,3], Pascal Omnes[1,4] and Pierre-Arnaud Raviart[2])

March 22, 2015

compressible Euler system

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \\ \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p = 0, \\ \\ \partial_t (\rho E) + \nabla \cdot [(\rho E + p)\mathbf{u}] = 0 \end{cases} \tag{1}$$

which can be simplified into the barotropic Euler system

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \\ \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p = 0 \end{cases} \tag{2}$$

when we suppose that the flow is isentropic. In (1) and (2), ρ is the density, p is the pressure, $\mathbf{u}$ is the velocity and $E := \frac{|\mathbf{u}|^2}{2} + \varepsilon$ is the total energy, ε being the internal energy. To close (1) and (2), p, ρ and ε are linked through the respective given functions $p(\rho, \varepsilon)$ and $p(\rho)$ which define the equations of state of the fluid. At last, $t \geq 0$ is the time variable and the spatial variable is defined by $\mathbf{x} \in \mathbb{R}^d$ ($d \in \{1, 2, 3\}$). Of course, as a nuclear core is a bounded domain Ω in $\mathbb{R}^d$, we also have to define boundary conditions on $\partial\Omega$ ($d \in \{1, 2, 3\}$ is the dimension of the space and is chosen as a function of the expected accuracy of the model).

To capture rarefaction and/or shock waves, a classical numerical approach is to discretize (1) or (2) using a Godunov type scheme. In this paper, a *Godunov type scheme* is a finite volume type scheme whose numerical fluxes are constructed by using an exact or an approximate 1D Riemann solver in the normal direction of the edges of the mesh (*e.g.* the Roe scheme [3] and the VFRoe scheme [4]). Nevertheless, it is now well known that Godunov type schemes applied to (1) or (2) are most of the time not accurate at low Mach number [5, 6, 7]: for the sake of simplicity, we name in the sequel "*low Mach number problem*" this loss of accuracy in the spatial periodic case for (1) or (2). When the mesh is cartesian and when the boundary conditions on $\partial\Omega$ are periodic – in other words, the physical space Ω is a torus $\mathbb{T}$ included in $\mathbb{R}^d$ –, it is shown in [8] that the *low Mach number problem* for (1) or (2) can be partially understood and can be cured by studying the *low Mach number problem* for the (dimensionless) wave equation

$$\begin{cases} \partial_t q + \frac{L}{M} q = 0, \\ \\ q(t = 0, \mathbf{x}) = q^0(\mathbf{x}) \end{cases} \tag{3}$$

where M is the Mach number ($0 < M \ll 1$), $q := (r, \mathbf{u})^T \in \mathbb{R}^{1+d}$ and $L := a_*(\nabla \cdot, \nabla)^T$ is the acoustic operator (a_* is a constant of order one) whose kernel is given by

$$KerL = \mathcal{E} \tag{4}$$

with

$$\mathcal{E} := \left\{ q \in (L^2(\mathbb{T}))^{1+d} : \nabla r = 0 \text{ and } \nabla \cdot \mathbf{u} = 0 \right\}.$$

More precisely, the *low Mach number problem* for (3) can be partially understood and can be cured by studying the linear equation

$$\begin{cases} \partial_t q + \frac{\mathcal{L}}{M} q = 0, \\ \\ q(t = 0, \mathbf{x}) = q^0(\mathbf{x}), \end{cases} \tag{5}$$

$\mathcal{L} := L + \delta L$ being the acoustic operator L perturbed by a partial differential operator δL coming from the first order truncation error of the Godunov scheme applied to the linear wave equation (3). Let us underline that $\mathcal{L}$ is a continuous operator in [8]; its discrete version $\mathbb{L}_h$ is studied in [9]. It is underlined in [8] that when (5) is well-posed and when the initial condition $q^0(\mathbf{x})$ is close to the incompressible subspace $\mathcal{E}$, the condition

$$\textit{The solution } q(t,\mathbf{x}) \textit{ of (5) remains close to its projection on } \mathcal{E} \textit{ at any time.} \tag{6}$$

is satisfied under the sufficient condition

$$q^0(\mathbf{x}) \in \mathcal{E} \quad \Longrightarrow \quad \forall t \geq 0,\ q(t,\mathbf{x}) \in \mathcal{E} \tag{7}$$

(see Point 2 of Theorem 2.2 in [8])[1]. The mathematical expression of (6) is recalled in the sequel (see (30)). Condition (6) – which is satisfied by the solution of (3) – means that the flow remains close to an incompressible flow at any time when it is initially the case.

When δL is the first order truncation error of the Godunov scheme applied on a cartesian mesh and when $d \in \{2,3\}$, we check that $\mathcal{E}$ is not an invariant subspace for (5) (see Point 2 of Lemma 4.2 in [8]) and that the kernel of $\mathcal{L}$ verifies

$$Ker\mathcal{L} \subsetneq \mathcal{E} \tag{8}$$

instead of (4) (see Point 3 of Lemma 4.3 in [8])[2]. As a consequence, (6) may not be satisfied ((7) is only a sufficient condition) and $q(t,\mathbf{x})$ may be far from an incompressible flow. Thus, we have proposed in [8] to modify the Godunov scheme in such a way (7) is satisfied. In the case of the Godunov scheme, the simplest choice to verify (7) is to center the discretization of the pressure gradient in the velocity equation (see also Point 2 of Lemma 4.2 in [8]) by deleting the upwinding stabilization term in this equation. Indeed, this *low Mach correction* – which defines the *low Mach Godunov scheme* – implies that

$$Ker\mathcal{L} = \mathcal{E} \tag{9}$$

(see Point 2 of Lemma 4.3 in [8]), which is stronger than (7). In the linear case (3), this theoretical approach gives a quite good understanding of the *low Mach number problem* and defines a simple and efficient *low Mach correction* for Godunov schemes. Moreover, numerical results proposed in [8] justify this correction in the non-linear case (*i.e.* for compressible Euler and Navier-Stokes systems) on meshes which are or are not cartesian. Nevertheless, the analysis proposed in [8] is partial and has been upgraded in [9] in fourth directions:

1. Condition (6) is too weak to characterize an accurate scheme at low Mach number for (3). Indeed, this condition does not exclude *a priori* a highly diffusive scheme in the incompressible space $\mathcal{E}$ that is to say a scheme for which $q(t,\mathbf{x})$ remains close to $\mathcal{E}$ at any time and goes to zero in short time. To exclude this possibility, (6) has to be replaced by the stronger condition

$$\textit{The solution } q(t,\mathbf{x}) \textit{ of (5) remains close to the projection on } \mathcal{E} \textit{ of the initial condition } q^0 \textit{ at any time.} \tag{10}$$

[1] The sufficient condition (7) means that $\mathcal{E}$ is an invariant subspace for (5).

[2] When $d = 1$, $\mathcal{E}$ is invariant – more precisely (9) is verified –, which underlines that the monodimensional case is particular (see Points 1 of Lemma 4.2 and 4.3 in [8]). In other words, the *low Mach number problem* does not exist when $d = 1$.

Condition (10) is justified since the solution of (3) verifies (10). This point is implicitly used in [9] and is detailed in this paper. More precisely, the mathematical expression of (10) will be specified in the sequel (see (32)) and we will also prove that

$$\mathcal{E} \subseteq Ker\mathcal{L} \tag{11}$$

is a sufficient condition to satisfy (10) (see Point 2 of Theorem 2.2 in the sequel). Since (9) is satisfied for the *low Mach Godunov scheme*, we obtain that (10) is also satisfied. This justifies the *low Mach correction* proposed in [8].

2. To explain the *low Mach number problem* on a cartesian mesh for (3), we have to prove that (10) is not verified in the case of the Godunov scheme. This result is proved in [9] in the continuous case by studying the short time behaviour of (5) and by using a Poincaré-Wirtinger inequality (see Proposition 4.1 and Corollary 4.1 in [9]).

3. It is also important to study the *discrete version* of the *low Mach correction* by analysing the discrete version $\mathbb{L}_h$ of $\mathcal{L}$. This is done in [9] where the link between $Ker\mathbb{L}_h$ and the discrete version $\mathcal{E}_h$ of $\mathcal{E}$ is studied on cartesian and triangular meshes (see Lemmae 5.1, 5.2 and 5.6 in [9]). As for the continuous case, we also study in [9] the short time behaviour of the (ordinary differential) equation

$$\begin{cases} \partial_t q_h + \dfrac{\mathbb{L}}{M} q_h = 0, \\ \\ q_h(t=0) = q_h^0 \end{cases} \tag{12}$$

which is the discrete version of the (partial differential) equation (5) (see Proposition 5.1 in [9]) by using a discrete Poincaré-inequality whose proof can be found in [10]. This allows us to explain in the semi-discrete case the *low Mach number problem* on a cartesian mesh.

4. The *low Mach number problem* does not exist when the mesh is triangular [11, 12]. We explain in [9] this particular behaviour by proving that the discrete version $Ker\mathbb{L}_h = \mathcal{E}_h$ of (9) is satisfied when the mesh is triangular (see Lemma 5.1 in [9]). Indeed, this result implies that the discrete version of (10) is satisfied without any *low Mach correction*.

The results proposed in [9] contribute to the understanding of the *low Mach number problem* and justify the *low Mach correction* proposed in [8]. Nevertheless, as this *low Mach correction* is obtained by deleting a part of the upwinding stabilization term in the Godunov scheme, the *low Mach Godunov scheme* may be unstable in the case of the non-linear systems (1) and (2) when the Mach number is of order one (although it is stable in the linear case when the Mach number is close to zero: this important point will be proved in this paper). Thus, we propose and we justify in this paper an *all Mach correction* which allows to recover the *low Mach correction* when the Mach number goes to zero, and the classical Godunov scheme when the Mach number is of order one. This allows to obtain a modified Godunov scheme that we name *all Mach Godunov scheme*. This *all Mach correction* is similar or identical to the one proposed in [13, 14]. The difficulty to justify it comes from the fact that the kernel of $\mathcal{L}$ is identical to the one obtained with the Godunov scheme – thus, it verifies (8) and does not verify (11) – and from the fact that (10) is not verified. In this paper, we prove that the *all Mach correction* is such that

The solution $q(t, \mathbf{x})$ of (5) remains close to the projection on $\mathcal{E}$ of the initial condition q^0 for short times. (13)

And we justify the use of the *short time* condition (13) instead of the *long time* condition (10) although the solution of (3) verifies (10). This point also underlines that (11) is too strong to characterize a scheme verifying (13). Then, we prove a linear stability result, we extend the *all Mach correction* in the non-linear case and we propose numerical results. All these results justify the proposed *all Mach correction* for Godunov type schemes.

At last, we underline that the theoretical approach proposed in this paper is general in the sense that it can be also used to analyze (and to eventually correct) the accuracy at low Mach number of schemes which are not at all of Godunov type (*e.g.* scheme on staggered grid: see [9]). That is why we recalled in this introduction the main steps studied in [8, 9] and that we try to explain in the following section the low Mach problem in a general framework not restricted to Godunov type schemes before applying it to this type of shemes.

The outline of this paper is the following. We recall and we clarify in Section 2 some results proposed in [8, 9]. To obtain an *all Mach Godunov scheme* in the case of the linear wave equation, we propose and we justify in Section 4 an *all Mach correction*. From this linear approach, we propose *all Mach Godunov type schemes* in Section 5 in the case of the barotropic Euler system (2). We propose in Section 6 a linear stability result for these non-linear schemes when the Godunov type scheme is a Roe scheme, and we justify in Section 7 the accuracy of this non-linear scheme with a formal asymptotic expansion. In Section 8, we extend the previous (barotropic) *all Mach Godunov type schemes* to the compressible Euler system (1). We introduce in Section 9 a cut-off in the *all Mach correction* to avoid possible non-entropic shock waves. We underline in Section 10 that the proposed approach to obtain *all Mach schemes* is not restricted to Godunov type schemes. At last, we propose numerical results in Section 11.

2. The low Mach number problem

We recall in this section some results obtained in [8, 9].

2.1. The low Mach asymptotics in the non-linear case

Let us define the Mach number $M := \dfrac{\overline{u}}{\overline{a}}$ where $\overline{u}$ and $\overline{a}$ are respectively an order of the magnitude of the fluid velocity and of the sound velocity in the domain Ω. Then, when M is close to zero and when the initial conditions are well-prepared in the following sense

$$\left\{\begin{array}{ll} \rho(t=0,\mathbf{x}) = \rho_*(\mathbf{x}), & \text{(a)} \\ p(t=0,\mathbf{x}) = p_* + O(M^2), & \text{(b)} \\ \mathbf{u}(t=0,\mathbf{x}) = \widehat{\mathbf{u}}(\mathbf{x}) + O(M) \quad \text{with} \quad \nabla\cdot\widehat{\mathbf{u}}(\mathbf{x}) = 0 & \text{(c)} \end{array}\right. \tag{14}$$

(the notation $O(f)$ means *of the order of* f), the solution $(\rho, \mathbf{u}, p)$ of the (dimensionless) compressible Euler system

$$\left\{\begin{array}{ll} \partial_t\rho + \nabla\cdot(\rho\mathbf{u}) = 0, & \text{(a)} \\ \partial_t(\rho\mathbf{u}) + \nabla\cdot(\rho\mathbf{u}\otimes\mathbf{u}) + \dfrac{\nabla p}{M^2} = 0, & \text{(b)} \\ \partial_t(\rho E) + \nabla\cdot[(\rho E + p)\mathbf{u}] = 0 & \text{(c)} \end{array}\right. \tag{15}$$

is close to $(\rho, \mathbf{u}, p)$ which satisfies $p = p_*$ and the incompressible Euler system

$$\begin{cases} \partial_t \rho + \mathbf{u} \cdot \nabla \rho = 0, \quad \rho(t = 0, \mathbf{x}) = \rho_*(\mathbf{x}), \\ \nabla \cdot \mathbf{u} = 0 \quad \text{and} \quad \mathbf{u}(t = 0, \mathbf{x}) = \widehat{\mathbf{u}}(\mathbf{x}), \\ \rho(t, \mathbf{x})(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = -\nabla \Pi. \end{cases} \tag{16}$$

In (16), Π is a new unknow which has the dimension of a pressure (it is sometimes named dynamic pressure). This pressure can formally be related to the pressure p through $p = p_* + M^2 \Pi + O(M^3)$. Let us note that we do not take into account any boundary conditions in [8, 9] and in the sequel. As a consequence, we suppose that the domain Ω in which (15) is solved is a torus $\mathbb{T}$ included in $\mathbb{R}^d$ where $d \in \{1, 2, 3\}$ is the dimension of the space.

2.2. *The low Mach asymptotics in the linear case*

The dimensionless barotropic Euler system is given by

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \dfrac{\nabla p(\rho)}{M^2} = 0. \end{cases} \tag{17}$$

The sound velocity in (17) is given by $\sqrt{p'(\rho)}/M$ (we suppose that $p'(\rho) > 0$), which is high at low Mach number (*i.e.* when $M \ll 1$). For smooth solutions, System (17) is equivalent to

$$\partial_t q + \mathcal{H}(q) + \frac{\mathcal{L}}{M}(q) = 0 \tag{18}$$

with

$$\begin{cases} q = \begin{pmatrix} r \\ \mathbf{u} \end{pmatrix}, \\ \mathcal{H}(q) = \begin{pmatrix} \mathbf{u} \cdot \nabla r \\ (\mathbf{u} \cdot \nabla)\mathbf{u} \end{pmatrix} := (\mathbf{u} \cdot \nabla) q, \\ \mathcal{L}(q) = \begin{pmatrix} (a_* + M r) \nabla \cdot \mathbf{u} \\ \dfrac{p'[\rho_*(1 + \frac{M}{a_*} r)]}{a_*(1 + \frac{M}{a_*} r)} \nabla r \end{pmatrix} \end{cases}$$

where $r(t, \mathbf{x})$ is such that

$$\rho(t, \mathbf{x}) := \rho_* \left[1 + \frac{M}{a_*} r(t, \mathbf{x}) \right] \tag{19}$$

with $a_* = \sqrt{p'(\rho_*)}$, ρ_* being a non-negative constant of order one. The operator $\mathcal{H}$ is the non-linear transport operator whose time scale is of order one; the operator $\mathcal{L}/M$ is the non-linear acoustic operator whose time scale is of order M. The linearized barotropic Euler system is thus given by

$$\partial_t q + H q + \frac{L}{M} q = 0 \tag{20}$$

with

$$\begin{cases} q = \begin{pmatrix} r \\ \mathbf{u} \end{pmatrix}, \\ \\ Hq = \begin{pmatrix} \mathbf{u}_* \cdot \nabla r \\ (\mathbf{u}_* \cdot \nabla)\mathbf{u} \end{pmatrix} := (\mathbf{u}_* \cdot \nabla)q, \\ \\ Lq = a_* \begin{pmatrix} \nabla \cdot \mathbf{u} \\ \\ \nabla r \end{pmatrix} \end{cases}$$

where $\mathbf{u}_* = \mathbf{C}_1^{st}$ and $a_* = C_2^{st}$ such that $O(|\mathbf{u}_*|) = O(a_*) = 1$. Let us underline that (20) can also be seen as a linearization of the compressible Euler System (15) with $p := p_*[1 + \frac{M}{a_*}r]$ when we replace the energy Equation (15)(c) by $s = C^{st}$ where s is the entropy. Thus, $r(t, \mathbf{x})$ can be considered as a pressure perturbation in the sequel.

Let us now introduce the sets

$$(L^2(\mathbb{T}))^{1+d} := \left\{ q := \begin{pmatrix} r \\ \mathbf{u} \end{pmatrix} : \int_{\mathbb{T}} r^2 dx + \int_{\mathbb{T}} |\mathbf{u}|^2 dx < +\infty \right\}$$

equipped with the inner product $\langle q_1, q_2 \rangle = \displaystyle\int_{\mathbb{T}} q_1 q_2 dx$ and

$$\begin{cases} \mathcal{E} &= \left\{ q \in (L^2(\mathbb{T}))^{1+d} : \nabla r = 0 \text{ and } \nabla \cdot \mathbf{u} = 0 \right\} \\ \\ &= \left\{ q \in (L^2(\mathbb{T}))^{1+d} : \exists (a, \mathbf{b}) \in \mathbb{R}^{1+d} \text{ and } \exists \psi \in H^1(\mathbb{T}) \text{ such that } r = a \text{ and } \mathbf{u} = \mathbf{b} + \nabla \times \psi \right\}, \\ \\ \mathcal{E}^\perp &= \left\{ q \in (L^2(\mathbb{T}))^{1+d} : \displaystyle\int_{\mathbb{T}} r dx = 0 \text{ and } \exists \phi \in H^1(\mathbb{T}) \text{ such that } \mathbf{u} = \nabla \phi \right\}. \end{cases}$$

The subspaces $\mathcal{E}$ and $\mathcal{E}^\perp$ are respectively called *incompressible subspace* and *acoustic subspace*. In the sequel, we use the following classical result:

Lemma 2.1.

$$\mathcal{E} \oplus \mathcal{E}^\perp = (L^2(\mathbb{T}))^{1+d} \qquad \textit{and} \qquad \mathcal{E} \perp \mathcal{E}^\perp.$$

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

$$q = \mathbb{P}q + q^\perp$$

where $(\mathbb{P}q, q^\perp) \in \mathcal{E} \times \mathcal{E}^\perp$.

The operator $\mathbb{P}$ is the Hodge projection, $q = \mathbb{P}q + q^\perp$ is the Hodge decomposition of q and we have $\langle \mathbb{P}q, q^\perp \rangle = 0$. With these tools, we can make explicit the low Mach asymptotics in the linear case (see Proposition 2.1 in [8]):

Proposition 2.1. *Let $q(t, \mathbf{x})$ be solution of*

$$\begin{cases} \partial_t q + Hq + \dfrac{L}{M} q = 0, \\ \\ q(t = 0, \mathbf{x}) = q^0(\mathbf{x}) \end{cases} \tag{21}$$

with $q^0 \in (L^2(\mathbb{T}))^{1+d}$*, and let* q_1 *be solution of*

$$\begin{cases} \partial_t q_1 + H q_1 = 0, \\ \\ q_1(t=0,\mathbf{x}) = \mathbb{P}q^0(\mathbf{x}). \end{cases} \tag{22}$$

Then, we have

$$q_1(t,\mathbf{x}) = \left(\mathbb{P}q^0\right)(\mathbf{x} - \mathbf{u}_* t) = \mathbb{P}q(t,\mathbf{x}) \tag{23}$$

and

$$\forall t \geq 0, \quad \|q - q_1\|(t) = \|q^0 - \mathbb{P}q^0\| \tag{24}$$

which implies

$$\|q^0 - \mathbb{P}q^0\| = CM \quad \Longrightarrow \quad \forall t \geq 0, \ \left\|q - \mathcal{T}_{\mathbf{u}_*,t}\left(\mathbb{P}q^0\right)\right\|(t) = CM \tag{25}$$

where $\mathcal{T}_{\mathbf{u}_*,t}$ *is the application defined by* $(\mathcal{T}_{\mathbf{u}_*,t} f)(\mathbf{x}) = f(\mathbf{x} - \mathbf{u}_* t)$ *and where* C *is a strictly positive constant, which is equivalent to*

$$\|q^0 - \mathbb{P}q^0\| = CM \quad \Longrightarrow \quad \forall t \geq 0, \ \|q - \mathbb{P}q\|(t) = CM. \tag{26}$$

In Proposition 2.1, $\|\cdot\|$ is the L^2-norm. Equality (25) allows to write that as soon as the initial condition q^0 is close to the incompressible subspace $\mathcal{E}$, the solution $q(t,\mathbf{x})$ of (21) remains close at any time to the solution $q_1(t,\mathbf{x})$ of (22). Thus, the transport Equation (22) defines the low Mach asymptotics of the linear Equation (21). Estimate (26) means that, as soon as the initial condition q^0 is close to the incompressible subspace $\mathcal{E}$, $q(t,\mathbf{x})$ remains close to $\mathcal{E}$.

Moreover, we can rewrite $\|q^0 - \mathbb{P}q^0\| = CM$ with the less accurate formulation $\|q^0 - \mathbb{P}q^0\| = O(M)$. By using (19), we easily obtain that the condition $\|q^0 - \mathbb{P}q^0\| = O(M)$ is equivalent to the well-prepared initial condition (14)(b,c) restricted to the case $p_* = p(\rho_*)$. Note that in the barotropic case, (14)(a) has to be replaced by $\rho(t=0,\mathbf{x}) = \rho_* + O(M^2)$ since $p = p(\rho)$.

The proof of Proposition 2.1 uses the linearity of (21), the fact that $\mathcal{E} = KerL$ and the conservation of the energy $E := \langle q, q\rangle$ [8]. At last, let us underline that Proposition 2.1 may also be seen as a simple application of a result by Schochet [15] obtained in the non-linear case (18).

Let us now suppose that $\mathbf{u}_* = 0$ or equivalently $H = 0$. Thus, Proposition 2.1 becomes:

Corollary 2.1. *Let* $q(t,\mathbf{x})$ *be solution of*

$$\begin{cases} \partial_t q + \dfrac{L}{M} q = 0, \\ \\ q(t=0,\mathbf{x}) = q^0(\mathbf{x}) \end{cases} \tag{27}$$

with $q^0 \in (L^2(\mathbb{T}))^{1+d}$*. Then, we have* $\mathbb{P}q = \mathbb{P}q^0$ *and*

$$\forall t \geq 0, \quad \|q - \mathbb{P}q^0\|(t) = \|q^0 - \mathbb{P}q^0\|$$

which allows to write that

$$\|q^0 - \mathbb{P}q^0\| = CM \quad \Longrightarrow \quad \forall t \geq 0, \ \|q - \mathbb{P}q^0\|(t) = CM \tag{28}$$

where C *is a strictly positive constant.*

As a consequence, the low Mach asymptotics of the linear wave Equation (27) is simply given by $\mathbb{P}q^0(\mathbf{x})$. Figure 1 represents schematically the solution of the linear wave equation.

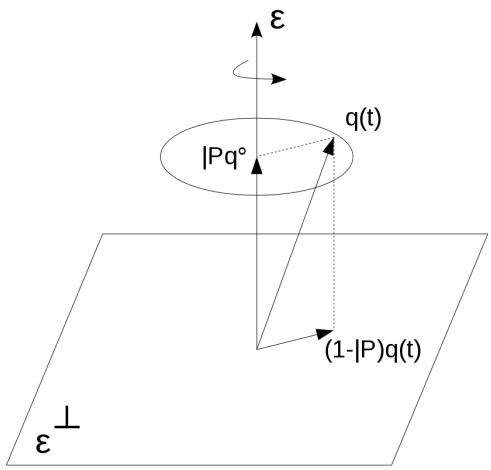

Fig. 1

2.3. The low Mach asymptotics in the case of the perturbed linear wave equation

The key points to obtain (28) are that $\mathcal{E} = KerL$ and that (27) conserves the energy. In fact, we can relax these two properties in the following way:

Theorem 2.2. *Let $\mathcal{L}$ be a linear operator and let $q(t, \mathbf{x})$ be solution of the linear equation*

$$\begin{cases} \partial_t q + \dfrac{\mathcal{L}}{M} q = 0, \\ q(t = 0) = q^0 \end{cases} \tag{29}$$

supposed to be well-posed in such a way that $\|q\|(t) \leq \widetilde{C}\|q^0\|$ for any $t \geq 0$, where $\widetilde{C}$ is a strictly positive constant (which does not depend on M). Let C be another strictly positive constant. Then:

1) When $\mathcal{E}$ is invariant for (29), we have

$$\|q^0 - \mathbb{P}q^0\| = CM \quad \Longrightarrow \quad \forall t \geq 0,\ \|q - \mathbb{P}q\|(t) \leq C\widetilde{C}M. \tag{30}$$

2) When $\mathcal{L}$ is such that

$$\mathcal{E} \subseteq Ker\mathcal{L}, \tag{31}$$

we have

$$\|q^0 - \mathbb{P}q^0\| = CM \quad \Longrightarrow \quad \forall t \geq 0,\ \|q - \mathbb{P}q^0\|(t) \leq C\widetilde{C}M. \tag{32}$$

This result is usefull to have a first understanding of the low Mach number problem. Indeed, let us consider that $\mathcal{L} := L + \delta L$ where δL is a perturbation (which may depend on M) deduced from the truncation error of a given numerical scheme applied to (27) on a cartesian mesh. Estimate (30) means that Equation (29) does not create any acoustic waves of order one in the acoustic subspace $\mathcal{E}^\perp$ when $\|q^0 - \mathbb{P}q^0\| = O(M)$ although the discretization introduces an error through δL. Estimate (32) characterizes the fact that the solution $q(t, \mathbf{x})$ of (29) remains close at any time to the low Mach asymptotics $\mathbb{P}q^0$ of the linear wave equation (27) when $\|q^0 - \mathbb{P}q^0\| = O(M)$ although the discretization introduces an error through the perturbation δL in $\mathcal{L}$.

Proof of Theorem 2.2: The proof of *Point 1* is written in [8] (see Theorem 2.2 in [8]). Nevertheless, we write again this proof below for reader's convenience. Indeed, the proof of *Point 2* uses the steps written in the proof of *Point 1*.

Point 1: Let us define $\widetilde{q}(t,\mathbf{x})$ and $\overline{q}(t,\mathbf{x})$ solutions of (29) with the respective initial conditions $\widetilde{q}^0 = \mathbb{P}q^0$ and $\overline{q}^0 = q^0 - \mathbb{P}q^0$. By linearity, we have $q = \widetilde{q} + \overline{q}$. Moreover

$$\begin{aligned} \|q - \mathbb{P}q\| &= \|\widetilde{q} - \mathbb{P}\widetilde{q} + \overline{q} - \mathbb{P}\overline{q}\| \\ &= \|\overline{q} - \mathbb{P}\overline{q}\| \end{aligned}$$

since $\mathcal{E}$ is invariant for (29). Then, we have

$$\|q - \mathbb{P}q\| \leq \|\overline{q}\| \tag{33}$$

since $(\mathbf{1} - \mathbb{P})$ is an orthogonal projection. On the other hand, we have $\|\overline{q}\| \leq \widetilde{C}\|\overline{q}^0\|$ and $\|\overline{q}^0\| = \|q^0 - \mathbb{P}q^0\| = CM$. Thus, we have

$$\|\overline{q}\| \leq C\widetilde{C}M \tag{34}$$

which allows to obtain $\|q - \mathbb{P}q\| \leq C\widetilde{C}M$ by using (33).

Point 2: Under Condition (31), we have $\widetilde{q} = \mathbb{P}q^0$. Thus, we have $q - \mathbb{P}q^0 = \overline{q}$ which allows to obtain $\|q - \mathbb{P}q^0\| \leq C\widetilde{C}M$ by using (34).□

2.4. A first definition of an accurate scheme at low Mach number

For a given operator $\mathcal{L}$ that does not necessarily satisfy (31), Estimate (32) leads us to say that the solution $q(t,\mathbf{x})$ of (29) is accurate at low Mach number in the incompressible regime of the linear wave equation if and only if the estimate

$$\forall C_1 \in \mathbb{R}_*^+,\ \exists C_2(C_1) \in \mathbb{R}_*^+ \textit{ such that } \|q^0 - \mathbb{P}q^0\| = C_1 M \implies \forall t \geq 0,\ \|q - \mathbb{P}q^0\|(t) \leq C_2 M \tag{35}$$

is satisfied, $C_2(C_1)$ being a positive parameter that depends on C_1 and does not depend on M. *Point 2* of Theorem 2.2 means that a sufficient condition to be accurate at low Mach number in the sense of (35) is that $\mathcal{E} \subseteq Ker\mathcal{L}$. Let us underline that when $\mathcal{E} \nsubseteq Ker\mathcal{L}$, we cannot say if the solution $q(t,\mathbf{x})$ is or is not accurate at low Mach number in the sense of (35) since (31) is only a *sufficient* condition. In that case, we have to study carrefully the time behaviour of (29) to verify if Estimate (35) is or is not satisfied.

In the same way, for a given operator $\mathcal{L}$ whose $\mathcal{E}$ is not necessarily invariant, Estimate (30) leads us to say that the solution $q(t,\mathbf{x})$ of (29) is free of any spurious acoustic wave if and only if the estimate

$$\forall C_1 \in \mathbb{R}_*^+,\ \exists C_2(C_1) \in \mathbb{R}_*^+ \textit{ such that } \|q^0 - \mathbb{P}q^0\| = C_1 M \implies \forall t \geq 0,\ \|q - \mathbb{P}q\|(t) \leq C_2 M. \tag{36}$$

Of course, (35) is stronger than (36) since for any q and q^0, we have $\|q - \mathbb{P}q\| \leq \|q - \mathbb{P}q^0\|$. *Point 1* of Theorem 2.2 underlines that the invariance of $\mathcal{E}$ in the energy space $(L^2(\mathbb{T}))^{1+d}$ is a sufficient condition to avoid spurious acoustic waves in the sense of (36). Nevertheless, the invariance of $\mathcal{E}$ is not sufficient to be accurate at low Mach number in the sense of (35).

Estimates (35) and (36) are usefull to analyze the accuracy of a given scheme at low Mach number and, in particular, to propose a *low Mach correction* when it is usefull as we will see it in §2.5 and §2.6 in the case of Godunov type schemes. Nevertheless, we will also see in Section 3 that in the case of Godunov type schemes, we will have to relax (35) and (36) in order to propose and to justify an *all Mach Godunov type scheme*.

2.5. A first explanation of the good or bad behaviour of Godunov type schemes at low Mach number through the study of discrete kernels

We show in this section that the low Mach number problem can be analyzed as we analyzed in §2.3 the low Mach asymptotics in the linear perturbed case (29).

Let us suppose that the domain Ω is discretized by N cells Ω_i. Let Γ_{ij} be the common edge of two neighboring cells Ω_i and Ω_j and $\mathbf{n}_{ij}$ be the unit vector normal to Γ_{ij} pointing from Ω_i to Ω_j. The semi-discrete Godunov scheme applied to the resolution of the linear wave equation (27) is given by

$$\left\{\begin{array}{ll} \dfrac{d}{dt}r_i + \dfrac{a_*}{M}\cdot\dfrac{1}{2|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|[(\mathbf{u}_i+\mathbf{u}_j)\cdot\mathbf{n}_{ij}+r_i-r_j]=0, & \text{(a)} \\ \\ \dfrac{d}{dt}\mathbf{u}_i + \dfrac{a_*}{M}\cdot\dfrac{1}{2|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|[r_i+r_j+\kappa(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{n}_{ij}=0 & \text{(b)} \end{array}\right. \tag{37}$$

with $\kappa = 1$. We introduce the parameter κ in (37)(b) for reasons that will appear in the sequel (let us note that (37) is the Godunov scheme if and only if $\kappa = 1$). This scheme can be written in the compact form

$$\left\{\begin{array}{l} \dfrac{d}{dt}q_h + \dfrac{\mathbb{L}_{\kappa,h}}{M}q_h = 0, \\ \\ q_h(t=0) = q_h^0 \end{array}\right. \quad \text{with} \quad q_h := \begin{pmatrix} r_i \\ \mathbf{u}_i \end{pmatrix} \tag{38}$$

where the subscript h recalls that (38) comes from a spatial discretization of (27) (h is a characteristic length of the mesh). The kernel $Ker\mathbb{L}_{\kappa,h}$ of the discrete acoustic operator $\mathbb{L}_{\kappa,h}$ is given by

$$Ker\mathbb{L}_{\kappa,h} := \left\{ \begin{pmatrix} r_i \\ \mathbf{u}_i \end{pmatrix} \in \mathbb{R}^{3N} \quad \text{such that} \quad \left\{\begin{array}{l} \displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|[(\mathbf{u}_i+\mathbf{u}_j)\cdot\mathbf{n}_{ij}+r_i-r_j]=0, \\ \\ \displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|[r_i+r_j+\kappa(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{n}_{ij}=0 \end{array}\right. \right\}. \tag{39}$$

We have the following result:

Lemma 2.2.

$$Ker\mathbb{L}_{\kappa=1,h} = \left\{ q_h := \begin{pmatrix} r_h \\ \mathbf{u}_h \end{pmatrix} \in \mathbb{R}^{3N} \quad \textit{such that} \quad \exists a\in\mathbb{R},\ \forall i:\ r_i = a \quad \textit{and} \quad (\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}=0 \right\} \tag{40}$$

and

$$Ker\mathbb{L}_{\kappa=0,h} = \left\{ q_h := \begin{pmatrix} r_h \\ \mathbf{u}_h \end{pmatrix} \in \mathbb{R}^{3N} \quad \textit{such that} \quad \exists a\in\mathbb{R},\ \forall i:\ r_i = a \quad \textit{and} \quad \sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\frac{\mathbf{u}_i+\mathbf{u}_j}{2}\cdot\mathbf{n}_{ij}=0 \right\}. \tag{41}$$

Moreover, we have

$$Ker\mathbb{L}_{\kappa=1,h} \subseteq Ker\mathbb{L}_{\kappa=0,h}. \tag{42}$$

By using the *Point 2* of Theorem 2.2 with Lemma 2.2, we obtain a first explanation of the good or bad behaviour of Godunov type schemes at low Mach number in 1D, 2D and 3D for different type of meshes.

Indeed, Lemma 2.2 shows that $Ker\mathbb{L}_{\kappa=1,h}$ – which is the kernel in the case of the Godunov scheme – may not be a good approximation of $\mathcal{E}$ because the continuity of $\mathbf{u}\cdot\mathbf{n}$ on each edge Γ_{ij} of the mesh could be too restrictive for particular meshes (*e.g.* when the mesh is cartesian). Nevertheless, it shows also that $Ker\mathbb{L}_{\kappa=0,h}$ may be a good approximation of $\mathcal{E}$ for any mesh type because

$$\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\frac{\mathbf{u}_i+\mathbf{u}_j}{2}\cdot\mathbf{n}_{ij} \simeq \int_{\Omega_i} \nabla\cdot\mathbf{u}dx. \tag{43}$$

Thus, by also using (42), we can say that at the discrete level, $Ker\mathbb{L}_{\kappa=1,h}$ may not satisfy (31) and that $Ker\mathbb{L}_{\kappa=0,h}$ may satisfy (31). These points are studied in [9] when the mesh is cartesian or triangular by showing that:

$$\begin{cases} \textit{on a 2D triangular mesh:} & Ker\mathbb{L}_{\kappa=1,h} = \mathcal{E}_h^{\Delta} \subset Ker\mathbb{L}_{\kappa=0,h}, \\ \textit{on a 1D cartesian mesh:} & Ker\mathbb{L}_{\kappa=1,h} = \mathcal{E}_h^{\square} = Ker\mathbb{L}_{\kappa=0,h}, \\ \textit{on a 2D or 3D cartesian mesh:} & Ker\mathbb{L}_{\kappa=1,h} \subsetneq \mathcal{E}_h^{\square} = Ker\mathbb{L}_{\kappa=0,h} \end{cases}$$

where $\mathcal{E}_h^{\Delta}$ and $\mathcal{E}_h^{\square}$ are *ad hoc* approximations of $\mathcal{E}$ which depend on the type of mesh (see Lemmae 5.1, 5.2 and 5.6 in [9]). Using these relations between the discrete kernels and the discrete incompressible spaces, using the sufficient condition (31), we obtain that in the sense of the definition (35) proposed in §2.4, we can say that:

$$\begin{cases} \textit{on a 2D triangular mesh:} & \textit{the Godunov scheme (i.e. (37) with } \kappa=1\textit{) is accurate at low Mach number}, \\ \textit{on a 1D cartesian mesh:} & \textit{the Godunov scheme (i.e. (37) with } \kappa=1\textit{) is accurate at low Mach number}, \\ \textit{on a 2D or 3D cartesian mesh:} & \textit{the Godunov scheme (i.e. (37) with } \kappa=1\textit{) is not accurate at low Mach number} \end{cases}$$

Let us underline that as (31) is only a sufficient condition, it remains to prove that the Godunov scheme (*i.e.* (37) with $\kappa = 1$) is not accurate at low Mach number. This result cannot be obtained by only studying the kernel of the scheme. It can be obtained by studying the short time behaviour of the first order modified equation associated to the Godunov scheme or by studying the short time behaviour of (37) when $\kappa = 1$ (see [9]).

Moreover, we also obtain that:

The *modified Godunov scheme* obtained with (37) and $\kappa = 0$ is accurate at low Mach number.

This last result leads us to define the *low Mach Godunov type schemes* in §2.6.

Proof of Lemma 2.2: The proof uses the fact that for any $q_h \in Ker\mathbb{L}_{\kappa,h}$ defined by (39), we have

$$\sum_{\Gamma_{ij}} |\Gamma_{ij}|\{(r_i-r_j)^2 + \kappa[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]^2\} = 0. \tag{44}$$

This relation was proven in [9] (see (88) in [9]). As a consequence, when $\kappa = 1$, we obtain that $\forall i : r_i = c$ and $(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij} = 0$. Let us now suppose that $\kappa = 0$. Thus, when $q_h \in Ker\mathbb{L}_{\kappa=0,h}$, we only deduce from (44) that $\forall i : r_i = c$. And, by injecting $r_i = c$ in (39), we find $\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_i + \mathbf{u}_j) \cdot \mathbf{n}_{ij} = 0$. The converse is obtained by using the fact that

$$\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\mathbf{n}_{ij} = 0. \tag{45}$$

We obtain (42) by using the fact that $\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\mathbf{u}_i \cdot \mathbf{n}_{ij} = \mathbf{u}_i \cdot \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\mathbf{n}_{ij} = 0.\square$

2.6. A low Mach Godunov type scheme in the linear and non-linear case

This approach leads us to modify the Godunov scheme by replacing $\kappa = 1$ in (37) with $\kappa = 0$ to recover the accuracy at low Mach number. This corresponds to center the discretization of ∇r in the acoustic operator.

The non-linear version of linear scheme (37) with $\kappa = 0$, applied to the compressible Euler system (1) or to the barotropic Euler system (2), consists in modifying any X scheme of Godunov type (*e.g.* X = Roe [3] or X = VFRoe [4]) in such a way that the discretization of the pressure gradient ∇p is centered. We named this class of schemes *low Mach X schemes* in [8]. Numerical low Mach number test-cases validate this approach in [8].

2.7. Toward an all Mach Godunov type scheme in the non-linear case

In the sequel of this paper, we modify the non-linear *low Mach X scheme* defined in §2.6 in such a way it is identical to the X scheme when the Mach number is greater than one. In other words, we introduce *all Mach Godunov type schemes* which are expected to be stable and accurate on any mesh type and for any Mach number.

3. Definition of an accurate scheme at low Mach number in the linear case

Estimate (35) is suggested by Estimate (28) of Corollary 2.1 which concerns the linearization (21) of the barotropic Euler System (18) with $H := 0$. But, when $H \neq 0$, Estimate (28) cannot be satisfied by the solution $q(t, \mathbf{x})$ of (21) and has to be replaced by Estimate (25) of Proposition 2.1. Nevertheless, we have the following result:

Lemma 3.1. *Let $q(t, \mathbf{x})$ be solution of*

$$\begin{cases} \partial_t q + Hq + \dfrac{L}{M} q = 0, \\ \\ q(t = 0, x) = q^0(x) \end{cases} \tag{46}$$

with $q^0 \in L^2(\mathbb{T}) \times (C^1(\mathbb{T}))^d$. Then, we have

$$\forall (C_1, C_2) \in (\mathbb{R}_*^+)^2,\ \exists C_3(C_1, C_2) \in \mathbb{R}_*^+ \text{ such that } \|q^0 - \mathbb{P}q^0\| = C_1 M \implies \forall t \in [0, C_2 M],\ \|q - \mathbb{P}q^0\|(t) \leq C_3 M, \tag{47}$$

$C_3(C_1, C_2)$ being a positive parameter that depends on (C_1, C_2) and does not depend on M.

As a consequence, the important point is to verify if Estimate (35) is satisfied or not satisfied *only for short times*. Thus, we relax (35) and we propose the following definition:

Definition 1. *The solution $q(t,\mathbf{x})$ of*

$$\begin{cases} \partial_t q + \dfrac{\mathcal{L}}{M} q = 0, \\ \\ q(t=0) = q^0 \end{cases} \tag{48}$$

is said to be accurate at low Mach number for short times in the incompressible regime of the linear wave equation if and only if the estimate

$$\forall (C_1, C_2) \in (\mathbb{R}_*^+)^2,\ \exists C_3(C_1,C_2) \in \mathbb{R}_*^+ \text{ such that } \|q^0 - \mathbb{P}q^0\| = C_1 M \implies \forall t \in [0, C_2 M],\ \|q - \mathbb{P}q^0\|(t) \leq C_3 M \tag{49}$$

is satisfied, $C_3(C_1,C_2)$ being a positive parameter that depends on (C_1,C_2) and does not depend on M.

In the sequel, we will construct a numerical scheme for which the solution of the associated first order modified equation is accurate at low Mach number in the sense of (49) but not in the sense of (35).

Let us note that we can keep (36) for the spurious acoustic waves when $H \neq 0$ because of Estimate (26) of Proposition 2.1. Nevertheless, when a solution $q(t,\mathbf{x})$ is accurate at low Mach number in the sense of Definition 1, we are sure that this solution is free of any spurious acoustic wave *in short time* (this is a consequence of the fact that for any q and q^0, we have $\|q - \mathbb{P}q\| \leq \|q - \mathbb{P}q^0\|$); but we can say nothing *a priori* in long time. Thus, we also relax (36) with:

Definition 2. *The solution $q(t,\mathbf{x})$ of*

$$\begin{cases} \partial_t q + \dfrac{\mathcal{L}}{M} q = 0, \\ \\ q(t=0) = q^0 \end{cases} \tag{50}$$

is said to be free of any spurious acoustic wave for short times if and only if the estimate

$$\forall (C_1, C_2) \in (\mathbb{R}_*^+)^2,\ \exists C_3(C_1,C_2) \in \mathbb{R}_*^+ \text{ such that } \|q^0 - \mathbb{P}q^0\| = C_1 M \implies \forall t \in [0, C_2 M],\ \|q - \mathbb{P}q\|(t) \leq C_3 M \tag{51}$$

is satisfied, $C_3(C_1,C_2)$ being a positive parameter that depends on (C_1,C_2) and does not depend on M.

Figures 2-4 describe three different behaviours based on Definitions 1 and 2: Figure 2 describes a solution $q(t,\mathbf{x})$ which is accurate at low Mach number; Figure 3 describes a solution $q(t,\mathbf{x})$ which is not accurate at low Mach number but which is free of any spurious acoustic wave; Figure 4 describes a solution $q(t,\mathbf{x})$ which is not accurate at low Mach number and which is not free of spurious acoustic waves. The numerical results proposed in §4.3 will be coherent with Figures 2-4.

Proof of Lemma 3.1: We have

$$\|q - \mathbb{P}q^0\|(t) \leq \left\| q - \mathcal{T}_{\mathbf{u}_*,t}\left(\mathbb{P}q^0\right) \right\|(t) + \left\| \mathcal{T}_{\mathbf{u}_*,t}\left(\mathbb{P}q^0\right) - \mathbb{P}q^0 \right\|(t)$$

where $\mathcal{T}_{\mathbf{u}_*,t}$ is the application defined by $(\mathcal{T}_{\mathbf{u}_*,t} f)(\mathbf{x}) = f(\mathbf{x} - \mathbf{u}_* t)$. Thus, by using (25), we obtain that

$$\|q^0 - \mathbb{P}q^0\| = C_1 M \quad \implies \quad \forall t \geq 0 : \|q - \mathbb{P}q^0\|(t) \leq C_1 M + \left\| \mathcal{T}_{\mathbf{u}_*,t}\left(\mathbb{P}q^0\right) - \mathbb{P}q^0 \right\|(t).$$

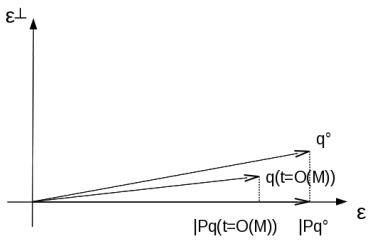

Fig. 2:
(49) and (51) are verified

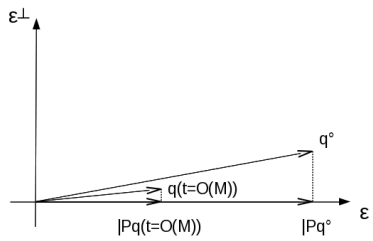

Fig. 3:
(49) is not verified, (51) is verified

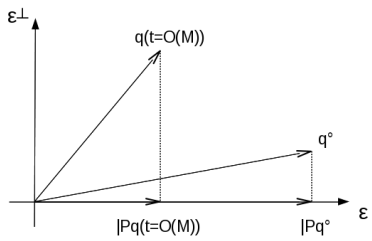

Fig. 4:
(49) and (51) are not verified

On the other hand, for any $\overline{q} := (\overline{r}, \overline{\mathbf{u}})^T \in \mathcal{E}$, we have

$$\left\|\mathcal{T}_{\mathbf{u}_*,t}\overline{q} - \overline{q}\right\|^2(t) = \int_{\mathbb{T}} |\overline{\mathbf{u}}(\mathbf{x} - \mathbf{u}_* t) - \overline{\mathbf{u}}(\mathbf{x})|^2 d\mathbf{x}.$$

But, for any $\overline{\mathbf{u}} \in \left(C^1(\mathbb{T})\right)^d$, we have

$$|\overline{\mathbf{u}}(\mathbf{x} - \mathbf{u}_* t) - \overline{\mathbf{u}}(\mathbf{x})| \leq |\mathbf{u}_*| t \max_{\mathbb{T}} |\nabla \overline{\mathbf{u}}|$$

with $|\nabla \overline{\mathbf{u}}|^2 := \sum_{k=1}^{d} |\nabla u_k|^2$ where $\mathbf{u} := (u_1, \ldots, u_d)^T$, d is the spatial dimension and $|\cdot|$ is the euclidian norm in $\mathbb{R}^d$. Thus

$$\forall t \in [0, C_2 M] : \ \left\|\mathcal{T}_{\mathbf{u}_*,t}\overline{q} - \overline{q}\right\|(t) \leq C_2 M |\mathbf{u}_*| \max_{\mathbb{T}} |\nabla \overline{\mathbf{u}}| \cdot |\mathbb{T}|$$

with $|\mathbb{T}| := \int_{\mathbb{T}} d\mathbf{x}$. This allows to write that

$$\forall t \in [0, C_2 M] : \ \left\|\mathcal{T}_{\mathbf{u}_*,t}\left(\mathbb{P}q^0\right) - \mathbb{P}q^0\right\|(t) \leq C_2 M |\mathbf{u}_*| \max_{\mathbb{T}} |\nabla \overline{\mathbf{u}}^0| \cdot |\mathbb{T}|$$

where $\mathbb{P}q^0 = (\overline{r}^0, \overline{\mathbf{u}}^0)^T$, which gives the result with

$$C_3 = C_1 + C_2 |\mathbf{u}_*| \max_{\mathbb{T}} |\nabla \overline{\mathbf{u}}^0| \cdot |\mathbb{T}|.$$

□

4. Construction and justification of an all Mach Godunov scheme in the linear case

In this section, we construct a modified Godunov type scheme which is asymptotically identical to the linear *low Mach Godunov scheme* (see (37) with $\kappa = 0$) when $M \ll 1$ and which is identical to the linear Godunov scheme (see (37) with $\kappa = 1$) when $M = O(1)$. We justify this construction by using the tools introduced in Section 2. We name this linear scheme *all Mach Godunov scheme*.

The non-linear version of this *all Mach Godunov scheme* will be directly obtained in Sections 5 and 8 from the linear approach proposed below.

4.1. The case of the linear wave equation on a cartesian mesh

This subsection is devoted to the cartesian case. This case is interesting because it allows to propose an *all Mach Godunov type scheme* through a simple study of the first order modified equation associated with the Godunov scheme applied to the linear wave equation (27).

Let us define the 2D system

$$\begin{cases} \partial_t q + \dfrac{\mathcal{L}_\nu}{M} q = 0, \\ \\ q(t=0,\mathbf{x}) = q^0(\mathbf{x}) \end{cases} \tag{52}$$

with $\mathbf{x} := (x, y)$, $q := (r, \mathbf{u})^T$, $\mathbf{u} := (u_x, u_y)^T$ and

$$\begin{cases} \mathcal{L}_\nu = L - MB_\nu, \\ \\ B_\nu q = \begin{pmatrix} \nu_r \Delta r \\ \\ \nu_{u_x} \dfrac{\partial^2 u_x}{\partial x^2} \\ \\ \nu_{u_y} \dfrac{\partial^2 u_y}{\partial y^2} \end{pmatrix} \end{cases} \tag{53}$$

where

$$\nu := (\nu_r, \nu_{\mathbf{u}}) \in (\mathbb{R}^+)^3 \qquad \text{and} \qquad \nu_{\mathbf{u}} := (\nu_{u_x}, \nu_{u_y}) \in (\mathbb{R}^+)^2.$$

Thus, (52) is a perturbed wave equation whose the perturbation is given by $\delta L_\nu = -MB_\nu$. In the 2D case (the 3D case is similar) [8], the first order modified equation of the Godunov scheme applied to the linear wave equation (27) is given by (52)(53) with $\nu = \nu^G$ where

$$\nu^G := (\nu_r^G, \nu_{\mathbf{u}}^G) \qquad \text{and} \qquad \nu_r^G := a_* \frac{\Delta x}{2M}, \quad \nu_{\mathbf{u}}^G := a_* \frac{\Delta x}{2M}(1,1)$$

(Δx is the mesh size supposed to be identical in the directions x and y for the sake of simplicity). We prove that (see Lemma 4.3 in [8]):

Lemma 4.1.
1) In 1D with $\nu_r \geq 0$, $\nu_x \geq 0$ and $\nu_y \geq 0$:

$$Ker\mathcal{L}_\nu = \mathcal{E}.$$

2) In 2D with $\nu_r \geq 0$, $\nu_{u_x} = \nu_{u_y} = 0$:

$$Ker\mathcal{L}_\nu = \mathcal{E}.$$

3) In 2D with $\nu_r \geq 0$, $\nu_{u_x} > 0$ and $\nu_{u_y} > 0$:

$$Ker\mathcal{L}_\nu = \left\{ q := \begin{pmatrix} r \\ \mathbf{u} \end{pmatrix} \in (L^2(\mathbb{T}))^3 \quad \textit{such that} \quad \exists c \in \mathbb{R} : \ r = c \quad \textit{and} \quad \partial_x u_x = \partial_y u_y = 0 \right\} \subsetneq \mathcal{E}. \tag{54}$$

The extension of Lemma 4.1 to the 3D case is straightforward.

We deduce from *Point 2* of Theorem 2.2 and from *Point 3* of Lemma 4.1 that the solution $q(t,\mathbf{x})$ of (52)(53) may not be accurate at low Mach number in the incompressible regime as soon as the spatial dimension is 2D (or 3D) and $\nu_{\mathbf{u}}$ is not equal to zero. Indeed, in that case, we do not have $\mathcal{E} \subseteq Ker\mathcal{L}_\nu$. Nevertheless, the situation is more complicate since $\mathcal{E} \subseteq Ker\mathcal{L}_\nu$ is only a *sufficient* condition: as a consequence, the knowledge of $Ker\mathcal{L}_\nu$ is not sufficient to have a good understanding of the behaviour at low Mach number of the Godunov scheme and of any modified Godunov scheme obtained by modifying the numerical viscosity ν^G. Moreover, for a particular choice of $\nu_{\mathbf{u}}$, we may expect that the short time estimate (49) is satisfied even if the long time estimate (35) is not satisfied. In that case, the solution $q(t,\mathbf{x})$ would be accurate at low Mach number in the sense of Definition 1.

This is illustrated by the following result:

Theorem 4.1. *Let $q(t,\mathbf{x})$ be the solution of the 2D equation (52)(53). Then, for any $\nu_r \geq 0$:*

1) When $\nu_{\mathbf{u}} = \nu^G_{\mathbf{u}}$, for almost all function $q^0 \in (L^2(\mathbb{T}))^3$, $q(t,\mathbf{x})$ verifies

$$\forall C_1 \in \mathbb{R}^+_* : \qquad \|q^0 - \mathbb{P}q^0\| = C_1 M \quad \Longrightarrow \quad \forall t \geq C_2 M,\ \|q - \mathbb{P}q^0\|(t) \geq C_3 \Delta x, \tag{55}$$

for any $M \leq \dfrac{C_3}{C_1}\Delta x$, C_2 and C_3 being strictly positive parameters that do not depend on M and Δx.

2) When $\nu_{\mathbf{u}} = \nu^G_{\mathbf{u}}$ and $\Delta x = C_0 M$, for any $q^0 \in H^2(\mathbb{T}^d)$, $q(t,\mathbf{x})$ verifies

$$\forall (C_0, C_1, C_2) \in (\mathbb{R}^+_*)^3 : \qquad \|q^0 - \mathbb{P}q^0\| = C_1 M \quad \Longrightarrow \quad \forall t \in [0, C_2 M],\ \|q - \mathbb{P}q^0\|(t) \leq C_3 M, \tag{56}$$

C_3 being a strictly positive parameter that does not depend on M.

3) When $\nu_{\mathbf{u}} = M\nu^G_{\mathbf{u}}$, for any $q^0 \in H^2(\mathbb{T}^d)$, $q(t,\mathbf{x})$ verifies

$$\forall (C_1, C_2) \in (\mathbb{R}^+_*)^2 : \qquad \|q^0 - \mathbb{P}q^0\| = C_1 M \quad \Longrightarrow \quad \forall t \in [0, C_2 M],\ \|q - \mathbb{P}q^0\|(t) \leq C_3 M, \tag{57}$$

C_3 being a strictly positive parameter that does not depend on M.

Again, we easily extend this result to the 3D case. This result shows that the short time behaviours of (52)(53) with $\nu = (\nu_r, \nu^G_{\mathbf{u}})$ and with $\nu = (\nu_r, M\nu^G_{\mathbf{u}})$ are different although the kernels of $\mathcal{L}_{(\nu_r,\nu^G_{\mathbf{u}})}$ and of $\mathcal{L}_{(\nu_r,M\nu^G_{\mathbf{u}})}$ are identical (see *Point 3* of Lemma 4.1). This is a consequence of the fact that Condition (31) is only a *sufficient* condition to be accurate at low Mach number.

More precisely:

- *Point 1* of Theorem 4.1 and its 3D version show that when the mesh is cartesian, for almost all $q^0 \in (L^2(\mathbb{T}))^{1+d}$, the Godunov scheme in 2D/3D is not accurate at low Mach number in the sense of Definition 1 when $M \ll \Delta x$.

  Let us note that we do not prove that the solution $q(t,\mathbf{x})$ of (52)(53) is not accurate at low Mach number *by producing* in short time spurious acoustic waves (see Definition 2 for the notion of spurious

acoustic wave). In other words, we prove that the short time behaviour of $q(t, \mathbf{x})$ is characterized by Figure 3 but we do not prove that it is characterized by Figure 4. Nevertheless, the numerical results proposed in §4.3 show that spurious acoustic waves are created in short time (at least at the discrete level), which corresponds to Figure 4.

- *Point 2* of Theorem 4.1 and its 3D version show that when the mesh is cartesian, for any $q^0 \in (L^2(\mathbb{T}))^{1+d}$, the Godunov scheme in 2D/3D is accurate at low Mach number in the sense of Definition 1 when $\Delta x = O(M)$, which is too expensive from a computational point of view.

- *Point 3* of Theorem 4.1 and its 3D version show that when the mesh is cartesian, for any $q^0 \in (L^2(\mathbb{T}))^{1+d}$, the modified Godunov scheme obtained by replacing $\nu_{\mathbf{u}}^G$ with $M\nu_{\mathbf{u}}^G$ is accurate at low Mach number in the sense of Definition 1 even when $M \ll \Delta x$. Thus, this scheme is also free of any spurious acoustic waves in the sense of Definition 2. This result is central in our way to construct an *all Mach Godunov scheme*. At last, we underline that all the results proposed in Theorem 4.1 are valid as soon as $\nu_r \geq 0$ that is to say not only when $\nu_r = \nu_r^G$.

Proof of Theorem 4.1: Let $q_1(t)$ be the solution of

$$\begin{cases} \partial_t q_1 + \dfrac{\mathcal{L}_\nu}{M} q_1 = 0, \\ \\ q_1(t = 0, \mathbf{x}) = (q^0 - \mathbb{P}q^0)(\mathbf{x}) \end{cases} \tag{58}$$

and $q_2(t)$ be the solution of

$$\begin{cases} \partial_t q_2 + \dfrac{\mathcal{L}_\nu}{M} q_2 = 0, \\ \\ q_2(t = 0, \mathbf{x}) = \mathbb{P}q^0(\mathbf{x}) \end{cases} \tag{59}$$

where $\mathcal{L}_\nu$ is defined as in (53). By linearity, the solution $q(t, \mathbf{x})$ of (52)(53) satisfies

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

Since $\|q - \mathbb{P}q^0\|(t) = \|q_1 + q_2 - \mathbb{P}q^0\|(t)$, we have

$$\forall t \geq 0 : \quad \|q - \mathbb{P}q^0\|(t) \geq \left| \|q_2 - \mathbb{P}q^0\|(t) - \|q_1\|(t) \right| \tag{60}$$

and

$$\forall t \geq 0 : \quad \|q - \mathbb{P}q^0\|(t) \leq \|q_1\|(t) + \|q_2 - \mathbb{P}q^0\|(t). \tag{61}$$

Moreover, since (58) is a dissipative equation when $\nu_r \geq 0$, $\nu_{u_x} \geq 0$ and $\nu_{u_y} \geq 0$ (see Lemma 4.1 in [8]), we obtain $\|q_1\|(t) \leq \|q^0 - \mathbb{P}q^0\|$ which implies that

$$\forall t \geq 0 : \quad \|q_1\|(t) \leq C_1 M \tag{62}$$

since $\|q^0 - \mathbb{P}q^0\| = C_1 M$. We will use below (60), (61) and (62) to prove (55), (56) and (57).

Proof of Point 1:

Let us define the orthogonal projection $\mathbb{P}_\nu$ on $Ker\mathcal{L}_\nu$ ($\mathbb{P}_\nu = \mathbb{P}$ if and only if $\nu_r \geq 0$ and $\nu_{u_x} = \nu_{u_y} = 0$; in particular, $\mathbb{P}_\nu \neq \mathbb{P}$ when $\nu = \nu^G$). In [9], we prove that

$$\forall t \geq \frac{ML_{\mathbb{T}}}{a_*} : \quad \|q_2 - \mathbb{P}q^0\|(t) \geq \frac{\Delta x}{3L_{\mathbb{T}}}\|\mathbb{P}q^0 - \mathbb{P}_\nu \mathbb{P}q^0\|$$

where $L_{\mathbb{T}}$ is a constant which only depends on $\mathbb{T}$ (see Estimate (50) of Corollary 4.1 in [9]). Hence

$$\forall t \geq C_2 M : \quad \|q_2 - \mathbb{P}q^0\|(t) \geq C\Delta x \tag{63}$$

with $C_2 = \dfrac{L_{\mathbb{T}}}{a_*}$ and $C = \dfrac{\|\mathbb{P}q^0 - \mathbb{P}_\nu \mathbb{P}q^0\|}{3L_{\mathbb{T}}}$. In the sequel, we suppose that C is strictly positive, which is the case for almost all function $q^0 \in (L^2(\mathbb{T}))^3$. Let us now suppose that

$$C_1 M \leq C\Delta x. \tag{64}$$

By using (62), (63) and (64), we obtain

$$\forall t \geq C_2 M : \quad \|q_2 - \mathbb{P}q^0\|(t) \geq C\Delta x \geq C_1 M \geq \|q_1\|(t). \tag{65}$$

And, by using (60) and (65), we obtain

$$\forall t \geq C_2 M : \quad \|q - \mathbb{P}q^0\|(t) \geq C\Delta x - C_1 M. \tag{66}$$

Let us now suppose that

$$C_1 M \leq C_3 \Delta x \qquad \text{with} \qquad C_3 = \frac{C}{2}. \tag{67}$$

We deduce from (66) and (67) that

$$\forall t \geq C_2 M : \quad \|q - \mathbb{P}q^0\|(t) \geq C_3 \Delta x$$

which allows to obtain (55).

Proof of Points 2 and 3:

Since $L\mathbb{P} = 0$, we deduce from (59) that

$$\partial_t(q_2 - \mathbb{P}q^0) + \frac{L}{M}(q_2 - \mathbb{P}q^0) = B_\nu(q_2 - \mathbb{P}q^0) + B_\nu \mathbb{P}q^0. \tag{68}$$

Then, by multypling (68) with $q_2 - \mathbb{P}q^0$ and by integrating, we obtain

$$\frac{1}{2} \cdot \frac{d}{dt}\|q_2 - \mathbb{P}q^0\|^2(t) = \langle q_2 - \mathbb{P}q^0, B_\nu(q_2 - \mathbb{P}q^0)\rangle + \langle q_2 - \mathbb{P}q^0, B_\nu \mathbb{P}q^0\rangle$$

since $\langle q_2 - \mathbb{P}q^0, L(q_2 - \mathbb{P}q^0)\rangle = 0$. And since

$$\begin{cases} \langle q_2 - \mathbb{P}q^0, B_\nu(q_2 - \mathbb{P}q^0)\rangle \leq 0, \\ \langle q_2 - \mathbb{P}q^0, B_\nu \mathbb{P}q^0\rangle \leq \|q_2 - \mathbb{P}q^0\| \cdot \|B_\nu \mathbb{P}q^0\|, \end{cases}$$

we can write that

$$\frac{d}{dt}\|q_2 - \mathbb{P}q^0\|(t) \leq \|B_\nu \mathbb{P}q^0\| \leq \max(|\nu_{u_x}|, |\nu_{u_y}|) \cdot \|\mathbb{P}q^0\|_{H^2}$$

(since $\|B_\nu \mathbb{P}q^0\| \leq \max(|\nu_{u_x}|, |\nu_{u_y}|) \cdot \|\mathbb{P}q^0\|_{H^2}$) which gives

$$\forall t \in [0, C_2 M] : \quad \|q_2 - \mathbb{P}q^0\|(t) \leq C_2 M \cdot \max(|\nu_{u_x}|, |\nu_{u_y}|) \cdot \|\mathbb{P}q^0\|_{H^2}$$

(since $\|q_2 - \mathbb{P}q^0\|(0) = 0$) that is to say

$$\forall t \in [0, C_2 M] : \quad \|q - \mathbb{P}q^0\|(t) \leq \left(C_1 + C_2 \max(|\nu_{u_x}|, |\nu_{u_y}|) \cdot \|\mathbb{P}q^0\|_{H^2}\right) M \tag{69}$$

by using (61) and (62). Let us now suppose that $\nu_{\mathbf{u}} = \nu_{\mathbf{u}}^G$. In that case, we have $\max(|\nu_{u_x}|, |\nu_{u_y}|) = \dfrac{a_* \Delta x}{2M}$ which implies that (69) is given by

$$\forall t \in [0, C_2 M] : \quad \|q - \mathbb{P}q^0\|(t) \leq \left(C_1 + \frac{C_2 a_* \Delta x}{2M} \|\mathbb{P}q^0\|_{H^2}\right) M$$

which allows to obtain (56) with $C_3 = C_1 + \dfrac{C_0 C_2 a_*}{2} \|\mathbb{P}q^0\|_{H^2}$ when $\Delta x = C_0 M$. We now suppose that $\nu_{\mathbf{u}} = M \nu_{\mathbf{u}}^G$. In that case, (69) is given by

$$\forall t \in [0, C_2 M] : \quad \|q - \mathbb{P}q^0\|(t) \leq \left(C_1 + \frac{C_2 a_* \Delta x}{2} \|\mathbb{P}q^0\|_{H^2}\right) M$$

which allows to obtain (57) with $C_3 = C_1 + \dfrac{C_2 a_* \Delta x}{2} \|\mathbb{P}q^0\|_{H^2}$.□

4.2. The case of the linear wave equation on any mesh type

To be accurate at low Mach number, *Point 2* of Theorem 4.1 leads us to modify the Godunov scheme applied to the linear wave equation

$$\begin{cases} \partial_t q + Hq + \dfrac{L}{M} q = 0, \\ \\ q(t = 0, x) = q^0(x) \end{cases} \tag{70}$$

by replacing $\kappa = 1$ (which is equivalent to $\nu_{\mathbf{u}} = \nu_{\mathbf{u}}^G$) with $\kappa = M$ (which is equivalent to $\nu_{\mathbf{u}} = M \nu_{\mathbf{u}}^G$) in (37). Thus, we propose the all Mach Godunov scheme

$$\begin{cases} \dfrac{d}{dt} r_i + \dfrac{a_*}{M} \cdot \dfrac{1}{2|\Omega_i|} \displaystyle\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| [(\mathbf{u}_i + \mathbf{u}_j) \cdot \mathbf{n}_{ij} + r_i - r_j] = 0, \\ \\ \dfrac{d}{dt} \mathbf{u}_i + \dfrac{a_*}{M} \cdot \dfrac{1}{2|\Omega_i|} \displaystyle\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| [r_i + r_j + \theta(M)(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}] \mathbf{n}_{ij} = 0 \end{cases} \tag{71}$$

with

$$\theta(M) = \min(M, 1) \tag{72}$$

which also allows to recover the Godunov scheme (37) when the Mach number is greater than one. In Section 9, we will modify (72) by introducing a cut-off (see (148) and Figures 27-28) to avoid the creation of non-entropic shock waves in the non-linear case when the Mach number is of order one.

Scheme (71) defines an *all Mach Godunov scheme* and may be rewritten with

$$\frac{d}{dt}\begin{pmatrix} r \\ \mathbf{u} \end{pmatrix}_i + \frac{1}{|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\Phi_{ij}^{\text{AM,Godunov}} = 0 \tag{73}$$

with the two following expressions for the numerical flux $\Phi_{ij}^{\text{AM,Godunov}}$ which are *equivalent* in this *linear* case[3]:

• **First expression:**

$$\Phi_{ij}^{\text{AM,Godunov}} = \Phi_{ij}^{\text{Godunov}} + [\theta(M)-1]\frac{a_*}{2M}\begin{pmatrix} 0 \\ [(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{n}_{ij} \end{pmatrix} \tag{74}$$

where $\Phi_{ij}^{\text{Godunov}}$ is the unmodified Godunov flux ($\Phi_{ij}^{\text{Godunov}}$ is easily deduced from (37)) and where $\theta(M)$ is defined by (72). Thus, the simple corrective flux

$$[\theta(M)-1]\frac{a_*}{2M}\begin{pmatrix} 0 \\ [(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{n}_{ij} \end{pmatrix} \tag{75}$$

defines an *all Mach correction* which is equal to zero when the Mach number is greater than one. This all Mach correction introduces numerical anti-diffusion since $\theta(M)-1\leq 0$. At last, we can note that the linear *all Mach Godunov scheme* (73)(74) may be seen as the Godunov scheme plus a *pressure correction* since the correction $[\theta(M)-1]\frac{a_*}{2M}[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{n}_{ij}$ in (75) is homogeneous to a pressure.

• **Second expression:** The flux (74) is equivalent to

$$\Phi_{ij}^{\text{AM,Godunov}} = \frac{a_*}{M}\begin{pmatrix} (\mathbf{u}\cdot\mathbf{n})^* \\ r^{**}\mathbf{n} \end{pmatrix}_{ij} \qquad \text{with} \qquad r_{ij}^{**} = \theta(M)r_{ij}^* + [1-\theta(M)]\frac{r_i+r_j}{2} \tag{76}$$

where $(r^*,(\mathbf{u}\cdot\mathbf{n})^*)$ is solution of the 1D linear Riemann problem in the $\mathbf{n}_{ij}$ direction

$$\begin{cases} \partial_t q_\zeta + \dfrac{L_\zeta}{M}q_\zeta = 0, \\[2ex] \zeta<0: \ q_\zeta(t=0,\zeta) = \begin{pmatrix} r_i \\ \mathbf{u}_i\cdot\mathbf{n}_{ij} \end{pmatrix}, \\[2ex] \zeta\geq 0: \ q_\zeta(t=0,\zeta) = \begin{pmatrix} r_j \\ \mathbf{u}_j\cdot\mathbf{n}_{ij} \end{pmatrix} \end{cases} \tag{77}$$

with $q_\zeta := \begin{pmatrix} r \\ u_\zeta \end{pmatrix}$ and $L_\zeta q_\zeta := a_*\partial_\zeta\begin{pmatrix} u_\zeta \\ r \end{pmatrix}$, ζ being the coordinate in the $\mathbf{n}_{ij}$ direction. This gives

$$\begin{cases} r_{ij}^* = \dfrac{r_i+r_j}{2} + \dfrac{(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}}{2}, \\[2ex] (\mathbf{u}\cdot\mathbf{n})_{ij}^* = \dfrac{(\mathbf{u}_i+\mathbf{u}_j)\cdot\mathbf{n}_{ij}}{2} + \dfrac{r_i-r_j}{2}. \end{cases}$$

[3]The notation AM in $\Phi_{ij}^{\text{AM,Godunov}}$ means that this flux defines an **A**ll **M**ach scheme.

The linear *all Mach Godunov scheme* (73)(76) – which is equivalent to (73)(74) – may be seen as a Godunov type scheme whose Riemann solver is corrected to be accurate at low Mach number.

4.3. Numerical results on a 2D cartesian mesh

We justify Theorem 4.1 and the linear *all Mach Godunov scheme* (73)(74) with numerical results obtained on a 2D cartesian mesh. The initial conditions $q^0 := (r^0, \mathbf{u}^0)^T$ is given by

$$\left\{ \begin{array}{rcl} r(t=0,x,y) & = & 1, \\ \\ u(t=0,x,y) & = & 2\sin^2(\pi x)\sin(4\pi y), \\ \\ v(t=0,x,y) & = & -\sin(2\pi x)\sin^2(2\pi y). \end{array} \right. \tag{78}$$

Thus, we have $q^0 \in \mathcal{E}$ (that is to say $q^0 = \mathbb{P}q^0$) which implies that

$$q = q^0 \tag{79}$$

is solution of the linear wave equation (27). We now study if (79) is or is not satisfied at the discrete level when we solve the linear wave equation (27) with the linear Godunov scheme (37) or with the *all Mach Godunov scheme* (73)(74).

We project q^0 on $\mathcal{E}_h^{\square}$ which gives q_h^0: thus, by construction, we have $q_h^0 = \mathbb{P}_h q_h^0$ where $\mathbb{P}_h$ is the discret Hodge projection on $\mathcal{E}_h^{\square}$.

We study on Figures 5-10 the linear Godunov scheme (37) and the *all Mach Godunov scheme* (73)(74). Figures 5 and 7 represent $\|q_h - q_h^0\|(t)$ for $0 \leq t \leq 0.5M$ (which is equal to $\|q_h - \mathbb{P}_h q_h^0\|(t)$ since $q_h^0 = \mathbb{P}_h q_h^0$ in the studied case). Figures 6 and 8 represent $\|\mathbf{u}_h^{\perp}\|(t)$ for $0 \leq t \leq 0.5M$ (where $\mathbf{u}_h^{\perp}$ is the velocity component of the projection of q_h in $(\mathcal{E}_h^{\square})^{\perp}$). Figures 9 and 10 represent $\|q_h - q_h^0\|(t)$ until an asymptotic state is reached. Thus, Figures 5-8 and Figures 9-10 describe the time behaviour of the schemes respectively in short time and in long time.

Figures 5-6 show that the linear Godunov scheme (37) is not accurate in the sense of Definition 1 and that it produces spurious acoustic waves in $(\mathcal{E}_h^{\square})^{\perp}$. At the opposite, Figures 7-8 show that the linear *all Mach Godunov scheme* (73)(74) is accurate in the sense of Definition 1 and, thus, is free of spurious acoustic waves in the sense of Definition 2. Figures 9-10 show that the linear Godunov scheme (37) and the linear *all Mach Godunov scheme* (73)(74) have the same long time behaviours: the asymptotic numerical solution is not accurate which means that both schemes are not accurate at low Mach number for long times. This last point justifies Definition 1 which only considers the short time behaviour.

5. Construction of all Mach Godunov type schemes in the barotropic case

We now extend the linear *all Mach Godunov schemes* (73)(74) and (73)(76) to the non-linear case when the linear wave equation (70) is replaced by the barotropic Euler system (2). This leads us to propose the non-linear *all Mach Godunov type scheme*

$$\frac{d}{dt}\left(\begin{array}{c} \rho \\ \rho\mathbf{u} \end{array} \right)_i + \frac{1}{|\Omega_i|} \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \Phi_{ij}^{\mathrm{AM,X}} = 0 \tag{80}$$

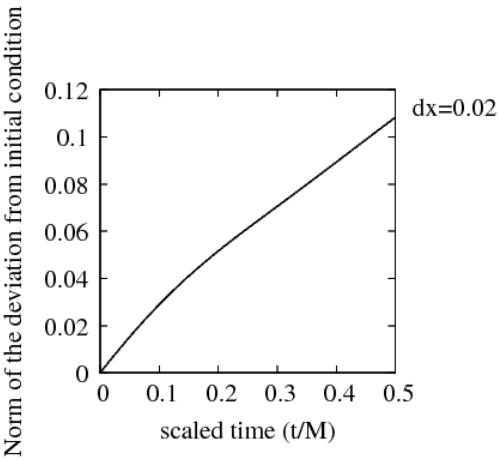

Fig. 5: $\|q_h - q_h^0\|(t)$ *when* $0 \leq t/M \leq 0.5$
Godunov scheme

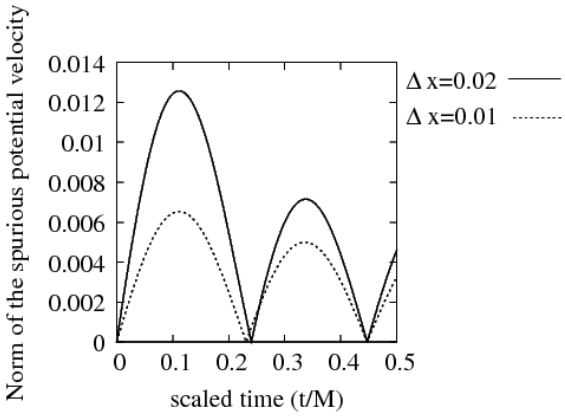

Fig. 6: $\|\mathbf{u}_h^{\perp}\|(t)$ *when* $0 \leq t/M \leq 0.5$
Godunov scheme

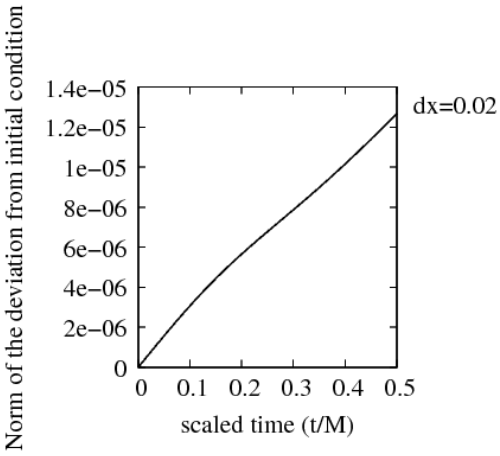

Fig. 7: $\|q_h - q_h^0\|(t)$ *when* $0 \leq t/M \leq 0.5$
All Mach Godunov scheme

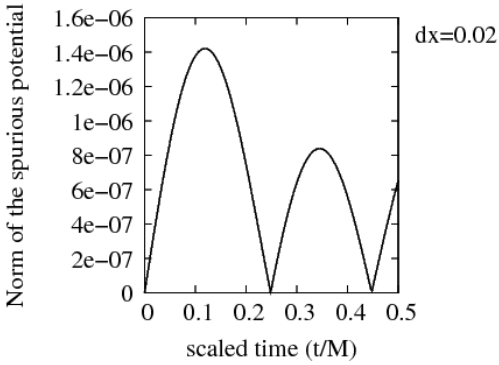

Fig. 8: $\|\mathbf{u}_h^{\perp}\|(t)$ *when* $0 \leq t/M \leq 0.5$
All Mach Godunov scheme

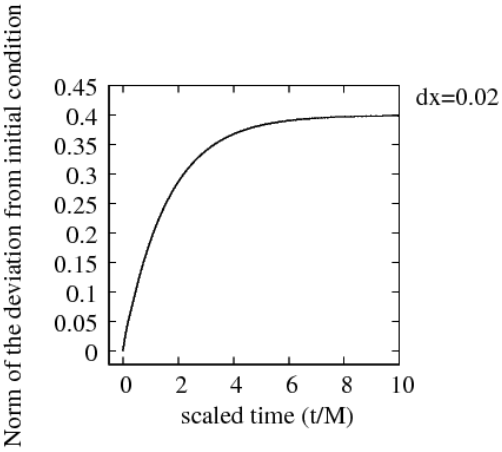

Fig. 9: $\|q_h - q_h^0\|(t)$ *when* $0 \leq t/M \leq 10$
Godunov scheme

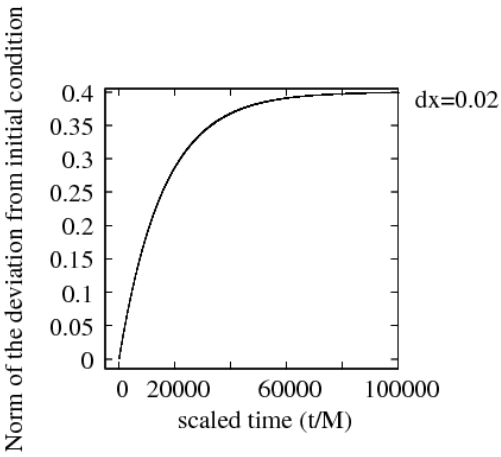

Fig. 10: $\|q_h - q_h^0\|(t)$ *when* $0 \leq t/M \leq 10^5$
All Mach Godunov scheme

with again two possible expressions for the numerical flux $\Phi_{ij}^{\mathrm{AM,X}}$. In (80), X is a Godunov type scheme: *e.g.* X = Roe [3], X = VFRoe [4] or X = Lagrange + Projection type scheme (see §6.4).

The two possible expressions for $\Phi_{ij}^{\mathrm{AM,X}}$ are the following:

• **First expression:** The non-linear version of (74) is given by

$$\Phi_{ij}^{\mathrm{AM,X}} = \Phi_{ij}^{\mathrm{X}} + (\theta_{ij} - 1)\frac{\rho_{ij}a_{ij}}{2}\begin{pmatrix} 0 \\ \Big[(\mathbf{u}_i - \mathbf{u}_j)\cdot\mathbf{n}_{ij}\Big]\mathbf{n}_{ij} \end{pmatrix} \tag{81}$$

where Φ_{ij}^{X} is the unmodified flux given by the X scheme and where

$$\theta_{ij} = \theta(M_{ij}) \qquad \text{with} \qquad \theta(M) = \min(M, 1), \tag{82}$$

M_{ij}, ρ_{ij} and a_{ij} being estimates at the edge Γ_{ij} respectively of the Mach number, the density and the sound velocity. Thus, the *all Mach correction* is now given by

$$(\theta_{ij} - 1)\frac{\rho_{ij}a_{ij}}{2}\begin{pmatrix} 0 \\ \Big[(\mathbf{u}_i - \mathbf{u}_j)\cdot\mathbf{n}_{ij}\Big]\mathbf{n}_{ij} \end{pmatrix} \tag{83}$$

and introduces anti-diffusion since $\theta_{ij} - 1 \leq 0$. The flux $\Phi_{ij}^{\mathrm{AM,Roe}}$ obtained with (81) and when X is the Roe scheme [3] is explicited in Annex B in the subsonic case (see (173)).

• **Second expression:** The non-linear version of (76) is given by

$$\Phi_{ij}^{\mathrm{AM,X}} = \begin{pmatrix} \rho^*(\mathbf{u}\cdot\mathbf{n})^* \\ \rho^*(\mathbf{u}^*\cdot\mathbf{n})\mathbf{u}^* + p^{**}\mathbf{n} \end{pmatrix}_{ij} \qquad \text{with} \qquad p_{ij}^{**} = \theta_{ij}p_{ij}^* + (1-\theta_{ij})\frac{p_i + p_j}{2} \tag{84}$$

where $(\rho^*, \mathbf{u}^*)$ is solution of a 1D (linearized or non-linearized) Riemann problem. Let us note that p^{**} in (84) replaces $p^* := p(\rho^*)$. As in the linear case (see (73)(76)), the non-linear *all Mach X scheme* (80)(84) may be seen as a Godunov type scheme whose Riemann solver is corrected to be accurate at low Mach number.

We underline that the non-linear *all Mach X schemes* (80)(81) (which is the non-linear version of (73)(74)) and (80)(84) (which is the non-linear version of (73)(76)) are not equivalent although the linear schemes (73)(74) and (73)(76) are equivalent.

We propose the following conjecture whose a better formulation will be proposed in §9.3:

Conjecture 5.1. *Let us suppose that the Godunov type scheme X applied to the barotropic Euler system (2) is stable for any Mach number lower than β (with $\beta \geq 1$)*[4]*, and is accurate when the Mach number belongs*

[4]The Mach number β depends on the X scheme, on the mesh and on the type of test-case. For example, the 1D Roe scheme is not always stable when the Mach number is too high; in that case, β is greater than one *but of order one*. On the other hand, we can choose $\beta = +\infty$ for the 1D exact Godunov scheme since it is stable (and entropic) by construction for any 1D test-case.

to $[\alpha, \beta]$ *(with* $\alpha \in]0, 1[$*)*[5]*. Then, in the periodic case and when the solution does not have any schock wave, the non-linear schemes (80)(81) and (80)(84) are stable and accurate on any mesh type and for any Mach number which belongs to* $[0, \beta]$.

In the two following sections, we (partly) justify Conjecture 5.1 when the X scheme is the Roe scheme [3]. More precisely:

- In Section 6, we (partly) justify the *stability question* by proposing a linear stability result in the subsonic case.
- In Section 7, we (partly) justify the *accuracy question* with a formal asymptotic expansion applied to (80)(81) when the X scheme is the Roe scheme [3].

We will see in Section 9 that the problem of accuracy is more complicate when the Mach number is of order one and when there are schock waves. Indeed, the *all Mach Roe scheme* (80)(81) has to be modified to avoid the creation of non-entropic shock waves when the Mach number is of order one. This modification will consist in introducing a cut-off in the definition (82) of θ_{ij} used in the all Mach correction (83). This underlines that the formulation of Conjecture 5.1 has to be improved, which will be done in Conjecture 9.1. Numerical results proposed in Section 9 justify this cut-off in the case of the compressible Euler system (1) for the Sod tube problem.

6. A linear stability result in the barotropic case

We now prove a linear stability result for the *all Mach Godunov type schemes* (80)(81) and (80)(84). This result (partly) justifies Conjecture 5.1 concerning the *stability question*.

We study this stability question by extending the linear Godunov scheme (73)(74) to the linear system

$$\begin{cases} \partial_t q + Hq + \dfrac{L}{M} q = 0, \\ \\ q(t = 0, x) = q^0(x) \end{cases} \tag{85}$$

where $\mathbf{u}_* \in \mathbb{R}^d$ ($d \in \{1, 2, 3\}$) is a constant velocity field. It is important to take into account the linear transport operator $Hq := (\mathbf{u}_* \cdot \nabla)q$ because the discretization of this operator has an impact on the stability of the all Mach Godunov scheme as we will see below. This is due to the fact that the Godunov approach does not split the material and acoustic waves (respectively described with $\partial_t q + Hq = 0$ and $\partial_t q + \dfrac{L}{M} q = 0$). This leads us to conclude this section with a remark on the Lagrange + Projection approach which splits the material and acoustic waves.

[5]The Mach number α is lower than one but of order one (*e.g.* $\alpha = 1/2$) since the Godunov type schemes are not accurate at low Mach number (when the dimension of the space is greater than one). Nevertheless, it is impossible to clearly define the Mach number α since it is impossible to define a clear boundary between the incompressible regime and the compressible regime. We can only say that α depends on the expected accuracy.

6.1. The linear all Mach Godunov scheme

When the Godunov scheme is applied to System (85) and when the flow is subsonic *i.e.*

$$|\mathbf{u}_*| < \frac{a_*}{M} \qquad \text{(subsonic condition)}, \tag{86}$$

the Godunov flux $\Phi_{ij}^{\text{Godunov}}$ is given by (see (158) in Annex A)

$$\Phi_{ij}^{\text{Godunov}} = \Phi_{ij}^{\text{Godunov,convection}} + \Phi_{ij}^{\text{Godunov,acoustic}} \tag{87}$$

where

$$\Phi_{ij}^{\text{Godunov,convection}} = \frac{1}{2}\begin{pmatrix} (\mathbf{u}_* \cdot \mathbf{n}_{ij})\Big[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\Big] \\ (\mathbf{u}_* \cdot \mathbf{n}_{ij})\Big[(\mathbf{u}_i + \mathbf{u}_j) + (r_i - r_j)\mathbf{n}_{ij}\Big] - |\mathbf{u}_* \cdot \mathbf{n}_{ij}|\Big[(\mathbf{u}_i - \mathbf{u}_j)\times \mathbf{n}_{ij}\Big]\times \mathbf{n}_{ij} \end{pmatrix} \tag{88}$$

and

$$\Phi_{ij}^{\text{Godunov,acoustic}} = \frac{a_*}{2M}\begin{pmatrix} (\mathbf{u}_i + \mathbf{u}_j)\cdot \mathbf{n}_{ij} + r_i - r_j \\ \Big[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\Big]\mathbf{n}_{ij} \end{pmatrix}. \tag{89}$$

Fluxes (88) and (89) discretize respectively the linear convection operator Hq and the linear acoustic operator $\frac{L}{M}q$. Flux (89) is of course identical to the Godunov flux in (37).

To obtain the *all Mach version* $\Phi_{ij}^{\text{AM,Godunov}}$ of $\Phi_{ij}^{\text{Godunov}}$ defined by (87), we just add the all Mach correction (75) to $\Phi_{ij}^{\text{Godunov}}$ as in (74). Thus, this consists in replacing the acoustic flux $\Phi_{ij}^{\text{Godunov,acoustic}}$ defined by (89) by the *all Mach acoustic flux*

$$\Phi_{ij}^{\text{AM,Godunov,acoustic}} = \Phi_{ij}^{\text{Godunov,acoustic}} + [\theta(M) - 1]\frac{a_*}{2M}\begin{pmatrix} 0 \\ \Big[(\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\Big]\mathbf{n}_{ij} \end{pmatrix}.$$

In other words, we do not correct the convective flux $\Phi_{ij}^{\text{Godunov,convection}}$ defined by (88), which is coherent with the fact that the low Mach number problem is only linked to a bad discretization of the acoustic operator at low Mach number.

To summarize, under the subsonic condition (86), the linear *all Mach Godunov scheme* applied to (85) is given by

$$\begin{cases} \dfrac{d}{dt}r_i + \dfrac{1}{2|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \quad \Big\{(\mathbf{u}_* \cdot \mathbf{n}_{ij})\Big[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\Big] + \dfrac{a_*}{M}\Big[(\mathbf{u}_i + \mathbf{u}_j)\cdot \mathbf{n}_{ij} + r_i - r_j\Big]\Big\} = 0, & \text{(a)} \\ \dfrac{d}{dt}\mathbf{u}_i + \dfrac{1}{2|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \quad \Big\{(\mathbf{u}_* \cdot \mathbf{n}_{ij})\Big[(\mathbf{u}_i + \mathbf{u}_j) + (r_i - r_j)\mathbf{n}_{ij}\Big] - |\mathbf{u}_* \cdot \mathbf{n}_{ij}|\Big[(\mathbf{u}_i - \mathbf{u}_j)\times \mathbf{n}_{ij}\Big]\times \mathbf{n}_{ij} & \text{(b)} \\ \qquad\qquad + \dfrac{a_*}{M}\Big[r_i + r_j + \theta(M)(\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\Big]\mathbf{n}_{ij}\Big\} = 0. \end{cases} \tag{90}$$

6.2. L^2-stability in the semi-discrete case

Let us define the energy

$$E_h = \sum_i |\Omega_i|(r_i^2 + |\mathbf{u}_i|^2).$$

We have the following L^2-stability result:

Theorem 6.1. *Let $(r, \mathbf{u})$ be solution of (90). Under the subsonic condition (86):*

1) For the Godunov scheme i.e. when $\theta(M) := 1$, we have:

$$\frac{d}{dt}E_h \leq 0. \tag{91}$$

2) For the all Mach Godunov scheme i.e. when $\theta(M) := \frac{|\mathbf{u}_*|}{a_*}M$, we have:

$$\frac{d}{dt}E_h \leq 0. \tag{92}$$

3) For the low Mach Godunov scheme i.e. when $\theta(M) := 0$, we have:

$$\frac{d}{dt}E_h \leq \sum_{\Gamma_{ij}} |\Gamma_{ij}|\, |\mathbf{u}_* \cdot \mathbf{n}_{ij}|\, |(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}|^2. \tag{93}$$

Inequality (91) confirms that the Godunov scheme is stable. Inequality (92) shows that the *all Mach Godunov scheme* is stable and, thus, justifies *from the stability point of view* the all Mach correction (83). It also underlines that the numerical dissipation of the linear (non-linear) Godunov scheme is (may be) too high when the flow is subsonic. Let us note that $\theta(M) := \frac{|\mathbf{u}_*|}{a_*}M$ is equal to the Mach number since a_*/M defined the sound velocity in (85). Inequality (93) avoids to obtain $\frac{d}{dt}E_h \leq 0$ when $\mathbf{u}_* \neq 0$. As a consequence, we may observe numerical instabilities except when $\mathbf{u}_* := 0$ *i.e.* when we restrict the stability analysis to the *low Mach Godunov scheme* applied to the linear wave equation (*i.e.* to the scheme (37) with $\kappa = 0$).

Proof of Theorem 6.1: Before proving *Points 1*, *2* and *3*, we perform some preliminary calculations.

Preliminary calculations: By multiplying (90)(a) with $2|\Omega_i|r_i$ and by summing with respect to i, we obtain

$$\frac{d}{dt}\sum_i |\Omega_i| r_i^2 = -\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \left\{ (\mathbf{u}_* \cdot \mathbf{n}_{ij})[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}]r_i + \frac{a_*}{M}[(\mathbf{u}_i + \mathbf{u}_j)\cdot \mathbf{n}_{ij} + r_i - r_j]r_i \right\}. \tag{94}$$

On the other hand, by using (45), we obtain

$$\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij}) r_i^2 = \sum_i \left(r_i^2 \mathbf{u}_* \cdot \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \mathbf{n}_{ij} \right) = 0.$$

Moreover

$$\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij}) r_i r_j = \sum_{\Gamma_{ij}} |\Gamma_{ij}|[\mathbf{u}_* \cdot (\mathbf{n}_{ij} + \mathbf{n}_{ji})] r_i r_j = 0$$

since $\mathbf{n}_{ij} + \mathbf{n}_{ji} = 0$. We deduce from the last two equalities that

$$\sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})(r_i + r_j) r_i = 0. \tag{95}$$

We have also

$$\sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}] r_i = \sum_{\Gamma_{ij}} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}] r_i + \sum_{\Gamma_{ij}} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ji})[(\mathbf{u}_j - \mathbf{u}_i) \cdot \mathbf{n}_{ji}] r_j$$

which gives

$$\sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}] r_i = \sum_{\Gamma_{ij}} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}](r_i - r_j). \tag{96}$$

Moreover

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|[(\mathbf{u}_i + \mathbf{u}_j) \cdot \mathbf{n}_{ij}] r_i = \frac{a_*}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_j \cdot \mathbf{n}_{ij}) r_i$$

by using again (45). We have also

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_j \cdot \mathbf{n}_{ij}) r_i = \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|(r_i \mathbf{u}_j \cdot \mathbf{n}_{ij} + r_j \mathbf{u}_i \cdot \mathbf{n}_{ji}),$$

which allows to write

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|[(\mathbf{u}_i + \mathbf{u}_j) \cdot \mathbf{n}_{ij}] r_i = \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|(r_i \mathbf{u}_j - r_j \mathbf{u}_i) \cdot \mathbf{n}_{ij}. \tag{97}$$

At last, we have

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(r_i - r_j) r_i = \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|(r_i - r_j) r_i + \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|(r_j - r_i) r_j.$$

Hence

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}|(r_i - r_j) r_i = \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}| \cdot |r_i - r_j|^2. \tag{98}$$

Thus, by using (94), (95), (96), (97) and (98), we find

$$\frac{d}{dt} \sum_i |\Omega_i| r_i^2 = - \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left\{ (\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}](r_i - r_j) + \frac{a_*}{M} \left[(r_i \mathbf{u}_j - r_j \mathbf{u}_i) \cdot \mathbf{n}_{ij} + |r_i - r_j|^2 \right] \right\}. \tag{99}$$

Let us now multiply (90)(b) with $2|\Omega_i|\mathbf{u}_i$. By summing with respect to i and by defining $\mathbf{t}_{ij}$ in such a way

$$\Big[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij} \Big] \mathbf{t}_{ij} = - \Big[(\mathbf{u}_i - \mathbf{u}_j) \times \mathbf{n}_{ij} \Big] \times \mathbf{n}_{ij} \qquad (|\mathbf{t}_{ij}| = 1 \text{ and } \mathbf{t}_{ij} \perp \mathbf{n}_{ij}) \tag{100}$$

(see (159) in Annex A), we obtain

$$\frac{d}{dt} \sum_i |\Omega_i| \, |\mathbf{u}_i|^2 = - \sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \quad \Big\{ (\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i + \mathbf{u}_j) + (r_i - r_j)\mathbf{n}_{ij}] \cdot \mathbf{u}_i + |\mathbf{u}_* \cdot \mathbf{n}_{ij}|[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij}](\mathbf{t}_{ij} \cdot \mathbf{u}_i)$$
$$+ \frac{a_*}{M} [r_i + r_j + \theta(M)(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}](\mathbf{n}_{ij} \cdot \mathbf{u}_i) \Big\}. \tag{101}$$

Let us note that in 3D, $\mathbf{t}_{ij}$ depends on $\mathbf{n}_{ij}$, $\mathbf{u}_i$ and $\mathbf{u}_j$. On the other hand, by using the arguments used to obtain (95) and (97), we respectively find

$$\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})(\mathbf{u}_i + \mathbf{u}_j) \cdot \mathbf{u}_i = 0 \tag{102}$$

and

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(r_i + r_j)(\mathbf{n}_{ij} \cdot \mathbf{u}_i) = \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|(r_j\mathbf{u}_i - r_i\mathbf{u}_j) \cdot \mathbf{n}_{ij}. \tag{103}$$

We have also

$$\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})(r_i - r_j)(\mathbf{n}_{ij} \cdot \mathbf{u}_i) = \sum_{\Gamma_{ij}} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})(r_i - r_j)(\mathbf{n}_{ij} \cdot \mathbf{u}_i) + \sum_{\Gamma_{ij}} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ji})(r_j - r_i)(\mathbf{n}_{ji} \cdot \mathbf{u}_j)$$

which allows to write

$$\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})(r_i - r_j)(\mathbf{n}_{ij} \cdot \mathbf{u}_i) = \sum_{\Gamma_{ij}} |\Gamma_{ij}|(\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}](r_i - r_j). \tag{104}$$

Moerover

$$\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\,[(\mathbf{u}_i - \mathbf{u}_j)\cdot\mathbf{n}_{ij}](\mathbf{n}_{ij}\cdot\mathbf{u}_i) = \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|\,[(\mathbf{u}_i - \mathbf{u}_j)\cdot\mathbf{n}_{ij}](\mathbf{n}_{ij}\cdot\mathbf{u}_i) + \frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|\,[(\mathbf{u}_j - \mathbf{u}_i)\cdot\mathbf{n}_{ji}](\mathbf{n}_{ji}\cdot\mathbf{u}_j)$$

and therefore

$$\theta(M)\frac{a_*}{M} \sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}](\mathbf{n}_{ij} \cdot \mathbf{u}_i) = \theta(M)\frac{a_*}{M} \sum_{\Gamma_{ij}} |\Gamma_{ij}|\,|(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}|^2. \tag{105}$$

At last, we obtain

$$\begin{aligned}\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\,|\mathbf{u}_* \cdot \mathbf{n}_{ij}|\,[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij}](\mathbf{t}_{ij} \cdot \mathbf{u}_i) \quad &= \quad \sum_{\Gamma_{ij}} |\Gamma_{ij}|\,|\mathbf{u}_* \cdot \mathbf{n}_{ij}|\,[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij}](\mathbf{t}_{ij} \cdot \mathbf{u}_i) \\ &\quad + \sum_{\Gamma_{ij}} |\Gamma_{ij}|\,|\mathbf{u}_* \cdot \mathbf{n}_{ji}|\,[(\mathbf{u}_j - \mathbf{u}_i) \cdot \mathbf{t}_{ji}](\mathbf{t}_{ji} \cdot \mathbf{u}_j)\end{aligned}$$

which gives

$$\sum_i \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\,|\mathbf{u}_* \cdot \mathbf{n}_{ij}|\,[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij}](\mathbf{t}_{ij} \cdot \mathbf{u}_i) = \sum_{\Gamma_{ij}} |\Gamma_{ij}|\,|\mathbf{u}_* \cdot \mathbf{n}_{ij}|\,|(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij}|^2 \tag{106}$$

(let us note that $\mathbf{t}_{ij} = -\mathbf{t}_{ji}$). Thus, by using (101), (102), (103), (104), (105) and (106), we find

$$\begin{aligned}\frac{d}{dt}\sum_i |\Omega_i|\,|\mathbf{u}_i|^2 \quad = \quad -\sum_{\Gamma_{ij}} |\Gamma_{ij}| \quad &\Big\{(\mathbf{u}_* \cdot \mathbf{n}_{ij})[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}](r_i - r_j) + |\mathbf{u}_* \cdot \mathbf{n}_{ij}|\,|(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{t}_{ij}|^2 \\ &+ \frac{a_*}{M}\left[(r_j\mathbf{u}_i - r_i\mathbf{u}_j) \cdot \mathbf{n}_{ij} + \theta(M)|(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij}|^2\right]\Big\}.\end{aligned} \tag{107}$$

Finally, by summing (99) and (107), we obtain

$$\begin{aligned}\frac{d}{dt}E_h &= -\sum_{\Gamma_{ij}}|\Gamma_{ij}| \quad \Big\{2(\mathbf{u}_*\cdot\mathbf{n}_{ij})[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}](r_i-r_j)+|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\,|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{t}_{ij}|^2 \\ &\qquad +\frac{a_*}{M}\left[|r_i-r_j|^2+\theta(M)|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2\right]\Big\}.\end{aligned} \tag{108}$$

Moreover, we have

$$-2(\mathbf{u}_*\cdot\mathbf{n}_{ij})[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}](r_i-r_j)\le|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\,\left[|\mathbf{u}_i-\mathbf{u}_j|^2+|r_i-r_j|^2\right].$$

Thus, by using the subsonic condition (86) and since $|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\le|\mathbf{u}_*|$, we obtain

$$-2(\mathbf{u}_*\cdot\mathbf{n}_{ij})[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}](r_i-r_j)\le|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\,|\mathbf{u}_i-\mathbf{u}_j|^2+\frac{a_*}{M}|r_i-r_j|^2.$$

By using (108), this allows to write

$$\frac{d}{dt}E_h\le-\sum_{\Gamma_{ij}}|\Gamma_{ij}|\left\{|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\,\left(|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{t}_{ij}|^2-|\mathbf{u}_i-\mathbf{u}_j|^2\right)+\frac{a_*}{M}\theta(M)|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2\right\}.$$

And by using (100), we obtain

$$\frac{d}{dt}E_h\le-\sum_{\Gamma_{ij}}|\Gamma_{ij}|\left\{-|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\,|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2+\frac{a_*}{M}\theta(M)|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2\right\}$$

which gives

$$\frac{d}{dt}E_h\le-\sum_{\Gamma_{ij}}|\Gamma_{ij}|\left\{\left[\frac{a_*}{M}\theta(M)-|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\right]|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2\right\}. \tag{109}$$

Proof of Points 1 and 2: Let us suppose that $\theta(M):=1$ (*cf. Point 1*). Since $|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\le|\mathbf{u}_*|$, we obtain $\frac{d}{dt}E_h\le 0$ by using the subsonic condition (86) and the inequality (109). When $\theta(M):=\frac{|\mathbf{u}_*|}{a_*}M$ (*cf. Point 2*), we deduce from (109) that

$$\frac{d}{dt}E_h\le-\sum_{\Gamma_{ij}}|\Gamma_{ij}|\left\{\left[|\mathbf{u}_*|-|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\right]|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2\right\}$$

which also gives $\frac{d}{dt}E_h\le 0$.

Proof of Point 3: When $\theta(M):=0$, we can only deduce from (109) that

$$\frac{d}{dt}E_h\le\sum_{\Gamma_{ij}}|\Gamma_{ij}|\,|\mathbf{u}_*\cdot\mathbf{n}_{ij}|\,|(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}|^2.$$

□

6.3. L^2-stability in the continuous case

To have a better understanding of the importance of the convection operator in Theorem 6.1, it is interesting to study the L^2-stability of the 1st order modified equation associated with (90). When the mesh is cartesian (we suppose for the sake of simplicity that the dimension is 2D), this equation is given by

$$\begin{cases} \partial_t q + \mathcal{H} q + \dfrac{\mathcal{L}_M}{M} q = 0, \\ \\ q(t=0,x) = q^0(x) \end{cases} \tag{110}$$

where, in the 2D case, $\mathcal{H}$ is the perturbed convection operator defined by

$$\mathcal{H} q = Hq - \frac{1}{2} \begin{pmatrix} u_{*,x} \Delta x \partial^2_{xx} u_x + u_{*,y} \Delta y \partial^2_{yy} u_y \\ \\ u_{*,x} \Delta x \partial^2_{xx} r + |u_{*,y}| \Delta y \partial^2_{yy} u_x \\ \\ |u_{*,x}| \Delta x \partial^2_{xx} u_y + u_{*,y} \Delta y \partial^2_{yy} r \end{pmatrix} \tag{111}$$

and where $\mathcal{L}_M$ is the perturbed linear acoustic operator defined by

$$\mathcal{L}_M q = Lq - \frac{a_*}{2} \begin{pmatrix} \Delta x \partial^2_{xx} r + \Delta y \partial^2_{yy} r \\ \\ \theta(M) \Delta x \partial^2_{xx} u_x \\ \\ \theta(M) \Delta y \partial^2_{yy} u_y. \end{pmatrix} \tag{112}$$

By defining the energy with

$$E := \langle q, q \rangle = \int_{\mathbb{T}^d} (r^2 + |\mathbf{u}|^2) dx,$$

we obtain the following result which is the continuous version of Theorem 6.1:

Theorem 6.2. *Let $q(t,x)$ be solution of (110). Under the subsonic condition (86), we have:*

1) When $\theta(M) := 1$:

$$\frac{d}{dt} E \leq 0. \tag{113}$$

2) When $\theta(M) := \dfrac{|\mathbf{u}_*|}{a_*} M$:

$$\frac{d}{dt} E \leq 0. \tag{114}$$

3) When $\theta(M) := 0$:

$$\frac{d}{dt} E \leq \Delta x |u_{*,x}| \left(\|\partial_x u_x\|^2 - \|\partial_x u_y\|^2 \right) + \Delta y |u_{*,y}| \left(\|\partial_y u_y\|^2 - \|\partial_y u_x\|^2 \right). \tag{115}$$

Proof of Theorem 6.2: The proof is similar to the proof of Theorem 6.1. Nevertheless, it is more simple since the operators are continuous.

Preliminary calculations: By multiplying (110) with q and by integrating over Ω, we obtain that

$$\begin{aligned}\frac{d}{dt}E &= -2\Delta x\, u_{*,x}\langle \partial_x r, \partial_x u_x\rangle - 2\Delta y\, u_{*,y}\langle \partial_y r, \partial_y u_y\rangle - \Delta y|u_{*,y}|\cdot\|\partial_y u_x\|^2 - \Delta x|u_{*,x}|\cdot\|\partial_x u_y\|^2 \\ &\quad -\frac{a_*}{M}\left(\Delta x\|\partial_x r\|^2 + \Delta y\|\partial_y r\|^2\right) - \frac{a_*}{M}\theta(M)\left(\Delta x\|\partial_x u_x\|^2 + \Delta y\|\partial_y u_y\|^2\right).\end{aligned} \tag{116}$$

Moreover, we have

$$-2\Delta x\, u_{*,x}\langle \partial_x r, \partial_x u_x\rangle \leq \Delta x|u_{*,x}|(\|\partial_x r\|^2 + \|\partial_x u_x\|^2).$$

Thus, under the subsonic condition (86), we can write that

$$-2\Delta x\, u_{*,x}\langle \partial_x r, \partial_x u_x\rangle \leq \Delta x\frac{a_*}{M}\|\partial_x r\|^2 + \Delta x|u_{*,x}|\,\|\partial_x u_x\|^2.$$

In the same way, we have

$$-2\Delta y\, u_{*,y}\langle \partial_y r, \partial_y u_y\rangle \leq \Delta y\frac{a_*}{M}\|\partial_y r\|^2 + \Delta y|u_{*,y}|\,\|\partial_y u_y\|^2,$$

Then, we deduce from (116) that

$$\frac{d}{dt}E \leq -\Delta x\left[\frac{a_*}{M}\theta(M) - |u_{*,x}|\right]\|\partial_x u_x\|^2 - \Delta y\left[\frac{a_*}{M}\theta(M) - |u_{*,y}|\right]\|\partial_y u_y\|^2 - \Delta y|u_{*,y}|\cdot\|\partial_y u_x\|^2 - \Delta x|u_{*,x}|\cdot\|\partial_x u_y\|^2.$$

Proof of Points 1, 2 and 3: We conclude the proof as in the semi-discrete case (see the proof of Theorem 6.1).□

6.4. A remark on the Lagrange + Projection approach

The potential loss of stability when $\theta(M) := 0$ (see *Point 3* of Theorems 6.1 and 6.2) is directly linked to

$$\mathcal{E}(q) := -\frac{1}{2}\begin{pmatrix} u_{*,x}\Delta x\partial^2_{xx}u_x + u_{*,y}\Delta y\partial^2_{yy}u_y \\ \\ u_{*,x}\Delta x\partial^2_{xx}r \\ \\ u_{*,y}\Delta y\partial^2_{yy}r \end{pmatrix}$$

in (111) which is the *non-disspipative part* of the truncation error of the Godunov scheme applied to the linear equation (85). The existence of this *non-dissipative* truncation error is a consequence of the fact that the Godunov scheme is built by taking into account *at the same time* the convective and acoustic waves (see Annex A). This suggests that a *Lagrange + Projection approach* – which consists in splitting the acoustic and convective waves – may not have any stability problem when $\theta(M) := 0$.

Indeed, a Lagrange + Projection approach applied to the linear equation (85) consists in computing an estimate of the solution by solving

$$\begin{cases} \partial_t q + \dfrac{L}{M}q = 0, \\ \\ q(t=0,x) = q^0(x) \end{cases} \qquad (\textit{Lagrange step}) \tag{117}$$

and, then, to correct this estimate by solving

$$\begin{cases} \partial_t q + Hq = 0, \\ \\ q(t=0,x) = q^0(x). \end{cases} \qquad (\textit{Projection step}) \tag{118}$$

Let us now suppose that we solve (117) with the *all Mach Godunov scheme* (73)(74) and that we solve (118) with the Godunov scheme (*i.e.* with the classical upwind scheme). For this particular Lagrange + Projection scheme, the 1st order modified equation is given by

$$\begin{cases} \partial_t q + \mathcal{H}q + \dfrac{\mathcal{L}_M}{M} q = 0, \\ \\ q(t=0,x) = q^0(x) \end{cases} \tag{119}$$

where, in the 2D case, $\mathcal{H}$ is the perturbed convective operator

$$\mathcal{H}q = Hq - \frac{1}{2} \begin{pmatrix} |u_{*,x}|\Delta x \partial^2_{xx} r_x + |u_{*,y}|\Delta y \partial^2_{yy} r \\ \\ |u_{*,x}|\Delta x \partial^2_{xx} u_x + |u_{*,y}|\Delta y \partial^2_{yy} u_x \\ \\ |u_{*,x}|\Delta x \partial^2_{xx} u_y + |u_{*,y}|\Delta y \partial^2_{yy} u_y \end{pmatrix} \tag{120}$$

and where $\mathcal{L}_M$ is the perturbed linear acoustic operator defined by (112). In that case, we easily obtain:

Theorem 6.3. *Let $q(t,x)$ be solution of (119). For any $\theta(M) \geq 0$, we have:*

$$\frac{d}{dt} E \leq 0.$$

It would be also easy to obtain the semi-discrete version of Theorem 6.3.

7. Formal asymptotic analysis in the barotropic case

We now (partly) justify Conjecture 5.1 *for the accuracy question* with a formal asymptotic analysis applied to the *all Mach Godunov type scheme* (80)(81) when X is the Roe scheme [3]. This analysis is classical [5, 12, 14]. The original point in the following calculus is that we clearly link this formal asymptotic analysis to the point of view proposed in §2.5 and we describe its weakeness.

When the X scheme is the Roe scheme [3], the dimensionless *all Mach Roe scheme* deduced from

(80)(81) and restricted to the subsonic case is given by (see (174) in Annex B)

$$\left\{\begin{array}{ll}\frac{d}{dt}\rho_i + \frac{1}{2|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left\{(\rho_i\mathbf{u}_i+\rho_j\mathbf{u}_j)\cdot\mathbf{n}_{ij} + M\frac{\rho_{ij}}{a_{ij}}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij} + \frac{a_{ij}}{M}(\rho_i-\rho_j)\right\} = 0, & \text{(a)}\\ \\ \frac{d}{dt}(\rho_i\mathbf{u}_i) + \frac{1}{2|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left\{\rho_i(\mathbf{u}_i\cdot\mathbf{n}_{ij})\mathbf{u}_i + \rho_j(\mathbf{u}_j\cdot\mathbf{n}_{ij})\mathbf{u}_j + \frac{a_{ij}}{M}(\rho_i-\rho_j)\left[\mathbf{u}_{ij}+(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})\mathbf{n}_{ij}\right]\right. & \text{(b)}\\ \\ \qquad -\rho_{ij}|\mathbf{u}_{ij}\cdot\mathbf{n}_{ij}|\left[(\mathbf{u}_i-\mathbf{u}_j)\times\mathbf{n}_{ij}\right]\times\mathbf{n}_{ij} + M\frac{\rho_{ij}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})}{a_{ij}}[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{u}_{ij} & \\ \\ \qquad \left.+\left[\frac{1}{M^2}(p_i+p_j)+\frac{\theta_{ij}}{M}\rho_{ij}a_{ij}(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}\right]\mathbf{n}_{ij}\right\} = 0 & \end{array}\right. \tag{121}$$

with $p_k = p(\rho_k)$, $a_{ij} = \dfrac{p_i - p_j}{\rho_i - \rho_j}$ and

$$\theta_{ij} = \theta(M_{ij}) \qquad \text{with} \qquad \theta(M_{ij}) = \min(M_{ij}, 1). \tag{122}$$

In (122), the local Mach number M_{ij} is given by $M_{ij} = M\dfrac{|\mathbf{u}_{ij}|}{a_{ij}}$. Thus, we can write that $M_{ij} = \mathcal{O}(M)$ which means that

$$\frac{\theta_{ij}}{M} \leq C \tag{123}$$

where C is a constant of order one (since $M_{ij} \ll 1$). Moreover, we impose periodic boundary conditions. Let us now assume the asymptotic expansion for $\phi = (\rho, \mathbf{u})$

$$\phi = \phi^{(0)} + M\phi^{(1)} + M\phi^{(2)} + \dots. \tag{124}$$

By plugging (124) in (121) and by separating the orders M^{-1} and M^0, we obtain:

- **Order M^{-1}:** We deduce from (121)(a) that

$$\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|a_{ij}^{(0)}\left(\rho_i^{(0)}-\rho_j^{(0)}\right) = 0.$$

Thus, we have $\sum_i \rho_i^{(0)}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|a_{ij}^{(0)}\left(\rho_i^{(0)}-\rho_j^{(0)}\right) = 0$ which gives

$$\sum_{\Gamma_{ij}}|\Gamma_{ij}|a_{ij}^{(0)}\left(\rho_i^{(0)}-\rho_j^{(0)}\right)\rho_i^{(0)} = 0. \tag{125}$$

By permuting i and j, we obtain

$$\sum_{\Gamma_{ij}}|\Gamma_{ij}|a_{ij}^{(0)}\left(\rho_j^{(0)}-\rho_i^{(0)}\right)\rho_j^{(0)} = 0. \tag{126}$$

By adding (125) and (126), we obtain

$$\sum_{\Gamma_{ij}} |\Gamma_{ij}| a_{ij}^{(0)} \left(\rho_j^{(0)} - \rho_i^{(0)}\right)^2$$

which implies that

$$\forall i: \quad \rho_i^{(0)} = \rho^{(0)}(t)$$

and, thus, $p_i^{(0)} = p^{(0)}(t)$ and $a_{ij}^{(0)} = a^{(0)}(t)$. Moreover, we deduce from (121)(b) and (123) that

$$\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left(p_i^{(1)} + p_j^{(1)}\right) = 0. \tag{127}$$

Let us note that Equation (127) is equivalent to

$$\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left[p_i^{(1)} + p_j^{(1)} + \kappa \rho^{(0)} a^{(0)} (\mathbf{u}_i^{(0)} - \mathbf{u}_j^{(0)}) \cdot \mathbf{n}_{ij}\right] \mathbf{n}_{ij} = 0 \tag{128}$$

with $\kappa = 0$. In the case of the Roe scheme – which is defined by (121) and $\theta_{ij} = 1$ instead of (122) –, we obtain (128) with $\kappa = 1$.

- **Order 0:** We deduce from (121)(a) that

$$\frac{d}{dt}\rho^{(0)}(t) + \frac{1}{2|\Omega_i|} \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left\{\rho^{(0)} \left(\mathbf{u}_i^{(0)} + \mathbf{u}_j^{(0)}\right) \cdot \mathbf{n}_{ij} + a^{(0)} \left(\rho_i^{(1)} - \rho_j^{(1)}\right)\right\} = 0. \tag{129}$$

On the other side, we have

$$\sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \rho^{(0)} \left(\mathbf{u}_i^{(0)} + \mathbf{u}_j^{(0)}\right) \cdot \mathbf{n}_{ij} = \rho^{(0)} \sum_{\Gamma_{ij}} \left[\left(\mathbf{u}_i^{(0)} + \mathbf{u}_j^{(0)}\right) \cdot \mathbf{n}_{ij} + \left(\mathbf{u}_j^{(0)} + \mathbf{u}_i^{(0)}\right) \cdot \mathbf{n}_{ji}\right] = 0$$

and

$$\sum_i \sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| a_{ij}^{(0)} \left(\rho_i^{(1)} - \rho_j^{(1)}\right) = \sum_{\Gamma_{ij}} |\Gamma_{ij}| \left[a^{(0)} \left(\rho_i^{(1)} - \rho_j^{(1)}\right) + a^{(0)} \left(\rho_j^{(1)} - \rho_i^{(1)}\right)\right] = 0.$$

Thus, by using (129), we obtain $\sum_i \left(2|\Omega_i| \frac{d}{dt}\rho^{(0)}(t)\right) = 0$ and therefore $\frac{d}{dt}\rho^{(0)}(t) = 0$. In other words, we have

$$\forall i: \quad \rho_i^{(0)} = C^{st} \tag{130}$$

and, thus, $p_i^{(0)} = C^{st}$ and $a_{ij}^{(0)} = C^{st}$. By plugging (130) in (129), we obtain that

$$\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left\{\rho^{(0)} \left(\mathbf{u}_i^{(0)} + \mathbf{u}_j^{(0)}\right) \cdot \mathbf{n}_{ij} + a^{(0)} \left(\rho_i^{(1)} - \rho_j^{(1)}\right)\right\} = 0$$

which gives

$$\sum_{\Gamma_{ij} \subset \partial\Omega_i} |\Gamma_{ij}| \left\{\rho^{(0)} a^{(0)} \left(\mathbf{u}_i^{(0)} + \mathbf{u}_j^{(0)}\right) \cdot \mathbf{n}_{ij} + \left(p_i^{(1)} - p_j^{(1)}\right)\right\} = 0 \tag{131}$$

by using the fact that $(a^{(0)})^2 = p_i^{(1)}/\rho_i^{(1)}$.

To summarize, we have proved that a necessary condition of validity of the expansion (124) is that $(p_i^{(1)}, \mathbf{u}_i^{(0)}) \in \mathbb{R}^{3N}$ satisfies

$$\left\{ \begin{array}{l} \displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \left\{ \rho^{(0)} a^{(0)} \left(\mathbf{u}_i^{(0)} + \mathbf{u}_j^{(0)} \right) \cdot \mathbf{n}_{ij} + \left(p_i^{(1)} - p_j^{(1)} \right) \right\} = 0, \\ \displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \left[p_i^{(1)} + p_j^{(1)} + \kappa \rho^{(0)} a^{(0)} (\mathbf{u}_i^{(0)} - \mathbf{u}_j^{(0)}) \cdot \mathbf{n}_{ij} \right] \mathbf{n}_{ij} = 0 \end{array} \right. \tag{132}$$

with $\kappa = 0$ in the case of the non-linear *all Mach Roe scheme* (121)(122). In the case of the Roe scheme, we obtain (132) with $\kappa = 1$. Thus, by defining $r_i := p_i^{(1)}/(\rho^{(0)} a^{(0)})$, we obtain that when $(r_i, \mathbf{u}_i^{(0)}) \in \mathbb{R}^{3N}$ satisfies (132), $(r_i, \mathbf{u}_i^{(0)})$ belongs to the kernel (39) of the discrete acoustic operator $\mathbb{L}_{\kappa,h}$. By using Theorem 2.2 and Lemma 2.2, we obtain that the non-linear *all Mach Roe scheme* (*i.e.* (121) with $\theta_{ij} = \theta(M_{ij})$) may be accurate at low Mach number and that the non-linear Roe scheme (*i.e.* (121) with $\theta_{ij} = 1$) may not be accurate at low Mach number.

8. Construction of all Mach Godunov type schemes for the compressible Euler system

We extend in this section the *all Mach Godunov type schemes* (80)(81) and (80)(84) obtained for the barotropic Euler system (2) to the compressible Euler system (1).

The previous sections show that the low Mach number inaccuracy can be studied and cured in the barotropic case, which underlines that the energy equation may not have any influence on this question. This leads us to test the all Mach correction obtained and justified without energy equation to the case with energy equation. In other words, we propose the *all Mach Godunov type scheme*

$$\frac{d}{dt} \begin{pmatrix} \rho \\ \rho \mathbf{u} \\ \rho E \end{pmatrix}_i + \frac{1}{|\Omega_i|} \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}| \Phi_{ij}^{\mathrm{AM,X}} = 0 \tag{133}$$

with the two possible expressions for the numerical flux $\Phi_{ij}^{\mathrm{AM,X}}$ (we recall that X is a Godunov type scheme):

• **First expression:**

$$\Phi_{ij}^{\mathrm{AM,X}} = \Phi_{ij}^{\mathrm{X}} + (\theta_{ij} - 1) \frac{\rho_{ij} a_{ij}}{2} \begin{pmatrix} 0 \\ \left[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij} \right] \mathbf{n}_{ij} \\ 0 \end{pmatrix} \tag{134}$$

where Φ_{ij}^{X} is the unmodified flux given by the X scheme and where

$$\theta_{ij} = \theta(M_{ij}) \qquad \text{with} \qquad \theta(M) = \min(M, 1). \tag{135}$$

Thus, the all Mach correction is now given by

$$(\theta_{ij} - 1) \frac{\rho_{ij} a_{ij}}{2} \begin{pmatrix} 0 \\ \left[(\mathbf{u}_i - \mathbf{u}_j) \cdot \mathbf{n}_{ij} \right] \mathbf{n}_{ij} \\ 0 \end{pmatrix}. \tag{136}$$

Let us note that we could replace (136) by

$$(\theta_{ij} - 1) \frac{a_{ij}}{2} \begin{pmatrix} 0 \\ \left[(\rho_i \mathbf{u}_i - \rho_j \mathbf{u}_j) \cdot \mathbf{n}_{ij} \right] \mathbf{n}_{ij} \\ 0 \end{pmatrix} \tag{137}$$

or by

$$(\theta_{ij}-1)\frac{1}{2}\begin{pmatrix} 0 \\ \Big[(\rho_i a_i \mathbf{u}_i - \rho_j a_j \mathbf{u}_j)\cdot \mathbf{n}_{ij}\Big]\mathbf{n}_{ij} \\ 0 \end{pmatrix}. \tag{138}$$

• **Second expression:**

$$\Phi_{ij}^{AM,X} = \begin{pmatrix} \rho^*(\mathbf{u}\cdot\mathbf{n})^* \\ \rho^*(\mathbf{u}^*\cdot\mathbf{n})\mathbf{u}^* + p^{**}\mathbf{n} \\ (\rho^* E^* + p^*)(\mathbf{u}\cdot\mathbf{n})^* \end{pmatrix}_{ij} \quad \text{with} \quad p_{ij}^{**} = \theta_{ij} p_{ij}^* + (1-\theta_{ij})\frac{p_i + p_j}{2} \tag{139}$$

where $(\rho^*, \mathbf{u}^*, E^*)$ is solution of a 1D (linearized or non-linearized) Riemann problem.

Concerning the stability of the all Mach schemes (133)(134) and (133)(139):

These all Mach schemes – directly deduced from the barotropic case – are justified to cure the accuracy problem at low Mach number. But, it is not obvious that the linear stability result obtained in Section 6 in the barotropic case when the X scheme is the Roe scheme remains valid. Indeed, the energy equation is as important as the two other equations in any stability analysis. This point will have to be studied carefully in a future work although the numerical results proposed in Section 11 justify (133)(134) when the X scheme is the Roe scheme.

9. Introduction of a cut-off in the all Mach correction

We show in this section that the linear *all Mach Godunov scheme* (71)(72) is responsible for the creation of spurious oscillations. A direct consequence of these spurious oscillations may be the creation of non-entropic shock waves in the non-linear case by the *all Mach Godunov type schemes* (133)(134). This leads us to modify the all Mach correction (136) by introducing a cut-off in (135). We will justify this cut-off with numerical results in Section 11.

9.1. Loss of a TVD property induced by the all Mach correction

We justify the possible creation of spurious oscillations by the linear *all Mach Godunov scheme* (71)(72) with the following simple property:

Property 9.1. *Let us suppose that* $\Omega = \mathbb{R}$ *and that* $TV(r^0 \pm u^0) < +\infty$ *with* $TV(f) := \sum_i |f_i - f_{i-1}|$. *Then, the explicit scheme*

$$\begin{cases} \dfrac{r_i^{n+1} - r_i^n}{\Delta t} + \dfrac{a_*}{2M\Delta x}(u_{i+1}^n - u_{i-1}^n) = \dfrac{a_*}{2M\Delta x}(r_{i+1}^n - 2r_i^n + r_{i-1}^n), \\ \dfrac{u_i^{n+1} - u_i^n}{\Delta t} + \dfrac{a_*}{2M\Delta x}(r_{i+1}^n - r_{i-1}^n) = \theta\dfrac{a_*}{2M\Delta x}(u_{i+1}^n - 2u_i^n + u_{i-1}^n) \end{cases} \tag{140}$$

verifies:

1) When $\theta = 1$ and under the CFL condition $\Delta t \leq \dfrac{M\Delta x}{a_*}$:

$$\forall n \geq 0 : \quad TV(r^n \pm u^n) \leq TV(r^0 \pm u^0). \tag{141}$$

Moreover, when $r^0 \pm u^0$ is a monotone discrete function, we have

$$\forall n \geq 0 : \quad TV(r^n \pm u^n) = TV(r^0 \pm u^0). \tag{142}$$

2) When $\theta \in [0, 1[$ and under the CFL condition $\Delta t \leq \dfrac{M\Delta x}{2a_*}$, we can just write that

$$\forall n \geq 0 : \quad TV(r^{n+1} \pm u^{n+1}) \leq TV(r^n \pm u^n) + TV(u^n). \tag{143}$$

The first point is classical. It comes from the fact that the characteristic variables $r \pm u$ are advected with an upwind scheme when (r_i^n, u_i^n) is given by (140) with $\theta = 1$, and that the upwind scheme is TVD under a CFL-like condition. The second point comes from the fact that when $\theta \neq 1$, $r \pm u$ are not advected with an upwind scheme.

A consequence of the first point of Property 9.1 is that when $\theta = 1$, the variables $r_i^n \pm u_i^n$ do not present any spurious oscillations since a TVD scheme is monotonicity preserving [16]. By cons, the second point means that when $\theta \in [0, 1[$, it can appears spurious oscillations on $r \pm u$, which are then transmitted to r and u.

To illustrate Property 9.1, we study the numerical solution obtained with (140) on $\Omega = [0, 1]$ when the initial conditions are given by

$$\left\{ \begin{array}{l} r^0(x \leq 0.5) = 1, \\ \\ u^0(x \leq 0.5) = 0 \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} r^0(x > 0.5) = 0.1, \\ \\ u^0(x > 0.5) = 0. \end{array} \right. \tag{144}$$

Moreover, we choose $a_* = M = 1$. We study the numerical solution before the waves reach the boundary $\partial\Omega$. Thus, we impose the boundary conditions

$$(r, u)(t \geq 0, x \in \partial\Omega) = (r^0, u^0)(x \in \partial\Omega).$$

In the sequel, we choose a number of cells equal to 100.

Figures 11-16 show the numerical results when $t = 0,25$ obtained with (140) when $\theta = 1$, which corresponds to the Godunov scheme. Figures 17-22 show the numerical results obtained with (140) when $\theta = 0$, which corresponds to the *low Mach Godunov scheme*. These numerical results confirm that (140) creates spurious oscillations when $\theta \neq 1$.

Proof of Property 9.1: By using (140), we easily obtain

$$\begin{aligned} (J_\pm)_{i+1}^{n+1} - (J_\pm)_i^{n+1} \quad = \quad & \left(1 - \frac{a_*\Delta t}{M\Delta x}\right)[(J_\pm)_{i+1}^n - (J_\pm)_i^n] + \frac{a_*\Delta t}{M\Delta x}[(J_\pm)_{i+1\mp 1} - (J_\pm)_{i\mp 1}^n] \\ & \pm(\theta - 1)\frac{a_*\Delta t}{2M\Delta x}[(u_{i+2}^n - u_{i+1}^n) - 2(u_{i+1}^n - u_i^n) + (u_i^n - u_{i-1}^n)] \end{aligned} \tag{145}$$

where $(J_\pm)_i^n := r_i^n \pm u_i^n$. This relation allows us to easily obtain (141) when $\theta = 1$ and $\Delta t \leq \dfrac{M\Delta x}{a_*}$. Let us now suppose that

$$\forall i : \ (J_\pm)_{i+1}^n \leq (J_\pm)_i^n \qquad \text{or} \qquad \forall i : \ (J_\pm)_{i+1}^n \geq (J_\pm)_i^n. \tag{146}$$

Thus, we deduce from (145) that

$$\forall i : \ (J_\pm)_{i+1}^{n+1} \leq (J_\pm)_i^{n+1} \qquad \text{or} \qquad \forall i : \ (J_\pm)_{i+1}^{n+1} \geq (J_\pm)_i^{n+1}$$

when $\theta = 1$ and $\Delta t \leq \dfrac{M\Delta x}{a_*}$. And, by using again (145), we finally obtain

$$\begin{aligned} TV\left((J_\pm)^{n+1}\right) &= \left(1 - \frac{a_*\Delta t}{M\Delta x}\right) TV\left((J_\pm)^n\right) + \frac{a_*\Delta t}{M\Delta x} TV\left((J_\pm)^n\right) \\ &= TV\left((J_\pm)^n\right). \end{aligned}$$

Let us now suppose that $\theta \neq 1$. In that case, (145) implies that

$$TV\left((J_\pm)^{n+1}\right) \leq \left(\left|1 - \frac{a_*\Delta t}{M\Delta x}\right| + \frac{a_*\Delta t}{M\Delta x}\right) TV\left((J_\pm)^n\right) + |\theta - 1| \frac{2a_*\Delta t}{M\Delta x} TV(u^n),$$

which allows to obtain (143) when $\theta \in [0, 1[$ and $\Delta t \leq \dfrac{M\Delta x}{2a_*}$. □

9.2. Introduction of a cut-off in the all Mach correction

The spurious oscillations created in the linear case when $\theta \neq 1$ will be also present in the non-linear case (1) (or (2)). For example, when the initial conditions of the compressible Euler system (1) are defined by the Riemann problem

$$\left\{ \begin{array}{l} \rho^0(x \leq 0) = \rho_L, \\ p^0(x \leq 0) = p_L, \\ u^0(x \leq 0) = 0 \end{array} \right. \qquad \text{and} \qquad \left\{ \begin{array}{l} \rho^0(x \leq 0) = \rho_R, \\ p^0(x > 0) = p_R, \\ u^0(x > 0) = 0, \end{array} \right. \tag{147}$$

the behaviour of the pressure and of the velocity in the vicinity of $(t = 0, x = 0)$ is conditionned by the linear wave equation. Thus, the numerical results in the vicinity of $(t = 0, x = 0)$ obtained with a non-linear *all Mach Godunov type schemes* (133)(134) will be similar to those obtained on Figures 11-16 with the linear *all Mach Godunov schemes* (71)(72), that is to say will include similar spurious oscillations.

When the Mach number remains close to zero (that is to say when $|p_R - p_L|/p_L \ll 1$ in the case of the Riemann problem (147)), these spurious oscillations are not a difficulty. But, when the Mach number increases, these spurious oscillations may become non-entropic shock waves. This leads us to replace $\theta(M) = \min(M, 1)$ in (135) by $\theta_\alpha(M)$ with

$$\left\{ \begin{array}{ll} \theta_\alpha(M) = M & \text{if} \quad M < \alpha, \\ \theta_\alpha(M) = 1 & \text{if} \quad M \geq \alpha \end{array} \right. \qquad \text{where} \qquad \alpha \in]0, 1[\tag{148}$$

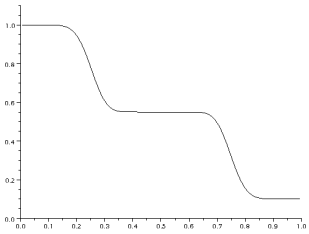

Fig. 11: $r(t = 0.25, x)$
$\theta = 1$

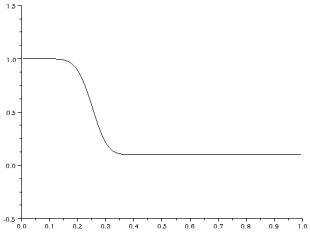

Fig. 12: $(r - u)(t = 0.25, x)$
$\theta = 1$

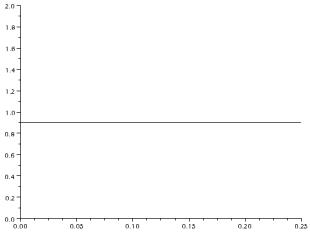

Fig. 13: $TV(r - u)(t)$
$\theta = 1$

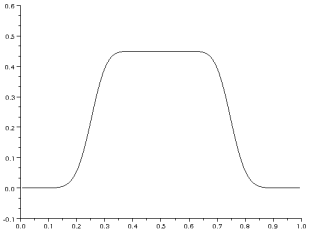

Fig. 14: $u(t = 0.25, x)$
$\theta = 1$

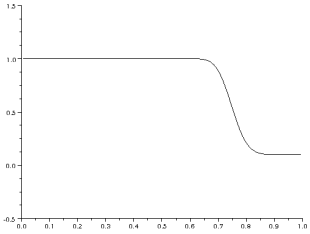

Fig. 15: $(r + u)(t = 0.25, x)$
$\theta = 1$

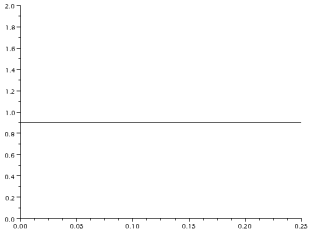

Fig. 16: $TV(r + u)(t)$
$\theta = 1$

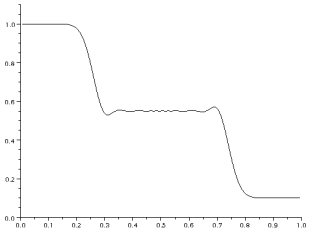

Fig. 17: $r(t = 0.25, x)$
$\theta = 0$

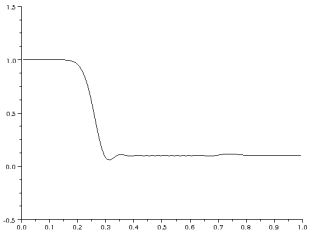

Fig. 18: $(r - u)(t = 0.25, x))$
$\theta = 0$

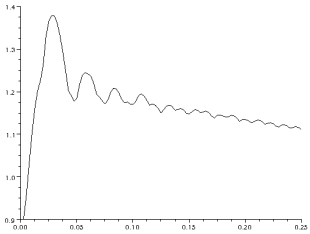

Fig. 19: $TV(r - u)(t)$
$\theta = 0$

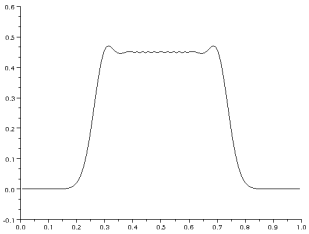

Fig. 20: $u(t = 0.25, x)$
$\theta = 0$

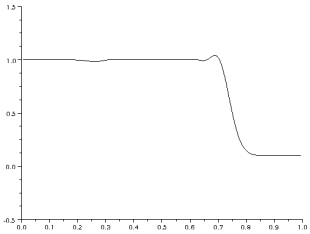

Fig. 21: $(r + u)(t = 0.25, x)$
$\theta = 0$

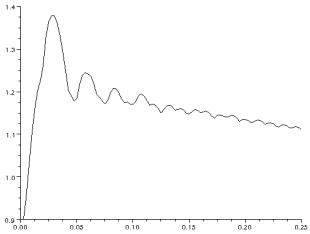

Fig. 22: $TV(r + u)(t)$
$\theta = 0$

(see Figures 23-24) in order to dump the spurious oscillations when the Mach number become larger than α. This means that we introduce a *cut-off* in the all Mach correction (136).

Let us note that α in (148) has to be of order one to keep the accuracy of the *all Mach Godunov type schemes* (133)(134) at low Mach number and, at the same time, has to be less than one to keep enough numerical viscosity to dump the spurious oscillations. The numerical results proposed in Section 11 justify this cut-off and show that $\alpha = 1/2$ is a good choice in the case of the Sod tube problem. Nevertheless, the choice of α remains heuristic at the present time.

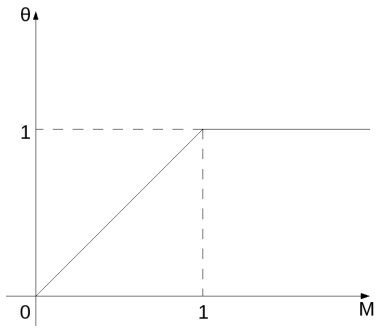

Fig. 23: $\theta(M)$ *(i.e. without cut-off)*

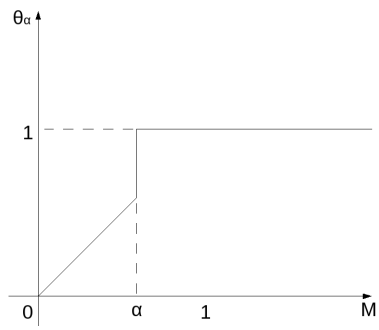

Fig. 24: $\theta_\alpha(M)$ *(i.e. with cut-off)*

9.3. Conjecture about the stability and accuracy of the all Mach Godunov type schemes

The previous numerical results incite us to improve Conjecture 5.1 in the following way:

Conjecture 9.1. *Let us suppose that the Godunov type scheme X applied to the barotropic Euler system (2) is stable for any Mach number lower than β (with $\beta \geq 1$), and is accurate when the Mach number belongs to $[\alpha, \beta]$ (with $\alpha \in]0, 1[$). Then, in the periodic case and with or without shock wave, the non-linear schemes (80)(81) and (80)(84) are stable and accurate on any mesh type and for any Mach number which belongs to $[0, \beta]$ when $\theta(M)$ is replaced by $\theta_\alpha(M)$ in (81) or (84).*

Let us note that we restrict this conjecture to the barotropic case. Indeed, the stability and the accuracy (at low Mach number) have been justified only in the barotropic case (2) (and when X is the Roe scheme [3]: see Sections 6 and 7). Thus, it remains to obtain similar results in the case of the compressible Euler system (1). Nevertheless, the numerical results proposed in Section 11 (and those already obtained in [8] when the Mach number is low) allow us to expect that Conjecture 9.1 is also valid in the case of the compressible Euler system (1) when we replace (80)(81) and (80)(84) with respectively (133)(134) and (133)(139).

10. Other all Mach schemes

The analysis used to justify the all Mach correction (136) is not limited to Godunov type schemes applied to the compressible Euler system (1). For example, the previous analysis applied to the Rusanov scheme [17] would lead us to use the all Mach correction (137) with a_{ij} replaced by $|\lambda_{ij}| := \max(|\mathbf{u}_{ij} \cdot \mathbf{n}_{ij} - a_{ij}|, |\mathbf{u}_{ij} \cdot \mathbf{n}_{ij} + a_{ij}|)$ in order to define the *all Mach Rusanov scheme*

$$\frac{d}{dt}\begin{pmatrix} \rho \\ \rho\mathbf{u} \\ \rho E \end{pmatrix}_i + \frac{1}{|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\Phi_{ij}^{\mathrm{AM,Rusanov}} = 0 \tag{149}$$

with

$$\Phi_{ij}^{\text{AM,Rusanov}} = \Phi_{ij}^{\text{Rusanov}} + (\theta_{ij} - 1)\frac{|\lambda_{ij}|}{2}\begin{pmatrix} 0 \\ \left[(\rho_i \mathbf{u}_i - \rho_j \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right]\mathbf{n}_{ij} \\ 0 \end{pmatrix} \tag{150}$$

where $\Phi_{ij}^{\text{Rusanov}}$ is the unmodified Rusanov flux and where $\theta_{ij} = \theta(M_{ij})$ or $\theta_{ij} = \theta_\alpha(M_{ij})$ ($\theta(M)$ and $\theta_\alpha(M)$ are defined by (135) and (148)). We could also formally justify the non-linear *all Mach Rusanov scheme* (149)(150) with a formal asymptotic analysis similar to the one used to justify the *all Mach Roe scheme* (121).

In the same way, when the mesh is 2D cartesian with $\Delta x = \Delta y$, the *all Mach Lax-Friedrichs scheme* is given by

$$\frac{d}{dt}\begin{pmatrix} \rho \\ \rho \mathbf{u} \\ \rho E \end{pmatrix}_i + \frac{1}{|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\Phi_{ij}^{\text{AM,LF}} = 0 \tag{151}$$

with

$$\Phi_{ij}^{\text{AM,LF}} = \Phi_{ij}^{\text{LF}} + (\theta_{ij} - 1)\frac{\Delta x}{2\Delta t}\begin{pmatrix} 0 \\ \left[(\rho_i \mathbf{u}_i - \rho_j \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right]\mathbf{n}_{ij} \\ 0 \end{pmatrix} \tag{152}$$

where Φ_{ij}^{LF} is the unmodified Lax-Friedrichs flux [18]. In (152), $\dfrac{\Delta x}{\Delta t}$ is equal to $\max\limits_i(|u_{x,i} \pm a_i|, |u_{y,i} \pm a_i|)/CFL$ with $CFL \leq 1$.

Concerning the stability of the all Mach schemes (149)(150) and (151)(152):

As in the case of the Godunov type schemes applied to the compressible Euler system (1) (see Section 8), the stability of these all Mach schemes will have to be carefully studied.

Let us also note that a Lagrange + Projection type scheme can also be corrected with a similar low Mach correction, and that the stability of this type of scheme should not be affected by the all Mach correction: see §6.4.

11. Numerical results

We study the behaviour of the *all Mach Roe scheme* (133)(134) (*i.e.* X is the Roe scheme [3]) when the initial conditions are those of the Sod tube problem that is to say

$$\begin{cases} \rho^0(x \leq 0.5) = 1, \\ p^0(x \leq 0.5) = 1, \\ u^0(x \leq 0.5) = 0 \end{cases} \quad \text{and} \quad \begin{cases} \rho^0(x \geq 0.5) = 0.125, \\ p^0(x \geq 0.5) = 0.1, \\ u^0(x \geq 0.5) = 0. \end{cases} \tag{153}$$

Moreover, we suppose that the fluid is a perfect gas whose adiabatic constant γ is equal to 1,4. The domain Ω is equal to $[0, 1]$ and we study the numerical solution before the waves reach the boundary $\partial\Omega$. Thus, we impose the boundary conditions

$$(\rho, p, u)(t \geq 0, x \in \partial\Omega) = (\rho^0, p^0, u^0)(x \in \partial\Omega).$$

We discretize the time operators in (133) with a first order Euler scheme, and the global scheme is explicit. Thus, the time step Δt is linked to the mesh size Δx through a classical CFL condition. The reference solution is obtained by using the Roe scheme with a number of cells equal to 10^4.

11.1. Numerical results without cut-off

We test the *all Mach Roe scheme* (133)(134) without cut-off *i.e.* with $\theta(M)$ given by (135).

Figures 25-28 show the results when $t = 0, 2$ and when the number of cells is equal to 10^4. These results show that the *all Mach Roe scheme* is stable and is accurate in a large part of the domain Ω. Nevertheless, the *all Mach Roe scheme* produces non-entropic shock waves in the vicinity of the foot of the rarefaction wave, where the Mach number is of order one (we recall that the Mach number is equal to $|u|/\sqrt{\gamma p/\rho}$ for a perfect gas). These non-entropic shock waves do not disappear when the mesh is refined.

These results show that we have to improve the *all Mach Roe scheme* (133)(134) to avoid the creation of non-entropic shock waves.

11.2. Numerical results with cut-off

The spurious oscillations obtained on Figures 17-22 – and justified by Property 9.1 – are responsible for the non-entropic shock waves obtained on Figures 25-28.

This leads us to replace in the *all Mach Roe scheme* (133)(134) the function $\theta(M) = \min(M, 1)$ by $\theta_\alpha(M)$ defined by (148) (see Figures 23-24) in order to avoid the non-entropic shock waves. This means that we introduce a cut-off in the all Mach correction (136).

Figures 29-32 show the numerical results for the Sod tube problem (when $t = 0, 2$ and with 10^4 cells) obtained when the *all Mach Roe scheme* (133)(134) uses $\theta_\alpha(M)$ with $\alpha = 1/2$. We see that the non-entropic shock waves have disappeared (compare with Figures 25-28).

Figures 33-36 compare the results obtained with the Roe scheme and with the *all Mach Roe scheme with cut-off* ($\alpha = 1/2$) when the number of cells is equal to 100. We see that the accuracy of the *all Mach Roe scheme* is better than the one of the Roe scheme in the rarefaction wave, which underlines that the numerical dissipation of the Roe scheme is too high in the rarefaction areas.

Let us note that the discontinuity of the derivatives where the Mach number is equal to $1/2$ with the *all Mach Roe scheme* (see Figures 33-36) is due to the non-regularity of the cut-off $\theta_\alpha(M)$ defined by (148) (we recall that $\alpha = 1/2$ in the present case). But, this does not affect the accuracy of the numerical results (this discontinuity disappears when the number of cells increases: see Figures 29-32). Nevertheless, we could make disappear this artefact when the number of cells is low by regularizing the cut-off.

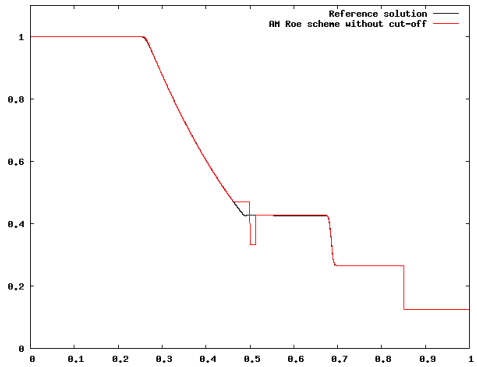

Fig. 25: $\rho(t = 0.2, x)$

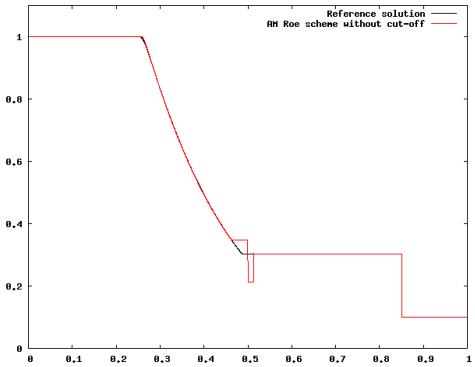

Fig. 26: $p(t = 0.2, x)$

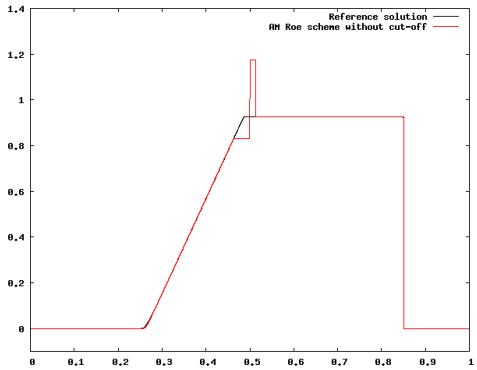

Fig. 27: $u(t = 0.2, x)$

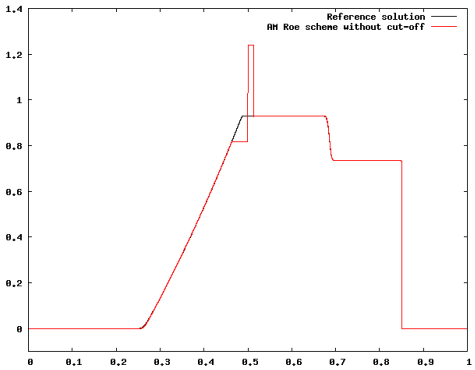

Fig. 28: $Mach(t = 0.2, x)$

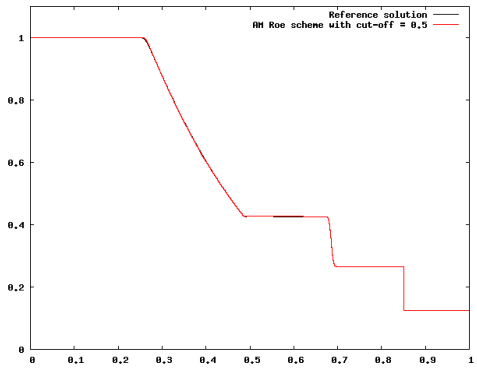

Fig. 29: $\rho(t = 0.2, x)$

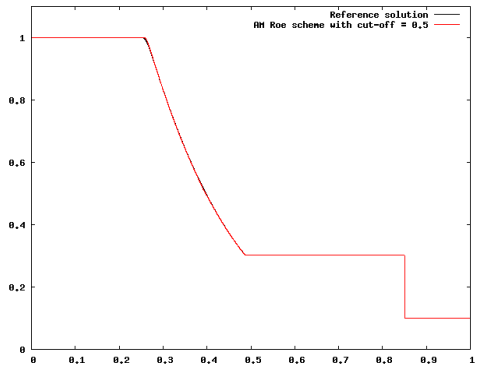

Fig. 30: $p(t = 0.2, x)$

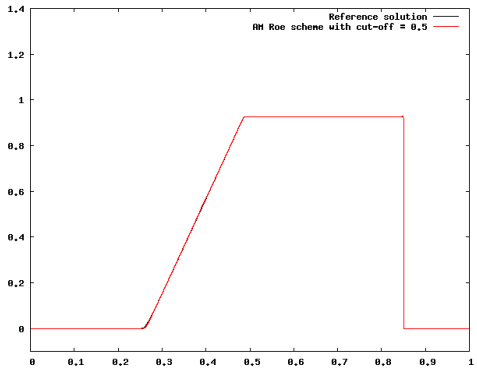

Fig. 31: $u(t = 0.2, x)$

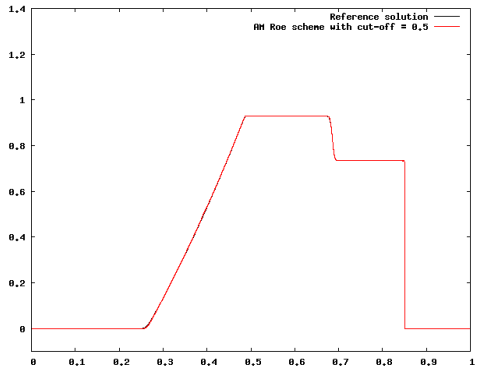

Fig. 32: $Mach(t = 0.2, x)$

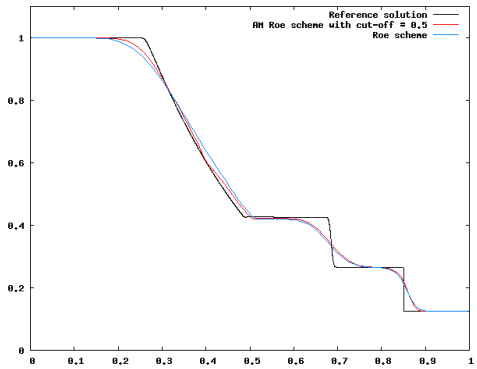

Fig. 33: $\rho(t = 0.2, x)$

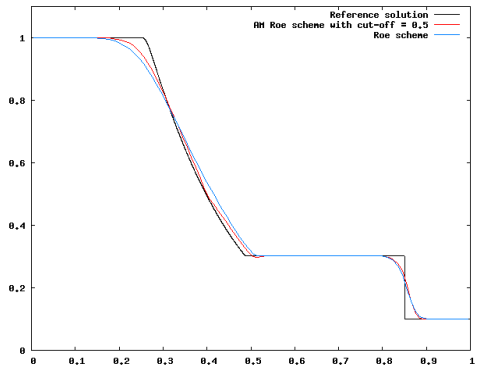

Fig. 34: $p(t = 0.2, x)$

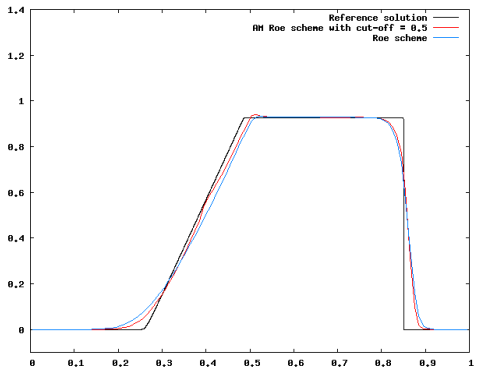

Fig. 35: $u(t = 0.2, x)$

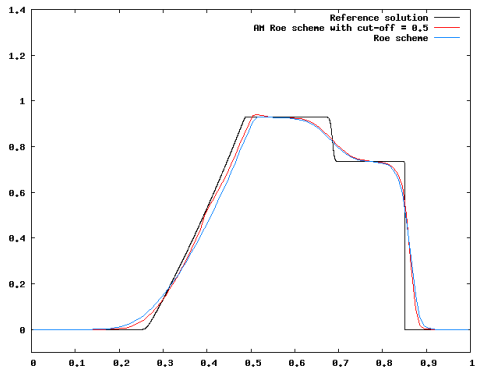

Fig. 36: $Mach(t = 0.2, x)$

12. Conclusion

Through the study of the linear wave equation discretized with a Godunov scheme, we have proposed a simple *all Mach correction* to apply to any Godunov type scheme solving the compressible Euler system to make this scheme accurate in the incompressible regime as well as in the compressible regime. We have named this modified scheme *all Mach Godunov type scheme*.

The short time behaviour of the solution of the first order equivalent equation associated with the *all Mach Godunov scheme* applied to the linear wave equation justifies this correction when the mesh is cartesian. In the non-linear barotropic case and when the Godunov type scheme is a Roe scheme, we justify this approach with a formal asymptotic expansion. At last, a linear stability result shows that this *all Mach Roe scheme* should be stable in the non-linear barotropic case.

We have proposed numerical results which justify this approach in the case of the compressible Euler system when the Godunov type scheme is a Roe scheme. Nevertheless, these numerical results underline also that when the Mach number is of order one, the *all Mach Roe scheme* may produce non-entropic shock waves when the proposed correction is not limited by a cut-off. On the other hand, the *all Mach Roe scheme* with or without cut-off remains accurate when the solution is smooth.

We justify this cut-off in the 1D linear case by showing that the all Mach correction has an impact on TVD properties. Nevertheless, it remains to justify it in the non-linear case by studying the entropic properties of the *all Mach Godunov type schemes* with or without cut-off.

At last, the proposed theoretical results have been obtained in the periodic case. Since the aim of this study is to obtain *all Mach Godunov type schemes* that can be applied to the modelling of a nuclear core and since a nuclear core is not a periodic domain, we will have to study the influence of non-periodic boundary conditions on the accuracy and stability of these *all Mach Godunov type schemes*.

A. The linear Godunov scheme in the subsonic case

The linear equation

$$\partial_t q + \mathbf{u}_* \cdot \nabla q + \frac{a_*}{M} L q = 0 \tag{154}$$

may be written with

$$\partial_t q + A_x \partial_x q + A_y \partial_y q + A_z \partial_z q = 0$$

where

$$A_x = \begin{pmatrix} u_{*,x} & \frac{a_*}{M} & 0 & 0 \\ \frac{a_*}{M} & u_{*,x} & 0 & 0 \\ 0 & 0 & u_{*,x} & 0 \\ 0 & 0 & 0 & u_{*,x} \end{pmatrix}, \quad A_y = \begin{pmatrix} u_{*,y} & 0 & \frac{a_*}{M} & 0 \\ 0 & u_{*,y} & 0 & 0 \\ \frac{a_*}{M} & 0 & u_{*,y} & 0 \\ 0 & 0 & 0 & u_{*,y} \end{pmatrix} \quad \text{and} \quad A_z = \begin{pmatrix} u_{*,z} & 0 & 0 & \frac{a_*}{M} \\ 0 & u_{*,z} & 0 & 0 \\ 0 & 0 & u_{*,z} & 0 \\ \frac{a_*}{M} & 0 & 0 & u_{*,z} \end{pmatrix}.$$

The one-dimensional Riemann problem associated with this equation in the direction $\mathbf{n} := (n_x, n_y, n_z)^T$ is defined by

$$\begin{cases} \partial_t q + A(\mathbf{n}) \partial_\zeta q = 0, \\ q(t=0,\zeta) = \begin{cases} q_L \text{ if } \zeta < 0, \\ q_R \text{ if } \zeta \geq 0 \end{cases} \end{cases} \tag{155}$$

where $A(\mathbf{n}) = A_x n_x + A_y n_y + A_z n_z$ that is to say

$$A(\mathbf{n}) = (\mathbf{u}_* \cdot \mathbf{n})\mathbf{1} + \begin{pmatrix} 0 & \frac{a_*}{M}\mathbf{n}^T \\ \frac{a_*}{M}\mathbf{n} & 0 \end{pmatrix},$$

$\mathbf{1}$ being the identity matrix in $\mathbb{R}^{4\times 4}$. The eigenvalues of $A(\mathbf{n})$ are given by

$$\lambda_1 = \mathbf{u}_* \cdot \mathbf{n}, \qquad \lambda_2 = \mathbf{u}_* \cdot \mathbf{n}, \qquad \lambda_3 = \mathbf{u}_* \cdot \mathbf{n} - \frac{a_*}{M} \qquad \text{and} \qquad \lambda_4 = \mathbf{u}_* \cdot \mathbf{n} + \frac{a_*}{M}.$$

The associated unit eigenvectors are given by

$$q_{\lambda_1} = \begin{pmatrix} 0 \\ \mathbf{t}_a \end{pmatrix}, \qquad q_{\lambda_2} = \begin{pmatrix} 0 \\ \mathbf{t}_b \end{pmatrix}, \qquad q_{\lambda_3} = \gamma \begin{pmatrix} 1 \\ -\mathbf{n} \end{pmatrix} \qquad \text{and} \qquad q_{\lambda_4} = \gamma \begin{pmatrix} 1 \\ \mathbf{n} \end{pmatrix}$$

where $\gamma^2 = 1/2$ and $(\mathbf{t}_a, \mathbf{t}_b, \mathbf{n})$ defines an orthonormal basis of $\mathbb{R}^3$. Thus, the eigenvector matrix P is given by

$$P = \begin{pmatrix} 0 & 0 & \gamma & \gamma \\ \mathbf{t}_a & \mathbf{t}_b & -\gamma\mathbf{n} & \gamma\mathbf{n} \end{pmatrix}$$

and we find that

$$P^{-1} = \begin{pmatrix} 0 & \mathbf{t}_a^T \\ 0 & \mathbf{t}_b^T \\ \frac{1}{2\gamma} & -\frac{\mathbf{n}^T}{2\gamma} \\ \frac{1}{2\gamma} & \frac{\mathbf{n}^T}{2\gamma} \end{pmatrix}.$$

Thus, we have

$$\Lambda(\mathbf{n}) := P^{-1}A(\mathbf{n})P = \begin{pmatrix} \mathbf{u}_* \cdot \mathbf{n} & 0 & 0 & 0 \\ 0 & \mathbf{u}_* \cdot \mathbf{n} & 0 & 0 \\ 0 & 0 & \mathbf{u}_* \cdot \mathbf{n} - \frac{a_*}{M} & 0 \\ 0 & 0 & 0 & \mathbf{u}_* \cdot \mathbf{n} + \frac{a_*}{M} \end{pmatrix}.$$

Then, by defining $w := P^{-1}q$ which gives

$$w := \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix} = \begin{pmatrix} \mathbf{t}_a \cdot \mathbf{u} \\ \mathbf{t}_b \cdot \mathbf{u} \\ \frac{1}{2\gamma}(r - \mathbf{u} \cdot \mathbf{n}) \\ \frac{1}{2\gamma}(r + \mathbf{u} \cdot \mathbf{n}) \end{pmatrix},$$

System (155) is equivalent to

$$\begin{cases} \partial_t w + \Lambda(\mathbf{n})\partial_\zeta q = 0, \\ \\ w(t=0,\zeta) = \begin{cases} w_L \text{ if } \zeta < 0, \\ w_R \text{ if } \zeta \geq 0. \end{cases} \end{cases} \tag{156}$$

By supposing that

$$\mathbf{u}_* \cdot \mathbf{n} - \frac{a_*}{M} < 0 \qquad \text{and} \qquad \mathbf{u}_* \cdot \mathbf{n} + \frac{a_*}{M} > 0$$

which is in particular satisfied when

$$|\mathbf{u}_*| < \frac{a_*}{M} \qquad \text{(subsonic condition)},$$

the solution $w^{RP} := w(t > 0, \zeta = 0)$ of (156) is given by

$$w^{RP} = \begin{pmatrix} w_{1,L} \text{ if } \mathbf{u}_* \cdot \mathbf{n} > 0 & \text{or} & w_{1,R} \text{ if } \mathbf{u}_* \cdot \mathbf{n} \leq 0 \\ w_{2,L} \text{ if } \mathbf{u}_* \cdot \mathbf{n} > 0 & \text{or} & w_{2,R} \text{ if } \mathbf{u}_* \cdot \mathbf{n} \leq 0 \\ w_{3,R} & & \\ w_{4,L} & & \end{pmatrix}.$$

Thus, since $A(\mathbf{n})q_{RP} = P\Lambda w_{RP}$, we obtain that

$$\begin{aligned} A(\mathbf{n})q_{RP} \quad = \quad & \begin{pmatrix} \gamma\left(\mathbf{u}_* \cdot \mathbf{n} - \dfrac{a_*}{M}\right) w_{3,R} + \gamma\left(\mathbf{u}_* \cdot \mathbf{n} + \dfrac{a_*}{M}\right) w_{4,L} \\ \\ \mathbf{t}_a[w_{1,L}(\mathbf{u}_* \cdot \mathbf{n})^+ + w_{1,R}(\mathbf{u}_* \cdot \mathbf{n})^-] + \mathbf{t}_b[w_{2,L}(\mathbf{u}_* \cdot \mathbf{n})^+ + w_{2,R}(\mathbf{u}_* \cdot \mathbf{n})^-] \end{pmatrix} \\ & + \begin{pmatrix} 0 \\ \\ -\gamma\left(\mathbf{u}_* \cdot \mathbf{n} - \dfrac{a_*}{M}\right) w_{3,R}\mathbf{n} + \gamma\left(\mathbf{u}_* \cdot \mathbf{n} + \dfrac{a_*}{M}\right) w_{4,L}\mathbf{n} \end{pmatrix}. \end{aligned}$$

By noting that

$$\begin{cases} (\mathbf{u}_L \cdot \mathbf{t}_a)\mathbf{t}_a + (\mathbf{u}_L \cdot \mathbf{t}_b)\mathbf{t}_b = \mathbf{u}_L - (\mathbf{u}_L \cdot \mathbf{n})\mathbf{n}, \\ \\ (\mathbf{u}_R \cdot \mathbf{t}_a)\mathbf{t}_a + (\mathbf{u}_R \cdot \mathbf{t}_b)\mathbf{t}_b = \mathbf{u}_R - (\mathbf{u}_R \cdot \mathbf{n})\mathbf{n}, \end{cases}$$

we finally obtain that

$$\begin{aligned} A(\mathbf{n})q_{RP} \quad = \quad & \begin{pmatrix} \dfrac{1}{2}(\mathbf{u}_* \cdot \mathbf{n})[r_L + r_R + (\mathbf{u}_L - \mathbf{u}_R) \cdot \mathbf{n}] \\ \\ \dfrac{1}{2}(\mathbf{u}_* \cdot \mathbf{n})\left[(\mathbf{u}_L + \mathbf{u}_R) \cdot \mathbf{n} + r_L - r_R\right]\mathbf{n} + (\mathbf{u}_* \cdot \mathbf{n})^+\left[\mathbf{u}_L - (\mathbf{u}_L \cdot \mathbf{n})\mathbf{n}\right] + (\mathbf{u}_* \cdot \mathbf{n})^-\left[\mathbf{u}_R - (\mathbf{u}_R \cdot \mathbf{n})\mathbf{n}\right] \end{pmatrix} \\ & + \frac{a_*}{2M}\begin{pmatrix} (\mathbf{u}_L + \mathbf{u}_R) \cdot \mathbf{n} + r_L - r_R \\ \\ [r_L + r_R + (\mathbf{u}_L - \mathbf{u}_R) \cdot \mathbf{n}]\,\mathbf{n} \end{pmatrix} \end{aligned}$$

which is equivalent to

$$A(\mathbf{n})q_{RP} = \frac{1}{2}\begin{pmatrix} (\mathbf{u}_* \cdot \mathbf{n})\,[r_L + r_R + (\mathbf{u}_L - \mathbf{u}_R)\cdot \mathbf{n}] \\ (\mathbf{u}_* \cdot \mathbf{n})\,[(\mathbf{u}_L + \mathbf{u}_R) + (r_L - r_R)\mathbf{n}] - |\mathbf{u}_* \cdot \mathbf{n}|[(\mathbf{u}_L - \mathbf{u}_R)\times \mathbf{n}]\times \mathbf{n} \end{pmatrix}$$

$$+\frac{a_*}{2M}\begin{pmatrix} (\mathbf{u}_L + \mathbf{u}_R)\cdot \mathbf{n} + r_L - r_R \\ [r_L + r_R + (\mathbf{u}_L - \mathbf{u}_R)\cdot \mathbf{n}]\,\mathbf{n} \end{pmatrix}$$

by noting that

$$(\mathbf{u}_* \cdot \mathbf{n})[(\mathbf{u}_L + \mathbf{u}_R)\cdot \mathbf{n}]\mathbf{n} + 2\left\{(\mathbf{u}_* \cdot \mathbf{n})^+[\mathbf{u}_L - (\mathbf{u}_L \cdot \mathbf{n})\mathbf{n}] + (\mathbf{u}_* \cdot \mathbf{n})^-[\mathbf{u}_R - (\mathbf{u}_R \cdot \mathbf{n})\mathbf{n}]\right\}$$

$$= (\mathbf{u}_* \cdot \mathbf{n})(\mathbf{u}_L + \mathbf{u}_R) + |\mathbf{u}_* \cdot \mathbf{n}|\left\{(\mathbf{u}_L - \mathbf{u}_R) - [(\mathbf{u}_L - \mathbf{u}_R)\cdot \mathbf{n}]\mathbf{n}\right\}$$

$$= (\mathbf{u}_* \cdot \mathbf{n})(\mathbf{u}_L + \mathbf{u}_R) - |\mathbf{u}_* \cdot \mathbf{n}|[(\mathbf{u}_L - \mathbf{u}_R)\times \mathbf{n}]\times \mathbf{n}$$

since

$$\mathbf{v} = (\mathbf{v}\cdot \mathbf{n})\mathbf{n} - (\mathbf{v}\times \mathbf{n})\times \mathbf{n} \qquad \text{for any} \qquad \mathbf{v} \in \mathbb{R}^3. \tag{157}$$

Moreover, by integrating (154) on Ω_i and by applying the Gauss law, we obtain

$$\frac{d}{dt}\int_{\Omega_i} q(t,\mathbf{x})d\mathbf{x} + \sum_{\Gamma_{ij}\subset\partial\Omega_i}\int_{\Gamma_{ij}} A(\mathbf{n}_{ij})q\,ds = 0.$$

By supposing that $q(t,\mathbf{x})$ is constant and equal to $q_i(t)$ in Ω_i and by approximating the flux $A(\mathbf{n})q$ with $A(\mathbf{n}_{ij})q_{RP,ij}$ on each edge Γ_{ij}, we obtain the (semi-discrete) Godunov finite volume scheme

$$|\Omega_i|\frac{d}{dt}q_i + \sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\Phi_{ij}^{God} = 0$$

where

$$\Phi_{ij}^{God} := \frac{1}{2}\begin{pmatrix} (\mathbf{u}_* \cdot \mathbf{n}_{ij})\left[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right] \\ (\mathbf{u}_* \cdot \mathbf{n}_{ij})\left[(\mathbf{u}_i + \mathbf{u}_j) + (r_i - r_j)\mathbf{n}_{ij}\right] - |\mathbf{u}_* \cdot \mathbf{n}_{ij}|\left[(\mathbf{u}_i - \mathbf{u}_j)\times \mathbf{n}_{ij}\right]\times \mathbf{n}_{ij} \end{pmatrix}$$
$$+\frac{a_*}{2M}\begin{pmatrix} (\mathbf{u}_i + \mathbf{u}_j)\cdot \mathbf{n}_{ij} + r_i - r_j \\ \left[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right]\mathbf{n}_{ij} \end{pmatrix}. \tag{158}$$

Let us remark that by using (157), we can define $\mathbf{t}_{ij}$ in such a way

$$\left[(\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{t}_{ij}\right]\mathbf{t}_{ij} = (\mathbf{u}_i - \mathbf{u}_j) - \left[(\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right]\mathbf{n}_{ij} = -\left[(\mathbf{u}_i - \mathbf{u}_j)\times \mathbf{n}_{ij}\right]\times \mathbf{n}_{ij} \qquad (|\mathbf{t}_{ij}| = 1 \text{ and } \mathbf{t}_{ij} \perp \mathbf{n}_{ij}). \tag{159}$$

Thus, (158) is equivalent to

$$\Phi_{ij}^{God} := \frac{1}{2}\begin{pmatrix} (\mathbf{u}_* \cdot \mathbf{n}_{ij})\left[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right] \\ (\mathbf{u}_* \cdot \mathbf{n}_{ij})\left[(\mathbf{u}_i + \mathbf{u}_j) + (r_i - r_j)\mathbf{n}_{ij}\right] + |\mathbf{u}_* \cdot \mathbf{n}_{ij}|\left[(\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{t}_{ij}\right]\mathbf{t}_{ij} \end{pmatrix} + \frac{a_*}{2M}\begin{pmatrix} (\mathbf{u}_i + \mathbf{u}_j)\cdot \mathbf{n}_{ij} + r_i - r_j \\ \left[r_i + r_j + (\mathbf{u}_i - \mathbf{u}_j)\cdot \mathbf{n}_{ij}\right]\mathbf{n}_{ij} \end{pmatrix}. \tag{160}$$

The problem with (160) is that in 3D, $\mathbf{t}_{ij}$ depends on $\mathbf{u}_i - \mathbf{u}_j$ which leads to think that (160) is non-linear. As a consequence, we prefer to use (158) in 3D. Nevertheless, $\mathbf{t}_{ij}$ only depends on $\mathbf{n}_{ij}$ in 2D. Thus, we can use (158) as well as (160) in 2D. At last, in 1D, the transverse diffusion does not exist, and we can use (160) with $\mathbf{t}_{ij} := 0$. In this paper, we use (158) except in the proof of Theorem 6.1 where (160) is used.

B. The all Mach Roe scheme in the barotropic and subsonic case

We firstly construct the Roe scheme applied to the barotropic Euler system (2) when the flow is subsonic. Then, we explicite the all Mach version of this scheme deduced from (133) and (134). Finally, we write the dimensionless version of this all Mach Roe scheme used in the asymptotic expansion proposed in Section 7.

B.1. The Roe scheme in the barotropic and subsonic case

Let us apply the finite volume scheme

$$\frac{d}{dt}\int_{\Omega_i} \mathcal{U}(t,\mathbf{x})d\mathbf{x} + \sum_{\Gamma_{ij}\subset\partial\Omega_i}\int_{\Gamma_{ij}} \mathbf{f}(\mathcal{U})\cdot\mathbf{n}ds = 0 \tag{161}$$

to the barotropic Euler system (2) written in 3D. In (161), $\mathcal{U} := (\rho, \rho\mathbf{u})^T$ and the flux $\mathbf{f}(\mathcal{U})$ is the 4×3 matrix

$$\mathbf{f}(\mathcal{U}) = \begin{pmatrix} \rho\mathbf{u}^T \\ \rho\mathbf{u}\otimes\mathbf{u} + p\mathbf{1} \end{pmatrix} =: (\mathbf{f}_x, \mathbf{f}_y, \mathbf{f}_z)$$

($\mathbf{1}$ is the identity matrix in $\mathbb{R}^{3\times 3}$). Thus, the flux in the direction $\mathbf{n}$ is defined by

$$\mathbf{f}(\mathcal{U})\cdot\mathbf{n} = (n_x\mathbf{f}_x + n_y\mathbf{f}_y + n_z\mathbf{f}_z)(\mathcal{U}) = \begin{pmatrix} \rho\mathbf{u}\cdot\mathbf{n} \\ \rho u_x\mathbf{u}\cdot\mathbf{n} + pn_x \\ \rho u_y\mathbf{u}\cdot\mathbf{n} + pn_y \\ \rho u_z\mathbf{u}\cdot\mathbf{n} + pn_z \end{pmatrix} = \begin{pmatrix} \rho\mathbf{u}\cdot\mathbf{n} \\ \rho\,(\mathbf{u}\cdot\mathbf{n})\,\mathbf{u} + p\mathbf{n} \end{pmatrix}.$$

The Roe scheme is an approximation of (161) given by

$$|\Omega_i|\frac{d}{dt}\mathcal{U}_i(t) + \sum_{\Gamma_{ij}\subset\partial\Omega_i} |\Gamma_{ij}|\Phi_{ij}^{\mathrm{Roe}} = 0 \tag{162}$$

where Φ_{ij}^{Roe} is an approximation on the interface Γ_{ij} of $\mathbf{f}(\mathcal{U})\cdot\mathbf{n}$. Since the Roe scheme is an upwind scheme [19], Φ_{ij}^{Roe} is given by

$$\Phi_{ij}^{\mathrm{Roe}} = \frac{\mathbf{f}(\mathcal{U}_i) + \mathbf{f}(\mathcal{U}_j)}{2}\cdot\mathbf{n}_{ij} - \frac{|A_{\mathbf{n}_{ij}}(\mathcal{U}_i,\mathcal{U}_j)|}{2}\cdot(\mathcal{U}_i - \mathcal{U}_j). \tag{163}$$

In (163), $A_{\mathbf{n}_{ij}}(\mathcal{U}_i, \mathcal{U}_j)$ is an approximation on Γ_{ij} of the jacobian matrix $A_{\mathbf{n}_{ij}}(\mathcal{U}) := \dfrac{D\mathbf{f}(\mathcal{U}) \cdot \mathbf{n}_{ij}}{D\mathcal{U}}$. More precisely, $A_{\mathbf{n}_{ij}}(\mathcal{U}_i, \mathcal{U}_j) = A_{\mathbf{n}_{ij}}(\mathcal{U}_{ij})$ where $\mathcal{U}_{ij}$ is an average state on Γ_{ij} which will be defined latter. Moreover, for any $(\mathcal{U}_L, \mathcal{U}_R, \mathbf{n})$, $|A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)| := \sum_{k=1}^{4} |\lambda_k| \mathbf{r}_k \otimes \mathbf{l}_k$ where λ_k are the eigenvalues of $A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)$, associated with the left eigenvectors $\mathbf{l}_k$ and to the right eigenvectors $\mathbf{r}_k$ such that $\mathbf{l}_m \cdot \mathbf{r}_n = \delta_{mn}$ (δ_{mn} is the Kronecker symbol). The jacobian matrix $A_{\mathbf{n}}(\mathcal{U})$ is given by

$$A_{\mathbf{n}}(\mathcal{U}) = \begin{pmatrix} 0 & \mathbf{n}^T \\ a^2\mathbf{n} - (\mathbf{u}\cdot\mathbf{n})\,\mathbf{u} & \mathbf{u}\otimes\mathbf{n} + (\mathbf{u}\cdot\mathbf{n})\,\mathbf{1} \end{pmatrix} =: A.$$

In order to find the eigen elements of A, let λ be an eigenvalue and $\mathbf{q} := (q_1, q_2, q_3, q_4)^T$ an associated eigenvector, and let us define $\mathbf{q_u} := (q_2, q_3, q_4)^T$. Then $A\mathbf{q} = \lambda\mathbf{q}$ is equivalent to

$$\mathbf{q_u}\cdot\mathbf{n} = \lambda q_1, \tag{164}$$

$$\left[a^2\mathbf{n} - (\mathbf{u}\cdot\mathbf{n})\,\mathbf{u}\right] q_1 + (\mathbf{q_u}\cdot\mathbf{n})\,\mathbf{u} + (\mathbf{u}\cdot\mathbf{n})\,\mathbf{q_u} = \lambda\mathbf{q_u}. \tag{165}$$

Taking the dot product of (165) with $\mathbf{n}$ and replacing $\mathbf{q_u}\cdot\mathbf{n}$ by its value from (164), we obtain

$$\left[a^2 - (\mathbf{u}\cdot\mathbf{n})^2 - \lambda^2\right] q_1 + 2\lambda q_1 \mathbf{u}\cdot\mathbf{n} = 0,$$

which implies that either $q_1 = 0$ or λ solves

$$\lambda^2 - 2\lambda\mathbf{u}\cdot\mathbf{n} - a^2 + (\mathbf{u}\cdot\mathbf{n})^2 = 0,$$

the solutions of which are $\lambda_1 = \mathbf{u}\cdot\mathbf{n} - a$ and $\lambda_4 = \mathbf{u}\cdot\mathbf{n} + a$. When $q_1 = 0$, then (164) implies that $\mathbf{q_u}\cdot\mathbf{n} = 0$ and then, (165) implies that $\lambda = \mathbf{u}\cdot\mathbf{n}$ is an eigenvalue with multiplicity two. Let $\mathbf{t}_a$ and $\mathbf{t}_b$ be such that $(\mathbf{n}, \mathbf{t}_a, \mathbf{t}_b)$ is an orthonormal basis of $\mathbb{R}^3$. Then, the condition $\mathbf{q_u}\cdot\mathbf{n} = 0$ implies that two eigenvectors associated with $\lambda_{2,3} = 0$ are $\mathbf{r}_2 = (0, \mathbf{t}_a)^T$ and $\mathbf{r}_3 = (0, \mathbf{t}_b)^T$. When $\lambda = \lambda_1$, then replacing $\mathbf{q_u}\cdot\mathbf{n}$ by $(\mathbf{u}\cdot\mathbf{n} - a)\,q_1$ in (165) yields $\left(a^2\mathbf{n} - a\mathbf{u}\right) q_1 + a\mathbf{q_u} = \mathbf{0}$, and thus an associated eigenvector is $\mathbf{r}_1 = (1, \mathbf{u} - a\mathbf{n})^T$. In the same way, we find that an eigenvector associated with λ_4 is $\mathbf{r}_4 = (1, \mathbf{u} + a\mathbf{n})^T$. Let P be the matrix of right eigenvectors

$$P = \begin{pmatrix} 1 & 0 & 0 & 1 \\ \mathbf{u} - a\mathbf{n} & \mathbf{t}_a & \mathbf{t}_b & \mathbf{u} + a\mathbf{n} \end{pmatrix}.$$

We have

$$P^{-1} = \begin{pmatrix} \dfrac{1}{2} + \dfrac{\mathbf{u}\cdot\mathbf{n}}{2a} & -\dfrac{\mathbf{n}^T}{2a} \\ -\mathbf{t}_a\cdot\mathbf{u} & \mathbf{t}_a^T \\ -\mathbf{t}_b\cdot\mathbf{u} & \mathbf{t}_b^T \\ \dfrac{1}{2} - \dfrac{\mathbf{u}\cdot\mathbf{n}}{2a} & \dfrac{\mathbf{n}^T}{2a} \end{pmatrix}.$$

Let us define $\Lambda = P^{-1}AP$, the diagonal matrix of the eigenvalues of A, and denote by $|\Lambda| := \mathrm{diag}(|\lambda_1|, |\lambda_2|, |\lambda_3|, |\lambda_4|)$. Thus, we have $|A| = P|\Lambda|P^{-1}$. In the subsonic case, it holds that $|\lambda_1| = a - \mathbf{u}\cdot\mathbf{n}$ and $|\lambda_4| = a + \mathbf{u}\cdot\mathbf{n}$, so that

$$|\Lambda|P^{-1} = \begin{pmatrix} \frac{a}{2} - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{2a} & -\frac{(a-\mathbf{u}\cdot\mathbf{n})}{2a}\mathbf{n}^T \\ -|\mathbf{u}\cdot\mathbf{n}|\mathbf{t}_a\cdot\mathbf{u} & |\mathbf{u}\cdot\mathbf{n}|\mathbf{t}_a^T \\ -|\mathbf{u}\cdot\mathbf{n}|\mathbf{t}_b\cdot\mathbf{u} & |\mathbf{u}\cdot\mathbf{n}|\mathbf{t}_b^T \\ \frac{a}{2} - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{2a} & \frac{(a+\mathbf{u}\cdot\mathbf{n})}{2a}\mathbf{n}^T \end{pmatrix}.$$

The first line of $|A|$ is easy to compute and is equal to $\left(a - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{a}, \frac{\mathbf{u}\cdot\mathbf{n}}{a}\mathbf{n}^T\right)$. The lower left 3×1 block of $|A|$ is equal to

$$\left[\frac{a}{2} - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{2a}\right](\mathbf{u} - a\mathbf{n}) - |\mathbf{u}\cdot\mathbf{n}|\,(\mathbf{t}_a\cdot\mathbf{u}\,\mathbf{t}_a + \mathbf{t}_b\cdot\mathbf{u}\,\mathbf{t}_b) + \left[\frac{a}{2} - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{2a}\right](\mathbf{u} + a\mathbf{n}) = \left[a - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{a}\right]\mathbf{u} + |\mathbf{u}\cdot\mathbf{n}|\,(\mathbf{u}\times\mathbf{n})\times\mathbf{n}$$

since every vector $\mathbf{v}\in\mathbb{R}^3$ may be written $\mathbf{v} = (\mathbf{v}\cdot\mathbf{n})\mathbf{n} + (\mathbf{v}\cdot\mathbf{t}_a)\mathbf{t}_a + (\mathbf{v}\cdot\mathbf{t}_b)\mathbf{t}_b = (\mathbf{v}\cdot\mathbf{n})\mathbf{n} - (\mathbf{v}\times\mathbf{n})\times\mathbf{n}$. Finally, the lower right 3×3 block of $|A|$ is equal to

$$-\frac{(a-\mathbf{u}\cdot\mathbf{n})}{2a}(\mathbf{u} - a\mathbf{n})\otimes\mathbf{n} + |\mathbf{u}\cdot\mathbf{n}|\,(\mathbf{t}_a\otimes\mathbf{t}_a + \mathbf{t}_b\otimes\mathbf{t}_b) + \frac{(a+\mathbf{u}\cdot\mathbf{n})}{2a}(\mathbf{u} + a\mathbf{n})\otimes\mathbf{n} = \frac{\mathbf{u}\cdot\mathbf{n}}{a}\mathbf{u}\otimes\mathbf{n} + a\mathbf{n}\otimes\mathbf{n} + |\mathbf{u}\cdot\mathbf{n}|(\mathbf{1} - \mathbf{n}\otimes\mathbf{n})$$

since it holds that $\mathbf{n}\otimes\mathbf{n} + \mathbf{t}_a\otimes\mathbf{t}_a + \mathbf{t}_b\otimes\mathbf{t}_b = \mathbf{1}$. The final expression of $|A_{\mathbf{n}}(\mathcal{U})| = |A|$ is thus

$$|A_{\mathbf{n}}(\mathcal{U})| = \begin{pmatrix} a - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{a} & \frac{\mathbf{u}\cdot\mathbf{n}}{a}\mathbf{n}^T \\ \left[a - \frac{(\mathbf{u}\cdot\mathbf{n})^2}{a}\right]\mathbf{u} + |\mathbf{u}\cdot\mathbf{n}|\,(\mathbf{u}\times\mathbf{n})\times\mathbf{n} & \frac{\mathbf{u}\cdot\mathbf{n}}{a}\mathbf{u}\otimes\mathbf{n} + (a - |\mathbf{u}\cdot\mathbf{n}|)\,\mathbf{n}\otimes\mathbf{n} + |\mathbf{u}\cdot\mathbf{n}|\mathbf{1} \end{pmatrix}.$$

The matrix $|A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)|$ in (163) is defined by

$$|A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)| = |A_{\mathbf{n}}(\mathcal{U}_{LR})|$$

where $\mathcal{U}_{LR}$ is computed with the Roe average state [19]

$$\begin{cases} \rho_{LR} = \sqrt{\rho_L \rho_R}, \\ u_{x,LR} = \dfrac{\sqrt{\rho_L} u_{x,L} + \sqrt{\rho_R} u_{x,R}}{\sqrt{\rho_L} + \sqrt{\rho_R}}, \\ u_{y,LR} = \dfrac{\sqrt{\rho_L} u_{y,L} + \sqrt{\rho_R} u_{y,R}}{\sqrt{\rho_L} + \sqrt{\rho_R}}, \\ u_{z,LR} = \dfrac{\sqrt{\rho_L} u_{z,L} + \sqrt{\rho_R} u_{z,R}}{\sqrt{\rho_L} + \sqrt{\rho_R}}, \\ a_{LR}^2 = \dfrac{\Delta p}{\Delta \rho} \end{cases}$$

with the notation $\Delta(\cdot) = (\cdot)_R - (\cdot)_L$. Now, we have to compute $|A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)| \cdot (\Delta\rho, \Delta(\rho\mathbf{u}))^T$. One of the features of the Roe scheme is that the mean states $(\rho_{LR}, \mathbf{u}_{LR})$ satisfy the relation

$$\Delta(\rho\mathbf{u}) = \rho_{LR}\Delta\mathbf{u} + (\Delta\rho)\mathbf{u}_{LR}.$$

Therefore, the first element of $|A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)| \cdot (\Delta\rho, \Delta(\rho\mathbf{u}))^T$ is equal to

$$\left(a_{LR} - \frac{(\mathbf{u}_{LR}\cdot\mathbf{n})^2}{a_{LR}}\right)\Delta\rho + \frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\Delta(\rho\mathbf{u})\cdot\mathbf{n} = a_{LR}\Delta\rho + \rho_{LR}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\Delta\mathbf{u}\cdot\mathbf{n} \tag{166}$$

while the last three elements of $|A_{\mathbf{n}}(\mathcal{U}_L, \mathcal{U}_R)| \cdot (\Delta\rho, \Delta(\rho\mathbf{u}))^T$ are equal to

$$\Delta\rho\left\{\left[a_{LR} - \frac{(\mathbf{u}_{LR}\cdot\mathbf{n})^2}{a_{LR}}\right]\mathbf{u}_{LR} + |\mathbf{u}_{LR}\cdot\mathbf{n}|\,(\mathbf{u}_{LR}\times\mathbf{n})\times\mathbf{n}\right\} \quad +$$
$$\Delta(\rho\mathbf{u})\cdot\mathbf{n}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\mathbf{u}_{LR} + \Delta(\rho\mathbf{u})\cdot\mathbf{n}\,(a_{LR} - |\mathbf{u}_{LR}\cdot\mathbf{n}|)\,\mathbf{n} + |\mathbf{u}_{LR}\cdot\mathbf{n}|\Delta(\rho\mathbf{u}). \tag{167}$$

Now, we use the following equalities

$$-\frac{(\mathbf{u}_{LR}\cdot\mathbf{n})^2}{a_{LR}}\Delta\rho\,\mathbf{u}_{LR} + \Delta(\rho\mathbf{u})\cdot\mathbf{n}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\mathbf{u}_{LR} = \rho_{LR}\Delta\mathbf{u}\cdot\mathbf{n}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\mathbf{u}_{LR} \tag{168}$$

on the one hand, and

$$\begin{aligned} \Delta\rho|\mathbf{u}_{LR}\cdot\mathbf{n}|\,(\mathbf{u}_{LR}\times\mathbf{n})\times\mathbf{n} + |\mathbf{u}_{LR}\cdot\mathbf{n}|\Delta(\rho\mathbf{u}) &= \Delta\rho|\mathbf{u}_{LR}\cdot\mathbf{n}|\,[(\mathbf{u}_{LR}\times\mathbf{n})\times\mathbf{n} + \mathbf{u}_{LR}] + \rho_{LR}|\mathbf{u}_{LR}\cdot\mathbf{n}|\Delta\mathbf{u} \\ &= \Delta\rho|\mathbf{u}_{LR}\cdot\mathbf{n}|\,(\mathbf{u}_{LR}\cdot\mathbf{n})\,\mathbf{n} + \rho_{LR}|\mathbf{u}_{LR}\cdot\mathbf{n}|\Delta\mathbf{u} \end{aligned} \tag{169}$$

and

$$\Delta(\rho\mathbf{u})\cdot\mathbf{n}\,(a_{LR} - |\mathbf{u}_{LR}\cdot\mathbf{n}|)\,\mathbf{n} = (\mathbf{u}_{LR}\cdot\mathbf{n}\Delta\rho + \rho_{LR}\Delta\mathbf{u}\cdot\mathbf{n})\,(a_{LR} - |\mathbf{u}_{LR}\cdot\mathbf{n}|)\,\mathbf{n} \tag{170}$$

on the other hand. From (168)–(170), we obtain that (167) is equal to

$$\begin{aligned} a_{LR}\Delta\rho\,\mathbf{u}_{LR} + \rho_{LR}\Delta\mathbf{u}\cdot\mathbf{n}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\mathbf{u}_{LR} + \rho_{LR}|\mathbf{u}_{LR}\cdot\mathbf{n}|\Delta\mathbf{u} + a_{LR}\mathbf{u}_{LR}\cdot\mathbf{n}\Delta\rho\,\mathbf{n} \\ -\rho_{LR}|\mathbf{u}_{LR}\cdot\mathbf{n}|\,(\Delta\mathbf{u}\cdot\mathbf{n})\,\mathbf{n} + a_{LR}\rho_{LR}\,(\Delta\mathbf{u}\cdot\mathbf{n})\,\mathbf{n} \\ = \left(a_{LR}\Delta\rho + \rho_{LR}\Delta\mathbf{u}\cdot\mathbf{n}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\right)\mathbf{u}_{LR} + a_{LR}\,(\mathbf{u}_{LR}\cdot\mathbf{n}\Delta\rho + \rho_{LR}\Delta\mathbf{u}\cdot\mathbf{n})\,\mathbf{n} \\ -\rho_{LR}|\mathbf{u}_{LR}\cdot\mathbf{n}|\,(\Delta\mathbf{u}\times\mathbf{n})\times\mathbf{n}. \end{aligned}$$

Gathering (166) and (171), we obtain that

$$
\begin{aligned}
|A_{\mathbf{n}}(\mathcal{U}_L,\mathcal{U}_R)|\cdot(\mathcal{U}_R-\mathcal{U}_L) \quad = \quad & a_{LR}\Delta\rho\begin{pmatrix}1\\ \mathbf{u}_{LR}\end{pmatrix}+a_{LR}(\mathbf{u}_{LR}\cdot\mathbf{n})\Delta\rho\begin{pmatrix}0\\ \mathbf{n}\end{pmatrix}+\rho_{LR}\frac{\mathbf{u}_{LR}\cdot\mathbf{n}}{a_{LR}}\Delta\mathbf{u}\cdot\mathbf{n}\begin{pmatrix}1\\ \mathbf{u}_{LR}\end{pmatrix}\\
& +\rho_{LR}a_{LR}\Delta\mathbf{u}\cdot\mathbf{n}\begin{pmatrix}0\\ \mathbf{n}\end{pmatrix}-\rho_{LR}|\mathbf{u}_{LR}\cdot\mathbf{n}|\begin{pmatrix}0\\ (\Delta\mathbf{u}\times\mathbf{n})\times\mathbf{n}\end{pmatrix}.
\end{aligned}
\tag{171}
$$

Thus, by using (162), (163) and (171)) we obtain

$$
\left\{
\begin{aligned}
&\frac{d}{dt}\rho_i+\frac{1}{2|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left\{(\rho_i\mathbf{u}_i+\rho_j\mathbf{u}_j)\cdot\mathbf{n}_{ij}+\frac{\rho_{ij}}{a_{ij}}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}+a_{ij}(\rho_i-\rho_j)\right\}=0, && \text{(a)}\\
&\frac{d}{dt}(\rho_i\mathbf{u}_i)+\frac{1}{2|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\Big\{\rho_i(\mathbf{u}_i\cdot\mathbf{n}_{ij})\mathbf{u}_i+\rho_j(\mathbf{u}_j\cdot\mathbf{n}_{ij})\mathbf{u}_j+a_{ij}(\rho_i-\rho_j)\Big[\mathbf{u}_{ij}+(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})\mathbf{n}_{ij}\Big]\\
&\qquad -\rho_{ij}|\mathbf{u}_{ij}\cdot\mathbf{n}_{ij}|\Big[(\mathbf{u}_i-\mathbf{u}_j)\times\mathbf{n}_{ij}\Big]\times\mathbf{n}_{ij}+\frac{\rho_{ij}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})}{a_{ij}}\Big[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}\Big]\mathbf{u}_{ij} && \text{(b)}\\
&\qquad +\Big[p_i+p_j+\rho_{ij}a_{ij}(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}\Big]\mathbf{n}_{ij}\Big\}=0
\end{aligned}
\right.
\tag{172}
$$

with $p_k=p(\rho_k)$ and $a_{ij}=\dfrac{p_i-p_j}{\rho_i-\rho_j}$.

B.2. The all Mach Roe scheme in the barotropic and subsonic case

We deduce from (133), (134) and (172) that the *all Mach Roe scheme* in the barotropic and subsonic case is given by

$$
\left\{
\begin{aligned}
&\frac{d}{dt}\rho_i+\frac{1}{2|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left\{(\rho_i\mathbf{u}_i+\rho_j\mathbf{u}_j)\cdot\mathbf{n}_{ij}+\frac{\rho_{ij}}{a_{ij}}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}+a_{ij}(\rho_i-\rho_j)\right\}=0, && \text{(a)}\\
&\frac{d}{dt}(\rho_i\mathbf{u}_i)+\frac{1}{2|\Omega_i|}\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\Big\{\rho_i(\mathbf{u}_i\cdot\mathbf{n}_{ij})\mathbf{u}_i+\rho_j(\mathbf{u}_j\cdot\mathbf{n}_{ij})\mathbf{u}_j+a_{ij}(\rho_i-\rho_j)\Big[\mathbf{u}_{ij}+(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})\mathbf{n}_{ij}\Big]\\
&\qquad -\rho_{ij}|\mathbf{u}_{ij}\cdot\mathbf{n}_{ij}|\Big[(\mathbf{u}_i-\mathbf{u}_j)\times\mathbf{n}_{ij}\Big]\times\mathbf{n}_{ij}+\frac{\rho_{ij}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})}{a_{ij}}\Big[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}\Big]\mathbf{u}_{ij} && \text{(b)}\\
&\qquad +\Big[p_i+p_j+\theta_{ij}\rho_{ij}a_{ij}(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}\Big]\mathbf{n}_{ij}\Big\}=0
\end{aligned}
\right.
\tag{173}
$$

with $\theta_{ij}=\theta(M_{ij}):=\min(M_{ij},1)$ and $M_{ij}=\dfrac{|\mathbf{u}_{ij}|}{a_{ij}}$. The difference between (172) and (173) is only in the last term of the left hand side of (172)(b) and (173)(b).

B.3. Dimensionless version of the all Mach Roe scheme in the barotropic and subsonic case

The dimensionless version of (173) is obtained by replacing in (173) p_i, p_j and a_{ij} respectively by p_i/M^2, p_j/M^2 and a_{ij}/M where M is an order of the local Mach number M_{ij}. This gives

$$\left\{\begin{array}{l} \dfrac{d}{dt}\rho_i + \dfrac{1}{2|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left\{(\rho_i\mathbf{u}_i+\rho_j\mathbf{u}_j)\cdot\mathbf{n}_{ij} + M\frac{\rho_{ij}}{a_{ij}}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij} + \frac{a_{ij}}{M}(\rho_i-\rho_j)\right\} = 0, \\ \dfrac{d}{dt}(\rho_i\mathbf{u}_i) + \dfrac{1}{2|\Omega_i|}\displaystyle\sum_{\Gamma_{ij}\subset\partial\Omega_i}|\Gamma_{ij}|\left\{\rho_i(\mathbf{u}_i\cdot\mathbf{n}_{ij})\mathbf{u}_i + \rho_j(\mathbf{u}_j\cdot\mathbf{n}_{ij})\mathbf{u}_j + \frac{a_{ij}}{M}(\rho_i-\rho_j)\Big[\mathbf{u}_{ij}+(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})\mathbf{n}_{ij}\Big]\right. \\ \qquad\qquad -\rho_{ij}|\mathbf{u}_{ij}\cdot\mathbf{n}_{ij}|\Big[(\mathbf{u}_i-\mathbf{u}_j)\times\mathbf{n}_{ij}\Big]\times\mathbf{n}_{ij} + M\dfrac{\rho_{ij}(\mathbf{u}_{ij}\cdot\mathbf{n}_{ij})}{a_{ij}}[(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}]\mathbf{u}_{ij} \\ \qquad\qquad \left. + \left[\dfrac{1}{M^2}(p_i+p_j) + \dfrac{\theta_{ij}}{M}\rho_{ij}a_{ij}(\mathbf{u}_i-\mathbf{u}_j)\cdot\mathbf{n}_{ij}\right]\mathbf{n}_{ij}\right\} = 0 \end{array}\right. \tag{174}$$

with $\theta_{ij} = \theta(M_{ij}) := \min(M_{ij}, 1)$ and $M_{ij} = M\dfrac{|\mathbf{u}_{ij}|}{a_{ij}}$.

References

[1] S. Dellacherie, On a low Mach nuclear core model, ESAIM:PROC 35 (2012) 79–106. 1

[2] G. F. G. B. G. B. M. Bernard M., S. Dellacherie, Y. Penel, Study of a low Mach nuclear core model for two-phase flows with phase transition I: stiffened gaz law., Math. Model. and Num. Anal. To appear. 1

[3] P. L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, J. Comp. Phys. 43 (1981) 357–372. 2, 13, 24, 25, 33, 41, 42

[4] T. Buffard, T. Gallouët, J.-M. Hérard, A sequel to a rough Godunov scheme: application to real gases, Computers and Fluids 29 (2000) 813–847. 2, 13, 24

[5] H. Guillard, C. Viozat, On the behavior of upwind schemes in the low Mach number limit, Computers and Fluids 28 (1999) 63–86. 2, 33

[6] H. Guillard, A. Murrone, On the behavior of upwind schemes in the low Mach number limit: II. Godunov type schemes, Computers and Fluids 33 (4) (2004) 655–675. 2

[7] H. Guillard, A. Murrone, Behavior of upwind scheme in the low Mach number limit: III. Preconditioned dissipation for a five equation two phase model, Computers and Fluids 37 (10) (2008) 1209–1224. 2

[8] S. Dellacherie, Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number, J. Comp. Phys. 229 (4) (2010) 978–1016. 2, 3, 4, 5, 6, 7, 8, 10, 13, 16, 18, 41

[9] S. Dellacherie, P. Omnes, F. Rieper, The influence of cell geometry on the Godunov scheme applied to the linear wave equation, J. Comp. Phys. 229 (14) (2010) 5315–5338. 3, 4, 5, 6, 12, 13, 19

[10] S. Dellacherie, J. Jung, P. Omnes, A low Mach correction for the Godunov scheme applied to the linear wave equation with porosity In preparation. 4

[11] F. Rieper, Influence of Cell Geometry on the Behaviour of the First-Order Roe Scheme in the Low Mach Number Regime, in: R. Eymard, J.-M. Hérard (Eds.), Finite Volumes for Complex Applications V, Wiley, 625–632, 2008. 4

[12] F. Rieper, G. Bader, The influence of cell geometry on the accuracy of upwind schemes in the low Mach number regime, J. Comp. Phys. 228 (8) (2009) 2918–2933. 4, 33

[13] F. Fillion, A. Chanoine, S. Dellacherie, A. Kumbaro, FLICA-OVAP: a new platform for core thermal-hydraulic studies, Proceedings of the 13th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-13) . 4

[14] F. Rieper, A low Mach number fix for Roe's approximate Riemann solver, J. Comp. Phys. To appear. 4, 33

[15] S. Schochet, Fast Singular Limits of Hyperbolic PDEs, Journal of Differential Equation 114 (1994) 476–512. 8

[16] A. Harten, High Resolution Schemes for Hyperbolic Conservation Laws, J. Comp. Phys. 135 (1997) 260–278. 38

[17] V. Rusanov, Calculation of intersection of non-steady shock waves with obstacles, Comput. Math. Phys. USSR 1 (1961) 267–279. 41

[18] P. Lax, Weak Solutions of Nonlinear Hyperbolic Equations and Their Numerical Computation, Comm. Pure. Appl. Math. VII (1954) 159–193. 42

[19] E. Godlewski, P.-A. Raviart, in: Numerical Approximation of Hyperbolic Systems of Conservation Laws, vol. 118 of *Applied Mathematical Sciences*, Springer-Verlag, New York, 215–220, 1996. 50, 53