



**HAL**  
open science

# A Mach-Sensitive Splitting Approach for Euler-like Systems

David Iampietro, Frédéric Daude, Pascal Galon, Jean-Marc Hérard

► **To cite this version:**

David Iampietro, Frédéric Daude, Pascal Galon, Jean-Marc Hérard. A Mach-Sensitive Splitting Approach for Euler-like Systems. ESAIM: Mathematical Modelling and Numerical Analysis, 2018, 52 (1), pp.207-253. 10.1051/m2an/2017063 . hal-01466827v2

**HAL Id: hal-01466827**

**<https://hal.science/hal-01466827v2>**

Submitted on 20 Dec 2017

**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.

# A Mach-Sensitive Splitting Approach for Euler-like Systems

D. Iampietro <sup>\*1,3,4</sup>, F. Daude <sup>†1,3</sup>, P. Galon <sup>‡3,5</sup>, and J-M Hérard <sup>§2,4</sup>

<sup>1</sup>*EDF lab Saclay, 7 boulevard Gaspard Monge 92120 Palaiseau, France*

<sup>2</sup>*EDF lab Chatou, 6 Quai Watier 78400 Chatou, France*

<sup>3</sup>*IMSIA, UMR EDF/CNRS/CEA/ENSTA 9219 Universit Paris Saclay 828 Boulevard des Marchaux 91762 Palaiseau Cedex, France*

<sup>4</sup>*I2M, UMR CNRS 7373 Technople Château-Gombert 39, rue F. Joliot Curie 13453 Marseille Cedex 13, France*

<sup>5</sup>*CEA DEN/DANS/DM2S/SEMT/DYN, D36 91190 Saclay, France*

December 20, 2017

Herein, a Mach-sensitive fractional step approach is proposed for Euler-like systems. The key idea is to introduce a time-dependent splitting which dynamically decouples convection from acoustic phenomenon following the fluctuations of the flow Mach number. By doing so, one seeks to maintain the accuracy of the computed solution for all Mach number regimes. Indeed, when the Mach number takes high values, a time-explicit resolution of the overall Euler-like system is entirely performed in one of the present splitting step. On the contrary, in the low-Mach number case, convection is totally separated from the acoustic waves production. Then, by performing an appropriate correction on the acoustic step of the splitting, the numerical diffusion can be significantly reduced. A study made on both convective and acoustic subsystems of the present approach has revealed some key properties as hyperbolicity and positivity of the density and internal energy in the case of an ideal gas thermodynamics. The one-dimensional results made on a wide range of Mach numbers using an ideal and a stiffened gas thermodynamics show that the present approach is as accurate and CPU-consuming as a state of the art Lagrange-Projection-type method.

## Introduction

Condensation Induced Water Hammer (CIWH) is a very specific two-phase fast transient phenomenon. It starts by the smooth deformation of a slow material interface between hot vapor and cooler liquid

---

\*[david.iampietro@edf.fr](mailto:david.iampietro@edf.fr); Corresponding author

†[frederic.daude@edf.fr](mailto:frederic.daude@edf.fr)

‡[pascal.galon@cea.fr](mailto:pascal.galon@cea.fr)

§[jean-marc.herard@edf.fr](mailto:jean-marc.herard@edf.fr)

water. Then, as time goes on, shear instabilities and steep temperature gradients entail the trapping and then the sudden condensation of vapor pockets leading to the production of strong shock waves in the liquid phase.

In the above description, two time-scales can be identified: a material scale linked to the slow interface deformation and a fast acoustic scale related to the propagation of the shock front. As the Euler compressible system produces similar multi-scale waves, it constitutes the framework of the present paper. Indeed, the Mach number of the flow  $M = |\mathbf{u}|/c$ , with  $\mathbf{u}$  the velocity field and  $c$  the fluid speed of sound, measures the gap between the material and the acoustic time-scales. The particularity of a CIWH results from the fact that, initially, since the deformation of the slow material wave is the leading process, the fluid behaves as a low Mach number compressible flow:  $0 < M \ll 1$ . Then, because of the stiffness of the liquid water thermodynamics, strong shock waves are triggered *even if*  $0 < M \ll 1$ . In order to enforce this fact, one can find in [39] a water-hammer experiment in liquid water for which  $|\mathbf{u}| \approx 0.2 \text{ m.s}^{-1}$ ,  $c_{\text{liquid}} \approx 1.5 \times 10^3 \text{ m.s}^{-1}$  leading to pressure jumps of several bars. Besides, the analytical solution of a symmetric double shock Riemann problem obtained with the Euler compressible system endowed with a stiffened gas thermodynamics is derived later on in this work. It allows to better understand this peculiar aspect.

Thus, the long-term objective of the present work is to set out a method involving compressible Godunov-like solvers in order to fulfill a two-fold aim:

- (I) Accurately follow the slow material waves when  $0 < M \ll 1$   
in the absence of fast transient phenomena.
- (II) Then, accurately follow the fast and strong acoustic waves even when  $0 < M \ll 1$ .

In the sequel, the different issues inherent to point (I) and point (II) are described. However, the present study is *entirely* dedicated to point (II) and we leave point (I) for future works.

Godunov-like schemes perform poorly in the context of point (I). A first issue is related to the first order time-discretization of such methods when  $0 < M \ll 1$ . Indeed, let us introduce  $\Delta x$  the cell size of a 1D mesh,  $\Delta t$  the time-step built using a time-explicit Godunov-like scheme,  $\lambda_{u+c}$  (respectively  $\lambda_u$ ) the non-dimensional acoustic eigenvalue (respectively the slow convective eigenvalue). Besides, define  $\mathcal{C}_{|u|+c} = \frac{1}{M} \frac{\Delta t \lambda_{u+c}}{\Delta x}$  (respectively  $\mathcal{C}_{|u|} = \frac{\Delta t \lambda_u}{\Delta x}$ ) the acoustic Courant number (respectively the convective Courant number). The stability of the fast acoustic dynamics requires  $\mathcal{C}_{|u|+c} = 1^-$  and then  $\mathcal{C}_{|u|} \approx M \mathcal{C}_{|u|+c} \approx M \ll 1$ . Hence, the slow material wave dynamics are largely depreciated because  $\mathcal{C}_{|u|}$ , associated with a first order time-discretization, is too small to compensate the numerical diffusion induced by the first-order space-discretization. In [31], the evolution of the stiffness of the profile of an isolated slow contact-discontinuity as function of  $\mathcal{C}_{|u|+c}$  supports this point. A solution, described in [17, 29] for the Euler isentropic system and in [34, 21] for the whole Euler compressible system endowed with an ideal gas thermodynamics, consists in applying a time-implicit discretization on the stiff pressure gradient. By doing so, one can replace the constraining time-explicit acoustic CFL condition  $\mathcal{C}_{|u|+c} = 1^-$  by a convective CFL one:  $\mathcal{C}_{|u|} = 1^-$ , the latter being adapted to accurately follow slow material waves. What is more, in the above references, additional time-implicit discretizations hold on the density convection term [17, 29] or on the total energy convection term [21]. The resulting implicit-explicit schemes are thus consistent, when  $M \rightarrow 0$  and  $\Delta t, \Delta x$  fixed, with a discrete incompressible solver. Schemes that are stable under the CFL condition  $\mathcal{C}_{|u|} = 1^-$  and have the above consistent behavior are said to be Asymptotic Preserving (AP) towards the Euler incompressible system.

A second issue, in the framework of point (I) is related to the spatial discretization of Godunov-like schemes. Indeed, in [28, 27] the authors show, using a 2D discrete asymptotic decomposition w.r.t  $M$  on a cartesian mesh, that Godunov-like schemes are not able to maintain an initial incompressible solution in the incompressible phase-space from one time-step to an other. In [18, 20, 19], the authors transpose the Schochet theory [36, 37] at the discrete level and highlight the same difficulty. Moreover, they point out that this issue only concerns 2D (respectively 3D) non-triangular (respectively non-tetrahedral) meshes for which the non-dimensional mesh size is bigger than  $M$ . Then, they concentrate on means to control the computed solution proximity towards the incompressible well-prepared space (see [19] for a definition) for non-dimensional time-scales lower than  $M$ . Without going into details, the key idea is that the discrete acoustic operator of Godunov-like schemes contains a non-centered diffusive part scaling as  $O(\Delta x/M)$  notably in the discrete momentum equation. When applied to an initial well-prepared solution, the latter produces new acoustic pressure waves scaling as  $O(M)$  instead of  $O(M^2)$ . In [22, 35] a low Mach number correction rescaling the momentum diffusive part as  $O(\Delta x)$  is proposed. It has since been re-used in [19, 11, 10].

The background of point (II) is more simple since we "only" want to be accurate on fast acoustic waves neglecting the slow material wave diffusion induced by the time-discretization as well as the proximity towards the incompressible phase space. However, as mentioned in the above paragraph, Godunov-like schemes suffer from a numerical diffusion of order  $O(\Delta x/M)$  due to the spatial discretization which can be observed even in 1D under the CFL condition  $\mathcal{C}_{|u|+c} = (1/2)^-$ .

In the light of the time-scale difficulties encountered in point (I), it seems relevant to introduce a fractional step method allowing to solve the slow convection process and the fast acoustic waves production separately. It is based on a Weighted Splitting Approach (WSA). The "Weighted" notion stems from the fact that the proposed splitting relies on a decomposition of the pressure balanced by a time-dependent splitting parameter. Inspired from [1, 8], a first part of the pressure decomposition, seen as a "small" pressure fluctuation if  $0 < M \ll 1$ , is sent to a convective subsystem  $\mathcal{C}$ . The remaining part goes into an acoustic subsystem  $\mathcal{A}$ . In the context of point (I), a time-implicit resolution of  $\mathcal{A}$  combined with a time-explicit resolution of  $\mathcal{C}$  moves the proposed approach towards the convective CFL condition  $\mathcal{C}_{u_0} = 1^-$ . The stability proof of the implicit-explicit version of the present method is examined a more recent companion paper [31]. As already mentioned, this paper focuses on the point (II). Thus, a time-explicit Godunov-like scheme is applied for both subsystems  $\mathcal{C}$  and  $\mathcal{A}$ . The spatial discretization comes from the relaxation schemes theory [33, 12, 4, 2, 14] and notably from the Suliciu-like relaxation solvers [42, 4]. It provides a simple approximate Riemann solver whose writing is independent of the equation of state. Besides, in order to reduce the numerical diffusion when  $0 < M \ll 1$ , an anti-diffusive term, directly taken from [35, 19, 11], but simply seen as a tool here, is added to the present acoustic pressure flux. In the sequel, the anti-diffusive term is referred to as  $\theta$ -correction.

Let us end the present WSA description by saying that, in the more common case where the production of shock waves is associated with a local rise of the Mach number  $M \approx 1$ , the present splitting is canceled and the overall Euler system is retrieved in the  $\mathcal{C}$  subsystem. Thus fast acoustic waves are accurately captured.

The paper is structured as follows: in section one, the dynamic splitting is firstly described at the continuous level. A study of each resulting conservative subsystem is done through hyperbolicity and positivity analyses involving ideal and stiffened gas equations of state. Section two deals with the

approximate Riemann solvers derived for the subsystems spatial discretization. Discrete positivity properties in the case of ideal gas thermodynamics are also derived under a non-restrictive condition. Following the steps of [11], section three is devoted to different truncation error analyses. The dependence in terms of Mach number as well as the impact of the  $\theta$ -correction on the numerical diffusive operator of the overall scheme is shown. Eventually, section four presents one-dimensional time-explicit results obtained for ideal and stiffened gas thermodynamics and for a wide panel of Mach numbers. It turns out that the presented method is as accurate and efficient as a Lagrange-Projection method presented in [11]. However in the case of a stiffened gas thermodynamics with  $0 < M \ll 1$ , the proposed method, although  $L^\infty$  stable, produces more oscillations than the Lagrange-Projection approach.

## 1 Convective and Acoustic Effects in Euler-like Systems

### 1.1 Homogeneous Equilibrium Model Equations

When the non-equilibrium effects are small, one way to model two-phase flows is to assume that the two phases have the same velocity, pressure, temperature and chemical potential. The conservation laws are then similar to the Euler compressible system. Define  $\mathbf{U} = [\rho, \rho \mathbf{u}, \rho e]^T$  the conservative variables vector with  $\rho$  the fluid density,  $\mathbf{u}$  its velocity vector and  $e$  its total specific energy. The mass, momentum and energy conservation then read:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (1a)$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + p \mathbf{I}) = \mathbf{0}, \quad (1b)$$

$$\partial_t (\rho e) + \nabla \cdot ((\rho e + p) \mathbf{u}) = 0, \quad (1c)$$

$$e = \frac{|\mathbf{u}|^2}{2} + \varepsilon, \quad \varepsilon = \varepsilon^{EOS}(\rho, p). \quad (1d)$$

Equality in (1d) is the equation of state for a single phase fluid. It relates  $\varepsilon$  the specific internal energy with density and pressure. Its strong level of nonlinearity is known to produce rarefaction or shock waves inside the flow. Recall that the Euler system is strictly hyperbolic, its eigenvalues being in one-dimension:  $\lambda_1^E = u - c < \lambda_2^E = u < \lambda_3^E = u + c$ ; with  $c$  the sound speed which strongly depends on the equation of state and can be defined as:

$$\rho c^2 = (\partial_p \varepsilon|_\rho)^{-1} \left( \frac{p}{\rho} - \rho \partial_\rho \varepsilon|_p \right). \quad (2)$$

What is more  $\lambda_1^E$  and  $\lambda_3^E$  are related to genuinely non-linear fields whereas  $\lambda_2^E$  is associated with a linearly degenerate one.

Eventually, let us write the second law of thermodynamics principle, introducing the specific entropy variable  $s = s^{EOS}(\rho, \varepsilon)$  related with  $\rho$  and  $\varepsilon$  by the differential equation:

$$d\varepsilon = T ds - p d\left(\frac{1}{\rho}\right), \quad (3)$$

$$\text{with: } T = T^{EOS}(\rho, s) = \partial_s \varepsilon|_\rho, \quad p = p^{EOS}(\rho, s) = \rho^2 \partial_\rho \varepsilon|_s.$$

Using equation (3), it can be easily verified that, for smooth EOS,  $s$  is also solution of the PDE:

$$\partial_\rho s|_p + c^2 \partial_p s|_\rho = 0. \quad (4)$$

Such a physical entropy is used to characterise admissible weak solutions of Euler system (1). Indeed, as proved in [26], the mapping  $(\rho, \rho \mathbf{u}, \rho e) \rightarrow -\rho s$  is a strictly convex function and  $(-\rho s, -\rho \mathbf{u} s)$  constitutes a mathematical entropy pair. Thus, any admissible weak solution of the Euler system should verify the inequality:

$$\partial_t (\rho s) + \nabla \cdot (\rho s \mathbf{u}) \geq 0. \quad (5)$$

Beyond conservativity and maximum principle, inequality (5) is a key theoretical property that one would like to obtain, at the discrete level, in a numerical scheme.

Let us end this subsection by defining the one-dimensional Riemann problem associated to system (1). Let  $\mathbf{U}_L$  and  $\mathbf{U}_R$  be two constant states of the one-dimensional Euler system (1). It reads:

$$\begin{aligned} \partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) &= \mathbf{0}, \\ \mathbf{U}(\cdot, t=0) &= \begin{cases} \mathbf{U}_L, & \text{if } x < 0 \\ \mathbf{U}_R, & \text{if } x > 0, \end{cases} \quad \text{with: } \mathbf{F}(\mathbf{U}) = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ (\rho e + p) u \end{bmatrix}. \end{aligned} \quad (6)$$

As proved in [40], in the case of ideal gas or in [26] under more general thermodynamical hypothesis, Riemann problem (6) admits a unique entropic solution made of contact waves, rarefaction waves and shock waves as long as  $\mathbf{U}_L$  and  $\mathbf{U}_R$  are close enough.

## 1.2 A Weighted Splitting Approach

As mentioned in the introduction in the context of point (I), two different physics are at stake inside Euler-like systems. In 1D, the first convects conservative variables using velocity  $u$ , the second contains pressure effects responsible for shock and rarefaction waves propagating at speed  $u \pm c$ . Thus, in the case of low-Mach compressible flows,  $|u| \ll c$  and the acoustic physics goes much faster than the convective one. Therefore, time-explicit schemes restricted by the acoustic CFL condition tend to diffuse material waves as time goes on.

Looking at the Euler compressible system (1), a first step to elaborate a cure consists in decoupling the convective from the acoustic physics and proceed to their resolution separately and successively. This can be done by splitting the conservation laws system into two new continuous subsystems:

$$\mathcal{C} : \begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + \mathcal{E}_0^2(t) p \mathbf{I}) = \mathbf{0}, \\ \partial_t (\rho e) + \nabla \cdot ((\rho e) + \mathcal{E}_0^2(t) p \mathbf{u}) = 0, \end{cases} \quad \mathcal{A} : \begin{cases} \partial_t \rho = 0, \\ \partial_t (\rho \mathbf{u}) + \nabla \cdot ((1 - \mathcal{E}_0^2(t)) p \mathbf{I}) = \mathbf{0}, \\ \partial_t (\rho e) + \nabla \cdot ((1 - \mathcal{E}_0^2(t)) p \mathbf{u}) = 0. \end{cases} \quad (8)$$

Here,  $\mathcal{E}_0(\cdot)$  is a time-dependent weighting factor belonging to interval  $]0, 1]$  and proportional to the maximal Mach number of the flow:

$$\begin{aligned} \mathcal{E}_0(t) &\propto \max(\mathcal{E}_{inf}, \min(M_{max}(t), 1)), \\ M_{max}(t) &= \sup_{x \in \Omega} \left( M(x, t) = \frac{|\mathbf{u}(x, t)|}{c(x, t)} \right), \end{aligned} \quad (9)$$

$\Omega$  being the computational domain. It has to be mentioned that in [1], Baraille and co-authors introduce the same kind of splitting for the Euler barotropic system with  $\mathcal{E}_0$  equal to zero. Since their resulting  $\mathcal{C}$ -subsystem is not hyperbolic, they re-introduce a similar pressure perturbation, with  $\mathcal{E}_0$  seen as a fixed parameter. They solve the Riemann problem associated with the perturbed subsystem and make  $\mathcal{E}_0$  tend towards zero in the obtained solution in order to derive their numerical scheme.

In the present splitting,  $\mathcal{E}_0(t)$  is strictly positive because of the lower bound  $\mathcal{E}_{inf}$  defined such that  $0 < \mathcal{E}_{inf} \ll 1$ . By doing so, the above loss of hyperbolicity on the jacobian of  $\mathcal{C}$  is avoided. Besides, here  $\mathcal{E}_0(t)$  has a physical interest since it measures the gap between the convective and the acoustic time-scales.

Indeed, assume that at instant  $t$  the flow is such that  $M_{max}(t)$  is close or superior to 1. Then,  $\mathcal{E}_0(t)$  will be close to 1, the subsystem  $\mathcal{C}$  formally converges towards the full Euler system while  $\mathcal{A}$  is a degenerated stationary subsystem. Hence, if  $\mathcal{C}$  is solved using a time-explicit Godunov-like scheme, Euler shocks related to a temporal rise of the Mach number would be correctly captured. On the contrary, In the case of a globally low-Mach number flow,  $M_{max}(t) \approx \mathcal{E}_0(t) \ll 1$ , and pressure terms completely disappear from  $\mathcal{C}$  which turns out to be a pure "convective" subsystem. Pressure terms are treated afterwards in  $\mathcal{A}$  which becomes an "acoustic" subsystem. Particularly, a time-implicit scheme ([15, 21]) applied on it would remove the most constraining part of the CFL condition. Let us end the splitting description by saying that both subsystems  $\mathcal{C}$  and  $\mathcal{A}$  are conservative, and that their formal summation allows to recover the original Euler system (1).

In the sequel,  $\mathcal{C}$  is referred as the convective subsystem and  $\mathcal{A}$  the acoustic one. Before going further into the numerical resolution of  $\mathcal{C}$  and  $\mathcal{A}$ , one has to study their basic mathematical properties: hyperbolicity and maximum principle. This is done in the next section.

### 1.2.1 Hyperbolicity of $\mathcal{C}$ and $\mathcal{A}$

Above all, the hyperbolicity of the two subsystems  $\mathcal{C}$  and  $\mathcal{A}$  is investigated. This ensures that solutions of  $\mathcal{C}$  and  $\mathcal{A}$  do not suffer from definition issues by producing waves with celerities evolving in the  $\mathbb{C}$  space. This is the object of the following proposition written in one space dimension but easily extendable to the multi-dimensional case:

**Proposition 1** (Hyperbolicity of convective and acoustic subsystems). *Let us introduce  $c_{\mathcal{C}}(\rho, p)$  and  $c_{\mathcal{A}}(\rho, p)$  two modified sound speeds such that:*

$$\begin{aligned} (\rho c_{\mathcal{C}}(\rho, p))^2 &= (\partial_p \varepsilon_{|\rho})^{-1} (\mathcal{E}_0^2 p - \rho^2 \partial_\rho \varepsilon_{|p}), \\ (\rho c_{\mathcal{A}}(\rho, p))^2 &= (\partial_p \varepsilon_{|\rho})^{-1} p. \end{aligned} \quad (10)$$

*In the case of a stiffened gas thermodynamics,  $c_{\mathcal{C}}^2 \geq 0$ . Besides, if pressure remains positive,  $c_{\mathcal{A}}^2 \geq 0$ . Under this condition, the subsystems  $\mathcal{C}$  and  $\mathcal{A}$  are hyperbolic. Their eigenvalues are:*

$$\begin{aligned} \lambda_1^{\mathcal{C}} = u - \mathcal{E}_0 c_{\mathcal{C}} &\leq \lambda_2^{\mathcal{C}} = u \leq \lambda_3^{\mathcal{C}} = u + \mathcal{E}_0 c_{\mathcal{C}}, \\ \lambda_1^{\mathcal{A}} = - (1 - \mathcal{E}_0^2) c_{\mathcal{A}} &\leq \lambda_2^{\mathcal{A}} = 0 \leq \lambda_3^{\mathcal{A}} = (1 - \mathcal{E}_0^2) c_{\mathcal{A}}, \end{aligned} \quad (11)$$

*the 1-wave and 3-wave of both subsystems are associated to genuinely non-linear fields whereas the 2-wave field are linearly degenerate. What is more,  $c_{\mathcal{C}}$ ,  $c_{\mathcal{A}}$  and  $c$  are related by:*

$$(c_{\mathcal{C}})^2 + (1 - \mathcal{E}_0^2) (c_{\mathcal{A}})^2 = c^2. \quad (12)$$

The proof of this proposition is written in Appendix B. Beside, using relation (12), it can be observed that, when  $\mathcal{E}_0$  is close to one,  $\mathcal{C}$  is approximately equivalent to the Euler system, and that is why:  $\forall k, \lim_{\mathcal{E}_0 \rightarrow 1} \lambda_k^{\mathcal{C}} = \lambda_k^E$ . Moreover, when  $\mathcal{E}_0$  tends towards zero,  $\lim_{\mathcal{E}_0 \rightarrow 0} \lambda_k^{\mathcal{C}} = \lambda_2^E = u$ , because of the pressure terms disappearance.  $\mathcal{C}$  then degenerates into a pure convective subsystem already exhibited in [1, 8]. However, thanks to equality (12), we have  $\forall k \in \{1, 3\} : |\lambda_k^{\mathcal{A}}| \leq c$  even when  $\mathcal{E}_0$  goes to zero. Thus, the weighted splitting approach always tends to underestimate acoustic wave speeds whatever the thermodynamics is. In **Section 2**, the transcription, at the discrete level, of these non physical wave speeds will be seen. A numerical way to bridge the gap between  $c_{\mathcal{A}}$  and  $c$  so that to follow the real physics will also be proposed. In order to make them less abstract, the expressions of  $c_{\mathcal{C}}$  and  $c_{\mathcal{A}}$  are provided below, in the case of ideal gas:

$$\begin{aligned} c_{\mathcal{C}} &= \sqrt{\frac{\gamma \mathcal{E}_0 P}{\rho}} < c, \quad \gamma_{\mathcal{E}_0} = \mathcal{E}_0^2 (\gamma - 1) + 1 < \gamma, \\ c_{\mathcal{A}} &= \sqrt{\frac{(\gamma - 1) p}{\rho}} < c = \sqrt{\frac{\gamma p}{\rho}}. \end{aligned} \tag{13}$$

In the following, positivity of the relevant quantities got from the thermodynamical phase-space is analyzed in both continuous subsystems  $\mathcal{C}$  and  $\mathcal{A}$ .

### 1.2.2 Positivity of Density and Internal Energy

Positivity requirements reflect the invariance of a given solution towards its thermodynamical phase space. In this study, one focuses on the ideal (IG) and the stiffened gas (SG) thermodynamics defined by the following sets:

$$\rho \varepsilon = \frac{p}{(\gamma - 1)}, \tag{14a}$$

$$\Phi_{PG} = \left\{ \mathbf{U}, \text{ s. t. } e = \frac{|\mathbf{u}|^2}{2} + \varepsilon, \rho > 0, \rho \varepsilon > 0 \right\}, \tag{14b}$$

$$= \left\{ \mathbf{U}, \text{ s. t. } e = \frac{|\mathbf{u}|^2}{2} + \varepsilon, \rho > 0, p > 0 \right\}, \tag{14c}$$

$$\rho \varepsilon = \frac{p + \gamma P_{\infty}}{(\gamma - 1)}, \quad P_{\infty} > 0, \tag{15a}$$

$$\Phi_{SG} = \left\{ \mathbf{U}, \text{ s. t. } e = \frac{|\mathbf{u}|^2}{2} + \varepsilon, \rho > 0, \rho \varepsilon - P_{\infty} > 0 \right\}, \tag{15b}$$

$$= \left\{ \mathbf{U}, \text{ s. t. } e = \frac{|\mathbf{u}|^2}{2} + \varepsilon, \rho > 0, p + P_{\infty} > 0 \right\}. \tag{15c}$$



For both thermodynamics, the construction of the above phase-spaces allows to guarantee that the Euler speed of sound  $c$  is real and strictly positive since:

$$\begin{aligned} (\Phi)_{IG} : c &= \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma(\gamma-1)(\rho\varepsilon)}{\rho}}, \\ (\Phi)_{SG} : c &= \sqrt{\frac{\gamma(p+P_\infty)}{\rho}} = \sqrt{\frac{\gamma(\gamma-1)(\rho\varepsilon - P_\infty)}{\rho}}. \end{aligned} \quad (16)$$

Thus, in the case of an ideal gas thermodynamics (respectively in the case of a stiffened gas thermodynamics), the positivity of  $\rho$  and  $\rho\varepsilon$  (respectively the positivity of  $\rho$  and  $\rho\varepsilon - P_\infty$ ) is studied. First of all, assume that  $\forall \phi \in \{\rho, \rho\varepsilon, \rho\varepsilon - P_\infty\}$  a positive inlet boundary condition as well as an admissible initial condition hold, namely:

$$\begin{aligned} \phi|_{\partial\Omega} &\geq 0 \text{ if } \mathbf{u} \cdot \mathbf{n}|_{\partial\Omega} \leq 0, \\ \phi(., t=0) &\geq 0, \end{aligned} \quad (17)$$

with  $\Omega$  the spatial domain of boundary  $\partial\Omega$  and  $\mathbf{n}$  the outward local normal vector of  $\partial\Omega$ .

Then, as recalled in [24] for sufficiently smooth solutions, positivity of density  $\rho$  is naturally obtained from mass equation in subsystem  $\mathcal{C}$  for both thermodynamics. Density is also stationary in subsystem  $\mathcal{A}$ . So, after having successively solved  $\mathcal{C}$  and  $\mathcal{A}$ , density remains positive.

In the case of an ideal gas thermodynamics, since  $\rho$  remains positive, the positivity of  $\rho\varepsilon$  is equivalent to the positivity of  $\varepsilon$ . For smooth solutions, the specific internal energy in subsystems  $\mathcal{C}$  and  $\mathcal{A}$  verifies:

$$\begin{cases} \partial_t \varepsilon + \mathbf{u} \cdot \nabla \varepsilon + \mathcal{E}_0^2(t) \frac{p}{\rho} \nabla \cdot \mathbf{u} = 0, & (\mathcal{C}) \\ \partial_t \varepsilon + (1 - \mathcal{E}_0^2(t)) \frac{p}{\rho} \nabla \cdot \mathbf{u} = 0. & (\mathcal{A}) \end{cases} \quad (18)$$

By making the same kind of regularity hypothesis than in [24], one can prove that, in the case of an ideal gas thermodynamics,  $\varepsilon$  remains positive on  $\Omega$  throughout time. See Appendix C for more details.

In the case of a stiffened gas thermodynamics, let us introduce  $P = \rho\varepsilon - P_\infty = \frac{p+P_\infty}{(\gamma-1)}$  which is the variable concerned by the positivity requirement. In  $\mathcal{C}$  and  $\mathcal{A}$ , it verifies the PDEs:

$$\begin{cases} \partial_t P + \nabla \cdot (P \mathbf{u}) + \mathcal{E}_0^2(t) (\gamma-1) P \nabla \cdot (\mathbf{u}) + (1 - \mathcal{E}_0^2(t)) P_\infty \nabla \cdot \mathbf{u} = 0, & (\mathcal{C}) \\ \partial_t P + (1 - \mathcal{E}_0^2(t)) (\gamma-1) P \nabla \cdot (\mathbf{u}) - (1 - \mathcal{E}_0^2(t)) P_\infty \nabla \cdot \mathbf{u} = 0. & (\mathcal{A}) \end{cases} \quad (19)$$

Let us first notice that  $\frac{P_\infty}{P} = (\gamma-1) \frac{P_\infty}{p+P_\infty}$  is not *a priori* bounded since the stiffened gas phase-space allows  $p$  to tend towards  $-P_\infty$ . Then, as shown in equation (103) of Appendix C, this prevents from controlling the operator  $(1 - \mathcal{E}_0^2(t)) P_\infty \nabla \cdot \mathbf{u}$ , and the positivity of  $P$  cannot *a priori* be ensured unless  $P_\infty = 0$  which is the ideal gas case or  $\mathcal{E}_0 = 1$  which corresponds to the resolution of Euler system with no splitting. More details are given in Appendix C. Thus, in the case of a stiffened gas thermodynamics, computations involving the discrete resolution of subsystems  $\mathcal{C}$  and  $\mathcal{A}$  in which the discrete pressure field is close to  $-P_\infty$  could potentially produce complex numbers for  $c$ . Nevertheless, in most cases, pressure remains positive and this difficulty can be avoided.

The next section is dedicated to the design of a time-explicit scheme to solve the above Mach-sensitive fractional step.

## 2 Relaxation Scheme Applied to the Weighted Splitting Approach

For the sake of simplicity and with no loss of generality, the scheme description is done in one dimension. Literature dealing with relaxation schemes is vast. Without being exhaustive, we refer to [33] for the derivation of relaxation schemes applied to abstract hyperbolic systems in which the whole flux is relaxed. In [12], the authors question the existence of solutions for the relaxation systems as well as their convergence towards a local equilibrium. A detailed study of the entropy-satisfying relaxation method applied to the isentropic gas dynamics system and extended to the fully compressible Euler system is given in [6]. It uses a Suliciu-like relaxation technique [42] which is also applied in [2] on a Drift-Flux model. Besides, the acoustic part of the Lagrange-Projection splitting derived in [25, 11] is solved the same way too. Eventually in [14], an extension of the Suliciu approach to general fluid systems is done. Following the same approach, we proceed to a Suliciu-like relaxation method on both subsystems  $\mathcal{C}$  and  $\mathcal{A}$ .

Let us recall that the Suliciu relaxation method applied on Euler-like systems consists in introducing a new pressure variable  $\Pi$  endowed with a "quasi-linear" dynamics converging towards the real pressure variable  $p$ . This convergence is ensured thanks to a source term whose timescale  $\mu \ll 1$ . The new system is still hyperbolic and has only linearly degenerate fields which makes the derivation of an exact Godunov solver easier. What is more, the high level of nonlinearity brought by the pressure variable *via* the equation of state (1d) is encapsulated in one single constant. As a consequence, the derivation of the numerical scheme can be done independently of the thermodynamics law. The cost to be paid is the increase of the system dimension through an additional equation for  $\Pi$ . What is more, one has to decide how to treat the equilibrium between  $\Pi$  and  $p$  numerically.

### 2.1 Suliciu Relaxation for the Weighted Splitting Approach

As previously mentioned, the Suliciu-like relaxation technique leans on a linearization of the pressure dynamics. Let us first derive the PDEs associated with  $p$  for both subsystems  $\mathcal{C}$  and  $\mathcal{A}$ . Using the mass equations, internal energy equations (18), and the fact that  $\forall D \in \{\partial_t, \partial_x\}$ ,  $Dp = (\partial_\rho p)|_\varepsilon D\rho + (\partial_\varepsilon p)|_\rho D\varepsilon$ , it yields:

$$\mathcal{C} : \partial_t p + u \partial_x p + \rho (c_{\mathcal{C}})^2 \partial_x u = 0, \quad (20) \quad \mathcal{A} : \partial_t p + (1 - \mathcal{E}_0^2) \rho (c_{\mathcal{A}})^2 \partial_x u = 0. \quad (21)$$

Then, replace pressure  $p(\rho, \varepsilon)$  by a new relaxation pressure variable  $\Pi$  which no longer depends of density and internal energy. One also expects  $\Pi$  to mimic the above physical pressure dynamics but with an additional linearization effect on the thermodynamics. This is done by introducing two constants  $a_{\mathcal{C}} > 0$  and  $a_{\mathcal{A}} > 0$  such that  $\Pi$  verifies:

$$\mathcal{C} : \partial_t \Pi + u \partial_x \Pi + \frac{a_{\mathcal{C}}^2}{\rho} \partial_x u = \frac{(p - \Pi)}{\mu}, \quad (22) \quad \mathcal{A} : \partial_t \Pi + (1 - \mathcal{E}_0^2) \frac{a_{\mathcal{A}}^2}{\rho} \partial_x u = \frac{(p - \Pi)}{\mu}. \quad (23)$$

Here,  $a_{\mathcal{C}}$  (respectively  $a_{\mathcal{A}}$ ) is homogeneous to a density times a velocity and encapsulates the non-linear effects brought by  $\rho c_{\mathcal{C}}(\rho, \varepsilon)$  (respectively  $\rho c_{\mathcal{A}}(\rho, \varepsilon)$ ). Besides, by using the mass equation and because  $a_{\mathcal{C}}$  and  $a_{\mathcal{A}}$  are constant, it is possible to rewrite equations (22) and (23) in a conservative way namely:

$$\mathcal{C} : \partial_t (\rho \Pi) + \partial_x ((\rho \Pi + a_{\mathcal{C}}^2) u) = \frac{\rho (p - \Pi)}{\mu}, \quad \mathcal{A} : \partial_t (\rho \Pi) + \partial_x ((1 - \mathcal{E}_0^2) a_{\mathcal{A}}^2 u) = \frac{\rho (p - \Pi)}{\mu}. \quad (25)$$

One can observe that, when  $\mu \rightarrow 0$  in (22) and (23), the relaxation pressure  $\Pi$  tends formally towards  $p$  at zeroth order in  $\mu$ . Hence  $(p - \Pi) / \mu$  can be formally interpreted as a correction term of time scale  $\mu$  forcing the relaxation pressure to converge towards the physical pressure instantaneously if  $\mu$  tends to zero.

Finally, the relaxation convective and acoustic systems read:

$$\mathcal{C}^\mu : \begin{cases} \partial_t \rho + \partial_x (\rho u) = 0, \\ \partial_t (\rho u) + \partial_x (\rho u^2 + \mathcal{E}_0^2(t) \Pi) = 0, \\ \partial_t (\rho e) + \partial_x ((\rho e + \mathcal{E}_0^2(t) \Pi) u) = 0, \\ \partial_t (\rho \Pi) + \partial_x ((\rho \Pi + a_{\mathcal{C}}^2) u) = \frac{\rho (p - \Pi)}{\mu}, \end{cases} \quad \mathcal{A}^\mu : \begin{cases} \partial_t \rho = 0, \\ \partial_t (\rho u) + \partial_x ((1 - \mathcal{E}_0^2(t)) \Pi) = 0, \\ \partial_t (\rho e) + \partial_x ((1 - \mathcal{E}_0^2(t)) \Pi u) = 0, \\ \partial_t (\rho \Pi) + \partial_x ((1 - \mathcal{E}_0^2(t)) a_{\mathcal{A}}^2 u) = \frac{\rho (p - \Pi)}{\mu}. \end{cases} \quad (26) \quad (27)$$

**Remark 1.** *It is worth noting that, the relaxation schemes framework can be formulated differently. Indeed, one can rewrite the subsystems  $\mathcal{C}^\mu$  and  $\mathcal{A}^\mu$  without the relaxation term  $\frac{\rho(p-\Pi)}{\mu}$ . Hence, the relaxation pressure  $\Pi$  is free to move away from the real pressure  $p$ . However in that case, the resolution of the homogeneous relaxation subsystems is completed by a projection sub-step on the equilibrium manifold:*

$$\left\{ (\rho, \rho u, \rho e, \rho \Pi) \text{ such that, } \Pi = p(\rho, \varepsilon), \text{ with } \varepsilon = e - \frac{\rho u^2}{2} \right\}. \quad (28)$$

As it will be seen in the sequel (see equation (47) dealing with the overall algorithm of the fractional step), in practice both subsystems  $\mathcal{C}^\mu$  and  $\mathcal{A}^\mu$  are solved without  $\frac{\rho(p-\Pi)}{\mu}$ . Then, at the end of each subsystem resolution, the updated relaxation state is projected instantaneously on the equilibrium manifold.

One way to calibrate the constant relaxation coefficient  $a_{\mathcal{C}}$  (respectively  $a_{\mathcal{A}}$ ) is to perform a Chapman-Enskog expansion by rewriting all the variables  $\phi \in \{\rho, u, e, \Pi\}$  in power of  $\mu$ :

$$\begin{aligned} \phi &= \phi_0 + \mu \phi_1 + \mu^2 \phi_2 + \dots, \\ \Pi_0 &= p. \end{aligned}$$

By doing so, one can exhibit a subcharacteristic-like condition, also called Whitham-like condition [44]. It allows to prevent  $\mathcal{C}^\mu$  (respectively  $\mathcal{A}^\mu$ ) from triggering instabilities when  $\mu \rightarrow 0$ . What is more, it can be used as a sufficient condition to build an entropy pair and an extended entropy inequality for the relaxation system (see [12, 5, 6]). As detailed in Appendix F, the subcharacteristic conditions obtained are:

$$\mathcal{C}^\mu : a_{\mathcal{C}} > \rho c_{\mathcal{C}}, \quad (29a)$$

$$\mathcal{A}^\mu : a_{\mathcal{A}} > \rho c_{\mathcal{A}}. \quad (29b)$$

**Remark 2.** By proceeding in the same manner, one could have obtained a Suliciu-like relaxation Euler system. The relaxation pressure PDE would have been:

$$\partial_t \Pi + u \partial_x \Pi + \frac{a_E^2}{\rho} \partial_x u = \frac{(p - \Pi)}{\mu}, \quad (30)$$

with  $a_E$  the constant relaxation coefficient constrained by the Euler subcharacteristic condition:

$$a_E > \rho c. \quad (31)$$

Recall that  $\lim_{\mathcal{E}_0 \rightarrow 1} c_C = c$ , and then (29a) becomes (31) in that case. More generally, the shape of such a Suliciu-like relaxation Euler system can be obtained by formally making  $\mathcal{E}_0$  tend towards one in  $\mathcal{C}^\mu$ .

**Remark 3.** In [13], [6], [9], [15] and [14], in order to solve the Euler system using relaxation techniques, the authors perform an inversion between the role played by total energy and entropy. The idea is to turn the total energy equation into a mathematical entropy constraint while injecting the pure transport entropy equation:

$$\partial_t s + u \partial_x s = 0. \quad (32)$$

By doing so, one can lean on good properties brought by relaxation methods applied on the barotropic Euler system and enforce the entropy inequality (5) in the numerical resolution of the full Euler system. More details can be found in the above references. In our splitting approach, such a strategy is avoided. Indeed, let us consider  $\mathbf{U}^C$  (respectively  $\mathbf{U}^A$ ) the conservative state solution of the subsystem  $C$  (respectively  $A$ ). It can be shown that the physical entropy function  $s(\mathbf{U}^C)$  (respectively  $s(\mathbf{U}^A)$ ) defined in equation (4) does no longer verify equation (32). For both subsystems, an additional non-conservative term appears and prevents from applying directly the barotropic-relaxation system results. That is why, in our case, a simple Suliciu-relaxation method is performed on the conservative system including total energy. Note that a similar relaxation treatment is done in [11] for the acoustic subsystem.

**Remark 4.** As previously noted in **Subsection 1.2.1**, the lower bound in the acoustic subcharacteristic condition (29b) uses  $c_A$  an artificial celerity naturally provided by subsystem  $A$ . In the case of an ideal or a stiffened gas thermodynamics,  $c_A < c$ , so that one can provide, at the discrete level, a constant  $a_A$  fulfilling a discrete version of inequality (29b) while violating the natural acoustic subcharacteristic condition (31) based on the real sound speed which is found in [9], [15] and [25]. According to a formal Chapman-Enskog expansion, the subcharacteristic condition (29b) provides a sufficient condition guaranteeing the stability of the time-explicit scheme for the resolution of  $\mathcal{A}^\mu$ . However, no theoretical result has been found to prove that it was also a sufficient condition to obtain the stability of the overall weighted splitting approach. In particular, in the case of low Mach number compressible flows with  $c_A \ll c$ , we think that it is relevant to numerically compare the effect of considering the more demanding condition (31) rather than (29b). This will be done in **Subsection 4.2**.

## 2.2 Derivation of the Relaxation Scheme

Let us define  $\Delta x$  (respectively  $\Delta t$ ) the space step (respectively the time step) of the scheme. For  $i \in [1, \dots, N_{\text{cells}}]$  let us set  $x_i = i \Delta x$ , the coordinate of the cell center  $i$  and  $x_{i+1/2} = x_i + \Delta x/2$ , the

coordinate of face  $i+1/2$ . Let us consider  $\mathbf{W} = [\rho, \rho u, \rho e, \rho \Pi]^T$  the extended relaxation conservative vector. Following **Remark 1**, the Riemann problem related to the *homogeneous* versions of (26) or (27) writes:

$$\partial_t \mathbf{W} + \partial_x \mathbf{F}^\mu(\mathbf{W}) = \mathbf{0}, \quad \mathbf{W}(t=0, \cdot) = \begin{cases} \mathbf{W}_L & \text{if } x < x_0, \\ \mathbf{W}_R & \text{if } x > x_0, \end{cases} \quad (33)$$

$$\text{with } \mathbf{F}^\mu \in \left\{ \mathbf{F}_C^\mu(\mathbf{W}) = \begin{bmatrix} \rho u \\ \rho u^2 + \mathcal{E}_0^2 \Pi \\ (\rho e + \mathcal{E}_0^2 \Pi) u \\ (\rho \Pi + (a_C)^2) u \end{bmatrix}, \mathbf{F}_A^\mu(\mathbf{W}) = (1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \Pi \\ \Pi u \\ (a_A)^2 u \end{bmatrix} \right\}. \text{ Let us introduce } \mathbf{U}_i^n \text{ the}$$

discrete approximation of  $\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} L(\mathbf{W})(x, t^n) dx$ ,  $\mathbf{W}$  verifying  $\partial_t \mathbf{W} + \partial_x \mathbf{F}^\mu(\mathbf{W}) = \mathbf{0} \forall (x, t)$ , and  $L : \mathbf{W} = [w_1, w_2, w_3, w_4]^T \in \mathbb{R}^4 \rightarrow [w_1, w_2, w_3]$ . Therefore  $\mathbf{U}_i^n$  represents the discrete approximation of the solution of the relaxation system without the component  $\rho \Pi$ . Then, the Godunov solver can be derived easily and reads:

$$\begin{aligned} \mathbf{U}_i^{n+1} &= \mathbf{U}_i^n - \frac{\Delta t}{\Delta x} \left( \mathbf{H}_{i+1/2}^n - \mathbf{H}_{i-1/2}^n \right), \\ \text{with: } \mathbf{H}_{i+1/2}^n &= L \left( \mathbf{F}^\mu \left( \mathbf{W}^{God} \left( 0; \mathbf{W}_i^n, \mathbf{W}_{i+1}^n \right) \right) \right) = \mathbf{H}_{i+1/2}^n(\mathbf{U}_i^n, \mathbf{U}_{i+1}^n), \end{aligned} \quad (34)$$

and  $(x, t) \rightarrow \mathbf{W}^{God} \left( \frac{x-x_{i+1/2}}{t}; \mathbf{W}_i^n, \mathbf{W}_{i+1}^n \right)$  the self similar solution of the Riemann problem (33) located at  $x_0 = x_{i+1/2}$ .

The study of  $\mathcal{C}^\mu$  and  $\mathcal{A}^\mu$  leading to the time-explicit expression of the Godunov flux has been done using the non-conservative variable  $\mathbf{Z}^T = [\rho, u, \Pi, e]^T$ . In the following, the structure of the fields, the eigenvalues and the Riemann invariants are described.

### 2.2.1 Convective Part

The relaxation system  $\mathcal{C}^\mu$  is strictly hyperbolic, its eigenvalues being:  $\lambda_1^{\mathcal{C}, \mu} = u - \mathcal{E}_0 a_C \tau$ ,  $\lambda_2^{\mathcal{C}, \mu} = \lambda_3^{\mathcal{C}, \mu} = u$ ,  $\lambda_4^{\mathcal{C}, \mu} = u + \mathcal{E}_0 a_C \tau$  with  $\tau = 1/\rho$ . Furthermore, each field is linearly degenerate and admits simple Riemann invariants:

$$\begin{aligned} \mathcal{I}_{\mathcal{E}_0, 1}^{\mathcal{C}, \mu} &= \left\{ u - \mathcal{E}_0 a_C \tau, \Pi + a_C^2 \tau, e + \frac{\mathcal{E}_0}{a_C} \Pi u \right\}, \\ \mathcal{I}_{\mathcal{E}_0, 2, 3}^{\mathcal{C}, \mu} &= \{u, \Pi\}, \\ \mathcal{I}_{\mathcal{E}_0, 4}^{\mathcal{C}, \mu} &= \left\{ u + \mathcal{E}_0 a_C \tau, \Pi + a_C^2 \tau, e - \frac{\mathcal{E}_0}{a_C} \Pi u \right\}. \end{aligned} \quad (35)$$

Let us notice that, for smooth solutions, mass equation in subsystem  $\mathcal{C}^\mu$  can be rewritten  $\partial_t \tau + u \partial_x \tau - \tau \partial_x u = 0$ . By multiplying this equation by  $a_C^2$  and summing it with the  $\Pi$  equation in (27), one obtains:

$$\partial_t (\Pi + a_C^2 \tau) + u \partial_x (\Pi + a_C^2 \tau) = 0. \quad (36)$$

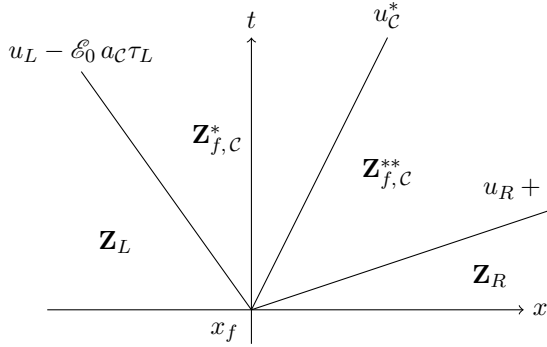
Thus,  $\Pi + a_C^2 \tau$  remains constant along the characteristic curves of speed  $u$ . Besides, it is a 2,3-strong Riemann invariant meaning that it is invariant through the 1-wave and the 4-wave. Eventually one

can notice that this quantity is solution of the following equation:

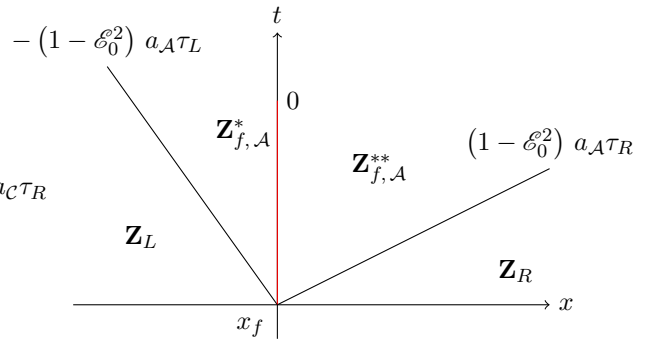
$$\partial_\rho (\psi)|_\Pi + \left(\frac{a_C}{\rho}\right)^2 \partial_\Pi (\psi)|_\rho = 0, \quad (37)$$

which can be related to the entropy equation (4). The pressure term linearization induced by the relaxation method has logically implied a linearization of the equation originally verified by entropy and  $\Pi + a_C^2 \tau$  seems to play the same role.

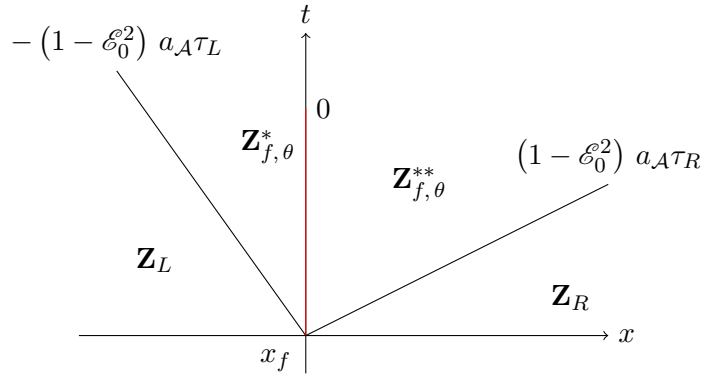
Besides, the knowledge of the Riemann invariants allow to easily solve the one-dimensional Riemann problem at a given face  $f$ , with  $\mathbf{Z}_L$  and  $\mathbf{Z}_R$  taken as initial conditions. Figure 1 describes the different states and waves produced.



**Figure 1:** Subsystem  $\mathcal{C}^\mu$ : waves and states



**Figure 2:** Subsystem  $\mathcal{A}^\mu$ : waves and states



**Figure 3:** Approximate Riemann Solver of Subsystem  $\mathcal{A}^\mu$ : waves and states

The two intermediate states  $\mathbf{Z}_{f,C}^*$  and  $\mathbf{Z}_{f,C}^{**}$  are:

$$\mathbf{Z}_{f,c}^* = \begin{bmatrix} \rho_{L,c}^* \\ u_C^* \\ \Pi_C^* \\ e_{L,c}^* \end{bmatrix}, \quad \mathbf{Z}_{f,c}^{**} = \begin{bmatrix} \rho_{R,c}^* \\ u_C^* \\ \Pi_C^* \\ e_{R,c}^* \end{bmatrix}, \quad (38) \text{ with: } \begin{cases} u_C^* = \frac{u_R + u_L}{2} - \frac{\mathcal{E}_0}{2a_C} (p_R - p_L), \\ \mathcal{E}_0^2 \Pi_C^* = \mathcal{E}_0^2 \frac{p_R + p_L}{2} - \frac{\mathcal{E}_0 a_C}{2} (u_R - u_L), \\ \rho_{k,c}^* = 1/\tau_{k,c}^*, \quad \tau_{k,c}^* = \tau_k + \frac{(-1)^{i_k+1}}{\mathcal{E}_0 a_C} (u_C^* - u_k), \\ e_{k,c}^* = e_k + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} (\Pi_C^* u_C^* - p_k u_k), \\ k \in \{L, R\}, \quad i_L = 1, \quad i_R = 2 \end{cases} \quad (39)$$

Define  $\mathbf{W}_{i+1/2}^{*,n}$  (respectively  $\mathbf{W}_{i+1/2}^{**,n}$ ) using  $\mathbf{Z}_{i+1/2,c}^*$  (respectively  $\mathbf{Z}_{i+1/2,c}^{**}$ ) and introduce  $\mathbf{U}_{i+1/2}^{*,n} = L(\mathbf{W}_{i+1/2}^{*,n})$ ,  $\mathbf{U}_{i+1/2}^{**,n} = L(\mathbf{W}_{i+1/2}^{**,n})$ . The convective numerical flux then reads:

$$\mathbf{H}_{c_{i+1/2}}^n = \begin{cases} L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n) & \text{if } u_i^n - \mathcal{E}_0^n (a_C^n)_{i+1/2} \tau_i^n > 0, \\ L(\mathbf{F}_C^\mu)(\mathbf{U}_{i+1/2}^{*,n}) & \text{if } u_i^n - \mathcal{E}_0^n (a_C^n)_{i+1/2} \tau_i^n \leq 0 < (u_C^*)_{i+1/2}^n, \\ L(\mathbf{F}_C^\mu)(\mathbf{U}_{i+1/2}^{**,n}) & \text{if } (u_C^*)_{i+1/2}^n \leq 0 < u_{i+1}^n + \mathcal{E}_0^n (a_C^n)_{i+1/2} \tau_{i+1}^n, \\ L(\mathbf{F}_C^\mu)(\mathbf{U}_{i+1}^n) & \text{if } u_{i+1}^n + \mathcal{E}_0^n (a_C^n)_{i+1/2} \tau_{i+1}^n \leq 0, \\ (a_C^n)_{i+1/2} = & K \max(\rho_i^n (c_C)_{i+1/2}^n, \rho_{i+1}^n (c_C)_{i+1/2}^n), \quad K > 1. \end{cases} \quad (40)$$

Furthermore, using the exact Godunov structure and the fact that all the fields are linearly degenerate, one can rewrite the relaxation flux in a more compact way (see [2, 3]) as:

$$\mathbf{H}_{c_{i+1/2}}^n = \begin{cases} \frac{1}{2} (L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n) + L(\mathbf{F}_C^\mu)(\mathbf{U}_{i+1}^n)) \\ - \frac{1}{2} |u_i^n - \mathcal{E}_0^n (a_C^n)_{i+1/2} \tau_i^n| (\mathbf{U}_{i+1/2}^{*,n} - \mathbf{U}_i^n) \\ - \frac{1}{2} |(u_C^*)_{i+1/2}^n| (\mathbf{U}_{i+1/2}^{**,n} - \mathbf{U}_{i+1/2}^{*,n}) \\ - \frac{1}{2} |u_{i+1}^n + \mathcal{E}_0^n (a_C^n)_{i+1/2} \tau_{i+1}^n| (\mathbf{U}_{i+1}^n - \mathbf{U}_{i+1/2}^{**,n}). \end{cases} \quad (41)$$

### 2.2.2 Acoustic Part

The acoustic system  $\mathcal{A}^\mu$  is also hyperbolic and its eigenvalues are:  $\lambda_1^{A,\mu} = -(1 - \mathcal{E}_0^2) a_A \tau$ ,  $\lambda_2^{A,\mu} = \lambda_3^{A,\mu} = 0$ ,  $\lambda_4^{A,\mu} = (1 - \mathcal{E}_0^2) a_A \tau$ . Once again the Riemann invariants can be easily found and read:

$$\begin{aligned} \mathcal{I}_{\mathcal{E}_0,1}^{A,\mu} &= \left\{ \rho, u + \frac{\Pi}{a_A}, e + \frac{\Pi u}{a_A} \right\}, \\ \mathcal{I}_{\mathcal{E}_0,2,3}^{A,\mu} &= \{u, \Pi\}, \\ \mathcal{I}_{\mathcal{E}_0,4}^{A,\mu} &= \left\{ \rho, u - \frac{\Pi}{a_A}, e - \frac{\Pi u}{a_A} \right\}. \end{aligned} \quad (42)$$

It can be noticed that:

$$\begin{aligned} \partial_t \left( u + \frac{\Pi}{a_{\mathcal{A}}} \right) + \lambda_4^{\mathcal{A}, \mu} \partial_x \left( u + \frac{\Pi}{a_{\mathcal{A}}} \right) &= 0, \\ \partial_t \left( u - \frac{\Pi}{a_{\mathcal{A}}} \right) + \lambda_1^{\mathcal{A}, \mu} \partial_x \left( u - \frac{\Pi}{a_{\mathcal{A}}} \right) &= 0. \end{aligned} \quad (43)$$

Thus,  $\omega_{\mathcal{A}}^+ = u + \frac{\Pi}{a_{\mathcal{A}}}$  (respectively  $\omega_{\mathcal{A}}^- = u - \frac{\Pi}{a_{\mathcal{A}}}$ ) remains constant along the characteristic curves of speed  $\lambda_4^{\mathcal{A}, \mu}$  (respectively  $\lambda_1^{\mathcal{A}, \mu}$ ). Besides,  $\omega_{\mathcal{A}}^+$  is a 4-strong Riemann invariant whereas  $\omega_{\mathcal{A}}^-$  is a 1-strong Riemann invariant. Following the steps of [15] and [25], equations (43) associated with the strong Riemann invariants natural properties provides a simple way to derive a time-implicit relaxation scheme for the acoustic subsystem. More details are given in [32, 31].

The one-dimensional Riemann problem can be solved exactly without difficulty. The solution is described on Figure 2.

$$\mathbf{Z}_{f, \mathcal{A}}^* = \begin{bmatrix} \rho_L \\ u_{\mathcal{A}}^* \\ \Pi_{\mathcal{A}}^* \\ e_{L, \mathcal{A}}^* \end{bmatrix}, \quad \mathbf{Z}_{f, \mathcal{A}}^{**} = \begin{bmatrix} \rho_R \\ u_{\mathcal{A}}^* \\ \Pi_{\mathcal{A}}^* \\ e_{R, \mathcal{A}}^* \end{bmatrix}, \quad (44) \text{ with: } \begin{cases} u_{\mathcal{A}}^* = \frac{u_R + u_L}{2} - \frac{1}{2a_{\mathcal{A}}} (p_R - p_L), \\ \Pi_{\mathcal{A}}^* = \frac{p_R + p_L}{2} - \frac{a_{\mathcal{A}}}{2} (u_R - u_L), \\ e_{k, \mathcal{A}}^* = e_k + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} (\Pi_{\mathcal{A}}^* u_{\mathcal{A}}^* - p_k u_k), \\ k \in \{L, R\}, \quad i_L = 1, \quad i_R = 2. \end{cases} \quad (45)$$

One can notice that the weighting parameter  $\mathcal{E}_0$  does not appear in the different intermediate quantities. Besides, the intermediate velocity, pressure and energy formulas are similar to those obtained using the Lagrange-Projection method [25, 11]. The only difference is that, in the present approach,  $a_{\mathcal{A}}$  is bounded by the modified acoustic subcharacteristic condition (29b) whereas in [25, 11] it is (31). The related numerical flux writes:

$$\begin{aligned} \mathbf{H}_{\text{ac}i+1/2}^n &= (1 - (\mathcal{E}_0^n)^2) \begin{bmatrix} 0 \\ (\Pi_{\mathcal{A}}^*)_{i+1/2}^n \\ (\Pi_{\mathcal{A}}^*)_{i+1/2}^n (u_{\mathcal{A}}^*)_{i+1/2}^n \end{bmatrix}, \\ (a_{\mathcal{A}}^n)_{i+1/2} &= K \max(\rho_i^n (c_{\mathcal{A}})_i^n, \rho_{i+1}^n (c_{\mathcal{A}})_{i+1}^n), \quad K > 1. \end{aligned} \quad (46)$$

### 2.2.3 General Remarks on the Splitting Operator Algorithm:

The overall algorithm updating the discrete solution from  $t^n$  to  $t^n + \Delta t$  is the following: starting from a given state  $\mathbf{U}^n$ , a given relaxation pressure  $\Pi^n = p^{EOS}(\mathbf{U}^n)$  and a given splitting parameter  $\mathcal{E}_0^n$ , the homogeneous versions of subsystems  $\mathcal{C}^\mu$  and  $\mathcal{A}^\mu$  are successively solved using the relaxation scheme fluxes presented in (40) and (46). At the end of each resolution, the new discrete states  $\mathbf{W}_i^{n+}$  (after the convective sub-step) and  $\mathbf{W}_i^{n+1}$  (after the acoustic sub-step) are projected on the equilibrium manifold (28). Such a projection procedure presented in [6] can be seen as an additional sub-step resolving  $\partial_t \mathbf{W} = \mathbf{S}^\mu(\mathbf{W})$  with  $\mathbf{S}^\mu(\mathbf{W}) = [0, 0, 0, (p - \Pi)/\mu]^T$  and  $\mu \rightarrow 0$ . It allows to provide the appropriate physical pressure for the flux construction between two sub-steps. Afterwards, the weighting factor  $\mathcal{E}_0$  is updated. One can notice that the overall operator splitting procedure is conservative since  $\mathcal{C}$  and  $\mathcal{A}$



are conservative subsystems and the resolution of  $\mathcal{C}^\mu$  and  $\mathcal{A}^\mu$  is performed using an exact conservative Godunov scheme. The global relaxation scheme including both steps writes:

$$\begin{cases} \mathbf{U}_i^{n+} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x} \left( \mathbf{H}_{\mathbf{c}_{i+1/2}}(\mathbf{U}_i^n, \mathbf{U}_{i+1}^n) - \mathbf{H}_{\mathbf{c}_{i-1/2}}(\mathbf{U}_{i-1}^n, \mathbf{U}_i^n) \right), \\ \mathbf{\Pi}_i^{n+} = p^{EOS}(\mathbf{U}_i^{n+}), \end{cases} \quad (47)$$

$$\begin{cases} \mathbf{U}_i^{n+1} = \mathbf{U}_i^{n+} - \frac{\Delta t}{\Delta x} \left( \mathbf{H}_{\mathbf{ac}_{i+1/2}}(\mathbf{U}_i^{n+}, \mathbf{U}_{i+1}^{n+}) - \mathbf{H}_{\mathbf{ac}_{i-1/2}}(\mathbf{U}_{i-1}^{n+}, \mathbf{U}_i^{n+}) \right), \\ \mathbf{\Pi}_i^{n+1} = p^{EOS}(\mathbf{U}_i^{n+1}). \end{cases}$$

Written in one single conservative step, the scheme reads:

$$\begin{aligned} \mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \frac{\Delta t}{\Delta x} & \left( \mathbf{H}_{\mathbf{c}_{i+1/2}}(\mathbf{U}_i^n, \mathbf{U}_{i+1}^n) - \mathbf{H}_{\mathbf{c}_{i-1/2}}(\mathbf{U}_{i-1}^n, \mathbf{U}_i^n) \right) \\ & - \frac{\Delta t}{\Delta x} \left( \mathbf{H}_{\mathbf{ac}_{i+1/2}}(\mathbf{U}_i^{n+}, \mathbf{U}_{i+1}^{n+}) - \mathbf{H}_{\mathbf{ac}_{i-1/2}}(\mathbf{U}_{i-1}^{n+}, \mathbf{U}_i^{n+}) \right). \end{aligned} \quad (48)$$

Following the continuous definition (9) of the parameter  $\mathcal{E}_0$ , one can introduce  $\mathcal{E}_0^n$  as:

$$\begin{aligned} \mathcal{E}_0^n &= \max(\mathcal{E}_{inf}, \min(M_{max}^n, 1)), \\ \text{with } M_{max}^n &= \max_{i \in [1, N_{cells}]} \left( \frac{|u_i^n|}{c_i^n} \right). \end{aligned} \quad (49)$$

Here,  $c_i^n = c(\rho_i^n, p_i^n)$  where  $c(.,.)$  is the sound speed function defined in equation (2). As already mentioned in **Subsection 1.2**,  $0 < \mathcal{E}_{inf} \ll 1$  is only a lower bound preventing  $\mathcal{E}_0^n$  from being exactly equal to zero if velocity is initially null everywhere.

### 2.3 CFL Condition Choice

**Definition 1** (CFL condition based on the Euler system). In order to adapt timesteps to the real waves produced by the Euler system, let us define the discrete time step at iteration  $n$  as:

$$\begin{aligned} \Delta t_E^n &= \frac{\sigma}{2} \frac{\Delta x}{\max_{i+1/2} \left( \max(|u_i^n| + c_i^n, |u_{i+1}^n| + c_{i+1}^n) \right)}, \\ 0 < \sigma < 1. \end{aligned} \quad (50)$$

CFL condition (50) is adapted to the resolution of the overall Euler system. However, because of the weighted splitting approach, one can exhibit two additional CFL conditions which would be sufficient to guarantee stability of both  $\mathcal{C}$  and  $\mathcal{A}$  subsystems if they were solved independently. These CFL conditions write:

$$\begin{aligned} \Delta t_{\mathcal{C}}^n &= \frac{\sigma}{2} \frac{\Delta x}{\max_{i+1/2} \left( \max \left( \left| u_i^n - \mathcal{E}_0^n (a_{\mathcal{C}})_{i+1/2}^n \tau_i^n \right|, \left| u_{i+1}^n + \mathcal{E}_0^n (a_{\mathcal{C}})_{i+1/2}^n \tau_{i+1}^n \right| \right) \right)}, \\ \Delta t_{\mathcal{A}}^n &= \frac{\sigma}{2} \frac{\Delta x}{(1 - (\mathcal{E}_0^n)^2) \max_{i+1/2} \left( (a_{\mathcal{A}})_{i+1/2}^n \max(\tau_i^n, \tau_{i+1}^n) \right)}, \\ 0 < \sigma < 1. \end{aligned} \quad (51)$$

One should keep in mind that it is absolutely not sufficient, in a fractional-step method, to constrain the time-step by only sub-steps CFL condition in order to guarantee the stability of the overall algorithm. A very simple hand-made but rather convincing example described in [16] shows that the CFL condition of the unsplit system has to be taken into account too. Hence, the final CFL condition reads:

$$\Delta t^n = \min(\Delta t_E^n, \Delta t_C^n, \Delta t_A^n). \quad (52)$$

We now study the discrete properties of our weighted splitting approach. Special attention will be paid to the positivity of both density and internal energy.

## 2.4 Discrete Properties of the Overall Scheme

### 2.4.1 Discrete Density Positivity

Let us first notice that the acoustic resolution step of  $\mathcal{A}^\mu$  does not modify density. Then, one has just to check that discrete density remains positive after the convective step. This is classically done in [2] by rewriting the convective relaxation scheme (34), (40) as:

$$\begin{aligned} \mathbf{U}_i^{n+1} &= \frac{\mathbf{U}^+(\mathbf{W}_i^n, \mathbf{W}_{i-1}^n) + \mathbf{U}^-(\mathbf{W}_{i+1}^n, \mathbf{W}_i^n)}{2}, \\ \text{with: } \mathbf{U}^+(\mathbf{W}_L, \mathbf{W}_R) &= \frac{2 \Delta t}{\Delta x} \int_0^{\frac{\Delta x}{2 \Delta t}} L(\mathbf{W}^{God})(\xi, \mathbf{W}_L, \mathbf{W}_R) d\xi, \\ \mathbf{U}^-(\mathbf{W}_L, \mathbf{W}_R) &= \frac{2 \Delta t}{\Delta x} \int_{-\frac{\Delta x}{2 \Delta t}}^0 L(\mathbf{W}^{God})(\xi, \mathbf{W}_L, \mathbf{W}_R) d\xi. \end{aligned} \quad (53)$$

Hence, positivity of the discrete density  $\rho_i^{n+1}$  is maintained if all the intermediate densities appearing in the Riemann problem described on Figure 1 and equalities (39) are positive. This can be done by adding an additional lower bound for constant  $a_C$  into the subcharacteristic condition (29a):

**Lemma 1** (Positivity of intermediate density). *Consider a given mesh face and the related local Riemann problem related to subsystem  $\mathcal{C}$ , producing waves described on Figure 1 with  $\mathbf{U}_L$  and  $\mathbf{U}_R$  as initial data. Assume that the global CFL condition (52) holds so that waves produced by local Riemann problems do not interact. Consider the intermediate densities  $\rho_{L,C}^*$  and  $\rho_{R,C}^*$  defined in (39); then:*

$$\begin{cases} \rho_{L,C}^* \geq 0, \\ \rho_{R,C}^* \geq 0, \end{cases} \Leftrightarrow \begin{cases} f_L(a_C) = a_C^2 + \frac{\rho_L(u_R - u_L)}{2\mathcal{E}_0} a_C - \frac{\rho_L(p_R - p_L)}{2} \geq 0, \\ f_R(a_C) = a_C^2 + \frac{\rho_R(u_R - u_L)}{2\mathcal{E}_0} a_C + \frac{\rho_R(p_R - p_L)}{2} \geq 0. \end{cases} \quad (54)$$

Define  $a_L^\rho$  (respectively  $a_R^\rho$ ) the highest positive root related to the second order polynomial function  $f_L(a_C)$  (respectively  $f_R(a_C)$ ). Thus, under the modified subcharacteristic condition:

$$\mathcal{C}^\mu : a_C \geq \max(\rho_L(c_C)_L, \rho_R(c_C)_R, a_L^\rho, a_R^\rho), \quad (55)$$

inequalities (54) hold. Furthermore, if  $\mathbf{U}_L$  and  $\mathbf{U}_R$  are such that  $a_L^\rho$  (respectively  $a_R^\rho$ ) is complex or negative,  $\rho_{L,C}^* \geq 0$  (respectively  $\rho_{R,C}^* \geq 0$ ) is automatically fulfilled and  $a_L^\rho$  (respectively  $a_R^\rho$ ) can be removed from (55). Eventually, it is equivalent to guarantee the intermediate density positivity and the ordering of the waves speeds:  $u_L - \mathcal{E}_0 a_C \tau_L \leq u_C^* \leq u_R + \mathcal{E}_0 a_C \tau_R$ .

The proof of this lemma, including the expressions of  $a_L^\rho$  and  $a_R^\rho$ , is written in Appendix D. The same kind of results can be found in [2] in order to enforce the mass fraction positivity. One has to mention that the non-dimensional expressions of  $a_L^\rho$  and  $a_R^\rho$  are of order  $O(1 + M/\mathcal{E}_0)$ . Therefore considering the discrete splitting parameter  $\mathcal{E}_0^n$  defined in (49), the non-dimensional roots  $a_L^\rho$  and  $a_R^\rho$  are of order one w.r.t the Mach number. Thus, their impact on the overall fractional step and notably on the numerical diffusion associated with the convective flux (41) is controlled in the sense that the modified subcharacteristic condition (55) does not imply that  $\lim_{M \rightarrow 0} a_C = +\infty$ .

## 2.4.2 Discrete Internal Energy Positivity for Ideal Gas Thermodynamics

As already presented in **Subsection 1.2.2**, in the case of an ideal gas thermodynamics, specific internal energy  $\varepsilon$  remains positive throughout space and time. Although  $\varepsilon$  is not a conservative variable, we can still consider equation (53) seen as a continuous convex combination of conservative states and notice that  $\Phi_{PG}$  defined in (14) is a convex set in the conservative phase-space (see [6] for a proof). Thus, similarly to density, a sufficient condition to guarantee the positivity of  $\varepsilon^{n+1}$  is to make sure that for  $k \in \{L, R\}$ ,  $\varepsilon_{k,C}^* = e_{k,C}^* - (u_C^*)^2/2$  as well as  $\varepsilon_{k,A}^* = e_{k,A}^* - (u_A^*)^2/2$  are positive. Such a sufficient condition is presented in the next lemma:

**Lemma 2** (Positivity of the intermediate internal energy). *Consider a given mesh face and the related local Riemann problem associated with subsystem  $\mathcal{C}$  (respectively  $\mathcal{A}$ ), producing waves described on Figure 1 (respectively Figure 2) with  $\mathbf{U}_L$  and  $\mathbf{U}_R$  as initial data. Assume that the global CFL condition (52) holds so that waves produced by local Riemann problems do not interact. For  $k \in \{L, R\}$ , consider the intermediate densities  $\varepsilon_{k,C}^*$  (respectively  $\varepsilon_{k,A}^*$ ) defined with quantities introduced in (39) (respectively (45));*

for the acoustic subsystem:

$$\begin{cases} f_{\mathcal{A},L}(a_C) = a_{\mathcal{A}}^2 - \rho_L^\varepsilon \frac{(u_R - u_L)}{2} a_{\mathcal{A}} + \rho_L^\varepsilon \frac{(p_R - p_L)}{2} \geq 0, \\ f_{\mathcal{A},R}(a_{\mathcal{A}}) = a_{\mathcal{A}}^2 - \rho_R^\varepsilon \frac{(u_R - u_L)}{2} a_{\mathcal{A}} - \rho_L^\varepsilon \frac{(p_R - p_L)}{2} \geq 0 \end{cases} \Rightarrow \begin{cases} \varepsilon_{L,\mathcal{A}}^* \geq 0, \\ \varepsilon_{R,\mathcal{A}}^* \geq 0, \end{cases} \quad (56)$$

with  $\rho_k^\varepsilon = \frac{p_k}{\varepsilon_k}$ .

for the convective subsystem:

$$\begin{cases} f_{\mathcal{C},L}(a_C) = a_C^2 - \mathcal{E}_0 \rho_L^\varepsilon \frac{(u_R - u_L)}{2} a_C + \mathcal{E}_0^2 \rho_L^\varepsilon \frac{(p_R - p_L)}{2} \geq 0, \\ f_{\mathcal{C},R}(a_C) = a_C^2 - \mathcal{E}_0 \rho_R^\varepsilon \frac{(u_R - u_L)}{2} a_C - \mathcal{E}_0^2 \rho_L^\varepsilon \frac{(p_R - p_L)}{2} \geq 0 \end{cases} \Rightarrow \begin{cases} \varepsilon_{L,C}^* \geq 0, \\ \varepsilon_{R,C}^* \geq 0, \end{cases} \quad (57)$$

Define  $(a_{\mathcal{A},L}^\varepsilon, a_{\mathcal{C},L}^\varepsilon)$  (respectively  $(a_{\mathcal{A},R}^\varepsilon, a_{\mathcal{C},R}^\varepsilon)$ ) the highest positive roots related to the couple of second order polynomial functions  $(f_{\mathcal{A},L}(a_{\mathcal{A}}), f_{\mathcal{C},L}(a_C))$  (respectively  $(f_{\mathcal{A},R}(a_{\mathcal{A}}), f_{\mathcal{C},R}(a_C))$ ). Then one can show that:

$$\begin{aligned} a_{\mathcal{C},L}^\varepsilon &= \mathcal{E}_0 a_{\mathcal{A},L}^\varepsilon, \\ a_{\mathcal{C},R}^\varepsilon &= \mathcal{E}_0 a_{\mathcal{A},R}^\varepsilon, \end{aligned} \quad (58)$$

so that under the modified subcharacteristic condition:

$$\begin{aligned} \mathcal{C}^\mu : a_C &\geq \max(\rho_L(cc)_L, \rho_R(cc)_R, a_{\mathcal{A},L}^\varepsilon, a_{\mathcal{A},R}^\varepsilon), \\ \mathcal{A}^\mu : a_{\mathcal{A}} &\geq \max(\rho_L(c_{\mathcal{A}})_L, \rho_R(c_{\mathcal{A}})_R, a_{\mathcal{A},L}^\varepsilon, a_{\mathcal{A},R}^\varepsilon), \end{aligned} \quad (59)$$

inequalities (56) and (57) hold. Eventually, if  $\mathbf{U}_L$  and  $\mathbf{U}_R$  are such that  $a_{\mathcal{A},L}^\varepsilon$  (respectively  $a_{\mathcal{A},R}^\varepsilon$ ) is complex or negative, the positivity of  $(\varepsilon_{L,C}^*, \varepsilon_{L,\mathcal{A}}^*)$  (respectively  $(\varepsilon_{R,C}^*, \varepsilon_{R,\mathcal{A}}^*)$ ) is automatically fulfilled and  $a_{\mathcal{A},L}^\varepsilon$  (respectively  $a_{\mathcal{A},R}^\varepsilon$ ) can be removed from (59).

The proof of this lemma and the formulas for the polynomial roots are provided in Appendix E. Once again, it can be shown that the non-dimensional expressions of  $a_{\mathcal{A},L}^\varepsilon$  and  $a_{\mathcal{A},R}^\varepsilon$  are of order one w.r.t the Mach number.

In the following section, a truncation error analysis performed on smooth solutions is derived in order to assess the effect of the splitting parameter  $\mathcal{E}_0$  in terms of numerical diffusion in the case of one-dimensional low Mach number compressible flows.

### 3 A Truncation Error Analysis

In [25, 11] a fractional step approach based on a Lagrange-Projection splitting [15] is proposed. The authors use a relaxation scheme, very similar to these introduced in (44), (45), (46), to discretize their corresponding acoustic flux. By performing a 1D non-dimensional truncation error analysis, they show that the dissipative part of the discrete acoustic momentum flux scales as  $O(\Delta x/M)$ . This prohibitive dissipation does not vanish through their transport sub-step and the resulting diffusive operator for the overall scheme is of the same order.

A detailed study in [18, 20, 19] points out that this pathology actually holds for all Godunov-like schemes and hides far more intricate spatial discretization issues if one is interested in maintaining the solution of compressible Riemann solvers close to its incompressible initial part.

As explained in the introduction, the present work evolves within the point (II) framework. Then, we are simply interested in reducing the numerical diffusion which could occur on acoustic wave fronts in the case where  $0 < M \ll 1$ . Hence, we only consider 1D non-dimensional truncation error analysis as a simple (although incomplete) tool to have a rough idea of the numerical diffusion produced by the spatial discretization of the present compressible solver when the Mach number is small compared with one. We notably want to measure the effect of the splitting parameter  $\mathcal{E}_0^n$  on the amplitude of the overall scheme numerical diffusion. For that purpose, we start by performing a truncation error analysis for each subsystem as if they were solved independently. Then, the additional numerical diffusion due to the composition between the discrete convective state update and the acoustic flux is analyzed.

#### 3.1 Truncation Error of the Weighted Splitting Subsystems

Each truncation error analysis is made on non-dimensional systems. Let us introduce  $t_r, l_r, \rho_r, u_r, p_r$  the reference time-scale, space-scale, density, material velocity and pressure. Besides, define a reference acoustic celerity  $c_r$  such that  $\rho_r c_r^2 = p_r$ . Finally, consider the Mach number  $M = u_r/c_r$  and the Strouhal

number  $St_r = l_r / (t_r u_r)$ . Then, the overall non-dimensional Euler compressible system reads:

$$\mathcal{E} : \begin{cases} St_r \partial_t \rho + \partial_x (\rho u) = 0, \\ St_r \partial_t (\rho u) + \partial_x \left( \rho u^2 + \frac{p}{M^2} \right) = 0, \\ St_r \partial_t (\rho e) + \partial_x ((\rho e + p) u) = 0, \end{cases} \quad (60)$$

with  $e = \varepsilon(\rho, \varepsilon) + M^2 \frac{u^2}{2}$ . Note that in the context of point (II), the reference time-scale is based on the fast acoustic waves:  $t_r = l_r / c_r$ . Thus,  $St_r = 1/M$ . In the sequel, the Mach number is *fixed* to a given low value:  $0 < M \ll 1$ . Then, by making the non-dimensional space-step  $\Delta x$  tend towards zero for smooth solutions, one seeks to find the amplitude of the diffusive operator induced by the spatial discretization of the overall fractional step. We notably want to identify the diffusion sources of order  $O(\Delta x/M)$ .

Here is the truncation error analysis performed on the convective subsystem  $\mathcal{C}$ :

**Proposition 2** (Truncation error analysis of the convective subsystem). *Consider the convective numerical scheme defined by equations (34), (40) and (41). Under the CFL condition (50), This scheme is consistent with the non-dimensional convective subsystem:*

$$\mathcal{C}^{trunc} : \begin{cases} St_r \partial_t \rho + \partial_x (\rho u) = O(St_r \Delta t) + O\left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x, \\ St_r \partial_t (\rho u) + \partial_x \left( \rho u^2 + \frac{\mathcal{E}_0^2(t)}{M^2} p \right) = O(St_r \Delta t) + O\left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x, \\ St_r \partial_t (\rho e) + \partial_x ((\rho e + \mathcal{E}_0^2(t) p) u) = O(St_r \Delta t) + O\left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x. \end{cases} \quad (61)$$

The proof is given in Appendix H. Let us start by mentioning that the first order time-discretization of the present approach produces a diffusion of order  $O(St_r \Delta t) = O(\Delta t/M)$  which can be important when  $0 < M \ll 1$ . This difficulty is not treated in the present paper which focuses exclusively on simple means to reduce the numerical diffusion induced by the spatial discretization of Godunov-like schemes. In order to better understand the above orders of magnitude, consider the diffusive part of the convective flux written in (41). Then one can rewrite the difference of states as:

$$\mathbf{U}_{i+1/2}^* - \mathbf{U}_i = \begin{bmatrix} \rho_{i,\mathcal{C}}^* - \rho_i \\ (\rho_{i,\mathcal{C}}^* - \rho_i) (u_{\mathcal{C}}^*)_{i+1/2} \\ (\rho_{i,\mathcal{C}}^* - \rho_i) e_{i,\mathcal{C}}^* \end{bmatrix} + \begin{bmatrix} 0 \\ \rho_i ((u_{\mathcal{C}}^*)_{i+1/2} - u_i) \\ \rho_i (e_{i,\mathcal{C}}^* - e_i) \end{bmatrix}, \quad (62)$$

$$\mathbf{U}_{i+1/2}^{**} - \mathbf{U}_{i+1} = \begin{bmatrix} \rho_{i+1,\mathcal{C}}^* - \rho_{i+1} \\ (\rho_{i+1,\mathcal{C}}^* - \rho_{i+1}) (u_{\mathcal{C}}^*)_{i+1/2} \\ (\rho_{i+1,\mathcal{C}}^* - \rho_{i+1}) e_{i+1,\mathcal{C}}^* \end{bmatrix} + \begin{bmatrix} 0 \\ \rho_{i+1} ((u_{\mathcal{C}}^*)_{i+1/2} - u_{i+1}) \\ \rho_{i+1} (e_{i+1,\mathcal{C}}^* - e_{i+1}) \end{bmatrix}, \quad (63)$$

where the expressions of  $(\rho_{i,\mathcal{C}}^*, e_{i,\mathcal{C}}^*)$  respectively  $(\rho_{i+1,\mathcal{C}}^*, e_{i+1,\mathcal{C}}^*)$  are provided in (39) with  $L = i$ ,  $R = i + 1$ . The diffusion of order  $O\left(\frac{M}{\mathcal{E}_0} \Delta x\right)$  has been produced by the density differences  $\rho_{i,\mathcal{C}}^* - \rho_i$  and  $\rho_{i+1,\mathcal{C}}^* - \rho_{i+1}$ . It stems from the volume contraction operator  $\rho \partial_x u$  which, contrary to [11], is

present in our convective subsystem to provide a conservative mass flux when associated with the transport operator  $u \partial_x \rho$ . The diffusion of order  $O\left(\frac{\mathcal{E}_0}{M} \Delta x\right)$  has been produced by the acoustic part of the non-dimensional relaxation eigenvalues  $u_i - \frac{\mathcal{E}_0}{M} (a_{\mathcal{C}})_{i+1/2} \tau_i$  and  $u_{i+1} + \frac{\mathcal{E}_0}{M} (a_{\mathcal{C}})_{i+1/2} \tau_{i+1}$ ; as well as the non-centered part of the intermediate velocity  $(u_{\mathcal{C}}^*)_{i+1/2} = \frac{u_{i+1} + u_i}{2} - (\mathcal{E}_0/M) \frac{(p_{i+1} - p_i)}{2(a_{\mathcal{C}})_{i+1/2}}$ . In both cases, the splitting parameter acts as a compensator of the strong diffusive effect of order  $O(1/M)$ .

Therefore the numerical diffusion produced by the convective subsystem  $\mathcal{C}$  is of order  $O(\Delta x)$  in every Mach regime. If the convective part of the present weighted splitting approach structurally avoids the numerical diffusion when  $0 < M \ll 1$ , the acoustic one continues to suffer from it. Indeed:

**Proposition 3** (Truncation error analysis of the acoustic subsystem). *Consider the acoustic numerical scheme defined by equations (34) and (46). Under the CFL condition (50), This scheme is consistent with the non-dimensional acoustic subsystem:*

$$\mathcal{A}^{trunc} : \begin{cases} St_r \partial_t \rho = O(St_r \Delta t), \\ St_r \partial_t (\rho u) + \partial_x \left( \frac{(1 - \mathcal{E}_0^2(t))}{M^2} p \right) = O(St_r \Delta t) + O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x\right), \\ St_r \partial_t (\rho e) + \partial_x \left( (1 - \mathcal{E}_0^2(t)) p u \right) = O(St_r \Delta t) + O\left((1 - \mathcal{E}_0^2)(M + \frac{1}{M}) \Delta x\right). \end{cases} \quad (64)$$

Here, the term of order  $O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x\right)$  in the momentum equation truncation error is directly produced by the dissipative part of the intermediate acoustic pressure:

$$\frac{(\Pi_{\mathcal{A}}^*)_{i+1/2}}{M^2} = (1/M^2) \frac{p_{i+1} + p_i}{2} - (1/M) \frac{(a_{\mathcal{A}})_{i+1/2}}{2} (u_{i+1} - u_i). \quad (65)$$

Besides, in the energy flux, the product between the centered part of  $(\Pi_{\mathcal{A}}^*)_{i+1/2}$  and the non-centered one of  $(u_{\mathcal{A}}^*)_{i+1/2} = \frac{u_{i+1} + u_i}{2} - (1/M) \frac{1}{2(a_{\mathcal{A}})_{i+1/2}} (p_{i+1} - p_i)$  provides the contribution  $-(1/M) \frac{1}{4(a_{\mathcal{A}})_{i+1/2}} (p_{i+1}^2 - p_i^2)$  also responsible for the  $O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x\right)$  dissipative term. One can notice that the splitting parameter  $\mathcal{E}_0$  does not allow to damp the above diffusive terms since it solely acts as a  $(1 - \mathcal{E}_0^2)$  factor.

What is more, the flux construction in the acoustic sub-step is fed by a modified conservative state  $\mathbf{U}^{n+}$  which is solution of the discrete convective scheme (34), (40) and (41). Such a modified state can hold perturbations which, once injected in the non-dimensional acoustic pressure flux (65), can potentially bring additional numerical diffusion of order  $O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x\right)$ . This is studied in the following paragraph.

### 3.2 Effect of the Convective and Acoustic Operators Composition on the Truncation Error

Let us consider a smooth initial state  $x_i \rightarrow \mathbf{U}(x_i)^{n+}$  solution of the non-dimensional discrete convective scheme. According to (61), under the CFL condition (50),  $\mathbf{U}_i^{n+}$  is such that:

$$\begin{aligned} \mathbf{U}_i^{n+} &= \mathbf{U}_i^n + \mathbf{B}_i^n + \underline{O} \left( M \left( 1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0} \right) \Delta x^2 \right), \\ \text{with: } \mathbf{B}_i^n &= -\frac{\Delta t^n}{St_r} \partial_x L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n) = -\Delta t^n M \partial_x L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n), \\ L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n) &= \left[ \rho_i^n u_i^n, \rho_i^n (u_i^n)^2 + (\mathcal{E}_0^n/M)^2 p_i^n, (\rho_i^n e_i^n + (\mathcal{E}_0^n)^2 p_i^n) u_i^n \right]^T, \\ \rho_i^n e_i^n &= \rho_i^n \varepsilon_i^n + M^2 \rho_i^n \frac{(u_i^n)^2}{2}, \\ \text{and } \Delta t^n &= \bar{A}^n \Delta x = \frac{\sigma}{2} \frac{\Delta x}{\max_{i+1/2} (M |u_i^n| + c_i^n, M |u_{i+1}^n| + c_{i+1}^n)}, \quad 0 < \sigma < 1. \end{aligned} \quad (66)$$

Neglecting the second order term  $\underline{O} \left( M \left( 1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0} \right) \Delta x^2 \right)$ , we want to study the influence of the perturbation  $\mathbf{B}_i^n$  in terms of additional numerical diffusion when injected inside the non-dimensional pressure flux of the acoustic momentum equation (65). According to (66),  $\mathbf{B}_i^n$  is of order  $\underline{O}(M\Delta x)$  since, according to formula (49),  $\mathcal{E}_0$  is proportional to the Mach number. Thus, it is only its contribution to the centered part of  $(\Pi_{\mathcal{A}}^*)_{i+1/2}/M^2$  that can potentially create an additional numerical diffusion scaling as a  $\underline{O}(\Delta x/M)$  term. Let us first rewrite  $p(\mathbf{U}_i^{n+})$  as:

$$\begin{aligned} p(\mathbf{U}_i^n) - \bar{A}^n M \Delta x \left[ \nabla_{\mathbf{U}} p(\mathbf{U}_i^n) \cdot \partial_x L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n) \right] + O((M\Delta x)^2), \\ \text{with: } p(\mathbf{U}_i^n) = p(\underbrace{\rho_i^n, \rho_i^n e_i^n - M^2 \frac{(\rho_i^n u_i^n)^2}{2\rho_i^n}}_{\rho_i^n \varepsilon_i^n}), \\ \text{and } \nabla_{\mathbf{U}} p(\mathbf{U}_i^n) = \begin{bmatrix} \partial_\rho p_{|\rho\varepsilon}(\rho_i^n, \rho_i^n \varepsilon_i^n) + M^2 \frac{(u_i^n)^2}{2} \partial_{\rho\varepsilon} p_{|\rho}(\rho_i^n, \rho_i^n \varepsilon_i^n) \\ -M^2 \frac{u_i^n}{2} \partial_{\rho\varepsilon} p_{|\rho}(\rho_i^n, \rho_i^n \varepsilon_i^n) \\ \partial_{\rho\varepsilon} p_{|\rho}(\rho_i^n, \rho_i^n \varepsilon_i^n) \end{bmatrix}. \end{aligned} \quad (67)$$

Then, the zeroth order terms w.r.t  $M$  of the product  $\nabla_{\mathbf{U}} p(\mathbf{U}_i^n) \cdot \partial_x L(\mathbf{F}_C^\mu)(\mathbf{U}_i^n)$  are only:

$$\left[ \partial_\rho p_{|\rho\varepsilon}(\rho_i^n, \rho_i^n \varepsilon_i^n) \right] \partial_x (\rho_i^n u_i^n) + \left[ \partial_{\rho\varepsilon} p_{|\rho}(\rho_i^n, \rho_i^n \varepsilon_i^n) \right] \partial_x (\rho_i^n \varepsilon_i^n u_i^n). \quad (68)$$

In the case of a stiffened gas thermodynamics (15):

$$\begin{aligned} \partial_\rho p_{|\rho\varepsilon}(\rho_i^n, \rho_i^n \varepsilon_i^n) &= 0, \\ \left[ \partial_{\rho\varepsilon} p_{|\rho}(\rho_i^n, \rho_i^n \varepsilon_i^n) \right] \partial_x (\rho_i^n \varepsilon_i^n u_i^n) &= \partial_x (\rho_i^n \varepsilon_i^n u_i^n) = u_i^n \partial_x \rho_i^n + \rho_i^n \partial_x u_i^n. \end{aligned} \quad (69)$$

Thus, if no supplementary hypothesis are made on the shape of  $\rho_i^n$  and  $u_i^n$ , the combination of both convective and acoustic sub-steps entails a spurious numerical diffusion because of the equation of state

relating the pressure with the internal energy  $\rho\varepsilon$ . Such a difficulty can be circumvented if one assumes that at time  $t^n$  the discrete solution  $\mathbf{U}_i^n$  lies into the discrete well-prepared space (see [38, 18, 19]):

$$\begin{aligned} u_i^n &= u_0 + O(M), \\ p_i^n &= p_0 + O(M^2), \end{aligned} \tag{70}$$

with:  $u_0, p_0$  constants of order one,

since in that case  $\partial_x(p_i^n u_i^n)$  becomes of order  $O(M)$ . Recall that the present work concentrates on the point (II) described in the introduction. Hence, in the case where the stiffness of the thermodynamics allows to generate high amplitude pressure jumps even if  $0 < M \ll 1$ , the well-prepared conditions (70) do not hold. However we can still consider the above analysis as a basic way to identify the main sources of numerical diffusion, try to remove them when it is possible and observe the impact of the corrections on the numerical results. In the sequel, we restrict our truncation error analysis to stiffened gas thermodynamics and initial well-prepared conditions.

In any case, a last special treatment has to be implemented to remove the  $O\left(\frac{(1-\varepsilon_0^2)\Delta x}{M}\right)$  diffusive terms brought by the acoustic non-centered part in the momentum flux.

### 3.3 Correction of the Acoustic Splitting Step

In [18, 25], facing at similar difficulties regarding the amplitude of the numerical diffusion brought by their acoustic sub-step, the authors apply a discrete correction to the non-centered part of the acoustic pressure. Such a correction, originally introduced in [22], has been also studied in [35, 19] as a way to control the accuracy of the computational solution towards its initial incompressible part when the Mach number is close to zero. Here, however, we only consider the correction as a tool which could potentially reduce the one-dimensional diffusion of our compressible solver and then provide a better accuracy towards the *compressible* solution of Riemann problems when the Mach number is small compared with one.

The correction consists in adding artificially a term of order  $O(M)$  in front of the non-centered part in the acoustic pressure. This new term can be built using the local velocity and sound speed. The modified acoustic flux reads:

$$\begin{aligned} \mathbf{H}_{\text{ac}i+1/2}^n &= (1 - (\varepsilon_0^n)^2) \begin{bmatrix} 0 \\ (\Pi_{\mathcal{A},\theta}^*)_{i+1/2}^n \\ (\Pi_{\mathcal{A},\theta}^* u_{\mathcal{A}}^*)_{i+1/2}^n \end{bmatrix}, \\ \text{with: } (\Pi_{\mathcal{A},\theta}^*)_{i+1/2}^n &= \frac{p_{i+1}^n + p_i^n}{2} - \frac{(a_{\mathcal{A}}\theta)_{i+1/2}^n}{2} (u_{i+1}^n - u_i^n), \\ \text{and } \theta_{i+1/2}^n &= \min\left(\frac{|(u_{\mathcal{A}}^*)_{i+1/2}^n|}{\max(c_{i+1}^n, c_i^n)}, 1\right). \end{aligned} \tag{71}$$

As noticed in [25], the introduction of this correction does not alter the consistency of the numerical scheme because it solely impacts the non-centered part in the momentum flux which is only responsible for the numerical diffusion. Furthermore, it is possible to build an approximate Riemann solver in the



sense of Harten, Lax and Van Leer [30] with the same eigenvalues than those produced by the exact Riemann problem associated with the acoustic relaxation system  $\mathcal{A}^\mu$ . Details on this approximate Riemann solver are given in Figure 3, and equations (72), (73). The insensitivity of the eigenvalues to the correction allows to maintain the same kind of CFL condition (52) for the modified acoustic scheme.

$$\mathbf{z}_{f,\theta}^* = \begin{bmatrix} \rho_L \\ u_{L,\theta}^* \\ \Pi_{\mathcal{A},\theta}^* \\ e_{L,\theta}^* \end{bmatrix}, \quad \mathbf{z}_{f,\theta}^{**} = \begin{bmatrix} \rho_R \\ u_{R,\theta}^* \\ \Pi_{\mathcal{A},\theta}^* \\ e_{R,\theta}^* \end{bmatrix}, \quad (72) \quad \text{with:} \quad \begin{cases} u_{k,\theta}^* = u_{\mathcal{A}}^* + (-1)^{i_k} (1-\theta) \frac{(u_R - u_L)}{2}, \\ e_{k,\theta}^* = e_{k,\mathcal{A}}^* + (-1)^{i_k} (1-\theta) \frac{(u_R - u_L) u_{\mathcal{A}}^*}{2}, \\ k \in \{L, R\}, \quad i_L = 1, \quad i_R = 2. \end{cases} \quad (73)$$

Thanks to this correction term, numerical diffusion of the subsystem  $\mathcal{A}$  is modified, namely:

**Proposition 4** (Truncation error analysis of the acoustic subsystem with correction). *Consider the acoustic numerical scheme defined by equations (34) with the corrected flux (71). Suppose that pressure follows the well-prepared initial condition written in (70). Then, under the CFL condition (50), This scheme is consistent with the non-dimensional acoustic subsystem:*

$$\mathcal{A}^{trunc} : \begin{cases} St_r \partial_t \rho = O(St_r \Delta t), \\ St_r \partial_t (\rho u) + \partial_x \left( \frac{(1 - \mathcal{E}_0^2(t))}{M^2} p \right) = O(St_r \Delta t) + O\left(\frac{(1 - \mathcal{E}_0^2)\theta}{M} \Delta x\right), \\ St_r \partial_t (\rho e) + \partial_x ((1 - \mathcal{E}_0^2(t)) p u) = O(St_r \Delta t) + O((1 - \mathcal{E}_0^2)(1 + \theta)M \Delta x). \end{cases} \quad (74)$$

Assume that there exists a smooth function  $(x, t) \rightarrow \theta(x, t)$  such that  $\forall(i, n), \theta(x_{i+1/2}, t^n) = \theta_{i+1/2}^n$ . Then the numerical diffusion contained in the term of order  $O\left(\frac{(1 - \mathcal{E}_0^2)\theta}{M} \Delta x\right)$  is actually of order  $O((1 - \mathcal{E}_0^2)\Delta x)$ . Moreover, the global truncation error analysis writes:

**Proposition 5** (Truncation error analysis of the overall scheme with correction). *Assume a fluid endowed with a stiffened gas thermodynamics (15). Consider the global relaxation scheme defined by equations (48) endowed with the corrected acoustic flux (71). Suppose that initial state  $\mathbf{U}_i^n$  follows the well-prepared initial conditions written in (70). Then, under the CFL condition (50), this scheme is consistent with the non-dimensional system:*

$$\mathcal{E}^{trunc} : \begin{cases} St_r \partial_t \rho + \partial_x (\rho u) = O(St_r \Delta t) + O\left(\left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x\right), \\ St_r \partial_t (\rho u) + \partial_x \left(\rho u^2 + \frac{p}{M^2}\right) = O(St_r \Delta t) + O\left(\left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x\right) \\ \quad + O\left((1 - \mathcal{E}_0^2)\left(1 + \frac{\theta}{M}\right) \Delta x\right), \\ St_r \partial_t (\rho e) + \partial_x ((\rho e + p) u) = O(St_r \Delta t) + O\left(\left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x\right) \\ \quad + O((1 - \mathcal{E}_0^2)(1 + \theta)M \Delta x). \end{cases} \quad (75)$$

The proofs of the above propositions are written in Appendix H. Since the local correction  $\theta_{i+1/2}^n$  as well as the splitting parameter  $\mathcal{E}_0^n$  are by construction of order  $M$ , the numerical diffusion for the overall fractional step becomes of order  $\underline{O}(\Delta x)$  even when the Mach number is small compared with one. In the sequel, the anti-diffusion introduced by the coefficient  $\theta$  is referred as to " $\theta$ -correction".

In the next section, one-dimensional numerical results of the present weighted splitting approach are presented. They are only made of 1D compressible test cases including ideal and stiffened gas thermodynamics. The main objective here is to assess the accuracy w.r.t the compressible analytical solution of the present approach as well as its efficiency for a wide panel of Mach numbers. In order to do so, we systematically compare the present work with the Lagrange-Projection method described in [11, 25].

## 4 Numerical Results

### 4.1 Ideal Gas Thermodynamics

Let us first consider a one-dimensional configurable shock-tube test-case. The fluid has been firstly endowed with an ideal gas thermodynamics (14a) with the heat capacity ratio  $\gamma = 7/5$ . The simulation has been conducted on a domain of length  $1\text{ m}$ , the initial discontinuity of the Riemann problem being located at  $x = 0.5\text{ m}$ . The initial inputs of the Riemann problem are summed up on **Table 1**: Recall that the analytical solution is made of a left-going 1-rarefaction wave, a 2-contact discontinuity

	Left state	Right state
$\rho$ ( $kg.m^{-3}$ )	$\rho_{0,L} = 1.$	$\rho_{0,R} = 0.125$
$u$ ( $m.s^{-1}$ )	$u_{0,L} = 0.$	$u_{0,R} = 0.$
$p$ ( $bar$ )	$p_{0,L} = p_{0,R} (1 + \Delta)$	$p_{0,R} = 0.1$

**Table 1:** Ideal gas shock tube initial conditions

propagating to the right and a right-going 3-shock wave. The maximal Mach number is reached at the tail of the 1-rarefaction wave and can be controlled by increasing or diminishing the parameter  $\Delta$ . When  $\Delta = 9$ , the classical Sod shock-tube described in [41] is retrieved, and the maximal Mach number  $M_{max}$  is about 0.92. We will refer to it as a Mach one case. When  $\Delta = 2 \times 10^{-1}$ ,  $M_{max} \approx 9.5 \times 10^{-2}$ . This will be considered as an intermediate regime. Finally, when  $\Delta = 8 \times 10^{-3}$ ,  $M_{max} \approx 4.2 \times 10^{-3}$  and we call it low-Mach case. Let us mention that in the above three test cases, the Mach number across the left-going 3-rarefaction wave evolves from 0 to  $M_{max}$ . Then, low Mach number regions are present in every test case. However, the definition of the splitting parameter  $\mathcal{E}_0(t)$ , based on  $M_{max}(t)$ , will provide a completely different behavior according to  $\Delta$ . Indeed, in the first case, after several time-steps,  $\mathcal{E}_0(t) \approx 1$  and the contribution of the acoustic subsystem  $\mathcal{A}$  is almost negligible. This test case allows to assess the quality of the present approach in the classical configuration where large pressure variations are related to a sudden rise of the Mach number. In the third case  $0 < \mathcal{E}_0(t) \ll 1$  and the Euler system is fully split. Besides, one can notice that, in this case, the amplitude of the 1-rarefaction and the 3-shock waves is small due to the ideal gas thermodynamics. Indeed, initial

conditions are such that:

$$\frac{|p_{0,R} - p_{0,L}|}{p_{0,R}} = O(M_{max}). \quad (76)$$

For every test case, the computation ends when the right-going 3-shock wave reaches the position  $x = 0.75m$ . The corresponding final times are  $T_{\text{end}} = 4.51 \times 10^{-4} s$  for  $M_{max} = 0.92$ ,  $T_{\text{end}} = 7.31 \times 10^{-4} s$  for  $M_{max} = 9.5 \times 10^{-2}$  and  $T_{\text{end}} = 7.43 \times 10^{-4} s$  for  $M_{max} = 4.2 \times 10^{-3}$ . Besides, transmissive boundary conditions have been considered. Finally, the CFL condition of our time-explicit scheme is the one written in (52) with  $\sigma = 0.9$ . In the sequel,  $M_{max}$  is rewritten  $M$  for the sake of simplicity.

As previously announced, three criteria have been involved in order to measure the quality of the present approach: mesh convergence in  $L^1$  norm, profiles of the different computed solutions and efficiency.

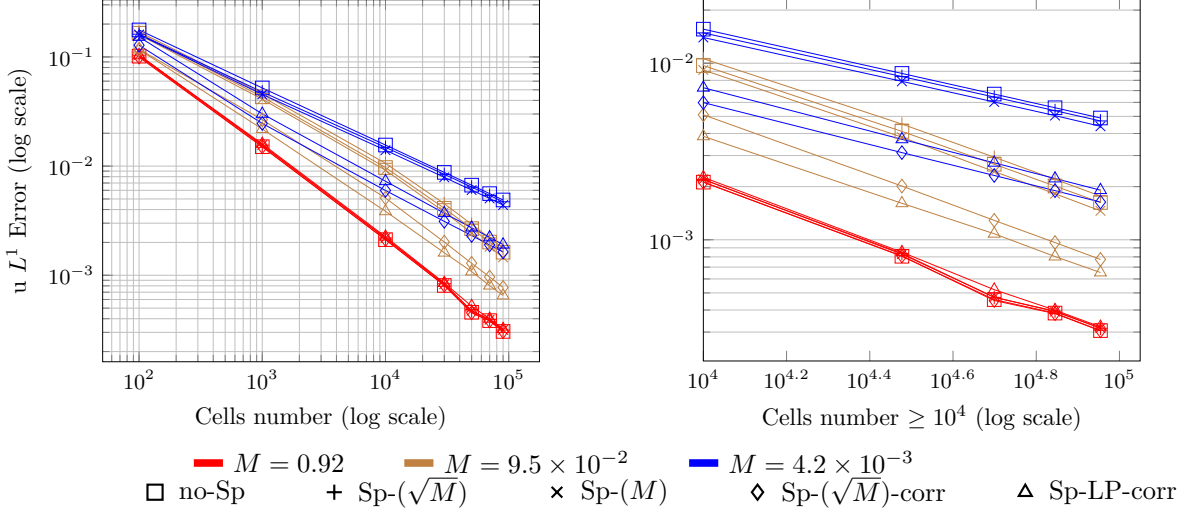
#### 4.1.1 Convergence Curves

Convergence curves have been built using a wide range of cells number:

$N_{cells} \in \{10^2, 10^3, 10^4, 3 \times 10^4, 5 \times 10^4, 7 \times 10^4, 9 \times 10^4\}$ . Convergence rates have been calculated using the error of the cases:  $N_{cells} = 7 \times 10^4$  and  $N_{cells} = 9 \times 10^4$ . For each variable of interest, three convergence curves are plotted according to the three different maximal Mach numbers defined above. Besides, five different schemes have been tested: "no-Sp" corresponds to the case where  $\mathcal{E}_0^n = 1$  is imposed along the simulation. Thus, the weighted splitting is not triggered. "Sp- $(\sqrt{M})$ " is the weighted splitting approach with  $\mathcal{E}_0^n = \max(\sqrt{\mathcal{E}_{inf}}, \min(\sqrt{M_{max}^n}, 1))$  while "Sp- $(M)$ " involves  $\mathcal{E}_0^n$  defined in formula (49). Although the asymptotic behavior w.r.t the Mach number is the same for both above definitions of  $\mathcal{E}_0^n$ , the convective flux of the second should provide a lower numerical diffusion in smooth areas according to **Proposition 2** and **Proposition 5**. Eventually, "LP" is the Lagrange-Projection splitting method, fully described in [11] and taken as a benchmark. The mention "-corr" means that the correction defined in (71) is triggered. Figure 4 corresponds to the velocity convergence curve while Figure 5 is associated with the pressure variable. Density convergence curve has intentionally not been plotted because results were extremely close.

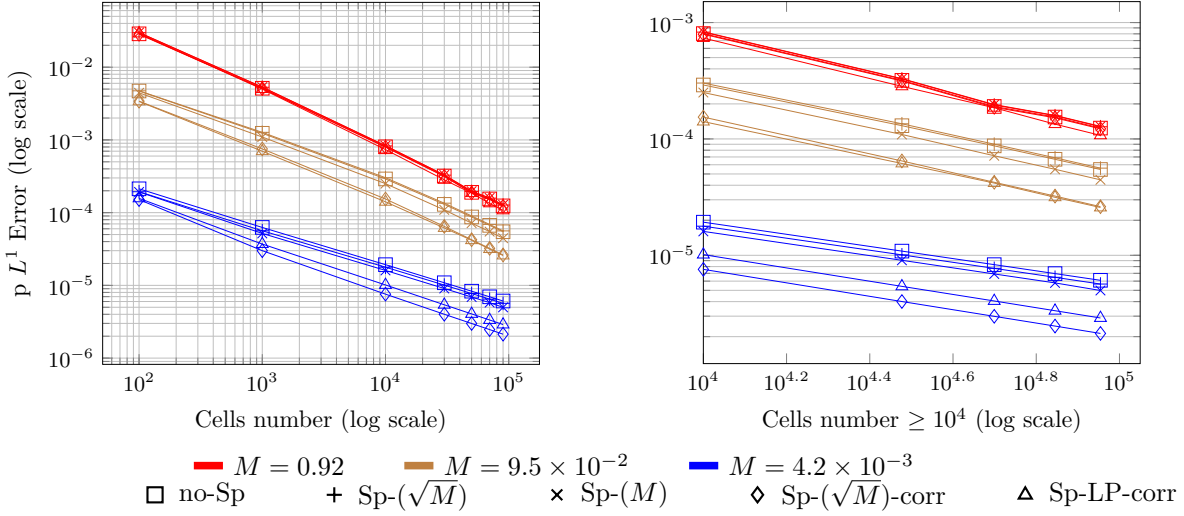
Every numerical scheme converges towards the analytical compressible solution as the mesh is refined. In the Mach one regime, convergence curves overlap quasi perfectly. It is coherent with the fact that, in such a regime, subsystem  $\mathcal{C}$  is almost similar to the full Euler system. By construction, the numerical contribution of the acoustic subsystem  $\mathcal{A}$  is negligible. Still, the proximity between the Lagrange-Projection scheme with correction and the weighted splitting schemes is less straightforward.

Furthermore, as  $M_{max}$  decreases one can observe that the schemes with the  $\theta$ -correction Sp- $(\sqrt{M})$ -corr and Sp-LP-corr are clearly more accurate than the other ones. For example on Figure 4, for  $M = 9.5 \times 10^{-2}$ , Sp- $(\sqrt{M})$ -corr reaches the precision level of  $2 \times 10^{-3}$  with a  $3 \times 10^3$  cells mesh whereas it requires more than  $7 \times 10^3$  for No-Sp. This is in agreement with the acoustic truncation error result of **Proposition 4** derived for a smooth and initially well-prepared solution. In 1D, the  $\theta$ -correction has an anti-diffusion effect which improves the global accuracy of the proposed approach in smooth regions of the computational solution. Although this was already observed in [19] for a Godunov scheme (without splitting), it is satisfying to notice that the present splitting does not alter this anti-diffusion effect. Moreover, as it can be seen on Figure 5, for  $M = 4.2 \times 10^{-3}$ , switching the weighting parameter  $\mathcal{E}_0^n$  from  $\sqrt{M_{max}^n}$  to  $M_{max}^n$  has only a very slight positive effect on the scheme accuracy. This is due to the fact that, in case of low-Mach number compressible flow, most of the



**Figure 4:** Velocity Convergence Curves

numerical diffusion is generated by the acoustic part of the weighted splitting approach. To complete this comparison, one could have wished to see the case  $\text{Sp}(M)\text{-corr}$  which, according to **Proposition 5**, is supposed to reduce the convective and acoustic numerical diffusion for a smooth solution initially in the well-prepared space. Unfortunately this case suffers from strong non-physical oscillations located in the left rarefaction wave area. Plots of these oscillations for different cells numbers can be seen on Appendix G.



**Figure 5:** Pressure Convergence Curves

They have already been observed for low-Mach corrected numerical schemes written in Eulerian coordinates (see [25], chapter 3, section 3.G). However, these spurious perturbations are damped in the sense of the  $L^\infty$  and  $L^1$  norms as the mesh is refined. So far finding the optimal choice for the couple  $(\mathcal{E}_0, \theta)$  in order to prevent the acoustic momentum flux from being completely centered and

thus triggering such oscillations is still an open issue.

Let us do a last remark about the convergence rates of the different curves as the Mach number tends towards zero. In [23], the authors show that, for Riemann problems whose maximal Mach number  $M_{max}$  is close to 1, the rate of convergence of variables varying through genuinely non-linear waves is around 1. However, for variables jumping through a contact wave, typically the density  $\rho$ , the rate is around 0.5. This numerical observation holds independently of the approximate Riemann solver at stake and is also mentioned in [19]. Thus, since  $u$  and  $p$  do not jump through the right-going 2-contact discontinuity of the above shock tubes, the expected rate of convergence should be 1.

	$M = 0.92$	$M = 9.5 \times 10^{-2}$	$M = 4.2 \times 10^{-3}$
No-Sp	0.870	0.803	0.530
Sp- $(\sqrt{M})$	0.868	0.814	0.531
Sp- $M$	0.860	0.829	0.597
Sp- $(\sqrt{M})$ -corr	0.868	0.833	0.580
Sp-LP-corr	0.882	0.806	0.572
HLLC	0.879	0.802	0.528

**Table 2:** Pressure Convergence Rate ( $L^1$  norm)

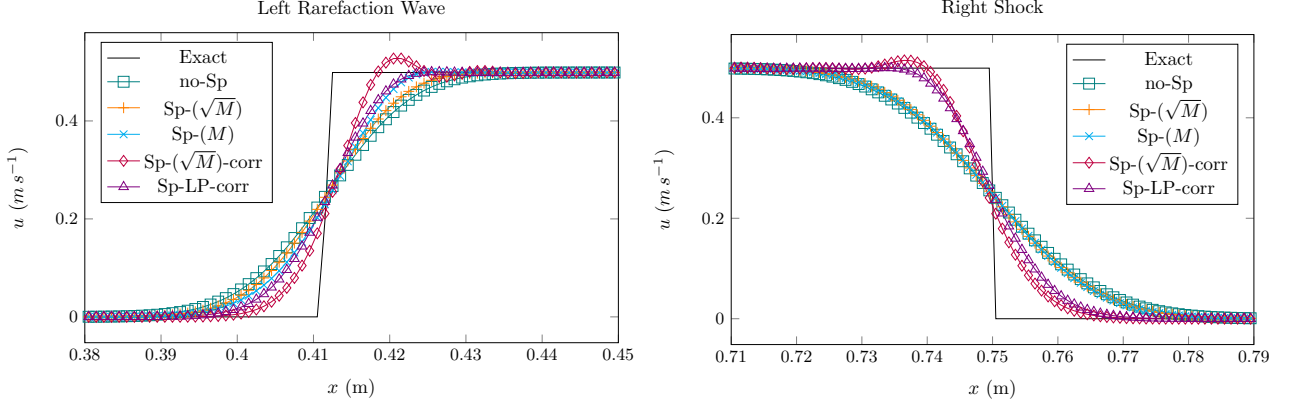
**Table 2** presents these orders of convergence for pressure. One can see that for *every* schemes, the order of convergence is depreciated as the maximal Mach number decreases. Indeed for pressure, it passes from 0.87 at  $M = 0.92$  (the expected order already obtained in [23]) to 0.82 at  $M = 9.5 \times 10^{-2}$  and 0.56 in the low-Mach number case. Seeking to confirm this behavior, the same test case has been computed using an HLLC-type scheme [43]; once again, at  $M = 4.2 \times 10^{-3}$  the convergence rate is 0.528. This unusual behavior can be summed up as: the lowest the maximal Mach number is, the slowest Godunov-like schemes are to reach the analytical compressible solution as the mesh is refined. The same convergence rate order can be found in [21] (page 20, Table 6.2) for a 1D double rarefaction wave problem performed on the Euler barotropic system with  $M \approx 3.1 \times 10^{-2}$ . The implicit-explicit AP scheme used to obtain this order is based on a Rusanov spatial discretization. Further investigations have to be undertaken in order to understand this numerical phenomenon.

Beyond convergence curves and rates, one must also have a look on the solution profile obtained with the different numerical schemes at a fixed mesh size. This is done in the next subsection.

#### 4.1.2 Solution Profiles

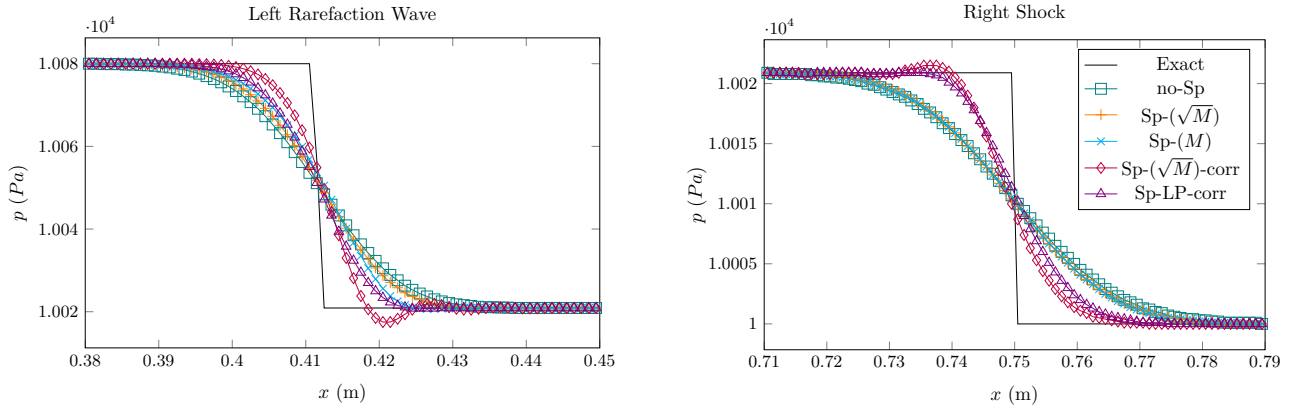
Figure 6 and Figure 7 show the velocity and pressure final profiles calculated with the different numerical solutions in the low-Mach regime. We only plot the left-going 1-rarefaction and the right-going 3-shock waves through which  $u$  and  $p$  change. Mesh is made of  $N_{cells} = 10^3$  cells. Let us point out that the stiffness of the rarefaction wave exact profile is only due to the fact that the width of the fan at a given instant  $t$  writes:  $|(u_{0,L} - c_{0,L}) - (u^* - c_L^*)| t = |c_L^* - c_{0,L} - u^*| t$ ; with  $u^*$  the intermediate analytical velocity and  $c_L^*$  the intermediate sound speed located at the left of the 2-contact discontinuity. Since the maximal Mach number is very low compared to one  $|u^*| \ll \min(c_{0,L}, c_L^*)$ , and since  $c_{0,L} \approx c_L^*$ , the width of the fan is approximately equal to  $|u^*| t$ . Thus, it is very small when acoustic time-scales such as  $0 < t < 10^{-3} s$  are considered.

One can notice that No-Sp is always the most diffusive scheme. Besides, the positive effect of the  $\mathcal{E}_0^n = \max(\mathcal{E}_{inf}, \min(M_{max}^n, 1))$  choice compared to  $\mathcal{E}_0^n = \max(\mathcal{E}_{inf}, \min(\sqrt{M_{max}^n}, 1))$  is exclusively located in the left rarefaction wave fan where the solution is continuous. In addition, No-Sp, Sp- $(\sqrt{M})$  and Sp- $(M)$  profiles overlap in the shock front region.



**Figure 6:** Velocity profile at  $M = 4.2 \times 10^{-3}$ , with  $N_{cells} = 10^3$

Eventually, the correction globally improves the computed solution accuracy. The Sp- $(\sqrt{M})$ -corr case produces profiles closer to the analytical solution than Sp-LP-corr at the cost of little overshoots in the tail of the left rarefaction wave and before the shock front.

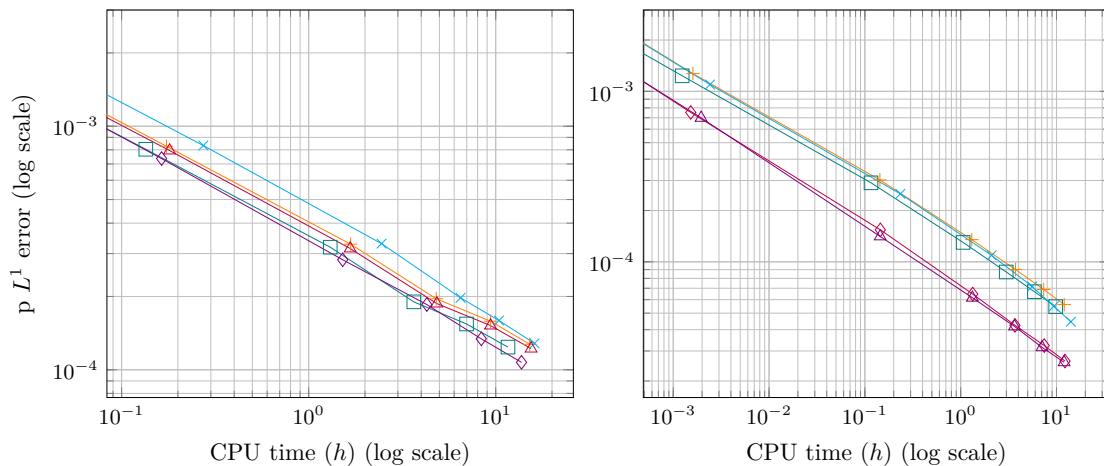


**Figure 7:** Pressure profile at  $M = 4.2 \times 10^{-3}$ , with  $N_{cells} = 10^3$

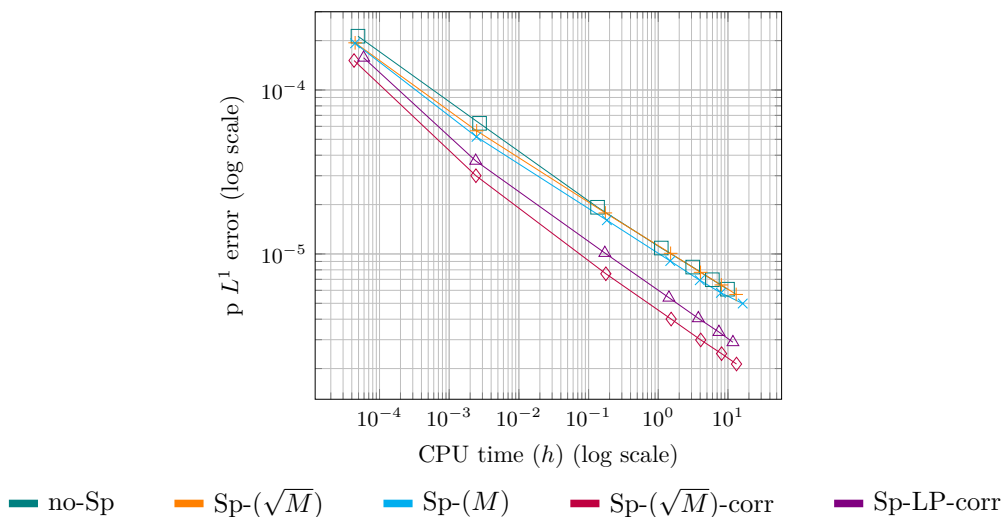
#### 4.1.3 Efficiency Curve

Computational cost at fixed accuracy level is now investigated. Figure 8 and Figure 9 describe the pressure efficiency curves of the different numerical schemes for the three Mach regimes. In the Mach one regime, every scheme seems to behave equivalently, the proposed weighted splitting approach requiring slightly more CPU time than No-Sp or Sp-LP-corr. When  $M = 9.5 \times 10^{-2}$ , the weighted

splitting approach is still slower than No-Sp, however the  $\theta$ -corrected schemes are clearly less time consuming, at fixed error than the other ones. Indeed Sp- $(\sqrt{M})$ -corr and Sp-LP-corr reach the precision of  $7 \times 10^{-5}$  in about one hour and a half whereas it requires six hours for No-Sp and more than seven hours for Sp- $(\sqrt{M})$ . Eventually, in the low-Mach case, Sp- $(\sqrt{M})$ -corr seems to produce better results than Sp-LP-corr. For a fixed precision of  $4 \times 10^{-6}$  the weighted splitting method needs about one hour and forty minutes whereas the Lagrange-Projection method requires a little less than three hours.



**Figure 8:** Pressure Efficiency Curves:  $M = 0.92$  (left),  $M = 9.5 \times 10^{-2}$  (right)



**Figure 9:** Pressure Efficiency Curves:  $M = 4.2 \times 10^{-3}$

## 4.2 Stiffened Gas Thermodynamics

In the above subsection, some elements seem to suggest that the weighted splitting approach produces satisfying results for a wide range of Mach number. The present method is notably able to capture strong shock waves associated with a sudden rise of the maximal Mach number. Nevertheless, as mentioned in point (II) in the introduction, we are interested in configurations where strong shock waves appear even if  $0 < M \ll 1$ . In Appendix A, the analytical solution of a symmetric double shock Riemann problem involving the Euler compressible system endowed with a stiffened gas thermodynamics (15) is derived. Starting with initial states of density  $\rho_0 = 10^3 \text{ kg.m}^{-3}$ , velocity  $|u_0| = 1 \text{ m.s}^{-1}$  and pressure  $p_0 = 3 \times 10^5 \text{ Pa}$ , one can analytically show that the non-dimensional pressure jump reads:

$$\frac{|p^* - p_0|}{p_0} = M_0 (1 + \alpha) \times O(1) \text{ w.r.t } M_0,$$

$$\text{with: } M_0 = \frac{|u_0|}{c_0}, \quad \alpha = \frac{P_\infty}{p_0}, \quad (77)$$

and  $p^*$  the intermediate pressure behind the shock fronts.

Then, one can notice that the stiffness of the non-dimensional thermodynamical coefficient  $\alpha$  can compensate the amplitude reduction effect of  $M_0$ . By choosing  $\gamma = 7.5$  and  $P_\infty = 3 \times 10^8 \text{ Pa}$ , one can obtain  $c_0 \approx 1.5 \times 10^3 \text{ m.s}^{-1}$  which is representative of liquid water at  $T_0 = 295 \text{ K}$ . A numerical application leads to:

$$M_0 \approx 7 \times 10^{-4},$$

$$\alpha = 10^3,$$

$$\frac{|p^* - p_0|}{p_0} \approx 5.2. \quad (78)$$

Hence, 15-bar amplitude shock waves are created while the Mach number is of order  $10^{-3}$ .

In the following, a shock tube Riemann problem using the above stiffened gas thermodynamics is tested. The initial conditions have been defined on Table 3, and we still have  $\gamma = 7.5$ ,  $P_\infty = 3 \times 10^8 \text{ Pa}$ . In this case, the maximal Mach number is about  $M_{max} \approx 4.6 \times 10^{-5}$ . Thus, we are still in a very low-Mach regime. The final time of the simulation is  $T_{end} = 1.58 \times 10^{-4} \text{ s}$ .

	Left state	Right state
$\rho \text{ (kg.m}^{-3}\text{)}$	$\rho_{0,L} = 10^3$	$\rho_{0,R} = 9 \times 10^2$
$u \text{ (m.s}^{-1}\text{)}$	$u_{0,L} = 0.$	$u_{0,R} = 0.$
$p \text{ (bar)}$	$p_{0,L} = 3$	$p_{0,R} = 1$

**Table 3:** Stiffened gas shock tube initial conditions

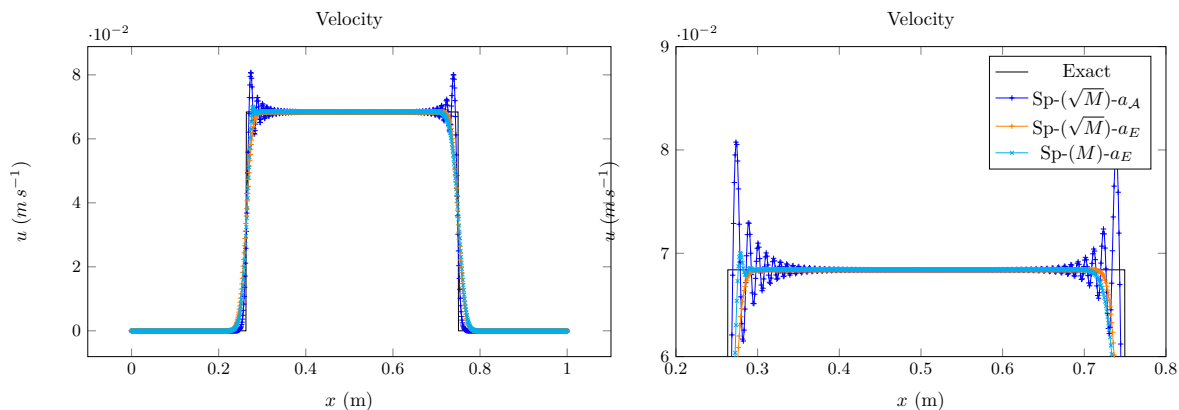
### 4.2.1 Acoustic Relaxation Coefficient Calibration

Let us go back to **Remark 4**. Testing a stiffened gas thermodynamics allows to compare the acoustic subcharacteristic condition (29b) derived from the weighted splitting approach and (31) obtained directly from the relaxation of the full Euler system and found in [25]. Let us recall that the inferior



bound of (29b) uses an artificial acoustic sound speed  $c_A$  whereas (31) is based on the physical sound speed  $c$ . In the previous ideal gas thermodynamics case  $c_A = \sqrt{(\gamma - 1)p/\rho} = \sqrt{(\gamma - 1)/\gamma} c$  and  $\sqrt{(\gamma - 1)/\gamma} \approx 0.53$  such that this non-physical acoustic celerity was of the same order that the real sound speed. However, with a stiffened gas thermodynamics,  $c_A$  does not change while  $c$  becomes  $\sqrt{\gamma(p + P_\infty)/\rho} \approx \sqrt{\gamma P_\infty/\rho}$ . Thus  $c_A/c \approx \sqrt{(\gamma - 1)/\gamma} \sqrt{p/P_\infty} \ll 1$ . One could wonder if considering the subcharacteristic condition (29b) based on a non-physical celerity rather than the one based on the real sound speed (31) has an effect on the overall scheme accuracy? So far, numerical arguments seem to go in favor of an acoustic relaxation coefficient based on the real sound speed. Indeed, Figure 10 shows two weighted splitting simulations of type Sp- $(\sqrt{M})$ . The first one, noted Sp- $(\sqrt{M})$ - $a_A$ , takes the subcharacteristic condition (29b) into account whereas the second one, Sp- $(\sqrt{M})$ - $a_E$ , involves (31). The mesh was composed of  $10^3$  cells.

It turns out that Sp- $(\sqrt{M})$ - $a_A$  produces non-physical oscillations inside the rarefaction fan and before the shock front. Things are even worse when Sp- $(M)$ - $a_A$  and Sp- $(M)$ - $a_E$  are compared. Indeed, even if the non-physical subcharacteristic condition (29b) is fulfilled, the amplitude of the spurious oscillations is such that pressure becomes negative after several timesteps. Simulation crashes because  $c_A$  becomes a complex number. On the contrary, Sp- $(M)$ - $a_E$  does not suffer from any oscillations or stability issues. Recall that the relaxation coefficient  $a_A$  multiplies the non-centered part of the acoustic momentum flux responsible for most of the numerical diffusion of the scheme. Hence, by considering subcharacteristic condition (31) rather than (29b) this coefficient has been considerably increased as well as the numerical diffusion coefficient. Non-physical oscillations are then removed.



**Figure 10:** Effect of the Estimation of the Relaxation Coefficient

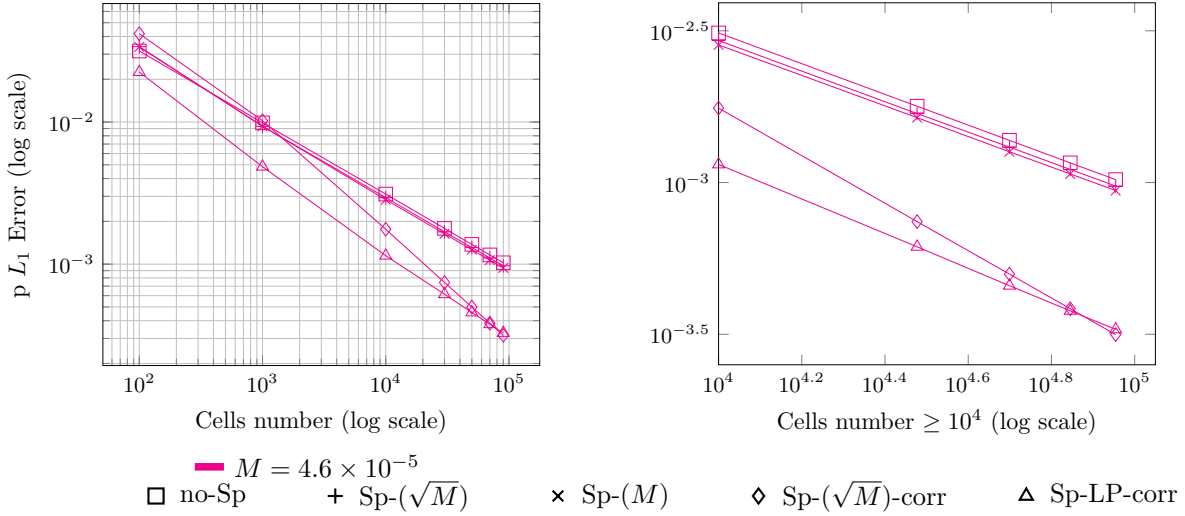
In the sequel, the acoustic relaxation coefficient  $a_A$  has been calculated using the physical sound speed:  $a_A > \rho c$ . The global CFL condition (52) is modified in consequence.

#### 4.2.2 Convergence Curves and Computed Solutions Profiles

Similarly to the ideal gas thermodynamics configuration, pressure convergence curve plotted on Figure 11 shows that the  $\theta$ -corrected schemes are the most accurate as the mesh is refined. However, one can notice that the Sp- $(\sqrt{M})$ -corr curve remains at the same level of accuracy than the non-corrected schemes until  $N_{cells} > 10^3$ . This can be explained by observing the solutions profiles drawn

on Figure 12. The correction centers the pressure flux since the Mach number is very small in every computational region. Hence, it triggers oscillations in areas where the solution is sharp. Such oscillations are present even in the case of Sp-LP-corr but their amplitude is smaller. In any case, the domain on which these oscillations are located as well as their relative height w.r.t the analytical solution are reduced as the mesh is refined. The present method is thus  $L^\infty$  stable. For  $N_{cells} \geq 3 \times 10^4$  the Sp- $(\sqrt{M})$ -corr becomes as accurate as the Sp-LP-corr one.

Once again let us point out an amazing numerical result already observed in the above ideal gas thermodynamics shock tube with  $M = 4.2 \times 10^{-3}$ . For every scheme (except Sp- $(\sqrt{M})$ -corr but additional points should be added to catch the asymptotic trend of its convergence curve), the observed pressure convergence rate written on **Table 4** is close to 0.5, the expected order for variables jumping through contact discontinuities; which was not the expected.



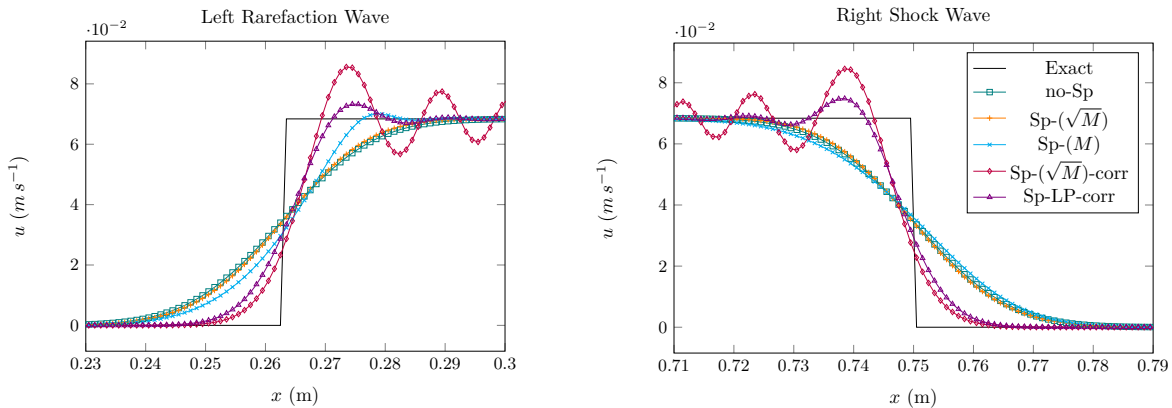
**Figure 11:** Pressure Convergence Curve

	$M = 4.6 \times 10^{-5}$
No-Sp	0.506
Sp- $(\sqrt{M})$	0.502
Sp- $M$	0.503
Sp- $(\sqrt{M})$ -corr	0.783
Sp-LP-corr	0.561

**Table 4:** Pressure Convergence Rate ( $L^1$  norm)

## Conclusion

In this work, a conservative fractional step approach based on a time-weighted splitting has been proposed for Euler-like models. The weighting parameter is proportional to the instantaneous maximal



**Figure 12:** Velocity Profile at  $M = 4.6 \times 10^{-5}$  with  $N_{cells} = 10^3$

flow Mach number  $M$ . When the latter takes high values the splitting allows to directly solve the overall Euler-like system in one step with an explicit time integration. Thus, shock waves are correctly captured without any diffusion or dispersion induced by the acoustic time-implicit discretization. On the contrary, if  $M$  is close to zero, convection is completely decoupled from acoustic. In that case, the acoustic discrete flux is modified by an anti-diffusive correction. If the fluid is endowed with a stiffened gas thermodynamics and if the solution is smooth as well as initially well-prepared, it results in a uniform truncation error with respect to  $M$ .

What is more, the Suliciu-like relaxation method used to discretize both convective and acoustic subsystems provides the density and internal energy positivity, in the case of an ideal gas thermodynamics, up to the introduction of new non-restrictive lower bounds for the relaxation constants. Besides, such relaxation constants encapsulate the thermodynamic nonlinearity and offer an easy way to deal with general equations of state.

The one-dimensional results performed with an ideal and a stiffened gas thermodynamics show that the time-explicit weighted splitting approach is as accurate and efficient as the time-explicit Lagrange-Projection method [11] for a wide range of Mach numbers. It can notably capture strong shock and rarefaction waves linked to a sudden rise of the Mach number of the flow. In the specific case where high amplitude shock waves appear even if  $0 < M \ll 1$ , the present method, completed with the anti-diffusive correction, remains  $L^\infty$  stable but suffer from stronger oscillations than the Lagrange-Projection method.

Besides, if one is interested in following the slow material waves of low-Mach number compressible flows when no fast transient phenomenon is present, the implicit-explicit version of the present approach can be relevant. Additional developments whose results are presented in [32, 31] deal with the adaptation of a time-implicit discretization technique proposed in [15, 11] to the present acoustic subsystem.

Among many areas of improvement, one would concern the definition of the discrete weighting parameter  $\mathcal{E}_0^n$ : in the same manner as for the relaxation constants, it could be transformed into a local weighting factor which would be only uniform for the Riemann problem solved at the interface. By doing so, the present weighted splitting approach could react to the spatial fluctuations of the flow Mach number and could improve even further the global accuracy of the method. Eventually, a reflec-

tion about the relevance of an extension of the present weighted splitting approach to homogeneous relaxed models will be undertaken.

## **Acknowledgements**

The first author received a financial support through the EDF-CIFRE contract 0561-2015. Computational facilities were provided by EDF.

The authors would also like to thank the reviewers who have spent time reading this paper. Indeed, the remarks and questions they have pointed out have considerably allowed the present work to be improved.

## Appendix A: A Symmetric Double Shock Riemann Problem with a Stiffened Gas Equation of State

Let us consider a fluid endowed with a stiffened gas equation of state described in (15). Let us define the 1D Riemann problem: The analytical solution is made of a left-going 1-shock, a steady 2-contact

	Left state	Right state
$\rho$ ( $kg.m^{-3}$ )	$\rho_{0,L} = \rho_0 = 10^3$	$\rho_{0,R} = \rho_0$
$u$ ( $m.s^{-1}$ )	$u_{0,L} = u_0 = 1.$	$u_{0,R} = -u_0$
$p$ (bar)	$p_{0,L} = p_0 = 3$	$p_{0,R} = p_0$

**Table 5:** Stiffened gas symmetric double shock initial conditions

discontinuity and a right-going 3-shock. The symmetry of the problem forces the intermediate velocity  $u^*$  related to the contact discontinuity speed to be equal to zero. This considerably simplifies the Rankine-Hugoniot relations. The intermediate pressure  $p^*$  and density  $\rho^*$  can be found analytically. Since  $u^* = 0$ , there is only one remaining Mach number that one can control through the initial conditions:  $M_0 = (|u_0|/c_0)$ . In the sequel, the analytical formula for the pressure jump  $|p^* - p_0| = p^* - p_0$  is derived.

Let us focus on the 3-shock Rankine-Hugoniot relations. Using the mass and the momentum equations, one obtains:

$$\begin{aligned}
 -\sigma [\rho_0 - \rho^*] + \left[ -\rho_0 u_0 - \rho^* \overbrace{u^*}^{=0} \right] &= 0 \Rightarrow \sigma = \frac{u_0}{\left(\frac{\tau_0}{\tau^*} - 1\right)}, \\
 \rho_0 u_0 \sigma + (\rho_0 u_0^2 + p_0 - p^*) &= 0 \Rightarrow \rho_0 u_0^2 \left( 1 + \frac{1}{\left(\frac{\tau_0}{\tau^*} - 1\right)} \right) + (p_0 - p^*) = 0,
 \end{aligned} \tag{79}$$

with:  $\tau = 1/\rho$ .

Recall that the energy equation reads:  $[\varepsilon] + \bar{p}[\tau] = 0$ ; with  $\bar{p} = (p^* + p_0)/2$  and  $\varepsilon = \frac{(p + \gamma P_\infty)\tau}{\gamma - 1}$ . Let us introduce  $P = p + P_\infty$  and  $P_0 = p_0 + P_\infty$ . As explained in subsection 1.2.2,  $P$  is the relevant variable in the case of a stiffened gas thermodynamics. It results that:

$$\begin{aligned}
 \tau^*(P^*) &= \tau_0 \frac{((P_0/\beta) + P^*)}{((P^*/\beta) + P_0)}, \\
 \text{with: } \beta &= \frac{\gamma - 1}{\gamma + 1}.
 \end{aligned} \tag{80}$$

After calculation,

$$1 + \frac{1}{\left(\frac{\tau_0}{\tau^*} - 1\right)} = 1 + \frac{\beta}{1 - \beta} \frac{((P_0/\beta) + P^*)}{(P^* - P_0)}. \tag{81}$$

When (81) is injected in (79) and after having multiplied by  $(P^* - P_0)$ , one obtains a second-order

polynomial function equation, namely:

$$\begin{aligned} X^2 - \frac{\gamma}{1-\beta} M_0^2 X - \gamma \frac{1+\beta}{1-\beta} M_0^2 &= 0, \\ \Leftrightarrow X^2 - \frac{\gamma(\gamma+1)}{2} M_0^2 X - \gamma^2 M_0^2 &= 0, \quad \text{with: } X = \frac{P^* - P_0}{P_0}. \end{aligned} \quad (82)$$

As  $P^* > P_0$ , we are looking for a strictly positive root, the solution writes:

$$\frac{P^* - P_0}{P_0} = M_0 \gamma \left( \frac{\gamma+1}{4} M_0 + \sqrt{1 + \frac{(\gamma+1)^2}{16} M_0^2} \right). \quad (83)$$

One can notice that if  $P_\infty = 0$  then,  $P^* = p^*$ ,  $P_0 = p_0$ , and equality (83) becomes similar to these obtained in [7] (page 845) for an isolated shock endowed with an ideal gas thermodynamics. Thus, in the case of an ideal gas thermodynamics:

$$\frac{p^* - p_0}{p_0} = M_0 \times O(1) \text{ w.r.t } M_0 \Rightarrow \lim_{M_0 \rightarrow 0} \frac{p^* - p_0}{p_0} = 0. \quad (84)$$

However, in the case of a stiffened gas thermodynamics, since  $P_0 = p_0 (1 + \alpha)$  with  $\alpha = (P_\infty/p_0)$ :

$$\begin{aligned} \frac{p^* - p_0}{p_0} &= M_0 (1 + \alpha) \gamma \left( \frac{\gamma+1}{4} M_0 + \sqrt{1 + \frac{(\gamma+1)^2}{16} M_0^2} \right), \\ &= M_0 (1 + \alpha) \times O(1) \text{ w.r.t } M_0. \end{aligned} \quad (85)$$

Here we can clearly see that, provided that the non-dimensional thermodynamical coefficient  $\alpha$  behaves as  $M_0^{-\delta}$  with  $\delta > 1$ , the stiffened gas thermodynamics is "stiff enough" to compensate the damping effect of  $M_0$  when  $M_0 \rightarrow 0$ . For example, if one considers  $\gamma = 7.5$  and  $P_\infty = 3 \times 10^8$  so that to obtain a speed of sound  $c_0 \approx 1.5 \times 10^3 \text{ m.s}^{-1}$  and a temperature of  $T_0 = 295 \text{ K}$ , a numerical calculation gives:

$$\begin{aligned} M_0 &\approx 7 \times 10^{-4}, \\ \alpha &= 10^3, \\ \frac{p^* - p_0}{p_0} &\approx 5.26. \end{aligned} \quad (86)$$

Hence, on this analytical solution got from the Euler compressible system endowed with a stiffened gas thermodynamics, a pressure jump of approximately 15 bars is created whereas the flow Mach number is of order  $10^{-3}$ .

## Appendix B: Subsystems Hyperbolicity

For the sake of simplicity, we prove **Proposition 1** in 1D. Let us consider the set of non conservative variables  $\mathbf{V} = [\rho, u, p]^T$ . If the solutions of subsystems  $\mathcal{C}$  and  $\mathcal{A}$  are smooth, one can rewrite them equivalently as:

$$\mathcal{C}^{NC} : \begin{cases} \partial_t \rho + u \partial_x \rho + \rho \partial_x u = 0, \\ \partial_t u + u \partial_x u + \frac{1}{\rho} \partial_x (\mathcal{E}_0^2(t) p) = 0, \\ \partial_t p + u \partial_x p + \rho c_C^2 \partial_x u = 0, \end{cases} \quad (87) \quad \mathcal{A}^{NC} : \begin{cases} \partial_t \rho = 0, \\ \partial_t u + \frac{1}{\rho} \partial_x ((1 - \mathcal{E}_0^2(t)) p) = 0, \\ \partial_t p + (1 - \mathcal{E}_0^2(t)) \rho c_A^2 \partial_x u = 0. \end{cases} \quad (88)$$

In variables  $\mathbf{V}$  the Jacobian matrices of subsystems  $\mathcal{C}^{NC}$  and  $\mathcal{A}^{NC}$  are:

$$\mathcal{C}^{NC} : \begin{bmatrix} u & \rho & 0 \\ 0 & u & \mathcal{E}_0^2/\rho \\ 0 & \rho c_C^2 & u \end{bmatrix} \quad (89) \quad \mathcal{A}^{NC} : \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & (1 - \mathcal{E}_0^2)/\rho \\ 0 & (1 - \mathcal{E}_0^2) \rho c_A^2 & 0 \end{bmatrix} \quad (90)$$

Supposing that  $c_C^2 \geq 0$  and  $c_A^2 \geq 0$ , the eigenvalues and eigenvectors can be easily obtained and read:

$$\mathcal{C}^{NC} : \begin{cases} \lambda_1^C = u - \mathcal{E}_0 c_C, \\ \lambda_2^C = u, \\ \lambda_3^C = u + \mathcal{E}_0 c_C, \end{cases} \quad \mathbf{r}_1^C = \begin{bmatrix} \rho \\ -\mathcal{E}_0 c_C \\ \rho c_C^2 \end{bmatrix}, \quad \mathbf{r}_2^C = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{r}_3^C = \begin{bmatrix} \rho \\ +\mathcal{E}_0 c_C \\ \rho c_C^2 \end{bmatrix}, \quad (91)$$

$$\mathcal{A}^{NC} : \begin{cases} \lambda_1^A = -(1 - \mathcal{E}_0^2) c_A, \\ \lambda_2^A = 0, \\ \lambda_3^A = (1 - \mathcal{E}_0^2) c_A, \end{cases} \quad \mathbf{r}_1^A = \begin{bmatrix} 0 \\ 1 \\ -\rho c_A \end{bmatrix}, \quad \mathbf{r}_2^A = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{r}_3^A = \begin{bmatrix} 0 \\ 1 \\ \rho c_A \end{bmatrix}. \quad (92)$$

Then, one can notice that, for the two subsystems, the 1-field and 3-field are genuinely non linear whereas the 2-field is linearly degenerate. Let us now study the sufficient conditions for which  $c_C^2 \geq 0$  and  $c_A^2 \geq 0$ . Consider the ideal gas thermodynamics presented in equation (14). Then:

$$\begin{aligned} c_C^2 &= (1 + \mathcal{E}_0^2(\gamma - 1)) \frac{p}{\rho} = \gamma_{\mathcal{E}_0} \frac{p}{\rho}, \\ c_A^2 &= (\gamma - 1) \frac{p}{\rho}. \end{aligned} \quad (93)$$

Since  $\mathcal{E}_0^2 \in [0, 1]$ ,  $\gamma_{\mathcal{E}_0} \in [1, \gamma]$ . What is more, by definition of the ideal gas phase-space (14), the pressure variable  $p$  is positive. Thus, in case of an ideal gas thermodynamics,  $c_C^2$  and  $c_A^2$  are naturally positive. On the contrary, when the stiffened gas thermodynamics is at stake one obtains:

$$\begin{aligned} c_C^2 &= \frac{\gamma_{\mathcal{E}_0} p + \gamma P_\infty}{\rho}, \\ c_A^2 &= (\gamma - 1) \frac{p}{\rho}. \end{aligned} \quad (94)$$

The stiffened gas phase-space (15) ensures that  $p > -P_\infty \Rightarrow \gamma_{\mathcal{E}_0} p + \gamma P_\infty > (\gamma - \gamma_{\mathcal{E}_0}) P_\infty$ . And  $\gamma - \gamma_{\mathcal{E}_0}$  is positive. Once again,  $c_C^2$  is positive without any condition. However  $c_A^2 \geq 0 \Leftrightarrow p \geq 0$  which is not guaranteed in the stiffened gas phase space since  $p + P_\infty > 0$ .

## Appendix C: Phase Space Invariance of the Continuous Subsystems

Consider  $\Omega$  a bounded spatial domain of  $\mathbb{R}^d$ ,  $d \in \{1, 2, 3\}$  which boundary is  $\partial\Omega$ .

The objective here is to study the positivity of  $\varepsilon$  (respectively  $P = \rho\varepsilon - P_\infty$ ) in the case of an ideal gas thermodynamics (respectively a stiffened gas thermodynamics). According to (14), (15), it corresponds to the phase-space invariance for both ideal and stiffened gas thermodynamics.

### Ideal Gas Thermodynamics

The specific internal energy of both subsystems verifies the following PDEs:

$$\begin{cases} \partial_t \varepsilon + \mathbf{u} \cdot \nabla \varepsilon + \mathcal{E}_0^2(t) \frac{p}{\rho} \nabla \cdot \mathbf{u} = 0, & (\mathcal{C}) \\ \partial_t \varepsilon + (1 - \mathcal{E}_0^2(t)) \frac{p}{\rho} \nabla \cdot \mathbf{u} = 0, & (\mathcal{A}) \end{cases} \quad (95)$$

which can also be rewritten:

$$\begin{cases} \partial_t \varepsilon + \mathbf{u} \cdot \nabla \varepsilon + \mathcal{E}_0^2(t) (\gamma - 1) \varepsilon \nabla \cdot \mathbf{u} = 0, & (\mathcal{C}) \\ \partial_t \varepsilon + (1 - \mathcal{E}_0^2(t)) (\gamma - 1) \varepsilon \nabla \cdot \mathbf{u} = 0, & (\mathcal{A}) \end{cases} \quad (96)$$

because  $p/\rho = (\gamma - 1)\varepsilon$ .

Define  $\varepsilon^- = \frac{\varepsilon - |\varepsilon|}{2}$  the negative part of the specific internal energy. Consider the following hypothesis about the solution's smoothness and the initial and boundary conditions:

$$\forall t \geq 0, \mathbf{u}(t, \cdot) \in L^\infty(\Omega), \nabla \cdot \mathbf{u}(t, \cdot) \in L^\infty(\Omega), \quad (97a)$$

$$\forall t \geq 0, \varepsilon(t, \cdot) \in L^\infty(\Omega), \nabla \varepsilon(t, \cdot) \in L^\infty(\Omega)^d, \quad (97b)$$

$$\forall \mathbf{x} \in \Omega, \varepsilon(t = 0, \mathbf{x}) > 0 \Leftrightarrow \varepsilon^-(0, \mathbf{x}) = 0, \quad (97c)$$

$$\varepsilon|_{\partial\Omega} \geq 0 \text{ if } \mathbf{u} \cdot \mathbf{n}|_{\partial\Omega} \leq 0. \quad (97d)$$

By multiplying equation (95) by  $\varepsilon^-$  and integrating over  $\Omega$  one obtains:

$$\begin{cases} \frac{d}{dt} \int_{\Omega} \frac{(\varepsilon^-)^2}{2} d\Omega + \int_{\Omega} \mathbf{u} \cdot \nabla \frac{(\varepsilon^-)^2}{2} d\Omega + \int_{\Omega} (\gamma - 1) \mathcal{E}_0^2(t) (\varepsilon^-)^2 \nabla \cdot \mathbf{u} d\Omega = 0 & (\mathcal{C}), \\ \frac{d}{dt} \int_{\Omega} \frac{(\varepsilon^-)^2}{2} d\Omega + \int_{\Omega} (\gamma - 1) (1 - \mathcal{E}_0^2(t)) (\varepsilon^-)^2 \nabla \cdot \mathbf{u} d\Omega = 0 & (\mathcal{A}). \end{cases} \quad (98)$$

By using Green's formula, the above equations can be transformed into:

$$\begin{cases} \frac{d}{dt} \frac{\|\varepsilon^-\|_{L^2}^2}{2} = \int_{\Omega} \nabla \cdot \mathbf{u} \left( \frac{1}{2} - (\gamma - 1) \mathcal{E}_0^2(t) \right) (\varepsilon^-)^2 d\Omega - \int_{\partial\Omega} \frac{(\varepsilon^-)^2}{2} \mathbf{u} \cdot \mathbf{n} d\Gamma & (\mathcal{C}), \\ \frac{d}{dt} \frac{\|\varepsilon^-\|_{L^2}^2}{2} = - \int_{\Omega} \nabla \cdot \mathbf{u} (\gamma - 1) (1 - \mathcal{E}_0^2(t)) (\varepsilon^-)^2 d\Omega & (\mathcal{A}). \end{cases} \quad (99)$$

Because of the admissible inlet boundary condition (97d),  $-\int_{\partial\Omega} \frac{(\varepsilon^-)^2}{2} \mathbf{u} \cdot \mathbf{n} d\Gamma$  is always negative so that we can derive the following inequalities:



$$\left\{ \begin{array}{l} \frac{d}{dt} \|\varepsilon^-\|_{L^2}^2 \leq \sup_{\Omega} \overbrace{|\nabla \cdot \mathbf{u} (1 - 2 \mathcal{E}_0^2(t) (\gamma - 1))|}^{L_{\mathcal{C}}(t)} \|\varepsilon^-\|_{L^2}^2 \quad (\mathcal{C}), \\ \frac{d}{dt} \|\varepsilon^-\|_{L^2}^2 \leq \sup_{\Omega} \overbrace{|2 \nabla \cdot \mathbf{u} (1 - \mathcal{E}_0^2(t) (\gamma - 1))|}^{L_{\mathcal{A}}(t)} \|\varepsilon^-\|_{L^2}^2 \quad (\mathcal{A}). \end{array} \right. \quad (100)$$

Thus, because of Grönwall's lemma:

$$\left\{ \begin{array}{l} \|\varepsilon^-\|_{L^2}^2(t) \leq \|\varepsilon^-\|_{L^2}^2(0) e^{\int_0^t L_{\mathcal{C}}(s) ds} = 0 \Rightarrow \|\varepsilon^-\|_{L^2}^2(t) = 0 \quad (\mathcal{C}), \\ \|\varepsilon^-\|_{L^2}^2(t) \leq \|\varepsilon^-\|_{L^2}^2(0) e^{\int_0^t L_{\mathcal{A}}(s) ds} = 0 \Rightarrow \|\varepsilon^-\|_{L^2}^2(t) = 0 \quad (\mathcal{A}). \end{array} \right. \quad (101)$$

One can notice that, beyond hypothesis presented in (97), a sufficient condition to derive inequalities (100) is  $\frac{p}{\rho} = K \varepsilon$  with  $K$  a bounded function on  $\Omega$ . Indeed, as previously seen, it allows to control the term  $\int_{\Omega} \varepsilon^- \frac{p}{\rho} \nabla \cdot \mathbf{u} d\Omega$  by  $\sup_{\Omega} |K \nabla \cdot \mathbf{u}| \|\varepsilon^-\|_{L^2}^2$ .

### Stiffened Gas Thermodynamics

When a stiffened gas thermodynamics defined by (15a) and (15b) is at stake, one is interested in the positivity of  $P = \rho \varepsilon - P_{\infty}$ . Such a variable follows the PDEs:

$$\left\{ \begin{array}{l} \partial_t P + \nabla \cdot (P \mathbf{u}) + \mathcal{E}_0^2(t) (\gamma - 1) P \nabla \cdot \mathbf{u} + P_{\infty} (1 - \mathcal{E}_0^2(t)) \nabla \cdot \mathbf{u} = 0, \quad (\mathcal{C}) \\ \partial_t P + (1 - \mathcal{E}_0^2(t)) (\gamma - 1) P \nabla \cdot \mathbf{u} - P_{\infty} (1 - \mathcal{E}_0^2(t)) \nabla \cdot \mathbf{u} = 0. \quad (\mathcal{A}) \end{array} \right. \quad (102)$$

By doing exactly the same kind of hypothesis (but replacing  $\varepsilon$  by  $P$ ) and calculations than for the ideal case, one can obtain:

$$\left\{ \begin{array}{l} \frac{d}{dt} \|P^-\|_{L^2}^2 \leq (2 \mathcal{E}_0^2 (\gamma - 1) + 1) \sup_{\Omega} |\nabla \cdot \mathbf{u}| \|P^-\|_{L^2}^2 - (1 - \mathcal{E}_0^2(t)) P_{\infty} \int_{\Omega} P^- \nabla \cdot \mathbf{u} d\Omega, \quad (\mathcal{C}) \\ \frac{d}{dt} \|P^-\|_{L^2}^2 \leq 2 (1 - \mathcal{E}_0^2) (\gamma - 1) \sup_{\Omega} |\nabla \cdot \mathbf{u}| \|P^-\|_{L^2}^2 + (1 - \mathcal{E}_0^2(t)) P_{\infty} \int_{\Omega} P^- \nabla \cdot \mathbf{u} d\Omega. \quad (\mathcal{A}) \end{array} \right. \quad (103)$$

Assume that there exists a function  $K \in L^{\infty}(\Omega)$  such that:

$$P^{\infty} = K P. \quad (104)$$

Then, it is possible to control the term  $P_{\infty} \int_{\Omega} P^- \nabla \cdot \mathbf{u} d\Omega$  with  $\|P^-\|_{L^2}^2$ . Inequalities (103) turn into:

$$\left\{ \begin{array}{l} \frac{d}{dt} \|P^-\|_{L^2}^2 \leq \sup_{\Omega} |((2 \mathcal{E}_0^2 (\gamma - 1) + 1) - K (1 - \mathcal{E}_0^2)) \nabla \cdot \mathbf{u}| \|P^-\|_{L^2}^2, \quad (\mathcal{C}) \\ \frac{d}{dt} \|P^-\|_{L^2}^2 \leq \sup_{\Omega} |(1 - \mathcal{E}_0^2) (2(\gamma - 1) + K) \nabla \cdot \mathbf{u}| \|P^-\|_{L^2}^2, \quad (\mathcal{A}) \end{array} \right. \quad (105)$$

and the Grönwall's lemma can be applied so that to obtain  $\|P^-\|_{L^2}^2(t) = 0$ . However, in the case of a stiffened gas thermodynamics  $P_{\infty}/P = \frac{(\gamma-1)P_{\infty}}{(p+P_{\infty})}$  and  $p$  is allowed to tend towards  $-P_{\infty}$ . Thus, hypothesis (104) does not hold *a priori* and we cannot ensure the positivity of  $P$  unless  $P_{\infty} = 0$  (ideal gas thermodynamics) or  $\mathcal{E}_0 = 1$  (splitting not triggered).

## Appendix D: Positivity of the Discrete Intermediate Density

Consider the Riemann problem presented on Figure 1 related to the convective subsystem. It produces intermediate states described in relations (38) and (39). Let us find a sufficient condition on the subcharacteristic coefficient  $a_C$  so that the intermediate densities  $\rho_{k,C}^*$ ,  $k \in \{L, R\}$  are positive.

$$\begin{aligned}
\rho_{k,C}^* \geq 0 &\Leftrightarrow \tau_{k,C}^* \geq 0, \\
&\Leftrightarrow \tau_k + \frac{(-1)^{i_k+1}}{\mathcal{E}_0 a_C} (u_C^* - u_k) \geq 0, \\
&\Leftrightarrow a_C^2 + \frac{\rho_k (u_R - u_L)}{2\mathcal{E}_0} a_C + \frac{(-1)^{i_k} \rho_k (p_R - p_L)}{2} \geq 0.
\end{aligned} \tag{106}$$

The second order polynomial function admits real roots if and only if  $\Delta_k^\rho \equiv \frac{\rho_k (u_R - u_L)^2}{8\mathcal{E}_0^2} + (-1)^{i_k+1} (p_R - p_L) \geq 0$ . Let us notice that  $\Delta_L^\rho < 0 \Rightarrow \Delta_R^\rho > 0$  and conversely. In that case the polynomial constraint (106) related to  $\Delta_L^\rho$  is automatically verified. Thus, consider the most demanding case where  $\Delta_L^\rho \geq 0$  and  $\Delta_R^\rho \geq 0$ , namely:

$$-\frac{\rho_L (u_R - u_L)^2}{8} \leq \mathcal{E}_0^2 (p_R - p_L) \leq \frac{\rho_R (u_R - u_L)^2}{8}. \tag{107}$$

If  $u_L \neq u_R$ , inequality (107) holds easily with low-Mach flows when  $\mathcal{E}_0$  tends toward zero.

Let us define  $a_k^\rho$ ,  $k \in \{L, R\}$  the highest roots related to the above polynomial functions:

$$\begin{aligned}
a_L^\rho &\equiv \frac{1}{2} \left( -\frac{\rho_L (u_R - u_L)}{2\mathcal{E}_0} + \sqrt{\frac{\rho_L^2 (u_R - u_L)^2}{4\mathcal{E}_0^2} + 2\rho_L (p_R - p_L)} \right), \\
a_R^\rho &\equiv \frac{1}{2} \left( -\frac{\rho_R (u_R - u_L)}{2\mathcal{E}_0} + \sqrt{\frac{\rho_R^2 (u_R - u_L)^2}{4\mathcal{E}_0^2} - 2\rho_R (p_R - p_L)} \right).
\end{aligned} \tag{108}$$

By noticing that  $\forall A \geq 0, -A + \sqrt{A^2 + B} > 0 \Leftrightarrow B > 0$ , one can build the following table which gives the sign of  $a_L^\rho$  and  $a_R^\rho$ : In practice, when either  $a_L^\rho$  or  $a_R^\rho$  are positive, we add it as an additional

	$u_R > u_L$	$u_R < u_L$
$p_R > p_L$	$a_L^\rho > 0, a_R^\rho < 0$	$a_L^\rho > 0, a_R^\rho > 0$
$p_R < p_L$	$a_L^\rho < 0, a_R^\rho > 0$	$a_L^\rho > 0, a_R^\rho > 0$

**Table 6:** Positivity Domain of  $a_L^\rho$  and  $a_R^\rho$

constraint into the subcharacteristic condition (29a) leading to the modified subcharacteristic condition (55).

The non-dimensional expressions of  $a_L^\rho$  and  $a_R^\rho$  are:

$$\begin{aligned} a_L^\rho &\equiv \frac{1}{2} \left( -\frac{M}{\mathcal{E}_0} \frac{\rho_L (u_R - u_L)}{2} + \sqrt{\left(\frac{M}{\mathcal{E}_0}\right)^2 \frac{\rho_L^2 (u_R - u_L)^2}{4} + 2 \rho_L (p_R - p_L)} \right), \\ a_R^\rho &\equiv \frac{1}{2} \left( -\frac{M}{\mathcal{E}_0} \frac{\rho_R (u_R - u_L)}{2} + \sqrt{\left(\frac{M}{\mathcal{E}_0}\right)^2 \frac{\rho_R^2 (u_R - u_L)^2}{4} - 2 \rho_R (p_R - p_L)} \right). \end{aligned} \quad (109)$$

Thus, if  $\mathcal{E}_0$  is proportional to the Mach number as defined in (49), the above non-dimensional roots are of order one.

Concerning the equivalence between the intermediate density positivity and the ordering of the eigenvalues of subsystem  $\mathcal{C}^\mu$ , one can notice that:

$$\begin{aligned} u_L - \mathcal{E}_0 a_C \tau_L &\leq u_C^*, \\ \Leftrightarrow 0 &\leq \mathcal{E}_0 a_C \left( \tau_L + \frac{1}{\mathcal{E}_0 a_C} (u_C^* - u_L) \right), \\ \Leftrightarrow 0 &\leq \tau_{L,C}^*. \end{aligned} \quad (110)$$

By doing the same calculation, one can see that  $u_C^* \leq u_R + \mathcal{E}_0 a_C \tau_R \Leftrightarrow \tau_{R,C}^* \geq 0$ .

Finally, let us recall that, in the acoustic Riemann problem presented on Figure 2,  $\rho_{L,\mathcal{A}}^* = \rho_L$  and  $\rho_{R,\mathcal{A}}^* = \rho_R$ . The intermediate densities are then already positive. No additional constraint on  $a_{\mathcal{A}}$  needs to be provided in order to preserve the density positivity.

## Appendix E: Positivity of the Discrete Intermediate Internal Energy

For an ideal gas thermodynamics, the specific internal energy is supposed to remain positive throughout space and time under admissible boundary conditions. Here a sufficient condition under the relaxation constants  $a_C$  and  $a_{\mathcal{A}}$  is looked for in order to obtain the positivity of the intermediate internal energies produced by the Riemann problems described on Figure 1 and Figure 2.

### Relaxed Convective Subsystem

Once again, for the Riemann problem related to the convective relaxed subsystem  $\mathcal{C}$ , the specific internal energy reads:

$$\begin{aligned} \varepsilon_{k,C}^* &= e_{k,C}^* - \frac{(u_C^*)^2}{2}, \\ &= e_k - \frac{(u_C^*)^2}{2} + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} (\Pi_C^* u_C^* - \Pi_k u_k), \\ &= \varepsilon_k + \frac{u_k^2 - (u_C^*)^2}{2} + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} (\Pi_C^* u_C^* - \Pi_k u_k), \\ &= \varepsilon_k + \frac{u_k^2 + (u_C^*)^2}{2} + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} u_C^* \left( \Pi_C^* + \frac{(-1)^{i_k+1} a_C}{\mathcal{E}_0} u_C^* \right) - \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} \Pi_k u_k. \end{aligned} \quad (111)$$

By combining,  $u + (-1)^{i_k} \mathcal{E}_0 a_C \tau$  and  $\Pi + (a_C)^2 \tau$  which both are 1-Riemann invariants of subsystem  $\mathcal{C}^\mu$ , one can build a new one:  $\Pi + \frac{(-1)^{i_k+1} a_C}{\mathcal{E}_0} u$ . Then, one can simplify the above expression of  $\varepsilon_{k,C}^*$ , namely:

$$\begin{aligned} \varepsilon_{k,C}^* &= \varepsilon_k + \frac{u_k^2 + (u_C^*)^2}{2} + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} u_C^* \left( \Pi_k + \frac{(-1)^{i_k+1} a_C}{\mathcal{E}_0} u_k \right) - \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} \Pi_k u_k, \\ &= \varepsilon_k + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} \Pi_k (u_C^* - u_k) + \frac{(u_C^* - u_k)^2}{2}. \end{aligned} \quad (112)$$

Thus, a sufficient condition which would guarantee that  $\forall k \in \{L, R\}$ ,  $\varepsilon_{k,C}^* \geq 0$  is:

$$\varepsilon_k + \mathcal{E}_0 \frac{(-1)^{i_k}}{a_C} p_k (u_C^* - u_k) \geq 0 \Leftrightarrow a_C^2 - \mathcal{E}_0 \rho_k^\varepsilon \frac{(u_R - u_L)}{2} a_C + (-1)^{i_k+1} \mathcal{E}_0^2 \rho_k^\varepsilon \frac{(p_R - p_L)}{2} \geq 0, \quad (113)$$

with  $\rho_k^\varepsilon = \frac{p_k}{\varepsilon_k}$ , and considering that  $p_k = \Pi_k$ . Inequality (113) is very similar to the one obtained for the intermediate density positivity. The most demanding case is the one where  $\forall k \in \{L, R\}$ ,  $\Delta_k^\varepsilon \equiv \rho_k^\varepsilon \frac{(u_R - u_L)^2}{8} + (-1)^{i_k} (p_R - p_L) \geq 0$ :

$$-\frac{\rho_R^\varepsilon (u_R - u_L)^2}{8} \leq p_R - p_L \leq \frac{\rho_L^\varepsilon (u_R - u_L)^2}{8}. \quad (114)$$

Once again the highest roots related to the polynomial functions written in (113) are:

$$\begin{aligned} a_{\mathcal{C},L}^\varepsilon &\equiv \frac{\mathcal{E}_0}{2} \left( \frac{\rho_L^\varepsilon (u_R - u_L)}{2} + \sqrt{\frac{(\rho_L^\varepsilon)^2 (u_R - u_L)^2}{4} - 2 \rho_L^\varepsilon (p_R - p_L)} \right), \\ a_{\mathcal{C},R}^\varepsilon &\equiv \frac{\mathcal{E}_0}{2} \left( \frac{\rho_R^\varepsilon (u_R - u_L)}{2} + \sqrt{\frac{(\rho_R^\varepsilon)^2 (u_R - u_L)^2}{4} + 2 \rho_R^\varepsilon (p_R - p_L)} \right). \end{aligned} \quad (115)$$

The sign of  $a_{\mathcal{C},L}^\varepsilon$  and  $a_{\mathcal{C},R}^\varepsilon$  is given by the following table: The non-dimensional version of these roots

	$u_R > u_L$	$u_R < u_L$
$p_R > p_L$	$a_{\mathcal{C},L}^\varepsilon > 0, a_{\mathcal{C},R}^\varepsilon > 0$	$a_{\mathcal{C},L}^\varepsilon < 0, a_{\mathcal{C},R}^\varepsilon > 0$
$p_R < p_L$	$a_{\mathcal{C},L}^\varepsilon > 0, a_{\mathcal{C},R}^\varepsilon > 0$	$a_{\mathcal{C},L}^\varepsilon > 0, a_{\mathcal{C},R}^\varepsilon < 0$

**Table 7:** Positivity Domain of  $a_{\mathcal{C},L}^\varepsilon$  and  $a_{\mathcal{C},R}^\varepsilon$

reads:

$$\begin{aligned} a_{\mathcal{C},L}^\varepsilon &\equiv \frac{\mathcal{E}_0}{2} \left( \frac{M \rho_L^\varepsilon (u_R - u_L)}{2} + \sqrt{\frac{M^2 (\rho_L^\varepsilon)^2 (u_R - u_L)^2}{4} - 2 \rho_L^\varepsilon (p_R - p_L)} \right), \\ a_{\mathcal{C},R}^\varepsilon &\equiv \frac{\mathcal{E}_0}{2} \left( \frac{M \rho_R^\varepsilon (u_R - u_L)}{2} + \sqrt{\frac{M^2 (\rho_R^\varepsilon)^2 (u_R - u_L)^2}{4} + 2 \rho_R^\varepsilon (p_R - p_L)} \right). \end{aligned} \quad (116)$$

Then, contrary to the non-dimensional roots involved in the intermediate density positivity they are of order  $O(\mathcal{E}_0)$ . When either  $a_{\mathcal{C},L}^\varepsilon$  or  $a_{\mathcal{C},R}^\varepsilon$  are positive, they are injected in the subcharacteristic condition (29a).

## Relaxed Acoustic Subsystem

The acoustic relaxed subsystem  $\mathcal{A}^\mu$  also produces intermediate specific internal energies  $\varepsilon_{k,\mathcal{A}}^* = e_{k,\mathcal{A}}^* - \frac{(u_{\mathcal{A}}^*)^2}{2} = \varepsilon_k + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} \Pi_k (u_{\mathcal{A}}^* - u_k) + \frac{(u_{\mathcal{A}}^* - u_k)^2}{2}$ . The proof is similar to the one done for the convective relaxed subsystem. Indeed:

$$\begin{aligned}
\varepsilon_{k,\mathcal{A}}^* &= e_{k,\mathcal{A}}^* - \frac{(u_{\mathcal{A}}^*)^2}{2}, \\
&= e_k - \frac{(u_{\mathcal{A}}^*)^2}{2} + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} (\Pi_{\mathcal{A}}^* u_{\mathcal{A}}^* - \Pi_k u_k), \\
&= \varepsilon_k + \frac{u_k^2 - (u_{\mathcal{A}}^*)^2}{2} + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} (\Pi_{\mathcal{A}}^* u_{\mathcal{A}}^* - \Pi_k u_k), \\
&= \varepsilon_k + \frac{u_k^2 + (u_{\mathcal{A}}^*)^2}{2} + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} u_{\mathcal{A}}^* (\Pi_{\mathcal{A}}^* + (-1)^{i_k+1} a_{\mathcal{A}} u_{\mathcal{A}}^*) - \frac{(-1)^{i_k}}{a_{\mathcal{A}}} \Pi_k u_k.
\end{aligned} \tag{117}$$

Recall that  $\omega_{\mathcal{A}}^- = u - \frac{\Pi}{a_{\mathcal{A}}}$  (respectively  $\omega_{\mathcal{A}}^+ = u + \frac{\Pi}{a_{\mathcal{A}}}$ ) introduced in **Subsection 2.2.2** is 1-Riemann invariant (respectively a 4-Riemann invariant). Then, one can replace  $\Pi_{\mathcal{A}}^* + (-1)^{i_k+1} a_{\mathcal{A}} u_{\mathcal{A}}^*$  by  $\Pi_k + (-1)^{i_k+1} a_{\mathcal{A}} u_k$ . Hence, equality (117) is equivalent to:

$$\varepsilon_{k,\mathcal{A}}^* = \varepsilon_k + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} \Pi_k (u_{\mathcal{A}}^* - u_k) + \frac{(u_{\mathcal{A}}^* - u_k)^2}{2}. \tag{118}$$

Thus, a sufficient condition which would guarantee that  $\forall k \in \{L, R\}$ ,  $\varepsilon_{k,\mathcal{A}}^* \geq 0$  is:

$$\varepsilon_k + \frac{(-1)^{i_k}}{a_{\mathcal{A}}} p_k (u_{\mathcal{A}}^* - u_k) \geq 0 \Leftrightarrow a_{\mathcal{A}}^2 - \rho_k^\varepsilon \frac{(u_R - u_L)}{2} a_{\mathcal{A}} + (-1)^{i_k+1} \rho_k^\varepsilon \frac{(p_R - p_L)}{2} \geq 0. \tag{119}$$

Sufficient conditions allowing to guarantee the intermediate specific energy positivity turns into the positivity of two polynomial functions of order two in  $a_{\mathcal{A}}$ . The most demanding case corresponds exactly to inequalities (114). Finally the roots above which the relaxation coefficient has to be are:

$$a_{\mathcal{A},L}^\varepsilon \equiv \frac{a_{\mathcal{C},L}^\varepsilon}{\mathcal{E}_0}; \quad a_{\mathcal{A},R}^\varepsilon \equiv \frac{a_{\mathcal{C},R}^\varepsilon}{\mathcal{E}_0}. \tag{120}$$

Since for  $k \in \{L, R\}$ ,  $a_{\mathcal{C},k}^\varepsilon = O(\mathcal{E}_0)$ , the non-dimensional expressions of  $a_{\mathcal{A},k}^\varepsilon$  are of order one. One can notice that, in case of low-Mach flows, the constraint imposed by the relaxation convective subsystem on the specific internal energy positivity is negligible compared to the one of the relaxed acoustic subsystem.

## Appendix F: Subcharacteristic Conditions for the Subsystems

The proof written below is only formal. It aims at exhibiting subcharacteristic conditions under which the relaxation subsystems contain diffusive operators. The latter would avoid instabilities which could prevent the convergence of the relaxation subsystems  $\mathcal{C}^\mu$  and  $\mathcal{A}^\mu$  towards  $\mathcal{C}$  and  $\mathcal{A}$ .

## Relaxed Convective Subsystem

Consider the relaxed convective subsystem  $\mathcal{C}^\mu$ :

$$\mathcal{C}^\mu : \begin{cases} \partial_t \rho + \partial_x (\rho u) = 0, \\ \partial_t \rho u + \partial_x (\rho u^2) + \partial_x (\mathcal{E}_0^2(t) \Pi) = 0, \\ \partial_t \rho e + \partial_x ((\rho e + \mathcal{E}_0^2(t) \Pi) u) = 0, \\ \partial_t \Pi + u \partial_x \Pi + \frac{a_c^2}{\rho} \partial_x u = \frac{1}{\mu} (p - \Pi). \end{cases} \quad (121)$$

Define  $\mathbf{W} = [\rho, \rho u, \rho e, \rho \Pi]^T$  and  $\mathbf{U} = [\rho, \rho u, \rho e]^T$ . Assume that one can perform a Chapman-Enskog expansion on  $\mathbf{U}$  and  $\Pi$  and write them in powers of  $\mu$ , namely:

$$\begin{aligned} \mathbf{U} &= \mathbf{U}_0 + \mu \mathbf{U}_1 + O(\mu^2), \\ \Pi &= p(\mathbf{U}_0) + \mu \Pi_1 + O(\mu^2), \end{aligned} \quad (122)$$

with  $\mathbf{U}_0$  and  $p(\mathbf{U}_0)$  solutions of subsystem  $\mathcal{C}$  and  $\mathbf{U}_1, \Pi_1$  of order one. Making  $\mu$  tend formally toward zero, the relaxed pressure equation becomes at order zero:

$$\partial_t p(\mathbf{U}_0) + u_0 \partial_x p(\mathbf{U}_0) + \frac{a_c^2}{\rho_0} \partial_x u_0 = -\Pi_1 \Leftrightarrow \left( \frac{a_c^2}{\rho_0} - \rho_0 c_c(\mathbf{U}_0)^2 \right) \partial_x u_0 = -\Pi_1. \quad (123)$$

In order to make  $\mathcal{C}^\mu$  converge towards  $\mathcal{C}$ , a basic step is to make sure that  $\mathbf{U}_1$  remains of order one throughout time. Its evolution is influenced by non linear convective effects which mix order zero and order one terms as well as pressure effects related to  $\mathcal{E}_0^2 \partial_x \Pi_1$  for the momentum equation and  $\mathcal{E}_0^2 \partial_x (\Pi_1 u_0 + p_0 u_1)$  for the energy equation. Using equation (123), one can notice that:

$$-\mathcal{E}_0^2 \partial_x \Pi_1 = \mathcal{E}_0^2 \partial_x \left( \left( \frac{a_c^2}{\rho_0} - \rho_0 c_c(\mathbf{U}_0)^2 \right) \partial_x u_0 \right). \quad (124)$$

Thus, under the convective subcharacteristic condition  $a_c > \rho_0 c_c(\mathbf{U}_0)$ , order zero terms results in a *diffusive* effect on the order one momentum equation. One can believe that this diffusion will be sufficient to prevent  $\mathbf{U}_1$  from exploding when  $\mu$  tends toward zero.

## Relaxed Acoustic Subsystem

The same argumentation can be done on the relaxed acoustic subsystem  $\mathcal{A}^\mu$ . It gives the expected subcharacteristic condition (29b).

## Appendix G: Bounded Oscillations of Sp-(M)-corr

According to the truncation error analyses derived in **Section 3**, the scheme Sp-( $M$ )-corr allows to reduce the spatial numerical diffusion in the convective as well as in the acoustic subsystem in the case of low-Mach number compressible flows. On Figure 13 velocity profiles are plotted for different meshes. The diffusion reduction results in non-physical oscillations in the tail of the left rarefaction wave. However, the  $L^\infty$  and the  $L^1$  norms of the induced error decay as the cells number increases. Hence, the scheme is stable and converges to the analytical solution.

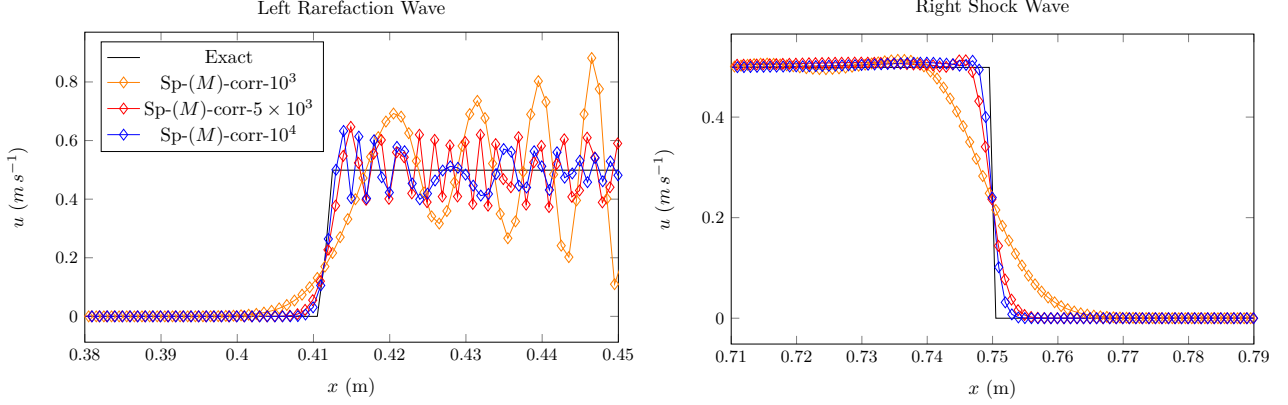


Figure 13: Velocity profile at  $M = 4.2 \times 10^{-3}$ ,  $\text{Sp-}(M)\text{-corr}$

## Appendix H: Truncation Error Analysis

### Truncation Error of the Convective Subsystem

Let us consider the convective numerical flux located at face indexed by  $i + 1/2$ . For the sake of notations simplicity and in order to adopt an unstructured formalism, let us rewrite the index  $i + 1/2$  as  $f$  for "face". Finally let us call  $L$  (respectively  $R$ ) the index of the left (respectively the right) neighbor cell of the face  $f$ . Typically, in 1D  $L = i$  and  $R = i + 1$ . As mentioned in **Subsection 2.2.1** the relaxation scheme for the convective subsystem can be written:

$$\begin{aligned}
 \mathbf{H}_{\mathbf{c}f}^n = \mathbf{H}_{\mathbf{c}}^n(\mathbf{U}_L, \mathbf{U}_R) &= \frac{1}{2} (L(\mathbf{F}_C^\mu)(\mathbf{U}_L) + L(\mathbf{F}_C^\mu)(\mathbf{U}_R)) - \frac{1}{2} |u_L - \mathcal{E}_0 a_C \tau_L| (\mathbf{U}_f^* - \mathbf{U}_L) \\
 &\quad - \frac{1}{2} |u_C^*| (\mathbf{U}_f^{**} - \mathbf{U}_f^*) \\
 &\quad - \frac{1}{2} |u_R + \mathcal{E}_0 a_C \tau_R| (\mathbf{U}_R - \mathbf{U}_f^{**}), \\
 L(\mathbf{F}_C^\mu)(\mathbf{U}) &= [\rho u, \rho u^2 + \mathcal{E}_0^2 p, (\rho e + \mathcal{E}_0^2 p) u]^T.
 \end{aligned} \tag{125}$$

Using the classical rescaling described in [11], the non-dimensional version of this numerical flux writes:

$$\begin{aligned}
 \mathbf{H}_{\mathbf{c}f}^n = \mathbf{H}^n(\mathbf{U}_L, \mathbf{U}_R) &= \frac{1}{2} (L(\mathbf{F}_C^\mu)(\mathbf{U}_L) + L(\mathbf{F}_C^\mu)(\mathbf{U}_R)) - \frac{1}{2} \left| u_L - \frac{\mathcal{E}_0}{M} a_C \tau_L \right| (\mathbf{U}_f^* - \mathbf{U}_L) \\
 &\quad - \frac{1}{2} |u_C^*| (\mathbf{U}_f^{**} - \mathbf{U}_f^*) \\
 &\quad - \frac{1}{2} \left| u_R + \frac{\mathcal{E}_0}{M} a_C \tau_R \right| (\mathbf{U}_R - \mathbf{U}_f^{**}), \\
 L(\mathbf{F}_C^\mu)(\mathbf{U}) &= [\rho u, \rho u^2 + (\mathcal{E}_0/M)^2 p, (\rho e + \mathcal{E}_0^2 p) u]^T, \quad \rho e = \rho \varepsilon + \frac{M^2}{2} \rho u^2.
 \end{aligned} \tag{126}$$

For the sake of notations, let us rewrite  $\mathbf{U}_f^*$  as  $\mathbf{U}_L^*$  and  $\mathbf{U}_f^{**}$  as  $\mathbf{U}_R^*$ . The non-dimensional intermediate states  $\mathbf{U}_k^*$ ,  $k \in \{L, R\}$  can be expressed as:

$$\mathbf{U}_k^* = \begin{bmatrix} \rho_{k,C}^* \\ \rho_{k,C}^* u_C^* \\ \rho_{k,C}^* e_{k,C}^* \end{bmatrix}, \quad (127)$$

with:

$$\left\{ \begin{array}{l} u_C^* = \frac{u_R + u_L}{2} - \frac{\mathcal{E}_0}{M} \frac{(p_R - p_L)}{2ac}, \\ p^* = \frac{p_R + p_L}{2} - \frac{M}{\mathcal{E}_0} \frac{ac}{2} (u_R - u_L), \\ \rho_{k,C}^* = 1/\tau_k^*, \quad \tau_k^* = \tau_k + \frac{M}{\mathcal{E}_0} \frac{(-1)^{i_k+1}}{ac} (u_C^* - u_k), \\ e_{k,C}^* = e_k + \mathcal{E}_0 M \frac{(-1)^{i_k}}{ac} (p^* u_C^* - p_k u_k), \\ ac = K \cdot \max(\rho_L cc(\rho_L, p_L), \rho_R cc(\rho_R, p_R)), \quad K > 1, \\ k \in \{L, R\}, \quad i_L \equiv 1, \quad i_R \equiv 2. \end{array} \right. \quad (128)$$

Let us define  $x_f$ ,  $x_L$  and  $x_R$  the positions of the face, left cell and right cell barycenters. In 1D:  $x_L = x_i$ ,  $x_R = x_{i+1}$  and  $x_f = x_L + \Delta x/2 = x_R - \Delta x/2$ . At a given time  $t$ , for a smooth function  $\phi(\cdot, t)$  let us write  $\phi_f$  for  $\phi(x_f, t)$  and  $\phi_L$  for  $\phi(x_L, t)$ . Particularly, we will consider  $x \rightarrow a(x, t)$  a smooth function such that  $a_C(x_{i+1/2}, t^n) = (a_C)_{i+1/2}^n$ .

Let us consider a non-dimensional smooth state  $(x, t) \rightarrow \mathbf{U}(x, t)$  verifying:

$$St_r \frac{\mathbf{U}(x_i, t^n + \Delta t) - \mathbf{U}(x_i, t^n)}{\Delta t} + \frac{\mathbf{H}_c(\mathbf{U}(x_i, t^n), \mathbf{U}(x_{i+1}, t^n)) - \mathbf{H}_c(\mathbf{U}(x_{i-1}, t^n), \mathbf{U}(x_i, t^n))}{\Delta x} = 0, \quad (129)$$

One wonders which partial differential equation does such a smooth solution verify?

It can be first noticed that  $\frac{\mathbf{U}(x_i, t^n + \Delta t) - \mathbf{U}(x_i, t^n)}{\Delta t}$  is consistent with  $\partial_t \mathbf{U}_i + \underline{Q}(\Delta t)$ .

Let us now focus on  $\mathbf{H}_c^n(\mathbf{U}_L, \mathbf{U}_R)$ : first,  $\forall k \in \{L, R\}$ :

$$\mathbf{U}_k^* - \mathbf{U}_k = \begin{bmatrix} \rho_{k,C}^* - \rho_k \\ (\rho_{k,C}^* - \rho_k) u_C^* + \rho_k (u_C^* - u_k) \\ (\rho_{k,C}^* - \rho_k) e_{k,C}^* + \rho_k (e_{k,C}^* - e_k) \end{bmatrix}. \quad (130)$$



Furthermore, performing a Taylor expansion around  $x_f$ , and setting  $i_L = 1$ ,  $i_R = 2$ , one obtains:

$$\left\{ \begin{array}{l} \rho_{k,c}^* - \rho_k = \left[ -\frac{\rho_f^2}{a_f} \left( \frac{M}{\mathcal{E}_0} \right) \partial_x u_{|f} + (-1)^{i_k+1} \left( \frac{\rho_f}{a_f} \right)^2 \partial_x p_{|f} \right] \frac{\Delta x}{2} + O \left( \left( 1 + \frac{M}{\mathcal{E}_0} \right) \Delta x^2 \right), \\ u_C^* - u_k = \left[ (-1)^{i_k+1} \partial_x u_{|f} - \left( \frac{\mathcal{E}_0}{M} \right) \frac{1}{a_f} \partial_x p_{|f} \right] \frac{\Delta x}{2} + O \left( \left( 1 + \frac{\mathcal{E}_0}{M} \right) \Delta x^2 \right), \\ e_{k,c}^* - e_k = M \mathcal{E}_0 \left[ -\frac{1}{a_f} \partial_x p u_{|f} + (-1)^{i_k+1} \left( \left( \frac{\mathcal{E}_0}{M} \right) \frac{p_f}{a_f^2} \partial_x p_{|f} + \left( \frac{M}{\mathcal{E}_0} \right) u_f \partial_x u_{|f} \right) \right] \frac{\Delta x}{2} \\ \quad + O \left( M \mathcal{E}_0 \left( 1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0} \right) \Delta x^2 \right), \\ u_C^* = u_f - \left( \frac{\mathcal{E}_0}{M} \right) \frac{1}{a_f} \partial_x p_{|f} \frac{\Delta x}{2} + O \left( \left( 1 + \frac{\mathcal{E}_0}{M} \right) \Delta x^2 \right), \\ e_{k,c}^* = e_f + (-1)^{i_k} \partial_x e_{|f} \frac{\Delta x}{2} + e_{k,c}^* - e_k. \end{array} \right. \quad (131)$$

Then:

$$\mathbf{U}_k^* - \mathbf{U}_k = \frac{\Delta x}{2} \left[ \begin{array}{l} -\frac{\rho_f^2}{a_f} \left( \frac{M}{\mathcal{E}_0} \right) \partial_x u_{|f} + (-1)^{i_k+1} \left( \frac{\rho_f}{a_f} \right)^2 \partial_x p_{|f} \\ \rho_f \left( (-1)^{i_k+1} - \frac{\rho_f u_f M}{a_f \mathcal{E}_0} \right) \partial_x u_{|f} + \frac{\rho_f}{a_f} \left( (-1)^{i_k+1} \frac{\rho_f u_f}{a_f} - \frac{\mathcal{E}_0}{M} \right) \partial_x p_{|f} \\ \rho_f e_f \left( \left( \frac{\rho_f M}{a_f \mathcal{E}_0} + (-1)^{i_k+1} \frac{u_f}{e_f} \mathcal{E}_0 \frac{M}{\mathcal{E}_0} \right) \partial_x u_{|f} + (-1)^{i_k+1} \left( \frac{\rho_f}{a_f^2} + \mathcal{E}_0 \frac{\mathcal{E}_0}{M} \frac{p_f}{a_f^2 e_f} \right) \partial_x p_{|f} \right) \\ \quad + \mathcal{E}_0 \frac{\mathcal{E}_0}{M} \frac{p_f}{a_f^2 e_f} \partial_x p u_{|f} \end{array} \right] + \underline{O} \left( \left( 1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0} \right) \Delta x^2 \right). \quad (132)$$

One can finally observe that, for the terms in order one in space, the Mach number is always compensated with the weighting parameter  $\mathcal{E}_0$ . Thus  $\forall k \in \{L, R\}$ :

$$\mathbf{U}_k^* - \mathbf{U}_k = \underline{O} \left( \left( 1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0} \right) \Delta x \right)_f. \quad (133)$$

Similarly:

$$\begin{aligned} \mathbf{U}_R^* - \mathbf{U}_L^* &= (\mathbf{U}_R^* - \mathbf{U}_R) + (\mathbf{U}_R - \mathbf{U}_L) + (\mathbf{U}_L - \mathbf{U}_L^*), \\ &= \underline{O} \left( \left( 1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0} \right) \Delta x \right)_f. \end{aligned} \quad (134)$$

Furthermore, one can easily see that  $\forall k \in \{L, R\}$ :

$$\begin{aligned} \left| u_k + (-1)^{i_k} \frac{\mathcal{E}_0}{M} a_C \tau_k \right| &= \left| u_f + (-1)^{i_k} \frac{\mathcal{E}_0}{M} a_f \tau_f \right| + O \left( \left( 1 + \frac{\mathcal{E}_0}{M} \right) \Delta x \right), \\ |u_C^*| &= |u_f| + O \left( \left( 1 + \frac{\mathcal{E}_0}{M} \right) \Delta x \right). \end{aligned} \quad (135)$$

Thus, at a given face  $f$  we have:

$$\mathbf{H}_{\mathbf{c}f}^n = \mathbf{H}_{\mathbf{c}}^n(\mathbf{U}_L, \mathbf{U}_R) = \frac{1}{2} (L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_L) + L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_R)) + \underline{O} \left( \left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x \right)_f. \quad (136)$$

Besides,  $\frac{1}{2} (L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_L) + L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_R))$  is consistent with  $L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_f) + \underline{O}(\Delta x^2)$ .

Finally  $\frac{\mathbf{H}_{i+1/2}^n - \mathbf{H}_{i-1/2}^n}{\Delta x}$  is consistent with  $\partial_x L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_i) + \underline{O} \left( \left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x \right)$ . Thus we have found that the smooth solution  $\mathbf{U}(x, t)$  verified the PDE  $\forall x_i, t^n$ :

$$St_r \partial_t \mathbf{U}_i^n + \partial_x L(\mathbf{F}_{\mathcal{C}}^\mu)(\mathbf{U}_i^n) = \underline{O}(St_r \Delta t) + \underline{O} \left( \left(1 + \frac{\mathcal{E}_0}{M} + \frac{M}{\mathcal{E}_0}\right) \Delta x \right). \quad (137)$$

### Truncation Error of the Acoustic Subsystem

Keeping the same notations than previously, the non-dimensional relaxation flux for the acoustic subsystem writes:

$$\mathbf{H}_{\mathbf{ac}f}^n = (1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \Pi_{\mathcal{A}}^* \\ \Pi_{\mathcal{A}}^* u_{\mathcal{A}}^* \end{bmatrix} = (1 - \mathcal{E}_0^2) \left( \begin{bmatrix} 0 \\ \frac{1}{M^2} \frac{p_R + p_L}{2} \\ \frac{p_R u_R + p_L u_L}{2} \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{1}{M} \frac{a_{\mathcal{A}}}{2} (u_R - u_L) \\ \frac{M a_{\mathcal{A}}}{4} (u_R^2 - u_L^2) + \frac{1}{4} \frac{1}{M a_{\mathcal{A}}} (p_R^2 - p_L^2) \end{bmatrix} \right),$$

with:  $a_{\mathcal{A}} = K \cdot \max(\rho_L c_{\mathcal{A}}(\rho_L, p_L), \rho_R c_{\mathcal{A}}(\rho_R, p_R))$ ,  $K > 1$ .

(138)

It is easy to check that  $(1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \frac{1}{M^2} \frac{p_R + p_L}{2} \\ \frac{p_R u_R + p_L u_L}{2} \end{bmatrix}$  is consistent with

$$(1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \frac{p_f}{M^2} \\ p_f u_f \end{bmatrix} + \begin{bmatrix} 0 \\ O((1 - \mathcal{E}_0^2)(\Delta x/M)^2) \\ O((1 - \mathcal{E}_0^2)\Delta x^2) \end{bmatrix}. \text{ Besides, } -(1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \frac{1}{M} \frac{a_{\mathcal{A}}}{2} (u_R - u_L) \\ \frac{M a_{\mathcal{A}}}{4} (u_R^2 - u_L^2) + \frac{1}{4} \frac{1}{M a_{\mathcal{A}}} (p_R^2 - p_L^2) \end{bmatrix}$$

is consistent with  $-(1 - \mathcal{E}_0^2) \frac{\Delta x}{2} \begin{bmatrix} 0 \\ \frac{1}{M} a_f \partial_x u|_f \\ M a_f u_f \partial_x u|_f + \frac{1}{M a_f} p_f \partial_x p|_f \end{bmatrix} + \begin{bmatrix} 0 \\ O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x^2\right) \\ O((1 - \mathcal{E}_0^2)(M + \frac{1}{M})\Delta x^2) \end{bmatrix}_f$ .

Finally, one obtains at order one in space:

$$\mathbf{H}_{\mathbf{ac}f}^n = (1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \frac{p_f}{M^2} \\ p_f u_f \end{bmatrix} + \begin{bmatrix} 0 \\ O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x\right) \\ O((1 - \mathcal{E}_0^2)(M + \frac{1}{M})\Delta x) \end{bmatrix}_f. \quad (139)$$

Thus we have found that the smooth solution  $\underline{U}(x, t)$  verified the PDE  $\forall x_i, t^n$ :

$$\begin{aligned} St_r \partial_t \rho &= O(St_r \Delta t), \\ St_r \partial_t (\rho u) + \partial_x ((1 - \mathcal{E}_0^2(t)) p) &= O(St_r \Delta t) + O\left(\frac{(1 - \mathcal{E}_0^2)}{M} \Delta x\right), \\ St_r \partial_t (\rho e) + \partial_x ((1 - \mathcal{E}_0^2(t)) p u) &= O(St_r \Delta t) + O\left((1 - \mathcal{E}_0^2)(M + \frac{1}{M}) \Delta x\right). \end{aligned} \quad (140)$$

## Truncation Error of the Acoustic Subsystem with low-Mach Correction

Endowed with the low-Mach correction described in equation (71), the acoustic flux at face  $x_f$  reads:

$$\mathbf{H}_{\text{ac}f}^n = (1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \Pi^* \\ \Pi^* u^* \end{bmatrix} = (1 - \mathcal{E}_0^2) \left( \begin{bmatrix} 0 \\ \frac{1}{M^2} \frac{\Pi_R + \Pi_L}{2} \\ \frac{\Pi_R u_R + \Pi_L u_L}{2} \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{\theta}{M} \frac{a_A}{2} (u_R - u_L) \\ \frac{M\theta a_A}{4} (u_R^2 - u_L^2) + \frac{1}{4} \frac{1}{M a_A} (\Pi_R^2 - \Pi_L^2) \end{bmatrix} \right). \quad (141)$$

The correction part is now consistent with:

$$-(1 - \mathcal{E}_0^2) \frac{\Delta x}{2} \begin{bmatrix} 0 \\ \frac{\theta_f}{M} a_f \partial_x u|_f \\ M\theta_f a_f u_f \partial_x u|_f + \frac{1}{M a_f} p_f \partial_x p|_f \end{bmatrix} + \begin{bmatrix} 0 \\ O\left(\frac{(1 - \mathcal{E}_0^2)\theta}{M} \Delta x^2\right) \\ O\left((1 - \mathcal{E}_0^2)\left(M\theta + \frac{1}{M}\right) \Delta x^2\right) \end{bmatrix}_f.$$

At first order w.r.t  $\Delta x$ :

$$\mathbf{H}_{\text{ac}f}^n = (1 - \mathcal{E}_0^2) \begin{bmatrix} 0 \\ \frac{p_f}{M^2} \\ p_f u_f \end{bmatrix} + \begin{bmatrix} 0 \\ O\left(\frac{(1 - \mathcal{E}_0^2)\theta}{M} \Delta x\right) \\ O\left((1 - \mathcal{E}_0^2)\left(M\theta + \frac{1}{M}\right) \Delta x\right) \end{bmatrix}_f. \quad (142)$$

For a smooth solution  $\underline{U}(x, t)$ , the truncation error analysis made on the acoustic scheme with low-Mach correction gives  $\forall x_i, t^n$ :

$$\begin{aligned} St_r \partial_t \rho &= O(St_r \Delta t), \\ St_r \partial_t (\rho u) + \partial_x ((1 - \mathcal{E}_0^2(t)) p) &= O(St_r \Delta t) + O\left(\frac{(1 - \mathcal{E}_0^2)\theta}{M} \Delta x\right), \\ St_r \partial_t (\rho e) + \partial_x ((1 - \mathcal{E}_0^2(t)) p u) &= O(St_r \Delta t) + O\left((1 - \mathcal{E}_0^2)\left(M\theta + \frac{1}{M}\right) \Delta x\right). \end{aligned} \quad (143)$$

## References

- [1] R. Baraille, G. Bourdin, F. Dubois, and A. Y. Le Roux. Une version à pas fractionnaires du schéma de Godunov pour l'hydrodynamique. *Compte Rendu de l'Académie des Sciences*, 314:147–152, 1992.
- [2] M. Baudin, C. Berthon, F. Coquel, R. Masson, and Q. H. Tran. A relaxation method for two-phase flow models with hydrodynamic closure laws. *Numerische Mathematik*, 99:411–440, 2005.
- [3] M. Baudin, F. Coquel, and Q. H. Tran. A semi-implicit relaxation scheme for modelling two-phase flow in a pipeline. *SIAM Journal of Scientific Computing*, 27:914–936, 2005.
- [4] F. Bouchut. Entropy satisfying flux vector splittings and kinetic BGK models. *Numerische Mathematik*, 94:623–672, 2003.
- [5] F. Bouchut. A reduced stability condition for nonlinear relaxation to conservation laws. *Journal of Hyperbolic Differential Equations*, 1:149–170, 2004.
- [6] F. Bouchut. *Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws*. Birkäser, 2004.
- [7] T. Buffard, T. Gallouët, and J-M. Hérard. A sequel to a rough Godunov scheme: application to real gases. *Computer & Fluids*, 29:813–847, 2000.
- [8] T. Buffard and J-M. Hérard. A conservative fractional step method to solve non-isentropic Euler equations. *Computer Methods in Applied Mechanics and Engineering*, 144:199–225, 1996.
- [9] C. Chalons and J.F. Coulombel. Relaxations approximation of the Euler equations. *Journal of Mathematical Analysis and Applications*, 348:872–893, 2008.
- [10] C. Chalons, M. Girardin, and S. Kokh. An all-regime Lagrange-Projection like scheme for 2D homogeneous models for two-phase flows on unstructured meshes. *Journal of Computational Physics*, 335:885–904, 2016.
- [11] C. Chalons, M. Girardin, and S. Kokh. An all-regime Lagrange-Projection like scheme for the gas dynamics equations on unstructured meshes. *Communications in Computational Physics*, 20:188–233, 2016.
- [12] G. Q. Chen, C. D. Levermore, and T.-P. Liu. Hyperbolic conservation laws with stiff relaxation terms and entropy. *Communications on Pure and Applied Mathematics*, 47:787–830, 1994.
- [13] F. Coquel, E. Godlewski, B. Perthame, A. In, and P. Rascle. Some new Godunov and relaxation methods for two-phase flow problems. *Kluwer Academic/Plenum Publishers (New York)*, pages 179–188, 2001.
- [14] F. Coquel, E. Godlewski, and N. Seguin. Relaxation of fluid systems. *Mathematical Models and Methods in Applied Science*, 22:43–95, 2012.

- [15] F. Coquel, Q. L. Nguyen, M. Postel, and Q. H. Tran. Entropy-satisfying relaxation method with large time-steps for Euler IBVPS. *Mathematics of Computation*, 79:1493–1533, 2010.
- [16] S. Dallet. *Simulation numérique d'écoulements diphasiques en régime compressible ou à faible nombre de Mach*. PhD thesis, Aix-Marseille Université, 2017.
- [17] P. Degond and M. Tang. All speed scheme for the low Mach number limit of the isentropic Euler equation. *Communications in Computational Physics*, 10:1–31, 2011.
- [18] S. Dellacherie. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *Journal of Computational Physics*, 229:978–1016, 2009.
- [19] S. Dellacherie, P. Omnes, J. Jung, and P.A. Raviart. Construction of modified Godunov type schemes accurate at any Mach number for the compressible Euler system. *Mathematical Models and Methods in Applied Science*, 26:2525–2615, 2016.
- [20] S. Dellacherie, P. Omnes, and F. Rieper. The influence of cell geometry on the Godunov scheme applied to the linear wave equation. *Journal of Computational Physics*, 229:5315–5338, 2010.
- [21] G. Dimarco, R. Loubère, and M-H. Vignal. Study of a new asymptotic preserving scheme for the Euler system in the low Mach number limit. *SIAM: Journal of Scientific Computing*, 39:2099–2128, 2017.
- [22] P. Fillion, A. Chanoine, S. Dellacherie, and A. Kumbaro. FLICA-OVAP: A new platform for core thermohydraulic studies. *Nuclear Engineering and Design*, 241:4348–4358, 2011.
- [23] T. Gallouët, J-M Hérard, and N. Seguin. Some recent finite volume schemes to compute Euler equations using real gas EOS. *International Journal for Numerical Methods in Fluids*, 39:1073–1138, 2002.
- [24] T. Gallouët, J-M Hérard, and N. Seguin. Numerical modeling of two-phase flows using the two-fluid two-pressure approach. *Mathematical Models and Methods in Applied Sciences*, 14:663–700, 2004.
- [25] M. Girardin. *Asymptotic preserving and all-regime Lagrange-Projection like numerical schemes: application to two-phase flows in low Mach regime*. PhD thesis, Université Pierre et Marie Curie, <https://tel.archives-ouvertes.fr/tel-01127428>, 2015.
- [26] E. Godlewski and P.A. Raviart. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Springer, 1996.
- [27] H. Guillard and A. Murrone. On the behavior of upwind schemes in the low Mach number limit: II Godunov type schemes. *Computers and Fluids*, 33:655–675, 2004.
- [28] H. Guillard and C. Viozat. On the behavior of upwind schemes in the low Mach number limit. *Computers and Fluids*, 28:63–86, 1999.
- [29] J. Haack, S. Jin, and J. G. Liu. An all-speed asymptotic-preserving method for the isentropic Euler and Navier-Stokes equations. *Communications in Computational Physics*, 12:955–980, 2012.

- [30] A. Harten, P. D. Lax, and B. Van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Review*, 25:35–61, 1983.
- [31] D. Iampietro, F. Daude, P. Galon, and J. M. Hérard. A Mach-sensitive implicit-explicit scheme adapted to compressible multi-scale flows. <https://hal.archives-ouvertes.fr/hal-01531306>, 2017.
- [32] D. Iampietro, F. Daude, P. Galon, and J. M. Hérard. A weighted splitting approach for low-Mach number flows. *Finite Volumes for Complex Applications VIII-Hyperbolic, Elliptic and Parabolic Problems: FVCA 8*, 200, 2017.
- [33] S. Jin and Z.-P. Xin. The relaxation schemes for systems of conservation laws in arbitrary dimensions. *Communications on Pure and Applied Mathematics*, 48:235–276, 1995.
- [34] S. Noelle, G. Bispen, K. R. Arun, M. Lukáčová-Medvid'ová, and C. D. Munz. A weakly asymptotic preserving low Mach number scheme for the Euler equations of gas dynamics. *SIAM Journal on Scientific Computing*, 36:B989–B1024, 2014.
- [35] F. Rieper. A low-Mach number fix for Roes approximate Riemann solver. *Journal of Computational Physics*, 230:5263–5287, 2011.
- [36] S. Schochet and G. Metivier. Fast singular limits of hyperbolic PDEs. *Journal of Differential Equations*, 114:476–512, 1994.
- [37] S. Schochet and G. Metivier. Limite incompressible des equations d' Euler non-isentropiques. *Preprint*, <https://www.math.u-bordeaux.fr/gmetivie/Preprints.html>, 2000.
- [38] S. Schochet and G. Metivier. The incompressible limit of Euler non-isentropic equations. *Archive for Rational Mechanics and Analysis*, 158:61–90, 2001.
- [39] A. R. Simpson. *Large water hammer pressures due to column separation in sloping pipes*. PhD thesis, Diss. University of Michigan, 1986.
- [40] J. Smoller. *Shock Waves and Reaction-Diffusion Equations*. Springer-Verlag, 1994.
- [41] G. A. Sod. *Numerical Methods in Fluid Dynamics, Initial and Initial-boundary Value Problems*. Cambridge University Press, 1985.
- [42] I. Suliciu. On the thermodynamics of fluids with relaxation and phase transitions. *International Journal of Engineering Science*, 36:921–947, 1998.
- [43] E.F. Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer, 1999.
- [44] J. B. Whitham. *Linear and Non Linear Waves*. John Wiley & Sons Inc, 1974.