Implicit discretization of Lagrangian gas dynamics

A. Plessier, S. del Pino, B. Després

Implicit discretization of Lagrangian gas dynamics

A. Plessier*[1,2], S. Del Pino[1,4], and B. Després[2,3]

[1]CEA, DAM, DIF, F-91297 Arpajon, France
[2]LJLL (UMR_7598) - Laboratoire Jacques-Louis Lions
[3]IUF - Institut Universitaire de France
[4]Université Paris-Saclay, Laboratoire en Informatique Haute Performance pour le Calcul et la simulation

July 12, 2021

Abstract

We construct an original framework based on convex analysis to prove the existence and uniqueness of a solution to a class of implicit numerical schemes. We propose an application of this general framework in the case of a new non linear implicit scheme for the 1D Lagrangian gas dynamics equations. We provide numerical illustrations that corroborate our proof of unconditional stability for this non linear implicit scheme.

Keywords

Implicit Finite Volume Scheme, Lagrangian Formalism, Entropy Stability, Convex Analysis

1 Introduction

To approach equations traducing the movements of fluids, explicit schemes are traditionally used, see,[16,29] because they are easy to implement. Explicit schemes need to satisfy a stability CFL condition: $c\Delta t \leqslant \Delta x$, where c is the speed of sound, Δt is the time step, and Δx is the size of the discretization in space of the mesh. The time step is constrained by the size of the smallest cell. Nevertheless, in some cases, a large speed of sound ($c \gg 1$) contributes to have such small time steps that it becomes unfavourable to use explicit methods, because the simulation cost is high.

Implicit in time schemes have always aroused interest in the literature. In particular, they are much less sensitive to the CFL number: for example a finite difference algorithm is proposed in Beam and Warming,[2] implicit and semi-implicit schemes are investigated in the Toth, Keppens and Bochev[33] and an experimentation on implicit upwind methods for Euler equations is done in Mulder and Van Leer.[25] Some other implicit algorithms are explained in the following articles, see for example.[9,24,28] A large part of them use the method of predictor-corrector scheme like in the works.[20,26,34–36] More recent works can be found in the articles.[6,27]

Nonetheless, some major technical difficulties appear for the numerical resolution of implicit non linear schemes. At a theoretical level, it is difficult to prove the existence and uniqueness of a solution to implicit schemes. Currently, a powerful theory is the one developed in Gallouët *et al.*,[15] for Navier-Stokes equations with viscosity, using the topological degree. The existence of a solution is proved in details, but the uniqueness requires restrictive hypothesis. Another strategy is explained in Brugano and Casulli[3] for piecewise linear systems in the case of a symmetrical structure. The existence of a solution is proved, and the uniqueness is also studied in the case of special hypothesis. A non-linear implicit-explicit strategy is found in Fryxell *et al.*,[14] that is second order in both space and time. There are no proof of existence or uniqueness, but the numerical results illustrate an unconditional stability of the method.

*alexiane.plessier@cea.fr

Our original contributions to this field are, firstly the elaboration of a general framework which allows to prove existence and uniqueness of implicit solution of some numerical schemes, and secondly the application of the general framework in the case of a non linear implicit scheme for the 1D model problem (1) written in semi-Lagrangian coordinates (semi-Lagrangian coordinates means Lagrange+update)

$$\begin{cases} \rho D_t \tau - \partial_x u = 0, \\ \rho D_t u + \partial_x p = 0, \\ \rho D_t E + \partial_x p u = 0. \end{cases} \tag{1}$$

One has $\rho = \frac{1}{\tau} > 0$ the mass density, p is the pressure, u is the velocity and E is the total energy density. The variables τ and p are taken positive to be physically admissible, see Serre[31] or Ern and Guermond[12] for more details. The material derivative is $D_t = \partial_t + u\partial_x$. The following set of equations is the isentropic Euler equations for the prediction step of the implicit scheme

$$\begin{cases} \rho D_t \tau - \partial_x u = 0, \\ \rho D_t u + \partial_x p = 0, \\ \rho D_t S = 0. \end{cases} \tag{2}$$

The first two equations are identical to (1), only the last one is different, with S denoting the physical entropy. These equations are equipped with a perfect gas equation of state

$$\begin{cases} p = (\gamma - 1)\dfrac{e}{\tau}, \\ e = C_v T, \\ S = C_v \log(e\tau^{\gamma-1}), \end{cases} \tag{3}$$

where C_v is the thermal capacity at constant volume, $\gamma > 1$ is the adiabatic index, $e = E - \frac{1}{2}u^2$ is the internal energy density, T is the temperature and c is the speed of sound given by $c^2 = \frac{\partial p}{\partial \rho}$.

The justification of using a prediction step based on the discretization of (2) comes from ideas in the work of Chalons, Coquel and Marmignon.[4] It will be explained in details in this article.

Consider a mesh $\mathcal{M}$ composed of N cells noted $j \in \{1, \ldots, N\}$. The time t is discretized with a time step Δt that corresponds to one iteration. The mass of the cell j is M_j. The boundary conditions are supposed periodic, and the fluxes are defined with the acoustic impedance $\alpha_{j+\frac{1}{2}} > 0$. The scheme has a predictor-corrector structure. The prediction step is written as

$$\text{Prediction step} \begin{cases} \overline{\tau_j} = \tau_j^n + \dfrac{\Delta t}{M_j}(\overline{u_{j+\frac{1}{2}}} - \overline{u_{j-\frac{1}{2}}}), \\ \overline{u_j} = u_j^n - \dfrac{\Delta t}{M_j}(\overline{p_{j+\frac{1}{2}}} - \overline{p_{j-\frac{1}{2}}}), \\ \overline{S_j} = S_j^n. \end{cases} \tag{4}$$

The correction step is given by

$$\text{Correction step} \begin{cases} \tau_j^{n+1} = \tau_j^n + \dfrac{\Delta t}{M_j}(\overline{u_{j+\frac{1}{2}}} - \overline{u_{j-\frac{1}{2}}}), \\ u_j^{n+1} = u_j^n - \dfrac{\Delta t}{M_j}(\overline{p_{j+\frac{1}{2}}} - \overline{p_{j-\frac{1}{2}}}), \\ E_j^{n+1} = E_j^n + \dfrac{\Delta t}{M_j}(\overline{p_{j+\frac{1}{2}}}\ \overline{u_{j+\frac{1}{2}}} - \overline{p_{j-\frac{1}{2}}}\ \overline{u_{j-\frac{1}{2}}}), \end{cases} \tag{5}$$

where only the total energy equation is modified. The correction step is explicit so the main difficulty is in the prediction step.

In (4) and (5), the fluxes are defined by

$$\overline{p_j} - \overline{p_{j+\frac{1}{2}}} = \alpha^n_{j+\frac{1}{2}}(\overline{u_{j+\frac{1}{2}}} - \overline{u_j}), \quad \overline{p_j} - \overline{p_{j-\frac{1}{2}}} = \alpha^n_{j-\frac{1}{2}}(\overline{u_j} - \overline{u_{j-\frac{1}{2}}}),$$

where the coefficient $\alpha^n_{j+\frac{1}{2}} > 0$ is for simplicity equal to a mean value of the acoustic impedance: $\alpha_{j+\frac{1}{2}} = \frac{1}{2}(\rho_j c_j + \rho_{j+1}c_{j+1})$. Another equivalent formula is

$$\begin{cases} \overline{p_{j+\frac{1}{2}}} = \dfrac{\rho_j c_j + \rho_{j+1}c_{j+1}}{4}(\overline{u_j} - \overline{u_{j+1}}) + \dfrac{1}{2}(\overline{p_j} + \overline{p_{j+1}}), \\ \overline{u_{j+\frac{1}{2}}} = \dfrac{1}{\rho_j c_j + \rho_{j+1}c_{j+1}}(\overline{p_j} - \overline{p_{j+1}}) + \dfrac{1}{2}(\overline{u_j} + \overline{u_{j+1}}). \end{cases}$$

In Lagrangian formalism, the mesh moves according to the velocity of the fluid: $x^{n+1}_{j+\frac{1}{2}} = x^n_{j+\frac{1}{2}} + \Delta t u^n_{j+\frac{1}{2}}$. The predictor-corrector scheme (4-5) is naturally conservative since it is expressed in terms of fluxes. Such a scheme can be proved to be weakly consistent as in Després.[10,11]

The predictor-corrector scheme is an adaptation of ideas from the article[4] which is dedicated to solve the classical Euler equations (6)

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x(\rho u^2 + p) = 0, \\ \partial_t \rho E + \partial_x(\rho E u + pu) = 0. \end{cases} \tag{6}$$

The authors explain that the difficulties of solving this system come from the flux terms in the second and third equations. Indeed, there is a strong non linearity due in particular to the pressure. To overcome this complexity, the authors propose a predictor-corrector strategy that we use also in this work. Firstly is to solve the isentropic Euler equations (7) during the prediction step

$$\begin{cases} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x(\rho u^2 + p) = 0, \\ \partial_t \rho S + \partial_x \rho S u = 0. \end{cases} \tag{7}$$

Secondly, the classical Euler equations (6) are solved in order to restore the conservation of the total energy. At a discrete level, the fluxes are expressed thanks to an isentropic scheme, and then inserted in the scheme associated to (6). For the prediction step, the authors use a relaxation scheme on the pressure. They prove the existence of a solution to the relaxation implicit scheme. Nonetheless, the robustness of the scheme depends on an extra equation as mentionned in a report, see.[30]

Let us now describe our results. For physically admissible data, the system (4) will be said *unconditionally stable* if there exists a unique solution to the implicit non-linear scheme. Our proof of the unconditional stability of (4) comes from a rewriting of the prediction step (4) under the form

$$\begin{cases} \text{Find } U \in \mathcal{D} \text{ such that} \\ \nabla J(U) = AU, \end{cases} \tag{8}$$

where U is a vector of real unknowns, J is a functional defined on a domain $\mathcal{D}$ and A is a matrix of real coefficients.

The proof that (8) has a unique solution relies on Theorem 4 that seems to be new considering the classical literature of convex analysis, see.[1,17,19] The ingredients to establish Theorem 4 are the following.

Hypothesis 1. *The open convex domain is* $\mathcal{D} =]-\infty, 0[^n \times \mathbb{R}^m \subset \mathbb{R}^n \times \mathbb{R}^m$, *where* $n > 0$ *and* $m \geqslant 0$ *are two integers. Its boundary is* $\partial\mathcal{D} = \{V \in \mathbb{R}^{n+m} : \exists\, j^* \in \{1, \ldots, n\}\ V_{j^*} = 0,\ \forall j \neq j^* \in \{1, \ldots, n\},\ V_j \leqslant 0\}$.

We made a slight abuse of notations by using the same letter n in the Hypothesis 1 as the iteration index in the scheme (4 -5). We believe this does not interfere with the readability. The case $m = 0$ corresponds to the Traffic flow equations studied in the Appendix. Otherwise $m = n$.

Hypothesis 2. *The function* $J : U \in \mathcal{D} \to J(U) \in \mathbb{R}$ *is* $\mathcal{C}^2$, *strictly convex and coercive in the sense that*

$$J(U) \to +\infty \text{ for } ||U|| \xrightarrow[U\in\mathcal{D}]{} +\infty. \tag{9}$$

Moreover for all $V \in \partial\mathcal{D}$ there exists a unit direction $\boldsymbol{d} \in \mathbb{R}^{n+m}$ which is outward from $\mathcal{D}$ such that

$$(\nabla J(V - \varepsilon \boldsymbol{d}), \boldsymbol{d}) \xrightarrow[V-\varepsilon \boldsymbol{d} \in \mathcal{D}]{\varepsilon \to 0^+} +\infty. \tag{10}$$

Also for all $V \in \partial\mathcal{D}$, one has

$$||\nabla J(W)|| \xrightarrow[W \in \mathcal{D}]{W \to V} +\infty. \tag{11}$$

The verification of (9), (10) and (11) will be obtained directly from the perfect gas law equations (3). For an isentropic gas, it can be simplified.

Hypothesis 3. *The matrix $A \in \mathcal{M}_{n+m}(\mathbb{R})$ is skew-symmetric and its kernel satisfies* $\ker(A) \cap \mathcal{D} \neq \emptyset$.

Theorem 4. *Under the Hypothesis 1, 2 and 3, the problem (8) has a unique solution.*

Applying this Theorem, we show that (4) is well defined for all $\Delta t > 0$.

Corollary 5. *Considering physically admissible data ($\tau_j^n > 0$ for all j), the prediction scheme (4) can be written under the form (8). Therefore, it is unconditionally stable.*

Moreover, the predictor-corrector scheme (4-5) satisfies two entropy inequalities.

Theorem 6. *For all $\Delta t > 0$, the solution of the prediction step (4) satisfies*

$$\forall j, \quad \frac{\overline{E_j} - E_j^n}{\Delta t} + \frac{\overline{pu}_{j+\frac{1}{2}} - \overline{pu}_{j-\frac{1}{2}}}{M_j} \leqslant 0, \text{ with } \overline{pu}_{j+\frac{1}{2}} = \overline{p_{j+\frac{1}{2}}}\ \overline{u_{j+\frac{1}{2}}}. \tag{12}$$

The solution of the correction step (5) verifies the entropy inequality

$$\forall j, \quad \frac{S_j^{n+1} - S_j^n}{\Delta t} \geqslant 0. \tag{13}$$

The organization of this article is as follows. In Section 2 we write the scheme (4) under the form (8). In Section 3, we prove Theorem 4. The proof is split in several parts. On the one side we rapidly prove the uniqueness of a solution, and on the other side we decompose the proof of the existence in different steps. In Section 4, we apply Theorem 4 for the isentropic Euler equations, and prove Corollary 5. In Section 5, the correction step is introduced and the complete scheme is fully justified. The proof of Theorem 6 concerning entropy inequalities for both steps is detailed. In Section 6, we provide a few numerical illustrations. The Appendix contains the application of Theorem 4 for the traffic flow problem, in the specific case where $m = 0$ in the definition of $\mathcal{D}$. It also contains a brief description of the modification to treat an isothermal equation of state.

2 Formulation under the form (8)

The objective of this Section is to provide the details of the transformation from the implicit scheme (4) to the form (8). The verification of Hypothesis 1, 2 and 3 will be performed in Section 4.

We consider Euler isentropic equations in one dimension (2) for compressible perfect gas with periodic boundary conditions. As the physical entropy S is constant during this prediction step, its equation is not necessary for the proof. Replacing the fluxes and rearranging the terms in (4), one obtains

$$\begin{aligned} &\frac{2M_j}{\Delta t}(\overline{\tau_j} - \tau_j^n) + \frac{1}{\alpha_{j+\frac{1}{2}}^n}(\overline{p_{j+1}} - \overline{p_j}) + \frac{1}{\alpha_{j-\frac{1}{2}}^n}(\overline{p_{j-1}} - \overline{p_j}) = \overline{u_{j+1}} - \overline{u_{j-1}}, \\ &\frac{2M_j}{\Delta t}(\overline{u_j} - u_j^n) + \alpha_{j+\frac{1}{2}}^n(\overline{u_j} - \overline{u_{j+1}}) + \alpha_{j-\frac{1}{2}}^n(\overline{u_j} - \overline{u_{j-1}}) = \overline{p_{j-1}} - \overline{p_{j+1}}. \end{aligned} \tag{14}$$

Let us define a vector of unknowns

$$U = ((-p_j)_{j\in\{1,\dots,N\}}, (u_j)_{j\in\{1,\dots,N\}}) \in \mathcal{D}, \tag{15}$$

where the domain $\mathcal{D}$ is defined as

$$\mathcal{D} = \{U \text{ such that } \forall j \in \{1,\dots,N\} \ -p_j < 0, \text{ and } u_j \in \mathbb{R}\}. \tag{16}$$

The matrix A is defined by

$$A = \left[\begin{array}{c|c} 0 & \mathrm{B} \\ \hline \mathrm{B} & 0 \end{array}\right] \text{ with } B = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & -1 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 0 & 1 \\ 1 & 0 & \cdots & 0 & -1 & 0 \end{bmatrix} \in \mathcal{M}_N(\mathbb{R}). \tag{17}$$

The functional $J : \mathcal{D} \to \mathbb{R}$ is defined as a sum of elementary functions

$$J(U) = \sum_{j=1}^{N} \frac{2M_j}{\Delta t} \left[L_j^1(-p_j) + L_j^2(u_j)\right] + \sum_{j=1}^{N} \left[Q_j^1(-p_j, -p_{j+1}) + Q_j^2(u_j, u_{j+1})\right], \tag{18}$$

where the elementary functions are

$$L_j^1(-p) = -C_j p^{1-\frac{1}{\gamma}} + \tau_j^n p, \text{ where } C_j = \gamma(\gamma-1)^{-1+\frac{1}{\gamma}} \exp\left(\frac{S_j}{C_v}\right)^{\frac{1}{\gamma}} > 0,$$

$$Q_j^1(-p,-q) = \frac{1}{\alpha_{j+\frac{1}{2}}} \frac{(q-p)^2}{2}, \ L_j^2(u) = \frac{u^2}{2} - u_j^n u, \ Q_j^2(u,v) = \alpha_{j+\frac{1}{2}} \frac{(u-v)^2}{2}.$$

Proposition 7. *The calculation of a solution $(\overline{\tau_j} > 0, \overline{u_j})_{1 \leqslant j \leqslant N}$ to the system (14) of scalar non-linear equations is equivalent to the calculation of a solution $U \in \mathcal{D}$ to the global non-linear equation $\nabla J(U) = AU$.*

Proof. For a perfect gas law, the correspondence between τ, p and S can be written as $\tau = (\gamma - 1)\exp\left(\frac{S}{C_v}\right)^{\frac{1}{\gamma}} p^{-\frac{1}{\gamma}}$. Therefore the equivalence between a solution of (14) and a solution (15) of $\nabla J(U) = AU$ is explicited by

$$\overline{\tau_j} = (\gamma-1)\exp\left(\frac{S_j}{C_v}\right)^{\frac{1}{\gamma}} p_j^{-\frac{1}{\gamma}} \text{ and } \overline{u_j} = u_j. \tag{19}$$

To finish the proof it is sufficient to calculate explicitly $\nabla J(U)$. The derivatives of L_j^1, L_j^2, Q_j^1 and Q_j^2 are

$$\frac{\partial L_j^1}{\partial(-p_j)} = C_j\left(\frac{\gamma-1}{\gamma}\right) p_j^{-\frac{1}{\gamma}} - \tau_j^n = \tau_j - \tau_j^n, \quad \frac{\partial L_j^2}{\partial u_j} = u_j - u_j^n,$$

$$\frac{\partial Q_j^1}{\partial(-p_j)} = \frac{1}{\alpha_{j+\frac{1}{2}}}(p_{j+1} - p_j) + \frac{1}{\alpha_{j-\frac{1}{2}}}(p_{j-1} - p_j),$$

$$\frac{\partial Q_j^2}{\partial u_j} = \alpha_{j+\frac{1}{2}}(u_j - u_{j+1}) + \alpha_{j-\frac{1}{2}}(u_j - u_{j-1}).$$

By (18), one calculates all the components of the vector $\nabla J(U) \in \mathbb{R}^{2N}$. With the definition (17) of the matrix A, one obtains immediately that the first N equations in the vectorial identity $\nabla J(U) = AU$ are

$$\frac{2M_j}{\Delta t}(\tau_j - \tau_j^n) + \frac{1}{\alpha_{j+\frac{1}{2}}^n}(p_{j+1} - p_j) + \frac{1}{\alpha_{j-\frac{1}{2}}^n}(p_{j-1} - p_j) = u_{j+1} - u_{j-1}, \quad 1 \leqslant j \leqslant N, \tag{20}$$

while the last N equations are

$$\frac{2M_j}{\Delta t}(u_j - u_j^n) + \alpha_{j+\frac{1}{2}}^n(u_j - u_{j+1}) + \alpha_{j-\frac{1}{2}}^n(u_j - u_{j-1}) = p_{j-1} - p_{j+1}, \quad 1 \leqslant j \leqslant N. \tag{21}$$

With the correspondence (19), one finds that (20 -21) is equal to (14). □

3 Proof of Theorem 4

In this Section, we prove Theorem 4 stated in the introduction under the Hypothesis 1, 2 and 3. Each Subsection corresponds to an intermediate result leading to the final outcome.
In convex analysis, see e.g.[Hirriart-Urruty and Lemarechal[18] Def. 3.2.5 p19], the closure of the function J is $\overline{J}$, defined as

$$\begin{aligned} \overline{J}:\mathbb{R}^{n+m} &\to \overline{\mathbb{R}} \\ U &\mapsto \begin{cases} \lim\limits_{V\to U}\ \inf\limits_{V\in\mathcal{D}} J(V) & \text{if } U\in\overline{\mathcal{D}}, \\ +\infty & \text{if not.} \end{cases} \end{aligned} \tag{22}$$

By construction, $\overline{J}$ is lower semi-continuous because J is continuous over $\mathcal{D}$. For a function J which is coercive on its domain $\mathcal{D}$ like in (9), the closure $\overline{J}$ is also coercive in the sense of the book of Hirriart-Urruty,[17] [Chapter 2, p 41]

$$\overline{J}(U)\to+\infty \text{ for } ||U|| \xrightarrow[U\in\mathbb{R}^{n+m}]{} +\infty.$$

3.1 Uniqueness

It relies on elementary considerations.

Lemma 8. *Assuming that the problem (8) admits a solution in $\mathcal{D}$, then it is unique.*

Proof. Let $U_1\in\mathcal{D}$ and $U_2\in\mathcal{D}$ be two solutions of the problem (8)

$$\begin{cases} \nabla J(U_1) = AU_1, \\ \nabla J(U_2) = AU_2. \end{cases}$$

One has

$$(\nabla J(U_1)-\nabla J(U_2), U_1-U_2) = (A(U_1-U_2), U_1-U_2).$$

Since A is a skew-symmetric matrix therefore $(A(U_1-U_2),U_1-U_2)=0$, that is

$$(\nabla J(U_1)-\nabla J(U_2), U_1-U_2) = 0.$$

Since J is strictly convex, this is only satisfied if $U_1=U_2$. □

3.2 Existence

The existence of a solution relies on a few intermediate results.

3.2.1 Existence of a minimum for J

The first result to prove is the existence of a minimum point for the function J, using a classical result from convex analysis.

Lemma 9. *The function J admits a unique minimum $U\in\mathcal{D}$.*

Proof. We apply [Theorem 1.1 p 48] of Dacorogna[7] to the function $f=\overline{J}$. So there exists $U\in\mathbb{R}^{n+m}$ such that $\overline{J}(U)\leqslant\overline{J}(V)$ for all $V\in\mathbb{R}^{n+m}$. Necessarily $\overline{J}(U)<\infty$ is finite so $U\in\overline{\mathcal{D}}$. It remains to show that $U\notin\partial\mathcal{D}$.

Let us assume on the contrary that $U\in\partial\mathcal{D}$. Thanks to the convexity of $\overline{J}$ and inequality (10) in Hypothesis 2, one can write

$$\overline{J}(U)\geqslant\overline{J}(U-\varepsilon\boldsymbol{d})+\varepsilon\left(\nabla J(U-\varepsilon\boldsymbol{d}),\boldsymbol{d}\right)>\overline{J}(U-\varepsilon\boldsymbol{d}).$$

It is a contradiction. Therefore $U\in\mathcal{D}$ is a minimum and U is unique thanks to the strict convexity of J on $\mathcal{D}$. □

Since $\mathcal{D}$ is an open set, the unique minimum $U\in\mathcal{D}$ of J satisfies the Euler equation, see Hirriart-Urruty,[17] [Chapter 2, p 41]

$$\nabla J(U)=0.$$

3.2.2 A continuation method

We prove in this section that, the problem (23) admits a solution in the domain $\mathcal{D}$ for all $0 \leqslant \varepsilon \leqslant 1$

$$\begin{cases} \text{Find } U_\varepsilon \in \mathcal{D} \text{ such that} \\ \nabla J(U_\varepsilon) = \varepsilon A U_\varepsilon. \end{cases} \tag{23}$$

For $\varepsilon = 0$, the problem is treated in Section 3.2.1. This allows to write (23) with a continuation method under the form of an initial value problem (24)

$$\begin{cases} \nabla^2 J(U_\varepsilon) \dfrac{dU_\varepsilon}{d\varepsilon} = AU_\varepsilon + \varepsilon A \dfrac{dU_\varepsilon}{d\varepsilon}, \\ U_0 = \underset{U \in \mathcal{D}}{\operatorname{argmin}}\ J(U). \end{cases} \tag{24}$$

A rearrangement yields $(\nabla^2 J(U_\varepsilon) - \varepsilon A) \frac{dU_\varepsilon}{d\varepsilon} = AU_\varepsilon$. For $U \in \mathcal{D}$, the matrix $\nabla^2 J(U) - \varepsilon A$ is invertible thanks to the following result.

Lemma 10. *Let A and B be two matrices of $\mathcal{M}_{N,N}(\mathbb{R})$ such that A is skew symmetric and B is positive definite. Then the matrix $C = A + B \in \mathcal{M}_{N,N}(\mathbb{R})$ is invertible.*

The Lemma 10 is applied with $-\varepsilon A$ as the skew symmetric matrix, and $\nabla^2 J(U_\varepsilon)$ as the positive definite matrix. Thus $(\nabla^2 J(U_\varepsilon) - \varepsilon A)^{-1}$ exists. The initial value problem can be rewritten as

$$\begin{cases} \dfrac{dU_\varepsilon}{d\varepsilon} = (\nabla^2 J(U_\varepsilon) - \varepsilon A)^{-1} AU_\varepsilon, \\ U_0 = \underset{U \in \mathcal{D}}{\operatorname{argmin}}\ J(U). \end{cases}$$

Let us define $\mathcal{I} = [0, +\infty[$ and the function $F : \mathcal{I} \times \mathcal{D} \to \mathbb{R}^{n+m}$ by $F(\varepsilon, V) = (\nabla^2 J(V) - \varepsilon A)^{-1} AV$. The problem is rewritten as

$$\begin{cases} \dfrac{dU_\varepsilon}{d\varepsilon} = F(\varepsilon, U_\varepsilon), \\ U_0 = \underset{U \in \mathcal{D}}{\operatorname{argmin}}\ J(U). \end{cases} \tag{25}$$

To obtain the existence of a maximal solution to (25), one can apply standard results from the theory of ODEs that we recall now.

Definition 11 (see Coddington and Levinson,[5] Chapter 1)**.** *Let $N \in \mathbb{N}$ and $F : \mathcal{I} \times \mathbb{R}^N \to \mathbb{R}^N$, let $\varepsilon_0 \in \mathcal{I}$, and $U_\varepsilon \in \mathbb{R}^N$ where $\mathcal{I}$ is a non empty interval of $\mathbb{R}$. A solution of the differential equation*

$$U'(\varepsilon) = F(\varepsilon, U(\varepsilon)), \tag{26}$$

is given by a non empty interval $I \subset \mathcal{I}$ and a differentiable function $U : I \to \mathbb{R}^N$ satisfying (26) for all $\varepsilon \in I$.

A solution of the initial value problem (or Cauchy problem) associated to (26) is a solution of (26) such that $\varepsilon_0 \in I$ and $U(\varepsilon_0) = U_0$.

Theorem 12 (Cauchy Lipschitz Theorem for locally Lipschitz functions,[5] Th. 3.1, p12)**.** *Let the function F be a $\mathcal{C}^1$ function, then, for all initial data $(\varepsilon_0, U_0) \in \mathcal{I} \times \mathbb{R}^N$, there exists an interval $I \in \mathcal{I}$ containing ε_0 such that there exists in I a unique solution to the associated initial value problem.*

In particular for all such data, there exists a unique maximal solution and all other solution verifying the condition of Cauchy is a restriction of the maximal solution.

Lemma 13. *There exists $0 < \varepsilon_{max}$ such that for all $\varepsilon \in [0, \varepsilon_{max}[$, the problem (25) admits a solution in $\mathcal{D}$. This solution satisfies (23).*

Proof. One applies the Cauchy Lipschitz Theorem. The function F is well defined, and differentiable in terms of ε. The derivative is continuous because A is a matrix of scalar coefficients, and J is a function of class $\mathcal{C}^2$ thanks to Hypothesis 2. F is then of class $\mathcal{C}^1$.

To prove the last part of the Lemma, one notes that

$$\frac{d}{d\varepsilon}\left(\nabla J(U_\varepsilon) - \varepsilon A U_\varepsilon\right) = (\nabla^2 J(U_\varepsilon) - \varepsilon A)\frac{d}{d\varepsilon}U_\varepsilon - AU_\varepsilon = 0.$$

Since $\nabla J(U_0) = 0$ it shows that $\nabla J(U_\varepsilon) - \varepsilon A U_\varepsilon = 0$ on the maximal interval. □

In the rest of this Section, we prove that $\varepsilon_{max} > 1$.

3.2.3 Upper bound of $J(U_\varepsilon)$

In this section, the objective is to prove that $J(U_\varepsilon)$ is bounded, which is necessary to conclude that the solution of the problem (23) stays in the domain $\mathcal{D}$.

Lemma 14. *There exists $\overline{U} \in \mathcal{D}$ such that the following inequality is satisfied on the maximal interval*

$$J(U_\varepsilon) \leqslant J(\overline{U}) < +\infty.$$

Proof. Let us take $\overline{U} \in \ker(A) \cap \mathcal{D}$ that is non empty by Hypothesis 3. The convexity of J implies that

$$J(U_\varepsilon) + (\nabla J(U_\varepsilon), \overline{U} - U_\varepsilon) \leqslant J(\overline{U}).$$

One finds

$$J(U_\varepsilon) + (\varepsilon A U_\varepsilon, \overline{U}) - (\varepsilon A U_\varepsilon, U_\varepsilon) \leqslant J(\overline{U}).$$

The matrix A is skew symmetric, hence $(AU_\varepsilon, U_\varepsilon) = 0$. So

$$J(U_\varepsilon) + \varepsilon(AU_\varepsilon, \overline{U}) \leqslant J(\overline{U}).$$

Using again the property of skew symmetry, one has $\varepsilon(AU_\varepsilon, \overline{U}) = -\varepsilon(A\overline{U}, U_\varepsilon)$. As $\overline{U} \in \ker(A) \cap \mathcal{D}$, hence $(\varepsilon A U_\varepsilon, \overline{U}) = 0$. So $J(U_\varepsilon) \leqslant J(\overline{U}) < +\infty$. □

In addition, since J is a coercive function by Hypothesis 2, there exists $K < +\infty$ such that

$$||U_\varepsilon|| < K \tag{27}$$

for all ε in the maximal interval.

3.2.4 End of the proof of Theorem 4

The end of the proof of Theorem 4 is based on the following standard result.

Theorem 15 (see Demailly,[8] Chapter 5, p 138)**.** *Let Ω be an open domain of $\mathbb{R} \times \mathbb{R}^m$ and $U : I = [t_0, b[\to \mathbb{R}^m$ a solution of the equation (E) $U' = F(t, U)$ where F is continuous on Ω. So $U(t)$ can be continuated further than b if and only if there exists a compact $C \subset \Omega$ such that the curve $t \mapsto (t, U(t)), t \in [t_0, b[$, stays in C.*

For our problem, one has the Property.

Proposition 16. *There exists a compact $C \subset \mathcal{D}$ such that $U_\varepsilon \in C$ for all $\varepsilon \in [0, \min(\varepsilon_{max}, 2)[$.*

Proof. Thanks to (27), U_ε is $\mathcal{D} \cap B(0, K)$. Moreover, $\nabla J(U_\varepsilon) = \varepsilon A U_\varepsilon$, so one has $||\nabla J(U_\varepsilon)|| \leqslant 2||A||K$. Let us consider

$$C = \mathcal{D} \cap B(0, K) \cap \{V \in \mathcal{D} \text{ such that } ||\nabla J(V)|| \leqslant 2||A||K\}.$$

It remains to prove that C is a compact of $\mathcal{D}$. Let us take a sequence $V_n \in C$ for $n \in \mathbb{N}$. Since (V_n) is bounded, there exists $V \in B(0, K)$ and a subsequence still denoted V_n such that $V_n \to V$. Necessarily $V \in \overline{\mathcal{D}}$, so either $V \in \partial\mathcal{D}$ or $V \in \mathcal{D}$.

Let us assume that $V \in \partial\mathcal{D}$. Thanks to Hypothesis 2, inequality (11), one has $||\nabla J(V_n)|| \to +\infty$. It is a contradiction with the definition of C. Therefore $V \in \mathcal{D}$. Since J is C^2, ∇J is a continuous function and $||\nabla J(V)|| \leqslant 2||A||K$. So $V \in C$, which shows that C is a compact of $\mathcal{D}$. □

Proof of Theorem 4. Thanks to Proposition 16 and Theorem 15, one has that $\varepsilon_{max} > 1$. Therefore, one takes $\varepsilon = 1$ which concludes the proof of existence of the solution of (8). □

4 Proof of the unconditional stability of Corollary 5

We prove the unconditional stability for the prediction step corresponding to the implicit discretization of the isentropic Euler equations. In this purpose we apply Theorem 4 after a precise check of all the different hypothesis. The definitions of J, A and U are given in Section 2.

4.1 Properties of J

We verify all the required properties of Hypothesis 2, that is the strict convexity of J, its coercivity and its limits at the boundary of the domain $\mathcal{D}$.

Property 1. *The function J (18) is continuous on $\mathcal{D}$.*

Property 2. *The function J (18) is strictly convex on $\mathcal{D}$.*

The proof is easily verified, but we detail the calculations.

Proof. The second derivatives, for all j and $k \in \{1, \dots, N\}$ are

$$\frac{\partial^2 L_j^1}{\partial(-p_j)^2} = C_j \left(\frac{\gamma-1}{\gamma^2}\right) p_j^{-1-\frac{1}{\gamma}}, \quad \frac{\partial^2 L_j^1}{\partial(-p_j)\partial(-p_k)} = 0, \quad \frac{\partial^2 L_j^2}{\partial u_j^2} = 1, \quad \frac{\partial^2 L_j^2}{\partial u_j \partial u_k} = 0.$$

$$\frac{\partial^2 Q_j^1}{\partial(-p_j)\partial(-p_k)} = \begin{cases} \dfrac{1}{\alpha_{j+\frac{1}{2}}} + \dfrac{1}{\alpha_{j-\frac{1}{2}}}, & \text{if } k = j, \\ -\dfrac{1}{\alpha_{j-\frac{1}{2}}}, & \text{if } k = j-1, \\ -\dfrac{1}{\alpha_{j+\frac{1}{2}}}, & \text{if } k = j+1, \\ 0, & \text{otherwise.} \end{cases},$$

$$\frac{\partial^2 Q_j^2}{\partial(u_j)\partial(u_k)} = \begin{cases} \alpha_{j+\frac{1}{2}} + \alpha_{j-\frac{1}{2}}, & \text{if } k = j, \\ -\alpha_{j-\frac{1}{2}}, & \text{if } k = j-1, \\ -\alpha_{j+\frac{1}{2}}, & \text{if } k = j+1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, for all $U \in \mathcal{D}$, and for all $Z \in \mathbb{R}^{2N}$, one has

$$\begin{aligned}(\nabla^2 J(U)Z, Z) = &\left(\frac{\partial^2 L_j^1}{\partial(-p_j)^2} + \frac{1}{2}\frac{\partial^2 Q_j^1}{\partial(-p_j)^2}\right)(-p_j^Z)^2 + \left(\frac{\partial^2 L_j^2}{\partial u_j^2} + \frac{1}{2}\frac{\partial^2 Q_j^2}{\partial u_j^2}\right)(u_j^Z)^2 \\ &+ \sum_{j=1}^{N} \frac{1}{2}\left(\frac{1}{\alpha_{j+\frac{1}{2}}} + \frac{1}{\alpha_{j-\frac{1}{2}}}\right)(p_{j+1}^Z - p_j^Z)^2 + \sum_{j=1}^{N} \frac{1}{2}(\alpha_{j+\frac{1}{2}} + \alpha_{j-\frac{1}{2}})(u_j^Z - u_{j+1}^Z)^2.\end{aligned}$$

For $Z \neq 0$, one has $(\nabla^2 J(U)Z, Z) > 0$. Therefore the function J is strictly convex on $\mathcal{D}$. □

Property 3. *The function J is coercive on $\mathcal{D}$.*

Proof. The function J (18) is the sum of elementary functions L_j^1, L_j^2, Q_j^1 and Q_j^2. The quadratic functions Q_j^1 and Q_j^2 are clearly bounded from below, as well as L_j^2. The function L_j^1 is also bounded from below because $C_j > 0$, $\tau_j^n > 0$, $\gamma > 1$ and $p^{1-\frac{1}{\gamma}}$ is dominated by p for $p \to +\infty$. So L_j^1 is coercive. It is evident that L_j^2 is coercive. Since J is defined by a sum over all indices $1 \leqslant j \leqslant N$, then J is coercive.

□

Property 4. *For all $V \in \partial\mathcal{D}$, J defined by (18) satisfies the limits (10) and (11).*

Proof. The first derivative of J with respect to $-p$ is

$$\frac{\partial J}{\partial(-p_j)} = C_j\left(\frac{\gamma-1}{\gamma}\right)p_j^{-\frac{1}{\gamma}} - \tau_j^n + \frac{1}{\alpha_{j+\frac{1}{2}}}(p_{j+1}-p_j) + \frac{1}{\alpha_{j-\frac{1}{2}}}(p_{j-1}-p_j).$$

Let $V \in \partial\mathcal{D}$. It means that there exists a non empty subset $K \subset \{1,\dots,N\}$ such that for all $k \in K$, $V_k = 0$. One takes as the unit outward direction $\boldsymbol{d} \in \mathbb{R}^{N+N}$, such that $\boldsymbol{d}_k = 1$ and for all $j \notin K$, $\boldsymbol{d}_j = 0$. The limit of the first derivative of J when $\varepsilon \to 0^+$ is

$$\begin{aligned}\lim_{\varepsilon\to 0^+}\frac{\partial J}{\partial(-p_k)}|_{V-\varepsilon\boldsymbol{d}} &= \lim_{\varepsilon\to 0^+} -\tau_k^n + C_k\left(\frac{\gamma-1}{\gamma}\right)\frac{1}{\varepsilon^{\frac{1}{\gamma}}} + \frac{1}{\alpha_{k+\frac{1}{2}}}(p_{k+1}-\varepsilon) + \frac{1}{\alpha_{k-\frac{1}{2}}}(p_{k-1}-\varepsilon),\\ &= +\infty.\end{aligned}$$

Indeed, $\lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon^{\frac{1}{\gamma}}} = +\infty$, and all other limits are finite. By summation over $k \in K$, and then over all indices for which the value of $\boldsymbol{d}$ is 0, one obtains (10). An evaluation of $\lim_{W\to V}\frac{\partial J}{\partial(-p_k)}|_W$ easily gives (11). □

4.1.1 Properties of matrix A

Let us prove Hypothesis 3 holds.

Property 5. *The matrix A defined by (17) is skew-symmetric (by construction).*

Property 6. *The kernel of A and the domain $\mathcal{D}$ intersect.*

Proof. We first evaluate the kernel of the matrix A. Let $X \in \mathbb{R}^{2N}$ satisfying $AX = 0$, so

$$\begin{cases} x_{N+2} - x_{2N} = 0,\\ -x_{N+1} + x_{N+3} = 0,\\ \vdots\\ -x_{2N-2} + x_{2N} = 0,\\ x_{N+1} - x_{2N-1} = 0,\\ x_2 - x_N = 0,\\ -x_1 + x_3 = 0,\\ \vdots\\ -x_{N-2} + x_N = 0,\\ x_1 - x_{N-1} = 0.\end{cases} \iff \begin{cases} x_{N+2} = x_{2N},\\ x_{N+1} = x_{N+3},\\ \vdots\\ x_{2N-2} = x_{2N},\\ x_{N+1} = x_{2N-1},\\ x_2 = x_N,\\ x_1 = x_3,\\ \vdots\\ x_{N-2} = x_N,\\ x_1 = x_{N-1}.\end{cases}$$

One obtains

$$\ker(A) = \begin{cases} \mathrm{Vect}\left\{\begin{pmatrix}\mathbb{1}_0\\ \mathbb{0}\end{pmatrix}, \begin{pmatrix}\mathbb{1}^0\\ \mathbb{0}\end{pmatrix}, \begin{pmatrix}\mathbb{0}\\ \mathbb{1}_0\end{pmatrix}, \begin{pmatrix}\mathbb{0}\\ \mathbb{1}^0\end{pmatrix}\right\}, & \text{if } N \in \mathbb{R} \text{ even},\\ \mathrm{Vect}\left\{\begin{pmatrix}\mathbb{1}\\ \mathbb{0}\end{pmatrix}, \begin{pmatrix}\mathbb{0}\\ \mathbb{1}\end{pmatrix}\right\}, & \text{if } N \in \mathbb{R} \text{ odd},\end{cases}$$

where $\mathbb{1}_0 = (1,0,1,\dots,0,1,0)$, $\mathbb{1}^0 = (0,1,0,\dots,1,0,1)$, $\mathbb{1} = (1,1,\dots,1,1)$, and $\mathbb{0} = (0,0,\dots,0,0)$.

One takes $U = (-1,\dots,-1,0,\dots,0)$, then U is in $\ker(A) \cap \mathcal{D}$. □

4.2 Proof of Corollary 5

Corollary 5 . *Considering physically admissible data ($\tau_j > 0$ and $p_j > 0$), the prediction scheme (4) can be written under the form (8). Therefore, it is unconditionally stable.*

Proof. Let $m = n = N$, $U = ((-p_j)_{j\in\{1,\dots,N\}}, (u_j)_{j\in\{1,\dots,N\}})$, A and J as defined by (17) and (18). All the hypothesis of Theorem 4 are satisfied. The existence and uniqueness of a solution to the implicit isentropic Euler scheme for all time step is proved. □

5 Proof of the entropy inequalities (Theorem 6)

In this Section, we study the two steps of the predictor-corrector scheme, and prove the stability inequality in each step.

We recall the complete scheme

$$\text{Prediction step}\begin{cases}\overline{\tau_j} = \tau_j^n + \dfrac{\Delta t}{M_j}(\overline{u_{j+\frac12}} - \overline{u_{j-\frac12}}),\\ \overline{u_j} = u_j^n - \dfrac{\Delta t}{M_j}(\overline{p_{j+\frac12}} - \overline{p_{j-\frac12}}),\\ \overline{S_j} = S_j^n,\end{cases}$$

$$\text{Correction step}\begin{cases}\tau_j^{n+1} = \tau_j^n + \dfrac{\Delta t}{M_j}(\overline{u_{j+\frac12}} - \overline{u_{j-\frac12}}),\\ u_j^{n+1} = u_j^n - \dfrac{\Delta t}{M_j}(\overline{p_{j+\frac12}} - \overline{p_{j-\frac12}}),\\ E_j^{n+1} = E_j^n + \dfrac{\Delta t}{M_j}(\overline{p_{j+\frac12}}\ \overline{u_{j+\frac12}} - \overline{p_{j-\frac12}}\ \overline{u_{j-\frac12}}).\end{cases}$$

Each inequality is studied separately.

5.1 Proof of stability inequality (12)

Actually, inequality (12) is a mathematical entropy inequality. Usually, entropy stability is defined for continuous in time problems, or for explicit in time schemes. A continuous definition is found in the book,[29] [Th. 3.3, p 27], and a discrete version of entropy stability is written in the book,[11] [Prop 2.25, p 66].

Definition 17. *The implicit conservative scheme (4) is said entropic if there exists a mathematical entropy pair (η, ψ) and a numerical entropy flux $\Phi(U, V)$ (such that $\Phi(U, U) = \psi(U)$) such that the following inequality holds for all j*

$$\frac{\eta(\overline{U}_j) - \eta(U_j^n)}{\Delta t} + \frac{\Phi(\overline{U}_j, \overline{U}_{j+1}) - \Phi(\overline{U}_{j-1}, \overline{U}_j)}{M_j} \leqslant 0. \tag{28}$$

We use this definition onto the scheme of the prediction step (4) where the vector of unknowns is $U = (\tau, u, S)^t$. The entropy corresponding to the isentropic system is $\eta(U) = E$. As η is convex, one gets $\eta(\overline{U_j}) - \eta(U_j^n) \leqslant \nabla\eta(\overline{U_j}) \cdot (\overline{U_j} - U_j^n)$, that is $\eta(\overline{U_j}) - \eta(U_j^n) \leqslant \frac{\Delta t}{M_j}\nabla\eta(\overline{U_j}) \cdot (\overline{f_{j+\frac12}} - \overline{f_{j-\frac12}})$.

In Lagrangian formalism, the calculus can be performed easily. Thanks to Gibbs formula, see,[29] [Chapter 4, p 318], one has

$$d\eta = udu - pd\tau + TdS.$$

So

$$\nabla\eta(\overline{U_j}) = \left(-\overline{p_j}, \overline{u_j}, \overline{T_j}\right)^t.$$

For the flux, the one state solver is

$$\overline{f_{j+\frac12}} = \begin{pmatrix}\overline{u_{j+\frac12}}\\ -\overline{p_{j+\frac12}}\\ 0\end{pmatrix} = \begin{pmatrix}\frac{1}{2\alpha_{j+\frac12}}(\overline{p_j} - \overline{p_{j+1}}) + \frac12(\overline{u_j} + \overline{u_{j+1}})\\ \frac{\alpha_{j+\frac12}}{2}(\overline{u_{j+1}} - \overline{u_j}) - \frac12(\overline{p_j} + \overline{p_{j+1}})\\ 0\end{pmatrix},$$

$$\overline{f_{j-\frac12}} = \begin{pmatrix}\overline{u_{j-\frac12}}\\ -\overline{p_{j-\frac12}}\\ 0\end{pmatrix} = \begin{pmatrix}\frac{1}{2\alpha_{j-\frac12}}(\overline{p_{j-1}} - \overline{p_j}) + \frac12(\overline{u_{j-1}} + \overline{u_j})\\ \frac{\alpha_{j-\frac12}}{2}(\overline{u_j} - \overline{u_{j-1}}) - \frac12(\overline{p_{j-1}} + \overline{p_j})\\ 0\end{pmatrix}.$$

End of the proof of inequality (12). First is the evaluation of $\nabla\eta(\overline{U_j})\cdot\overline{f_{j+\frac{1}{2}}}$, and second $\nabla\eta(\overline{U_j})\cdot\overline{f_{j-\frac{1}{2}}}$. Then the difference between the two expressions is performed. In the rest of the calculations, the time dependence will be omitted to simplify the writing. One has

$$\begin{aligned}
\nabla\eta(U_j)\cdot f_{j+\frac{1}{2}} &= -p_j\left(\frac{1}{2\alpha_{j+\frac{1}{2}}}(p_j-p_{j+1})+\frac{1}{2}(u_j+u_{j+1})\right)\\
&\quad + u_j\left(\frac{\alpha_{j+\frac{1}{2}}}{2}(u_{j+1}-u_j)-\frac{1}{2}(p_j+p_{j+1})\right),\\
&= p_j\frac{1}{2\alpha_{j+\frac{1}{2}}}(p_j+\alpha_{j+\frac{1}{2}}u_j)-u_j\frac{1}{2}(p_j+\alpha_{j+\frac{1}{2}}u_j)\\
&\quad - p_j\frac{1}{2\alpha_{j+\frac{1}{2}}}(-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1})+u_j\frac{1}{2}(-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1}),\\
&= (p_j+\alpha_{j+\frac{1}{2}}u_j)(p_j+\alpha_{j+\frac{1}{2}}u_j)\frac{-1}{2\alpha_{j+\frac{1}{2}}}\\
&\quad + (-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1})(-p_j+\alpha_{j+\frac{1}{2}}u_j)\frac{1}{2\alpha_{j+\frac{1}{2}}},\\
&= -\frac{1}{2\alpha_{j+\frac{1}{2}}}(p_j+\alpha_{j+\frac{1}{2}}u_j)^2+\frac{1}{2\alpha_{j+\frac{1}{2}}}(-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1})(-p_j+\alpha_{j+\frac{1}{2}}u_j).
\end{aligned}$$

and with the same method of calculus

$$\nabla\eta(U_j)\cdot f_{j-\frac{1}{2}} = \frac{1}{2\alpha_{j-\frac{1}{2}}}(-p_j+\alpha_{j-\frac{1}{2}}u_j)^2-\frac{1}{2\alpha_{j-\frac{1}{2}}}(p_{j-1}+\alpha_{j-\frac{1}{2}}u_{j-1})(p_j+\alpha_{j-\frac{1}{2}}u_j).$$

The difference is

$$\begin{aligned}
\nabla\eta(U_j)\cdot(f_{j+\frac{1}{2}}-f_{j-\frac{1}{2}}) &= \nabla\eta(U_j)\cdot f_{j+\frac{1}{2}}-\nabla\eta(U_j)\cdot f_{j-\frac{1}{2}},\\
&= -\frac{1}{2\alpha_{j+\frac{1}{2}}}(p_j+\alpha_{j+\frac{1}{2}}u_j)^2+\frac{1}{2\alpha_{j+\frac{1}{2}}}(-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1})(-p_j+\alpha_{j+\frac{1}{2}}u_j)\\
&\quad -\frac{1}{2\alpha_{j-\frac{1}{2}}}(-p_j+\alpha_{j-\frac{1}{2}}u_j)^2+\frac{1}{2\alpha_{j-\frac{1}{2}}}(p_{j-1}+\alpha_{j-\frac{1}{2}}u_{j-1})(p_j+\alpha_{j-\frac{1}{2}}u_j).
\end{aligned}$$

One has $ab\leqslant\frac{a^2}{2}+\frac{b^2}{2}$, for a and b reals. So

$$\begin{aligned}
\nabla\eta(U_j)\cdot(f_{j+\frac{1}{2}}-f_{j-\frac{1}{2}}) &\leqslant -\frac{1}{2\alpha_{j+\frac{1}{2}}}(p_j+\alpha_{j+\frac{1}{2}}u_j)^2-\frac{1}{2\alpha_{j-\frac{1}{2}}}(-p_j+\alpha_{j-\frac{1}{2}}u_j)^2\\
&\quad +\frac{1}{2\alpha_{j+\frac{1}{2}}}\left[\frac{1}{2}(-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1})^2+\frac{1}{2}(-p_j+\alpha_{j+\frac{1}{2}}u_j)^2\right]\\
&\quad +\frac{1}{2\alpha_{j-\frac{1}{2}}}\left[\frac{1}{2}(p_{j-1}+\alpha_{j-\frac{1}{2}}u_{j-1})^2+\frac{1}{2}(p_j+\alpha_{j-\frac{1}{2}}u_j)^2\right],\\
&\leqslant -\frac{1}{2\alpha_{j+\frac{1}{2}}}(p_j+\alpha_{j+\frac{1}{2}}u_j)^2+\frac{1}{4\alpha_{j+\frac{1}{2}}}(-p_j+\alpha_{j+\frac{1}{2}}u_j)^2\\
&\quad +\frac{1}{4\alpha_{j+\frac{1}{2}}}(-p_{j+1}+\alpha_{j+\frac{1}{2}}u_{j+1})^2\\
&\quad -\frac{1}{2\alpha_{j-\frac{1}{2}}}(-p_j+\alpha_{j-\frac{1}{2}}u_j)^2+\frac{1}{4\alpha_{j-\frac{1}{2}}}(p_j+\alpha_{j-\frac{1}{2}}u_j)^2\\
&\quad +\frac{1}{4\alpha_{j-\frac{1}{2}}}(p_{j-1}+\alpha_{j-\frac{1}{2}}u_{j-1})^2.
\end{aligned}$$

One has the equalities $\frac{1}{2\alpha}(p+\alpha u)^2-\frac{1}{4\alpha}(p-\alpha u)^2=\frac{1}{4\alpha}(p+\alpha u)^2+pu$ and $\frac{1}{2\alpha}(p-\alpha u)^2-\frac{1}{4\alpha}(p+\alpha u)^2=$

$\frac{1}{4\alpha}(p-\alpha u)^2 - pu$. These results are injected in the previous inequality

$$\begin{aligned}\nabla\eta(U_j)\cdot(f_{j+\frac12}-f_{j-\frac12}) \leqslant & -\frac{1}{4\alpha_{j+\frac12}}(p_j+\alpha_{j+\frac12}u_j)^2+\frac{1}{4\alpha_{j+\frac12}}(p_{j+1}-\alpha_{j+\frac12}u_{j+1})^2+p_ju_j\\ & -\frac{1}{4\alpha_{j-\frac12}}(p_j-\alpha_{j-\frac12}u_j)^2+\frac{1}{4\alpha_{j-\frac12}}(p_{j-1}+\alpha_{j-\frac12}u_{j-1})^2-p_ju_j,\\ \leqslant & -\frac{1}{4\alpha_{j+\frac12}}(p_j+\alpha_{j+\frac12}u_j)^2+\frac{1}{4\alpha_{j+\frac12}}(p_{j+1}-\alpha_{j+\frac12}u_{j+1})^2\\ & -\frac{1}{4\alpha_{j-\frac12}}(p_j-\alpha_{j-\frac12}u_j)^2+\frac{1}{4\alpha_{j-\frac12}}(p_{j-1}+\alpha_{j-\frac12}u_{j-1})^2.\end{aligned}$$

One then denotes

$$\Phi(U_j,U_{j+1})=\Phi_{j+\frac12}=\frac{1}{4\alpha_{j+\frac12}}(p_j+\alpha_{j+\frac12}u_j)^2-\frac{1}{4\alpha_{j+\frac12}}(p_{j+1}-\alpha_{j+\frac12}u_{j+1})^2,$$

and

$$\Phi(U_{j-1},U_j)=\Phi_{j-\frac12}=\frac{1}{4\alpha_{j-\frac12}}(p_{j-1}+\alpha_{j-\frac12}u_{j-1})^2-\frac{1}{4\alpha_{j-\frac12}}(p_j-\alpha_{j-\frac12}u_j)^2.$$

Hence

$$\eta(U_j^{n+1})-\eta(U_j^n)\leqslant-\frac{\Delta t}{M_j}(\Phi_{j+\frac12}^{n+1}-\Phi_{j-\frac12}^{n+1}),$$

also written differently as

$$\frac{\eta(U_j^{n+1})-\eta(U_j^n)}{\Delta t}+\frac{\Phi_{j+\frac12}^{n+1}-\Phi_{j-\frac12}^{n+1}}{M_j}\leqslant 0. \tag{29}$$

This corresponds exactly to the entropy inequality because Φ is the entropic flux. As a matter of fact, one remarks

$$\begin{aligned}\Phi(U,U)=\psi(U)= & =\frac{1}{4\alpha}(p+\alpha u)^2-\frac{1}{4\alpha}(p-\alpha u)^2,\\ & =\frac{1}{4\alpha}(p^2+2\alpha pu+\alpha^2u^2-p^2+2\alpha pu-\alpha^2u^2),\\ & =pu.\end{aligned}$$

Therefore as $\eta=E$ and $\Phi=pu$, one rewrites (29) as $\frac{\overline{E_j}-E_j^n}{\Delta t}+\frac{(\overline{pu})_{j+\frac12}-(\overline{pu})_{j-\frac12}}{M_j}\leqslant 0$.

□

5.2 Proof of entropy inequality (13)

To show (13), one starts with (12).

Proof of inequality (13). Let us now denote $r^n=\frac{\overline{E_j}-E_j^n}{\Delta t}+\frac{(\overline{pu})_{j+\frac12}-(\overline{pu})_{j-\frac12}}{M_j}\leqslant 0$. During the correction step, the discretization of the total energy E is

$$\frac{E_j^{n+1}-E_j^n}{\Delta t}=-\frac{((\overline{pu})_{j+\frac12}-(\overline{pu})_{j-\frac12})}{M_j}.$$

The variation of total energy is evaluated between the correction and the prediction step as

$$\begin{aligned}\frac{E_j^{n+1}-\overline{E_j}}{\Delta t} & =\frac{E_j^{n+1}-E_j^n+E_j^n-\overline{E_j}}{\Delta t}\\ & =-\frac{((\overline{pu})_{j+\frac12}-(\overline{pu})_{j-\frac12})}{M_j}-r^n+\frac{((\overline{pu})_{j+\frac12}-(\overline{pu})_{j-\frac12})}{M_j}\\ & =-r^n\geqslant 0.\end{aligned}$$

Since $u^{n+1} = \overline{u}$, one has

$$\frac{E_j^{n+1} - \overline{E_j}}{\Delta t} = \frac{e_j^{n+1} - \overline{e}_j}{\Delta t} = -r^n \geqslant 0. \quad (30)$$

To conclude, it is important to have in mind the Gibbs formula $TdS = de + pd\tau$, where the variable τ is fixed because $\tau^{n+1} = \overline{\tau} = \frac{1}{\overline{\rho}}$. Therefore $S(e, \rho)$ is a growing function of e. Moreover, one has

$$\frac{S_j^{n+1} - S_j^n}{\Delta t} = \frac{S_j^{n+1} - \overline{S}_j}{\Delta t} + \frac{\overline{S}_j - S_j^n}{\Delta t}.$$

During the prediction step $\overline{S_j} = S_j^n$, so

$$\begin{aligned} \frac{S_j^{n+1} - S_j^n}{\Delta t} &= \frac{S_j^{n+1} - \overline{S}_j}{\Delta t}, \\ &= \frac{S(e_j^{n+1}, \overline{\rho}_j) - S(\overline{e}_j, \overline{\rho}_j)}{\Delta t}. \end{aligned}$$

Thus, thanks to (30), one concludes that $\frac{S_j^{n+1} - S_j^n}{\Delta t} \geqslant 0$. □

6 Numerical illustrations

In this Section, we provide some numerical illustrations which show that the theoretical properties of robustness and existence of a solution are transferred to discrete methods.

6.1 Euler equations

For each example, the implicit scheme is compared to the explicit acoustic scheme (31) which provides a reference solution. Some explanations on finite volume methods are found in,[22] or[13] for instance. The implicit scheme is solved using a Newton algorithm. For all our test problems, we have observed that this algorithm converges without any difficulties in only few iterations (less than 10) and we do not comment this issue further.

$$\text{Explicit scheme} \begin{cases} \tau_j^{n+1} = \tau_j^n + \dfrac{\Delta t}{M_j}(u_{j+\frac{1}{2}}^n - u_{j-\frac{1}{2}}^n), \\ u_j^{n+1} = u_j^n - \dfrac{\Delta t}{M_j}(p_{j+\frac{1}{2}}^n - p_{j-\frac{1}{2}}^n), \\ E_j^{n+1} = E_j^n + \dfrac{\Delta t}{M_j}(p_{j+\frac{1}{2}}^n \; u_{j+\frac{1}{2}}^n - p_{j-\frac{1}{2}}^n \; u_{j-\frac{1}{2}}^n). \end{cases} \quad (31)$$

6.1.1 Sod shock tube

For this problem, the initial conditions are

$$p_0(x) = \begin{cases} 1 & x < 0.5 \\ 0.1 & x > 0.5 \end{cases}, \quad \rho_0(x) = \begin{cases} 1 & x < 0.5 \\ 0.125 & x > 0.5 \end{cases}, \quad u_0(x) = 0.$$

The boundary conditions are $u_{left} = u_{right} = 0$. The adiabatic index is $\gamma = 1.4$. The final time of the simulation is $t = 0.2$. Several CFL (0.4, 40, 80, 537) are taken to evaluate the robustness of the scheme. The number of cells is 100.

For a CFL equal to 0.4 the curves are quasi identical in Figure 1.

For an implicit CFL 100 times larger, the numerical smearing is visible in Figure 2 for the rarefaction waves as well as the shock. On the contrary, the contact discontinuity is still at the correct location.

As it can be seen in Figure 3, one observes that when the CFL tends to be very large, there is more numerical dissipation on shocks and rarefaction waves but the contact discontinuity still seems to be at the right position.

The solution of Figure 4 shows the unconditional stability of the implicit scheme.

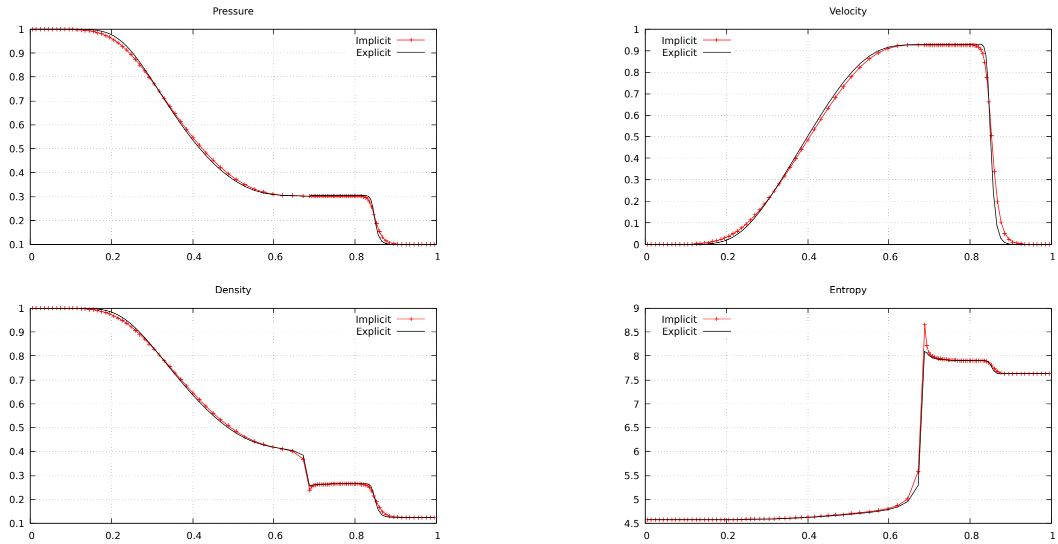

Figure 1: Sod shock tube for Euler equations. The CFL are $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 0.4$.

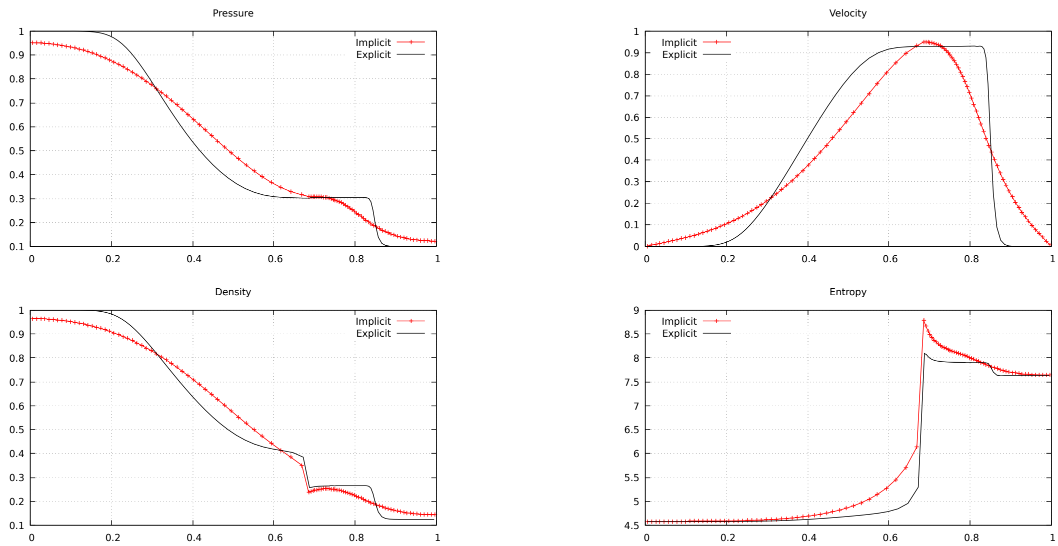

Figure 2: Sod shock tube for Euler equations. The CFL are $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 40$.

A remark can be done on the position of the contact discontinuity that is slightly shifted. To explain the origin of this misplacement, we can evoke the fact that the interaction with the boundaries of the domain is very important. To validate this hypothesis, we performed another calculation (same initial conditions and final time) on a domain 9 times larger. The results are visible in Figure 5, and the contact discontinuity is once again well positioned.

6.2 Position of the contact discontinuity

We develop hereafter a possible explanation for the precision of the position for the contact discontinuity in Figure 5. It deals with the integration of the Riemann problem.

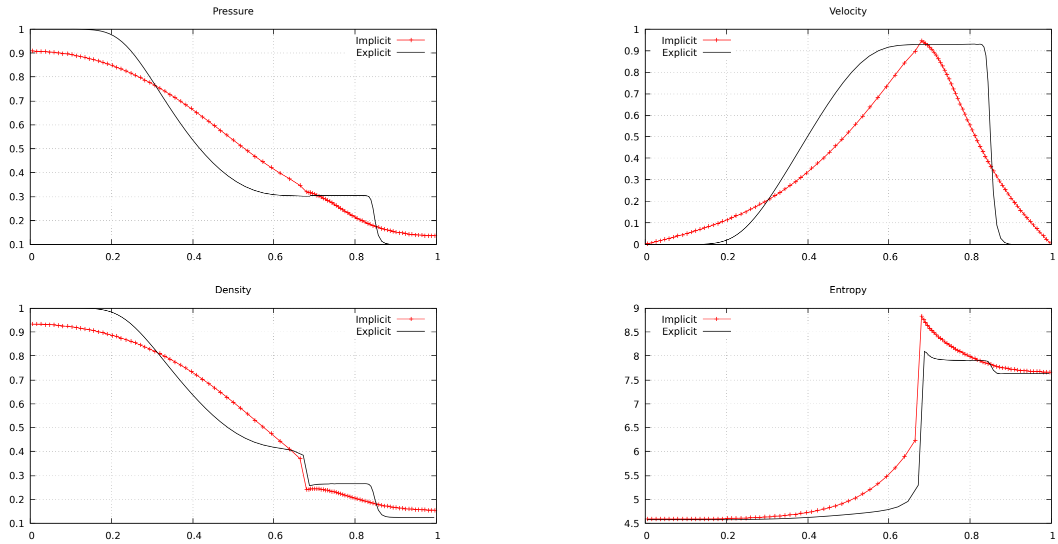

Figure 3: Sod shock tube for Euler equations. The CFL are $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 80$.

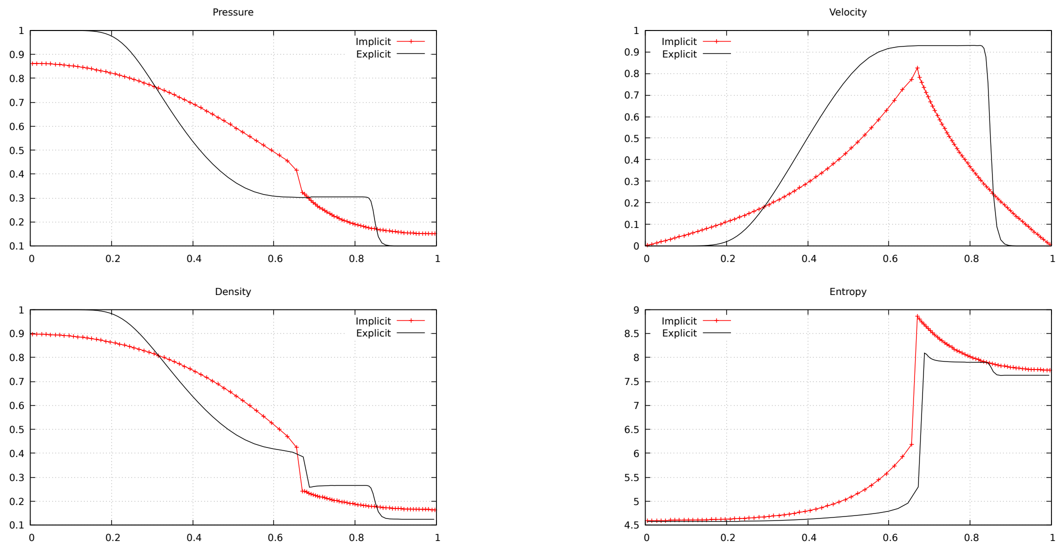

Figure 4: Sod shock tube for Euler equations. The CFL are $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 537$ (only one time step).

Indeed, consider the following initial conditions

$$\tau_0(x) = \begin{cases} \tau_L, & x < 0, \\ \tau_R, & x > 0, \end{cases} \qquad u_0(x) = \begin{cases} u_L, & x < 0, \\ u_R, & x > 0, \end{cases} \qquad S_0(x) = \begin{cases} S_L, & x < 0, \\ S_R, & x > 0. \end{cases}$$

Lagrangian isentropic equations are

$$\begin{cases} \partial_t \tau(x) - \partial_m u(x) = 0, \\ \partial_t u(x) + \partial_m p(\tau(x), S(x)) = 0, \\ \partial_t S(x) = 0. \end{cases} \tag{32}$$

where $dm = \rho dx$.

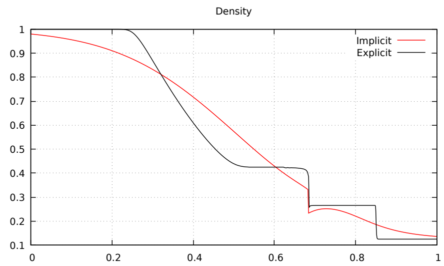

Figure 5: Sod shock tube for Euler equations. Mesh of 9000 cells, domain between $[-4, 5]$. The CFL are $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 537$.

We are looking for a solution of class $\mathcal{C}^0(\mathbb{R}) \cap (\mathcal{C}^1(\mathbb{R}^+) \cap \mathcal{C}^1(\mathbb{R}^-))$, for the variables τ and u. The variable S is discontinuous. The natural boundary conditions are

$$\begin{aligned} \lim_{x\to-\infty} p(x) &= \frac{(\gamma-1)e^{S_L}}{\tau_L^\gamma}, \\ \lim_{x\to+\infty} p(x) &= \frac{(\gamma-1)e^{S_R}}{\tau_R^\gamma}, \\ \lim_{x\to-\infty} \partial_x p(x) &= 0, \\ \lim_{x\to+\infty} \partial_x p(x) &= 0. \end{aligned} \tag{33}$$

The equations (32) are discretized in time but the space part is left continuous, $\Delta t > 0$, $x \in \mathbb{R}$. This mimics the implicit scheme, indeed Δt can be taken extremely big, but the space step Δx is very small. This corresponds of having a discrete Δt compared to a small continuous Δx.

$$\begin{cases} \dfrac{\tau(x)-\tau_0(x)}{\Delta t} - \partial_m u(x) = 0, \\ \dfrac{u(x)-u_0(x)}{\Delta t} + \partial_m p(x) = 0, \\ \qquad\qquad \dfrac{S(x)-S_0(x)}{\Delta t} = 0. \end{cases} \tag{34}$$

Lemma 18. *The system (34) is self similar in* $\frac{x}{\Delta t}$.

Proof. Writing $y = \frac{x}{\Delta t}$, then $dx = \Delta t dy$. One has

$$\begin{aligned} \tau(x) &= \hat{\tau}(\frac{x}{\Delta t}), \text{ so } \hat{\tau}(y) - \tau_0(y) - \frac{1}{\rho}\partial_y \hat{u}(y) = 0, \\ u(x) &= \hat{u}(\frac{x}{\Delta t}), \text{ so } \hat{u}(y) - u_0(y) + \frac{1}{\rho}\partial_y \hat{p}(\hat{\tau}(y), \hat{S}(y)) = 0, \\ S(x) &= \hat{S}(\frac{x}{\Delta t}), \text{ so } \hat{S}(y) - S_0(y) = 0. \end{aligned}$$

□

The method of calculation of a solution to (34) is detailed after.

For $x < 0$ the solution satisfies the equations

$$\begin{cases} \tau(x) - \tau_L - \dfrac{1}{\rho_L}\partial_x u(x) = 0, \\ u(x) - u_L + \dfrac{1}{\rho_L}\partial_x p(x) = 0. \end{cases}$$

The variable u is derived from the second equation and injected in the first equation to obtain

$$\tau(x) - \tau_L + \frac{1}{\rho_L^2} p''(x) = 0.$$

Introducing the enthalpy $H(p,S) = e + p\tau$, one has $dH = de + pd\tau + \tau dp = TdS + \tau dp$, hence

$$\frac{\partial H}{\partial p}(p, S_L) - \tau_L + \frac{1}{{\rho_L}^2} p''(x) = 0.$$

Factorizing, one gets

$$\frac{1}{{\rho_L}^2} p''(x) + \frac{\partial}{\partial p}\left(H(p, S_L) - p\tau_L\right) = 0.$$

One obtains

$$\frac{\partial}{\partial_x}\left(\frac{1}{{\rho_L}^2}\frac{(p'(x))^2}{2} + H(p, S_L) - p\tau_L\right) = 0.$$

Therefore

$$\frac{1}{{\rho_L}^2}\frac{(p'(x))^2}{2} + H(p, S_L) - p\tau_L = K_L.$$

Using the boundary conditions (33), the integration constant is $K_L = e_L$.

$$\frac{1}{{\rho_L}^2}\frac{(p'(x))^2}{2} + H(p, S_L) - p\tau_L - e_L = 0.$$

So

$$\frac{p'(x)^2}{2{\rho_L}^2} = -H(p, S_L) + p\tau_L + e_L.$$

One finally obtains

$$p'(x) = \pm\rho_L\sqrt{-2H(p, S_L) + 2p\tau_L + 2e_L}. \tag{35}$$

For $x > 0$, the solution verifies the equations

$$\begin{cases} \tau(x) - \tau_R - \dfrac{1}{\rho_R}\partial_x u(x) = 0, \\ u(x) - u_R + \dfrac{1}{\rho_R}\partial_x p(x) = 0. \end{cases}$$

We apply the same method than in the case $x < 0$, and check that the solution is of the same kind. One finally finds an expression for p'

$$p'(x) = \pm\rho_R\sqrt{-2H(p, S_R) + p\tau_R + e_R}. \tag{36}$$

At the interface, when $x = 0$, the continuity conditions are

$$p(0^-) = p(0^+) = p^\star, \quad u(0^-) = u(0^+) = u^\star,$$

with $p^\star \in \mathbb{R}$. One gets

$$u^\star = u_L - \frac{1}{\rho_L}p'(0^-) = u_R - \frac{1}{\rho_R}p'(0^+).$$

That is

$$-u_L + \frac{1}{\rho_L}p'(0^-) = -u_R + \frac{1}{\rho_R}p'(0^+).$$

Using (35) and (36), one finds the scalar equation

$$-u_L \pm \sqrt{-2H(p^\star, S_L) + 2p^\star\tau_L + 2e_L} = -u_R \pm \sqrt{-2H(p^\star, S_R) + 2p^\star\tau_R + 2e_R}$$

where $p^\star$ is the unknown.

In the numerical examples, we took initial conditions of a Sod shock tube, that are recalled hereafter.

$$u_L = u_R = 0, \quad \rho_L = \frac{1}{\tau_L} = 1, \quad \rho_R = \frac{1}{\tau_R} = \frac{1}{8}, \quad p_L = 1, \quad p_R = 0.1, \quad \gamma = 1,4,$$

$$\mathrm{e}^{S_L} = \frac{10}{4}, \quad \mathrm{e}^{S_R} = \frac{8^{1.4}}{4}.$$

As $u_L = u_R = 0$, one has

$$-H(p^\star, S_L) + p^\star \tau_L + e_L = -H(p^\star, S_R) + p^\star \tau_R + e_R. \tag{37}$$

Using the perfect gas law, one can rewrite the equation in terms of p and S.

$$\tau = \frac{((\gamma-1)\mathrm{e}^S)^{\frac{1}{\gamma}}}{p^{\frac{1}{\gamma}}}, \quad e = \frac{p\tau}{\gamma-1} = (\gamma-1)^{\frac{1}{\gamma}-1}\mathrm{e}^{\frac{S}{\gamma}} p^{1-\frac{1}{\gamma}}.$$

One obtains

$$\mathrm{e}^{\frac{S_L}{\gamma}} \left[-\frac{\gamma}{\gamma-1} p^{\star 1-\frac{1}{\gamma}} + p_L^{-\frac{1}{\gamma}} p^\star + \frac{p_L^{1-\frac{1}{\gamma}}}{\gamma-1} \right] = \mathrm{e}^{\frac{S_R}{\gamma}} \left[-\frac{\gamma}{\gamma-1} p^{\star 1-\frac{1}{\gamma}} + p_R^{-\frac{1}{\gamma}} p^\star + \frac{p_R^{1-\frac{1}{\gamma}}}{\gamma-1} \right].$$

Lemma 19. *The equation (37) admits a unique positive solution $p^\star \in [p_R, p_L]$.*

Proof. Let us denote $f_L(p) = -H(p, S_L) + p\tau_L + e_L$, and $f_R(p) = -H(p, S_R) + p\tau_R + e_R$, so that (37) is rewritten as $f_L(p^\star) - f_R(p^\star) = 0$.

The properties of the function f_L are the following. One has $f_L(p_L) = 0$, $f_L'(p_L) = -\frac{\partial H(p_L, S_L)}{\partial p} + \tau_L = -\tau_L + \tau_L = 0$, and $f_L''(p) = -\frac{\partial^2 H(p,S)}{\partial p^2} = -\frac{\partial \tau}{\partial p} = \frac{1}{\rho^2 c^2} > 0$. With the same calculations, one finds $f_R(p_R) = 0$, $f_R'(p_R) = 0$ and $f_R''(p) > 0$. The two functions f_L and f_R are strictly convex, with a minimum value equal to 0, obtained respectively for p_L and p_R.

Let us denote $f(p) = f_L(p) - f_R(p)$. One analyzes the function f in the case of the Sod shock tube, that is for $p_R \leqslant p \leqslant p_L$. One obtains $f(p_R) = f_L(p_R) > 0$, and $f(p_L) = -f_R(p_L) < 0$. The function f changes sign, so it takes at least once the value 0 in between p_R and p_L, which validates the existence of a solution. To have the uniqueness, one needs to prove the monotonicity of f. One has $f'(p) = f_L'(p) - f_R'(p)$. For all $p_R < p < p_L$, one finds $f_L'(p) < 0$ and $f_R'(p) > 0$, so $f'(p) < 0$ which concludes to the monotonicity of f, and the uniqueness of the solution to (37). □

Numerically, we calculated with a Newton method that the solution to (37) is approximately equal to $p^\star = 0.2559$. It corresponds to a velocity of $u^\star = 0.8789$. The exact value of the velocity for the Riemann problem at the contact discontinuity is $u^{exact} = 0.9275$. This value is found in Toro,[32] [Table 4.3, p 131]. The difference between $u^\star$ and u^{exact} is equal to 5.2%, which is a satisfying accuracy considering that the implicit simulation performs with only one time step. In our mind, this small relative error of 5.2% is the reason why the contact discontinuity of the implicit solver is approximately superimposed with the reference one in Figure 5.We observed a similar behavior for all the other test problems and we believe it is a strong asset of this family of implicit Lagrangian schemes.

7 Conclusion

We have used a strategy of predictor-corrector scheme, based on the previous work[4] in Eulerian coordinates, to solve numerically the Euler equations. We have defined an abstract frame in order to analyze a family of implicit schemes written under the peculiar form (8). We have proved the existence and uniqueness of a solution to the prediction step of our implicit scheme. We provided two examples using this result and led numerical tests that have indeed corroborated the theoretical statements of stability. The second example is in the Appendix. The numerical illustrations compared the implicit scheme to an explicit scheme of reference, and showed the precision of this new algorithm in these cases.

In a future work, it would be interesting to generalize this method to the case of thin elasto-plastic structures using Kluth and Després[21] or Maire *et al.*[23] We could also try to improve this work by using a more elaborate flux such as a two state solver flux, or increase the scheme order at the order 2. It will probably ameliorate the precision, but it stays to evaluate the cost of simulation that it would generate. The multi-dimensional version would need a more advanced management for the displacement of the mesh, but the principal ingredients of Theorem 4 should remain similar.

Other numerical examples that are more realistic have to be performed to evaluate the pertinance of this algorithm. Theoretically as well, it would be great to have an explanation on the rapid convergence of the Newton algorithm for the prediction step, and to have a more elegant proof of the entropy inequalities using the frame (8).

A Traffic flow problem

A.1 Formulation under the form (8)

As another application of Theorem 4, consider the traffic flow problem in Lagrangian formalism

$$\partial_t \tau + \partial_m f(\tau) = 0, \quad f(\tau) = -u(\tau) = \tau^{-1} - 1, \tag{38}$$

where $\rho = \frac{1}{\tau} > 0$ is the density, f is the flux and u is the velocity. The mass variable is $dm = \rho_0(x)dx$. After discretizing this problem in time and space on a mesh $\mathcal{M}$ composed of $N > 0$ cells denoted by j, the following implicit scheme is obtained

$$\frac{\tau_j^{n+1} - \tau_j^n}{\Delta t} + \frac{\frac{1}{\tau_{j+1}^{n+1}} - \frac{1}{\tau_j^{n+1}}}{\Delta m} = 0,$$

which can be rewritten as

$$\tau_j^{n+1} - \tau_j^n + \frac{\Delta t}{\Delta m}\left(\frac{1}{\tau_{j+1}^{n+1}} - \frac{1}{\tau_j^{n+1}}\right) = 0. \tag{39}$$

This scheme is provided with periodic boundary conditions. This problem only needs one step to be solved implicitly, the unconditional stability of (39) will be proved further on.

Let us suppose that τ_j is positive for all $j \in \{1, \dots, N\}$ and let us denote $\nu = \frac{\Delta t}{\Delta m}$. In the rest of this Section one uses the fact that $\rho_j^{n+1} = \frac{1}{\tau_j^{n+1}}$ and the dependence in time is omitted for the sake of simplicity. The scheme becomes

$$\frac{1}{\rho_j} - \frac{1}{\rho_j^n} + \nu(\rho_{j+1} - \rho_j) = 0.$$

After reorganization, one gets

$$-\frac{1}{\rho_j} + \frac{1}{\rho_j^n} + \frac{\nu}{2}(-\rho_{j-1} + 2\rho_j - \rho_{j+1}) = \frac{\nu}{2}(\rho_{j+1} - \rho_{j-1}). \tag{40}$$

One can now express the implicit scheme under the form (8). The vector of unknowns is defined by

$$U = (\rho_j)_{1 \leqslant j \leqslant N}. \tag{41}$$

The matrix of coefficients for the right handside of the equations (40) reads

$$A = \frac{\nu}{2}\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & -1 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 0 & 1 \\ 1 & 0 & \cdots & 0 & -1 & 0 \end{bmatrix}. \tag{42}$$

The terms -1 on the first row and 1 on the last row are due to the periodic boundary conditions. The functional $J : \mathcal{D} \to \mathbb{R}$, is defined on the domain

$$\mathcal{D} = \{U \text{ such that } \forall j \in \{1, \dots, N\},\ \rho_j > 0\}. \tag{43}$$

J is evaluated as follows

$$J(U) = \sum_{j=1}^{N}\left(-\log(\rho_j) + \frac{\rho_j}{\rho_j^n}\right) + \sum_{j=1}^{N}\frac{\nu}{4}(\rho_j - \rho_{j+1})^2. \tag{44}$$

By construction the j^{th} partial derivative of J, is

$$\frac{\partial J}{\partial \rho_j} = -\frac{1}{\rho_j} + \frac{1}{\rho_j^n} + \frac{\nu}{2}(-\rho_{j-1} + \rho_j + \rho_j - \rho_{j+1}),$$

which is exactly the left handside of (40).

A.2 Application of Theorem 4 for the traffic flow

The function J defined by (44) satisfies Hypothesis 2, and the matrix A given by (42) verifies Hypothesis 3.

Proposition 20. *The implicit scheme (39) for the traffic flow has a unique solution.*

Proof. Let $m = 0$, $n = N$, $U = (\rho_j)_{j\in\{1,\dots,N\}}$, A and J defined as in (42) and (44). All the hypothesis of Theorem 4 are satisfied, so the existence and uniqueness of a solution to the implicit traffic flow scheme (39) is deduced. □

A.3 Numerical illustrations

The precision of the implicit scheme is illustrated, compared to the explicit one, see,[11] [Chapter 2], both with Upwind fluxes. The aim of this example is to show the well posedness of the scheme and its robustness. In this case, this is now contact discontinuity, rarefaction fan and shocks for which the accuracy is depicted for large time step.

$$\text{Explicit scheme: } \frac{\tau_j^{n+1}-\tau_j^n}{\Delta t} + \frac{\frac{1}{\tau_{j+1}^n} - \frac{1}{\tau_j^n}}{\Delta m} = 0,$$
$$\text{Implicit scheme: } \frac{\tau_j^{n+1}-\tau_j^n}{\Delta t} + \frac{\frac{1}{\tau_{j+1}^{n+1}} - \frac{1}{\tau_j^{n+1}}}{\Delta m} = 0.$$

The example is performed on a mesh of 100 cells, for a final time $t = 1.2$ and velocity boundary conditions. The first experiment has the following initial conditions

$$\tau_0(x) = \begin{cases} 2, & x < 0.3, \\ 1.4, & 0.3 \leqslant x < 0.7, \\ 1.1, & 0.7 \leqslant x. \end{cases}$$

The boundary condition is $u_{right} = 0.5$.

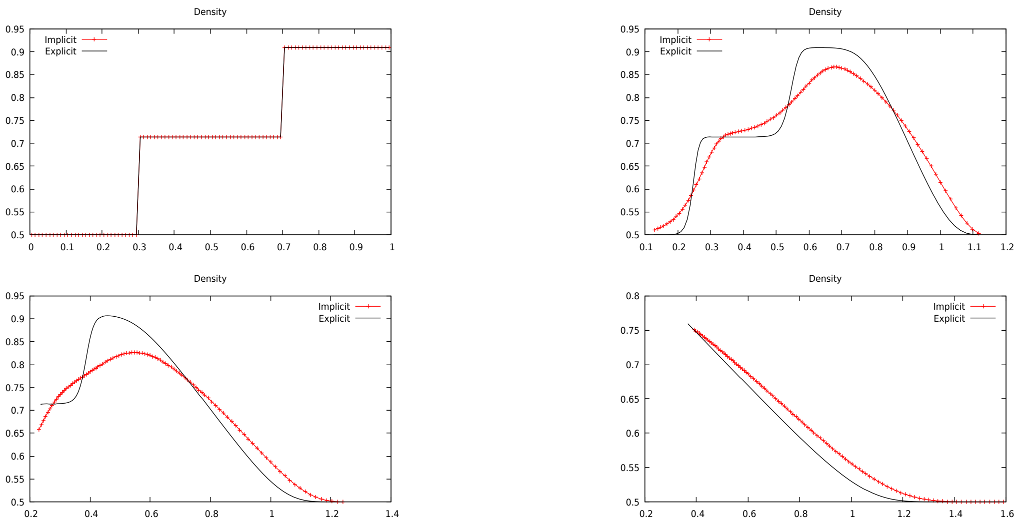

Figure 6: Traffic flow problem at different time steps, on a mesh of 100 cells: $t = 0$, $t = 0.25$, $t = 0.5$ and $t = 1.2$. $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 40$.

One clearly sees in Figure 6 that the implicit scheme is numerically more dissipated than the explicit scheme, both on the rarefaction waves and the shocks. It is explained by the larger CFL, but the results are nonetheless similar.

Another example is a sinusoïdal initial condition, with velocity boundary condition.

$$\tau_0(x) = \frac{1}{0.5 - 0.3\sin(10x)}, \qquad u_{right} = \frac{\sin(1) - 0.5}{\sin(1) + 0.5}.$$

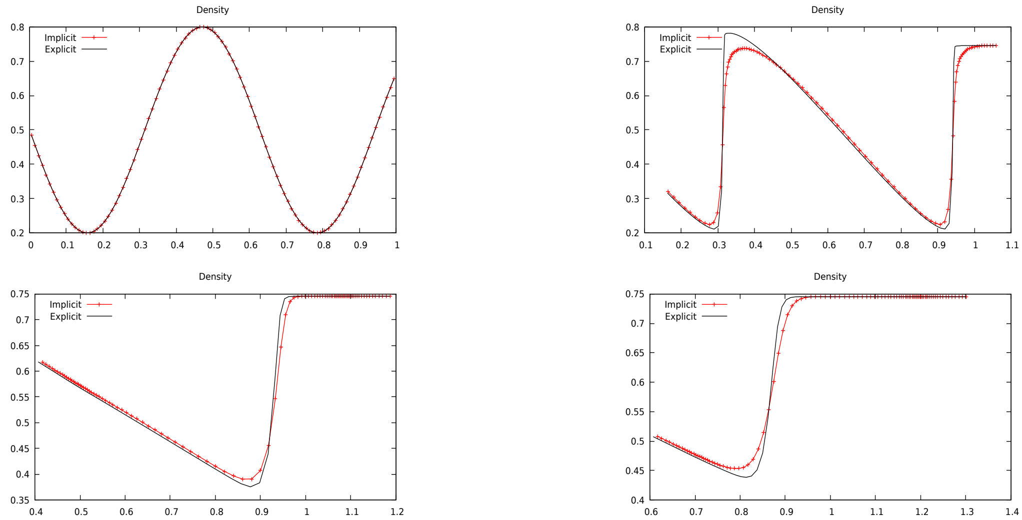

Figure 7: Traffic flow problem at different time steps, on a mesh of 100 cells: $t = 0$, $t = 0.25$, $t = 0.75$ and $t = 1.2$. $CFL_{explicit} = 0.4$ and $CFL_{implicit} = 20$.

The curves are following the same movement, and it can only be mentionned a dissipation at the end of the rarefaction waves. Figure 7 shows the precision of the implicit scheme considering a larger CFL.

B Isothermal equation of state

We briefly describe the modification of the method to treat an isothermal equation of state. An isothermal equation of state $p = \frac{C_T}{\tau}$ can be analyzed by letting $\gamma \to 1$ in the perfect gaz equation of state (3). Nevertheless since this method is singular, it is simpler to directly perform the required modification. Actually, we only need to modify the function L_j^1 in the definition of J (18).

The function L_j^1 is designed to verify the equation $\frac{\partial L_j^1}{\partial(-p_j)} = \tau_j - \tau_j^n$, see the proof of Proposition 7. With the isothermal equation of state, one takes $L_j^1(-p_j) = -C_T \log(p_j) + p_j \tau_j^n$. Because of the logarithmic term in L_j^1, the hypothesis 2 of the Theorem 4 is satisfied in a stronger form. For $V \in \partial\mathcal{D}$, one has $J(W) \xrightarrow[W\in\mathcal{D}]{W\to V} +\infty$.

References

[1] D. Azé and J.-B. Hirriart-Urruty. *Analyse Variationnelle et Optimisation*. Cépadues, 2010.

[2] R. M. Beam and R. F. Warming. An implicit finite-difference algorithm for hyperbolic systems in conservation-law form. *Journal of Computational Physics*, 22(1):87–110, 1976.

[3] L. Brugnano and V. Casulli. Iterative solution of piecewise linear systems. *SIAM J. Sci. Comput.*, 30:pp. 463–472, 2008.

[4] C. Chalons, F. Coquel, and C. Marmignon. Time-Implicit Approximation of the Multipressure Gas Dynamics Equations in Several Space Dimensions. *SIAM*, 48:1678–1706, 2010.

[5] E. Coddington and N. Levinson. *Theory of Ordinary Differential Equations.* Tata McGraw-Hill Education, 1955.

[6] F. Coulette, E. Franck, P. Helluyn, A. Ratnani, and E. Sonnendrücker. Implicit time schemes for compressible fluid models based on relaxation methods. *Computers and Fluids*, 188:70–85, 2019.

[7] B. Dacorogna. *Direct Methods in the Calculus of Variations.* Applied Mathematical Sciences,78. Springer-Verlag, Berlin, 1989.

[8] J.-P. Demailly. *Analyse Numérique et Equations Différentielles-4ème Ed.* EDP sciences, 2016.

[9] I. Demirdzic, Z. Lilek, and Peric M. A collocated finite volume method for predicting flows at all speeds. *Journal for Numerical Methods in Fluids*, 16:1029 – 1050, 1993.

[10] B. Després. Weak consistency of the cell-centered Lagrangian GLACE scheme on general meshes in any dimension. *Computer Methods in Applied Mechanics and Engineering*, 199(41):2669–2679, 2010.

[11] B. Després. *Numerical Methods for Eulerian and Lagrangian Conservation Laws.* Birkhäuser Basel, 2017.

[12] A. Ern and J.-L. Guermond. *Éléments Finis: Théorie, Applications, Mise en Oeuvre*, volume 36. Springer Science & Business Media, 2002.

[13] R. Eymard, T. Gallouët, and R. Herbin. *Handbook of Numerical Analysis - Finite Volume Methods*, volume 7. Elsevier, 2000.

[14] B. A. Fryxell, P. R. Woodward, P. Colella, and K.-H. Winkler. An implicit-explicit hybrid method for lagrangian hydrodynamics. *Journal of Computational Physics*, 63:283–310, 1986.

[15] T. Gallouët, L. Gastaldo, R. Herbin, and J.-C. Latché. An unconditionally stable pressure correction scheme for the compressible barotropic Navier-Stokes equations. *ESAIM: M2AN*, 42:303–331, 2008.

[16] S. K. Godunov. Difference methods of solving equations of gas dynamics. *Izd-vo Novosibirsk, un-ta*, 1962.

[17] J.-B. Hirriart-Urruty. *Optimisation et Analyse Convexe.* EDP Sciences, 1998.

[18] J.-B. Hirriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization.* Springer-Verlag, 1996.

[19] J.-B. Hirriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis.* Springer-Verlag, 2004.

[20] R. I. Issa. Solution of the implicitly discretised fluid flow equations by operator-splitting. *Journal of Computational Physics*, 62:40–65, 1985.

[21] G. Kluth and B. Després. Discretization of hyperelasticity on unstructured mesh with a cell-centered Lagrangian scheme. *Journal of Computational Physics*, 229(24):9092–9118, 2010.

[22] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*, volume 31. Cambridge university press, 2002.

[23] P-H. Maire, R. Abgrall, J. Breil, R. Loubère, and B. Rebourcet. A nominally sceond-order cell-centered Lagrangian scheme for simulating elastic-plastic flows on two-dimensional unstructured grids. *Journal Of Computational Physics*, 235:626–665, 2013.

[24] F. Moukalled and Darwish M. A high-resolution pressure-based algorithm for fluid flow at all speeds. *Journal of Computational Physics*, 168:101–130, 2001.

[25] W. A. Mulder and B. Van Leer. Experiments with implicit upwind methods for the Euler equations. *Journal of Computational Physics*, 59:232 – 246, 1985.

[26] G. Patnaik, R. H. Guirguis, J. P. Boris, and Oran E. S. A barely implicit correction for flux-corrected transport. *Journal of Computational Physics*, 71:1–20, 1987.

[27] S. Peluchon, G. Gallice, and L. Mieussens. A robust implicit-explicit acoustic-transport splitting scheme for two-phase flows. *Journal of Computational Physics*, 339:328–355, 2017.

[28] E. S. Politis and Giannakoglou K. C. A pressure-based algorithm for high speed turbomachinery flows. *Journal for Numerical Methods in Fluids*, 25:63–80, 1997.

[29] P.-A. Raviart and E. Godlewski. *Numerical Approximation of Hyperbolic Systems of Conservation Laws.* Springer, 1996.

[30] N. Seguin, F. Coquel, and E. Godlewski. Approximation par relaxation de systèmes hyperboliques. *Séminaire d'analyse appliquée, Université Paris 13*, 2008.

[31] D. Serre. *Systems of Conservation Laws 1: Hyperbolicity, Entropies, Shock Waves.* Cambridge University Press, 1999.

[32] E. F. Toro. *Riemann Solvers and Numeriacal Methods for Fluid Dynamics.* Springer Verlag, 1999.

[33] G. Toth, R. Keppens, and M. A. Bochev. Implicit and semi-implicit schemes in the versatile advectio code: Numerical tests. *Astronomy and Astrophysics*, 332:1159 – 1170, 1998.

[34] D. R. Van der Heuk, C. Vuik, and P. Wesseling. Stability analysis of segregated solution methods for compressible flows. *Applied Numerical Mathematics*, 38:257 – 274, 2001.

[35] D. R. Van der Heuk, C. Vuik, and P. Wesseling. A conservative pressure-correction method for flow at all speed. *Computers and Fluids*, 32:1113 – 1132, 2003.

[36] D. Vidovic, A. Segal, and P. Wesseling. A superlinearly convergent Mach-uniform finite volume method for Euler equations on staggered unstructured grids. *Journal of Computational Physics*, 217:277–294, 2006.