SMAI-JCM
SMAI Journal of Computational Mathematics

A family of second-order dissipative finite volume schemes for hyperbolic systems of conservation laws

Mehdi Badsi, Christophe Berthon & Ludovic Martaud

Volume 9 (2023), p. 31-60.
https://doi.org/10.5802/smai-jcm.94

The SMAI Journal of Computational Mathematics is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/
Submissions at https://smai-jcm.centre-mersenne.org/ojs/submission
e-ISSN: 2426-8399

SMAI Journal of Computational Mathematics
Vol. 9, 31-60 (2023)

A family of second-order dissipative finite volume schemes for hyperbolic systems of conservation laws

Mehdi Badsi [1]
Christophe Berthon [2]
Ludovic Martaud [3]

[1] Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes, France
E-mail address: mehdi.badsi@univ-nantes.fr
[2] Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes, France
E-mail address: christophe.berthon@univ-nantes.fr
[3] Laboratoire de Mathématiques Jean Leray, CNRS UMR 6629, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes, France
E-mail address: ludovic.martaud@univ-nantes.fr.

Abstract. We propose and study a family of formally second-order accurate schemes to approximate weak solutions of hyperbolic systems of conservation laws. Theses schemes are based on a dissipative property satisfied by the second-order discretization in space. They are proven to satisfy a global entropy inequality for a generic strictly convex entropy. These schemes do not involve limitation techniques. Numerical results are provided to illustrate their accuracy and stability.

2020 Mathematics Subject Classification. 65N08, 35L65, 35L67.

Keywords. Systems of conservation laws, Second-order finite Volume schemes, Explicit schemes, Global entropy inequality.

1. Introduction

This work is concerned with the numerical approximation of weak solutions of systems made of $d \geq 1$ conservation laws in one space dimension given by the following Cauchy problem

$$\begin{cases} \partial_t w + \partial_x f(w) = 0, & t > 0, \quad x \in \mathbb{R}, \\ w(x, t = 0) = w_0(x). \end{cases} \tag{1.1}$$

The unknown state vector $w(x,t)$ is assumed to belong to Ω, a non-empty convex open subset of $\mathbb{R}^d$. Here, $f : \Omega \to \mathbb{R}^d$ is a given smooth flux function. For all $w \in \Omega$, the $d \times d$ Jacobian matrix $\nabla f(w)$ is assumed to be diagonalizable in $\mathbb{R}$. The system (1.1) is then a hyperbolic system of conservation laws. The initial data $w_0 : \mathbb{R} \to \Omega$ is a given measurable function in $L^1_{\text{loc}}(\mathbb{R})$.

Even though the initial data is smooth, it is well-known that the solutions to (1.1) may develop, in a finite time, discontinuities [29, 30, 31, 38]. Weak solutions are in general non unique and one has to select a physically admissible solution among many others. In this regard, the system (1.1) is usually endowed with entropy inequalities. We consider in this work to be given a strictly convex function $\eta \in C^2(\Omega, \mathbb{R})$, called entropy function, and an entropy flux function $G \in C^1(\Omega, \mathbb{R})$ such that we have

$$\forall w \in \Omega, \quad \nabla \eta(w)^T \nabla f(w) = \nabla G(w)^T. \tag{1.2}$$

The authors acknowledge financial support from the ANR project MUFFIN ANR-19-CE46-0004.
https://doi.org/10.5802/smai-jcm.94

It is then required that weak solutions verify an entropy inequality given in the weak sense by

$$\partial_t \eta(w) + \partial_x G(w) \leq 0. \tag{1.3}$$

Instead of (1.3), we shall focus in this work on the weaker inequality

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}} \eta(w(x,t))\,\mathrm{d}x \leq 0, \tag{1.4}$$

which results from an integration in space of the inequality (1.3) for compactly supported solutions. This inequality will be called throughout this work a global entropy inequality. Our main concern is to design a family of schemes that are formally second-order accurate in space and that verifies a discrete analogue to (1.4). We consider uniform meshes in space and time $(x_{i+\frac{1}{2}})_{i\in\mathbb{Z}} \subset \mathbb{R}$ and $(t^n)_{n\in\mathbb{N}} \subset [0,+\infty)$ of respectively constant size $\Delta x > 0$ and $\Delta t > 0$. We have $x_{i+\frac{1}{2}} = x_{i-\frac{1}{2}} + \Delta x$ for all $i \in \mathbb{Z}$ and $t^{n+1} = t^n + \Delta t$ for all $n \in \mathbb{N}$. Weak solutions to (1.1) are approximated within the standard finite volume framework. At time t^n, we consider the following piecewise constant function

$$w_\Delta(x,t^n) = w_i^n \quad \text{ if } x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}), \tag{1.5}$$

where the quantities w_i^n are approximations of the average of the solution over the cell $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$,

$$w_i^n \approx \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} w(x,t^n)\,\mathrm{d}x,$$

where $w(x,t^n)$ naturally belongs to $L^1_{\mathrm{loc}}(\mathbb{R})$.

Many strategies are known to define an updated sequence $(w_i^{n+1})_{i\in\mathbb{Z}}$ starting from $(w_i^n)_{i\in\mathbb{Z}}$ (for instance, see [10, 22, 36, 43] and references therein). The crux of the problem lies in the discrete analogue to (1.3) which is known to ensure the stability of the scheme and to avoid non-admissible discontinuous waves (for instance, see [8, 26]). Several first-order accurate finite volume schemes are known to satisfy (1.3): Godunov [17, 22], the kinetic schemes [3, 27], the HLL scheme[21, 22], the HLLC scheme[44], Suliciu relaxation approaches [2, 4, 7, 9, 12, 24, 25] or many other schemes [7, 12, 12, 14, 15, 23, 24, 25] based on the strategy introduced by Tadmor [41, 42]. As far as second-order accurate in space schemes are concerned, the literature is less exhaustive. Several results have nevertheless been obtained in the semi discrete case [6, 41] or in the fully discrete case using an entropy averaging technique [1, 5, 34, 35]. Both strategies are however not sufficient to rule out non-admissible discontinuities in the converged solutions. Other strategies such as MUSCL technique [45] or ENO/WENO schemes [33, 39, 40, 46] or DG schemes [11, 24, 40] which require limitation techniques are also of common use. However efficient, in this work we shall not use limitation techniques.

We propose here to study a family of formally second-order accurate schemes that verifies a discrete analogue to (1.4). Although such a global criteria is clearly not sufficient to avoid non-entropic solutions, it provides an upper bound of the global discrete entropy and thus guarantees a form of computational stability for the approximated solution. The here designed schemes are derived starting from the HLL scheme [22] (also called Rusanov scheme [37]) complemented with a second-order correction in space which is asked to verify a dissipative property. In Section 3, a quadratic stability study in the simpler case of the scalar linear transport equation, using both algebraic and Fourier approaches is performed. The purpose of this section is twofold: to explicit the dissipative property in a simple setting, and then to justify the use of a second-order in time discretization to recover the quadratic stability under a hyperbolic CFL (Courant, Friedrichs,Levy) condition. In Section 4, we establish the global entropy inequality (1.4) in the general case of non linear systems (1.1) for a given single entropy-entropy-flux pair (η, G). The proof relies on the following ingredients: a second-order Taylor expansion of the entropy, a reformulation of the global entropy dissipation, the use of the dissipative property granted by the second-order discretization in space and the choice of a large enough viscosity coefficient. Because the proof is also intended to be made in the general case of an open convex subset $\Omega \subset \mathbb{R}^d$, a topological restriction on the approximate solution is also needed. Namely, it is required to belong to a compact subset of Ω. It is somehow restrictive but it is up to our knowledge very standard in the literature, notably in the case of the Euler equations where it consists in assuming the solutions to be bounded and uniformly far away from the vacuum [38]. In Section 5, we perform numerical experiments that assess the stability and the accuracy of our schemes. Eventually, a short conclusion is given at the very end of this paper.

2. **Unlimited second-order space HLL type schemes**

The starting point is the original symmetric first-order HLL scheme [22] also called Rusanov scheme [37], that reads as follows:

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}_\lambda^{O1}(w_i^n, w_{i+1}^n) - \mathcal{F}_\lambda^{O1}(w_{i-1}^n, w_i^n)\right), \tag{2.1}$$

where the numerical flux function is given by

$$\mathcal{F}_\lambda^{O1}(w_i^n, w_{i+1}^n) = \frac{1}{2}\left(f(w_i^n) + f(w_{i+1}^n)\right) - \frac{\lambda}{2}\left(w_{i+1}^n - w_i^n\right). \tag{2.2}$$

Here, $\lambda > 0$ stands for the numerical viscosity coefficient. Under the following CFL condition:

$$\frac{\lambda \Delta t}{\Delta x} \leq \frac{1}{2} \quad \text{with} \quad \lambda \geq \lambda^{\mathrm{HLL}} = \max_{i\in\mathbb{Z}}\left(|\mu(w_i^n)|\right), \tag{2.3}$$

where $\mu(w)$ denotes the spectral radius of the jacobian matrix $\nabla f(w)$, the scheme (2.1) is first-order in space, entropy preserving and convergent (see [22]). Equipped with the first-order scheme (2.1), we are now going to increase the order of accuracy in space. For the sake of clarity of this work, we remind of the notion of second-order consistency in space (some details can be found in [4] and [13]). We consider here only five-points finite volume schemes.

Definition 2.1 (Second-order weak consistency [4])**.** Let $\mathcal{F} : (\mathbb{R}^d)^4 \to \mathbb{R}^d$ be a continuous function which verifies the consistency relation:

$$\forall u \in \Omega, \quad \mathcal{F}(u, u, u, u) = f(u). \tag{2.4}$$

Consider the semi-discrete numerical scheme defined for $t \in [0, T)$ with $T > 0$ and for $i \in \mathbb{Z}$ by:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t}(t) = -\frac{1}{\Delta x}\left(\mathcal{F}(u_{i-1}, \ldots, u_{i+2}) - \mathcal{F}(u_{i-2}, \ldots, u_{i+1})\right). \tag{2.5}$$

Assume for all $i \in \mathbb{Z}$, $u_i(t) = w_i(t) = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} w(x,t)\,\mathrm{d}x$ where w is a smooth solution to (1.1) defined on $\mathbb{R} \times [0, T)$. The semi-discrete scheme is second-order consistent in the weak sense, if for all times $t \in [0, T)$ and for any smooth compactly supported function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$ we have

$$\int_{\mathbb{R}} \frac{\mathrm{d}w_\Delta(t)}{\mathrm{d}t}\varphi(x)\mathrm{d}x = \int_{\mathbb{R}} \left(\partial_t w(t,x)\right)\varphi(x)\mathrm{d}x + \mathcal{O}(\Delta x^2), \tag{2.6}$$

where $w_\Delta(t) = \sum_{i\in\mathbb{Z}} w_i(t)\mathbf{1}_{[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})}$.

This definition is based on the notion of weak convergence in space and does not prevent a semi-discrete scheme from being at a lower order of space accuracy when assessing the error in a strong topology. We now give the following result which gives a sufficient condition for the second-order consistency in the weak sense. It is somehow a standard result but we give a proof for the sake of completeness of this work.

Lemma 2.2 (Weak consistency [4]). *Let $\mathcal{F} : (\mathbb{R}^d)^4 \to \mathbb{R}^d$ be a continuous function which verifies the consistency relation:*

$$\forall u \in \Omega, \quad \mathcal{F}(u,u,u,u) = f(u).$$

Consider a semi-discrete numerical scheme defined for $t \in [0,T)$ with $T > 0$ and for $i \in \mathbb{Z}$ by:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t}(t) = -\frac{1}{\Delta x}\left(\mathcal{F}(u_{i-1},\dots,u_{i+2}) - \mathcal{F}(u_{i-2},\dots,u_{i+1})\right). \tag{2.7}$$

Let $t \in [0,T)$ being fixed. Assume for all $i \in \mathbb{Z}$,

$$\mathcal{F}(u_{i-1},\dots,u_{i+2}) = f(u(x_{i+\frac{1}{2}},t)) + \mathcal{O}(\Delta x^2) \tag{2.8}$$

where the $\mathcal{O}(\Delta x^2)$ term is uniform with respect to $i \in \mathbb{Z}$ and where we have set for a given smooth enough function $u(x,t)$,

$$u_i(t) = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x,t)\,\mathrm{d}x, \tag{2.9}$$

then the scheme is second-order consistent in the weak sense.

Proof. Let w be a smooth solution on $\mathbb{R} \times [0,T)$ to (1.1) and consider the semi-discrete scheme defined by (2.7) associated with w. Let $t \in [0,T)$ and set for all $i \in \mathbb{Z}$, $\mathcal{F}_{i+\frac{1}{2}} = \mathcal{F}(w_{i-1},\dots,w_{i+2})$ where $w_i(t) = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} w(x,t)\,\mathrm{d}x$. Let $\varphi \in C_c^\infty(\mathbb{R})$ then we have,

$$\begin{aligned}\int_{\mathbb{R}} \frac{\mathrm{d}w_\Delta(t)}{\mathrm{d}t}\varphi(x)\mathrm{d}x &= \sum_{i\in\mathbb{Z}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{\mathrm{d}w_i}{\mathrm{d}t}(t)\varphi(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} -\frac{1}{\Delta x}(\mathcal{F}_{i+\frac{1}{2}} - \mathcal{F}_{i-\frac{1}{2}})\varphi(x)\mathrm{d}x \\ &= -\sum_{i\in\mathbb{Z}} (\mathcal{F}_{i+\frac{1}{2}} - \mathcal{F}_{i-\frac{1}{2}})\frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \varphi(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} \mathcal{F}_{i+\frac{1}{2}}(\varphi_{i+1} - \varphi_i),\end{aligned}$$

where we have set $\varphi_i = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \varphi(x)\mathrm{d}x$. Using the consistency of the numerical flux (2.8), we obtain

$$\int_{\mathbb{R}} \frac{\mathrm{d}w_\Delta(t)}{\mathrm{d}t}\varphi(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} f(w(x_{i+\frac{1}{2}},t))(\varphi_{i+1} - \varphi_i) + \mathcal{O}(\Delta x^2)\sum_{i\in\mathbb{Z}}(\varphi_{i+1} - \varphi_i).$$

To conclude the proof, it suffices to prove that $\sum_{i\in\mathbb{Z}} f(w(x_{i+\frac{1}{2}},t))(\varphi_{i+1}-\varphi_i) = \int_{\mathbb{R}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x + \mathcal{O}(\Delta x^2)$. Decomposing the integral as a sum of integrals over each interval (x_i, x_{i+1}) we have

$$\int_{\mathbb{R}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} \int_{x_i}^{x_{i+1}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x.$$

Using the smoothness of the functions f, w and φ and a Taylor–Lagrange expansion at the second order around $x_{i+\frac{1}{2}}$, we have the existence of a function $\xi : x \in [x_i, x_{i+1}) \mapsto \xi(x) \in (\min(x, x_{i+\frac{1}{2}}), \max(x, x_{i+\frac{1}{2}}))$ such that

$$\begin{aligned}&\int_{x_i}^{x_{i+1}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x \\ &\qquad = f(w(x_{i+\frac{1}{2}},t))\partial_x\varphi(x_{i+\frac{1}{2}})\Delta x + \int_{x_i}^{x_{i+1}} \frac{(x - x_{i+\frac{1}{2}})^2}{2}\partial_{xx}\left(f(w(\,\cdot\,,t))\partial_x\varphi(\,\cdot\,)\right)(\xi(x))\mathrm{d}x,\end{aligned}$$

where the first-order term has vanished because $x_{i+\frac{1}{2}}$ is the mid point of the interval (x_i, x_{i+1}) and the function $x \mapsto (x - x_{i+\frac{1}{2}})$ is symmetric with respect to $x_{i+\frac{1}{2}}$. By summation, we thus have

$$\int_{\mathbb{R}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x$$
$$= \sum_{i\in\mathbb{Z}} f\left(w(x_{i+\frac{1}{2}},t)\right)\partial_x\varphi(x_{i+\frac{1}{2}})\Delta x + \sum_{i\in\mathbb{Z}}\int_{x_i}^{x_{i+1}} \frac{(x-x_{i+\frac{1}{2}})^2}{2}\partial_{xx}\left(f(w(\,\cdot\,,t))\partial_x\varphi(\,\cdot\,)\right)(\xi(x))\mathrm{d}x.$$

Since φ is compactly supported the second sum is a $\mathcal{O}(\Delta x^2)$ term, so one has

$$\int_{\mathbb{R}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} f\left(w(x_{i+\frac{1}{2}},t)\right)\partial_x\varphi(x_{i+\frac{1}{2}})\Delta x + \mathcal{O}(\Delta x^2).$$

Using again a Taylor–Lagrange expansion, we obtain that $\varphi_{i+1} - \varphi_i = \partial_x\varphi(x_{i+\frac{1}{2}})\Delta x + \mathcal{O}(\Delta x^3)$ and by summation we eventually deduce

$$\int_{\mathbb{R}} f(w(x,t))\partial_x\varphi(x)\mathrm{d}x = \sum_{i\in\mathbb{Z}} f\left(w(x_{i+\frac{1}{2}},t)\right)(\varphi_{i+1}-\varphi_i) + \mathcal{O}(\Delta x^2)$$

where the $\mathcal{O}(\Delta x^2)$ term is uniform with respect to $i \in \mathbb{Z}$ because φ is compactly supported. Using the fact that w is a smooth solution to (1.1) concludes the proof. ∎

The numerical HLL flux function (2.2) is first-order consistent. Indeed, with (2.9), a standard Taylor expansion in a neighborhood of $x_{i+\frac{1}{2}}$ gives

$$\mathcal{F}_\lambda^{O1}(w_i^n, w_{i+1}^n) = f\left(w(x_{i+\frac{1}{2}},t^n)\right) - \frac{\lambda\Delta x}{2}\partial_x w(x_{i+\frac{1}{2}},t^n) + \mathcal{O}\left(\Delta x^2\right). \tag{2.10}$$

From the above Taylor expansion, we see that it is sufficient to propose a correction which is consistent with $+\frac{\lambda}{2}\partial_x w(x_{i+\frac{1}{2}},t^n)$. We therefore consider second-order space accuracy numerical flux functions of the form

$$\mathcal{F}_\lambda^{O2}(w_{i-1}^n,\dots,w_{i+2}^n) = \mathcal{F}_\lambda^{O1}(w_i^n,w_{i+1}^n) + \frac{1}{2}\left(\alpha_i^n + \alpha_{i+1}^n\right), \tag{2.11}$$

where the corrective terms α_i^n (which will be defined here after) must satisfy the following consistency

$$\alpha_i^n = \frac{\lambda\Delta x}{2}\partial_x w(x_i,t^n) + \mathcal{O}(\Delta x^2), \tag{2.12}$$

for a smooth function w. Therefore the second-order numerical flux function $\mathcal{F}_\lambda^{O2}$ will be composed of both an approximation of the term $-\frac{\lambda\Delta x}{2}\partial_x w$ which inherits from the HLL flux function $\mathcal{F}_\lambda^{O1}$ and an approximation of the same term but with the opposite sign $+\frac{\lambda\Delta x}{2}\partial_x w$ which comes from the corrective term α_i. Of course, at the continuous level the sum of these two terms is equal to zero. However at the discrete level, since these two terms will not be discretized within the same stencil, they will not generally cancel. The difference will be asked to control the numerical dissipativity of the scheme and thus its stability. This is why our family of schemes will be called dissipative according to the following definition.

Definition 2.3 (Entropy dissipative numerical flux)**.** Let an integer $l \in \mathbb{N}$ and $\mathcal{F} : (\mathbb{R}^d)^{2l+2} \to \mathbb{R}^d$ be a continuous function. Consider a numerical scheme in the conservative form:

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}\left(w_{i-l}^n,\dots,w_{i+l+1}^n\right) - \mathcal{F}\left(w_{i-l-1}^n,\dots,w_{i+l}^n\right)\right), \quad \forall i \in \mathbb{Z}.$$

Let $\eta \in C^1(\Omega, \mathbb{R})$ be a convex entropy function. The numerical flux $\mathcal{F}$ is said globally dissipative relatively to the entropy function η if the following inequality holds

$$\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n) \cdot \frac{\mathcal{F}\left(w_{i-l}^n, \dots, w_{i+l+1}^n\right) - \mathcal{F}\left(w_{i-l-1}^n, \dots, w_{i+l}^n\right)}{\Delta x} > 0, \tag{2.13}$$

for any non-constant sequence $(w_i^n)_{i\in\mathbb{Z}} \subset \Omega$.

In view of the previous definition, we propose the following definition of the second-order correction term α_i^n,

$$\alpha_i^n := \frac{\lambda}{2}\Delta x\overline{\partial_x w_i^n} := \frac{\lambda}{2}\left[\Theta_i^n\left(w_{i+1}^n - w_i^n\right) + \left(I - \Theta_i^n\right)\left(w_i^n - w_{i-1}^n\right)\right], \tag{2.14}$$

where I stands for $d \times d$ identity matrix while Θ_i^n is a free $d \times d$ diagonal matrix parameter to be defined in such a way that the scheme is globally dissipative. More precisely, the numerical flux (2.11) will be asked to satisfy (2.13) which in our proof of stability reformulates as the inequality (4.4). For the sake of the simplicity, we consider only diagonal matrices, but more general matrices could be considered. Equipped with the corrective terms α_i^n, one can establish the following.

Proposition 2.4 (Weak second-order consistency)**.** *Let a smooth function $u(x,t)$ compactly supported in space and let u_i defined for all $i \in \mathbb{Z}$ by* (2.9)*. Consider the numerical flux $\mathcal{F}_\lambda^{O2}$ defined by* (2.11)*,* (2.14)*. Assume the sequence of matrices $(\Theta_i^n)_{i\in\mathbb{Z}}$ be uniformly bounded with respect to $i \in \mathbb{Z}$ and bounded as $\Delta x \to 0$ then the scheme is second-order consistent in the weak-sense.*

Proof. Invoking Lemma (2.2), it suffices to use a Taylor expansion of $\mathcal{F}_\lambda^{O2}$ given by (2.11), (2.14) and the fact that u is compactly supported in space. ■

To conclude this section, the numerical scheme to be studied throughout this work writes for all $n \in \mathbb{N}$ and $i \in \mathbb{Z}$:

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{F}_{i+\frac{1}{2}}^n - \mathcal{F}_{i-\frac{1}{2}}^n\right), \tag{2.15}$$

with,

$$\begin{aligned} \mathcal{F}_{i+\frac{1}{2}}^n &= \mathcal{F}_\lambda^{O1}\left(w_i^n, w_{i+1}^n\right) + \frac{1}{2}\left(\alpha_{i+1}^n + \alpha_i^n\right), \\ \alpha_i^n &= \frac{\lambda}{2}\left(\Theta_i^n\delta_{i+\frac{1}{2}}^n + (I - \Theta_i^n)\delta_{i-\frac{1}{2}}^n\right), \\ \delta_{i+\frac{1}{2}}^n &= w_{i+1}^n - w_i^n, \end{aligned} \tag{2.16}$$

where $\mathcal{F}_\lambda^{O1}$ is defined in (2.2).

3. Quadratic stability for the scalar linear transport equation

The purpose of this section is to explicit (in a simple case) the dissipative property (2.13) required by the second-order discretization in space, and to establish a global entropy stability property (1.4) under a suitable CFL condition. We therefore consider momentarily the scalar linear transport equation with velocity $a \neq 0$,

$$\begin{cases} \partial_t w + a\partial_x w = 0, & t > 0, \quad x \in \mathbb{R}, \\ w(x, t = 0) = w_0(x). \end{cases} \tag{3.1}$$

In this particular case, our numerical scheme (2.15) reads

$$w_i^{n+1} = w_i^n - \frac{\nu}{2}\left(\frac{a}{\lambda}\left(\delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n\right) - \delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n + \frac{1}{\lambda}\left(\alpha_{i+1}^n - \alpha_{i-1}^n\right)\right), \tag{3.2}$$

where $\nu = \lambda\Delta t/\Delta x$, α_i^n and $\delta_{i+\frac{1}{2}}^n$ are defined in (2.16). For the stability analysis, it will be convenient to consider the standard discrete Sobolev spaces defined by

$$l^2(\mathbb{Z}) = \left\{ v \in \mathbb{R}^{\mathbb{Z}} \,\middle|\, \sum_{i\in\mathbb{Z}} |v_i|^2 \Delta x < +\infty \right\}, \tag{3.3}$$

$$h^2(\mathbb{Z}) = \left\{ v \in l^2(\mathbb{Z}) \,\middle|\, \left(\frac{v_{i+1}-v_i}{\Delta x}\right)_{i\in\mathbb{Z}} \in l^2(\mathbb{Z}),\ \left(\frac{v_{i+1}-2v_i+v_{i-1}}{\Delta x^2}\right)_{i\in\mathbb{Z}} \in l^2(\mathbb{Z}) \right\}. \tag{3.4}$$

We begin with the following easy preliminary lemma.

Lemma 3.1 (Existence of dissipative corrections)**.** *For any non zero sequence $(w_i^n)_{i\in\mathbb{Z}}$ in $h^2(\mathbb{Z})$, there exists $(\Theta_i^n)_{i\in\mathbb{Z}}$ such that the scheme* (3.2) *is globally dissipative in the sense of Definition 2.3 relatively to the quadratic entropy $\eta(s) = s^2$. Moreover, the dissipative inequality* (2.13) *is equivalent to the following inequality*

$$\sum_{i\in\mathbb{Z}} \left(\left(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n\right)^2 - 2\Theta_i^n \left((\delta_{i+\frac{1}{2}}^n)^2 - (\delta_{i-\frac{1}{2}}^n)^2 \right) \right) > 0. \tag{3.5}$$

Proof. Using Definition 2.3, we compute the discrete l^2-scalar product of the divergence of the flux with $(w_i^n)_{i\in\mathbb{Z}}$. We then obtain,

$$\begin{aligned}
\sum_{i\in\mathbb{Z}} w_i^n \left(\frac{a}{\lambda}\left(\delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n\right) - \delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n + \frac{1}{\lambda}\left(\alpha_{i+1}^n - \alpha_{i-1}^n\right) \right) \\
= \frac{1}{2}\sum_{i\in\mathbb{Z}} (\delta_{i+\frac{1}{2}}^n)^2 - \frac{1}{2}\sum_{i\in\mathbb{Z}} \Theta_i^n \left((\delta_{i+\frac{1}{2}}^n)^2 - (\delta_{i-\frac{1}{2}}^n)^2 \right) - \frac{1}{2}\sum_{i\in\mathbb{Z}} \delta_{i+\frac{1}{2}}^n \delta_{i-\frac{1}{2}}^n, \\
= \frac{1}{4}\sum_{i\in\mathbb{Z}} \left(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n\right)^2 - \frac{1}{2}\sum_{i\in\mathbb{Z}} \Theta_i^n \left((\delta_{i+\frac{1}{2}}^n)^2 - (\delta_{i-\frac{1}{2}}^n)^2 \right).
\end{aligned}$$

Now, the case $\Theta_i^n = 0$ for all $i \in \mathbb{Z}$ shows the existence because the sequence $(w_i^n)_{i\in\mathbb{Z}}$ is non constant and in $h^2(\mathbb{Z})$. ■

The inequality (3.5) is the mathematical expression of the dissipative property relatively to the quadratic entropy that we want to ensure for the second-order in space discretization. It takes a more general form when considering non linear hyperbolic systems with a general entropy η (it is the inequality (4.4)). Several choices of the sequence $(\Theta_i^n)_{i\in\mathbb{Z}}$ are possible. For example the choice $\Theta_i^n = -\operatorname{sign}((\delta_{i+\frac{1}{2}}^n)^2 - (\delta_{i-\frac{1}{2}}^n)^2)$ is another possibility. Some are very promising as we shall see in the numerical examples. We now prove the following.

Proposition 3.2 (Quadratic stability with a parabolic CFL condition)**.** *Let $(w_i^n)_{i\in\mathbb{Z}}$ be a non zero sequence in $h^2(\mathbb{Z})$ and $(\Theta_i^n)_{i\in\mathbb{Z}}$ be a sequence such that the inequality* (3.5) *holds. Let $\lambda > 0$ and set $\nu = \frac{\lambda\Delta t}{\Delta x}$ such that*

$$0 < \nu \le \frac{\sum_{i\in\mathbb{Z}} \left(\left(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n\right)^2 - 2\Theta_i^n \left((\delta_{i+\frac{1}{2}}^n)^2 - (\delta_{i-\frac{1}{2}}^n)^2 \right) \right)}{\sum_{i\in\mathbb{Z}} \left(\frac{a}{\lambda}\left(\delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n\right) - \delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n + \frac{1}{\lambda}\left(\alpha_{i+1}^n - \alpha_{i-1}^n\right) \right)^2}, \tag{3.6}$$

then the first-order in time scheme (3.2) *is L^2-stable, $\sum_{i\in\mathbb{Z}} |w_i^{n+1}|^2 \Delta x \le \sum_{i\in\mathbb{Z}} |w_i^n|^2 \Delta x$.*

Proof. Using the scheme (3.2), we develop the square

$$\sum_{i\in\mathbb{Z}}(w_i^{n+1})^2\Delta x = \sum_{i\in\mathbb{Z}}(w_i^n)^2\Delta x - \nu\sum_{i\in\mathbb{Z}} w_i^n\left(\frac{a}{\lambda}\left(\delta^n_{i+\frac12}+\delta^n_{i-\frac12}\right)-\delta^n_{i+\frac12}+\delta^n_{i-\frac12}+\frac{1}{\lambda}\left(\alpha^n_{i+1}-\alpha^n_{i-1}\right)\right)\Delta x$$
$$+\frac{\nu^2}{4}\sum_{i\in\mathbb{Z}}\left(\frac{a}{\lambda}\left(\delta^n_{i+\frac12}+\delta^n_{i-\frac12}\right)-\delta^n_{i+\frac12}+\delta^n_{i-\frac12}+\frac{1}{\lambda}\left(\alpha^n_{i+1}-\alpha^n_{i-1}\right)\right)^2\Delta x.$$

Now, we develop the second term of the above equation and recall that $\delta^n_{i+\frac12} = w^n_{i+1}-w^n_i$, so we have

$$\sum_{i\in\mathbb{Z}} w_i^n\left(\frac{a}{\lambda}\left(\delta^n_{i+\frac12}+\delta^n_{i-\frac12}\right)-\delta^n_{i+\frac12}+\delta^n_{i-\frac12}+\frac{1}{\lambda}\left(\alpha^n_{i+1}-\alpha^n_{i-1}\right)\right)$$
$$=\frac{a}{\lambda}\sum_{i\in\mathbb{Z}} w_i^n\left(w^n_{i+1}-w^n_{i-1}\right)-\sum_{i\in\mathbb{Z}} w_i^n\delta^n_{i+\frac12}+\sum_{i\in\mathbb{Z}} w_i^n\delta^n_{i-\frac12}-\frac{1}{\lambda}\sum_{i\in\mathbb{Z}}\alpha_i^n\left(\delta^n_{i+\frac12}+\delta^n_{i-\frac12}\right),$$

where we have used a translation of indices for the third term. Using now the fact that $\frac{2\alpha_i^n}{\lambda} = \Theta_i^n\delta^n_{i+\frac12}+(1-\Theta_i^n)\delta^n_{i-\frac12}$ and again a translation of indices, we have

$$\sum_{i\in\mathbb{Z}} w_i^n\left(\frac{a}{\lambda}\left(\delta^n_{i+\frac12}+\delta^n_{i-\frac12}\right)-\delta^n_{i+\frac12}+\delta^n_{i-\frac12}+\frac{1}{\lambda}\left(\alpha^n_{i+1}-\alpha^n_{i-1}\right)\right)$$
$$=\frac12\sum_{i\in\mathbb{Z}}(\delta^n_{i+\frac12})^2-\frac12\sum_{i\in\mathbb{Z}}\Theta_i^n\left((\delta^n_{i+\frac12})^2-(\delta^n_{i-\frac12})^2\right)-\frac12\sum_{i\in\mathbb{Z}}\delta^n_{i+\frac12}\delta^n_{i-\frac12},$$
$$=\frac14\sum_{i\in\mathbb{Z}}\left(\delta^n_{i+\frac12}-\delta^n_{i-\frac12}\right)^2-\frac12\sum_{i\in\mathbb{Z}}\Theta_i^n\left((\delta^n_{i+\frac12})^2-(\delta^n_{i-\frac12})^2\right).$$

As a consequence, we obtain

$$\sum_{i\in\mathbb{Z}}(w_i^{n+1})^2\Delta x = \sum_{i\in\mathbb{Z}}(w_i^n)^2\Delta x+\frac{\nu}{4}\Bigg(-\sum_{i\in\mathbb{Z}}\left(\left(\delta^n_{i+\frac12}-\delta^n_{i-\frac12}\right)^2-2\Theta_i^n\left((\delta^n_{i+\frac12})^2-(\delta^n_{i-\frac12})^2\right)\right)\Delta x$$
$$+\frac{\nu^2}{4}\sum_{i\in\mathbb{Z}}\left(\frac{a}{\lambda}\left(\delta^n_{i+\frac12}+\delta^n_{i-\frac12}\right)-\delta^n_{i+\frac12}+\delta^n_{i-\frac12}+\frac{1}{\lambda}\left(\alpha^n_{i+1}-\alpha^n_{i-1}\right)\right)^2\Delta x\Bigg),$$

Because of the inequality (3.5) the second sum after the equality is negative, therefore the inequality (3.6) yields the result. ■

A consistency analysis of the CFL condition (3.6) for smooth solutions yields an order of magnitude which is such that $\frac{\Delta t}{\Delta x^2} \underset{\Delta x\to 0}{=} \mathcal{O}(1)$. The CFL condition of the scheme (3.2) is then of parabolic type. Such a CFL condition has already been obtained in [46] for the TECNO schemes. It is possible to recover a hyperbolic CFL magnitude $\frac{\Delta t}{\Delta x} \underset{\Delta x\to 0}{=} \mathcal{O}(1)$ using a second-order in time extension of our scheme. Indeed, if we adopt an explicit second-order in time discretization based on the SSP Runge–Kutta methods introduced in [18, 19, 20] then we are led to the following scheme:

$$\begin{aligned} w_i^{n+\frac12} &= w_i^n-\frac{\nu}{2}\left(\frac{a}{\lambda}(\delta^n_{i+\frac12}+\delta^n_{i-\frac12})-\delta^n_{i+\frac12}+\delta^n_{i-\frac12}+\frac{1}{\lambda}(\alpha^n_{i+1}+\alpha^n_{i-1})\right),\\ w_i^{n+1} &= \frac{w_i^n}{2}+\frac12\left(w_i^{n+\frac12}-\frac{\nu}{2}\left(\frac{a}{\lambda}(\delta^{n+\frac12}_{i+\frac12}+\delta^{n+\frac12}_{i-\frac12})-\delta^{n+\frac12}_{i+\frac12}+\delta^{n+\frac12}_{i-\frac12}+\frac{1}{\lambda}(\alpha^{n+\frac12}_{i+1}+\alpha^{n+\frac12}_{i-1})\right)\right).\end{aligned} \tag{3.7}$$

Chosing the sequences $(\Theta_i^m)_{i\in\mathbb{Z}}$ for $m\in\left\{n, n+\frac12\right\}$ to be constant in the above scheme, then yields the following.

Proposition 3.3 (Quadratic stability with hyperbolic CFL). *Consider* $(\Theta_i^n)_{i\in\mathbb{Z}}$ *and* $(\Theta_i^{n+\frac{1}{2}})_{i\in\mathbb{Z}}$ *be constant sequences both equal to a constant* $\Theta \in \mathbb{R}$ *and* $(w_i^n)_{i\in\mathbb{Z}} \in l^2(\mathbb{Z})$. *Let* $\lambda > 0$ *and* $\nu = \frac{\lambda\Delta t}{\Delta x}$ *such that*

$$0 < \nu \max\left(1, \left(\frac{a}{\lambda}\right)^2, \left(1 - 2\Theta + \frac{a}{\lambda}\right)^2\right) \leq 1, \tag{3.8}$$

then the second-order in time scheme (3.7) *is* L^2*-stable,* $\sum_{i\in\mathbb{Z}} |w_i^{n+1}|^2 \Delta x \leq \sum_{i\in\mathbb{Z}} |w_i^n|^2 \Delta x$.

Proof. We denote $\hat{w}_\Delta(\xi, t^n)$, the Fourier transform of $w_\Delta(x, t^n)$ given by (1.5). We have

$$\hat{w}_\Delta(\xi, t^n) = \frac{\sin(\pi\xi\Delta x)}{\pi\xi} \sum_{j\in\mathbb{Z}} w_j^n e^{-2\pi \mathrm{i}\xi j\Delta x},$$

where $\mathrm{i}^2 = -1$. Using the scheme (3.7), we deduce $\hat{w}_\Delta(\xi, t^{n+1}) = g(\xi\Delta x)\hat{w}_\Delta(\xi, t^n)$ where $g(\xi\Delta x)$ is the amplification factor at the frequency ξ associated with the scheme (3.7). Thanks to the Fourier isometry property, to prove the L^2 stability of the scheme (3.7) it is sufficient to prove $|g(\xi\Delta x)| \leq 1$ for all $\xi \in \mathbb{R}$. For the sake of conciseness in the notation, we set $\phi_\xi = \sin^2(\pi\xi\Delta x)$ and $h_\Theta = (1-2\Theta)\sin(\pi\xi\Delta x) + \frac{a}{\lambda}$. It can be shown that the amplification factor $g(\xi\Delta x)$ verifies

$$\begin{aligned}
|g(\xi\Delta x)|^2 &= \left(1 - 2\nu\phi_\xi\left(\phi_\xi - \nu\phi_\xi^3 + \nu(1-\phi_\xi)h_\Theta^2\right)\right)^2 + 4\phi_\xi(1-\phi_\xi)\left(\nu h_\Theta\left(1 - 2\nu\phi_\xi^2\right)\right)^2, \\
&= 1 - 4\nu\phi_\xi\left(\phi_\xi - \nu\phi_\xi^3 + \nu(1-\phi_\xi)h_\Theta^2\right) + 4\nu^2\phi_\xi^2\left(\phi_\xi - \nu\phi_\xi^3 + \nu(1-\phi_\xi)h_\Theta^2\right)^2 \\
&\quad + 4\phi_\xi(1-\phi_\xi)\left(\nu h_\Theta\left(1 - 2\nu\phi_\xi^2\right)\right)^2, \\
&= 1 - 4\nu\phi_\xi^2\left(1 - \nu\phi_\xi^2\right) + 4\nu^2\phi_\xi(1-\phi_\xi)h_\Theta^2\left(-1 + \left(1 - 2\nu\phi_\xi^2\right)^2\right) \\
&\quad + 4\nu^2\phi_\xi^2\left(\phi_\xi - \nu\phi_\xi^3 + \nu(1-\phi_\xi)h_\Theta^2\right)^2.
\end{aligned}$$

We then notice that $-1 + \left(1 - 2\nu\phi_\xi^2\right)^2 = -4\nu\phi_\xi^2\left(1 - \nu\phi_\xi^2\right)$, so that

$$|g(\xi\Delta x)|^2 = 1 - 4\nu\phi_\xi^2\left(1 - \nu\phi_\xi^2\right)\left(1 + 4\nu^2\phi_\xi(1-\phi_\xi)h_\Theta^2\right) + 4\nu^2\phi_\xi^2\left(\phi_\xi - \nu\phi_\xi^3 + \nu(1-\phi_\xi)h_\Theta^2\right)^2.$$

We remark that the restriction of $|g(\xi\Delta x)|$ in the case where $\phi_\xi = 1$, yields

$$|g(\xi\Delta x)|^2_{|\phi_\xi=1} = 1 - 4\nu(1-\nu) + 4\nu^2(1-\nu)^2 = (1 - 2\nu(1-\nu))^2.$$

We see that if $\nu \geq 1$ then $|g(\xi\Delta x)|^2_{|\phi_\xi=1} \geq 1$, therefore it is necessary to consider $\nu \leq 1$. So we now consider $0 \leq \nu \leq 1$. Since for all $\xi \in \mathbb{R}$, $0 \leq \phi_\xi \leq 1$ we have

$$0 \leq (1-\phi_\xi) \leq \left(1 - \nu\phi_\xi^2\right) \quad \text{and} \quad \left(1 - \nu\phi_\xi^2\right)^2 \leq \left(1 - \nu\phi_\xi^2\right). \tag{3.9}$$

Consequently,

$$\begin{aligned}
\left(\phi_\xi - \nu\phi_\xi^3 + \nu(1-\phi_\xi)h_\Theta^2\right)^2 &= \left(\phi_\xi\left(1 - \nu\phi_\xi^2\right) + \nu(1-\phi_\xi)h_\Theta^2\right)^2, \\
&= \left(1 - \nu\phi_\xi^2\right)^2\phi_\xi^2 + 2\nu\phi_\xi\left(1 - \nu\phi_\xi^2\right)(1-\phi_\xi)h_\Theta^2 + \nu^2(1-\phi_\xi)^2 h_\Theta^4, \\
&\leq \left(1 - \nu\phi_\xi^2\right)\left(\phi_\xi^2 + 2\nu\phi_\xi(1-\phi_\xi)h_\Theta^2 + \nu^2\left(1 - \nu\phi_\xi^2\right)h_\Theta^4\right), \\
&\leq \left(1 - \nu\phi_\xi^2\right)\left(\phi_\xi^2 + 2\nu\phi_\xi(1-\phi_\xi)h_\Theta^2 + \nu\left(1 - \nu\phi_\xi^2\right)h_\Theta^4\right),
\end{aligned}$$

where we have used (3.9) and the fact that $\nu^2 \leq \nu$. We eventually obtain

$$|g(\xi\Delta x)|^2 \leq 1 - 4\nu\phi_\xi^2\left(1-\nu\phi_\xi^2\right)\left(1+2\nu^2\phi_\xi\left(1-\phi_\xi\right)h_\Theta^2 - \nu\phi_\xi^2 - \nu^2\left(1-\nu\phi_\xi^2\right)h_\Theta^4\right),$$
$$\leq 1 - 2\nu^3\phi_\xi^2\phi_{2\xi}\left(1-\nu\phi_\xi^2\right)h_\theta^2 - 4\nu\phi_\xi^2\left(1-\nu\phi_\xi^2\right)^2\left(1-\nu^2 h_\Theta^4\right),$$

which is upper bounded by 1 because of the CFL condition (3.8) and the definition of h_Θ. ∎

In this proof, the sequences $(\Theta_i^n)_{i\in\mathbb{Z}}$ and $(\Theta_i^{n+\frac{1}{2}})_{i\in\mathbb{Z}}$ are taken to be constant which is convenient for the Fourier analysis. However, in this very particular case, the scheme does not enjoy the extra dissipation in the inequality (3.5): the second sum vanishes. We believe that taking $(\Theta_i^n)_{i\in\mathbb{Z}}$ non-constant and ensuring (3.5) improves the stability.

4. Global entropy inequality in the general case

In this section, we establish the global entropy inequality (1.4) for the second-order in space and first-order in time scheme (2.15) in the general case of the system (1.1) with a given pair of entropy, entropy flux (η, G) defined in Ω. For the forthcoming developments, it is convenient to condense the scheme (2.15) in the form

$$w_i^{n+1} = w_i^n + \frac{\Delta t}{\Delta x}\mathcal{R}_i^n, \tag{4.1}$$

where

$$\mathcal{R}_i^n = -\frac{1}{2}\left(f(w_{i+1}^n) - f(w_{i-1}^n)\right) + \frac{\lambda}{2}\left(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n\right) - \frac{1}{2}\left(\alpha_{i+1}^n - \alpha_{i-1}^n\right). \tag{4.2}$$

Note that in the general case Ω is not necessarily equal to $\mathbb{R}^d$. Typically, for the Euler equations we have $\Omega = \{w = (\rho, \rho u, \rho E) \in \mathbb{R}^3 : \rho > 0, \rho E - \rho u^2/2 > 0\}$. Since our proof is based on Taylor expansion of the entropy of the updated state $\eta(w_i^{n+1})$, we have, at the very first, to make sure it belongs to Ω. For second-order accurate in space scheme, it is a challenging issue to establish it without using limitation techniques. We thus use a very simple topological argument which is also quite restrictive from the CFL point of view. It is summarized in the following lemma.

Lemma 4.1 (Scheme robustness)**.** *Let Ω be a strict non empty convex open subset of $\mathbb{R}^d$. Let $(w_i^n)_{i\in\mathbb{Z}} \subset K^n$, where K^n is a compact set of Ω and assume the sequence of matrices $(\Theta_i^n)_{i\in\mathbb{Z}}$ be bounded. There exists a constant $c^n > 0$ such that if $0 < \frac{\Delta t}{\Delta x} \leq c^n$ then the sequence $(w_i^{n+1})_{i\in\mathbb{Z}}$ given by the scheme* (4.1)*, or equivalently* (2.15)*, is contained in Ω.*

Proof. We recall that the signed distance function to the boundary of Ω associated with the Euclidean norm $\|\cdot\|_2$ on $\mathbb{R}^d$ is defined for all $w \in \mathbb{R}^d$ by:

$$\operatorname{dist}(w, \partial\Omega) = \begin{cases} \inf_{y\in\partial\Omega}\|w-y\|_2, & \text{if } w \in \overline{\Omega}, \\ -\inf_{y\in\partial\Omega}\|w-y\|_2 & \text{if } w \in \mathbb{R}^d \setminus \overline{\Omega}. \end{cases}$$

It is a 1-Lipschitz continuous function on $\mathbb{R}^d$. Therefore using the scheme (4.1), we have for all $i \in \mathbb{Z}$,

$$\left|\operatorname{dist}(w_i^{n+1}, \partial\Omega) - \operatorname{dist}(w_i^n, \partial\Omega)\right| \leq \|w_i^{n+1} - w_i^n\|_2 = \frac{\Delta t}{\Delta x}\|\mathcal{R}_i^n\|_2. \tag{4.3}$$

We then obtain

$$\operatorname{dist}(w_i^{n+1}, \partial\Omega) \geq \operatorname{dist}(w_i^n, \partial\Omega) - \frac{\Delta t}{\Delta x}\|\mathcal{R}_i^n\|_2.$$

Since $(w_i^n)_{i\in\mathbb{Z}} \subset K^n$ and the signed distance function is continuous on the compact set K^n, one has $\operatorname{dist}(w_i^n, \partial\Omega) \geq \min_{w\in K^n}\operatorname{dist}(w, \partial\Omega) > 0$ where the minimum is positive because K^n is a compact set into an open set. Using the continuity of the physical flux f on the compact set K^n and the fact that the sequence $(\Theta_i^n)_{i\in\mathbb{Z}}$ is bounded, there exists a constant R^n such that for all $i \in \mathbb{Z}$, $\|\mathcal{R}_i^n\|_2 \leq R^n$. We thus have for all $i \in \mathbb{Z}$,

$$\operatorname{dist}\left(w_i^{n+1}, \partial\Omega\right) \geq \min_{w\in K^n} \operatorname{dist}\left(w, \partial\Omega\right) - \frac{\Delta t}{\Delta x} R^n.$$

If $0 < \frac{\Delta t}{\Delta x} \leq \frac{\min_{w\in K^n} \operatorname{dist}(w,\partial\Omega)}{2R^n}$ then for all $i \in \mathbb{Z}$, $\operatorname{dist}\left(w_i^{n+1}, \partial\Omega\right) \geq \frac{1}{2} \min_{w\in K^n} \operatorname{dist}\left(w, \partial\Omega\right) > 0$ which exactly means that $w_i^{n+1} \in \Omega$. ■

Our main result is the following.

Theorem 4.2 (Global entropy inequality)**.** *Let Ω be a non-empty convex open subset of $\mathbb{R}^d$. Consider $(\eta, G) \in C^2(\Omega, \mathbb{R}) \times C^1(\Omega, \mathbb{R})$ a pair of strictly convex entropy-entropy-flux which satisfies* (1.2)*. Let $(w_i^n)_{i\in\mathbb{Z}}$ be a non zero sequence in $h^2(\mathbb{Z})$ compactly supported and moreover such that $\sum_{i\in\mathbb{Z}} \eta(w_i^n)\Delta x$ is finite. Let $(\Theta_i^n)_{i\in\mathbb{Z}}$, a sequence of bounded matrices such that*

$$\mathcal{S}^n := \int_0^1 \sum_{i\in\mathbb{Z}} N_i^n(s) P_i^n(s) D_i^n \cdot D_i^n \mathrm{d}s > 0, \tag{4.4}$$

where the block matrices $(D_i^n, N_i^n, P_i^n) \in \mathbb{R}^{2d} \times (\mathcal{M}_{2d}(\mathbb{R}))^2$ are respectively defined by

$$\begin{aligned} D_i^n &= \begin{pmatrix} \delta_{i-\frac{1}{2}}^n \\ \delta_{i+\frac{1}{2}}^n \end{pmatrix}, \\ N_i^n(s) &= \begin{pmatrix} \nabla^2\eta\left(w_i^n - s\delta_{i-\frac{1}{2}}^n\right) & 0 \\ 0 & \nabla^2\eta\left(w_i^n + s\delta_{i+\frac{1}{2}}^n\right) \end{pmatrix}, \\ P_i^n(s) &= \begin{pmatrix} (1-2s)I + \Theta_i^n & -\Theta_i^n \\ \Theta_i^n - I & 2(1-s)I - \Theta_i^n \end{pmatrix}, \end{aligned} \tag{4.5}$$

for all $s \in [0, 1]$. Also assume there exists a compact set $K^n \subset \Omega$ such that $(w_i^n)_{i\in\mathbb{Z}} \subset K^n$. Let the numerical diffusion λ be such that

$$\lambda > \lambda^n, \tag{4.6}$$

where

$$\begin{aligned} \lambda^n &= \frac{2\max\left(0, \sum_{i\in\mathbb{Z}} \int_0^1 s\left(\int_0^1 N_i^n(us)\mathrm{d}u\right) F_i^n(s) D_i^n \cdot D_i^n \mathrm{d}s\right)}{\sum_{i\in\mathbb{Z}} \int_0^1 N_i^n(s) P_i^n(s) D_i^n \cdot D_i^n \mathrm{d}s} \geq 0, \\ F_i^n(s) &= \begin{pmatrix} -\nabla f\left(w_i^n - s\delta_{i-\frac{1}{2}}^n\right) & 0 \\ 0 & \nabla f\left(w_i^n + s\delta_{i+\frac{1}{2}}^n\right) \end{pmatrix}, \quad \forall s \in [0, 1], \end{aligned} \tag{4.7}$$

and $\frac{\Delta t}{\Delta x}$ be such that

$$0 < \frac{\Delta t}{\Delta x} \leq \frac{-\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n) \cdot \mathcal{R}_i^n}{\int_0^1 (1-s) \sum_{i\in\mathbb{Z}} \nabla^2\eta\left(w_i^n + s\frac{\Delta t}{\Delta x}\mathcal{R}_i^n\right) \mathcal{R}_i^n \cdot \mathcal{R}_i^n \mathrm{d}s}. \tag{4.8}$$

If $\Omega = \mathbb{R}^d$, then one has the global entropy inequality,

$$\sum_{i\in\mathbb{Z}} \eta(w_i^{n+1})\Delta x \leq \sum_{i\in\mathbb{Z}} \eta(w_i^n)\Delta x. \tag{4.9}$$

If $\Omega \neq \mathbb{R}^d$ then the global entropy inequality (4.9) still holds if moreover $0 < \frac{\Delta t}{\Delta x} \leq c^n$ where c^n is a positive non explicit constant given by the Lemma 4.1.

Few remarks are in order about this result:

- The numerical diffusion λ^n given by (4.7) has an explicit formula and can be implemented.
- The inequality (4.4) ensures the dissipative property of the scheme in the sense of Definition 2.3. It is an inequality to be satisfied by the matrix parameter (Θ_i^n) and we shall prove that it has always a solution (see Proposition 4.4).
- The condition (4.6) on λ ensures that there is enough diffusion in the scheme so that it can be globally stable.
- The CFL condition (4.8) is a priori non linear (except when considering the quadratric entropy) but it can be easily solved numerically.

Before proving this theorem, we shall need the first following Lemma.

Lemma 4.3 (Poincaré inequality)**.** *Let $(w_i^n)_{i\in\mathbb{Z}} \subset \Omega$ be a compactly supported sequence in $h^2(\mathbb{Z})$. Let $\eta \in C^2(\Omega, \mathbb{R})$ a strictly convex entropy. Then there exists a constant C which depends on $(w_i^n)_{i\in\mathbb{Z}}$ and η such that*

$$\sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\dots,d\}}} |(\nabla\eta(w_{i+1}^n) - \nabla\eta(w_{i-1}^n)))_j\,(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j| \geq C \sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\dots,d\}}} |\delta_{i+\frac{1}{2}}^n|^2,$$

with $\delta_{i+\frac{1}{2}}^n = w_{i+1}^n - w_i^n$.

Proof. At first, we use a Taylor expansion with an integral form of the remainder to write

$$\sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\dots,d\}}} \left|(\nabla\eta(w_{i+1}^n) - \nabla\eta(w_{i-1}^n))_j\,(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j\right|$$
$$= \sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\dots,d\}}} \left|\left(\int_0^1 \nabla^2\eta(w_{i-1}^n + s(w_{i+1}^n - w_{i-1}^n))(\delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n)\mathrm{d}s\right)_j (\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j\right|.$$

Since η is a strictly convex entropy and as $(w_i^n)_{i\in\mathbb{Z}}$ is compactly supported in $h^2(\mathbb{Z})$ there exists a constant $C(\eta, w^n) > 0$ such that

$$\sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\dots,d\}}} \left|\left(\int_0^1 \nabla^2\eta(w_{i-1}^n + s(w_{i+1}^n - w_{i-1}^n))(\delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n)\mathrm{d}s\right)_j (\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j\right|$$
$$\geq C(\eta, w^n) \sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\dots,d\}}} \left|(\delta_{i+\frac{1}{2}}^n + \delta_{i-\frac{1}{2}}^n)_j(\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j\right|.$$

Now, since $(w_i^n)_{i\in\mathbb{Z}} \subset \Omega$ is a compactly supported sequence, for all i in $\mathbb{Z}$ and for all $j \in \{1,\dots,d\}$ we have

$$\left|\delta_{i+\frac{1}{2}}^n\right|_j^2 = \sum_{l\leq i} (\delta_{l+\frac{1}{2}}^n + \delta_{l-\frac{1}{2}}^n)_j(\delta_{l+\frac{1}{2}}^n - \delta_{l-\frac{1}{2}}^n)_j \leq \sum_{l\leq i} \left|(\delta_{l+\frac{1}{2}}^n + \delta_{l-\frac{1}{2}}^n)_j(\delta_{l+\frac{1}{2}}^n - \delta_{l-\frac{1}{2}}^n)_j\right|.$$

Let $\mathcal{T}$ be the support of $(\delta^n_{i+\frac{1}{2}})_{i\in\mathbb{Z}}$ which is of finite cardinal since (w^n_i) is compactly supported. Summing over $i \in \mathbb{Z}$ we obtain,

$$\sum_{i\in\mathbb{Z}} \left|\delta^n_{i+\frac{1}{2}}\right|_j^2 \leq \sum_{i\in\mathcal{T}} \sum_{l\leq i} \left|(\delta^n_{l+\frac{1}{2}} + \delta^n_{l-\frac{1}{2}})_j (\delta^n_{l+\frac{1}{2}} - \delta^n_{l-\frac{1}{2}})_j\right|,$$
$$\leq \mathrm{Card}\,(\mathcal{T}) \sum_{l\in\mathbb{Z}} \left|(\delta^n_{l+\frac{1}{2}} + \delta^n_{l-\frac{1}{2}})_j (\delta^n_{l+\frac{1}{2}} - \delta^n_{l-\frac{1}{2}})_j\right|.$$

Finally, summing over the components $j \in \{1, \ldots, d\}$, we eventually deduce,

$$\sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\ldots,d\}}} \left|\delta^n_{i+\frac{1}{2}}\right|^2 \leq \frac{\mathrm{Card}\,(\mathcal{T})}{C(\eta, w^n)} \sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\ldots,d\}}} \left|(\nabla\eta(w^n_{i+1}) - \nabla\eta(w^n_{i-1}))_j (\delta^n_{i+\frac{1}{2}} - \delta^n_{i-\frac{1}{2}})_j\right|.$$

■

Proposition 4.4 (Existence of dissipative corrections)**.** *Let $(w^n_i)_{i\in\mathbb{Z}} \subset \Omega$ be non constant sequence compactly supported in $h^2(\mathbb{Z})$ and let $\eta \in C^2(\Omega, \mathbb{R})$ a strictly convex entropy. If $(\Theta^n_i)_{i\in\mathbb{Z}}$ verifies*

$$\Theta^n_i = -\theta\,\mathrm{diag}_{1\leq j\leq d} \left(\mathrm{sign}\left((\nabla\eta(w^n_{i+1}) - \nabla\eta(w^n_{i-1}))_j\, (\delta^n_{i+\frac{1}{2}} - \delta^n_{i-\frac{1}{2}})_j \right)\right), \tag{4.10}$$

with

$$\begin{aligned} \theta &> \frac{-\min\left(0, \int_0^1 \sum_{i\in\mathbb{Z}} Q^n_i(s) D^n_i \cdot D^n_i \mathrm{d}s\right)}{\sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\ldots,d\}}} \left|\left(\nabla\eta(w^n_{i+1}) - \nabla\eta(w^n_{i-1})\right)\right)_j (\delta^n_{i+\frac{1}{2}} - \delta^n_{i-\frac{1}{2}})_j\right|}, \\ Q^n_i(s) &= \begin{pmatrix} (1-2s)\nabla^2\eta(w^n_i - s\delta^n_{i-\frac{1}{2}}) & -\nabla^2\eta(w^n_i + s\delta^n_{i+\frac{1}{2}}) \\ 0 & 2(1-s)\nabla^2\eta(w^n_i + s\delta^n_{i+\frac{1}{2}}) \end{pmatrix}, \end{aligned} \tag{4.11}$$

for all $s \in [0, 1]$, then the dissipative inequality (4.4) *holds.*

Proof. We first remark that θ is always well-defined because, by virtue of Lemma 4.3, the denominator in (4.11) cannot be zero since the sequence $(w^n_i)_{i\in\mathbb{Z}}$ is non-constant. Moreover since (w^n_i) is compactly supported, the ratio in (4.11) is finite. Next, since the matrix $(\Theta^n_i)_{i\in\mathbb{Z}}$ defined by(4.10) are symmetric (because they are diagonal) we have $\Theta^n_i a \cdot b = a \cdot \Theta^n_i b$, for all vectors $(a, b) \in (\mathbb{R}^d)^2$. As a consequence, from the definition of the matrices P^n_i, N^n_i given by (4.5) and the definition of $\mathcal{S}^n$ given in (4.4), we have

$$\mathcal{S}^n = \int_0^1 \sum_{i\in\mathbb{Z}} Q^n_i(s) D^n_i \cdot D^n_i \mathrm{d}s - \sum_{i\in\mathbb{Z}} \Theta^n_i \left(\nabla\eta(w^n_{i+1}) - \nabla\eta(w^n_{i-1})\right) \cdot \left(\delta^n_{i+\frac{1}{2}} - \delta^n_{i-\frac{1}{2}}\right).$$

Using the $(\Theta^n_i)_{i\in\mathbb{Z}}$ formula (4.10), we eventually obtain

$$\mathcal{S}^n = \int_0^1 \sum_{i\in\mathbb{Z}} Q^n_i(s) D^n_i \cdot D^n_i \mathrm{d}s + \theta \sum_{\substack{i\in\mathbb{Z}\\ j\in\{1,\ldots,d\}}} \left| \left((\nabla\eta(w^n_{i+1}) - (\nabla\eta(w^n_{i-1}))\right)_j (\delta^n_{i+\frac{1}{2}} - \delta^n_{i-\frac{1}{2}})_j\right|,$$

which is positive with θ verifying the inequality (4.11). ■

4.1. Reformulation of the global dissipation

We have the following.

Lemma 4.5 (Reformulation of the global entropy flux)**.** *Let the sequence* $(w_i^n)_{i\in\mathbb{Z}} \subset \Omega$ *be in* $h^2(\mathbb{Z})$ *and let* $(\eta, G) \in C^2(\Omega,\mathbb{R}) \times C^1(\Omega,\mathbb{R})$ *a pair of strictly convex entropy-entropy-flux which satisfies* (1.2)*. Then*

$$-\frac{1}{2}\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n)\cdot\left(f(w_{i+1}^n)-f(w_{i-1}^n)\right) = \sum_{i\in\mathbb{Z}}\int_0^1 \frac{s}{2}\left(\int_0^1 N_i^n(us)\mathrm{d}u\right) F_i^n(s) D_i^n\cdot D_i^n \mathrm{d}s,$$

where D_i^n*,* $N_i^n(s), F_i^n(s)$ *are defined in* (4.5)–(4.7) *respectively.*

Proof. Since $\eta\in C^2(\Omega,\mathbb{R})$, $G\in C^1(\Omega,\mathbb{R})$, f is assumed to be smooth and $(w_i^n)_{i\in\mathbb{Z}}$ belongs to Ω, a Taylor expansion with an integral form of the remainder gives,

$$\begin{aligned}
G(w_{i+1}^n) &= G(w_i^n) + \int_0^1 \nabla G\left(w_i^n + s\delta_{i+\frac12}^n\right)\cdot \delta_{i+\frac12}^n \mathrm{d}s,\\
\nabla\eta(w_i^n + s\delta_{i+\frac12}^n) &= \nabla\eta(w_i^n) + s\int_0^1 \nabla^2\eta\left(w_i^n + us\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n \mathrm{d}u,\\
f(w_{i+1}^n) &= f(w_i^n) + \int_0^1 \nabla f\left(w_i^n + s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n \mathrm{d}s,
\end{aligned}$$

from which we deduce

$$\begin{aligned}
&\nabla\eta\left(w_i^n\right)\cdot\left(f(w_{i+1}^n)-f(w_i^n)\right)\\
&\qquad= \int_0^1 \nabla\eta(w_i^n)\cdot\nabla f\left(w_i^n + s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n \mathrm{d}s,\\
&\qquad= \int_0^1 \nabla G\left(w_i^n + s\delta_{i+\frac12}^n\right)\cdot\delta_{i+\frac12}^n \mathrm{d}s\\
&\qquad\quad - \int_0^1\int_0^1 s\nabla^2\eta\left(w_i^n + us\delta_{i+\frac12}^n\right)\nabla f\left(w_i^n + s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n\cdot\delta_{i+\frac12}^n \mathrm{d}u\mathrm{d}s,\\
&\qquad= G(w_{i+1}^n) - G(w_i^n)\\
&\qquad\quad - \int_0^1\int_0^1 s\nabla^2\eta\left(w_i^n + us\delta_{i+\frac12}^n\right)\nabla f\left(w_i^n + s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n\cdot\delta_{i+\frac12}^n \mathrm{d}u\mathrm{d}s.
\end{aligned}\tag{4.12}$$

In the same way, we have

$$\begin{aligned}
&\nabla\eta(w_i^n)\cdot\left(f(w_i^n)-f(w_{i-1}^n)\right)\\
&\quad= G(w_i^n) - G(w_{i-1}^n) + \int_0^1\int_0^1 s\nabla^2\eta\left(w_i^n - us\delta_{i-\frac12}^n\right)\nabla f\left(w_i^n - s\delta_{i-\frac12}^n\right)\delta_{i-\frac12}^n\cdot\delta_{i-\frac12}^n \mathrm{d}u\mathrm{d}s,
\end{aligned}\tag{4.13}$$

Using the equations (4.12), (4.13) and the definitions (4.5), (4.7) we deduce the expected result. ∎

Lemma 4.6 (Reformulation of the global dissipation)**.** *Let the sequences* $(w_i^n)_{i\in\mathbb{Z}}\subset\Omega$ *be in* $h^2(\mathbb{Z})$ *and let* $\eta\in C^2(\Omega,\mathbb{R})$*. Then*

$$\frac{1}{2}\sum_{i\in\mathbb{Z}}\left(\lambda\nabla\eta(w_i^n)\cdot(\delta_{i+\frac12}^n - \delta_{i-\frac12}^n) - \nabla\eta(w_i^n)\cdot(\alpha_{i+1}^n - \alpha_{i-1}^n)\right) = -\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\int_0^1 N_i^n(s)P_i^n(s)D_i^n\cdot D_i^n\mathrm{d}s.$$

where the block matrices D_i^n*,* N_i^n*,* P_i^n *are defined in the equation* (4.5)*.*

Proof. Since $(w_i^n)_{i\in\mathbb{Z}}$ is included in Ω, using a Taylor expansion with an integral form of the remainder, we get

$$\begin{aligned}
\eta(w_{i+1}^n) &= \eta(w_i^n) + \nabla\eta(w_i^n)\cdot\delta_{i+\frac12}^n + \int_0^1 (1-s)\nabla^2\eta\left(w_i^n + s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n\cdot\delta_{i+\frac12}^n\mathrm{d}s,\\
\eta(w_{i-1}^n) &= \eta(w_i^n) - \nabla\eta(w_i^n)\cdot\delta_{i-\frac12}^n + \int_0^1 (1-s)\nabla^2\eta\left(w_i^n - s\delta_{i-\frac12}^n\right)\delta_{i-\frac12}^n\cdot\delta_{i-\frac12}^n\mathrm{d}s,
\end{aligned}\tag{4.14}$$

which yields

$$\begin{aligned}\frac{\lambda}{2}\sum_{i\in\mathbb{Z}}\nabla\eta(w_i^n)\cdot\left(\delta_{i+\frac12}^n-\delta_{i-\frac12}^n\right)&=\frac{\lambda}{2}\sum_{i\in\mathbb{Z}}\left(\eta(w_{i+1}^n)-2\eta(w_i^n)+\eta(w_{i-1}^n)\right)\\&\quad-\frac{\lambda}{2}\int_0^1(1-s)\sum_{i\in\mathbb{Z}}\nabla^2\eta\left(w_i^n+s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n\cdot\delta_{i+\frac12}^n\mathrm{d}s\\&\quad-\frac{\lambda}{2}\int_0^1(1-s)\sum_{i\in\mathbb{Z}}\nabla^2\eta\left(w_i^n-s\delta_{i-\frac12}^n\right)\delta_{i-\frac12}^n\cdot\delta_{i-\frac12}^n\mathrm{d}s,\\&=-\frac{\lambda}{2}\int_0^1(1-s)\sum_{i\in\mathbb{Z}}N_i^n(s)D_i^n\cdot D_i^n\mathrm{d}s.\end{aligned}\tag{4.15}$$

For the second-order correction term, using a translation of indices, we have

$$\begin{aligned}-\frac12\sum_{i\in\mathbb{Z}}\nabla\eta(w_i^n)\cdot\left(\alpha_{i+1}^n-\alpha_{i-1}^n\right)&=\frac12\sum_{i\in\mathbb{Z}}\left(\nabla\eta(w_{i+1}^n)-\nabla\eta(w_{i-1}^n)\right)\cdot\alpha_i^n,\\&=\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\left(\nabla\eta(w_{i+1}^n)-\nabla\eta(w_{i-1}^n)\right)\cdot\Theta_i^n\delta_{i+\frac12}^n\\&\qquad+\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\left(\nabla\eta(w_{i+1}^n)-\nabla\eta(w_{i-1}^n)\right)\cdot(I-\Theta_i^n)\delta_{i-\frac12}^n.\end{aligned}$$

But, using once again a Taylor expansion with an integral remainder, we have

$$\begin{aligned}\nabla\eta(w_{i+1}^n)&=\nabla\eta(w_i^n)+\int_0^1\nabla^2\eta\left(w_i^n+s\delta_{i+\frac12}^n\right)\delta_{i+\frac12}^n\mathrm{d}s,\\\nabla\eta(w_{i-1}^n)&=\nabla\eta(w_i^n)-\int_0^1\nabla^2\eta\left(w_i^n-s\delta_{i-\frac12}^n\right)\delta_{i-\frac12}^n\mathrm{d}s,\end{aligned}$$

from which we infer

$$\begin{aligned}-\frac12\sum_{i\in\mathbb{Z}}\nabla\eta(w_i^n)\cdot\left(\alpha_{i+1}^n-\alpha_{i-1}^n\right)&=\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\int_0^1\nabla^2\eta\left(w_i^n+s\delta_{i+\frac12}^n\right)\Theta_i^n\delta_{i+\frac12}^n\cdot\delta_{i+\frac12}^n\mathrm{d}s,\\&\quad+\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\int_0^1\nabla^2\eta\left(w_i^n+s\delta_{i+\frac12}^n\right)\left(I-\Theta_i^n\right)\delta_{i-\frac12}^n\cdot\delta_{i+\frac12}^n\mathrm{d}s\\&\quad+\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\int_0^1\nabla^2\eta\left(w_i^n-s\delta_{i-\frac12}^n\right)\Theta_i^n\delta_{i+\frac12}^n\cdot\delta_{i-\frac12}^n\mathrm{d}s\\&\quad+\frac{\lambda}{4}\sum_{i\in\mathbb{Z}}\int_0^1\nabla^2\eta\left(w_i^n-s\delta_{i-\frac12}^n\right)\left(I-\Theta_i^n\right)\delta_{i-\frac12}^n\cdot\delta_{i-\frac12}^n\mathrm{d}s.\end{aligned}\tag{4.16}$$

Rewriting (4.15) and (4.16) with the definitions (4.5) we deduce the result. ■

4.2. **Proof of the main result**

Let $(\eta, G)\in C^2(\Omega,\mathbb{R})\times C^1(\Omega,\mathbb{R})$ a pair of strictly convex entropy, entropy-flux which satisfies (1.2). Consider the sequence $(w_i^n)_{i\in\mathbb{Z}}$ verifying the assumptions of Theorem 4.2. Consider the CFL condition $\frac{\Delta t}{\Delta x}$ given in Theorem 4.2. Therefore the sequence $(w_i^{n+1})_{i\in\mathbb{Z}}$ is contained in Ω. Since $\eta\in C^2(\Omega,\mathbb{R})$, using a Taylor expansion, in the above equations, we deduce

$$\eta(w_i^{n+1})=\eta(w_i^n)+\frac{\Delta t}{\Delta x}\nabla\eta(w_i^n)\cdot\mathcal{R}_i^n+\left(\frac{\Delta t}{\Delta x}\right)^2\int_0^1(1-s)\nabla^2\eta\left(w_i^n+s\frac{\Delta t}{\Delta x}\mathcal{R}_i^n\right)\mathcal{R}_i^n\cdot\mathcal{R}_i^n\mathrm{d}s.$$

It implies that,

$$\sum_{i\in\mathbb{Z}} \eta(w_i^{n+1})\Delta x = \sum_{i\in\mathbb{Z}} \eta(w_i^n)\Delta x + \frac{\Delta t}{\Delta x}\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n)\cdot \mathcal{R}_i^n \Delta x$$
$$+\left(\frac{\Delta t}{\Delta x}\right)^2 \int_0^1 (1-s)\sum_{i\in\mathbb{Z}} \nabla^2\eta\left(w_i^n + s\frac{\Delta t}{\Delta x}\mathcal{R}_i^n\right)\mathcal{R}_i^n\cdot\mathcal{R}_i^n \mathrm{d}s\Delta x.$$

To establish the global entropy inequality (4.9), it is sufficient to prove the following inequality

$$\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n)\cdot \mathcal{R}_i^n + \frac{\Delta t}{\Delta x}\int_0^1 (1-s)\sum_{i\in\mathbb{Z}} \nabla^2\eta\left(w_i^n + s\frac{\Delta t}{\Delta x}\mathcal{R}_i^n\right)\mathcal{R}_i^n\cdot\mathcal{R}_i^n \,\mathrm{d}s \le 0. \tag{4.17}$$

Thanks to the Lemmas 4.5 and 4.6, we deduce that the dissipation reformulates as

$$\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n)\cdot \mathcal{R}_i^n = \sum_{i\in\mathbb{Z}}\int_0^1 \left(\frac{s}{2}\left(\int_0^1 N_i^n(us)\mathrm{d}u\right)F_i^n(s) - \frac{\lambda}{4}N_i^n(s)P_i^n(s)\right)D_i^n\cdot D_i^n \mathrm{d}s. \tag{4.18}$$

But, as the sequence of matrices $(\Theta_i^n)_{i\in\mathbb{Z}}$ are selected in order to satisfy the inequality (4.4), and λ is such that the inequality (4.6) is verified, we have

$$\sum_{i\in\mathbb{Z}} \nabla\eta(w_i^n)\cdot \mathcal{R}_i^n < 0,$$

which is the definition of a dissipative flux given by Definition 2.3. We may remark at this point that the sequence $(w_i^n)_{i\in\mathbb{Z}}$ is in $h^2(\mathbb{Z})$, it is in particular bounded. The Hessian of the entropy $\nabla^2\eta$ is positive (because η is strictly convex) and continuous on a compact set that depends on the sequence $(w_i^n)_{i\in\mathbb{Z}}$. So for λ^n defined in (4.7) the numerator is finite because $(w_i^n)_{i\in\mathbb{Z}}$ is in $h^2(\mathbb{Z})$. The denominator is non zero because $(w_i^n)_{i\in\mathbb{Z}}$ is a non zero sequence in $h^2(\mathbb{Z})$, so λ^n is finite. The ratio $\frac{\Delta t}{\Delta x}$ verifies the inequality (4.8). In this inequality, the right hand side is positive because the numerator is positive by construction of the sequences $(\Theta_i^n)_{i\in\mathbb{Z}}$ and the denominator is finite because the sequence $(w_i^n)_{i\in\mathbb{Z}}$ is compactly supported so the sum is convergent. We thus deduce that (4.17) is satisfied. It ends the proof of Theorem 4.2.

To conclude this section, we mention that a formal consistency analysis (which is laborious) for a smooth compactly supported solution of the CFL condition (4.8) yields formally $\frac{\Delta t}{\Delta x^2} \underset{\Delta x\to 0}{=} \mathcal{O}(1)$.

5. Numerical results

In this section, we give several numerical examples to assess the accuracy and the stability of the schemes (2.15). We consider the scalar Burgers equation, the Euler equations and a non-convex scalar flux. For all these test cases, we propose several choices of the matrix parameter Θ_i^n. For the Burgers equation, we compare our formula for the viscosity coefficient and the CFL number (4.7)–(4.8) with the standard hyperbolic ones [16]. For the Burgers and the Euler equations, we measure the error in L^1, L^2 and L^∞ norms between the numerical solutions and the exact solution. To compute the errors, we increase the time accuracy of the scheme (2.15) using the well-known SSP Rung–Kutta methods introduced in [18, 19, 20]. Since this high-order time approach is based on convex combination of first-order time sub-steps, our global entropy stability result 4.2 is preserved. Plots of the obtained numerical solutions and the global entropy are also given.

5.1. Burgers equation

This section deals with the Burgers equation. It consists in taking $\Omega = \mathbb{R}$ and the flux function is given by $f(w) = w^2/2$ for all $w \in \Omega$. We consider the entropy function $\eta(w) = w^2/2$. We shall present several tests with the following parameters:

$$\Theta^n_{a,i} = -\theta\,\mathrm{sign}\left((\delta^n_{i+\frac{1}{2}})^2 - (\delta^n_{i-\frac{1}{2}})^2\right), \qquad \Theta^n_{b,i} = -\theta\tanh\left((\delta^n_{i+\frac{1}{2}})^2 - (\delta^n_{i-\frac{1}{2}})^2\right),$$
$$\Theta^n_{c,i} = \frac{\left((\delta^n_{i-\frac{1}{2}})^2 - (\delta^n_{i+\frac{1}{2}})^2\right)\left((\delta^n_{i-\frac{1}{2}})^2 + (\delta^n_{i+\frac{1}{2}})^2\right)}{\left((\delta^n_{i-\frac{1}{2}})^2 + (\delta^n_{i+\frac{1}{2}})^2\right)^2 + \varepsilon}, \qquad \Theta^n_{d,i} = \frac{1}{2}, \tag{5.1}$$

where θ satisfies (4.11) and we fix $\varepsilon = 10^{-12}$. We numerically verify that these choices of the matrix parameter $(\Theta^n_i)_{i\in\mathbb{Z}}$ always satisfy the inequality (4.4). The numerical viscosity is taken equal to $\lambda = \max\left(\lambda^n, \lambda^{\mathrm{HLL}}\right)$ where λ^{HLL}, λ^n are respectively defined in (2.3) and (4.7), and the time step Δt is selected according to (4.8). In addition, we also compare our schemes (2.15) to the standard Rusanov scheme given by (2.2) coupled to the standard MUSCL reconstruction [45] with the minmod slope limiter.

5.1.1. *Smooth solution*

We take a smooth initial data $w_0(x) = 0.25 + 0.5\sin(\pi x)$ over a periodic domain $[-1, 1)$. With a final time small enough, here given by $t = 0.3$, the exact solution remains smooth so that the order of accuracy can be evaluated.

The numerical solutions are displayed in Figure 5.1. We notice the good behavior of the approximations. This remark is completed by Table 5.1 where the order of accuracy are evaluated for different L^p-norms. We observe asymptotically the expected second-order of accuracy whereas the Rusanov MUSCL scheme yields the following rate of convergence given by: 1.89 for the L^1-norm, 1.65 for the L^2-norm and 1.29 for the L^∞-norm. The bottom of the Figure 5.1 shows the time evolution of the numerical viscosity and the time step given by (4.8). We also compute τ being the ratio between the CPU time needed to run the simulation with our scheme (2.15) and the CPU time required to run the standard MUSCL scheme. Then we obtain

$$\tau_{\Theta^n_{i,a}} = 3111.71, \quad \tau_{\Theta^n_{i,b}} = 3250.71, \quad \tau_{\Theta^n_{i,c}} = 1633.14, \quad \tau_{\Theta^n_{i,d}} = 3053.29.$$

According to these results, and as shown in the Figure 5.1, the scheme presented in this work coupled to the formula (4.7)–(4.8) has a larger computation time. However, a direct measure shows that for all the choices of the matrices Θ^n_i given in (5.1), around 97% of the CPU time is only devoted to the computation of the formula (4.7)–(4.8). To illustrate this, we proceed to the same measure of the CPU time but using a second-order SSP time discretization for which we used the usual hyperbolic CFL condition (as adopted for the MUSCL Russanov scheme). Then we obtain

$$\tau_{\Theta^n_{i,a}} = 0.78, \quad \tau_{\Theta^n_{i,b}} = 0.94, \quad \tau_{\Theta^n_{i,c}} = 0.61, \quad \tau_{\Theta^n_{i,d}} = 0.59.$$

In this case, the scheme introduced in this work is comparable to the standard Rusanov MUSCL scheme and still, we observe the decrease of the global entropy.

5.1.2. *Discontinuous solution*

We take a discontinuous initial data over the periodic domain $[-1, 1)$ defined by

$$w_0(x) = \begin{cases} 1 & \text{if } \quad -0.25 \le x \le 0.25, \\ 0 & \text{otherwise.} \end{cases}$$

The exact solution is made of both rarefaction and shock waves. With a final time $t = 0.3$, the waves do not interact. The numerical simulations are presented in Figure 5.2. The Table 5.2 gives the evaluated orders of accuracy.

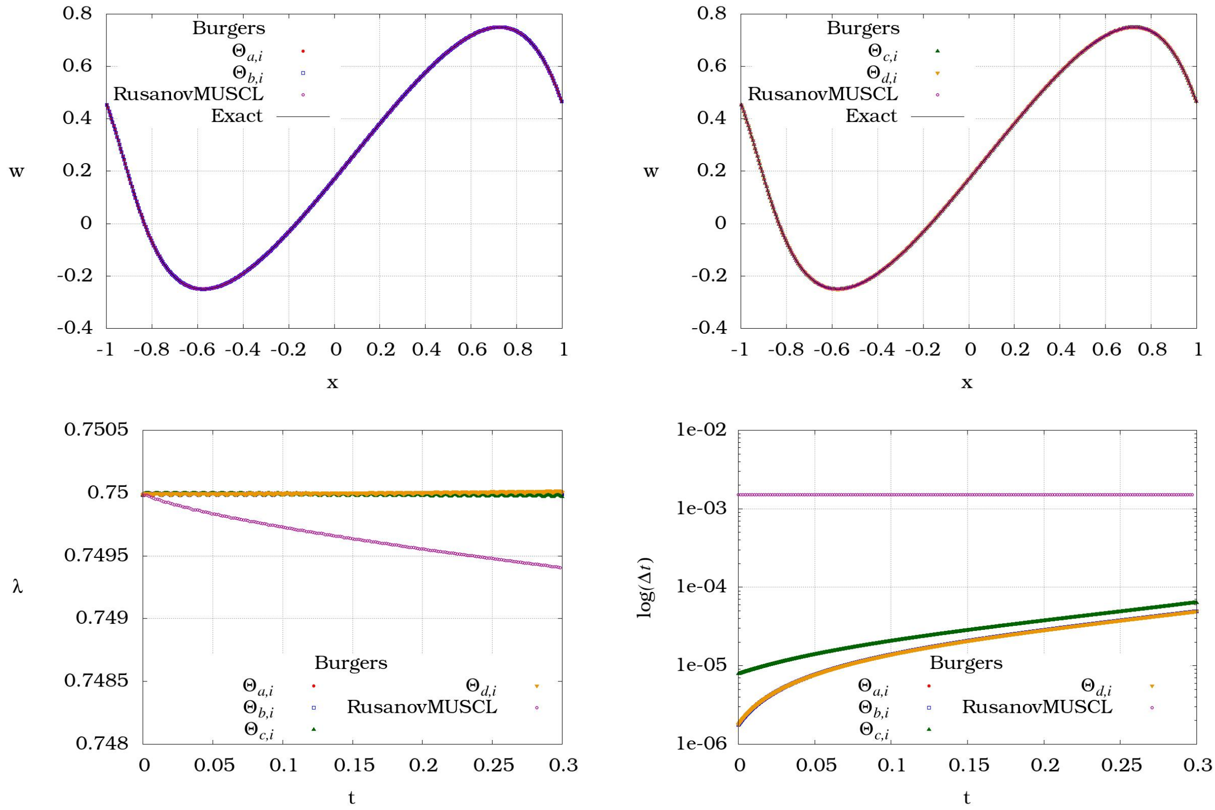

FIGURE 5.1. On the top, second-order accurate approximations of the smooth Burgers solutions at time $t = 0.3$ with a mesh made of 400 cells and with the Θ_i^n matrices given by (5.1). On the bottom, comparison between $(\lambda, \Delta t)$ given by our results (4.6)–(4.8) and the standard parameters used for the Rusanov MUSCL scheme endowed with the minmod slope limiter.

With the specific choice $\Theta_{c,i}$, we notice a nice behavior of the approximate solution. It seems that the presence of a discontinuity improves the CFL condition (4.8). We also compute τ being the ratio between the CPU time needed to run the simulation with our scheme (2.15) and the CPU time required to run the standard MUSCL Rusanov scheme. Then, we have

$$\tau_{\Theta_{i,a}^n} = 79.87, \quad \tau_{\Theta_{i,b}^n} = 77.24, \quad \tau_{\Theta_{i,c}^n} = 36.01, \quad \tau_{\Theta_{i,d}^n} = 69.33.$$

In the case of a discontinuous solution, our scheme has a large computation time. But, as in Section 5.1.1, a direct measure shows that for all the choices of the matrices Θ_i^n given in (5.1), around 97% of the CPU time is spent on the computation of the formula (4.7)–(4.8). Once again, if we use a second-order SSP time discretization then we have the following ratios:

$$\tau_{\Theta_{i,a}^n} = 0.89, \quad \tau_{\Theta_{i,b}^n} = 0.93, \quad \tau_{\Theta_{i,c}^n} = 0.65, \quad \tau_{\Theta_{i,d}^n} = 0.73.$$

5.2. **Euler system**

We consider the Euler system where $\Omega = \{(\rho, \rho u, \rho E) \in \mathbb{R}^3 \; : \; \rho > 0, \rho E - \rho u^2/2 > 0\}$, for a perfect diatomic gas where the unknown vector is $w = (\rho, \rho u, \rho E)^T$ and the flux function is $f(w) = (\rho u, \rho u^2 + p, \rho E u + pu)^T$, with $p = (\gamma - 1)(\rho E - \frac{\rho u^2}{2})$. We fix $\gamma = 1.4$ and we consider the entropy

TABLE 5.1. Errors and order evaluations for the second-order accurate scheme with the smooth Burgers solution for Θ_i^n defined in (5.1).

	Second-order scheme errors $\Theta_i^n = \Theta_{a,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	5.7E-04	-	6.8E-04	-	1.8E-03	-
200	1.4E-04	2.0	1.7E-04	2.0	4.5E-04	2.0
400	3.5E-05	2.0	4.1E-05	2.0	1.1E-04	2.0
800	8.8E-06	2.0	1.0E-05	2.0	2.8E-05	2.0
1600	2.2E-06	2.0	2.6E-06	2.0	6.9E-06	2.0
	Second-order scheme errors $\Theta_i^n = \Theta_{b,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	5.7E-04	-	6.8E-04	-	1.8E-03	-
200	1.4E-04	2.0	1.7E-04	2.0	4.5E-04	2.0
400	3.5E-05	2.0	4.1E-05	2.0	1.1E-04	2.0
800	8.8E-06	2.0	1.0E-05	2.0	2.8E-05	2.0
1600	2.2E-06	2.0	2.6E-06	2.0	6.9E-06	2.0
	Second-order scheme errors $\Theta_i^n = \Theta_{c,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	1.4E-03	-	1.6E-03	-	4.1E-03	-
200	2.4E-04	2.5	2.8E-04	2.6	7.6E-04	2.4
400	3.9E-05	2.7	4.3E-05	2.7	1.1E-04	2.8
800	8.7E-06	2.1	1.0E-05	2.1	2.7E-05	2.0
1600	2.2E-06	2.0	2.6E-06	2.0	6.8E-06	2.0
	Second-order scheme errors $\Theta_i^n = \Theta_{d,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	4.4E-04	-	4.8E-04	-	1.1E-03	-
200	1.1E-04	2.0	1.1E-04	2.1	2.7E-04	2.1
400	2.6E-05	2.0	2.8E-05	2.0	6.6E-05	2.0
800	6.5E-06	2.0	6.8E-06	2.0	1.6E-05	2.0
1600	1.6E-06	2.0	1.7E-06	2.0	4.0E-06	2.0

$\eta(w) = -\rho \ln\left(\frac{p}{\rho^\gamma}\right)$. For the Euler problem, we perform four numerical simulations: one with a smooth solution to measure the accuracy of the scheme, two with a shock tube solution and one with the solution close to the vacuum. We use the following matrix parameter

$$\begin{aligned}
\Theta_{a,i}^n &= -\theta \operatorname{diag}_{1\le j\le d}\left(\operatorname{sign}\left(\left(\nabla\eta(w_{i+1}^n) - \nabla\eta(w_{i-1}^n)\right)_j (\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j\right)\right),\\
\Theta_{b,i}^n &= -\theta \operatorname{diag}_{1\le j\le d}\left(\tanh\left(\left(\nabla\eta(w_{i+1}^n) - \nabla\eta(w_{i-1}^n)\right)_j (\delta_{i+\frac{1}{2}}^n - \delta_{i-\frac{1}{2}}^n)_j\right)\right),\\
\Theta_{c,\varepsilon,i}^n &= \operatorname{diag}_{1\le j\le d}\left(\frac{\left((\delta_{i-\frac{1}{2}}^n)_j^2 - (\delta_{i+\frac{1}{2}}^n)_j^2\right)\left((\delta_{i-\frac{1}{2}}^n)_j^2 + (\delta_{i+\frac{1}{2}}^n)_j^2\right)}{\left((\delta_{i-\frac{1}{2}}^n)_j^2 + (\delta_{i+\frac{1}{2}}^n)_j^2\right)^2 + \varepsilon}\right),
\end{aligned} \tag{5.2}$$

where θ is defined in (4.11) and ε is equal to 10^{-12}. To perform all the following numerical experiments we increase the time accuracy of the scheme (2.15) using the SSP Rung–Kutta methods introduced

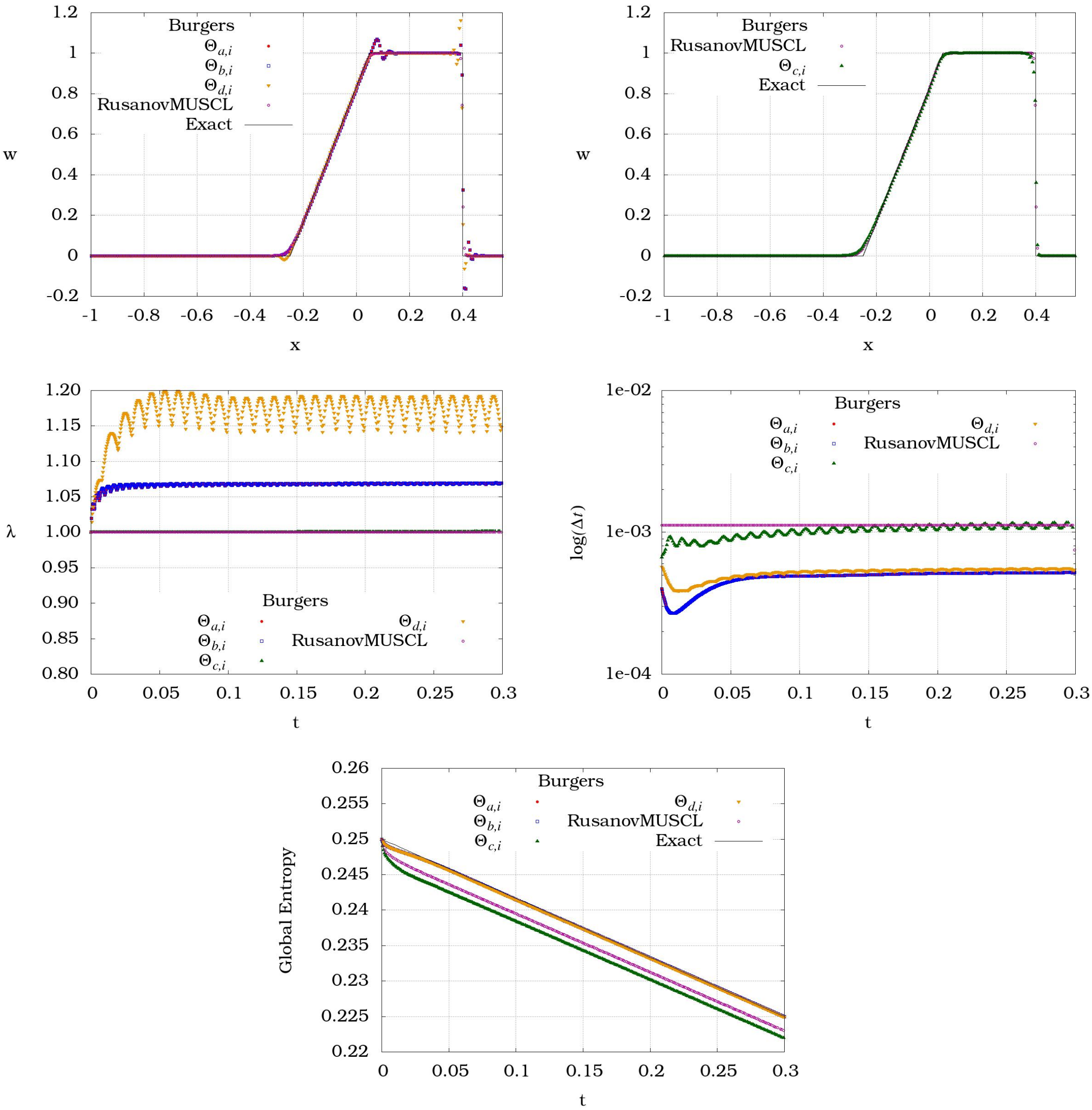

FIGURE 5.2. On the top, second-order accurate approximations of the Burgers solution made of rarefaction and shock waves at time $t = 0.3$ with a mesh made of 400 cells with periodic boundary conditions. In the middle, compararison between $(\lambda, \Delta t)$ given by our results (4.6)–(4.8) and the standard parameters used for the Rusanov MUSCL scheme endowed with the minmod slope limiter. On the bottom, the decrease of the global entropy.

in [18, 19, 20]. The numerical viscosity λ verifies $\lambda = \lambda^{\text{HLL}}$ where λ^{HLL} is defined in (2.3). The time step Δt satisfies the usual CFL condition $\frac{\lambda \Delta t}{\Delta x} \leq \frac{1}{2}$ and we verify numerically at each time step the decrease of the global entropy given by (4.9). As a global entropy reference, we use a numerical solution computed with the standard first-order Rusanov scheme (2.2) on a fine grid having 50 000

TABLE 5.2. Errors and order evaluations for the second-order accurate schemes with the Burgers solution made of rarefaction and shock waves, for Θ_i^m defined in (5.1).

	Second-order scheme errors $\Theta_i^n = \Theta_{a,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	3.4E-02	-	6.4E-02	-	3.3E-01	-
200	1.7E-02	1.0	4.3E-02	0.6	3.3E-01	0.0
400	8.5E-03	1.0	3.0E-02	0.5	3.2E-01	0.0
800	4.3E-03	1.0	2.1E-02	0.5	3.2E-01	0.0
1600	2.1E-03	1.0	1.5E-02	0.5	3.2E-01	0.0
	Second-order scheme errors $\Theta_i^n = \Theta_{b,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	5.4E-02	-	9.0E-02	-	3.8E-01	-
200	2.7E-02	1.0	5.9E-02	0.6	3.8E-01	0.0
400	1.4E-02	1.0	4.0E-02	0.6	3.8E-01	0.0
800	7.0E-03	1.0	2.7E-02	0.6	3.8E-01	0.0
1600	3.5E-03	1.0	1.8E-02	0.5	3.8E-01	0.0
	Second-order scheme errors $\Theta_i^n = \Theta_{c,\varepsilon,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	3.5E-02	-	7.1E-02	-	3.6E-01	-
200	1.8E-02	0.9	4.9E-02	0.5	3.6E-01	0.0
400	9.2E-03	1.0	3.4E-02	0.5	3.6E-01	0.0
800	4.6E-03	1.0	2.3E-02	0.5	3.6E-01	0.0
1600	2.3E-03	1.0	1.6E-02	0.5	3.6E-01	0.0
	Second-order scheme errors $\Theta_i^n = \Theta_{d,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	3.1E-02	-	5.8E-02	-	2.8E-01	-
200	1.4E-02	1.1	4.0E-02	0.6	2.8E-01	0.0
400	7.1E-03	1.0	2.8E-02	0.5	2.8E-01	0.0
800	3.5E-03	1.0	1.9E-02	0.5	2.8E-01	0.0
1600	1.7E-03	1.0	1.4E-02	0.5	2.8E-01	0.0

cells. We graphically compare our results to the standard Rusanov MUSCL scheme [45] endowed with the minmod slope limiter and extended with the same second-order in time discretization.

5.2.1. *Smooth solution*

The initial data is given as follows over the periodic domain $[-1, 1)$:

$$\rho_0(x) = 1 + 0.5\sin^2(\pi x), \quad u_0(x) = 0.5, \quad p_0(x) = 1.$$

For a such initial data the Euler equations reduce to a linear transport problem and the solution remains smooth for all time $t > 0$.

The numerical solutions are displayed in Figure 5.3. The second-order accuracy is observed in Table 5.3 whereas the Rusanov MUSCL scheme with the minmod slope limiter provides the following orders of accuracy: 1.88 for the L^1-norm, 1.63 for the L^2-norm and 1.27 for the L^∞-norm.

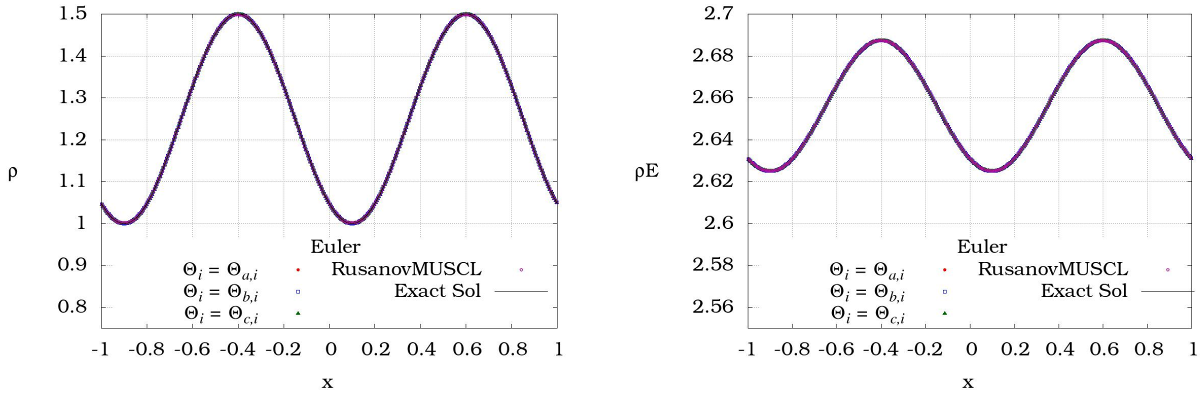

FIGURE 5.3. Second-order accurate approximation of the smooth Euler solution and entropy at time $t = 0.2$ with a mesh made of 400 cells.

TABLE 5.3. Errors and order evaluations for the second-order accurate schemes with the continuous Euler solution and for Θ_i^n described in (5.2).

	Second-order scheme errors $\Theta_i^n = \Theta_{a,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	2.1E-03	-	1.7E-03	-	1.7E-03	-
200	5.3E-04	2.0	4.2E-04	2.0	4.2E-04	2.0
400	1.3E-04	2.0	1.1E-04	2.0	1.1E-04	2.0
800	3.3E-05	2.0	2.6E-05	2.0	2.6E-05	2.0
1600	8.4E-06	2.0	6.6E-06	2.0	6.6E-06	2.0
	Second-order scheme errors $\Theta_i^n = \Theta_{b,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	2.1E-03	-	1.7E-03	-	1.7E-03	-
200	5.3E-04	2.0	4.2E-04	2.0	4.2E-04	2.0
400	1.3E-04	2.0	1.1E-04	2.0	1.1E-04	2.0
800	3.3E-05	2.0	2.6E-05	2.0	2.6E-05	2.0
1600	8.4E-06	2.0	6.6E-06	2.0	6.6E-06	2.0
	Second-order scheme errors $\Theta_i^n = \Theta_{c,\varepsilon,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	7.3E-03	-	6.8E-03	-	1.1E-02	-
200	1.5E-03	2.3	1.5E-03	2.2	2.9E-03	1.9
400	2.1E-04	2.8	1.8E-04	3.0	3.7E-04	3.0
800	3.7E-05	2.5	2.8E-05	2.7	2.6E-05	3.8
1600	8.4E-06	2.1	6.6E-06	2.1	6.5E-06	2.0

5.2.2. *Shock tube solution*

We consider the initial data given by

$$\rho_0(x) = \begin{cases} 1 & \text{if } x < 0.5, \\ 0.125 & \text{otherwise,} \end{cases} \qquad u_0(x) = 0, \qquad p_0(x) = \begin{cases} 1 & \text{if } x < 0.5, \\ 0.1 & \text{otherwise,} \end{cases}$$

Table 5.4. Errors and order evaluations for the second-order accurate schemes with the Sod shock tube Euler solution and for Θ_i^n described in (5.2).

	Second-order scheme errors $\Theta_i^n = \Theta_{a,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	1.9E-02	-	2.8E-02	-	9.5E-02	-
200	1.1E-02	0.8	1.9E-02	0.6	8.6E-02	0.1
400	6.3E-03	0.8	1.3E-02	0.5	9.2E-02	0.1
800	3.5E-03	0.8	9.1E-03	0.5	9.7E-02	0.1
1600	2.0E-03	0.8	7.0E-03	0.4	1.0E-01	0.1
	Second-order scheme errors $\Theta_i^n = \Theta_{b,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	1.9E-02	-	2.8E-02	-	9.6E-02	-
200	1.1E-02	0.8	1.9E-02	0.6	8.7E-02	0.1
400	6.3E-03	0.8	1.3E-02	0.5	9.2E-02	0.1
800	3.5E-03	0.8	9.2E-03	0.5	9.7E-02	0.1
1600	2.0E-03	0.8	7.0E-03	0.4	1.0E-01	0.1
	Second-order scheme errors $\Theta_i^n = \Theta_{c,\varepsilon,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	1.5E-02	-	2.2E-02	-	7.9E-02	-
200	8.2E-03	0.8	1.5E-02	0.6	6.7E-02	0.2
400	4.3E-03	0.9	9.9E-03	0.6	6.9E-02	0.0
800	2.3E-03	0.9	6.9E-03	0.5	7.6E-02	0.1
1600	1.3E-03	0.9	5.4E-03	0.4	8.4E-02	0.1

over the domain $[0,1]$. The final time is 0.2. To respect the periodic conditions on the boundaries, we work on the domain $[-1,1]$ and we symmetrize the shock tube problem on $[-1,0]$. The numerical solutions are displayed Figure 5.4.

The second discontinuous test case deals with the Lax shock tube problem [28] for which the initial condition is

$$\rho_0(x) = \begin{cases} 0.445 & \text{if } x < 0.5, \\ 0.5 & \text{otherwise}, \end{cases} \qquad u_0(x) = \begin{cases} 0.698 & \text{if } x < 0.5, \\ 0 & \text{otherwise}, \end{cases} \qquad p_0(x) = \begin{cases} 3.528 & \text{if } x < 0.5, \\ 0.5710 & \text{otherwise}, \end{cases}$$

over the domain $[0,1]$. As previously, in order to respect the periodic conditions on the boundaries, we work on the domain $[-1,1]$ and we symmetrize the shock tube problem on $[-1,0]$. The final time is 0.1. The Figure 5.5 shows our results and the Table 5.5 assesses the convergence of the solution.

Our matrix $\Theta_{i,b}^n$ given by the second equation of (5.2) does not ensure the decrease of the global entropy for this test case. The matrix $\Theta_{c,\varepsilon,i}^n$ yields very satisfying results.

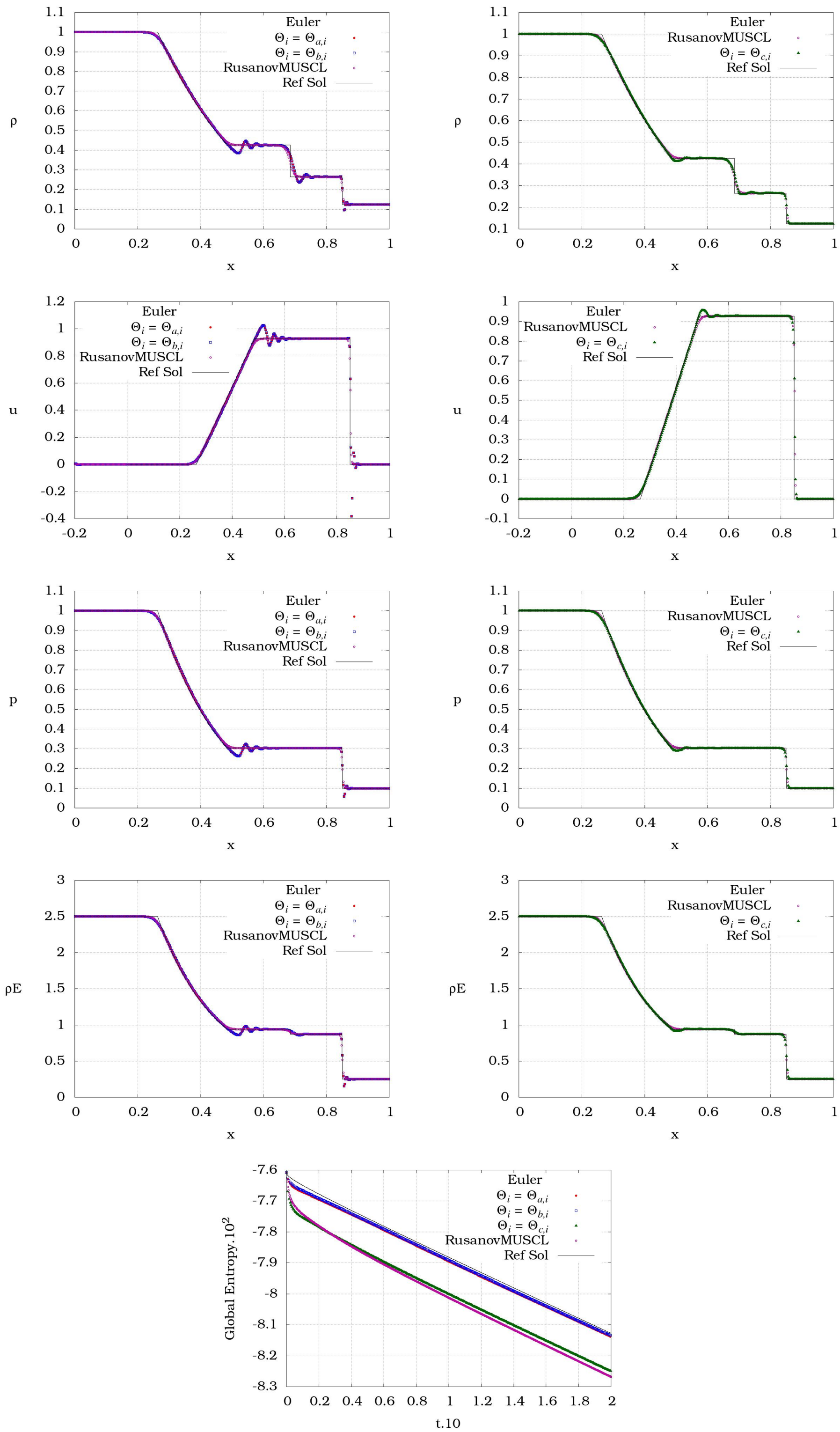

FIGURE 5.4. On the top, second-order accurate approximation of the Euler solution for the Sod shock tube problem at time $t = 0.2$ with a mesh made of 400 cells. On the bottom, the decrease of the global entropy integrated over the grid.

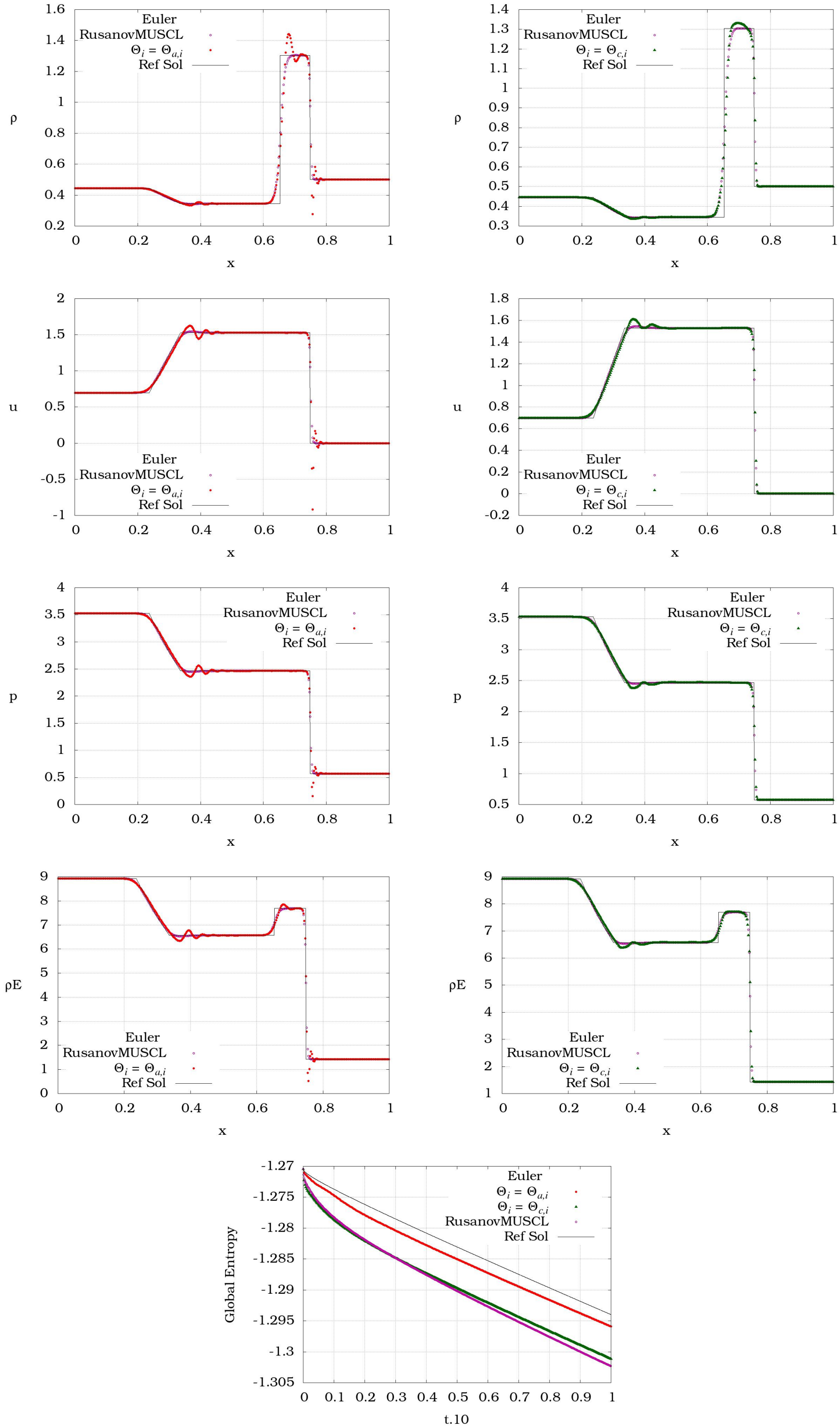

FIGURE 5.5. Second-order accurate approximation of the Euler solution for the Lax shock tube problem at time $t = 0.1$ with a mesh made of 400 cells: on the top wave profiles, on the bottom the decrease of the global entropy.

Table 5.5. Errors and order evaluations for the second-order accurate schemes with the Lax shock tube Euler solution and for $\Theta_i^n \in \{\Theta_{a,i}^n, \Theta_{c,\varepsilon,i}^n\}$ described in (5.2).

	Second-order scheme errors $\Theta_i^n = \Theta_{a,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	4.0E-02	-	1.0E-01	-	5.7E-01	-
200	2.9E-02	0.5	9.2E-02	0.1	6.1E-01	0.1
400	1.7E-02	0.7	6.6E-02	0.5	5.9E-01	0.1
800	1.1E-02	0.7	5.3E-02	0.3	6.0E-01	0.0
1600	6.8E-03	0.7	4.4E-02	0.3	6.2E-01	0.0
	Second-order scheme errors $\Theta_i^n = \Theta_{c,\varepsilon,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	4.9E-02	-	1.3E-01	-	6.3E-01	-
200	2.7E-02	0.9	9.7E-02	0.4	6.6E-01	0.1
400	1.5E-02	0.8	6.6E-02	0.5	5.9E-01	0.2
800	8.9E-03	0.8	5.1E-02	0.4	6.0E-01	0.0
1600	5.2E-03	0.8	4.0E-02	0.3	6.2E-01	0.0

5.2.3. *Solution near the vacuum*

The test case of this section concerns the solutions of the Euler equations near the vacuum. We consider the following initial condition

$$\rho_0(x) = 1, \qquad u_0(x) = \begin{cases} -2.7 & \text{if } x < 0.5, \\ 2.7 & \text{otherwise,} \end{cases} \qquad p_0(x) = 1,$$

over the domain $[0, 1]$. The final time is 0.05. Under these conditions, the solution is made of two rarefaction waves. As previously, to respect the periodic conditions on the boundaries, we work on the domain $[-1, 1]$ and we symmetrize the problem on $[-1, 0]$. For this test case, we slightly increase the numerical viscosity such that $\lambda = 1.05\lambda^{\text{HLL}}$ with λ^{HLL} is given by (2.3) and the time step Δt is selected according to $\frac{\lambda \Delta t}{\Delta x} \leq \frac{1}{2}$ with a CFL number equals to 0.1. The Figure 5.6 shows our results on several meshes and the Table 5.6 evaluates the convergence of the numerical solution toward the exact solution.

Table 5.6. Errors and order evaluations for the second-order accurate schemes with the double rarefaction wave solution and for the sequence $(\Theta_{c,\varepsilon,i}^n)_{i\in\mathbb{Z}}$ given in (5.2).

	Second-order scheme errors $\Theta_i^n = \Theta_{c,\varepsilon,i}^n$					
cells	L^1	order	L^2	order	L^∞	order
100	3.1E-02	-	5.4E-02	-	2.0E-01	-
200	2.5E-02	0.3	4.7E-02	0.2	2.1E-01	0.0
400	1.4E-02	0.8	2.9E-02	0.7	1.4E-01	0.6
800	7.9E-03	0.9	1.6E-02	0.9	7.3E-02	0.9
1600	4.1E-03	1.0	8.2E-03	0.9	4.3E-02	0.8

For this test case, our choices $(\Theta_{a,i}^n)_{i\in\mathbb{Z}}$, $(\Theta_{b,i}^n)_{i\in\mathbb{Z}}$ generate spurious oscillations which lead to negative pressures. Therefore these matrices do not work for this numerical test. The results for the matrices $(\Theta_{c,\varepsilon,i}^n)_{i\in\mathbb{Z}}$ are robust. At the final time, the pressure stays positive (min $\approx 10^{-3}$) without spurious oscillations.

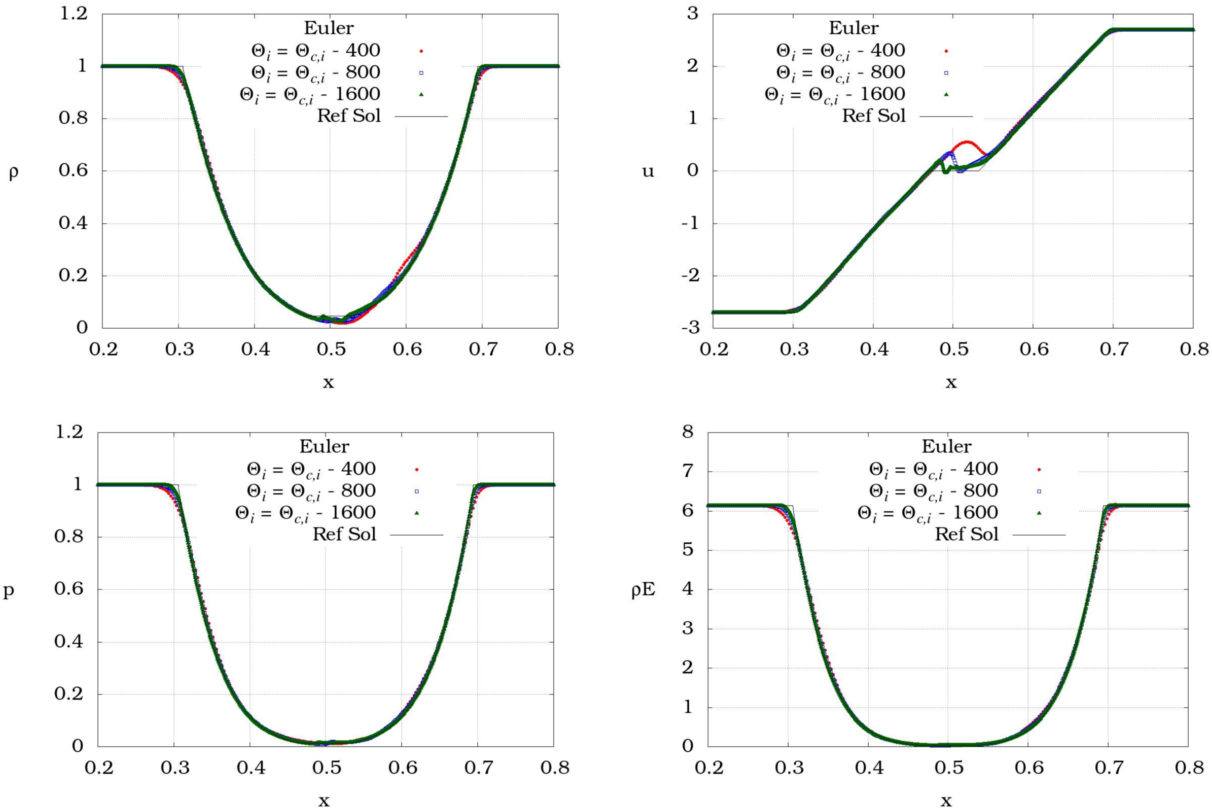

FIGURE 5.6. Second-order accurate approximation of the Euler solution for the double rarefaction waves problem at time $t = 0.05$ with a refined mesh made of 400 cells to 1600 cells.

5.3. **Non-convex scalar flux**

In this section, we study the case of the non-convex scalar flux f. We reproduce the numerical test case given in [32, Section 3.1]. In this regard, we consider $f(w) = w^3$ endowed with the quadratic entropy $\eta(w) = w^2/2$. The domain is $(-1, 1)$ and we consider the initial condition $w_0(x) = -\sin(\pi x)$. We prescribe periodic boundary conditions on both sides of the domain. We consider the matrices $(\Theta_i^n)_{i\in\mathbb{Z}}$ given in (5.1). The Figure 5.7 shows the results at the final time 1.

For this test case, the choice of $(\Theta_{c,\varepsilon,i}^n)_{i\in\mathbb{Z}}$ yields a classical solution. But we observe, depending on the choice of the matricies $(\Theta_i^n)_{i\in\mathbb{Z}}$, that the numerical solution given by the genuinely second-order scheme (2.15) may converge to a non classical solution. This behavior has already been observed in the recent paper by LeFLoch *et al.* [32] where the authors shown that a genuinely high-order scheme may arbitrary converge to a non classical solution in the case of non convex flux.

6. **Conclusion**

We have presented a family of unlimited and formally second-order accurate in space finite volume schemes. They are designed in such a way that a dissipative property is required for the second-order discretization in space. We have given a proof of a global entropy inequality under a CFL which is not standard and probably far from being optimal. The proof requires the strict convexity of the entropy. At a numerical level, we assessed the behavior of the schemes both for smooth and discontinuous

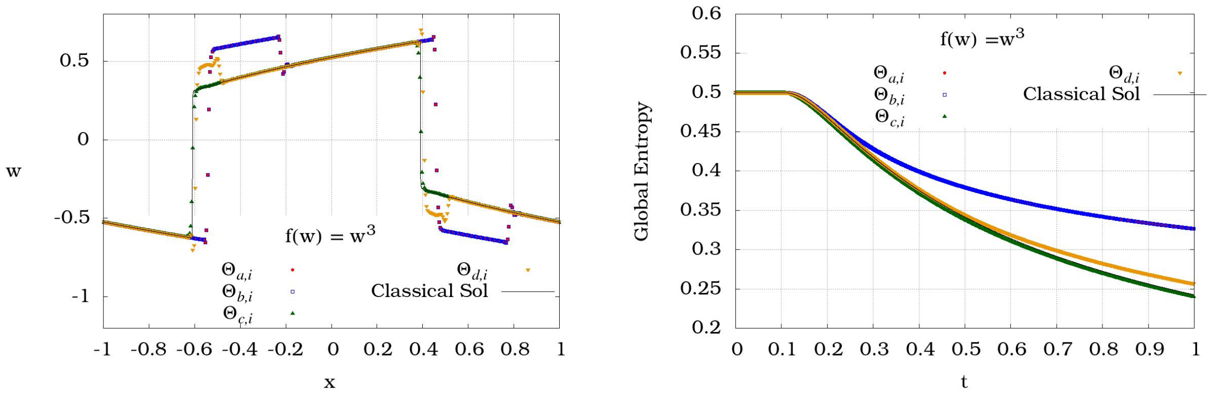

FIGURE 5.7. Second-order accurate approximation of the scalar conservation law with non convex flux ($f(w) = w^3$) solution at time $t = 1$ with a mesh made of 400 cells: on the left non classical wave profiles, on the right the decrease of the global entropy integrated over the grid.

solutions in the case of convex fluxes. In the case of non convex fluxes, the numerical solution may converge to a non-classical solution. In both context, we observed that there is one choice of the second-order discretization in space which yields good stability results without losing the order of accuracy. The thorough analysis of this particular second-order discretization in space will be the purpose of a future work.

Acknowledgements

The authors thank J. P. Vila for stimulating discussions and the ANR project MUFFIN ANR-19-CE46-0004 for the financial support. The authors thank the anonymous referee that suggests the study of non-convex fluxes.

References

[1] Christophe Berthon. Stability of the MUSCL schemes for the Euler equations. *Comm. Math. Sci*, 3(2):133–158, 2005.

[2] Christophe Berthon. Numerical approximations of the 10-moment Gaussian closure. *Math. Comput.*, 75(256):1809–1832, 2006.

[3] François Bouchut. Construction of BGK models with a family of kinetic entropies for a given system of conservation laws. *J. Stat. Phys.*, 95(1-2):113–170, 1999.

[4] François Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Frontiers in Mathematics. Birkhäuser, 2004.

[5] François Bouchut, Christian Bourdarias, and Benoit Perthame. A MUSCL method satisfying all the numerical entropy inequalities. *Math. Comput.*, 65(216):1439–1462, 1996.

[6] Manuel J. Castro, Ulrik S. Fjordholm, Siddhartha Mishra, and Carlos Parés. Entropy conservative and entropy stable schemes for nonconservative hyperbolic systems. *SIAM J. Numer. Anal.*, 51(3):1371–1391, 2013.

[7] Christophe Chalons and Philippe G. LeFloch. A fully discrete scheme for diffusive-dispersive conservation laws. *Numer. Math.*, 89(3):493–509, 2001.

[8] Bernardo Cockburn, Frédéric Coquel, and Philippe G. LeFloch. Convergence of the finite volume method for multidimensional conservation laws. *SIAM J. Numer. Anal.*, 32(3):687–705, 1995.

[9] Frédéric Coquel, Jean-Marc Hérard, and Khaled Saleh. A splitting method for the isentropic baer-nunziato two-phase flow model. *ESAIM, Proc.*, 38:241–256, 2012.

[10] Frédéric Coquel and Benoit Perthame. Relaxation of Energy and Approximate Riemann Solvers for General Pressure Laws in Fluid Dynamics. *SIAM J. Numer. Anal.*, 35(6):2223–2249, 1998.

[11] Arnaud Duran and Fabien Marche. Recent advances on the discontinuous Galerkin method for shallow water equations with topography source terms. *Comput. Fluids*, 101:88–104, 2014.

[12] Arnaud Duran, Jean-Paul Vila, and Rémy Baraille. Semi-implicit staggered mesh scheme for the multi-layer shallow water system. *C. R. Acad. Sci. Paris*, 355(12):1298–1306, 2017.

[13] Robert Eymard, Thierry Gallouët, and Raphaèle Herbin. *Finite Volume Methods*, volume 7 of *Handbook of Numerical Analysis*. North-Holland, 2000.

[14] Thierry Gallouët, Raphaèle Herbin, Jean-Claude Latché, and Nicolas Therme. Consistent Internal Energy Based Schemes for the Compressible Euler Equations. In *Numerical Simulation in Physics and Engineering: Trends and Applications: Lecture Notes of the XVIII 'Jacques-Louis Lions' Spanish-French School*, volume 24 of *SEMA SIMAI Springer Series*, pages 119–154. Springer, 2021.

[15] Laura Gastaldo, Raphaèle Herbin, Jean-Claude Latché, and Nicolas Therme. A MUSCL-type segregated–explicit staggered scheme for the Euler equations. *Comput. Fluids*, 175:91–110, 2018.

[16] Edwige Godlewski and Pierre-Arnaud Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118 of *Applied Mathematical Sciences*. Springer, 1996.

[17] Sergeĭ K. Godunov. A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb., N. Ser.*, 47(89):271–306, 1959.

[18] Sigal Gottlieb. On high order strong stability preserving Runge-Kutta and multi step time discretizations. *J. Sci. Comput.*, 25(1):105–128, 2005.

[19] Sigal Gottlieb and Chi-Wang Shu. Total variation diminishing Runge–Kutta schemes. *Math. Comput.*, 67(221):73–85, 1998.

[20] Sigal Gottlieb, Chi-Wang Shu, and Eitan Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43(1):89–112, 2001.

[21] Amiram Harten and Peter D. Lax. A random choice finite difference scheme for hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 18(2):289–315, 1981.

[22] Amiram Harten, Peter D. Lax, and Bram Van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.*, 25:35–61, 1983.

[23] Raphaèle Herbin, Jean-Claude Latché, and Trung T. Nguyen. Consistent explicit staggered schemes for compressible flows Part I: the barotropic Euler equations. 2013.

[24] Andreas Hiltebrand and Siddhartha Mishra. Entropy stable shock capturing space–time discontinuous Galerkin schemes for systems of conservation laws. *Numer. Math.*, 126(1):103–151, 2014.

[25] Andreas Hiltebrand, Siddhartha Mishra, and Carlos Parés. Entropy-stable space–time DG schemes for non-conservative hyperbolic systems. *ESAIM, Math. Model. Numer. Anal.*, 52(3):995–1022, 2018.

[26] Farzad Ismail and Philip L. Roe. Affordable, entropy-consistent Euler flux functions II: Entropy production at shocks. *J. Comput. Phys.*, 228(15):5410–5436, 2009.

[27] Brahim Khobalatte and Benoit Perthame. Maximum principle on the entropy and second-order kinetic schemes. *Math. Comput.*, 62(205):119–131, 1994.

[28] Peter D. Lax. Weak solutions of nonlinear hyperbolic equations and their numerical computation. *Commun. Pure Appl. Math.*, 7:159–193, 1954.

[29] Peter D. Lax. Shock waves and entropy. In *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, pages 603–634. Academic Press Inc., 1971.

[30] Peter D. Lax. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*, volume 11 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1973.

[31] Philippe G. LeFloch. *Hyperbolic systems of conservation laws. The theory of classical and nonclassical shock waves*. Lectures in Mathematics. Birkhäuser, 2002.

[32] Philippe G. LeFloch and Hendrik Ranocha. Kinetic Functions for Nonclassical Shocks, Entropy Stability, and Discrete Summation by Parts. *J. Sci. Comput.*, 87(2), 2021.

[33] Sebastian Noelle, Yulong Xing, and Chi-Wang Shu. High-order well-balanced finite volume WENO schemes for shallow water equation with moving water. *J. Comput. Phys.*, 226(1):29–58, 2007.

[34] Benoit Perthame and Youchum Qiu. A variant of Van Leer's method for multidimensional systems of conservation laws. *J. Comput. Phys.*, 112(2):370–381, 1994.

[35] Benoit Perthame and Chi-Wang Shu. On positivity preserving finite volume schemes for Euler equations. *Numer. Math.*, 73(1):119–130, 1996.

[36] Philip L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.*, 43(2):357–372, 1981.

[37] Viktor V. Rusanov. The calculation of the interaction of non-stationary shock waves and obstacles. *USSR Comput. Math. Math. Phys.*, 1(2):304–320, 1962.

[38] Denis Serre. *Systems of conservation laws. I: Hyperbolicity, entropies, shock waves.* Cambridge University Press, 1999. translated from the 1996 French original by I. N. Sneddon.

[39] Jing Shi, Yong-Tao Zhang, and Chi-Wang Shu. Resolution of high order WENO schemes for complicated flow structures. *J. Comput. Phys.*, 186(2):690–696, 2003.

[40] Chi-Wang Shu. High-order finite difference and finite volume WENO schemes and discontinuous Galerkin methods for CFD. *International Journal of Computational Fluid Dynamics*, 17(2):107–118, 2003.

[41] Eitan Tadmor. The numerical viscosity of entropy stable schemes for systems of conservation laws. I. *Math. Comput.*, 49(179):91–103, 1987.

[42] Eitan Tadmor. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numer.*, 12(1):451–512, 2003.

[43] Eleuterio F. Toro. *Riemann solvers and numerical methods for fluid dynamics. A practical introduction.* Springer, 3rd edition, 2009.

[44] Eleuterio F. Toro, M. Spruce, and W. Speares. Restoration of the contact surface in the HLL-Riemann solver. *Shock Waves*, 4(1):25–34, 1994.

[45] Bram van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. *J. Comput. Phys.*, 32(1):101–136, 1979.

[46] Hamed Zakerzadeh and Ulrik S. Fjordholm. High-order accurate, fully discrete entropy stable schemes for scalar conservation laws. *IMA J. Numer. Anal.*, 36(2):633–654, 2015.