

## NUMERICAL RESOLUTION OF SCALAR CONVEX EQUATIONS: EXPLICIT STABILITY, ENTROPY AND CONVERGENCE CONDITIONS

FREDERIC LAGOUTIERE<sup>1</sup>

**Abstract.** In this paper we extend to scalar convex equations a previous work done for linear advection numerical resolution (cf. [3], [11]). We do a study of finite volume schemes, giving some explicit conditions on the fluxes for such a scheme to be convergent. These conditions are of three types: consistency, stability, entropy. Then we propose one particular scheme, taking the most **downwind** possible fluxes under the convergence conditions, applying the idea of [3] to derive anti-dissipative schemes. The proposed scheme is, as expected, low dissipative. We show some numerical results.

**AMS Subject Classification.** 65M06, 65N06, 82C40 .

### 1. INTRODUCTION

The present work deals with the numerical resolution of one-dimensional convex scalar equations. This is a very “popular” topic, because studying these equations is a way to understand well non-linear properties, non-linear effects, being able to have complete proofs of convergence for schemes rather simply in general (at least in comparison with non-linear systems).

For the theoretical analysis of these equations (existence, uniqueness of solutions, entropy criterion...), we refer to [10], [12], and for the numerical point of view, some important contributions are [7] (monotonicity properties), [8] (TVD properties and incremental analysis), [18] (flux limiters), [9] (high order schemes and the number of extrema), [16] and [17] (ENO schemes), [1] and [2] (several space dimensions), [15] (convergence of MUSCL scheme) and [5] (numerical entropy conditions).

Here we are interested in non-dissipative or anti-dissipative schemes, and particularly in sharp shock profiles. In this article we will follow the ideas developed in our recent works on linear advection and Euler (multicomponent) system (cf. [3], [11], [4]). We developed a new formalism to ensure  $L^\infty$ -stability and Total Variation Diminishing (TVD) for a finite-volume scheme. This formalism is equivalent to the slope-limiters and flux-limiters formalisms (see [8], [18] and [19] for example). The final result was the study of a particular scheme (equivalent to the Ultrabee limiter in [19]). We proved for this scheme a **uniform-in-time error estimate** for a set of initial conditions that is dense in  $L^1$ . We conjectured the same result for any initial condition (see [3], [11]). This scheme was understood as the **downwind choice of the fluxes** under some stability conditions, instead of the classical upwind choice. We will not repeat here how to obtain this non-dissipative scheme but the general ideas will appear in the following (in fact all can be retrieved replacing the flux function with the identity function in the general scalar equation (1) below).

During the CEMRACS 1999, the goal was to derive a similar analysis for **non-linear** scalar conservation laws, typically Burgers equation. The following is devoted to the analysis of convergence of finite-volume schemes

---

*Keywords and phrases:* Scalar Equations, Convex Equations, Numerical Schemes, Entropy Conditions, Convergence

<sup>1</sup> Centre d'études de Bruyères-le-Châtel, CEA/DIF/DCSA/SNEC/LPNN, BP 12, 91680 BRUYÈRES-LE-CHÂTEL, FRANCE

to entropy solutions of scalar conservation laws. More precisely, we are interested in some **explicit entropy convergence conditions**.

The problem we want to solve is the partial differential equation

$$\partial_t u + \partial_x f(u) = 0 \quad \forall x \in \mathbb{R}, \quad \forall t \in \mathbb{R}^+ \quad (1)$$

associated with the initial condition

$$u(0, x) = u^0(x) \quad \forall x \in \mathbb{R}. \quad (2)$$

The function  $f$  is the flux function and is assumed to be in  $C^2(\mathbb{R})$  and strictly convex. The initial condition  $u^0$  is taken in  $BV(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$  for theoretical reasons (see [10]). In all the following we do the assumption that we are away from sonic points, that is that  $f'(u) > 0$  in all the domain, or  $f'(u) < 0$ .

Now let  $S(u)$  be a strictly convex function in  $C^1(\mathbb{R})$  and  $G(u)$  a function in  $C^1(\mathbb{R})$  such that

$$S' f' = G'.$$

The function  $S$  is called an entropy function and  $G$  is the associated entropy flux. Any  $u$  solution of (1), (2) is said to be an entropy solution if and only if it verifies (in the sense of distributions)

$$\partial_t(S(u)) + \partial_x(G(u)) \leq 0. \quad (3)$$

Results about existence and uniqueness of entropy solutions can be found in [10] or [13].

We now discretize the problem (1), (2) in a finite-volume formalism. We define a regular mesh of  $\mathbb{R}$  with cell size  $\Delta x$  and a time increment  $\Delta t$ . The numerical initial condition is given by

$$u_j^0 = \int_{x_{j-1/2}}^{x_{j+1/2}} \frac{u^0(x)}{\Delta x} dx \quad \text{with } x_{j-1/2} = (j-1/2)\Delta x. \quad (4)$$

The partial differential equation (1) is approximated by

$$u_j^{n+1} = u_j^n - \lambda(f_{j+1/2}^n - f_{j-1/2}^n) \quad (5)$$

where  $\lambda = \Delta t/\Delta x$ ,  $u_j^n$  is the approximate value of  $u(n\Delta t, j\Delta x)$  and the values  $(f_{j+1/2}^n)_{j \in \mathbb{Z}}$  are the fluxes at the cell edges at time  $n\Delta t$ . The fluxes are to be determined (the choice of the scheme relies on the choice of these fluxes). In order to simplify notations, we now examine the scheme over one time step only, so that we can rewrite it

$$\hat{u}_j = u_j - \lambda(f_{j+1/2} - f_{j-1/2}). \quad (6)$$

In the second section, we give some general definitions.

In the third section, we try to transpose directly the analysis done for the linear case. This idea appears to be too naïve: the scheme does not select the entropy solution.

In section 4, we do the necessary entropy analysis in order to obtain explicit inequalities for the fluxes to give an entropic and convergent scheme. The final result is given by theorem 4.4.

Section 5 is devoted to numerical examples for Burgers equation (the chosen scheme is the constrained downwind scheme).

We then conclude.

## 2. BASIC DEFINITIONS AND RESULTS

**Definition 2.1.** The scheme (6) is said to be explicit if and only if there exists an integer  $k \geq 1$  and a function  $g$  of  $2k$  variables such that

$$f_{j+1/2} = g(u_{j-k+1}, u_{j-k+2}, \dots, u_j, \dots, u_{j+k}) \quad \forall j \in \mathbb{Z}. \quad (7)$$

In the following, we will only consider explicit schemes.

**Definition 2.2.** An explicit scheme (6) is said to be consistent if and only if the function  $g$  of definition 2.1 verifies

$$g(u, u, \dots, u) = f(u). \quad (8)$$

**Definition 2.3.** The scheme (5) is said to be  $L^\infty$ -stable if and only if there exists  $C \in \mathbb{R}$  (independent of  $n$  and  $\Delta t$ ) such that

$$\max_{j \in \mathbb{Z}} |u_j^n| \leq C \quad \forall n \in \mathbb{N}. \quad (9)$$

**Definition 2.4.** The total variation at the time step  $n$  of the discrete solution  $u_j^n$  is

$$TV^n = \sum_{j \in \mathbb{Z}} |u_{j+1}^n - u_j^n|. \quad (10)$$

The scheme (6) is said to be Total Variation Diminishing (TVD) if and only if for any given  $(u_j)_{j \in \mathbb{Z}}$ ,  $\sum_{j \in \mathbb{Z}} |\widehat{u_{j+1}} - \widehat{u_j}| \leq \sum_{j \in \mathbb{Z}} |u_{j+1} - u_j|$ , that is  $\widehat{TV} \leq TV$ .

Let us denote

$$\overset{\Delta t, \Delta x}{u}(t, x) = \sum_{n \in \mathbb{N}} \sum_{j \in \mathbb{Z}} u_j^n 1_{[x_{j-1/2}, x_{j+1/2}]}(x) 1_{[n\Delta t, (n+1)\Delta t]}(t). \quad (11)$$

**Definition 2.5.** The scheme (4), (6) with  $\lambda$  fixed is said to be convergent (in  $L^1(\mathbb{R})$ ) toward  $u(t, x)$  if and only if there exists a sequence  $(\Delta x_m)_{m \in \mathbb{N}}$  such that

- $\lim_{m \rightarrow \infty} \Delta x_m = 0$  ;
- $\lim_{m \rightarrow \infty} \|\overset{\Delta t_m, \Delta x_m}{u}(t, \cdot) - u(t, \cdot)\|_{L^1(\mathbb{R})} = 0 \quad \forall t \in \mathbb{R}^+$ ,  $\Delta t_m$  being defined by  $\Delta t_m = \lambda \Delta x_m$ .

We have now the following convergence theorem (cf. [6]).

**Theorem 2.6.** Assume  $u^0 \in BV(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ .

Consider the explicit scheme (4), (6).

Assume that this scheme is consistent,  $L^\infty$ -stable and TVD, and that its fluxes  $f_{j+1/2}$  are locally Lipschitz-continuous.

Then it is convergent toward a weak solution of (1), (2).

Note that it is not necessarily convergent toward the entropy solution. We will have to put a numerical entropy condition in order to get such a convergence.

## 3. THE NAÏVE IDEA

Here we propose to follow similar steps as those we have developed in the linear case (see [3], [11]). Knowing that the exact solution obeys a local maximum principle with decreasing in time total variation, we ask the

numerical solution to verify these properties. Doing that, we know that any resulting scheme will converge toward a weak solution (if we choose some Lipschitz continuous fluxes).

Assume that  $f'(u_j) > 0 \quad \forall j \in \mathbb{Z}$  (we are away from sonic points). We now focus on the cell labelled  $j$  and write

$$\begin{cases} m_j = \min(u_{j-1}, u_j), \\ M_j = \max(u_{j-1}, u_j), \\ \phi_j = \min(f(u_{j-1}), f(u_j)), \\ \Phi_j = \max(f(u_{j-1}), f(u_j)), \\ b_j = \frac{u_j - M_j}{\lambda} + \Phi_j, \\ B_j = \frac{u_j - m_j}{\lambda} + \phi_j. \end{cases}$$

The following proposition gives a sufficient condition on  $f_{j+1/2}$  for the scheme (6) to be  $L^\infty$ -stable and TVD.

**Proposition 3.1.** *Consider the scheme (6) and assume it verifies  $\forall j$ ,*

$$\begin{aligned} b_j &\leq f_{j+1/2} \leq B_j, \\ \phi_{j+1} &\leq f_{j+1/2} \leq \Phi_{j+1}. \end{aligned} \tag{12}$$

*Then, it is consistent,  $L^\infty$ -stable and TVD.*

The second double inequality is the consistency inequality.

*Proof.* By assumption, the inequalities are true for any  $j$ , so we know that

$$\phi_j \leq f_{j-1/2} \leq \Phi_j.$$

Reporting this in the first double inequality of the proposition, we get

$$\frac{u_j - M_j}{\lambda} + f_{j-1/2} \leq f_{j+1/2} \leq \frac{u_j - m_j}{\lambda} + f_{j-1/2},$$

equivalent to

$$u_j - M_j \leq \lambda(f_{j+1/2} - f_{j-1/2}) \leq u_j - m_j,$$

and to

$$m_j \leq u_j - \lambda(f_{j+1/2} - f_{j-1/2}) \leq M_j.$$

Finally, using the equation of the scheme (6), we see that

$$m_j \leq \hat{u}_j \leq M_j. \tag{13}$$

So we have  $\hat{u}_j \in [u_{j-1}, u_j]$ ; so there exists  $\kappa_{j-1/2} \in [0, 1]$  such that  $\hat{u}_j = \kappa_{j-1/2}u_j + (1 - \kappa_{j-1/2})u_{j-1} = u_j - (1 - \kappa_{j-1/2})(u_j - u_{j-1})$ . We thus can write  $\hat{u}_j = u_j - D_{j-1/2}(u_j - u_{j-1})$  with  $D_{j-1/2} \in [0, 1]$ . This leads to  $L^\infty$ -stability and TVD property (result due to Harten, cf. [8]).  $\square$

The analysis is not finished since we don't know if the inequalities (12) can be enforced (are they compatible?). The following proposition gives a positive answer to this question, under the classical CFL condition (Courant-Friedrichs-Lewy).

**Proposition 3.2.** *Assume the CFL condition  $\lambda f'(u_j) \leq 1$  is verified.*

*Then,*

$$f(u_j) \in [\phi_{j+1}, \Phi_{j+1}] \cap [b_j, B_j] \text{ which is so not empty.} \tag{14}$$

*Furthermore, if the flux  $f_{j+1/2}$  is chosen in this interval, the scheme (6) verifies inequalities (13).*

If these assumptions are verified for each  $j$  and if the fluxes are Lipschitz continuous, the scheme converges toward a weak solution.

*Proof.* The proof of this result relies exactly on the same ideas as for the linear case. The linear case can be retrieved taking  $f(u) = u$ .

Of course  $f(u_j) \in [\phi_{j+1}, \Phi_{j+1}]$ , so we only have to prove that  $f(u_j) \in [b_j, B_j]$ . We prove that  $B_j \geq f(u_j)$  and the inequality  $b_j \leq f(u_j)$  can be shown using the same steps. There are two possibilities:  $u_j \leq u_{j-1}$  and  $u_j > u_{j-1}$ .

- If  $u_j \leq u_{j-1}$ , we have  $m_j = u_j$ . So  $B_j = \phi_j$ . We assumed  $f'(u_j) > 0$ , and  $f$  is strictly convex, so  $f$  is increasing in the interval  $[u_j, u_{j-1}]$ . Finally we get  $\phi_j = f(u_j)$ , and the inequality  $B_j \geq f(u_j)$  is true since it means in this case  $f(u_j) \geq f(u_j)$ .
- If  $u_j > u_{j-1}$ , we have  $m_j = u_{j-1}$ . The inequality  $B_j \geq f(u_j)$  is equivalent to

$$\lambda \frac{f(u_j) - \phi_j}{u_j - u_{j-1}} \leq 1.$$

There are two different cases. Either  $\phi_j = f(u_j)$  and the above inequality is trivial ( $0 \leq 1$ ), or  $\phi_j = f(u_{j-1})$  and we have to prove that

$$\lambda \frac{f(u_j) - f(u_{j-1})}{u_j - u_{j-1}} \leq 1.$$

The convexity assumption leads to conclude, since

$$\lambda \frac{f(u_j) - f(u_{j-1})}{u_j - u_{j-1}} \leq \lambda f'(u_j) \leq 1$$

thanks to the CFL condition.

So the interval  $[\phi_{j+1}, \Phi_{j+1}] \cap [b_j, B_j]$  contains  $f(u_j)$  (the upwind value). It is not empty.  $\square$

The idea now is to define the fluxes as the most downwind possible in the stability interval  $[\phi_{j+1}, \Phi_{j+1}] \cap [b_j, B_j]$ , as we did in the linear case in [11]. Let  $\omega_{j+1/2}$  (resp.  $\Omega_{j+1/2}$ ) denote  $\max(b_j, \phi_{j+1})$  (resp.  $\min(B_j, \Phi_{j+1})$ ). The scheme is then defined by

$$f_{j+1/2} = \begin{cases} \omega_{j+1/2} & \text{if } f(u_{j+1}) \leq \omega_{j+1/2}, \\ f(u_{j+1}) & \text{if } \omega_{j+1/2} < f(u_{j+1}) \leq \Omega_{j+1/2}, \\ \Omega_{j+1/2} & \text{if } \Omega_{j+1/2} < f(u_{j+1}). \end{cases}$$

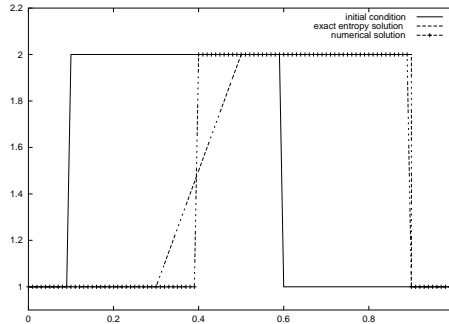
We can observe the numerical result for Burgers equation ( $f(u) = u^2/2$ ) on the interval  $[0, 1]$  with periodic boundary conditions (figure 1).

We see that the numerical solution seems to converge toward a weak solution (accordingly to the theoretical results above), but this weak solution is not the ‘‘physical’’ solution, the entropy solution: a shock has been computed instead of a decompression wave in the increasing domain.

$L^\infty$ -stability and diminishing of the total variation are not sufficient conditions for the convergence toward the right entropy solution. It is necessary to add a new criterion to select the good numerical solution. This will be done in the next section.

#### 4. DISCRETE ENTROPY PROPERTIES

To ensure that the scheme (4), (6) converges toward the entropy solution of (2), (1), we have the theorem 4.2 below (that can be found in [6]). This theorem needs the definition of discrete entropy flux and discrete entropy inequality.

FIGURE 1. Initial condition, numerical and exact solution for  $t = 0.2$ .

**Definition 4.1.** The scheme (6) is said to be consistent with the entropy inequality (3) if and only if there exists a continuous function  $H$  of  $2k$  variables such that

- $H$  is consistent with  $G$ :

$$H(u, u, \dots, u) = G(u); \quad (15)$$

- $H$  verifies the discrete entropy inequality

$$\frac{S(\widehat{u}_j) - S(u_j)}{\Delta t} \leq - \frac{H(u_{j-k+1}, \dots, u_{j+k}) - H(u_{j-k}, \dots, u_{j+k-1})}{\Delta x}. \quad (16)$$

$H$  is the numerical (or discrete) entropy flux.

**Theorem 4.2.** Assume  $u^0 \in BV(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ .

Consider the scheme (4), (6) (assumed to be explicit).

Assume that this scheme is consistent,  $L^\infty$ -stable, TVD, and that its fluxes are locally Lipschitz-continuous.

Assume moreover that this scheme is consistent with the entropy inequality (3) for one couple entropy-entropy flux.

Then, it is convergent toward the entropy solution of (1), (2).

Under the above assumptions, the fact that only one discrete entropy inequality is needed is due to the strict convexity hypothesis on  $f$  (see [6]).

Instead of inequality (16), we will further use a more explicit one (inequality (19)). For this one we introduce the following quantities.

$$\begin{cases} \widehat{u_{j+1/2,L}} = u_j - 2\lambda(f_{j+1/2} - f(u_j)) \\ \widehat{u_{j+1/2,R}} = u_{j+1} - 2\lambda(f(u_{j+1}) - f_{j+1/2}). \end{cases} \quad (17)$$

These quantities can be viewed as updated values in semi-cells Left and Right, and verify

$$\widehat{u}_j = \frac{\widehat{u_{j+1/2,L}} + \widehat{u_{j-1/2,R}}}{2}. \quad (18)$$

These quantities are extensively used in [1], [2]..., and allow to write very simple discrete entropy inequalities that involve only the exact (and not numerical) entropy flux function.

**Lemma 4.3.** Consider the notations (17). Assume that the scheme (6) is explicit, consistent, and that the fluxes  $f_{j+1/2}$  are continuous.

Assume that it verifies

$$\frac{S(\widehat{u_{j+1/2,L}}) + S(\widehat{u_{j+1/2,R}})}{2} - \frac{S(u_j) + S(u_{j+1})}{2} \leq -\lambda(G(u_{j+1}) - G(u_j)). \quad (19)$$

Then it is consistent with the entropy inequality (3) (cf. definition 4.1).

*Proof.* The proof consists in finding a numerical entropy flux that is continuous, consistent with  $G$  and verifying the discrete entropy inequality (16).

The inequality (19) is equivalent to

$$\frac{S(\widehat{u_{j+1/2,L}}) + S(\widehat{u_{j-1/2,R}})}{2} - S(u_j) + \frac{S(u_j) - S(\widehat{u_{j-1/2,R}}) - S(u_{j+1}) + S(\widehat{u_{j+1/2,R}})}{2} \leq -\lambda(G(u_{j+1}) - G(u_j)),$$

that is

$$\begin{aligned} \frac{S(\widehat{u_{j+1/2,L}}) + S(\widehat{u_{j-1/2,R}})}{2} - S(u_j) \\ \leq -\lambda(G(u_{j+1}) - \frac{S(u_{j+1}) - S(\widehat{u_{j+1/2,R}})}{2\lambda} - (G(u_j) - \frac{S(u_j) - S(\widehat{u_{j-1/2,R}})}{2\lambda})). \end{aligned}$$

Since  $S$  is a convex function, we have

$$\frac{S(\widehat{u_{j+1/2,L}}) + S(\widehat{u_{j-1/2,R}})}{2} \geq S(\frac{\widehat{u_{j+1/2,L}} + \widehat{u_{j-1/2,R}}}{2}) = S(\widehat{u}_j).$$

For every  $j \in \mathbb{Z}$  we define

$$G_{j+1/2} = G(u_{j+1}) - \frac{S(u_{j+1}) - S(\widehat{u_{j+1/2,R}})}{2\lambda}. \quad (20)$$

One has

$$S(\widehat{u}_j) - S(u_j) \leq -\lambda(G_{j+1/2} - G_{j-1/2}),$$

that is

$$\frac{S(\widehat{u}_j) - S(u_j)}{\Delta t} \leq -\frac{G_{j+1/2} - G_{j-1/2}}{\Delta x}.$$

Now assume the fluxes  $f_{j+1/2}$  are explicit, continuous and consistent. Then there exists an integer  $k$  and a function  $H$  (given by equation (20)) such that

$$G_{j+1/2} = H(u_{j-k+1}, \dots, u_{j+k})$$

and  $H$  is continuous and consistent (equation (15)). This ends the proof.  $\square$

The interesting point is that this lemma provides a **shifted** entropy condition using only the fluxes  $f_{j+1/2}$  besides the exact entropy pair  $(S, G)$ . We will be able to transpose this local condition to the fluxes  $f_{j+1/2}$  directly, and so to obtain an explicit entropy condition (on the fluxes).

Here comes the main result of this paper. We assume as in the previous section that  $f'(u_j) > 0 \quad \forall j \in \mathbb{Z}$ . Let us recall some definitions already introduced,

$$\begin{cases} m_j = \min(u_{j-1}, u_j), \\ M_j = \max(u_{j-1}, u_j), \\ \phi_j = \min(f(u_{j-1}), f(u_j)), \\ \Phi_j = \max(f(u_{j-1}), f(u_j)), \\ b_j = \frac{u_j - M_j}{\lambda} + \Phi_j, \\ B_j = \frac{u_j - m_j}{\lambda} + \phi_j, \end{cases} \quad (21)$$

and define some new quantities:

$$\begin{cases} s_{m_{j+1/2}} = \min_{[m_{j+1}, M_{j+1}]} (S''(u)), \\ v_{m_{j+1/2}} = \min_{[m_{j+1}, M_{j+1}]} (f'(u)), \\ s_{M_{j+1/2}} = \max_{[m_{j+1}, M_{j+1}]} (S''(u)), \\ v_{M_{j+1/2}} = \max_{[m_{j+1}, M_{j+1}]} (f'(u)), \end{cases} \quad (22)$$

and

$$\begin{cases} a_{j+1/2} = 2\lambda s_{M_{j+1/2}}, \\ c_{j+1/2} = S'(u_{j+1}) - S'(u_j) - 2\lambda s_{M_{j+1/2}}(f(u_{j+1}) + f(u_j)), \\ d_{j+1/2} = G(u_{j+1}) - f(u_{j+1})S'(u_{j+1}) - (G(u_j) - f(u_j)S'(u_j)) \\ \quad + \lambda s_{M_{j+1/2}}(f(u_{j+1})^2 + f(u_j)^2), \\ r_{j+1/2} = \frac{-c_{j+1/2} - \sqrt{c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}}}{2a_{j+1/2}}, \\ R_{j+1/2} = \frac{-c_{j+1/2} + \sqrt{c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}}}{2a_{j+1/2}}. \end{cases} \quad (23)$$

**Theorem 4.4.** *Assume  $u^0 \in BV(\mathbb{R}) \cap L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ .*

*Consider the scheme (4), (6) (assumed to be explicit).*

*Assume the fluxes  $f_{j+1/2}$  are locally Lipschitz-continuous.*

*Assume  $f'(u_j) > 0 \quad \forall j \in \mathbb{Z}$ .*

*Assume the CFL condition*

$$\lambda \leq \min_{j \in \mathbb{Z}} \left( \frac{1}{v_{M_{j+1/2}}} \left( \min \left( \frac{2}{\sqrt{5}} \frac{s_{m_{j+1/2}}^2}{s_{M_{j+1/2}}^2}, \frac{1}{2} \frac{v_{m_{j+1/2}} s_{m_{j+1/2}}}{v_{M_{j+1/2}} s_{M_{j+1/2}}} \right) \right) \right) \quad (24)$$

*is verified (considering notations (21), (22), (23)).*

*Assume*

$$f_{j+1/2} \in [\phi_{j+1}, \Phi_{j+1}] \cap [b_j, B_j] \cap [r_{j+1/2}, R_{j+1/2}] \quad \forall j \in \mathbb{Z}$$

*( $r_{j+1/2}$  and  $R_{j+1/2}$  are reals, the intersection of the intervals is not empty and contains  $f(u_j)$ ).*

*Then, the scheme is convergent toward the weak entropy solution of (1), (2).*

This convergence result is the consequence of the two following lemmas.



**Lemma 4.5.** *Considering the notations (21), (22) and (23), assume  $f'(u_j) > 0$  and  $f'(u_{j+1}) > 0$ . Assume the CFL condition*

$$\lambda \leq \frac{2}{\sqrt{5}v_{M_{j+1/2}}} \frac{s_{m_{j+1/2}}^2}{s_{M_{j+1/2}}^2}$$

is verified.

Then,

$$c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} \geq 0.$$

A consequence of that is that  $r_{j+1/2}$  and  $R_{j+1/2}$  are real numbers.

*Proof.* We are looking for a condition on  $\lambda$  such that the discriminant  $c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}$  is positive. To do that we first simplify the expression of this discriminant. We write for each  $j \in \mathbb{Z}$

$$\begin{cases} f_j = f(u_j), \\ S'_j = s'(u_j), \\ G_j = g(u_j). \end{cases}$$

Then the discriminant can be written as

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &= (S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} + f_j))^2 \\ &\quad - 8\lambda s_{M_{j+1/2}}(G_{j+1} - G_j - (f_{j+1}S'_{j+1} - f_jS'_j) \\ &\quad \quad + \lambda s_{M_{j+1/2}}(f_{j+1}^2 + f_j^2)). \end{aligned}$$

We first develop it.

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &= (S'_{j+1} - S'_j)^2 + 4\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} + f_j)^2 \\ &\quad - 4\lambda s_{M_{j+1/2}}(S'_{j+1} - S'_j)(f_{j+1} + f_j) \\ &\quad - 8\lambda s_{M_{j+1/2}}(G_{j+1} - G_j - (f_{j+1}S'_{j+1} - f_jS'_j)) \\ &\quad - 8\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1}^2 + f_j^2). \end{aligned}$$

It is convenient to recast the above identity as follows.

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &= (S'_{j+1} - S'_j)^2 + 4\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} + f_j)^2 \\ &\quad + 4\lambda s_{M_{j+1/2}}(S'_{j+1} - S'_j)(f_{j+1} - f_j) \\ &\quad - 8\lambda s_{M_{j+1/2}}(G_{j+1} - G_j - (f_{j+1} - f_j)S'_j) \\ &\quad - 8\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1}^2 + f_j^2) \\ &= (S'_{j+1} - S'_j)^2 + 4\lambda s_{M_{j+1/2}}(S'_{j+1} - S'_j)(f_{j+1} - f_j) \\ &\quad - 4\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} + f_j)^2 + 16\lambda^2 s_{M_{j+1/2}}^2 f_{j+1}f_j \\ &\quad - 8\lambda s_{M_{j+1/2}}(G_{j+1} - G_j - (f_{j+1} - f_j)S'_j), \end{aligned}$$

and

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &= (S'_{j+1} - S'_j)^2 - 4\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} - f_j)^2 \\ &\quad + 4\lambda s_{M_{j+1/2}}(S'_{j+1} - S'_j)(f_{j+1} - f_j) \\ &\quad - 8\lambda s_{M_{j+1/2}}(G_{j+1} - G_j - (f_{j+1} - f_j)S'_j). \end{aligned} \tag{25}$$

So we have the following equivalent form

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &= \left( \int_{u_j}^{u_{j+1}} S''(x) dx \right)^2 - 4\lambda^2 s_{M_{j+1/2}}^2 \left( \int_{u_j}^{u_{j+1}} f'(x) dx \right)^2 \\ &\quad + 4\lambda s_{M_{j+1/2}} \int_{u_j}^{u_{j+1}} S''(x) dx \int_{u_j}^{u_{j+1}} f'(x) dx \\ &\quad - 8\lambda s_{M_{j+1/2}} \int_{u_j}^{u_{j+1}} (G'(x) - f'(x)S'_j) dx. \end{aligned}$$

By definition of the couple entropy-entropy flux, we have  $G'(x) = f'(x)S'(x)$ , so  $G'(x) - f'(x)S'_j = f'(x)(S'(x) - S'_j)$  and we get

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &= \left( \int_{u_j}^{u_{j+1}} S''(x) dx \right)^2 - 4\lambda^2 s_{M_{j+1/2}}^2 \left( \int_{u_j}^{u_{j+1}} f'(x) dx \right)^2 \\ &\quad + 4\lambda s_{M_{j+1/2}} \int_{u_j}^{u_{j+1}} S''(x) dx \int_{u_j}^{u_{j+1}} f'(x) dx \\ &\quad - 8\lambda s_{M_{j+1/2}} \int_{u_j}^{u_{j+1}} f'(x) \int_{u_j}^x S''(y) dy dx. \end{aligned}$$

Now recall that  $f'(u_j) > 0$  and that  $f'(u_{j+1}) > 0$  by assumption. The function  $f$  being convex,  $f'(x)$  stays between  $f'_j$  and  $f'_{j+1}$  for each  $x$  located between  $u_j$  and  $u_{j+1}$ . Consequently,  $f'(x) > 0$  in each integral above. And  $S''(x) > 0$  too, because  $S$  is convex too. So with the definition of  $s_{M_{j+1/2}}$  given in (23),

$$\int_{u_j}^{u_{j+1}} f'(x) \int_{u_j}^x S''(y) dy dx < \int_{u_j}^{u_{j+1}} f'(x) \left( \int_{u_j}^x s_{M_{j+1/2}} dy \right) dx,$$

in both cases  $u_{j+1/2} > u_j$  or  $u_{j+1/2} \leq u_j$ . That is why

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &> \left( \int_{u_j}^{u_{j+1}} S''(x) dx \right)^2 - 4\lambda^2 s_{M_{j+1/2}}^2 \left( \int_{u_j}^{u_{j+1}} f'(x) dx \right)^2 \\ &\quad - 4\lambda s_{M_{j+1/2}} \int_{u_j}^{u_{j+1}} f'(x) \left( \int_{u_j}^x s_{M_{j+1/2}} dy \right) dx. \end{aligned}$$

Replacing the terms  $f'$  and  $S''$  with their minimal or maximal values in the interval  $[m_{j+1}, M_{j+1}]$  we obtain a new lower bound of the discriminant:

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &> (u_{j+1} - u_j)^2 (s_{m_{j+1/2}}^2 - 4\lambda^2 s_{M_{j+1/2}}^2 v_{M_{j+1/2}}^2) \\ &\quad - 4\lambda s_{M_{j+1/2}} \frac{(u_{j+1} - u_j)^2}{2} s_{M_{j+1/2}} v_{M_{j+1/2}}. \end{aligned}$$

Here we have proved that

$$\begin{aligned} c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} &> (u_{j+1} - u_j)^2 (s_{m_{j+1/2}}^2 - 4\lambda^2 s_{M_{j+1/2}}^2 v_{M_{j+1/2}}^2 \\ &\quad - 2\lambda s_{M_{j+1/2}}^2 v_{M_{j+1/2}}). \end{aligned}$$

It suffices now to find a condition under which the right term of the above inequality is positive in order to have a sufficient condition for the discriminant to be positive. Let us find a condition forcing

$$s_{m_{j+1/2}}^2 - 4\lambda^2 s_{M_{j+1/2}}^2 v_{M_{j+1/2}}^2 - 2\lambda s_{M_{j+1/2}}^2 v_{M_{j+1/2}} \geq 0. \quad (26)$$

The left term just above is a second order polynomial expression in  $\lambda$  whose second order coefficient is negative. This term is positive when and only when  $\lambda$  is located between the two roots of the polynomial. The discriminant of this polynomial is

$$\Delta = 4s_{M_{j+1/2}}^4 v_{M_{j+1/2}}^2 + 16s_{M_{j+1/2}}^2 v_{M_{j+1/2}}^2 s_{m_{j+1/2}}^2.$$

It is positive and its square root is greater than the absolute value of the first order coefficient of the polynomial. Thus the polynomial has one positive root and one negative root. The positive one is the only one we are interested in because we know that  $\lambda > 0$ . This root, denoted as  $\lambda_+$ , is

$$\lambda_+ = \frac{2s_{M_{j+1/2}}^2 v_{M_{j+1/2}} - 2\sqrt{s_{M_{j+1/2}}^4 v_{M_{j+1/2}}^2 + 4s_{M_{j+1/2}}^2 v_{M_{j+1/2}}^2 s_{m_{j+1/2}}^2}}{-8s_{M_{j+1/2}}^2 v_{M_{j+1/2}}^2},$$

or can be written

$$\lambda_+ = \frac{1}{4v_{M_{j+1/2}}} \left( -1 + \sqrt{1 + \frac{4s_{m_{j+1/2}}^2}{s_{M_{j+1/2}}^2}} \right).$$

We try to find a lower bound for  $\lambda_+$ . Let us write (temporarily)

$$x = \frac{4s_{m_{j+1/2}}^2}{s_{M_{j+1/2}}^2}.$$

This number is positive, and a classical argument ensures the existence of a number  $\tilde{x} \in [0, x]$  such that

$$\sqrt{1+x} = 1 + \frac{x}{2\sqrt{1+\tilde{x}}}.$$

Moreover, since  $x \leq 4$ , we have  $\tilde{x} \leq 4$  and

$$\sqrt{1+x} \geq 1 + \frac{x}{2\sqrt{5}}.$$

Replacing  $x$  with its expression, we find an upper bound of the positive root  $\lambda_+$ :

$$\lambda_+ \geq \frac{1}{v_{M_{j+1/2}}} \frac{2s_{m_{j+1/2}}^2}{\sqrt{5}s_{M_{j+1/2}}^2}.$$

Recall that if  $0 \leq \lambda \leq \lambda_+$ , the expression (26) is positive and  $c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}$  is positive too. So we can conclude, writing that a sufficient condition for the positivity of the discriminant is

$$\lambda \leq \frac{2}{\sqrt{5}v_{M_{j+1/2}}} \frac{s_{m_{j+1/2}}^2}{s_{M_{j+1/2}}^2}.$$

This is the CFL condition of the lemma. □

**Lemma 4.6.** *Consider the scheme (6), (4) and the notations (21), (22) and (23). Assume the CFL condition*

$$\lambda \leq \frac{2}{\sqrt{5}v_{M_{j+1/2}}} \frac{s_{m_{j+1/2}}^2}{s_{M_{j+1/2}}^2}$$

is verified.

Assume the other CFL condition

$$\lambda \leq \frac{1}{2v_{M_{j+1/2}}} \frac{v_{m_{j+1/2}} s_{m_{j+1/2}}}{v_{M_{j+1/2}} s_{M_{j+1/2}}}$$

is verified too.

Then,

$$f(u_j) \in [\phi_{j+1}, \Phi_{j+1}] \cap [b_j, B_j] \cap [r_{j+1/2}, R_{j+1/2}] \text{ which is not empty.}$$

Furthermore, if the flux  $f_{j+1/2}$  is chosen in this interval, the scheme verifies the two inequalities (13) and (19).

*Proof.* The CFL condition of the lemma 4.5 being verified,  $r_{j+1/2}$  and  $R_{j+1/2}$  are real numbers. Furthermore, the two CFL conditions are stricter than the CFL condition of proposition 3.2. As a consequence of this proposition, inequalities (13) are verified. We only need to verify that  $f(u_j) \in [r_{j+1/2}, R_{j+1/2}]$ . As for the proof of the proposition 3.2, we only show that  $R_{j+1/2} \geq f(u_j)$  (the other inequality can be shown equivalently). We are looking for a condition under which

$$\sqrt{c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}} \geq 2a_{j+1/2}f_j + c_{j+1/2},$$

that is

$$\sqrt{c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}} \geq 4\lambda s_{M_{j+1/2}}f_j + S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} + f_j),$$

or

$$\sqrt{c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2}} \geq S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} - f_j). \quad (27)$$

We first show that the right term sign is the same as the sign of  $(u_{j+1} - u_j)$  (under the CFL condition). We easily get

$$S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} - f_j) = \int_{u_j}^{u_{j+1}} S''(x) dx - 2\lambda s_{M_{j+1/2}} \int_{u_j}^{u_{j+1}} f'(x) dx.$$

So

- either  $(u_{j+1} - u_j) > 0$  and

$$S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} - f_j) > (u_{j+1} - u_j)(s_{m_{j+1/2}} - 2\lambda s_{M_{j+1/2}}v_{M_{j+1/2}});$$

- or  $(u_{j+1} - u_j) \leq 0$  and

$$S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} - f_j) \leq (u_{j+1} - u_j)(s_{m_{j+1/2}} - 2\lambda s_{M_{j+1/2}}v_{M_{j+1/2}}).$$

We have to study the sign of  $(s_{m_{j+1/2}} - 2\lambda s_{M_{j+1/2}}v_{M_{j+1/2}})$ . The CFL condition implies

$$\lambda \leq \frac{1}{2v_{M_{j+1/2}}} \frac{v_{m_{j+1/2}} s_{m_{j+1/2}}}{v_{M_{j+1/2}} s_{M_{j+1/2}}} \leq \frac{1}{2v_{M_{j+1/2}}} \frac{s_{m_{j+1/2}}}{s_{M_{j+1/2}}}.$$

Then

$$\lambda \leq \frac{1}{2v_{M_{j+1/2}}} \frac{s_{m_{j+1/2}}}{s_{M_{j+1/2}}},$$

that is equivalent to

$$(s_{m_{j+1/2}} - 2\lambda s_{M_{j+1/2}}v_{M_{j+1/2}}) \geq 0.$$

Consequently, the right term in (27) has the same sign as  $(u_{j+1} - u_j)$ . Then we remark that the inequality (27) is unconditionally (on  $\lambda$ ) verified if  $(u_{j+1} - u_j) \leq 0$ . We only have now to investigate the case  $(u_{j+1} - u_j) > 0$ . In this case, the right term in (27) is positive and the inequality is equivalent to

$$c_{j+1/2}^2 - 4a_{j+1/2}d_{j+1/2} \geq (S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} - f_j))^2.$$

We use the expression (25) of the left term, that we found during the proof of the lemma 4.5:

$$\begin{aligned} & (S'_{j+1} - S'_j)^2 - 4\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} - f_j)^2 \\ & + 4\lambda s_{M_{j+1/2}} (S'_{j+1} - S'_j) (f_{j+1} - f_j) \\ & - 8\lambda s_{M_{j+1/2}} (G_{j+1} - G_j - (f_{j+1} - f_j) S'_j) \geq (S'_{j+1} - S'_j)^2 \\ & \quad + 4\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} - f_j)^2 \\ & \quad - 4\lambda s_{M_{j+1/2}} (f_{j+1} - f_j) (S'_{j+1} - S'_j). \end{aligned}$$

This is equivalent to

$$\begin{aligned} & -8\lambda^2 s_{M_{j+1/2}}^2 (f_{j+1} - f_j)^2 \\ & + 8\lambda s_{M_{j+1/2}} (f_{j+1} - f_j) (S'_{j+1} - S'_j) \\ & - 8\lambda s_{M_{j+1/2}} (G_{j+1} - G_j - (f_{j+1} - f_j) S'_j) \geq 0, \end{aligned}$$

or

$$-\lambda s_{M_{j+1/2}} (f_{j+1} - f_j)^2 + (f_{j+1} - f_j) S'_{j+1} - (G_{j+1} - G_j) \geq 0$$

(we remind that we are looking for a condition for this inequality to be verified). As for the proof of the lemma 4.5, we get equivalent inequalities:

$$\begin{aligned} & -\lambda s_{M_{j+1/2}} \left( \int_{u_j}^{u_{j+1}} f'(x) dx \right)^2 \\ & + \int_{u_j}^{u_{j+1}} f'(x) S'_{j+1} dx - \int_{u_j}^{u_{j+1}} f'(x) S'(x) dx \geq 0, \\ & -\lambda s_{M_{j+1/2}} \left( \int_{u_j}^{u_{j+1}} f'(x) dx \right)^2 + \int_{u_j}^{u_{j+1}} f'(x) \left( \int_x^{u_{j+1}} S''(y) dy \right) dx \geq 0. \end{aligned}$$

It is sufficient to verify that

$$-(u_{j+1} - u_j)^2 \lambda s_{M_{j+1/2}} v_{M_{j+1/2}}^2 + \int_{u_j}^{u_{j+1}} f'(x) (u_{j+1} - x) s_{m_{j+1/2}} dx \geq 0$$

(because we only consider the case  $u_{j+1} > u_j$ ). It is sufficient that  $\lambda$  verifies

$$-(u_{j+1} - u_j)^2 \lambda s_{M_{j+1/2}} v_{M_{j+1/2}}^2 + \frac{(u_{j+1} - u_j)^2}{2} v_{m_{j+1/2}} s_{m_{j+1/2}} \geq 0,$$

or endly

$$\lambda s_{M_{j+1/2}} v_{M_{j+1/2}}^2 + \frac{1}{2} v_{m_{j+1/2}} s_{m_{j+1/2}} \geq 0.$$

It naturally leads to the CFL condition (presented in the lemma)

$$\lambda \leq \frac{1}{2v_{M_{j+1/2}}} \frac{v_{m_{j+1/2}} s_{m_{j+1/2}}}{v_{M_{j+1/2}} s_{M_{j+1/2}}}.$$

Now we have to prove that the scheme verifies the discrete entropy inequality (16) in the cell  $j$  for one function  $H$  consistent with  $G$ . We will show that the scheme verifies the (sufficient) inequality (19) of the lemma 4.3.

Let us consider the second order polynomial of  $x$   $P(x) = a_{j+1/2}x^2 + c_{j+1/2}x + d_{j+1/2}$ . We have already shown that under the CFL condition of this lemma, this polynomial has two real roots  $r_{j+1/2}$  and  $R_{j+1/2}$ , and that  $r_{j+1/2} \leq f(u_j) \leq R_{j+1/2}$ . The second order coefficient  $a_{j+1/2}$  is positive, so that  $P(x)$  is negative between  $r_{j+1/2}$  and  $R_{j+1/2}$ . Let  $f_{j+1/2} \in [r_{j+1/2}, R_{j+1/2}]$  be fixed. This value verifies  $P(f_{j+1/2}) \leq 0$ . We write

$$\begin{aligned} & f_{j+1/2}^2 2\lambda s_{M_{j+1/2}} \\ & + f_{j+1/2}(S'_{j+1} - S'_j - 2\lambda s_{M_{j+1/2}}(f_{j+1} + f_j)) \\ & + G_{j+1} - f_{j+1}S'_{j+1} - (G_j - f_j S'_j) \\ & + \lambda s_{M_{j+1/2}}(f_{j+1}^2 + f_j^2) \leq 0. \end{aligned}$$

Equivalently,

$$\begin{aligned} & -(f_{j+1/2} - f_j)S'_j - (f_{j+1} - f_{j+1/2})S'_{j+1} \\ & + \lambda(f_{j+1/2} - f_j)^2 s_{M_{j+1/2}} + \lambda(f_{j+1} - f_{j+1/2})^2 s_{M_{j+1/2}} \leq -(G_{j+1} - G_j). \end{aligned}$$

Multiplying all by  $\lambda > 0$ , we can rewrite:

$$\begin{aligned} & \frac{-2\lambda(f_{j+1/2} - f_j)S'_j + 2\lambda^2(f_{j+1/2} - f_j)^2 s_{M_{j+1/2}}}{2} \\ & + \frac{-2\lambda(f_{j+1} - f_{j+1/2})S'_{j+1} + 2\lambda^2(f_{j+1} - f_{j+1/2})^2 s_{M_{j+1/2}}}{2} \leq -\lambda(G_{j+1} - G_j). \end{aligned}$$

The convexity of  $S$  implies for any couple  $(a, b)$ :

$$S(b) \leq S(a) + (b - a)S'(a) + \frac{(b - a)^2}{2} \max_{[a;b]}(S''),$$

from where

$$\begin{cases} -2\lambda(f_{j+1/2} - f_j)S'_j + 2\lambda^2(f_{j+1/2} - f_j)^2 s_{M_{j+1/2}} \\ \geq S(u_j - 2\lambda(f_{j+1/2} - f_j)) - S(u_j) \\ -2\lambda(f_{j+1} - f_{j+1/2})S'_{j+1} + 2\lambda^2(f_{j+1} - f_{j+1/2})^2 s_{M_{j+1/2}} \\ \geq S(u_{j+1} - 2\lambda(f_{j+1} - f_{j+1/2})) - S(u_{j+1}). \end{cases}$$

So  $f_{j+1/2}$  verifies the inequality

$$\begin{aligned} & \frac{S(u_j - 2\lambda(f_{j+1/2} - f_j)) - S(u_j)}{2} + \frac{S(u_{j+1} - 2\lambda(f_{j+1} - f_{j+1/2})) - S(u_{j+1})}{2} \\ & \leq -\lambda(G_{j+1} - G_j). \end{aligned}$$

Replacing  $u_j - 2\lambda(f_{j+1/2} - f_j)$  by  $\widehat{u_{j+1/2,L}}$  and  $u_{j+1} - 2\lambda(f_{j+1} - f_{j+1/2})$  by  $\widehat{u_{j+1/2,R}}$  we obtain

$$\frac{S(\widehat{u_{j+1/2,L}}) + S(\widehat{u_{j+1/2,R}})}{2} - \frac{S(u_j) + S(u_{j+1})}{2} \leq -\lambda(G(u_{j+1}) - G(u_j)).$$

That is exactly the entropy inequality (19) of the lemma 4.3.  $\square$

Now we can think about a particular scheme. As in the previous work [11], the choice we do is to take the most downwind possible value. Here, ‘‘possible’’ means that the flux  $f_{j+1/2}$  will be in every case taken in the above mentioned interval  $[\phi_{j+1}, \Phi_{j+1}] \cap [b_j, B_j] \cap [r_{j+1/2}, R_{j+1/2}]$ , the CFL condition (24) being verified. Let us denote  $[\omega_{j+1/2}, \Omega_{j+1/2}]$  this interval. This means

$$\begin{cases} \omega_{j+1/2} = \max(\phi_{j+1}, b_j, r_{j+1/2}), \\ \Omega_{j+1/2} = \min(\Phi_{j+1}, B_j, R_{j+1/2}). \end{cases}$$

Always considering that  $f'(u_j) > 0 \forall j$ , we define the scheme defining  $f_{j+1/2}$  as follows

$$f_{j+1/2} = \begin{cases} \omega_{j+1/2} & \text{if } f_{j+1} < \omega_{j+1/2}, \\ f_{j+1} & \text{if } \omega_{j+1/2} \leq f_{j+1} \leq \Omega_{j+1/2}, \\ \Omega_{j+1/2} & \text{if } \Omega_{j+1/2} < f_{j+1}. \end{cases} \quad (28)$$

This is equivalent to the choice of  $f_{j+1/2}$  that minimizes  $|f_{j+1/2} - f(u_{j+1})|$  under the constraint  $\omega_{j+1/2} \leq f_{j+1/2} \leq \Omega_{j+1/2}$ .

This scheme is a 4-points scheme:  $f_{j+1/2}$  depends on the three variables  $u_{j-1}$ ,  $u_j$  and  $u_{j+1}$  so that  $\hat{u}_j$  is a function of the four variables  $u_{j-2}$ ,  $u_{j-1}$ ,  $u_j$  and  $u_{j+1}$ . It is not difficult to prove that the fluxes are Lipschitz-continuous functions of their variables (they are continuous and piecewise  $C^\infty$ ). Applying theorem 4.4, we deduce that this scheme is convergent toward the entropy solution.

**Remark 4.7.** All the above work concerns the case where  $f'(u_j) > 0 \forall j$ . In the opposite case where  $f'(u_j) < 0 \forall j$ , equivalent conditions and results can be found, changing the definitions (21), (22) and (23) a little. The problem occurs when there exists  $j$  such that  $f'(u_j) > 0$  and  $f'(u_{j+1}) \leq 0$  for example. This is a typical problem for most of schemes (even Godunov or Roe schemes) due to the presence of a sonic point. It is possible in general to modify a scheme so that it becomes entropic even at the sonic point, but we did not do this work.

**Remark 4.8.** The same work could be done for a strictly concave flux  $f(u)$ , for example the flux for the traffic flow equation (see [14]).

## 5. SOME NUMERICAL RESULTS

We insert now a few numerical results obtained with the scheme described above. We consider inviscid Burgers equation

$$\partial_t u + \partial_x \frac{u^2}{2} = 0. \quad (29)$$

An example of entropy-entropy flux couple is  $(u^2/2, u^3/3)$ . We used this couple to define the fluxes (recall that  $S$  and  $G$  play a role in (28)).

We present two results: one in short time (figure 2) and the other in long time (figure 3). The chosen initial condition is the same as for the previous result (figure 1), allowing the solution to develop a shock and a rarefaction wave. We put some periodic boundary conditions on the interval  $[0, 1]$  in order to be able to “follow” the waves in very long time. We can compare the upwind scheme (Roe scheme) and the one we propose here.

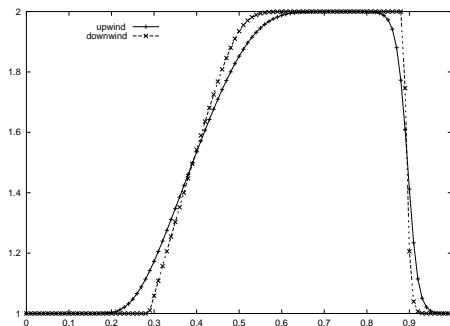
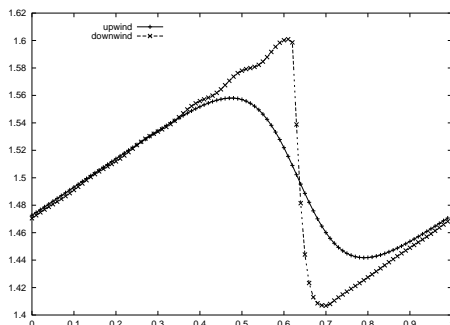
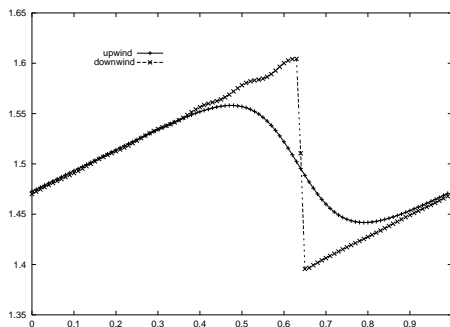


FIGURE 2. Solutions for  $t = 0.2$ .

For both results, we remark that the scheme developed here is less dissipative. However, the results are quite far away from the results obtained for long time linear advection (cf. infinite time convergence conjectured

FIGURE 3. Solutions for  $t = 4.7$ .

in [3]). The entropy conditions pollute a little the sharpness of the shock. To remedy this lack of resolution, we can do the same remark as LeFloch and Liu in [5]: the entropy condition is not necessary in the decreasing regions. So we can ask for the scheme to be only  $L^\infty$ -stable in the decreasing regions (and take interval (14) of proposition 3.2). With this released scheme we obtain in long time the better result of figure 4.

FIGURE 4. Solutions for  $t = 4.7$ .

At last the shock is perfectly restored! This result is more similar to the ones for linear advection (exact advection of Heavyside functions).

## 6. CONCLUSION

The first part of this work is devoted to the display of some explicit conditions for a finite-volume scheme to be convergent toward the entropy solution of a strictly convex scalar partial differential equation. We found conditions for the fluxes that guarantee such a convergence. The display of an explicit convergence interval for the fluxes is of practical interest: this interval being given, we can now write convergent schemes easily.

The second part concerns the numerical behavior of one particular convergent scheme. In the given convergence interval, we choose the most downwind flux, following the ideas of [11], [4]. As expected, the resulting scheme appears to be quite anti-dissipative. However, the entropy conditions introduce a little numerical dissipation and do not allow to recover exact discontinuity profiles as for advection equation, Euler equations (contact discontinuities) or multi-fluids equation (interfaces). In a last step, a little modification, as in [5], gives a scheme that preserves perfect shock profiles.

**Acknowledgments.** I deeply thank Frédéric Coquel for all his helpful suggestions.



## REFERENCES

- [1] F. COQUEL and P. G. LE FLOCH. Convergence of finite difference schemes for conservation laws in several space dimensions: the corrected antidiffusive approach. *Mathematics of Computation*, 57(195):169–210, 1991.
- [2] F. COQUEL and P. G. LE FLOCH. Convergence of finite difference schemes for conservation laws in several space dimensions: a general theory. *SIAM Journal of Numerical Analysis*, 30(3):675–700, 1993.
- [3] B. DESPRÉS and F. LAGOUTIÈRE. Un schéma non linéaire anti-dissipatif pour l'équation d'advection linéaire. *Comptes Rendus de l'Académie des Sciences, série 1*, 328:939–944, 1999.
- [4] B. DESPRÉS and F. LAGOUTIÈRE. Contact discontinuity capturing schemes for linear advection and compressible gas dynamics. *Preprint*, 2001.
- [5] P. G. LE FLOCH and J.-G. LIU. Generalized monotone schemes, discrete paths of extrema, and entropy conditions. *Mathematics of Computation*, 68(227):1025–1055, 1999.
- [6] E. GODLEWSKI and P.-A. RAVIART. *Hyperbolic systems of conservation laws*. Ellipses, 1991.
- [7] A. HARTEN. High resolution schemes for conservation laws. *Journal of Computational Physics*, 49:357–393, 1983.
- [8] A. HARTEN. On a class of high resolution total-variation-stable finite-difference schemes. *SIAM Journal of Numerical Analysis*, 21(1):1–23, 1984.
- [9] A. HARTEN and S. OSHER. Uniformly high-order accurate nonoscillatory schemes. *SIAM Journal of Numerical Analysis*, 24(2):279–309, 1987.
- [10] S. KRUKOV. First order quasi-linear equations in several independent variables. *English Translation in Math. USSR. Sb.*, 10:217–243, 1970.
- [11] F. LAGOUTIÈRE. *Modélisation mathématique et résolution numérique de problèmes de fluides compressibles à plusieurs constituants*. PhD thesis, Université Paris VI, 2000.
- [12] P. D. LAX. Hyperbolic systems of conservation laws, II. *Communications on Pure and Applied Mathematics*, 10:537–566, 1957.
- [13] P. D. LAX. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*. Philadelphia, 1973. SIAM.
- [14] R. J. LEVEQUE. *Numerical methods for conservation laws*. Birkhäuser, 1992.
- [15] P.-L. LIONS and P. E. SOUGANIDIS. Convergence of MUSCL and filtered schemes for scalar conservation laws and Hamilton-Jacobi equations. *Numerische Mathematik*, 69:441–470, 1995.
- [16] C.-W. SHU and S. OSHER. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77:439–471, 1988.
- [17] C.-W. SHU and S. OSHER. Efficient implementation of essentially non-oscillatory shock-capturing schemes, II. *Journal of Computational Physics*, 83:32–78, 1989.
- [18] P. K. SWEBY. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM Journal of Numerical Analysis*, 21(5):995–1011, 1984.
- [19] E. F. TORO. *Riemann solvers and numerical methods for fluid dynamics. A practical introduction*. Springer, 1997.