An antidiffusive entropy scheme for monotone scalar conservation laws

F. Bouchut

Département de Mathématiques et Applications
Ecole Normale Supérieure et CNRS, UMR 8553
45, rue d'Ulm
75230 Paris cedex 05, France
e-mail: Francois.Bouchut@ens.fr
http://www.dma.ens.fr/~fbouchut/

Abstract

In a recent work J. Sci. Comput. 16 (2001), 479-524, B. Després and F. Lagoutière introduced a new approach to derive numerical schemes for hyperbolic conservation laws. Its most important feature is the ability to perform an exact resolution for a single traveling discontinuity. However their scheme is not entropy satisfying and can keep nonentropic discontinuities. The purpose of our work is, starting from the previous one, to introduce a new class of schemes for monotone scalar conservation laws, that satisfy an entropy inequality, while still resolving exactly the single traveling shocks or contact discontinuities. We show that it is then possible to have an excellent resolution of rarefaction waves, and also to avoid the undesirable staircase effect. In practice, our numerical experiments show second-order accuracy.

Key-words: Scalar conservation laws – entropy scheme – entropy indicator – self-adaptive antidiffusion – step functions – smoothness indicator

Mathematics Subject Classification: 65M05

Contents

1 Introduction **2**

2 Self-adaptive antidiffusion via entropy indicator **3**

2.1 Nonlinearly stable antidiffusive schemes 4

2.2 Entropy indicator 6

2.3 Algorithm 9

2.4 Limitation of the entropy indicator 10

3 Accuracy **11**

3.1 Exact resolution of step functions 11

3.2 Smoothness indicator 16

4 Numerical results **18**

1 Introduction

It has always been a main issue in numerical analysis to conciliate stability and accuracy. Usually in hyperbolic conservation laws, stability properties such as maximum principle, TVD property or entropy inequalities impose a minimum amount of diffusion, and this leads to the spreading of discontinuities. Precise analysis can nevertheless lead to stable schemes that preserve exactly *stationary* discontinuities. However, for *traveling* discontinuities, it has always been a challenging problem to derive stable schemes that keep very sharp discontinuities. Even for linear advection with constant coefficient, almost all schemes completely flatten the profile after a sufficiently long time. Some cures have been proposed, such as artificial compression methods [6], [11], subcell resolution [7], or more recently [1]. A usual way to understand the problem is as follows: we need to put some antidiffusion (a second-order derivative in space with wrong sign) at the level of the linear expansion of the scheme, but only at the locations where this is possible without violating the stability criterion. This vision has been used in particular in the analysis of flux limiters for second-order schemes [10], [5].

In recent works [4], [8], [9], B. Després and F. Lagoutière have proposed a new approach that enables to put antidiffusion while keeping the nonlinear stability. In the case of linear advection their main scheme coincides with a previously known scheme called Ultrabee. However, even in this case their analysis is original, and enables to prove exact resolution of step functions, a property that had not been remarked previously. A corollary of this property is that the long time behavior is much better than in usual schemes, indeed after a transitional period the initial data is somehow projected to a class of piecewise constant functions, which shape are no more modified by the evolution. However, a defect of their scheme is that it is not entropy satisfying, and produces nonentropic discontinuities in nonlinear conservation laws. A correction is proposed in [8], [9], but puts rather strong dissipation in rarefaction waves.

The purpose of this paper is, starting from the analysis of [4], to introduce a new class of schemes for monotone scalar conservation laws, that satisfy an entropy inequality, while still resolving exactly the single traveling shocks or contact discontinuities. This is done by performing an overall limitation of the numerical flux obtained in [4], according to an evaluation of the entropy dissipation. The antidiffusion here performs in a self-adaptive manner. A very special feature of our schemes is that they evolve not only the cell average of the approximate solution but also another quantity, the *entropy indicator* that approximates the cell average of the entropy. This idea was present in the method of [2], [3], where this was a crucial point to obtain really second-order accuracy with all entropy inequalities (however the other evolved quantity was the slope there).

By construction, our method is only first-order accurate. However, as we show by numerical experiments, if we choose a suitably well-balanced self-adaptive antidiffusion procedure involving a smoothness indicator telling roughly where the function is discontinuous, we can obtain second-order accuracy. In particular, we obtain an exceptionally good resolution of rarefaction waves. Also, we avoid the staircase effect present in the method of [4], due to the projection to piecewise constant functions. A counterpart is that then a small dissipation process is reintroduced in the long time for functions that are far from step functions.

The paper is organized as follows. In Section 2 we present our framework of self-adaptive antidiffusion and entropy indicator. In Section 3 we examine under which conditions the schemes obtained in the preceding section exactly resolve traveling discontinuities for step functions, and introduce our main scheme involving the smoothness indicator. Numerical results are finally presented in Section 4.

2 Self-adaptive antidiffusion via entropy indicator

In all the paper we consider a one-dimensional scalar conservation law

$$\partial_t u + \partial_x F(u) = 0, \tag{2.1}$$

with $t > 0$, $x \in \mathbb{R}$ and $u(t, x) \in \mathbb{R}$. The nonlinearity F is assumed to be lipschitz continuous and nondecreasing,

$$F' \geq 0. \tag{2.2}$$

We look for solutions satisfying an entropy inequality

$$\partial_t S(u) + \partial_x G(u) \leq 0, \tag{2.3}$$

where S, the entropy, is a given convex function and G is the associated entropy flux, defined by

$$G' = S'F'. \tag{2.4}$$

Our approach can handle only a single entropy at a time, therefore the function S is assumed fixed once for all.

2.1 Nonlinearly stable antidiffusive schemes

We recall here the main condition derived in [4], [8], [9], that allows to build nonlinearly stable, in the sense of maximum principle and TVD property, but linearly antidiffusive schemes.

As usual, we approximate the solution $u(t,x)$ to (2.1) by discrete values u_i^n, $i \in \mathbb{Z}$, $n \in \mathbb{N}$, over a grid of points $x_{i+1/2}$, $i \in \mathbb{Z}$,

$$\ldots < x_{-1/2} < x_{1/2} < x_{3/2} < \ldots . \tag{2.5}$$

We define the cells $]x_{i-1/2}, x_{i+1/2}[$ and their lengths

$$\Delta x_i = x_{i+1/2} - x_{i-1/2} > 0, \tag{2.6}$$

and we consider a constant timestep $\Delta t > 0$. We define the discrete times $t_n = n\Delta t$ (variable timesteps could be used also), and

$$\lambda_i = \frac{\Delta t}{\Delta x_i}. \tag{2.7}$$

The discrete values u_i^n intend to be approximations of the averages of the exact solutions over the cells,

$$u_i^n \simeq \frac{1}{\Delta x_i} \int_{x_{i-1/2}}^{x_{i+1/2}} u(t_n, x)\, dx. \tag{2.8}$$

We shall often denote u_i instead of u_i^n, whenever there is no ambiguity.
We consider conservative schemes of the form

$$u_i^{n+1} - u_i^n + \lambda_i (F_{i+1/2} - F_{i-1/2}) = 0, \tag{2.9}$$

where the numerical flux $F_{i+1/2}$ needs to be defined, and in order to fix the ideas one can think of a dependency in u_{i-1}, u_i, u_{i+1}. The usual CFL condition will be used,

$$0 \leq \lambda_i F'(u) \leq 1. \tag{2.10}$$

A basic result from [4], [8], [9] is the following (with slightly different notations).

Proposition 2.1 *Assume that the CFL condition (2.10) holds, and that the numerical flux satisfies*

$$F_{i+1/2} \in [F_{i+1/2}^{diss}, F_{i+1/2}^{\ell}] \cap [F_{i+1/2}^{diss}, F_{i+1/2}^{r}], \tag{2.11}$$

where

$$\begin{aligned} F_{i+1/2}^{diss} &= F(u_i), \\ F_{i+1/2}^{\ell} &= \frac{u_i - u_{i-1}}{\lambda_i} + F(u_{i-1}), \\ F_{i+1/2}^{r} &= F(u_{i+1}). \end{aligned} \tag{2.12}$$

Then the scheme (2.9) satisfies the maximum principle and is total variation diminishing (TVD).

Proof. We have to precise that $[a,b]$ stands for the closed interval between a and b, independently of their ordering. In other words, $[a,b] = \{(1-X)a + Xb,\ 0 \leq X \leq 1\}$. Condition (2.11) allows to write

$$F_{i+1/2} = (1 - X^{\ell}_{i+1/2})F^{diss}_{i+1/2} + X^{\ell}_{i+1/2}F^{\ell}_{i+1/2}, \tag{2.13}$$

$$F_{i+1/2} = (1 - X^{r}_{i+1/2})F^{diss}_{i+1/2} + X^{r}_{i+1/2}F^{r}_{i+1/2}, \tag{2.14}$$

with $0 \leq X^{\ell}_{i+1/2}, X^{r}_{i+1/2} \leq 1$. Because of the CFL condition (2.10), we can write

$$F(u_i) - F(u_{i-1}) = v_{i-1/2}(u_i - u_{i-1}), \qquad 0 \leq \lambda_i v_{i-1/2} \leq 1. \tag{2.15}$$

Then, we compute

$$\begin{aligned}
&F_{i+1/2} - F_{i-1/2} \\
&= [F^{diss}_{i+1/2} + X^{\ell}_{i+1/2}(F^{\ell}_{i+1/2} - F^{diss}_{i+1/2})] - [F^{diss}_{i-1/2} + X^{r}_{i-1/2}(F^{r}_{i-1/2} - F^{diss}_{i-1/2})] \\
&= [F(u_i) + X^{\ell}_{i+1/2}(\frac{u_i - u_{i-1}}{\lambda_i} + F(u_{i-1}) - F(u_i))] \\
&\quad -[F(u_{i-1}) + X^{r}_{i-1/2}(F(u_i) - F(u_{i-1}))] \\
&= X^{\ell}_{i+1/2}\frac{u_i - u_{i-1}}{\lambda_i} + (1 - X^{\ell}_{i+1/2} - X^{r}_{i-1/2})(F(u_i) - F(u_{i-1})) \\
&= \left[(1 - \lambda_i v_{i-1/2})X^{\ell}_{i+1/2} + \lambda_i v_{i-1/2}(1 - X^{r}_{i-1/2})\right] \frac{u_i - u_{i-1}}{\lambda_i}.
\end{aligned} \tag{2.16}$$

Now, call $D_{i-1/2}$ the last term between brackets. We have $F_{i+1/2} - F_{i-1/2} = D_{i-1/2}(u_i - u_{i-1})/\lambda_i$, which yields by (2.9)

$$u^{n+1}_i = u_i - D_{i-1/2}(u_i - u_{i-1}). \tag{2.17}$$

Since $0 \leq D_{i-1/2} \leq 1$, the classical Harten criterion enables to conclude that the TVD property and the maximum principle hold. □

Remark 2.1 The *dissipative flux* $F^{diss}_{i+1/2}$ is the classical upwind flux, and the fluxes $F^{\ell}_{i+1/2}$, $F^{r}_{i+1/2}$ are the extremal leftwind and rightwind fluxes, in the sense that

$$\begin{aligned}
u_i - \lambda_i(F^{\ell}_{i+1/2} - F^{diss}_{i-1/2}) &= u_{i-1}, \\
u_i - \lambda_i(F^{diss}_{i+1/2} - F^{r}_{i-1/2}) &= u_i.
\end{aligned} \tag{2.18}$$

We can now define the numerical flux of Després and Lagoutière.

Definition 2.2 *The antidiffusive flux $F^{a}_{i+1/2}$ of [4] is defined by*

$$[F^{diss}_{i+1/2}, F^{\ell}_{i+1/2}] \cap [F^{diss}_{i+1/2}, F^{r}_{i+1/2}] = [F^{diss}_{i+1/2}, F^{a}_{i+1/2}]. \tag{2.19}$$

We notice that $F^a_{i+1/2}$ differs from $F^{diss}_{i+1/2}$ if and only if

$$\Big(F(u_{i+1}) - F(u_i)\Big)\Big(u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1}))\Big) > 0. \tag{2.20}$$

According to the CFL condition (2.10), this can happen only if

$$(u_{i+1} - u_i)(u_i - u_{i-1}) > 0, \tag{2.21}$$

which means that u_{i-1}, u_i, u_{i+1} are strictly ordered.

In the remaining of the paper, we always consider numerical fluxes $F_{i+1/2}$ satisfying (2.11), or equivalently

$$F_{i+1/2} \in [F^{diss}_{i+1/2}, F^a_{i+1/2}]. \tag{2.22}$$

A leading idea is to remark that $F^{diss}_{i+1/2}$ is a first-order *linearly diffusive* flux satisfying entropy inequalities, while $F^a_{i+1/2}$ is a first-order *linearly antidiffusive* flux, that exactly resolves step functions. Our strategy will be then to find a numerical flux between them that keeps the good properties of both extreme fluxes, and that is indeed second-order.

2.2 Entropy indicator

This subsection is devoted to the introduction of a limitation procedure applied to $F^a_{i+1/2}$, in order to find a flux $F_{i+1/2}$ satisfying (2.22), and an entropy inequality. This limitation means somehow to precise the amount of antidissipation according to an estimate of the entropy dissipation. This needs of course to be sharp in order to avoid putting too much dissipation.

The entropy inequality (2.3) is discretized via a new quantity S^n_i (or S_i for short), that we shall call *entropy indicator*, that approximates the cell average of the entropy,

$$S^n_i \simeq \frac{1}{\Delta x_i}\int_{x_{i-1/2}}^{x_{i+1/2}} S(u(t_n, x))\,dx. \tag{2.23}$$

We shall look for a discrete entropy inequality

$$S^{n+1}_i - S^n_i + \lambda_i(G_{i+1/2} - G_{i-1/2}) \le 0, \tag{2.24}$$

for some consistent numerical entropy flux $G_{i+1/2}$. We impose the following constraint on S^n_i,

$$S^n_i \ge S(u^n_i), \tag{2.25}$$

which is coherent with (2.8), (2.23) and Jensen's inequality. This constraint ensures some kind of stability, because according to (2.24), $\sum_i \Delta x_i S^n_i$ decreases with time, and

$$\sum_i \Delta x_i S(u^n_i) \le \sum_i \Delta x_i S^n_i \le \sum_i \Delta x_i S^0_i. \tag{2.26}$$

The entropy indicators S^n_i will evolve through another equation

$$S^{n+1}_i - S^n_i + \lambda_i(G_{i+1/2-} - G_{i-1/2+}) = 0. \tag{2.27}$$

Indeed, the whole scheme has an unusual structure of a system: knowing the values $(u_i^n, S_i^n)_{i\in\mathbb{Z}}$ at time t_n, one has to compute the values $(u_i^{n+1}, S_i^{n+1})_{i\in\mathbb{Z}}$ at time t_{n+1}. As a consequence, the computation of the numerical flux $F_{i+1/2}$ and of the numerical left/right entropy fluxes $G_{i+1/2\pm}$ will involve the values of u_i^n and S_i^n, $i \in \mathbb{Z}$.

In order to satisfy (2.24), we ask the inequality

$$G_{i+1/2+} \leq G_{i+1/2-}. \tag{2.28}$$

Obviously, (2.27) and (2.28) yield (2.24) with any numerical entropy flux $G_{i+1/2}$ satisfying $G_{i+1/2+} \leq G_{i+1/2} \leq G_{i+1/2-}$.

Our strategy can now be resumed as follows. Starting from values $(u_i^n, S_i^n)_{i\in\mathbb{Z}}$ satisfying the constraint (2.25), we are going to define some numerical left/right entropy fluxes $G_{i+1/2\pm}$ expressed explicitly in terms of $(u_i^n, S_i^n)_{i\in\mathbb{Z}}$ and of $F_{i+1/2}$. Writing the condition (2.28) and the constraints (2.25) at time t_{n+1}, we are going to derive an explicit limitation on the numerical flux $F_{i+1/2}$, of the form $|F_{i+1/2} - F^{diss}_{i+1/2}|$ less than something explicit.

Our choice for $G_{i+1/2\pm}$ is done as follows. Since $F_{i+1/2}$ satisfies (2.11), we can write again (2.13) and (2.14), or more explicitly

$$\begin{aligned} F_{i+1/2} &= (1 - X^\ell_{i+1/2})F(u_i) + X^\ell_{i+1/2}\left(\frac{u_i - u_{i-1}}{\lambda_i} + F(u_{i-1})\right) \\ &= (1 - X^r_{i+1/2})F(u_i) + X^r_{i+1/2}F(u_{i+1}). \end{aligned} \tag{2.29}$$

This defines $X^\ell_{i+1/2}$, $X^r_{i+1/2} \in [0,1]$, as functions of $F_{i+1/2}$. In the case when $F^a_{i+1/2} = F^{diss}_{i+1/2} \equiv F(u_i)$, these equations could not define unique values, therefore we just set $X^\ell_{i+1/2} = X^r_{i+1/2} = 0$. Then, by analogy with (2.29), we define

$$\begin{aligned} G_{i+1/2-} &= (1 - X^\ell_{i+1/2})(G(u_i) + E_{i+1/2-}) + X^\ell_{i+1/2}\left(\frac{S_i - S(u_{i-1})}{\lambda_i} + G(u_{i-1})\right), \\ G_{i+1/2+} &= (1 - X^r_{i+1/2})(G(u_i) - E_{i+1/2+}) + X^r_{i+1/2}G(u_{i+1}), \end{aligned} \tag{2.30}$$

where $E_{i+1/2\pm}$ will be chosen later explicitly, independently of $F_{i+1/2}$.

Lemma 2.3 *Assume that the CFL condition (2.10) holds, and that the numerical flux $F_{i+1/2}$ satisfies (2.11). Define the left/right numerical entropy fluxes by (2.29)-(2.30), and the scheme by (2.9), (2.27).*
(i) If $E_{i+1/2\pm}$ satisfy

$$E_{i+1/2-} \leq \frac{S_i - S(u_i)}{\lambda_i}, \qquad E_{i+1/2+} \leq 0, \tag{2.31}$$

then the scheme preserves the constraint (2.25).
(ii) The entropy condition (2.28) holds if and only if

$$|F_{i+1/2} - F^{diss}_{i+1/2}|K_{i+1/2} \leq E_{i+1/2-} + E_{i+1/2+}, \tag{2.32}$$

where $K_{i+1/2}$ *is given whenever* $F^a_{i+1/2} \neq F^{diss}_{i+1/2}$ *by*

$$K_{i+1/2} = \frac{G(u_{i+1}) - G(u_i) + E_{i+1/2+}}{|F(u_{i+1}) - F(u_i)|} - \frac{S_i - S(u_{i-1}) - \lambda_i(G(u_i) - G(u_{i-1}) + E_{i+1/2-})}{|u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1}))|}. \tag{2.33}$$

Proof. Let us express the constraint (2.25) evaluated at the next time level, *i.e.* $S(u_i^{n+1}) - S_i^{n+1} \leq 0$. According to (2.27) and (2.30), we have

$$\begin{aligned} &S(u_i^{n+1}) - S_i^{n+1} \\ &= S(u_i^{n+1}) - S_i + X^\ell_{i+1/2}(S_i - S(u_{i-1})) \\ &\quad + (1 - X^\ell_{i+1/2} - X^r_{i-1/2})\lambda_i(G(u_i) - G(u_{i-1})) \\ &\quad + (1 - X^\ell_{i+1/2})\lambda_i E_{i+1/2-} + (1 - X^r_{i-1/2})\lambda_i E_{i-1/2+}. \end{aligned} \tag{2.34}$$

Noticing that from (2.16),

$$u_i^{n+1} = u_i^n - X^\ell_{i+1/2}(u_i - u_{i-1}) - \lambda_i(1 - X^\ell_{i+1/2} - X^r_{i-1/2})(F(u_i) - F(u_{i-1})), \tag{2.35}$$

we observe that $S(u_i^{n+1}) - S_i^{n+1}$, seen as a function of $X^\ell_{i+1/2}$, $X^r_{i-1/2}$, is convex. Therefore, it is enough to prove that this function is nonpositive at the corners of the square $[0,1] \times [0,1]$.

– At $(X^\ell_{i+1/2}, X^r_{i-1/2}) = (0,1)$, we must have

$$S(u_i) - S_i + \lambda_i E_{i+1/2-} \leq 0, \tag{2.36}$$

– at $(X^\ell_{i+1/2}, X^r_{i-1/2}) = (1,0)$, we must have

$$\lambda_i E_{i-1/2+} \leq 0, \tag{2.37}$$

– at $(X^\ell_{i+1/2}, X^r_{i-1/2}) = (0,0)$, we must have

$$S\Big(u_i - \lambda_i(F(u_i) - F(u_{i-1}))\Big) - S_i + \lambda_i\Big(G(u_i) - G(u_{i-1}) + E_{i+1/2-} + E_{i-1/2+}\Big) \leq 0, \tag{2.38}$$

– at $(X^\ell_{i+1/2}, X^r_{i-1/2}) = (1,1)$, we must have

$$S\Big(u_{i-1} - \lambda_i(F(u_{i-1}) - F(u_i))\Big) - S(u_{i-1}) - \lambda_i(G(u_i) - G(u_{i-1})) \leq 0. \tag{2.39}$$

The two first conditions give (2.31). Then, we observe that (2.39) is always true under the CFL condition (2.10), this is just the entropy inequality for the upwind scheme. Finally, the same entropy inequality for the upwind scheme gives that

$$S\Big(u_i - \lambda_i(F(u_i) - F(u_{i-1}))\Big) - S(u_i) + \lambda_i(G(u_i) - G(u_{i-1})) \leq 0, \tag{2.40}$$

and therefore (2.38) is a consequence of (2.36)-(2.37). This concludes the (i).

In order to prove (ii), we write $G_{i+1/2+} - G_{i+1/2-}$ with the expressions (2.30). In the case where $F^a_{i+1/2} = F^{diss}_{i+1/2}$, we have $X^\ell_{i+1/2} = X^r_{i+1/2} = 0$ and $G_{i+1/2+} - G_{i+1/2-} = -E_{i+1/2+} - E_{i+1/2-}$, thus this is nonpositive if and only if $E_{i+1/2-} + E_{i+1/2+} \geq 0$. Let us now assume that $F^a_{i+1/2} \neq F^{diss}_{i+1/2}$, *i.e.* (2.20). Then we can express $X^\ell_{i+1/2}$ and $X^r_{i+1/2}$ in terms of $F_{i+1/2}$ by (2.29),

$$\begin{aligned} X^\ell_{i+1/2} &= \lambda_i \frac{F_{i+1/2} - F(u_i)}{u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1}))}, \\ X^r_{i+1/2} &= \frac{F_{i+1/2} - F(u_i)}{F(u_{i+1}) - F(u_i)}. \end{aligned} \tag{2.41}$$

Since $X^\ell_{i+1/2}, X^r_{i+1/2} \geq 0$, we can put absolute values in the numerators and denominators of the right-hand sides of (2.41). Putting these values in (2.30) and writing that $G_{i+1/2+} - G_{i+1/2-} \leq 0$, we easily get (2.32)-(2.33). □

Remark 2.2 If we neglect $E_{i+1/2\pm}$ and replace S_i by $S(u_i)$ in the expression (2.33) of $K_{i+1/2}$, we get something nonnegative. Indeed, one has

$$\begin{aligned} \frac{G(u_{i+1}) - G(u_i)}{|F(u_{i+1}) - F(u_i)|} &= S'(\xi_{i+1/2+}) \operatorname{sgn}(u_{i+1} - u_i), \\ \frac{S(u_i) - S(u_{i-1}) - \lambda_i(G(u_i) - G(u_{i-1}))}{|u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1}))|} &= S'(\xi_{i+1/2-}) \operatorname{sgn}(u_i - u_{i-1}), \end{aligned} \tag{2.42}$$

for some $\xi_{i+1/2+} \in (u_i, u_{i+1})$ and $\xi_{i+1/2-} \in (u_{i-1}, u_i)$. Since $K_{i+1/2}$ is defined only when $F^a_{i+1/2} \neq F^{diss}_{i+1/2}$, which implies (2.21), this gives by the convexity of S that $S'(\xi_{i+1/2+}) \operatorname{sgn}(u_{i+1} - u_i) - S'(\xi_{i+1/2-}) \operatorname{sgn}(u_i - u_{i-1}) \geq 0$. It is even strictly positive if S is strictly convex.

Remark 2.3 When $F^a_{i+1/2} \neq F^{diss}_{i+1/2}$, or equivalently when (2.20) holds, we can always define $X^\ell_{i+1/2}$ and $X^r_{i+1/2}$ by (2.41). Then the property $F_{i+1/2} \in [F^{diss}_{i+1/2}, F^a_{i+1/2}]$ is equivalent to $0 \leq X^\ell_{i+1/2} \leq 1$ and $0 \leq X^r_{i+1/2} \leq 1$. Also, $\max(X^\ell_{i+1/2}, X^r_{i+1/2}) = 1$ means that $F_{i+1/2} = F^a_{i+1/2}$.

2.3 Algorithm

The algorithm for our scheme can now be described as follows. We need first to define some values $E_{i+1/2\pm}$ in terms of $(u^n_i, S^n_i)_{i\in\mathbb{Z}}$ (supposed to satisfy the constraint (2.25)). These values need to satisfy (2.31) and

$$E_{i+1/2-} + E_{i+1/2+} \geq 0. \tag{2.43}$$

Then, we define the numerical flux $F_{i+1/2} \in [F^{diss}_{i+1/2}, F^a_{i+1/2}]$ by taking the value which is the closest to $F^a_{i+1/2}$ satisfying (2.32). In particular, if $K_{i+1/2} \leq 0$, there is no need of limitation and $F_{i+1/2} = F^a_{i+1/2}$. On the contrary, if $F^a_{i+1/2} =$

$F^{diss}_{i+1/2}$, we need to take $F_{i+1/2} = F^{diss}_{i+1/2}$. Then, we compute $X^{\ell}_{i+1/2}$ and $X^{r}_{i+1/2}$ by (2.41) (or $X^{\ell}_{i+1/2} = X^{r}_{i+1/2} = 0$ if $F^{a}_{i+1/2} = F^{diss}_{i+1/2}$), and the numerical entropy fluxes $G_{i+1/2\pm}$ by (2.30). Finally, we update the values u_i^n and S_i^n by (2.9), (2.27). According to Proposition 2.1 and Lemma 2.3, this scheme preserves the constraint (2.25), is TVD, satisfies the maximum principle and the entropy inequality (2.24).

We have now to comment the possible choices of $E_{i+1/2\pm}$, that need to satisfy (2.31) and (2.43).

We can observe first that if $S_i - S(u_i)$=0, the only possible choice is $E_{i+1/2-} = E_{i+1/2+} = 0$, and thus by (2.32) $F_{i+1/2} = F^{diss}_{i+1/2}$. Consequently, in general, the nonnegative quantity $S_i - S(u_i)$ represents somehow an antidiffusion potential: larger values of it allow to obtain a more antidiffusive numerical flux $F_{i+1/2}$. In this respect, one could think of choosing the largest possible values for $E_{i+1/2\pm}$.

However, we observe in (2.27), (2.30) that larger values of $E_{i+1/2\pm}$ induce smaller values of S_i^{n+1}. In particular, if too much antidiffusion is performed at time t_n, the antidiffusion potential will be very small at time t_{n+1} and no more antidiffusion will be possible at this time. Therefore, we conclude that the choice of $E_{i+1/2\pm}$ need be done very carefully in order to consume some reasonable part of the antidiffusion potential, not too little in order to prevent dissipation at time t_n, and not too much in order to prevent dissipation at time t_{n+1}.

The efficiency of the algorithm, that is observed in numerical experiments, can be explained by the following argument. If, for some reason, the entropy potential is small at some time t_n, then, it will probably increase at lot at the next time t_{n+1}, and what is dissipated at time t_n will be recovered by antidissipation at time t_{n+1}, indeed S_i^n plays the role of the memory of the past. This mechanism is what we call self-adaptive antidiffusion.

2.4 Limitation of the entropy indicator

We intend to introduce here a variant of the algorithm defined in the previous subsection, that enables to fully justify the convergence of the scheme.

Let us define, for any sequence $(u_i)_{i\in\mathbb{Z}}$, the following values,

$$\widehat{S}_i = \begin{cases} \dfrac{u_{i+1}-u_i}{u_{i+1}-u_{i-1}} S(u_{i-1}) + \dfrac{u_i - u_{i-1}}{u_{i+1}-u_{i-1}} S(u_{i+1}) & \text{if } u_i \in (u_{i-1}, u_{i+1}), \\ S(u_i) \quad \text{otherwise.} \end{cases} \tag{2.44}$$

By convexity of S, we have obviously $\widehat{S}_i \geq S(u_i)$, thus these values are canonical candidates for the entropy indicators S_i.

Our modified algorithm is defined with the following overall limitation of the entropy indicator: when applying (2.27), we call the result S_i^{n+1-}. Then, the updated values S_i^{n+1} are defined by

$$S_i^{n+1} = \min\Big(S_i^{n+1-}, \widehat{S}_i^{n+1}\Big), \tag{2.45}$$

where the values $\widehat{S}_i^{n+1}$ are computed with the sequence $(u_i^{n+1})_{i\in\mathbb{Z}}$. Obviously, these values S_i^{n+1} still satisfy the constraint (2.25). Indeed one has now

$$S(u_i^n) \leq S_i^n \leq \widehat{S}_i^n, \tag{2.46}$$

and the inequality (2.24) still holds. In practice we observe only very small differences between this algorithm and the previous one. However, (2.46) guarantees the consistency of S_i^n, and allows to prove the following result, whatever is the choice of $E_{i+1/2\pm}$ satisfying (2.31), (2.43).

Proposition 2.4 *For initial data in BV, the modified scheme with the limitation (2.45) on the entropy indicator converges to a weak solution to (2.1) satisfying the entropy inequality (2.3).*

Proof. The TVD property and the conservative formula (2.9) guarantee the strong compactness. In order to use the usual Lax-Wendroff approach, it is enough to have consistency of the fluxes $F_{i+1/2}$, which is obvious from (2.11). Similarly, in order to obtain the entropy inequality, we use (2.24), thus we just have to prove the consistency of S_i^n and $G_{i+1/2}$. The consistency of S_i^n comes from (2.46). For the entropy fluxes, it is enough to have the consistency of $G_{i+1/2\pm}$ since $G_{i+1/2}$ lies between them. This comes from (2.30) since $0 \leq X_{i+1/2}^\ell, X_{i+1/2}^r \leq 1$, S_i is consistent and $E_{i+1/2\pm}$ are consistent with 0 by (2.31), (2.43). $\square$

3 Accuracy

This section is devoted to the introduction of criteria for selecting the values of $E_{i+1/2\pm}$ that define the scheme described in Subsection 2.3. The first criterion, explained in Subsection 3.1, is the exact resolution of step functions for a linear equation. It puts some further conditions on these coefficients. A choice is then proposed in Subsection 3.2, involving a smoothness indicator that detects the location of discontinuities.

3.1 Exact resolution of step functions

As was done in [4], we are going to introduce some special simple functions for which we are able to compute the effect of the algorithm.

Definition 3.1 *(i) A discrete step function is a sequence $(u_i)_{i\in\mathbb{Z}}$ such that there exists an integer k and two values u_ℓ, u_r for which we have*

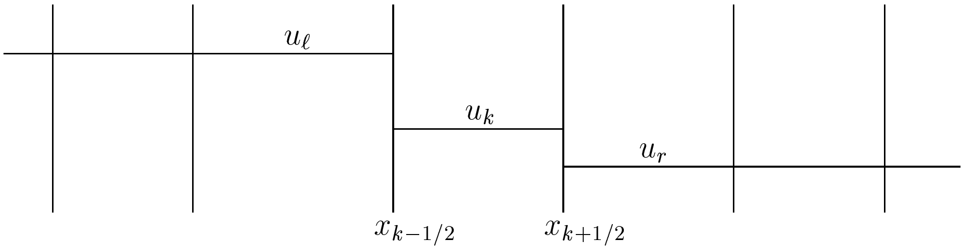
$$\begin{cases} u_i = u_\ell & \text{if } i \leq k-1, \\ u_i = u_r & \text{if } i \geq k+1, \end{cases} \tag{3.1}$$

and

$$u_k \in [u_\ell, u_r]. \tag{3.2}$$

(ii) A continuous step function is a function defined over $\mathbb{R}$ and taking at most two values with a discontinuity.

Figure 1: A discrete step function



A discrete step function is represented in Figure 1. The integer k indexing the intermediate cell may be not unique if the discrete step function takes only one or two values, instead of three in general.

Lemma 3.2 *(i) The sequence of cell averages of a continuous step function form a discrete step function.*
(ii) Given a discrete step function (u_i), there exists a unique continuous step function having u_i as cell averages.

Proof. The only nontrivial statement is (ii). If $u_\ell = u_r$, then the unique solution is the constant. Otherwise, the continuous step function $\widetilde{u}(x)$ must have a discontinuity at a point $x_{k-1/2} \leq \widetilde{x} \leq x_{k+1/2}$, with values u_ℓ and u_r on each side. The average of $\widetilde{u}$ in the cell $(x_{k-1/2}, x_{k+1/2})$ must be u_k, thus

$$\int_{x_{k-1/2}}^{\widetilde{x}} u_\ell \, dx + \int_{\widetilde{x}}^{x_{k+1/2}} u_r \, dx = (x_{k+1/2} - x_{k-1/2}) u_k, \tag{3.3}$$

which gives

$$\widetilde{x} = \frac{u_k - u_\ell}{u_r - u_\ell} x_{k-1/2} + \frac{u_r - u_k}{u_r - u_\ell} x_{k+1/2}. \tag{3.4}$$

This value $\widetilde{x}$ lies in $[x_{k-1/2}, x_{k+1/2}]$ because of (3.2). $\square$

The values $\widehat{S}_i$ defined in (2.44) have been chosen in order to be compatible with step functions, in the following sense.

Lemma 3.3 *If $(u_i)_{i\in\mathbb{Z}}$ is a discrete step function, then $(\widehat{S}_i)_{i\in\mathbb{Z}}$ defined in (2.44) is also a discrete step function, formed by the cell averages of $S(\widetilde{u})$, where $\widetilde{u}$ is the continuous step function having u_i as cell averages.*

Proof. If (u_i) is constant, $\widehat{S}_i$ is also constant. Assume on the contrary that (u_i) is not constant. We have $\widehat{S}_i = S(u_i)$ for any $i \neq k$, thus $\widehat{S}_i = S(u_l)$ for $i \leq k-1$, and $\widehat{S}_i = S(u_r)$ for $i \geq k+1$. The value $\widehat{S}_k$ is given by the first line in (2.44), so that $\widehat{S}_k \in [S(u_l), S(u_r)]$, which yields that $(\widehat{S}_i)$ is a discrete step function. In the special case when $S(u_l) = S(u_r)$, $(\widehat{S}_i)$ is constant. Otherwise, according to Lemma 3.2 and (3.4), the discontinuity of the continuous step function $\widetilde{S}(x)$ associated to $(\widehat{S}_i)$ is located at

$$\widetilde{x}_S = \frac{S_k - S(u_\ell)}{S(u_r) - S(u_\ell)} x_{k-1/2} + \frac{S(u_r) - S_k}{S(u_r) - S(u_\ell)} x_{k+1/2} = \widetilde{x}. \tag{3.5}$$

Therefore, $\widetilde{S}$ is a continuous step function with a discontinuity at $\widetilde{x}$ and with values $S(u_l)$ and $S(u_r)$ on each side, *i.e.* $\widetilde{S} = S(\widetilde{u})$. $\square$

With the above definitions, we can now clearly define the *exact translation* of a discrete step function, corresponding the the resolution of the linear conservation law, $F(u) = au$. Indeed, given a discrete step function, we can associate its continuous step function, that we can translate by $a\Delta t$, and then compute its cell averages. This gives the *exact translation* of the initial discrete step function.

Proposition 3.4 *Assume that the flux function is linear $F(u) = au$, $a \geq 0$, and consider a discrete step function $(u_i)_{i\in\mathbb{Z}}$ with intermediate cell k, together with entropy indicators given by their canonical values $S_i = \widehat{S}_i$, $i \in \mathbb{Z}$. Then, a scheme defined as in Subsection 2.3 with coefficients $E_{i+1/2\pm}$ satisfying (2.31), (2.43), sends the couple $(u_i)_{i\in\mathbb{Z}}, (S_i)_{i\in\mathbb{Z}}$ to its exact translation by $a\Delta t$ together with entropy indicators equal again to their canonical values if and only if the coefficients $E_{k+1/2\pm}$ satisfy the following conditions.*
(i) Incell evolution. If $\lambda_k a|u_{k+1} - u_k| < (1 - \lambda_k a)|u_k - u_{k-1}|$, then

$$\begin{cases} E_{k+1/2-} = a(S_k - S(u_k)), \\ -a(S_k - S(u_k)) \leq E_{k+1/2+} \leq 0. \end{cases} \tag{3.6}$$

(ii) Trans-cell evolution. If $\lambda_k a|u_{k+1} - u_k| > (1 - \lambda_k a)|u_k - u_{k-1}|$, then

$$\begin{cases} E_{k+1/2+} = -a(S_k - S(u_k)), \\ a(S_k - S(u_k)) \leq E_{k+1/2-} \leq \dfrac{S_k - S(u_k)}{\lambda_k}. \end{cases} \tag{3.7}$$

In the case of equality, no condition is required.

Proof. We recall that by the CFL condition, $0 \leq \lambda_k a \leq 1$. Since $F(u) = au$, we can take $G(u) = aS(u)$.
– First, if $u_\ell = u_r = u$, then the discrete and continuous initial step functions are constant, and $F_{i+1/2} = F^a_{i+1/2} = F^{diss}_{i+1/2} = F(u)$ for all i. Also, $S_i = \widehat{S}_i = S(u)$ for all i, thus (2.31) and (2.43) impose that $E_{i+1/2\pm} = 0$ for all i. Formulas (2.30) give also $G_{i+1/2\pm} = G(u)$, and (2.9), (2.27) show that there is no change in u_i and S_i, which is the exact solution. This is in accordance with the fact that in this case there is no condition to impose.
– Next, assume that $u_\ell \neq u_r$. We observe that $F^\ell_{i+1/2} = F^{diss}_{i+1/2}$ for $i < k$, and $F^r_{i+1/2} = F^{diss}_{i+1/2}$ for $i > k$, thus $F_{i+1/2} = F^a_{i+1/2} = F^{diss}_{i+1/2} = F(u_i)$ for all $i \neq k$. Since by (2.44), $S_i = S(u_i)$ for all $i \neq k$, (2.31), (2.43) impose that $E_{i+1/2\pm} = 0$, and $G_{i+1/2\pm} = G(u_i)$ for all $i \neq k$. We deduce by (2.9), (2.27) that $u_i^{n+1} = u_i$ and $S_i^{n+1} = S_i$ for all $i \leq k-1$ or $i \geq k+2$, which corresponds to the exact solution. It remains now to examine the values of u_k^{n+1}, S_k^{n+1}, u_{k+1}^{n+1} and S_{k+1}^{n+1}.

The initial continuous step function has a discontinuity at $\widetilde{x}$ given by (3.4), while the exact continuous step function at time t_{n+1} has a discontinuity at $\widetilde{x}^{n+1} = \widetilde{x} + a\Delta t$. In order to determine if $\widetilde{x}^{n+1}$ lies in the cell k or $k+1$, we observe that $\widetilde{x}^{n+1} \leq x_{k+1/2}$ means by (3.4) that $a\Delta t \leq x_{k+1/2} - \widetilde{x} = \Delta x_k(u_k - u_\ell)/(u_r - u_\ell)$. By (3.2) we can put absolute values, thus this is equivalent to $\lambda_k a|u_r - u_\ell| \leq |u_k - u_\ell|$. Using again (3.2), we have $|u_r - u_\ell| = |u_r - u_k| + |u_k - u_\ell|$, thus this gives the condition determining the incell or trans-cell evolution,

$$\lambda_k a|u_{k+1} - u_k| \leq (1 - \lambda_k a)|u_k - u_{k-1}|. \tag{3.8}$$

(i) Case of incell evolution. The average of the exact continuous step function at time t_{n+1} is computed by (3.3) where we replace $\widetilde{x}$ by $\widetilde{x}^{n+1}$, thus u_k^{n+1} must be given by $u_k^{n+1} = u_k - a\Delta t(u_r - u_\ell)/\Delta x_k$. Comparing to (2.9), we deduce that the flux must be given by $F_{k+1/2} = au_{k+1} = F^r_{k+1/2}$. This is also the value to get the right average u_{k+1}^{n+1}. Similarly, the values needed for the entropy fluxes are $G_{k+1/2+} = G_{k+1/2-} = G(u_{k+1})$, because with (2.44),

$$\begin{aligned}
\widehat{S}_k^{n+1} &= \widehat{S}_k + (u_k^{n+1} - u_k)\frac{S(u_{k+1}) - S(u_{k-1})}{u_{k+1} - u_{k-1}} \\
&= \widehat{S}_k - \lambda_k a\Big(S(u_{k+1}) - S(u_{k-1})\Big) \\
&= \widehat{S}_k - \lambda_k\Big(G(u_{k+1}) - G(u_{k-1})\Big).
\end{aligned} \tag{3.9}$$

We notice now that (3.8) ensures that $F^a_{k+1/2} = F^r_{k+1/2}$. In the case $F^a_{k+1/2} = F^{diss}_{k+1/2}$, *i.e.* $a(u_{k+1} - u_k) = 0$, we have $a(S(u_{k+1}) - S(u_k)) = 0$, and $a(S_k - S(u_k)) = 0$. The flux is necessarily the right one, and (2.30) gives that the conditions are $E_{k+1/2\pm} = 0$, which identify to (3.6). Otherwise, $F^a_{k+1/2} \neq F^{diss}_{k+1/2}$, and the condition is that the choice $F_{k+1/2} = F^r_{k+1/2}$ gives the right value for $G_{k+1/2\pm}$ (*i.e.* $G(u_{k+1})$), because then (2.28) holds automatically. Since then $X^r_{k+1/2} = 1$, the second line of (2.30) gives $G_{k+1/2+} = G(u_{k+1})$, thus the only condition is to write that the first line in (2.30) gives the right value $G(u_{k+1})$. We have by (2.41)

$$X^\ell_{k+1/2} = \frac{\lambda_k a}{1 - \lambda_k a}\frac{u_{k+1} - u_k}{u_k - u_{k-1}}, \tag{3.10}$$

thus

$$X^\ell_{k+1/2} + \lambda_k a(1 - X^\ell_{k+1/2}) = \lambda_k a\frac{u_{k+1} - u_{k-1}}{u_k - u_{k-1}}, \tag{3.11}$$

and with (2.44),

$$\begin{aligned}
&X^\ell_{k+1/2}\frac{S_k - S(u_{k-1})}{\lambda_k} + a(1 - X^\ell_{k+1/2})(S_k - S(u_{k-1})) \\
&= a(S(u_{k+1}) - S(u_{k-1})).
\end{aligned} \tag{3.12}$$

Therefore, identifying the first line in (2.30) to $G(u_{k+1})$ gives

$$(1 - X^\ell_{k+1/2})\Big(aS(u_k) + E_{k+1/2-} - aS_k\Big) = 0, \tag{3.13}$$

concluding the (i) because the second line of (3.6) comes from (2.31), (2.43).

(ii) Case of trans-cell evolution. A computation gives that the value of the numerical flux to get the right averages u_k^{n+1}, u_{k+1}^{n+1} is now $F_{k+1/2}=F^\ell_{k+1/2}$, and the values needed for the entropy fluxes are $G_{k+1/2+}=G_{k+1/2-}=G(u_{k-1})+(S_k-S(u_{k-1}))/\lambda_k$. We notice now that the trans-cell inequality gives that $F^a_{k+1/2}=F^\ell_{k+1/2}$. In the case $F^a_{k+1/2}=F^{diss}_{k+1/2}$, *i.e.* $(1-\lambda_k a)(u_k-u_{k-1})=0$, we have $(1-\lambda_k a)(S(u_k)-S(u_{k-1}))=0$, and $(1-\lambda_k a)(S_k-S(u_k))=0$. The flux is necessarily the right one, and (2.30) gives that the conditions are $\mp E_{k+1/2\pm}=(S_k-S(u_k))/\lambda_k$, which identify to (3.7). Otherwise, $F^a_{k+1/2}\neq F^{diss}_{k+1/2}$, and the condition is that the choice $F_{k+1/2}=F^\ell_{k+1/2}$ gives the right value for $G_{k+1/2\pm}$, because then (2.28) holds automatically. Since then $X^\ell_{k+1/2}=1$, the first line of (2.30) gives the right value for $G_{k+1/2-}$. Thus the only condition is on $G_{k+1/2+}$. We have by (2.41)

$$X^r_{k+1/2}=\frac{1-\lambda_k a}{\lambda_k a}\frac{u_k-u_{k-1}}{u_{k+1}-u_k}, \tag{3.14}$$

thus

$$1-\lambda_k a(1-X^r_{k+1/2})=(1-\lambda_k a)\frac{u_{k+1}-u_{k-1}}{u_{k+1}-u_k}, \tag{3.15}$$

and with (2.44),

$$\begin{aligned}&\frac{S_k-S(u_{k+1})}{\lambda_k}-a(1-X^r_{k+1/2})(S_k-S(u_{k+1}))\\ &=-(\frac{1}{\lambda_k}-a)(S(u_{k+1})-S(u_{k-1})).\end{aligned} \tag{3.16}$$

Therefore, identifying the second line in (2.30) to $G(u_{k-1})+(S_k-S(u_{k-1}))/\lambda_k$ gives

$$(1-X^r_{k+1/2})\Big(aS(u_k)-E_{k+1/2+}-aS_k\Big)=0, \tag{3.17}$$

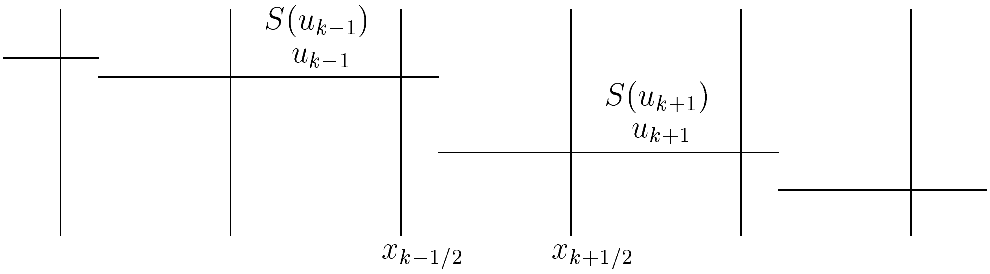
concluding the (ii), because the second line of (3.7) comes from (2.31), (2.43). Finally, in the limit case between incell and trans-cell, if $F^a_{k+1/2}\neq F^{diss}_{k+1/2}$, we have $X^\ell_{k+1/2}=X^r_{k+1/2}=1$, thus either in (3.13) or in (3.17) there is no condition. □

Remark 3.1 In Proposition 3.4, the numerical fluxes are always given by $F_{i+1/2}=F^a_{i+1/2}$, $i\in\mathbb{Z}$, and the entropy inequalities (2.24) vanish identically.

An exact resolution is also possible for double-cell constant functions.

Proposition 3.5 *Assume that the flux function is linear $F(u)=au$, $a\geq 0$, and consider a scheme defined as in Subsection 2.3, that satisfies the conditions of Proposition 3.4 for all discrete step functions. Assume also that $E_{i+1/2\pm}$ depend only on the values of the data in cells $i-1$, i, $i+1$, and that the mesh is uniform with size Δx. Consider an initial data (u_i^0) which is double-cell constant: $u^0_{2j}=u^0_{2j+1}$ for any $j\in\mathbb{Z}$, and take $S_i^0=S(u_i^0)$. Then the scheme gives values (u_i^n) at time t_n which are the cell averages of the exact translation by at of the initial data.*

Figure 2: Exact translation



Proof. Define a function $u^0(x) = u_i^0$ for $x_{i-1/2} < x < x_{i+1/2}$. We argue by induction, and prove that if at time t_n, u_i and S_i are given by the cell averages of $u^0(x-at_n)$ and $S(u^0(x-at_n))$ respectively, then we have the same properties at time t_{n+1}. According to Lemma 3.3, we observe that in such a situation, S_i is given by

$$\begin{cases} S_{k+2j+1} = S(u_{k+2j+1}) \quad \text{for any } j \in \mathbb{Z}, \\ S_{k+2j} = \widehat{S}_{k+2j} \quad \text{for any } j \in \mathbb{Z}, \end{cases} \tag{3.18}$$

where k is chosen such that the exact translation $u^0(x - at_n)$ has no discontinuity in $(x_{k+1/2}, x_{k+3/2})$, see Figure 2. Then, the inequalities (2.31), (2.43) impose that $E_{k+2j+3/2\pm} = 0$. Thus, by (2.32) and Remark 2.2, we have $F_{k+2j+3/2} = F^{diss}_{k+2j+3/2}$, and therefore by (2.30) $G_{k+2j+3/2\pm} = G(u_{k+2j+1})$. Next, the fluxes $F_{k+1/2}$, $G_{k+1/2\pm}$ depend only on the data u_i, S_i in cells $k-1$, k, $k+1$, where these values are the values of a discrete step function. Therefore, Proposition 3.4 gives the values of these fluxes, $F_{k+1/2} = F^a_{k+1/2}$, and similar values for $G_{k+1/2\pm}$. As a consequence, we see that u_k^{n+1}, u_{k+1}^{n+1} and S_k^{n+1}, S_{k+1}^{n+1} take the right values, and it is enough to consider these locations since we have invariance by translation of the index by two units. □

3.2 Smoothness indicator

Our main goal is here to introduce a particular choice of the values $E_{i+1/2\pm}$ defining the scheme described in Subsection 2.3. These values need to satisfy (2.31), (2.43), and a quite natural dependency of these coefficients is in u_{i-1}, u_i, u_{i+1}, S_{i-1}, S_i, S_{i+1}. In order to give an exact resolution of step functions in the linear case, we need also to fulfill the conditions of Proposition 3.4 for any discrete step function.

Let us first define the smoothness indicators $B_{i+1/2\pm}$,

$$\begin{aligned} B_{i+1/2-} &= \min\Big(S_{i-1} - S(u_{i-1}), S_i - S(u_i)\Big), \\ B_{i+1/2+} &= \min\Big(S_i - S(u_i), S_{i+1} - S(u_{i+1})\Big). \end{aligned} \tag{3.19}$$

The idea in (3.19) is that for a discrete step function, since $S_i - S(u_i)$ can be nonzero only for one value $i = k$, we have $B_{i+1/2\pm} = 0$ for all i. Therefore in general, $B_{i+1/2\pm}$ can be nonzero only where the approximate solution is roughly continuous, and this enables to prescribe a different behavior at these locations. Then, our choice is the following,

$$
\begin{aligned}
E_{i+1/2-} = &\Big(a_{i+1/2} + (1-\theta_{i+1/2-})(\frac{1}{\lambda_i} - a_{i+1/2})\Big)\Big(S_i - S(u_i)\Big)\\
&+ \psi(\lambda_i a_{i+1/2})\theta_{i+1/2-} a_{i+1/2} B_{i+1/2-},\\
E_{i+1/2+} = &\Big(-a_{i+1/2} + (1-\theta_{i+1/2+}) a_{i+1/2}\Big)\Big(S_i - S(u_i)\Big)\\
&+ \widehat{\psi}(\lambda_i a_{i+1/2})\theta_{i+1/2+} a_{i+1/2} B_{i+1/2+},
\end{aligned}
\tag{3.20}
$$

where $a_{i+1/2}$ is an approximation of $F'(u)$, $0 \leq \lambda_i a_{i+1/2} \leq 1$, and

$$
\begin{aligned}
\theta_{i+1/2-} &= \min\left(1, \frac{|u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1}))|}{\lambda_i |F(u_{i+1}) - F(u_i)|}\right),\\
\theta_{i+1/2+} &= \min\left(1, \frac{\lambda_i |F(u_{i+1}) - F(u_i)|}{|u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1}))|}\right),
\end{aligned}
\tag{3.21}
$$

except in the special case when $F(u_{i+1}) - F(u_i) = u_i - u_{i-1} - \lambda_i(F(u_i) - F(u_{i-1})) = 0$, where we let $\theta_{i+1/2-} = \theta_{i+1/2+} = 1$ for example. The functions ψ, $\widehat{\psi}$ are assumed to satisfy for $0 \leq z \leq 1$

$$
0 \leq \psi(z) \leq \frac{1}{z} - 1, \qquad 0 \leq \widehat{\psi}(z) \leq 1. \tag{3.22}
$$

Since $0 \leq \theta_{i+1/2-}, \theta_{i+1/2+} \leq 1$, the inequalities (2.31) and (2.43) obviously hold. The values (3.21) come from the ratio $X^\ell_{i+1/2}/X^r_{i+1/2}$ in (2.41), and enable to detect whether a possible discontinuity is going to stay in the cell i, or to travel to the next cell $i+1$. These definitions ensure that the conditions of Proposition 3.4 are satisfied (if the canonical value $a_{i+1/2} = a$ is chosen in the linear case).

Before choosing ψ and $\widehat{\psi}$, we first claim that it is natural to take

$$
\widehat{\psi}(z) = \frac{1-z}{z}\psi(1-z). \tag{3.23}
$$

Indeed, it is the condition for the scheme to satisfy an invariance property, which is that in the linear case $F(u) = au$ and with uniform grid, if we symmetrize the scheme in order to solve the equation with negative velocity $-a$, and then translate the result by one unit, we have to recover the same scheme, except that the Courant number λa is replaced by $1 - \lambda a$. This gives (3.23). Then, our choice is the following,

$$
\psi(z) = 4\frac{(1-z)^3}{2-z}, \qquad \widehat{\psi}(z) = 4\frac{(1-z)z^2}{1+z}. \tag{3.24}
$$

We have no justification of this choice except that it gives excellent accuracy in numerical computations. We need finally to define $a_{i+1/2}$, and a simple choice is

$$
a_{i+1/2} = \sup_{u \in [u_i, u_{i+1}]} F'(u). \tag{3.25}
$$

Other possibilities are proposed in Test 4.

Remark 3.2 For our scheme defined above, it is not possible to make an analysis of order of accuracy in the classical sense of expansions. Indeed, a natural thing would be to look for an expansion $S_i - S(u_i) \simeq \beta S''(u_{i+1}-u_i)(u_i-u_{i-1})$ for some smooth coefficient $\beta(t,x) \geq 0$. From the point of view of (2.8) and (2.23), one should have $\beta = 1/24$. However, we rather observe in practice values of β between 0.1 and 0.2 for smooth functions. For discontinuous functions, the value $S_i = \widehat{S}_i$ is rather observed, and we have from (2.44), if $(u_{i+1}-u_i)(u_i-u_{i-1}) > 0$,

$$\begin{aligned}\widehat{S}_i &= \frac{u_{i+1}-u_i}{u_{i+1}-u_{i-1}} S(u_{i-1}) + \frac{u_i - u_{i-1}}{u_{i+1}-u_{i-1}} S(u_{i+1}) \\ &= S(u_i) + \frac{1}{2} S'' \left[\frac{u_{i+1}-u_i}{u_{i+1}-u_{i-1}} (u_i - u_{i-1})^2 + \frac{u_i - u_{i-1}}{u_{i+1}-u_{i-1}} (u_{i+1}-u_i)^2 \right] \\ &\quad + o(u_{i+1}-u_i)^2 + o(u_i - u_{i-1})^2 \\ &= S(u_i) + \frac{1}{2} S''(u_{i+1}-u_i)(u_i-u_{i-1}) + o(u_{i+1}-u_i)^2 + o(u_i-u_{i-1})^2,\end{aligned} \tag{3.26}$$

which yields $\beta = 1/2$ (and indeed there is no correction in (3.26) if $S(u) = u^2/2$). Also, a common behavior is to have a coefficient β quickly oscillating between 0 and 1/2.

4 Numerical results

The numerical results presented here intend to provide a comparison between the antidiffusive scheme $F^a_{i+1/2}$ of [4], [8], [9], our antidiffusive entropy scheme with the definitions of Subsection 3.2, and a classical second-order ENO scheme.

For our scheme, we choose $S(u) = u^2/2$, and the entropy indicators S_i are initialized as $S_i = S(u_i)$. Other choices are possible, like $S_i = \widehat{S}_i$, but it does not make a big difference in the results. We have to mention that at the practical level, when applying the limitation of the numerical flux according to (2.32), care must be taken when dividing by $K_{i+1/2}$. Indeed, if $K_{i+1/2}$ and $E_{i+1/2-} + E_{i+1/2+}$ are both small, one has to keep the antidiffusive flux $F^a_{i+1/2}$, otherwise we observe instabilities, especially in the case when $\theta_{i+1/2-} = \theta_{i+1/2+} = 1$.

In all tests, $x \in (0,1)$, periodic boundary conditions are applied, and the mesh is uniform with 48 cells. An arbitrary CFL condition of 0.32179 is used.

Test 1: linear advection with smooth data

We solve a linear equation $F(u) = au$, $a = 1$, with initial data $u^0(x) = \sin(2\pi x)$, $0 < x < 1$. The results are plotted on Figures 3 and 4 for $t = 1$ and $t = 100$ respectively. We observe in Figure 3 that our antidiffusive entropy scheme gives results comparable to the ENO scheme, avoiding the staircases of the antidiffusive flux. For large times (Figure 4), the diffusion is lower than that of ENO, the extrema are better resolved. For the same test, Figure 7 represents the coefficient β discussed in Remark 3.2,

$$\beta_i = \begin{cases} \dfrac{S_i - S(u_i)}{(u_{i+1} - u_i)(u_i - u_{i-1})} & \text{if } (u_{i+1} - u_i)(u_i - u_{i-1}) > 0, \\ 0 \quad \text{otherwise.} \end{cases} \tag{4.1}$$

Test 2: linear advection with unsmooth data

We still take $F(u) = au$, $a = 1$, but with the initial data of Harten,

$$u^0(x) = \begin{cases} -z\sin(1.5\,\pi z^2) & \text{if } z < -1/3, \\ |\sin(2\pi z)| & \text{if } |z| \leq 1/3, \\ 2z - 1 - \sin(3\pi z)/6 & \text{otherwise,} \end{cases} \tag{4.2}$$

where $z = 2x - 1$, $-1 < z < 1$. In this test only, we take 0.8 for the CFL number. The results are plotted at $t = 1$ on Figure 5 and at $t = 100$ on Figure 6. Again the antidiffusive entropy scheme gives accurate results without undesirable staircases. We observe on Figure 6 that some dissipation appears after a long time, that prevents from keeping the extrema. It is however much better than the ENO scheme. The coefficient β of (4.1) is plotted on Figure 8 for $t = 300$.

Test 3: Burgers equation

In this test we take $F(u) = u^2/2$, with initial data $u^0(x) = 1$ if $0 < x \leq 1/4$, $u^0(x) = 2$ if $1/4 < x < 1$. The final time is $t = 1/6$, and the solution develops a shock and a rarefaction. As shown on Figure 9, the antidiffusive and antidiffusive entropy schemes give the exact shock, which is here located at the interface of two cells. Notice that this holds with the choice (3.25) of the interface velocity $a_{i+1/2}$, which is not the one of the Rankine-Hugoniot relation. Also, this works for any value of the CFL number. The rarefaction is not computed by the antidiffusive scheme, which gives a nonentropy discontinuity, but the antidiffusive entropy scheme gives a sharp computation of this rarefaction, there is absolutely no smearing of the forward and backward fronts.

Test 4: Non convex flux

In this test we take a non convex flux

$$F(u) = \begin{cases} \dfrac{1}{5 - 4u} & \text{if } u \leq 1, \\ \dfrac{4}{5 - u} & \text{if } 1 \leq u < 5. \end{cases} \tag{4.3}$$

The initial data is $u^0(x) = 2$ if $0 < x \leq 1/4$, $u^0(x) = 0$ if $1/4 < x < 1$, and the result is represented on Figure 10 for $t = 5/8$. Since the flux is not convex, the entropy inequality (2.3) does not select a unique solution. We experiment several choices of the interface velocity $a_{i+1/2}$ in (3.20),

$$\begin{aligned}
a_{i+1/2} &= \max(F'(u_i), F'(u_{i+1})),\\
a_{i+1/2} &= \sup F'(u),\ \ u \in [u_i, u_{i+1}],\\
a_{i+1/2} &= \max(F'(u_{i-1}), F'(u_i), F'(u_{i+1})),\\
a_{i+1/2} &= \sup F'(u),\ \ u \in [u_{i+1}, u_{i-1}].
\end{aligned} \tag{4.4}$$

The last choice seems to be the most diffusive, and gives the Kruzkov solution.

Numerical order of accuracy

We indicate in Table 1 the L^1 errors for Tests 1, 2, 3 (with $t = 1$ for Tests 1, 2), when the number of cells is successively doubled. We observe roughly an order of accuracy of 2 in Test 1, an order 1.5 in Test 2, and an order 1 in Test 3. For comparison, we provide the L^1 errors for the ENO scheme in Table 2.

Table 1: L^1 error, Antidiffusive Entropy scheme

Number of cells	Test 1	Test 2	Test 3
48	3.89E-3	2.73E-2	3.36E-3
96	9.77E-4	8.73E-3	1.78E-3
192	2.21E-4	2.62E-3	1.03E-3
384	4.66E-5	8.99E-4	5.15E-4

Table 2: L^1 error, ENO scheme

Number of cells	Test 1	Test 2	Test 3
48	1.48E-3	7.71E-2	1.91E-2
96	3.81E-4	4.02E-2	9.69E-3
192	9.44E-5	2.26E-2	4.87E-3
384	2.37E-5	1.30E-2	2.44E-3

Figure 3:

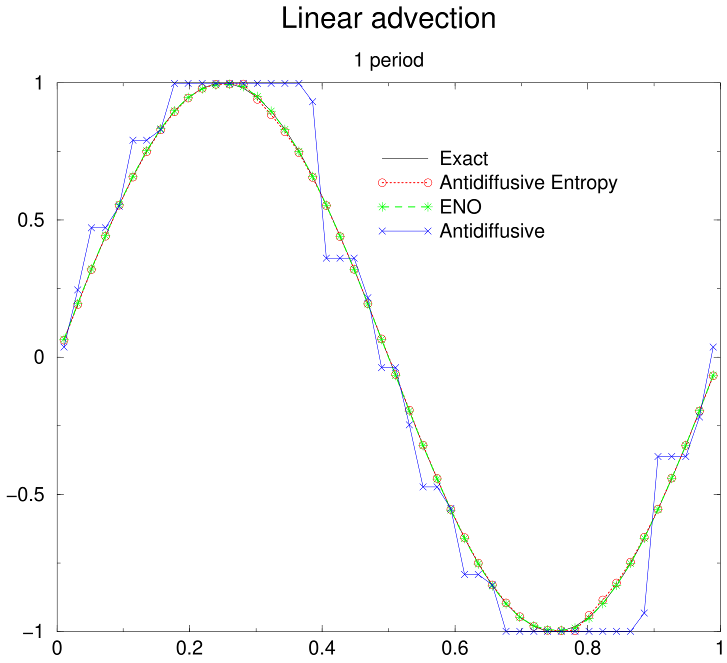


Figure 4:

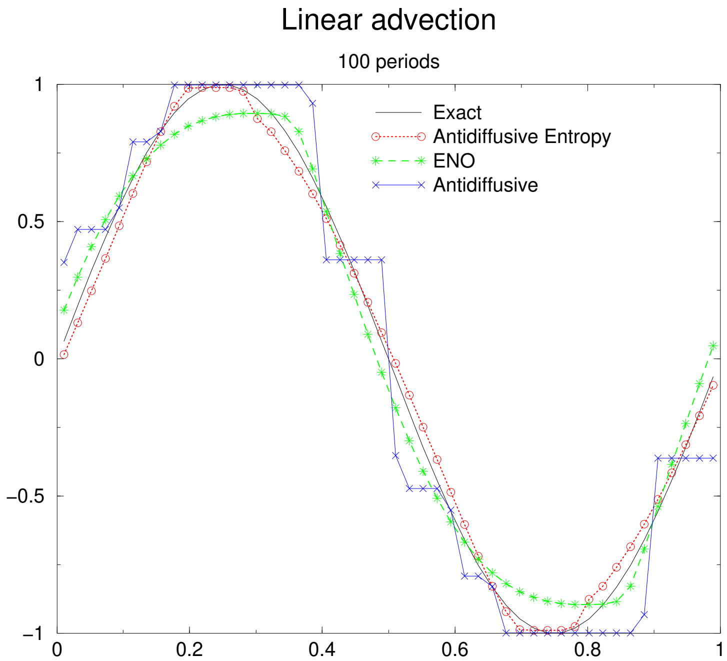


Figure 5:

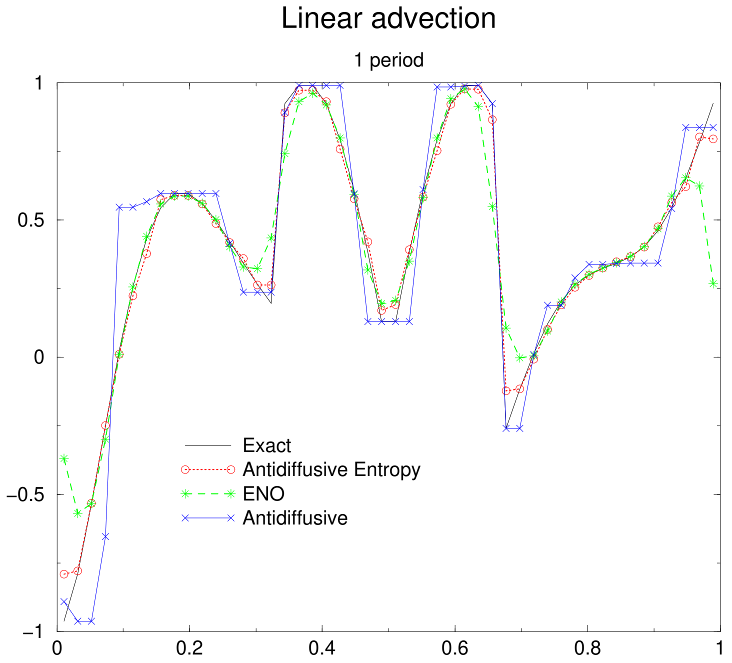


Figure 6:

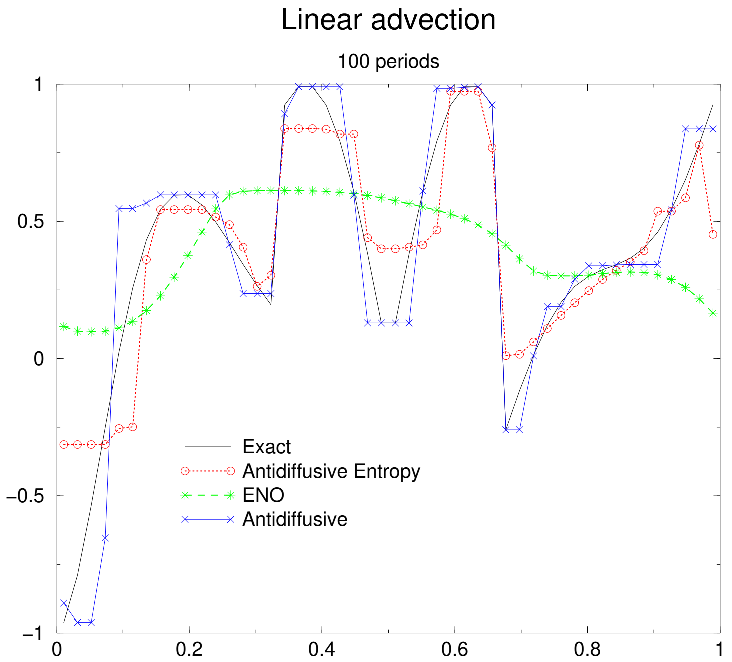


Figure 7:

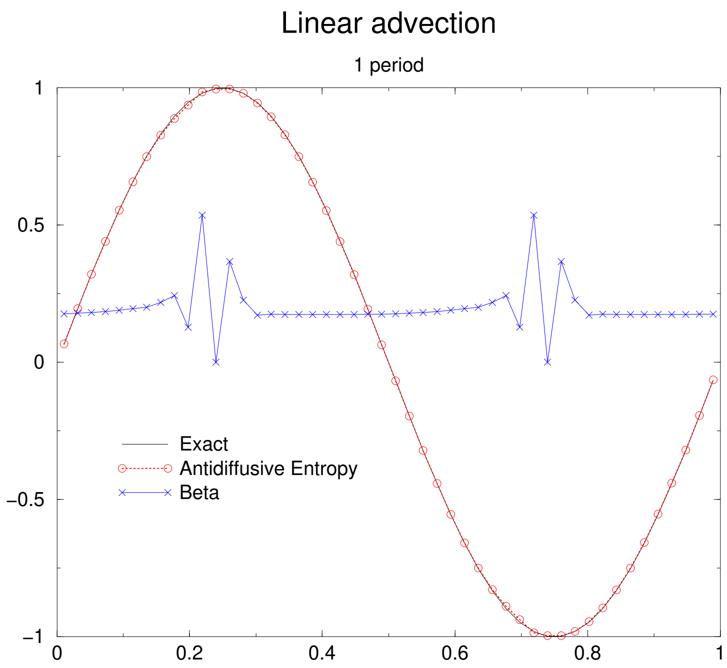


Figure 8:

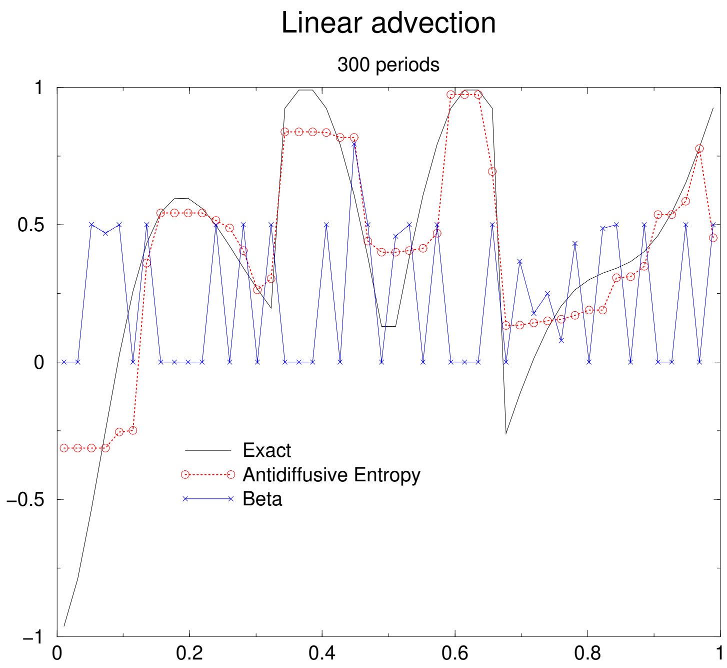


Figure 9:

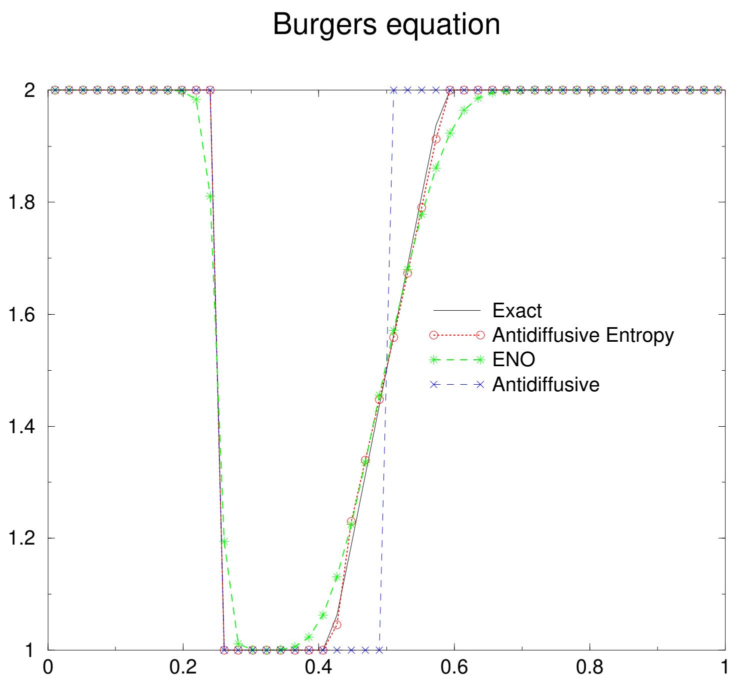
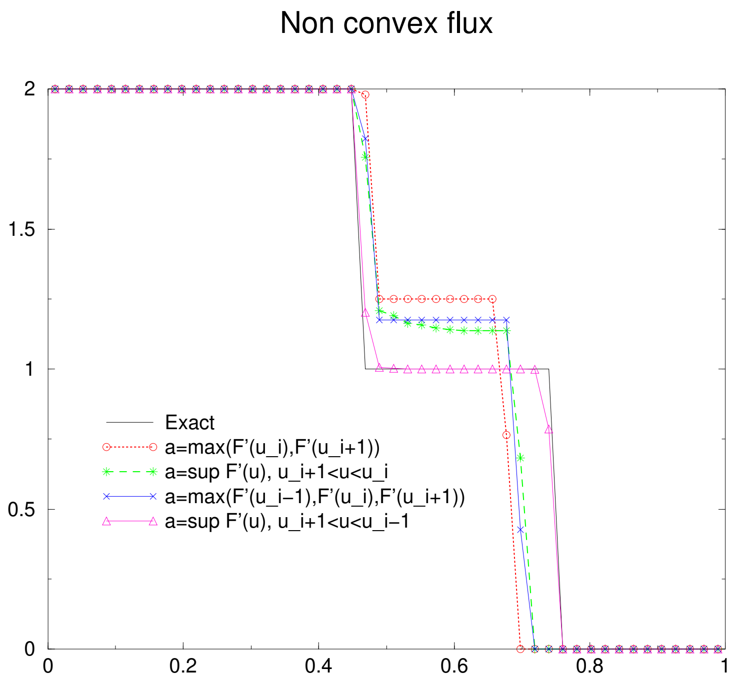


Figure 10:



References

[1] F. Alouges, F. De Vuyst, G. Le Coq, E. Lorin, Un procédé de réduction de la diffusion numérique des schémas à différence de flux d'ordre un pour les systèmes hyperboliques non linéaires, C. R. Acad. Sci. Paris, Ser. I 335 (2002), 627-632.

[2] F. Bouchut, Ch. Bourdarias, B. Perthame, A MUSCL method satisfying all the numerical entropy inequalities, Math. of Comp. 65 (1996), no. 216, 1439-1461.

[3] F. Coquel, P. LeFloch, An entropy satisfying MUSCL scheme for systems of conservation laws, Numer. Math. 74 (1996), 1-33.

[4] B. Després, F. Lagoutière, Contact discontinuity capturing schemes for linear advection and compressible gas dynamics, J. Sci. Comput. 16 (2001), 479-524.

[5] E. Godlewski, P.-A. Raviart, *Hyperbolic systems of conservation laws*, Mathématiques et Applications 3/4, Ellipses, Paris, 1991.

[6] A. Harten, The artificial compression method for computation of shocks and contact discontinuities. III. Self-adjusting hybrid schemes, Math. of Comp. 32 (1978), no. 142, 363-389.

[7] A. Harten, ENO schemes with subcell resolution, J. Comp. Phys. 83 (1989), 148-184.

[8] F. Lagoutière, Modélisation mathématique et résolution numérique de problèmes de fluides compressibles à plusieurs constituants, PhD thesis, Université Paris VI, 2000.

[9] F. Lagoutière, Numerical resolution of scalar convex equations: explicit stability, entropy and convergence conditions, CEMRACS 1999 (Orsay), 183-199 (electronic), ESAIM Proc. 10, Soc. Math. Appl. Indust., Paris, 1999.

[10] P.K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, SIAM J. Numer. Anal. 21 (1984), 995-1011.

[11] H. Yang, An artificial compression method for ENO schemes: the slope modification method, J. Comp. Phys. 89 (1990), 125-160.