Non-dissipative entropic discontinuous reconstruction schemes for hyperbolic conservation laws.

Frédéric Lagoutière

1 Introduction

This paper deals with the finite volume discretization of scalar partial differential equations in dimension 1 in space,

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

where $f \in \mathcal{C}^1(\mathbb{R})$, with initial condition

$$u(0, \cdot) = u^0(\cdot) \in L^\infty(\mathbb{R}). \tag{2}$$

For the sake of simplicity and as in [2], we restrict the discussion to the case without sonic points: $f'(u) > 0$ for every u.

We are interested in an *entropy* solution of this problem, that is to say in a weak solution that satisfies the additional partial differential inequation

$$\partial_t S(u) + \partial_x G(u) \leq 0, \quad t \in \mathbb{R}^+, x \in \mathbb{R}, \tag{3}$$

for one entropy-entropy flux pair (S, G), i.e. a pair of $C^1(\mathbb{R})$ functions (S, G) such that S is convex and $G' = S'f'$. It is known that this solution is unique if f and S are strictly convex and that this solution is the Krushkov entropy solution, cf. [8, 14] for example. This solution belongs to $L^\infty(]0, T[\times\mathbb{R})$ $\forall T \in \mathbb{R}^+$ and verifies furthermore $u(t, \cdot) \in BV(\mathbb{R})$ $\forall t \in \mathbb{R}^+$ if $u^0 \in BV(\mathbb{R})$.

We are here concerned with the numerical approximation of these entropy solutions in the standard framework of finite volume schemes. Let $\Delta x \in \mathbb{R}$ and $\Delta t \in \mathbb{R}$ be given positive reals. We replace equation (1) with the discrete in time and space equation

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}(f_{j+1/2}^n - f_{j-1/2}^n) \quad \forall n \in \mathbb{N}, \forall j \in \mathbb{Z},$$

and replace (2) with the with numerical initial condition $(u_j^0)_{j\in\mathbb{Z}}$,

$$u_j^0 = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u^0(x)\,dx \quad \forall j \in \mathbb{Z}.$$

The numerical fluxes $f_{j+1/2}^n$ $(j \in \mathbb{Z}, n \in \mathbb{N})$ have to be computed in such a manner that the numerical approximation

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

converges towards the (an) entropy solution of (1,2) as Δt and Δx tend to 0 (in a norm to be specified). Our previous work [16] presented some general convergence conditions. The present paper uses this analysis to derive some new antidissipative convergent schemes. As in [16], we consider reconstruction schemes. The originality of this work is to analyze *discontinuous reconstruction* schemes, which means that the reconstruction of the solution is discontinuous in each cell. We do not intend to obtain a high-order scheme but a scheme that is precise on discontinuous solutions. This seems not to have been the aim in previous works on reconstruction schemes, usually looking for second order or high order approximations. See for example the geometric limiters theory (slope limiters) in [9], in relation with [23] for the flux limiter theory. A general reference on discrete entropy conditions is [10]. The paper [3] focus on entropy conditions for second order geometric reconstruction schemes. MUSCL schemes and entropy conditions are studied in [5] and [20], and [18] deals with high order algorithms. Let us also recall [22] for a study of convergence and order in general.

The paper organizes as follows.

We first recall (section 2) the framework and the stability and convergence results obtained in [16] for reconstruction schemes. Reconstruction schemes can be decomposed into 3 steps. The first step consists in reconstructing the given constant-by-cell solution; the second step is a resolution of the exact PDE with the reconstructed solution; the last step is on projection on the mesh.

Equipped with this, we explore a new class of reconstruction schemes: discontinuous reconstruction schemes. Section 3 is thus devoted to the transposition of the former general results to this particular reconstruction schemes. Discontinuous reconstruction schemes are based on the reconstruction of the approximate solution as a discontinuous-in-cell functions with one discontinuity in each cell and one constant value on the left and one other

constant value on the right of the discontinuity. The position of the discontinuity is not fixed (in particular, it is not a priori the middle of the cell). The goal of such a reconstruction is to compute good approximate *discontinuous* solutions and to avoid numerical dissipation. In this aim, we are led, in a first stage, to choose the reconstruction that maximizes the total variation of the solution among reconstructions guaranteeing stability and decreasing of the entropy. For this particular scheme, we prove that shocks are computed exactly. We finally analyze more precisely the cases of linear advection and the case of a convex flux: Burgers' equation. For the sake of computing simplicity, we then explore a way to perform the previous computations in an approximate way, replacing the exact computation of the second step of the algorithm by an approximate one. This simplified procedure has to guarantee the same properties. We compare the results for different reconstructions on Burgers' equation and on an equation with a non-convex flux.

These different cases are illustrated with various numerical results.

2 Framework

This section is devoted to recalling the main ingredients of [16].

As mentioned in the introduction, we consider finite volume approximations of (1) of the form

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}(f_{j+1/2}^n - f_{j-1/2}^n) \quad \forall j \in \mathbb{Z}, \forall n \in \mathbb{N}, \tag{4}$$

where u_j^n is ought to represent the value of solution u in space cell $C_j = [(j-1/2)\Delta x, (j+1/2)\Delta x]$. The numerical initial condition u_j^0 is given by

$$u_j^0 = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u^0(x)\,dx \quad \forall j \in \mathbb{Z}. \tag{5}$$

We propose to compute the numerical fluxes $(f_{j+1/2}^n)_{n\in\mathbb{N},j\in\mathbb{Z}}$ using a three-step procedure:

- given a constant-by-cell function, compute a reconstructed function that contains more details,
- compute the exact (entropy) solution at time Δt of (1) with the reconstructed function as initial condition,

- "project" this exact solution on the mesh in order to obtain a constant-by-cell function for the following time step.

Note that the last two steps are equivalent to computing the *fluxes* of the exact solution, which shows the finite volume form of the algorithm: see equation (7).

Each of these steps can be represented by an operator: we shall call $\mathcal{R}$, $\mathcal{E}$ and $\mathcal{P}$ respectively the reconstruction, the exact and the projection operators. Let us provide a more precise definition of them.

Definition 1 1 Let $u : \mathbb{R} \longrightarrow \mathbb{R}$ be a constant-by-cell function.

$$\mathcal{R}u : \mathbb{R} \longrightarrow \mathbb{R} \text{ denotes the reconstruction of } u.$$

2 Let $t \in \mathbb{R}$ and $u : \mathbb{R} \longrightarrow \mathbb{R}$ be a function in $L^\infty(\mathbb{R})$.

$$\mathcal{E}(t)u : \mathbb{R} \longrightarrow \mathbb{R} \text{ denotes the exact entropy solution at time } t \text{ of equation (1) with initial condition } u.$$

3 Let $u : \mathbb{R} \longrightarrow \mathbb{R}$ be a function in $L^\infty(\mathbb{R})$.

$$\mathcal{P}u : \mathbb{R} \longrightarrow \mathbb{R} \text{ denotes the "projection" of } u \text{ on the mesh:}$$

$$\mathcal{P}u(x) = \sum_{j\in\mathbb{Z}} u_j \mathbb{1}_{[(j-1/2)\Delta x,(j+1/2)\Delta x)}(x)$$

with

$$u_j = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u(x)\,dx \quad \forall j \in \mathbb{Z}.$$

We now define the approximate solution $\overline{u}^n : \mathbb{R} \longrightarrow \mathbb{R}$ at time step n by

$$\overline{u}^n(x) = \sum_{j\in\mathbb{Z}} u_j^n \mathbb{1}_{[(j-1/2)\Delta x,(j+1/2)\Delta x)}(x).$$

The reconstruction scheme is defined by

$$\overline{u}^{n+1} = \mathcal{P}\mathcal{E}(\Delta t)\mathcal{R}\overline{u}^n. \tag{6}$$

The reconstruction scheme can be viewed as a finite volume one by solving the exact computation and the projection in one, taking as numerical fluxes in equation (4)

$$f_{j+1/2}^n = \frac{1}{\Delta t}\int_0^{\Delta t} f\left(\mathcal{E}(s)\mathcal{R}\overline{u}^n((j+1/2)\Delta x)\right)\,ds \quad \forall n \in \mathbb{N}, \forall j \in \mathbb{Z}. \tag{7}$$

In this paper, we essentially focus on the reconstruction operator. We at first assume to be able to compute the exact solution and the projection. This of course strongly depends on the reconstruction and on the flux of the PDE; it will be done in the case of linear advection and Burgers' equation. For arbitrary fluxes, we propose at the end of the paper to replace the exact computation by an approximate one.

The equivalent formulations (6) and (4,7) will be alternatively used.

The Godunov scheme corresponds to $\mathcal{R}u = u$ (no reconstruction). If $\mathcal{R}u$ is an affine-by-cell function for every constant-by-cell function u, the resulting scheme is a MUSCL scheme; a precise study of such a MUSCL scheme has been performed in [3].

We now briefly recall the principal results shown in [16] for general reconstructions. These results will be applied to the particular case of discontinuous reconstructions in section 3.

2.1 Conservativity

We consider only conservative reconstructions such that

$$\frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} \mathcal{R}u(x)\,dx = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u(x)\,dx = \mathcal{P}u(j\Delta x). \qquad (8)$$

The exact operator $\mathcal{E}(t)$ and the projection $\mathcal{P}$ being conservative, the scheme defined by (6) is consequently conservative.

2.2 L^∞-decreasing and decreasing of the total variation

We here recall an L^∞ stability result extracted form [16]. Let us introduce the notations

$$\left\{ \begin{array}{l} m = \inf_{j\in\mathbb{Z}} u_j^0, \\ M = \sup_{j\in\mathbb{Z}} u_j^0. \end{array} \right. \qquad (9)$$

Proposition 1 Assume the CFL (Courant-Friedrichs-Lewy) condition $\max_{u\in[m,M]} f'(u)\Delta t \leq \Delta x$ is fulfilled. Assume the reconstructed solution $\mathcal{R}\overline{u}^n$ verifies, $\forall n \in \mathbb{N}$, $\forall j \in \mathbb{Z}$,

$$\min(u_{j-1}^n, u_j^n) \leq \frac{1}{\Delta x}\int_{(j-1/2-\theta)\Delta x}^{(j+1/2-\theta)\Delta x} \mathcal{R}\overline{u}^n(x)\,dx \leq \max(u_{j-1}^n, u_j^n) \qquad \forall\theta \in [0,1]. \quad (10)$$

Then, the scheme given by (4, 7), or (6), is L^∞-decreasing and TVD.

Recall that in the whole paper, f' is supposed to be positive.

The proof relies on a stability property of $\frac{1}{\Delta x}\mathbb{1}_{[-\Delta x/2,\Delta x/2]} * u(t,x) = \frac{1}{\Delta x}\int_{x-\Delta x/2}^{x+\Delta x/2} u(t,x)\,dx$ and is done in [16]. This result gives a constraint on the reconstruction for the scheme to be stable.

2.3 Numerical entropy inequalities

It is known that the TVD property does not imply the convergence toward the entropy solution of (1,2). To ensure this, a usual criterion is the existence of *numerical entropy fluxes*. Here is the definition from [3].

Definition 2 Let (S,G) be an entropy-entropy flux pair. It is said that scheme (4, 7) has discrete entropy fluxes relatively to (S,G) if and only if $\forall (u_j^n)_{j\in\mathbb{Z}}\ \exists (G_{j+1/2}^n)_{j\in\mathbb{Z}}$ such that

- $G_{j+1/2}^n$ is consistent with G (in the classical sense of finite volume);
- $$S_j^{n+1} \leq S_j^n - \frac{\Delta t}{\Delta x}\left(G_{j+1/2}^n - G_{j-/2}^n\right) \forall j \in \mathbb{Z}, n \in \mathbb{N} \tag{11}$$
$$\text{with} \quad S_j^n = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} S(\mathcal{R}\overline{u}^n(x))\,dx. \tag{12}$$

It seems reasonable, the exact resolution being used, to take, as entropy flux $G_{j+1/2}^n$, the exact flux (similarly to eq. (7))

$$G_{j+1/2}^n = \frac{1}{\Delta t}\int_0^{\Delta t} G\left(\mathcal{E}(s)\mathcal{R}\overline{u}^n((j+1/2)\Delta x)\right)\,ds \quad \forall n \in \mathbb{N}, \forall j \in \mathbb{Z}. \tag{13}$$

This choice will be done in the following.

With these numerical entropy fluxes specified, equation (11) acts like a new constraint on the reconstruction procedure.

Remark 1 Thus the fluxes $f_{j+1/2}^n$ depend on the values u_j^n at time step n and on quantities

$$S_j^{n-1} - \frac{\Delta t}{\Delta x}\left(G_{j+1/2}^{n-1} - G_{j-1/2}^{n-1}\right).$$

To compute the fluxes by following a supplementary unknown, the entropy, is already the key ingredient of the non-dissipative schemes of [2]. The non-dissipative reservoir scheme of [1] also relies on the computation of additional variables (not entropies in this case, but a reservoir and a CFL counter).

3 Discontinuous-in-cell reconstruction schemes

We now propose a new kind of reconstruction schemes that enters the stable schemes derived in [16]. These schemes are derived to capture discontinuous solutions of linear and non-linear scalar equations. For this purpose, we reconstruct the unknown inside each cell as a discontinuous function with one and only one discontinuity separating two constant values (in order the analysis to be simple). The adopted framework is thus different from the classical one of reconstruction schemes whose goal is to elaborate some high order approximate solutions: [3], [5], [9], [25] using a piecewise affine reconstruction, [4], [18] using piecewise polynomials (of desired order) reconstructions. The discontinuous reconstruction schemes are comparable to the so-called "sub-cell resolution schemes" developed in [13] by Harten. The difference is that we here use only a first order reconstruction (which is not coupled with a polynomial smooth one) and that this reconstruction is done in every cell. Furthermore, we derive some explicit stability and entropy conditions on this reconstruction. From an other point of view, such discontinuous reconstructions aiming at preserving discontinuities are related to moving mesh methods. Indeed, putting a discontinuity in the reconstructed solution can be viewed as putting an additional cell. We refer to [11] and [21] for this type of algorithms.

Let us describe the reconstruction operator $\mathcal{R}$. Given a value u_j^n, we reconstruct it as a function taking the value $\overline{u_j^n}^l$ on an interval of length $d_j^n \Delta x \in [0, \Delta x]$ on the left side of the cell and the value $\overline{u_j^n}^r$ on an interval of length $(1 - d_j^n)\Delta x \in [0, \Delta x]$ on the right side of the cell (see figure 1):

$$\mathcal{R}\overline{u}^n(x) = \begin{cases} \overline{u_j^n}^l \text{ for } x \in [(j-1/2)\Delta x, (j-1/2+d_j^n)\Delta x), \\ \overline{u_j^n}^r \text{ for } x \in [(j-1/2+d_j^n)\Delta x, (j+1/2)\Delta x), \end{cases} \quad j \in \mathbb{Z}. \quad (14)$$

where $d_j^n \in [0,1]$, $\overline{u_j^n}^l$ and $\overline{u_j^n}^r$ are to be specified. In all the following, $\mathcal{R}\overline{u}^n(x)$ is of the form (14). The conservativity, stability and entropy requirements will provide some constraints on the 3 parameters just introduced.

3.1 Conservativity

This section is the translation of section 2.1 in terms of discontinuous-in-cell reconstructions. The conservativity constraint (8) reads

$$d_j^n \overline{u_j^n}^l + (1 - d_j^n)\overline{u_j^n}^r = u_j^n. \quad (15)$$

Thus among the 3 parameters, only 2 are free.

3.2 L^∞ and total variation-decreasing

We here examine a simple way to adapt results of section 2.2 to discontinuous-in-cell reconstructions. We find in theorem 1 a condition that ensures the stability. In the following, $\lfloor a, b \rceil$ denotes $\mathrm{conv}(\{a\}, \{b\})$, i.e. $[a, b]$ if $a \leq b$, $[b, a]$ if $b \leq a$.

Theorem 1 Assume that $\max_{u \in [m,M]} f'(u)\Delta t \leq \Delta x$. Assume that the reconstruction operation is conservative (eq. (15)) and that $\mathcal{R}\overline{u}^n(x)$ verifies

$$\overline{u_j^n}^l \in \lfloor u_{j-1}^n, u_j^n \rceil \quad \text{and} \quad \overline{u_j^n}^r \in \lfloor u_j^n, u_{j+1}^n \rceil \quad \forall j \in \mathbb{Z} \quad \text{(see fig. 1)} \qquad (16)$$

Then, the scheme (6, 14) is L^∞-decreasing and TVD. As Δt and Δx converge to 0, the numerical solution $\overline{u}_{\Delta x}^{\Delta t}$ converges in $L^\infty(]0, +\infty[, L^1(\mathbb{R}))$, up to a subsequence, toward a weak solution of the initial value problem (1-2).

Remark 2 Condition (16) together with (15) implies that there is no reconstruction on a local extremum:

$$(u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) \leq 0 \Longrightarrow \overline{u_j^n}^l = \overline{u_j^n}^r = u_j^n.$$

Remark 3 The result in theorem 1 is not trivial. Indeed, the condition there proposed for the reconstruction does not at all imply that the total variation of the reconstruction is smaller than the one of the initial function u. Consider for example the monotone case $u_j^n < u_{j+1}^n$ and the reconstruction

$$\begin{cases} \overline{u_j^n}^l = u_{j-1}^n, \\ \overline{u_j^n}^r = u_{j+1}^n, \\ d_j^n = \frac{u_{j+1}^n - u_j^n}{u_{j+1}^n - u_{j-1}^n}, \end{cases}$$

(this choice will be studied in section 3.4). In this case, the total variation of the reconstructed solution is three times greater than the one of u. What the theorem states is that after an exact resolution and a projection, the total variation will be reduced enough.

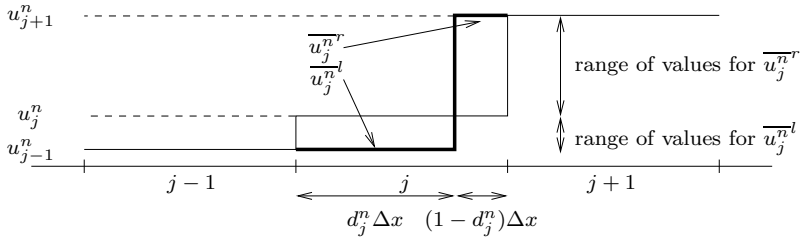

Figure 1: Discontinuous reconstruction satisfying conservativity and stability requirements.

Proof The principle of the proof is to show that condition (16) implies (10). Let us examine the convolution of the discontinuous reconstruction by $\frac{1}{\Delta x}\mathbb{1}_{[-\Delta x/2,\Delta x/2]}$ and let us denote it $[\mathcal{R}\overline{u}^n]_{\Delta x}$:

$$[\mathcal{R}\overline{u}^n]_{\Delta x}\left((j-\theta)\Delta x\right) = \frac{1}{\Delta x}\int_{(j-1/2-\theta)\Delta x}^{(j+1/2-\theta)\Delta x} \mathcal{R}\overline{u}^n(x)\,dx, \theta \in [0,1].$$

It is a convex combination of the 4 local values of the reconstruction $\overline{u^n_{j-1}}^l$, $\overline{u^n_{j-1}}^r$, $\overline{u^n_j}^l$ and $\overline{u^n_j}^r$:

$$[\mathcal{R}\overline{u}^n]_{\Delta x}((j-\theta)\Delta x) = \alpha_1\overline{u^n_{j-1}}^l + \alpha_2\overline{u^n_{j-1}}^r + \alpha_3\overline{u^n_j}^l + \alpha_4\overline{u^n_j}^r \qquad (17)$$

where the coefficients are given by

$$\begin{cases} \alpha_1 = \max(0, d^n_{j-1} + \theta - 1), \\ \alpha_2 = \min(\theta, 1 - d^n_{j-1}), \\ \alpha_3 = \min(d^n_j, 1-\theta), \\ \alpha_4 = \max(0, 1 - d^n_j - \theta). \end{cases}$$

Let us assume that $u^n_{j-1} \leq u^n_j$. Then, $u^n_{j-1} \leq \overline{u^n_{j-1}}^r$ and the conservativity equation,

$$d^n_{j-1}\overline{u^n_{j-1}}^l + (1-d^n_{j-1})\overline{u^n_{j-1}}^r = u^n_{j-1},$$

implies that $\overline{u^n_{j-1}}^l \leq u^n_{j-1}$, so that $\overline{u^n_{j-1}}^l \leq \overline{u^n_{j-1}}^r$ and $\overline{u^n_{j-1}}^l \leq \overline{u^n_j}^l$, and finally $\overline{u^n_{j-1}}^l \leq \overline{u^n_j}^r$. In conclusion, $\overline{u^n_{j-1}}^l \leq \min(\overline{u^n_{j-1}}^r, \overline{u^n_j}^l, \overline{u^n_j}^r)$, so that $[\mathcal{R}\overline{u}^n]_{\Delta x}((j-\theta)\Delta x)$ verifies, by equation (17),

$$[\mathcal{R}\overline{u}^n]_{\Delta x}((j-\theta)\Delta x) \geq [\mathcal{R}\overline{u}^n]_{\Delta x}((j-1)\Delta x) = u^n_{j-1} \quad \forall\theta \in [0,1].$$

One can prove, on the same manner, that $[\mathcal{R}\overline{u}^n]_{\Delta x}((j-\theta)\Delta x) \leq u^n_j\ \forall\theta \in [0,1]$. A similar result is proved under the assumption $u^n_{j-1} \geq u^n_j$, leading to

the inequalities (10). Proposition 1 ends the proof of stability. The convergence toward a weak solution of the problem is a classical result combining compactness in L^1 and the Lax-Wendroff theorem, see [17].

3.3 Numerical entropy inequalities

3.3.1 Decrease of one strictly convex entropy

Let (S, G) be an entropy-entropy flux pair. We consider the entropy condition (11), where S_j^n is simply given by

$$S_j^n = d_j^n S(\overline{u_j^n}^l) + (1 - d_j^n) S(\overline{u_j^n}^r).$$

Thus, the three parameters d_j^n, $\overline{u_j^n}^l$ and $\overline{u_j^n}^r$ are to be chosen such that they allow the existence of numerical entropy fluxes with equation (11). Some particular choices will be done in sections 3.4, 3.8 and 3.9.

The next sections are devoted to the choice of schemes that satisfy some of the previous conditions, for some particular scalar equations: advection equation, Burgers' equation and, finally, a non-convex example.

3.4 A non-entropic scheme for linear advection with constant velocity

Here is considered the linear flux $f(u) = au$ where $a > 0$ is a given constant. The weak solution to the Cauchy problem is unique and given by $u(t, x) = u^0(x - at)$, so that no entropy condition is needed. Thus, in a first step, we propose a scheme that does not ensure the decreasing of any entropy. The second scheme we propose is entropic for the quadratic entropy.

The ideas underlying the present study is to derive some non-dissipative schemes. In the classical "upwind" scheme, the dissipation is due to the projection operation $\mathcal{P}$: there is no reconstruction procedure (to say it differently, $\overline{u_j^n}^l = \overline{u_j^n}^r = u_j^n$ and the value of d_j^n does not matter, $\forall j \in \mathbb{Z}$, $\forall n \in \mathbb{N}$). One natural idea is then to choose the reconstruction in such a manner that it maximizes the variation inside each cell, under the conservativity and stability constraints (15) and (16). This defines all the parameters

of the discontinuous reconstruction as

$$\left.\begin{array}{l} \overline{u_j^n}^l = u_{j-1}^n, \\ \overline{u_j^n}^r = u_{j+1}^n, \\ d_j^n = \dfrac{u_{j+1}^n - u_j^n}{u_{j+1}^n - u_{j-1}^n}. \end{array}\right\} \quad \text{if } (u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) > 0, \qquad (18)$$

$$\left.\begin{array}{l} \overline{u_j^n}^l = \overline{u_j^n}^r = u_j^n, \\ d_j^n \text{ does not matter} \end{array}\right\} \quad \text{if } (u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) \leq 0,$$

The following result establishes a link with other theories.

Proposition 2 Assume the CFL condition $a\Delta t/\Delta x \leq 1$. The scheme (6, 14, 18) with $f(u) = au$ and $a > 0$ is equivalent to the Ultrabee limiter (cf. [24]) and to the limited downwind scheme (cf. [6]).

Proof We prove the equivalence with the limited downwind scheme. Equivalence of the Ultrabee limiter and the limited downwind scheme is proved in [6]. For this task we recall the limited downwind scheme. It is based on formulation (4) ($f(u) = au$, $a > 0$) and relies on the intervals $I_{j+1/2}^n = [ab_{j+1/2}^n, aB_{j+1/2}^n]$ with

$$b_{j+1/2}^n = \max(m_{j+1}^n, M_j^n + \frac{\Delta x}{a\Delta t}(u_j^n - M_j^n)),$$
$$B_{j+1/2}^n = \min(M_{j+1}^n, m_j^n + \frac{\Delta x}{a\Delta t}(u_j^n - m_j^n)),$$

where

$$m_j^n = \min(u_{j-1}^n, u_j^n),$$
$$M_j^n = \max(u_{j-1}^n, u_j^n).$$

These intervals are shown to be such that if $f_{j+1/2}^n \in I_{j+1/2}^n \ \forall j \in \mathbb{Z}$ and if $a\Delta t/\Delta x \leq 1$, $u_j^{n+1} \in [m_j^n, M_j^n] \ \forall j \in \mathbb{Z}$. This condition is known to be sufficient to have the TVD property, by a classical argument of incremental analysis of Le Roux in [19] and Harten in [12]. The limited downwind scheme is obtained taking $f_{j+1/2}^n$ as close as possible to u_{j+1}^n in $I_{j+1/2}^n$. It leads to the formula

$$f_{j+1/2}^n = \begin{cases} ab_{j+1/2}^n & \text{if } u_{j+1}^n < b_{j+1/2}^n, \\ au_{j+1}^n & \text{if } b_{j+1/2}^n \leq u_{j+1}^n \leq B_{j+1/2}^n, \\ aB_{j+1/2}^n & \text{if } B_{j+1/2}^n < u_{j+1}^n. \end{cases}$$

Let us now compute the flux $f_{j+1/2}^n$ with the discontinuous reconstruction scheme (6, 14, 18) with formula (7), assuming $(u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) > 0$

(in the other case, the equivalence of the two schemes is obvious because they both degenerate to the upwind scheme $f^n_{j+1/2} = au^n_j$). For the sake of simplicity, we assume the data is locally increasing: $\overline{u^n_j}^l = u^n_{j-1} < u^n_j < u^n_{j+1} = \overline{u^n_j}^r$. Then:

- If $a\Delta t \leq (1-d^n_j)\Delta x$, $f^n_{j+1/2} = a\overline{u^n_j}^r = au^n_{j+1}$. On the other hand, one has $d^n_j = \frac{u^n_{j+1}-u^n_j}{u^n_{j+1}-u^n_{j-1}}$, so that $a\Delta t \leq (1-d^n_j)\Delta x$ is equivalent to $a\frac{\Delta t}{\Delta x}(u^n_{j+1}-u^n_{j-1}) \leq (u^n_j - u^n_{j-1})$ (the data is increasing), and $a\frac{\Delta t}{\Delta x}u^n_{j+1} \leq (u^n_j - u^n_{j-1}) + a\frac{\Delta t}{\Delta x}u^n_{j-1}$ and finally to $u^n_{j+1} \leq \frac{\Delta x}{a\Delta t}(u^n_j - u^n_{j-1}) + u^n_{j-1}$, so that

$$\begin{aligned} B^n_{j+1/2} &= \min(M^n_{j+1}, m^n_j + \frac{\Delta x}{a\Delta t}(u^n_j - m^n_j)) \\ &= \min(u^n_{j+1}, u^n_{j-1} + \frac{\Delta x}{a\Delta t}(u^n_j - u^n_{j-1})) = u^n_{j+1}. \end{aligned}$$

  Thus $f^n_{j+1/2}$ and the limited downwind flux are the same under the same condition $b^n_{j+1/2} \leq u^n_{j+1} \leq B^n_{j+1/2}$ (condition $b^n_{j+1/2} \leq u^n_{j+1}$ being automatically verified because $b^n_{j+1/2} = u^n_j < u^n_{j+1}$).

- If $a\Delta t > (1-d^n_j)\Delta x$, $f^n_{j+1/2} = (1-d^n_j)\frac{\Delta x}{\Delta t}\overline{u^n_j}^r + (a-(1-d^n_j)\frac{\Delta x}{\Delta t})\overline{u^n_j}^l$. Once more, let us replace d^n_j by $\frac{u^n_{j+1}-u^n_j}{u^n_{j+1}-u^n_{j-1}}$. The value of the flux is then $f^n_{j+1/2} = \frac{u^n_j-u^n_{j-1}}{u^n_{j+1}-u^n_{j-1}}\frac{\Delta x}{\Delta t}u^n_{j+1} + (a - \frac{u^n_j-u^n_{j-1}}{u^n_{j+1}-u^n_{j-1}}\frac{\Delta x}{\Delta t})u^n_{j-1}$ which finally leads to $f^n_{j+1/2} = a(u^n_{j-1} + \frac{\Delta x}{a\Delta t}(u^n_j - u^n_{j-1}))$. Furthermore, the hypothesis that the data is increasing implies that $B^n_{j+1/2} = \min(u^n_{j+1}, u^n_{j-1} + \frac{\Delta x}{a\Delta t}(u^n_j - u^n_{j-1}))$, and hypothesis $a\Delta t > (1-d^n_j)\Delta x$ says that this minimum is $u^n_{j-1} + \frac{\Delta x}{a\Delta t}(u^n_j - u^n_{j-1})$, so that $f^n_{j+1/2} = aB^n_{j+1/2}$. We conclude that $f^n_{j+1/2} = aB^n_{j+1/2}$ if $B^n_{j+1/2} < u^n_{j+1}$, which again coincides with the limited downwind scheme.

To complete the proof, one would have to consider the case of a decreasing data $u^n_{j-1} > u^n_j > u^n_{j+1}$. This can be done on the same manner.

Let us now recall an interesting property of the limited downwind scheme for advection: the exact advection of step functions (see [6] for the proof, in the limited downwind formalism).

Proposition 3 Let $(u^n_j)_{j\in\mathbb{Z},n\in\mathbb{N}}$ be computed with scheme (6, 14, 18) with $f(u) = au$ and $a > 0$, and $a\frac{\Delta t}{\Delta x} \leq 1$. Let us assume that $\exists n \in \mathbb{N}$ such that

$(u_j^n)_{j\in\mathbb{Z}}$ verifies: $\exists \alpha^n \in [0,1[$ such that $\forall j \in \mathbb{Z}$,

$$u_{3j+1}^n = u_{3j}^n \text{ and } u_{3j+2}^n = \alpha^n u_{3j+1}^n + (1-\alpha^n) u_{3j+3}^n.$$

Then

- either $0 \leq \alpha^n + a\frac{\Delta t}{\Delta x} < 1$ and for all j

  $$u_{3j+1}^{n+1} = u_{3j}^{n+1} = u_{3j}^n \text{ and } u_{3j+2}^{n+1} = (\alpha^{n+1}) u_{3j+1}^{n+1} + (1-\alpha^{n+1}) u_{3j+3}^{n+1}$$

  with $0 \leq \alpha^{n+1} = \alpha^n + a\frac{\Delta t}{\Delta x} < 1$;

- or $1 \leq \alpha^n + a\frac{\Delta t}{\Delta x} < 2$ and for all j

  $$u_{3j+2}^{n+1} = u_{3j+1}^{n+1} = u_{3j+1}^n \text{ and } u_{3j+3}^{n+1} = (\alpha^{n+1}) u_{3j+2}^{n+1} + (1-\alpha^{n+1}) u_{3j+4}^{n+1}$$

  with $0 \leq \alpha^{n+1} = \alpha^n + a\frac{\Delta t}{\Delta x} - 1 < 1$.

This means that the set of step functions is preserved by the scheme and advected with the right velocity. In particular, this proposition applies for discontinuous functions such as Heavyside profiles. See [6] for extension to compressible gas dynamics and [7] for extension to multi-fluids computations. We again refer to [6] for numerical results and for a conjecture of non-diffusion in infinite time. The drawback of this scheme is that it is not entropic. This is particularly striking (and harmful) for non-linear equations, as noticed in [15]. The next sections are devoted to schemes verifying the entropy conditions above.

We propose two different entropic schemes. The first one, detailed in section 3.5, is based on the ideas of the limited downwind scheme: it consists in maximizing the variation of the reconstructed solution in each cell, under the conservativity (15), the stability (16) and the entropy (11) constraints. The result, as shown below in the numerical results section, is a non-dissipative scheme that computes exact shocks (located on only one cell) and that produces some staircase in smooth profiles such as rarefaction waves. In order to avoid these staircases, we then propose another way of reconstruction that preserves both exact shocks and regularity inside rarefaction waves, section 3.6.

3.5 A non-dissipative entropic scheme that maximizes the total variation

The limited downwind scheme has been shown to be equivalent to the discontinuous reconstruction scheme with parameters defined by (18), and this can

be viewed as choosing the parameters such that they maximize $|\overline{u_j^n}^r - \overline{u_j^n}^l|$ under constraints (15) and (16). This is not an entropic choice. We here will add some entropy condition for one particular strictly convex entropy. We here propose to maximize $|\overline{u_j^n}^r - \overline{u_j^n}^l|$ under constraints (15), (16) and (11) with numerical entropy fluxes given by (13), $\forall j \in \mathbb{Z}$, $\forall n \in \mathbb{N}$. The choice to maximize the total variation is motivated by the observation that numerical diffusion leads to the diminution of the total variation. Thus, maximizing the total variation during the reconstruction step appears as a natural way to avoid numerical diffusion. Another choice is proposed in section 3.6.

The numerical entropy fluxes are defined by

$$G_{j+1/2}^n = \frac{1}{\Delta t}\int_0^{\Delta t} G(\mathcal{E}(s)\mathcal{R}\overline{u}^n((j+1/2)\Delta x))\,ds \quad \forall n \in \mathbb{N}, \forall j \in \mathbb{Z}.$$

The following of this section is devoted to the computation of parameters $\overline{u_j^n}^l$, $\overline{u_j^n}^r$ and d_j^n under the assumption $(u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) > 0$. For every $n \in \mathbb{N}, j \in \mathbb{Z}$, let us define $\Sigma_j^n = S_j^{n-1} - \frac{\Delta t}{\Delta x}(G_{j+1/2}^{n-1} - G_{j-/2}^{n-1})$. So Σ_j^n is the upper limit of S_j^n, because the numerical entropy inequality reads $S_j^n - \Sigma_j^n \leq 0$. Let us define

$$C_j^n = \left\{ (u^r, u^l, d) \in \mathbb{R}^2 \times [0,1] \text{ s.t. } \left\{ \begin{array}{l} du^l + (1-d)u^r = u_j^n, \\ u^l \in \lfloor u_{j-1}^n, u_j^n \rceil, \\ u^r \in \lfloor u_j^n, u_{j+1}^n \rceil, \\ dS(u^l) + (1-d)S(u^r) \leq \Sigma_j^n \end{array} \right. \right\}.$$

The retained reconstruction $(\overline{u_j^n}^l, \overline{u_j^n}^r, d_j^n)$ is the solution of the maximization problem

$$|\overline{u_j^n}^r - \overline{u_j^n}^l| = \max_{(u^r,u^l,d)\in C_j^n} |u^r - u^l|. \tag{19}$$

Lemma 1 Problem (19) does have a solution.

Proof By definition,

$$\begin{aligned}\Sigma_j^n &= S_j^{n-1} - \frac{\Delta t}{\Delta x}(G_{j+1/2}^{n-1} - G_{j-1/2}^{n-1}) \\ &= S_j^{n-1} - \frac{1}{\Delta x}\left(\int_0^{\Delta t} G(\mathcal{E}(s)\mathcal{R}\overline{u}^{n-1}((j+1/2)\Delta x))\,ds \right. \\ &\qquad \left. - \int_0^{\Delta t} G(\mathcal{E}(s)\mathcal{R}\overline{u}^{n-1}((j-1/2)\Delta x))\,ds\right),\end{aligned}$$

thus the entropy property of the exact operator implies

$$\frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} S(\mathcal{E}(\Delta t)\mathcal{R}\overline{u}^{n-1}(x)\,dx \leq \Sigma_j^n.$$

Now, by virtue of Jensen's inequality, one recovers $S(u_j^n) \leq \Sigma_j^n$. This shows that C_j^n is a non-empty set containing at least (u_j^n, u_j^n, d) for every $d \in [0,1]$. C_j^n is a closed non-empty set and the function to maximize is continuous, so that there exists a solution to the maximization problem.

The interest of numerical entropy inequalities is to ensure the convergence toward an entropy solution.

Theorem 2 Consider a scheme in the form (6, 14). Assume that $\max_{u\in[m,M]} f'(u)\Delta t \leq \Delta x$. Assume the reconstruction operation is conservative (eq. (15)) and that $\mathcal{R}\overline{u}^n(x)$ verifies

$$\left(\overline{u_j^n}^l, \overline{u_j^n}^r, d_j^n\right) \in C_j^n \quad \forall n \in \mathbb{N}, \forall j \in \mathbb{Z}.$$

Then, the scheme is L^∞-decreasing and TVD and owns numerical entropy fluxes. As Δt and Δx converge to 0, the numerical solution $\overline{u}_{\Delta x}^{\Delta t}$ converges in $L^\infty(]0,+\infty[, L^1(\mathbb{R}))$, up to a subsequence, toward a weak solution of the initial value problem (1-2) that verifies the entropy inequality

$$\partial_t S(u) + \partial_x G(u) \leq 0, \quad t \in \mathbb{R}^+, x \in \mathbb{R}.$$

The proof is a classical consequence of the Lax-Wendroff theorem. Remark that if f and S are strictly convex, the entropy solution is unique, thus the whole sequence of numerical approximations converges to the Krushkov solution.

As for the limited downwind scheme in the linear case, we now state a result of exact computation of pure shock solutions.

Proposition 4 Assume initial condition is $u^0(x) = u_L\ \forall x \in]-\infty, -\Delta x/2]$, $u^0(x) = u_R\ \forall x \in]-\Delta x/2, +\infty[$, discretized as $u_j^0 = u_L\ \forall j < 0$, $u_j^0 = u_R\ \forall j \geq 0$, such that $u^0(t, x-\sigma t)$ is a shock solution of (1) verifying entropy inequality (3) for a given entropy S. Then scheme (6, 14, 19) is exact in the sense that $\forall j \in \mathbb{Z}$, $\forall n \in \mathbb{N}$, $u_j^n = \frac{1}{\Delta x}\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u^0(t, x-\sigma n\Delta t)\,dx$.

Proof There is no reconstruction for the first iterate. Assume for simplicity that $\sigma\Delta t \leq \Delta x$, which is less strong than a usual CFL condition. After one

iterate with time step Δt, the exact solution is $\mathcal{E}(\Delta t)\overline{u}^0 = u^0(t, x - \sigma\Delta t)$, so that $u_0^1 = \sigma\frac{\Delta t}{\Delta x}u_L + (1 - \sigma\frac{\Delta t}{\Delta x})u_R$, and, because of the entropy inequality, $\Sigma_0^1 \geq \sigma\frac{\Delta t}{\Delta x}S(u_L) + (1 - \sigma\frac{\Delta t}{\Delta x})S(u_R)$. Thus the solution of (19) is

$$\begin{cases} \overline{u_1^1}^l = u_L, \\ \overline{u_1^1}^r = u_R, \\ d_1^1 = \sigma\frac{\Delta t}{\Delta x}. \end{cases}$$

The reconstructed solution is the exact solution $u^0(t, x-\sigma\Delta t)$ and the shock will be propagated for each time step. This proves the result.

For the sake of simplicity, we focus on the entropy

$$S(u) = \frac{u^2}{2}.$$

Without the entropy constraint, the parameter d_j^n solution to the maximization problem is $d_j^n = \frac{u_{j+1}^n - u_j^n}{u_{j+1}^n - u_{j-1}^n}$. With constraint (11), the maximization is more intricate, but the following lemma states that the maximal value $|\overline{u_j^n}^r - \overline{u_j^n}^l|$ is attained for either $\overline{u_j^n}^l = u_{j-1}^n$ or $\overline{u_j^n}^r = u_{j+1}^n$.

Lemma 2 Assume $(u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) > 0$ and let $(\overline{u_j^n}^r, \overline{u_j^n}^l, d_j^n)$ be such that

$$|\overline{u_j^n}^r - \overline{u_j^n}^l| = \max_{(u^l, u^r, d) \in C_j^n} |u^r - u^l|.$$

Then, either $\overline{u_j^n}^l = u_{j-1}^n$ or $\overline{u_j^n}^r = u_{j+1}^n$ and the solution is unique.

Proof Let us recall that $2S_j^n = d_j^n \overline{u_j^n}^{l\,2} + (1 - d_j^n)\overline{u_j^n}^{r\,2}$ and that $u_j^n = d_j^n\overline{u_j^n}^l + (1 - d_j^n)\overline{u_j^n}^r$. So

$$\begin{aligned} 2S_j^n &= d_j^n \left(u_j^n - (1 - d_j^n)(\overline{u_j^n}^r - \overline{u_j^n}^l)\right)^2 + (1 - d_j^n)\left(u_j^n + d_j^n(\overline{u_j^n}^r - \overline{u_j^n}^l)\right)^2 \\ &= u_j^{n\,2} + d_j^n(1 - d_j^n)(\overline{u_j^n}^r - \overline{u_j^n}^l)^2. \end{aligned}$$

Thus

$$(\overline{u_j^n}^r - \overline{u_j^n}^l)^2 = \frac{2S_j^n - u_j^{n\,2}}{d_j^n(1 - d_j^n)}$$

and, for a given value of the entropy s_j^n of the reconstructed solution, $|\overline{u_j^n}^r - \overline{u_j^n}^l|$ is maximal when $d_j^n(1 - d_j^n)$ is minimal, i.e. when d_j^n is minimal or maximal. Thus, let S_j^n denote the entropy of the solution of the

maximization problem: this solution is such that the associated d_j^n is either minimal or maximal under constraints $\overline{u_j^n}^l \in [u_{j-1}^n, u_j^n]$ and $\overline{u_j^n}^r \in [u_j^n, u_{j+1}^n]$. Furthermore, the minimal d_j^n is attained for $\overline{u_j^n}^l = u_{j-1}$, whereas the maximal d_j^n is associated to $\overline{u_j^n}^r = u_{j+1}$. The result is established.

Effective computations

It remains to describe briefly the scheme. This is now only a matter of computation. The reconstruction of the solution at time step n in cell j follows the algorithm here described. Assume u_{j-1}^n, u_j^n, u_{j+1}^n and $\Sigma_j^n = S_j^{n-1} - \frac{\Delta t}{\Delta x}(G_{j+1/2}^{n-1} - G_{j-/2}^{n-1})$ are known.

- If $(u_{j+1}^n - u_j^n)(u_j^n - u_{j-1}^n) \leq 0$, $\overline{u_j^n}^l = \overline{u_j^n}^r = u_j^n$ and d_j^n does not matter (case without reconstruction).

- Else, we have to compute (before comparing the resulting local variations) $(\overline{u_j^n}^l, \overline{u_j^n}^r, d_j^n)$ in the two cases: $\overline{u_j^n}^l = u_{j-1}^n$ and $\overline{u_j^n}^r = u_{j+1}^n$. Straightforward and uninteresting computations lead to the following algorithm.

  - For the case $\overline{u_j^n}^l = u_{j-1}^n$, compute first

$$l_j^n = 1 - \frac{(u_j^n - u_{j-1})^2}{2\Sigma_j^n - 2u_{j-1}^n u_j^n + {u_{j-1}^n}^2},$$

    who is a candidate to be d_j^n, and

$$v_j^{nr} = \frac{u_j^n - l_j^n u_{j-1}^n}{1 - l_j^n},$$

    who is a candidate to be $\overline{u_j^n}^r$. Then, either $v_j^{nr} \in [u_j^n, u_{j+1}^n]$ and all the constraints are satisfied, or $v_j^{nr} \notin [u_j^n, u_{j+1}^n]$ and then we take finally

$$v_j^{nr} = u_{j+1}^n$$

    and

$$l_j^n = \frac{u_{j+1}^n - u_j^n}{u_{j+1}^n - u_{j-1}^n},$$

    the entropy constraint in there inactive.

– For the case $\overline{u_j^n}^r = u_{j+1}^n$, compute first

$$m_j^n = \frac{(u_{j+1}^n - u_j)^2}{2\Sigma_j^n - 2u_{j+1}^n u_j^n + {u_{j+1}^n}^2},$$

who is another candidate to be d_j^n, and

$$w_j^{nl} = \frac{u_j^n - (1-m)u_{j+1}^n}{m},$$

who is a candidate to be $\overline{u_j^n}^l$. Then, either $w_j^{nl} \in [u_{j-1}^n, u_j^n]$ and all the constraints are satisfied, or $w_j^{nl} \notin [u_{j-1}^n, u_j^n]$ and then we take finally

$$w_j^{nl} = u_{j-1}^n$$

and

$$m_j^n = \frac{u_{j+1}^n - u_j^n}{u_{j+1}^n - u_{j-1}^n},$$

the entropy constraint in there inactive.

- Now, compare the two possible reconstructions in term of total variation:

  – either $|v_j^{nr} - u_{j-1}^n| > |u_{j+1}^n - w_j^{nl}|$ and

  $$\overline{u_j^n}^l = u_{j-1}^n, \quad \overline{u_j^n}^r = v_j^{nr}, \quad d_j^n = l_j^n,$$

  – or $|v_j^{nr} - u_{j-1}^n| \leq |u_{j+1}^n - w_j^{nl}|$ and

  $$\overline{u_j^n}^l = w_j^{nl}, \quad \overline{u_j^n}^r = u_{j+1}^n, \quad d_j^n = m_j^n.$$

The algorithm is complete.

Remark 4 By Jensen's inequality, quantities $2\Sigma_j^n - 2u_{j-1}^n u_j^n + {u_{j-1}^n}^2$ and $2\Sigma_j^n - 2u_{j+1}^n u_j^n + {u_{j+1}^n}^2$ involved in the definitions of possible d_j^n are non-negative and smaller than 1. They unfortunately can be 0, but this means that $\mathcal{E}(\Delta t)\mathcal{R}\overline{u}^{n-1} = u_j^n = u_{j-1}^n$ in $[(j-1/2)\Delta x, (j+1/2)\Delta x]$, which naturally leads to the choice of $\overline{u_j^n}^l = \overline{u_j^n}^r = u_j^n$, and the value of d_j^n does not matter. This has to be taken in consideration carefully on the numerical point of view.

3.6 A more regularizing entropic choice

The algorithm previously described, maximizing the total variation of the reconstruction (under some constraints), presents the drawback of creating stair-cases (see the sections devoted to numerical experiments in the following). We here propose, among all possibilities, another reconstruction, which is justified a posteriori by the shape of the numerical results it produces. With the same definitions as in the algorithm above, let us follow

- either $|v_j^{nr} - u_{j-1}^n| > |u_{j+1}^n - w_j^{nl}|$ and

$$\overline{u_j^n}^l = w_{j\,l}^n, \quad \overline{u_j^n}^r = u_{j+1}^n, \quad d_j^n = m_j^n,$$

- or $|v_j^{nr} - u_{j-1}^n| \leq |u_{j+1}^n - w_j^{nl}|$ and

$$\overline{u_j^n}^l = u_{j-1}^n, \quad \overline{u_j^n}^r = v_j^{nr}, \quad d_j^n = l_j^n.$$

Thus this reconstruction consists in taking either $\overline{u_j^n}^l = u_{j-1}^n$ or $\overline{u_j^n}^r = u_{j+1}^n$, the choice being the one leading to the *smallest* total variation. This produces more regularized solutions without any spreading of discontinuities. Of course proposition 4 remains true with this new scheme.

3.7 Application to advection equation with constant velocity

The schemes developed above are here used to solve the advection equation with unit velocity $\partial_t u + \partial_x u = 0$ with periodic boundary conditions. The entropy flux is $G(u) = u^2/2$. The test-case is Harten's one (cf. [13]) in $[-1, 1]$:

$$u^0(x) = \begin{cases} 2x + 2 - \sin(3\pi(x - 1/2))/6 \text{ if } -1 \leq x < -1/2, \\ (1/2 - x)\sin(3/2\pi(x - 1/2)^2) \text{ if } -1/2 \leq x < 1/6, \\ |\sin(2\pi(x - 1/2))| \text{ if } 1/6 \leq x < 5/6, \\ 2x - 2 - \sin(3\pi(x - 1/2))/6 \text{ if } 5/6 \leq x < 1. \end{cases}$$

Here are the results after 100 periods ($t = 100$) with $\Delta t = 0.8\Delta x$ with 50 cells and 200 cells. We compare the results of the classical Minmod limiter (cf. [8, 23, 24]), the self-adaptive antidiffusive scheme of [2] and the two entropic reconstructions proposed in the paper.

Results show that discontinuities are kept by the two proposed versions of discontinuous reconstruction schemes. The behavior in smooth region is nevertheless different. Small steps are created during the first time iterates

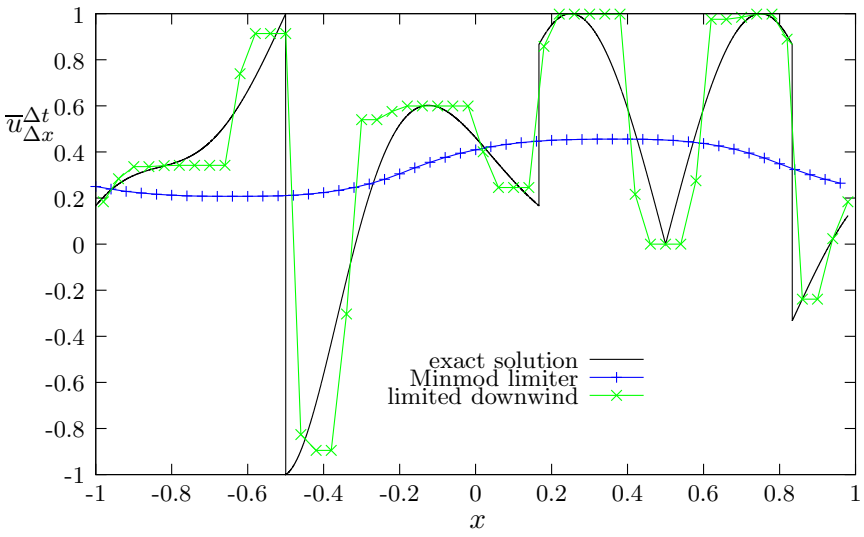

Figure 2: With 50 cells.

by the first discontinuous reconstruction scheme, and are then perfectly advected. This behavior is similar to the one of the limited downwind scheme, but here with much better accuracy. Indeed, the length of the steps does not exceed 2 cells. The results with the self-adaptive antidiffusive scheme of [2] have a similar shape in certain regions, see for example figures 6 and 7. The second reconstruction provides more smoothness but has the light drawback of deteriorating a little the solution in very long time, as is seen on figure 5 near $x = -0.8$ and $x = -0.2$. For those results in the linear case, the first reconstruction provides better results in long time. Nevertheless, one shall see in the following that it creates much more steps in rarefaction waves of non-linear equations.

3.8 Application to Burgers' equation

The limited downwind scheme, here reinterpreted as the discontinuous reconstruction scheme (6, 14, 18) with $f(u) = \frac{u^2}{2}$, has been shown to produce non-entropic shocks for Burgers' equation in [15]. In this reference, a modification of the limited downwind scheme is proposed that leads to entropy inequalities, but with too much dissipation in rarefaction waves. We will see in the present section that the entropy conditions derived above are large enough to obtain non dissipative entropic schemes. In the present paper, we only exploit the condition obtained for one entropy, which is sufficient

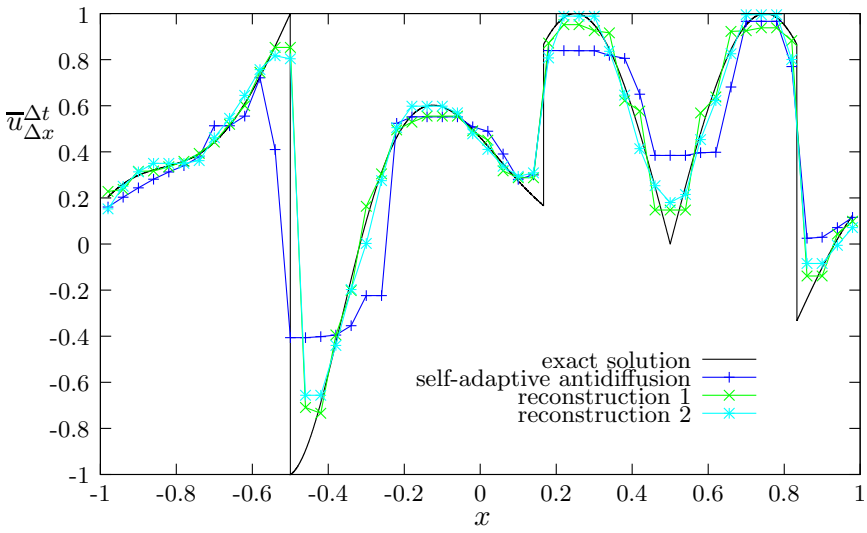

Figure 3: With 50 cells.

if this entropy is strictly convex (since the flux f is). The entropy flux is $G(u) = \frac{u^3}{3}$.

We first present results combining a shock and a rarefaction wave, obtained with $1 + \mathbb{1}_{[0.1,0.6]}$ as initial condition. The final time is 0.2. The exact solution is

$$u(0.2, x) = \begin{cases} 1 \text{ if } x \leq 0.3, \\ 1 + 5(x - 0.3) \text{ if } 0.3 \leq x \leq 0.5, \\ 2 \text{ if } 0.5 \leq x \leq 0.9. \end{cases}$$

Numerical computations have been done with a Courant number

$$\max_j u_j^n \Delta t / \Delta x = 0.3.$$

We shall compare the results given by the Minmod limiter, the self-adaptive antidiffusive entropy scheme of [2] and the two reconstruction schemes discussed above. Figures 8 and 9 present results at time $t = 0.2$ with 50 and 200 cells. We then observe the long-time behavior of the numerical solutions on figures 10 and 11. The initial condition is the same as above, but the final time is now 100. Both test-cases show the antidissipative behavior of the proposed reconstruction algorithms. What is remarkable is that the stair-cases effect with the first reconstruction, that maximizes the total variation in the reconstruction, does not disagree with the (stated) convergence

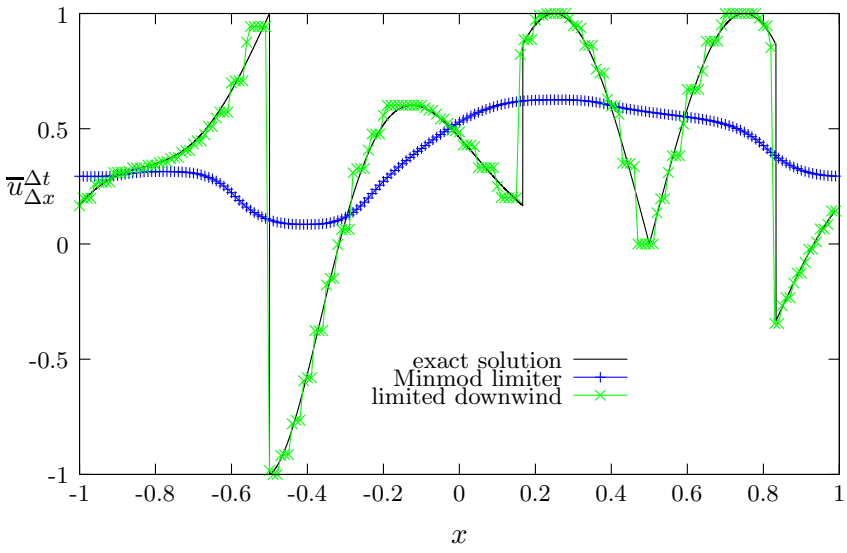

Figure 4: With 200 cells.

toward the entropy solution. This is numerically verified on the results with 200 cells. In particular, we notice that the stair-cases due to the first reconstruction, similar at the first glance to the ones of the Ultrabee limiter or limited downwind scheme (cf [24], [6]), are here controlled. We note that the second proposed reconstruction gives excellent results. It allows to obtain more smooth numerical solutions that are comparable to those of [2].

3.9 Approximate resolution and application to other scalar equations

It may be difficult to compute the exact evolution of solutions with such discontinuous reconstructed initial conditions, because there is no positive lower bound on d_j^n nor $1 - d_j^n$, so that there is no CFL condition ensuring the absence of interactions of waves generated at each discontinuity. We propose one way to compute easily approximate solutions with the discontinuous reconstruction. The reconstruction procedure is the same as in the preceding. The way to compute an approximate solution is based on a Godunov scheme on a non-regular mesh. The principle is to use the Godunov scheme on each sub-cell; but, as already noticed, the sub-cell are of arbitrary small length. Thus, the CFL condition of the Godunov scheme leads to an

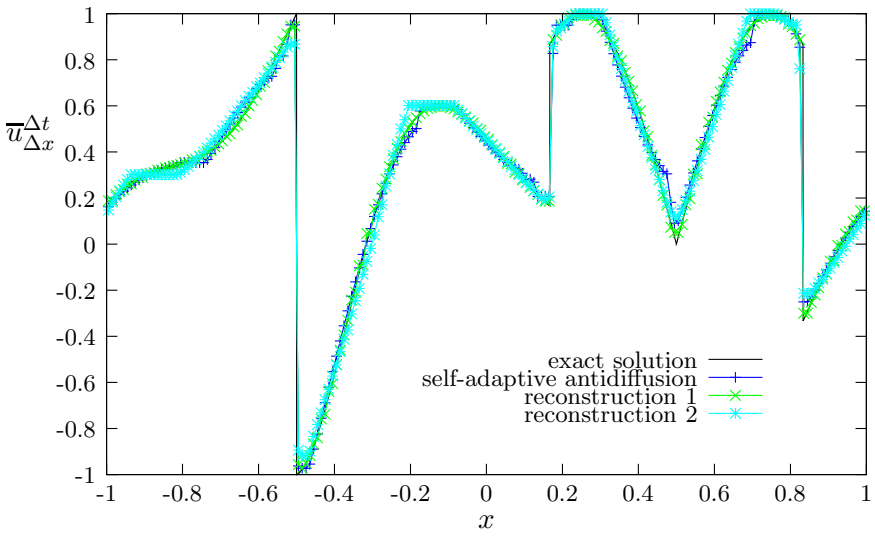

Figure 5: With 200 cells.

arbitrarily small time step. We first consider a time step verifying

$$\frac{\Delta t}{\Delta x} \max_{u \in [\inf_{j\in\mathbb{Z}} u_j^0, \sup_{j\in\mathbb{Z}} u_j^0]} f'(u) \leq 1/2$$

(the classical upper bound is 1). The procedure to compute the flux $f^n_{j+1/2}$ is the following.

- Perform the reconstruction of u as in the rest of the paper.
- If $d_j^n \leq 1/2$, take $f^n_{j+1/2} = f(\overline{u_j^n}^r)$ and $g^n_{j+1/2} = G(\overline{u_j^n}^r)$ (these are Godunov fluxes): the CFL condition above indeed guaranties that the discontinuity in cell j does not interact with the cell edge $j+1/2$.
- If $d_j^n > 1/2$, compute a maximal local time step Δt_j^n for which there is no interaction between the reconstructed discontinuity in cell j and cell edge $j+1/2$: take Δt_j^n such that

  $$\Delta t_j^n \sigma_j^n = (1 - d_j^n)\Delta x.$$

  where σ_j^n is the maximal wave (shock or rarefaction) velocity for the Riemann problem $(\overline{u_j^n}^l, \overline{u_j^n}^r)$ inside cell number j.

  - If $\Delta t_j \geq \Delta t$, take $f^n_{j+1/2} = f(\overline{u_j^n}^r)$ and $g^n_{j+1/2} = G(\overline{u_j^n}^r)$.

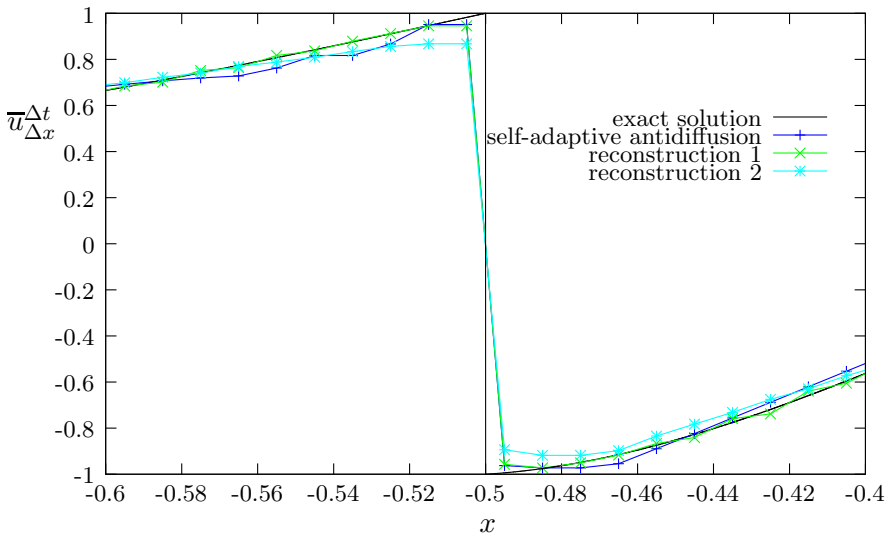

Figure 6: Zoom of figure 5.

– Else, that is to say, if $\Delta t_j < \Delta t$ (which does occur for example if $d_j^n \approx 1$) the flux $f_{j+1/2}^n$ is the combination of flux $f(\overline{u_j^n}^r)$ during Δt_j^n and a residual flux for the residual time $\Delta t - \Delta t_j^n$. The residual flux is evaluated by computing the new (after little time step Δt_j^n) projected value $\widetilde{u_j^n}$ of u in the right half-cell j:

$$\widetilde{u_j^n} = 2\left((d_j^n - 1/2)\overline{u_j^n}^l + (1-d_j^n)\overline{u_j^n}^r\right) - \frac{\Delta t_j^n}{\Delta x/2}\left(f(\overline{u_j^n}^r) - f(\overline{u_j^n}^l)\right).$$

Note that the unknown is projected on a whole half-cell $[j\Delta x, (j+1/2)\Delta x]$ and not on $[(j-1/2+d_j^n)\Delta x, (j+1/2)\Delta x]$. This guarantees that the allowed time step in the second stage will be lower bounded. Thus the global fluxes are taken as

$$f_{j+1/2}^n = \frac{1}{\Delta t}\left(\Delta t_j^n f(\overline{u_j^n}^r) + (\Delta t - \Delta t_j^n) f(\widetilde{u_j^n})\right).$$

and

$$g_{j+1/2}^n = \frac{1}{\Delta t}\left(\Delta t_j^n G(\overline{u_j^n}^r) + (\Delta t - \Delta t_j^n) G(\widetilde{u_j^n})\right).$$

Remark 5 The computation of wave velocity σ_j^n can be achieved in the following manner. First compute the shock speed associated with discontinuity

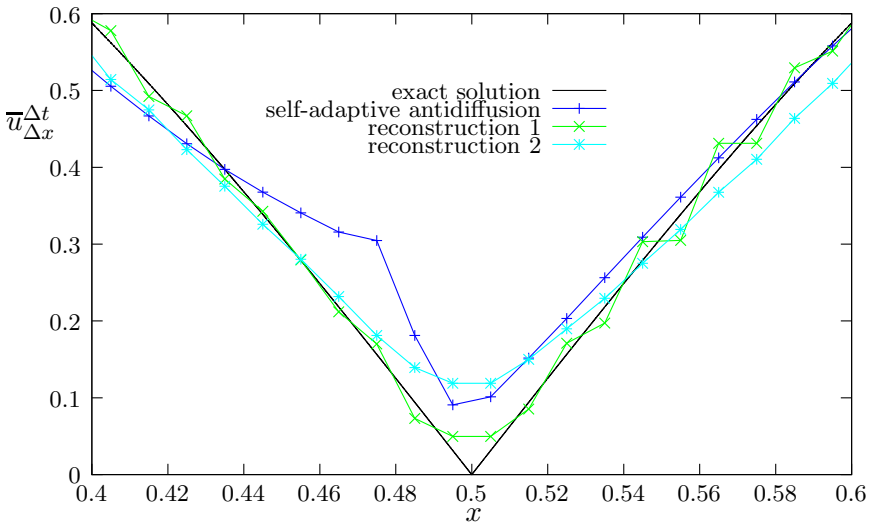

Figure 7: Another zoom of figure 5.

inside cell j, *via* Rankine-Hugoniot jump relations:

$$s_j^n = \frac{f(\overline{u_j^n}^r) - f(\overline{u_j^n}^l)}{\overline{u_j^n}^r - \overline{u_j^n}^l}.$$

Then, if $-s_j^n(S(\overline{u_j^n}^r) - S(\overline{u_j^n}^l)) + G(\overline{u_j^n}^r) - G(\overline{u_j^n}^l) \leq 0$, the shock is an admissible shock and thus one takes $\sigma_j^n = s_j^n$, else

$$\sigma_j^n = \max_{u \in \lfloor \overline{u_j^n}^l, \overline{u_j^n}^r \rceil} f'(u).$$

The present scheme is thus based on the Godunov scheme. The only difference is the computation of a reconstructed solution and of a flux in two steps. The use of a global half time step (in regard to the usual Godunov time step) allows to recover the properties of the Godunov scheme: stability and existence of entropy fluxes. The precise proof is skipped (note that the second part of the flux is nothing else than a Godunov flux for half-cells).

Let us now just present some results. First, we compare the above approximate schemes with the previous ones based on exact resolution of the equation with the reconstructed condition. We then present some results obtained for other fluxes than Burgers' one. What is remarkable is the resemblance between the solutions obtained with the exact method and the approximate one.

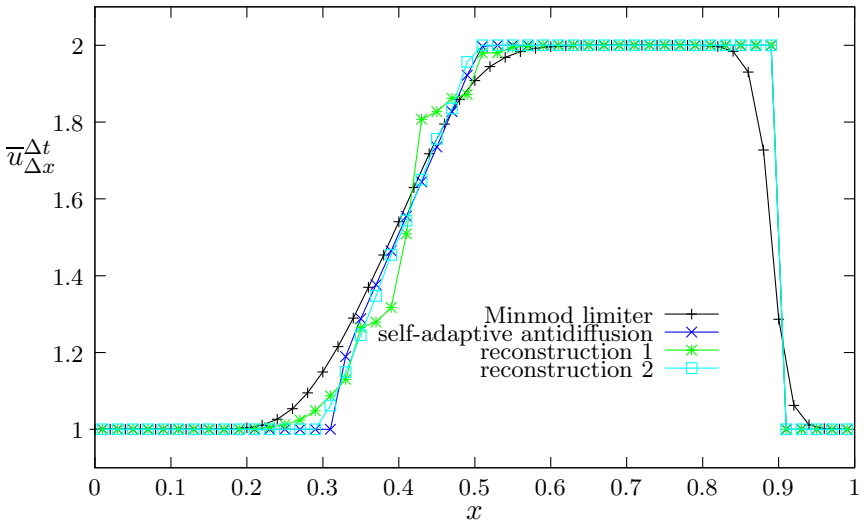

Figure 8: Final time 0.2, 50 cells.

For initial condition $1 + \mathbb{1}_{[0.1,0.6]}$ and time 0.2 with a Courant number 0.3, results with first and second reconstruction are reported on figures 12, 13, 14 and 15.

The schemes with approximate resolution give results that are very close to the exact resolution scheme. It is quite surprising that the second part of the step, which is a pure Godunov one, does not bring numerical diffusion, or, to say it differently, that the reconstruction procedure manages to reduce that much the numerical diffusion due to the Godunov part of the algorithm.

Equipped with this simple algorithm, we are able to consider various scalar equations. We now will consider the following flux:

$$f(x) = \begin{cases} \dfrac{1}{5-4x} \text{ if } x \in [0,1], \\ \dfrac{4}{5-x} \text{ if x } \in]1,5[, \end{cases}$$

as in [2]. This flux is not convex, thus the numerical decrease of one entropy does not guaranty convergence toward the entropy solution (this entropy criterion does not select a unique solution). The considered entropy is $S(u) = u^2/2$ so that the entropy flux is

$$G(u) = \begin{cases} \dfrac{\log(|5-4x|)}{4} + \dfrac{5}{4(5-4x)} + 4\log(4) + \dfrac{15}{4} \text{ if } x \in [0,1], \\ 4\log(|5-x|) + \dfrac{20}{5-x} \text{ if x } \in]1,5[, \end{cases}$$

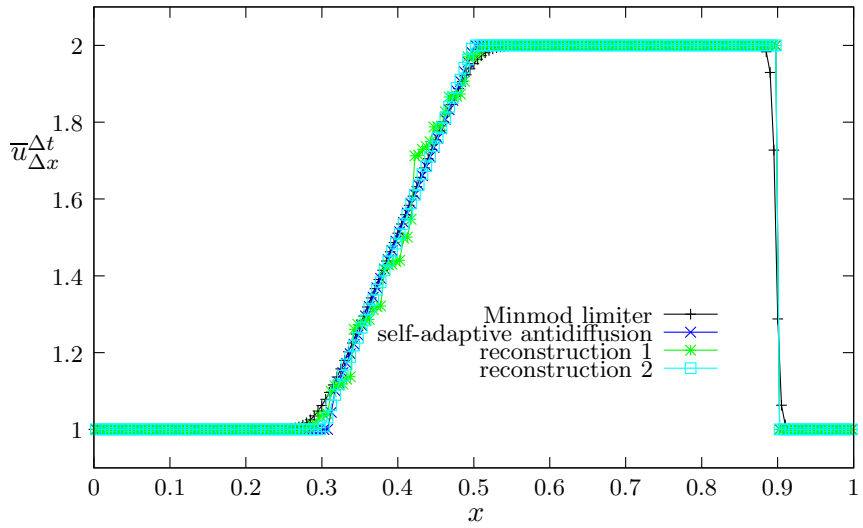

Figure 9: Final time 0.2, 200 cells.

The initial condition is $2 \times \mathbb{1}_{[0.1,0.4]}$ and the final time is 5/8, for a Courant number of 0.3. We compare results with a reference solution computed with the Godunov scheme and 10000 cells (thus close to the Krushkov solution). The schemes used are a self-adaptive entropy scheme of [2], and the first and the second reconstructions. We see this 3 different schemes provide numerical approximations converging toward different weak solutions of the PDE.

References

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

[2] F. Bouchut, An antidiffusive entropy scheme for monotone scalar conservation laws. J. Sci. Comput. 21 (2004), no. 1: 1–30.

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

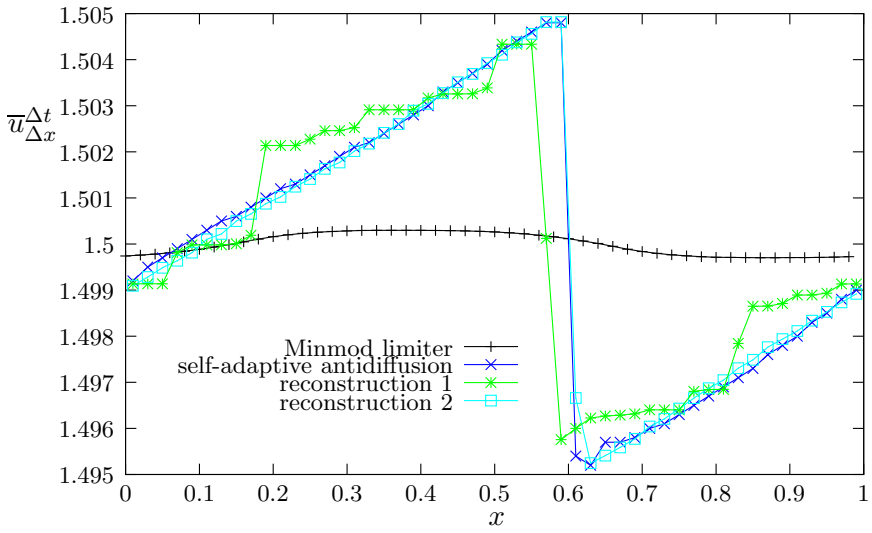

Figure 10: Final time 100, 50 cells.

[4] P. Collela, P. R. Woodward, The piecewise parabolic method (PPM) for gas-dynamical simulations, J. Comp. Phys. 54 (1984): 174–201.

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

[6] B. Després, F. Lagoutière, Contact discontinuity capturing schemes for linear advection and compressible gas dynamics, J. Sci. Comput. 16 (2001), no. 4: 479–524 (2002).

[7] B. Després, F. Lagoutière, Numerical resolution of a two-component compressible fluid model with interfaces, to appear in Progress in Computational Fluid Dynamics.

[8] E. Godlewski, P.-A. Raviart, Hyperbolic systems of conservation laws, Ellipses (1991).

[9] J. J. Goodman, R. J. LeVeque, A geometric approach to high resolution TVD schemes., SIAM J. Numer. Anal. 25 (1988), no. 2: 268–284.

[10] A. Harten, J. M. Hyman and P. D. Lax, On finite-difference approximations and entropy conditions for shocks, CPAM XXIX (1976): 297–322.

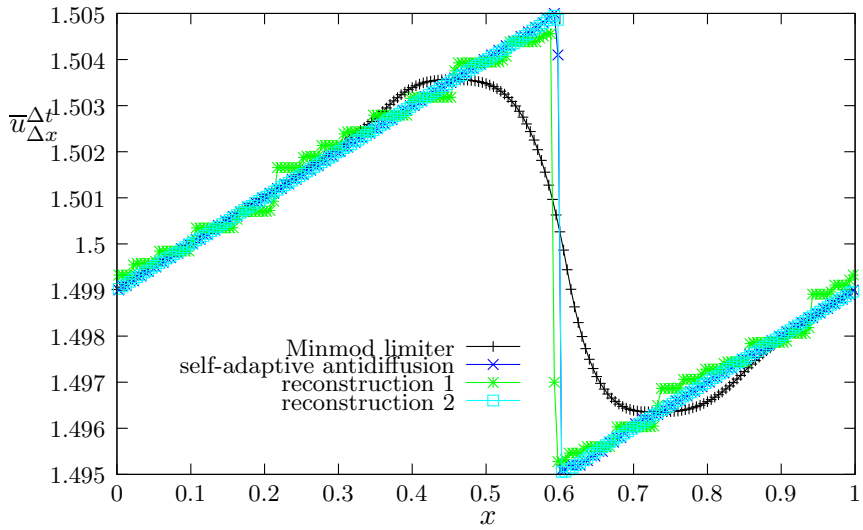

Figure 11: Final time 100, 200 cells.

[11] A. Harten and J. M. Hyman, Self adjusting grid methods for one-dimensional hyperbolic conservation laws, J. Comput. Phys. 50 (1983): 235–269.

[12] A. Harten, On a class of high resolution total-variation-stable finite-difference schemes, SIAM J. Numer Anal. 21 (1984), no. 1: 1–23.

[13] A. Harten, ENO schemes with subcell resolution, J. Comput. Phys. 83 (1989): 148–184.

[14] S. Krushkov, First-order quasilinear equations in several independent variables, Math. USSR Sb. 10 (1970): 217–243.

[15] F. Lagoutière, Numerical resolution of scalar convex equations: explicit stability, entropy and convergence conditions, CEMRACS 1999 (electronic), ESAIM Proc. 10 (1999): 183–199.

[16] F. Lagoutière, Stability of reconstruction schemes for scalar hyperbolic conservation laws, preprint.

[17] P. D. Lax and B. Wendroff, Systems of conservation laws, Comm. Pure Appl. Math. 23 (1960): 217–237.

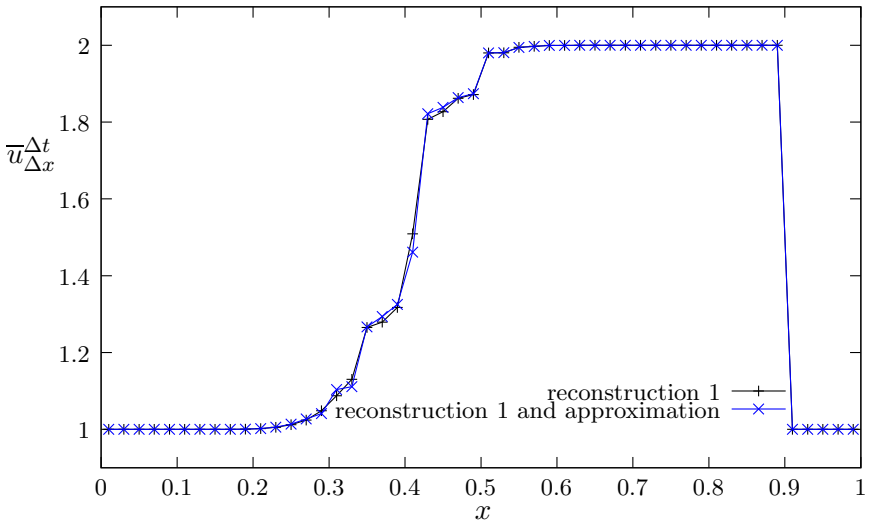

Figure 12: First reconstruction, 50 cells.

[18] P. G. LeFloch, J. M. Mercier, C. Rohde, Fully discrete, entropy conservative schemes of arbitrary order, SIAM J. Numer. Anal. 40 (2002), no. 5: 1968–1992.

[19] A. Y. Le Roux, A numerical conception of entropy for quasi-linear equations, Math. of Comp. 31 (1977), no. 140: 848–872.

[20] P.-L. Lions and P. E. Souganidis, Convergence of MUSCL and filtered schemes for scalar conservation laws and Hamilton-Jacobi equations, Numer. Math. 69 (1995): 441–470.

[21] K. Lipnikov and M. Shashkov, The error-minimization-based strategy for moving mesh methods, Commun. Comput. Phys. 1 (2006) no. 1: 53–80.

[22] S. Osher and E. Tadmor, On the convergence of difference approximations to scalar conservation laws, Math. of Comp. 50 (1988), no. 181: 19–51.

[23] P. K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, SIAM Journal of Numerical Analysis 21, 5 (1984): 995–1011.

[24] E. F. Toro, Riemann solvers and numerical methods for fluid dynamics, Springer-Verlag (1997).

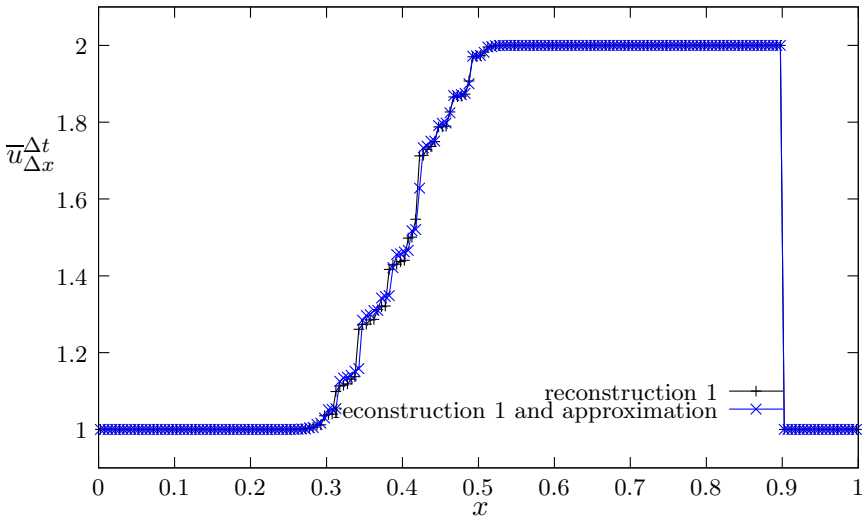

Figure 13: First reconstruction, 200 cells.

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

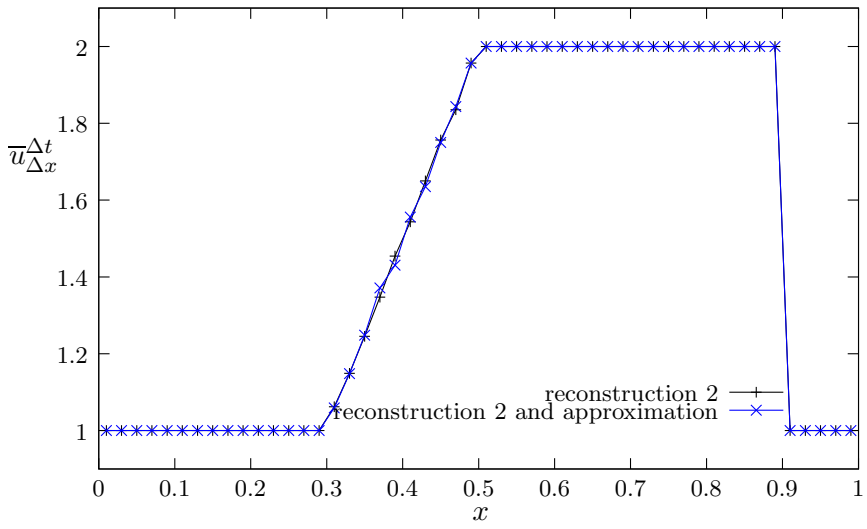

Figure 14: Second reconstruction, 50 cells.

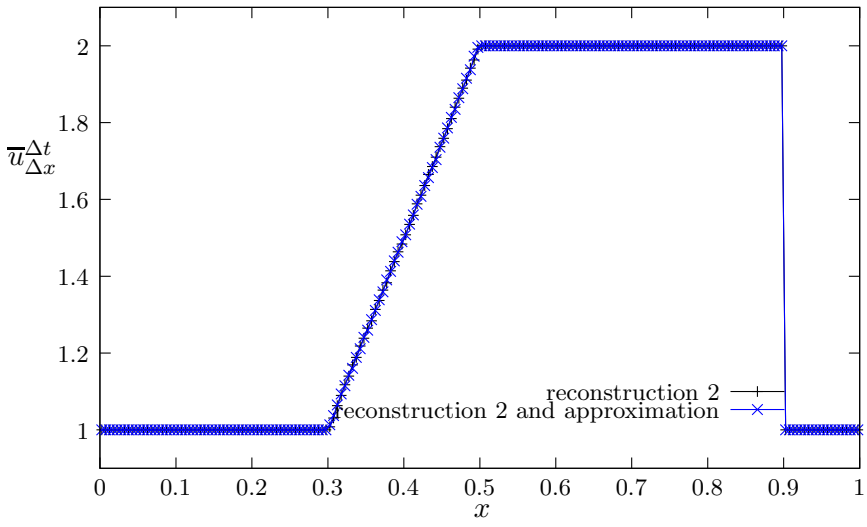

Figure 15: Second reconstruction, 200 cells.

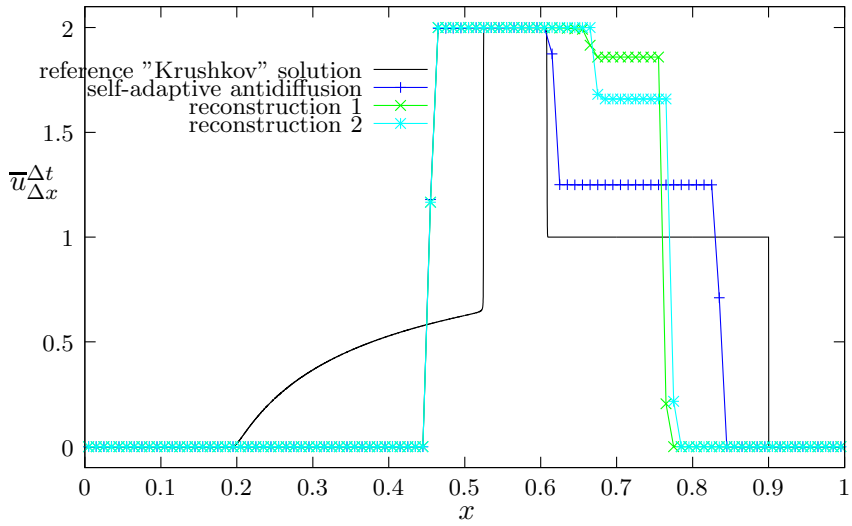

Figure 16: Different weak solutions with different algorithms.