Numerical approximations of one-dimensional linear conservation equations with discontinuous coefficients

Laurent Gosse, Francois James

Numerical approximations of one-dimensional linear conservation equations with discontinuous coefficients [1]

Laurent GOSSE[2] **& François JAMES**[3]

Abstract

Conservative linear equations arise in many areas of application, including continuum mechanics or high-frequency geometrical optics approximations. This kind of equations admits most of the time solutions which are only bounded measures in the space variable known as duality solutions. In this paper, we study the convergence of a class of finite-differences numerical schemes and introduce an appropriate concept of consistency with the continuous problem. Some basic examples including computational results are also supplied.

Key words Linear conservation equations – Duality solutions – Finite difference schemes – Weak consistency – Nonconservative product

1991 mathematical subjects classification 65M06 – 65M12 – 35F10

[1] Work partially supported by TMR project HCL #ERBFMRXCT960033

[2] Foundation for Research and Technology Hellas / Institute of applied and Computational Mathematics – P.O. Box 1527 – 71110 Heraklion, Crete – GREECE

[3] MAPMO – UMR CNRS 6628 – Université d'Orléans – BP 6759 – 45067 Orléans Cedex 2 – FRANCE

1 Introduction

This paper is devoted to rather general numerical approximations of the following linear conservation equation:

$$\begin{cases} \partial_t \mu + \partial_x (a\mu) = 0 \qquad \text{for} \quad (t,x) \in]0,T[\times\mathbb{R}, \\ \mu(0,.) = \mu_0 \in \mathcal{M}_{loc}(\mathbb{R}), \end{cases} \tag{1}$$

when the coefficient a satisfies

$$a \in L^\infty(]0,T[\times\mathbb{R}), \qquad \partial_x a \leq \alpha \quad \text{in }]0,T[\times\mathbb{R}, \quad \alpha \in L^1(]0,T[). \tag{2}$$

We shall also consider briefly the corresponding transport equation

$$\begin{cases} \partial_t u + a\partial_x u = 0 \qquad \text{for} \quad (t,x) \in]0,T[\times\mathbb{R}, \\ u(0,.) = u_0 \in BV_{loc}(\mathbb{R}). \end{cases} \tag{3}$$

This kind of equations is encountered for example in the field of non linear hyperbolic systems. The transport equation appears in the context of nonconservative products involved for instance in multispecies chemical reacting models, and in several numerical methods for hyperbolic systems (see e.g. [9, 20]). The conservation equation arises when considering systems with measure-valued solutions, see for instance [19, 21, 29]. Another field of application is the so-called pressureless gases model: [1, 10, 16, 4, 5]. Equation (1) appears also explicitly when linearizing a nonlinear hyperbolic equation

$$\partial_t u + \partial_x f(u) = 0 \tag{4}$$

with respect to the initial data or the flux f. Concerning the first case, we refer to the numerical application by Olazabal [24], where a 2-dimensional perturbation of a 1-dimensional shock is studied (see also [15]). A simplified model for this is the linearized equation

$$\partial_t \mu + \partial_x [f'(u)\mu] = 0, \tag{5}$$

and we refer to [6] for a theoretical study of this problem when f is convex. In the context of the flux identification for convex scalar conservation laws, we obtain the same equation, with a measure-valued right-hand side. We refer to [18], where the adopted point of view is very close to the one in this paper.

One may consider also the high-frequency geometrical optics approximations for the two-dimensional Helmholtz equation in a non-homogeneous medium. If one looks for planar wave solutions in the form $A(x,y)\mathbf{e}^{i\omega\varphi(x,y)}$, where A is the amplitude, ω the time frequency, and φ the phase of the wave, then φ satisfies a steady eikonal equation with a source term on the right-hand side and the "energy" $\tilde{A} = A^2/2$ might be sought as the solution of $\mathrm{div}_{x,y}(\tilde{A}\cdot\nabla\varphi) = 0$ (cf [11, 12]). Most of the numerical approximations one can get for this stationary problem are obtained by a time dependant scheme iterated up to the convergence. The following one-dimensional equation can therefore be considered as a simplified model for this process:

$$\partial_t \tilde{A} + \partial_x(\partial_x\varphi \cdot \tilde{A}) = 0. \tag{6}$$

Since φ is usually defined in the sense of the viscosity theory [23], it is only endowed with a Lipschitz smoothness in space. This matches the context in which we propose our work.

An appropriate theoretical framework for (1) has been recently introduced by Bouchut and James [2, 3] (see also Poupaud and Rascle [26] for another approach in the multidimensional case). It turns out that, in most of the cases, μ is a measure in the space variable. So, because of the very low regularity imposed on the coefficient a, one cannot treat *a priori* this Cauchy

problem in the theory of distributions. One way out is to understand the solution of (1) in the *duality sense*. For this purpose, it will be useful to write down the dual problem

$$\begin{cases} \partial_t p + a\partial_x p = 0, \qquad (t,x) \in \ ]0,T[\times\mathbb{R}, \\ p(T,.) = p^T \in \mathrm{Lip}(\mathbb{R}) \text{ with compact support.} \end{cases} \tag{7}$$

It is known that this backward problem admits a Lipschitz continuous solution under condition (2), and this fact has been already used to obtain uniqueness for (4) (see [25, 8, 17, 28, 22]). The point is that there is no uniqueness for (7), and one of the main results in [2, 3] is to characterize a class of solutions, known as *reversible solutions*, for which existence and uniqueness hold. The duality solution of (1) is then the unique element of the space $C([0,T];\mathcal{M}_{loc}(\mathbb{R}))$ satisfying for all reversible p's

$$\frac{d}{dt}\int_{\mathbb{R}} p(t,x)\,\mu(t,dx) = 0. \tag{8}$$

A similar notion can be introduced for (3). Equipped with this characterization, it is therefore possible to give a precise interpretation of the ambiguous product $(a\mu)$ in the distributional framework.

We want now to make more precise what we mean by numerical approximation of (1). We consider for $K \in \mathbb{N}$ conservative algorithms of the type:

$$\begin{cases} \mu_j^{n+1} = \mu_j^n - \dfrac{\Delta t}{\Delta x}\left(< \mathbf{A}^n_{j+\frac12}, \vec{\mu}^n_{j+\frac12} >_{\mathbb{R}^{2K}} - < \mathbf{A}^n_{j-\frac12}, \vec{\mu}^n_{j-\frac12} >_{\mathbb{R}^{2K}} \right) \\ \vec{\mu}^n_{j+\frac12} = (\mu^n_{j-K+1}, ..., \mu^n_{j+K}) \in \mathbb{R}^{2K} \\ \mathbf{A}^n_{j+\frac12} = (a^n_{j+\frac12,-K+1}, ..., a^n_{j+\frac12,K}) \in \mathbb{R}^{2K} \end{cases} \tag{9}$$

where μ_j^n and $\mathbf{A}^n_{j+\frac12}$ denote respectively some approximations of $\mu(n\Delta t, j\Delta x)$ and $a(n\Delta t, j\Delta x)$. At this numerical level, the main difficulty is to handle the lack of *a priori* estimates satisfied by (9). Consequently, most of the work is done estimating what we called the *dual scheme* which is obtained by a summation by parts (as it is done for the continuous equations). Because of the smoothness of the reversible solutions of (7), it seems more hopeful to seek strong properties such as BV, L^∞ or Lipschitz-like bounds for these backward approximations. We prove that, under some CFL-type conditions on the space-time grid, we have compactness results and convergence towards the reversible solution associated to every smooth final data. Moreover, property (8) is automatically enforced by the definition of our dual scheme. Finally, as a consequence of the conservative character of (9), we have also a uniform bound on the total mass of the approximate solution of (1). Putting all these arguments together gives easily the expected convergence result towards the duality solution of the problem (1).

Consequently, this paper is organized as follows: in Section 2, we recall the specific characterizations of duality solutions for (1) and (3), with the existence and uniqueness results. We present also the derivation of the universal representative $\hat{a}$ of a, which gives a meaning to the product $a\mu$ in the distribution theory. In Section 3, we develop our theory for conservative $(2K+1)$-points schemes for (1) and (3). We define the associated dual scheme and analyse its behaviour by checking the sign of some appropriate coefficients. The cornerstone of our convergence proofs are some positivity requirements for these coefficients, which give bounds on the amplitude and the total variation of the approximations, as well as monotonicity and monotonicity preserving properties, together with a convenient notion of weak consistency with the time-continuous equation (1). In Section 4, we use these general results to establish the convergence of some very classical numerical schemes developed in the context of scalar conservation laws belonging to the Lax-Friedrichs (LxF) and upwind families. Finally, in Section 5,

we present some numerical computations obtained with three-points schemes taken from both these classes.

2 Some features about duality solutions

In this section we recall the definitions of the *duality solutions* to the direct problems (1) and (3), introduced by Bouchut and James [2, 3]. As mentioned before, a key tool is the *adjoint equation*, (7) for the conservative case, and

$$\begin{cases} \partial_t \pi + \partial_x (a\pi) = 0, & (t,x) \in \ ]0,T[\times\mathbb{R} \\ \pi(T,.) = \pi^T \in L^\infty_{loc}(\mathbb{R}) \end{cases} \tag{10}$$

for the transport equation. We first introduce the notion of *reversible solutions* to the backward problems (7) and (10). Since one of the aims of this paper is to characterize the approximations of (1) and (3) for which the afore mentioned dual scheme mimics these properties, we state precisely the most important properties of these solutions. Next, we give the definitions and fundamental properties of duality solutions.

In all Section 2 we consider a coefficient $a \in L^\infty(\Omega)$, $\Omega =]0,T[\times\mathbb{R}$, satisfying the one-sided Lipschitz condition (2). Notice that (2) actually implies some regularity on a: indeed for almost every $t \in]0,T[$, $a(t,.) \in BV_{loc}(\mathbb{R})$ and for any $x_1 < x_2$

$$TV_{[x_1,x_2]}(a(t,.)) \leq 2\left(|\alpha(t)|(x_2 - x_1) + \|a\|_{L^\infty}\right). \tag{11}$$

Following [3], we introduce the following four spaces:

$$\begin{array}{l} \mathcal{S}_{\mathcal{M}} = C([0,T], \mathcal{M}_{loc}(\mathbb{R}) - \sigma(\mathcal{M}_{loc}(\mathbb{R}), C_c(\mathbb{R}))), \\ \mathcal{S}_{\mathrm{Lip}} = \mathrm{Lip}_{loc}([0,T]\times\mathbb{R}), \\ \mathcal{S}_{BV} = C([0,T], L^1_{loc}(\mathbb{R})) \bigcap \mathcal{B}([0,T], BV_{loc}(\mathbb{R})), \\ \mathcal{S}_{L^\infty} = C([0,T], L^\infty_{loc}(\mathbb{R}) - \sigma(L^\infty_{loc}(\mathbb{R}), L^1_c(\mathbb{R}))). \end{array} \tag{12}$$

We are here interested in solutions $p \in \mathcal{S}_{\mathrm{Lip}}$ to (7), $\mu \in \mathcal{S}_{\mathcal{M}}$ to (1) and to solutions $\pi \in \mathcal{S}_{L^\infty}$ to (10), $u \in \mathcal{S}_{BV}$ to (3).

Detailed proofs of all the theorems in this section are to be found in [3].

2.1 Reversible solutions of the dual backward problems

We shall denote by $\mathcal{L}$ the space of Lipschitz solutions to (7). The key problem here is that there is no uniqueness for solutions in this class, as is it evidenced by the following example (Conway [8]). Consider $a(x) = -\operatorname{sgn}(x)$. Then any solution to (7) is of the following form:

$$p(t,x) = \begin{cases} p^T(x - (T-t)\operatorname{sgn} x) & \text{if} \quad T - t \leq |x|, \\ h(T - t - |x|) & \text{if} \quad T - t \geq |x|, \end{cases} \tag{13}$$

for some $h \in \mathrm{Lip}([0,T])$ such that $h(0) = p^T(0)$. Notice that there is a "canonical choice" for the above h, namely $h \equiv p^T(0)$. If p^T has a finite total variation, then it is preserved for this solution. Motivated by these observations, we introduce the following definition.

Definition 1 (reversible solutions) *(i) We call exceptional solution any function $p_e \in \mathcal{L}$ such that $p_e(T,.) = 0$. We denote by $\mathcal{E}$ the vector space of exceptional solutions.*
(ii) We call domain of support of exceptional solutions the open set

$$\mathcal{V}_e = \Big\{ (t,x) \in \Omega;\ \exists\ p_e \in \mathcal{E},\ p_e(t,x) \neq 0 \Big\}.$$

(iii) Any $p \in \mathcal{L}$ is called reversible if p is locally constant in $\mathcal{V}_e$. The vector space of reversible solutions to (7) will be denoted by $\mathcal{R}$.

In the preceding example, the exceptional solutions are given by $p_e(t,x) = h((T-t-|x|)^+)$ with $h \in \mathrm{Lip}([0,T])$, $h(0)=0$, and we have $\mathcal{V}_e = \{(t,x) \in \Omega;\ |x| < T-t\}$.

Theorem 1 (backward Cauchy problem) *Let $p^T \in \mathrm{Lip}_{loc}(\mathbb{R})$. Then there exists a unique $p \in \mathcal{L}$ reversible solution to (7) such that $p(T,.) = p^T$. This solution satisfies for any $x_1 < x_2$ and $t \in [0,T]$*

$$\|p(t,.)\|_{L^\infty(I)} \leq \|p^T\|_{L^\infty(J)}, \tag{14}$$

$$\|\partial_x p(t,.)\|_{L^\infty(I)} \leq e^{\int_t^T \alpha(s).ds} \|\partial_x p^T\|_{L^\infty(J)}, \tag{15}$$

with $I =]x_1, x_2[$ and $J =]x_1 - \|a\|_\infty(T-t), x_2 + \|a\|_\infty(T-t)[$. Moreover, $\partial_x p(t,.) \geq 0$ if $\partial_x p^T \geq 0$.

Equipped with this class of solutions, we shall be able now to give a precise meaning to the formal definition given by (8). But before that, we state some very important properties of reversible solutions.

First, more handable characterizations of reversible solutions are given by their specific behaviour with respect to monotonicity and total variation properties, which are of course related.

Theorem 2 *Let $p \in \mathcal{L}$.*
1–. Characterization by total variation.

(i) If p is reversible then $t \mapsto \int_{\mathbb{R}} |\partial_x p(t,x)|\, dx$ is constant in $[0,T]$.

(ii) If the above function is constant and finite, then p is reversible.

2–. Characterization by monotonicity.

p is reversible if and only if there exists p_1, $p_2 \in \mathcal{L}$ such that $\partial_x p_1 \geq 0$, $\partial_x p_2 \geq 0$ and $p = p_1 - p_2$.

Next, another important feature of reversible solutions is the following stability result with respect to perturbations of the coefficient and final data.

Theorem 3 (stability) *Let (a_n) be a bounded sequence in $L^\infty(\Omega)$, with $a_n \rightharpoonup a$ in $L^\infty(\Omega) - w\star$. Assume $\partial_x a_n \leq \alpha_n(t)$, where (α_n) is bounded in $L^1(]0,T[)$, $\partial_x a \leq \alpha \in L^1(]0,T[)$. Let (p_n^T) be a bounded sequence in $\mathrm{Lip}_{loc}(\mathbb{R})$, $p_n^T \to p^T$, and denote by p_n the reversible solution to*

$$\begin{cases} \partial_t p_n + a_n \partial_x p_n = 0 \quad in \quad \Omega, \\ p_n(T,.) = p_n^T. \end{cases}$$

Then $p_n \to p$ in $C([0,T] \times [-R,R])$ for any $R > 0$, where p is the reversible solution to

$$\begin{cases} \partial_t p + a \partial_x p = 0 \quad in \quad \Omega, \\ p(T,.) = p^T. \end{cases}$$

We turn now to the resolution of (10). The following definition and properties actually follow by differentiating the reversible solutions of (7). More precisely, if $\pi \in \mathcal{S}_{L^\infty}$ solves (10), there exists a unique (up to an additive constant) $p \in \mathcal{S}_{\mathrm{Lip}}$ which solves (7) (see Lemma 2.2.1 in [3]). Thus we can state the following definition.

Definition 2 *We say that $\pi \in \mathcal{S}_{L^\infty}$ solving (10) is a reversible solution if the corresponding p is reversible.*

The reversible conservative solutions therefore enjoy the following properties.

Theorem 4 (Conservative reversible solutions) *The following three properties are equivalent for $\pi \in \mathcal{S}_{L^\infty}$ solution to (10).*
(i) π is reversible,
(ii) $\pi = 0$ in $\mathcal{V}_e$,
(iii) $\pi = \pi_1 - \pi_2$, for some $\pi_i \in \mathcal{S}_{L^\infty}$ solutions to (10), such that $\pi_i \geq 0$.

From the existence and uniqueness theorem 1 for the nonconservative Cauchy problem, we have immediately

Theorem 5 (Conservative backward Cauchy problem) *Let $\pi^T \in L^\infty_{loc}(\mathbb{R})$. Then there exists a unique $\pi \in \mathcal{S}_{L^\infty}$ reversible solution to (10) such that $\pi(T,.) = \pi^T$. This solution satisfies for any $x_1 < x_2$ and $t \in [0,T]$*

$$\|\pi(t,.)\|_{L^\infty(I)} \leq e^{\int_t^T \alpha} \|\pi^T\|_{L^\infty(J)},$$

where $I =]x_1, x_2[$ and $J =]x_1 - \|a\|_\infty (T-t), x_2 + \|a\|_\infty (T-t)[$.
Moreover, $\pi \geq 0$ if $\pi^T \geq 0$.

2.2 Duality solutions

Without any further comment, we turn to the forward problem (1), and state the following

Definition 3 (conservative duality solutions) *We say that $\mu \in \mathcal{S}_{\mathcal{M}}$ is a duality solution to (1) if for any $0 < \tau \leq T$, and any reversible solution p to (7) with compact support in x, the function $t \mapsto \displaystyle\int_{\mathbb{R}} p(t,x)\mu(t,dx)$ is constant on $[0,\tau]$.*

Theorem 6 (forward conservative Cauchy problem) *Given $\mu^0 \in \mathcal{M}_{loc}(\mathbb{R})$, there exists a unique $\mu \in \mathcal{S}_{\mathcal{M}}$ duality solution to (1), such that $\mu(0,.) = \mu^0$. This solution satisfies for any $x_1 < x_2$ and $t \in [0,T]$*

$$\int_{[x_1,x_2]} |\mu(t,dx)| \leq \int_{[x_1 - \|a\|_\infty t, x_2 + \|a\|_\infty t]} |\mu^0(dx)|. \tag{16}$$

Moreover, $t \mapsto \displaystyle\int_{\mathbb{R}} |\mu(t,dx)|$ is nonincreasing on $[0,T]$.

Once again, the similar notion of duality solution for the transport equation (3) follows by analogy to the conservative case.

Definition 4 (nonconservative duality solutions) *We say that $u \in \mathcal{S}_{BV}$ is a duality solution to (3) if for any $0 < \tau \leq T$, and any reversible solution π to (10) with compact support in x, the function $t \mapsto \displaystyle\int_{\mathbb{R}} \pi(t,x)u(t,x)\,dx$ is constant on $[0,\tau]$.*

Theorem 7 (forward nonconservative Cauchy problem) *Given $u^0 \in BV(\mathbb{R})$, there exists a unique $u \in \mathcal{S}_{BV}$ duality solution to (3), such that $u(0,.) = u^0$. This solution satisfies for any $x_1 < x_2$ and $t \in [0,T]$*

$$TV_I(u(t,.)) \leq TV_J(u^0), \tag{17}$$

$$\|u(t,.)\|_{L^\infty(I)} \leq \|u^0\|_{L^\infty(J)}, \tag{18}$$

with $I =]x_1, x_2[$ and $J =]x_1 - \|a\|_\infty t, x_2 + \|a\|_\infty t[$. Moreover, $u \in \mathrm{Lip}([0,T], L^1_{loc}(\mathbb{R}))$.

Notice that the formal result which allows formally to pass from the conservative equation (1) to the nonconservative one (3) by integration holds true in the duality sense. More precisely, we have the following proposition, which will be useful in the sequel.

Proposition 1 *(i) Let $u \in \mathcal{S}_{BV}$ be a duality solution to $\partial_t u + a\partial_x u = 0$. Then $\mu = \partial_x u \in \mathcal{S}_{\mathcal{M}}$ is a duality solution to $\partial_t \mu + \partial_x(a\mu) = 0$.*
(ii) Let $\mu \in \mathcal{S}_{\mathcal{M}}$ be a duality solution to $\partial_t \mu + \partial_x(a\mu) = 0$. Then there exists $u \in \mathcal{S}_{BV}$ duality solution to $\partial_t u + a\partial_x u = 0$, such that $\mu = \partial_x u$. Moreover, u is unique up to an additive constant.

Up to now, the major drawback of duality solutions is that they are not defined as distributional solutions, since the product $a\mu$ or $a\partial_x u$ is not defined. The purpose of the next section is to give some indications about that, and to state a stability result with respect to perturbations of a and initial data, which is an important feature of duality solutions.

2.3 Definition of the $(a\mu)$ product and stability

First we have to introduce a notion of flux, which defines the product $a\mu$ in a rather simple way, through the equation.

Definition 5 (Generalized flux) *Let $\mu \in \mathcal{S}_{\mathcal{M}}$ be a duality solution to (1). We define the flux corresponding to μ by*

$$a \vartriangle \mu = -\partial_t u, \tag{19}$$

where $\mu = \partial_x u$ and $u \in \mathcal{S}_{BV}$ is a duality solution to the nonconservative problem (cf Proposition 1(ii)). We have therefore

$$\partial_t \mu + \partial_x(a \vartriangle \mu) = 0 \quad in \quad \mathcal{D}'(\Omega). \tag{20}$$

The application $\mu \mapsto a \vartriangle \mu$ is of course linear, and since $u \in \mathrm{Lip}([0,T], L^1_{loc}(\mathbb{R}))$, one can prove that $a \vartriangle \mu \in L^\infty(]0,T[, \mathcal{M}_{loc}(\mathbb{R}))$, and for any $x_1 < x_2$,

$$\|a \vartriangle \mu\|_{L^\infty(]0,T[,\mathcal{M}(]x_1,x_2[))} \leq \|a\|_\infty \int\limits_{]x_1 - \|a\|_\infty T, x_2 + \|a\|_\infty T[} |\mu(0,dx)|.$$

The following stability theorem is a consequence of Proposition 1 and Theorem 3.

Theorem 8 (Weak stability) *Let (a_n) be a bounded sequence in $L^\infty(]0,T[\times\mathbb{R})$, with $a_n \to a$ in $L^\infty(]0,T[\times\mathbb{R}) - w*$. Assume $\partial_x a_n \leq \alpha_n(t)$, where (α_n) is bounded in $L^1(]0,T[)$, $\partial_x a \leq \alpha \in L^1(]0,T[)$. Consider a sequence $(\mu_n) \in \mathcal{S}_{\mathcal{M}}$ of duality solutions to*

$$\partial_t \mu_n + \partial_x(a_n \mu_n) = 0 \quad in \quad \Omega,$$

such that $\mu_n(0,.)$ is bounded in $\mathcal{M}_{loc}(\mathbb{R})$, and $\mu_n(0,.) \rightharpoonup \mu^0 \in \mathcal{M}_{loc}(\mathbb{R})$.
Then $\mu_n \to \mu$ in $\mathcal{S}_{\mathcal{M}}$, where $\mu \in \mathcal{S}_{\mathcal{M}}$ is the duality solution to

$$\partial_t \mu + \partial_x(a\mu) = 0 \quad in \quad \Omega, \qquad \mu(0,.) = \mu^0.$$

Moreover, $a_n \vartriangle \mu_n \rightharpoonup a \vartriangle \mu$ weakly in $\mathcal{M}_{loc}(\Omega)$.

As it stands, the definition of the flux depends on the solution we consider, and thus is not completely satisfactory. We have actually the following result, which is proved through the study of the backward flow associated to (1). The proof is much more delicate than the previous result, in particular for the last assertion of the theorem.

Theorem 9 (Universal representative) *There exists a bounded Borel function* $\widehat{a}:]0,T[\times\mathbb{R}\to\mathbb{R}$ *such that for any conservative duality solution* μ*, one has*

$$a\,{}_\triangle\,\mu=\widehat{a}\mu. \tag{21}$$

We call such a function a universal representative of a*.*

Moreover, one can choose $\widehat{a}$ *such that*

$$a.e.\ t\in]0,T[,\ \forall x\in\mathbb{R},\ \widehat{a}(t,x)\in[a(t,x+),a(t,x-)]\,. \tag{22}$$

In particular, we have

$$\widehat{a}(t,x)=a(t,x)=a(t,x+)=a(t,x-)\quad a.e.\ in\ ]0,T[\times\mathbb{R}. \tag{23}$$

3 Numerical approximation

3.1 Some conservative linear numerical schemes

Starting from here, we introduce a uniform grid defined by the two parameters Δx and Δt denoting respectively the mesh-size and the time-step. As usual, the parameter λ will refer to $\Delta t/\Delta x$, and we shall write for short $\Delta\to 0$ when $\Delta t,\Delta x\to 0$ with a fixed λ. Moreover, the following notations will be of constant use in the sequel:

$$\forall j\in\mathbb{Z},\quad \mu_j^0=\frac{1}{\Delta x}\int_{\mathbb{R}}\mathbf{1}_{[(j-\frac12)\Delta x,(j+\frac12)\Delta x[}(x)d\mu_0(x).$$

The aim of this work is to derive numerical algorithms able to compute a sequence $(\mu_j^n)_{j\in\mathbb{Z}}^{n\in\mathbb{N}}$ of approximations of local averages:

$$\forall(j,n)\in(\mathbb{Z}\times\mathbb{N}_*),\quad \mu_j^n\simeq\frac{1}{\Delta x}\int_{\mathbb{R}}\mathbf{1}_{[(j-\frac12)\Delta x,(j+\frac12)\Delta x[}(x)\mu(n\Delta t,dx).$$

We will also frequently use the vectors $\vec{\mu}^n_{j+\frac12}$ and $\mathbf{A}^n_{j+\frac12}$ in $\mathbb{R}^{2K}$ introduced in (9). In the whole section, the notation a_j^n will stand for an approximation of the coefficient a which can vary from one scheme to another. The letter N will also stand for the quantity $T/\Delta t$. We give at once several examples directly inspired by standard algorithms used in the context of scalar conservation laws.

Lax-Friedrichs type schemes. A sequence of nonnegative viscosity coefficients $\varepsilon^n_{j+\frac12}$ being given, this class of schemes writes

$$\mu_j^{n+1}=\mu_j^n-\lambda\left[\frac12(a_{j+1}^n\mu_{j+1}^n-a_{j-1}^n\mu_{j-1}^n)-\frac{1}{2\lambda}\left[\varepsilon^n_{j+\frac12}(\mu_{j+1}^n-\mu_j^n)-\varepsilon^n_{j-\frac12}(\mu_j^n-\mu_{j-1}^n)\right]\right]. \tag{24}$$

In this case, we have

$$\mathbf{A}^n_{j+\frac12}=\frac12\left(a_j^n+\frac{\varepsilon^n_{j+\frac12}}{\lambda},a_{j+1}^n-\frac{\varepsilon^n_{j+\frac12}}{\lambda}\right).$$

The classical LxF scheme corresponds to the constant value $\varepsilon^n_{j+\frac12}\equiv 1$. For the case where $\varepsilon^n_{j+\frac12}\equiv\frac12$, we get the modified "à la Tadmor" version [27]. Notice that, for this kind of schemes, we have $K=1$, but more than three points may be involved through the viscosity coefficients $\varepsilon^n_{j+\frac12}$.

Upwind type schemes. We first define for each $z \in \mathbb{R}$ its positive and negative parts:

$$z^+ = \max(0, z), \qquad z^- = \min(z, 0).$$

We introduce the following discretization:

$$\mu_j^{n+1} = \mu_j^n - \lambda \left[[(a^n_{j+\frac{1}{2}})^+ \mu_j^n - (a^n_{j-\frac{1}{2}})^+ \mu^n_{j-1})] + [(a^n_{j+\frac{1}{2}})^- \mu^n_{j+1} - (a^n_{j-\frac{1}{2}})^- \mu^n_j] \right]. \tag{25}$$

In this case, we have

$$\mathbf{A}^n_{j+\frac{1}{2}} = \left((a^n_{j+\frac{1}{2}})^+, (a^n_{j+\frac{1}{2}})^- \right).$$

We will present in Section 4 some possible choices for the values $a^n_{j+\frac{1}{2}}$.

Notice that we can rewrite the scheme (25) in the following form:

$$\mu_j^{n+1} = \mu_j^n - \lambda \left[(a^n_{j+\frac{1}{2}})^- (\mu^n_{j+1} - \mu^n_j) + (a^n_{j+\frac{1}{2}})^+ (\mu^n_j - \mu^n_{j-1}) + (a^n_{j+\frac{1}{2}} - a^n_{j-\frac{1}{2}}) \mu^n_j \right],$$

which appears as a natural upwind discretization of

$$\partial_t \mu + a \partial_x \mu + \partial_x a \cdot \mu = 0.$$

Remark 1 *We would like to emphasize that the approximation of a (namely, the choice of the vector $\mathbf{A}^n_{j+\frac{1}{2}}$) may not be totally arbitrary. For instance, concerning the linearized equation (5), it depends on the approximation used for (4). In the geometrical optics setting, (6), it is given by a discretization of $\partial_x \varphi$, which is definitely not straightforward to choose.*

3.2 Working out the associated dual scheme

An important tool for the study of the numerical schemes for (1) is the *dual* algorithm.

Definition 6 *For every direct scheme (9) operating on $(\mu_j^n)_{j\in\mathbb{Z}}^{0\le n\le N}$, we define the* **dual scheme** *as the relation operating on the real-valued sequence $(p_j^n)_{j\in\mathbb{Z}}^{0\le n\le N}$ and satisfying the formal equality:*

$$\forall\ 1 \le n \le N, \qquad \sum_{j\in\mathbb{Z}} \mu_j^n p_j^n = \sum_{j\in\mathbb{Z}} \mu_j^{n-1} p_j^{n-1}. \tag{26}$$

This equality is of course the discrete analogue of (8) which characterizes the duality solutions of (1). Now, we detail the structure of this dual scheme: by its definition, we have

$$\sum_{j\in\mathbb{Z}} \left[\mu_j^{n-1} p_j^n - \lambda (< \mathbf{A}^{n-1}_{j+\frac{1}{2}}, \vec{\mu}^{n-1}_{j+\frac{1}{2}} >_{\mathbb{R}^{2K}} - < \mathbf{A}^{n-1}_{j-\frac{1}{2}}, \vec{\mu}^{n-1}_{j-\frac{1}{2}} >_{\mathbb{R}^{2K}}) - \mu_j^{n-1} p_j^{n-1} \right] = 0. \tag{27}$$

A summation by parts gives therefore

$$\sum_{j\in\mathbb{Z}} p_j^n < \mathbf{A}^{n-1}_{j+\frac{1}{2}}, \vec{\mu}^{n-1}_{j+\frac{1}{2}} >_{\mathbb{R}^{2K}} = \sum_{j\in\mathbb{Z}} \sum_{k=-K+1}^{K} p_j^n a^{n-1}_{j+\frac{1}{2},k} \mu^{n-1}_{j+k} = \sum_{j\in\mathbb{Z}} \sum_{k=-K+1}^{K} p^n_{j-k} a^{n-1}_{j-k+\frac{1}{2},k} \mu_j^{n-1},$$

so that, for $1 \le n \le N$, we get

$$\sum_{j\in\mathbb{Z}} \mu_j^{n-1} \left[p_j^n - \lambda \sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k} (p^n_{j-k} - p^n_{j-k+1}) - p_j^{n-1} \right] = 0. \tag{28}$$

That gives the expression of the dual scheme:

$$p_j^{n-1} = p_j^n - \lambda \sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k}(p^n_{j-k} - p^n_{j-k+1}). \tag{29}$$

Expressions (27) and (28) imply respectively the two *discrete weak formulations*:

$$\sum_{j\in\mathbb{Z}} p_j^0\mu_j^0 + \sum_{n=1}^{N-1}\sum_{j\in\mathbb{Z}} p_j^{n+1}\left[\mu_j^{n+1} - \mu_j^n + \lambda(< \mathbf{A}^n_{j+\frac{1}{2}}, \vec{\mu}^n_{j+\frac{1}{2}} >_{\mathbb{R}^{2K}} - < \mathbf{A}^n_{j-\frac{1}{2}}, \vec{\mu}^n_{j-\frac{1}{2}} >_{\mathbb{R}^{2K}})\right] = 0,$$

and

$$\sum_{n=0}^{N-1}\sum_{j\in\mathbb{Z}} \mu_j^n\left[p_j^n - p_j^{n+1} + \lambda \sum_{k=-K+1}^{K} a^n_{j-k+\frac{1}{2},k}(p^{n+1}_{j-k} - p^{n+1}_{j-k+1})\right].$$

At this point, it is convenient to introduce some other notations. We first rewrite the scheme (29) in order to emphasize boundedness and monotonicity. Let us introduce the following coefficients:

$$\begin{array}{lcll} B^n_{j,k} & = & \lambda(a^n_{j-k-\frac{1}{2},k+1} - a^n_{j-k+\frac{1}{2},k}), & k \notin \{-K, 0, K\} \\ B^n_{j,-K} & = & \lambda a^n_{j+K-\frac{1}{2},-K+1} & \\ B^n_{j,K} & = & -\lambda a^n_{j-K+\frac{1}{2},K} & \\ B^n_{j,0} & = & 1 + \lambda(a^n_{j-\frac{1}{2},1} - a^n_{j+\frac{1}{2},0}) & \end{array} \tag{30}$$

We notice that by construction

$$\forall (j,n) \in \mathbb{Z}\times\mathbb{N}, \quad \sum_{k=-K}^{K} B^n_{j,k} = 1, \tag{31}$$

and that (29) is equivalent to

$$p_j^{n-1} = \sum_{k=-K}^{K} B^{n-1}_{j,k} p^n_{j-k}. \tag{32}$$

Next, to study the TVD and monotonicity preservation properties of the scheme, we introduce

$$\Delta p^n_{j+\frac{1}{2}} = p^n_{j+1} - p^n_j,$$

and another set of coefficients, namely

$$\begin{array}{lcll} C^n_{j,k} & = & \lambda(a^n_{j-k+\frac{1}{2},k+1} - a^n_{j-k+\frac{1}{2},k}), & k \notin \{-K, 0, K\} \\ C^n_{j,-K} & = & \lambda a^n_{j+K+\frac{1}{2},-K+1} & \\ C^n_{j,K} & = & -\lambda a^n_{j-K+\frac{1}{2},K} & \\ C^n_{j,0} & = & 1 + \lambda(a^n_{j+\frac{1}{2},1} - a^n_{j+\frac{1}{2},0}) & \end{array} \tag{33}$$

for which we have

$$\forall (j,n) \in \mathbb{Z}\times\mathbb{N}, \quad \sum_{k=-K}^{K} C^n_{j+k,k} = 1. \tag{34}$$

Notice that the coefficients $B^n_{j,k}$ and $C^n_{j,k}$ satisfy

$$\begin{array}{ll} C^n_{j,k} & = B^n_{j,k} + \lambda(a^n_{j-k+\frac{1}{2},k+1} - a^n_{j-k-\frac{1}{2},k+1}), \quad \text{for } -K \le k \le K-1, \\ C^n_{j,K} & = B^n_{j,K}. \end{array} \tag{35}$$

Writing (29) for the indexes j and $j+1$, and making the difference, we obtain

$$\Delta p^{n-1}_{j+\frac{1}{2}} = \sum_{k=-K}^{K} C^{n-1}_{j,k} \Delta p^{n}_{j-k+\frac{1}{2}}. \tag{36}$$

Coefficients $B^n_{j,k}$ and $C^n_{j,k}$ characterize various stability properties for the adjoint scheme (29), which are given in the following two lemmas.

Lemma 1 *Assume that the coefficients $a^n_{j+\frac{1}{2},k}$ introduced in (9) are uniformly bounded, and that*

$$\forall (j,n) \in \mathbb{Z} \times \mathbb{N}, \qquad B^n_{j,k} \geq 0. \tag{37}$$

Then the following estimates hold for all $n \in \{0, \dots, N-1\}$:

$$\sup_{j\in\mathbb{Z}} |p^n_j| \quad \leq \quad \sup_{j\in\mathbb{Z}} |p^N_j|; \tag{38}$$

$$\forall\ J>0, \qquad \sum_{|j|\leq J} |p^{n-1}_j - p^n_j| \quad \leq \quad C\lambda \sum_{|j|\leq J} |p^n_j - p^n_{j-1}|. \tag{39}$$

Moreover, the scheme (29) is monotone.

Proof. Because of the formulation (32), the uniform bound on the size of the p^n_j is a straightforward consequence of relation (31) and the sign requirement (37). Now, for the equicontinuity in time, we notice that

$$|p^{n-1}_j - p^n_j| = \left| \sum_{k=-K}^{K} B^{n-1}_{j,k} p^n_{j-k} - \left(\sum_{k=-K}^{K} B^{n-1}_{j,k} \right) p^n_j \right| \leq \sum_{k=-K}^{K} B^{n-1}_{j,k} |p^n_{j-k} - p^n_j|.$$

We use now the standard triangular inequalities:

$$\left\{ \begin{array}{l} k>0: \ |p^n_{j-k} - p^n_j| \leq \sum_{l=1}^{k} |p^n_{j-l+1} - p^n_{j-l}| \\ k<0: \ |p^n_{j-k} - p^n_j| \leq \sum_{l=0}^{k-1} |p^n_{j-l+1} - p^n_{j-l}| \end{array} \right.$$

We plug this in the time variation expression and switch the j and l indices to get:

$$\sum_{|j|\leq J} |p^{n-1}_j - p^n_j| \leq \sum_{k=-K}^{-1} \sum_{l=0}^{k-1} \sum_{|j|\leq J} B^{n-1}_{j+l,k} |p^n_{j+1} - p^n_j| + \sum_{k=1}^{K} \sum_{l=1}^{k} \sum_{|j|\leq J} B^{n-1}_{j+l,k} |p^n_{j+1} - p^n_j|.$$

We move now the sum over the j's:

$$\begin{aligned} \sum_{|j|\leq J} |p^{n-1}_j - p^n_j| \quad &\leq \quad \sum_{|j|\leq J} \left(\sum_{l=0}^{-K+1} \sum_{k=-K}^{-1} B^{n-1}_{j+l,k} + \sum_{l=1}^{K} \sum_{k=l}^{K} B^{n-1}_{j+l,k} \right) |p^n_{j+1} - p^n_j| \\ &\leq \quad \lambda \sum_{|j|\leq J} \left(\sum_{l=0}^{-K+1} a^{n-1}_{j+\frac{1}{2},l} - \sum_{l=1}^{K} a^{n-1}_{j+\frac{1}{2},l} \right) |p^n_{j+1} - p^n_j|. \end{aligned}$$

Finally, this gives:

$$\sum_{|j|\leq J} |p^{n-1}_j - p^n_j| \leq 2K\lambda \sup_{k,j,n} |a^n_{j+\frac{1}{2},k}| \sum_{|j|\leq J} |p^n_{j+1} - p^n_j|.$$

Concerning monotonicity, we introduce the operator $H : \mathbb{R}^{2K} \to \mathbb{R}$ such that:

$$p^{n-1}_j = H(p^n_{j-K}, p^n_{j-K+1}, ..., p^n_{j+K})$$

Then, the partial derivatives of H are just given by the $B^n_{j,k}$ coefficients. Consequently, H is a monotone increasing function of each of its arguments under requirement (37). □

Lemma 2 *Assume that the coefficients $a^n_{j+\frac{1}{2},k}$ are uniformly bounded and that*

$$\forall (j,n) \in \mathbb{Z} \times \mathbb{N}, \qquad B^n_{j,k} \geq 0, \quad C^n_{j,k} \geq 0. \tag{40}$$

Then in addition to properties of Lemma 1, the dual scheme (29) satisfies the **backward** *TVD estimate*

$$\sum_{j\in\mathbb{Z}} |p^n_{j+1} - p^n_j| \leq \sum_{j\in\mathbb{Z}} |p^N_{j+1} - p^N_j|, \tag{41}$$

and preserves monotonicity.

Proof. The TVD property follows easily from the formulation (36), (34) and the sign requirement (40). Moreover, if we assume that each $\Delta p^n_{j+\frac{1}{2}} \geq 0$, then the formulation (36) implies that $\Delta p^{n-1}_{j+\frac{1}{2}} \geq 0$ as a convex combination of some positive quantities. This proves the two announced statements. □

Remark 2 *All the properties in Lemmas 1 and 2 are discrete analogues of those of reversible solutions. Schemes satisfying only (37) do not enjoy all the properties, in particular they lack the monotonicity preservation.*

3.3 Notion of consistency and convergence

We turn now to the definition of a notion of consistency for our schemes. Let us denote by μ^Δ, p^Δ the piecewise constant functions defined for all (t,x) respectively by μ^n_j, and p^n_j on each cell

$$T^n_j \stackrel{\text{def}}{=} [n\Delta t, (n+1)\Delta t[\times [(j-\frac{1}{2})\Delta x, (j+\frac{1}{2})\Delta x[.$$

We define also the following vector-valued function:

$$\begin{array}{rcll} \mathbf{A}^\Delta = (a^\Delta_k)_{k=-K+1,\dots,K} : [0,T]\times\mathbb{R} & \to & \mathbb{R}^{2K} & \\ (t,x) & \mapsto & \mathbf{A}^n_{j+\frac{1}{2}} & \text{for } (t,x) \in T^n_j, \end{array} \tag{42}$$

and we assume that, for a given $p^T \in \mathrm{Lip}(\mathbb{R})$, the discretization $(p^N_j)_{j\in\mathbb{Z}}$ satisfies

$$\sup_{j\in\mathbb{Z}} |\Delta p^N_{j+\frac{1}{2}}| \leq \Delta x \ \mathrm{Lip}(p^T). \tag{43}$$

This is achieved for instance by taking the local averages of p^T on cells. Finally, we shall need the functions a^Δ and b^Δ defined for $(x,t) \in T^n_j$ by

$$\begin{array}{rcl} a^\Delta(t,x) & = & \displaystyle\sum_{k=-K+1}^{K} a^\Delta_k(t,x), \\ b^\Delta(t,x) & = & \displaystyle\frac{1}{\Delta x}\sum_{k=-K+1}^{K} \left[a^\Delta_k(t,x+(k-1)\Delta x) - a^\Delta_k(t,x+(k-2)\Delta x)\right]. \end{array} \tag{44}$$

We can state now the most important definition:

Definition 7 *The scheme (9) is said to be* **weakly consistent** *with the continuous equation (1) if the coefficients $a^n_{j+\frac{1}{2},k}$ are uniformly bounded and*

(i) $$a^\Delta \rightharpoonup a \qquad \text{in } L^\infty - w\star \qquad \text{as } \Delta \to 0;$$

(ii) for each Δ, there exists $\alpha^\Delta \in L^1(]0,T[)$, with $\|\alpha^\Delta\|_1 \leq C$ uniformly in Δ, such that $b^\Delta(t,.) \leq \alpha^\Delta(t)$ for a.e. $t \in]0,T[$.

These assumptions are the discrete analogues of those in the stability result for reversible solutions (Theorem 3). From assertion (i) it follows by an easy computation that $b^\Delta \to \partial_x a$ in the sense of distributions. Assumption (ii) allows to precise this convergence: provided a satisfies (2), we have actually $b^\Delta \to \partial_x a$ for the weak topology of measures. This leads to the weak consistency for the backward problem and therefore to the following result.

Theorem 10 *Let p^T be a Lipschitz continuous function with Lipschitz constant $\mathrm{Lip}(p^T)$. Assume that the adjoint scheme is consistent and satisfies the positivity requirements (40). Then the sequence (p^Δ) converges as $\Delta \to 0$ in the strong topology of $L^1_{loc}(\Omega)$ and almost everywhere towards the reversible solution of the problem (7).*

Proof. We begin by a discrete analogue of the Lipschitz estimate (15). From (36) and the non negativity of the $C^{n-1}_{j,k}$'s, it follows

$$|\Delta p^{n-1}_{j+\frac{1}{2}}| \leq \sum_{k=-K}^{K} C^{n-1}_{j,k} |\Delta p^n_{j-k+\frac{1}{2}}| \leq M^n_K \sum_{k=-K}^{K} C^{n-1}_{j,k},$$

where for $q \in \mathbb{N}$, $M^n_q \equiv \sup_{-q\leq k\leq q} |\Delta p^n_{j-k+\frac{1}{2}}|$. Using now (35) and (37), we have

$$|\Delta p^{n-1}_{j+\frac{1}{2}}| \leq \left(1 + \lambda \sum_{k=-K}^{K-1} (a^{n-1}_{j-k+\frac{1}{2},k+1} - a^{n-1}_{j-k-\frac{1}{2},k+1})\right) M^n_K.$$

Going back to the definitions of a^Δ_k and b^Δ, the preceding inequality rewrites

$$|\Delta p^{n-1}_{j+\frac{1}{2}}| \leq \left(1 + \frac{\lambda}{\Delta t} \int_{t^{n-1}}^{t^n} \int_{\mathbb{R}} b^\Delta(t,x) \mathbf{1}_{]x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}[}(x) dx\, dt\right) M^n_K.$$

Assumption (ii) in Definition 7 gives therefore after an immediate induction

$$\begin{aligned} |\Delta p^{n-1}_{j+\frac{1}{2}}| &\leq \left(1 + \int_{t^{n-1}}^{t^n} \alpha^\Delta(t) dt\right) M^n_K \\ &\leq \prod_{q=n-1}^{N} \left(1 + \int_{t^{q-1}}^{t^q} \alpha^\Delta(t) dt\right) M^N_{(N-n+1)K}. \end{aligned}$$

But $\prod_{q=n-1}^N \left(1 + \int_{t^{q-1}}^{t^q} \alpha^\Delta(t) dt\right) \leq e^{\int_{t^{n-1}}^{t^N} \alpha^\Delta(t) dt} \leq e^C$ by the consistency assumption (ii). Thus we obtain the desired estimate: if $Q > 0$ is a given integer,

$$\sup_{-Q\leq j\leq Q} \frac{1}{\Delta x} |\Delta p^{n-1}_{j+\frac{1}{2}}| \leq e^C \sup_{-Q-(N-n+1)K\leq \ell\leq Q+(N-n+1)K-1} \frac{|\Delta p^N_{j-\ell+\frac{1}{2}}|}{\Delta x}. \tag{45}$$

Letting $\Delta \to 0$, $N \to +\infty$ and $\lim_n \prod_q \left(1 + \int_{t^{q-1}}^{t^q} \alpha^\Delta(t) dt\right) = e^{\int \alpha^\Delta(t) dt}$, so that we recover at the limit an analogue of (15). Thus, provided the sequence (p^n_j) converges, its limit is Lipschitz continuous.

We turn now to relative compactness. The former estimate readily gives, for any given $a < b$,

$$\|p^\Delta(t,.) - p^\Delta(t,. + \Delta x)\|_{L^1(]a,b[)} \leq \Delta x (b-a) e^C \mathrm{Lip}(p^T).$$

In the same way, we get from (39)

$$\|p^\Delta(t,.) - p^\Delta(t + \Delta t,.)\|_{L^1(]a,b[)} \leq \Delta x \Delta t (b-a) e^C \mathrm{Lip}(p^T).$$

Thus the sequence (p^Δ) is relatively compact in $L^1_{loc}(\Omega)$, so we have convergence, up to a subsequence, to some p which is Lipschitz continuous.

Next, p solves the backward equation. Indeed, if

$$p_t^\Delta = \frac{p_j^n - p_j^{n-1}}{\Delta t} \qquad \text{for} \quad (t,x) \in T_j^n,$$

it follows from the definition of the adjoint scheme and (45) that p_t^Δ is bounded in L^∞, so $p_t^\Delta \rightharpoonup \partial_t p$ in $L^\infty - w\star$. Then, we have

$$\begin{aligned} \sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k} \frac{p^n_{j-k} - p^n_{j-k-1}}{\Delta x} &= \frac{\displaystyle\sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k} p^n_{j-k} - \sum_{k=-K+1}^{K} a^{n-1}_{j-k-\frac{1}{2},k} p^n_{j-k-1}}{\Delta x} \\ &\quad - \sum_{k=-K+1}^{K} \frac{a^{n-1}_{j-k+\frac{1}{2},k} - a^{n-1}_{j-k-\frac{1}{2},k}}{\Delta x} p^n_{j-k-1} \end{aligned} \tag{46}$$

Setting $(ap)^\Delta = \sum_{j,n} \sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k} p^n_{j-k} \mathbf{1}_{T_j^n}$, we rewrite the first term in the right-hand side of (46) as $[(ap)^\Delta(t,x) - (ap)^\Delta(t,x-\Delta x)]/\Delta x$, which converges to $\partial_x(ap)$ provided $(ap)^\Delta \rightharpoonup ap$. But

$$(ap)^\Delta(t,x) = \sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k} p^n_{j-k} + \sum_{k=-K+1}^{K} a^{n-1}_{j-k+\frac{1}{2},k} (p^n_{j-k} - p^n_j) \text{ for } (t,x) \in T_j^n$$

The first term tends to ap in $\mathcal{D}'$ by the consistency assumption on a^Δ and the bounds on p^Δ, the second tends to 0 because of the Lipschitz estimate

$$|p^n_{j-k} - p^n_j| \leq K \Delta x \, e^C \mathrm{Lip}(p^T), \tag{47}$$

and the boundedness of $a^{n-1}_{j-k+\frac{1}{2},k}$. The same trick allows us to rewrite the second term in (46) as

$$(b^\Delta p^\Delta)(t,x) + \sum_{k=-K+1}^{K} \frac{a^{n-1}_{j-k+\frac{1}{2},k} - a^{n-1}_{j-k-\frac{1}{2},k}}{\Delta x} (p^n_{j-k} - p^n_j).$$

Assumption (ii) in Definition 7 leads to $b^\Delta \to \partial_x a$ in the weak sense of measures, and p^Δ is a uniformly bounded Borel function, so $b^\Delta p^\Delta \to \partial_x a \cdot p$ in the sense of distributions. The second term is handled in the same way, since (47) holds for any (t,x) and the remaining coefficient is a bounded measure.

So far, we proved that, up to a subsequence, $(p_j^n)_{j,n}$ converges strongly to a Lipschitz continuous solution to (7). To prove that p is reversible, which will lead by uniqueness to the convergence of the whole sequence, we remark that by construction the adjoint scheme preserves monotonicity (Lemma 2). Thus, if we split $p^T = p_1^T - p_2^T$, with $\partial_x p_i^T \geq 0$, and denote p_i^Δ the discrete solution computed by (9), then

(i) $p_i^\Delta \to p_i$ Lipschitz solution to (7);
(ii) $\partial_x p_i^\Delta \geq 0$ by Lemma 2, so that $\partial_x p_i \geq 0$;
(iii) $p^\Delta = p_1^\Delta - p_2^\Delta \to p = p_1 - p_2$ by linearity.

So p is reversible by the second characterization of Theorem 2. □

Remark 3 *Theorem 10 actually gives an alternative proof for the existence of reversible solutions to (7).*

Theorem 11 *Assume that the adjoint scheme is consistent and satisfies the positivity requirements (40). Then the sequence* (μ^Δ) *converges as* $\Delta \to 0$ *in the weak topology of* $\mathcal{M}(\Omega)$ *towards the duality solution of the problem (1).*

Proof. By its formulation, the scheme (9) is conservative and consequently μ^Δ is endowed with a uniform bound in $\mathcal{M}(\Omega)$. So, up to a subsequence, we have:

$$\begin{cases} \mu^\Delta \rightharpoonup \mu \text{ in the weak } \star \text{ topology of } \mathcal{M}(\Omega), \\ \dfrac{d}{dt}\displaystyle\int_{\mathbb{R}} p^\Delta(t,x)\mu^\Delta(t,dx) = 0, \end{cases}$$

which means that μ^Δ converges towards the unique duality solution of (1). By the classical uniqueness argument, the whole sequence is convergent. □

3.4 Convergence for the associated transport equation

This subsection is devoted to the study of some numerical schemes for the transport equation (3). We introduce some schemes which are in a way "integrated versions" of the conservative schemes (9), and prove the convergence to the duality solution to (3). As a corollary, we shall recover some convergence results of the "discrete product" of a by μ towards the product $\hat{a}\mu$.

Concerning the proofs, we shall limit ourselves to the nice case where the coefficients $C^n_{j,k}$ defined by (33) are nonnegative. The Lax-Friedrichs type schemes do not fall in this category, but for the sake of brevity, and in view of their poor numerical behaviour, we do not wish to state the proofs here. Let us now be more specific.

We consider the following scheme

$$u_j^{n+1} = u_j^n - \lambda < \mathbf{A}^n_{j+\frac{1}{2}}, \vec{\Delta} u_j^n >_{\mathbb{R}^{2K}}, \qquad \text{with} \quad \vec{\Delta} u_j^n = \left(u^n_{j+k} - u^n_{j+k-1}\right)_{k=-K+1,\dots,K}, \tag{48}$$

and denote by u^Δ the corresponding constant by cell function. We first notice that, setting

$$\mu_j^n = \frac{u^n_{j+1} - u^n_j}{\Delta x}, \tag{49}$$

a simple computation shows that μ_j^n is given by the conservative scheme (9). This is the discrete analogue of Proposition 1. Thus, formally, we pass from nonconservative to conservative by discrete differentiation, and interpret μ_j^n as a numerical approximation of $(u^n_{j+1} - u^n_j)\delta_{x_{j+\frac{1}{2}}}$, which is related to $\partial_x u^\Delta$.

Theorem 12 *Assume that the positivity and the consistency requirements of Lemma 2 and Definition 7 are met, then the sequence* (u^Δ) *converges as* $\Delta \to 0$ *towards the unique duality solution of the equation (3) in the strong topology of* $L^1_{loc}(\Omega)$.

Proof. We merely give a sketch of the proof, since the arguments used here are very similar to those in the proof of Theorems 10 and 11.

First, the scheme (48) is by construction endowed with a uniform BV bound, so that the function u^Δ belongs to $L^\infty(0,T;BV(\mathbb{R}))$ as soon as we assume the initial sequence $(u_j^0)_{j\in\mathbb{Z}}$ to be bounded in total variation. We immediately deduce that the family $(u^\Delta)_{\Delta\to 0}$ is relatively compact in the strong topology of $L^1_{loc}(]0,T[\times\mathbb{R})$ and almost everywhere convergent up to the extraction of a subsequence. Therefore, we are done as soon as we prove that $t \mapsto \int u\pi\, dx$ is constant for any reversible compactly supported π.

Therefore, as for the conservative case, we introduce the adjoint scheme by imposing

$$\forall \quad 1 \le n \le N, \qquad \sum_{j\in\mathbb{Z}} \pi_j^n u_j^n = \sum_{j\in\mathbb{Z}} \pi_j^{n-1} u_j^{n-1}. \tag{50}$$

A straightforward computation leads to the following scheme

$$\begin{aligned}\pi_j^{n-1} &= \pi_j^n - \lambda\Big(\sum_{k=-K+1}^{K} a^{n-1}_{j-k-\frac{1}{2},k}\pi^n_{j-k} - \sum_{k=-K}^{K-1} a^{n-1}_{j-k-\frac{1}{2},k+1}\pi^n_{j-k}\Big)\\ &= \sum_{k=-K}^{K} C^{n-1}_{j-1,k}\pi^n_{j-k}.\end{aligned}$$

Under the boundedness and nonnegativity requirements on $C^{n-1}_{j-1,k}$, this scheme is clearly bounded in L^∞ and preserves nonnegativity. Since the corresponding constant by cell function π^Δ is L^∞ bounded, up to a subsequence, π^Δ converges to some π in $L^\infty - w\star$. The consistency requirements imply that π solves the backward equation (10), as in the proof of Theorem 10. Finally, π is reversible, since the positivity is preserved, and using the third characterization in Theorem 4.

Passing to the limit in (50), we obtain that u^Δ converges to the duality solution. □

Corollary 1 *Set* $(a\mu)^\Delta(t,x) = \sum_{j,n} < \mathbf{A}^n_{j+\frac{1}{2}}, \vec{\mu}^n_{j+\frac{1}{2}} >_{\mathbb{R}^{2K}} \mathbf{1}_{T^n_j}(t,x)$. *Then, under the assumptions of Theorem 12,*

$$(a\mu)^\Delta \longrightarrow a \vartriangle \mu = \hat{a}\mu \qquad in \quad \mathcal{D}'(\Omega).$$

Proof. First notice that $u^\Delta_x \equiv \sum_{j,n}(u^n_{j+1} - u^n_j)/\Delta x \mathbf{1}_{T^n_j}$ converges in $\mathcal{D}'(\Omega)$ to $\partial_x u$, and that, by Proposition 1, $\partial_x u$ solves (1) in the sense of duality. On the other hand, by construction, μ^Δ defined by (9) tends to μ, which is also duality solution to (1). Since, at $t = 0$, $\mu(0,\cdot) = \partial_x u(0,\cdot)$, we have by uniqueness $\mu = \partial_x u$. This justifies the "discrete differentiation" of the scheme.

Finally, we notice that $u^\Delta_t \equiv \sum_{j,n}(u^{n+1}_j - u^n_j)/\Delta t \mathbf{1}_{T^n_j}$ converges in $\mathcal{D}'(\Omega)$ to $\partial_t u$. But, on the one hand, by definition of the flux, $\partial_t u = -a \vartriangle \mu = \hat{a}\mu$, and on the other hand, $u^\Delta_t = -(a\mu)^\Delta$ by construction of the scheme. Thus we are done. □

4 Some classical examples

The aim of this section is to illustrate the preceding results on a few examples from the usual literature. Obviously, we do not pretend to exhaustivity. In the following, we choose for a^n_j

$$a^n_j = \frac{1}{\Delta x \Delta t}\iint_{\mathbb{R}_+\times\mathbb{R}} a(t,x)\mathbf{1}_{T^n_j} dx\, dt. \tag{51}$$

This is justified and natural since the only assumption on a is an L^∞ bound.

Remark 4 *Notice that for the function* $\bar{a}^\Delta(t,x) = \sum_{j,n} a^n_j \mathbf{1}_{T^n_j}(t,x)$ *converges a.e. to* a *and is bounded in* L^∞, *so that* $\bar{a}^\Delta \rightharpoonup a$ *in* $L^\infty - w\star$. *Moreover, since for a.e.* t, $\partial_x a(t,.)$ *is a locally bounded measure, we have for a.e.* x

$$a(t,x) - a(t,x-\Delta x) = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \partial_x a(t,d\xi) \le \Delta x\, \alpha(t).$$

Thus $\forall\, j \in \mathbb{Z}$ *and a.e.* $t \in]t^n, t^{n+1}[$, $a^n_j - a^n_{j-1} \le \Delta x\, \alpha(t)$, *and also* $(a^n_j - a^n_{j-1})^+ \le \Delta x\, \alpha(t)$ *(*α *in (2) can always be chosen nonnegative).*

4.1 Lax-Friedrichs type schemes

The most encountered first-order discretizations belonging to this family correspond to constant values for the viscosity coefficient $\varepsilon^n_{j+\frac{1}{2}}$. We first give a general consistency result. We recall from the preceding section that we have:

$$\mathbf{A}^n_{j+\frac{1}{2}} = \frac{1}{2}\left(a^n_j + \frac{\varepsilon^n_{j+\frac{1}{2}}}{\lambda}, a^n_{j+1} - \frac{\varepsilon^n_{j+\frac{1}{2}}}{\lambda}\right)$$

This choice leads to the following coefficients

$$\left\{\begin{array}{l} B^n_{j,-1} = C^n_{j,-1} = \dfrac{\lambda}{2}\left(a^n_{j+1} + \dfrac{\varepsilon^n_{j+\frac{3}{2}}}{\lambda}\right) \\ B^n_{j,0} = 1 - \dfrac{\varepsilon^n_{j+\frac{1}{2}} + \varepsilon^n_{j-\frac{1}{2}}}{2} \\ C^n_{j,0} = 1 + \dfrac{\lambda}{2}\left(a^n_{j+1} - a^n_j\right) - \varepsilon^n_{j+\frac{1}{2}} \\ B^n_{j,1} = C^n_{j,1} = -\dfrac{\lambda}{2}\left(a^n_j - \dfrac{\varepsilon^n_{j-\frac{1}{2}}}{\lambda}\right) \end{array}\right.$$

Lemma 3 *The Lax-Friedrichs type schemes (24) are weakly consistent in the sense of Definition 7 under the condition*

$$\exists M > 0, \qquad \forall \quad 0 \le n \le N, \quad j \in \mathbb{Z}, \qquad -\varepsilon^n_{j+\frac{1}{2}} + 2\varepsilon^n_{j-\frac{1}{2}} - \varepsilon^n_{j-\frac{3}{2}} \le M\Delta x. \tag{52}$$

Proof. We have, for $(t,x) \in T^n_j$, $a^\Delta(t,x) = \frac{1}{2}(a^n_j + a^n_{j+1})$, which tends to a in $L^\infty w\star$ by construction of the a^n_j's, so that the first requirement of Definition 7 is met. Next, for $(t,x) \in T^n_j$, a simple computation gives

$$b^\Delta(t,x) = \frac{1}{2\Delta x}\left[(a^n_{j+1} - a^n_j) + (a^n_j - a^n_{j-1}) + \frac{1}{\lambda}(-\varepsilon_{j+\frac{1}{2}} + 2\varepsilon_{j-\frac{1}{2}} - \varepsilon_{j-\frac{3}{2}})\right].$$

From condition (52) and Remark 4, we obtain that b^Δ satisfies the second requirement of Definition 7, with $\alpha^\Delta = \alpha + M/(2\lambda)$. This concludes the proof. □

We are going to state two convergence theorems. The first one is a direct consequence of the general results of the previous section, but needs a restrictive CFL condition. In order to relax this assumption, we have to strengthen the constraints on $\varepsilon^n_{j+\frac{1}{2}}$. We present the proofs of these results for the sake of completeness, but we do not wish to search for optimal conditions, since there is a numerical evidence of the bad quality achieved by Lax-Friedrichs type schemes in this context (see Section 5).

Proposition 2 *The scheme (24) converges towards the duality solution of (1) as $\Delta \to 0$, under the consistency condition (52), provided (51) is chosen and the following conditions are met:*

$$\forall (j,n) \in \mathbb{Z} \times \mathbb{N}, \qquad \lambda|a^n_j| \le \varepsilon^n_{j\pm\frac{1}{2}} \le 1, \qquad \lambda(a^n_j - a^n_{j+1})/2 \le 1 - \varepsilon^n_{j+\frac{1}{2}}. \tag{53}$$

Proof. The proof is an immediate consequence of Theorem 11, since the conditions in (53) exactly imply the positivity requirements on the coefficients $B^n_{j,k}$ and $C^n_{j,k}$. □

The second requirement of (53) cannot be met if, for instance, $\varepsilon^n_{j+\frac{1}{2}} \equiv 1$ and $x \mapsto a(x,t)$ is a decreasing function. To fix this drawback, we also propose an alternative result

Proposition 3 *The scheme (24) converges towards the duality solution of (1) as $\Delta \to 0$, under the consistency condition (52), provided (51) is chosen and the following conditions are met:*

$$\forall (j,n) \in \mathbb{Z} \times \mathbb{N}, \qquad \lambda |a_j^n| \leq \varepsilon^n_{j\pm\frac{1}{2}} \leq 1, \qquad \varepsilon^n_{j+\frac{1}{2}} \equiv \varepsilon^n. \tag{54}$$

Proof. First we notice that (54) implies $B^n_{j,k} \geq 0$, so that Lemma 1 applies. However, as noticed before, we do not have $C^n_{j,0} \geq 0$. We have therefore to prove first that the discrete Lipschitz estimate holds, then that we can recover monotonicity preservation.

Concerning the Lipschitz estimate, we have from (36) that

$$\begin{aligned}\Delta p^{n-1}_{j+\frac{1}{2}} &= \frac{\lambda}{2}\left(a^{n-1}_{j+1} + \frac{\varepsilon^{n-1}_{j+\frac{3}{2}}}{\lambda}\right)\Delta p^n_{j+\frac{3}{2}} + \left[1 + \frac{\lambda}{2}(a^{n-1}_{j+1} - a^{n-1}_j) - \varepsilon^{n-1}_{j+\frac{1}{2}}\right]\Delta p^n_{j+\frac{1}{2}} \\ &\qquad - \frac{\lambda}{2}\left(a^{n-1}_j - \frac{\varepsilon^{n-1}_{j-\frac{1}{2}}}{\lambda}\right)\Delta p^n_{j-\frac{1}{2}}.\end{aligned}$$

Using the first requirement in (54), we can write

$$\begin{aligned}|\Delta p^{n-1}_{j+\frac{1}{2}}| &\leq \frac{\lambda}{2}\left(a^{n-1}_{j+1} + \frac{\varepsilon^{n-1}_{j+\frac{3}{2}}}{\lambda}\right)|\Delta p^n_{j+\frac{3}{2}}| + \left|1 + \frac{\lambda}{2}(a^{n-1}_{j+1} - a^{n-1}_j) - \varepsilon^{n-1}_{j+\frac{1}{2}}\right||\Delta p^n_{j+\frac{1}{2}}| \\ &\qquad - \frac{\lambda}{2}\left(a^{n-1}_j - \frac{\varepsilon^{n-1}_{j-\frac{1}{2}}}{\lambda}\right)|\Delta p^n_{j-\frac{1}{2}}| \\ &\leq \left[\frac{\lambda}{2}\left(|a^{n-1}_{j+1} - a^{n-1}_j| + (a^{n-1}_{j+1} - a^{n-1}_j)\right) + \left(1 - \varepsilon^{n-1}_{j+\frac{1}{2}} + \frac{\varepsilon^{n-1}_{j+\frac{3}{2}} + \varepsilon^{n-1}_{j-\frac{1}{2}}}{2}\right)\right] M^n_1,\end{aligned}$$

with the notations of Theorem 10. We have for a.e. t

$$\frac{1}{2}\left(|a^{n-1}_{j+1} - a^{n-1}_j| + (a^{n-1}_{j+1} - a^{n-1}_j)\right) = (a^{n-1}_{j+1} - a^{n-1}_j)^+ \leq \Delta x\, \alpha(t).$$

We can proceed as in the proof of Theorem 10 if for all j

$$\varepsilon^{n-1}_{j+\frac{3}{2}} + \varepsilon^{n-1}_{j-\frac{1}{2}} \leq 2\,\varepsilon^{n-1}_{j+\frac{1}{2}}.$$

An easy computation shows that this is possible only if the sequence is constant, because $\varepsilon^{n-1}_{j+\frac{1}{2}} \geq 0$. If the second condition in (54) holds, we have therefore

$$|\Delta p^{n-1}_{j+\frac{1}{2}}| \leq \left(1 + \int_{t^{n-1}}^{t^n} \alpha(t)\,dt\right) M^n_1,$$

and we obtain the final estimate exactly as in the proof of Theorem 10.

So far, we know that, up to a subsequence, p^Δ tends to p, Lipschitz solution to (7). We want now to prove that p is the reversible solution. Since the scheme does not preserve monotonicity *a priori*, we have to be a bit more careful.

We have by (36) and (35)

$$\Delta p^{n-1}_{j+\frac{1}{2}} = \sum_{k=-K}^{K} B^{n-1}_{j,k}\Delta p^n_{j-k+\frac{1}{2}} + \sum_{k=-K}^{K-1} \lambda\left(a^{n-1}_{j-k+\frac{1}{2},k+1} - a^{n-1}_{j-k-\frac{1}{2},k+1}\right)\Delta p^n_{j-k+\frac{1}{2}}. \tag{55}$$

First consider (55) for $n = N$. Assuming $\Delta p^N_{j+\frac{1}{2}} \geq 0$ for all j's (which is achieved by a suitable discretization if $\partial_x p^T \geq 0$), since $B^{N-1}_{j,k} \geq 0$, we have

$$\Delta p^{N-1}_{j+\frac{1}{2}} \geq \lambda \left(a^{N-1}_{j-k+\frac{1}{2},k+1} - a^{N-1}_{j-k-\frac{1}{2},k+1} \right) \Delta p^N_{j-k+\frac{1}{2}}. \tag{56}$$

But, by (43), $\sup_j |\Delta p^N_{j+\frac{1}{2}}| = O(\Delta x)$, and, on the other hand, $|a^{N-1}_{j-k+\frac{1}{2},k+1} - a^{N-1}_{j-k-\frac{1}{2},k+1}| \leq C$ since the coefficients are bounded. This implies that the right-hand side of (55) is larger than a $O(\Delta x)$.

We proceed now by induction, and assume that for some $n \leq N$, $\inf_j \Delta p^n_{j+\frac{1}{2}} \geq O(\Delta x)$. The first term in the right-hand side of (55) is larger than a $O(\Delta x)$ since $B^{n-1}_{j,k} \geq 0$ and $\sum_k B^{n-1}_{j,k} = 1$. The second term is treated exactly as above, since the Lipschitz estimate gives for all n's

$$\sup_{j\in\mathbb{Z}} |\Delta p^n_{j+\frac{1}{2}}| \leq \sup_{j\in\mathbb{Z}} |\Delta p^N_{j+\frac{1}{2}}| = O(\Delta x).$$

We can conclude now, because if we set for $(t,x) \in T^n_j$, $p^\Delta_x \equiv \Delta p_{j-\frac{1}{2}}/\Delta x$, then, up to a subsequence, $p^\Delta_x \rightharpoonup \partial_x p$ in $L^\infty - w\star$ as $\Delta \to 0$, so that the limit p of p^Δ satisfies $\partial_x p \geq 0$. □

4.2 Upwind schemes

From the expression (25), one sees that the keypoint is in the determination of the vector $\mathbf{A}^n_{j+\frac{1}{2}}$ once the a^n_j's are fixed. The simplest choice is as follows:

$$\mathbf{A}^n_{j+\frac{1}{2}} = \left((a^n_j)^+, (a^n_{j+1})^- \right). \tag{57}$$

This scheme can be interpreted as an adaptation of the classical Engquist-Osher scheme [13] to the linear case. One notices that the corresponding scheme is not consistent with the continuous problem in the usual sense of Taylor expansions as soon as the coefficient a encounters a change of its sign. Anyway, we have the following consequence of Theorem 11.

Proposition 4 *The upwind discretization given by (25), (57) is consistent with the continuous equation (1) provided (51) is chosen. Moreover, it converges towards its unique duality solution as Δ goes to zero under the CFL condition:*

$$\forall (j,n) \in \mathbb{Z}\times\mathbb{N}, \qquad \lambda |a^n_j| \leq \frac{1}{2}. \tag{58}$$

Proof. We check the sign of the following coefficients:

$$\left\{ \begin{array}{rcll} B^n_{j,-1} & = & C^n_{j,-1} = \lambda (a^n_j)^+ & \geq 0 \\ B^n_{j,0} & = & 1 + \lambda[(a^n_j)^- - (a^n_j)^+] & \\ C^n_{j,0} & = & 1 + \lambda[(a^n_{j+1})^- - (a^n_j)^+] & \\ B^n_{j,1} & = & C^n_{j,1} = -\lambda (a^n_j)^- & \geq 0 \end{array} \right.$$

The second and third expressions are positive under the restriction (58). On the other hand, the consistency requirements of Definition 7 are met with for instance $\alpha^\Delta = 2\alpha$. □

According to [11, 12], another possibility is to use the following average values which correspond to Vol'pert's superposition product [30] (or the straight lines regularization in [9]):

$$\mathbf{A}^n_{j+\frac{1}{2}} = \left(\frac{(a^n_j + a^n_{j+1})^+}{2}, \frac{(a^n_j + a^n_{j+1})^-}{2} \right) \tag{59}$$

We shall consider more general upwind schemes defined by, for any given number $\theta \in [0,1]$,

$$\mathbf{A}^n_{j+\frac{1}{2}} = \left(((1-\theta)a^n_j + \theta a^n_{j+1})^+, ((1-\theta)a^n_j + \theta a^n_{j+1})^-\right). \tag{60}$$

This definition has to be compared with the last assertion of Theorem 9. The definition of the scheme defines in some way the value of a everywhere, and for $\theta \in [0,1]$ this is coherent with (22).

Proposition 5 *The upwind discretizations given by (25), (60) are consistent with the continuous equation (1) provided (51) is chosen. Moreover, they converge towards its unique duality solution as Δ goes to zero under the CFL condition:*

$$\forall (j,n) \in \mathbb{Z} \times \mathbb{N}, \qquad \lambda |a^n_j| \leq \frac{1}{2}. \tag{61}$$

Proof. In this case, we have the following quantities:

$$\left\{ \begin{array}{rcll} B^n_{j,-1} & = & C^n_{j,-1} = \lambda((1-\theta)a^n_j + \theta a^n_{j+1})^+ & \geq 0 \\ B^n_{j,0} & = & 1 + \lambda[((1-\theta)a^n_{j-1} + \theta a^n_j)^- - ((1-\theta)a^n_j + \theta a^n_{j+1})^+] & \\ C^n_{j,0} & = & 1 - \lambda|\theta a^n_{j+1} + (1-\theta)a^n_j| & \\ B^n_{j,1} & = & C^n_{j,1} = -\lambda((1-\theta)a^n_{j-1} + \theta a^n_j)^- & \geq 0 \end{array} \right.$$

The second and third expressions are positive under the restriction (61). The two consistency requirements of Definition 7 are again met for $\alpha^\Delta = 2\alpha$. □

We mention a variant of the preceding schemes, which is used by Olazabal [24] and Godlewski et *al.* [15]. They consider the convex nonlinear equation (4) with an entropy initial datum, for which a Roe type scheme is used. In this context, it is well-known that the scheme converges almost everywhere towards the entropy solution of the problem; moreover it satisfies a uniform discrete one-sided Lipschitz condition (see [7]). Next, they linearize this equation, obtaining (5), and propose the following "linearized Roe scheme" to solve it: let us denote by $\bar{a}$ a Roe linearized of f, and set $\bar{a}^n_{j+\frac{1}{2}} = \bar{a}(u^n_j, u^n_{j+1})$. Then the scheme is exactly the preceding one, with $\bar{a}^n_{j+\frac{1}{2}}$ playing the same role as $a^n_{j+\frac{1}{2}}$. Thus

$$\mathbf{A}^n_{j+\frac{1}{2}} = \left(\frac{\bar{a}(u^n_j, u^n_{j+1})^+}{2}, \frac{\bar{a}(u^n_j, u^n_{j+1})^-}{2} \right). \tag{62}$$

The stability analysis (nonnegativity of $B^n_{j,k}, C^n_{j,k}$) follows exactly as before. Concerning the consistency, the strong convergence of u^n_j implies the convergence of a^Δ, and the one-sided Lipschitz property provides the required bound on b^Δ.

5 Numerical results

We illustrate in this section the behaviour of the four schemes studied in Section 3, on five test cases. For three of them (sections 5.1, 5.2 and 5.3), uniqueness is ensured, and all the schemes converge towards the duality solution. Then, considering the associated transport equation, one can compute explicitly the exact solution. In the last two cases (sections 5.4 and 5.5), uniqueness does not hold, and it is clearly evidenced that each scheme chooses its own solution.

All the computations have been performed using a CFL condition of $\frac{1}{2}$, except the last case, where other values are interesting to consider. In all the figures, we shall have the following conventions

upw	upwind scheme (25)
EFO	modified upwind scheme (59)
LxF	standard Lax-Friedrichs scheme ($\varepsilon^n_{j+\frac{1}{2}} \equiv 1$)
Tad	modified Lax-Friedrichs scheme ($\varepsilon^n_{j+\frac{1}{2}} \equiv \frac{1}{2}$)

Finally, the numerical approximation of a Dirac mass has been chosen as $1/\Delta x$ on the appropriate cell.

5.1 Approximation of a Dirac mass in the compressive case

We consider here $a(t,x) = -\operatorname{sgn}(x-\frac{1}{2})$ for all t. The initial datum is $\mu_0(x) = \mathbf{1}_{x\leq\frac{1}{2}}$. In this case, the exact solution is $\frac{1}{2}\delta_{x=\frac{1}{2}}$. We choose $\Delta x = 0.002$. The approximate solutions are displayed in Figure 1 and we also present the numerical primitives in Figure 2 in order to show that the weight of the Dirac mass is correctly computed.

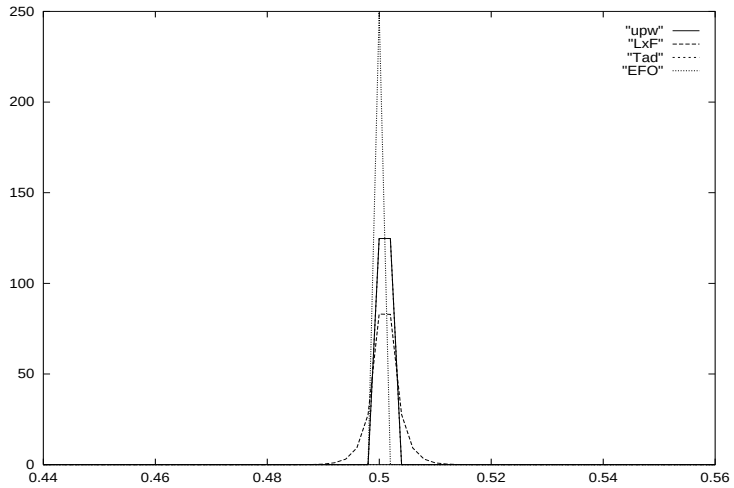

Figure 1: Numerical solutions in the case $a(t,x) = -\operatorname{sgn}(x-\frac{1}{2})$

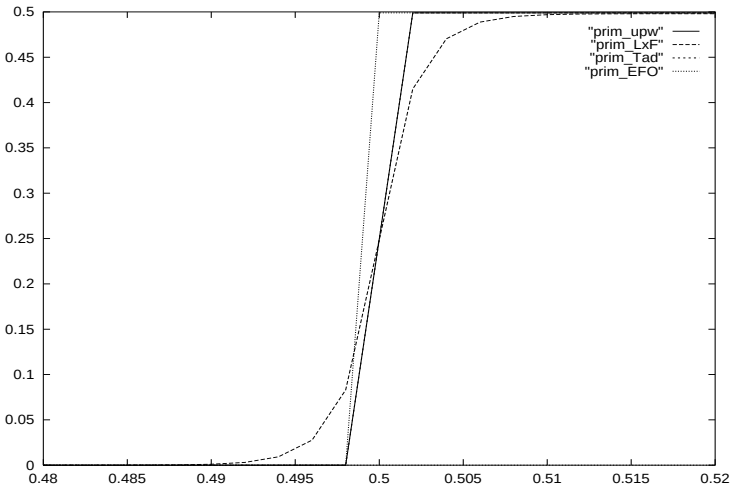

Figure 2: Numerical primitives in the case $a(t,x) = -\operatorname{sgn}(x-\frac{1}{2})$

5.2 Lipschitz expansive coefficient and smooth initial datum

We turn now to a smooth coefficient $a(t,x) = x - \frac{1}{2}$ for all t, and $\mu_0(x) = \sin(\pi x)\mathbf{1}_{x\in[0,1]}$. The exact solution is given by

$$\mu(t,x) = \frac{\mu_0(x + \frac{t}{2})}{1+t}. \tag{63}$$

This example clearly evidences the lack of "strong" consistency in this theory. Indeed the Engquist-Osher upwind scheme (57) exhibits a spurious spike at the point where a changes sign. This spike is concentrated on one cell, and is of bounded amplitude. Thus we clearly have only a weak convergence. A similar phenomenon was observed by Engquist and Runborg in the simulation of two-dimensional geometrical optics (see [14]). The modified version proposed in ([12, 11]) is better suited in this case. The Lax-Friedrichs type schemes behave in the same way as the scheme (59), so that we only display the results for the upwind type schemes. The solution is given at time $T = 3$ in figure 3.

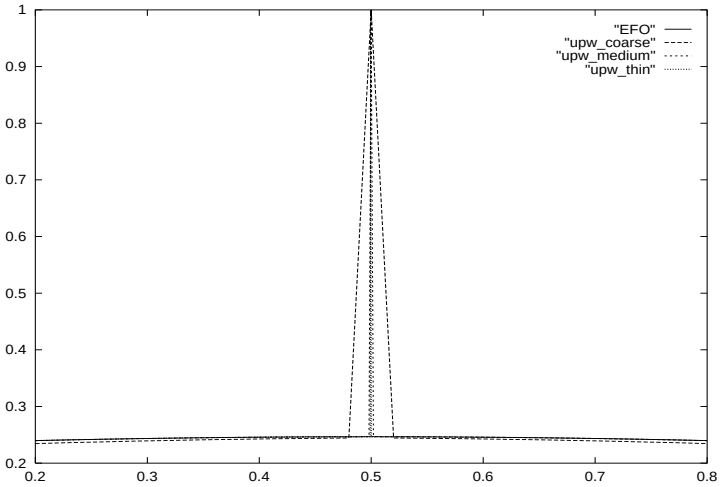

Figure 3: Upwind schemes in the case $a(t,x) = x - \frac{1}{2}$ for $\Delta x = 0.02, 0.002, 0.0005$

5.3 Lipschitz expansive coefficient and Riemann initial datum

We keep on using a smooth coefficient $a(t,x) = x - \frac{1}{2}$ for all t, but we consider now a Riemann initial datum $\mu_0(x) = \mathbf{1}_{x\leq\frac{1}{2}}$. The exact solution is again given by (63). We display the results obtained by both upwind schemes (57) and (59) with $\Delta x = 0.002$ in Figure 4. The solution is given at time $T = 3$ and is free from any spurious oscillation or numerical diffusion.

However, considering the results obtained by the LxF schemes displayed in Figure 5, one notices an excessive numerical dissipation creating an artificial profile which length shrinks to zero as $\Delta x \to 0$. Moreover, the approximate solution generated by the LxF scheme is endowed with oscillations whose amplitudes decrease also to zero as we refine the grid.

5.4 Spreading of a Dirac mass by a rarefaction

We turn now to the nonuniqueness cases starting with the conservative version of the first example presented in [2], Section 3.1. This corresponds to the following problem:

$$a(t,x) = \begin{cases} -1 & \text{if} \quad x - \frac{1}{2} \leq -t, \\ \dfrac{x-\frac{1}{2}}{t} & \text{if} \quad -t \leq x - \frac{1}{2} \leq 0, \\ 0 & \text{if} \quad x - \frac{1}{2} \geq 0, \end{cases}$$

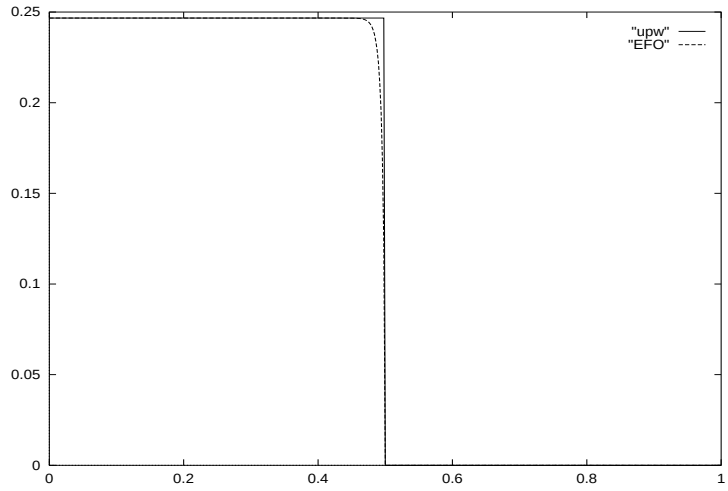

Figure 4: Numerical solutions with upwind schemes in the case $a(t,x) = x - \frac{1}{2}$

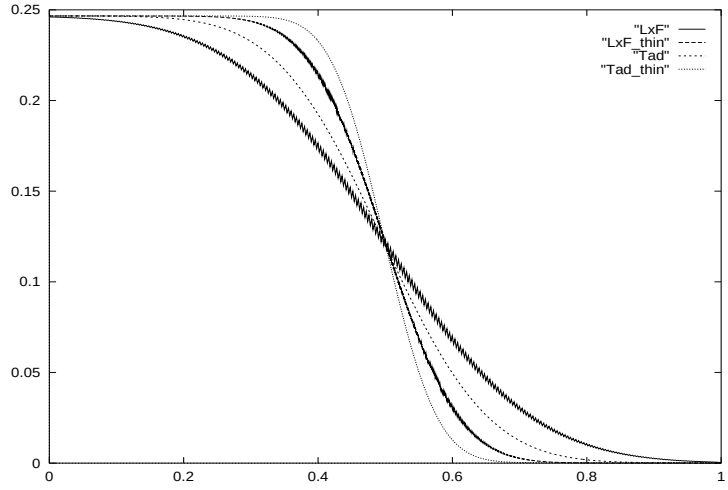

Figure 5: Numerical solutions with LxF schemes in the case $a(t,x) = x - \frac{1}{2}$ and $\Delta x = 0.002, 0.0005$

with the initial datum $\mu_0(x) = \delta_{x=\frac{1}{2}}$. For any $\varphi \in BV(]-1,0[)$ we define for $t > 0$

$$u(t,x) = \begin{cases} -1 & \text{if} \quad x - \frac{1}{2} < -t, \\ \varphi(\dfrac{x-\frac{1}{2}}{t}) & \text{if} \quad -t < x - \frac{1}{2} < 0, \\ 0 & \text{if} \quad x > 0. \end{cases}$$

Then $\mu = \partial_x u$ belongs to $\mathcal{S}_{\mathcal{M}}$ for any $T > 0$ and solves (1) in $]0,\infty[\times\mathbb{R}$.

Two computations are displayed here at time $T = 0.1$, the first on a medium mesh ($\Delta x = 0.002$), the second on a refined mesh ($\Delta x = 0.0005$). The first remark is that the solution generated by the standard Lax-Friedrichs scheme is highly oscillating, while Tadmor's modification behaves nicely (see Figure 6). The Dirac mass is spread in a more or less symmetric way. The upwind type schemes are not displayed here: they give a good approximation of a solution which is the Dirac mass at $x = \frac{1}{2}$.

It is more interesting to show the primitives of the solutions, especially to understand the behaviour of the schemes when we refine the mesh, see Figure 7. It becomes clear that, on the medium grid, the most important phenomenon for Lax-Friedrichs type schemes is the numerical

diffusion. Indeed, since the velocity on the right is zero, no information should be present for $x > \frac{1}{2}$, and the profiles are symmetrical. When refining the mesh, this phenomenon disappears, but it is not clear at all that the schemes converge to the Dirac mass at $x = \frac{1}{2}$; it is not even clear that they converge to the same solution.

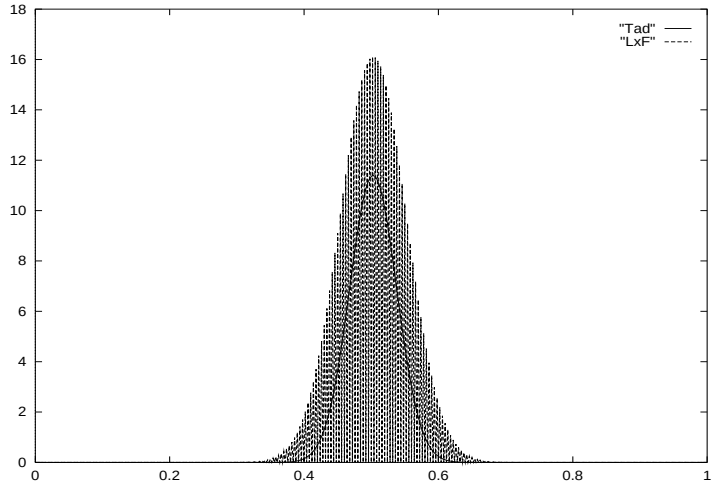

Figure 6: Numerical solutions for a rarefaction, LxF schemes, $\Delta x = 0.002$

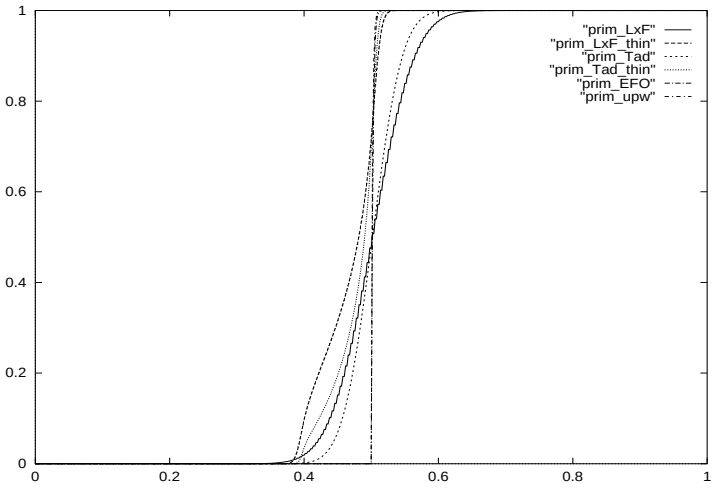

Figure 7: Numerical primitives for a rarefaction with $\Delta x = 0.002, 0.0005$

5.5 Spreading of a Dirac mass by a wildly expansive coefficient

By a wildly expansive coefficient, we mean a discontinuous coefficient which does not satisfy the OSLC condition (2). A typical example is $a(t,x) = \mathrm{sgn}(x - \frac{1}{2})$, and we take for initial datum $\mu_0 = \delta_{x=\frac{1}{2}}$. First we present a set of numerical solutions with $\Delta x = 0.0025$, and $\Delta t = 0.001$ in Figure 8. When refining the grid, the oscillations in Lax-Friedrichs remain as it might be expected considering the proof of Proposition 2, where the role of the OSLC condition is crucial.

Next, we play with the value of the Courant number for the modified Tadmor scheme [27]. It turns out that each CFL number determines a spreading of the Dirac mass, which clearly illustrate the lack of uniqueness in this problem, see Figure 9.

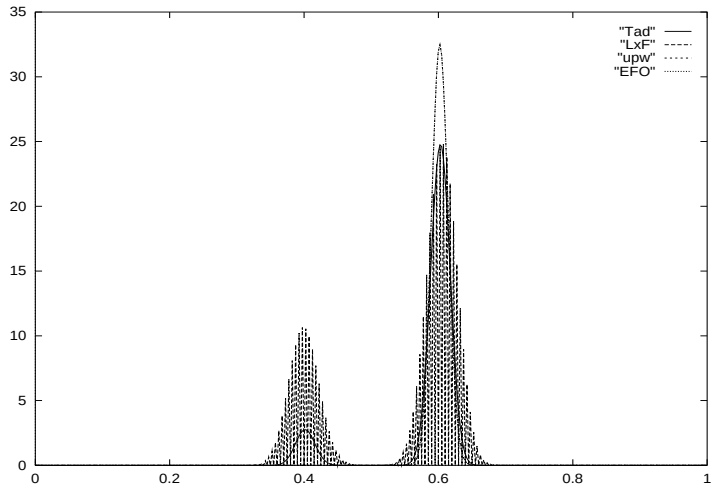

Figure 8: Numerical solutions with $a(t,x) = \mathrm{sgn}(x - \frac{1}{2})$ and $\Delta x = 0.0025$

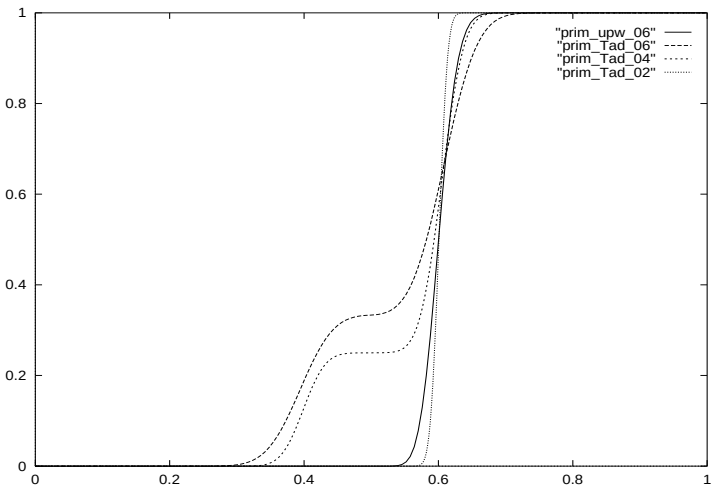

Figure 9: Numerical primitives with $a(t,x) = \mathrm{sgn}(x - \frac{1}{2})$ and $\Delta x = 0.002, 0.004, 0.006$

References

[1] F. Bouchut, *On zero pressure gas dynamics* Advances in Kinetic theory and computing, Selected papers, Series on advances in mathematics for applied sciences, **22**, pp171-190, World Scientific, 1994

[2] F. Bouchut and F. James, *Équations de transport unidimensionnelles à coefficients discontinus*, C.R. Acad. Sci. Paris, Série I, **320** (1995), 1097-1102.

[3] F. Bouchut and F. James, *One-dimensional transport equations with discontinuous coefficients*, Nonlinear Analysis, TMA, **32** (1998), n° 7, 891-933.

[4] F. Bouchut and F. James, *Solutions en dualité pour les gaz sans pression*, C.R. Acad. Sci. Paris, Série I, **326** (1998), 1073-1078.

[5] F. Bouchut and F. James, *Duality solutions for pressureless gases, monotone scalar conservation laws, and uniqueness*, Prépublication MAPMO, université d'Orléans, juin 1998; to appear in Comm. Partial Diff. Eq. (1998).

[6] F. Bouchut and F. James, *Differentiability with respect to initial data for a scalar conservation law*, to appear in the Proceedings of the 7-th Conference on Hyperbolic Problems, Zürich, 1998.

[7] Y. Brenier and S. Osher, *The discrete one-sided Lipschitz condition for convex scalar conservation laws*, SIAM J. Numer. Anal., **25** (1988), 8-23.

[8] E.D. Conway, *Generalized Solutions of Linear Differential Equations with Discontinuous Coefficients and the Uniqueness Question for Multidimensional Quasilinear Conservation Laws*, J. of Math. Anal. and Appl., **18** (1967), 238-251.

[9] G. Dal Maso, P. LeFloch and F. Murat, *Definition and Weak Stability of Nonconservative Products*, J. Math. Pures Appl., **74** (1995), 483-548

[10] W. E, Y.G. Rykov and Y.G. Sinai, *Generalized Variational Principles, Global Weak Solutions and Behavior with Random Initial Data for Systems of Conservation Laws Arising in Adhesion Particle Dynamics*, Comm. Math. Phys., **177** (1996), 349-380

[11] B. Engquist, E. Fatemi and S. Osher, *Numerical solution of the high frequency asymptotic expansion for the scalar wave equation*, J. Comp. Physics, **120** (1995), 145-155.

[12] B. Engquist, E. Fatemi and S. Osher, *Finite difference methods for geometrical optics and related nonlinear PDEs approximating the high frequency Helmoltz equation*, Department report UCLA, March 1995.

[13] B. Engquist and S. Osher, *Stable and entropy satisfying approximations for transonic flow calculations*, Math. Comp. **34** (1980), 45-75.

[14] B. Engquist and O. Runborg, *Multi-phase computations in geometrical optics*, Journal of Computational and Applied Mathematics, **74** (1996), 175-192.

[15] E. Godlewski, M. Olazabal and P.-A. Raviart, *On the linearization of hyperbolic systems of conservation laws. Application to stability*, Équations aux dérivées partielles et applications, articles dédiés à J.-L. Lions, Gauthier-Villars, Paris, 1998, 549-570.

[16] E. Grenier, *Existence globale pour le système des gaz sans pression* C.R. Acad. Sci. Paris, **321** (1995), 171-174.

[17] D. Hoff, *The Sharp Form of Oleinik's Entropy Condition in Several Space Variables*, Trans. of the A.M.S., **276** (1983), 707-714.

[18] F. James and M. Sepúlveda, *Convergence results for the flux identification in a scalar conservation law*, to appear in SIAM J. Cont. Opt. (1998)

[19] H.C. Kranzer and B.L. Keyfitz, *A Strictly Hyperbolic System of Conservation Laws Admitting Singular Shocks*, Nonlinear Evolution Equations that Change Type, IMA Volumes in Mathematics and its Applications, **27** (1990), Springer-Verlag.

[20] B. Larrouturou, *How to preserve the mass fractions positivity when computing multicomponent flows*, J. Comp. Phys, **95** (1991), 59-84.

[21] P. LeFloch, *An existence and uniqueness result for two nonstrictly hyperbolic systems*, Nonlinear evolution equations that change type, IMA volumes in mathematics and its applications, **27** (1990), Springer Verlag.

[22] P. LeFloch and Z. Xin, *Uniqueness via the Adjoint Problems for Systems of Conservation Laws*, Comm. on Pure and Applied Maths., XLVI (11), 1499-1533 (1993).

[23] P.-L. Lions, Generalized solutions of Hamilton-Jacobi equations, Research notes in mathematics, **69**, Pitman, 1982.

[24] M. Olazabal, *Résolution numérique du système des perturbations linéaires d'un écoulement MHD*, Thèse université Paris 6, 1998.

[25] O.A. Oleinik, *Discontinuous Solutions of Nonlinear Differential Equations*, Amer. Math. Soc. Transl. (2), **26** (1963), 95-172.

[26] F. Poupaud and M. Rascle, *Measure solutions to the linear multi-dimensional transport equation with non-smooth coefficients*, Comm. partial diff. eq. **22** (1997), 337-358.

[27] E. Tadmor, *Numerical viscosity and the entropy condition for conservative difference schemes*, Math. Comp. **43** (1984), 995-1011.

[28] E. Tadmor, *Local Error Estimates for Discontinuous Solutions of Nonlinear Hyperbolic Equations*, SIAM J. Numer. Anal. **28** (1991), 891-906.

[29] D. Tan, T. Zhang and Y. Zheng, *Delta shock-waves as limits of vanishing viscosity for hyperbolic systems of conservation laws*, J. Diff. Eq. **112** (1994), 1-32.

[30] A.I. Vol'pert, *The spaces BV and quasilinear equations*, Math. USSR Sb., **2** (1967), n° 2, 225-267.