

## EQUILIBRIUM SCHEMES FOR SCALAR CONSERVATION LAWS WITH STIFF SOURCES

RAMAZ BOTCHORISHVILI, BENOIT PERTHAME, AND ALEXIS VASSEUR

ABSTRACT. We consider a simple model case of stiff source terms in hyperbolic conservation laws, namely, the case of scalar conservation laws with a zeroth order source with low regularity. It is well known that a direct treatment of the source term by finite volume schemes gives unsatisfactory results for both the reduced CFL condition and refined meshes required because of the lack of accuracy on equilibrium states. The source term should be taken into account in the upwinding and discretized at the nodes of the grid. In order to solve numerically the problem, we introduce a so-called *equilibrium schemes* with the properties that (i) the maximum principle holds true; (ii) discrete entropy inequalities are satisfied; (iii) steady state solutions of the problem are maintained. One of the difficulties in studying the convergence is that there are no  $BV$  estimates for this problem. We therefore introduce a kinetic interpretation of upwinding taking into account the source terms. Based on the kinetic formulation we give a new convergence proof that only uses property (ii) in order to ensure desired compactness framework for a family of approximate solutions and that relies on minimal assumptions. The computational efficiency of our *equilibrium schemes* is demonstrated by numerical tests that show that, in comparison with an usual upwind scheme, the corresponding *equilibrium* version is far more accurate. Furthermore, numerical computations show that equilibrium schemes enable us to treat efficiently the sources with singularities and oscillating coefficients.

### 1. INTRODUCTION

Two kinds of singular sources arise classically in the theory of hyperbolic equations. A class of such problems are relaxation terms (see [4], [5], [19]) which force some algebraic relations between unknowns in such a way to ensure the system with a smaller number of unknowns in the relaxation limit. Another class consists in low regularity, possibly concentrations, in the source. The motivation of this paper comes from such a phenomenon in the Saint Venant model for shallow water where bottom (bathymetry) can be irregular in practical situations. Note that even in case of sufficiently smooth bottom these terms are very dominating and important in the process of creation of equilibriums, i.e., steady state solutions.

As a simplified model problem, we consider the scalar conservation law

$$(1.1) \quad \frac{\partial u}{\partial t} + \frac{\partial A(u)}{\partial x} + B(x, u) = 0, \quad t \geq 0, \quad x \in \mathbb{R},$$

---

Received by the editor March 29, 2000 and, in revised form, January 3, 2001.

2000 *Mathematics Subject Classification*. Primary 65M06, 65M12, 35L65.

*Key words and phrases*. Hyperbolic conservation laws, kinetic formulation, stiff source terms, upwind schemes, convergence.

$$(1.2) \quad u(x, t = 0) = u_0(x), \quad u_0(x) \in L^1 \cap L^\infty(\mathbb{R}),$$

with a smooth flux function  $A(\cdot)$ , where the unknown function  $u(t, x)$  belongs to  $\mathbb{R}$ . Also, the equation (1.1) is endowed with the full family of entropy inequalities

$$(1.3) \quad \frac{\partial S(u)}{\partial t} + \frac{\partial \eta(u)}{\partial x} + S'(u)B(x, u) \leq 0,$$

for all convex entropy functions  $S(\cdot)$  and corresponding entropy fluxes  $\eta(\cdot)$  defined in accordance with the relation

$$(1.4) \quad \eta'(u) = S'(u)A'(u)$$

(see Kruzkov [13], Lax [16] for more details).

We start our consideration with the special form of the source term

$$B(x, u) = z'(x)b(u), \quad b(u) \in C^1(\mathbb{R}).$$

We set and assume

$$a(u) = A'(u) \in C^1(\mathbb{R}), \quad D(u) = \int_0^u \frac{a(s)}{b(s)} ds,$$

$$(1.5) \quad 0 < \frac{a(u)}{b(u)} < \infty, \quad D(\pm\infty) = \pm\infty,$$

$$(1.6) \quad z'(x), z(x) \in L^\infty(\mathbb{R}).$$

The assumption (1.5) is restrictive because it implies that  $D(\cdot)$  is increasing, and therefore it implies that the equilibria are continuous. This assumption will be relaxed in Section 5 and in the Appendix. However it simplifies greatly the presentation and we first state our results in this framework.

A numerical difficulty which arises in such situations is to preserve, at a discrete level, the “equilibrium”, i.e., steady states given by

$$\frac{\partial A(u)}{\partial x} + z'(x)b(u) = 0,$$

which can be made explicit in the form of the algebraic relation

$$(1.7) \quad D(u) + z(x) = \text{const.}$$

Note that condition (1.5) ensures the existence of a unique Lipschitz continuous solution to the equation (1.7) because the function  $D$  is increasing. We now consider the discrete version of (1.7)

$$(1.8) \quad D(u_j) + z_j = \text{const.},$$

where  $z_j$  represents the average of  $z(x)$  over cell  $C_j$ ,  $C_j = ]x_{j-1/2}, x_{j+1/2}[$ , corresponding to a uniform grid with nodal points  $x_j$  and space discretization step  $\Delta x = x_{j+1/2} - x_{j-1/2}$ , and we denote by  $u_j^n$  the average at time  $t_n = n\Delta t$  of  $u(x, t_n)$  on the cell  $C_j$ , as usual in the so-called finite volume approach. We consider the *equilibrium schemes* as a solver with the properties:

$$(1.9) \quad \text{steady state initial data corresponding to (1.8) are maintained;}$$

$$(1.10) \quad \text{all the discrete entropy inequalities are valid;}$$

$$(1.11) \quad \text{approximate solutions are, locally in time, } L^\infty \text{ bounded.}$$

In this paper (see Sections 3 and 5 for details) we introduce a class of finite volume schemes which have properties (1.9)–(1.11). In the simplest case, e.g., under the condition (1.5), the scheme has the usual simple form of an upwind finite volume scheme

$$(1.12) \quad u_j^{n+1} - u_j^n + \frac{\Delta t}{\Delta x} \left( A(u_{j+1,-}^n, u_j^n) - A(u_j^n, u_{j-1,+}^n) \right) = 0.$$

Here  $A(.,.)$  is the usual Engquist-Osher numerical flux function (see [7])

$$(1.13) \quad A(u, v) = \int_0^u a_-(\xi) d\xi + \int_0^v a_+(\xi) d\xi,$$

with  $a_{\pm}(\xi)$  the positive and negative parts of  $a(\xi)$  defined by  $a(\xi) = a_+(\xi) + a_-(\xi)$ ,  $|a(\xi)| = a_+(\xi) - a_-(\xi)$ . And the so-called discrete equilibrium states  $u_{j\pm 1, \mp}^n$  are defined according to the relations (see (1.8) as well)

$$(1.14) \quad \begin{aligned} D(u_{j+1,-}) + z_j &= D(u_{j+1}) + z_{j+1}, \\ D(u_{j-1,+}) + z_j &= D(u_{j-1}) + z_{j-1}. \end{aligned}$$

Notice that for a general source term  $B(x, u)$  with a unique equilibrium we still can define the equilibrium scheme in the simple form (1.12). Namely we introduce the following initial value problem:

$$(1.15) \quad \frac{\partial A(v)}{\partial x} + B(x, v) = 0,$$

$$(1.16) \quad v(x_j) = u_j.$$

If (1.15), (1.16) is well posed on the intervals  $(x_j, x_{j+1}]$  and  $(x_j, x_{j-1}]$ , i.e., the unique solution to (1.15), (1.16) exists, then we can define discrete equilibrium states for the numerical equation (1.12) as

$$u_{j,+} = v(x_{j+1}), \quad u_{j,-} = v(x_{j-1}).$$

Clearly, when no simple explicit formulas are available, for numerical purposes one can apply suitable ODE solvers to (1.15), (1.16) for the definition of discrete equilibrium states.

The presented scheme is a particularly simple and efficient variant of the usual Engquist-Osher scheme which plays a particular role here because of its kinetic interpretation ([3], [12]) which allows a very simple interpretation of the relations (1.14). In the special subcase which is presented in the introduction of a single equilibrium, it is not completely original, and combined with a Riemann solver it can be interpreted as a well balanced scheme following denominations and ideas of Greenberg et al. [11]. Then a proof of convergence relying on much stronger assumptions can be found in Gosse [10]. More generally it falls in a class which has been advocated recently by several authors (see Greenberg et al., for source terms which only depend on  $x$ , Gosse and Leroux [9] for sources depending on  $u$  only, LeVeque [17], Vazquez-Cendon [26] on upwind discretization of source terms for the Saint-Venant system, Bermudez et al. [1] for upwind treatment of sources for 2D shallow water equations on unstructured meshes). The source  $z'(x)b(u)$  is discretized at the nodes of the grid, while the conservative quantities are cell centered. Notice that other strategies to preserve equilibriums also exist (see for instance Russo [22] for a central scheme). Notice also that, as well known and as we will see later, the natural discretization of the source, as  $z'_j b(u_j^n)$  is very inefficient, although it was proved that the splitting algorithm converges with a rate  $\Delta t$  (see

[15], [2]). Concerning the case of double equilibria, presented in Section 5, it seems that the question was not addressed before.

The main result of this paper is the following convergence theorem.

**Theorem 1.1** (Convergence of the scheme). *We assume (1.5), (1.6) and the CFL condition stated in (3.4), (3.8) below, and we define the approximate solution  $u_{\Delta x}(t, x) = u_j^n$  for  $t \in [t_n, t_{n+1})$  and  $x \in C_j$ . Then the scheme (1.12)–(1.14) satisfies the properties (1.9)–(1.11) as  $\Delta x \rightarrow 0$ ,  $u_{\Delta x}(t, x)$  converges in  $L^p([0, T] \times \mathbb{R})$ , for all  $1 \leq p < \infty$ , and all  $T > 0$ , towards the unique entropy solution to (1.1), (1.2).*

Especially in this theorem we avoid BV estimates, which are not available for the exact or approximate solutions due to the low regularity of the source term. Also they are known to be limited to one space dimension, and the scheme and theorem extend obviously to higher dimensions on rectangular grids. In order to avoid the need of BV bounds, we design a new method of investigation of numerical schemes. It is based on a new tool for the identification of solutions to kinetic equations corresponding to entropy solutions of the problem under consideration (see Section 2). It has the advantage that the same strategy extends to multidimensions without any difficulties. The analysis of the properties of the schemes is performed in Section 3 and the strong convergence towards entropy solution is proved in Section 4. A variant of the scheme for more general cases of functions  $D(u)$  is designed in Section 5. High computational efficiency of our schemes is demonstrated by numerical tests in the last section.

## 2. KINETIC FORMULATION. UNIQUENESS OF KINETIC SOLUTIONS

In this section we introduce a general tool for studying the convergence of numerical schemes for nonlinear scalar conservation laws. One of the most crucial steps in studying the convergence of numerical schemes for nonlinear problems is the derivation of a priori estimates ensuring compactness of the family of approximate solutions. For hyperbolic conservation laws the suitable compactness framework traditionally is ensured by BV estimates (see [14],[23]). When no BV bounds are available, the convergence study of the schemes has to be based on weaker compactness arguments. Such arguments were introduced by DiPerna [6] with the notion of measure valued solutions. At the numerical level Szepessy [24] has shown all the interest of the method which has been widely used especially in the two-dimensional case on unstructured grids.

A more powerful and easy to use approach is the so-called kinetic formulation of the problem (Lions, Perthame, Tadmor [18]). Especially, this approach allows a very simple uniqueness proof (Perthame [20]) which simplifies the convergence analysis of numerical schemes. The equation (1.1) and the family of entropy inequalities (1.3) can be written equivalently as a single kinetic equation with a “density” function  $\chi(\xi; u(t, x))$  (but we drop the dependency upon  $t$  and  $x$  below).

**Definition 2.1.** The function  $u(t, x)$  is called a kinetic solution if

$$(2.1) \quad \frac{\partial \chi(\xi; u)}{\partial t} + a(\xi) \cdot \frac{\partial \chi(\xi; u)}{\partial x} - b(\xi) z'(x) \frac{\partial \chi(\xi; u)}{\partial \xi} = \frac{\partial m(t, x, \xi)}{\partial \xi}$$

for some nonnegative bounded measure  $m(t, x, \xi)$  which satisfies

$$(2.2) \quad m(t, x, \xi) = 0 \quad \text{for} \quad |\xi| > \|u(t, \cdot)\|_{L^\infty},$$

and

$$(2.3) \quad \chi(\xi; u) = \begin{cases} +1, & 0 < \xi \leq u, \\ -1, & u \leq \xi < 0, \\ 0, & \text{otherwise.} \end{cases}$$

Evidently, equation (2.1) is linear and more attractive for investigation than the nonlinear equations (1.1) or (1.3). It is clear as well that to prove the convergence of schemes for linear equations like (2.1) there is no need in hard estimates like BV (see Section 4 for details). But nonlinearities are hidden in the structure of the function  $\chi(\xi; u)$  and in the measure valued source term  $m(t, x, \xi)$ . The relation between a function  $u(t, x)$  and the related equilibrium function  $\chi(\xi; u(t, x))$  is given by the formula

$$(2.4) \quad u(t, x) = \int_{\mathbb{R}} \chi(\xi; u(t, x)) d\xi.$$

With the same formula and (2.3), we can recover weak distributional solutions to (1.1) just by integrating (2.1) in  $\xi$ . Since  $m(t, x, \xi)$  is nonnegative in addition, the family of entropy inequalities (1.3) for all convex entropy functions can be recovered just multiplying (2.1) by  $S'(\xi)$  and integrating in  $\xi$ . On the other hand, the relation between atomic measure valued solutions by DiPerna [6], or Young measures,  $\nu_{t,x}(\xi)$ , a probability measure in  $\xi$ , and the kinetic equilibrium function can be made explicit as well (see also [21]):

$$(2.5) \quad \nu_{t,x}(\xi) = \delta(\xi) - \frac{\partial \chi(\xi; u)}{\partial \xi}.$$

So in frames of kinetic formulation one can recover all existing properties of entropy solutions to (1.1). Especially concerning weak limits of entropy solutions, we can naturally arrive at the notion of *kinetic solutions* to scalar conservation laws.

**Definition 2.2.** Let the function  $f(t, x, \xi)$  belong to  $L^\infty(0, T; L^\infty(\mathbb{R}_{x,\xi}^2) \cap L^1(\mathbb{R}_{x,\xi}^2))$  for all  $T \geq 0$ . It is called a generalized kinetic solution to equation (1.1), if

$$(2.6) \quad \frac{\partial f(t, x, \xi)}{\partial t} + a(\xi) \cdot \frac{\partial f(t, x, \xi)}{\partial x} - b(\xi) z'(x) \frac{\partial f(t, x, \xi)}{\partial \xi} = \frac{\partial m(t, x, \xi)}{\partial \xi},$$

in the sense of distribution for some nonnegative measure  $m(t, x, \xi)$  bounded on  $[0, T] \times \mathbb{R}_x \times \mathbb{R}_\xi$  for all  $T > 0$  which satisfies

$$(2.7) \quad 0 \leq \text{sign}(\xi) f(t, x, \xi) = |f(t, x, \xi)| \leq 1,$$

$$(2.8) \quad \frac{\partial f}{\partial \xi} = \delta(\xi) - \nu(t, x, \xi),$$

with  $\nu(t, x, \xi)$  some nonnegative measure such that  $\int_{\mathbb{R}} \nu(t, x, \xi) d\xi = 1$  for all  $t, x$ .

Note that it is relatively easy to arrive at *generalized kinetic solutions* departing from numerical schemes, e.g., see Section 4. But we are interested only in the entropy solutions of the problem (1.1), (1.2). For this purpose we prove that *generalized kinetic solutions* are unique and have the form  $f = \chi(\xi; u)$ . To do so, we adapt the approach by Perthame [20] for studying the uniqueness for scalar conservation laws. It is simpler than the usual uniqueness proof of Kruzkov [13] or the one by Diperna [6] used for the similar identification problem of entropy measure valued solutions, since it does not require doubling of variables. Furthermore, we arrive

at simple explicit requirements that can be naturally adapted for investigation of numerical schemes.

**Theorem 2.1** (Uniqueness of generalized kinetic solutions). *Let  $f(t, x, \xi)$  be a generalized kinetic solution to (1.1), (1.2), such that for a.e.  $t > 0$ ,*

$$(2.9) \quad \int_{\mathbb{R}^2} f(t, x, \xi) \varphi(x) S'(\xi) dx d\xi + \int_0^t \int_{\mathbb{R}^2} a(\xi) \varphi'(x) S'(\xi) f(\tau, x, \xi) dx d\xi d\tau - \int_0^t \int_{\mathbb{R}^2} (b(\xi) S'(\xi))_\xi \varphi'(x) f(\tau, x, \xi) dx d\xi d\tau \leq \int_{\mathbb{R}^2} \chi(\xi; u^0(x)) \varphi(x) S'(\xi) dx d\xi.$$

$$(2.10) \quad \int_R u(t, x) \varphi(x) dx \longrightarrow \int_R u^0(x) \varphi(x) dx \text{ as } t \rightarrow 0,$$

for any convex entropy functions  $S(\xi)$  and all nonnegative test functions  $\varphi \in D(R)$ . Then  $f(t, x, \xi) = \chi(\xi; u)$  where  $u(t, x)$  is the entropy solution of (1.1), (1.2) and  $f(t, x, \xi) \longrightarrow \chi(\xi; u_0)$  in  $L^1(\mathbb{R}^2)$  as  $t \rightarrow 0$ .

*Proof of Theorem 2.2.* The proof is based on a comparison of  $|f|$  and  $|f|^2$ . To justify the equations on these quantities, we first regularize by setting  $f_\varepsilon(t, x, \xi) = f * \omega_\varepsilon(t, x)$ ,  $m_\varepsilon(t, x, \xi) = m * \omega_\varepsilon(t, x)$ , where  $\omega_\varepsilon(t, x) = \frac{1}{\varepsilon_1} \omega_1(\frac{t}{\varepsilon_1}) \frac{1}{\varepsilon_2} \omega_2(\frac{x}{\varepsilon_2})$  with  $\varepsilon_1, \varepsilon_2 > 0$  and  $\omega_1, \omega_2$  two nonnegative regularizing kernels with  $\text{supp}(\omega_1(\cdot)) \subset [-\infty, 0]$  ensuring the possibility to regularize for any  $t > 0$ . We have

$$(2.11) \quad \frac{\partial f_\varepsilon}{\partial t} + a(\xi) \cdot \frac{\partial f_\varepsilon}{\partial x} - \xi z'(x) \frac{\partial f_\varepsilon}{\partial \xi} = \frac{\partial m_\varepsilon}{\partial \xi} + \Psi_\varepsilon,$$

$$\Psi_\varepsilon = \xi \frac{\partial}{\partial \xi} [(z'(x)f) * \omega_\varepsilon - z'(x)f * \omega_\varepsilon].$$

Then following [20], we divide the proof into several steps.

(i) We prove that

$$(2.12) \quad \frac{\partial}{\partial t} \int |f| d\xi + \frac{\partial}{\partial x} \int a(\xi) |f| d\xi + z'(x) \int |f| d\xi = -2m(t, x, 0).$$

This is simply obtained in multiplying (2.11) by  $\text{sign}(\xi)$  and letting  $\varepsilon_1, \varepsilon_2$  vanish, after noting that  $\text{sign}(\xi) f_\varepsilon = |f_\varepsilon|$  by Definition 2.2.

(ii) We prove that

$$(2.13) \quad \frac{\partial}{\partial t} \int \frac{f^2}{2} d\xi + \frac{\partial}{\partial x} \int a(\xi) \frac{f^2}{2} d\xi + z'(x) \int \frac{f^2}{2} d\xi \geq -2m(t, x, 0).$$

To do this, we multiply (2.11) by  $f_\varepsilon$  and notice that because  $f$  is bounded

$$\int \frac{f_\varepsilon^2}{2} d\xi \longrightarrow \int \frac{f^2}{2} d\xi, \quad \text{as } \varepsilon \rightarrow 0.$$

It is therefore enough to study the right-hand side appearing from (2.11). It contains two terms. The first one is treated using the property (2.8)

$$\begin{aligned} \int \frac{\partial}{\partial \xi} m_\varepsilon f_\varepsilon d\xi &= - \int m_\varepsilon(t, x, \xi) \frac{\partial}{\partial \xi} f_\varepsilon d\xi \\ &= -m_\varepsilon(t, x, 0) + \int m_\varepsilon \nu * \omega_\varepsilon d\xi \geq -m_\varepsilon(t, x, 0). \end{aligned}$$

The second term is estimated as

$$\begin{aligned} & \left| \int f_\varepsilon \Psi_\varepsilon d\xi \right| = \left| \int (f_\varepsilon + \xi \partial_\xi f_\varepsilon) ((z' f) * \omega_\varepsilon - z' f_\varepsilon) d\xi \right| \\ &= \left| \int (f_\varepsilon - \nu(t, x, \xi) * \omega_\varepsilon) \left( \int (z'(y) - z'(x)) f(t, y, \xi) \omega_{1\varepsilon} \omega_{2\varepsilon}(x - y) dy \right) d\xi \right| \\ &\leq \int |f_\varepsilon - \nu * \omega_\varepsilon| d\xi \int |z'(x) - z'(y)| \omega_{2\varepsilon}(x - y) dy. \end{aligned}$$

Evidently, this term vanishes a.e. as  $\varepsilon \rightarrow 0$  since  $z'$  belongs to  $L^\infty$  according to (1.6) and thus we arrive at (2.13).

(iii) Strong continuity at  $t = 0$ . From Brenier's lemma [3], we deduce that  $u(t, x) = \int_{\mathbb{R}} f(t, x, \xi) d\xi$  satisfies, for all convex functions  $S$  with  $S(0) = 0$ ,

$$S(u(t, x)) \leq \int_{\mathbb{R}} S'(\xi) f(t, x, \xi) d\xi.$$

As  $t \rightarrow 0$ , we deduce from (2.9) and the above inequality that

$$\text{w-lim } S(u(t, x)) \leq \int_{\mathbb{R}} S'(\xi) \chi(u^0(x)).$$

This inequality for the weak convergence of convex function implies that

$$u(t, x) \rightarrow u^0(x), \quad \text{strongly in all } L^p, \quad 1 \leq p < \infty,$$

and also

$$f(t, x, \xi) \rightarrow \chi(u^0(x)), \quad \text{strongly in all } L^p, \quad 1 \leq p < \infty.$$

(iv) Subtracting (2.12) and (2.13), and estimating the terms with  $z'$  taking account of (1.6) we have

$$(2.14) \quad \frac{\partial}{\partial t} \int (|f| - f^2) d\xi + \frac{\partial}{\partial x} \int a(\xi) (|f| - f^2) d\xi \leq \|z'\|_\infty \int (|f| - f^2) d\xi.$$

Next, we now use the strong continuity of  $f(t)$  at  $t = 0$  (point (iii) of the proof), and we deduce that

$$(2.15) \quad |f(t=0)| - f^2(t=0) = |\chi(\xi; u^0)| - |\chi(\xi; u^0)|^2 = 0.$$

Finally, since  $|f| - f^2 \geq 0$  by the assumption (2.7), we deduce from (2.14), (2.15) that

$$|f(t)| - f^2(t) = 0, \quad \text{for all } t \geq 0.$$

Therefore  $f$  takes the values  $-1, 0$  or  $+1$ . From this and the property  $\int \nu d\xi = 1$  in (2.8), we deduce that  $f = \chi(\xi; u)$  for some  $u(t, x)$ . Therefore  $u$  is the entropy solution (see [18]) and thus the proof is completed.  $\square$

*Remark 2.1.* The time continuity result at  $t = 0$  is the kinetic version of a similar statement discovered by Eymard et al. [8]. It is essential to get a simple proof of convergence of numerical schemes and avoid the possible initial layers that can occur for oscillating initial data. Also for nondegenerate fluxes, the time continuity is proved by A.Vasseur [25] with weaker assumptions. Namely one does not need the weak continuity at  $t = 0$ .

*Remark 2.2.* It is clear that the proof of the theorem is independent of the number of spatial variables and thus it holds true in multidimensions as well.

*Remark 2.3.* It is also easy to deduce uniqueness of entropy solutions of the problem (1.1), (1.2) using the same approach as above or the approach in [20].

*Remark 2.4.* It is also easy to see that under the assumptions of Theorem 2.2, the measure valued source term in kinetic equation (2.6) satisfies

$$\int_0^t \int_{R^2} m(\tau, x, \xi) d\tau dx d\xi \longrightarrow 0 \text{ as } t \rightarrow 0.$$

*Remark 2.5.* Kinetic solutions are more general compared to entropy solutions because they make sense even for  $u_0 \in L^1(R)$  which does not allow us to define  $A(u)$  when  $A$  is superlinear.

### 3. MAIN PROPERTIES OF THE SCHEME

In this section, we focus on the derivation of the quantitative properties (bounds and CFL condition, entropy inequality) and we give a first connection to the kinetic formulation. We also indicate a formal reason for consistency. For the sake of simplicity we now consider the simpler case  $b(\xi) = \xi$ .

First of all, we can prove the *well-balanced* property (1.9) of the scheme (1.12)–(1.14), i.e., that *it preserves the equilibrium*. Indeed, at the equilibrium state then

$$(3.1) \quad u_{j+1,-}^n = u_j^n, \quad u_{j-1,+}^n = u_j^n,$$

thus resulting in  $u_j^{n+1} = u_j^n$  from (1.12). If we are not at the equilibrium state, then the choice of  $u_{j,\pm}^n$  creates itself the states to which the averaged values  $u_{j+1}, u_{j-1}$  could be connected via equilibrium's relation from left and right respectively. That is why we call  $u_{j+1,-}, u_{j-1,+}$  the discrete equilibria associated with  $u_{j+1}$  on the left and the right.

We now explain the *consistency*. Compared to the classical formula (for the homogeneous problem) with fluxes  $A(u_{j+1}^n, u_j^n)$  at the interface  $x_{j+1/2}$ , we have in mind that the  $u_{j+1,-}^n$  is computed so that  $A(u_{j+1,-}^n, u_j^n) - A(u_j^n, u_{j+1}^n)$  discretizes the term  $z'(x)u$ , in a split way, at the interface  $x_{j+1/2}$  for waves incoming in the cell  $C_j$ . We can precise this point and prove the consistency of the discretization

$$(3.2) \quad \Delta x \widetilde{(z'(x)u)}_j^n = [A(u_{j+1,-}^n, u_j^n) - A(u_{j+1}^n, u_j^n)] + [A(u_j^n, u_{j-1}^n) - A(u_j^n, u_{j-1,+}^n)].$$

This can be seen with the help of the definition (1.14) and the Taylor expansions

$$u_{j+1,-} - u_{j+1} = \frac{z_{j+1} - z_j}{D'(\Theta_{j+1,-} u_{j+1,-} + (1 - \Theta_{j+1,-}) u_{j+1})}, \quad 0 \leq \Theta_{j+1,-} \leq 1,$$

$$u_{j-1} - u_{j-1,+} = \frac{z_j - z_{j-1}}{D'(\Theta_{j-1,+} u_{j-1,+} + (1 - \Theta_{j-1,+}) u_{j-1})}, \quad 0 \leq \Theta_{j-1,+} \leq 1,$$

which imply that

$$u_{j+1,-}^n \rightarrow u_{j+1}^n, \quad u_{j-1,+}^n \rightarrow u_{j-1}^n, \quad \text{as } \Delta x \rightarrow 0,$$

from which we deduce that for some intermediate state  $\xi$  we have

$$\begin{aligned} A(u_{j+1,-}^n, u_j^n) - A(u_{j+1}^n, u_j^n) &\sim a_-(\xi)(u_{j+1,-} - u_{j+1}) \\ &\sim \frac{\xi}{a(\xi)} a_-(\xi)(z_{j+1} - z_j), \end{aligned}$$

thus showing the consistency relation (3.2) for incoming waves in the first bracket.

With the above notation, (1.12) is therefore written in terms of the usual Engquist-Osher numerical flux function  $A_{j+1/2} = A(u_j^n, u_{j+1}^n)$ ,

$$u_j^{n+1} - u_j^n + \frac{\Delta t}{\Delta x}(A_{j+1/2}^n - A_{j-1/2}^n) + \frac{\Delta t}{\Delta x}(\widetilde{z'(x)u})_j^n = 0.$$

But this differs deeply from the standard discretization

$$(3.3) \quad u_j^{n+1} - u_j^n + \frac{\Delta t}{\Delta x}(A_{j+1/2}^n - A_{j-1/2}^n) + \frac{\Delta t}{\Delta x}(z'(x)u)_j^n = 0,$$

that does not preserve the equilibrium although it converges, but very slowly, to steady states (see Section 6).

Finally, it remains to establish the *stability properties* (1.10), (1.11). Evidently, under the usual CFL condition for homogeneous equation

$$(3.4) \quad C^n \Delta t \leq \Delta x,$$

$$C^n = \sup_j (|a(u_j^n)|, |a(u_{j+1,-}^n)|, |a(u_{j-1,+}^n)|),$$

the numerical scheme (1.12) gives the usual estimate

$$(3.5) \quad \min(u_{j-1,+}^n; u_j^n; u_{j+1,-}^n) \leq u_j^{n+1} \leq \max(u_{j-1,+}^n; u_j^n; u_{j+1,-}^n).$$

Notice however that in general the inequality  $|u_{j,\pm}^n| \leq |u_j^n|$  is not true and, naturally, we cannot have the same maximum principle as in the homogeneous case.

Because we use the Engquist-Osher numerical flux function, we can reformulate the scheme under a kinetic form that is more suitable for investigation. We set

$$(3.6) \quad u_j^{n+1} = \int f_j^{n+1} d\xi,$$

where the “density function”  $f_j^{n+1}$  is defined by the identity

$$(3.7) \quad f_j^{n+1}(\xi) - \chi_j^n(\xi) + \frac{\Delta t}{\Delta x} \left( a_-(\xi) \chi_{j+1,-}^n(\xi) + a_+(\xi) \chi_j^n(\xi) - a_-(\xi) \chi_j^n(\xi) - a_+(\xi) \chi_{j-1,+}^n(\xi) \right) = 0,$$

with the notation  $\chi_j^n(\xi) := \chi(\xi; u_j^n)$ ,  $\chi_{j-1,+}^n(\xi) := \chi(\xi; u_{j-1,+}^n)$ ,  $\chi_{j+1,-}^n(\xi) := \chi(\xi; u_{j+1,-}^n)$ . Indeed, integrating (3.7) in  $\xi$  and using (3.6) gives exactly (1.12), (1.13).

**Lemma 3.1.** *Under the CFL condition (3.4) with*

$$(3.8) \quad C^n = \max_{|u| \leq K_\infty} |a(u)|,$$

$$(3.9) \quad K_\infty = \exp(2T \|z'(x)\|_{L^\infty}) \|u^0\|_{L^\infty},$$

*the scheme (1.12)–(1.14) satisfies the maximum principle*

$$(3.10) \quad |u_j^n| \leq K_\infty, \quad \forall n \Delta t \leq T, \quad \forall j \in \mathbb{Z},$$

*and the in-cell entropy inequality for all convex entropy functions  $S$*

$$(3.11) \quad S(u_j^{n+1}) - S(u_j^n) + \frac{\Delta t}{\Delta x} \left( \eta(u_{j+1,-}^n, u_j^n) - \eta(u_j^n, u_{j-1,+}^n) \right) \leq 0,$$

which can be written more precisely as

$$(3.12) \quad \chi_j^{n+1}(\xi) - \chi_j^n(\xi) + \frac{\Delta t}{\Delta x} \left( a_-(\xi) \chi_{j+1,-}^n(\xi) + a_+(\xi) \chi_j^n(\xi) - a_-(\xi) \chi_j^n(\xi) - a_+(\xi) \chi_{j-1,+}^n(\xi) \right) = \frac{\partial m_j^{n+1}(\xi)}{\partial \xi},$$

for some bounded nonnegative measures  $m_j^{n+1}(\xi)$ .

*Proof of Lemma 3.1.* We first prove the  $L^\infty$  bound (3.10). One can estimate

$$(3.13) \quad \begin{aligned} \left| \int a_-(\xi) (\chi_{j+1,-}^n(\xi) - \chi_{j+1}^n(\xi)) d\xi \right| &= \left| \int \xi D'(\xi) (\chi_{j+1,-}^n(\xi) - \chi_{j+1}^n(\xi)) d\xi \right| \\ &\leq \max(|u_{j+1}^n|, |u_{j+1,-}^n|) |D(u_{j+1,-}^n) - D(u_{j+1}^n)| \\ &\leq \Delta x \|z'(x)\|_{L^\infty} \sup_j |u_j^n|. \end{aligned}$$

By analogy, the following estimate holds true:

$$(3.14) \quad \left| \frac{\Delta t}{\Delta x} \int a_+(\xi) (\chi_{j-1,+}^n(\xi) - \chi_{j-1}^n(\xi)) d\xi \right| \leq \Delta t \|z'(x)\|_{L^\infty} \sup_j |u_j^n|.$$

Next, we rewrite (3.7) as

$$(3.15) \quad \begin{aligned} f_j^{n+1}(\xi) &= \chi_j^n(\xi) \left( 1 - \frac{\Delta t}{\Delta x} |a(\xi)| \right) - \frac{\Delta t}{\Delta x} a_-(\xi) \chi_{j+1}^n(\xi) + \frac{\Delta t}{\Delta x} a_+(\xi) \chi_{j-1}^n(\xi) \\ &\quad - \frac{\Delta t}{\Delta x} a_-(\xi) (\chi_{j+1,-}^n(\xi) - \chi_{j+1}^n(\xi)) + \frac{\Delta t}{\Delta x} a_+(\xi) (\chi_{j-1,+}^n(\xi) - \chi_{j-1}^n(\xi)) \end{aligned}$$

and we use (3.13), (3.14). This yields

$$\sup_j |u_j^{n+1}| \leq (1 + 2\Delta t \|z'(x)\|_{L^\infty}) \sup_j |u_j^n|,$$

which results in (3.10).

Next, we turn to the entropy inequalities. Notice that by definition (see (2.3)),  $0 \leq \text{sign}(\xi) \chi(\xi; v) \leq 1$  for any  $v$ . Since  $f_j^{n+1}(\xi)$  in (3.7) is a convex combination of such  $\chi$ , we also have

$$(3.16) \quad 0 \leq \text{sign}(\xi) f_j^{n+1}(\xi) = |f_j^{n+1}| \leq 1.$$

Therefore, for some nonnegative compactly supported measure  $m_j^{n+1}(\xi)$ , we have

$$(3.17) \quad \chi_j^{n+1}(\xi) - f_j^{n+1}(\xi) = \frac{\partial m_j^{n+1}(\xi)}{\partial \xi}, \quad m_j^{n+1}(\xi) \geq 0,$$

as a consequence of (3.16) and Brenier's lemma [3]. Notice that the boundedness of the support of the measure  $m_j^{n+1}(\xi)$  can be obtained from (3.17) taking into account the uniform  $L^\infty$  estimates (3.10). Also its boundedness is a consequence of  $L^2$  bounds on  $u_j^n$  (see [18]) after multiplying by  $\xi$  and integrating (3.17) over  $R_\xi$ . Putting (3.17) in (3.7), multiplying by  $S'(\xi)$  and integrating in  $\xi$  result in the entropy property (3.12) that concludes the proof.  $\square$

*Remark 3.1.* The CFL condition we used is rather restrictive. A better choice of  $C^n$  in (3.4) is to use the maximum speed of propagation at time  $t_n$  rather than the uniform upper bound in the interval  $[0, T]$ . Notice that this approach results in more suitable timestepping for different time levels which is usually used in numerical computations.

*Remark 3.2.* One can derive better uniform  $L^\infty$  as well. Namely, we set

$$K_0 = \frac{\|z'(x)\|_{L^\infty}}{\min_u D'(u) \max_i |u_i^0|},$$

$$K'_\infty = \exp((1 + K_0 \Delta x) T \|z'(x)\|_{L^\infty}) \|u^0\|_{L^\infty}.$$

Then Lemma 3.1 can be proved exactly in the same way with  $K_\infty$  replaced by  $K'_\infty$ . Notice that, in the limit  $\Delta t \rightarrow 0$ , it results in the natural and more precise  $L^\infty$ -estimate  $\|u\|_{L^\infty} \leq \exp(T \|z'(x)\|_{L^\infty}) \|u^0\|_{L^\infty}$ .

#### 4. PROOF OF THE CONVERGENCE THEOREM

We now conclude the convergence proof. Based on the kinetic formulation, it is rather simple. Namely, we first adapt the uniqueness theorem of kinetic solutions and formulate it in another form better adapted to the numerical scheme and called the ‘Main Convergence Theorem’. For this purpose we formulate the requirements (consistency of the scheme,  $L^\infty$  and entropy stability, continuous dependency upon initial data) ensuring strong convergence of a family of approximate solutions. In a second step we show that the scheme satisfies the assumption of the Main Convergence Theorem.

**Theorem 4.1** (Main Convergence Theorem). *Let the family of approximate solutions  $u_{\Delta x}(t, x) \in L^\infty(0, T; L^1(\mathbb{R}))$  satisfy, for some constants  $K_m, K_1, K_\infty$ , some distribution  $\Psi_{\Delta x}(t, x, \xi)$ , some measure  $m_{\Delta x}(t, x, \xi)$  and some function  $\Psi_{0, \Delta x}(t)$ ,*

$$(4.1) \quad \frac{\partial \chi(\xi; u_{\Delta x})}{\partial t} + a(\xi) \frac{\partial \chi(\xi; u_{\Delta x})}{\partial x} - \xi z'(x) \frac{\partial \chi(\xi; u_{\Delta x})}{\partial \xi} = \frac{\partial m_{\Delta x}(t, x, \xi)}{\partial \xi} + \Psi_{\Delta x},$$

$$(4.2) \quad \Psi_{\Delta x}(t, x, \xi) \rightarrow 0 \text{ in } D' \text{ as } \Delta x \rightarrow 0;$$

$$(4.3) \quad m_{\Delta x}(t, x, \xi) \geq 0, \quad \|m_{\Delta x}(t, x, \xi)\|_{M^1} \leq K_m,$$

$$(4.4) \quad \|u_{\Delta x}\|_{L^1} \leq K_1, \quad \|u_{\Delta x}\|_{L^\infty} \leq K_\infty,$$

$$(4.5) \quad \int_{\mathbb{R}^2} \chi(\xi; u_{\Delta x}) \varphi(x) S'(\xi) dx d\xi + \int_0^t \int_{\mathbb{R}^2} a(\xi) \varphi'(x) S'(\xi) \chi(\xi; u_{\Delta x}) dx d\xi d\tau \\ \leq \int_{\mathbb{R}^2} \chi(\xi; u^0(x)) \varphi(x) S'(\xi) dx d\xi + \Psi_{0, \Delta x}(t).$$

$$(4.6) \quad \int u_{\Delta x} \varphi(x) dx d\xi = \int u^0(x) \varphi(x) dx d\xi + \Psi_{1, \Delta x}(t),$$

for all nonnegative test functions  $\varphi \in D(\mathbb{R})$  and smooth convex entropy functions  $S$ , where  $\Psi_{i, \Delta x}(t)$ , are bounded functions such that, for  $i = 0, 1$ ,

$$(4.7) \quad \Psi_{i, \Delta x}(t) \rightarrow \Psi_i(t) \text{ in } L^\infty - w^*, \quad \Psi_i(t) \text{ is continuous and } \Psi_i(0) = 0.$$

Then, as  $\Delta x \rightarrow 0$ ,  $u_{\Delta x}$  converges strongly in  $L^p([0, T] \times \mathbb{R})$ ,  $1 \leq p < \infty$ , to the unique entropy solution to (1.1), (1.2).

*Proof of Theorem 4.1.* The proof consists in proving that, passing to the limit in the above system, we obtain a function  $f(t, x, \xi)$  which satisfies the assumptions of the uniqueness theorem (Theorem 2.2) for kinetic solutions. It is decomposed into four steps: extracting a subsequence, continuity at  $t = 0$ , uniqueness for the limiting problem and conclusion of the proof.

(i) *Extracting a subsequence.* Due to the uniform estimates (4.3), (4.4) we can extract subsequences and we obtain

$$\chi(\xi; u_{\Delta x}) \rightarrow f(t, x, \xi) \text{ in } L^\infty\text{-w}^*,$$

$$m_{\Delta x}(t, x, \xi) \rightarrow m(t, x, \xi) \text{ in } M^1 \text{ weak.}$$

Hence we can pass to the limit as  $\Delta x \rightarrow 0$  in the linear equation (4.1) and we obtain the equation (2.6). Notice that this strategy for passing to the limit is also used by Perthame and Tzavaras [21], with applications to compensated compactness arguments.

(ii) *Continuity at  $t = 0$ .* Passing to the limit in (4.5)–(4.7) as  $\Delta x \rightarrow 0$  results in the conditions (2.9)–(2.10) for initial time continuity of Theorem 2.2.

(iii). *Uniqueness, identification of limiting function.* By definition of the  $\chi$  function one has

$$(4.8) \quad 0 \leq \text{sign}(\xi)\chi(\xi; u_{\Delta x}) = |\chi(\xi; u_{\Delta x})| \leq 1;$$

$$(4.9) \quad \frac{\partial \chi(\xi; u_{\Delta x})}{\partial \xi} = \delta(\xi) - \nu_{\Delta x},$$

and thus  $\nu_{\Delta x} = \delta(\xi - u_{\Delta x})$  is a nonnegative measure satisfying  $\int \nu_{\Delta x} d\xi = 1$ . Passing to the limit in these relations, we arrive at the requirements (2.7), (2.8) of Definition 2.1 for the *kinetic solution*  $f$ . It is easy to see that all requirements of the uniqueness Theorem 2.2 of kinetic solutions are satisfied, and thus we have

$$\chi(\xi; u_{\Delta x}) \rightarrow \chi(\xi; u) \text{ in } L^\infty\text{-w}^* \text{ as } \Delta x \rightarrow 0,$$

where  $u$  is the unique entropy solution to (1.1), (1.2).

(iv) *Conclusion.* From the nonlinearity of  $\chi$ , and more precisely from the identity  $\int S'(\xi)\chi(\xi; u)d\xi = S(u) - S(0)$ , we deduce that

$$\chi(\xi; u_{\Delta x}) \rightarrow \chi(\xi; u) \text{ a.e.,}$$

$$u_{\Delta x} \rightarrow u \text{ a.e.,}$$

and this concludes the proof of the main convergence theorem.  $\square$

*Remark 4.1.* One can relax the uniform  $L^1$  boundedness of approximate solutions in (4.4), by means of consequent constriction of approximate solutions, on domains of finite measure and then artificial continuation by zero outside. In this case the conclusion of the theorem is strong convergence of  $u_{\Delta x}$  in  $L^p_{loc}([0, T] \times \mathbb{R})$ ,  $1 \leq p < \infty$ , to the unique entropy solution to (1.1), (1.2) as  $\Delta x \rightarrow 0$ .

*Remark 4.2.* One of the essential features of the main convergence theorem is that it does not require bounded variations on approximate solutions. Also, it does not depend on the space dimension, and the method by means of which the family of approximate solutions is constructed. Note that the uniform BV-estimate is known to be an obstacle in proving convergence in the case of insufficiently smooth source terms, as here, but also in multidimensions on irregular meshes. Notice also that all the requirements of the main convergence theorem are necessary.

We are now ready to prove the convergence theorem stated in the introduction.

*Proof of the Convergence Theorem.* The proof is based on the verification of the requirements of the main convergence theorem departing from the formulation (3.12) of the scheme. It is decomposed into four steps.

(i) *Derivation of (4.1), (4.2).* We set  $\chi_{\Delta x} := \chi(\xi; u_{\Delta x})$ ,  $\varphi_j^n(\xi) = \varphi(t_n, x_j, \xi)$ , with  $\varphi(t, x, \xi)$  a test function. Notice that

$$(4.10) \quad \begin{aligned} \sum_n (\chi_j^{n+1} - \chi_j^n) \varphi_j^n &= - \sum_n (\varphi_j^{n+1} - \varphi_j^n) \chi_j^{n+1} \\ &= \int \chi_{\Delta x} \varphi_t dt + \Psi_1(\varphi, \Delta x, u_{\Delta x}), \end{aligned}$$

where  $\Psi_1(\varphi, \Delta x, u_{\Delta x}) \rightarrow 0$  as  $\Delta x \rightarrow 0$ ;

$$(4.11) \quad \begin{aligned} \sum_j a_-(\xi) (\chi_{j+1}^n - \chi_j^n) \varphi_j^n &= - \sum_j a_-(\xi) (\varphi_j^n - \varphi_{j-1}^n) \chi_j^n \\ &= \int a_-(\xi) \chi_{\Delta x} \varphi_x dx + \Psi_2(\varphi, \Delta x, u_{\Delta x}), \end{aligned}$$

$$(4.12) \quad \begin{aligned} \sum_j a_+(\xi) (\chi_j^n - \chi_{j-1}^n) \varphi_j^n &= - \sum_j a_+(\xi) (\varphi_{j+1}^n - \varphi_j^n) \chi_j^n \\ &= \int a_+(\xi) \chi_{\Delta x} \varphi_x dx + \Psi_3(\varphi, \Delta x, u_{\Delta x}), \end{aligned}$$

where  $\Psi_2(\varphi, \Delta x, u_{\Delta x}) \rightarrow 0$ ,  $\Psi_3(\varphi, \Delta x, u_{\Delta x}) \rightarrow 0$  as  $\Delta x \rightarrow 0$ ; taking into account the left-hand side of (3.13) one has

$$(4.13) \quad \begin{aligned} \int a_-(\xi) \frac{\chi_{j+1}^n - \chi_j^n}{\Delta x} \varphi_j^n d\xi &= - \frac{z_{j+1} - z_j}{\Delta x} \int \text{sgn}(a_-(\xi)) \xi \varphi_j^n \frac{\partial \chi(\xi, u_{\Delta x, -})}{\partial \xi} d\xi \\ &= z'(x) \int \varphi \xi \text{sgn}(a_-(\xi)) \frac{\partial \chi_{\Delta x}}{\partial \xi} d\xi + \Psi_4(\varphi, \Delta x, u_{\Delta x}, z'), \end{aligned}$$

where

$$u_{\Delta x, -} := \Theta_{j+1, -}^n u_{j+1, -}^n + (1 - \Theta_{j+1, -}^n) u_{j+1}^n, \quad (t, x) \in [t_n, t_{n+1}) \times C_j, \\ \Psi_4(\varphi, \Delta x, u_{\Delta x}, z') \rightarrow 0$$

as  $\Delta x$  vanishes, since  $u_{j+1, -} \rightarrow u_{j+1}$ . Because of the same reasons

$$(4.14) \quad \int a_+(\xi) \frac{\chi_{j-1}^n - \chi_j^n}{\Delta x} \varphi_j^n d\xi = z'(x) \int \varphi \xi \text{sgn}(a_+(\xi)) \frac{\partial \chi_{\Delta x}}{\partial \xi} d\xi + \Psi_5(\cdot, \cdot, \cdot),$$

where  $\Psi_5(\varphi, \Delta x, u_{\Delta x}, z') \rightarrow 0$  as  $\Delta x \rightarrow 0$ . Multiplying (3.15) by  $\varphi_j^n$  with account of (3.17), (4.9)–(4.13) (or what is the same, summing (4.9), (4.13)) one arrives at the suitable distributional form of the equation that results in the validity of (4.1), (4.2).

(ii) *Derivation of (4.4).* Uniform  $L^\infty$  estimates are already obtained in Lemma 3.1. The  $L^1$ -estimate is derived from (3.16) after multiplying (3.15) by  $\text{sign}(\xi) \Delta x$ , integrating in  $\xi$  and summing in  $j$ . Namely one arrives at the estimate

$$\sum_j \Delta x |u_j^{n+1}| \leq (1 + 2\Delta t \|z'(x)\|_{L^\infty}) \sum_j \Delta x |u_j^n|,$$

which results in  $K_1 = \exp(1 + 2t \|z'(x)\|_{L^\infty}) \|u^0\|_{L^1}$  in (4.4).

(iii) *Derivation of (4.3).* The sign of  $m_{\Delta x}$  is already provided by Lemma 3.1. To prove the bound on  $m$  in (4.3) one can use the uniform  $L^1$  and  $L^\infty$  boundedness of the approximate solutions  $u_{\Delta x}$ . We multiply (3.15) by  $\xi \Delta x$ . Then after integration over  $\mathbb{R}_\xi$  and summing in  $j$  one arrives simply at the following rough (but sufficient for our purposes) estimate  $K_m = K_1 K_\infty$ .

(iv) *Derivation of (4.5)–(4.7).* After multiplying of (3.15) by  $\varphi_j \Delta x$ ,  $\varphi := \varphi(x_j, \xi)$ ,  $\varphi \in D(\mathbb{R}^2)$ , then integrating in  $\xi$  over  $R_\xi$  and summing in  $j$  and summing in  $n$  until any  $k, k\Delta t \leq T$  we obtain the following expression:

$$\Delta x \sum_j \int_{R_\xi} \chi^{k+1} \varphi_j S'(\xi) d\xi = \Delta x \sum_j \int_{R_\xi} \chi^0 \varphi_j S'(\xi) d\xi + \psi_{0\Delta x}(t_{k+1}, \varphi, S),$$

where

$$\begin{aligned} \psi_{0\Delta x}(t_{k+1}, \varphi, S) &= -\Delta t \sum_{i=0}^k \sum_j \int \varphi_j S'(\xi) [a_-(\xi)(\chi_{j+1}^i S'(\xi) - \chi_j^i(\xi)) \\ &\quad + a_+(\xi)(\chi_j^i(\xi) - \chi_{j-1}^i(\xi)) + a_-(\xi)(\chi_{j+1,-}^i(\xi) - \chi_{j+1}^i(\xi)) \\ &\quad + a_+(\xi)(\chi_{j-1}^i(\xi) - \chi_{j-1,+}^i(\xi)) - \Delta x \frac{\partial m_j^{i+1}(\xi)}{\partial \xi}] d\xi \\ &= \Psi_{0\Delta x}(t_{k+1}, \varphi) + \Delta t \Delta x \sum_{i=0}^k \sum_j \int \varphi_j \frac{dS'(\xi)}{d\xi} m_j^{i+1}(\xi) d\xi \leq \Psi_{0\Delta x}(t_{k+1}, \varphi), \end{aligned}$$

since  $S''(\xi) \geq 0$ ,  $\varphi_j \geq 0$ ,  $m_j^i(\xi) \geq 0$ , and

$$\begin{aligned} \Psi_{0\Delta x}(t_{k+1}, \varphi) &= \\ &\Delta t \sum_{i=0}^k \sum_j \int [a_-(\xi)(\varphi_{j+1} - \varphi_j) + a_+(\xi)(\varphi_j - \varphi_{j-1})] S'(\xi) \chi_j^i(\xi) d\xi \\ &\quad - \Delta t \sum_{i=0}^k \sum_j \int \varphi_j S'(\xi) [a_-(\xi)(\chi_{j+1,-}^i(\xi) - \chi_{j+1}^i(\xi)) \\ &\quad \quad \quad + a_+(\xi)(\chi_{j-1}^i(\xi) - \chi_{j-1,+}^i(\xi))] d\xi. \end{aligned}$$

Then for this expression of  $\Psi_{0\Delta x}$  one can estimate the right-hand side using (3.15), uniform boundedness of approximate solutions  $u_{\Delta x}$  and the measures  $m_{\Delta x}$  that yields for  $t_k \leq t \leq t_{k+1}$

$$\begin{aligned} |\Psi_{0\Delta x}| &\leq t_{k+1} \max_{|u| \leq K_\infty} |a(u)| \sum_j \int |\varphi_{j+1} - \varphi_j| d\xi \\ &\quad + t_{k+1} 2K_\infty \|z'(x)\|_{L^\infty} \sum_j \max_\xi |\varphi_j(\xi)| + t_{k+1} K_m \sum_j \int \left| \frac{\Delta x d\varphi_j(\xi)}{d\xi} \right| d\xi. \end{aligned}$$

Then for this expression of  $\Psi_{0\Delta x}$  one can estimate the right-hand side using (3.15), uniform boundedness of approximate solutions  $u_{\Delta x}$  and the measures  $m_{\Delta x}$  that yields for  $t_k \leq t \leq t_{k+1}$

$$\begin{aligned} |\Psi_{0\Delta x}| &\leq t_{k+1} \max_{|u| \leq K_\infty} |a(u)| \sum_j \int |\varphi_{j+1} - \varphi_j| d\xi \\ &\quad + t_{k+1} 2K_\infty \|z'(x)\|_{L^\infty} \sum_j \max_\xi |\varphi_j(\xi)| + t_{k+1} K_m \sum_j \int \left| \frac{\Delta x d\varphi_j(\xi)}{d\xi} \right| d\xi. \end{aligned}$$

Clearly  $\Psi_{0\Delta x}(t_{k+1}, \varphi)$  vanishes together with  $t_{k+1}$  for any  $\varphi \in D(\mathbb{R}^2)$  that results in the validity of (4.5), (4.7) with taking into account the formulas given above. Evidently, with  $S'(\xi) = 1$  and by use of the same technique as above one can easily recover the weak continuity requirements (4.6), (4.7) of approximate solutions  $u_{\Delta x}$  at  $t = 0$ .  $\square$

Finally, applying the main convergence theorem results in the strong convergence of the equilibrium scheme that concludes the proof.

## 5. EXTENSION TO MORE GENERAL FLUXES $A$

In this section, we extend the scheme to a more realistic situation where the function  $D$  is not monotone, motivated by the Saint-Venant system (SVS in short). This creates an additional difficulty because some discontinuous equilibria are unstable (nonentropic), and they should not be preserved by the scheme. Again we restrict our attention to the case  $b(\xi) = \xi$ .

We assume that the initial data  $u^0$  is nonnegative (and thus  $u(t, x) \geq 0$  also, because it plays the role of the height of water in the SVS), that  $a(u_0) = 0$  for some  $u_0 > 0$ . As before we define

$$(5.1) \quad A(u) = \int_0^u a(\xi) d\xi, \quad D(u) = \int_{u_0}^u \frac{a(\xi)}{\xi} d\xi,$$

and, our assumptions are summarized by

$$(5.2) \quad \begin{aligned} & a(\xi) \text{ is increasing on } ]0, +\infty[ \text{ and } u_0 \geq 0, \\ & D \text{ is nonnegative, decreasing on } ]0, u_0[, \text{ increasing on } ]u_0, +\infty[, \\ & D(u) \rightarrow +\infty, \quad \text{as } u \rightarrow 0 \text{ or } +\infty. \end{aligned}$$

Finally, for the sake of simplicity we only consider an increasing bottom  $z$  in this section. The general formulas are given in the Appendix. Setting  $\Delta z_{j+1/2} = z_{j+1} - z_j \geq 0$ , the scheme has this form:

$$(5.3) \quad u_j^{n+1} - u_j^n + \frac{\Delta t}{\Delta x} [ A(u_{j+1,-}^n, u_j^n) - A(u_j^n, u_{j-1,+}^n)$$

$$(5.4) \quad + A(u_0) - A(\tilde{u}_{j,-}^n) ] = 0,$$

where  $A(\cdot, \cdot)$  still denotes the Engquist-Osher scheme defined by (1.13).

We set

$$(5.5) \quad \underline{u}_j^n = \inf(u_0, u_j^n),$$

$$(5.6) \quad \bar{u}_j^n = \sup(u_0, u_j^n);$$

then  $u_{j-1,+}^n, u_{j+1,-}^n, \tilde{u}_{j,-}^n, \tilde{u}_{j,+}^n$  are defined by

$$(5.7) \quad \begin{cases} D(u_{j-1,+}^n) = \sup(0, D(\bar{u}_{j-1}^n) - \Delta z_{j-1/2}), \\ u_{j-1,+}^n \geq u_0, \end{cases}$$

$$(5.8) \quad \begin{cases} D(u_{j+1,-}^n) = D(\underline{u}_{j+1}^n) + \Delta z_{j+1/2}, \\ u_{j+1,-}^n \leq u_0, \end{cases}$$

$$(5.9) \quad \begin{cases} D(\tilde{u}_{j,-}^n) = \inf(\Delta z_{j+1/2}, D(\bar{u}_j^n)), \\ \tilde{u}_{j,-}^n \leq u_0. \end{cases}$$

First, let us explain the derivation of the flux functions described above. Equations (1.14) cannot be solved since they have either zero or two solutions in general (as

in the corresponding equilibrium equations for SVS). However, we can define a natural scheme at the kinetic level. It is based on the kinetic formulation of the scalar conservation law in (2.1). For a regular solution  $m = 0$  and the equation (2.1) can be solved with the help of the method of characteristics. Then, depending on the bottom's jump particles lose or gain energy when crossing the cell interfaces. More precisely, the characteristics are given by

$$(5.10) \quad \frac{dx}{dt} = a(\xi), \quad \frac{d\xi}{dt} = -z'(x)\xi.$$

Notice that this system admits the energy  $D(\xi) + z(x)$ ; in other words,

$$(5.11) \quad \frac{d}{dt} \left( D(\xi(t)) + z(x(t)) \right) = 0,$$

and therefore the function  $\chi(\xi; u)$  is constant along the trajectories  $D(\xi(t)) + z(x(t)) = \text{Cst}$ . This method allows us to construct  $u_{j,-}^n$  and  $u_{j,+}^n$  by an ‘‘energy barrier’’ (the particles pass through the interface with a loss or gain of velocity), and provides the additional term  $\tilde{u}_{j,-}^n$  from a reflection effect (some particles do not pass through the interface and are reflected into the cell).

Notice also that, with the kinetic functions  $f_{j,+}^n(\xi) = \chi(\xi, u_{j,+}^n)$ ,  $f_{j,-}^n(\xi) = \chi(\xi, u_{j,-}^n)$ , and  $\tilde{f}_{j,-}^n(\xi) = \mathbf{1}_{\{\tilde{u}_{j,-}^n \leq \xi \leq u_0\}}$ , the scheme corresponds at the kinetic level to the discrete equation

$$(5.12) \quad f_j^{n+1}(\xi) - \chi_j^n(\xi) + \frac{\Delta t}{\Delta x} \left( a(\xi)_- [f_{j+1,-}^n(\xi) + \tilde{f}_{j,-}^n(\xi)] + a(\xi)_+ \chi_j^n(\xi) - a(\xi)_- \chi_j^n(\xi) - a(\xi)_+ f_{j-1,+}^n(\xi) \right) = 0,$$

which replaces equation (3.7) in this nonmonotonic case.

We show in this section the following convergence theorem.

**Theorem 5.1** (Convergence of the general scheme). *With the notation and assumptions of Theorem 1.1 except that (1.5) is replaced by (5.2), the above scheme satisfies the properties (1.10), (1.11), maintains the equilibrium initial data*

$$\begin{aligned} D(u_j) + z_j &= \text{const} \\ \text{sign}(u_j - u_0) &= \text{const}. \end{aligned}$$

As  $\Delta x \rightarrow 0$ ,  $u_{\Delta x}(t, x)$  converges in  $L^p([0, T] \times \mathbb{R})$ , for all  $1 \leq p < \infty$ , and all  $T > 0$ , towards the unique entropy solution to (1.1), (1.2).

*Proof of Theorem 5.1.* The statement concerning equilibrium preservation is readily proved by a direct computation of fluxes and we skip the proof. We only point out the difference with the proof for the previous simpler case. First notice that  $D$  is decreasing for  $\xi \leq u_0$  so that  $\tilde{u}_{j,-}^n \geq u_{j+1,-}^n$  and the support of  $f_{j+1,-}^n$  and  $\tilde{f}_{j,-}^n$  are disconnected. Therefore  $f_{j-1,+}^n$ ,  $f_{j+1,-}^n + \tilde{f}_{j,-}^n$  are valued in  $[0, 1]$  and Lemma 3.1 is still valid (with the obvious modifications of the flux functions in the entropy inequalities). In order to show Theorem 4.1 using Theorem 2.2 we only have to show that the expression

$$(5.13) \quad T_j^n(\xi) = \frac{1}{\Delta x} (a(\xi)_- (f_{j+1,-}^n - \chi_{j+1}^n) + a(\xi)_+ (\chi_{j-1}^n - f_{j-1,+}^n) + a(\xi)_- \tilde{f}_{j,-}^n)$$

converges in the sense of distribution to  $z'\xi\frac{\partial f}{\partial \xi}$  which corresponds to (4.13), (4.14). As before, we denote  $\chi_{\Delta x}$  the function defined by

$$(5.14) \quad \chi_{\Delta x}(t, x, \xi) = \chi_j^n(\xi)$$

for  $(t, x) \in ]n\Delta t, (n+1)\Delta t[ \times ]j\Delta x, (j+1)\Delta x[$ . We introduce a test function  $\phi(t, x, \xi)$  and denote

$$(5.15) \quad \phi_j^n(\xi) = \frac{1}{\Delta t \Delta x} \int_{j\Delta x}^{(j+1)\Delta x} \int_{n\Delta t}^{(n+1)\Delta t} \phi(t, x, \xi) dt dx.$$

So when we multiply (5.13) by  $\phi$  and integrate and sum up, we find

$$(5.16) \quad \begin{aligned} & \Delta t \Delta x \sum_{j,n} \int T_j^n(\xi) \phi_j^n(\xi) d\xi \\ &= \Delta t \Delta x \sum_{j,n} \frac{1}{\Delta x} \left[ \int a(\xi)_- [f_{j,-}^n(\xi) - \chi_j^n(\xi)] \phi_{j-1}^n(\xi) d\xi \right. \\ & \quad + \int a(\xi)_+ [\chi_j^n(\xi) - f_{j,+}^n(\xi)] \phi_{j+1}^n(\xi) d\xi \\ & \quad \left. + \int a(\xi)_- \tilde{f}_{j,-}^n(\xi) \phi_j^n(\xi) d\xi \right] \\ &= \Delta t \Delta x \sum_{j,n} \int R_j^n(\xi) \phi_j^n(\xi) d\xi + \psi_0, \end{aligned}$$

where  $|\psi_0| \leq \Delta x C(\partial_x \phi) (\sup_{j,n} (\int |T_j^n| d\xi))$ , and

$$(5.17) \quad \int R_j^n(\xi) \phi_j^n(\xi) d\xi = \frac{1}{\Delta x} \int_0^{u_0} a(\xi) [f_{j,-}^n(\xi) - \chi_j^n(\xi)] \phi_j^n(\xi) d\xi$$

$$(5.18) \quad + \frac{1}{\Delta x} \int_{u_0}^{+\infty} a(\xi) [\chi_j^n(\xi) - f_{j,+}^n(\xi)] \phi_j^n(\xi) d\xi$$

$$(5.19) \quad + \frac{1}{\Delta x} \int_0^{u_0} a(\xi) \tilde{f}_{j,-}^n(\xi) \phi_j^n(\xi) d\xi.$$

To analyze these terms, we denote by  $\mathcal{D}_1 : \mathbb{R}^+ \rightarrow ]0, u_0[$  (respectively  $\mathcal{D}_2 : \mathbb{R}^+ \rightarrow ]u_0, +\infty[$ ) the inverse of  $D$  on  $]0, u_0[$  (respectively on  $]u_0, +\infty[$ ). The right-hand side term of (5.17) can be written as

$$\begin{aligned} & \frac{1}{\Delta x} \int_0^{u_0} a(\xi) [f_{j,-}^n(\xi) - \chi_j^n(\xi)] \phi_j^n(\xi) d\xi \\ &= \frac{1}{\Delta x} \int_{\underline{u}_j^n}^{u_{j,-}^n} D'(\xi) \xi \phi_j^n(\xi) d\xi \\ &= \frac{1}{\Delta x} \int_{D(\underline{u}_j^n)}^{D(u_{j,-}^n)} \mathcal{D}_1(\zeta) \phi_j^n(\mathcal{D}_1(\zeta)) d\zeta \\ &= \frac{1}{\Delta x} \int_{D(\underline{u}_j^n)}^{D(\underline{u}_j^n) + \Delta z_{j-1/2}} \mathcal{D}_1(\zeta) \phi_j^n(\mathcal{D}_1(\zeta)) d\zeta \\ &= \frac{\Delta z_{j-1/2}}{\Delta x} \underline{u}_{j,-}^{n*} \phi_j^n(\underline{u}_{j,-}^{n*}) \end{aligned}$$

where  $D(\underline{u}_{j,-}^{n*})$  is given by the mean value theorem. In particular,

$$(5.20) \quad \underline{u}_{j,-}^n \leq \underline{u}_{j,-}^{n*} \leq \underline{u}_j^n.$$

But since  $D$  is equicontinuous on  $]0, \sup u[$ , from (5.20) we deduce

$$(5.21) \quad |\underline{u}_{j,-}^{n*} - \underline{u}_j^n| \leq \varepsilon(\Delta z_{j-1/2})$$

where  $\varepsilon$  is continuous, independent on  $n, j, \Delta x$  and  $\varepsilon(0) = 0$ . So the right-hand side of (5.17) is equal to

$$(5.22) \quad \frac{\Delta z_{j+1/2}}{\Delta x} \underline{u}_j^n \phi_j^n(\underline{u}_j^n) + \psi_{1,j}^n + \psi_{2,j}^n$$

with

$$\begin{aligned} |\psi_{1,j}^n| &\leq \|z'\|_{L^\infty} C(\phi) \varepsilon(\|z'\|_{L^\infty} \Delta x), \\ \psi_{2,j}^n &\leq C(\phi) |z'_{\Delta x}(j\Delta x) - z'_{\Delta x}(j\Delta x + \Delta x)|, \end{aligned}$$

where we denote by  $z_{\Delta x}$  the function which is equal to  $z_j + (x - j\Delta x)(z_{j+1} - z_j)/\Delta x$  on  $]j\Delta x, (j+1)\Delta x[$ .

We consider now two cases:  $D(\bar{u}_j^n) \geq \Delta z_{j+1/2}$  and  $D(\bar{u}_j^n) \leq \Delta z_{j+1/2}$ .

(i) *The case  $D(\bar{u}_j^n) \leq \Delta z_{j+1/2}$ .* We have  $|D(\underline{u}_j^n) - D(\bar{u}_j^n)| \leq \Delta z_{j+1/2}$  so from (5.22), the right-hand side term of (5.17) is equal to

$$(5.23) \quad \frac{\Delta z_{j+1/2}}{\Delta x} u_j^n \phi_j^n(u_j^n) + \psi_{1,j}^n + \psi_{2,j}^n + \psi_{3,j}^n$$

with  $|\psi_{3,j}^n| \leq \|z'\|_{L^\infty} C(\phi) \varepsilon(\|z'\|_{L^\infty} \Delta x)$ . Thanks to (5.7),  $u_{j,+}^n = u_0$ , so the first term of (5.18) is vanishing. Thanks to (5.9),  $D(\tilde{u}_{j,-}^n) = D(\bar{u}_j^n)$ , so the terms (5.18) and (5.19) together give

$$\begin{aligned} \psi_{4,j}^n &= \frac{1}{\Delta x} \int_{u_0}^{\bar{u}_j^n} D'(\xi) \xi \phi_j^n(\xi) d\xi + \frac{1}{\Delta x} \int_{\bar{u}_{j,-}^n}^{u_0} D'(\xi) \xi \phi_j^n(\xi) d\xi \\ &= \frac{1}{\Delta x} \int_0^{D(\bar{u}_j^n)} [\mathcal{D}_2(\zeta) \phi_j^n(\mathcal{D}_2(\zeta)) - \mathcal{D}_1(\zeta) \phi_j^n(\mathcal{D}_1(\zeta))] d\zeta \\ &\leq \frac{\Delta z_{j+1/2}}{\Delta x} \sup_{0 \leq \zeta \leq \Delta z_{j+1/2}} |\mathcal{D}_2(\zeta) \phi_j^n(\mathcal{D}_2(\zeta)) - \mathcal{D}_1(\zeta) \phi_j^n(\mathcal{D}_1(\zeta))| \\ &\leq \|z'\|_{L^\infty} C(\phi) \varepsilon(\|z'\|_{L^\infty} \Delta x) \end{aligned}$$

since  $D(\bar{u}_j^n) \leq \Delta z_{j+1/2}$  and  $\mathcal{D}_1, \mathcal{D}_2$  are continuous at 0 with the same value.

Finally we find

$$(5.24) \quad \int R_j^n(\xi) \phi_j^n(\xi) d\xi = \frac{\Delta z_{j+1/2}}{\Delta x} u_j^n \phi_j^n(u_j^n) + \psi_j^n$$

$$(5.25) \quad = -z'_{\Delta x}(x) \int \partial_\xi(\xi \phi_j^n(\xi)) \chi_{\Delta x}(\xi, u_j^n) d\xi + \psi_j^n.$$

(ii) *The case  $D(\bar{u}_j^n) \geq \Delta z_{j+1/2}$ .* In this case  $\underline{u}_j^n = u_0$ , so because of (5.22), the right-hand side of (5.17) is equal to

$$(5.26) \quad \frac{\Delta z_{j+1/2}}{\Delta x} u_0 \phi_j^n(u_0) + \psi_{1,j}^n + \psi_{2,j}^n.$$

Thanks to (5.9),  $D(\tilde{u}_{j,-}^n) = \Delta z_{j+1/2}$  so (5.19) is equal to

$$\begin{aligned} \frac{1}{\Delta x} \int_{\tilde{u}_{j,-}^n}^{u_0} D'(\xi) \xi \phi_j^n(\xi) d\xi &= \frac{1}{\Delta x} \int_{\Delta z_{j+1/2}}^0 \mathcal{D}_1(\zeta) \phi_j^n(\mathcal{D}_1(\zeta)) d\zeta \\ &= -\frac{\Delta z_{j+1/2}}{\Delta x} \tilde{u}_{j,-}^{n*} \phi_j^n(\tilde{u}_{j,-}^{n*}) \\ &= -\frac{\Delta z_{j+1/2}}{\Delta x} u_0 \phi_j^n(u_0) + \psi_{5,j}^n, \end{aligned}$$

where  $D(\tilde{u}_{j,-}^{n*})$  is obtained by the mean value theorem, so  $|D(\tilde{u}_{j,-}^{n*}) - D(u_0)| \leq \Delta z_{j+1/2}$  and consequently  $|\psi_{5,j}^n| \leq \|z'\|_\infty C(\phi) \varepsilon (\|z'\|_{L^\infty} \Delta x)$ . Thanks to (5.7),  $D(u_{j,+}^n) = D(\bar{u}_j^n) - \Delta z_{j+1/2}$  so the term (5.18) gives

$$\frac{1}{\Delta x} \int_{u_{j,+}^n}^{\bar{u}_j^n} D'(\xi) \xi \phi_j^n(\xi) d\xi = \frac{1}{\Delta x} \int_{D(\bar{u}_j^n) - \Delta z_{j+1/2}}^{D(\bar{u}_j^n)} \mathcal{D}_2(\zeta) \phi_j^n(\mathcal{D}_2(\zeta)) d\zeta.$$

Using the main value theorem and  $u_j^n = \bar{u}_j^n$ , we obtain for (5.18)

$$\frac{\Delta z_{j+1/2}}{\Delta x} \bar{u}_{j,+}^{n*} \phi_j^n(\bar{u}_{j,+}^{n*}) = \frac{\Delta z_{j+1/2}}{\Delta x} u_j^n \phi_j^n(u_j^n) + \psi_{6,j}^n.$$

We have  $|D(\bar{u}_{j,+}^{n*}) - D(u_j^n)| \leq \Delta z_{j-1}$ , so  $|\psi_{6,j}^n| \leq \|z'\|_\infty C(\phi) \varepsilon (\|z'\|_{L^\infty} \Delta x)$ . So, we still recover (5.24) (5.25).

Finally, whatever the sign of  $D(\bar{u}_j^n) - \Delta z_{j+1/2}$  is, putting (5.25) in (5.16) and noticing that the sum is taken on  $-R/\Delta x \leq j \leq R/\Delta x$  and  $0 \leq n \leq T/\Delta t$  if  $\text{Supp}\phi \subset [0, T] \times [-R, R]$  we find

$$\begin{aligned} \int T(t, x, \xi) \phi(t, x, \xi) dt dx d\xi &= - \int z'(x) \partial_\xi (\xi \phi(t, x, \xi)) \chi_{\Delta x}(\xi, u_j^n) d\xi dt dx \\ &\quad + \tilde{\psi}_1 + \tilde{\psi}_2 + \tilde{\psi}_3 \end{aligned}$$

with

$$\begin{aligned} |\tilde{\psi}_1| &\leq C(\phi, z) \varepsilon (\|z'\|_{L^\infty} \Delta x), \\ |\tilde{\psi}_2| &\leq C(\phi) \int_{-R}^R |z_{\Delta x}(x) - z(x)| dx, \\ |\tilde{\psi}_3| &\leq C(\phi) \int_{-R}^R |z'(x) - z'(x + \Delta x)| dx. \end{aligned}$$

Therefore  $\tilde{\psi}_1, \tilde{\psi}_2, \tilde{\psi}_3$  converge to 0 when  $\Delta x$  tends to 0, which gives the desired result.  $\square$

## 6. NUMERICAL TESTS

In order to study the computational capability of the approach developed above, we have considered several test problems.

(i) We use the Burgers-Hopf equation with source term describing bathymetry in the SVS model,

$$(6.1) \quad \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \frac{u^2}{2} + z'(x)u = 0,$$

$$(6.2) \quad u(x, t = 0) = 0 \text{ for } x > 0, \quad u(x = 0, t) = 2 \text{ for } t > 0.$$

TABLE 1. Comparison of errors

Number of nodes	Engquist-Osher $L^\infty$ -error	scheme $L^1$ -error	Equilibrium $L^\infty$ -error	scheme $L^1$ -error
101	0.1651	0.4880	$6.434 \cdot 10^{-5}$	$1.263 \cdot 10^{-5}$
201	$8.452 \cdot 10^{-2}$	0.2625	-	-
401	$4.272 \cdot 10^{-2}$	0.1360	-	-
801	$2.146 \cdot 10^{-2}$	$6.847 \cdot 10^{-2}$	-	-
1001	$1.718 \cdot 10^{-3}$	$5.532 \cdot 10^{-2}$	-	-
2501	$6.864 \cdot 10^{-3}$	$2.217 \cdot 10^{-2}$	-	-
5001	$3.520 \cdot 10^{-3}$	$1.115 \cdot 10^{-2}$	-	-
10001	$3.694 \cdot 10^{-3}$	$5.561 \cdot 10^{-3}$	-	-

The steady state solution of this problem is given by the simple relation

$$(6.3) \quad u + z = 2,$$

and the discrete equilibrium states are defined by the simple formulas

$$u_{j+1}^- = u_{j+1} - \Delta z_{j+1/2}, \quad u_{j-1}^+ = u_{j-1} + \Delta z_{j-1/2}.$$

For a continuous bottom the function  $z(x)$  is chosen as

$$(6.4) \quad z(x) = \begin{cases} \cos(\pi x), & 4.5 \leq x \leq 5.5, \\ 0, & \text{otherwise.} \end{cases}$$

For a discontinuous “bottom”, we choose

$$(6.5) \quad z(x) = \begin{cases} \cos(\pi x), & 5 < x < 6, \\ 0, & \text{otherwise.} \end{cases}$$

Numerical computations of the problem (6.1), (6.4) are performed by the standard Engquist-Osher scheme with centered finite difference approximation for  $z'(x)$ , see (3.3), and we compare these results with those obtained by our equilibrium version. For each case we give the steady state obtained when the shock front corresponding to the initial and boundary values (6.2) leave the frame. The steady state solution calculated by the Engquist-Osher scheme with 101 nodal points is given in Figure 1. Notice that the errors are presented in large domains ( $x \geq 4.5$ ). Steady state solutions calculated by the corresponding equilibrium version of the scheme with 101 nodes in space are given in Figure 2. The advantage of the equilibrium scheme is evident. It should be emphasized that the standard scheme converges very slowly and mesh refinement does not help much to improve the accuracy of computations. We can see in Table 1 the errors between the approximated solutions given by the schemes and the steady state associated to the discretized bottom.

Thus even for this simple test problem numerical results obtained by the equilibrium schemes are far better than those obtained by centered source terms in the Engquist-Osher scheme with 100 times more nodes in space. Emphasizing from the experience of computations that the errors are usually amplified with the complexity and irregularity of bottoms, we can conjecture that the efficiency of equilibrium schemes will be more significant in higher dimensions and with more complicated bottoms. To demonstrate capabilities of the method to treat bottoms with singularities, computations are performed with discontinuous function  $z$  defined according to (6.5). In the case of computations by 101 nodes in space for the equilibrium scheme, deviations from steady state solution associated to the discretized bottom

at the same nodes are  $6.50883 \cdot 10^{-5}$  and  $1.28746 \cdot 10^{-5}$  in  $L^1$  and  $L^\infty$ , respectively (see Figure 3). Notice that the errors are of the same order as in the case of continuous bottom. Also notice that with centered source terms, the scheme with 101 nodes in space is unstable in this case of a discontinuous bottom. The situation improves with 1001 nodes in space and one can perform computations by standard Engquist-Osher version but the errors are too high (see Figure 4).

Thus one can conclude that computational efficiency of our equilibrium scheme is independent of the complexity of the bottom functions. Notice that the standard version with 101 nodes is unstable while equilibrium version demonstrates excellent computational capabilities and high accuracy.

(ii) The next numerical test shows the efficiency of the developed equilibrium approach even in case of simple linear flux function  $A(u) = u$ . We show that the developed approach can be successfully applied for the numerical treatment of another type of sources, namely to sources containing oscillatory coefficients. As a test problem we consider the linear equation

$$(6.6) \quad \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = z(x)u,$$

$$(6.7) \quad u(0, x) = 0,$$

$$(6.8) \quad u(t, 0) = 2,$$

$$(6.9) \quad z(x) = \cos\left(\frac{x}{\varepsilon}\right),$$

where  $\varepsilon > 0$  is a parameter. For a numerical solution of (6.6)–(6.9) the following scheme is applied:

$$(6.10) \quad \frac{u^{n+1} - u_j^n}{\Delta t} + \frac{u_j^n - u_{j-1,+}^n}{\Delta x} = 0$$

where  $u_{j-1,+}^n$  is defined from the equilibrium relation as solution of the problem

$$\frac{du}{dx} = z(x)u, \quad u(x_{j-1}) = u_{j-1}^n$$

calculated at  $x = x_j$ ; calculations are performed with 101 nodal points in space. Numerical results are given for two different times and  $\varepsilon = 0.05$ . The two graphs in Figure 5 show numerical results by standard scheme before and after steady state should be reached. They show that this method deviates greatly from the exact solution. The two graphs in Figure 6 show how the right steady state is reached by the equilibrium scheme. One can conclude as well that the equilibrium scheme gives maximum possible accuracy even with extremely rough grids. Compare the resolution provided by the grid for  $z(x)$  and the corresponding exact and numerical solution (see Figure 6).

More numerical results are given in an internal research paper, available on the web site [www.inria.fr](http://www.inria.fr).

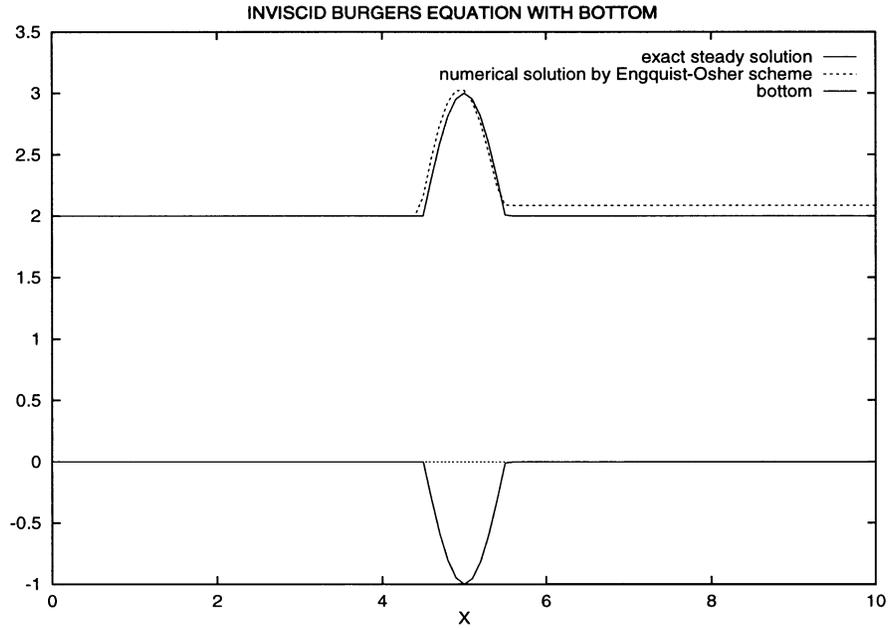


FIGURE 1. Standard Engquist-Osher scheme, 101 nodes in space

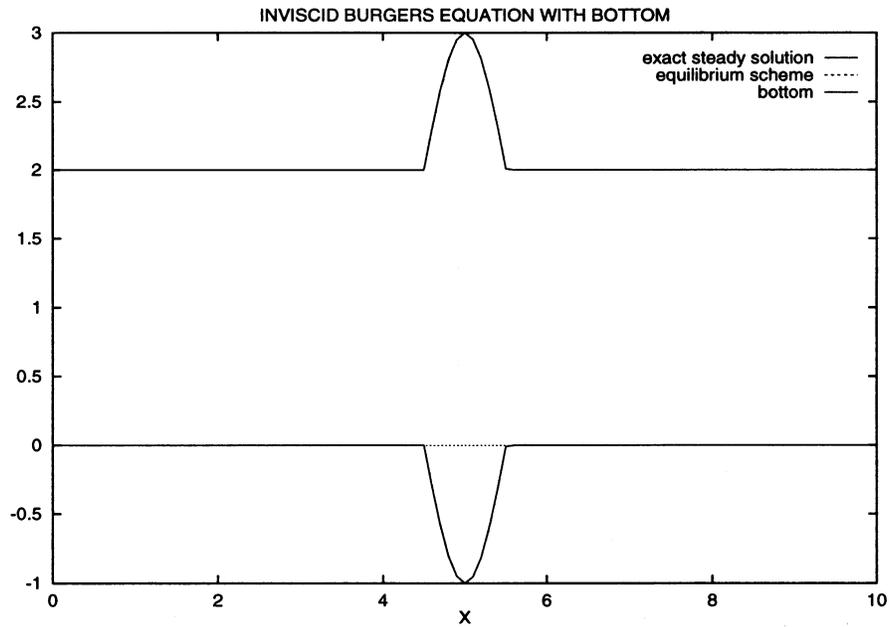


FIGURE 2. Equilibrium version of Engquist-Osher scheme, 101 nodes in space

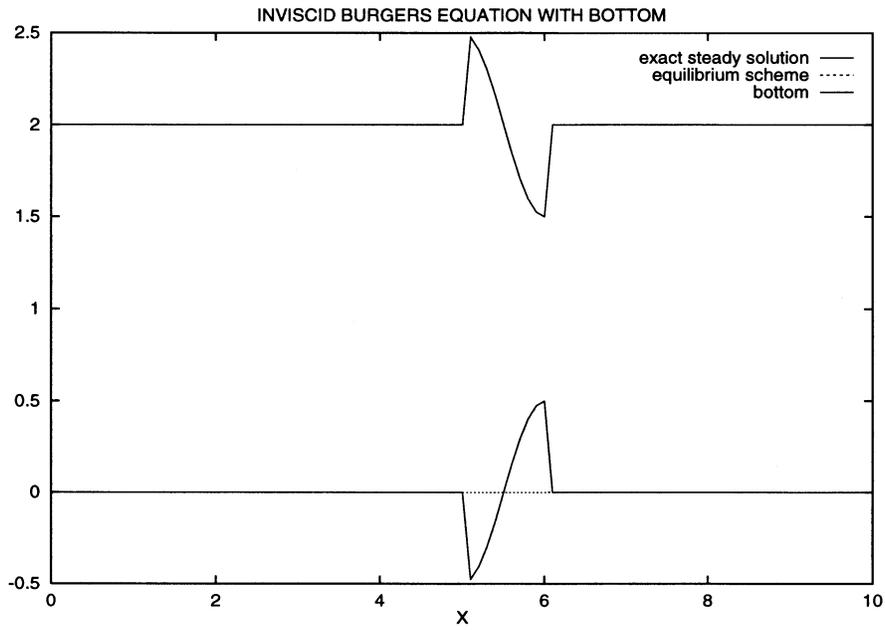


FIGURE 3. Equilibrium version of Engquist-Osher scheme, 101 nodes in space

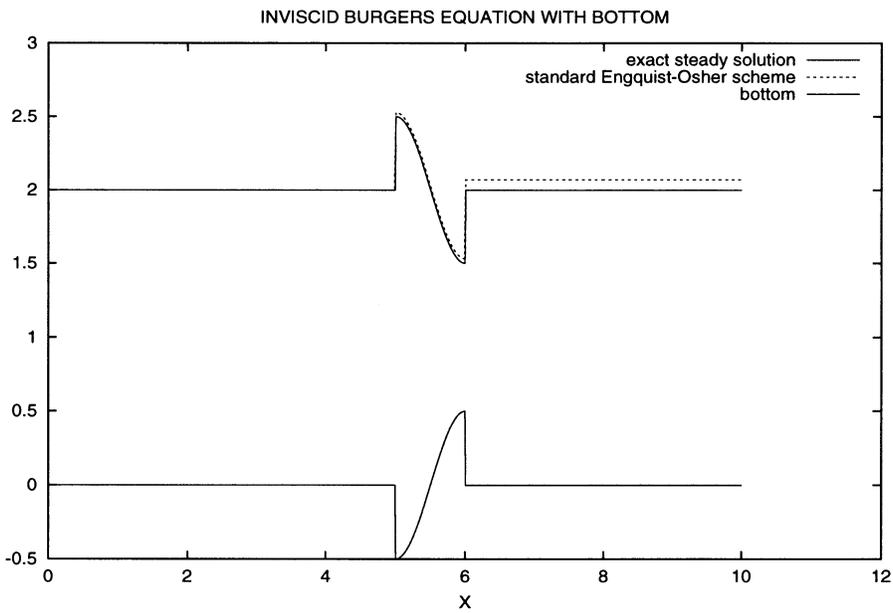
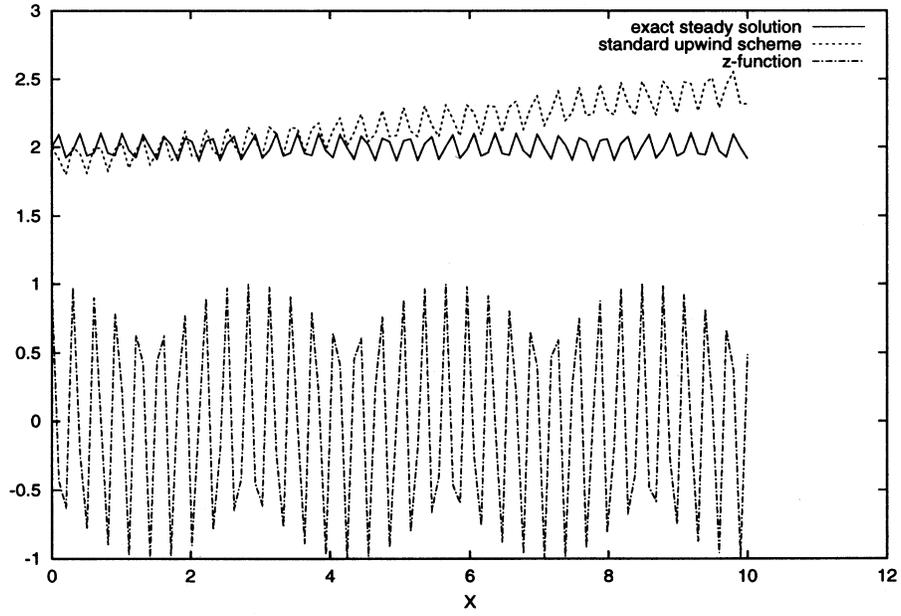
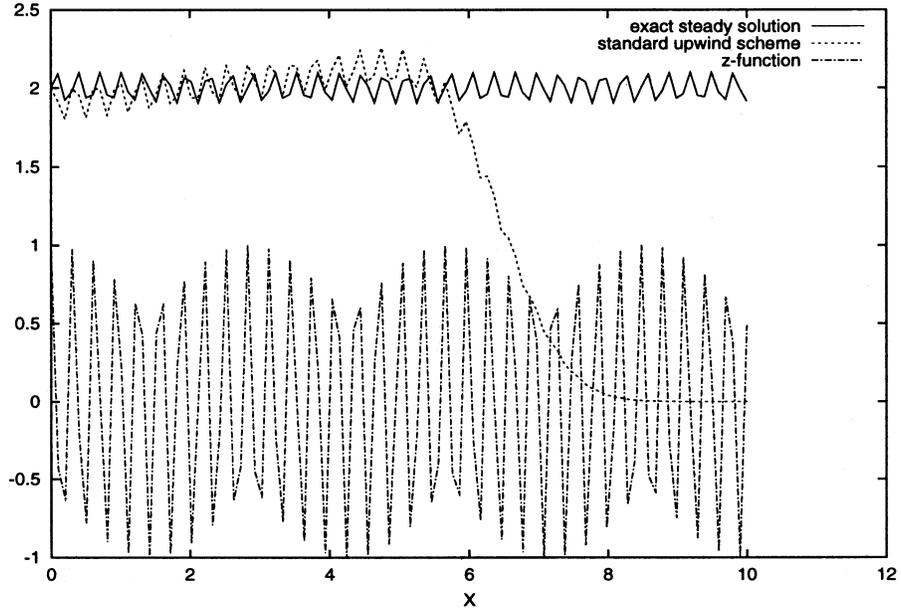
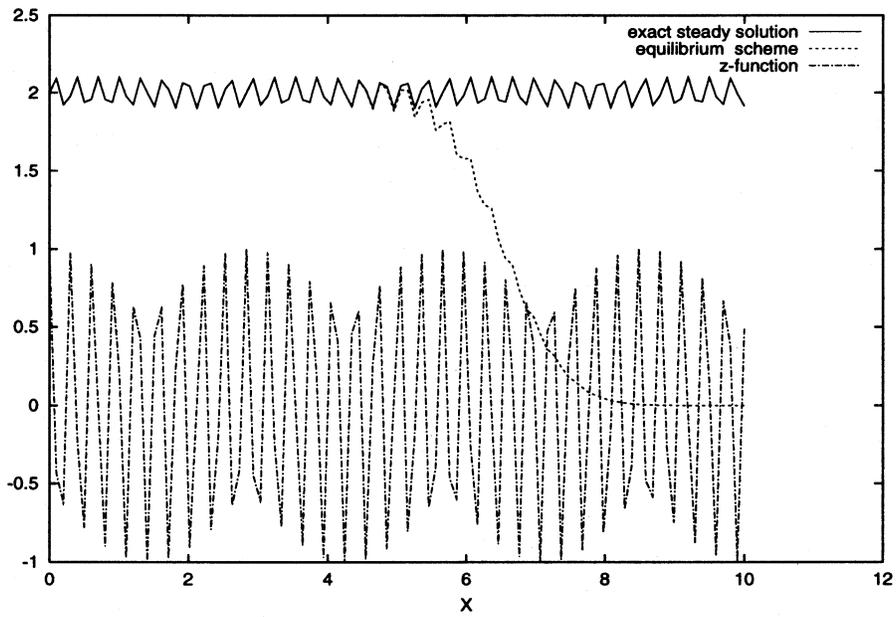
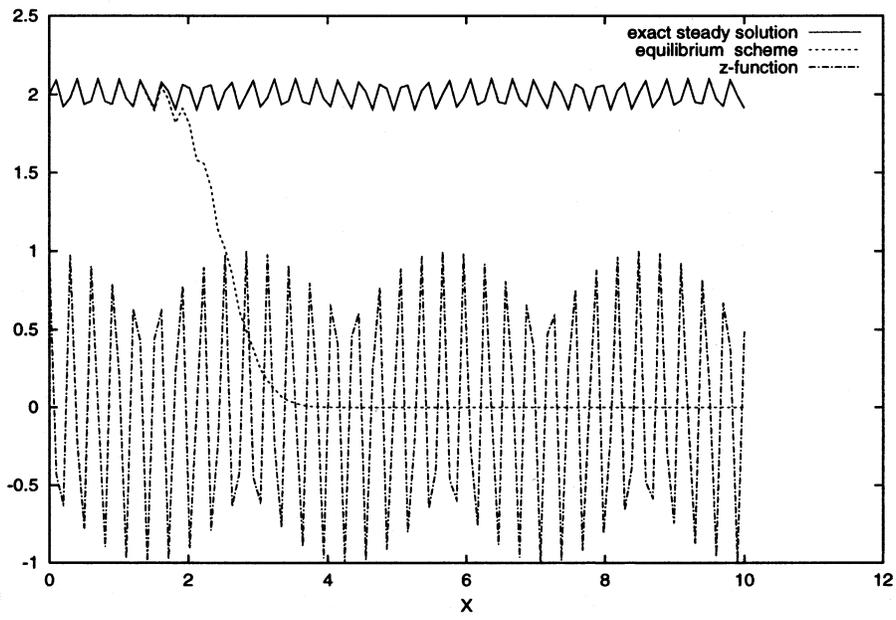


FIGURE 4. Discontinuous bottom, 1000 nodes in space

FIGURE 5. Linear advection equation, 101 nodes,  $\varepsilon = 0.05$

FIGURE 6. Linear advection equation, 101 nodes,  $\varepsilon = 0.05$

7. APPENDIX. THE SCHEME FOR NONMONOTONIC BOTTOM  $z$ 

We give here the complete scheme of Section 5 for a nonmonotonic bottom  $z$ . Again, we assume that the initial data  $u^0$  is nonnegative,  $a(\xi)$  is increasing on  $]0, +\infty[$  with  $a(u_0) = 0$  for some  $u_0 > 0$  and we still assume (5.2). Then the scheme has the form

$$(7.1) \quad u_j^{n+1} - u_j^n + \frac{\Delta t}{\Delta x} [A(u_{j+1,-}^n, u_j^n) - A(u_j^n, u_{j-1,+}^n)$$

$$(7.2) \quad + A(\tilde{u}_{j,+}^n) - A(\tilde{u}_{j,-}^n)] = 0.$$

Here  $A(\cdot, \cdot)$  still denotes the Engquist-Osher scheme

$$(7.3) \quad A(u_1, u_2) = \int_0^{u_1} a(\xi)_- d\xi + \int_0^{u_2} a(\xi)_+ d\xi,$$

with  $a(\xi)_+ = a(\xi)\mathbf{1}_{\{\xi \geq u_0\}}$ ,  $a(\xi)_- = a(\xi)\mathbf{1}_{\{0 \leq \xi \leq u_0\}}$ . And, setting

$$(7.4) \quad \underline{u}_j^n = \inf(u_0, u_j^n),$$

$$(7.5) \quad \overline{u}_j^n = \sup(u_0, u_j^n),$$

then the values  $u_{j-1,+}^n, u_{j+1,-}^n, \tilde{u}_{j,-}^n, \tilde{u}_{j,+}^n$  are defined by

$$(7.6) \quad \begin{cases} D(u_{j-1,+}^n) = \sup(0, D(\overline{u}_{j-1}^n) - \Delta z_{j-1/2}), \\ u_{j-1,+}^n \geq u_0, \end{cases}$$

$$(7.7) \quad \begin{cases} D(u_{j+1,-}^n) = \sup(0, D(\underline{u}_{j+1}^n) + \Delta z_{j+1/2}), \\ u_{j+1,-}^n \leq u_0, \end{cases}$$

$$(7.8) \quad \begin{cases} D(\tilde{u}_{j,-}^n) = \sup(0, \inf(\Delta z_{j+1/2}, D(\overline{u}_j^n))), \\ \tilde{u}_{j,-}^n \leq u_0, \end{cases}$$

$$(7.9) \quad \begin{cases} D(\tilde{u}_{j,+}^n) = \sup(0, \inf(-\Delta z_{j-1/2}, D(\underline{u}_j^n))), \\ \tilde{u}_{j,+}^n \geq u_0. \end{cases}$$

## ACKNOWLEDGMENT

The research of R.Botchorishvili was supported by NATO Fellowship No. 10/C/98/FR-EST while he was in residence at INRIA-Rocquencourt.

## REFERENCES

1. Bermudez A., Dervieux A., Desideri J-A., Vazquez M.E., Upwind schemes for two-dimensional shallow water equations with variable depth using unstructured meshes, *Comput. Methods Appl. Mech. Engrg.*, **155**(1998) 49-72. MR **99b**:76060
2. Bouchut F., Perthame B. Kruzkov's estimates for scalar conservation laws revisited, *Trans. A.M.S.* **350**(7) (1998) 2847-2870. MR **98m**:65156
3. Brenier Y., Résolution d'équations d'évolution quasilinéaires en dimensions N d'espace à l'aide d'équations linéaires en dimensions N+1, *J. Diff. Eq.* **50**(3) (1982) 375-390. MR **85f**:35117
4. Chen G.-Q., Levermore C.D., Liu T.P., Hyperbolic conservation laws with stiff relaxation terms and entropy, *Comm. Pure Appl. Math.* **48**(6) (1995) 787-830. MR **95h**:35133
5. Coquel F., Perthame B., Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics, *SIAM J. Num. Anal.* **35**(6) (1998) 2223-2249. MR **2000a**:76129
6. DiPerna R.J., Measure valued solutions to conservation laws, *Arch. Rat. Mech. Anal.* **88** (1985) 223-270. MR **86g**:35121
7. Engquist B., Osher S., Stable and entropy satisfying approximations for transonic flow calculations, *Math. Comp.* **34** (1980) 45-75. MR **81b**:65082

8. Eymard R., Gallouët T., Herbin R., Existence and uniqueness of the entropy solution to a nonlinear hyperbolic equation, *Chinese Ann. Math. Ser. B* **16** (1) (1995) 1–14. MR **96m**:35202
9. Gosse L., Leroux A.-Y., A well-balanced scheme designed for inhomogeneous scalar conservation laws, *C. R. Acad. Sc., Paris, Sér. I Math.* **323** (1996) 543–546. MR **97i**:35112
10. Gosse L., Localization effects and measure source terms in numerical schemes for balance laws, Preprint.
11. Greenberg J. M., Leroux A.-Y., Baraille R., Noussair A., Analysis and approximation of conservation laws with source terms, *SIAM J. Numer. Anal.* **34** (5)(1997) 1980–2007. MR **98k**:65049
12. Giga Y., Miyakawa T., A kinetic construction of global solutions of first-order quasilinear equations, *Duke Math. J.* **50** (1983) 505–515. MR **85g**:35017
13. Kruzkov S.N., Generalized solutions of the Cauchy problem in the large for nonlinear equations of first order, *Dokl. Akad. Nauk. SSSR* **187**(1) (1970) 29–32; English trans, *Soviet Math. Dokl.* **10** (1969). MR **40**:3046
14. Kuznetsov N.N., Finite difference schemes for multidimensional first order quasilinear equation in classes of discontinuous functions, in: “*Probl. Math. Phys. Vych. Mat.*”. Moscow: Nauka (1977) 181–194. MR **80b**:65121
15. J.O. Langseth, A. Tveito and R. Winther, On the convergence of operator splitting applied to conservation laws with source terms, *SIAM J. Num. Anal.* **33** (1996) 843–863. MR **97b**:65106
16. Lax P., Shock waves and entropy, in: “*Contributions to Nonlinear Functional Analysis.*” E.H. Zarantonello, ed. New York: Academic Press (1971) 603–634. MR **52**:14677
17. Leveque R., *Numerical Methods for Conservation Laws, Lectures in Mathematics*, ETH Zurich, Birkhauser (1992). MR **92m**:65106
18. Lions P.L., Perthame B., Tadmor E., A kinetic formulation of multidimensional scalar conservation laws and related equations, *J. Amer. Math. Soc.* **7** (1994) 169–191. MR **94d**:35100
19. Natalini R., Convergence to equilibrium for the relaxation approximations of conservation laws, *Comm. Pure Appl. Math.* **49** (1996) 795–823. MR **97c**:35131
20. Perthame B., Uniqueness and error estimates in first order quasilinear conservation laws via the kinetic entropy defect measure, *J. Math. P. et Appl.* **77** (1998) 1055–1064. MR **2000e**:35141
21. Perthame B., Tzavaras A., Kinetic formulation for systems of two conservation laws and elastodynamics, *Arch. Ration. Mech. Anal.* **155** (2000) 1–48. MR **2001h**:74038
22. Russo G., personal communication.
23. Sanders R., On the convergence of monotone finite difference schemes with variable spatial differencing, *Math. Comp.*, **40** (161), (1983) 91–106. MR **84a**:65075
24. Szepessy A., Convergence of a streamline diffusion finite element method for conservation law with boundary conditions, *RAIRO Model. Math. et Anal. Num.* **25** (1991) 749–782. MR **92g**:65115
25. Vasseur A., Time regularity for the system of isentropic gas dynamics with  $\gamma = 3$ , *Comm. in P.D.E.* **24** (1999) 1987–1997. MR **2000i**:35171
26. Vazquez-Cendon M.E., Improved treatment of source terms in upwind schemes for shallow water equations in channels with irregular geometry, *J. Comput. Phys.*, **148**(2) (1999) 497–526. MR **99i**:76102

VIAM, TBILISSI STATE UNIVERSITY, 2 UNIVERSITY STREET, 380043 TBILISSI, GEORGIA  
*E-mail address:* `rdboch@viam.hepi.edu.ge`

INRIA, M3N, DOMAINE DE VOLUCEAU, BP 105, F78153 LE CHESNAY  
*Current address:* ENS, DMA, 45, rue d’Ulm, F75230 Paris cédex 05, France  
*E-mail address:* `benoit.perthame@ens.fr`

LABORATOIRE J.A. DIEUDONNÉ, UMR 6621, UNIVERSITÉ NICE-SOPHIA ANTIPOLIS, PARC VAL-ROSE, F-06108 NICE CEDEX 02, FRANCE  
*E-mail address:* `vasseur@math3.unice.fr`