SIAM J. NUMER. ANAL.
Vol. 44, No. 1, pp. 300–321

NUMERICAL METHODS FOR NONCONSERVATIVE HYPERBOLIC SYSTEMS: A THEORETICAL FRAMEWORK.*

CARLOS PARÉS†

Abstract. The goal of this paper is to provide a theoretical framework allowing one to extend some general concepts related to the numerical approximation of 1-d conservation laws to the more general case of first order quasi-linear hyperbolic systems. In particular this framework is intended to be useful for the design and analysis of well-balanced numerical schemes for solving balance laws or coupled systems of conservation laws. First, the concept of path-conservative numerical schemes is introduced, which is a generalization of the concept of conservative schemes for systems of conservation laws. Then, we introduce the general definition of approximate Riemann solvers and give the general expression of some well-known families of schemes based on these solvers: Godunov, Roe, and relaxation methods. Finally, the general form of a high order scheme based on a first order path-conservative scheme and a reconstruction operator is presented.

Key words. nonconservative products, finite volume method, well-balanced schemes, approximate Riemann solvers, Godunov methods, Roe methods, relaxation methods, high order methods

AMS subject classifications. 74S10, 65M06, 35L60, 35L65, 35L67

DOI. 10.1137/050628052

1. Introduction.

The motivating question of this paper was the design of numerical schemes for P.D.E. systems that can be written under the form

$$\partial_t w + \partial_x F(w) = \mathcal{B}(w) \cdot \partial_x w + S(w)\partial_x \sigma, \tag{1}$$

where the unknown $w(x,t)$ takes values on an open convex set D of $\mathbb{R}^N$; F is a regular function from D to $\mathbb{R}^N$; $\mathcal{B}$ is a regular matrix function from D to $\mathcal{M}_{N\times N}(\mathbb{R})$; S, a function from D to $\mathbb{R}^N$; and $\sigma(x)$, a known function from $\mathbb{R}$ to $\mathbb{R}$.

System (1) includes as particular cases: systems of conservation laws ($\mathcal{B} = 0$, $S = 0$), systems of conservation laws with source term or balance laws ($\mathcal{B} = 0$), and coupled system of balance laws as defined in [7].

More precisely, the discretization of the shallow water systems that govern the flow of one shallow layer or two superposed shallow layers of immiscible homogeneous fluids was focused (see http://www.damflow.org). The corresponding systems can be written, respectively, as a balance law or a coupled system of two balance laws. Systems with similar characteristics also appear in other flow models such as two-phase flows.

It is well known that standard methods that solve correctly systems of conservation laws can fail in solving systems of balance laws, specially when approaching equilibria or near to equilibria solutions. Moreover, they can produce unstable methods when they are applied to coupled systems of conservation or balance laws. In the context of the numerical analysis of systems and coupled systems of balance laws, many authors have studied the design of well-balanced schemes, that is, schemes that

*Received by the editors March 30, 2005; accepted for publication (in revised form) July 22, 2005; published electronically March 7, 2006. This research has been partially supported by the Spanish Government Research project BFM2003-07530-C02-02.

http://www.siam.org/journals/sinum/44-1/62805.html

†Departamento de Análisis Matemático, Facultad de Ciencias, Universidad de Málaga, 29071-Málaga, Spain (pares@anamat.cie.uma.es).

preserve some equilibria: see [2], [3], [5], [7], [10], [11], [12], [17], [18], [23], [24], [27], [29], [31], [32], [36], [37], [38], [39], ...

Among the main techniques used in the derivation of well-balanced numerical schemes, one of them consists of choosing first a standard conservative scheme for the discretization of the flux terms and then discretizing the source and the coupling terms in order to obtain a consistent scheme which solves correctly a predetermined family of equilibria. If this first procedure is followed, the calculation of the correct discretization of the source and the coupling terms depends both on the specific problem and the conservative numerical scheme chosen, and it may become rather cumbersome. In [11] it was shown that the technique of modified equations can be helpful in this procedure.

Another technique consists of considering (1) as a particular case of one-dimensional quasi-linear hyperbolic system

$$\frac{\partial W}{\partial t} + \mathcal{A}(W)\frac{\partial W}{\partial x} = 0, \quad x \in \mathbb{R}, t > 0, \tag{2}$$

by adding to the system the trivial equation

$$\frac{\partial \sigma}{\partial t} = 0.$$

Once the system is rewritten under this form, piecewise constant approximations of the solutions are considered, then are updated by means of approximate Riemann solvers at the intercells.

If this second procedure is followed, the main difficulty both from the mathematical and the numerical points of view comes from the presence of nonconservative products, which makes difficult even the definition of weak solutions. Many papers have been devoted to the definition and stability of nonconservative products, and its application to the definition of weak solutions of nonconservative hyperbolic systems; see [1], [4], [6], [9], [13], [14], [26], [34], [41].

In this article we assume the definition of nonconservative products as Borel measures given by Dal Maso, LeFloch, and Murat in [14]. This definition, which depends on the choice of a family of paths in the phases space, allows one to give a rigorous definition of weak solutions of (2). Together with the definition of weak solutions, a notion of *entropy* has to be chosen as the usual Lax's concept or one related to an entropy pair. The classical theory of simple waves of hyperbolic systems of conservation laws and the results concerning the solutions of Riemann problems can then be extended to systems (2).

The choice of the family of paths may be, in general, a difficult task. The goal of this article is, once the choice is done, to provide a theoretical framework for the numerical approximation of the corresponding weak solutions of a strictly hyperbolic system (2) whose characteristic fields are either genuinely nonlinear or linearly degenerate.

The organization of the article is as follows: in section 2, a brief resume of the theory developed in [14] is presented, together with some remarks concerning the choice of paths and some properties of weak solutions.

In section 3 we introduce the concept of path-conservative numerical schemes, which is a generalization of that of conservative schemes for systems of conservation laws: a scheme will be said to be path-conservative if it conserves to some extent the Borel measure related to the nonconservative products.

Section 4 is devoted to the well-balance property: we recall the general definition of a well-balanced numerical scheme proposed in [29] and show that the well-balance property of a scheme is strongly related to its ability to approach stationary contact discontinuities.

In section 5, a general definition of approximate Reimann solvers for (2) is presented. We verify that the generalizations of the classical methods of Roe [35] and Godunov [16] presented respectively in [40] and [30] are particular cases of path-conservative methods based on approximate Riemann solvers fitting this general definition. We give also some guidelines about how to construct relaxation schemes.

Section 6 is devoted to high order methods based on reconstruction techniques. The general form of a scheme based on a first order path-conservative scheme and a reconstruction operator is presented. The schemes constructed in [8] are particular cases in which the first order method is of the Roe type. Some general results concerning the order and well-balance properties of these methods are finally presented.

2. Weak solutions. Consider the problem

$$\frac{\partial W}{\partial t} + \mathcal{A}(W)\frac{\partial W}{\partial x} = 0, \quad x \in \mathbb{R},\ t > 0, \tag{3}$$

where $W(x,t)$ belongs to Ω, an open convex subset of $\mathbb{R}^N$, and $W \in \Omega \mapsto \mathcal{A}(W) \in \mathcal{M}_{N\times N}(\mathbb{R})$ is a smooth locally bounded map. We suppose that system (3) is strictly hyperbolic, that is, for each $W \in \Omega$, $\mathcal{A}(W)$ has N real distinct eigenvalues $\lambda_1(W) < \cdots < \lambda_N(W)$, with associated eigenvectors $R_1(W),\ldots,R_N(W)$. We also suppose that for each $i = 1,\ldots,N$, the characteristic field $R_i(W)$ is either genuinely nonlinear,

$$\nabla\lambda_i(W)\cdot R_i(W) \neq 0, \quad \forall\ W \in \Omega,$$

or linearly degenerate,

$$\nabla\lambda_i(W)\cdot R_i(W) = 0, \quad \forall\ W \in \Omega.$$

The theory developed by Dal Maso, LeFloch, and Murat (see [14]) allows one to give a rigorous definition of nonconservative products associated with the choice of a family of paths in Ω.

DEFINITION 2.1. *A* family of paths *in* $\Omega \subset \mathbb{R}^N$ *is a locally Lipschitz map*

$$\Phi\colon [0,1]\times\Omega\times\Omega \mapsto \Omega,$$

such that:

- $\Phi(0;W_L,W_R) = W_L$ *and* $\Phi(1;W_L,W_R) = W_R$*, for any* $W_L, W_R \in \Omega$*;*
- *for every arbitrary bounded set* $\mathcal{O} \subset \Omega$*, there exists a constant* k *such that*

$$\left|\frac{\partial\Phi}{\partial s}(s;W_L,W_R)\right| \leq k|W_R - W_L|,$$

for any $W_L, W_R \in \mathcal{O}$ *and almost every* $s \in [0,1]$*;*

- *for every bounded set* $\mathcal{O} \subset \Omega$*, there exists a constant* K *such that*

$$\left|\frac{\partial\Phi}{\partial s}(s;W_L^1,W_R^1) - \frac{\partial\Phi}{\partial s}(s;W_L^2,W_R^2)\right| \leq K(|W_L^1 - W_L^2| + |W_R^1 - W_R^2|),$$

for any $W_L^1, W_R^1, W_L^2, W_R^2 \in \mathcal{O}$ *and almost every* $s \in [0,1]$.

Suppose that a family of paths Φ in Ω has been chosen. Then, for $W \in (L^\infty(\mathbb{R} \times \mathbb{R}^+) \cap BV(\mathbb{R} \times \mathbb{R}^+))^N$, the nonconservative product can be interpreted as a Borel measure denoted by $[\mathcal{A}(W)W_x]_\Phi$. If the family of segments is chosen, this interpretation is equivalent to the definition of nonconservative product proposed by Volpert in [41].

Across a discontinuity with speed ξ a weak solution must satisfy the generalized Rankine–Hugoniot condition

$$\int_0^1 \left(\xi\mathcal{I} - \mathcal{A}(\Phi(s; W^-, W^+))\right) \frac{\partial \Phi}{\partial s}(s; W^-, W^+)\, ds = 0, \tag{4}$$

where $\mathcal{I}$ is the identity matrix and W^-, W^+ are the left and right limits of the solution at the discontinuity. In the particular case of a system of conservation laws (that is, if $\mathcal{A}(W)$ is the Jacobian matrix of some flux function $F(W)$), (4) is independent of the family of paths and it reduces to the usual Rankine–Hugoniot condition.

As it occurs in the conservative case, not every discontinuity is admissible. Therefore, a concept of entropic solution has to be assumed, as one of the following definitions.

DEFINITION 2.2. *A weak solution is said to be an entropic solution in the Lax sense if, at each discontinuity, there exists $i \in \{1, \dots, N\}$ such that*

$$\lambda_i(W^+) < \xi < \lambda_{i+1}(W^+) \quad \textit{and} \quad \lambda_{i-1}(W^-) < \xi < \lambda_i(W^-)$$

if the ith characteristic field is genuinely nonlinear or

$$\lambda_i(W^-) = \xi = \lambda_i(W^+)$$

if the ith characteristic field is linearly degenerate.

DEFINITION 2.3. *Given an entropy pair (η, G) for (3), i.e., a pair of regular functions from Ω to $\mathbb{R}$ such that*

$$\nabla G(W) = \nabla \eta(W) \cdot \mathcal{A}(W), \quad \forall\ W \in \Omega,$$

a weak solution is said to be entropic if it satisfies the inequality

$$\partial_t \eta(W) + \partial_x G(W) \le 0,$$

in the distributions sense.

The choice of the family of paths is important as it determines the speed of propagation of discontinuities. For scalar balance laws, rigorous justifications of the choice of the family of paths can be given using different techniques based on weak limits; see [19], [20]. In general, this choice has to be based on the physical background (see [25], [33] for instance). In any case, it is natural from the mathematical point of view to require this family to satisfy some hypotheses concerning the relation of the paths with the integral curves of the characteristic fields. Following [30], here we will assume that the family of paths satisfies the following hypotheses:

(H1) Given two states, W_L and W_R, belonging to the same integral curve γ of a linearly degenerate field, the path $\Phi(s; W_L, W_R)$ is a parameterization of the arc of γ linking W_L and W_R.

(H2) Given two states, W_L and W_R, belonging to the same integral curve γ of a genuinely nonlinear field, R_i, such that $\lambda_i(W_L) < \lambda_i(W_R)$, the path $\Phi(s; W_L, W_R)$ is a parameterization of the arc of γ linking W_L and W_R.

(H3) Let us denote by $\mathcal{RP} \subset \Omega \times \Omega$ the set of pairs (W_L, W_R) such that the Riemann problem

$$\left\{ \begin{array}{l} \dfrac{\partial W}{\partial t} + \mathcal{A}(W)\dfrac{\partial W}{\partial x} = 0, \\ \\ W(x,0) = \left\{ \begin{array}{ll} W_L & \text{if } x < 0, \\ \\ W_R & \text{if } x > 0, \end{array} \right. \end{array} \right. \tag{5}$$

has a unique self-similar solution $W(x,t) = V(x/t; W_L, W_R)$ (where the function V is piecewise regular) composed by at most N simple waves: rarefaction waves, contact discontinuities, or shocks (i.e., discontinuities satisfying the jump condition (4) and the entropy condition given by Definition 2.2 or 2.3). These simple waves connect $J+1$ intermediate states

$$W_0 = W_L; W_1, \ldots, W_{J-1}; W_J = W_R;$$

with $J \leq N$. We assume that, given two states $(W_L, W_R) \in \mathcal{RP}$, the curve described by the path $\Phi(s; W_L, W_R)$ in Ω is equal to the union of those corresponding to the paths $\Phi(s; W_j, W_{j+1})$, $j = 0, \ldots, J-1$.

If the definition of weak solutions of (3) is based on a family of paths satisfying these hypotheses, the following natural properties hold (see [30]).

PROPOSITION 2.4. *Let us suppose that the concept of weak solutions of* (3) *is defined on the basis of a family of paths satisfying hypotheses* (H1)–(H3). *Then*

(i) *Given two states W_L and W_R belonging to the same integral curve of a linearly degenerate field, the contact discontinuity given by*

$$W(x,t) = \left\{ \begin{array}{ll} W_L & \text{if } x < \sigma t, \\ W_R & \text{if } x > \sigma t, \end{array} \right.$$

where σ is the (constant) value of the corresponding eigenvalue through the integral curve, is a weak solution of (3).

(ii) *Let (W_L, W_R) be a pair belonging to $\mathcal{RP}$ and let W be the solution of the corresponding Riemann problem* (5). *The following equality holds:*

$$\left\langle \left[\mathcal{A}(W(\cdot,t))W_x(\cdot,t)\right]_\Phi, 1 \right\rangle = \int_0^1 \mathcal{A}\left(\Phi(s; W_L, W_R)\right) \frac{\partial \Phi}{\partial s}(s; W_L, W_R)\, ds.$$

Consequently, the total mass of the Borel measure $[\mathcal{A}(W(\cdot,t))W_x(\cdot,t)]_\Phi$ does not depend on t.

(iii) *Let (W_L, W_R) be a pair belonging to $\mathcal{RP}$ and let W_j be any of the intermediate states appearing in the solution of the Riemann problem* (5). *Then*

$$\begin{aligned} &\int_0^1 \mathcal{A}\left(\Phi(s; W_L, W_R)\right) \frac{\partial \Phi}{\partial s}(s; W_L, W_R)\, ds \\ &\quad = \int_0^1 \mathcal{A}\left(\Phi(s; W_L, W_j)\right) \frac{\partial \Phi}{\partial s}(s; W_L, W_j)\, ds \\ &\quad + \int_0^1 \mathcal{A}\left(\Phi(s; W_j, W_R)\right) \frac{\partial \Phi}{\partial s}(s; W_j, W_R)\, ds. \end{aligned}$$

Some general guidelines to construct a family of paths satisfying these hypotheses (at least for pairs $(W_L, W_R) \in \mathcal{RP}$) have been presented in [30].

In the following proposition we establish a property of the solution of a Riemann problem that will be of importance in the definition of generalized approximate Riemann solvers for (3).

PROPOSITION 2.5. *Given $(W_L, W_R) \in \mathcal{RP}$, the solution $W(x,t) = V(x/t; W_L, W_R)$ of the Riemann problem* (5) *satisfies the following equality:*

$$\int_0^1 \mathcal{A}\left(\Phi(s; W_L, W_R)\right) \frac{\partial \Phi}{\partial s}(s; W_L, W_R)\, ds + \int_0^\infty \left(V(v; W_L, W_R) - W_R\right)\, dv + \int_{-\infty}^0 \left(V(v; W_L, W_R) - W_L\right)\, dv = 0. \tag{6}$$

Proof. Let A, T be two positive numbers such that

$$V(x/T; W_L, W_R) = W_L, \quad \text{if } x < -A,$$
$$V(x/T; W_L, W_R) = W_R, \quad \text{if } x > A.$$

Integrating (3) in $[-A, A] \times [0, T]$, we obtain

$$\int_{-A}^{A} V(x/T; W_L, W_R)\, dx - AW_L - AW_R + \int_0^T \left\langle \left[\mathcal{A}(W(\cdot,t))W_x(\cdot,t)\right]_\Phi, 1 \right\rangle\, dt = 0.$$

Then (6) is easily obtained by taking into account (ii) of Proposition 2.4 and making the change of variables $v = x/T$ in the integral at the right-hand side. ∎

Remark 1. If the concept of entropic solution is related to an entropy pair (η, G) with convex η, the following inequality can also be proved for the solution of a Riemann problem:

$$G(W_R) + \int_0^\infty \left(\eta\big(V(v; W_L, W_R)\big) - \eta(W_R)\right)\, dv \le G(W_L) - \int_{-\infty}^0 \left(\eta\big(V(v; W_L, W_R)\big) - \eta(W_L)\right)\, dv. \tag{7}$$

The proof is identical to that corresponding to systems of conservation laws.

3. Path-conservative numerical schemes. The central concept of the theory developed in this article is that of *path-conservative* numerical scheme, which is a generalization of conservative schemes for systems of conservation laws. We recall that, given a system of conservation laws

$$\partial_t W + \partial_x F(W) = 0, \quad x \in \mathbb{R},\ t > 0, \tag{8}$$

the expression of a conservative numerical scheme is as follows:

$$W_i^{n+1} = W_i^n + \frac{\Delta t}{\Delta x}\left(G_{i-1/2} - G_{i+1/2}\right), \tag{9}$$

where Δt and Δx are the time step and the space step, which are supposed to be constant for simplicity; W_i^n represents the approximation of the average of the exact solution at the ith cell $I_i = [x_{i-1/2}, x_{i+1/2}]$ at time $t^n = n\Delta t$,

$$W_i^n \cong \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} W(x, t^n)\, dx,$$

and $G_{i+1/2} = G(W^n_{i-q}, \dots, W^n_{i+p})$ is the numerical flux at the intercell $x_{i+1/2}$,

$$G_{i+1/2} \cong \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} F(W(x_{i+1/2}, t))\, dt. \tag{10}$$

This expression is usually motivated as follows: a weak solution of (8) satisfies the equality

$$\int_a^b W(x, t_1)\, dx = \int_a^b W(x, t_0)\, dx + \int_{t_0}^{t_1} F(W(a, t))\, dt - \int_{t_0}^{t_1} F(W(b, t))\, dt, \tag{11}$$

for every rectangle $[a, b] \times [t_0, t_1]$ in $\mathbb{R} \times [0, \infty)$, and (9) is the discrete analogue of the equality (11) corresponding to the rectangle $I_i \times [t^n, t^{n+1}]$.

Let us give a reinterpretation of (9) in terms of measures in order to motivate its generalization to nonconservative problems. A weak solution can be understood as a function that satisfies the equality (8) in the sense of distributions. In the particular case of a piecewise regular weak solution, given $t > 0$ the distribution $[F(W(\cdot, t))_x]$ is defined by

$$\begin{aligned} \langle [F(W(\cdot, t))_x], \phi \rangle &= \int_{\mathbb{R}} F(W(x, t))_x \phi(x)\, dx \\ &\quad + \sum_l \left(F(W_l^+) - F(W_l^-) \right) \phi(x_l(t)), \quad \forall\ \phi \in \mathcal{D}(\mathbb{R})^N, \end{aligned} \tag{12}$$

where the derivative appearing in the integral term has to be understood in the pointwise sense; the index l of the sum runs in the number of discontinuities appearing in the solution; $x_l(t)$ is the location at time t of the lth discontinuity; W_l^- and W_l^+ the limits of the solution to the left and right of the lth discontinuity at time t; finally, $\mathcal{D}(\mathbb{R})$ represents the set of functions of class $\mathcal{C}^\infty(\mathbb{R})$ with compact support. The distribution $[F(W(\cdot, t))_x]$ can be interpreted as a Borel measure having the Lebesgue decompostion $\mu_a + \mu_s$, where μ_a is given by

$$\mu_a(E) = \int_E F(W(x, t))_x\, dx,$$

for every Borel set E, and

$$\mu_s = \sum_l \left(F(W_l^+) - F(W_l^-) \right) \delta_{x = x_l(t)}, \tag{13}$$

being $\delta_{x=a}$ the Dirac measure placed at $x = a$. Given a Borel set E, we will denote its measure by

$$\langle [F(W(\cdot, t))_x], 1_E \rangle.$$

Using this notation, (11) can be rewritten as follows:

$$\int_a^b W(x, t_1)\, dx = \int_a^b W(x, t_0)\, dx - \int_{t_0}^{t_1} \left\langle [F(W(\cdot, t))_x], 1_{[a,b]} \right\rangle\, dt. \tag{14}$$

If we now define the piecewise constant function W^n whose value at the cell I_i is the approximation W_i^n, the discrete analogue of (14) would be

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x} \langle [F(W^n)_x], 1_{I_i} \rangle, \tag{15}$$

but this equality is not equivalent to (9): notice that the measure $[F(W^n)_x]$ consists only of its singular part

$$\sum_i \left(F(W_{i+1}^n) - F(W_i^n)\right) \delta_{x=x_{i+1/2}},$$

and that the cells I_i have been defined as closed intervals. Therefore, in (15) the punctual mass placed at $x_{i+1/2}$ contribute both to cells I_i and I_{i+1}. In this sense, the conservative numerical scheme (9) can be interpreted as follows: $G_{i+1/2}$ can be considered as an *intermediate flux* that is used to split the Dirac measures placed at the intercells

$$\begin{aligned}\left(F(W_{i+1}^n) - F(W_i^n)\right) \delta_{x=x_{i+1/2}} &= \left(F(W_{i+1}^n) - G_{i+1/2}\right) \delta_{x=x_{i+1/2}} \\ &+ \left(G_{i+1/2} - F(W_i^n)\right) \delta_{x=x_{i+1/2}},\end{aligned}$$

and then, the first summand contributes to cell I_{i+1} and the second one to I_i, i.e.,

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\left((F(W_i^n) - G_{i-1/2})) + (G_{i+1/2} - F(W_i^n))\right), \tag{16}$$

which is obviously equivalent to (9).

Let us now come back to nonconservative systems (3) and suppose that a family of paths Φ has been chosen to define the weak solutions. If W is again a piecewise regular weak solution, for a given time t the Borel measure related to the nonconservative product is defined as follows:

$$\begin{aligned}\langle [\mathcal{A}(W(\cdot,t))W_x(\cdot,t)]_\Phi, \phi\rangle &= \int_{\mathbb{R}} \mathcal{A}(W(x,t))W_x(x,t)\phi(x)\,dx \\ &+ \sum_l \left(\int_0^1 \mathcal{A}(\Phi(s;W_l^-,W_l^+))\frac{\partial\Phi}{\partial s}(s;W_l^-,W_l^+)\,ds\right)\phi(x_l(t)), \\ &\forall\,\phi\in\mathcal{C}_0(\mathbb{R}),\end{aligned} \tag{17}$$

which is obviously a generalization of (12). In the above equality, the expression W_x appearing in the first integral represents again the pointwise derivative of $W(\cdot,t)$; $x_l(t)$, W_l^-, W_l^+ are like in (12); and $\mathcal{C}_0(\mathbb{R})$ is the set of continuous maps with compact support.

Notice that again this measure can be decomposed as a sum $\mu_a^\Phi + \mu_s^\Phi$ where

$$\mu_a^\Phi(E) = \int_E \mathcal{A}(W(x,t))W_x(x,t)\,dx$$

for every Borel set E, and:

$$\mu_s^\Phi = \sum_l \left(\int_0^1 \mathcal{A}(\Phi(s;W_l^-,W_l^+))\frac{\partial\Phi}{\partial s}(s;W_l^-,W_l^+)\,ds\right)\delta_{x=x_l(t)}. \tag{18}$$

Given a rectangle $[a,b]\times[t_0,t_1]$ in $\mathbb{R}\times[0,\infty)$, a weak solution of (3) satisfies the equality

$$\int_a^b W(x,t_1)\,dx = \int_a^b W(x,t_0)\,dx - \int_{t_0}^{t_1} \langle [\mathcal{A}(W(\cdot,t))W_x(\cdot,t)]_\Phi, 1_{[a,b]}\rangle\,dt \tag{19}$$

that generalizes (11).

The discrete analogue of (19) is now

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\langle [\mathcal{A}(W^n)W_x^n]_\Phi \, , 1_{I_i}\rangle, \tag{20}$$

where, again, W^n is the piecewise constant function taking the value W_i^n at cell I_i. Newly, the measure $[\mathcal{A}(W^n)W_x^n]_\Phi$ consists only of its singular part,

$$\sum_i \Big(\int_0^1 \mathcal{A}(\Phi(s; W_i^n, W_{i+1}^n)) \frac{\partial \Phi}{\partial s}(s; W_i^n, W_{i+1}^n)\, ds \Big) \delta_{x=x_{i+1/2}}.$$

Therefore, the punctual masses placed at the intercells have to be decomposed into two terms $D_{i+1/2}^{\pm}$, one contributing to the cell I_i and the other to the cell I_{i+1}. This idea leads to the following definition.

DEFINITION 3.1. *Given a family of paths Ψ, a numerical scheme is said to be Ψ-conservative if it can be written under the form*

$$W_i^{n+1} = W_i^n - \frac{\Delta t}{\Delta x}\big(D_{i-1/2}^+ + D_{i+1/2}^-\big), \tag{21}$$

where

$$D_{i+1/2}^{\pm} = D^{\pm}(W_{i-q}^n, \dots, W_{i+p}^n),$$

D^- and D^+ being two continuous functions from Ω^{p+q+1} to Ω satisfying:

$$D^{\pm}(W, \dots, W) = 0, \quad \forall\ W \in \Omega, \tag{22}$$

and

$$\begin{aligned} D^-(W_{-q}, \dots, W_p) + D^+(W_{-q}, \dots, W_p) \\ = \int_0^1 \mathcal{A}(\Psi(s; W_0, W_1)) \frac{\partial \Psi}{\partial s}(s; W_0, W_1)\, ds, \end{aligned} \tag{23}$$

for every $W_i \in \Omega$, $i = -q, \dots, p$.

This definition generalizes the usual concept of a conservative numerical scheme for a system of conservation laws:

PROPOSITION 3.2. *Let us suppose that* (3) *is a system of conservation laws, i.e., $\mathcal{A}$ is the Jacobian of a flux function F. Then, every numerical scheme which is Ψ-conservative for some family of paths Ψ is consistent and conservative in the usual sense. Conversely, a consistent conservative numerical scheme is Ψ-conservative for every family of paths Ψ.*

Proof. Observe first that, in the case of a conservative system, (23) reduces to

$$D^-(W_{-q}, \dots, W_p) + D^+(W_{-q}, \dots, W_p) = F(W_1) - F(W_0).$$

Therefore, given a Ψ-conservative numerical scheme (21) we can define a numerical flux function G as follows:

$$\begin{aligned} G(W_{-q}, \dots, W_p) &= D^-(W_{-q}, \dots, W_p) + F(W_0) \\ &= -D^+(W_{-q}, \dots, W_p) + F(W_1). \end{aligned} \tag{24}$$

Then, (21) is equivalent to the conservative scheme (9) corresponding to the numerical flux G. Moreover, from (22) we easily deduce

$$G(W,\dots,W) = F(W).$$

Conversely, given a consistent conservative numerical scheme with numerical flux function G, it can be written under the form (21) by defining

$$\begin{aligned} D^-(W_{-q},\dots,W_p) &= G(W_{-q},\dots,W_p) - F(W_0),\\ D^+(W_{-q},\dots,W_p) &= -G(W_{-q},\dots,W_p) + F(W_1). \end{aligned}$$

It can be easily verified that (22) and (23) are satisfied for every family of paths Ψ. ∎

Remark 2. According to Proposition 3.2, a path-conservative numerical scheme applied to a conservative problem is just a conservative scheme formulated in the so-called *wave propagation form* (see [28]). It is important to notice that, in despite of its form, a path-conservative numerical scheme (21) is not a *nonconservative numerical scheme* in the usual sense: a numerical scheme for solving a *conservative problem* is said to be nonconservative if it cannot be written under the form (9).

Notice that condition (23) plays a double role. On the one hand, it is used to approximate the punctual masses associated to discontinuities. On the other hand, together with (22), it is a consistency requirement for regular solutions and smooth data. In effect, if W is a regular enough solution and $\mathcal{A}(W)$, $D^{\pm}(W_{-q},\dots,W_p)$ are also regular, from (22) and (23) it can be deduced that

$$\begin{aligned} &\frac{1}{\Delta x}\big(D^+(W(x_{i-q-1},t),\dots,W(x_{i+p-1},t)) + D^-(W(x_{i-q},t),\dots,W(x_{i+p},t))\big)\\ &\qquad = \mathcal{A}(W(x_i,t))W_x(x_i,t) + O(\Delta x). \end{aligned}$$

Path-conservative numerical schemes satisfy a certain *conservation* property. In effect, let W be a weak solution of (3) corresponding to an initial condition W_0 such that

$$W_0(x) = W_L, \quad \forall\, x < -A; \quad W_0(x) = W_R, \quad \forall\, x > A; \tag{25}$$

for some $A > 0$. Given $0 \le t_0 < t_1 < \infty$, W satisfies

$$\int_{\mathbb{R}} \big(W(x,t_1) - W(x,t_0)\big)\,dx = -\int_{t_0}^{t_1} \langle [\mathcal{A}(W(\cdot,t))W_x(\cdot,t)]_\Phi, 1\rangle\,dt. \tag{26}$$

Let us suppose now that a Ψ-conservative scheme is applied to approach this solution and let W^n be the piecewise constant function whose value at the cell I_i is W_i^n. Summing up in (21) and taking into account (17) and (23), we deduce the equality

$$\int_{\mathbb{R}} \big(W^{n+1}(x) - W^n(x)\big)\,dx = -\Delta t\,\langle [\mathcal{A}(W^n)W_x^n]_\Psi, 1\rangle, \tag{27}$$

which is clearly an approximation of (26).

As it was remarked in [29] in the context of Roe schemes, the best choice of the family of paths Ψ appearing in Definition 3.1 is the family Φ selected for the definition of weak solutions: in this case, (26) and (27) makes reference to the same

Borel measure and the jump conditions of weak solutions and numerical solutions are consistent.

In fact, a Lax–Wendroff theorem can be conjectured: if the numerical solutions obtained with a Ψ-conservative converge in an adequate sense, its limit has to be a weak solution whose definition is also related to the family of paths Ψ.

We stress that such a theorem would not be in contradiction with the negative results shown in [22] or [15]; in these works the failure of the convergence of nonconservative schemes to weak solutions of *conservative* problems was studied. But in our case, if the system is conservative, a path-conservative numerical scheme is not a nonconservative scheme (see Remark 2). Nevertheless, this kind of negative results are also expectable if a path-conservative numerical scheme based on a family Ψ is used to approach weak solutions based on a different family of paths Φ: in that case, the consistency for smooth solutions is still provided by (22) and (23) but discontinuities can be incorrectly treated. In fact, a negative result of this type was observed in [29] in the context of the approximation of shallow water systems with source term.

Unfortunately, the construction of Φ-conservative schemes can be difficult or very costly in practice. In this case, a simpler family of paths Ψ has to be chosen, as the family of segments:

$$\Psi(s; W_L, W_R) = W_L + s(W_R - W_L). \tag{28}$$

4. Well-balancing. Well-balancing is related to the numerical approximation of equilibria, i.e., steady state solutions. Notice that system (3) can only have nontrivial steady state solutions if it has some linearly degenerate fields; let $W(x)$ be a regular steady state solution

$$\mathcal{A}(W(x)) \cdot W'(x) = 0 \quad \forall\, x \in \mathbb{R}.$$

If $W'(x) \neq 0$, then 0 is an eigenvalue of $\mathcal{A}(W(x))$ and $W'(x)$ is an associated eigenvector. Therefore, $x \mapsto W(x)$ can be interpreted as a parameterization of an integral curve of a linearly degenerate characteristic field whose corresponding eigenvalue takes the value 0 through the curve. In order to define the concept of well-balancing, let us introduce the set Γ of all the integral curves γ of a linearly degenerate field of $\mathcal{A}(W)$ such that the corresponding eigenvalue vanishes on Γ. According to [29] we introduce the following definitions.

DEFINITION 4.1. *Given a curve $\gamma \in \Gamma$, a numerical scheme for solving* (3)

$$W_j^{n+1} = W_j^n + \frac{\Delta t}{\Delta x} H(W_{j-q}^n, \dots, W_{j+p}^n) \tag{29}$$

is said to be exactly well-balanced for γ if, given any $\mathcal{C}^1$ function $x \in (\alpha, \beta) \subset \mathbb{R} \mapsto W(x) \in \Omega$ such that

$$W(x) \in \gamma, \quad \forall\, x \in (\alpha, \beta), \tag{30}$$

and $p + q + 1$ points in (α, β) $x_{-q}, \dots, x_p$ such that

$$x_{-q} < \cdots < x_p; \quad x_{i+1} - x_i = \Delta x,\ i = -q, \dots, p-1, \tag{31}$$

then

$$H(W(x_{-q}), \dots, W(x_p)) = 0. \tag{32}$$

The scheme is said to be well-balanced with order k for γ if, given any $\mathcal{C}^{k+1}$ function W and any set of points $\{x_{-q}, \dots, x_p\}$ satisfying (30), (31), *then*

$$|H(W(x_{-q}), \dots, W(x_p))| = O\left(\Delta x^{k+1}\right). \tag{33}$$

Finally, the scheme is said to be exactly well-balanced or well-balanced with order k if these properties are satisfied for any curve of Γ.

We have only considered 1-level schemes and uniform meshes in order to avoid an excess of notation, but the definition can be easily extended to more general schemes.

The well-balance property of a scheme is strongly related to its ability to approximate stationary contact discontinuities. We can state for instance the following proposition.

PROPOSITION 4.2. *Given a numerical scheme of the form* (21) *with $q = 0$ and $p = 1$ and a curve γ of Γ, the numerical scheme is exactly well-balanced for γ if and only if it solves exactly every stationary contact discontinuity linking two states belonging to γ.*

Proof. Both properties are satisfied if and only if

$$D^{\pm}(W_0, W_1) = 0, \quad \forall\, W_0, W_1 \in \gamma. \qquad \square$$

Remark 3. For numerical schemes with arbitrary values of p and q the direct implication of the proposition is also valid. To see this, observe first that a numerical scheme is exactly well-balanced for γ if and only if

$$H(W_{-q}, \dots, W_p) = 0,$$

for any given *ordered* set of states $\{W_{-q}, \dots, W_p\}$ of γ, where some of the states can be repeated. Then it can be easily shown that this property implies that the numerical scheme solves exactly stationary contact discontinuities linking two states belonging to γ.

5. Approximate Riemann solvers. This section is devoted to generalize the notion of approximate Riemann solvers introduced in [21] for conservative systems (8) and extended in [5] for balance laws. The organization of this section closely follows Bouchut's book.

DEFINITION 5.1. *Given a family of paths Ψ, a Ψ-approximate Riemann solver for* (3) *is a function $\widetilde{V} : \mathbb{R} \times \Omega \times \Omega \mapsto \Omega$ satisfying the following:*

(i) *for every $W \in \Omega$,*

$$\widetilde{V}(v; W, W) = W \quad \forall\, v \in \mathbb{R}; \tag{34}$$

(ii) *for every W_L, $W_R \in \Omega$ there exist $\lambda_{min}(W_L, W_R)$, $\lambda_{max}(W_L, W_R)$ in $\mathbb{R}$ such that,*

$$\widetilde{V}(v; W_L, W_R) = W_L, \quad \text{if } v < \lambda_{min}(W_L, W_R),$$
$$\widetilde{V}(v; W_L, W_R) = W_R, \quad \text{if } v > \lambda_{max}(W_L, W_R);$$

(iii) *for every $W_L, W_R \in \Omega$,*

$$\begin{aligned} &\int_0^1 \mathcal{A}\left(\Psi(s; W_L, W_R)\right) \frac{\partial \Psi}{\partial s}(s; W_L, W_R)\, ds \\ &\quad + \int_0^{\infty} \left(\widetilde{V}(v; W_L, W_R) - W_R\right) dv \\ &\quad + \int_{-\infty}^0 \left(\widetilde{V}(v; W_L, W_R) - W_L\right) dv = 0. \end{aligned} \tag{35}$$

Notice that (35) is a generalization of the property (6) satisfied by the exact solution of a Riemann problem (5).

Given a Ψ-approximate Riemann solver for (3) a numerical scheme can be constructed as follows:

$$W_i^{n+1} = \frac{1}{\Delta x}\left(\int_{x_{i-1/2}}^{x_i} \widetilde{V}\left(\frac{x - x_{i-1/2}}{\Delta t}; W_{i-1}^n, W_i^n\right) dx + \int_{x_i}^{x_{i+1/2}} \widetilde{V}\left(\frac{x - x_{i+1/2}}{\Delta t}; W_i^n, W_{i+1}^n\right) dx\right). \tag{36}$$

Under a CFL condition 1/2, the numerical scheme can also be written under the form (21) with

$$D_{i+1/2}^- = -\int_{-\infty}^0 \left(\widetilde{V}(v; W_i^n, W_{i+1}^n) - W_i^n\right) dv, \tag{37}$$

$$D_{i+1/2}^+ = -\int_0^{\infty} \left(\widetilde{V}(v; W_i^n, W_{i+1}^n) - W_{i+1}^n\right) dv. \tag{38}$$

PROPOSITION 5.2. *A numerical scheme* (21) *based on a Ψ-approximate Riemann solver is Ψ-conservative.*

Proof. The proof is straightforward from (37), (38), and Definition 5.1. ∎

Remark 4. If the numerical scheme is intended to solve only weak solutions with *small* discontinuities, i.e., discontinuities linking pairs of states (W_L, W_R) belonging to $\mathcal{RP}$, then it is enough for the approximate Riemann solver $\widetilde{V}$ to be defined in $\mathbb{R} \times \mathcal{RP}$.

A numerical scheme (21) based on a Ψ-approximate Riemann solver is well-balanced for a curve γ of the set Γ, if and only if, given two states W_L and W_R in γ the following equalities hold:

$$\int_{-\infty}^0 \left(\widetilde{V}(v; W_L, W_R) - W_L\right) dv = 0,$$
$$\int_0^{\infty} \left(\widetilde{V}(v; W_L, W_R) - W_R\right) dv = 0.$$

These equalities are trivially satisfied if

$$\widetilde{V}(v; W_L, W_R) = \begin{cases} W_L & \text{if } v < 0, \\ W_R & \text{if } v > 0, \end{cases}$$

i.e., if the approximate Riemann solver is exact for pairs of states (W_L, W_R) belonging to γ.

We recall hereafter some classical choices of approximate Riemann solvers.

5.1. Godunov methods. Godunov methods correspond to the choice of the exact Riemann solver, i.e.,

$$\widetilde{V}(v; W_L, W_R) = V(v; W_L, W_R),$$

being $V(x/t; W_L, W_R)$ the exact solution of the Riemann problem (5). This is clearly a Φ-approximate Riemann solver. Moreover, if the concept of entropic solution is

related to an entropy pair (η, G) with convex η, according to Remark 1 it is *dissipative* for this pair (see [5]).

In [30] it has been shown that if the family of paths satisfies the hypotheses (H1)–(H3) stated in section 2, Godunov methods can be written under the form (21) with

$$D^-_{i+1/2} = \int_0^1 \mathcal{A}\left(\Phi(s; W^n_i, W^n_{i+1/2})\right) \frac{\partial \Phi}{\partial s}(s; W^n_i, W^n_{i+1/2})\, ds,$$
$$D^+_{i+1/2} = \int_0^1 \mathcal{A}\left(\Phi(s, W^n_{i+1/2}, W^n_{i+1})\right) \frac{\partial \Phi}{\partial s}(s; W^n_{i+1/2}, W^n_{i+1})\, ds,$$

where $W^n_{i+1/2}$ is the (constant) value at $x = x_{i+1/2}$ of the solution of the Riemann problem related to the states W^n_i and W^n_{i+1}. If the solution is discontinuous at $x = x_{i+1/2}$ the limit to the left or the right can be chosen indifferently.

Godunov methods are exactly well-balanced (see [30]).

5.2. Roe methods. Approximate Riemann solvers are often constructed as follows: $\widetilde{V}(x/t; W_L, W_R)$ is the solution of a linear Riemann problem

$$\text{(39)} \qquad \begin{cases} \dfrac{\partial U}{\partial t} + \mathcal{A}(W_L, W_R)\dfrac{\partial U}{\partial x} = 0, \\ \\ U(x,0) = \begin{cases} W_L & \text{if } x < 0, \\ W_R & \text{if } x > 0, \end{cases} \end{cases}$$

where $\mathcal{A}(W_L, W_R)$ is a linearization of $\mathcal{A}(W)$. It can be easily shown that this is a Ψ-approximate Riemann solver, if and only if, $\mathcal{A}(W_L, W_R)$ is a Roe linearization in the sense defined by Toumi in [40].

DEFINITION 5.3. *Given a family of paths Ψ, a function $\mathcal{A}_\Psi : \Omega \times \Omega \mapsto \mathcal{M}_{N\times N}(\mathbb{R})$ is called a* Roe linearization *if it verifies the following properties:*

1. *for each $W_L, W_R \in \Omega$, $\mathcal{A}_\Psi(W_L, W_R)$ has N distinct real eigenvalues,*
2. *$\mathcal{A}_\Psi(W, W) = \mathcal{A}(W)$, for every $W \in \Omega$,*
3. *for any $W_L, W_R \in \Omega$,*

$$\text{(40)} \qquad \mathcal{A}_\Psi(W_L, W_R)(W_R - W_L) = \int_0^1 \mathcal{A}(\Psi(s; W_L, W_R)) \frac{\partial \Psi}{\partial s}(s; W_L, W_R)\, ds.$$

Once a Roe linearization $\mathcal{A}_\Psi$ has been chosen, some straightforward calculations allow one to show that, under a CFL condition 1/2, the numerical scheme can be written under the form (21) with

$$D^-_{i+1/2} = \mathcal{A}^-_{i+1/2}(W^n_{i+1} - W^n_i),$$
$$D^+_{i+1/2} = \mathcal{A}^+_{i+1/2}(W^n_{i+1} - W^n_i),$$

where

$$\mathcal{A}_{i+1/2} = \mathcal{A}_\Psi(W^n_i, W^n_{i+1}),$$

and, as usual,

$$\text{(41)} \quad \mathcal{L}^\pm_{i+1/2} = \begin{bmatrix} (\lambda_1^{i+1/2})^\pm & & 0 \\ & \ddots & \\ 0 & & (\lambda_N^{i+1/2})^\pm \end{bmatrix}, \ \mathcal{A}^\pm_{i+1/2} = \mathcal{K}_{i+1/2}\mathcal{L}^\pm_{i+1/2}\mathcal{K}^{-1}_{i+1/2}$$

being $\mathcal{L}_{i+1/2}$ the diagonal matrix whose coefficients are the eigenvalues of $\mathcal{A}_{i+1/2}$

$$\lambda_1^{i+1/2} < \lambda_2^{i+1/2} < \cdots < \lambda_N^{i+1/2},$$

and $\mathcal{K}_{i+1/2}$ is a $N \times N$ matrix whose columns are associated eigenvectors.

As in the case of systems of conservation laws, a CFL condition 1 is used in practice, as this condition ensures the linear stability of the method. An entropy-fix technique also has to be added to the numerical scheme.

In [29] it has been shown that a Roe scheme based on a family of paths Ψ is exactly well-balanced for a curve $\gamma \in \Gamma$ if, given two states W_L and W_R in γ, the path $\Psi(s; W_L, W_R)$ is a parameterization of the arc of γ linking these states. In particular, if the family of path Ψ coincides with the family Φ used in the definition of weak solutions, the numerical scheme is exactly well-balanced. The numerical scheme is well-balanced with order k if $\Psi(s; W_L, W_R)$ approximates with order $k+1$ a regular parameterization of the arc of γ linking the states. In particular, a Roe scheme based on the family of segments (28) is always well-balanced with order 2. Moreover, it is exactly well-balanced for curves of Γ that are straight lines (see [29] for details).

The construction of Roe methods for systems of the form (1) has been studied in [29].

5.3. Relaxation methods. The goal of this paragraph is to give some guidelines about the construction of approximate Riemann solvers for nonconservative systems based on the relaxation technique. This has been done for balance laws in [5].

The idea is as follows. First of all, a new nonconservative hyperbolic system is considered,

$$\frac{\partial \widetilde{W}}{\partial t} + \mathcal{B}(\widetilde{W}) \frac{\partial \widetilde{W}}{\partial x} = 0, \quad x \in \mathbb{R},\ t > 0, \tag{42}$$

where $\widetilde{W}$ now takes values in an open convex $\widetilde{\Omega}$ of $\mathbb{R}^{\widetilde{N}}$, with $\widetilde{N} > N$. Again, $\mathcal{B}$ is a smooth locally bounded map from $\widetilde{\Omega}$ to $\mathcal{M}_{\widetilde{N}\times\widetilde{N}}(\mathbb{R})$.

Let us suppose that there exist two linear operators $\mathcal{L} : \widetilde{\Omega} \mapsto \Omega$ and $\mathcal{M} : \Omega \mapsto \widetilde{\Omega}$ such that

$$\mathcal{L}\mathcal{M}(W) = W \quad \forall\ W \in \Omega.$$

In practice, system (42) has to be chosen in such a way that it is possible to easily construct an approximate Riemann solver with good properties (this is the case, for instance, if Riemann problems related to (42) are easy to solve). Then, an approximate Riemann solver for (3) is deduced.

The main difference with the conservative case comes from the fact that, in this case, together with system (42) a family of paths in $\widetilde{\Omega}$ also has to be chosen in order to define the approximate Riemann solver for this system.

The following lemma, whose demonstration is straightforward, gives a sufficient condition to obtain a Ψ-approximate Reimann solver for (3) from a $\widetilde{\Psi}$-approximate Reimann solver for (42).

LEMMA 5.4. *Let Ψ and $\widetilde{\Psi}$ be two families of paths in Ω and $\widetilde{\Omega}$, respectively, such that*

$$\begin{aligned} &\int_0^1 \mathcal{L}\mathcal{B}\big(\widetilde{\Psi}(s; \mathcal{M}(W_L), \mathcal{M}(W_R))\big) \frac{\partial \widetilde{\Psi}}{\partial s}(s; \mathcal{M}(W_L), \mathcal{M}(W_R))\, ds \\ &\qquad = \int_0^1 \mathcal{A}\big(\Psi(s; W_L, W_R)\big) \frac{\partial \Psi}{\partial s}(s; W_L, W_R)\, ds. \end{aligned} \tag{43}$$

Then, if $\mathcal{R}(v;\widetilde{W}_L,\widetilde{W}_R)$ *is a* $\widetilde{\Psi}$*-approximate Reimann solver for* (42)*, the function*

$$\widetilde{V}(v;W_L,W_R)=\mathcal{L}\mathcal{R}(v;\mathcal{M}(W_L),\mathcal{M}(W_R)),$$

gives a Ψ*-approximate Reimann solver for* (3).

Remark 5. It can also be easily shown that, if (η,G) is an entropy pair for (3) and $(\widetilde{\eta},\widetilde{G})$ is an entropy extension to (42) (see [5]), and both η and $\widetilde{\eta}$ are convex functions, then, if $\mathcal{R}$ is dissipative for $(\widetilde{\eta},\widetilde{G})$, $\widetilde{V}$ is dissipative for (η,G).

6. High order schemes based on reconstruction of states. The goal of this section is to obtain a high order scheme for (3) based on a first order path-conservative numerical scheme (21) with $q=0$ and $p=1$, that is,

$$D^{\pm}_{i+1/2}=D^{\pm}(W^n_i,W^n_{i+1}),$$

and a reconstruction operator of order s, i.e., an operator that associates to a given sequence $\{W_i\}$ two new sequences $\{W^-_{i+1/2}\}$, $\{W^+_{i+1/2}\}$ in such a way that, whenever

$$W_i=\frac{1}{\Delta x}\int_{I_i}W(x)\,dx,\quad\forall\ i\in\mathbb{Z},$$

for some smooth function W, then

$$W^{\pm}_{i+1/2}=W(x_{i+1/2})+O\left(\Delta x^s\right),\quad\forall\ i\in\mathbb{Z}.$$

In the case of a system of conservation laws (8), high order methods based on the reconstruction of states can be built using the following procedure: a first order conservative scheme with numerical flux function $G(U,V)$ and a reconstruction operator of order s are first chosen. Next, the method of lines is used: the system is discretized only in space, leaving the problem continuous in time. Let us denote by $\overline{W}_i(t)$ the cell average of solution W of (3) over the cell I_i at time t,

$$\overline{W}_i(t)=\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}}W(x,t)\,dx.$$

The following equation can be easily obtained from (8):

$$\overline{W}'_i(t)=\frac{1}{\Delta x}\left(F(W(x_{i-1/2},t))-F(W(x_{i+1/2},t))\right). \tag{44}$$

Now, (44) is approached as follows:

$$W'_i(t)=\frac{1}{\Delta x}\big(\widetilde{G}_{i-1/2}-\widetilde{G}_{i+1/2}\big), \tag{45}$$

with

$$\widetilde{G}_{i+1/2}=G(W^-_{i+1/2}(t),W^+_{i+1/2}(t)), \tag{46}$$

$W_i(t)$ being the approximation to $\overline{W}_i(t)$, and $\{W^{\pm}_{i+1/2}(t)\}$ the reconstructions associated to the sequence $\{W_i(t)\}$. It can be shown that (45)–(46) give a semidiscrete method of order s for (8).

Notice that (45) is a system of ordinary differential equations which is solved using a standard numerical method.

Let us introduce an interpretation of (45) in terms of measures, as it was done in section 2, in order to generalize it to nonconservative systems. First, notice that (44) can also be written under the form

$$\overline{W}_i'(t) = -\frac{1}{\Delta x}\left\langle [F(W(\cdot,t))_x], 1_{I_i}\right\rangle. \tag{47}$$

Next, let us choose at every cell I_i and at every time $t>0$ a regular function P_i^t such that

$$\lim_{x\to x_{i-1/2}^+} P_i^t(x) = W_{i-1/2}^+(t), \qquad \lim_{x\to x_{i+1/2}^-} P_i^t(x) = W_{i+1/2}^-(t). \tag{48}$$

If we consider now the approximation of $W(\cdot,t)$ given by the piecewise regular function $\mathcal{W}^t$ whose restriction to I_i is P_i^t, the discrete analogue of (47) would be

$$W_i' = -\frac{1}{\Delta x}\left\langle [F(\mathcal{W}^t)_x], 1_{I_i}\right\rangle, \tag{49}$$

but, again, (49) is not equivalent to (45). In this case, $[F(\mathcal{W}^t)_x]$ is the sum of a regular measure, whose Radon–Nykodim derivative at the cell I_i is $F(P_i^t)_x$, and the singular measure

$$\sum_i \left(F(W_{i+1/2}^+(t)) - F(W_{i+1/2}^-(t))\right)\delta_{x=x_{i+1/2}}.$$

If, again, the numerical flux of the first order scheme is used to split the Dirac measures placed at the intercells

$$\begin{aligned}&\left(F(W_{i+1/2}^+(t)) - F(W_{i+1/2}^-(t))\right)\delta_{x=x_{i+1/2}}\\ &\quad = \left(F(W_{i+1/2}^+(t)) - \widetilde{G}_{i+1/2}\right)\delta_{x=x_{i+1/2}} + \left(\widetilde{G}_{i+1/2} - F(W_{i+1/2}^-(t))\right)\delta_{x=x_{i+1/2}},\end{aligned}$$

and the first and second summands are assigned, respectively, to the cells I_{i+1} and I_i, we obtain from (49),

$$\begin{aligned}W_i' = -\frac{1}{\Delta x}\Bigg(&F(W_{i-1/2}^+(t)) - \widetilde{G}_{i-1/2} + \widetilde{G}_{i+1/2} - F(W_{i+1/2}^-(t))\\ &+ \int_{x_{i-1/2}}^{x_{i+1/2}} F(P_i^t(x))_x\,dx\Bigg),\end{aligned} \tag{50}$$

which is obviously equivalent to (45).

We go now to the general case (3). In this case, the equation for the cell averages is the following:

$$\overline{W}_i' = -\frac{1}{\Delta x}\langle [\mathcal{A}(W(\cdot,t))W(\cdot,t)_x]_\Phi, 1_{I_i}\rangle. \tag{51}$$

The natural extension of (50) is then

$$W_i' = -\frac{1}{\Delta x}\left(\widetilde{D}_{i-1/2}^+ + \widetilde{D}_{i+1/2}^- + \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}[P_i^t(x)]\frac{dP_i^t}{dx}(x)\,dx\right), \tag{52}$$

with

$$\widetilde{D}^{\pm}_{i+1/2} = D^{\pm}(W^{-}_{i+1/2}(t), W^{+}_{i+1/2}(t)). \tag{53}$$

In (52) the integral terms are approximations of the regular measure of the Lebesgue decomposition of $[\mathcal{A}(W(\cdot,t))W_x(\cdot,t)]_\Phi$ while the terms $\widetilde{D}^{\pm}_{i-1/2}$ are related to its singular part.

Notice that there is an important difference between the conservative and nonconservative case: while in the conservative case the numerical scheme is independent of the functions P^t_i chosen at the cells (only the property (48) is important), this is not the case for nonconservative systems. As a consequence, while the numerical scheme (45) has order s, in the case of the scheme (52) the order will depend on the choice of the functions P^t_i.

In practice, the definition of the reconstruction operator gives the natural choice of the functions P^t_i, as the usual procedure is the following: given a sequence $\{W_i\}$ of values at the cells, an approximation function is calculated at every cell I_i using the values W_j at a *stencil*,

$$P_i(x; W_{i-l}, \dots, W_{i+r}),$$

with l, r being two natural numbers. The reconstructions $W^{\pm}_{i+1/2}$ are then calculated by taking the limits of these functions at the intercells. These approximations functions are usually calculated by means of interpolation or approximation techniques. The natural choice of P^t_i is thus

$$P^t_i(x) = P_i(x; W_{i-l}(t), \dots, W_{i+r}(t)).$$

Let us now investigate the order of the numerical scheme (52). Notice first that, for regular solutions W, the differential equation (51) can be written as follows:

$$\overline{W}'_i(t) = -\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}(W(x,t))W_x(x,t)\,dx. \tag{54}$$

THEOREM 6.1. *Let us suppose that $\mathcal{A}$, $D^{\pm}$ are regular and with bounded derivatives. Let us suppose also that the reconstruction operator is of order s and that, given the sequence defined by*

$$W_i = \frac{1}{\Delta x}\int_{I_i} W(x)\,dx,$$

for any smooth function W, the following approximation properties are satisfied:

$$P_i(x; W_{i-l}, \dots, W_{i+r}) = W(x) + O\left(\Delta x^{s_1}\right) \quad \forall\, x \in I_i,$$
$$\frac{d}{dx}P_i(x; W_{i-l}, \dots, W_{i+r}) = W'(x) + O\left(\Delta x^{s_2}\right) \quad \forall\, x \in I_i.$$

Then (52) *is an approximation of order at least $\bar{s} = \min(s, s_1+1, s_2+1)$ to the system* (54) *in the following sense:*

$$\begin{aligned}\widetilde{D}^{+}_{i-1/2} - \widetilde{D}^{-}_{i+1/2} + \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}(P^t_i(x))\frac{dP^t_i}{dx}(x)\,dx \\ = \int_{x_{i-1/2}}^{x_{i+1/2}} \mathcal{A}(W(x,t))W_x(x,t)\,dx + O\left(\Delta x^{\bar{s}}\right),\end{aligned} \tag{55}$$

for every smooth enough solution W, being $\{W^{\pm}_{i+1/2}(t)\}$ the reconstructions corresponding to the sequence $\{\overline{W}_i(t)\}$ and P^t_i the functions defined by

$$P^t_i(x) = P_i(x; \overline{W}_{i-l}(t), \dots, \overline{W}_{i+r}(t)).$$

The proof is identical to that of the particular case studied in [8], where general high order numerical schemes based on first order Roe methods were introduced.

Remark 6. For the usual reconstruction techniques one has $s_2 \leq s_1 < s$ and the order of (52) is thus $s_2 + 1$ for nonconservative systems and s for systems of conservation laws. Therefore a loss of accuracy can be observed when a technique of reconstruction is applied to a nonconservative problem. This effect has been detected and verified numerically for WENO-Roe methods in [8].

Notice that (52) can also be written under a form similar to (21),

$$W_i' = -\frac{1}{\Delta x}\left(E^+_{i-1/2} + E^-_{i+1/2}\right), \tag{56}$$

with

$$\begin{aligned} E^+_{i+1/2} &= \widetilde{D}^+_{i+1/2} + \int_{x_{i+1/2}}^{x_{i+1}} \mathcal{A}(P^t_{i+1}(x))\frac{dP^t_{i+1}}{dx}(x)\,dx, \\ E^-_{i+1/2} &= \widetilde{D}^-_{i+1/2} + \int_{x_i}^{x_{i+1/2}} \mathcal{A}(P^t_i(x))\frac{dP^t_i}{dx}(x)\,dx. \end{aligned} \tag{57}$$

Using this notation, the following equality holds:

$$\begin{aligned} E^+_{i+1/2} + E^-_{i+1/2} = &\int_{x_i}^{x_{i+1/2}} \mathcal{A}(P^t_i(x))\frac{dP^t_i}{dx}(x)\,dx \\ &+ \int_0^1 \mathcal{A}(\Psi(s; W^-_{i+1/2}, W^+_{i+1/2}))\frac{\partial \Psi}{\partial s}(s; W^-_{i+1/2}, W^+_{i+1/2})\,ds \\ &+ \int_{x_{i+1/2}}^{x_{i+1}} \mathcal{A}(P^t_{i+1}(x))\frac{dP^t_{i+1}}{dx}(x)\,dx, \end{aligned} \tag{58}$$

with Ψ being the family of paths for which the first order numerical scheme is path-conservative.

This latter equality can be understood as a path-conservation property similar to (23), where now the path linking $W_i(t)$ and $W_{i+1}(t)$ is the composition of three paths:

$$x \in [x_i, x_{i+1/2}] \mapsto P^t_i(x), \tag{59}$$

linking $W_i(t)$ and $W^-_{i+1/2}(t)$;

$$s \in [0,1] \mapsto \Psi(s; W^-_{i+1/2}(t), W^+_{i+1/2}(t)), \tag{60}$$

linking $W^-_{i+1/2}(t)$ and $W^+_{i+1/2}(t)$; and finally,

$$x \in [x_{i+1/2}, x_{i+1}] \mapsto P^t_{i+1}(x), \tag{61}$$

linking $W^+_{i+1/2}(t)$ and $W_{i+1}(t)$. Nevertheless, this family of paths does not depend only on the states $W_i(t)$ and $W_{i+1}(t)$ (as was the case in Definition 3.1) but on the values at the stencil

$$W_{i-l}(t), \ldots, W_{i+r}(t).$$

The definition of a well-balanced scheme can be easily extended for semidiscrete methods (see [8]).

DEFINITION 6.2. *Let us consider a semidiscrete method for solving* (3)*:*

$$\begin{cases} W_i' = \frac{1}{\Delta x}\mathcal{H}(\mathbf{W}(t); i), \quad i \in \mathbb{Z}, \\ \mathbf{W}(0) = \mathbf{W}_0, \end{cases} \tag{62}$$

where $\mathbf{W}(t) = \{W_i(t)\}$ *represents the vector of approximations to the cell averages of the exact solution, and* $\mathbf{W}_0 = \{W_i^0\}$ *is the vector of initial data. Let* γ *be a curve of* Γ*. The numerical method* (62) *is said to be exactly well-balanced for* γ *if, given a regular stationary solution* W*, such that*

$$W(x) \in \gamma \quad \forall\, x \in \mathbb{R},$$

the vector $\mathbf{W} = \{W(x_i)\}$*, where* x_i *denotes the center of the cell* I_i*, is a critical point for the system of differential equations* (62)*, i.e.,*

$$\mathcal{H}(\mathbf{W}; i) = 0 \quad \forall\, i,$$

and it is said to be well-balanced with order k *if:*

$$\mathcal{H}(\mathbf{W}; i) = O\left(\Delta x^k\right) \quad \forall\, i.$$

Finally, the semidiscrete method (62) *is said to be exactly well-balanced or well-balanced with order* k *if these properties are satisfied for every curve* γ *of the set* Γ*.*

We give hereafter two results concerning the well-balanced property of this scheme generalizing those presented in [8] for the particular case of Roe-based reconstruction methods, but before then we introduce a new definition.

DEFINITION 6.3. *The reconstruction operator is said to be* exactly well-balanced *for a curve* $\gamma \in \Gamma$ *if, given a sequence* $\{W_i\}$ *in* γ*, the approximation functions satisfy*

$$P_i(x; W_{i-l}, \ldots, W_{i+r}) \in \gamma \quad \forall\, x \in [x_{i-1/2}, x_{i+1/2}], \tag{63}$$

for every i*.*

THEOREM 6.4. *Let* γ *belong to* Γ*. Let us suppose that both the first order scheme and the reconstruction operator are exactly well-balanced for* γ*. Then, the numerical scheme* (52) *is also exactly well-balanced for* γ*.*

THEOREM 6.5. *Under the hypothesis of Theorem* 6.1*, the scheme* (52) *is well-balanced with an order of at least* $\bar{s} = \min(s, s_1 + 1, s_2 + 1)$*.*

The proofs of these results are identical to the corresponding theorems stated in [8].

Aknowledgments. The author wishes to thank M. J. Castro, J. M. Gallardo, and M. L. Muñoz for their helpful comments; and to F. Bouchut for stimulating discussions.

REFERENCES

[1] F. Alouges and B. Merlet, *Approximate shock curves of nonconservative hyperbolic systems in one space dimension*, J. Hyperbolic Differ. Equ., 1 (2004), pp. 769–788.

[2] E. Audusse, F. Bouchut, M. O. Bristeau, R. Klein, and B. Perthame, *A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows*, SIAM J. Sci. Comp., 25 (2004), pp. 2050–2065.

[3] A. Bermúdez and M. E. Vázquez, *Upwind methods for hyperbolic conservation laws with source terms*, Comput. & Fluids, 23 (1994), pp. 1049–1071.

[4] S. Bianchini and A. Bressan, *Vanishing viscosity solutions of nonlinear hyperbolic systems*, Ann. of Math., 161 (2005), pp. 223–342.

[5] F. Bouchut, *Nonlinear Stability of Finite Volume Methods for Hyperbolic Conservation Laws and Well-Balanced Schemes for Sources*, Birkhäuser, Basel, Switzerland, 2004.

[6] F. Bouchut and F. James, *One-dimensional transport equations with discontinuous coefficients*, Nonlinear Anal., 32 (1998), pp. 891–933.

[7] M. J. Castro, J. Macías, and C. Parés, *A Q-scheme for a class of systems of coupled conservation laws with source term. Application to a two-layer* 1*-D shallow water system*, M2AN Math. Mod. Numer. Anal., 35 (2001), pp. 107–127.

[8] M. J. Castro, J. M Gallardo, and C. Parés, *Finite volume schemes based on WENO reconstruction of states for solving nonconservative hyperbolic systems. Applications to shallow water systems*, Math. Comp., 2005 to appear.

[9] J. J. Cauret, J. F. Colombeau, and A. Y. LeRoux, *Discontinuous generalized solutions of nonlinear nonconservative hyperbolic equations*, J. Math. Anal. Appl., 139 (1989), pp. 552–573.

[10] T. Chacón, A. Domínguez, and E. D. Fernández, *A family of stable numerical solvers for shallow water equations with source terms*, Comput. Methods Appl. Mech. Eng., 192 (2003), pp. 203–225.

[11] T. Chacón, A. Domínguez, and E. D. Fernández, *Asymptotically balanced schemes for nonhomogeneous hyperbolic systems–application to the shallow water equations*, C.R. Math. Acad. Sci. Paris, 338 (2004), pp. 85–90.

[12] T. Chacón, E. D. Fernández, M. J. Castro, and C. Parés, *On well-balanced finite volume methods for nonhomogeneous nonconservative hyperbolic systems,* preprint, 2005.

[13] J. F. Colombeau and A. Heibig, *Nonconservative products in bounded variation functions*, SIAM J. Math. Anal., 23 (1992), pp. 941–949.

[14] G. Dal Maso, P. G. LeFloch, and F. Murat, *Definition and weak stability of nonconservative products*, J. Math. Pures Appl., 74 (1995), pp. 483–548.

[15] F. De Vuyst, *Schémas Nonconservatifs et Schémas Cinétiques Oour la Simulation Numérique D'écoulements Hypersoniques Non Visqueux en Déséquilibre Thermochimique*, Thése de Doctorat de l'Université Paris VI, Paris, France, 1994.

[16] S. K. Godunov, *A finite difference method for the computation of discontinuous solutions of the equations of fluid dynamics*, Mat. Sb., 47 (1959), pp. 357–393.

[17] L. Gosse, *A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms*, Comput. Math. Appl., 39 (2000), pp. 135–159.

[18] L. Gosse, *A well-balanced scheme using nonconservative products designed for hyperbolic systems of conservation laws with source terms*, Math. Models Methods Appl. Sci., 11 (2001), pp. 339–365.

[19] L. Gosse, *Localization effects and measure source terms in numerical schemes for balance laws*, Math. Comp., 71 (2002), pp. 553–582.

[20] G. Guerra, *Well-posedness for a scalar conservation law with singular nonconservative source*, J. Differential Equations, 206 (2004), pp. 438–469.

[21] A. Harten, P. D. Lax, and B. Van Leer, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev., 25 (1983), pp. 35–61.

[22] T. Hou and P. G. LeFloch, *Why nonconservative schemes converge to wrong solutions: Error analysis*, Math. Comp., 62 (1994), pp. 497–530.

[23] J. M. Greenberg and A. Y. LeRoux, *A well-balanced scheme for the numerical processing of source terms in hyperbolic equations*, SIAM J. Numer. Anal., 33 (1996), pp. 1–16.

[24] J. M. Greenberg, A. Y. LeRoux, R. Baraille, and A. Noussair, *Analysis and approximation of conservation laws with source terms*, SIAM J. Numer. Anal., 34 (1997), pp. 1980–2007.

[25] P. G. LeFloch, *Propagating phase boundaries. Formulation of the problem and existence via the Glimm method*, Arch. Rational Mech. Anal., 123 (1993), pp. 153–197.
[26] P. G. LeFloch and A. E. Tzavaras, *Representation of weak limits and definition of nonconservative products*, SIAM J. Math. Anal., 30 (1999), pp. 1309–1342.
[27] R. LeVeque, *Balancing source terms and flux gradients in high-resolution Godunov methods: The quasi-steady wave-propagation algorithm*, J. Comput. Phys., 146 (1998), pp. 346–365.
[28] R. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, Cambridge, UK, 2002.
[29] C. Parés and M. J. Castro, *On the well-balanced property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems*, M2AN Math. Model. Numer. Anal., 38 (2004), pp. 821–852.
[30] C. Parés, J. M. Gallardo, M. L. Muñoz, and M. J. Castro, *Godunov's method for nonconservative hyperbolic systems. Application to linear balance laws*, preprint, 2005.
[31] B. Perthame and C. Simeoni, *A kinetic scheme for the Saint–Venant system with a source term*, Calcolo, 38 (2001), pp. 201–231.
[32] B. Perthame and C. Simeoni, *Convergence of the upwind interface source method for hyperbolic conservation laws*, in Hyperbolic Problems: Theory, Numerics, Applications, Thou and Tadmor, ed., Springer, Berlin, 2003.
[33] P. A. Raviart and L. Sainsaulieu, *A nonconservative hyperbolic system modeling spray dynamics.* I. *Solution of the Riemann problem*, Math. Models Methods Appl. Sci., 5 (1995), pp. 297–333.
[34] J. P. Raymond, *A new definition of nonconservative products and weak stability results*, Boll. Un. Mat. Ital. B, 10 (1996), pp. 681–699.
[35] P. L. Roe, *Approximate Riemann solvers, parameter vectors, and difference schemes*, J. Comput. Phys., 43 (1981), pp. 357–372.
[36] P. L. Roe, *Upwinding difference schemes for hyperbolic conservation laws with source terms*, in Proceedings of the Conference on Hyperbolic Problems, Carasso, Raviart, and Serre, eds., Springer, 1986, pp. 41–51.
[37] H. Tang, T. Tang, and K. Xu, *A gas-kinetic scheme for shallow-water equations with source terms*, Z. Angew. Math. Phys., 55 (2004), pp. 365–382.
[38] T. Tang and Z. H. Teng, *Error bounds for fractional step methods for conservation laws with source terms*, SIAM J. Numer. Anal., 32 (1995), pp. 110–127.
[39] E. F. Toro, *Shock-Capturing Methods for Free-Surface Shallow Flows*, Wiley, Chichester, UK, 2001.
[40] I. Toumi, *A weak formulation of Roe's approximate Riemann solver*, J. Comput. Phys., 102 (1992), pp. 360–373.
[41] A. I. Volpert, *Spaces BV and quasilinear equations*, Math. USSR Sbornik, 73 (1967), pp. 255–302.